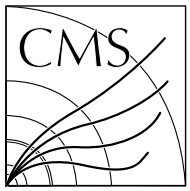


CERN-LHCC-2015-10
LHCC-P-008
CMS-TDR-15-02
ISBN 978-92-9083-417-5
1 June 2015

TECHNICAL PROPOSAL FOR THE PHASE-II UPGRADE OF THE COMPACT MUON SOLENOID

This Technical Proposal presents the upgrades foreseen to prepare the CMS experiment for the High Luminosity LHC. In this second phase of the LHC physics program, the accelerator will provide to CMS an additional integrated luminosity of about 2500 fb^{-1} over 10 years of operation, starting in 2025. This will substantially enlarge the mass reach in the search for new particles and will also greatly extend the potential to study the properties of the Higgs boson discovered at the LHC in 2012. In order to meet the experimental challenges of unprecedented p-p luminosity, the CMS collaboration will need to address the aging of the present detector and to improve the ability of the apparatus to isolate and precisely measure the products of the most interesting collisions. This document describes the conceptual designs and the expected performance of the upgrades, along with the plans to develop the appropriate experimental techniques. The infrastructure upgrades and the logistics of the installation in the experimental area are also discussed. Finally, the initial cost estimates of the upgrades are presented.

CERN-LHCC-2015-010 / LHCC-P-008
18/09/2015

Editors

J. Butler, D. Contardo, M. Klute, J. Mans, L. Silvestris

Chapter Editors

D. Abbaneo, A. Ball, P. Bloch, O. Buchmueller, F. Cavallari, A. Dabrowski, P. de Barbaro, I. Fisk, M. Girone, F. Hartmann, J. Hauser, C. Jessop, D. Lange, F. Meijers, C. Seez, W. Smith

Cover Design

S. Cittolin

Acknowledgments

We would like to thank the technical staff from the various institutions for the design, R&D and testing of the many components of this upgrade.

Yerevan Physics Institute, Yerevan, Armenia
V. Khachatryan, A.M. Sirunyan, A. Tumasyan

Institut für Hochenergiephysik der OeAW, Wien, Austria
W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, M. Flechl, M. Friedl, R. Frühwirth[1], V.M. Ghete, C. Hartl, N. Hörmann, J. Hrubec, M. Jeitler[1], V. Knünz, A. König, M. Krammer[1], I. Krätschmer, D. Liko, T. Matsushita, I. Mikulec, D. Rabady[2], B. Rahbaran, H. Rohringer, J. Schieck[1], R. Schöfbeck, J. Strauss, W. Treberer-Treberspurg, W. Waltenberger, C.-E. Wulz[1]

National Centre for Particle and High Energy Physics, Minsk, Belarus
V. Mossolov, N. Shumeiko, J. Suarez Gonzalez

Universiteit Antwerpen, Antwerpen, Belgium
S. Alderweireldt, T. Cornelis, E.A. De Wolf, X. Janssen, A. Knutsson, J. Lauwers, S. Luyckx, S. Ochesanu, R. Rougny, M. Van De Klundert, H. Van Haevermaet, P. Van Mechelen, N. Van Remortel, A. Van Spilbeeck

Vrije Universiteit Brussel, Brussel, Belgium
S. Abu Zeid, F. Blekman, J. D'Hondt, N. Daci, I. De Bruyn, K. Deroover, N. Heracleous, J. Keaveney, S. Lowette, L. Moreels, A. Olbrechts, Q. Python, D. Strom, S. Tavernier, W. Van Doninck, P. Van Mulders, G.P. Van Onsem, I. Van Parijs

Université Libre de Bruxelles, Bruxelles, Belgium
P. Barria, C. Caillol, B. Clerbaux, G. De Lentdecker, H. Delannoy, G. Fasanella, L. Favart, A.P.R. Gay, A. Grebenyuk, G. Karapostoli, T. Lenzi, A. Léonard, T. Maerschalk, A. Marinov, L. Perniè, A. Randle-conde, T. Reis, S. Rugovac, T. Seva, C. Vander Velde, P. Vanlaer, Y. Yang, R. Yonamine, F. Zenoni, F. Zhang[3]

Ghent University, Ghent, Belgium
K. Beernaert, L. Benucci, A. Cimmino, S. Crucy, D. Dobur, A. Fagot, G. Garcia, M. Gul, J. Mccartin, A.A. Ocampo Rios, D. Poyraz, D. Ryckbosch, S. Salva, M. Sigamani, N. Strobbe, M. Tytgat, W. Van Driessche, E. Yazgan, N. Zaganidis

Université Catholique de Louvain, Louvain-la-Neuve, Belgium
S. Basegmez, C. Beluffi[4], O. Bondu, S. Brochet, G. Bruno, R. Castello, A. Caudron, L. Ceard, G.G. Da Silveira, C. Delaere, D. Favart, L. Forthomme, B. Francois, A. Giammanco[5], J. Hollar, A. Jafari, P. Jez, M. Komm, V. Lemaitre, A. Mertens, C. Nuttens, L. Perrini, A. Pin, K. Piotrzkowski, A. Popov[6], L. Quertenmont, M. Selvaggi, M. Vidal Marono

Université de Mons, Mons, Belgium
N. Beliy, G.H. Hammad

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil
W.L. Aldá Júnior, G.A. Alves, L. Brito, M. Correa Martins Junior, M. Hamer, C. Hensel, C. Mora Herrera, A. Moraes, M.E. Pol, P. Rebello Teles

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
E. Belchior Batista Das Chagas, W. Carvalho, J. Chinellato[7], A. Custódio, E.M. Da Costa, D. De Jesus Damiao, C. De Oliveira Martins, S. Fonseca De Souza, L.M. Huertas Guativa, H. Malbouisson, D. Matos Figueiredo, L. Mundim, H. Nogima, W.L. Prado Da Silva, A. Santoro, A. Sznajder, E.J. Tonelli Manganote[7], A. Vilela Pereira

Universidade Estadual Paulista [a], Universidade Federal do ABC [b], São Paulo, Brazil
S. Ahuja[a], C.A. Bernardes[b], A. De Souza Santos[b], S. Dogra[a], T.R. Fernandez Perez Tomei[a], E.M. Gregores[b], P.G. Mercadante[b], C.S. Moon[a,8], S.F. Novaes[a], Sandra S. Padula[a], D. Romero Abad, J.C. Ruiz Vargas

Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria
A. Aleksandrov, R. Hadjiiska, P. Iaydjiev, M. Rodozov, S. Stoykova, G. Sultanov, M. Vutova

University of Sofia, Sofia, Bulgaria
A. Dimitrov, I. Glushkov, L. Litov, B. Pavlov, P. Petkov

Institute of High Energy Physics, Beijing, China
M. Ahmad, J.G. Bian, G.M. Chen, H.S. Chen, M. Chen, T. Cheng, R. Du, C.H. Jiang, R. Plestina[9], F. Romeo, S.M. Shaheen, J. Tao, C. Wang, Z. Wang, H. Zhang

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China
C. Asawatangtrakuldee, Y. Ban, Q. Li, S. Liu, Y. Mao, S.J. Qian, D. Wang, Z. Xu, W. Zou

Universidad de Los Andes, Bogota, Colombia
C. Avila, A. Cabrera, L.F. Chaparro Sierra, C. Florez, J.P. Gomez, B. Gomez Moreno, J.C. Sanabria

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia
N. Godinovic, D. Lelas, I. Puljak, P.M. Ribeiro Cipriano

University of Split, Faculty of Science, Split, Croatia
Z. Antunovic, M. Kovac

Institute Rudjer Boskovic, Zagreb, Croatia
V. Brigljevic, K. Kadija, J. Luetic, S. Micanovic, L. Sudic

University of Cyprus, Nicosia, Cyprus
A. Attikis, G. Mavromanolakis, J. Mousa, C. Nicolaou, F. Ptochos, P.A. Razis, H. Rykaczewski

Charles University, Prague, Czech Republic
M. Bodlak, M. Finger[10], M. Finger Jr.[10]

Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt
M. El Sawy[11,12], E. El-khateeb[13], S. Elgammal[12], T. Elkafrawy[13], A. Mohamed[14], E. Salama[13,12]

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
B. Calpas, M. Kadastik, M. Murumaa, M. Raidal, A. Tiko, C. Veelken

Department of Physics, University of Helsinki, Helsinki, Finland
P. Eerola, J. Pekkanen, M. Voutilainen

Helsinki Institute of Physics, Helsinki, Finland
J. Härkönen, V. Karimäki, R. Kinnunen, T. Lampén, K. Lassila-Perini, S. Lehti, T. Lindén, P. Luukka, T. Mäenpää, T. Peltola, E. Tuominen, J. Tuominiemi, E. Tuovinen, L. Wendland

Lappeenranta University of Technology, Lappeenranta, Finland
J. Talvitie, T. Tuuva

DSM/IRFU, CEA/Saclay, Gif-sur-Yvette, France
M. Besancon, F. Couderc, M. Dejardin, D. Denegri, B. Fabbro, J.L. Faure, C. Favaro, F. Ferri, S. Ganjour, A. Givernaud, P. Gras, G. Hamel de Monchenault, P. Jarry, E. Locci, M. Machet, J. Malcles, J. Rander, A. Rosowsky, M. Titov, A. Zghiche

Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France
I. Antropov, S. Baffioni, F. Beaudette, P. Busson, L. Cadamuro, E. Chapon, C. Charlot, T. Dahms, O. Davignon, N. Filipovic, A. Florent, R. Granier de Cassagnac, S. Lisniak, L. Mastrolorenzo, P. Miné, I.N. Naranjo, M. Nguyen, C. Ochando, G. Ortona, P. Paganini, P. Pigard, S. Regnard, R. Salerno, J.B. Sauvan, Y. Sirois, T. Strebler, Y. Yilmaz, A. Zabi

Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, Université de Haute Alsace Mulhouse, CNRS/IN2P3, Strasbourg, France
J.-L. Agram[15], J. Andrea, A. Aubin, R. Baumann, D. Bloch, C. Bonnin, J.-M. Brom, M. Buttignol, E.C. Chabert, N. Chanon, L. Charles, C. Collard, E. Conte[15], X. Coubez, J.-C. Fontaine[15], D. Gelé, U. Goerlach, C. Goetzmann, L. Gross, J. Hosselet, A.-C. Le Bihan, J.A. Merlin[2], M. Richer, K. Skovpen, P. Van Hove

Centre de Calcul de l'Institut National de Physique Nucleaire et de Physique des Particules, CNRS/IN2P3, Villeurbanne, France
S. Gadrat

Université de Lyon, Université Claude Bernard Lyon 1, CNRS-IN2P3, Institut de Physique Nucléaire de Lyon, Villeurbanne, France
S. Beauceron, C. Bernet, G. Boudoul, E. Bouvier, C.A. Carrillo Montoya, J. Chasserat, R. Chierici, D. Contardo, B. Courbon, P. Depasse, H. El Mamouni, J. Fan, J. Fay, S. Gascon, M. Gouzevitch, B. Ille, F. Lagarde, I.B. Laktineh, M. Lethuillier, L. Mirabito, A.L. Pequegnot, S. Perries, J.D. Ruiz Alvarez, D. Sabes, L. Sgandurra, V. Sordini, M. Vander Donckt, P. Verdier, S. Viret, H. Xiao

Georgian Technical University, Tbilisi, Georgia
T. Toriashvili[16]

Tbilisi State University, Tbilisi, Georgia
Z. Tsamalaidze[10]

RWTH Aachen University, I. Physikalisches Institut, Aachen, Germany
C. Autermann, S. Beranek, M. Edelhoff, L. Feld, A. Heister, M.K. Kiesel, K. Klein, M. Lipinski, A. Ostapchuk, M. Preuten, F. Raupach, S. Schael, J.F. Schulte, T. Verlage, H. Weber, B. Wittmer, V. Zhukov[6]

RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
M. Ata, M. Brodski, E. Dietz-Laursonn, D. Duchardt, M. Endres, M. Erdmann, S. Erdweg, T. Esch, R. Fischer, A. Güth, T. Hebbeker, C. Heidemann, K. Hoepfner, D. Klingebiel, S. Knutzen, P. Kreuzer, V. Kutzner, M. Merschmeyer, A. Meyer, P. Millet, M. Olschewski, K. Padeken, P. Papacz, T. Pook, M. Radziej, H. Reithler, M. Rieger, F. Scheuch, L. Sonnenschein, D. Teyssier, S. Thüer

RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany
V. Cherepanov, Y. Erdogan, G. Flügge, H. Geenen, M. Geisler, F. Hoehle, B. Kargoll, T. Kress, Y. Kuessel, A. Künsken, J. Lingemann[2], A. Nehrkorn, A. Nowack, I.M. Nugent, C. Pistone, O. Pooth, A. Stahl

Deutsches Elektronen-Synchrotron, Hamburg, Germany
M. Aldaya Martin, I. Asin, N. Bartosik, O. Behnke, U. Behrens, A.J. Bell, K. Borras, A. Burgmeier, A. Cakir, L. Calligaris, A. Campbell, S. Choudhury, F. Costanza, C. Diez Pardos, G. Dolinska, S. Dooling, T. Dorland, G. Eckerlin, D. Eckstein, T. Eichhorn, G. Flucke, E. Gallo[17], J. Garay Garcia, A. Geiser, A. Gizhko, P. Gunnellini, J. Hauk, M. Hempel[18], H. Jung, A. Kalogeropoulos, O. Karacheban[18], M. Kasemann, P. Katsas, J. Kieseler, C. Kleinwort, I. Korol, W. Lange, J. Leonard, K. Lipka, A. Lobanov, W. Lohmann[18], R. Mankel, I. Marfin[18], I.-A. Melzer-Pellmann, A.B. Meyer, G. Mittag, J. Mnich, A. Mussgiller, S. Naumann-Emme, A. Nayak, E. Ntomari, H. Perrey, D. Pitzl, R. Placakyte, A. Raspereza, B. Roland, M.Ö. Sahin, P. Saxena, T. Schoerner-Sadenius, M. Schröder, C. Seitz, S. Spannagel, K.D. Trippkewitz, C. Wissing

University of Hamburg, Hamburg, Germany
V. Blobel, M. Centis Vignali, A.R. Draeger, J. Erfle, E. Garutti, K. Goebel, D. Gonzalez, M. Görner, J. Haller, M. Hoffmann, R.S. Höing, A. Junkes, R. Klanner, R. Kogler, T. Lapsien, T. Lenz, I. Marchesini, D. Marconi, M. Meyer, D. Nowatschin, J. Ott, F. Pantaleo[2], T. Peiffer, A. Perieanu, N. Pietsch, J. Poehlsen, D. Rathjens, C. Sander, H. Schettler, P. Schleper, E. Schlieckau, A. Schmidt, J. Schwandt, M. Seidel, V. Sola, H. Stadie, G. Steinbrück, H. Tholen, D. Troendle, E. Usai, L. Vanelderen, A. Vanhoefer, B. Vormwald

Institut für Experimentelle Kernphysik, Karlsruhe, Germany
M. Akbiyik, C. Barth, C. Baus, J. Berger, C. Böser, E. Butz, T. Chwalek, F. Colombo, W. De Boer, A. Descroix, A. Dierlamm, S. Fink, F. Frensch, M. Giffels, A. Gilbert, F. Hartmann[2], S.M. Heindl, U. Husemann, I. Katkov[6], A. Kornmayer[2], P. Lobelle Pardo, B. Maier, H. Mildner, M.U. Mozer, T. Müller, Th. Müller, M. Plagge, G. Quast, K. Rabbertz, S. Röcker, F. Roscher, H.J. Simonis, F.M. Stober, R. Ulrich, J. Wagner-Kuhr, S. Wayand, M. Weber, T. Weiler, C. Wöhrmann, R. Wolf

Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
G. Anagnostou, G. Daskalakis, T. Geralis, V.A. Giakoumopoulou, A. Kyriakis, D. Loukas, A. Psallidas, I. Topsis-Giotis

University of Athens, Athens, Greece
A. Agapitos, S. Kesisoglou, A. Panagiotou, N. Saoulidou, E. Tziaferi

University of Ioánnina, Ioánnina, Greece
I. Evangelou, G. Flouris, C. Foudas, P. Kokkas, N. Loukas, N. Manthos, I. Papadopoulos, E. Paradas, J. Strologas

Wigner Research Centre for Physics, Budapest, Hungary
G. Bencze, C. Hajdu, A. Hazi, P. Hidas, D. Horvath[19], F. Sikler, V. Veszpremi, G. Vesztergombi[20], A.J. Zsigmond

Institute of Nuclear Research ATOMKI, Debrecen, Hungary
N. Beni, S. Czellar, J. Karancsi[21], J. Molnar, Z. Szillasi

University of Debrecen, Debrecen, Hungary
M. Bartók[22], A. Makovec, P. Raics, Z.L. Trocsanyi, B. Ujvari

National Institute of Science Education and Research, Bhubaneswar, India
P. Mal, K. Mandal, N. Sahoo, S.K. Swain

Panjab University, Chandigarh, India
S. Bansal, S.B. Beri, V. Bhatnagar, R. Chawla, R. Gupta, U.Bhawandeep, A.K. Kalsi, A. Kaur, M. Kaur, R. Kumar, A. Mehta, M. Mittal, J.B. Singh, G. Walia

University of Delhi, Delhi, India
Ashok Kumar, A. Bhardwaj, B.C. Choudhary, R.B. Garg, A. Kumar, S. Malhotra, M. Naimuddin, N. Nishu, K. Ranjan, R. Sharma, V. Sharma

Saha Institute of Nuclear Physics, Kolkata, India
S. Banerjee, S. Bhattacharya, K. Chatterjee, S. Dey, S. Dutta, Sa. Jain, N. Majumdar, A. Modak, K. Mondal, S. Mukherjee, S. Mukhopadhyay, A. Roy, D. Roy, S. Roy Chowdhury, S. Sarkar, M. Sharan

Bhabha Atomic Research Centre, Mumbai, India
A. Abdulsalam, R. Chudasama, D. Dutta, V. Jha, V. Kumar, A.K. Mohanty[2], L.M. Pant, P. Shukla, A. Topkar

Tata Institute of Fundamental Research, Mumbai, India
T. Aziz, S. Banerjee, S. Bhowmik[23], R.M. Chatterjee, R.K. Dewanjee, S. Dugad, S. Ganguly, S. Ghosh, M. Guchait, A. Gurtu[24], G. Kole, S. Kumar, B. Mahakud, M. Maity[23], G. Majumder, K. Mazumdar, S. Mitra, G.B. Mohanty, B. Parida, T. Sarkar[23], K. Sudhakar, N. Sur, B. Sutar, N. Wickramage[25]

Indian Institute of Science Education and Research (IISER), Pune, India
S. Chauhan, S. Dube, S. Sharma

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
H. Bakhshiansohi, H. Behnamian, S.M. Etesami[26], A. Fahim[27], R. Goldouzian, M. Khakzad, M. Mohammadi Najafabadi, M. Naseri, S. Paktinat Mehdiabadi, F. Rezaei Hosseinabadi, B. Safarzadeh[28], M. Zeinali

University College Dublin, Dublin, Ireland
M. Felcini, M. Grunewald

INFN Sezione di Bari [a], Università di Bari [b], Politecnico di Bari [c], Bari, Italy
M. Abbrescia[a,b], C. Calabria[a,b], C. Caputo[a,b], A. Colaleo[a], D. Creanza[a,c], L. Cristella[a,b], N. De Filippis[a,c], M. De Palma[a,b], G. De Robertis[a], G. Donvito[a], L. Fiore[a], G. Iaselli[a,c], F. Loddo[a], G. Maggi[a,c], M. Maggi[a], G. Miniello[a,b], S. My[a,c], S. Nuzzo[a,b], A. Pompili[a,b], G. Pugliese[a,c], R. Radogna[a,b], A. Ranieri[a], G. Selvaggi[a,b], L. Silvestris[a,2], V. Spinoso[a], R. Venditti[a,b], P. Verwilligen[a]

INFN Sezione di Bologna [a], Università di Bologna [b], Bologna, Italy
G. Abbiendi[a], C. Battilana[2], A.C. Benvenuti[a], D. Bonacorsi[a,b], S. Braibant-Giacomelli[a,b], L. Brigliadori[a,b], R. Campanini[a,b], P. Capiluppi[a,b], A. Castro[a,b], F.R. Cavallo[a], S.S. Chhibra[a,b], G. Codispoti[a,b], M. Cuffiani[a,b], G.M. Dallavalle[a], F. Fabbri[a], A. Fanfani[a,b], D. Fasanella[a,b], P. Giacomelli[a], C. Grandi[a], L. Guiducci[a,b], S. Marcellini[a], G. Masetti[a], A. Montanari[a], F.L. Navarria[a,b], A. Perrotta[a], A.M. Rossi[a,b], T. Rovelli[a,b], G.P. Siroli[a,b], N. Tosi[a,b], R. Travaglini[a,b]

INFN Sezione di Catania [a], Università di Catania [b], CSFNSM [c], Catania, Italy
G. Cappello[a], M. Chiorboli[a,b], S. Costa[a,b], F. Giordano[a,b], R. Potenza[a,b], A. Tricomi[a,b], C. Tuve[a,b]

INFN Sezione di Firenze [a], Università di Firenze [b], Firenze, Italy
G. Barbagli[a], M. Brianzi[a], V. Ciulli[a,b], C. Civinini[a], R. D'Alessandro[a,b], E. Focardi[a,b], S. Gonzi[a,b], V. Gori[a,b], P. Lenzi[a,b], M. Meschini[a], S. Paoletti[a], E. Scarlini[a,b], G. Sguazzoni[a], A. Tropiano[a,b], L. Viliani[a,b]

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, F. Fabbri, D. Piccolo, F. Primavera

INFN Sezione di Genova [a], Università di Genova [b], Genova, Italy
V. Calvelli[a,b], F. Ferro[a], M. Lo Vetere[a,b], M.R. Monge[a,b], E. Robutti[a], S. Tosi[a,b]

INFN Sezione di Milano-Bicocca [a], Università di Milano-Bicocca [b], Milano, Italy
L. Brianza, M.E. Dinardo[a,b], S. Fiorendi[a,b], S. Gennai[a], R. Gerosa[a,b], A. Ghezzi[a,b], P. Govoni[a,b], S. Malvezzi[a], R.A. Manzoni[a,b], B. Marzocchi[a,b,2], D. Menasce[a], L. Moroni[a], M. Paganoni[a,b], D. Pedrini[a], S. Ragazzi[a,b], N. Redaelli[a], T. Tabarelli de Fatis[a,b]

INFN Sezione di Napoli [a], Università di Napoli 'Federico II' [b], Napoli, Italy, Università della Basilicata [c], Potenza, Italy, Università G. Marconi [d], Roma, Italy
S. Buontempo[a], F. Cassese, N. Cavallo[a,c], S. Di Guida[a,d,2], M. Esposito[a,b], F. Fabozzi[a,c], A.O.M. Iorio[a,b], G. Lanza[a], L. Lista[a], S. Meola[a,d,2], M. Merola[a], P. Paolucci[a,2], G. Passeggio[a], C. Sciacca[a,b], F. Thyssen

INFN Sezione di Padova [a], Università di Padova [b], Padova, Italy, Università di Trento [c], Trento, Italy
P. Azzi[a,2], N. Bacchetta[a], M. Bellato[a], L. Benato[a,b], D. Bisello[a,b], A. Boletti[a,b], R. Carlin[a,b], P. Checchia[a], M. Dall'Osso[a,b,2], T. Dorigo[a], U. Dosselli[a], F. Gasparini[a,b], U. Gasparini[a,b], A. Gozzelino[a], M. Gulmini[a,29], S. Lacaprara[a], M. Margoni[a,b], A.T. Meneguzzo[a,b], F. Montecassiano[a], D. Pantano[a], J. Pazzini[a,b], N. Pozzobon[a,b], P. Ronchese[a,b], F. Simonetto[a,b], E. Torassa[a], M. Tosi[a,b], S. Vanini[a,b], S. Ventura[a], M. Zanetti, P. Zotto[a,b], A. Zucchetta[a,b,2]

INFN Sezione di Pavia [a], Università di Pavia [b], Pavia, Italy
A. Braghieri[a], A. Magnani[a], M. Manghisoni[a], P. Montagna[a,b], L. Ratti[a], S.P. Ratti[a,b], V. Re[a], C. Riccardi[a,b], P. Salvini[a], G. Traversi[a], I. Vai[a], P. Vitulo[a,b]

INFN Sezione di Perugia [a], Università di Perugia [b], Perugia, Italy
L. Alunni Solestizi[a,b], M. Biasini[a,b], G.M. Bilei[a], L. Bissi[a], B. Checcucci[a], D. Ciangottini[a,b,2], L. Fanò[a,b], P. Lariccia[a,b], D. Magalotti[a], G. Mantovani[a,b], S. Marconi, M. Menichelli[a], A. Morozzi, F. Moscatelli[a], D. Passeri[a,b], P. Placidi[a,b], A. Saha[a], M. Salvatore[a,b], A. Santocchia[a,b], L. Servoli[a], A. Spiezia[a,b], L. Storchi[a]

INFN Sezione di Pisa [a], Università di Pisa [b], Scuola Normale Superiore di Pisa [c], Pisa, Italy
K. Androsov[a,30], S. Arezzini[a], P. Azzurri[a], G. Bagliesi[a], A. Basti[a], R. Beccherle[a], J. Bernardini[a], T. Boccali[a], F. Bosi[a], G. Broccolo[a,c], A. Caratelli, R. Castaldi[a], M. Ceccanti[a], A. Ciampa[a], M.A. Ciocci[a,30], R. Dell'Orso[a], S. Donato[a,c,2], G. Fedi, L. Foà[a,c†], A. Giassi[a], M.T. Grippo[a,30], F. Ligabue[a,c], T. Lomtadze[a], G. Magazzu[a], P. Mammini[a], L. Martini[a,b], E. Mazzoni[a], A. Messineo[a,b], M. Minuti, A. Moggi[a], F. Morsani[a], F. Palla[a], A. Profeti[a], F. Raffaelli[a], A. Rizzi[a,b], A. Savoy-Navarro[a,31], A.T. Serban[a], P. Spagnolo[a], P. Squillacioti[a,30], R. Tenchini[a], G. Tonelli[a,b], A. Venturi[a], P.G. Verdini[a]

INFN Sezione di Roma [a], Università di Roma [b], Roma, Italy
L. Barone[a,b], F. Cavallari[a], G. D'imperio[a,b,2], I. Dafinei[a], D. Del Re[a,b], M. Diemoz[a], S. Gelli[a,b], C. Jorda[a], E. Longo[a,b], F. Margaroli[a,b], P. Meridiani[a], M. Nuccetelli[a], G. Organtini[a,b], R. Paramatti[a], F. Pellegrino[a], F. Preiato[a,b], S. Rahatlou[a,b], C. Rovelli[a], F. Santanastasio[a,b], I.G. Talamo[a], P. Traczyk[a,b,2]

INFN Sezione di Torino [a], Università di Torino [b], Torino, Italy,
Università del Piemonte Orientale [c], Novara, Italy
N. Amapane[a,b], R. Arcidiacono[a,c,2], S. Argiro[a,b], M. Arneodo[a,c], R. Bellan[a,b], C. Biino[a], N. Cartiglia[a], M. Costa[a,b], R. Covarelli[a,b], A. Degano[a,b], N. Demaria[a], L. Finco[a,b,2], B. Kiani[a,b], C. Mariotti[a], S. Maselli[a], E. Migliore[a,b], V. Monaco[a,b], E. Monteil[a,b], M. Musich[a], M.M. Obertino[a,b], L. Pacher[a,b], N. Pastrone[a], M. Pelliccioni[a], G.L. Pinna Angioni[a,b], F. Ravera[a,b], A. Romero[a,b], M. Ruspa[a,c], R. Sacchi[a,b], A. Solano[a,b], A. Staiano[a], U. Tamponi[a]

INFN Sezione di Trieste [a], Università di Trieste [b], Trieste, Italy
S. Belforte[a], V. Candelise[a,b,2], M. Casarsa[a], F. Cossutti[a], G. Della Ricca[a,b], B. Gobbo[a], C. La Licata[a,b], M. Marone[a,b], A. Schizzi[a,b], A. Zanetti[a]

Kangwon National University, Chunchon, Korea
A. Kropivnitskaya, S.K. Nam

Kyungpook National University, Daegu, Korea
D.H. Kim, G.N. Kim, M.S. Kim, D.J. Kong, S. Lee, Y.D. Oh, A. Sakharov, D.C. Son

Chonbuk National University, Jeonju, Korea
J.A. Brochero Cifuentes, H. Kim, T.J. Kim, M.S. Ryu

Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Korea
S. Song

Korea University, Seoul, Korea
S. Choi, Y. Go, D. Gyun, B. Hong, M. Jo, H. Kim, Y. Kim, B. Lee, K. Lee, K.S. Lee, S. Lee, S.K. Park, Y. Roh

Seoul National University, Seoul, Korea
H.D. Yoo

University of Seoul, Seoul, Korea
M. Choi, H. Kim, J.H. Kim, J.S.H. Lee, I.C. Park, G. Ryu

Sungkyunkwan University, Suwon, Korea
Y. Choi, Y.K. Choi, J. Goh, D. Kim, E. Kwon, J. Lee, I. Yu

Vilnius University, Vilnius, Lithuania
A. Juodagalvis, J. Vaitkus

National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
I. Ahmed, Z.A. Ibrahim, J.R. Komaragiri, M.A.B. Md Ali[32], F. Mohamad Idris[33], W.A.T. Wan Abdullah, M.N. Yusli

Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico
E. Casimiro Linares, H. Castilla-Valdez, E. De La Cruz-Burelo, I. Heredia-de La Cruz[34], A. Hernandez-Almada, R. Lopez-Fernandez, A. Sanchez-Hernandez

Universidad Iberoamericana, Mexico City, Mexico
S. Carrillo Moreno, F. Vazquez Valencia

Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
I. Pedraza, H.A. Salazar Ibarguen

Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico
A. Morelos Pineda

University of Auckland, Auckland, New Zealand
D. Krofcheck

University of Canterbury, Christchurch, New Zealand
P.H. Butler

National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
A. Ahmad, M. Ahmad, Q. Hassan, H.R. Hoorani, W.A. Khan, T. Khurshid, M. Shoaib

National Centre for Nuclear Research, Swierk, Poland
H. Bialkowska, M. Bluj, B. Boimska, T. Frueboes, M. Górski, M. Kazana, K. Nawrocki, K. Romanowska-Rybinska, M. Szleper, P. Zalewski

Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland
G. Brona, K. Bunkowski, K. Doroba, A. Kalinowski, M. Konecki, J. Krolikowski, M. Misiura, M. Olszewski, M. Walczak

Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
P. Bargassa, C. Beirão Da Cruz E Silva, A. Di Francesco, P. Faccioli, P.G. Ferreira Parracho, M. Gallinaro, N. Leonardo, L. Lloret Iglesias, F. Nguyen, J. Rodrigues Antunes, J. Seixas, O. Toldaiev, D. Vadruccio, J. Varela, P. Vischia

Joint Institute for Nuclear Research, Dubna, Russia
S. Afanasiev, P. Bunin, M. Gavrilenko, I. Golutvin, I. Gorbunov, A. Kamenev, V. Karjavin, V. Konoplyanikov, A. Lanev, A. Malakhov, V. Matveev[35], P. Moisenz, V. Palichik, V. Perelygin, S. Shmatov, S. Shulha, N. Skatchkov, V. Smirnov, A. Zarubin

Petersburg Nuclear Physics Institute, Gatchina (St. Petersburg), Russia
V. Golovtsov, Y. Ivanov, V. Kim[36], E. Kuznetsova, P. Levchenko, V. Murzin, V. Oreshkin, I. Smirnov, V. Sulimov, L. Uvarov, S. Vavilov, A. Vorobyev

Institute for Nuclear Research, Moscow, Russia
Yu. Andreev, A. Dermenev, S. Gninenko, N. Golubev, A. Karneyeu, M. Kirsanov, N. Krasnikov, A. Pashenkov, D. Tlisov, A. Toropin

Institute for Theoretical and Experimental Physics, Moscow, Russia
V. Epshteyn, V. Gavrilov, N. Lychkovskaya, V. Popov, I. Pozdnyakov, G. Safronov, A. Spiridonov, E. Vlasov, A. Zhokin

National Research Nuclear University 'Moscow Engineering Physics Institute' (MEPhI), Moscow, Russia
A. Bylinkin

P.N. Lebedev Physical Institute, Moscow, Russia
V. Andreev, M. Azarkin[37], I. Dremin[37], M. Kirakosyan, A. Leonidov[37], G. Mesyats, S.V. Rusakov, A. Vinogradov

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
A. Baskakov, A. Belyaev, E. Boos, M. Dubinin[38], L. Dudko, A. Ershov, A. Gribushin, A. Kaminskiy[39], V. Klyukhin, O. Kodolova, I. Lokhtin, I. Myagkov, S. Obraztsov, S. Petrushanko, V. Savrin

State Research Center of Russian Federation, Institute for High Energy Physics, Protvino, Russia
I. Azhgirey, I. Bayshev, S. Bitioukov, V. Kachanov, A. Kalinin, D. Konstantinov, V. Krychkine, I. Kurochkin, V. Petrov, R. Ryutin, A. Sobol, L. Tourtchanovitch, S. Troshin, N. Tyurin, A. Uzunian, A. Volkov

University of Belgrade, Faculty of Physics and Vinca Institute of Nuclear Sciences, Belgrade, Serbia
P. Adzic[40], M. Ekmedzic, J. Milosevic, V. Rekovic

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
J. Alcaraz Maestre, E. Calvo, M. Cerrada, M. Chamizo Llatas, N. Colino, B. De La Cruz, A. Delgado Peris, D. Domínguez Vázquez, A. Escalante Del Valle, C. Fernandez Bedoya, J.P. Fernández Ramos, J. Flix, M.C. Fouz, P. Garcia-Abia, O. Gonzalez Lopez, S. Goy Lopez, J.M. Hernandez, M.I. Josa, E. Navarro De Martino, A. Pérez-Calero Yzquierdo, J. Puerta Pelayo, A. Quintario Olmeda, I. Redondo, L. Romero, M.S. Soares

Universidad Autónoma de Madrid, Madrid, Spain
C. Albajar, J.F. de Trocóniz, M. Missiroli, D. Moran

Universidad de Oviedo, Oviedo, Spain
H. Brun, J. Cuevas, J. Fernandez Menendez, S. Folgueras, I. Gonzalez Caballero, E. Palencia Cortezon, J.M. Vizan Garcia

Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
I.J. Cabrillo, A. Calderon, J.R. Castiñeiras De Saa, P. De Castro Manzano, J. Duarte Campderros, M. Fernandez, J. Garcia-Ferrero, G. Gomez, A. Lopez Virto, J. Marco, R. Marco, C. Martinez Rivero, F. Matorras, F.J. Munoz Sanchez, J. Piedra Gomez, T. Rodrigo, A.Y. Rodríguez-Marrero, A. Ruiz-Jimeno, L. Scodellaro, I. Vila, R. Vilar Cortabitarte

CERN, European Organization for Nuclear Research, Geneva, Switzerland
D. Abbaneo, P. Aspell, E. Auffray, G. Auzinger, M. Bachtis, P. Baillon, A.H. Ball, S.L. Bally, L. Bardo, D. Barney, A. Benaglia, J. Bendavid, L. Benhabib, G.M. Berruti, G. Blanchot, P. Bloch, A. Bocci, A. Bonato, C. Botta, T. Camporesi, D. Ceresa, G. Cerminara, A. Charkiewicz, J. Christiansen, K. Cichy, S. Colafranceschi[41], A. Conde Garcia, E. Conti, T. Cooijmans, B. Curé, M. D'Alfonso, D. d'Enterria, A. Dabrowski, M. Dabrowski, J. Daguin, V. Daponte, A. David, M. De Gruttola, F. De Guio, A. De Roeck, S. De Visscher, E. Di Marco, M. Dobson, M. Dordevic, B. Dorney, T. du Pree, M. Dünser, N. Dupont, A. Elliott-Peisert, R. Favre-felix, G. Franzoni, W. Funk, A. Gaddi, T. Gadek, M. Gastal, H. Gerwig, D. Gigi, K. Gill, M. Girone, F. Glege, R. Guida, S. Gundacker, M. Guthoff, J. Hammer, M. Hansen, P. Harris, J. Hegeman, A. Honma, V. Innocente, P. Janot, J. Kaplon, H. Kirschenmann, K. Kloukinas, M.J. Kortelainen, K. Kousouris, M.I. Kovács, K. Krajczar, P. Lecoq, R. Loos, C. Lourenço, M.T. Lucchini, N. Magini, L. Malgeri, M. Mannelli, A. Marchioro, A. Martelli, L. Masetti, F. Meijers, S. Mersi, E. Meschi, F. Moortgat, S. Morovic, M. Mulders, M.V. Nemallapudi, H. Neugebauer, A. Onnela, S. Orfanelli[42], L. Orsini, L. Pape, E. Perez, M. Peruzzi, A. Petrilli, G. Petrucciani, A. Pfeiffer, D. Piparo, H. Postema, A. Racz, G. Rolandi[43], P. Rose, M. Rovere, M. Ruan, V. Ryjov, H. Sakulin, C. Schäfer, C. Schwick, A. Sharma, P. Silva, P. Sphicas[44], J. Steggemann, B. Stieger, M. Stoye, Y. Takahashi, D. Treille, A. Triossi, P. Tropea, A. Tsirou, F. Vasey, G.I. Veres[20], P. Vichoudis, N. Wardle, P. Wertelaers, H.K. Wöhri, A. Zagozdzinska[45], W.D. Zeuner

Paul Scherrer Institut, Villigen, Switzerland
W. Bertl, K. Deiters, W. Erdmann, R. Horisberger, Q. Ingram, H.C. Kaestli, D. Kotlinski,

U. Langenegger, D. Renker, T. Rohe

Institute for Particle Physics, ETH Zurich, Zurich, Switzerland
F. Bachmair, L. Bäni, L. Bianchini, M.A. Buchmann, B. Casal, G. Dissertori, M. Dittmar, M. Donegà, P. Eller, C. Grab, C. Heidegger, D. Hits, J. Hoss, G. Kasieczka, W. Lustermann, B. Mangano, M. Marionneau, P. Martinez Ruiz del Arbol, M. Masciovecchio, D. Meister, F. Micheli, P. Musella, F. Nessi-Tedaldi, F. Pandolfi, J. Pata, F. Pauss, L. Perrozzi, M. Quittnat, M. Rossini, A. Starodumov[46], M. Takahashi, V.R. Tavolaro, K. Theofilatos, R. Wallny

Universität Zürich, Zurich, Switzerland
T.K. Aarrestad, C. Amsler[47], L. Caminada, M.F. Canelli, V. Chiochia, A. De Cosa, C. Galloni, A. Hinzmann, T. Hreus, B. Kilminster, C. Lange, J. Ngadiuba, D. Pinna, P. Robmann, F.J. Ronga, D. Salerno, Y. Yang

National Central University, Chung-Li, Taiwan
M. Cardaci, K.H. Chen, T.H. Doan, Sh. Jain, R. Khurana, M. Konyushikhin, C.M. Kuo, W. Lin, Y.J. Lu, S.S. Yu

National Taiwan University (NTU), Taipei, Taiwan
Arun Kumar, R. Bartek, P. Chang, Y.H. Chang, Y.W. Chang, Y. Chao, K.F. Chen, P.H. Chen, C. Dietz, F. Fiori, U. Grundler, W.-S. Hou, Y. Hsiung, Y.F. Liu, R.-S. Lu, M. Miñano Moya, E. Petrakou, J.F. Tsai, Y.M. Tzeng

Chulalongkorn University, Faculty of Science, Department of Physics, Bangkok, Thailand
B. Asavapibhop, K. Kovitanggoon, G. Singh, N. Srimanobhas, N. Suwonjandee

Cukurova University, Adana, Turkey
A. Adiguzel, S. Cerci[48], Z.S. Demiroglu, C. Dozen, S. Girgis, G. Gokbulut, Y. Guler, E. Gurpinar, I. Hos, E.E. Kangal[49], A. Kayis Topaksu, G. Onengut[50], K. Ozdemir[51], S. Ozturk[52], B. Tali[48], H. Topakli[52], M. Vergili, C. Zorbilmez

Middle East Technical University, Physics Department, Ankara, Turkey
I.V. Akin, B. Bilin, S. Bilmis, B. Isildak[53], G. Karapinar[54], M. Yalvac, M. Zeyrek

Bogazici University, Istanbul, Turkey
E.A. Albayrak[55], E. Gülmez, M. Kaya[56], O. Kaya[57], T. Yetkin[58]

Istanbul Technical University, Istanbul, Turkey
K. Cankocak, S. Sen[59], F.I. Vardarlı

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkov, Ukraine
B. Grynyov

National Scientific Center, Kharkov Institute of Physics and Technology, Kharkov, Ukraine
L. Levchuk, P. Sorokin

University of Bristol, Bristol, United Kingdom
R. Aggleton, F. Ball, L. Beck, J.J. Brooke, E. Clement, D. Cussans, H. Flacher, J. Goldstein, M. Grimes, G.P. Heath, H.F. Heath, J. Jacob, L. Kreczko, C. Lucas, Z. Meng, D.M. Newbold[60], S. Paramesvaran, A. Poll, T. Sakuma, S. Seif El Nasr-storey, S. Senkin, D. Smith, V.J. Smith

Rutherford Appleton Laboratory, Didcot, United Kingdom
D. Barducci, K.W. Bell, A. Belyaev[61], C. Brew, R.M. Brown, D. Cieri, D.J.A. Cockerill, J.A. Coughlan, K. Harder, S. Harper, E. Olaiya, D. Petyt, C.H. Shepherd-Themistocleous, A. Thea, L. Thomas, I.R. Tomalin, T. Williams, W.J. Womersley, S.D. Worm

Imperial College, London, United Kingdom
M. Baber, R. Bainbridge, O. Buchmuller, A. Bundock, D. Burton, S. Casasso, M. Citron, D. Colling, L. Corpe, N. Cripps, P. Dauncey, G. Davies, A. De Wit, M. Della Negra, P. Dunne, A. Elwood, W. Ferguson, J. Fulcher, D. Futyan, G. Hall, G. Iles, M. Kenzie, R. Lane, R. Lucas[60], L. Lyons, A.-M. Magnan, S. Malik, J. Nash, A. Nikitenko[46], J. Pela, M. Pesaresi, K. Petridis, D.M. Raymond, A. Richards, A. Rose, C. Seez, A. Tapper, K. Uchida, M. Vazquez Acosta[62], T. Virdee, S.C. Zenz

Brunel University, Uxbridge, United Kingdom
J.E. Cole, P.R. Hobson, A. Khan, P. Kyberd, D. Leggat, D. Leslie, I.D. Reid, P. Symonds, L. Teodorescu, M. Turner

Baylor University, Waco, USA
A. Borzou, K. Call, J. Dittmann, K. Hatakeyama, A. Kasmi, H. Liu, N. Pastika

The University of Alabama, Tuscaloosa, USA
O. Charaf, S.I. Cooper, C. Henderson, P. Rumerio

Boston University, Boston, USA
A. Avetisyan, T. Bose, C. Fantasia, D. Gastler, P. Lawson, D. Rankin, C. Richardson, J. Rohlf, J. St. John, L. Sulak, D. Zou

Brown University, Providence, USA
J. Alimena, E. Berry, S. Bhattacharya, D. Cutts, N. Dhingra, A. Ferapontov, A. Garabedian, U. Heintz, E. Laird, G. Landsberg, Z. Mao, M. Narain, S. Piperov, S. Sagir, T. Sinthuprasith, R. Syarif

University of California, Davis, Davis, USA
R. Breedon, G. Breto, M. Calderon De La Barca Sanchez, S. Chauhan, M. Chertok, J. Conway, R. Conway, P.T. Cox, R. Erbacher, M. Gardner, W. Ko, R. Lander, M. Mulhearn, D. Pellett, J. Pilot, F. Ricci-Tam, S. Shalhout, J. Smith, M. Squires, D. Stolp, M. Tripathi, S. Wilbur, R. Yohay

University of California, Los Angeles, USA
R. Cousins, P. Everaerts, C. Farrell, J. Hauser, M. Ignatenko, D. Saltzberg, E. Takasugi, V. Valuev, M. Weber

University of California, Riverside, Riverside, USA
K. Burt, R. Clare, J. Ellison, J.W. Gary, G. Hanson, J. Heilman, M. Ivova PANEVA, P. Jandir, E. Kennedy, F. Lacroix, O.R. Long, A. Luthra, M. Malberti, M. Olmedo Negrete, A. Shrinivas, H. Wei, S. Wimpenny

University of California, San Diego, La Jolla, USA
J.G. Branson, G.B. Cerati, S. Cittolin, R.T. D'Agnolo, A. Holzner, R. Kelley, D. Klein, J. Letts, I. Macneill, D. Olivito, S. Padhi, M. Pieri, M. Sani, V. Sharma, S. Simon, M. Tadel, A. Vartak, S. Wasserbaech[63], C. Welke, F. Würthwein, A. Yagil, G. Zevi Della Porta

University of California, Santa Barbara, Santa Barbara, USA
D. Barge, J. Bradmiller-Feld, C. Campagnari, A. Dishaw, V. Dutta, K. Flowers, M. Franco Sevilla, P. Geffert, C. George, F. Golf, L. Gouskos, J. Gran, J. Incandela, C. Justus, N. Mccoll, S.D. Mullin, J. Richman, D. Stuart, I. Suarez, W. To, C. West, J. Yoo

California Institute of Technology, Pasadena, USA
D. Anderson, A. Apresyan, A. Bornheim, J. Bunn, Y. Chen, J. Duarte, A. Mott, H.B. Newman, C. Pena, M. Pierini, M. Spiropulu, J.R. Vlimant, S. Xie, R.Y. Zhu

Carnegie Mellon University, Pittsburgh, USA
V. Azzolini, A. Calamba, B. Carlson, T. Ferguson, M. Paulini, J. Russ, M. Sun, H. Vogel, I. Vorobiev

University of Colorado Boulder, Boulder, USA
J.P. Cumalat, W.T. Ford, A. Gaz, F. Jensen, A. Johnson, M. Krohn, T. Mulholland, U. Nauenberg, K. Stenson, S.R. Wagner

Cornell University, Ithaca, USA
J. Alexander, A. Chatterjee, J. Chaves, J. Chu, S. Dittmer, N. Eggert, N. Mirman, G. Nicolas Kaufman, J.R. Patterson, A. Rinkevicius, A. Ryd, L. Skinnari, L. Soffi, W. Sun, S.M. Tan, W.D. Teo, J. Thom, J. Thompson, J. Tucker, Y. Weng, P. Wittich

Fermi National Accelerator Laboratory, Batavia, USA
S. Abdullin, M. Albrow, J. Anderson, G. Apollinari, L.A.T. Bauerdick, A. Beretvas, J. Berryhill, P.C. Bhat, G. Bolla, K. Burkett, J.N. Butler, H.W.K. Cheung, F. Chlebana, S. Cihangir, V.D. Elvira, I. Fisk, J. Freeman, E. Gottschalk, L. Gray, D. Green, S. Grünendahl, O. Gutsche, J. Hanlon, D. Hare, R.M. Harris, J. Hirschauer, B. Hooberman, Z. Hu, S. Jindariani, M. Johnson, U. Joshi, A.W. Jung, B. Klima, B. Kreis, S. Kwan†, S. Lammel, J. Linacre, D. Lincoln, R. Lipton, T. Liu, R. Lopes De Sá, J. Lykken, K. Maeshima, J.M. Marraffino, V.I. Martinez Outschoorn, S. Maruyama, D. Mason, P. McBride, P. Merkel, K. Mishra, S. Mrenna, S. Nahn, C. Newman-Holmes, V. O'Dell, K. Pedro, O. Prokofyev, G. Rakness, E. Sexton-Kennedy, A. Soha, W.J. Spalding, L. Spiegel, L. Taylor, S. Tkaczyk, N.V. Tran, L. Uplegger, E.W. Vaandering, C. Vernieri, M. Verzocchi, R. Vidal, H.A. Weber, A. Whitbeck, F. Yang

University of Florida, Gainesville, USA
D. Acosta, P. Avery, P. Bortignon, D. Bourilkov, A. Carnes, M. Carver, D. Curry, S. Das, G.P. Di Giovanni, R.D. Field, I.K. Furic, J. Hugon, J. Konigsberg, A. Korytov, J.F. Low, P. Ma, K. Matchev, H. Mei, P. Milenovic[64], G. Mitselmakher, D. Rank, R. Rossin, L. Shchutska, M. Snowball, D. Sperka, J. Wang, S. Wang, J. Yelton

Florida International University, Miami, USA
S. Hewamanage, S. Linn, P. Markowitz, G. Martinez, J.L. Rodriguez

Florida State University, Tallahassee, USA
A. Ackert, J.R. Adams, T. Adams, A. Askew, J. Bochenek, B. Diamond, J. Haas, S. Hagopian, V. Hagopian, K.F. Johnson, A. Khatiwada, H. Prosper, V. Veeraraghavan, M. Weinberg

Florida Institute of Technology, Melbourne, USA
M.M. Baarmand, V. Bhopatkar, M. Hohlmann, H. Kalakhety, D. Noonan, T. Roy, F. Yumiceva

University of Illinois at Chicago (UIC), Chicago, USA
M.R. Adams, L. Apanasevich, D. Berry, R.R. Betts, I. Bucinskaite, R. Cavanaugh, O. Evdokimov, L. Gauthier, C.E. Gerber, D.J. Hofman, P. Kurt, C. O'Brien, I.D. Sandoval Gonzalez, C. Silkworth, P. Turner, N. Varelas, Z. Wu, M. Zakaria

The University of Iowa, Iowa City, USA
B. Bilki[65], W. Clarida, K. Dilsiz, S. Durgut, R.P. Gandrajula, M. Haytmyradov, V. Khristenko, J.-P. Merlo, H. Mermerkaya[66], A. Mestvirishvili, A. Moeller, J. Nachtman, H. Ogul, Y. Onel, F. Ozok[55], A. Penzo, C. Snyder, P. Tan, E. Tiras, J. Wetzel, K. Yi

Johns Hopkins University, Baltimore, USA
I. Anderson, B.A. Barnett, B. Blumenfeld, D. Fehling, L. Feng, A.V. Gritsan, P. Maksimovic, C. Martin, M. Osherson, M. Swartz, M. Xiao, Y. Xin, C. You

The University of Kansas, Lawrence, USA
P. Baringer, A. Bean, G. Benelli, C. Bruner, R.P. Kenny III, D. Majumder, M. Malek, M. Murray, S. Sanders, R. Stringer, Q. Wang, J.S. Wood

Kansas State University, Manhattan, USA
A. Ivanov, K. Kaadze, S. Khalil, M. Makouski, Y. Maravin, A. Mohammadi, L.K. Saini, N. Skhirtladze, S. Toda

Lawrence Livermore National Laboratory, Livermore, USA
D. Lange, F. Rebassoo, D. Wright

University of Maryland, College Park, USA
C. Anelli, A. Baden, O. Baron, A. Belloni, B. Calvert, S.C. Eno, C. Ferraioli, J.A. Gomez, N.J. Hadley, S. Jabeen, R.G. Kellogg, T. Kolberg, J. Kunkle, Y. Lu, A.C. Mignerey, Y.H. Shin, A. Skuja, M.B. Tonjes, S.C. Tonwar

Massachusetts Institute of Technology, Cambridge, USA
A. Apyan, R. Barbieri, A. Baty, K. Bierwagen, S. Brandt, W. Busza, I.A. Cali, Z. Demiragli, L. Di Matteo, G. Gomez Ceballos, M. Goncharov, D. Gulhan, Y. Iiyama, G.M. Innocenti, M. Klute, D. Kovalskyi, Y.S. Lai, Y.-J. Lee, A. Levin, P.D. Luckey, A.C. Marini, C. Mcginn, C. Mironov, X. Niu, C. Paus, D. Ralph, C. Roland, G. Roland, J. Salfeld-Nebgen, G.S.F. Stephans, K. Sumorok, M. Varma, D. Velicanu, J. Veverka, J. Wang, T.W. Wang, B. Wyslouch, M. Yang, V. Zhukova

University of Minnesota, Minneapolis, USA
B. Dahmes, A. Finkel, E. Frahm, W.J. Gilbert, A. Gude, P. Hansen, S. Kalafut, S.C. Kao, K. Klapoetke, Y. Kubota, Z. Lesko, J. Mans, S. Nourbakhsh, N. Ruckstuhl, R. Rusack, N. Tambe, J. Turkewitz

University of Mississippi, Oxford, USA
J.G. Acosta, S. Oliveros

University of Nebraska-Lincoln, Lincoln, USA
E. Avdeeva, K. Bloom, S. Bose, D.R. Claes, A. Dominguez, C. Fangmeier, R. Gonzalez Suarez, R. Kamalieddin, J. Keller, D. Knowlton, I. Kravchenko, J. Lazo-Flores, F. Meier, J. Monroy, F. Ratnikov, J.E. Siado, G.R. Snow

State University of New York at Buffalo, Buffalo, USA
M. Alyari, J. Dolen, J. George, A. Godshalk, C. Harrington, I. Iashvili, J. Kaisen, A. Kharchilava, A. Kumar, S. Rappoccio

Northeastern University, Boston, USA
G. Alverson, E. Barberis, D. Baumgartel, M. Chasco, A. Hortiangtham, B. Knapp, A. Massironi, D.M. Morse, D. Nash, T. Orimoto, R. Teixeira De Lima, D. Trocino, R.-J. Wang, D. Wood, J. Zhang

Northwestern University, Evanston, USA
K.A. Hahn, A. Kubik, N. Mucia, N. Odell, B. Pollack, A. Pozdnyakov, M. Schmitt, S. Stoynev, K. Sung, M. Trovato, M. Velasco

University of Notre Dame, Notre Dame, USA
A. Brinkerhoff, N. Dev, M. Hildreth, C. Jessop, D.J. Karmgard, N. Kellams, K. Lannon, S. Lynch, N. Marinelli, F. Meng, C. Mueller, Y. Musienko[35], T. Pearson, M. Planer, A. Reinsvold, R. Ruchti, G. Smith, S. Taroni, N. Valls, M. Wayne, M. Wolf, A. Woodard

The Ohio State University, Columbus, USA
L. Antonelli, J. Brinson, B. Bylsma, L.S. Durkin, S. Flowers, A. Hart, C. Hill, R. Hughes, K. Kotov, T.Y. Ling, B. Liu, W. Luo, D. Puigh, M. Rodenburg, B.L. Winer, H.W. Wulsin

Princeton University, Princeton, USA
O. Driga, P. Elmer, J. Hardenbrook, P. Hebda, S.A. Koay, P. Lujan, D. Marlow, T. Medvedeva, M. Mooney, J. Olsen, C. Palmer, P. Piroué, X. Quan, H. Saka, D. Stickland, C. Tully, J.S. Werner, A. Zuranski

University of Puerto Rico, Mayaguez, USA
S. Malik

Purdue University, West Lafayette, USA
V.E. Barnes, D. Benedetti, D. Bortoletto, L. Gutay, M.K. Jha, M. Jones, K. Jung, M. Kress, D.H. Miller, N. Neumeister, B.C. Radburn-Smith, X. Shi, I. Shipsey, D. Silvers, J. Sun, A. Svyatkovskiy, F. Wang, W. Xie, L. Xu

Purdue University Calumet, Hammond, USA
N. Parashar, J. Stupak

Rice University, Houston, USA
A. Adair, B. Akgun, Z. Chen, K.M. Ecklund, F.J.M. Geurts, M. Guilbaud, W. Li, B. Michlin, M. Northup, B.P. Padley, R. Redjimi, J. Roberts, J. Rorie, Z. Tu, J. Zabel

University of Rochester, Rochester, USA
B. Betchart, A. Bodek, P. de Barbaro, R. Demina, Y. Eshaq, T. Ferbel, M. Galanti, A. Garcia-Bellido, P. Goldenzweig, J. Han, A. Harel, O. Hindrichs, A. Khukhunaishvili, G. Petrillo, M. Verzetti

The Rockefeller University, New York, USA
L. Demortier

Rutgers, The State University of New Jersey, Piscataway, USA
S. Arora, A. Barker, J.P. Chou, C. Contreras-Campana, E. Contreras-Campana, D. Duggan, D. Ferencek, Y. Gershtein, R. Gray, E. Halkiadakis, D. Hidas, E. Hughes, S. Kaplan, R. Kunnawalkam Elayavalli, A. Lath, K. Nash, S. Panwalkar, M. Park, S. Salur, S. Schnetzer, D. Sheffield, S. Somalwar, R. Stone, S. Thomas, P. Thomassen, M. Walker

University of Tennessee, Knoxville, USA
M. Foerster, G. Riley, K. Rose, S. Spanier, A. York

Texas A&M University, College Station, USA
O. Bouhali[67], A. Castaneda Hernandez[67], A. Celik, M. Dalchenko, M. De Mattia, A. Delgado, S. Dildick, R. Eusebi, W. Flanagan, J. Gilmore, T. Huang, T. Kamon[68], V. Krutelyov, R. Montalvo, R. Mueller, I. Osipenkov, Y. Pakhotin, R. Patel, A. Perloff, J. Roe, A. Rose, A. Safonov, A. Tatarinov, K.A. Ulmer[2]

Texas Tech University, Lubbock, USA
N. Akchurin, C. Cowden, J. Damgov, C. Dragoiu, P.R. Dudero, J. Faulkner, S. Kunori, K. Lamichhane, S.W. Lee, T. Libeiro, S. Undleeb, I. Volobouev

Vanderbilt University, Nashville, USA
E. Appelt, A.G. Delannoy, S. Greene, A. Gurrola, R. Janjam, W. Johns, C. Maguire, Y. Mao, A. Melo, H. Ni, P. Sheldon, B. Snook, S. Tuo, J. Velkovska, Q. Xu

University of Virginia, Charlottesville, USA
M.W. Arenton, S. Boutle, B. Cox, B. Francis, J. Goodell, R. Hirosky, A. Ledovskoy, H. Li, C. Lin, C. Neu, E. Wolfe, J. Wood, F. Xia

Wayne State University, Detroit, USA
C. Clarke, R. Harr, P.E. Karchin, C. Kottachchi Kankanamge Don, P. Lamichhane, J. Sturdy

University of Wisconsin, Madison, USA
D.A. Belknap, D. Carlsmith, M. Cepeda, A. Christian, S. Dasu, L. Dodd, S. Duric, E. Friis, B. Gomber, R. Hall-Wilton, M. Herndon, A. Hervé, P. Klabbers, A. Lanaro, A. Levine, K. Long, R. Loveless, A. Mohapatra, I. Ojalvo, T. Perry, G.A. Pierro, G. Polese, I. Ross, T. Ruggles, T. Sarangi, A. Savin, A. Sharma, N. Smith, W.H. Smith, D. Taylor, N. Woods

†: Deceased
1: Also at Vienna University of Technology, Vienna, Austria
2: Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
3: Also at State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China
4: Also at Institut Pluridisciplinaire Hubert Curien, Université de Strasbourg, Université de Haute Alsace Mulhouse, CNRS/IN2P3, Strasbourg, France
5: Also at National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
6: Also at Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University, Moscow, Russia
7: Also at Universidade Estadual de Campinas, Campinas, Brazil
8: Also at Centre National de la Recherche Scientifique (CNRS) - IN2P3, Paris, France
9: Also at Laboratoire Leprince-Ringuet, Ecole Polytechnique, IN2P3-CNRS, Palaiseau, France
10: Also at Joint Institute for Nuclear Research, Dubna, Russia
11: Also at Beni-Suef University, Bani Sweif, Egypt
12: Now at British University in Egypt, Cairo, Egypt
13: Now at Ain Shams University, Cairo, Egypt
14: Also at Zewail City of Science and Technology, Zewail, Egypt
15: Also at Université de Haute Alsace, Mulhouse, France
16: Also at Tbilisi State University, Tbilisi, Georgia
17: Also at University of Hamburg, Hamburg, Germany
18: Also at Brandenburg University of Technology, Cottbus, Germany
19: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
20: Also at Eötvös Loránd University, Budapest, Hungary
21: Also at University of Debrecen, Debrecen, Hungary
22: Also at Wigner Research Centre for Physics, Budapest, Hungary
23: Also at University of Visva-Bharati, Santiniketan, India
24: Now at King Abdulaziz University, Jeddah, Saudi Arabia
25: Also at University of Ruhuna, Matara, Sri Lanka
26: Also at Isfahan University of Technology, Isfahan, Iran
27: Also at University of Tehran, Department of Engineering Science, Tehran, Iran
28: Also at Plasma Physics Research Center, Science and Research Branch, Islamic Azad University, Tehran, Iran
29: Also at Laboratori Nazionali di Legnaro dell'INFN, Legnaro, Italy

30: Also at Università degli Studi di Siena, Siena, Italy
31: Also at Purdue University, West Lafayette, USA
32: Also at International Islamic University of Malaysia, Kuala Lumpur, Malaysia
33: Also at Malaysian Nuclear Agency, MOSTI, Kajang, Malaysia
34: Also at Consejo Nacional de Ciencia y Tecnología, Mexico city, Mexico
35: Also at Institute for Nuclear Research, Moscow, Russia
36: Also at St. Petersburg State Polytechnical University, St. Petersburg, Russia
37: Also
at National Research Nuclear University 'Moscow Engineering Physics Institute' (MEPhI), Moscow, Russia
38: Also at California Institute of Technology, Pasadena, USA
39: Also at INFN Sezione di Padova; Università di Padova; Università di Trento (Trento), Padova, Italy
40: Also at Faculty of Physics, University of Belgrade, Belgrade, Serbia
41: Also at Facoltà Ingegneria, Università di Roma, Roma, Italy
42: Also at National Technical University of Athens, Athens, Greece
43: Also at Scuola Normale e Sezione dell'INFN, Pisa, Italy
44: Also at University of Athens, Athens, Greece
45: Also at Warsaw University of Technology, Institute of Electronic Systems, Warsaw, Poland
46: Also at Institute for Theoretical and Experimental Physics, Moscow, Russia
47: Also at Albert Einstein Center for Fundamental Physics, Bern, Switzerland
48: Also at Adiyaman University, Adiyaman, Turkey
49: Also at Mersin University, Mersin, Turkey
50: Also at Cag University, Mersin, Turkey
51: Also at Piri Reis University, Istanbul, Turkey
52: Also at Gaziosmanpasa University, Tokat, Turkey
53: Also at Ozyegin University, Istanbul, Turkey
54: Also at Izmir Institute of Technology, Izmir, Turkey
55: Also at Mimar Sinan University, Istanbul, Istanbul, Turkey
56: Also at Marmara University, Istanbul, Turkey
57: Also at Kafkas University, Kars, Turkey
58: Also at Yildiz Technical University, Istanbul, Turkey
59: Also at Hacettepe University, Ankara, Turkey
60: Also at Rutherford Appleton Laboratory, Didcot, United Kingdom
61: Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
62: Also at Instituto de Astrofísica de Canarias, La Laguna, Spain
63: Also at Utah Valley University, Orem, USA
64: Also at University of Belgrade, Faculty of Physics and Vinca Institute of Nuclear Sciences, Belgrade, Serbia
65: Also at Argonne National Laboratory, Argonne, USA
66: Also at Erzincan University, Erzincan, Turkey
67: Also at Texas A&M University at Qatar, Doha, Qatar
68: Also at Kyungpook National University, Daegu, Korea

Contents

Figures

Tables

Chapter 1

The Path to New Discoveries for CMS at the High Luminosity LHC

I am on the edge of mysteries and the veil is getting thinner and thinner – Louis Pasteur

1.1 Introduction

The Physics Program of the CMS experiment at the Large Hadron Collider (LHC) is off to a remarkable start! It is aimed at answering fundamental questions in particle physics. What is the origin of elementary particle masses? What is the nature of the dark matter we observe in the Universe? Are the fundamental forces unified? How does QCD behave under extreme conditions? Do matter and antimatter properties differ? In the first major physics run in 2011 and 2012, the collider reached a peak luminosity of $7.7 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$, more than 75% of its design luminosity, and delivered an integrated luminosity of $\sim$25 fb^{-1} to each of its two general purpose experiments, ATLAS and CMS. This data has yielded a vast quantity of physics results, summarized by the CMS collaboration in more than 300 publications. The highlight has been the observation in 2012 of a new particle of mass $\sim$125 GeV by the ATLAS and CMS collaborations [1, 2], which partially provides an answer to the first question. This particle was identified as a Higgs Boson. Figure 1.1 shows Higgs boson candidates from CMS. Figure 1.2 shows the signals from CMS that contributed to establishing the existence of this new particle.

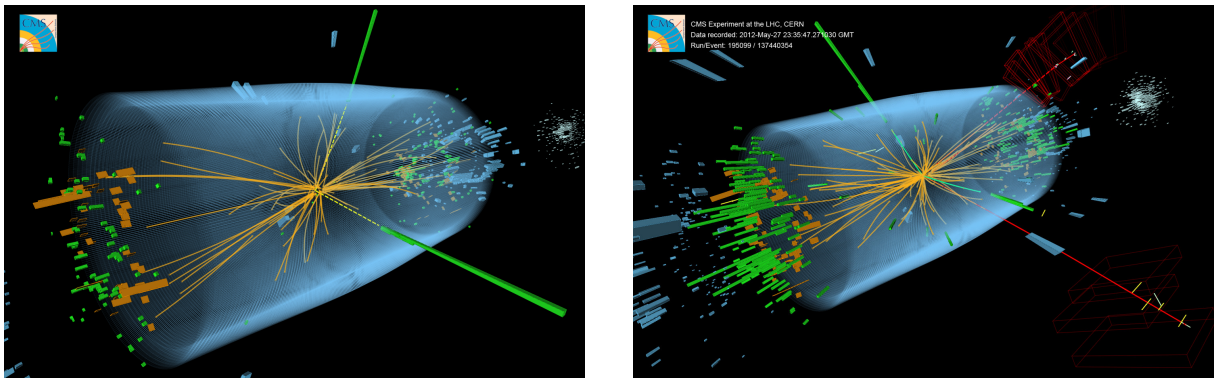
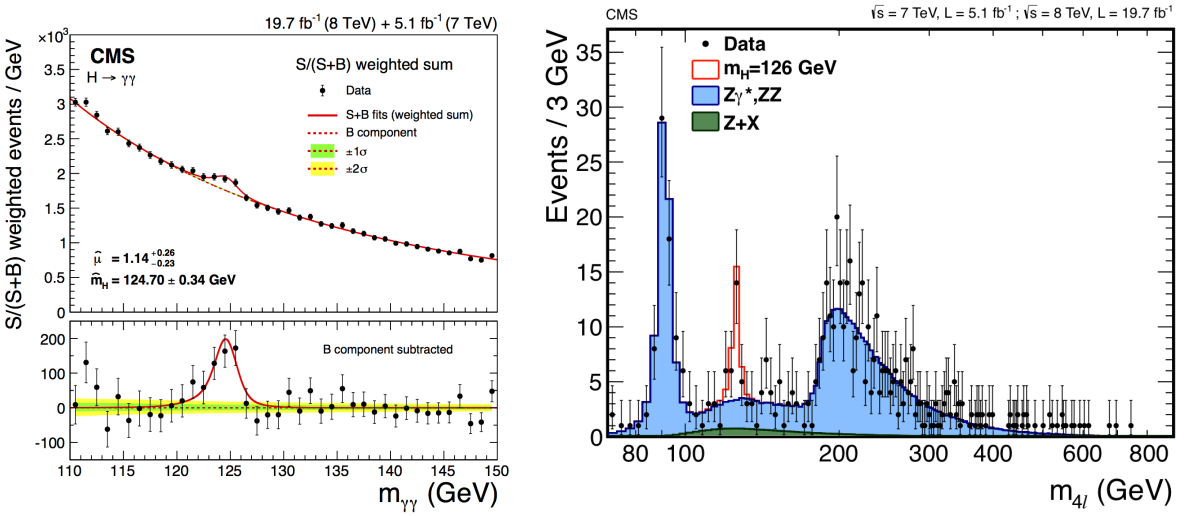


Figure 1.1: (a) Candidate for the decay Higgs $\to \gamma\gamma$, where the green lines are the two photons; and (b) Candidate for the decay Higgs $\to ZZ^*(ee\mu\mu)$, where here the green lines towards the center of the picture are the two electrons and the red lines in the center and at the upper right of the detector are the two muons.

In addition to discovering the new boson, CMS was able to begin the detailed study of its properties to show that it was indeed a Higgs boson. The decays of the new boson to the gauge bosons of the SM, the W, Z, and the photon, were established, each with more than 5



(a) (b)

Figure 1.2: The left hand side upper plot shows the $\gamma\gamma$ mass spectrum with a peak near 125 GeV. The left hand lower plot shows this distribution with the background subtracted. The right hand plot show the mass spectrum of four leptions. The three peaks are, in order of increasing mass, the decay of the Z boson, the Higgs bosons decaying into $Z(l^+l^-)Z^*(l^{'+}l^{'-})$ and di-boson production of two Z(l^+l^-), where l and $l^{'}$ are either a muon or electron

.

standard deviation significance. By using a combination of theory predictions for the decays and production, the couplings of the new boson to these particles have been determined and are shown in Figure 1.3 to follow the mass dependence uniquely characteristic of the Higgs field. Moreover, searches for the decay to fermions, $\tau^+\tau^-$ and $b\bar{b}$, are hovering at the edge of significance and their corresponding couplings are consistent with Standard Model (SM) expectations for the Higgs. Studies of the properties of Higgs decay have provided compelling evidence that the spin and parity of the new boson are indeed 0^+. Since the Higgs could decay to low mass particles that have not been observed, the "invisible" width of the Higgs has been studied and limits have been placed on it. The coupling to the top quark through the $t\bar{t}H$ process, appears to be within reach. A remarkable example for a new analysis technique is the measurement of the total width of the Higgs boson. Using the properties of off-shell Higgs production at masses of a few hundred GeV, CMS was recently able to constrain the Higgs boson width to 5.4 times the expected value in the standard model [3] of 4.1 MeV, a 200 times more stringent constraint than that reached in previous "direct" measurements.

The SM does not provide answers to the remaining questions. Those require new physics. In fact, the scalar nature of the particle, presents theoretical challenges. Radiative corrections to the Higgs should cause the mass to increase to very high values. New physics must appear at masses not too far from 1 TeV to cancel this growth. Although the 125 GeV Higgs behaves like a SM Higgs, measurement of its properties are still not very precise. Deviations from perfect SM behavior because of its interaction with other forms of matter, including dark matter, could answer some very fundamental questions, such as the origin of the matter-antimatter asymmetry of the universe. The detailed study of the 125 GeV Higgs is a scientific imperative that must be pursued to a much higher level of statistical precision than is available today.

Many searches have been undertaken with the data taken in 2011 and 2012 but they have not

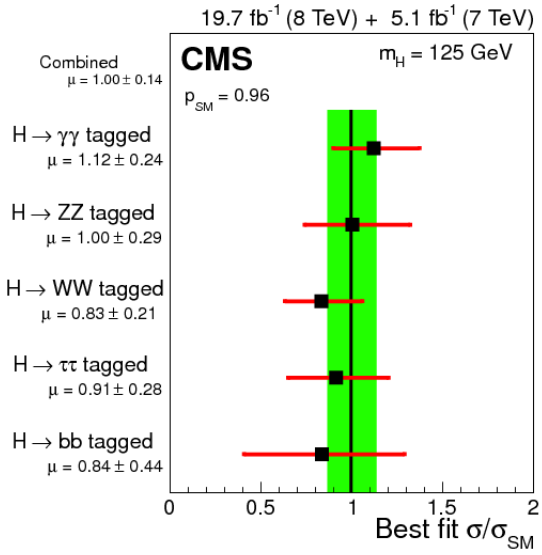


Figure 1.3: Higgs couplings divided by SM prediction from the full dataset from the 2011 and 2012 runs

yet revealed evidence of "beyond the standard model (BSM)" physics. The theory known as supersymmetry (SUSY) contains a partner for every SM particle but with spin that is less by 1/2 unit. SUSY is known to be a "broken symmetry" because SUSY partners are not observed yet and so, if they exist, would have to have higher masses than have been accessible so far at particle accelerators. These particles, if the lightest ones are not too much heavier than 1 TeV, can cancel the growth of the Higgs mass from radiative corrections. SUSY, under certain assumptions, can also provide a dark matter particle. The lightest SUSY particle will not interact with ordinary matter and will also result in events with large missing E_T, which is taken as one of the main experimental signatures of SUSY. The rate of production and the characteristics of decays of SUSY particles depend on their mass spectrum, which is not predicted, so the search has to investigate many possibilities. Figure 1.4 shows the cross sections for various types of supersymmetric particles from MSSM as a function of their masses. So far, SUSY has not been seen at the LHC. In simplified models of supersymmetry, the SUSY partners of the gluons, the "gluinos", and quarks, the "squarks", with masses below about 1 TeV are excluded, while scenarios with 3rd generation squarks, the sbottoms and the stops, with masses below 1 TeV are still compatible with the data. SUSY also predicts several more Higgs-type particles. Searches for these have also been undertaken but so far no additional Higgs bosons have been found. Figure 1.5 shows the exclusion limits for "stop" production. A summary of many SUSY limits is shown in Fig. 1.6.

Many other proposals for new physics exist, which try to address at least some of questions posed at the beginning of this chapter. Searches have been carried out with the data from 2011 and 2012 for many of them. For example, new gauge bosons with standard model couplings are also excluded below 2.5 TeV.

Another approach to discovering new physics is to make precision measurements of rare decays that are well-predicted in the SM. If new physics is present it might either enhance or suppresses the decay rate. An example of such a precision measurement made in Run 1 is the observation of the very rare decay $B_s \to \mu^+\mu^-$, whose signal is shown in Fig. 1.7. This decay is very highly suppressed in the SM but it can receive additional contributions from new

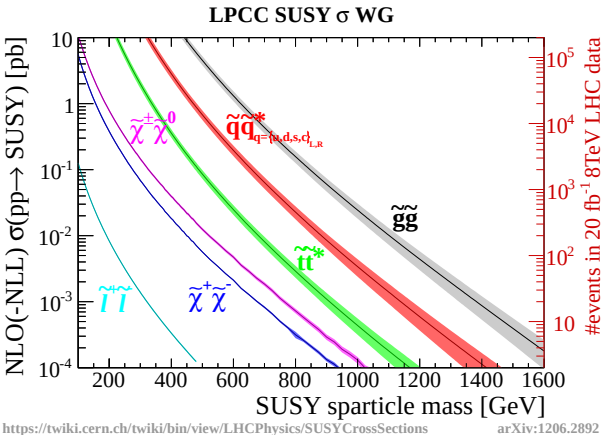


Figure 1.4: Cross section for SUSY particles as a function of their masses

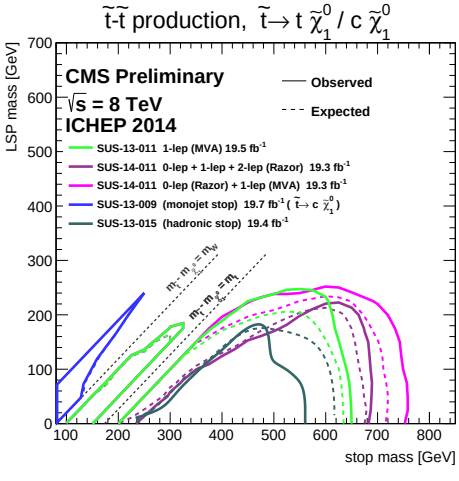


Figure 1.5: Mass limits for "stop", the Supersymmetric partner of the top quark

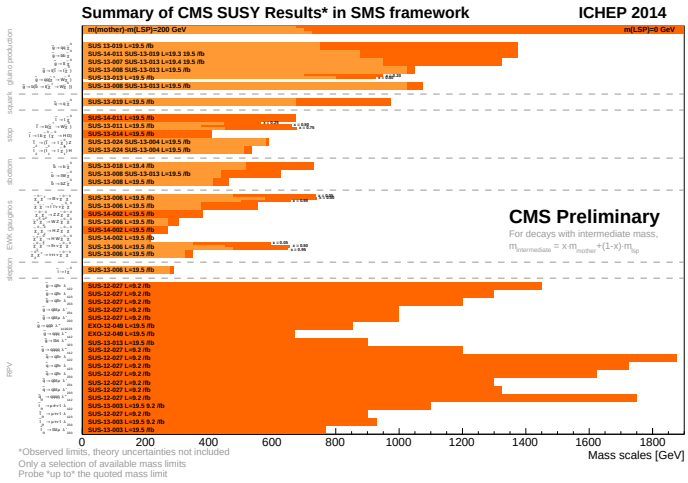


Figure 1.6: SUSY mass limits from the full dataset of the 2011 and 2012 runs

physics. The observed branching fraction is $3.0^{+1.0}_{-0.9} \times 10^{-9}$, consistent with the expectations of the SM. This result places very strict constraints on models of new physics. The companion decay, $B_d \to \mu^+\mu^-$, is predicted to have a branching fraction that is a factor of 20 lower than the B_s. Improving the precision of the B_s measurement and observing the B_d decay and measuring its branching fraction are important goals for the HL-LHC.

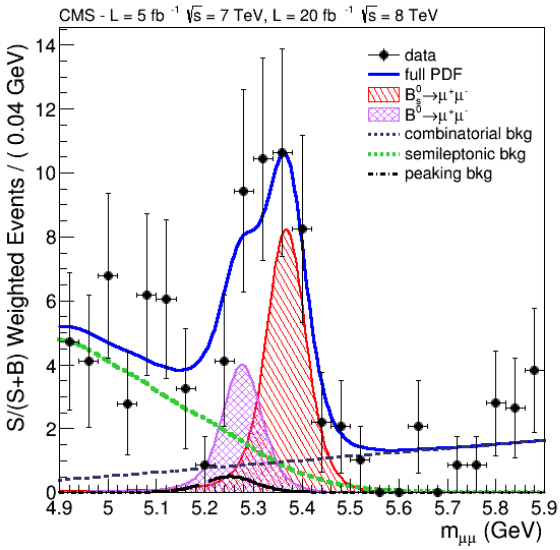


Figure 1.7: The mass distribution for $\mu^+\mu^-$ particles decaying downstream of the primary vertex. The purple and red curves show the B^0 and B_s signals, respectively, while the dashed line, and the green and black shapes show three different types of background. The solid curve shows the sum of all of the fit components.

Direct searches for exotic processes is another approach to discovering new physics. An example is the direct search for dark matter in states with a single jet from initial state radiation followed by the production of two weakly interacting massive particles (WIMPs), characterized by large missing E_T. While no dark matter particles are observed, the sensitivity of CMS in certain ranges of the mass of the dark matter particles and the cross section exceeds that of many other search techniques. The limits set by this analysis are shown in Fig. 1.8 and compared with dedicated "direct" searches for dark matter.

While no new physics has yet been seen at the LHC, the results so far have dramatically changed the landscape for these searches. Some theories and models have been largely eliminated; others have had their parameters sharply limited; and, for others, search strategies have been refined to be more sensitive to hard-to-identify configurations. For example, attention is being paid to "stealth" SUSY, in which new particles look very similar to SM backgrounds and to "compressed" SUSY, where particles have very similar masses making some decay configurations very hard to observe. Generic approaches to the searches have been developed that are somewhat insensitive to the details of specific production mechanisms or decay patterns. The sensitivity of CMS for observing SUSY and many other new physics signals improves with increasing luminosity and is an important goal for the HL-LHC.

Precision Higgs studies and the search for new physics and its study, when found, provide a powerful demand for higher luminosity. The LHC machine group now has a plan for achieving higher peak and integrated luminosity, well above those for which CMS was designed. The CMS detector requires upgrades to preserve the efficiency, resolution, and background rejec-

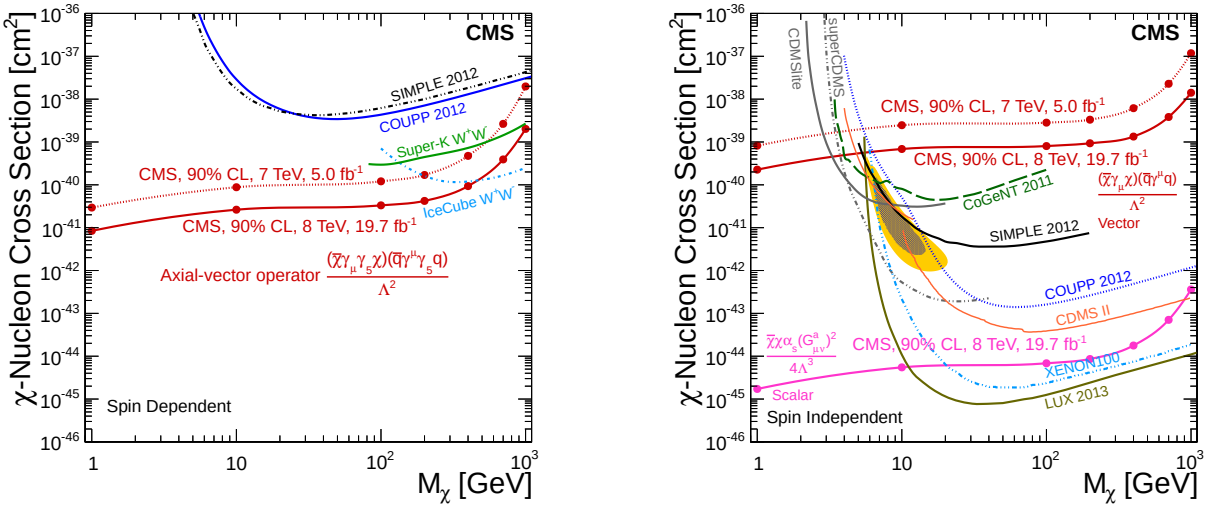


Figure 1.8: left-hand plot: Limits for spin-dependent detection of dark matter; and right-hand plot: Limits for spin-independent detection of dark matter

tion of the detector at these high luminosities. The physics case and conceptual design of an upgraded detector are the subject of this Technical Proposal.

1.2 The Luminosity Plans and Goals of the HL-LHC

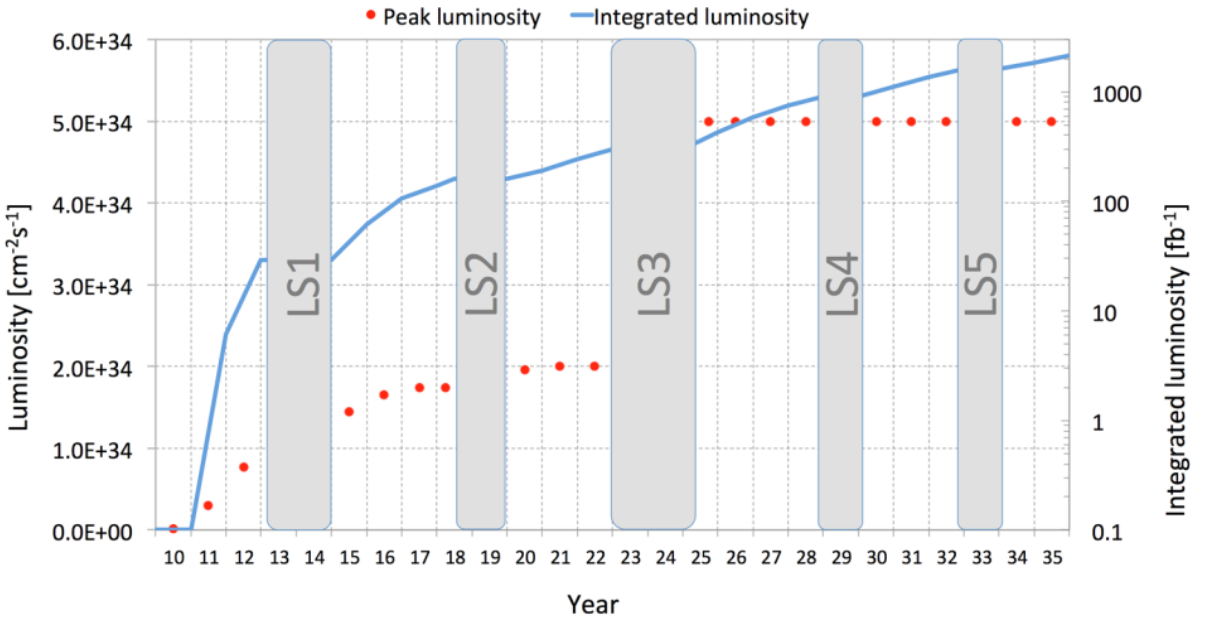


Figure 1.9: Projected LHC performance through 2035, showing preliminary dates for long shutdowns of LHC and projected luminosities.

CERN began planning to increase the luminosity of the LHC above the original design even before the machine went into operation. Major revisions to the machine or the experiments require access to the accelerator tunnels and the experimental areas that can only be accomplished efficiently during long shutdown periods. The current plan calls for a series of long

periods of data-taking, referred to as Run-I, Run-II, etc. interleaved with long shutdowns, designated LS1, LS2, LS3.

Run-I is the name given to the completed data-taking period in 2011 and 2012. During the first long shutdown, LS1, which started in 2013 and ended at the beginning of 2015, modifications were made to the LHC to enable it to run at the center-of-mass energy of 13 TeV. The energy will be raised closer to the design energy of 14 TeV over time.

Run-II is just beginning. It is expected that the bunch spacing will be reduced to 25 ns from 50 ns. The original performance goal for the LHC, to operate at an instantaneous luminosity of $1 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ with 25 ns bunch spacing, is likely to be achieved relatively soon after the startup. Under these conditions, early in Run-II, CMS will experience an average of about 25 inelastic interactions per bunch crossing, referred to as event pileup (PU). This is the operating scenario for which the CMS experiment was designed.

A new scheme to form the bunch trains in the Proton Synchrotron (PS) will allow the luminosity to exceed the original design before the second long shutdown,LS2, planned for mid-2018-2019. In LS2, the injector chain will be further improved and upgraded to deliver very bright bunches (high intensity and low emittance). It is anticipated that the peak luminosity could reach $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ in this first phase of the LHC program, providing an integrated luminosity of over $300\,\text{fb}^{-1}$ by 2023. To maintain its present performance in this period, the CMS detector will undergo an initial series of staged upgrades in the period from LS1 through LS2. This program, known as the CMS Phase-I Upgrade, has been documented in a Phase-I Technical Proposal [4] and three Technical Design reports (TDRs) describing the upgrades of the Pixel detector, the Hadron Calorimeter and hardware Trigger [5–7].

By 2023, the quadrupoles that focus the beams at the ATLAS and CMS collision regions are expected to be close to the end of their lives due to radiation exposure. There will be another long shutdown, LS3, to replace them with new low-β quadrupole triplets. In addition, crab-cavities will be added to optimize the bunch overlap at the interaction region. These changes will produce a significant increase in the LHC luminosity.

The high luminosity period that follows LS3 with the upgraded LHC is referred to here as HL-LHC or Phase-II. The proposed operating scenario is to level the instantaneous luminosity at $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ from a potential peak value of $2 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$ at the beginning of fills, and to deliver $250\,\text{fb}^{-1}$ per year for a further 10 years of operation. Under these conditions the event PU will rise substantially to become a major challenge for the experiments, and the performance degradation due to integrated radiation dose will need to be addressed. This Technical Proposal presents the CMS upgrade program for Phase-II.

The schedule of beam operations and long shutdowns, together with projections of the peak and integrated luminosities, is shown in Fig. 1.9, and is, of course, subject to change.

1.3 The Physics Opportunities at the HL-LHC

The CMS physics program at the HL-LHC will build on the experience acquired, and the results obtained, in the first phase of the LHC operation, with more than $300\,\text{fb}^{-1}$ of accumulated data and will continue the quest to answer fundamental questions in particle physics, on one hand with precision measurements, incuding detailed study of any new phenomena discovered before LS3, and on the other hand by direct searches for new physics.

In this section, a few highlights of the accomplishments possible with the full $3\,\text{ab}^{-1}$ are pre-

sented. The results take into account the impact of radiation damage and pileup, as well as the improvement to the detector planned to preserve its capabilities. More complete and detailed projections are given in chapter 10, which also presents the assumptions that went into producing each result.

The study of the Higgs boson will continue to be central to the program. It will include precise measurements of the Higgs boson couplings, probing of its tensor structure, and the search for rare SM and BSM decays. The enormous dataset will give access to all the p-p production processes and decays of the Higgs boson. Figure 1.10 shows the current CMS results (left) and a projection for the measurement of Higgs boson couplings in a dataset of 3 ab^{-1} at 14 TeV center-of-mass energy (right) as a function of the boson or fermion masses [8, 9]. Compared to a precision of about 20% on Higgs boson couplings today, percent-level precision can be reached for most coupling measurements. The coupling to the second-generation fermions will be probed for the first time measuring the Higgs boson decay to two muons. Measurements of di-Higgs production with a cross section of about 40 fb will allow the study of the Higgs boson self coupling.

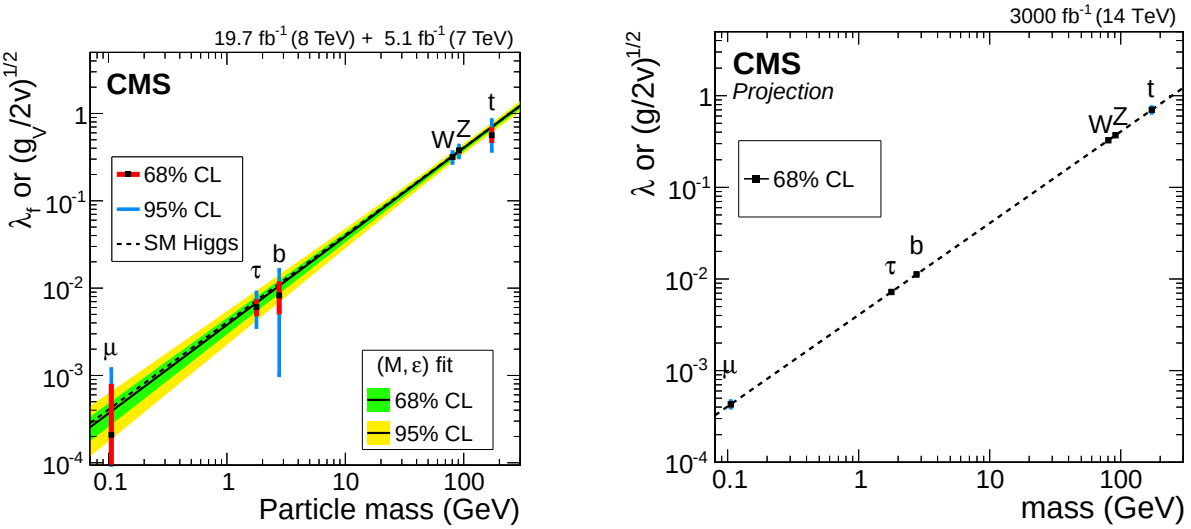


Figure 1.10: Observed and projected precision on Higgs boson couplings as a function of boson or fermion masses.

Figure 1.11 shows the predicted mass distribution for $H \rightarrow ZZ^*$ to four leptons for the full Phase-II luminosity with the upgraded CMS detector. This may be compared with the Run-I result shown in Figure 1.2 of this report, which has less than 1% of the integrated luminosity shown here.

Higgs boson coupling to charged leptons is a crucial measurement. The coupling to electrons is too small to measure, but the coupling to τ-leptons will be well-measured by the end of Phase-II. The couplings to the muons, whose branching fraction is only $\sim 10^{-4}$ will become accessible at the HL-LHC.

In addition, the role of the Higgs boson in the electroweak symmetry breaking will be tested in studies of the vector boson scattering processes. These measurements could also be sensitive to new physics through the triple-gauge couplings (TGCs) and quartic-gauge couplings (QGCs). Forward-jet tagging will be crucial to efficiently identify these processes. In general, precision measurements of electroweak observables have played a key role in validating the SM and in putting indirect constraints on BSM physics.

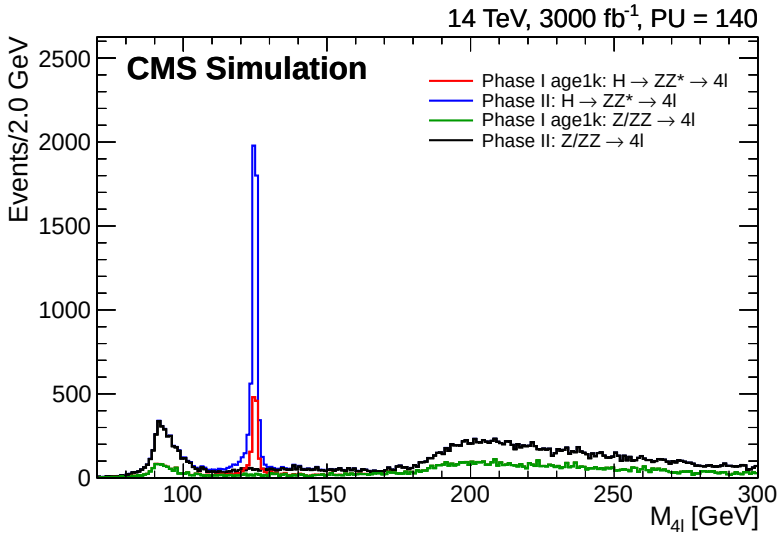


Figure 1.11: Four lepton mass distributions for the full Phase-II luminosity of 3000 fb^{-1} for the signal, $H \to ZZ \to 4\ell$, and for the irreducible $ZZ \to 4\ell$ background. Both processes have been simulated with the aged Phase-I detector with pileup of 140 and the Phase-II detector with pileup of 140.

New particles are expected at the TeV scale but have not yet been seen. This could mean that they exist at masses above the current level of sensitivity. It could also mean that they could be present at lower masses but their cross sections are lower than expected or their experimental signatures are more difficult to observe. In either case, sensitivity for searches of new particles grows with increased luminosity. Figure 1.12 shows the reach for selected searches for supersymmetry for a dataset of 300 fb^{-1} and 3 ab^{-1} at 14 TeV. Improvements in analysis and large samples of data will likely open the possibility for searches for even more rare signals with very low background that could result in our exceeding these projections.

The very high integrated luminosity collected at HL-LHC will extend the reach of the search for new heavy gauge bosons, to 6 TeV or more, for standard model couplings, or in case for very narrow width resonances, probe regions of 0.5-1 TeV. Similarly searches for extra dimensions, compositeness, leptoquarks etc. can be extended in range by a few TeV. Many signatures for exotica scenarios include the production of heavy semi-stable particles that will either traverse, or decay in, the CMS detector. For these searches it is imperative to keep the detector capabilities at the same level of quality as for the present data taking, if not to enhance these to increase sensitivity, in particular by use of precise timing.

The most prominent potential signal for new physics within the reach of the LHC is the evidence for dark matter, as has been reported by astrophysical and cosmic microwave background experiments. The origin of dark matter is one of the most intriguing enigmas of our time, and if it is caused by a particle, it is definitely not a member of our present standard model catalogue. One of the next priorities for the physics program at the LHC will therefore be the hunt for dark matter. The discovery of supersymmetric particles would likely be a big step forward in our understanding of dark matter. However, several specific searches for dark matter have been developed in recent years using for example mono-object (jet, photon, vector boson, top, etc.) signatures. These searches turn out to be competitive with the direct search experiments in the appropriate regions of comparison, and projections indicate that the high luminosity upgrade of the LHC can drive this search below the neutrino coherent scattering

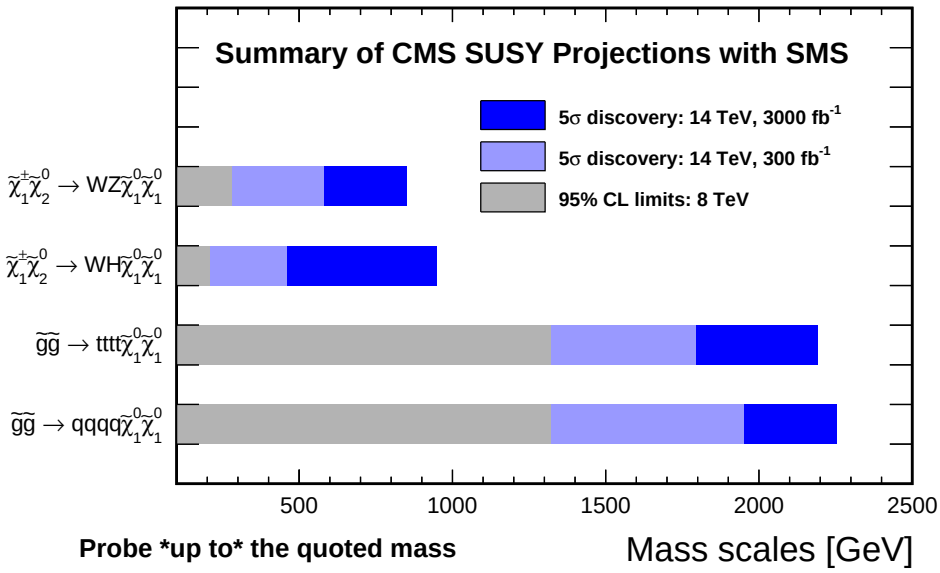


Figure 1.12: Mass reach of SUSY searches from selected 8 TeV results (masses excluded at 95% CL) and from projections for 14 TeV running at high luminosities (highest masses for 5σ observation). Simplified models are used for the interpretations in each case.

limit, which will be a concern for the direct experiments [10]. For this program to be successful it is essential that, in particular, the quality of the missing E_T measurement is kept at a similar level as for the present data.

The approach of observing new physics through the study of rare decays also benefits from the increased statistics. The decay $B_s \to \mu^+\mu^-$ will become a precision measurement and the decay $B_d \to \mu^+\mu^-$ will be established and will be measured with reasonable accuracy. The projected mass distribution is shown in Fig. 1.13. The significance of the B_d, predicted to be 2.2σ after 300 fb^{-1}, will improve to 6.8 σ with 3000 fb^{-1}.

A particularly exciting possibility is to use the Higgs boson as a search tool for dark matter at the LHC. The Higgs boson may well be a portal connecting the standard model with other new physics sectors, such as the dark sector. In that case, and if the dark matter particle is relatively light, the search for dark matter in the decay of Higgs particles, via the so called 'invisible decay' channel will be important. A new channel proposed to search for dark matter is mono-Higgs production[11], similar to the mono-jet signature, except that anobservable Higgs boson is emitted in the final state from the produced dark matter particles. For this channel the high luminosity of the HL-LHC will be essential.

In the event of a discovery during the first phase of the LHC, the large dataset of the HL-LHC will be critical to unveil the nature of the observed new particles. This will require precise measurement of their properties, such as production cross sections, masses, and spin-parity. It will also be essential to extend the searches of other new physics signals.

In addition and in support of these discovery topics, very many measurements of SM phenomena will be made at the HL-LHC. In addition to providing high statistics measurements that can provide insight into many models of these processes, they will also help define SM "backgrounds" that must be known and well-modelled to carry out he discovery portion of the program. For example, parton distribution functions (PDFs) of the proton are crucial in-

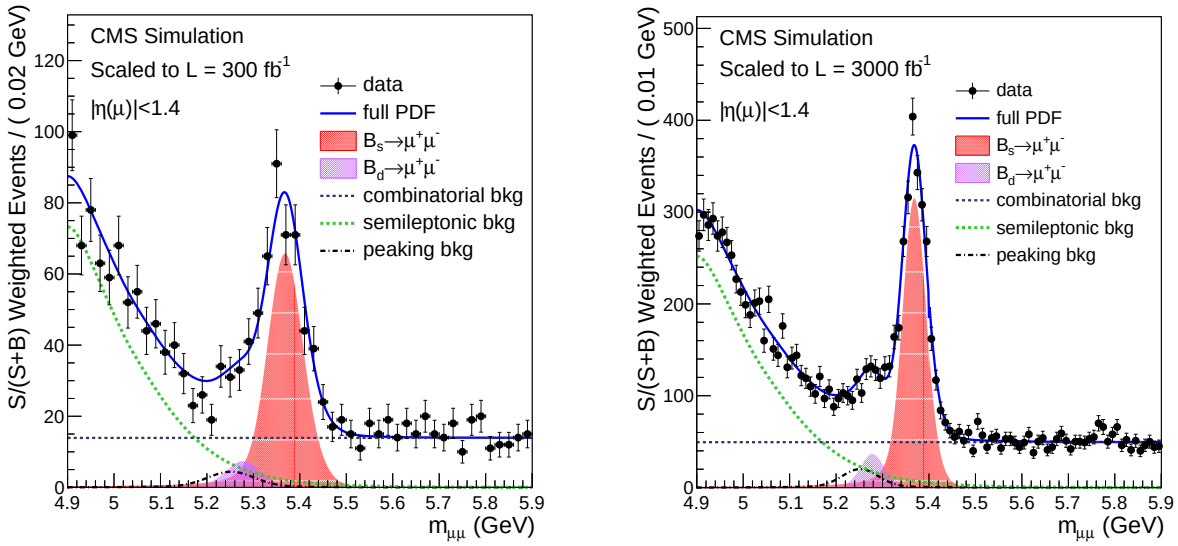


Figure 1.13: Projections of the mass fits for the $B_{s,d} \to \mu^+\mu^-$ measurement to 300 fb^{-1} (left) and 3000 fb^{-1} (right) of integrated luminosity, respectively assuming the expected performances of Phase-I and Phase-II CMS detector. Both plots are for barrel only.

gredients of measurements at the LHC. Future Higgs boson coupling measurements will be limited by their uncertainties unless significant progress is made. Other precision measurements, like the measurement of the W boson mass, the effective lepton mixing angle, and the strong coupling constant α_S, have large uncertainties from PDFs. If new physics phenomena are discovered, their characterization will also suffer from PDF uncertainties, e.g. for gluino or squark production in the few TeV range, uncertainties can be as large as 100% since they probe PDFs at very large values of x. Improvements are needed from experimental data, theoretical calculations, and methodological framework. With the high luminosity data, CMS will contribute to this program by precision measurements of inclusive, differential, and double-differential cross sections of events with jets, photons, W and Z bosons, and top quarks. This requires excellent trigger and pileup-mitigation capabilities. The charm and strange PDFs can be constrained by measurements of charm-tagged jets in events with an electroweak bosons. This will also require excellent vertex reconstruction capabilities. The search for new physics builds on our knowledge of SM physics.

1.4 Challenges of High Luminosity to the Physics Reach of CMS

The basic goal of the Phase-II upgrade is to maintain the excellent performance of the CMS detector in terms of efficiency, resolution, and background rejection for all the physics objects used in the analysis of the data. The main challenges that must be overcome to achieve this goal are radiation damage to the CMS detector from the high integrated luminosity of the HL-LHC and the very high "pileup" that comes from the high instantaneous luminosity. In this section, each of these challenges is described in general terms. Details of the problems for each subsystem of CMS are given in the subsequent chapters along with the proposed remedies.

1.4.1 Radiation Damage to the CMS Detector at the HL-LHC

The LHC will produce collisions at a rate of about 5×10^9/s. The particles emerging from these collisions and the radioactivity they induce in the material of the detectors and the on-board electronics will cause significant damage and could result in a progressive degradation of the detector performance. Maintaining the detector performance in the harsh conditions of the LHC was a major consideration in the initial design of CMS and of the Phase-I upgrade. When one considers that the annual dose delivered to the detector per year in the HL-LHC era will be similar to the total dose of all operations from the beginning of the LHC program to the start of LS3, the magnitude of the problem becomes clear.

The main source of radiation is from the particles produced in the proton-proton collisions. The charged particles, mainly pions, produce ionization in the detectors they pass through. They also undergo nuclear interactions that produce cascades of particles that add to the radiation load. Photons, mainly from π^0 decays, interact in the material of the beam pipe or tracking systems to form e^+e^- pairs or reach the calorimeters where they produce electromagnetic cascades. Particles are also backscattered from the calorimeters or escape from cascades within them. These particles spread out and interact with other detector components. Neutrons, in particular, may travel long distances, slowing down and scattering many times in the CMS detector, which is largely hermetic. When the neutrons interact, they can also produce photons and electrons. This results in a mixed field of very low energy neutrons, photons, and electrons that have lost any correlation with the bunch structure of the original collisions and form a relatively uniform background in space and time within the detector volume. Since personnel will need to work on and around the detector when the beam is off, simulations are also performed to predict material activation and dose to personnel.

In order to design a detector that will maintain its performance at the HL-LHC and continue to perform well as the integrated luminosity approaches 3ab^{-1}, predictions of the dose rate and particle fluence for each type of particle is needed. Simulations are used to predict the magnitude and composition of radiation as a function of luminosity. The radiation simulations are performed with Monte Carlo transport codes MARS'109 [12] and FLUKA 2011.2b.6 [13] [14]. A description of the codes, their functionality, the CMS geometry models and settings are given in Appendix A. The impact of each type of particle on the various candidates for material used in the detector must also be known. Information on the performance of the current detector is obtained from test beam measurements, special radiation exposures, and from Run-I experience. The beginnings of radiation damage are clearly observed in Run-I and the results are used to benchmark the simulations.

For most simulations, a p-p collision is used as the primary event, except for machine-induced background (MIB) simulations, which are described in Appendix A.1.7. The event generator DPMJET III [15] (version 3.0-6) is used to create the primary proton-proton events in radiation simulations. It is directly linked to the FLUKA code and used as the default event generator for high energy hadronic interactions. The pseudorapidity and transverse momentum distributions are shown in Fig. 1.14. While the multiplicity decreases with increasing $|\eta|$, the mean momentum increases rapidly.

The normalization used for the prediction of dose and fluence depends on the inelastic collision rate and the duration of LHC operation (total radiation period). For an instantaneous luminosity of $5\times10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, and an inelastic cross section of $\sim$80 mb, an average of 4×10^9 inelastic p-p events per second are produced[1]. For consistency with previous reports, cumu-

[1]The inelastic non-diffractive cross section of $\sqrt{s} = 14$ TeV protons is predicted by various event generators.

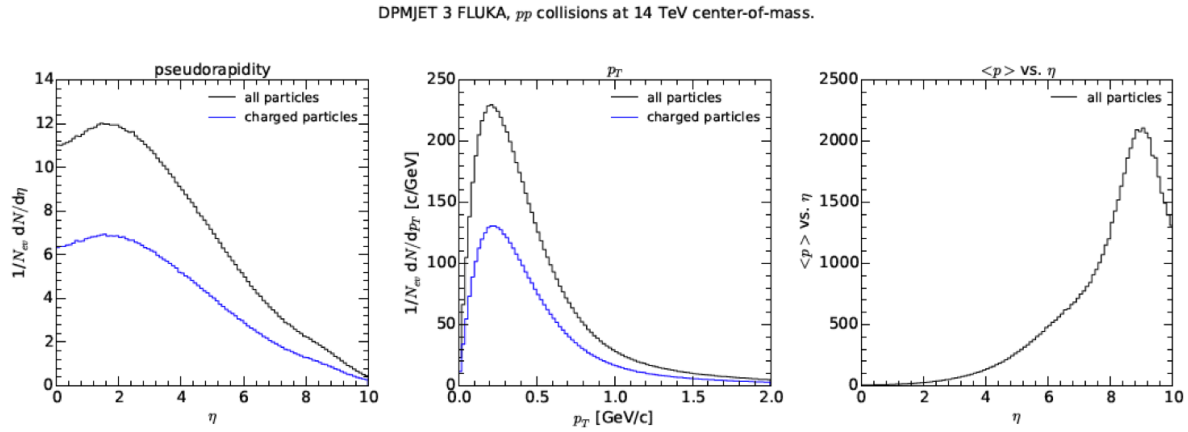


Figure 1.14: Pseudorapidity and transverse momentum distributions of the particles generated by the DPMJET III generator. The black line is for all particles, the blue line is for charged particles.

lative estimates of particle fluences and dose are based on a previous LHC schedule, where it was assumed that the CMS experiment will receive 3000 fb^{-1} by 2035. However, estimates of the residual activity, where the time structure of LHC operation is significant, are based on a pp collision time profile with a total of 2827.5 fb^{-1} until 2035 [17].

Two main FLUKA CMS geometries are used to perform simulations for HL-LHC conditions; the "TP Baseline" geometry, derived from the latest nominal Phase-I CMS geometry [18] with the presently installed electromagnetic and hadron endcap calorimeters; and a modified geometry that includes specific elements to represent the Phase-II detector upgrades. A full description of both geometries, with corresponding diagrams can be found in Appendix A.

An example of the predictions of expected radiation levels for HL-LHC conditions in the CMS detector is given in Figure 1.15, which shows the distribution of absorbed dose over the CMS detector for an integrated luminosity of 3000 fb^{-1}.

The damage produced in the detectors by this radiation varies from sub-detector to subdetector. For silicon detectors, radiation produces defects in the silicon lattice that change the bulk electrical properties of the silicon. One consequence is that leakage currents will increase. Trapping centers for charge carriers reduce the Charge Collection Efficiency (CCE), which in turn lowers the signals from charged particles. The full depletion voltage (the voltage required to make the full thickness of the detector depleted from charges, hence making it fully sensitive to a crossing ionizing particle) increases, eventually reaching unsustainable levels and forcing operation with only partial depletion, resulting in lower signals. At the HL-LHC, some detectors will be subjected to fluences as high as 10^{16} particles/cm^2, which will make efficient operation difficult. For calorimeters, which in CMS are mainly scintillating $PbWO_4$ crystals or plastic scintillating tiles with wavelength-shifting fibers embedded in them, the main problem is the loss of transmission of the media through which the scintillation light or wavelength-shifted light must pass. The actual scintillation method does not appear to be harmed by the radiation. This results in a reduction in the signal that can be quite large, in some cases more than 90%, and a corresponding reduction in the resolution. Strategies for mitigating the effects of radiation vary from sub-detector to sub-detector and are described in subsequent chapters.

Being the only event generator tuned with LHC data, the results of EPOSLHC is used, which predicts 80 mb inelastic cross section [16].

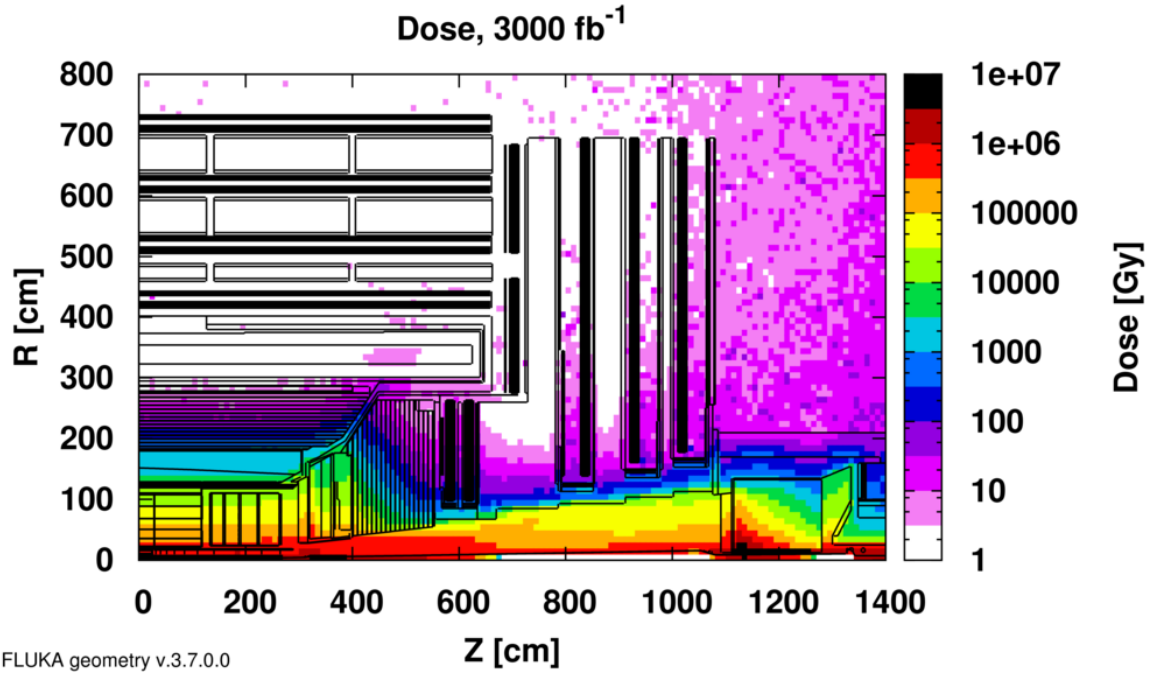


Figure 1.15: Absorbed dose in the CMS cavern after an integrated luminosity of 3000 fb^{-1}. R is the transverse distance from the beamline and Z is the distance along the beamline from the Interaction Point at Z=0.

1.4.2 High Pileup

Each of the colliding beams at the LHC consists of many intense "bunches" of protons. Each bunch has a length with rms of ∼5 cm, transverse dimensions of about 10 microns, and contains a few$\times 10^{11}$ protons. Bunches will be separated in time by 25ns, corresponding to a spatial separation of approximately 750 cm. There are ∼2800 filled bunches in each beam and this number cannot be substantially increased. The collision of two bunches is called a "bunch crossing" or "BX" and these occur at a rate of 40MHz. At the nominal luminosity of the HL-LHC, the average number of interactions in a single crossing is approximately 140. Most of these interactions are "soft" or "peripheral" collisions that, if not well understood, are at least well-characterized and do not contribute to the search for new physics at the 0.1-few TeV scale. As shown clearly in Fig. 1.14, they contain low P_T particles and produce little energy in the CMS detector. A relatively small fraction of all collisions are "hard" collisions that contain high transverse momentum particles that may come from new high mass objects. Nevertheless, the presence of some tracks and energy from 140 (on average) extra collisions can confuse or degrade the triggers and the offline reconstruction of the hard scatter.

There are, in fact, three kinds of pileups based on the time at which energy is deposited in the tracking detectors to form hits and in the calorimeters to form "showers". In-time pileup (IT) refers to "hits" or energy deposits from the "extra" p-p collisions in the current bunch-crossing other than those from the collision containing the hard scatter of interest. This is the largest source of hits in the tracking system and also produces significant energy deposition in the calorimeters. In addition, there is out-of-time pileup (OOT), which comes in two varieties: early out-of-time pileup, which refers to energy left in calorimeters from previous BXs in the crossing of interest; and late out-of-time pileup, which refers to energy from later BXs that is integrated along with the trailing portion of the pulse from the BX of interest.

Since the number of bunches cannot be increased, luminosity increases at the LHC result in higher pileup. Pileup produces many more hits in the tracking detectors, leading to mismeasured or misidentified tracks. It also adds extra energy to the calorimeter measurements, such as jet energies, associated with the collision that contained a hard scatter. Electroweak phenomena, which are of special interest, are often characterized by having "isolated" leptons, that is leptons or photons with very little activity around them. Energy or tracks from pileup can contribute to an activity that is not due to the collision containing the leptons or photons and cause them to appear non-isolated. Pileup confuses the trigger and also the offline reconstruction and interpretation of events. It increases the amount of data that has to be read out in each BX that contains a hard scatter. In fact, at the HL-LHC, most of the data read out will be associated with the "pile-up" collisions rather than the collision containing hard scatters. It also increases the execution time for the reconstruction of events in the High Level Trigger and the offline analysis.

In-time pileup can be observed in a single bunch-crossing by the many collision vertices that are reconstructed by the tracking system. A relatively high-pileup crossing that was produced in a special data run in 2012 is shown in Fig. 1.16. There are 78 reconstructed vertices. The total number of pileup collisions is actually somewhat larger because some vertices have too few tracks to be reconstructed. The upgraded tracking system can be designed with enough additional segmentation to associate charged particles with the correct interaction vertices most of the time, even for PU of 140 or 200. This enables the collision containing the hard scatter to be correctly reconstructed and for isolated leptons to be correctly identified in most cases.

The calorimeters in CMS do not have "pointing capability" so it is not possible to associate showers in them with particular vertices. However, the particle flow techniques that are now employed will associate charged tracks in the shower with a particular vertex and this helps to arrive at the correct interpretation of events even in the presence of very high pileup. Moreover, the many simultaneous "typical" or "minimum bias" collisions in the bunch crossing produce a rather smooth energy distribution that can be extrapolated into jet cones or isolation cones to enable a subtraction of their effect.

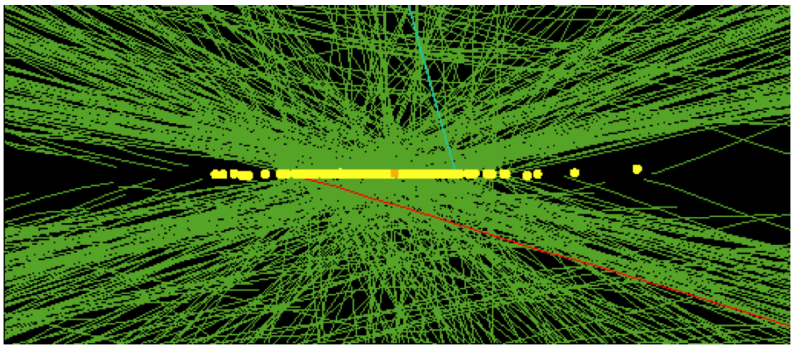


Figure 1.16: High pileup event with 78 reconstructed vertices taken in 2012

Out-of-time pileup is illustrated in Fig 1.17. The degree of OOT depends on the intrinsic time spread and jitter of the pulses produced in each detector by particles passing through it and by shaping times and other characteristics of the readout electronics. Tracking systems typically respond in times short compared to the inter-bunch spacing of 25 ns and are not very sensitive to OOT. Calorimeters may produce longer signals and may need longer shaping times and so may suffer from problems with OOT. If the detectors and their readout electronics provide timing or pulse shape information, it is possible to use it to correct the energy deposition associated with a bunch crossing for the energy leakage into that crossing from OOT. Timing

measurements are likely to play a more significant role in the HL-LHC era than they have in Run-I, which had 50 ns bunch spacing, in order to cope with out-of-time pileup.

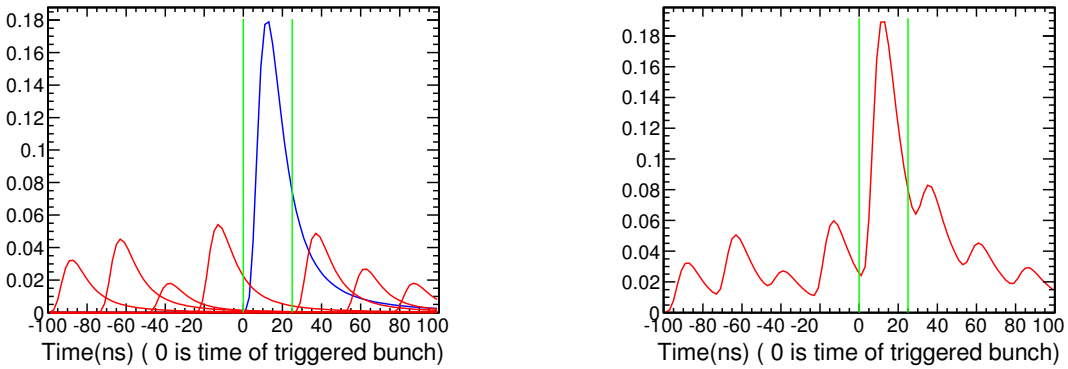


Figure 1.17: Illustration of out-of-time pileup: (left) Individual pulses from a detector as a function of time in ns. The "triggered" or "signal" BX is located at "0" and is bounded by the green vertical lines. The "signal" pulse appears in blue and extends into the next two 25ns "buckets", which contain some pulse height from the later pileup collisions. Some energy from the preceding "bucket" also falls into the "signal" BX. (right) The sum of the pulse heights as a function of time. The challenge is to correct the "summed" pulse height or charge (related to area under the pulse height distribution) to derive the pulse height or the total charge in the blue signal pulse.

1.5 A Brief Overview of the CMS Phase-II Upgrade

1.5.1 A brief introduction to the CMS detector

An exploded view of the CMS detector is shown in Figure 1.18. At the heart of the experiment is a 13 m long, 6 m diameter, 4T superconducting solenoid providing large bending power for momentum measurements and whose return field is large enough to saturate the iron plates in the return yoke, enabling it to be used for muon momentum reconstruction. The gaps between the plates provide slots for four muon tracking stations, each of which consists of several layers of aluminum drift tubes (DT) in the barrel region and cathode strip chambers (CSCs) in the endcap region. Each system is complemented by resistive plate chambers (RPCs).

The bore of the magnet is large enough to accommodate the tracking and calorimetry systems. The tracking volume is contained in a cylinder of 5.8 m length and 2.6 m in diameter. CMS employs ten layers of silicon microstrip detectors, which provide the required granularity and precision to reconstruct efficiently charged tracks in high multiplicity events. The silicon microstrip tracker with its long bending path, combined with the strong solenoidal field, provides excellent momentum resolution. In addition three layers of silicon pixel detectors in the barrel region, complemented by two forward disks at each end, seed track reconstruction and improve impact parameter measurements, as well as providing points with sufficient resolution to reconstruct secondary vertices from decays of particles containing b and c quarks.

The electromagnetic calorimeter (ECAL) provides coverage up to pseudorapidity $|\eta| = 3$ and uses blocks of lead tungstate crystals whose scintillation light is detected by silicon avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. A preshower

system is installed in front of the endcap ECAL for π^0 rejection. The ECAL is surrounded by a brass/scintillator sampling hadron calorimeter (HCAL) with coverage up to $|\eta| = 3$. The light is converted by wavelength shifting (WLS) fibres embedded in the scintillator tiles and channeled via clear fibres for readout to hybrid photodiodes (HPDs) that can operate in high axial magnetic fields. This central calorimetry is complemented by a "tail-catcher (HO)" in the barrel region insuring that hadronic showers are sampled with nearly eleven interaction lengths. Coverage up to $|\eta| = 5$ is provided by an iron/quartz-fibre calorimeter (HF). The Cherenkov light emitted in the quartz fibres is detected by photomultipliers. The HF ensures nearly full geometric coverage for measurement of the transverse energy in the event.

CMS is triggered by dedicated custom electronics in the Underground Control Room (USC55) which form various partial triggers using trigger primitives from the front ends of the calorimeters and muon detectors. These are then sent to the Global Level 1 trigger which is designed to handle up to 100 kHz rate; the latency is 3.6 μs. Data must be stored on detectors during Level 1 processing. When a Level 1 accept occurs, data fragments from individual detectors are sent to the High Level Trigger (HLT), operating on a large computer cluster, to build complete events. The HLT performs a lean version of the offline reconstruction using full event data and uses the result to decide if the event should be written, together with trigger information, to mass storage for subsequent analysis. In Run-I CMS wrote out about 500-1000 events/second. A detailed description of the CMS detector is given in reference [19].

Since the LHC luminosity will exceed the orignal luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ for which CMS was designed, there is an ongoing upgrade that will allow CMS to perform well until the start of the HL-LHC. Modifications to the detector described above include the completion of the fourth layer of endcap muon detectors, which was only partially implemented in Run-I; improvement in the muon electronics; the replacement of the 3 barrel layer, 2 endcap disk pixel detector with a 4 barrel-layer, 3 endcap disk pixel detector; the provision of longitudinal segmentation in the barrel and endcap hadron calorimeters; replacement of the HPDs for the barrel and endcap hadron calorimeters with Silicon Photomultipliers (SiPMs); the replacement of single anode photomultipliers in the HF with multianode photomultipliers; and the modernization and upgrade of the trigger and data acquisition systems to handle higher data volumes.

1.5.2 General Considerations for the Phase-II upgrade

For the HL-LHC, the brightness of beams and the new focusing/crossing scheme at the interaction point will enable the accelerator to potentially deliver a luminosity of $2 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$ at the beginning of each fill. This would increase the interaction rate and collision PU beyond the capabilities of existing and envisioned detector and trigger technologies. It is therefore proposed to maintain a lower, but stable instantaneous luminosity by continuously tuning the beam focus and crossing profile throughout the duration of beam fills in a process referred to as luminosity levelling. The nominal scenario is to operate at a leveled luminosity of $5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$, corresponding to a mean pileup of 140 interactions per beam crossing [2]. The primary goal of the Phase-II upgrade program is therefore to maintain the excellent performance of the Phase-I detector under these challenging conditions throughout the extended operation of HL-LHC.

A major focus of CMS has been to identify changes that are mandatory for the beam conditions of HL-LHC and significant effort has been expended to understand the effect of radiation damage. Performance projections are based on a combination of detailed measurements using the

[2] An average of 140 p-p collisions per beam bunch crossing assumes a high value of the total interaction cross section extrapolation to 14 TeV, and takes into account bunch to bunch fluctuations as discussed in ref [20]

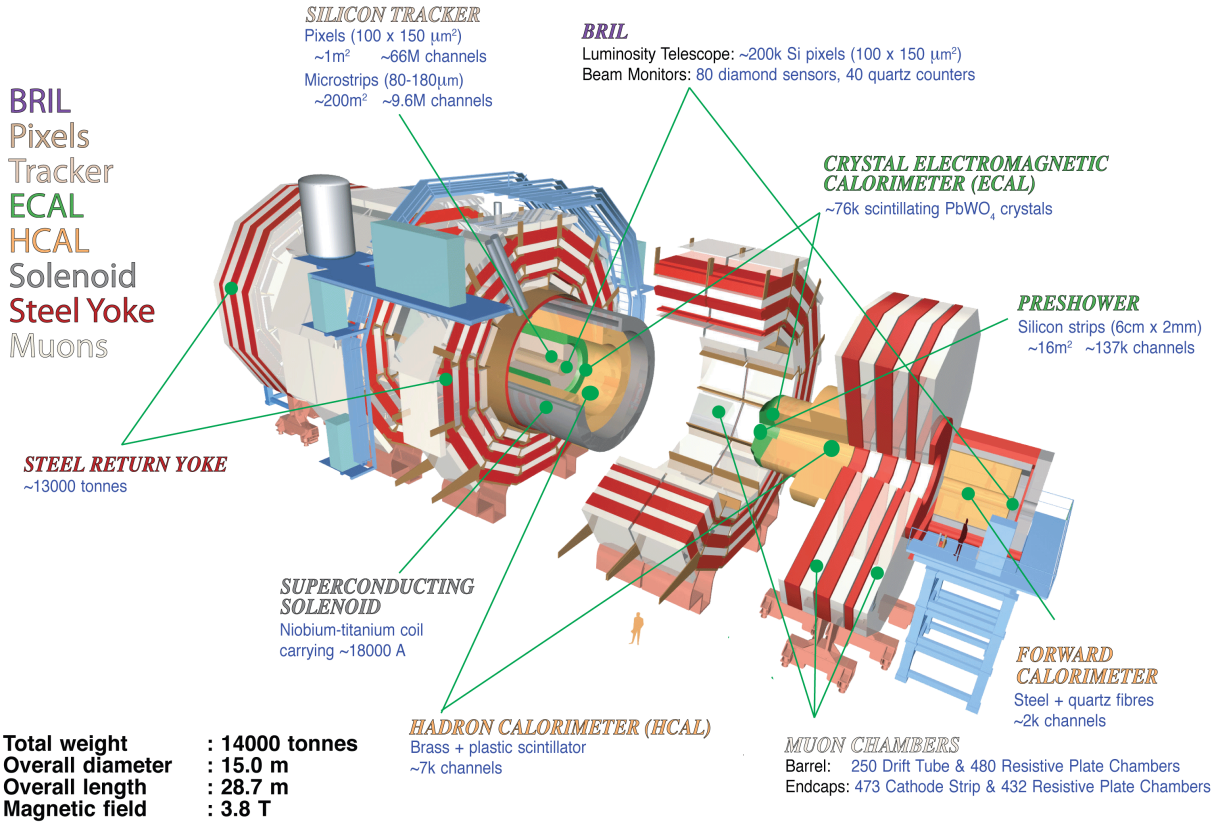


Figure 1.18: A schematic representation of the CMS Detector in Run 1, with its various sections in retracted positions. The central yoke block is YB0. The first block (YB+1, and corresponding YB-1) is shown partially moved away from YB0. The second yoke block (YB+2, which has a corresponding YB-2) is shown fully moved past the solenoid vacuum tank. The endcap calorimeters are attached to the first endcap disk YE+1, then endcap CSCs and RPCs, then YE+2, more muon chambers, and YE+3, with additional muon chambers on the front and back. Eventually, another disk, YE+4 will be added at the end to provide shielding from beam related backgrounds. This configuration is repeated on the other end. In operation, the detector is closed by moving all pieces together.

data taken in the experiment throughout the period 2011-2012 and the exposure of test components to radiation levels matching anticipated HL-LHC doses. From these studies it is very clear that the tracker and the endcap calorimeters must be replaced for Phase-II.

With these required changes, the performance issues associated with high PU, that are also the most pronounced in the inner and forward detector regions, can be addressed. Pile-up mitigation in CMS heavily relies upon particle-flow event reconstruction. To this end, the tracker granularity can be increased to maintain the excellent tracking efficiency to enable the determination of the original p-p collision points for all charged particles. New endcap calorimeter configurations will also provide the opportunity to optimize segmentation and improve energy resolution, particularly for jets.

The ability to ensure efficient event selection for data acquisition is a key prerequisite to fully benefit from increased luminosity. The precise study of the relatively low-mass Higgs boson discovered in 2012, and the search for new particles occurring in cascade decays will require continued use of low transverse momentum, p_{T}, trigger thresholds. To achieve this, the trigger electronics (i.e. the L1 trigger) must be upgraded. A sufficient reduction in trigger rate can

only be accomplished by improving p_T resolution to obtain lower rates without loss of efficiency, and by mitigating the effect of the combinatorial backgrounds arising from PU. A new approach is therefore required, namely the introduction of tracking information at L1, providing the capability to implement trigger algorithms similar to that of the current HLT, including the use of precise momentum measurements. Facilitating tracking in the L1 trigger is an important driver of the design of the Phase-II Tracker. The upgraded L1 "track trigger" will require a new hardware architecture to incorporate the tracking information. While the addition of track information in the L1 trigger provides significant gains in rate reduction with good efficiency, it will nevertheless be necessary to increase the trigger accept rate in order to maintain the required efficiency for all of the important physics channels. This is particularly the case for triggers involving hadrons and photons, for which the sensitivity to PU is higher and the track trigger is somewhat less efficient.

The measurement of processes with small production cross-sections and/or decay branching ratios is a major goal of the HL-LHC physics program. This requires specific upgrades in the forward regions of the detector to maximize the physics acceptance over the largest solid angle possible. To ensure proper trigger performance within the present coverage, the muon system will be completed with new chambers. The new endcap calorimeter configuration offers the opportunity to extend the muon coverage with a tagging station up to $|\eta| \approx 3$ or more, with significant acceptance gain for multi-muon final states. To mitigate PU effects in jet identification and energy measurement, the tracker will be extended up to $|\eta| \approx 4$, thereby also covering the peak production region of jets accompanying Vector Boson Fusion (VBF) and Vector Boson Scattering (VBS) processes, which are among the highest priorities of the physics program. With this extension, measurements of total energy and missing energy will be greatly improved, and b-tagging acceptance will be increased.

As the luminosity integrated over the Phase-II operation period will not be limited by the accelerator performance but by the ability of the detector to sustain high PU, the upgrades of the readout electronics will be designed with some margin to allow efficient data taking up to a PU of 200. This will also provide some flexibility for the luminosity levelling process to use instantaneous luminosity and corresponding beam lifetime information to maximize the useful integrated luminosity in order to obtain the best possible physics performance of the experiment. It is expected that the sustainable luminosity limit will be driven by the performance of sub-detectors that are not going to be replaced for Phase-II. Further simulation studies will include possible alternatives to the luminosity levelling schemes and PU beyond 140 to further optimize designs and to determine the best operating scenarios. Among the studies foreseen, CMS will investigate whether or not precise measurements of the production time of particles, which vary with a rms of $\sim$200 ps within a single bunch crossing, would enable a valuable improvement in PU mitigation, particularly for the contributions of neutral particles, which are not detected in the tracker.

1.5.3 Elements of the Phase-II Upgrades

Tracker The Tracker will suffer significant radiation damage by LS3 and must be completely replaced for Phase-II. To maintain adequate track reconstruction performance at the much higher PU levels of the HL-LHC, the granularity of both the outer tracker and the pixel systems will be increased by roughly a factor 4. In the outer tracker, this will be achieved by shortening the lengths of silicon sensor strips relative to those in the current detector, without changing the pitch very significantly. A number of design improvements will lead to a much lighter Outer Tracker providing significantly improved p_T resolution and a lower rate of γ-conversions compared to the present detector. In addition, the module design will be capable of providing

track-stub information to the L1 trigger at 40 MHz for tracks with $p_T \geq 2\,\text{GeV}$. This will ensure powerful background rejection at the earliest stage of the event selection. The pixel system will implement smaller pixels and thinner sensors for improved impact parameter resolution and better two-track separation. This will improve b-tagging as well as τ-hadronic decay and track reconstruction efficiencies within boosted jets. With up to 10 additional pixel disks in each of the forward regions the system coverage will be extended to close to $|\eta| = 4$, to better match the range of coverage of the calorimetry.

Calorimeter endcaps The electromagnetic and hadronic endcap calorimeters will also suffer significant radiation damage by LS3, and so must be replaced.

The replacement is called the High Granularity Calorimeter (HGC) and has electromagnetic and hadronic sections with excellent transverse and longitudinal segmentation. It will provide detailed three dimensional images of showers. The electromagnetic section consists of $\sim$30 tungsten and copper plates interleaved with silicon sensors as the active material. The sensors have pads of variable sizes of less than $\sim$ 1.0cm^2. The electromagnetic section has $25X_0$ and one interaction length (λ). The hadronic part has a front section of 12 brass and copper plates interleaved with silicon sensors for a depth of 3.5λ. This covers the hadronic shower maximum measurement. It is followed by a "backing hadron calorimeter" of similar design to the current HE detector, brass plates interleaved with plastic scintillating tiles read out with a wavelength shifting fiber, to provide an overall depth of $\sim 10\lambda$ for the full calorimeter. The design of the High Granularity Calorimeter draws upon the ILC/CALICE[21] concepts for 3D measurement of shower topologies.

Muon endcaps The muon system in the region $1.5 \leq |\eta| \leq 2.4$ currently consists of four stations of Cathode Strip Chambers (CSC). It is the only region of the muon detector that lacks redundant coverage despite the fact that it is a challenging region for muons in terms of backgrounds and momentum resolution. To maintain good L1 muon trigger acceptance in this region it is therefore proposed to enhance these four stations with additional chambers that make use of new detector technologies with higher rate capability, along the lines of what was planned in the original design of CMS. The two first stations are in a region where the magnetic field is still reasonably high and so will use Gas Electron Multiplier (GEM) chambers for good position resolution in order to improve momentum resolution for the standalone muon trigger and to improve the matching with tracks in the global muon trigger. The two last stations will use low-resistivity Resistive Plate Chambers (RPC) with lower granularity but good timing resolution to mitigate background effects. In addition, the implementation of a GEM station in the space that becomes free behind the new endcap calorimeters is being proposed in order to increase the coverage for muon detection to $|\eta| \approx 3$.

Beam radiation protection and luminosity measurement The systems that provide protection against beam background and measurement of the luminosity will require work in several areas to manage the high radiation levels of the HL-LHC. The protection systems will be upgraded with new poly-crystalline diamond sensors that will be read out using the standard LHC Beam Loss Monitor hardware and software and fully integrated into the LHC control system. The Machine Induced Background (MIB) and Luminosity measuring systems in the Pixel volume must also be replaced.

Trigger The latency of the present L1 trigger is limited to 3.4 μs by the tracker readout. For Phase-II operation, it will be increased to 12.5 μs to provide sufficient time for the hardware track reconstruction and matching of tracks to muons and calorimeter information. This change will require upgrades of the readout electronics in some of the existing sub-detectors that will be kept for Phase-II. A proper design of the front-end electronics for these systems will allow

latency limitations to be overcome and at the same time to eliminate L1-trigger rate restrictions. Based on the expected performance of the trigger with track information, the proposed L1-trigger acceptance rate is 500 kHz for beam conditions yielding 140 PU. This will allow CMS to maintain thresholds comparable to those that will be used in a typical Phase-I trigger menu. To retain comparable performance in beam conditions that result in 200 PU, the L1 rate must increase to 750 kHz, and so all detectors will have readout capabilities compatible with this possibility. Studies are underway to optimize scenarios for the trigger menu and to determine if a higher acceptance rate would further improve the exploitation of the higher luminosity for key physics signals. Any further increase of the L1 readout rate would require an increase of the Pixel readout bandwidth.

Specific sub-detector upgrades also required for CMS to meet these trigger requirements are the front-end electronics of the barrel calorimeter; the Muon readout electronics in the CSCs of the inner rings in stations 2 to 4; and the DT readout.

Data Acquisition and Trigger Control The Data Acquisition (DAQ) system will be upgraded to implement the increase of bandwidth and computing power that will be required to accommodate the larger event size and L1-trigger rate, and the greater complexity of the reconstruction at high PU. Compared to Phase-I, the bandwidth and the computing power requirements would respectively increase by factors of about 10(15) and 15(30) for operation at PU of 140(200). This is well within the projected network and computing technology capabilities expected at the time of Phase-II. Assuming an online event selection of 1/100 event at the HLT, as is the case in the current system, the subsequent rate of recorded data will increase at PU of 140(200) to 5(7.5) kHz from LHC Run-I levels of roughly a few hundred Hz.

Software and Computing Assuming only technology improvements and maintaining existing techniques, the offline software and computing areas would fall short by a factor of 4(12) of the resources needed for the challenging conditions expected in Phase-II at 140(200) pileup. To minimize the computing needs, both at the online and offline levels, a significant R&D program has started as part of the upgrade effort to improve the algorithms and approaches used for data reconstruction, analysis, storage, and access and to adapt the CMS software and computing model to new technologies and resources.

Experimental Area and Shutdown Considerations During long shutdowns CMS is highly configurable to allow access to the various sub-systems, as shown in Fig. 1.18, but the access to different areas must often be sequential because of the limited overall size of the experimental cavern. Shutdown planning for Phase-II is still at an early stage, but an initial evaluation of the work sequence and time estimates indicates that the full scope of work can be accomplished in a shutdown of approximately 30 months duration, from end to re-start of beam operations. In order to gain flexibility in scheduling the work during LS3 while reducing overall costs, consideration is being given to advancing some specific tasks to LS2, if funding is available. Radiation protection and dose to personnel will be a primary concern in planning the upgrades and the shutdown work. This will require development of special shielding, tooling, and work procedures.

1.5.4 Upgrade Performance Studies

Full simulations of detector signals using GEANT 4 have been produced in order to develop the CMS scope for Phase-II and to evaluate the performance of the proposed upgrades. It is assumed that the luminous region has a Gaussian shape along the beam (z) axis with a 5 cm RMS. The configurations that have been simulated are the following:

- the Phase-I detector operated at 50 PU ($10^{34}\text{cm}^{-2}\text{s}^{-1}$) without radiation aging; to establish a benchmark for the required performance of the Phase-II upgrades;
- the Phase-I detector operated at 140 PU ($5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$) with modelling of the effects of radiation damage after integrated luminosities of $1000\,\text{fb}^{-1}$, and in a few cases up to $3000\,\text{fb}^{-1}$, for the outer tracker and hadron and electromagnetic calorimetry in order to identify the key areas to be addressed by the upgrades, and to also determine when the need to install upgrades will become critical; and
- the Phase-II detector operated at 140 PU ($5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$); to evaluate the performance reach for the new concepts. It is assumed that the performance of the new sub-detectors will not degrade with radiation while an intermediate aging of $1000\,\text{fb}^{-1}$ is included for the barrel calorimetry since their active elements will not be replaced.

Since the pixel design is still being developed, the Phase-I configuration has been used for the Phase-II simulations, implementing the new disks in the forward regions but not incorporating expected resolution improvements. Software development for the reconstruction of physics objects (leptons, photons, jets, total hadronic and missing energy), including the sophisticated particle flow techniques for global event reconstruction, is a long-term endeavor. The CMS collaboration has made a large effort to adapt existing software to the new detector geometries and to improve or tune algorithms for the higher PU conditions. This work is still ongoing. Taking these considerations into account, it follows that the performance presented in this document should be considered to be conservative.

It is also important to recognize that the second configuration does not describe the full aging of the detector at 1000 fb^{-1}. In reality, the first two layers of the pixel detector will be too damaged to operate and the forward electromagnetic and hadronic calorimetry will also be non-functional. The outer tracker will also be at the point at $1000\,\text{fb}^{-1}$ where it will soon become completely disfunctional.

Simulations based on DELPHES [22], implementing simplified detector geometries and parameterization of efficiencies and resolutions applied at the generator level, have also been used to produce samples with sufficient statistical precision for studies of the physics backgrounds. The DELPHES parameterizations and results have been tuned and verified by comparison of various DELPHES samples with corresponding samples that have been produced with the full CMS GEANT4-based simulation.

1.6 Organization and Content of this Technical Proposal

Each sub-detector or sub-system upgrade is presented in its own dedicated chapter. These chapters contain the motivations and requirements for each specific upgrade, the description of the conceptual design of the new system and the proposed or considered experimental techniques that will be used to fully exploit it. The simulation studies included in each chapter are those that best demonstrate the deficiencies of current system for Phase-II operation and the intrinsic detector performance of the upgrade. The plans for research and development are either distributed among the various chapter subsections or they are summarized at the end of the chapter. The CORE cost estimate (Material and Services) are presented in the last section of each chapter.

Chapter 8 presents estimates of the computing resources, based on the current software and computing model, that would be needed to analyze the very complex events obtained at the

HL-LHC at the proposed HLT output rate. The elements of an R&D program are presented that will enable CMS to adapt its current computing model to cope with these conditions and the anticipated changes in computing technology.

The identification and reconstruction of the physics objects - namely electrons, γ's, muons, taus, jets, MET and b tags - is a multi-detector process that is achieved through global event reconstruction. For each of the aforementioned objects, performance comparisons for the different simulation conditions are discussed in Chapter 9. Sensitivity and reach for key physics signals that demonstrate the benefit of specific upgrades are presented in Chapter 10.

The infrastructure upgrades, together with the scope and logistics of work during LS3 are presented in Chapter 11. Schedule considerations and the assumptions for the cost estimate of the Phase-II upgrades are discussed in Chapter 12.

To assist the reader, a glossary of the acronyms used in this document can be found in Appendix B.

This Technical Proposal presents the conceptual designs for cost-effective upgrades that maximize the physics potential of CMS over the full HL-LHC period. All major sub-detector and sub-system upgrades will be documented in Technical Design Reports (TDRs). These reports are scheduled for completion in the period from the end of 2016 to end of 2017 [3], when it is expected that all of the necessary experimental techniques will have been developed and studied. They will also provide updated cost and schedule information. This process of producing a Technical Proposal followed by more detailed Technical Design Reports has been the standard approach that CMS has taken throughout its history.

[3]The DAQ system does not need similar production time as the other sub-systems and the corresponding TDR will therefore be published at a later stage, in 2020, to also fully build on the most recent technology progress

Chapter 2

Tracker

2.1 The Phase-II Tracker Upgrade

In order to maintain or improve the physics performance of the CMS detector in the high pileup conditions of the HL-LHC, the entire tracking system must be replaced with new detectors featuring higher radiation tolerance and enhanced functionality.

In this Section the limitations of the present tracking system are briefly reviewed, followed by a discussion of the requirements for the upgraded tracking detectors and an overview of the design concepts. Sections 2.2 and 2.3 describe in more detail the ongoing developments for the Pixel and the Outer Tracker detectors, respectively. In Section 2.4 the expected performance of the upgraded detector is presented and Section 2.5 describes the project organization, provides an overview of the schedule and the most important next steps, and presents a preliminary estimate of the cost.

2.1.1 Limitations of the CMS Tracker

The present Outer Tracker was designed to operate without any loss of efficiency up to an integrated luminosity of $500\,\mathrm{fb}^{-1}$, and an average pileup (PU) of less than 50 collisions per bunch crossing. The pixel detector will be replaced with the Phase-I upgrade [6] during the Extended Technical Stop at the end of 2016. The Outer Tracker and Phase-I Pixel detector would be unable to cope with even the first few years of operation of the HL-LHC. They will have to be replaced and the only available time to do this will be during the long shutdown (LS3) needed to install the machine components for the LHC upgrade. In addition to their intrinsic limitations, the Phase-I tracking detectors restrict the CMS Data Acquisition to a maximum Level-1 (L1) accept rate of about 100 KHz, with an available latency of 4 μs for the trigger decision. Operation at high luminosity requires a substantial upgrade of the trigger system, with significantly higher rate capability and longer latency.

Accumulated radiation damage in the pixel sensors reduces the charge collection as well as the Lorentz angle, leading to lower charge sharing among neighbouring pixels and hence worse spatial resolution. The effects are modelled in the detailed PixelAV [23] simulation. A worse hit-resolution directly translates to degraded precision in primary vertex reconstruction, track impact parameter resolution, and b-tagging performance. After $500\,\mathrm{fb}^{-1}$ the aging model predicts an impact parameter degradation of more than 50% while the longitudinal impact parameter resolution degrades by a factor of two. In addition, for a PU of 140, bandwidth limitations in the readout electronics would lead to an irreducible data loss of approximately 7% in the first pixel barrel layer, which is the crucial layer for primary and secondary vertexing.

For the Outer Tracker, the most prominent consequence of irradiation is the increase of leakage current, which can be mitigated, to some extent, by lowering the operating temperature of the

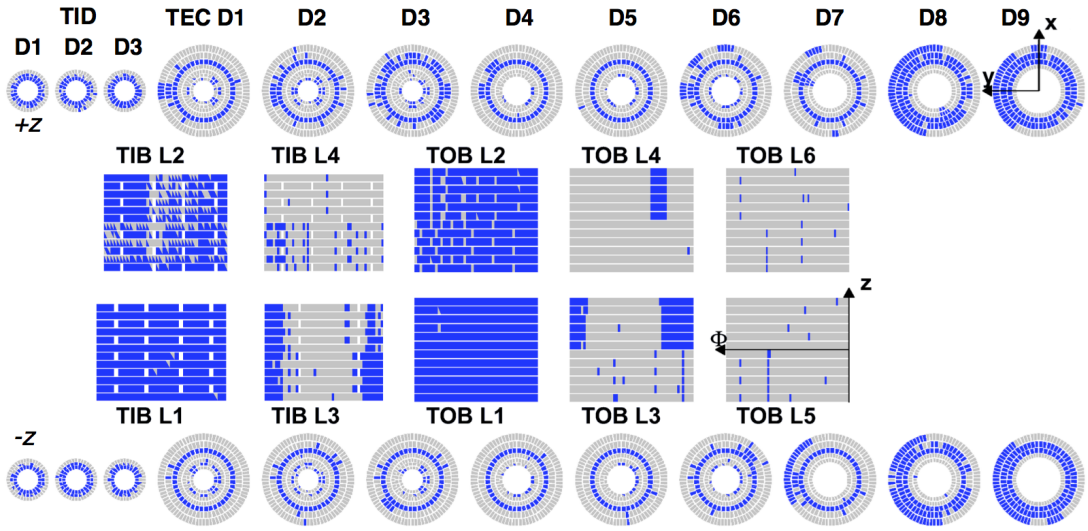


Figure 2.1: Map of non-functional modules (in blue) after an integrated luminosity of 1000 fb^{-1}, for the achievable minimum coolant temperature of $-20°$C.

cooling system to achieve a lower silicon sensor temperature. The leakage current is monitored by the detector control system for each of the 15148 detector modules of the tracker, and its evolution with irradiation and annealing is accurately reproduced by well-established models of radiation damage in silicon [24]. The evolution of the leakage current of the tracker sensors is predicted by a detailed model that takes into account the estimated luminosity profile, the position and size of each module, the expected particle fluence at specific module locations (obtained from FLUKA simulations [25]) and the expected temperature versus time scenario that includes annealing periods. The model also implements a map of the efficiency of the module thermal contacts derived from data.

Figure 2.1 displays the map of the non-functional modules after 1000 fb^{-1} for the achievable minimum coolant temperature of $-20°$C. Almost all the stereo modules in the barrel (Inner Barrel layers 1 and 2 and Outer Barrel layers 1 and 2), as well as in the endcap (rings 1, 2 and 5), are no longer operational.

The loss of hits on many layers of the tracker results in a significant degradation of track reconstruction performance, as seen in Fig. 2.2. Within $|\eta| < 2.4$, the track finding efficiency for $p_T = 10$ GeV muons decreases from 100% for the strip tracker without aging at 50 PU to $75-95\%$ after 1000 fb^{-1} at 140 PU. The efficiency for tracks from $t\bar{t}$ events with $p_T > 0.9$ GeV drops from above $85-95\%$ to $50-80\%$, while the single track fake rate increases from less than 5% to $12-45\%$. Reducing the fake rate can only be achieved by requiring more hits on each track, thereby further reducing the efficiency for finding real tracks.

The efficiency loss decreases the physics reach of most searches for new physics, diminishes the effectiveness of high-p_T lepton isolation criteria, and degrades jet energy and missing transverse energy (MET) resolution. Fake tracks cause biases and resolution degradation in jet energy measurements, increase background levels, and adversely affect high-p_T lepton isolation criteria.

CMS also makes profitable use of tracks below 0.9 GeV in the Global Event Description [26], although the track finding efficiency progressively decreases at very low momentum. For $p_T <$ 0.9 GeV the performance degradation in the aged detector is even larger, with the fake rate

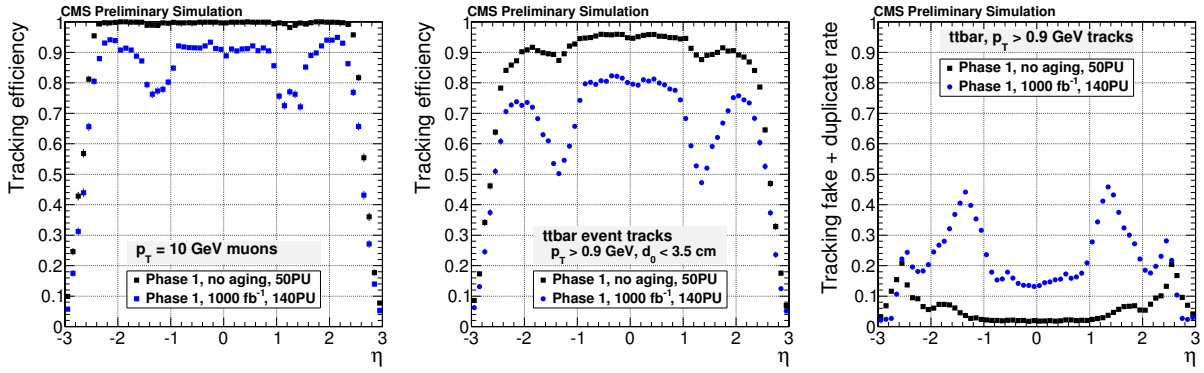


Figure 2.2: Left: efficiency for $p_T = 10$ GeV muons as a function of pseudorapidity for the Phase-I detector before and after the Outer Tracker has been aged by an equivalent of 1000 fb^{-1}. Center: same plot for charged particles from $t\bar{t}$ events for which the particles have $p_T > 0.9$ GeV and are produced within 3.5 cm of the interaction region (in the transverse direction). Right: fake rate, the fraction of reconstructed tracks that are not matched to a simulated charged particle, for the same selection of particles in $t\bar{t}$ events.

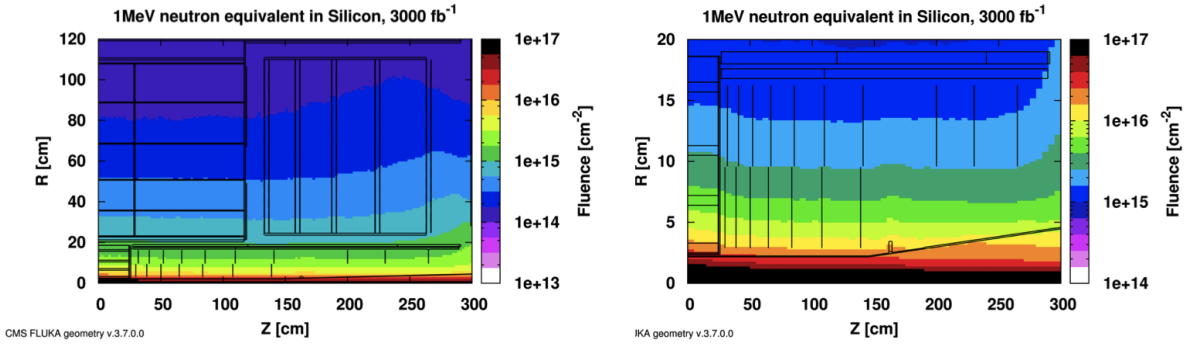


Figure 2.3: Left: map of the expected particle fluence in the Tracker volume corresponding to an integrated luminosity of 3000 fb^{-1}, expressed in terms of 1 MeV neutron equivalent fluence. Right: detail of the fluence in the pixel volume. The expected fluence has a strong dependence on radius, while it is almost independent of the z coordinate.

reaching close to 70% in the rapidity regions around $|\eta| = 1.5$.

2.1.2 Requirements for the Tracker Upgrade

The main requirements for the Tracker Upgrade can be summarized as follows:

• Radiation tolerance. The upgraded Tracker must be able to operate efficiently up to an integrated luminosity of 3000 fb^{-1}. The expected particle fluences that must be tolerated are shown in Fig. 2.3. This requirement must be fulfilled without any maintenance intervention for the Outer Tracker, while for the Pixel detector it is envisaged to keep the present concept of easy accessibility, offering the option to replace modules and other elements in the innermost regions.

• Increased granularity. In order to ensure efficient tracking performance at high pileup, the channel occupancy must be maintained near or below the 1% level in all tracker regions, which requires higher channel density. An average of 140 collisions per bunch crossing is taken as the target number of pileup events to benchmark the performance of the detector.

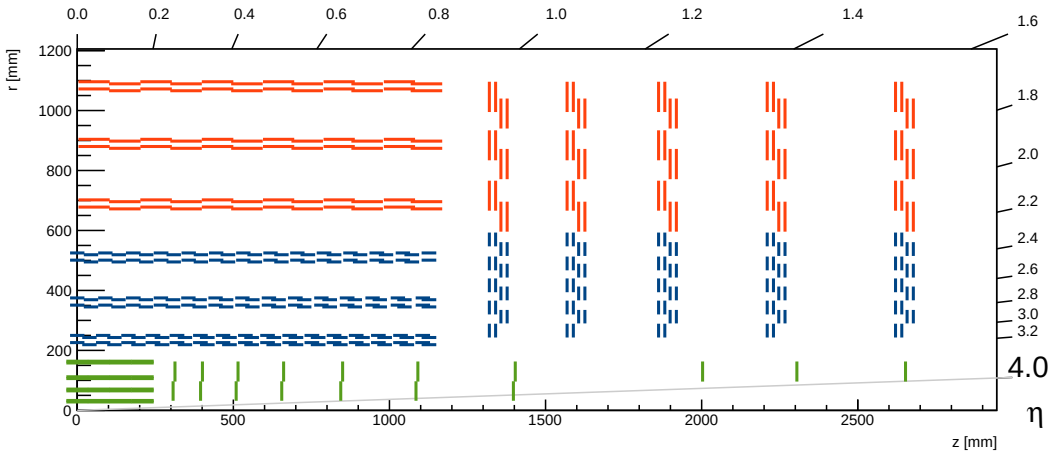


Figure 2.4: Sketch of one quarter of the Tracker layout. Outer Tracker: blue lines correspond to PS modules, red lines to 2S modules (see text). The Inner Pixel detector, with forward extension, is shown in green.

• Improved two-track separation. The present Tracker has degraded track finding performance in high-energy jets, due to hit merging in the Pixel detector. In order to optimally exploit the statistics of the high-luminosity operation, the ability to distinguish two close-by tracks needs to be improved.

• Reduced material in the tracking volume. The performance of the current Tracker is significantly limited by the amount of material, which also affects the performance of the calorimeters and of the overall event reconstruction in CMS. Operation at high luminosity will greatly benefit from a tracker with significantly less material in the fiducial volume.

• Robust pattern recognition. Track finding at high pileup becomes increasingly more difficult and time consuming. The layout of the upgraded Tracker should enable fast and efficient track finding, which is particularly important for the high-level trigger (HLT).

• Compliance with the L1 trigger upgrade. The selection of interesting physics events at L1 becomes extremely challenging at high luminosity, not only because of the rate increase, but also because selection algorithms become inefficient at high pileup. Therefore, in order to preserve and possibly enhance the performance of CMS in a wide spectrum of physics channels, it is foreseen to increase the maximum L1 rate and latency to 750 kHz and 12.5 μs, and to add tracking information in the trigger decision, moving to L1 part of the reconstruction that is today performed in the high-level trigger.

• Extended tracking acceptance. It has been demonstrated that the overall CMS physics capabilities would greatly benefit from an extended coverage of the tracker and calorimeters in the forward region (see 10). For the Tracker, this requirement mostly concerns the layout of the Pixel detector.

In Sections 2.1.3 and 2.1.4 the main features of the upgraded Pixel and Outer Tracker are introduced. Detailed descriptions of the two designs are presented in Sections 2.2 and 2.3. The boundary between the two detectors is at around $R = 20$ cm, the same location as the interface between the Pixel detector and the Strip detector in the present configuration of CMS. A sketch of one quadrant of the Phase-II Tracker layout is shown in Fig. 2.4.

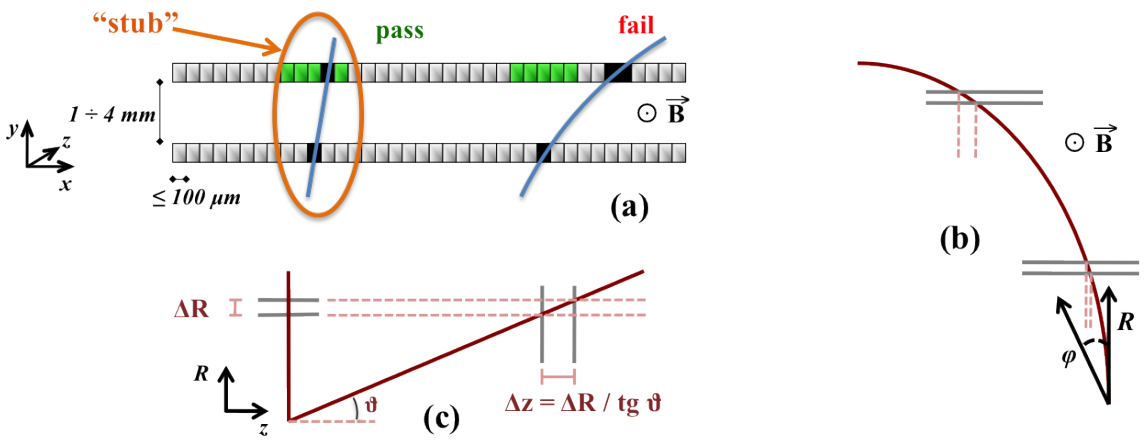


Figure 2.5: (a) Correlation of signals in closely-spaced sensors enables rejection of low-p_T particles; the channels shown in light green represent the "selection window" to define an accepted "stub". (b) The same transverse momentum corresponds to a larger distance between the two signals at large radii for a given sensor spacing. (c) For the end-cap disks, a larger spacing between the sensors is needed to achieve the same discriminating power as in the barrel at the same radius. The acceptance window can therefore be tuned along with the sensor spacing to achieve the desired p_T filtering in different regions of the detector.

2.1.3 Overview of the Pixel detector design

The requirement of radiation tolerance is particularly demanding for the Pixel detector, as shown above in Fig. 2.3. Preliminary studies show that good results can be obtained by using thin planar silicon sensors, segmented into very small pixels. With such a configuration the detector resolution is much more robust with respect to radiation damage than the present detector, where the precision relies on the ability to reconstruct the tails of the charge deposited in a 300 micron-thick sensor. At the same time the required improvement in two-track separation mentioned above is also obtained. Pixel sizes of 25×100 μm^2 or 50×50 μm^2 are being considered, representing a factor of 6 reduction in surface area compared to the present pixel cells. For the readout chip, such a small pixel size can be achieved with the use of 65 nm CMOS technology and an architecture where a group of channels (pixel region) shares digital electronics for buffering, control, and data formatting.

An alternative option that is being actively pursued is the possibility to use 3D silicon sensors, offering intrinsically higher radiation resistance because of the shorter charge collection distance. As the production process is more expensive and so not suitable for large volumes, the use of 3D sensors could be limited to the small regions of highest particle fluence.

The research on sufficiently radiation tolerant sensors and the design of the readout chip are the key activities during this initial phase of the detector development. They are discussed in Sections 2.2.1 and 2.2.2.

The new design will preserve the ease-of-access of the current detector that enables the possibility to replace degraded parts over an Extended Technical Stop. The geometry of the Phase-I detector [6] with 4 barrel layers and 3 forward disks is taken as a starting point. The forward extension could be most simply realized by increasing the number of forward disks from 3 to 10, out of which the last 3 consist of the outer part only, in order to be compatible with the conical section of the beam pipe. Such an extended pixel detector will have an active surface of approximately 4 m^2, compared to 2.7 m^2 for the Phase-I detector. The time required for the

construction of the pixel detector is considerably shorter than for the Outer Tracker, hence the time available to design the detector is longer by approximately two years.

2.1.4 Overview of the Outer Tracker detector design

The Outer Tracker provides data both for the L1 reconstruction (for each bunch crossing), and for the global event processing upon reception of a L1 trigger decision. The L1 functionality depends upon local data reduction in the front-end readout electronics, in order to reduce the required bandwidth of the L1 data stream. This is achieved with modules that are themselves capable of rejecting signals from particles below a certain p_{T} threshold, that are referred to as "p_{T} modules" [27]. A threshold of around 2 GeV corresponds to a data volume reduction of roughly one order of magnitude, which is sufficient for the purposes of L1 data transmission.

The modules are composed of two closely spaced silicon sensors read out by a common front-end. The front-end ASICs correlate the signals collected in the two sensors, and select pairs that form "stubs" compatible with particles above the chosen p_{T} threshold. The strong magnetic field of CMS provides sufficient sensitivity to measure p_{T} over the small sensor separation, enabling the use of p_{T} modules in the entire radial range above $R \approx 20$ cm. Stub data are sent out at every bunch crossing, while all other signals are stored in the front-end pipelines for reading out when a trigger is received. In order to implement the same p_{T} threshold for the stubs throughout the tracking volume, the acceptance window must be programmable in the front-end ASICs, and different sensor spacings must be implemented in different regions of the Tracker (see sketches in Fig. 2.5).

In order to achieve the required radiation tolerance it is critical to choose appropriate sensor material and processing technology. An extensive R&D program has been carried out to identify viable options and sensor designs for the Tracker and to define the requirements on the operating temperature (Section 2.3.1).

Two types of p_{T} modules are under development. "2S" modules are composed of two superimposed strip sensors of approximately 10×10 cm^2, mounted with the strips parallel to one another. They populate the outer regions, above $R \approx 60$ cm (in red in the sketch of Fig. 2.4), which corresponds to a sensor surface area of approximately 150 m^2. Wire bonds at opposite ends of the sensor provide the connectivity of both sensors to the readout hybrid. A single "service hybrid" carries a 5 Gb/s data link, an optical converter, and the DC/DC converter that provides power to the module electronics. The use of one optical link per module provides the bandwidth needed for the trigger functionality, and at the same time offers significant advantages in the overall system design by avoiding additional electrical interconnectivity in the tracking volume. "PS" modules are composed of two sensors of approximately 5×10 cm^2, one segmented in strips, and the other segmented in "macro-pixels" of size 100 μm $\times$ 1.5 mm. The chosen pixel size permits the use of the "C4" bump-bonding technology, an industrial process that is expected to be affordable for a large-scale production. As for the 2S module, wire bonds provide the connections from the strip sensor and from the macro-pixel readout chip to the front-end hybrid, and, in turn, to the auxiliary electronics for powering and readout, all of which is integrated in the module assembly. PS modules are deployed in the radial range between $R \approx 20$ cm and $R \approx 60$ cm (blue in the sketch of Fig. 2.4), resulting in a sensor surface area of about 60 m^2 (30 m^2 short-strip sensors and 30 m^2 macro-pixel sensors). The pixelated sensors provide sufficiently precise measurements of the z coordinate for tracking to enable primary vertex discrimination at L1. At the same time, three additional layers of unambiguous 3D coordinates each with an associated estimate of the particle p_{T}, are of particular use for track finding, offering enhanced robustness for the pattern recognition in a more cost effective way

than, for instance, an extension of the Pixel detector to include additional layers at larger radii.

Enabling trigger functionality at L1 requires an additional step in the back-end processing beyond what is needed in the present system. At each bunch crossing the "stub data" are processed to form "L1 tracks". The L1 tracks are tracking primitives that are combined with information from the other sub-detectors to form L1 triggers. The cabling of the detector and the overall architecture of the back-end system must be optimized for efficient track finding with an affordable amount of data traffic. The development of the electronics system is discussed in Section 2.3.2. More details on module designs are given in Section 2.3.3.

To remove heat from electronics and sensors, CO_2 two-phase cooling will be used. This choice of cooling technology helps to reduce the amount of passive material in the tracking volume.

The layout of the Outer Tracker has been the subject of extensive studies and detailed modelling, exploring several variants, including geometries with barrel layers only, and geometries with different numbers of barrel layers, and/or different numbers and size of end-cap disks. The version shown in Fig. 2.4 has been adopted as the baseline design, as it provides efficient use of the silicon sensors while providing good tracking performance while minimizing both cost and material in the tracking volume. Further optimization of the inner barrel region may be possible, and is being explored. It should be noted that all end-cap disks are equipped down to the lowest radius, to be compatible with an extension of the tracking acceptance up to $|\eta| = 4$, while in the present tracker, rings located beyond $|\eta| \leq 2.5$ are not equipped.

The designs of mechanical structures that are consistent with the chosen tracker designs are described in Section 2.3.4.

In summary, in the baseline layout the Outer Tracker consists of 15508 detector modules (8424 2S and 7084 PS), with a total active surface of 218 m^2, 47.8 million strips and 218 million macropixels.

2.2 The Pixel Detector

The target integrated luminosity of 3000 fb^{-1} corresponds to a hadron fluence[1] of 2×10^{16} n_{eq} cm^{-2} at 3 cm from the interaction region, roughly where the first layer of the Pixel Detector will be located as is the case for the Phase-I detector. The fluence decreases rapidly with distance and is about 3×10^{15} n_{eq} cm^{-2} at a radius of 11 cm. The latter is similar to the fluence foreseen for the ATLAS IBL [28] and the innermost layers of the CMS Phase-I pixel detector [6] after 500 fb^{-1}.

The extreme radiation levels of the innermost regions of the pixel detector pose major challenges for the development of sensors and electronics, and different options are under study. The possibility of replacing parts of the detector during an Extended Technical Stop remains as an option in the event that some of the detector components will not survive the full radiation dose provided by the HL-LHC.

2.2.1 Development of silicon sensors

The n-in-n planar technology, developed for the current pixel detector and its Phase-I replacement, is a valid baseline for instrumenting the outer layers of the barrel and the outer regions of the forward disks. Collection of electrons is advantageous because of their higher mobility

[1] It is common practice to express the fluence as the number of 1 MeV neutrons per unit surface area that would produce an equivalent damage in silicon.

compared to holes, leading to lower trapping probability and hence larger collection efficiency after heavy irradiation. Several development lines are ongoing to further improve the performance of planar sensors, and possibly reduce the cost for their deployment in the extended pixel geometry (Fig. 2.4) and operation at the HL-LHC.

Adoption of n-in-p sensors could reduce the cost, as they are produced with single-sided photolithography which is cheaper than the double-sided processing required for n-in-n sensors. The n-in-p option requires the development of a robust scheme to protect against micro-discharges, in order to ensure safe operation at the high bias voltage required after heavy irradiation (in n-in-p sensors the sensor edges are at bias voltage).

Thin sensors (150 μm or less) offer advantages in terms of lower bias voltage and lower leakage current. Moreover, the shorter drift distance results in smaller clusters which, when combined with a smaller pixel cell and a reduced signal threshold for the smaller charge that is produced (see below Section 2.2.2), can achieve good resolution and improved two-track separation in high-energy jets. The enhanced radiation tolerance should make it possible to extend the use of planar sensors at small radii.

A variety of n-in-p planar pixel sensors of various thicknesses are under study to identify the most suitable combination of material and thickness, sensor technology and design, and to assess the ultimate radiation tolerance achievable with planar sensors.

For the innermost regions of the detector the 3D silicon sensor technology could offer advantages. For instance, an intrinsically lower bias voltage and shorter drift distance leads to lower trapping probability. For small pixel cells, however, 3D sensors may suffer from significant charge loss in the implant region. A dedicated development program is ongoing to optimize cell configuration, sensor thickness, number of implants per cell, and implant aspect ratio, for operation at the HL-LHC. Once the sensor design has been optimized it will be possible to carry out a thorough assessment of the performance and radiation tolerance for comparison with the planar (i.e. surface implant) option.

While continuing to follow the generic R&D of other technologies such as diamond sensors, HV-CMOS, and monolithic active pixel sensors (MAPs, foreseen for the ALICE upgrade [29]), the development plan outlined above, based on the comparison of planar silicon and 3D silicon sensors, will provide a solid baseline for the pixel detector. Figure 2.6 displays a compilation of results of measured charge in planar silicon strip structures (left) and 3D silicon pixel structures from different vendors (right), after heavy irradiations (see also [30]); such results suggest that sufficient charge collection may be retained even at the innermost radii, and show the clear advantage of 3D silicon in regard to low bias voltage. The final choice will take into account performance and radiation tolerance, as well as the expected sensor cost and yield.

2.2.2 Readout electronics

The general concept for the implementation of the pixel front-end electronics system is illustrated in the sketch of Fig. 2.7 (left), for a typical barrel layout. Hybrid pixel detector modules are made from multiple pixel Read-Out Chips (ROC) bump-bonded to a single pixel sensor. Readout and control signals plus power are connected to the ROCs with wire bonding to a thin and light Printed Circuit Board (PCB) glued on the back side of the pixel sensor. Low-mass cables connect the pixel modules to the global readout, control and powering systems. The heat generated by the pixel chips and sensor is removed via a heat distribution layer to thin CO_2 cooling pipes. The optional use of Through Silicon Vias (TSV), to get readout signals and power of the ROC out from its back side will also be investigated to build more compact pixel detector

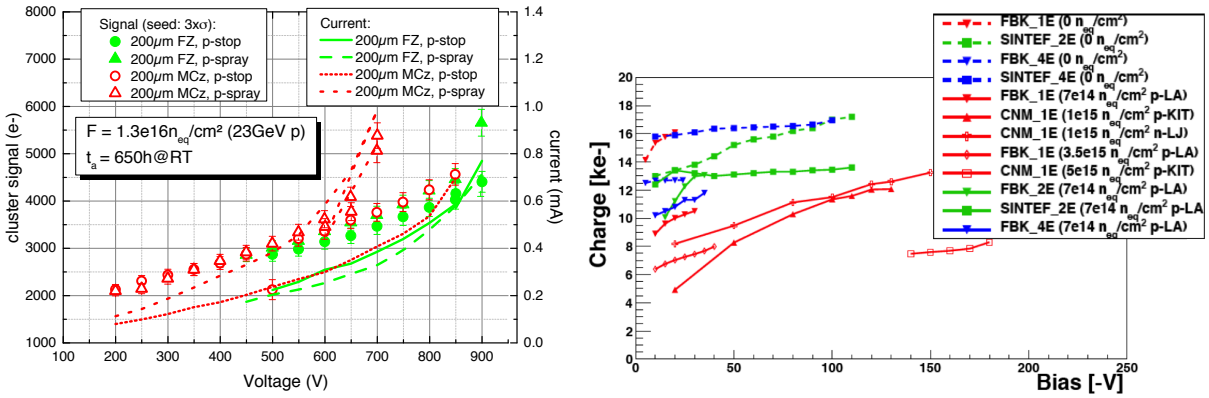


Figure 2.6: Left: signal charge and leakage current measured in planar n-in-p strip structures in different materials and technologies, after irradiation to $1.3 \cdot 10^{16}$ n_{eq} cm^{-2} with 23 GeV protons and annealing equivalent to 650 hours at room temperature. Right: signal charge in 3D n-in-p pixel structures from different vendors and with different column configurations, before irradiation (dashed lines) and after irradiation up to $5 \cdot 10^{15}$ n_{eq} cm^{-2} (solid lines).

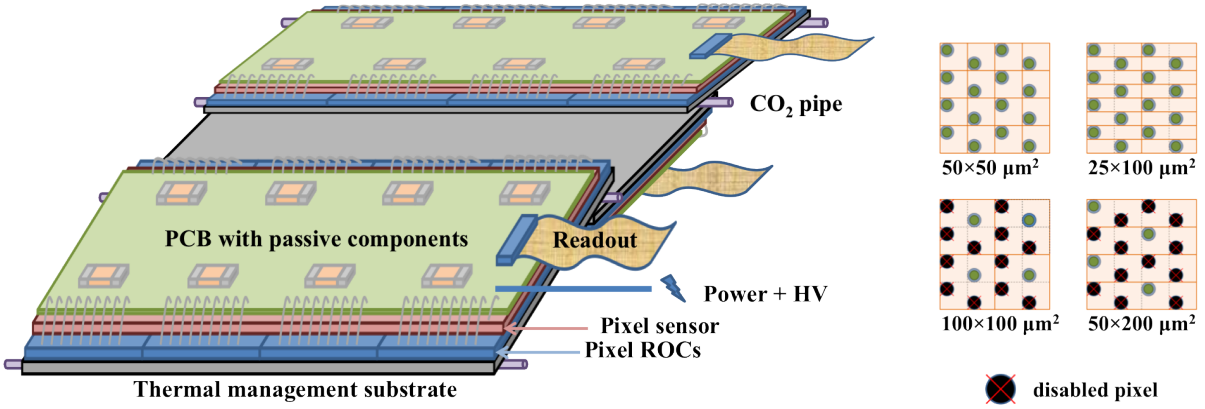


Figure 2.7: Left: sketch of the on-detector pixel electronics system, shown for the barrel configuration. Right: bump-bonding grid applied to different pixel aspect ratios and sizes.

modules with less inactive surface and material. In that case the PCB would be located on the back side of the pixel ROCs, to which it would be connected via coarse pitch bump-bonding.

A pixel surface of 2500 μm^2 has been chosen as a target, to be implemented in two aspect ratios: 50×50 μm^2 and 25×100 μm^2. These sizes are the result of a compromise between the requirement of maintaining and possibly improving tracking and vertexing performance with thinner sensors, and the expected technology limitations in sensors, readout chips and bump-bonding technologies. In the outer part of the detector, where the particle density is significantly smaller, a larger pixel size may bring advantages in terms of reduced power consumption. With an appropriate bump-bonding pattern the same ROC can be used for the two different pixel aspect ratios, and also for larger pixels with compatible dimensions by disabling the unused channels, as illustrated in the sketches of Fig. 2.7 (right). Such a scheme will allow an optimization of aspect ratio and pixel size in the different regions of the detector with one single readout chip. Since the sensor technologies and pixel sizes will be defined at a later stage, the chip will implement the possibility to tune the bias currents of the analogue front-end in order to accommodate an appropriate range of values for collected charge and channel capacitance corresponding to the sensor choices currently envisaged (planar or 3D silicon, different

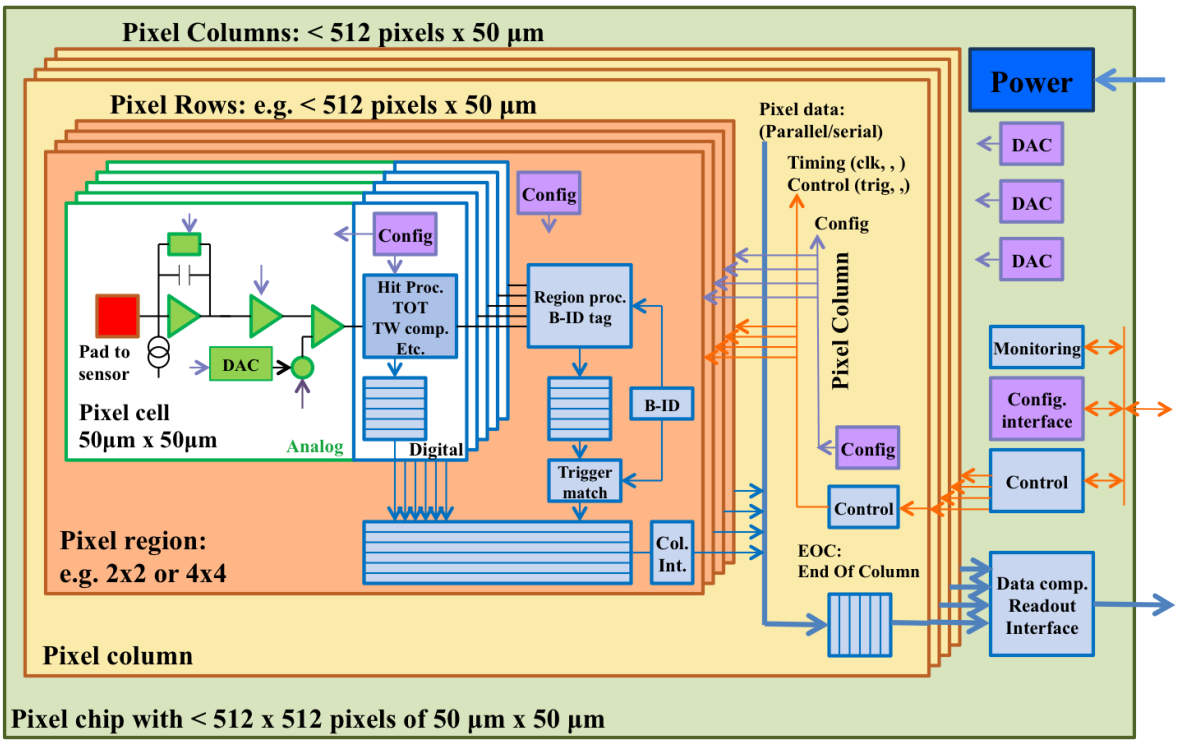


Figure 2.8: RD53 digital pixel chip architecture.

sensor thicknesses and pixel sizes).

2.2.2.1 The Readout Chip

The pixel readout chip (ROC) is the core building block of the pixel electronics system. Compared to the Phase-I implementation, the Phase-II ROC will feature six times smaller pixels, will have to cope with about five times higher hit rates, five or ten times higher trigger rates, as well as longer trigger latency. To address the challenges associated with the anticipated hit density and corresponding needs for I/O and buffering capabilities, ATLAS and CMS have chosen 65 nm CMOS technology. The new ROC will have unprecedented complexity for an HEP ASIC, with up to 1 billion transistors, and will be exposed to 10 times higher radiation levels than the current LHC ASICs. As a consequence, the development will require substantial dedicated effort to qualify the technology and optimize the design.

In order to optimize resources for such a demanding project, the cross-experiment collaboration RD53 [31] was formed, today comprising 19 institutes with a balanced participation between ATLAS and CMS. The work program of RD53 for the next three years includes the radiation qualification of the selected 65 nm CMOS IC technology, the development of all the basic circuits required to build the HL-LHC pixel ROCs, and the production of a large prototype chip with 50×50 μm^2 pixels, which will serve as a common basis for the final designs of the ATLAS and CMS pixel ROCs. In addition, RD53 is developing a pixel chip and system simulation and verification framework with the defined readout and control protocols.

The results from initial radiation tests are promising. However, some significant radiation effects have been observed for total dose levels above a level of 2 MGy [32], which require additional studies. In particular it is important to understand and quantify the interplay of operating temperature (nominally -20°C) and possible periods of annealing at room temperature.

A common pixel chip architecture, that is fully digital after the basic threshold detection and charge digitization in the analogue pixel cell, has been defined as shown in Fig. 2.8. Digital hit processing, including the critical trigger latency buffer, is implemented within the pixel array in local pixel regions (e.g. 2×2 or 4×4 pixels) followed by data merging, data formatting and readout after the Level-1 trigger accept. Buffering requirements have been analysed with a statistical model [33] and simulations and have been verified to be compatible with the proposed extended CMS trigger latency of 12.5 μs. A buffer depth of 16 pixel clusters for a 4×4 pixel region is sufficient to guarantee a hit loss probability below 10^{-3} for the highest hit rate of 2 GHz/cm^2 .

The physical size of the pixel chip is determined via an optimization for the chosen detector layout, while also taking into account the implications of the expected yield on the pixel chip fabrication and chip-to-sensor bump-bonding. A chip size of about 20×20 mm^2 with about 90% active pixel area (with the remaining 10% being used for end-of-column readout, control interface and wire bonding pads), is currently believed to be an appropriate choice for the detector layout, compatible with reliable production in 65 nm technology and available bump-bonding technologies. A chip of similar size, albeit with lower channel density, has been successfully deployed in the construction of the ATLAS IBL [28].

2.2.2.2 System aspects

The pixel detector module concept sketched in Fig. 2.7 will be realized in a minimal number of different versions in order to accommodate the requirements of the different detector regions. For the inner barrel layers, narrow modules of 1×4 chips (about 20×80 mm^2) are needed to form hermetic surfaces at the lowest radii. In the outer layers, it will probably be advantageous to deploy wider 2×4 chip modules (about 40×80 mm^2), to reduce the amount of assembly work and optimize the detector geometry. In the forward disks, it is envisaged to use rectangular modules as in the barrel, to minimize the differences and the required design and development effort. However half-size modules (1×2 and 2×2 chips) may be required to cover effectively the disk surface.

A second generation of the radiation hard GBT link chip [34], the LP-GBT (see also [35]), will operate the conversion to high-speed optical links at 10 Gb/s. The LP-GBT chip and associated laser and driver are not expected to have sufficient radiation tolerance to be used in the central part of the pixel detector. They will be located on the support cylinders at the outer boundary of the detector volume, as in the Phase-I detector. Electrical links (e-links) with lengths up to 2 m will connect the modules to the LP-GBTs. The chip and module design will allow the implementation of a variable number of e-links per ROC, and also to merge 2 or 4 ROCs into a single e-link, to match the expected hit rate and required readout bandwidth in different locations (see Fig. 2.9). For the highest expected hit rates in the first barrel layer a readout bandwidth of up to about 4 Gb/s per ROC may be required.

It is envisaged to implement 1.2 Gb/s e-links on lightweight twisted-pair cables, with optimized cable drivers and receivers in order to minimize cable mass. Active cable equalization[2] enables link rates several times higher than the basic analog bandwidth of the cable (this is an extensively used technique for commercial high speed networking via twisted pair cables). Given the large number of e-links that will be implemented in the detector volume, the optimization of this part of the system is a key element of the project.

[2]The basic concept involves implementing an active receiver that applies (a good aproximation of) the inverse filter function of the cable. See for example `http://www.analog.com/library/analogdialogue/archives/38-07/equalizer.pdf`

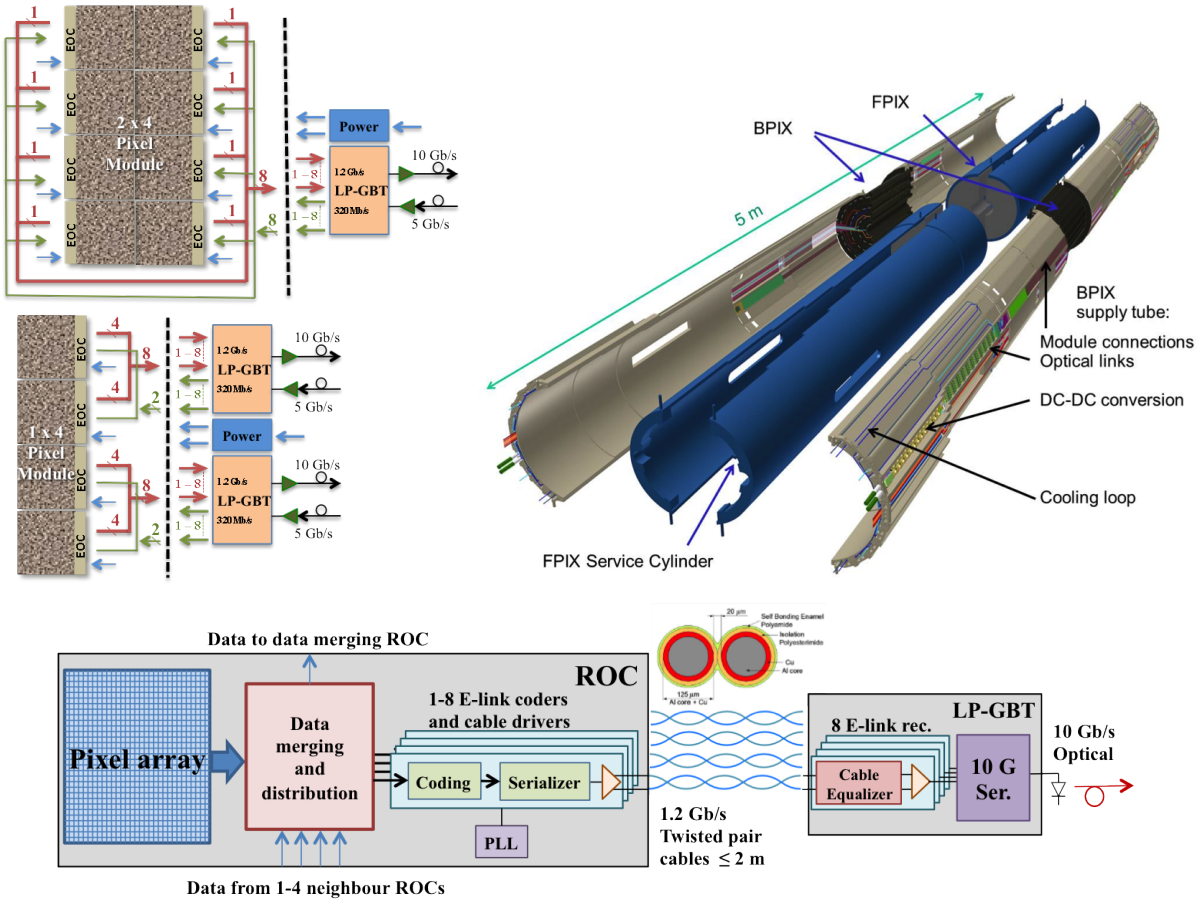


Figure 2.9: Top left: two examples of possible configurations of pixel modules and electrical readout links; in one scheme a 2×4 chip module is equipped with one e-link per chip, and served by one LP-GBT (this configuration can be used in the outer part of the barrel, with relatively low hit density); in the other scheme a 1×4 chip module is equipped with four e-links per chip, and served by two LP-GBTs (this configuration may be needed in the first layer of the barrel, with the highest hit density). Top right: exploded view of the Phase-I pixel detector. Power electronics and optical readout electronics are located on support cylinders at the outer boundary of the detector volume. A similar arrangement will be adopted for the Phase-II detector. Bottom: the ROC chip will have a highly flexible readout interface, with the possibility of implementing up to 8 e-links per chip, or merging data from 2 or 4 chips into a single e-link.

It is expected that the ROC will have a power consumption of the order of $0.3 - 0.5$ W/cm^2, similar or up to a factor two higher than the current CMS pixel chip. The exact power density of the final chip cannot yet be predicted, as many delicate and detailed design choices still have to be made, and a detailed characterization of radiation-induced transistor leakage is on-going. The relatively large power supply currents at low-voltage pose significant challenges for the power delivery to the pixel chips. A passive power distribution system from the low voltage power supplies located outside the experiment is excluded, as it would require an unacceptably large conductor cross-section. Moreover, DC/DC power conversion based on inductors cannot be integrated on the pixel modules because of the excessive radiation levels and tight constraints on space and material budget. The most promising options to be explored are: (i) A distributed DC/DC conversion scheme with air-core inductor-based DC/DC conversion on

the support cylinders (about 2 m away from the load) in combination with switched capacitor power conversion on the pixel chip itself; (ii) Serial powering with on-chip shunt regulators; and (iii) A combination of serial powering with on-chip switched capacitor power conversion and regulation. The choice of an appropriate power distribution scheme is another crucial element for the development of a high-performance pixel detector.

The cooling system will be an extension of the Phase-I system, based on evaporative CO_2. Optimization of the coolant distribution and of the on-detector thermal contacts will proceed in parallel with the engineering design of the pixel detector modules.

2.2.3 Development plans

With the ROC design and sensor R&D well underway, the main goal for the next two years is to address the development of the electronics architecture, in particular the two most crucial elements of the system: the power distribution and the low-mass electrical links. A flexible detector modelling and simulation framework is being developed, to enable quantitative comparisons of the different options for the ROC, the sensors and the electronics system, to optimize the detector performance. Finally, conceptual designs of modules and support structures will be developed, providing realistic mass estimates and enabling more refined detector modeling and simulations, allowing the overall detector layout to be optimised.

2.3 The Outer Tracker

2.3.1 Development of silicon sensors

The sensitive elements of the Outer Tracker are planar silicon sensors, which will be exposed to particle fluences up to $1.5 \cdot 10^{15}$ n_{eq} cm^{-2}, a factor of ten larger than the design requirement for the present Tracker. It has been shown [36] that the performance of p-in-n float-zone sensors degrades too much for such high fluence levels. Different types of substrates have to be utilized to achieve higher radiation tolerance. The choice of thin (< 300 μm) sensors could offer advantages in terms of reduced leakage current and less material in the tracking volume.

The baseline design of the Outer Tracker requires the following three types of sensors:

• 2S sensor: strip sensor for the 2S module. AC coupled sensor of approximately 10×10 cm^2, with two rows of 5 cm long strips with 90 μm pitch.

• PS-s sensor: strip sensor for the PS module. AC coupled sensor of approximately 5×10 cm^2, with two rows of 2.5 cm long strips with 100 μm pitch.

• PS-p sensor: "macro-pixel" sensor for the PS module. DC coupled sensor of approximately 5×10 cm^2, with 32 rows of macro-pixels 1.5 mm long with 100 μm pitch.

2.3.1.1 Research of suitable silicon material

A comprehensive program has been carried out to identify suitable silicon materials for the Outer Tracker at HL-LHC [37, 38]. Several different test-structures have been implemented in a mask design, which was used to process a number of wafers made from different silicon materials, technologies and thicknesses (ranging from 50 μm to 320 μm). To minimize systematic differences stemming from variations in proprietary technologies associated with specific vendors, a single vendor, Hamamatsu Photonics K.K. (HPK) was engaged to produce all the wafers, and the measurement setups of the participating groups have been carefully inter-calibrated.

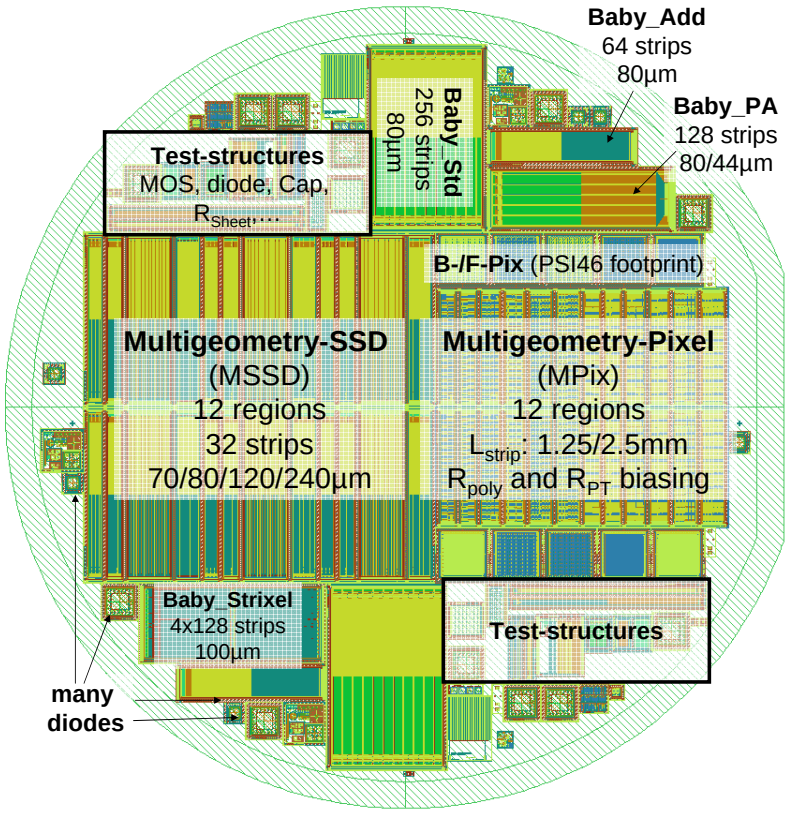


Figure 2.10: Overview of the wafer mask with labels indicating the main test-structures used in the measurement program.

The test-structures on the wafers (Fig. 2.10) include:

• Simple diodes to study macroscopic properties (leakage current, full depletion voltage, charge collection) [39–41] and microscopic properties (defect spectroscopy) of the base material [42].

• Mini-strip sensors to study sensor parameters (electrical parameters of the strips, charge collection, resolution, efficiency and Lorentz angle.) [41, 43–46].

• Multi-geometry structures to study effects related to geometry, such as strip pitch and width. These include a strip structure (MSSD) with about 3 cm-long strips [47, 48], relevant for the development of the 2S and PS-s sensors, and a macro-pixel structure (MPix) with about 1.5 mm macro-pixels, relevant for the development of the PS-p sensor. For the pixelated structure, punch-through and poly-silicon bias connections have been implemented and studied.

• Process control test-structures, to measure process and material parameters. These will be used as standardized tools to monitor the process quality during production [49–51].

• Special sensor designs such as sensors with integrated pitch adapter in the first or second metal layer, and a four-fold segmented sensor with read-out at the sensor edges [52].

The materials investigated are:

• Float-zone (FZ). The most widely used material, which was the first material available with the high quality and high resistivity needed for particle detectors. With n-type silicon it serves as reference to compare to the sensors of the present Tracker.

Table 2.1: Summary of the particle fluences used to assess the radiation tolerance of the samples studied. The values were obtained from FLUKA simulations [25].

Radius cm	Protons 10^{14} n_{eq} cm^{-2}	Neutrons 10^{14} n_{eq} cm^{-2}	Total 10^{14} n_{eq} cm^{-2}	Ratio p/n
60	3	4	7	0.75
20	10	5	15	2.0
15	15	6	21	2.5

• Magnetic Czochralski (MCz). This growth technique results in a high oxygen concentration in the silicon, which was shown to be beneficial in terms of radiation hardness by RD48[3] and RD50[4]. See also [53].

• Epitaxial. This method enables the production of very thin (25 μm $-$ 100 μm) active sensors by chemical vapour deposition of silicon on a carrier wafer, also resulting in high oxygen content. This material is of particular interest for the upgrade of the Pixel detector, which receives the highest fluences.

During the measurement program it was discovered that the FZ material from HPK has a particularly high oxygen concentration (above $8 \cdot 10^{16}$ cm^{-3}), and therefore the results obtained may not be valid for FZ material with substantially lower oxygen content. Studies of material with lower oxygen concentration from a different sensor vendor are underway to clarify this point, and disentangle the specific effect of oxygen content from other effects linked to the production process.

The samples have been fully characterized and then irradiated to particle fluences corresponding to those expected in the Tracker at radii of 15 cm, 20 cm and 60 cm for an integrated luminosity of 3000 fb^{-1}. Irradiations with protons and neutrons have been combined to simulate the hadron mixture expected in the Tracker at the different radii (Table 2.1).

Measurements have been performed after each irradiation step with a very short annealing time, and after the full irradiation (with both protons and neutrons) in several longer annealing periods. In addition, to investigate the energy dependence of proton damage, several proton sources have been used: 23 MeV from the compact cyclotron at KIT[5], 800 MeV from the LINAC at LANSCE[6] and 23 GeV from the PS at CERN[7]. Neutron irradiations have been performed at the nuclear reactor at JSI[8] [54]. For the operation after irradiation, it is assumed that the cooling system will keep the temperature of sensors at -20°C or lower: such a temperature ensures sufficient margin from sensor thermal run-away for an achievable thermal contact efficiency (Section 2.3.3).

2.3.1.2 Summary of main results

• Charge collection. The measurements of the charge collection (CC) of 300 μm thick FZ sensors confirmed the observation from [36] that strip sensors with collection of holes for signal generation have greater degradation than sensors with collection of electrons for signal generation. This has also been found to be true for thin sensors (Fig. 2.11). In addition, irradiated

[3] http://rd48.web.cern.ch/rd48/
[4] http://rd50.web.cern.ch/rd50/
[5] http://www.ekp.kit.edu/english/irradiation_center.php
[6] http://lansce.lanl.gov/about/linac.shtml
[7] https://irradiation.web.cern.ch/irradiation/irrad1.htm
[8] http://www.rcp.ijs.si/ric/description-a.html

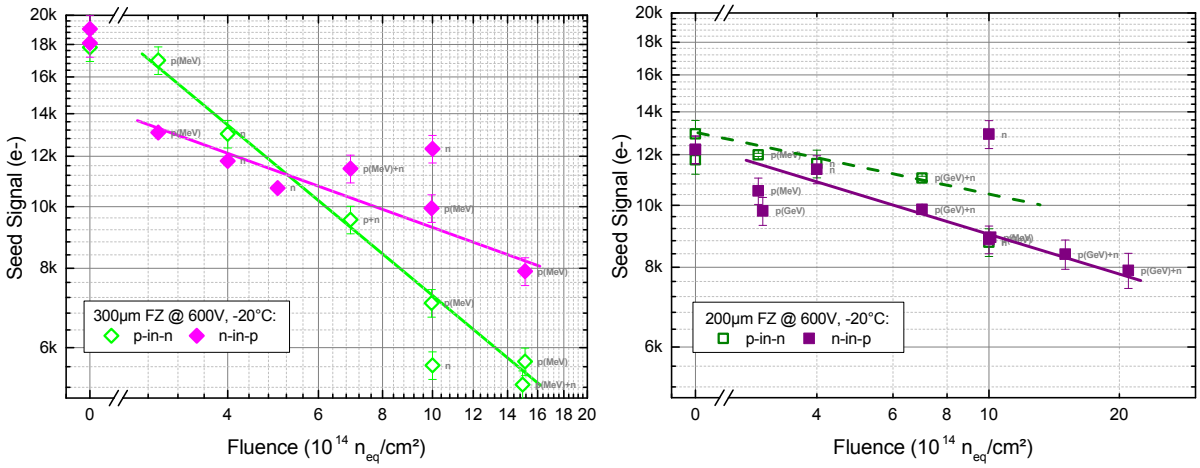


Figure 2.11: Charge collected on the seed strip (strip with the highest signal in a cluster) vs. fluence for 600 V biasing at -20° C after short annealing (50 h to 250 h) at room temperature, for sensor thicknesses of 320 μm (left) and 200 μm (right). Lines are drawn to guide the eye.

sensors with n-type bulk have shown non-Gaussian noise, resulting in an irreducible rate of fake hits, that would in some cases be as large as the expected signal occupancy. Device simulations [55] have shown that irradiated p-in-n strip sensors develop high electric fields at the strip edges that intensify with increasing accumulation of oxide charge. The electric fields around the strips in n-in-p sensors are instead reduced by higher oxide charge, which makes them more robust with respect to effects such as breakdown, noise or micro-discharge after heavy irradiation with charged particles.

• Annealing. As shown in Fig. 2.12, the CC for thin sensors biased to 600 V does not degrade with the annealing time. It is therefore possible to exploit the annealing of the volume generated current, with no significant loss of signal. There is still higher CC at $1.5 \cdot 10^{15}$ n_{eq} cm^{-2} with thicker sensors, but the additional charge decreases with annealing time. One striking result is that MCz material shows very small annealing effects, both in CC and full depletion voltage. Samples indicated as "dd-FZ" in the figure (deep-diffused FZ) are made of FZ material with very deep back-side doping, in this case to a depth of 120 μm, reducing the active thickness to 200 μm, for a physical thickness of 320 μm.

• Strip isolation. In order to implement electron readout of n-doped strips, an isolation technique involving p-doped elements between the strips has to be defined (or developed). The solutions that have been investigated are: a moderate p-doped layer ("p-spray") and highly p-doped structures surrounding the strips ("p-stops"). After a fluence of $1.5 \cdot 10^{15}$ n_{eq} cm^{-2} both solutions were shown to preserve a high inter-strip resistance of more than ten times the bias resistance as is required for good localization of the signal charge.

All other strip parameters studied showed no significant deterioration after irradiation.

2.3.1.3 Conclusions and outlook

For the choice of the sensor polarity, the results collected thus far show that sensors with electron readout are more robust in terms of high-field effects after irradiation, and also provide higher CC than p-in-n sensors. Moreover, n-in-p sensors evade the higher complication and cost of production associated with n-in-n sensors which involves a double-sided process that requires more masks. However, the use of n-in-p sensors requires special attention in the design of the area around the sensor edges, which are at high potential.

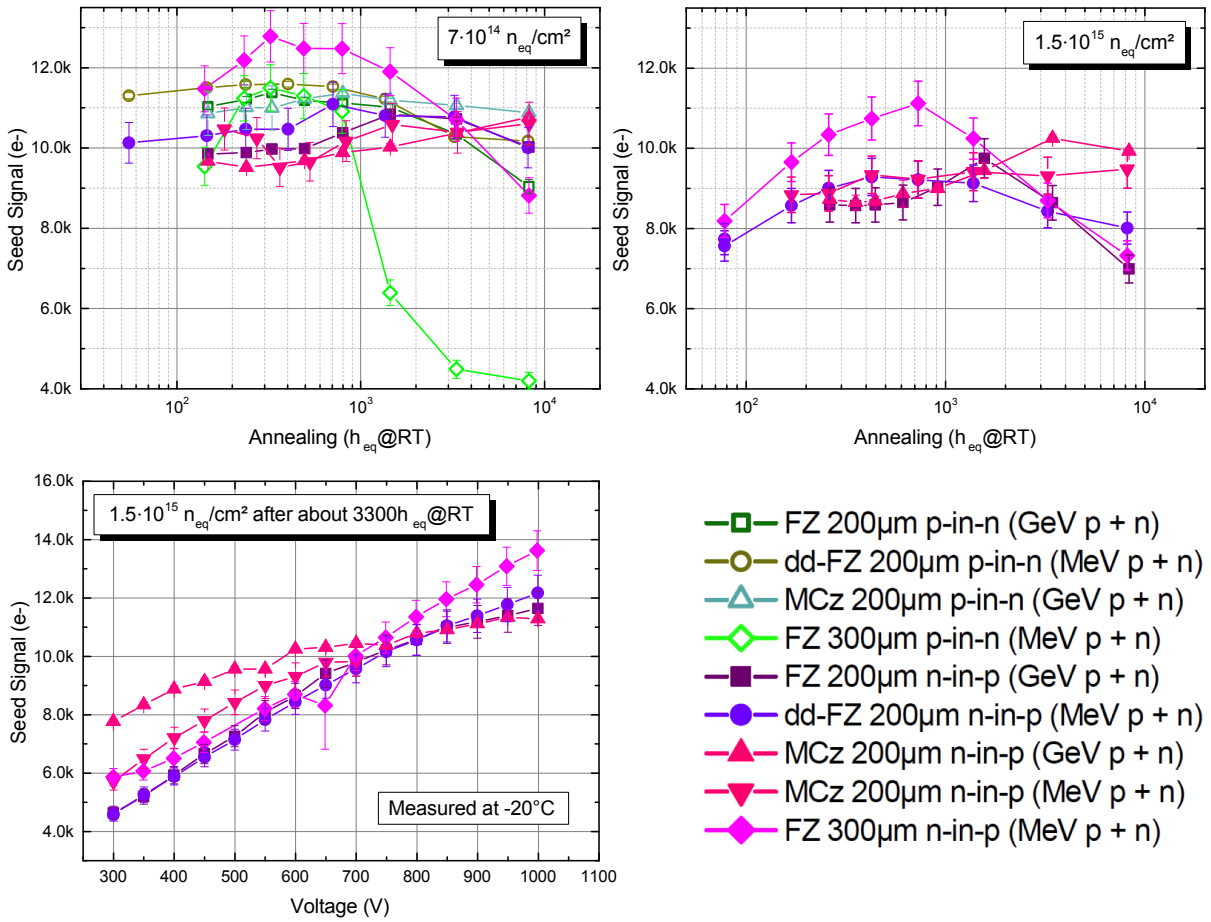


Figure 2.12: Top: charge collection at 600 V and -20° C vs. equivalent annealing time at room temperature (RT), for two different irradiation levels. Bottom: charge collection vs. bias voltage after $1.5 \cdot 10^{15}$ n_{eq} cm^{-2} and 3300 h equivalent annealing at room temperature.

The p-type materials investigated thus far show comparable CC at the highest expected fluence (innermost layer of the Outer Tracker). In addition, 300 μm sensors show higher CC than 200 μm sensors during annealing up to about 20 weeks (3360 hours) at room temperature, beyond which time the CC drops to the value of 200 μm sensors and below. It is possible to obtain more charge from 300 micron-thick sensors by applying higher bias voltage (e.g. 900 V). The CC of 200 μm sensors also shows some increase (of about 30%) at higher bias voltage after the highest fluence and 20 weeks of annealing at room temperature (Fig. 2.12).

The MCz material investigated shows lower full depletion voltages after irradiation compared to the FZ material. This results in a stable CC during annealing at 600 V and a higher CC at low bias voltage (Fig. 2.12). The latter effect is more pronounced after GeV proton irradiation. Studies of diodes made of these materials show a proton-energy-dependent damage resulting in different electric field profiles and full depletion voltage which also depend on the oxygen content of the materials [39]. The quantification of these effects is the subject of further studies. The observed beneficial properties of MCz material needs to be weighed against the potentially higher cost.

Based on the choice of n-in-p sensors, the collaboration is developing the specifications to launch a market survey in 2014, that will identify companies with adequate production capabilities and quality suitable to meet the needs of the Tracker. In parallel, studies will continue to

optimize the details of the sensor designs, and quantify more precisely the advantages and disadvantages of the possible options. The possibility of producing the sensors on 8-inch wafers (rather than 6-inch) is also being pursued in the interest of potential cost savings.

2.3.2 The Electronics System

The electronics for the Phase-II Tracker relies on custom development of front-end ASICs and back-end processing cards, and on common, LHC-wide developments for services such as data transfer (GBT and VL projects) and low-voltage powering (DC/DC converter project). Compared to the present Tracker, five main technological leaps will be exploited to transform the electronic system of the Phase-II Tracker:

• The Phase-II tracker has substantially larger channel count, and contributes to the L1 trigger decision at bunch-crossing rate. Most of the data generated by the Tracker are in fact trigger data. Use of binary data has been chosen to cope with those requirements. The new electronic system is fully digital, with the exception of the analogue ASIC front-ends.

• Use of a common data transfer system for clock, data and control, instead of the separate readout and control systems implemented in the present detector.

• Use of point-of-load DC/DC converters. This will make it possible to power the detector at a higher voltage than is the case for the present detector, thus minimizing ohmic losses in the cables.

• Use of high-density flexible hybrid circuits as low-mass, dense front-end interconnects.

• Use of a flip-chip bump-bonding technique for all front-end (FE) ASICs, enabling the assembly of the FE Hybrids on high-volume production lines. The wire bonds from the Hybrid to the readout ASICs are eliminated, thereby reducing the total wire bond count by about 50%.

The developments of electronics for 2S and PS modules are now proceeding in parallel although the development of the 2S system started earlier and is therefore more advanced. Technologies, architectures and data formats are kept similar for both module types wherever possible, to maximize the use of identical components, particularly at the back end. Commercial manufacturing and assembly processes are targeted, both to expedite the schedule and to profit from the latest technical capabilities that will be available in industry.

2.3.2.1 System Overview

The electronic system block diagram is shown in Fig. 2.13 (left). The system is designed to deliver trigger and L1 readout data with high efficiency up to a L1 accept rate of 750 kHz, and to cope with latencies up to 12.5 μs.

At the front end, the electronic system is built around the sensor modules (2S or PS). Electrically, it consists of two FE Hybrids interfacing to the two rows of strips of the silicon sensor(s) (2S and PS modules), plus, in the case of the PS module, of a Macro-Pixel-Sub-Assembly block (MaPSA) integrating the pixelated sensor with its readout chips. Data generated by eight FE chips are buffered, aggregated and formatted by the Concentrator IC (CIC) that acts as a data hub for the Service Hybrid. The latter hosts all services to/from the counting room: data transfer (4.8 Gb/s, bidirectional multi-mode optical link based on the low-speed/low-power grade of the LP-GBT chip), low-voltage powering (LV, $10 - 12$ V) and high-voltage biasing (HV, up to -800 V). It connects to the FE Hybrid through wire bonds made at the same time as those that connect the FE hybrid to the sensors. The LP-GBT ASIC serialises (deserialises) data sent to (received from) the VTRx+ optoelectronic transceiver. It also acts as I2C master of the module, controlling

the FE ASICs and contains monitoring functions that will be used to check environment and functionality. Both FE and Service Hybrids are multilayer high-density flexible kapton circuits that will be tightly folded during module assembly to provide connectivity to both top and bottom sensors and/or to sensor backsides, while limiting impact on material budget. Both 130 nm and 65 nm technologies are likely to be used for the FE ASICs in order to achieve adequate performance while minimizing development and production costs.

At the backend, the Data, Trigger & Control Board (DTC) sends and receives data to/from multiple modules (typically $50-70$ modules per card). This board is a custom development based on off-the-shelf commercial FPGAs and multi-channel optoelectronic transceivers. It processes three data streams to/from the detector: DAQ, Trigger and Timing&Control with a global usable bandwidth of 3.2 Gb/s in each direction (raw bandwidth, including Forward Error Correction, is 4.8 Gb/s). The format of the crates housing the DTC board will be either μTCA or ATCA depending on achievable channel density and constraints imposed by the other subsystems to which it will interface.

The low- and high-voltage power supply system will evolve from that of the current tracker. Power will be supplied to the FE at a voltage of $10-12$ V while the HV supply will be designed to reach -800 V. The granularity of the distribution is still under consideration. Individual supply lines to each detector module would best match the Tracker modularity and alleviate the need for complex fail-safe redundancy schemes.

2.3.2.2 Front-End Electronics for 2S modules

Sensor strips are wire bonded to the flexible FE Hybrid on each end of the 2S sensor module. The top sensor pads are bonded to the FE hybrid pads on the top Hybrid layer. Similarly, the bottom sensor pads are also bonded to the top Hybrid layer, but folded over a carbon fiber spacer matching the separation between the two sensor planes (see below Section 2.3.3.1). The polyimide Hybrid provides four copper high-density routing layers, with a total thickness of approximately 150 μm.

Each FE Hybrid hosts eight CMS Binary Chips (CBC, 130 nm CMOS) plus a CIC (65 nm CMOS). The CBC processes data from 254 strips (127 bottom and 127 top sensor strips), identifies clusters up to a programmable maximum width, performs top-to-bottom correlations over configurable windows to identify high$-p_{\mathrm{T}}$ stubs at LHC Bunch Crossing (BX) rate, and provides unsparsified binary readout data at the L1 trigger rate. The CBCs connect to the FE Hybrid via approximately 800 bumps on a 250 μm pitch. They are flipped and soldered ("flip-chip" bonded) to the FE Hybrid during the same reflow step as all other passive components on the circuit. CBCs exchange data with their neighbours to identify clusters spreading over chip boundaries, and with the CIC when sending out trigger data and L1 accept readout data. The data flow, shown schematically in Fig. 2.13 (right), separates L1 readout (DAQ) from Trigger (TRIG) paths. Assuming a data transfer rate of 160 Mb/s (other rates are currently under study), each CBC chip sends 2 bits of DAQ and 10 bits of TRIG data to the CIC every 6.2 ns. This bandwidth is compatible with transferring up to 3 stubs from each CBC every BX, and sending unsparsified readout data from each CBC pipeline up to a 1 MHz L1 rate.

The CIC buffers, sparsifies and aggregates the data from all CBCs and funnels them to the Service Hybrid that provides a limited 1.6 Gb/s bandwidth to each module end, i.e. 10 bits at 160 Mb/s per FE Hybrid. This limited bandwidth is shared by eight CBCs served by one CIC, making best possible use of the statistical distribution of hits in time and space. The CBC L1 readout data are sparsified in the CIC, resulting in 2 bits at 160 Mb/s being needed for the DAQ channel out of the CIC. Sparsification implies however that the DAQ channels of

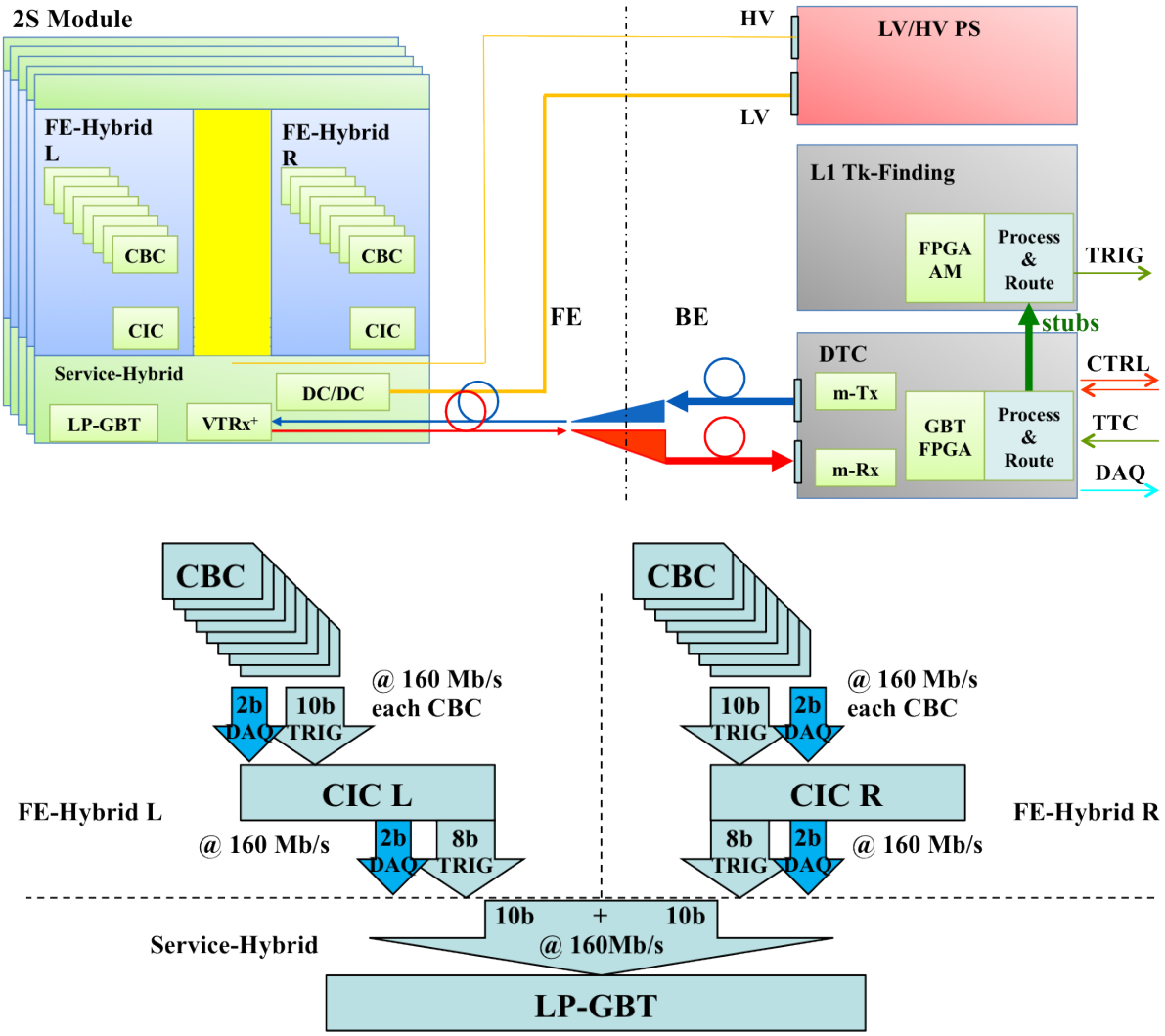


Figure 2.13: Electronic system block diagram (top), and Front-End data flow (bottom). Both are shown for the case of 2S modules.

different modules will run asynchronously. Synchronous unsparsified operation will remain possible by selecting the corresponding CIC mode, but only up to a 100 kHz L1 rate. Trigger data are sent out of the CIC in block synchronous mode, aggregated over a fixed number of bunch crossings. Each block is 8 BXs long and synchronous to all other CIC blocks in the system. Up to a maximum of 12 stubs produced by any of the CBC on the FE Hybrid during this 8 BX window are stored in the data block. Additional stubs are discarded. This block transfer scheme makes it possible to redistribute the limited bandwidth in time (i.e. over 8 consecutive 25 ns BXs, for a total of 256 bits), while limiting the latency of the TRIG channel.

Preliminary simulations indicate that this approach will be at least 99% efficient for the nominal high pileup conditions of 140 collisions per bunch crossing. As mentioned above, some level of configurability will be included in the CIC so as to make the system adaptable to different operation conditions.

The two FE Hybrids on the 2S module are estimated to draw a total power of about 2.2 W from the low-voltage supply rails.

2.3.2.3 Front-End Electronics for PS modules

Since the PS module is constructed using one pixelated and one strip sensor, two front-end chips (in 65 nm CMOS) must be developed: the Short Strip ASIC (SSA) and the Macro-Pixel ASIC (MPA). The correlation of strip and macro-pixel sensor hits takes place in the MPA. The SSA, much like the CBC in the 2S module case, sits on the FE Hybrid. It processes the sensor signals, and sends unsparsified strip data to the corresponding MPA, at BX frequency. The MPA is bump-bonded to the macro-pixel sensor (Fig. 2.16). Due to the large pixel size, a standard bump pitch of 200 μm can be used, relaxing the assembly requirements. Two rows of eight MPAs are bumped to each macro-pixel sensor, resulting in an assembly of approximately 32000 pixels per module (2000 bumps per MPA). The MPA processes and sparsifies the signals from each pixel. It correlates the bottom macro-pixel sensor hits with the data received from the SSA strips and builds clusters and stubs. Connections to/from the FE Hybrid are made with wire bonds. As in the 2S module case, a CIC located on the FE Hybrid buffers and aggregates the stub and cluster data received from the MPA, and sends them to the Service Hybrid. The transfer scheme and data formats are very similar to those used in the 2S module, allowing the same CIC to be used for both module types.

Assuming a data transfer rate of 160 Mb/s (other rates are currently under study), each MPA sends 2 bits of DAQ data and 10 bits of TRIG data to the CIC every 6.2 ns. From the CIC, 2 bits of DAQ and 8 bits of TRIG data are sent to the LP-GBT. This bandwidth is compatible with transferring to the CIC up to 4 stubs from each MPA every 2 BX, and sending cluster data up to a 750 kHz L1 rate with good margin. However, the limited bandwidth available from the LP-GBT constrains the amount of data that the CIC can pass on to the Service Hybrid. Preliminary simulation results (at 140 average number of collisions per bunch crossing) indicate that DAQ data can be transferred efficiently to the backend for L1 rates up to 750 kHz, with a gradual degradation of performance above this rate (occasional buffer overflows), especially for the first barrel layer. In the second and third barrel layers, TRIG stubs can be transferred with full efficiency and ample headroom in the available bandwidth; in the first layer the bandwidth budget is tight, but additional margin can be gained, if required, by increasing the p_{T} threshold from 2 GeV to about 3 GeV (by narrowing the correlation window width).

The binary readout of the upgraded Outer Tracker does not provide information on the amount of charge deposited in the sensors, leaving only the four pixel layers to discriminate minimal ionizing particles from particles with higher specific ionization. An option under study to recover sufficient sensitivity to identify particle with high ionization (which is useful for some searches of physics beyond the standard model, e.g. Section 10.4.3) consists in implementing a second, higher threshold in the SSA chip, and add one more bit to the data stream recording whether the detected signal was above or below that higher threshold. Preliminary studies indicate that such functionality can be implemented in the chip with limited added complexity, and would provide the required performance.

The FE electronics of PS modules is expected to draw between 4 W and 5.4 W of power from the LV rails, where the range reflects the uncertainties in the current estimates of power dissipation by the MPA and SSA.

2.3.2.4 Front-End Services

Each CMS Tracker module is connected to its own set of services and is thus an autonomous element in the system. The Service Hybrid hosts all of the following connections and associated electronic devices: (i) the converters generating the necessary voltages for the FE electronics (1.2 V) and opto-electronics (2.5 V); (ii) a miniature HV connector for the sensor bias lines; (iii)

the LP-GBT chip serializing (deserializing) data to (from) the backend, controlling the FE ASICs via I2C, receiving and distributing clock and trigger signals and providing local monitoring capability; and (iv) the VTRx+ optoelectronic transceiver converting all upstream (downstream) data to optical (electrical) signals. All existing services connecting the detector to the backend electronics located on the balconies of the experimental cavern or in the service cavern will be replaced by new LV, HV electrical cables, and multi-mode fiber-optic cables.

The DC/DC converter, LP-GBT and VTRx+ devices are all derived from common developments for LHC experiment upgrades. More details about these components can be found in the presentations at the ACES 2014 workshop [35]. The performance level of these developments is currently well matched to the Phase-I upgrade needs, but will need to progress further to comply with the high density, low power dissipation and small physical volume requirements of the Phase-II Tracker. The CMS Tracker relies heavily on the continuation of these common projects into Phase-II and on their successful completion.

The same flex hybrid technology used for the FE Hybrids is foreseen to be used for the Service Hybrid, with wire-bonded connections to the FE Hybrid. Due to the factor of two difference in length, the Service Hybrid for the PS module must be split into two separate circuits for powering and data transfer, implemented on each module side.

The required Service Hybrid power levels are 4 W (6.4 W to 8.3 W) from the LV supply rails for 2S (PS) modules, assuming a 75% DC/DC conversion efficiency, and 1 W from the HV supply rails, at the end of life of the innermost sensors.

2.3.2.5 Back-End Electronics

Development of the Back-End (BE) electronics is at an early stage since most components will be based on commercial off-the-shelf parts and limited environmental qualification is needed, as they are located in the service cavern, or, in the worst case, on the balconies in the experimental cavern. Moreover, significant technological evolution can be expected in the coming years, from which CMS can obtain full benefit with a timely development schedule. Therefore only a preliminary design, sufficient to establish a realistic cost scale, is presented here.

The BE powering and data I/O functions that can be considered as evolving from the current Tracker are briefly described below. The concepts underlying the new L1 track finding functionality are also explained.

Powering. The existing Tracker power supply bank will be entirely replaced. Based on estimates presented in Section 2.3.2.4 above, the low voltage supply system will need to provide about 34 kW (45 kW to 59 kW) of FE power for 2S (PS) modules at a voltage of $10 - 12$ V. The HV system will need to reach -800 V of sensor bias voltage, with close to 1 mA of current per module at end of life in the most exposed areas of the Tracker. Ideally, both LV and HV distribution systems will have module-level granularity that fits the autonomous module building-block concept and avoids complicated failure mitigation schemes.

Data Input/Output. All data transfer to and from each front-end module is collected into one fiber pair as required for the GBT protocol. Multi-fiber cables (typically $96 - 144$ fibers) connect to the DTC board via high density optical connectors ($12 - 72$ fibers). The fibers are then fanned out to multi-channel optoelectronic transceiver engines placed close to the processing FPGAs that code/decode GBT data carrying TTC, Control, DAQ and Trigger information. The DTC board implements both readout and control functionalities, and in addition provides TRIG data to the L1 track finding system.

The hardware format of the DTC board is not yet defined. It could follow the AMC-μTCA format standardized by CMS for its Phase-I upgrade program, but may instead take the form of an 8U-ATCA board if this matches better the infrastructure at hand for the DAQ or L1 track finding systems. In fact, most of the data processed by the DTC board is trigger data for L1 track finding, and its hardware will thus most likely follow the format used for the L1 trigger.

L1 Track Finding. The DTC board (Fig. 2.13, left) sends a stream of stubs selected by the FE ASICs correlation logic to the L1 track finding system. Each bunch crossing produces on the order of 10,000 stubs. Only about 5 to 10% of these stubs actually belong to primary tracks with $p_T > 2$ GeV (approximately 125 per bunch crossing). The remainder are random combinations of hits, or are due to lower-p_T particles and secondary particles produced by interactions of primary particles with the material in the detector. The goal of the L1 Track Finding system is to perform pattern recognition to reconstruct the tracks of primary particles with $p_T > 2$ GeV, and discard as many as possible of all the other stubs.

In the past, hadron collider experiments have successfully made use of information from silicon-based trackers in L2 triggers based on Associative Memories (AM) [56]. However, in those applications the event rate was much lower and the latency much longer than foreseen in the CMS case. Similarly, FPGA-based applications have been used in L1 trigger applications, but at much lower channel count and precision. Thus, the challenge of developing a L1 Track Finding system for CMS is still to be met. The project will benefit from the rapid technological progress expected in the areas of FPGA processing power and data link bandwidth in coming years. Keeping up with the high stub rate will require the implementation of both time and regional multiplexing to enable parallel processing of data from different bunch crossings and detector regions. Note that the multiplexing ratios proposed below are only indicative and are likely to change during the course of the development program.

Three L1 Track Finding approaches are being pursued by different teams with the goal of demonstrating at least one viable option for the Technical Design Report.

The Associative Memory (AM) approach makes use of a massively parallel architecture to quickly tackle the intrinsically complex combinatorics of track finding algorithms, avoiding the typical power law dependence of execution time on occupancy. The time required to perform pattern recognition is approximately linearly proportional to the number of hits. In this approach, the Tracker is divided into 48 angular regions (regional multiplexing) called trigger towers (6 in η by 8 in ϕ). Approximately 200 stubs per bunch crossing are expected in each trigger tower at 140 pileup, to be matched against about 2 million reference track patterns stored in custom-designed AM chips (an estimated 100 million reference patterns are needed for the full Tracker). Multiple processing engines are assigned to each tower, resulting in typical time multiplexing ratios of 20 to 40. Efficient data dispatching for time and regional multiplexing is achieved by using a full-mesh backplane ATCA platform, which provides high bandwidth, low latency, and flexible real time communication among processing nodes. One ATCA crate will be required per trigger tower. A custom ATCA board (Pulsar II) based on a Xilinx Virtex-7 FPGA has been prototyped with the goal of creating a scalable platform to demonstrate the AM-based approach. In addition, a pattern recognition mezzanine card will be developed, to host a powerful FPGA and a new custom-designed AM chip.

The second development effort on L1 Track Finding uses an approach in which FPGAs are used to implement a conventional road-based track search. Tracks are seeded by finding pairs of stubs in neighbouring layers (tracklets). These tracklets are then projected to other layers to search for matching stubs. The track parameters are obtained using a linearized χ^2 fit. The Tracker stub processing load is shared among 168 non-projective trigger towers (6 in z, and

28 in ϕ), and time multiplexed by a factor of four in all ϕ slices. This approach has been conceptually demonstrated in software and does not rely on the development of dedicated ASICs. In order for it to be translated into hardware, however, it requires the availability of a large number of memories interconnected with powerful Digital Signal Processing nodes in FPGAs.

The third approach under study is a Time Multiplexed Trigger (TMT) design similar to the one which is currently being implemented in the CMS calorimeter trigger as part of the Phase-I trigger upgrade [7]. In the TMT approach multiple sources send data to a single destination for complete event processing as is currently done in the CMS HLT. It requires two layers with a passive switching network between them, implemented in the form of an optical-fiber network. By assembling all data from an individual event in each bunch crossing into a single processor the TMT avoids replication of data shared across regional boundaries within the detector. In practice however, for a system the size of the CMS Tracker it is impossible at present to avoid subdividing the detector into a small number of regions. It is proposed to limit this division to 5 sectors in ϕ, with a time multiplexing ratio of 24. A μTCA processing board based on a Xilinx Virtex-7 FPGA is in production for the TMT L1 Calorimeter Trigger. Its capabilities are well matched to what is needed to demonstrate feasibility of a high-throughput processing application such as L1 Track Finding.

Vertical slice demonstration test benches are being assembled in several labs to demonstrate track finding using the technology and hardware available today. They will process simulated data with HL-LHC occupancy and rate, and will allow system performance evaluations (such as latency, efficiency and fake rate) and thus comparisons of different technology or component choices. The goal is to demonstrate in the next two years that the L1 Track Finding can be achieved, to compare the different approaches being proposed so that the system design can be better understood and optimized, and to allow a cost-benefit analysis to be completed.

2.3.2.6 System Demonstrator and Near Term Development Plans

The development of electronics for the 2S module has already completed several prototype cycles. The CBC chip family is a prime example: the CBC1 chip, delivered in 2011, demonstrated a noise of 1000 electrons for a 5 cm-long strip, with 0.3 mW per channel power dissipation in 130 nm CMOS technology [57]. The CBC2 chip, delivered in 2013, demonstrated dual sensor readout capability (2×127 channels), correlation logic functionality, and C4 bump-bonding connectivity (250 μm pitch) [58].

In mid-2013, CBC2 chips were for the first time flip-chipped to 2×CBC2 high-density Hybrid circuit prototypes [59], demonstrating the feasibility of bump-bonding FE chips onto hybrid substrates, and later allowing dual-sensor mini modules to be successfully assembled and tested (Fig. 2.14 left), thereby validating flipped-chip bonding of the FE CBC ASIC and production of the high-density FE hybrid. By the end of 2013, a test beam at DESY successfully demonstrated a full readout chain [60] based on the GLIB multi-purpose AMC board [61], effectively validating many of the concepts proposed for the Phase-II system including correlation of hits from dual sensor modules, optical data transfer using the GBT protocol, μTCA BE electronics, and integration into full CMS X-DAQ framework, among other things.

This system-level demonstration gives us confidence that the proposed concepts are sound and can be used to scale the system up and/or to develop prototype PS modules. Prototyping plans for 2014-15 include: a full-size 8×CBC2 flex FE Hybrid and full readout chain for 2S modules; six small-size MPA prototypes ("MPA-Light") in 65 nm CMOS technology, consisting of 16 rows of 3 pixels, to be bump-bonded onto a pixelated sensor as a first step towards a PS module prototype ("MaPSA-Light" assembly, see Fig. 2.14 right); development of a fully func-

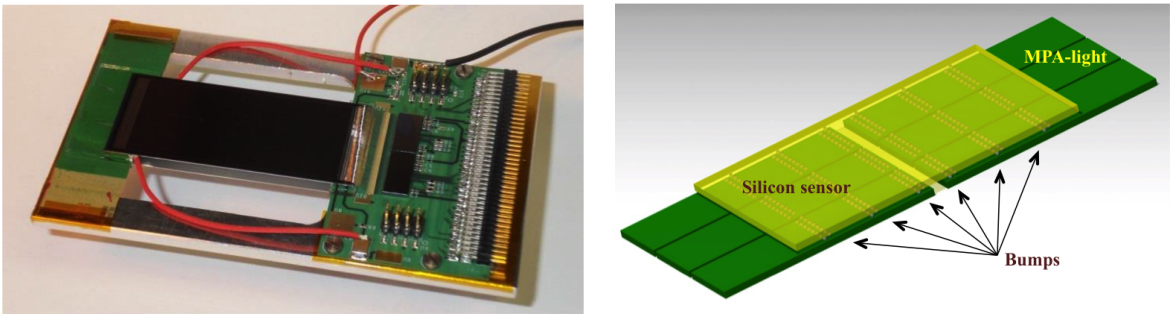


Figure 2.14: Dual sensor, mini 2S module, based on 2 CBC2 chips (left). CAD view of MaPSA-Light assembly prototype (right).

tional Service Hybrid demonstrator based on a DC/DC converter and optoelectronic devices developed for Phase-I.

2.3.3 Module design

As discussed above, the Phase-II Tracker will utilize two types of p_{T} modules, 2S modules and PS modules. To achieve efficient rejection of low-p_{T} particles throughout the Tracker volume, modules in different regions will make use of a few different sensor spacings. For 2S (PS) modules, spacings of 1.8 and 4 mm (1.6, 2.6 and 4 mm) are foreseen. These modules will be used in the end-cap disks as well as the central barrel region of the Tracker.

The design of the modules should provide for efficient removal of the heat generated by the electronics and sensors, accurate geometrical positioning, minimal mass, as well as a simple and reproducible assembly procedure. For the thermal performance, the design requirement is to achieve a sensor temperature of -20°C or lower with a coolant temperature of -30°C for modules irradiated with the full HL-LHC integrated luminosity. No explicit margin is required in the thermal performance of the modules because the cooling system will be designed for a nominal operating temperature of -30°C, but with the capability of reaching -40°C, to provide a comfortable margin with respect to the design specifications of the overall system.

Despite the different sizes and functionalities of 2S and PS modules, several common solutions are used in both. The sensors are glued to carbon-fiber reinforced aluminium (Al-CF) elements that serve as supports and spacer while also playing a key role in the thermal management of the module. The Al-CF material has low density, high thermal conductivity, and most importantly a low coefficient of thermal expansion (CTE) of 4 ppm/K, which, in combination with the symmetric design of the sensor sub-assembly, results in minimal stress and deformation of the sensors when operated at low temperatures. Since Al-CF is electrically conductive, all surfaces of the supports will be electrically insulated by means of parylene coating.

The sensor sub-assembly is then combined with the module electronics as outlined below for the two module types. The development of the designs has been guided by detailed Finite Element Analysis (FEA) of their thermal performance. The concepts chosen enable the different sensor spacings to be incorporated with minimal or no impact on the design of the active components.

2.3.3.1 Design of the 2S Module

An exploded CAD view of the 2S module is shown in Fig. 2.15, for the version with 4.0 mm sensor spacing. The sensors are glued to two long Al-CF spacers. The ends of the spacers,

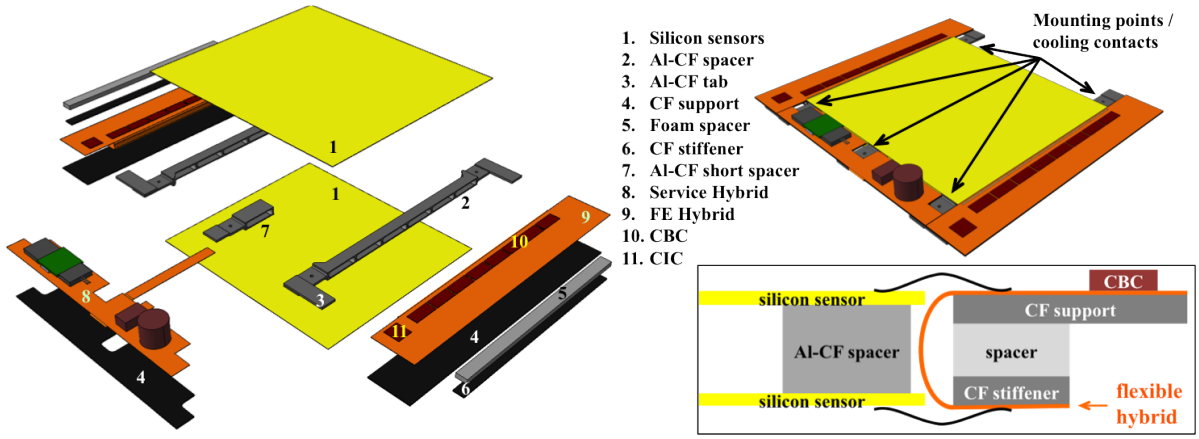


Figure 2.15: Exploded view of the 2S module components (left), 3D view of the assembled module (upper right), and a sketch of the FE Hybrid folded around its support (lower right). The module is mounted on five cooling elements on the mechanical support structure.

which extend beyond the sensors, house the mounting holes of the module, and additional Al-CF tabs that provide support and cooling contacts to the FE Hybrids and the Service Hybrid. All hybrids use flex circuits that are glued onto carbon-fiber supports (see Section 2.3.2). As shown in the sketch of Fig. 2.15, the FE Hybrid is folded around a foam spacer enabling wire-bond connections from both sensors to the FE Hybrid. The FE Hybrid support is designed to give minimal clearance between the Hybrid and the sensors, to facilitate the encapsulation of the bond wires.

Biasing of the sensors is provided by flexible tails that are glued and wire bonded to the back side of each sensor. These tails extend out from the sensors and are plugged into miniature connectors on the bottom of the Service Hybrid. Due to the power dissipation of the active components on the Service Hybrid, an extra Al-CF short spacer is placed between the sensors on the Service Hybrid side of the module to optimize the cooling performance. The 2S module is therefore mounted on five cooling blocks, of which two are used for precise positioning of the module on the support structure (see Section 2.3.4).

2.3.3.2 Design of the PS Module

The PS module has several specific features that impact the module design. The pixelated sensor has the readout chips bump-bonded onto it, covering the entire sensor surface. The sensor and chips will be bump-bonded first, and then handled as a single component in the module assembly. The pixelated sensor plus chips has a large heat dissipation, spread over the entire sensor surface. Therefore, unlike the 2S module, the PS module cannot be efficiently cooled through a few small-area contacts. As shown in Fig. 2.16, a carbon fiber reinforced polymer (CFRP) base plate is employed as a carrier for all module components and as a more robust interface to the cooling structure. The sensor sub-assembly is attached to this base plate via a large area glue joint to provide the thermal path for the heat dissipated in the sensors and the MPA chips.

The sketch of Fig. 2.16 shows a schematic detail of the flex FE Hybrid. The Hybrid is supported by CFRP strips and is folded around Al-CF spacers in order to bring the surface of the Hybrids with the bond pads to the same level as the bond pads on the strip sensor and on the periphery of the MPAs. The Al-CF spacer with the attached FE Hybrids are glued to the tabs of the base plate.

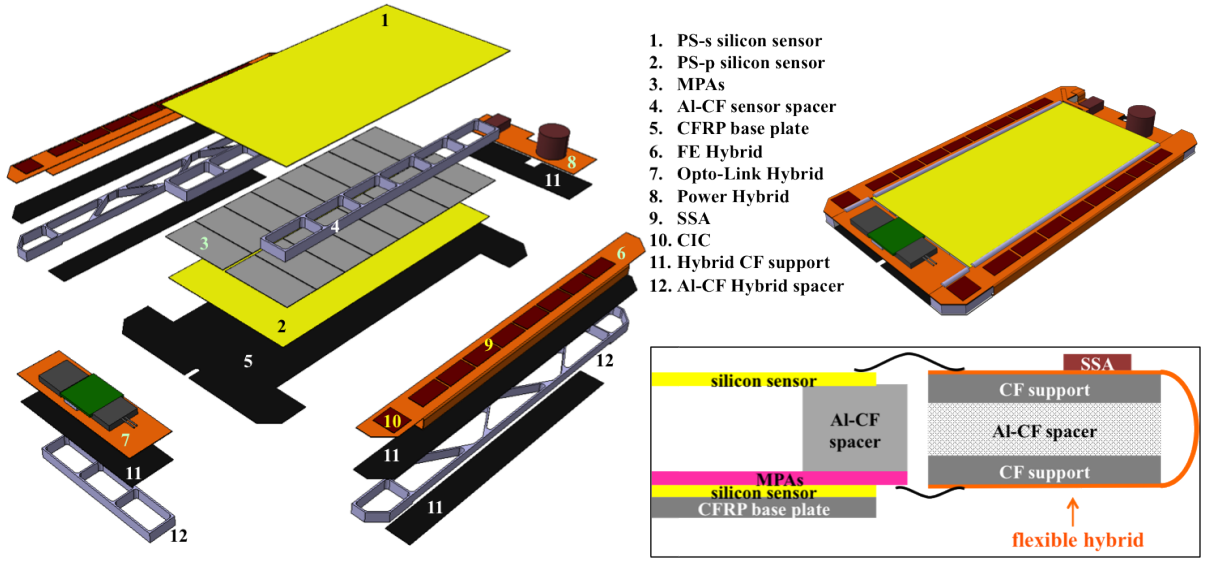


Figure 2.16: Exploded view of the PS module components (left), 3D view of the assembled module(upper right), sketch of the FE Hybrid folded around its support (lower right). The base plate is glued onto a flat surface on the support structure that is kept cold during operation.

As explained in Section 2.3.2.4, the powering and readout link functions of the PS module are implemented in two small flex circuits, located on opposite sides of the sensors. The Power Hybrid is glued directly to the surface of the base plate in order to maximize the efficiency of the cooling contact, while the Optical Link Hybrid is mounted on a spacer in order to raise its surface to the level of the FE Hybrids to facilitate wire bonding. The power for the FE Hybrids and the bias voltage are distributed from the Power Hybrid via flexible tails.

The CFRP base plate is equipped with two holes that are used to precisely position the module on the support and cooling structure. The low-impedance thermal interface between the module and the cooling structure is established via phase-change thermal interface glues that allow for the dismounting of modules for repair or replacement prior to the final installation in the Tracker.

2.3.3.3 Thermal performance and mass estimates

The development of both module designs has been guided by finite element analysis (FEA) of thermal performance and mechanical deformations that occur when they are cooled to low temperatures. Such studies have guided several basic aspects of the module assembly and cooling concepts including the choice of material and the sizes of the thermal contacts.

The FEA models include the module geometry, its support and cooling structure, as well as a realistic description of the power dissipation of the front-end electronics and the sensors. The sensor power dissipation is estimated for each module type and version (2S or PS module, and specific sensor spacing), based upon the expected maximum fluence at the different locations in the Tracker volume (see Fig. 2.3). A summary of the estimated power dissipation in the sensors after an irradiation corresponding to 3000 fb^{-1} of integrated luminosity and typical power consumption values for the othe FE components are collected in Table 2.2 (for PS modules the chosen values are at the high end of the ranges quoted in Section 2.3.2.3). The heat generated by the power converter is calculated from the sum of the power consumed by the front-end electronics under the assumption that the power converter operates at an efficiency of 75%,

Table 2.2: Left: summary of the sensor power dissipation at 600 V bias voltage for a sensor temperature of $-20°C$ after an irradiation corresponding to an integrated luminosity of 3000 fb^{-1}. Right: power dissipation of the front-end electronics used in the finite element analysis of 2S and PS modules.

Sensor power		
Module	Spacing [mm]	Power [mW]
PS	1.6	360
PS	2.6	680
PS	4.0	480
2S	1.8	370
2S	4.0	570

Front-End power			
#	Component	Module	Power [mW]
2	CIC	2S & PS	400
1	Opto Package	2S & PS	800
2×8	MPA	PS	4000
2×8	SSA	PS	1000
2×8	CBC	2S	1800

Table 2.3: Mass estimates for the 4.0 mm variants of both module designs, excluding the opto package, power converter and passive components. The sensors are assumed to be 200 μm thick.

Component	2S Module [g]	PS Module [g]
Sensors	9.0	4.5
Hybrids	1.7	1.9
ASICs	0.9	2.7
Al-CF	5.8	7.4
CFRP	6.8	4.4
Misc. (glues, coatings, etc.)	1.0	2.3
Total	25	23

which results in a dissipation of 1.0 W for the 2S module and 2.0 W for the PS module, because of the higher front-end power.

The thermal performance of the designs is characterized in terms of the coolant temperature necessary to reach the sensor operation temperature of $-20°C$, and the difference between the nominal CO_2 operation temperature of $-30°C$ and the coolant temperature for which the sensors would go into thermal runaway. For the 4.0 mm (1.8 mm) 2S module, the operation point is reached at a coolant temperature of $-27.3°C$ ($-27°C$) with a margin of approximately $4.0°C$ ($7.6°C$) to thermal runaway.

For the PS module, with a structure with two cooling pipes routed parallel to the strip direction and a heat spreader that covers the full area under the base plate, the operation point for the 2.6 mm version is reached at $-25.6°C$ coolant temperature with a margin of $11.7°C$ to thermal runaway.

Table 2.3 shows a mass breakdown for the 4.0 mm variants of the 2S and PS modules excluding the mass of the optical link, the power converter and passive components. The Al-CF spacers contribute significantly to the overall mass of the modules. An optimization of these parts will allow minimization of the mass while ensuring the required thermal performance.

2.3.3.4 Conclusions and outlook

The designs of both modules appear to be effective in providing adequate support and cooling to sensors and electronics in a lightweight assembly, sufficiently simple and modular to be produced in large quantities. The designs also offer good flexibility to improve the cooling

performance in order to accommodate possible changes in the power of the final electronics without having to modify the assembly concepts. The next stages of development will involve laboratory tests of full-size thermal mock-ups to validate the results of the FEA and to optimize size and shape of the different module parts, also taking into account ease-of-production and cost aspects.

2.3.4 Support structures and cooling

The mechanical concept of the Phase-II Tracker is the same as for the current Tracker. The Tracker is constructed and assembled in surface facilities and installed into CMS in two units: the Outer Tracker and the Pixel detector. The structural backbone of the Outer Tracker is a support tube at the outermost radius of the Tracker and spanning both ends. The Pixel detector is supported from the Outer Tracker, and it is installed inside CMS with the central beam-pipe in place. The Pixel detector can be removed for maintenance or replacement during Extended Technical Stops, whilst the Outer Tracker is expected to remain in place for the duration of its operational lifetime. For improved radiation resistance the Tracker will be operated and maintained at -20°C or lower using two-phase CO_2 cooling. All cold volumes are sealed from surrounding air moisture and continuously flushed with dry air or nitrogen.

2.3.4.1 The Tracker Support Tube (TST)

The design of the TST is the same as in the current Tracker. The TST is 5.4 m long and 2.4 m in diameter. The 30 mm-thick wall is a sandwich construction with carbon-fiber/epoxy face sheets and a honeycomb core. Metallic reinforcements at the tube ends are used to attach the tube to supports and to hold the end bulkheads. Rails at the horizontal plane inside the TST support the Outer Tracker barrels and end-cap disks. A thermal screen is attached to the inner wall of the TST.

2.3.4.2 The Tracker Barrel with 2S modules (TB2S)

The TB2S mechanics reuses proven solutions from the current Tracker, adapted to the new layout and to the new type of modules and electronics. The TB2S has 4464 detector modules of type 2S. The modules are mounted onto 372 ladders, each containing 12 modules. The ladders are installed on 3 layers in a support wheel that consists of 4 vertical disks joined by cylinders at the inner and outer radii. The ladders, which are half the length of the wheel, are installed from the two ends of the barrel and have a small overlap at the center to avoid gaps in coverage (Fig. 2.17).

The ladders are made of two carbon-fiber C-shaped profiles, joined by several orthogonal elements, also made of carbon-fiber. The modules are cooled via a cooling pipe that transits the full length of the ladder and back, forming a U shaped circuit. Wires and optical fibers are routed inside the profiles of the ladders.

2.3.4.3 The Tracker Barrel with PS modules (TBPS)

The TBPS has 4130 PS type modules on 3 layers. For efficient cooling the modules are glued directly onto 74 flat, cooled support plates. Each plate is a sandwich structure made of two high-conductivity carbon-fiber skins and a foam core which embeds a U-shaped cooling pipe. Wires and fibers are routed along the edge of the plate. To form the complete cylindrical layers the plates are joined and supported by vertical rings (Fig. 2.18, upper row). The complete TBPS is supported from rails attached to the inner surface of the TB2S wheel.

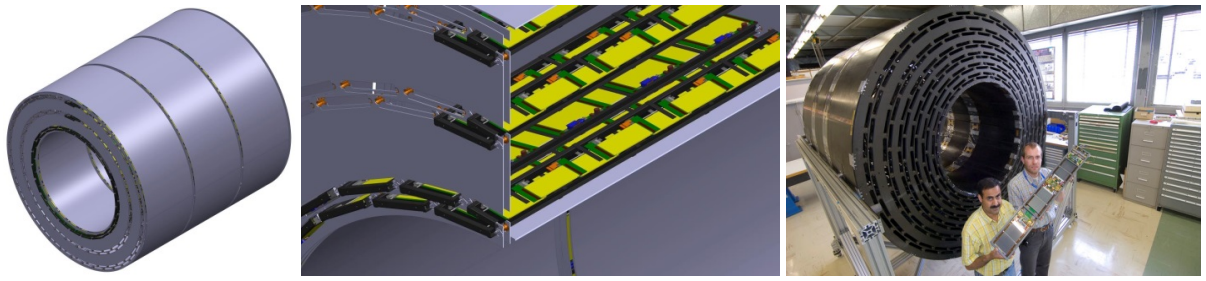


Figure 2.17: The TB2S wheel (left), composed by three inner and three outer cylinders joined by four disks, supports 372 ladder units, each with 12 detector modules (center). The same concept, with 6 layers instead of the 3 layers, was already used in the current Tracker Outer Barrel (right).

Figure 2.18: Upper row: sketches illustrating the baseline TBPS design; modules are mounted on support plates, that are joined to form three concentric layers. Lower row: sketches illustrating the alternative TBPS geometry; in the parts with tilted modules, support is provided by rings with embedded cooling pipes.

Alternative TBPS geometry. To reduce the number of modules while maintaining the same geometrical coverage, an alternative BPS layout with tilted geometry is being studied. In this layout the modules closest to the center are parallel to the beam while while those nearer to the ends of the barrel are at tilt angles ranging from 35 to 75 degrees. The short flat barrel part uses the same support Plate concept as the baseline TBPS design. The tilted modules are supported by rings that include circular cooling pipes. Several successive rings have the same tilt angle, thus limiting the number of different ring geometries to be produced (Fig. 2.18, lower row). Cooling manifolds, wires and optical fibers are routed along guide rails on the outer edge of the Rings.

2.3.4.4 The Tracker End-Cap Double Disks (TEDD)

The TEDD uses exactly the same rectangular 2S and PS modules that are used in the barrel sections (TB2S and TBPS) of the Outer Tracker. The modules are mounted on flat disks, which

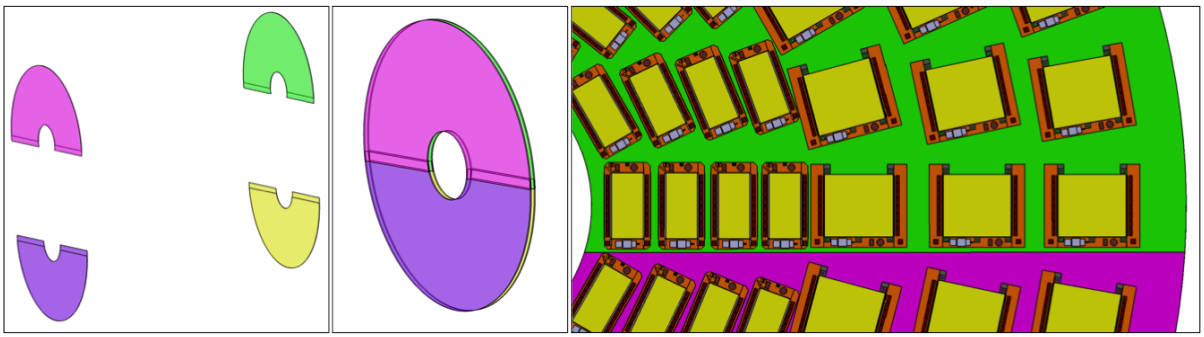


Figure 2.19: Four TEDD dee structures (left) are joined to form two disks and then a Double-Disk (center). Right: arrangement of modules on a disk surface.

for assembly reasons are split in half-disks, or "dees". Two disks are grouped to form one Double-Disk, which provides one hermetic detector plane (Fig. 2.19); ten Double-Disk units will be produced, five for each end-cap. The cooling pipes run inside the sandwich structure of each dee. PS modules, located in the low radius regions, are attached directly to the cooled front and back surfaces of the dee, while 2S modules, located in larger radius regions, are cooled via module support inserts which are connected to the cooling pipes. Wires and optical fibers are routed radially outwards on each dee, above and between the modules.

2.3.4.5 Cooling

The total power dissipated in the tracking volume is estimated at about 90 kW for the Outer Tracker and 20 kW for the Pixel detector. The cooling system must remove this heat load and maintain the silicon sensors at a temperature of $-20°C$ or lower. The system will be designed for a nominal coolant operating temperature of $-30°C$, but having the capability to lower the temperature to about $-40°C$ for additional margin.

The Tracker is cooled using two-phase CO_2. The CO_2 cooling plants, located in the service cavern (USC) provide a flow of CO_2 in liquid-phase. The flow is distributed through manifolds located in the experimental cavern (UXC) to about 50 cooling loops that serve different sections of the Tracker. The pressure of the CO_2 drops in capillaries at the entry of the Tracker, allowing the fluid to evaporate in the detector. To avoid risks of dry-out and loss of cooling, only about 50% of the fluid is evaporated. The return flow is a mixture of liquid and gaseous CO_2. This cooling system concept has been successfully employed in the LHCb VELO since 2008 [62], and is also the choice for the CMS Phase-I Pixel upgrade.

The CO_2 cooling plants currently in construction for the Phase-I Pixel detector have a total cooling capacity of 15 kW (Fig. 2.20). This can be increased up to 45 kW. The basic concept is to have several cooling plants that can backup one another. In case of a failure or during maintenance periods neighboring plants can take over the load. This is achieved by opening/closing supply lines between the cooling plants and the manifolds in the experimental cavern from which the final cooling circuits go to the detector.

A thermal screen with heating foils is used to maintain the periphery of the Tracker at the temperature of the surrounding detectors (expected to be $+18°C$ or lower), and all external surfaces above the maximum dew point of the cavern ($+13°C$). A $10-20$ mm thick layer of insulation is placed below the heating foils. No active cold panels are placed inside the insulation layer. The cooling system will always be in operation, at least at the level required to avoid incurring additional damage from radiation effects as a result of reverse annealing.

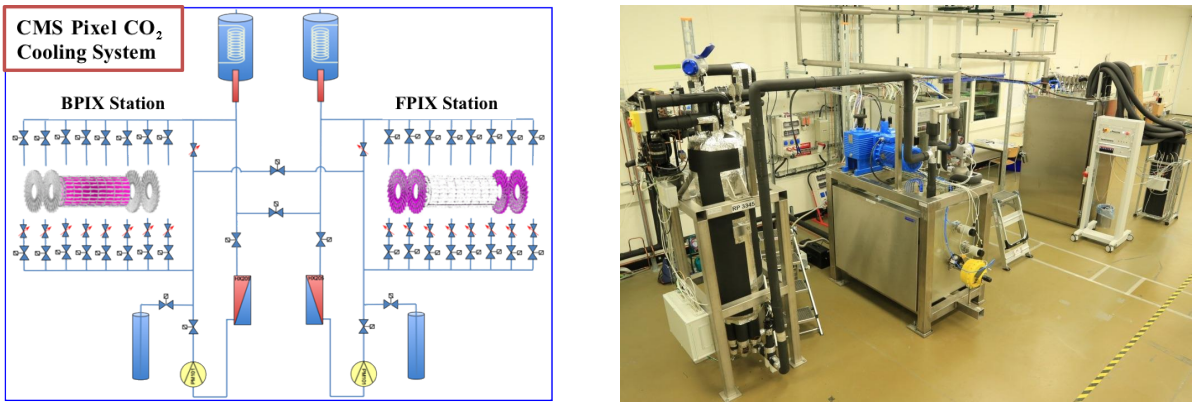


Figure 2.20: Schematic of the CMS Pixel Phase-I cooling system (left), and its prototype plant at CERN (right). A similar system, but larger in scale, will be used for the Phase-II Tracker.

Diesel-powered electrical supplies are used to ensure continuous operation even during power cuts. Particular care will be devoted to the engineering of the bulkheads and the feedthroughs for the service lines, in order to ensure a proper sealing of the cold volume that will be flushed continuously with nitrogen (or dry air when CMS is open) as it is done for the present detector.

2.4 Performance estimates

A full Monte Carlo simulation of LHC events with high pileup gives the best prediction of the detector performance and overall physics capabilities of the upgraded CMS, once the detector response is known and correctly simulated (including sensing elements, electronics and inactive material), and the reconstruction algorithms have been fully optimized for the new detector. In the initial design stage, when many aspects of the detector are undefined and a large phase-space of options needs to be explored, it is useful to complement the full simulation with simpler and more flexible modelling tools, to facilitate many aspects of the design optimization.

While the conceptual design of the Pixel detector is being finalized, the performance of the Phase-I detector in moderate pileup and before radiation damage is taken to be a minimal conservative target for the Phase-II detector to match or exceed in high-pileup conditions. To this end, the amount of material expected for the Phase-I Pixel detector is provisionally assumed to correspond to that of the Phase-II Pixel detector. Moreover, the Phase-II Pixel detector implemented in the simulation is identical to the Phase-I detector with 7 additional forward disks on each end, bringing the total to 10 disks per end, and the dynamic inefficiency is removed, to reflect the higher readout capabilities foreseen in the Phase-II ROC.

2.4.1 Detector standalone modeling: “tkLayout”

A dedicated software package “tkLayout” [63] has been developed to support the detector design effort. The software creates a 3D model of the Tracker from simple configuration files. Materials are assigned to the active and inactive volumes, and services are routed automatically, according to user-defined rules. Based on the geometry generated and on the coordinate resolution expected from each sensing surface, the tracking precision is calculated with multiple scattering effects taken into account via correlated errors between consecutive measurement points along the track. The expected precision on the track parameters is calculated both for “offline” tracks (i.e. including Pixel hits) and for tracks reconstructed at Level-1 (using only the

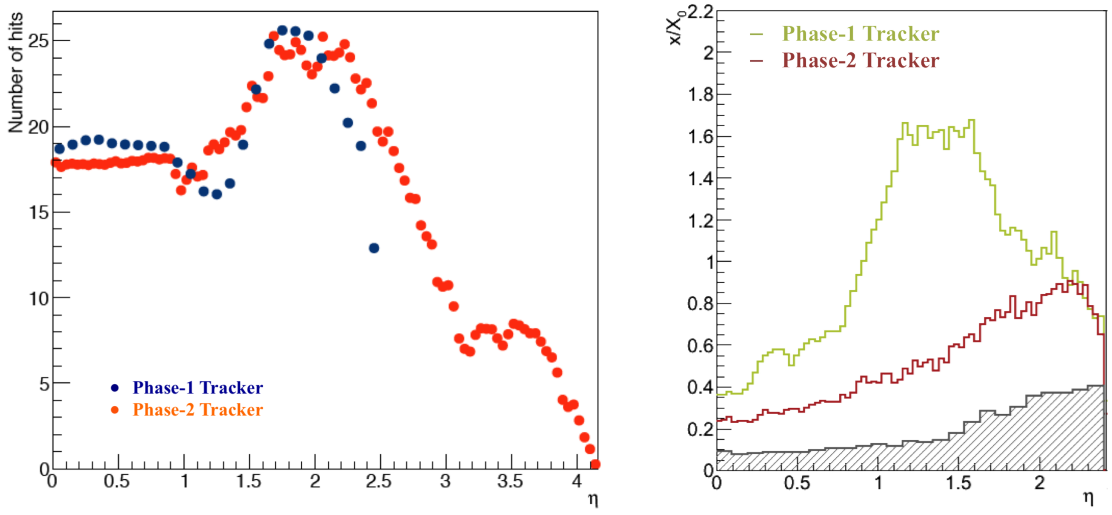


Figure 2.21: Number of hits (left) and radiation length (right) versus η for the Phase-II Tracker and the Phase-I Tracker. The radiation length distribution is shown for the tracking acceptance of the Phase-I Tracker, and reflects only the material inside the tracking volume; the expected contribution of the Phase-I pixel detector (hashed histogram) is provisionally used also for the Phase-II Tracker.

Outer Tracker high-p_T stubs). The tkLayout program provides summaries and statistics such as total number of modules, active surface, number of channels, power consumption, total weight, etc. The software also calculates the total radiation length and interaction length as a function of pseudorapidity, and hence the expected fraction of interacting particles and photon conversions.

The Phase-II Tracker offers a significantly extended tracking acceptance compared to the Phase-I Tracker, with slightly fewer hits per track over the common η range (see Fig. 2.21 left).

The inactive material inside the tracking volume is substantially reduced, as shown in the right plot of Fig. 2.21, where the distribution of radiation length versus η for the Phase-I Tracker is compared to that expected for the current model of the Phase-II Tracker.

For the current Tracker, the peak observed in the rapidity region $1.0 < \eta < 1.5$ has two main components: the printed circuit boards and the related electrical interconnections implementing the control electronics, that are located at the end of the barrel regions, and the conductors bringing the current to the FE electronics, that have a large cross section and are crossed five times in that rapidity range due to the specific detector layout, with a shorter Inner Barrel complemented by Inner Disks.

In the Phase-II Tracker there are no dedicated control electronics, since the control functionalities are integrated with the readout in a single optical data link; the cross section of the conductors for the Front-End powering is substantially reduced all the way to the individual modules, thanks to implementation of DC/DC converters on the Service Hybrids; moreover, all barrel layers have the same length. In the Phase-II Tracker the material of the modules is the main contributor to the total material of the detector, and the gradual increase in the amount of material up to $\eta \approx 2.3$ is mostly due to the inner layers of the TBPS, which are traversed at low incident angles by high-rapidity tracks emerging from the LHC collisions: this motivates the study of the tilted TBPS geometry, which could mitigate that effect.

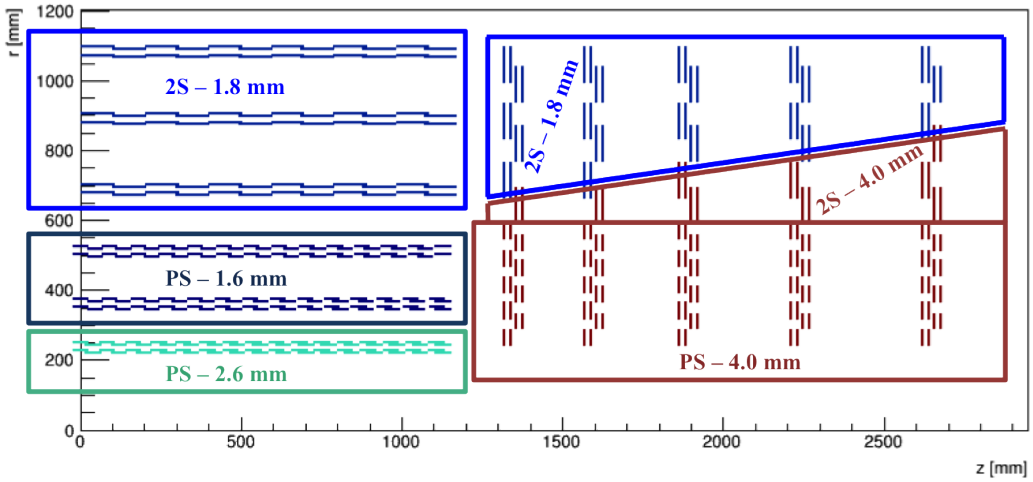


Figure 2.22: Map of the sensor spacing adopted in the Tracker. The sensor spacing is optimized at the same time as the acceptance window using tkLayout, to obtain a p_T threshold of 2 GeV for the stub selection in all module locations. The optimal acceptance window is then recalculated and fine-tuned using the Monte Carlo simulation.

The software also provides the map of expected leakage current in the sensors after irradiation, taking fluence maps generated with FLUKA [25]) as input. This map is used for the optimization of the module cooling (see Section 2.3.3). The software also calculates the p_T resolution within a module for the on-board stub selection, which (as explained in Section 2.1.4) depends on the module location and orientation. This facilitates the optimization of sensor spacings and acceptance windows in various regions of the Tracker (see Fig. 2.22).

The output of tkLayout is organized into a mini-website[9]. Besides a comprehensive set of plots and tables, tkLayout produces input files for the 3D modelling of the mechanics, and for use in the CMS full simulation. The software package has been developed so far mostly to support the design of the Outer Tracker where it has been used to optimize the detector layout and to provide the input geometry for the engineering design and the Monte Carlo simulation. The focus is now shifting towards the Pixel detector.

2.4.2 Stub Finding Performance

The performance of the stub finding capability of p_T modules is characterized in terms of the efficiency achieved for particles above the chosen threshold, and the rate reduction obtained, which enables the transmission of the stub data at 40 MHz with the available bandwidth.

In Fig. 2.23 the data reduction is demonstrated for all regions of the detector, using the sensor spacings indicated in Fig. 2.22, and acceptance windows tuned to efficiently select stubs from particles with p_T above 2 GeV. The stub rate remains by far the largest in barrel layer 1, generating high data volumes near to the limit of the available bandwidth. If needed, the acceptance window could be shrunk, effectively increasing the p_T threshold for that layer.

The resulting stub finding efficiency is shown in Fig. 2.24 (top) for muons in all the tracker detecting layers, demonstrating that the efficiency plateau is reached at 2 GeV in all locations, as intended. In the bottom left plot the efficiency is shown, for barrel layer 1, for muons, electrons and pions. The slightly lower efficiency for electrons and pions is due to interactions with detector material.

[9]See for example `http://mersi.web.cern.ch/mersi/layouts/.current/`

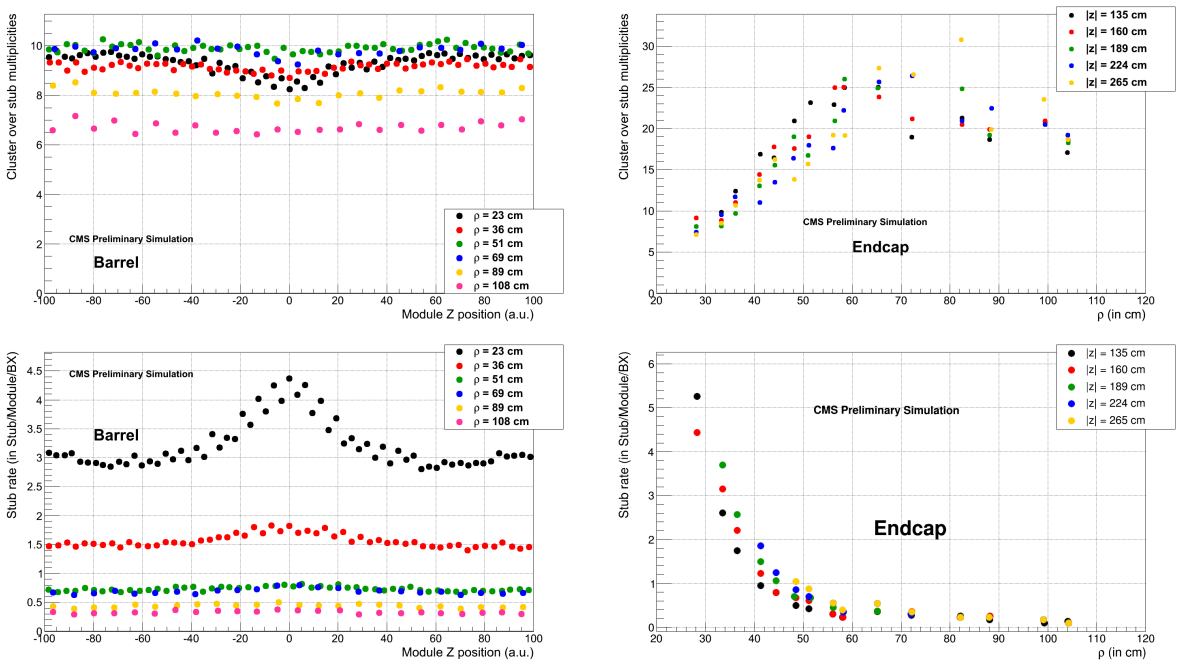


Figure 2.23: In the top row the cluster-to-stub multiplicity ratio is shown for barrel modules (left) and end-cap modules (right) as a function of z and radius, respectively, for a 140 PU minimum-bias sample processed through the CMS full simulation. In the bottom row the average stub multiplicity per module per bunch crossing is shown for barrel modules (left) and end-cap modules (right), in the same conditions. It should be noted that 2S modules (populating barrel layer $4-6$ and end-cap rings $10-15$) have $2\times$ larger surface than PS modules.

As mentioned in Section 2.3.2.6, the operating principle of p_{T}-modules has been demonstrated in a test beam at DESY, using mini 2S module prototypes (shown in Fig. 2.14). Different values of the local curvature of tracks due to different p_{T} have been emulated by changing the incident angle between the 2 GeV positron beam and the prototype p_{T}-module under test. The plot shown in the bottom right corner of Fig. 2.24 has been obtained by stepping through several incident angles and translating the value of the angle to the equivalent p_{T} at a radius of 75 cm. The resulting efficiency curve corresponds to a measured effective p_{T} threshold of 2.2 ± 0.1 GeV, which compares well with the nominal threshold of 2.14 GeV, as calculated from the module parameters (sensor spacing, acceptance window and strip pitch).

2.4.3 Level-1 track finding

The three approaches currently under consideration to process stub data in the back-end electronics and reconstruct tracks for the Level-1 trigger have been introduced in Section 2.3.2.5. For the AM-based track finding and for the "tracklet" pattern recognition, substantial effort has already been devoted to optimizing and qualifying the algorithms, thus deriving some preliminary estimates of the performance that could be achieved.

In the AM approach, coarse patterns ("roads") stored in the chips are filled with matching stub data: when a road has stubs in all detecting layers or all but one, it is taken as a track candidate. The full-resolution coordinates of the stubs associated with a track candidate are then used in a subsequent step to determine the helix parameters and to evaluate the quality of the fit. The road width is a key parameter in the optimization of this method: narrow roads lead to a larger number of patterns, and therefore a larger and more expensive system, while wide roads make

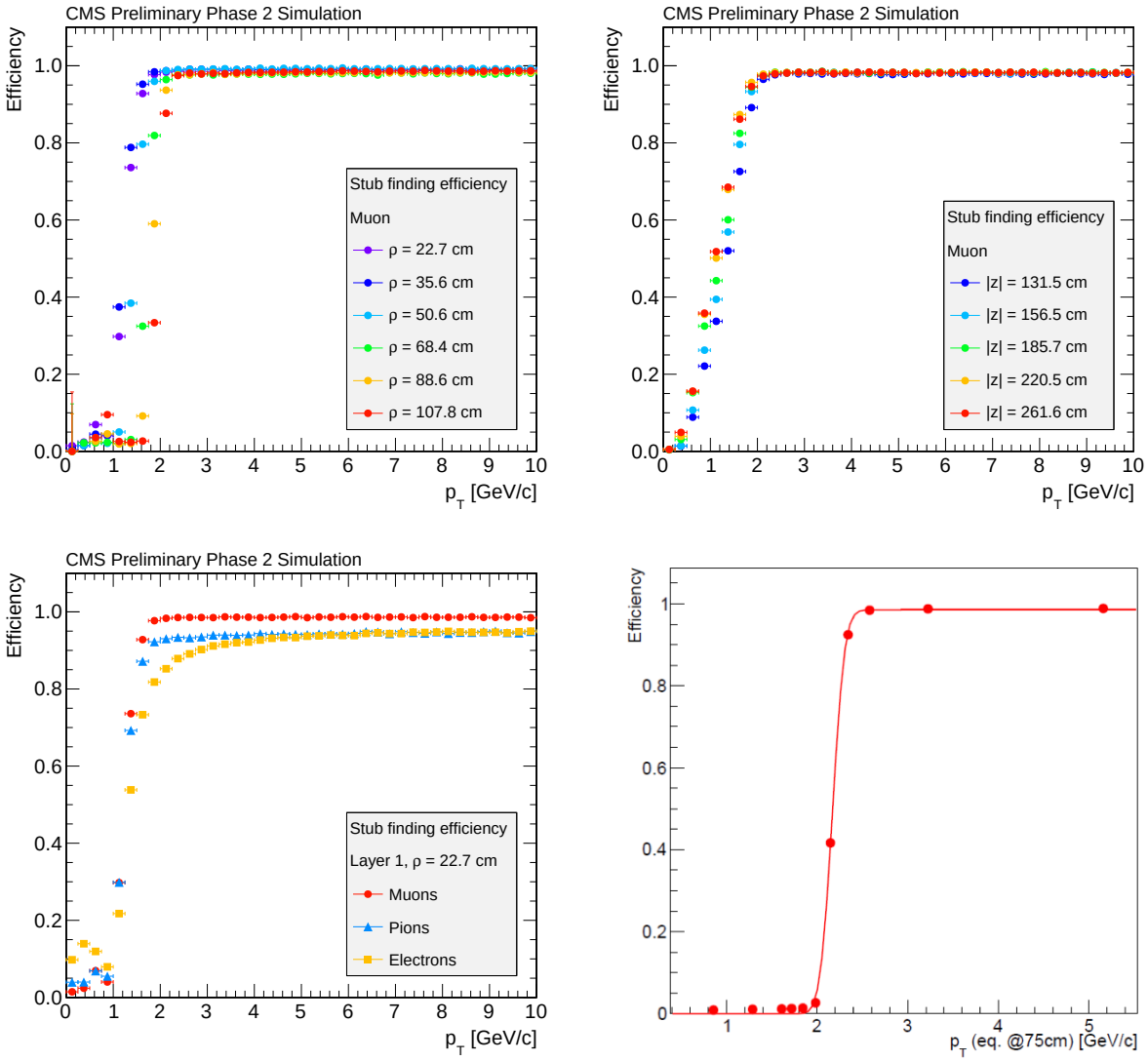


Figure 2.24: Upper row: stub finding efficiency for muons, in all barrel layers (left) and end-cap disks (right). The efficiency plateau is reached at p_{T} of 2 GeV in all detecting layers. Bottom left: efficiency in the barrel layer 1 for muons, electrons and pions. Bottom right: stub finding efficiency measured from real data taken on a test beam at DESY with mini 2S module prototypes; the angle between the beam and the module under test is translated into an equivalent p_{T} at a radius of 75 cm in the barrel as discussed in the text.

the fitting step more complicated, increasing the required processing power and latency. The track fitting algorithms being evaluated are: the Principal Component Analysis (PCA) algorithm originally developed for the CDF SVX trigger [64], the Hough transform algorithm [65], and the Retina algorithm [66].

The "tracklet" algorithm is a more conventional track finding method, in which compatible pairs of stubs in neighbouring layers are combined to form track seeds ("tracklets") and extrapolated to other layers to find matching stubs. Track finding and track fitting are part of the same process, and a subsequent step eliminates duplicates. The performance for single particles is evaluated from samples of single muons, pions and electrons, uniformly distributed in ϕ, η and p_{T}, overlaid with an average of 140 PU events. The track finding efficiency is shown

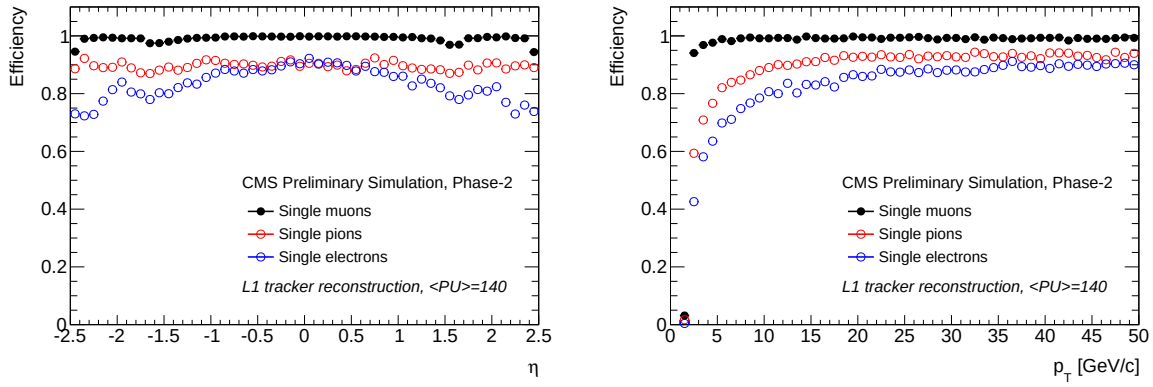


Figure 2.25: Efficiency for L1 track reconstruction as a function of η (left) and p_{T} (right) for muons, pions, and electrons.

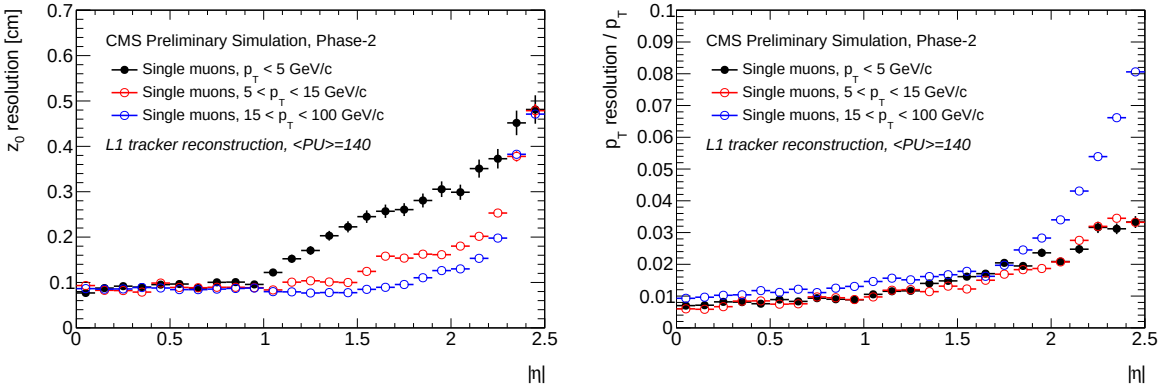


Figure 2.26: Resolution in z_0, and relative resolution in p_{T} for the L1 track reconstruction of single muons as function of η for different p_{T} ranges.

in Fig. 2.25 for the tracklet algorithm, where the efficiency is defined with respect to truth-level tracks and corresponds to loose track quality criteria for muons, pions and electrons. Muons have an overall efficiency of about 99%, while pions and electrons have lower efficiency due to interactions with detector material. It is expected that those figures (notably for electrons) can be further improved. The performance obtained with the AM-based pattern recognition shows similar features.

The pattern recognition performance has also been verified in even more dense environments, namely events with four high energy top jets overlaid with 140 pileup events, and single particles overlaid with 200 pileup events. In both cases the efficiency was found to remain approximately the same, demonstrating the robustness of the tracking.

The L1 track z_0 and p_{T} resolutions are shown for single muons in Fig. 2.26 as a function of $|\eta|$ for three ranges of p_{T}. The momentum resolution decreases as expected beyond $|\eta| > 1.5$ due to the lower lever arm of the Tracker. A z_0 resolution of about 1 mm is achieved in the central region, similar to the average separation of pileup vertices.

2.4.4 Offline Tracking

The CMS track reconstruction is described in Ref. [67]. The tracking configuration used for the Phase-II detector is based on the high-PU configuration for the Phase-I detector, where

track finding starts with seeds formed in the pixel detector. The Phase-II Outer Tracker has the capability of providing high quality seeds from the p_T module stubs, that are currently not yet exploited in the reconstruction. Improved performance should therefore be expected in future from a global reoptimization of the tracking software for the new detector. In addition, the Phase-II pixel detector is expected to have improved hit resolution, and further improved coverage in the high-η extension. For all these reasons the preliminary results shown below should be considered to be conservative estimates of what will be eventually achievable with the Phase-II Tracker.

Some highlights of the expected tracking performance are shown in Fig. 2.27. The track reconstruction efficiency for the Phase-I detector at 50 PU and the Phase-II detector at 140 PU is shown in the upper row. The muon tracking efficiency is 100% over the acceptance of each detector, dropping off for $|\eta| > 2.4$ for the Phase-I detector and $|\eta| > 3.8$ for the Phase-II detector. For tracks from $t\bar{t}$ events, tracking efficiencies for the Phase-I and Phase-II detectors are nearly identical for $|\eta| < 2.1$, where the Phase-I efficiency starts to decrease. The Phase-II efficiency begins to drop for $|\eta| > 3.7$. These results indicate that the Phase-II efficiency at 140 PU is as good as the Phase-I efficiency at 50 PU and extends the η range by about 1.5 units on each side.

The fake rate, defined as the fraction of reconstructed tracks that are not associated with a simulated charged particle, is shown in the left plot of the lower row for tracks with $p_T > 0.9$ GeV from $t\bar{t}$ events. Although the fake rate for Phase-II at 140 PU is higher than for Phase-I at 50 PU, it is still adequate for the CMS event reconstruction. Also, as mentioned above, there are improvements expected for the performance of the Phase-II detector once the track reconstruction software is fully optimized to take advantage of the additional capabilities of the new detector.

The better hit resolution of the Phase-II detector and the reduction in material results in a significantly improved p_T resolution as shown in the right plot of the lower row, for $p_T = 10$ GeV and $p_T = 1$ GeV muons.

The Phase-II Tracker also provides adequate performance in the low end of the p_T spectrum, as required for the Global Event Description. At pileup higher than nominal, the tracking finding efficiency remains basically unaffected up to 200 PU, while the fake rate increases moderately (by less than a factor of two), demonstrating the robustness of the detector.

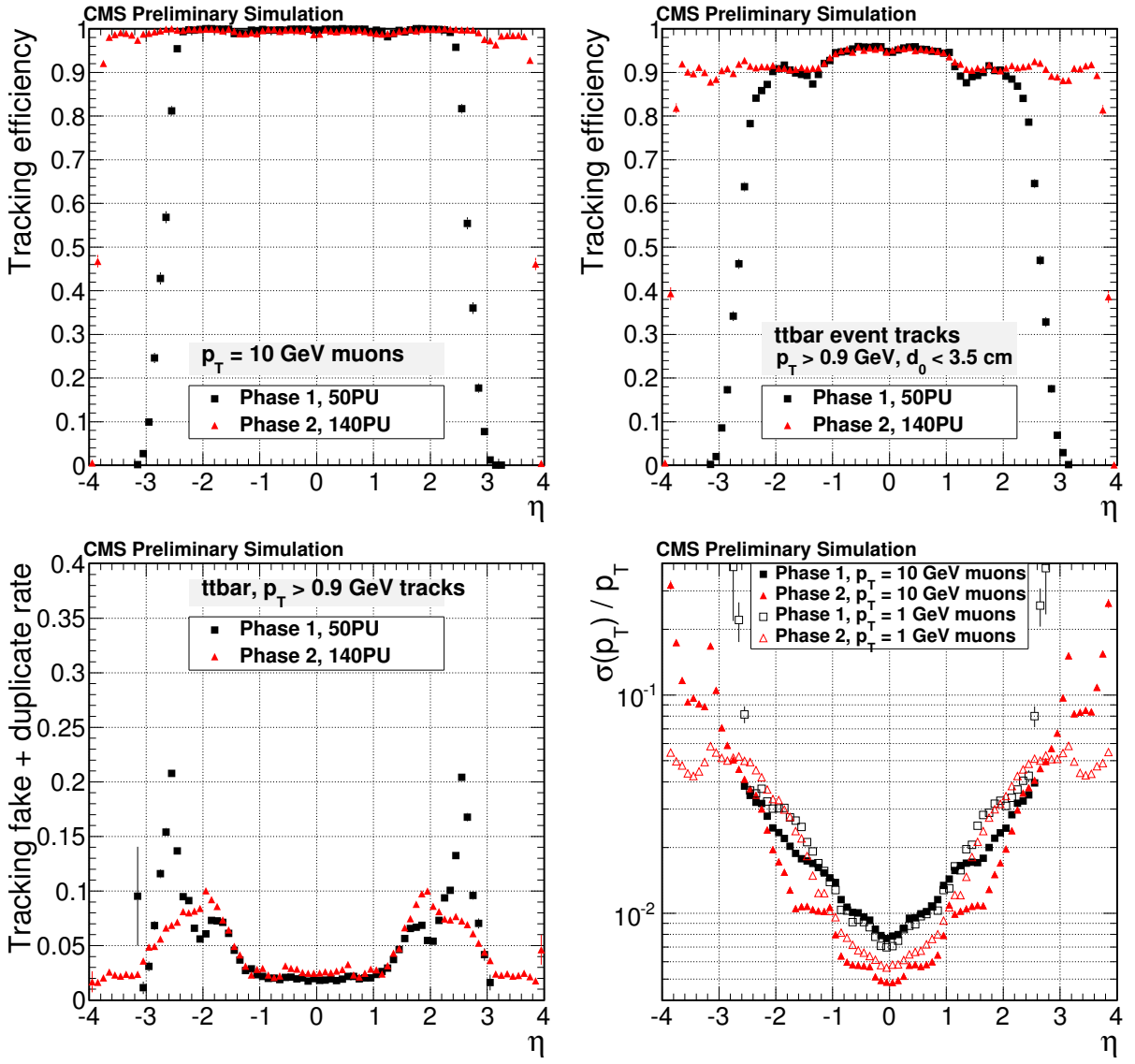


Figure 2.27: Upper row: track reconstruction efficiency as a function of η for the Phase-I detector without aging at 50 PU and the Phase-II detector at 140 PU; the left plot shows the efficiency for $p_{\mathrm{T}} = 10$ GeV muons and the right plot shows the efficiency for charged particles from $t\bar{t}$ events with $p_{\mathrm{T}} > 0.9$ GeV and originating within 3.5 cm of the beam line (in the transverse direction). Lower row, left plot: fraction of reconstructed tracks that are not matched to simulated charged particles as a function of η for the Phase-I detector at 50 PU (without aging) and the Phase-II detector at 140 PU. Lower row, right plot: same comparison for the p_{T} resolution of $p_{\mathrm{T}} = 10$ GeV and $p_{\mathrm{T}} = 1$ GeV muons.

2.5 Project planning and cost estimates

2.5.1 Project organization

The Phase-2 Tracker Upgrade project is part of the Tracker organization, which also comprises the Maintenance and Operation of the present detectors (Pixel Detector and Strip Tracker), and the Phase-1 Upgrade of the Pixel Detector.

A top-level management team composed of the Tracker Project Manager, his or her Deputies, and the Tracker Resources Manage oversees all three activities, with the support of the Management Board and Finance Board, and reports to the Tracker Institution Board. The Tracker Editorial Board and Conference Committee oversee the preparation of tracker-specific papers and participation in conferences.

The Phase-2 Upgrade of the Outer Tracker and Pixel Detector is steered by a Coordinator and Deputy, assisted by a Phase-2 Management Board that includes regional representatives of the countries participating in current developments, the likely future stakeholders of the construction. The composition of the board is revised once per year to keep up with the evolution of the project. The coordination of Phase-2 developments is organized in Working Groups. For the Outer Tracker, the R&D started with two key activities: sensor development and Front-End ASICs design. With the advancement of the project, more working groups (OT module design, OT mechanics) and subgroups (under OT electronics) have been formed, and a few more are expected to be formed in the near future (e.g. power, cooling, overall mechanical integration, etc.). The Phase-2 Pixel Detector Upgrade project, which is in a less mature stage, is currently composed of Sensors and Electronics Working Groups; its organizational structure will undergo a similar evolution as the Outer Tracker. A small group developing the detector modelling software package (see Section 2.4.1) serves both upgrade projects. The simulation and reconstruction group, under the supervision of the Tracker Detector Performance Group, also serves both projects.

2.5.2 Schedule and next steps

The installation of the Phase-II Outer Tracker and Pixel Detector is expected in late 2024. Sufficient contingency in the schedule is reserved for a thorough commissioning of the assembled detectors prior to the installation in CMS in order to insure optimal performance at the restart of accelerator operation. A highly simplified view of the project schedule is shown in Fig. 2.28. Module construction is the driver for both the Outer Tracker and the Pixel detector projects. Module components entail many difficult technological challenges as well as expensive and complex procurements; module assembly and quality control are labour-intensive and crucial to the long-term health of the detectors. The schedule for the Pixel Detector project is somewhat more relaxed due to the shorter construction time.

Major steps planned for the years $2015 - 2017$ are listed in Table 2.4. The target dates are established for the purpose of project planning. Actual dates can be several months later in all cases without affecting technical progress. The milestones shown are planned to cover the developments that need to be done up to the engineering design review (EDR).

The Tracker Technical Design Report is planned to be completed in Q1 of 2017 (milestone TP.TK.CM.3), so that the procurement phase for the outer tracker can begin by the end of 2017. Most of the Outer Tracker milestones are related to module construction. As discussed in Section 2.3.1, the range of products for the sensors that can be used in the detector is already defined. The final choice will be done in the context of the procurement, taking also cost into account. A major global objective is to fully finalize and validate the design of both module types,

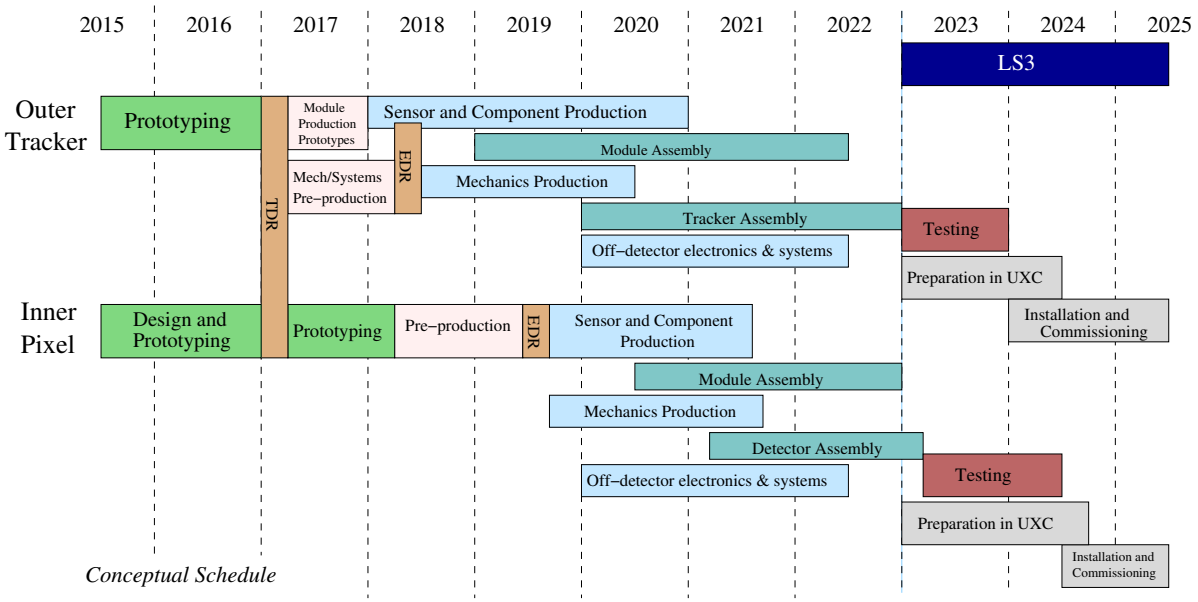


Figure 2.28: Conceptual schedule for the Outer Tracker and Inner Pixel portions of the Phase-II upgrade.

up to the point of building final prototypes (milestones TP.TK.OT.5, TP.TK.OT.9 – TP.TK.OT.10, TP.TK.OT.11 - TP.TK.OT.12). At the same time, efforts will be made to qualify vendors for the production of all the module components including hybrids, mechanical parts and assembly work (TP.TK.OT.8), to minimize risks at the beginning of the procurement/production phase. For the mechanical structures, the first goal is to define the geometry of the TBPS, choosing between the baseline layout and the one with progressively tilted modules (TP.TK.OT.1, discussed in Section 2.3.4.3), and then finalize and validate the engineering design of all three subsystems (TB2S, TBPS and TEDD – TP.TK.OT.6); the fine-tuning of the details can be completed after the start of the module production. The back-end electronics can be developed later, taking advantage of the likely availability of hardware with improved performance and reduced cost. A medium-term milestone is related to the L1 track finding, which is a completely new functionality of the upgraded tracker. A hardware demonstrator is planned to assess feasibility and performance, refine the estimate of required latency and establish a solid basis to design the final system (TP.TK.OT.7). The Engineering Design Review for Outer Tracker is planned for 2018 (TP.TK.OT.13).

The baseline pixel system detector layout will be defined (TP.TK.PX.1) in early 2016 to allow for detailed simulations of the full tracking system performance in the tracker TDR. In preparation for the TDR, the ROC technical specifications will also be defined (TP.TK.PX.2) as will the overall system concept (TP.TK.PX.4) and the specific outstanding issue of the powering scheme to be adopted (TP.TK.PX.3). The choice of the sensor technology for each layer (TP.TK.PX.5) is also planned to be made in time for the finalization of the TDR. Following the TDR, engineering design of detector modules and support structures will be carried out (TP.TK.PX.7). A full-size ROC demonstrator will be produced by the RD53 collaboration (TP.TK.PX.6) and will serve as basis for the design of the final ROCs of CMS and ATLAS. A set of prototypes will be constructed leading up the Engineering Design Review planned in 2019 (TP.TK.PX.8).

The Cooling systems, as well as the development of the mechanical and thermal interfaces of the tracking volume to the rest of the CMS detector, will be treated as common projects. Conceptual designs will be developed as soon as the basic parameters of both the Outer Tracker

Table 2.4: Planned next steps for the Outer Tracker and Pixel upgrade projects.

Project	Milestone Identifier	Target date	Description
Outer Tracker	TP.TK.OT.1	Q3 2015	Choice of TBPS geometry
	TP.TK.OT.2	Q4 2015	Finalization of FE ASICs specifications
	TP.TK.OT.3	Q4 2015	Full-size protoptypes of 2S modules
	TP.TK.OT.4	Q1 2016	Small-size prototypes of PS modules
	TP.TK.OT.5	Q3 2016	Final prototype CBC chip
	TP.TK.OT.6	Q3 2016	Baseline engineering design of support structures
	TP.TK.OT.7	Q4 2016	System demonstrator for L1 track finding
	TP.TK.OT.8	Q4 2016	Qualification of vendors for module components
	TP.TK.OT.9	Q1 2017	Final prototypes of 2S module hybrids
	TP.TK.OT.10	Q2 2017	Final prototypes of PS module hybrids
	TP.TK.OT.11	Q3 2017	Final prototypes of MPA, SSA and CIC chips (all FE chips finalized)
	TP.TK.OT.12	Q4 2017	Final prototypes 2S and PS modules
	TP.TK.OT.13	Q2 2018	Outer Tracker Engineering Design Review
Pixel	TP.TK.PX.1	Q1 2016	Definition of detector baseline layout
	TP.TK.PX.2	Q2 2016	Finalization of ROC specifications
	TP.TK.PX.3	Q3 2016	Choice of baseline powering scheme
	TP.TK.PX.4	Q3 2016	Definition of Front-End Electronics system concept
	TP.TK.PX.5	Q4 2016	Choice of sensor technology for each layer
	TP.TK.PX.6	Q3 2017	Full-size ROC demonstrator available (RD53 deliverable)
	TP.TK.PX.7	Q2 2018	Engineering design of detector modules and support structures
	TP.TK.PX.8	Q3 2019	Pixel Engineering Design Review
Common	TP.TK.CM.1	Q3 2016	Conceptual design of Cooling Systems
	TP.TK.CM.2	Q3 2016	Conceptual design of mechanical interfaces, bulkheads and feedthroughs
	TP.TK.CM.3	Q1 2017	Submission of Technical Design Report

and the Pixel are fully defined and will be described in the TDR.

2.5.3 Cost estimates

The construction of the present Tracker, as well as the Phase-I Upgrade of the Pixel Detector, are used as a guide in establishing a cost estimate for the Phase-II Upgrades. The cost estimates provided are based on the detailed parts list of the detectors, according to the current versions of the designs. For all relevant procurements, the scale of the production is a key factor that determines the unit cost, hence the existing detector is the most significant precedent, while costs of prototypes or catalogue prices of components are often not relevant for large procurements. Where appropriate, and especially for novel materials and technologies not used in the present detectors, potential vendors have been requested to provide estimates of the likely cost scale for the quantity of components that are needed. A summary of the present cost estimates is shown in Table 2.5.

Table 2.5: CORE cost estimates for the Phase-II Upgrades of the Pixel Detector and Outer Tracker.

	Estimated CORE cost in MCHF (2014)		
	Outer Tracker		Pixel Detector
Modules	2S	27.9	11.8
	PS	22.8	
Mechanics		6.6	3.6
BE electronics		12.4	2.0
Power system		6.1	1.4
Services cables and pipes		5.5	1.1
Cooling system		4.2	0.6
Infrastructure and installation		3.5	2.7
TOTAL		**89**	**23**

Chapter 3

Calorimetry

3.1 Introduction

The CMS calorimeters are designed to provide hermetic coverage between $-5 < |\eta| < 5$ with a set of electromagnetic and hadronic subdetectors. Within the planned HL-LHC physics program, the calorimeters are critical for the identification and reconstruction of photons and electrons and the measurement of jets and missing transverse momentum. For Higgs physics, high resolution and efficient identification for photons is required for the $H \rightarrow \gamma\gamma$ decay process, which is important for Higgs parameter and self-coupling measurements. Electrons are important for Higgs physics and for many physics topics beyond the Standard Model. Efficient separation of electrons from hadrons requires information from the front portion of the hadron calorimeter as well as information from the tracker and ECAL. Jets and missing transverse momentum in CMS are measured using particle flow techniques which combine information from the tracker and calorimeters. At high pileup, as expected for HL-LHC, particle flow requires good segmentation in the transverse and longitudinal directions to correctly associate tracks and calorimeter energy deposits.

The existing CMS calorimeters were designed to meet these challenges for 500 fb^{-1} at the LHC over 10 years of data-taking at a maximum instantaneous luminosity of $1 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$. With the Phase-I upgrades of the hadron calorimeter, the detector is capable of operating up to $2 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$. The HL-LHC performance criteria, however, require a re-examination of the ability of the detector active material and electronics to meet the requirements of 3000 fb^{-1}. Based on these studies, upgrades to the calorimeters are proposed including the replacement of the endcap calorimetry systems.

3.1.1 Review of the Existing CMS Calorimeters

The CMS electromagnetic calorimeter (ECAL) is a homogeneous calorimeter made of 75848 lead tungstate ($PbWO_4$) scintillating crystals, located inside the CMS superconducting solenoid magnet [19]. It is made of a barrel part (EB) covering the region of pseudorapidity $|\eta| < 1.48$ and two endcaps (EE), which extend the coverage up to $|\eta| < 3.0$. The photo-detectors are Avalanche Photo-Diodes (APD) in the barrel and Vacuum Phototriodes (VPT) in the endcaps. A preshower detector (ES), based on lead absorbers equipped with silicon strip sensors, is placed in front of the endcap crystals. Electrons and photons are typically reconstructed up to $|\eta| < 2.5$, the region covered by the silicon tracker, while jets are reconstructed up to $|\eta| < 3.0$. The ECAL energy resolution achieved during 2010-2011 is described in [68] and ranges from 1.1 to 2.6% in the barrel and 2.2 to 5% in the endcaps for photons from the Higgs boson decay.

The CMS hadron calorimetry system (HCAL) has four major sections: the HCAL Barrel (HB), HCAL Endcap (HE), HCAL Outer (HO), and HCAL Forward (HF). The HB and HE calorime-

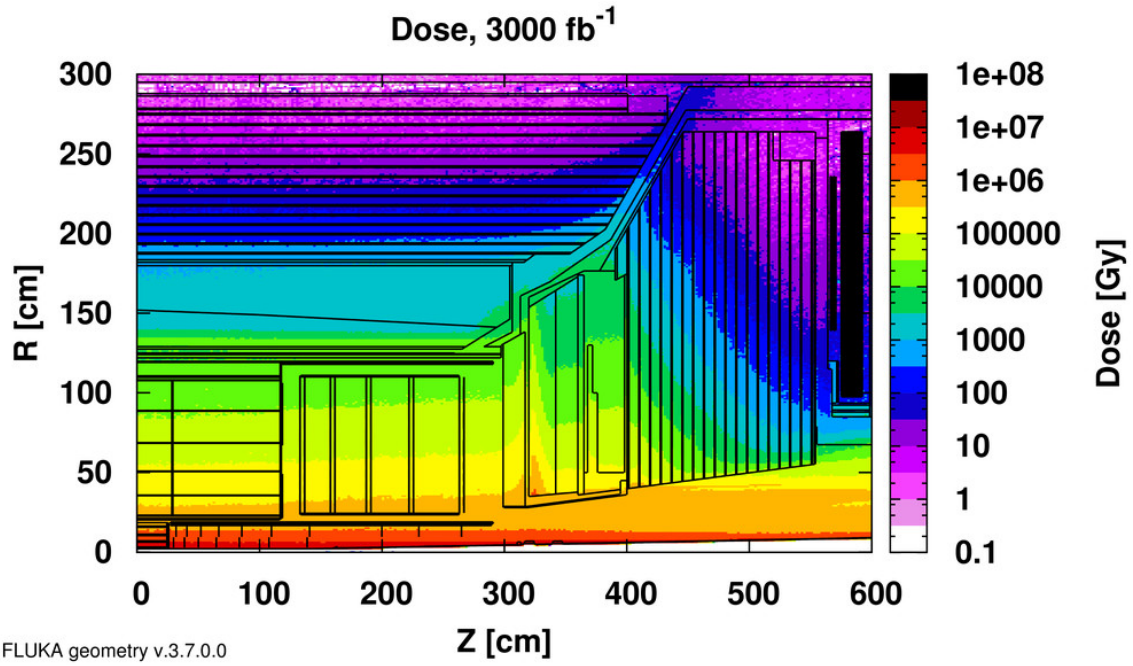


Figure 3.1: Expected radiation dose in the central calorimeters (EB, HB, EE, and HE) after an integrated luminosity of 3000 fb^{-1}.

ters are sampling calorimeters which use a brass absorber and plastic scintillator as the active material. Light from the plastic scintillator is wavelength-shifted (WLS) and captured in fibers. Fibers from the calorimeter tiles are grouped to form projective towers which were read out originally by hybrid photodetectors and then by SiPM devices after the Phase-I upgrade of CMS. The HCAL Outer calorimeter (HO) functions as a tail-catcher for hadronic showers and sits outside the CMS solenoid magnet. The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers which run longitudinally (parallel to the beam) through the absorber and collect Cherenkov light, primarily from the electromagnetic component of showers which develop in the calorimeter. Together, these detectors cover $0 < |\eta| < 5.0$ and are important for the measurement of jets and missing transverse momentum.

3.1.2 Radiation, Pileup, and Upgrade Scope

The HL-LHC conditions are a significant challenge to both detector longevity and detector performance. The challenge is particularly strong in the endcap region, where the radiation levels have a strong gradient along η and typically change by a factor 100 between $|\eta| = 1.48$ and $|\eta| = 3.0$ as shown in Figure 3.1.

For the full 3000 fb^{-1} planned for HL-LHC, the high-energy charged hadron fluence at $|\eta|$ =2.6 in the ECAL will be $2{\cdot}10^{14}/\text{cm}^2$ and the absorbed dose will be 300 kGy with an expected dose rate of 30 Gy/h. In the HCAL, the radiation dose at the location of the HE scintillating tiles will reach up to 300 kGy. As discussed below, these dose and fluence levels would result in significant loss to detector performance and motivate the replacement of the endcap calorimeters of CMS for HL-LHC.

In the barrel ECAL, the expected radiation damage in the highest η region of the EB after 3000 fb^{-1} is about the same as in parts of the EE after 30 fb^{-1}, and thus radiation damage is not an issue for EB crystals. However, the EB detector faces significant issues from high

pileup, increased photodetector noise, and trigger efficiency at $5 \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1}$, while the electronics must be upgraded to meet the Level-1 readout rate and latency requirements. For the HB detector, some portions of the detector active material will need to be replaced, while the photodetector and electronics longevity issues have been resolved by the Phase-I upgrade.

The outer calorimeter (HO) will see relatively little dose and the forward calorimeter (HF) is designed to withstand the full 3000 fb^{-1} without significant loss of performance. The electronics of these detectors can meet the requirements of the Level-1 readout rate and latency and no upgrade is planned for these subsystems.

3.1.3 Upgrade Outline

This chapter on calorimetry is organized in the following way. Section 3.2 presents studies of the ageing of the barrel and endcap calorimeters, the preshower, the ECAL electronics, and the forward HCAL (HF). Section 3.3 describes how the performance of the calorimeters changes as a function of the ageing and pileup. These two sections establish the basis for the proposed upgrades, demonstrating which components require upgrade and which are acceptable for the full HL-LHC run.

The barrel upgrades are described in Section 3.4. The full replacement of the EB electronics is required to meet the Level-1 requirements and this opportunity will be taken to adjust the shaping time to reduce the impact of out-of-time pileup and APD noise. The rejection of neutron-induced fakes in the Level-1 trigger algorithm will also be improved. For the HB, the replacement of a subset of the scintillator tiles with more radiation-tolerant designs will be required to maintain performance, particularly for electron and photon identification.

A complete replacement of the endcap calorimeters with a new high-granularity sampling calorimeter is proposed in Section 3.5. The proposed design encorporates a silicon/tungsten electromagnetic section followed by two hadronic sections, both using brass as the primary absorber material. In the front section, the active material is silicon while the back section uses plastic scintillator. The design is targetted to achieve very high performance for physics objects reconstructed in the presence of high levels of pileup. The high transverse and longitudinal granularity of the detector allows tracking and separatation of signal and pileup particles inside the calorimeter.

Detector-level performances for the upgraded calorimeters are presented in Sections 3.4 and 3.5. The full object-level performance, which combines the calorimeter response with the tracker information, is presented in Chapter 9.

3.1.4 Calorimetry timing

The high spatial granularity of the current and proposed CMS calorimeters is a critical tool to manage the high pileup levels expected in the CMS detector during HL-LHC. While precision time measurements have not traditionally been a key aspect of calorimeters at hadron colliders, such measurements may allow the subtraction of the neutral energy that enters the electron and photon clusters from pileup, help identify pileup jets, and improve the estimation of the production vertex location for di-photon events. This may allow improved performance at 140 PU and CMS is therefore considering the possibility of implementing such capabilities in the EB and endcap calorimeters.

In the forward direction, a measurement of cluster time with a precision of 20-30 ps would allow the association of energy deposits with vertices to within 1 cm, resulting in a significant reduction of fake objects and energy deposits. Precise timing measurements could be achieved

by different means and with likely different performance in the barrel and endcap regions. In the EB, this capability will be considered in the specifications and design of the new front-end electronics. As discussed below, the front-end electronics of the endcap calorimeter will have an inherent timing capability. In both cases, timing capabilities could be enhanced by adding a dedicated timing detector in front of the existing one, with generic R&D ongoing for micro-channel plates for direct ionization measurement [69, 70], gas electron multipliers [71], and avalanche photodiode detectors [72].

The software to use precise timing measurement in the framework of the particle flow global event reconstruction, including use of the tracker information, is currently being developed. Generic R&D is ongoing to develop precise timing-measurement devices and investigate system issues for timing measurements in the framework of calorimetry for future HEP experiments or for medical applications [73]. CMS is building on these efforts to determine if practical and cost-effective solutions exist for precise time measurement capabilities in a large-scale hadron-collider detector environment.

3.2 Longevity of the Phase-I Detector

3.2.1 ECAL Longevity

The following sections describe in detail the expected behavior of the ECAL detector components at the HL-LHC. This includes the degradation of the crystal light transmission and the increase in APD noise due to radiation damage. In addition the aging of the readout electronics, preshower and endcap VPTs is presented.

3.2.1.1 Radiation damage in lead tungstate

ECAL lead tungstate crystals were optimized for their application in CMS, particularly with respect to radiation hardness through many studies with ionizing electromagnetic radiation and hadrons [74]. The effects of ionizing electromagnetic radiation on lead tungstate were studied in depth using γ radiation from ^{60}Co sources. The crystal light transmission is reduced through the formation of color centers, while the scintillation mechanism is unchanged. The crystal transparency largely recovers through spontaneous annealing at room temperature [74]. The light output tends towards an equilibrium level, which depends on the dose rate. A precise monitoring system using injected laser light at 447 nm is used to track and correct for response changes during LHC operation.

The dominant concern for HL-LHC is the loss of response due to hadron irradiation. In CMS the flux of fast hadrons is dominated by charged pions with energies of order 1 GeV. The effect of charged hadrons has been studied on crystals irradiated with pions and protons from the CERN Proton Synchrotron [75, 76]. Figure 3.2 shows that the band edge of the transmission spectrum is shifted to higher wavelengths and the transmission is attenuated. There is no damage to the scintillation mechanism [77], however the light output is seriously affected by the transmission loss in the scintillation spectrum range. This damage is cumulative and does not anneal out at room temperature. A complete annealing is observed when heating the crystals up to 350°C. A small recovery can be achieved at lower temperature (50-70^{o}C) or with optical bleaching, however this recovery is insufficient to remediate the expected damage at the HL-LHC [76].

The crystal transmission loss is quantified by the radiation-induced absorption coefficient: $\mu_{\text{ind}}(\lambda)$, where the attenuation of the optical transmission LTO along the crystal length L, with respect to its initial value LTO_0, is given by:

$$\frac{LTO(\lambda)}{LTO_0(\lambda)} = e^{-\mu_{\text{ind}}(\lambda)L} \tag{3.1}$$

The absorption coefficient μ_{ind} is found to be linearly dependent on the hadron fluence in the studied range of fluence ($5 \cdot 10^{11}$ to $1.3 \cdot 10^{14}$protons/cm^2) [75–78].

Extensive beam tests of hadron-irradiated matrices of crystals [78] were performed to determine the expected μ_{ind} values, light collection efficiency, and energy resolution of the calorimeter during HL-LHC operation, as depicted in Figure 3.2 (right) and 3.3 (left). Based on these measurements, a model of crystal aging has been developed. The radiation damage is modeled using the MARS simulation for the hadron fluence and γ dose estimate, and the experimental results described above for the attenuation of the optical transmission. A full simulation of the electromagnetic shower development in the crystals is performed with Geant4 [79]. The ray-tracing program SLitrani [80] is used to model the light output as a function of μ_{ind}. The γ damage is simulated as a function of the instantaneous luminosity, by assuming the equilibrium value of the light transmission is reached. The hadron damage is simulated by modeling the value of μ_{ind} measured in [75, 77] as a function of the hadron fluence. A comparison with the beam test measurements shows good agreement between the expected light output and the simulation (see Figure 3.2 right). Figure 3.3 (right) shows the predicted response reduction, S/S_0, for an electron shower of 50 GeV as a function of η for various integrated luminosities. The uncertainty in this prediction is about 30%, and it comes from uncertainty in the hadron fluence and γ dose estimates, spread in the crystal response to γ radiation, uncertainty in the effects of the hadron radiation on the crystals, and extrapolation of these effects from low to high fluence. Based on the aging model, the barrel ($|\eta| < 1.48$) will suffer less than a factor of 2 light output loss up to 3000 fb^{-1}, while in the endcaps there will be a factor of more than 10 reduction after 3000 fb^{-1} for $\eta > 2$.

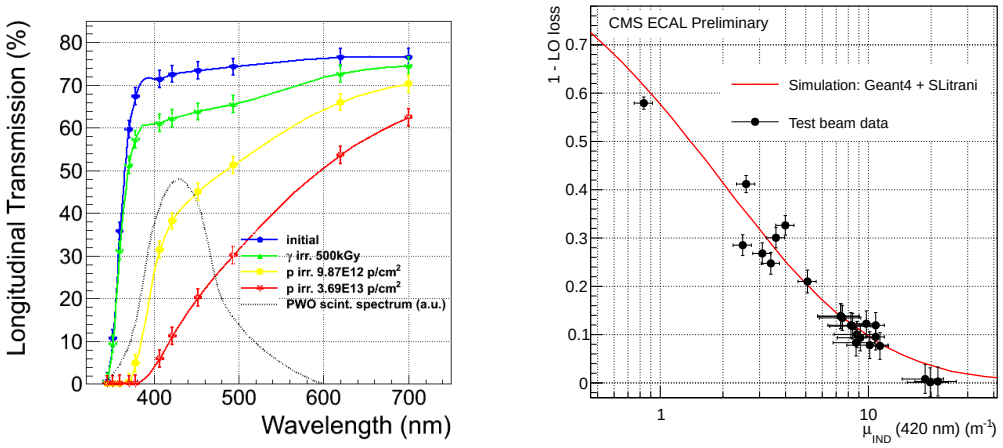


Figure 3.2: Left) Light transmission curves of lead tungstate crystals after γ and proton irradiation. Overlaid (black dotted line) is the lead tungstate scintillation spectrum. Right) Test-beam measurements of the crystal light output as a function of the induced absorption coefficient μ_{ind}. The aging model expectation is superimposed on the data.

The reduction of scintillation light output with respect to undamaged crystals affects the energy resolution causing:

- a degradation of the stochastic term due to the reduced light output, proportional to

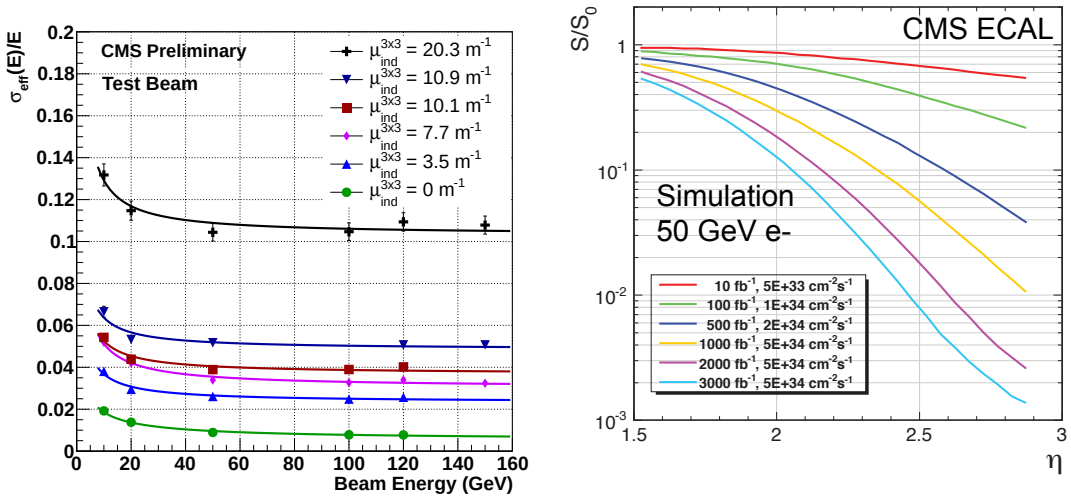


Figure 3.3: Left) Test-beam measurements of the energy resolution as a function of electron energy for hadron irradiated crystal matrices. The various data points refer to crystal matrices with different μ_{ind}. Right) Expectation of the relative light output S/S_0 of the ECAL crystals for electron showers of 50 GeV as a function of the pseudorapidity η for various aging conditions, corresponding to the integrated luminosities as written in the figure legend.

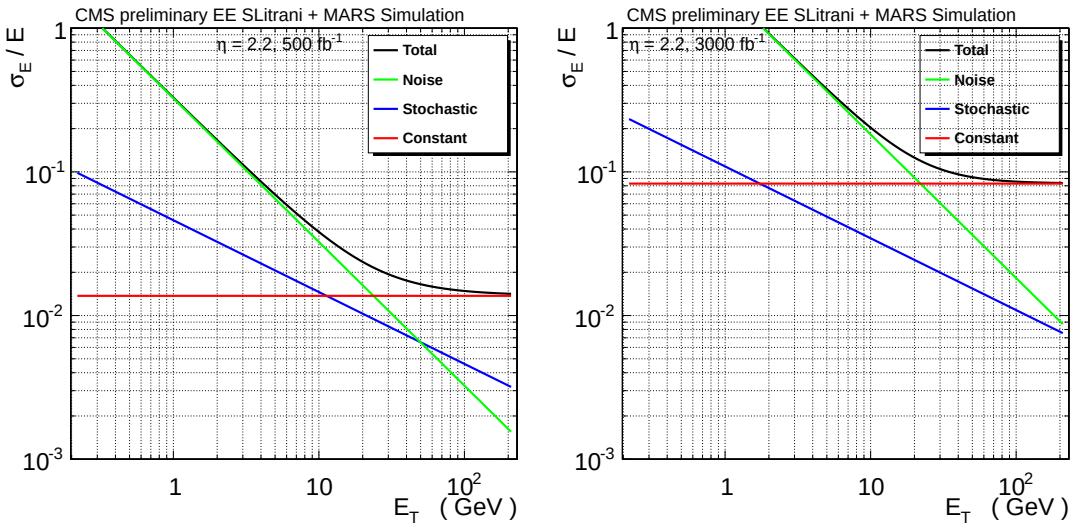


Figure 3.4: The energy resolution including aging effects at $|\eta|$=2.2 (left) after 500 fb^{-1} and (right) after 3000 fb^{-1} versus E_T

$\sqrt{S_0/S}$

- a worsening of the noise term due to the amplification of the energy equivalent noise, proportional to S_0/S,
- a worsening of the constant term due to the increased light collection non-uniformity along the crystal depth z, proportional to $\partial(S/S_0)/\partial z$.

Figure 3.3 left shows the electron energy resolution measured in a test beam for 3×3 crystal matrices with different μ_{ind}. The photo-detectors and electronics of the test-beam setup were different than in CMS, so the electronics noise term cannot be directly compared with the model. The progressive degradation of the constant term of the energy resolution from these data is in good agreement with the model prediction. The beam test has also shown that the linearity and timing response of the calorimeter are also degraded at high values of μ_{ind}.

The aging effects on the ECAL energy resolution are illustrated by the model prediction in Figure 3.4 for crystals located in the endcaps at $|\eta| = 2.2$ after 500 and 3000 fb^{-1}. In the endcaps the response loss will seriously impact the energy resolution, with a progressive worsening of the constant term and the energy equivalent electronics noise. The consequences for the detection of photons and electrons are studied quantitatively in Section 3.3.1.

3.2.1.2 Radiation-induced Aging of Avalanche Photodiodes

The ECAL barrel APDs are silicon photodiodes with internal gain, developed and produced by Hamamatsu Photonics in collaboration with CMS [81]. They are operated at gain $G = 50$. As with all silicon devices, they are sensitive to γ and hadron damage [82]. The γ radiation causes an increase in the surface current and decreases the quantum efficiency. The principal concern for the HL-LHC is hadron damage. Neutrons dominate the hadron spectrum behind the ECAL crystals. Hadrons create defects in the silicon bulk which induce an increase in dark current of the detector and consequently in the levels of electronic noise. The creation of defects in the APD is linear with the neutron fluence, an expectation which has been confirmed by observation of the dark current in the detector during Run-I. The neutron fluence predicted with the FLUKA simulation is about $\Phi_n = 1.2 - 2.4 \cdot 10^{14}$ n/cm^2 (1 MeV neutron equivalent) at $\eta = 0 - 1.45$ for a delivered luminosity of 3000 fb^{-1}. The expected dark current of a pair of APDs at gain 50 at $\eta = 1.45$, considering only the permanent bulk damage, is 200-250 μA. Other effects due to neutron damage at high fluences are: a shift of the bias voltage required to obtain a gain of 50 and a reduction of quantum efficiency.

The preamplifier noise depends on the square root of the APD bulk current times the shaping time τ [83], which is currently 43 ns. Figure 3.5 shows the noise measured with the present electronics, as a function of the APD dark current. The noise is expected to increase up to 9-10 ADC counts for a current of 200-250 μA. Figure 3.5 right shows the prediction of the noise increase as a function of the dark current with the Phase-I ECAL electronics. The noise levels in EB will grow as a function of the integrated luminosity reaching almost 350 MeV per channel at $\eta = 1.45$ after 3000 fb^{-1}, neglecting changes in crystal light output and quantum efficiency. Considering that electromagnetic showers in ECAL are typically reconstructed summing the energy of at least a 5×5 crystal matrix, the level of electronics noise will be the limiting factor of the EB energy resolution for photons from the Higgs boson decay after 1500 fb^{-1}. More quantitative studies with full detector simulations based on this aging model are described in Section 3.3.1. The impact of the noise increase on anomalous signal rejection is described in Section 3.4.5.1.

The APD dark current (I_D) is strongly dependent on the temperature and it follows the relation:

$I_D \sim T^{\frac{3}{2}} e^{-E/2K_BT}$, where $E = 1.2$ eV. A reduction of a factor 2.5(2) in dark current is expected if the ECAL operating temperature is reduced from 18 to 8(10) °C, which translates into a reduction of the noise by 35(30)%. The dark current annealing depends on the temperature and is slower at 8-10°C, but the difference is minimal. Figure 3.5 also presents the expected noise levels in the detector at lower temperature and it shows that operating the APDs at 8-10°C after 3000 fb^{-1} corresponds in terms of noise to operating at 18°C after about 1200 fb^{-1}.

Another potentially dangerous effect observed on highly irradiated APDs is self-heating due to dark current increase: the dark current becomes so high that the internal temperature of the device grows and it cannot reach high gain. For currents above 150 μA the working point and consequently the gain cannot be precisely fixed, with undesirable effects on detector stability and thus on energy resolution. Preliminary tests indicate that the contact with the crystal helps stabilize the APD temperature. Further tests are ongoing. The risk of APD self-heating is reduced at lower temperature, where the current is smaller.

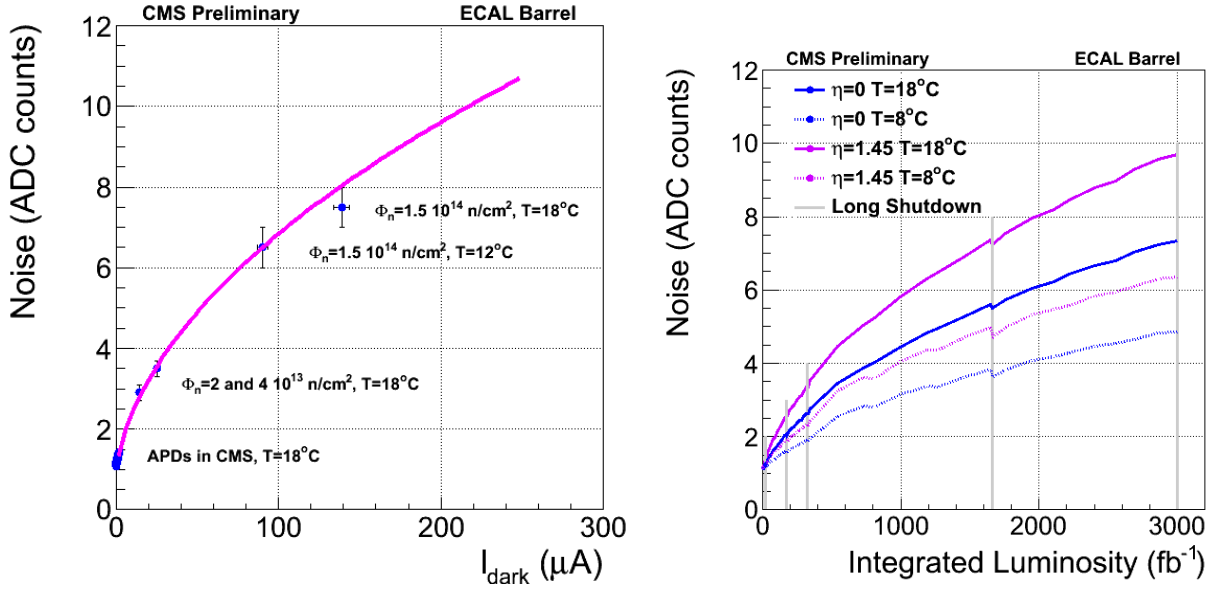


Figure 3.5: Left: noise as a function of the dark current with the Phase-I ECAL electronics. Right: APD noise as a function of the integrated luminosity. Also shown are curves for different ECAL operating temperatures.

3.2.1.3 Vacuum phototriodes response losses with aging

Endcap VPTs are single-stage photomultiplier tubes manufactured by Research Institute Electron, St. Petersburg, Russia, following a period of development in collaboration with CMS and PNPI Gatchina, Russia [84]. VPT response losses arise from VPT conditioning and faceplate transparency loss. An extensive set of tests has been carried out to quantify long-term conditioning losses using dedicated 4 T test facilities. Conditioning losses are thought to arise from residual gasses inside the tubes that are ionized during operation and are accelerated to the photo-cathode where their impacts cause a loss of quantum efficiency. Losses have been measured for 26 VPTs as a function of the integrated photo-cathode current drawn from VPTs under illumination with blue LED light. The losses show expected asymptotic behavior. The average saturation value is about 73%. Most of the losses occur by the time an integrated charge of 5-10 mC has been drawn from the photo-cathode. VPTs in the region $2.7 < \eta < 3.0$ are expected to have reached full conditioning by the end of Run-I in 2012. Full conditioning of VPTs in the

regions $2.4 < \eta < 2.7$ and $2.1 < \eta < 2.4$ is expected by the end of 2016 and by mid 2018 (the end of Run-II), respectively, and for $1.8 < \eta < 2.1$ by the end of Run 3 (at the end of 2022). A small degree of recovery is expected during periods of no beam, at the level of a few percent of the initial loss.

The transparency of the VPT faceplates (UV borosilicate glass) decreases in the presence of ionizing radiation. The transparency loss for the glass batches used for faceplate production was required to be less than 10% for a received dose of 20 kGy. Further faceplate tests have been carried out with a Co^{60} source and with a 24 GeV proton beam for proton fluences up to $4.3{\cdot}10^{13}$ cm^{-2}. All the indications to date show that the transparency losses are asymptotic and level out at ∼10%. The VPT faceplate transparency losses in the region $2.7 < \eta < 3.0$ are expected to saturate by the end of Run-II.

Another consideration of relevance for the signal loss in the endcaps is that the gain of the VPTs cannot be further increased to compensate for the crystal light output reduction. The VPT signal loss, even though of minor relevance with respect to the crystal radiation damage, is taken into account in the aging model described in Section 3.3.1.

3.2.1.4 Present ECAL Electronics and its longevity

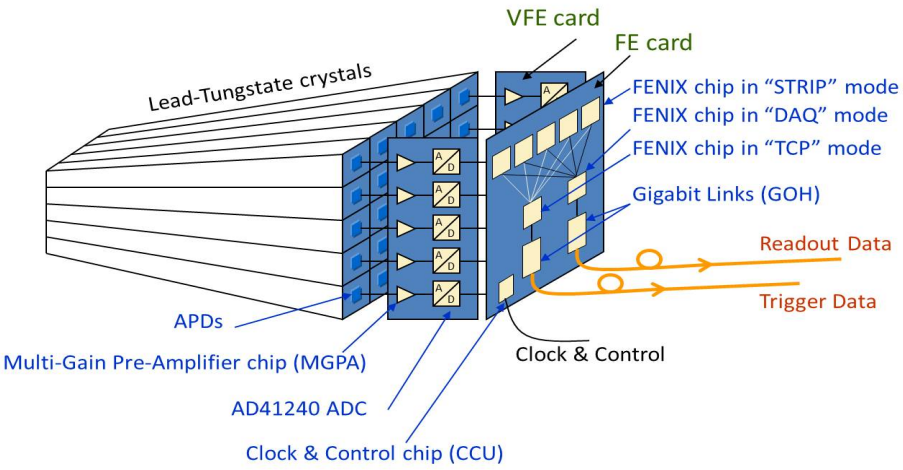


Figure 3.6: The ECAL Barrel on-detector electronics. The Mother Board (MB)(not shown) connects the APDs to the very-front-end card (VFE) which contains pulse amplification, shaping and digitization functions. The VFE connects to the front end card (FE) which forms the trigger primitives and contains a trigger latency buffer. The EE electronics have a similar architecture.

The present ECAL on-detector electronics is shown in Figure 3.6. Two APDs are mounted on the rear face of the crystal and connected in parallel to form one readout per crystal. The APDs are connected to the passive motherboard (MB) which distributes high voltage and low voltage as well as interconnecting the APDs to the very-front-end (VFE) cards. Each VFE card has five readout channels consisting of a multi-gain pre-amplifier (MPGA) with a 43 ns shaping time and a 12-bit analog-to-digital converter (ADC) (least significant bit 40 MeV). The digitized signals from five VFE cards are passed to an FE card. The FE card forms the trigger primitive for the 5x5 crystal array and contains a digital latency buffer and the primary event buffer.

The trigger primitive is the transverse energy deposited in a trigger tower and a single bit to qualify the energy deposit along η. The trigger primitives are transmitted optically to a trigger concentrator card in the underground service cavern. The per-crystal information is buffered

in the FE for transmission to the data concentrator card, with a maximum latency in the event buffer of 6.4 μs and a maximum Level-1 accept rate of about 150 kHz.

The LVR,VFE and FE are all fabricated with 0.25 μm radiation-hard CMOS technology. All the boards were tested to resist the radiation environment of the ECAL endcaps and for ten-year operation. During LHC Run-I very few failures occurred in the front-end electronics, resulting in less than 1% of non-operational channels. No failure was observed on the MB, LVR or VFE cards; a few faults were observed on the FE cards, which can be mostly traced back to bad contacts between the FE board and the Gigabit Optical Hybrid (GOH). These occurred soon after the installation and the failure rate is extremely small since then (1-2 per year). No failures have yet been attributed to radiation.

Accelerated aging tests on the present electronics have been carried out on few boards to achieve a total estimated lifetime of 40 years. The tests include ON/OFF cycling, temperature cycling, and maintaining the boards at high temperature. After every two-years-equivalent of these accelerated aging tests, the electronics functionality is verified. After 40 years of accelerated aging, no failures have been seen in the FE, VFE, and motherboards tested.

3.2.2 Pre-Shower Longevity

Each endcap preshower (ES) of the ECAL comprises 2 orthogonal planes of silicon strip sensors (320 μm thick, 2 mm$\times$61 mm) and associated front-end electronics, absorbers and services, covering the fiducial region $1.653 < |\eta| < 2.6$. It was designed and tested to withstand neutron fluences up to 4×10^{14}n/cm^{-2} and an integrated dose of 6 Mrads (corresponding to 500 fb^{-1}) and be operational for at least 10 years. The ES front-end electronics perform amplification and shaping of the signals from the silicon, sampling the amplified/shaped signal at 25 ns intervals and storing these voltages in a 192-cell-deep analogue circular buffer (i.e. up to 4.8 μs). Upon reception of an L1 trigger signal, three consecutive voltage samples corresponding to that event are read-out through optical links to the off-detector electronics. The electronics used for controlling and monitoring the ES, as well as the optical links, are identical to those used in the rest of ECAL (and the CMS Tracker).

In terms of longevity, production samples of the front-end electronics underwent a rigorous campaign of accelerated aging and thermal/power cycling during the construction phase in 2007/2008, to mimic operation during at least 10 years. No problems were found, but more tests would be required to extrapolate operation until 2035 (end of HL-LHC). Triplicated logic in the front-end chips minimize the susceptibility to radiation-induced single-event effects (SEE).

The main effect of radiation is on the silicon sensors. Hadrons damage the silicon crystal lattice, causing defects that trap charge, resulting in a decrease of the charge collection efficiency (CCE) and thus a decrease of the signal-to-noise. The voltage required for full depletion also increases. If it goes above the 500 V rating of the on-detector components the full depletion cannot be achieved and the CCE will decrease. Finally, the sensor leakage current, and thus the noise, increases with fluence. The DC-coupled sensors are protected by a series resistor, providing a hardware limit of 2 mA. The ES is presently operated such that the maximum leakage current allowed by software is 0.7 mA per sensor. The leakage current can be decreased by lowering the bias voltage and/or the silicon temperature. Lowering the bias has the undesirable effect of decreasing the CCE, as mentioned previously. An 8°C drop results in a factor 2 lower current and no adverse affects. The ES was designed for a minimum operating temperature of around -15°C. Figure 3.7 shows the expected evolution of leakage current for the sensors at the highest pseudorapidity, as a function of time, with the expected luminosity profile for LHC Phase-I

and Phase-II. The operating temperature is assumed to be -15°C, except during the shutdown periods, when the ES is at 20°C. It can be seen that after 2027 (corresponding to an integrated luminosity of approx 1000 fb^{-1}) the only way of lowering the current to below 0.7 mA/sensor is to lower the bias.

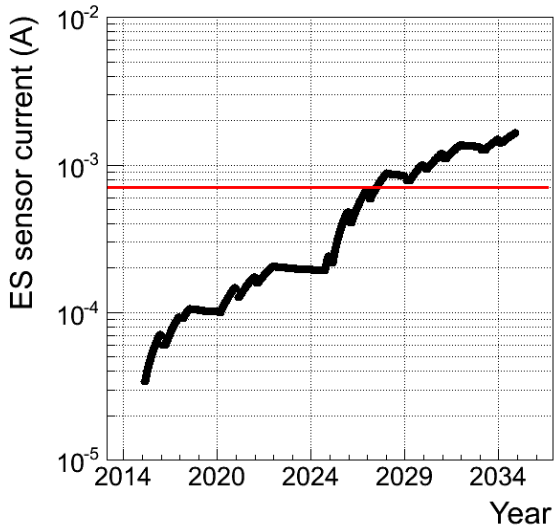


Figure 3.7: Expected evolution of the ES leakage current for the sensors at the highest rapidity as a function of time.

As the ES is calibrated using single minimum-ionizing particles (MIPs), the signal-to-noise for single MIPs will not be sufficient for accurate calibration from about 1000 fb^{-1} onwards, due to decreased CCE. Although the ES could still be operational at the end of HL-LHC, its overall performance will be significantly degraded.

As mentioned previously, the ES front-end has a limited buffer length. The specifications for the CMS Level-1 trigger for HL-LHC operation includes a latency greater than 12.5 μs and a Level-1 accept rate up to 750 kHz. Upgrading the ES electronics to cope with these specifications would require a complete rebuild of the detector. With the aforementioned degraded performance it is not practical to consider this rebuild. The ES will thus be removed from CMS in LS3, prior to HL-LHC operation. The improved granularity of the proposed upgrade endcap calorimeter will compensate for the absence of the ES.

3.2.3 HCAL Longevity

3.2.3.1 Radiation damage in scintillator tiles

As discussed in the introduction, for an integrated luminosity of 3000 fb^{-1}, the radiation dose received by the HE scintillating tiles will reach up to 3×10^5 Gy, and the HB tiles will accumulate up to 2×10^3 Gy. During the 2012 run, a decrease in light output from the HE tiles of up to 30% was observed for a dose of $\sim 10^3$ Gy. Figure 3.8 shows the observed decrease of light signal in the first layer of the HE as a function of the accumulated luminosity. The decrease is measured by injecting 337 nm laser light into the center of each tile [85]. The reduction in light was confirmed by comparing the HE signals during pp collisions from April 2012 to those from December 2012. Jet energy calibration studies also noticed this decline in light output.

The observed light reduction substantially exceeds expectations based on previous studies of

radiation damage using sources and electron linacs [86]. The reduction of the signal could be due to a decrease in the light emission of the polystyrene substrate of the scintillator or the two dopants, by a decrease in tile transparency, or by combination of these effects. The WLS fiber could also have either type of damage, and the clear fiber could have a decrease in transparency. The contributions from these various sources are currently under study for different kinds of scintillator tiles and WLS fibers, including those used for the HE. So far, however, the cause of the anomalously large light reduction is not understood, and can not be reliably reproduced outside of the collision hall environment.

The longevity estimates for the detector are therefore based on a parameterized fit of the data from the detector. The observed light attenuation, r, was found to be well described as a function of accumulated luminosity by $r \sim e^{\frac{-L}{D}}$, where L is the integrated luminosity, and D is a parameter which depends on the tile position. In general, D is expected to depend on tile size, the dose rate, the manufacturing process used for the tile, the handling of the tile, and on conditions of the irradiation, such as the gas content of the nearby atmosphere and the temperature. Note that both the total dose and the dose rate are much smaller for tiles near the outer radius of the HE than for those at the inner radius.

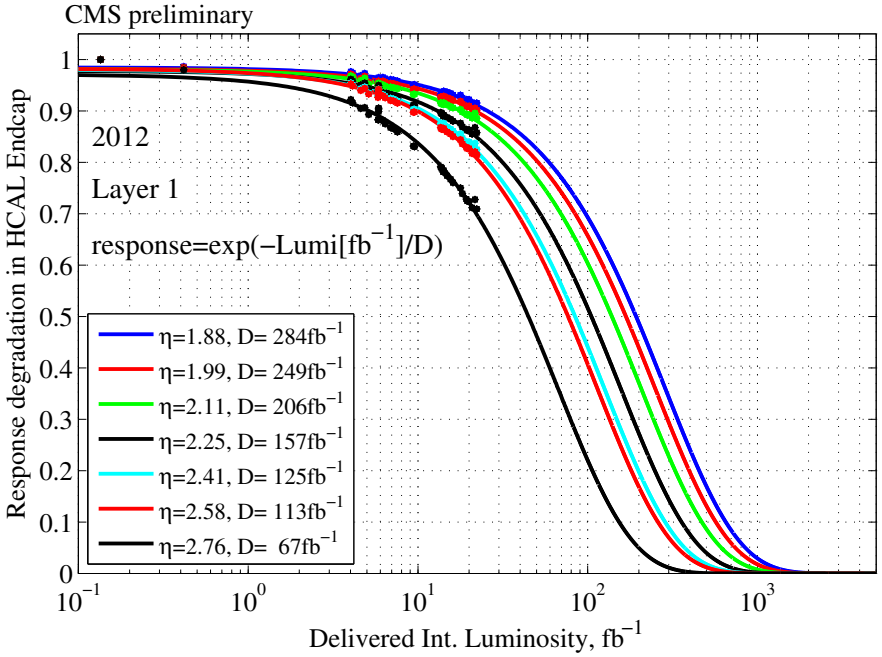


Figure 3.8: Fraction decrease of light signal from the first layer of HE as a function of accumulated luminosity for different values of the tile position (η), along with a fit to an exponential

Using the parameterization from the detector laser data and the results of the FLUKA simulation, we calculate for each tile the expected reduction of light output after an accumulation of $L = 1000\,\mathrm{fb}^{-1}$. The result is shown in Fig. 3.9.

The expected attenuation of light signals from different HB scintillating tiles after accumulation of $L = 3000\,\mathrm{fb}^{-1}$ is presented in Figure 3.10, using the parameterization derived from the HE laser data and translating the dependence on luminosity to dose using the FLUKA simulation. The HB predictions have significant uncertainties due to the large tile size of the HB tiles, the higher nitrogen/oxygen ratio in the HB environment compared with HE, and the difficulties

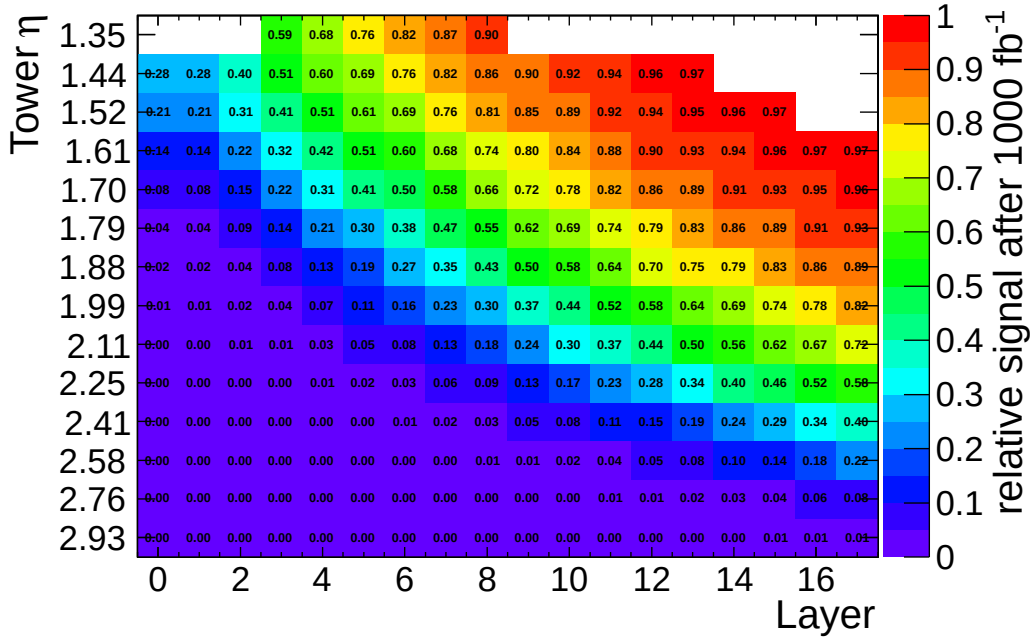


Figure 3.9: Expected attenuation of light signals from different HE scintillating tiles after accumulation of $L = 1000\,\text{fb}^{-1}$

in extrapolating from the current low-damage levels in the outer portions of the HE. The HO scintillating tiles are located outside the solenoid magnet where their radiation damage after accumulation of $L = 3000\,\text{fb}^{-1}$ will be negligible.

3.2.3.2 Radiation damage in the HF quartz fibers

The radiation dose at the location of the HF quartz fibers will reach up to 10^7Gy after an integrated luminosity of $3000\,\text{fb}^{-1}$ (Fig. 3.11). The optical transmission of the HF fibers was monitored with the measurement of an attenuation of the light reflected from the front edge of the fibers [87]. The attenuation of the HF light output during 2011-2012 LHC runs is presented in Figure 3.12. These measurements are in a good agreement with a direct measurement of a decrease of the HF signals in pp collision data with the increase of the accumulated LHC luminosity.

The observed decrease of light output from the HF cells is in agreement with the expectations based on quartz fiber radiation damage measurements [88] and on the calculated dose map in the HF shown in Figure 3.11. While the light output from the HF fibers will decrease by factor of 2.5 at $|\eta| < 4.5$, the monitoring of signal attenuation and appropriate re-calibration of the HF channels will allow acceptable performance without replacement of the quartz fibers. It should be noted that the loss of the optical transmission of the HF fibers will causes some longitudinal non-uniformity of the HF response and will increase the constant term in the F energy resolution. However this constant term is estimated to be much less than the intrinsic constant term of a quartz fiber calorimeter, caused by its extreme non-compensating nature.

3.2.3.3 Longevity of photodetectors

As part of the Phase-I upgrade of CMS, the HB and HE HPDs will be replaced with SiPMs for all channels. The specifications on radiation tolerance and longevity of the SiPMs are established to provide reliable operation up to an integrated luminosity of $3000\,\text{fb}^{-1}$ [5]. The result of a

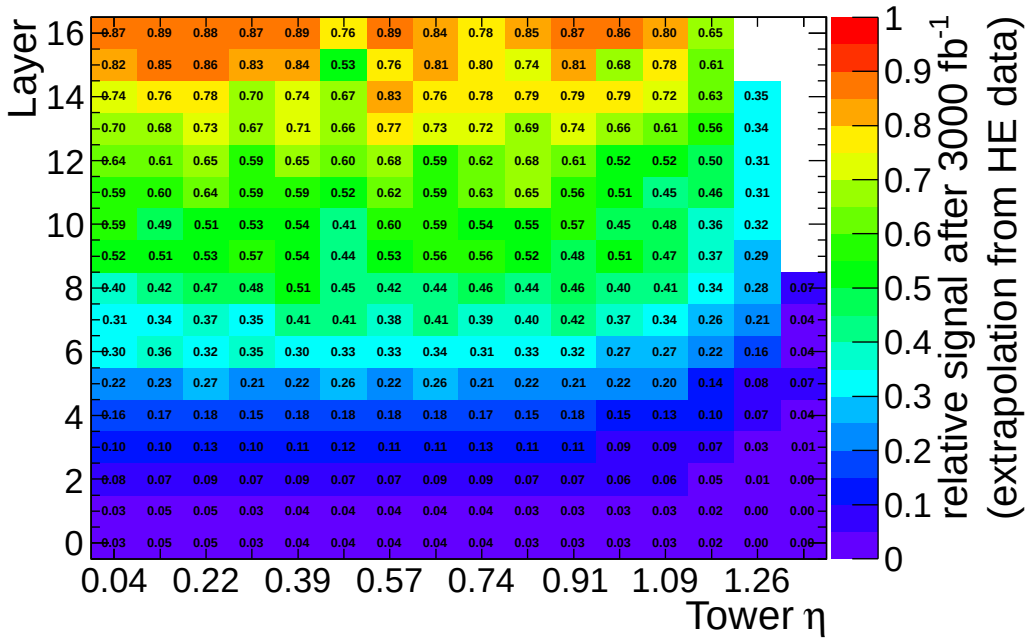


Figure 3.10: Expected attenuation of light signals from different HB scintillating tiles after accumulation of $L = 3000\,\mathrm{fb}^{-1}$

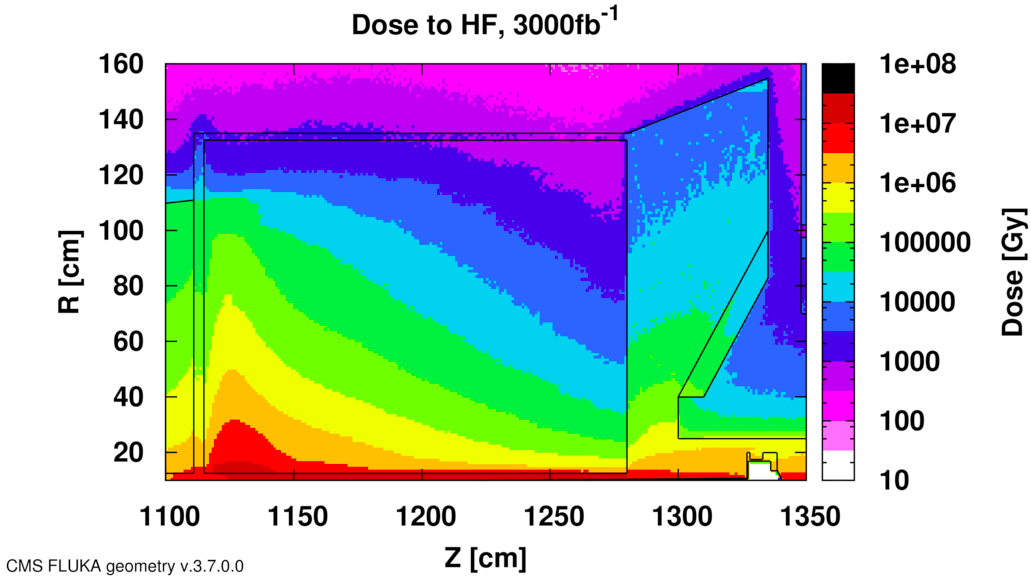


Figure 3.11: Expected radiation dose in the HF after accumulation of 3000 fb^{-1} LHC luminosity

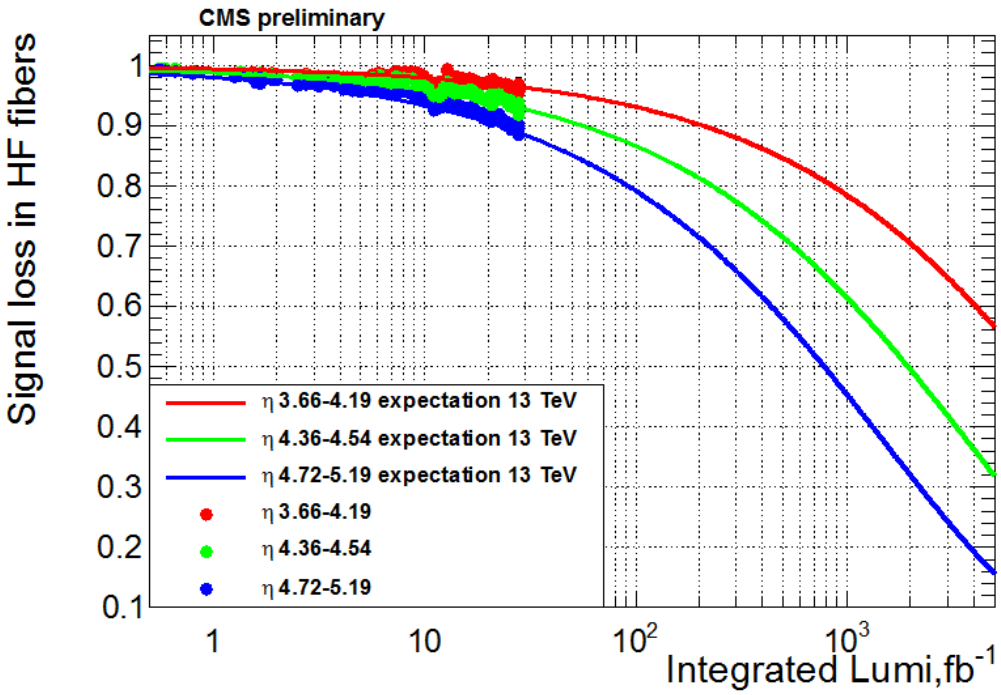


Figure 3.12: Attenuation of light signals from different HF cells as function of accumulated luminosity

recent SiPM radiation tolerance study at the operating temperature of 24^oC is shown in Figure 3.13. It was observed that the SiPM devices from multiple vendors all demonstrate an adequate radiation tolerance.

During LS1, HAMAMATSU R7600-200-M4 photomultiplier tubes (PMT) were installed in the HF. Several tests of longevity of this type of PMT have been performed prior to installation. The measured decrease of the PMT gain was less than 50% after accumulation of total anode charge of 1500 C. The maximum anode charge will be about of 500 C for operation with the gain of 7×10^5 and $L = 3000\,\mathrm{fb}^{-1}$. After irradiation with a nucleon fluence of $10^{13}/\mathrm{cm}^2$, a darkening of the PMT window led to attenuation of the PMT signal to the level of 0.82 with respect to the initial signal level. Therefore, the full accumulation of $L = 3000\,\mathrm{fb}^{-1}$ should result in only moderate attenuation of the HF PMT signals which can be corrected during signal reconstruction without any impact on the HF performance.

3.2.3.4 Longevity of FE electronics

The HB/HE Front-end electronics are located in the Readout Boxes (RBX) where photo-detectors are installed while the HF front-end electronics is located in racks on the HF platforms. All HCAL front-end electronics will be replaced as part of the Phase-I Upgrade with components designed to operate up to an integrated luminosity of 3000 fb^{-1} [5]. As an example, the Charge Integrator and Encoder (QIE) ASIC, the core component of the front-end electronics, has recently completed its radiation qualification. The results of the radiation tolerance study of the QIE10 ASIC have shown no ionization effects up to dose of 50 Gy for the low-rate irradiation with 230 MeV protons and no observable effects due to hadron damage to the level required.

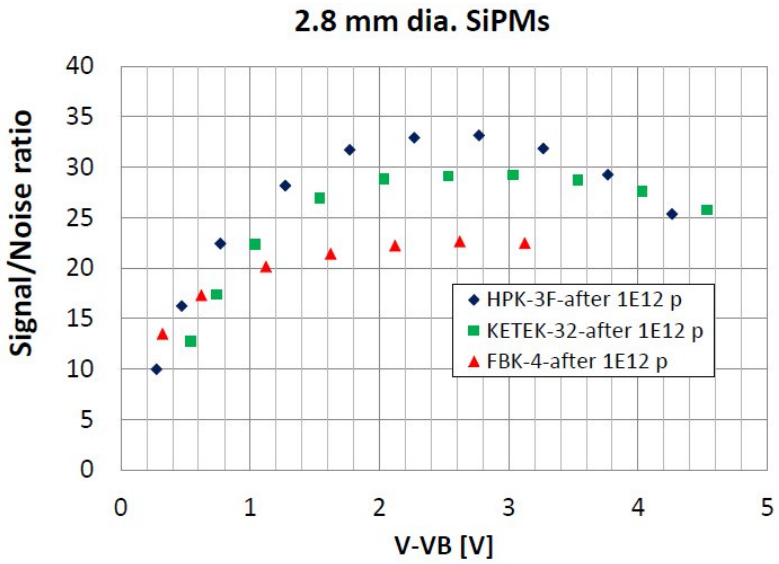


Figure 3.13: SiPM signal/noise ratio after irradiation with 62 MeV protons to a fluence of $10^{12}/\text{cm}^2$ (equivalent dose was about 140 Gy)

3.3 Performance of Existing CMS Calorimeters at High Pileup and Integrated Luminosity

3.3.1 ECAL Performance with respect to aging and pileup

In order to evaluate the ECAL performance under aging/pileup conditions, photons from $H \to \gamma\gamma$ and electrons from $Z \to ee$ decays, generated with Pythia [89], are simulated with the CMS full detector simulation. Detector aging, tuned on results from beam tests of irradiated crystals as described in Section 3.2.1, is applied to simulate a reduced crystal light output and the corresponding degradation of the stochastic, noise and constant terms of the energy resolution. Photodetector aging is also simulated, in particular VPT response loss for the endcaps and APD noise increase for the barrel. For the crystals, aging is applied as a function of the delivered luminosity (to compute the μ_{ind} due to the hadron damage) and the instantaneous luminosity (to compute the electromagnetic damage at the equilibrium level given by the dose rate). The simulated scenarios correspond to integrated luminosities of 0, 1000 and 3000 fb^{-1} with average pileup conditions of 50 for the first sample and 140 for the others, corresponding respectively to the conditions of the unaged detector with Phase-I maximum pileup, a typical point of Phase-II, and the end of Phase-II. The instantaneous luminosity used for the aging model is assumed to be $5{\cdot}10^{34}\text{cm}^{-2}\text{s}^{-1}$ for the last two samples. The Run-I detector geometry is used in all cases.

An electromagnetic shower from a photon or an electron is spread over several ECAL crystals. The presence of material in front of the ECAL (corresponding to 1–2 X_0 depending on the η region) causes conversion of photons and bremsstrahlung from electrons and positrons. The strong magnetic field of the experiment tends to spread the radiated energy along ϕ within the tracker volume. Dynamic clustering algorithms [68] are used to sum together energy deposits in crystals belonging to the same electromagnetic shower and to recover the radiated energy. They proceed first with the formation of "basic clusters", corresponding to local maxima of energy deposits. These are then merged to form a supercluster (SC), which is extended in ϕ, to recover the radiated energy.

Energy deposits from pileup events have an effect on both pulse amplitude estimation and cluster reconstruction. The signal pulse shape is distorted by low-amplitude pileup pulses that overlay it. In-time and out-of-time pileup both contribute as if they were additional noise. The mean level of the pedestal is shifted to higher values due to the pileup. The pedestal is presently calculated from the 3 samples before the pulse, therefore the constant level of the pileup energy deposit can be subtracted, but the event-to-event fluctuations remain. The contribution to the single-crystal energy fluctuation has been estimated in terms of σ_{eff} (defined as half of the width of the smallest interval of the distribution containing 68.2% of the events) with minimum bias simulated events, and is shown in Fig. 3.14 as a function of η. It ranges from 80 MeV at $|\eta| = 0$ to 2 GeV at $|\eta| = 2.5$. For comparison, at the start of Run-I the electronics noise levels were 42 MeV and 120 MeV in EB and EE respectively. After 1000 fb^{-1}, the APD noise dominates over the PU contribution in the EB.

The performance of the dynamic clustering algorithm after aging and with pileup has also been studied. The cluster size is extended because of low-energy deposits surrounding the shower core that mimic radiated energy. The number of crystals included in a supercluster is 45% larger at 140 pileup than at the present Run-I levels. This effect is also visible in shower shape variables, which show wider clusters with increasing pileup or noise. The pileup energy fluctuations also degrade the energy resolution which is particularly significant in the forward region. The clustering algorithms require a complex tuning dependent on the aging and pileup to achieve optimal performance. This will be shown in Chapter 9. However we can illustrate the principal effects on identifying photons and electrons by using a simplified clustering algorithm. In a typical shower most of the energy is located in a 3×3 matrix of crystals about the shower center. We use a fixed 5×5 matrix to fully capture the shower energy. This algorithm is known to provide a good approximation to an optimal dynamic clustering as it is less sensitive to the fine structure of the pileup. In the following studies the reconstructed 5×5 crystal photons and electrons are required to be matched in (η, ϕ)-space within a cone of 0.15 to the generated particles, to have $p_{\mathrm{T}} > 30$ GeV and either $|\eta| < 1.444$ or $1.566 < |\eta| < 3.0$ to avoid the transition region between the barrel and endcap.

Fig. 3.15 shows the energy resolution for photons as a function of the pseudo-rapidity η for different aging and pileup conditions. The energy resolution in the EB is sensitive to the APD noise increase, which becomes the dominating factor after 1500 fb^{-1}. In the EE the energy resolution degrades substantially after Phase-I, particularly for the region $|\eta| > 2$.

Both noise increase and pileup also affect the reconstruction efficiency and selection criteria, e.g. isolation variables, which are typically utilized for photon and electron selections in physics events. The selection efficiency decreases with increasing noise or pileup, particularly in the high-η endcap regions, and the fake rate for photons increases at high aging and pileup. Figure 3.16 shows the selection and reconstruction efficiency for photons from the H boson decay versus background efficiency estimated using the jet in γ+jet events. The selection criteria were chosen utilizing standard selection variables for photons (shape variables, ratio of energy in the ECAL with respect to the HCAL and isolation). The specific selection cuts for each sample (0, 1000, and 3000 fb^{-1}) were optimized by scanning the phase space of possible cut values in order to have a selection efficiency of 80% for the signal photons. The corresponding efficiency for jets is shown in the plot. Low background efficiency values can be achieved for jets in the Barrel, even in the presence of high pileup. In the endcaps a jet efficiency of 2% is obtained for a selection efficiency of 80%, and it almost doubles with ageing and pileup. The effects of aging and pileup severely degrade the ability to discriminate between photons and jets in the entire forward region. The clusters become broader, less isolated and have a larger hadronic component which enhances the probability of a jet that has fluctuated to have a large

electromagnetic energy component to be mis-identified as a photon. The increase in fake rate observed in the endcaps for 140 PU and Phase-II conditions compromises their exploitation in physics analyses with photons and electrons due to the high background.

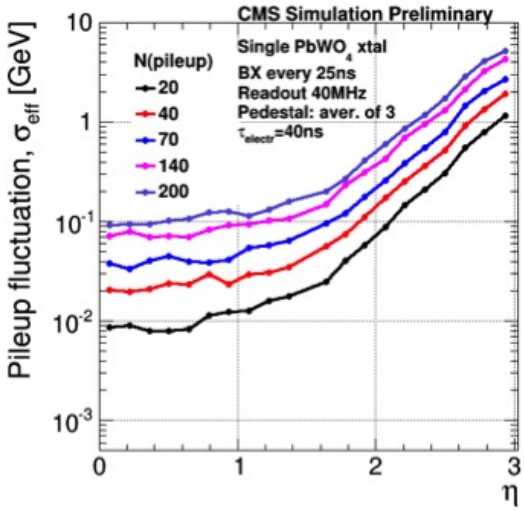


Figure 3.14: Pileup contribution to energy resolution in terms of an effective noise term(σ_{eff}) as a function of η for different N(pileup) scenarios. Both in-time and out-of-time pileup is taken into account.

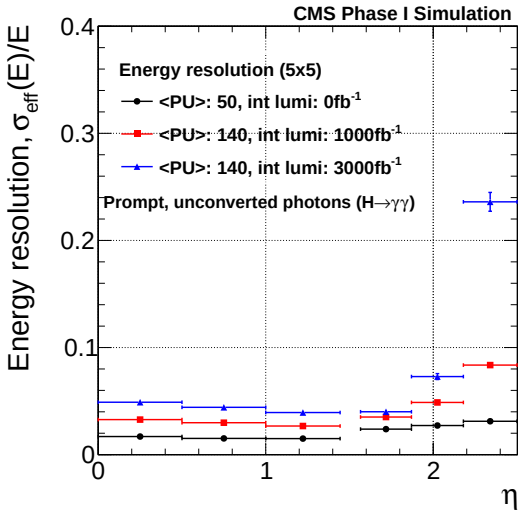


Figure 3.15: The energy resolution $\sigma_{eff}(E)/E$ for photons from the Higgs boson decay for different integrated luminosities and pileup as indicated in the figure.

3.3.2 HCAL Performance

Based on the extrapolation of the results of the in-situ measurement of light loss in HE, the expected attenuation of light signals from different HE scintillating tiles after the accumulation of various amounts of luminosity up $L = 1000\,\text{fb}^{-1}$ is included in the CMS detector simulation.

The signal decrease can be compensated in reconstruction by adjustment of the calibration factors for each channel, at the cost of increasing the electronics noise by the same factor. For

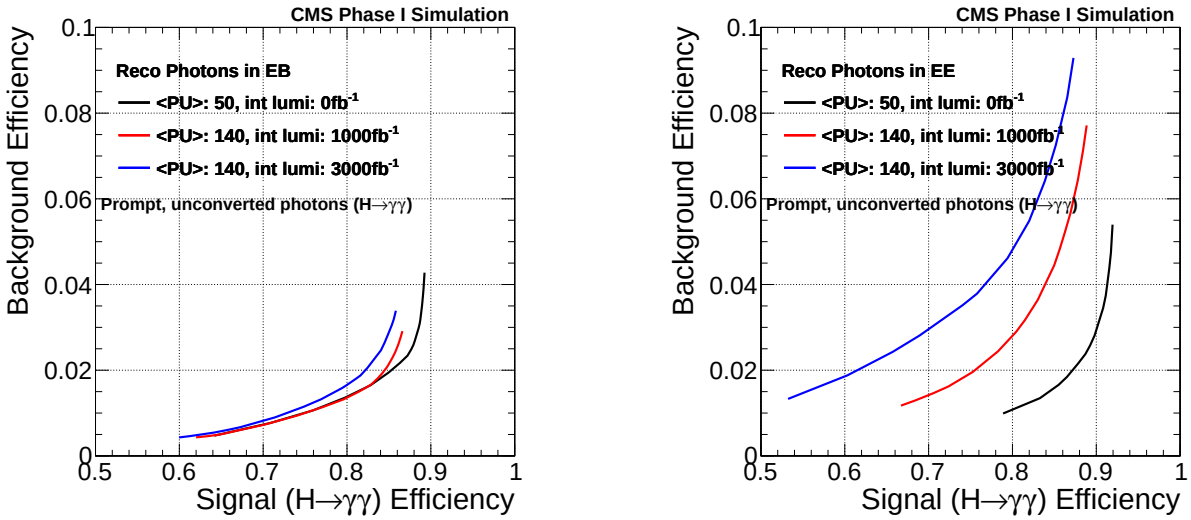


Figure 3.16: Selection efficiency for photons from the H boson decay versus background efficiency estimated with the jet in γ+jet events for the ECAL Barrel (left) and Endcap (right) for three MC samples which simulate: unaged detector at the end of Phase-I pileup conditions, conditions early in HL-LHC operation, end of Phase-II conditions.

calibration factors greater than 100, the electronic noise contribution to the reconstructed signals becomes large compared to the typical energies deposited.

The effects of the expected severe radiation damage to the HE on jets reconstructed using information from the calorimeter system is shown in Fig. 3.17 for 1000 fb^{-1}. Jets deposit energy in both the ECAL and HCAL, however to isolate the effect of the HCAL this plot has been produced with the EE with its LHC Run-I performance. Dijet events, requiring one jet in the endcap pseudorapidity region, were simulated without pileup. The zero-suppression thresholds for the silicon photomultipliers (SiPMs) in the HE were modified to keep the occupancy close to 20%. Recalibration factors, up to a maximum factor of 100, were applied to the HE reconstructed energy deposits. During the jet reconstruction procedure, energy deposits in projective towers were first summed. Thresholds on the tower energies were modified to account for the increased SiPM pedestal widths due to the radiation damage and the resulting HE recalibration factors. Standard CMS jet energy scale corrections can not be used for a detector under these conditions. We recalculated the noise offset, relative, and absolute jet energy corrections and applied them to the jets in order to ensure that the average p_{T} response is unity. The jet p_{T} resolution is significantly degraded after 1000 fb^{-1}, indicating the necessity of rebuilding the HE for Phase-II.

3.3.3 Impact of Aging and Instantaneous Luminosity on Triggering in ECAL

The radiation damage of the crystals and the noise increase of the APDs will affect the trigger performance of the ECAL as well as the offline reconstruction. The trigger algorithms of the ECAL are established in the front-end electronics and there is very limited flexibility on the algorithm and no ability to suppress individual crystals from the sum.

In the endcaps, the trigger sums are calculated in fixed sets of crystals. With the loss of light transmission, the per-crystal noise will rise up to the level of 40 GeV in the forward parts of the detector, leading to a trigger tower noise of 200 GeV which corresponds to an E_{T} noise of 25 GeV. The noise distribution is shown on the left side of Fig. 3.18. Such a high noise level will

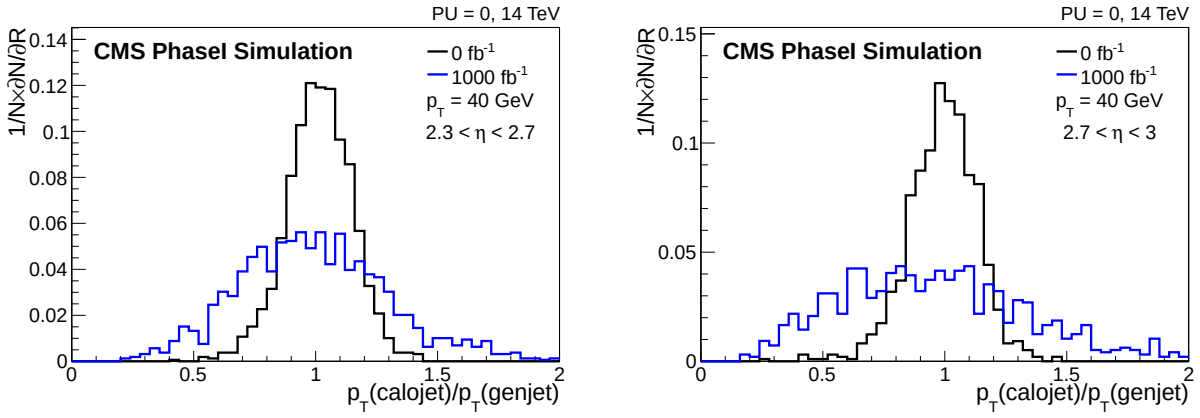


Figure 3.17: Comparison between the performance of calorimeter jet reconstruction using the undamaged detector at 0 fb^{-1} to the aged HCAL endcap detector after 1000 fb^{-1}. The comparison is made for $p_{\mathrm{T}} = 40\,\mathrm{GeV}$ quark jets in the endcap calorimeters, generated with $2.3 < \eta < 2.7$ (left) and $2.7 < \eta < 3.0$ (right). The ECAL endcap is not aged in this comparison. Noise offset, relative, and absolute corrections are applied to the jets. No pileup is included.

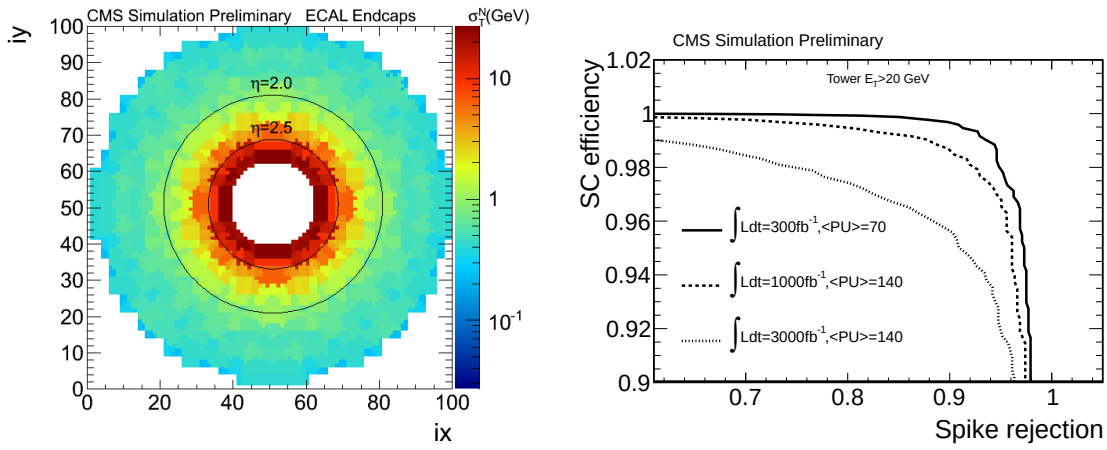


Figure 3.18: Left: Equivalent transverse energy of the electronics noise (σ_T^N) per trigger tower in the endcaps after 3000 fb^{-1}. Right: Predicted efficiency of the online spike killing algorithm versus signal acceptance (EM showers Z→ee events) for a range of detector ageing and event pileup conditions. Only towers with $E_{\mathrm{T}} > 20$ GeV are considered.

severely affect both the jet and electromagnetic triggers and have a significant impact on any missing-E_T triggers.

In the barrel detector, the trigger primitives are affected by "spike" events. Spikes are anomalous isolated signals, with equivalent energies that can exceed 100 GeV in CMS, observed in the barrel during LHC proton-proton (pp) collisions, with a rate that is proportional to the collision rate of the proton beams. These deposits are understood to be associated with particles (produced in pp collisions) striking the APDs and very occasionally interacting to produce secondaries that cause large anomalous signals through direct ionization of the silicon. Rejection of spikes in the present CMS Level-1 trigger is achieved by exploiting additional functionality of the ECAL front-end electronics, the strip Fine-Grained Veto Bit (sFGVB). This provides a single feature-bit for a given trigger tower (5x5 crystal array, corresponding to a single readout unit of the ECAL front-end electronics) and is configured to return zero for an isolated energy deposit (a single channel above threshold) or one for an EM-like energy deposit (multiple channels above threshold). If the sFGVB is set to zero, and the trigger tower transverse energy is greater than a configurable threshold, the energy deposition is considered spike-like and the tower will not contribute to the triggering of CMS for the corresponding event.

The performance of the sFGVB algorithm has been studied for the conditions expected during HL-LHC operation using Monte Carlo events where a simulation of the APD geometry has been implemented into the standard CMS Monte Carlo event generator. A digitization model has been implemented, taking as input the GEANT energy depositions recorded in the two APD active layers. The model was tuned on early minimum-bias triggered data at 7 TeV and the qualitative features of the spikes observed in data are well reproduced. Figure 3.18 (right) shows the predicted performance of the online spike killing algorithm for conditions expected at the end of LHC Phase-I, and during HL-LHC. It shows the efficiency of the current sFGVB algorithm in rejecting spikes versus the efficiency for accepting true EM showers (superclusters from Z→ee events), for Level-1 electron/photon candidates with transverse energy greater than 20 GeV. The target performance for this algorithm is a greater than 95% efficiency for rejecting spikes (to minimize the rate of spurious triggers) and a $\leq$ 1% inefficiency for EM showers (to maintain high signal acceptance). The simulations predict that this level of performance can be achieved for Phase-I conditions (integrated luminosity of 300 fb^{-1}), but cannot be maintained during HL-LHC operation.

3.4 Barrel Calorimeter Upgrades

3.4.1 ECAL Barrel Calorimeter Upgrade Overview

The primary driver of the ECAL barrel calorimeter upgrade is the trigger requirements for a trigger latency of 12.5 μs and a Level 1 trigger rate of 750 kHz. At the same time, the opportunity will be taken to optimize the electronics and operating temperature of the system to adjust for ongoing radiation-induced noise increase in the APDs.

The upgrade conceptual design for the electronics has been developed using the specifications and constraints given in Table 3.1. The basic performance specifications for the new electronics are the same as the existing system as the crystals/APDs will not change. The dynamic range should be between a few tens of MeV to 2 TeV equivalent electron/photon energies. The lower bound is defined by the intrinsic noise from the APDs, currently around 40 MeV. This will increase with radiation, as discussed in Section 3.2.1.2, but the upgrade will limit the noise to around 200 MeV through a lowering of the temperature from 18 to 8°C and an optimization of the preamplifier shaping time from the present 43 ns to around 20 ns.

The cooling of the crystals and APDs is discussed in more detail below in Section 3.4.2. The new electronics is described in detail in Section 3.4.3. It will allow improvements in the triggering and also reductions in the effect of APD noise and out-of-time pileup. The electromagnetic trigger improvements are described in Section 3.4.5.1.

3.4.2 ECAL Barrel APD noise mitigation by cooling

As shown in Section 3.2.1.2, due to the increase in APD dark current, the noise in the ECAL barrel will progressively become the largest contribution to the energy resolution for photons from the Higgs boson decay. A substantial reduction in the APD dark current can be achieved by cooling the APDs, thus reducing the electronic noise contribution to the resolution. In particular a temperature decrease from 18°C to 8–10°C will halve the APD dark current and consequently the noise will be reduced by about 30%.

The ECAL cooling system provides temperature stabilization for the photodetectors, as well as for the crystals and electronics. $PbWO_4$ crystals produce $\sim$ 30% higher light output at 8–10°C with respect to 18°C, and the scintillation decay time increases by about 20% [90, 91]. The Panda experiment has studied radiation-induced light output loss in a range of temperature between -25°C and 20°C [92]. Even though their measurements were aimed at running the detector at -25°C, the approximate behavior of $PbWO_4$ crystals at intermediate temperature can be established from their studies. The spontaneous $PbWO_4$ light output recovery from gamma-irradiation at 8–10°C is less effective: the light output loss is larger (25% at 8–10°C compared to 10–15% at 18–20°C) and the time evolution of the radiation-induced light output becomes slower. However, since the light output is larger at lower temperature, even in the presence of a larger damage, overall the light output is larger at 8–10°C by about 10% compared to 18°C. The dynamics of the $PbWO_4$ radiation damage at 8–10°C will be studied in detail with the Panda experimental setup, in order to optimize the operating temperature of the calorimeter for the Phase-II operation.

Table 3.1: Basic principles and constraints for the upgraded EB electronics

Principle/Constraint	**Implications**
Single crystal readout at 40 MHz	• no deep on-detector pipeline memories • no on-detector trigger-primitive generation • no L1 trigger latency or rate restrictions
No change to on-detector mechanics and use 5x5 groups of crystals as basic unit	• power dissipation $\leq$ present electronics ($\approx$2.7 W/crystal) • new VFE/FE: same form-factors as existing boards
Local LV regulation: DC-DC converters	• low on-detector power dissipation
Use ongoing developments if possible	• GBT/Versatile links for control/readout • xTCA as baseline for off-detector boards

The ECAL barrel requires a temperature stability of $\pm 0.05^{\circ}$C, which is obtained through a multi-stage cooling system. The Phase-I operating temperature of the calorimeter, 18°C, is obtained using "mixed water" (demineralized water at 13°C) as cooling fluid. The water is then first regulated at 17.8°C and then adjusted by a heating system which provides the fine regulation. A reduction in temperature below 13°C is possible by replacing the "mixed water" with a chilled water source (demineralized water at 5°C). This should allow the detector to be cooled to about 8–10°C, leaving a margin for temperature regulation in the pumping station and heater. This change will require replacing the heat exchanger, pump system, control system and heater.

All cooling pipes must be properly insulated to avoid condensation as the typical dew-point in the cavern is around 14°C. In particular, the cooling pipes from the experimental cavern balconies to the detector must be replaced. Other measures to reduce condensation may be necessary, such as dry air/N_2 flow through the pipe ducts, similar to the CMS Tracker. The dew point for the detectors inside the vacuum tank is presently very low (around 0°C) when the Tracker dry air/N_2 system is active. The present humidity monitoring system will be upgraded in order to be able to safely operate the detector at lower temperature. Further engineering studies are planned to understand the various boundary effects at the EB/HB boundary and the external end of the supermodules. We are studying, with a test setup, the risk of damages on the alveoli in case of condensation. Mechanical stresses inside the supermodules will be simulated and understood, taking into account the higher ΔT between the new crystal temperature and the required 18°C external surface temperature of the supermodules. It is expected, however, that these issues can be satisfactorily resolved with appropriate engineering attention.

3.4.3 ECAL Barrel electronics upgrade

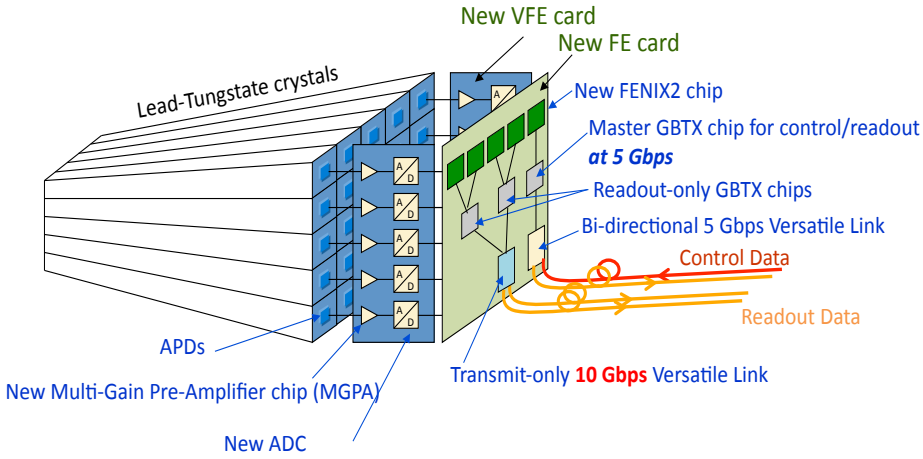


Figure 3.19: A possible new architecture for the upgraded ECAL Barrel Electronics.

The ECAL barrel electronics will be replaced to meet the latency and bandwidth requirements of the Phase-II Level 1 trigger system. This upgrade will also allow the provision of individual-crystal information to the Level 1 calorimeter trigger. The impact of this improvement is discussed in more detail in Section 3.4.5.1. By including the VFE in the upgrade, the pulse shaping and sampling can be optimized to mitigate the increase in APD noise and pileup, as well as enhance online spike suppression and improve timing resolution. Finally it will reduce the risk of using electronics for 30 years and 3000 fb^{-1} that were designed for 10 years running and 500 fb^{-1}.

The APDs are connected to the VFE through a passive motherboard (MB). Neither the APDs nor the MBs will be replaced as there is no identified concern with these items (from accelerated aging or irradiation tests to 3000 fb^{-1}) which would justify the risk of disconnecting the cooling block and cooling system joints. The performance goals of the EB upgrade can be achieved through the replacement of the VFE and FE cards, together with the associated low-voltage distribution system and optical links.

Figure 3.19 shows a possible new architecture. The individual boards will follow the same configuration and form factor as the present electronics shown in Figure 3.6 in order to fit into the same physical space and use the existing services as far as possible.

The VFE will be upgraded to give more optimal noise performance and to help discriminate against spikes using timing information. The shaping time will be made shorter to significantly reduce the effective noise and out-of-time pileup effect. A shorter shaping time will allow better discrimination between the different pulse shapes of signals from scintillation light in the crystals and those from "spikes". Shorter shaping time, combined with improved timing distribution, will allow the realization of the asymptotic precision of the existing detector for large-energy hits [93]. This will also help in the determination of pulse arrival time for spike rejection, pileup suppression and vertex association. Alternatively it may be possible to use charge integration employing an evolved version of the QIE10 chip being used in the Phase-I CMS HCAL upgrade [5] which also incorporates a TDC functionality. The increase in APD noise also means that the energy associated with least-significant bit may increase and this will reduce the dynamic range requirement on a new ADC.

In the original design of the present ECAL electronics it was planned to provide the trigger with single-crystal information. This requires 16 bits (14 bits plus 2 bits for timing and error checking) times 25 crystals times 40 MHz = 16 Gb/s. However this was not possible due to the limited 800 Mb/s bandwidth of the optical links that transmit the trigger information from the FE to the off-detector trigger concentrator card (TCC). It would have required twenty links per trigger tower and the financial expense, power consumption and space constraints were prohibitive. Currently rad-hard 10 Gb/s optical links that are radiation hard are under development at CERN and it is envisioned that it will be quite feasible to transmit all the crystal information to the off-detector electronics at 40 MHz.

The system will possibly use 0.13 μm CMOS chips for the replacement ASICs in the VFE/FE cards. This technology is naturally radiation tolerant and provides a power consumption of up to 75% less than the equivalent 0.25 μm device along with a 20% reduction in size [94]. New ASICs will require a different bias voltage to that currently supplied by the low-voltage regulator card (LVR). The new LVR should exploit the radiation-hard "point of load" DC-DC converters that have recently been developed for use in the LHC experiments [95]. These compact devices are highly efficient and have low noise output. They will allow the power consumption of the front end to be reduced to about 50 kW (around half of the present value) and significantly reduce the amount of copper cabling required for power transmission.

For the control and distribution of the clock, the trigger commands, and FE and VFE settings and configurations, the GBTX control bus will be used, simplifying the existing scheme which uses a specific control board (CCS). Each control link will serve a group of 25 crystals in the barrel. An additional redundant link would also be highly desirable to facilitate recovery from a bad link.

All precautions must be taken during the design phase of chips and boards, in order to make sure that both the chosen architecture is intrinsically radiation-hard and the design itself is

resistant with respect to radiation and single-event-upsets. Chip pin layout and power dissipation must be designed with the goal of long-term performance (more than 10 years). All chips and boards will be tested also with respect to radiation hardness and accelerated ageing during the validation phase.

The off-detector electronics will also need to be upgraded to accommodate the higher transfer rates and to shift the trigger primitive generation off-detector. These processors do not need to be radiation-hard, as they will be located in the service cavern, so commercially-available FPGAs will be employed. These processors will be powerful enough to accommodate all functions presently performed by the ECAL off-detector electronics. For example the MP7 [96] μTCA board being developed for the CMS Phase-I trigger upgrade could fulfill these requirements with some customization.

The specific architecture will be chosen to maximise overlap with choices made elsewhere in CMS and will be customised to fit specific ECAL needs, for instance it is desirable to fit all the input channels from one supermodule in a single board. The chosen board must have a powerful programmable processing unit to analyze the received data and transmit to the next level of the trigger a pre-processed set of trigger primitives. The algorithms that must be implemented in this board include the rejection of anomalous ECAL signals and basic clustering of localised energy.

3.4.4 Technical Aspects for the Barrel Electronics Upgrade

The barrel is constructed of 18 super-modules (SM) in each half of the detector. It is necessary to extract each SM in order to replace the electronics. A fixture known as an "enfourneur" is used to facilitate fast extraction and re-installation of the super-modules. One of these fixtures currently exists and a second will be built to allow simultaneous SM extraction or insertion at both ends of CMS.

Once extracted the SM will be placed on a support frame, which allows the opening of the stainless steel cover at the back to access the electronics. Six support frames are already available from the construction time. Simulations of radiation exposure including precise dimensions and material content have been performed to assess the activation levels of the SM. A very light shielding on the front part of the SM, around the crystals, may be necessary, but the electronics located at the back can be manipulated safely.

The time required to disassemble and refurbish the electronics up to the level of the VFE and LVR boards is estimated to be 7 weeks per SM. Once completed, the SM electronics must be tested and commissioned for at least 2 weeks. This operation will be done via several parallel lines working on distinct SM, such that the 36 SM will be fully refurbished and tested in 52 weeks. Each line typically requires three workers.

The low voltage and high voltage system must be upgraded. In particular the low voltage system will have different characteristics, due to the different power consumption. The requirements for the high voltage system will be similar to the present one, allowing for a slightly higher APD current. The cooling plant, including the cooling apparatus, the pipes and all the sensors may be designed starting from 2018. Cooling simulation work may start already now, to assess the feasibility and implications of operating the detector at a lower temperature. The Detector Control System (DCS) must be upgraded in order to take care of all safety aspects of the operation at lower temperature, in particular dewpoint and temperature change in time. More humidity sensors must be installed in the electronics compartement and on the front of the supermodules.

3.4.5 Performance of the Upgraded Barrel Electromagnetic Calorimeter

3.4.5.1 ECAL Trigger Improvements

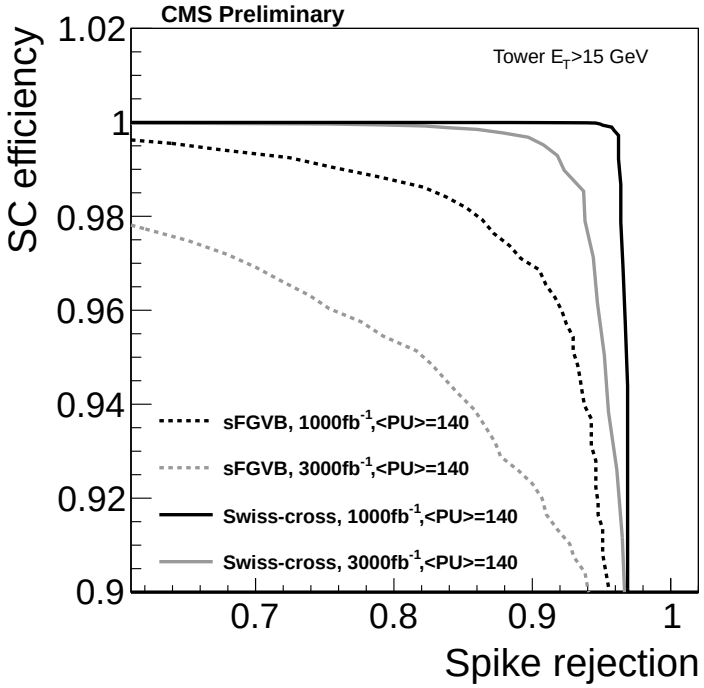


Figure 3.20: Predicted improvement in efficiency/rejection performance at $1000\,\mathrm{fb}^{-1}$ and $3000\,\mathrm{fb}^{-1}$ when single-crystal information is used to reject spikes via the "Swiss-cross" algorithm compared with the sFGVB. Only towers with $E_\mathrm{T} > 15$ GeV are considered.

As discussed above, the upgraded ECAL on-detector electronics will provide single-crystal information to the calorimeter trigger calculations rather than the current 5×5 crystal sums. Figure 3.20 shows the improvement in spike rejection possible with the single-crystal information provided to the L1 trigger by the upgraded electronics. The plot is computed for spike and Z→ee samples with an integrated luminosity of $3000\,\mathrm{fb}^{-1}$, a mean pileup of 140, and for L1 candidates with an E_T threshold of 15 GeV. It shows that the online spike-killer can be significantly improved when single crystal information is used to compute an energy-sharing variable that discriminates between spikes and electromagnetic showers. The "Swiss-cross" algorithm compares the energetic crystal with energy from the four crystals that each share a side with the energetic crystal. For a true shower, these neighboring crystals should have $\approx 20\%$ of the energy of the seed crystal while for a spike the fraction is usually much smaller as the crystals have only noise deposits. Efficiencies that are close to the target level of performance can be attained using this algorithm. However, the Swiss-cross algorithm degrades at high integrated luminosity as shown in Figure 3.20 due to increasing APD noise.

Additional measures are being explored to provide further robustness of the online spike killer against pileup and aging. By cooling the APDs as described in Section 3.4.2 the noise due to the dark current can be reduced. In terms of spike killing performance, this is roughly equivalent to the difference between the $3000\,\mathrm{fb}^{-1}$ and $1000\,\mathrm{fb}^{-1}$ curves in Figure 3.20.

Further suppression of spikes is possible by exploiting differences in spike and EM pulse shapes (spike signals have no scintillation component) to provide improved discrimination for conditions with high pileup and larger electronics noise. A similar technique, using the reconstructed timing of the spike and EM signals, is already used in the CMS offline reconstruction

to reject spikes. That algorithm is predicted to provide $\sim$ 99% spike rejection with a $\sim$ 99% acceptance for EM signals, assuming an integrated luminosity of 3000 fb^{-1}, a mean pileup of 140, and for energy deposits with transverse energy greater than 10 GeV. Implementing this algorithm online will require the modification to the pulse shaping as proposed for the new VFE card.

3.4.5.2 Energy resolution of the Upgraded Barrel Electromagnetic Calorimeter

Figure 3.21 left shows the expected noise levels in the ECAL barrel as a function of the integrated luminosity at $\eta = 1.45$ if operating the detector at 18°C (red curves) or at 8°C (blue curve), with the present electronics (continuous line, shaping time τ =43 ns), or the upgraded electronics (dotted line, shaping time τ =20 ns). The upgrade plan for the ECAL Barrel mitigates the radiation induced noise increase by rejuvenating the calorimeter of a factor of more than three in integrated luminosity.

Figure 3.21 right shows the energy resolution in the ECAL Barrel for unconverted photons from the Higgs boson decay. This study has been performed with full simulation of the detector effects including the ageing and the planned upgrade. The various curves correspond to different levels of aging and pileup, as it is also shown in Fig. 3.15. The green curve represents the effect of the aging due to 1000 fb^{-1} of integrated luminosity mitigated by the barrel electromagnetic calorimeter upgrade (EB operated at 8°C and shorter shaping time). The apparent energy resolution improvement versus η is due to the fact that the average energy of the photons from the Higgs boson decay increases with η, and thus the importance of the noise effect on the energy resolution decreases. For a fixed energy spectrum, instead of fixed transverse energy, the energy resolution worsens slightly with η.

These figures show how the planned upgrade mitigates the radiation-induced noise increase, such that the noise does not dominate the energy resolution for photons from the Higgs boson decay throughout the duration of the HL-LHC program.

3.4.6 HCAL Barrel Calorimeter Upgrade

As shown in Figure 3.10, the HCAL Barrel Calorimeter will have significant light loss by the end of HL-LHC. This light loss, particularly in the front layers, will affect electron and photon identification significantly as well as reducing the resolution of the detector for the reconstruction of neutral hadrons in jets. The Phase-I upgrade will provide a significant improvement in longevity, as the SiPMs have a photon-detection efficiency which is at least a factor of 2.5 greater than the original HPD photodetectors. However, the replacement of some of the megatiles closest to the beam-pipe is planned to occur during LS3 when the central detectors are cleared away and there is sufficient access for this activity.

The proposed technology for replacement of the megatiles is doubly-doped plastic scintillator. In typical applications of plastic scintillator technology, the primary dopant concentration is tuned to balance the competing effects of light production and self-absorption processes which both increase with increasing concentration. In the presence of radiation damage, a higher dopant concentration can result in a more-stable, though lower, light output. This hypothesis has been tested using specially produced variations of a commercial scintillator, EJ-200, a plastic scintillator with polyvyniltoluene as a base. The emission spectra, as a function of wavelength, of the scintillator samples with different concentrations of the primary dopant are shown in Fig. 3.22, before any irradiation and after an irradiation (Co-60) up to a dose of 50 Mrad. It is important to note that the samples used are 1x1x5 cm^3 rods, too small for observing effects related with modifications of the attenuation length of each sample. Increasing

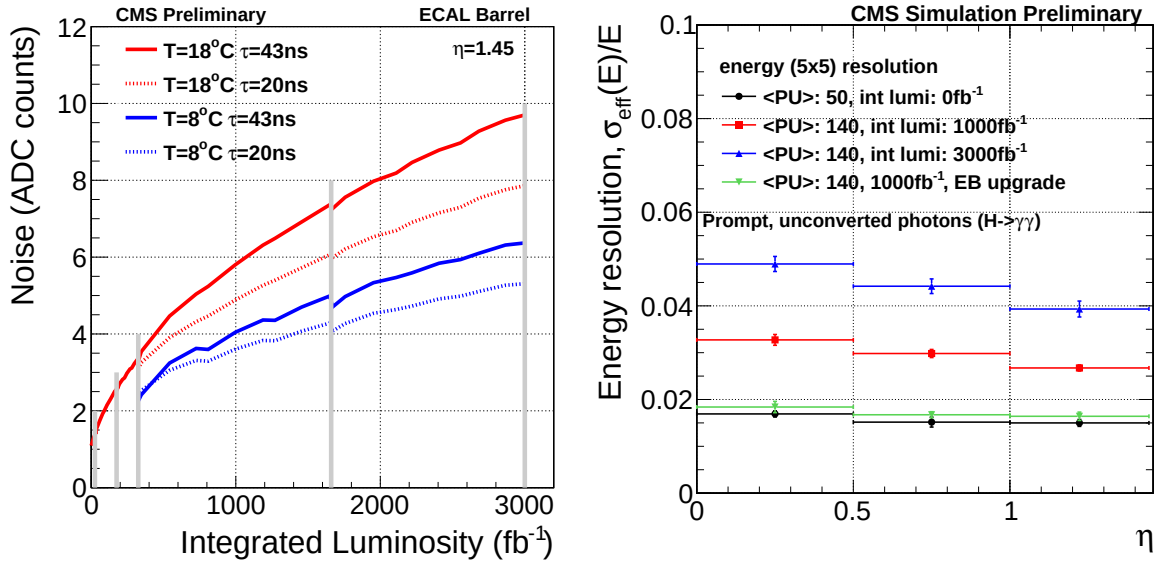


Figure 3.21: (Left) Expected noise level in the ECAL Barrel versus integrated luminosity at $\eta = 1.45$ if operating the detector at 18°C (red curves) or at 8°C (blue curve), with the present electronics (continuous line, shaping time τ =43 ns), or the upgraded electronics (dotted line, shaping time τ =20 ns). (Right) Energy resolution $\sigma_{eff}(E)/E$ for photons from the Higgs boson decay for different integrated luminosities and pileup, showing the resolution improvement provided by the upgrade to the barrel electromagnetic calorimeter (EB operated at 8°C and shaping time τ =20 ns).

the concentration of the primary dopant clearly decreases the light output of a non-irradiated sample (self-absorption is in fact enhanced), but allows the sample to maintain a more-constant light output after being irradiated.

To understand the effect of self-absorption over longer distances, 10x10x0.4 cm^3 tiles were built with over-doped plastic scintillator, equipped with Y11 fibers, and their light output (using cosmic rays) compared with similarly sized SCSN-81 tiles. The light output is very similar, thus demonstrating that over-doping is not affecting the light attenuation enough to be an issue.

The replacement megatiles will be compatible with the Phase-I HCAL Upgrade front-end electronics and photodetectors[5]. It may be necessary to adjust the optical decoder units in the EB readout modules to tune the segmentation depending on exactly which megatiles are replaced. The Phase-I off-detector electronics may not be compatible with the full trigger bandwidth, but the readout and trigger primitive generation capabilities will likely be merged with the ECAL barrel off-detector electronics. The inclusion of the HB data represents only a 17% fiber capacity increase (or a 10% bandwidth increase) for the EB electronics.

3.5 Endcap Calorimeter Upgrade

For an integrated luminosity of 3000 fb^{-1} and in the region $\eta \sim 3$, the electromagnetic calorimetry near shower max will sustain integrated doses of 1.5 MGy (150 Mrads) and neutron fluences of 10^{16} n/cm^2. Integrated doses at the location of the front layers of the existing HE are expected to reach 300 kGy (30 Mrads). At the same time, the effects of pileup will become ever more severe, making the identification of electromagnetic objects more challenging and swamping the

relatively-isolated VBF and VBS jets with increasing QCD multijet background.

To address these challenges, CMS proposes the replacement of the endcap calorimeters, which cover $1.5 < |\eta| < 3.0$, with a new high-granularity sampling calorimeter. The proposed design encorporates a silicon/tungsten electromagnetic section followed by two hadronic sections, both using brass as the primary absorber material. In the front section, the active material is silicon while the back section uses plastic scintillator. The design is targetted to achieve very high performance for physics objects reconstructed in the presence of high levels of pileup. The high transverse and longitudinal granularity of the detector allows tracking and separatation of signal and pileup particles inside the calorimeter.

Recent advances in silicon detectors in terms of cost per unit area and radiation tolerance, and advances in electronics and data transmission suggest the possibility of their use in an HL-LHC calorimeter. The challenges for high-luminosity LHC operation are mainly in the area of engineering (mechanical and thermal), data transmission and Level-1 trigger formation. A dense high-granularity calorimeter offers the opportunity of high performance in the presence of high pileup. It is possible to apply the methods of particle flow within the calorimeter to optimize the jet energy resolution. Similar calorimeters are proposed for future ILC/CLIC detectors, for which they have been shown to provide very high resolving power for single particles in dense jet environments, with energies of several hundred GeV [97]. For high-luminosity running in the LHC, the proposed calorimeter offers the prospect of resolving single particles and jets in the dense pileup environment of the forward (endcap) regions. The CALICE collaboration has been studying high-granularity calorimetry for several years and have made much progress in understanding this type of calorimeter. There are many areas where their experience is valuable—the most relevant being the mechanical design and the already existing test beam results.

3.5.1 Design overview

The overall design of the high-granularity calorimeter, HGC, is shown in Fig. 3.23. Moving outwards from the interaction region, the volume begins near the location of the front face of the current electromagnetic endcap calorimeter with a tungsten/copper-silicon sampling electromagnetic calorimeter, the EE, with a depth of about $26\,X_0$ and $1.5\,\lambda$. This is followed by a brass-silicon hadron calorimeter, $3.5\,\lambda$ deep, the FH, which is followed by a $5\,\lambda$ brass-scintillator sampling backing calorimeter, the BH, to bring the total calorimeter depth to about $10\,\lambda$, as measured in the longitudinal direction. The depth measured along trajectories from the interaction point is between 0.5% and 10% greater, depending on the pseudorapidity.

The EE and FH will use planes of silicon as the active medium, while the BH, situated where the radiation levels are low, can be constructed with plastic scintillator, as is the current HE.

In this structure there will be only a small gap between the electromagnetic and hadron calorimeters, which improves the jet energy measurements. Additionally, by bringing the hadron calorimeter into the space currently used for the EE electronics, there will be space behind the BH where the radiation levels will be sufficiently low to allow space to situate not only HGC services, but also additional muon chambers.

For EE a variable longitudinal sampling is proposed, determined by the absorber thickness:

- 10 layers: $0.65\,X_0$ thickness absorber followed by a plane of silicon,
- 10 layers: $0.88\,X_0$ thickness absorber followed by a plane of silicon,
- 8 layers: $1.26\,X_0$ thickness absorber followed by a plane of silicon.

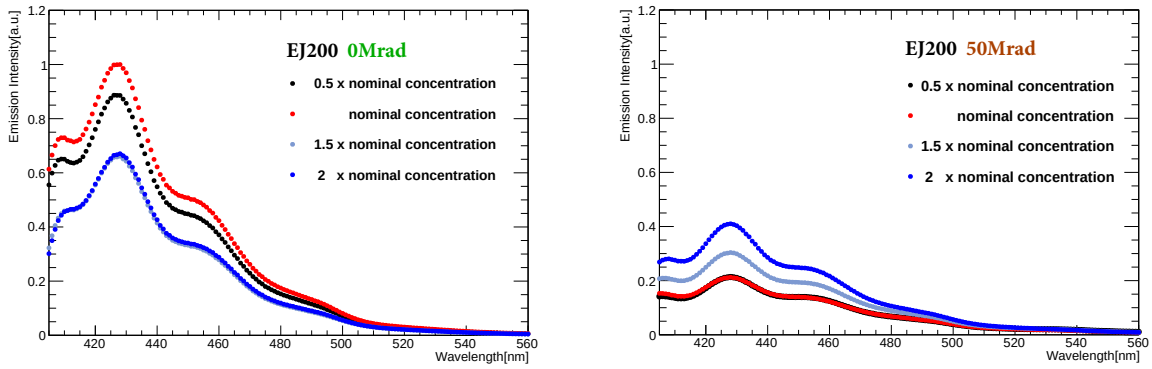


Figure 3.22: Emission spectra, as a function of the emitted wavelength, of samples of EJ-200 with different levels of primary-dopant concentration, un-irradiated samples (left) and Co-60 irradiated samples (right). The concentration varies between 0.5 and 2 times the concentration found in commercial EJ-200. The spectra are normalized with the same normalization factor.

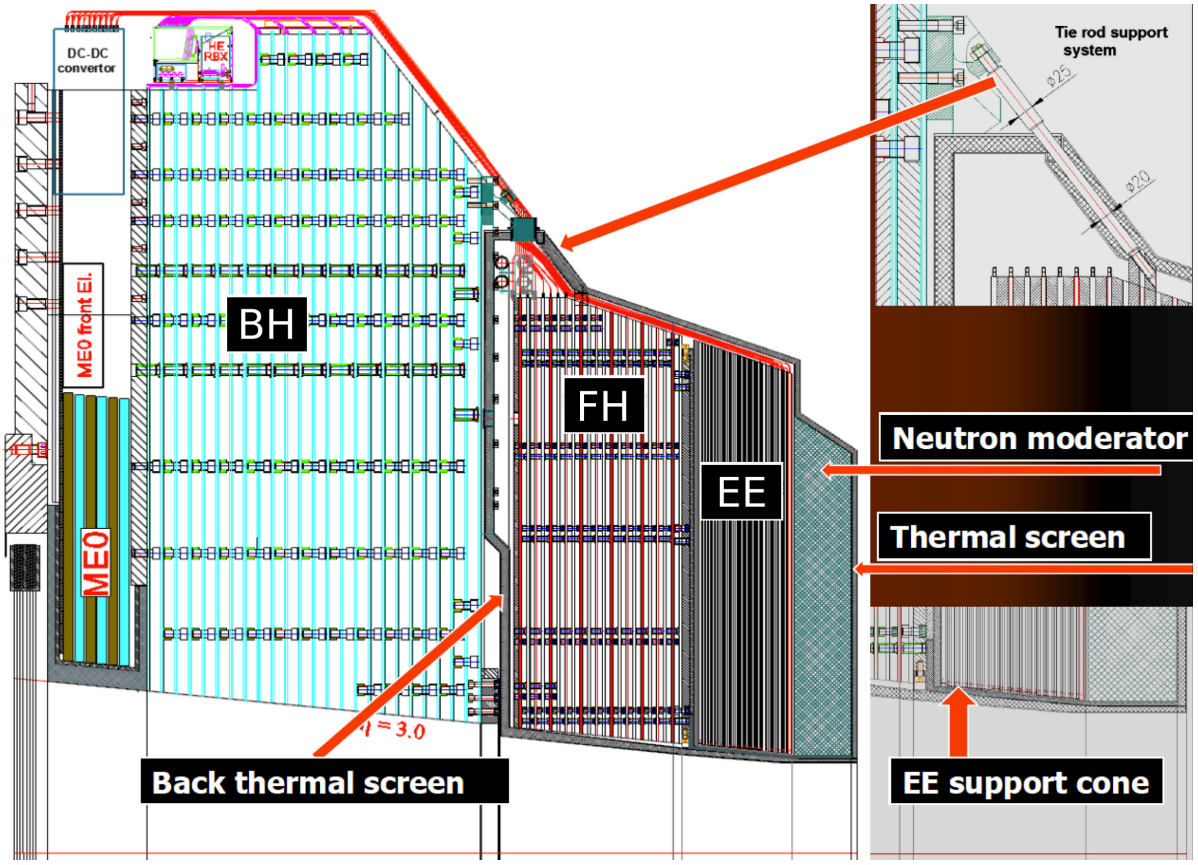


Figure 3.23: Technical drawing of the overall structure. The EE has its front face at the same location as the front face of the current EE. Directly behind it there is the FH, which is a 3.5 λ silicon-brass calorimeter. Behind that is a 5 λ backing hadron calorimeter (BH), which, since the radiation levels are low, uses a similar technology to that of the current HE.

The absorber layers are composed of tungsten in an alveolar mechanical structure, and tungsten/copper in the units which slide into the alveolar structure. The organisation, which is described in Section 3.5.2, is such that the absorber layers alternate: tungsten, tungsten/copper, tungsten etc. A similar configuration, but with tungsten only as the absorbing material, has been studied in test beams by the CALICE collaboration and, when instrumented with 525 μm-thick silicon sensors, gave an energy resolution of around $16.5\%/\sqrt{E}$ with a linear energy response [21].

The FH has a thickness of 3.5 λ with 12 samplings, and a total thickness of about 60 cm. The absorber material is brass. Each sampling layer in the FH, as in the EE, is read out independently.

The lateral granularity in EE and FH, corresponding to the cell structure on the silicon planes, is varied with the thickness of the active layer of the silicon so as to limit the cell capacitance to 60 pF. Most of the calorimeter uses cells of 1.05 cm^2. Table 3.2 shows the main parameters of the EE and FH.

The BH has a depth of 5 λ, to give a total depth for the whole calorimeter of about 10 λ. The aim is to arrive at a maximum radiation dose at the front face of the BH and at the highest $|\eta|$ of $\lesssim 5$ Mrads for an integrated luminosity of 3000 fb^{-1}, allowing the use of plastic scintillator. The proposal is for the BH to have 24 staggered layers in the same geometry as the current HE, resulting in 12 longitudinal samplings which are grouped into two longitudinal readout segments.

Table 3.2: Parameters of the EE and FH.

	EE	FH	Total
Area of silicon (m^2)	380	209	589
Channels	4.3M	1.8M	6.1M
Detector modules	13.9k	7.6k	21.5k
Weight (one endcap) (tonnes)	16.2	36.5	52.7
Number of Si planes	28	12	40

3.5.2 Mechanical design

The overall mechanical design of the EE is inspired by that prototyped for the ILD and described in Vol. 4 of the ILC TDR [97] but with a modified, radial, geometry with twelve 30° sectors making up each endcap. A carbon-fibre structure integrates tungsten absorber plates which alternate with empty slots into which "cassettes" corresponding to 30° sectors are inserted. The sector geometry is designed so that the radial sector boundaries can be arranged to occur in three different orientations, each rotated by 10° with respect to the others, so that the radial boundaries are in different locations in different layers. The design also eases the routing of services by better matching the 12-fold symmetry of the HGC with the 18-fold symmetry of the CMS barrel hadron calorimeter. Each cassette consists of two active planes and their associated front-end electronics on either side of absorber material. The absorber material provides both mechanical structure and support, and cooling. The structure is illustrated in Fig. 3.24.

The cassette is built up on either side of a 6 mm-thick copper plate containing cooling channels. Pairs of hexagonal silicon sensor wafers are mounted on a printed circuit board (Fig. 3.25 (left)). The other faces of the sensors are glued to a tungsten-copper composite (75% W, 25% Cu) baseplate of appropriate thickness to make up, when mounted either side of the copper cooling plate, the required total absorber thickness in terms of radiation lengths. The W/Cu composite has a coefficient of thermal expansion that is well-matched to silicon. The baseplates

are mounted on either side of the central copper cooling-plate and in intimate thermal contact with it, and in a way that allows for movement due to the mismatching of coefficients of thermal expansion. A thin Kapton layer between the baseplate and the silicon wafer ensures high voltage insulation of the sensor back-plane (which may be biased up to 900 V).

The thermal deformation of this configuration has been tested by cooling a test setup from 23° to –41° C. The modules are screwed and pinned in place as they would be for installation of the cassette. The distortion is checked by measuring the capacitance between the baseplate and the copper cooling-plate. The screws and pins are then released to examine the distortion of the cold module. A maximum deflection of 0.8 mm is measured using precision shims, and this distortion is such as to press the module, when it is held by the screws and pins, against the copper cooling-plate. The resulting stress in the Si sensor is $\approx$ 6 MPa, giving a safety factor of $>$ 1000 since the ultimate tensile strength of silicon is 7 GPa.

The full thickness of the cassette which will be inserted in the slot is defined by two stainless steel side bars, and a thin shielding cover plate, and leaves an air-gap of 2 mm in which the power and signals are routed. The structure is illustrated in Fig. 3.25 (right). The sensor pads are connected to the PCB with wire bonds which are made through holes in the PCB and are protected by a thin cover as shown in Fig. 3.26.

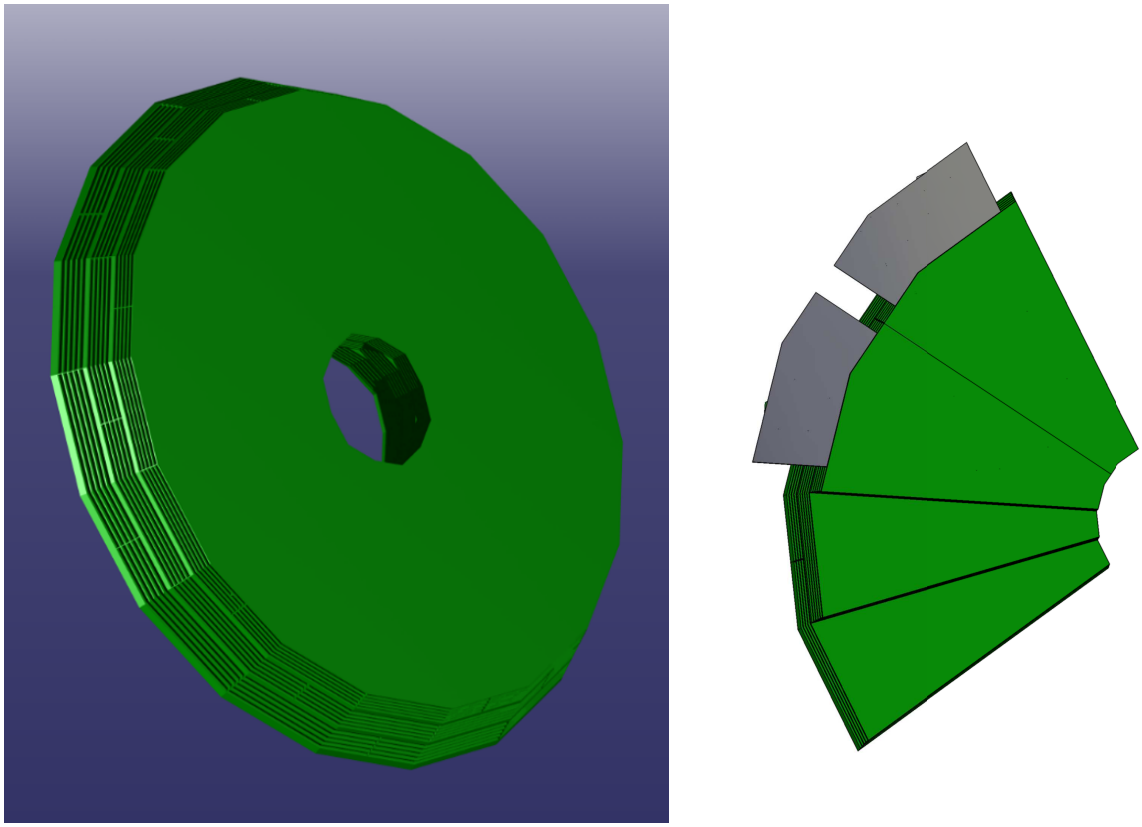


Figure 3.24: (Left) EE carbon-fibre structure integrating tungsten absorber plates alternating with empty slots. (Right) Insertion of a cassette into a slot the structure.

A first set of FEA computations has been made to study the design of the EE structure, investigating the number of carbon-fibre layers required for good stability, the method of attachment onto FH and similar related questions. The computations have demonstrated, among other results, that the maximum rigidity is achieved when the wedges are assembled into a monolithic

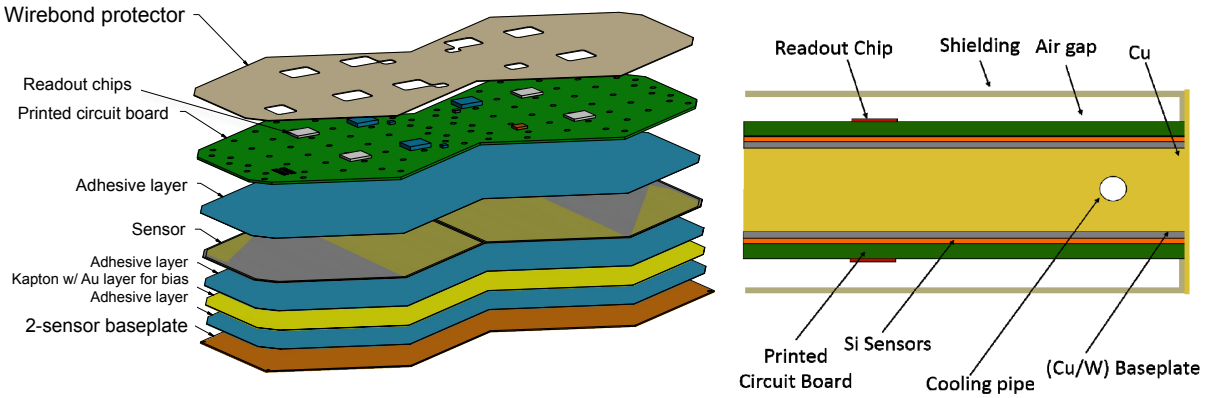


Figure 3.25: (Left) Module, consisting of printed circuit board, silicon sensors, and baseplate. (Right) Sketch of modules mounted either side of a copper and tungsten absorber/cooling plate, showing the longitudinal arrangement of a double layer.

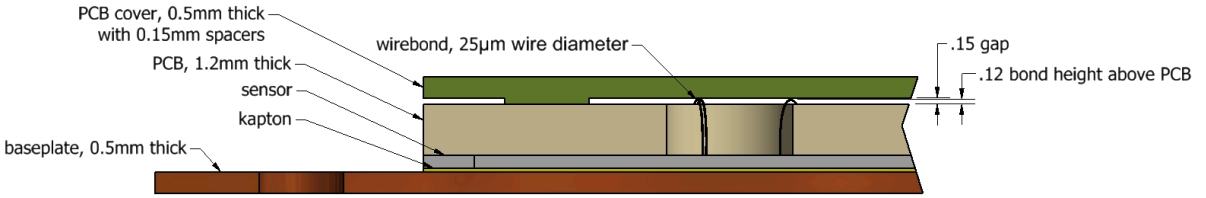


Figure 3.26: Protection of the wire bonds which connect the sensor pads to the PCB through holes in the PCB.

structure and attached to a thin aluminium inner support cone. The inner support cone is an important part of the design as it allows a better distribution of the load thus permitting the reduction of the carbon-fibre wall thickness of the wedges. The use of an aluminium cone thus allows a reduction of the number of carbon-fibre plies with respect to a solution without this cone.

The overall mechanical structure of FH is a bolted structure following the mechanical structure of the current HE, but with cassettes, similar to those for EE but carrying only a single sensor plane, slid into the slots between the brass plates. In these cassettes, the baseplate, between the sensor wafer and the Cu cooling plate, is not required to function as an absorber and a low-cost alternative to the W/Cu composite, that meets the thermal and mechanical requirements, will be chosen. The structure of FH is shown in Fig. 3.27. Each absorber plate has a 30° wide and 11 mm deep milled pocket to house a 10 mm thick cassette and the plates are assembled together using bolts. To allow for handling of the structure during assembly, rotation and installation, a 30 mm thick brass disk is installed in front of FH to allow for mounting brackets.

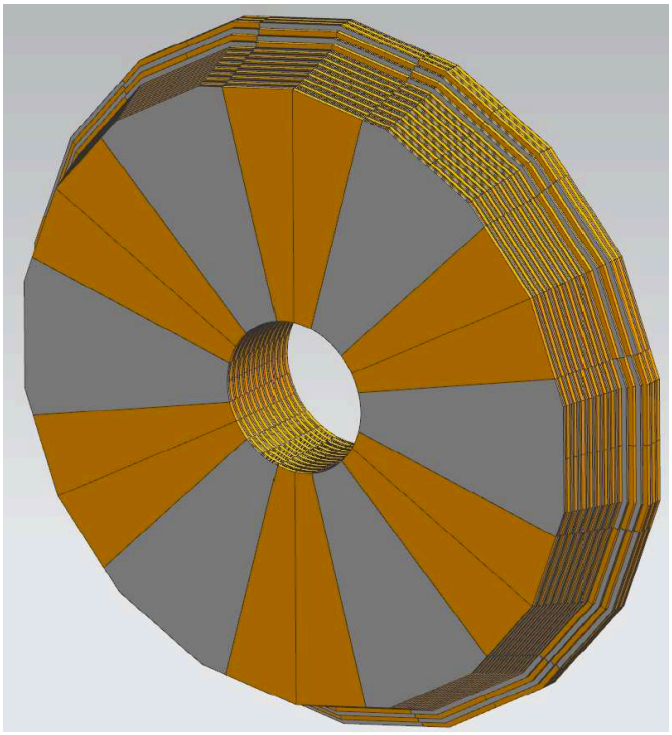


Figure 3.27: Bolted structure of FH having two configurations of brass absorber and cassette for each layer, similar to the mechanical structure of the current HE. Cassettes are inserted into 30° slots between the brass plates.

Evaporative CO_2 cooling will be used. Such a system is described in the Phase-I pixels TDR, and is currently being designed for the Phase-II CMS tracker. CO_2 evaporates from its liquid phase between –56° C and +31° C and a practical range of application is from –45° C to +25° C. The cooling fluid would pass through stainless steel pipes embedded in the copper plates. The pressure reduction capillary, required before the active cooling part of the circuit begins, is also embedded in the copper plates. In CO_2-based cooling, evaporation takes place at much higher pressures than other two-phase refrigerants. In general, the volume of vapour created stays low while it remains compressed, which means that it flows more easily through small channels, and it is possible to use smaller cooling pipes. The evaporation temperature of high-

pressure CO_2 in small cooling lines is also more stable—the pressure drop has a limited effect on the boiling temperature. Additional benefits of using CO_2 are low cost and environmental friendliness.

The total power dissipated by the front-end ASIC is estimated to be 10.5 mW/channel (details are given in Section 3.5.4), resulting in a total of 65 kW for both endcaps. To this must be added the power needed to drive the data links for the trigger, control and readout. The total is scaled up by 10% to account for the power dissipated in voltage regulation, and by a further 5% to account for dissipation in cables inside the cold volume, giving a total of 100 kW. A further 25 kW must be added to account for the power dissipated by the leakage accross the sensors after 3000 fb^{-1}, giving a total of 125 kW of power dissipated by the system inside the cold volume. Each endcap is serviced by eight CO_2 transfer lines, each capable of removing 15kW of power at $-30°C$, giving a security margin of about a factor of two. The cooling plants will be dimensioned in order to maintain sufficient operational margin throughout the lifetime of the detector.

3.5.3 Module design

The silicon sensors for the HGC will be simple, large area, single-sided, and its readout cells will be DC-coupled. The sensors will have an active thickness of 300, 200 or 100 μm, determined by the neutron fluence expected where they are located. As mentioned previously, the cell sizes are adjusted so as to limit the cell capacitance to about 60 pF. Table 3.3 shows the regions of the EE and FH where the different active thicknesses are used, together with the maximum neutron fluence expected after 3000 fb^{-1}, the cell sizes, and the expected signal-to-noise ratio, S/N, for a mimimum-ionizing particle signal, MIP, before and after 3000 fb^{-1}. The regions are specified by the radius, R, measured from the beam axis. In the calculation of S/N the charge collection efficiencies used to estimate the signal after 3000 fb^{-1} are taken from the measurements (Section 3.5.12) of structures whose sensitive thicknesses are generally larger than nominal. The leakage currents and capacitances used to estimate the noise are for sensors of nominal sensitive thickness.

Table 3.3: Silicon sensor arrangement: thickness of active silicon layer in the EE and FH, with the associated cell size and S/N for a MIP before and after an integrated luminosity of 3000 fb^{-1}.

Thickness	300 μm	200 μm	100 μm
Maximum dose (Mrad)	3	20	100
Maximum n fluence (cm^{-2})	6×10^{14}	2.5×10^{15}	1×10^{16}
EE region	$R > 120$ cm	$120 > R > 75$ cm	$R < 75$ cm
FH region	$R > 100$ cm	$100 > R > 60$ cm	$R < 60$ cm
Si wafer area (m^2)	290	203	96
Cell size (cm^2)	1.05	1.05	0.53
Cell capacitance (pF)	40	60	60
Initial S/N for MIP	13.7	7.0	3.5
S/N after 3000 fb^{-1}	6.5	2.7	1.7

The physical thickness of the silicon wafers will be 320 μm to allow production on high-volume commercial lines. The preferred sensor type is n-in-p, the type chosen by the tracker and anticipated to be more radiation tolerant. The sensors and the cells will be hexagonally shaped, so as to make best use of the wafer surface (a factor 1.3 gain with respect to a square sensor),

while providing a geometry that can be tiled. Sensors will be fabricated on 6" production lines, and a full size hexagonal sensor will cover an area about 137 cm^2.

Two sensors will be assembled on a baseplate, and connected to a printed circuit board to form a module. The 300 and 200 μm active-thickness sensors form 256-channel modules, while the 100 μm sensors form 512-channel modules.

The front-end readout chips are 70-channels wide, and include for each channel an amplifier, a 40 MHz low-power ADC and a TDC for signal digitization, as well as logic for digital data handling. The 70 channels comprise 63 standard cells, plus a further cell subdivided into seven small calibration pads (see Section 3.5.11). There will be either 4 or 8 front-end chips on a module, according to the number of readout channels. A large area multi-layer PCB covering most of the sensor surface will route signals from the sensor to the front-end chips. The connections to the cells on the sensor will be made with wire bonds, through suitable openings in the PCB, while the front-end chip will probably be flip-chip bonded to the PCB. It is envisaged that the trigger data will be generated from sums of adjacent channels, and that both zero suppression as well as a simple data compression scheme will be applied to reduce the data rates of both the Level-1 trigger and full resolution data streams.

Each module will produce up to 6 Gb/s of Level-1 trigger data, and up to 3.2 Gb/s of full resolution data, assuming a 1 MHz Level-1 accept rate. The radiation levels at the inner radius of the EE is likely to preclude the use of on-module optical links for the 200 and 100 μm modules. As a result, the present design uses three separate 5 Gb/s electrical links to transfer the data from the 200 and 100 μm modules to the back of the BH, about 4 m away, where optical links will be deployed. A fourth link provides the clock and control function. The electrical links will use Twinax cables, arranged as a ribbon in a thin, flexible support strip. These cables have a profile of less than 1 mm, are foldable and can be edge soldered to a PCB, allowing them to fit within the available space inside the calorimeter. As described in Section 3.5.7 10 Gb/s optical links are used directly by the 300 μm modules.

The radiation levels, and stringent space constraints, will make it difficult to house a DC-DC converter directly on the module. The present design therefore foresees the deployment of DC-DC converters at the back of the BH, and thus a substantial copper cross-section is required to limit the power loss over the 4 m cable length to the modules. The power cables will be low-profile Kapton PCBs, with wide copper traces providing the required cross-section. The sensor bias will be provided with a small-diameter high-voltage cable (rated for 1 kV).

Cooling circuits are embedded in the copper plates sandwiched in the centre of each cassette. Heat generated by the module electronics flows through the sensor to this cooled plate. The power estimate for the front-end 10.5 mW/chan (see Section 3.5.4). For the links the power estimate is based on the CERN LpGBT. This version is proposed to have only twenty 160 Mb/s E-link I/O ports and requires 50% of the power of the 'generic' version of the GBT with 500 mW for a bi-directional data/CTRL link running at 5 Gb/s, and 300 mW for a unidirectional data-only link.

There are approximately 22 000 modules to be produced. In comparison with the existing (and planned) tracker modules, the modules described here are simpler, with fewer wire bonds and component parts, and do not have the tight specifications for mechanical precision that are essential for a precision tracker. The design of the detector module will be optimized for ease of assembly and with a possible industrialization in mind.

3.5.4 Front-end electronics

Two options are under consideration for the preamplifier and signal digitization. The baseline design uses a preamplifier and shaper DC-coupled to the sensors, and a time-over-threshold (ToT) measurement with a TDC for digitization of large signals. The ToT front-end has fast shaping with a peaking time of 15 ns after the first shaping stage and 20 ns after the second stage. The gain is around 25 mV/fC and is linear up to 100 fC, which approximately corresponds to the largest signal in cells within a 5 GeV electromagnetic shower in the region using wafers with 300 μm active thickness. This is approximately equal to a transverse energy, E_{T}, of 1.5 GeV. Because of the way the cell sizes and thicknesses of the active layers of the silicon are varied, this approximate correspondence between signal charge and E_{T} holds across the full η range of the detector. A 10-bit ADC is used for measurement of small pulses, and saturates for pulses of $\gtrsim$ 100 fC. For signals $\gtrsim$ 80 fC the ToT comparator, using the differential signal after the second shaper, starts and stops a TDC, thus providing a digitization measuring the magnitude of large signals. Table 3.4 shows the noise performance for different combinations of detector cell capacitance and leakage current, corresponding to different cell locations, and before and after exposure to hadron fluences corresponding to an integrated luminosity of 3000 fb^{-1}. A schematic circuit diagram of the system is shown in Fig. 3.28. The ASICs will be fabricated using the 130 nm TSMC CMOS process. The total power requirements of the ToT system, including both the digital and analogue parts is estimated to be 8.5 mW/channel if a TDC with 100 ps step size is used. The power requirements are itemized in Table 3.5. A further 2 mW/channel needs to be added to this to account for additional digital elements to deal with the readout, such as data compression and a pipeline, bringing the total front-end power budget to 10.5 mW/channel.

Table 3.4: Simulated noise, N, of the ToT front-end serving a cell with capacitance C and leakage current I_{d}. The leakage currents correspond to pre-irradiation, and to the maximum, after an integrated luminosity of 3000 fb^{-1}, for each cell type. The cell types are designated by their active thickness, T.

T (μm)	C (pF)	I_{d} (μA)	N (e$^-$)
300	40	≈ 0	1600
	40	3.5	1750
200	60	≈ 0	2100
	60	5.2	2250
100	60	≈ 0	2100
	60	10.5	2400

Table 3.5: Power requirements for elements of a single channel of the ToT front-end.

	Power (mW)
Preamp.	2.0
Shaper	1.5
ADC	1.0
TDC	4.0

The backup design uses a low gain preamplifier in linear mode connected to two fast shapers. The preamplifier and shapers are DC-coupled to detector. The gains are 100 and 800 μV/fC respectively, with noise around 1.8 fC. The high noise due to the low preamplifier gain does not permit MIP sensitivity. Whereas with the ToT system the channels connected to special small subdivided pads on each wafer merely provide added security and redundancy for calibration with MIP signals, with the backup front-end design they provide the only way to achieve MIP calibration. In addition, the possibility of precision timing (see Section 3.5.5) is absent.

3.5.5 Precision timing

The jitter of the start time for a time-over-threshold measurement is found in simulation of the ToT front-end circuit to have an RMS value of 50 ps. Thus the ToT architecture potentially provides a high precision time measurement for each and every cell with an energy deposit

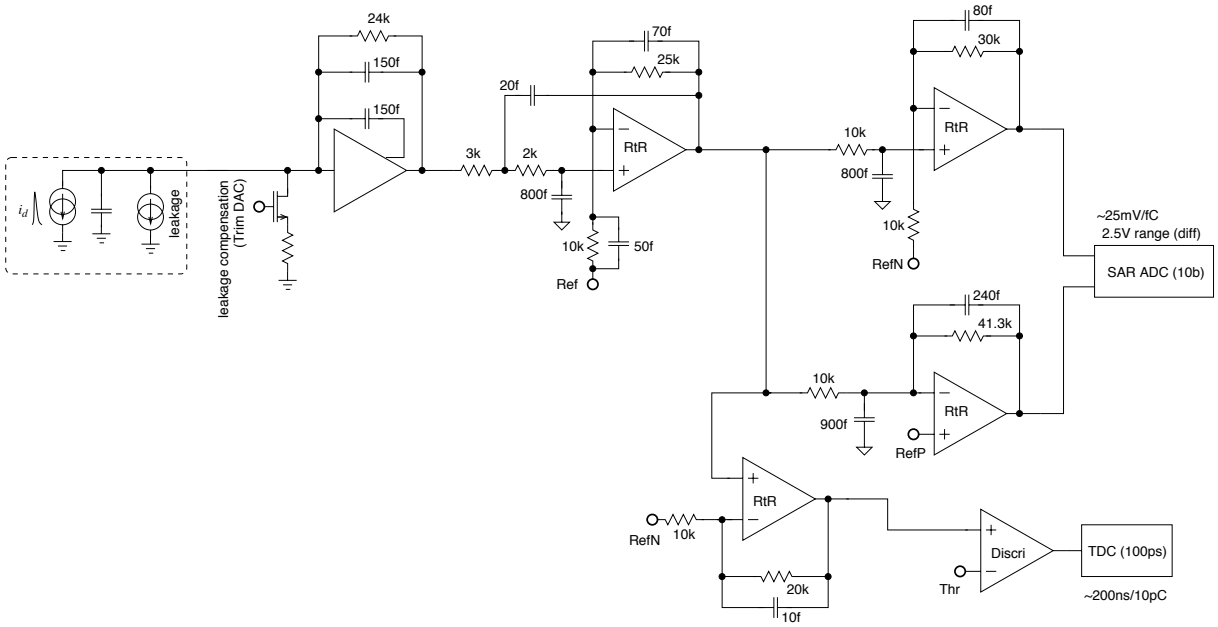


Figure 3.28: Schematic circuit diagram of ToT system.

larger than $\approx$ 80 fC, which approximately corresponds to the largest signal in the cells in an electromagnetic shower of $E_{\mathrm{T}} \sim 1.5$ GeV. More energetic showers would benefit from multiple measurements, potentially reducing the uncertainty. For example, showers typical of photons from $\mathrm{H} \to \gamma\gamma$ events have about 80 cells above the ToT threshold. The intrinsic time jitter of a silicon detector with signals of this size remains to be determined, but it is likely to be small compared to the single cell resolution. The slew of the time as a function of signal amplitude would need to be corrected for, and a constraint on overall timing precison could arise from the precision and stability achievable for this correction. The bin size and jitter of the TDC used in the ToT front-end provides a further constraint. The TDC has been measured to have an effective resolution of 50 ps. The system aspects of realizing the required precision for clock distribution seem challenging, but existing experiments have successfully achieved comparable timing precision in large-scale distributed systems. For example, the ALICE time-of-flight system uses glass-RPCs with 160 000 channels and a detector area of 150 m^2. It has achieved an overall timing resolution of 80 ps stable over many years of operation [98], and measurement of multiple tracks appear to improve an overall measurement of event time in a way that implies that the jitter attributable to clock distribution is 25–50 ps.

Precision timing of showers is a very attractive possibility, with potential to assist in the removal of pileup and the location of interaction vertices, and its realization is being actively studied.

3.5.6 Power distribution

For each endcap 50 kW of low-voltage power must be supplied. At the 1.2 V level required by the electronics, this amounts to over 40 kA of current for each endcap. DC-DC converters are used to reduce the required cross-section of the cables between the power supplies and the detector. In the present design, based on existing technology, these are located outside the cold volume, in the low radiation region at the back of the endcap calorimeter structure on its outer circumference. This has the drawback of requiring a cable length of 3 or 4 m between the DC-DC converters and the modules. The cross-section of these cables has been dimensioned to limit the voltage drop to between 10% and 15% (total), and voltage regulators on the detector

modules are foreseen (and accounted for in the power estimates). Studies indicate that the DC-DC converter modules can be made sufficiently compact to be accommodated directly on the detector modules, with only a marginal increase in the gap between absorbers. Should sufficiently radiation hard ASIC technologies for use in the DC-DC converters become available we will investigate this option further.

In the present design, we assume that the DC-DC converters are located behind the BH, together with the optical links and other HGC services. In such a configuration, the cable runs from the power supply racks on the balconies to the DC-DC converters have a typical length of 10 m, and the longest path from the DC-DC converters to the HGC modules is about 4 m. Based on the ongoing development of DC-DC converters for the CMS pixels and outer tracker, we assume a conversion ratio of 8/1 with a 65% efficiency.

Under these assumptions, and allowing for a power loss over the cables consistent with an overall 50% power delivery efficiency (i.e. including loss in DC-DC converters), the copper cross-section from the power supplies to the DC-DC converters and through the services choke points across the ME1/1 chambers can be reduced by about an order of magnitude, to around 30 cm^2 for each endcap. On the other hand, in this configuration a copper cross-section of about 200 cm^2 is required for the cables from the DC-DC converters to the front-end modules (about 1–2 mm^2 per module).

A serial-powering scheme would allow a substantial reduction in the copper cross-section of the cables from the DC-DC converters to the front-end modules, and is a possible option for further study.

3.5.7 Data transfer

The current design envisages a hybrid solution for the transfer of data off the detector. For the 300 μm modules (i.e. EE $R > 120$ cm, FH $R > 100$ cm), 10 Gb/s optical data links will be used directly on the detector and modules will be grouped together to share the high-bandwidth links. Each set of four modules is serviced by a bidirectional 10 Gb/s optical link for full readout data as well as for clock and controls. For the layers which provide trigger data, two 10 Gb/s optical links are used to carry the trigger data for the four modules.

For the 200 and 100 μm modules, optical drivers may not be sufficiently radiation-hard for operation within the detector. In this region, 5 Gb/s Twinax cables will be used for the data links. Each 200 μm module will have one bidirectional link for readout and control, and two links for trigger data on those layers which provide input to the trigger. As the channel count per sensor is higher on the 100 μm modules, two readout/control links and four trigger links will be used for these modules. The Twinax cables transport data from the front-end ASIC to the back of the calorimeter, where a conversion to optical links is made. Table 3.6 summarizes the number of different types of module, corresponding to different active thicknesses of the sensors, and whether the module sends trigger data for each bunch crossing as well as full readout data on receipt of a Level-1 accept ("Trigger"), or whether it only sends data on receipt of a Level-1 accept ("Data only"). Twinax cables manufactured by TempFlex have been tested by members of the ATLAS pixel detector group who have shown that they can transfer data over 5 m at rates up to 8 Gbits/sec. Micro-twisted-pair cables could provide an alternative solution. The change from electrical to optical signals would take place behind the calorimeter using a radiation-hard FPGA, like the Microsemi Igloo2 that will be used in the Phase-I upgrade of the HCAL. The FPGA would be used only to deserialize the data, compress it, and retransmit it with 10 Gbps optical links to the service cavern.

Table 3.6: Number of different types of module, for each active thickness of sensor. "Trigger" modules send trigger data for each bunch crossing as well as full readout data on receipt of a Level-1 accept. "Data only" modules only send data on receipt of a Level-1 accept.

Thickness	cell size (cm^2)	Trigger	Data only	Total
300 μm	1.05	8 544	2 256	10 800
200 μm	1.05	4 584	2 832	7 416
100 μm	0.53	2 040	1 464	3 504

3.5.8 Services

At the periphery of each cassette there is a space of about 80 cm $\times$ 2 mm. The services that need to taken in or out of the cassettes comprise:

Low Voltage: Heavy duty copper tracks on a 0.5 mm thick PCB are used to take the low voltage power from the periphery to individual modules within the cassette, as shown in Fig. 3.29. A section of about 50 mm^2 of copper is needed.

High Voltage: Up to 10 mA/module will have to be supplied at a voltage of 900 V (10 mA is the current for the modules subjected to the largest flux of neutrons, and after 3000 fb^{-1}). Cables having a 1 mm diameter are used in current CMS tracker modules to supply the HV. We envisage arranging these cables at five locations, each set with a cross section of 1 cm $\times$1 mm.

Data Links: Three 10 Gb/s optical readout links for each four 300 μm modules; three 5 Gb/s Twinax cables for each 200 and 100 μm module.

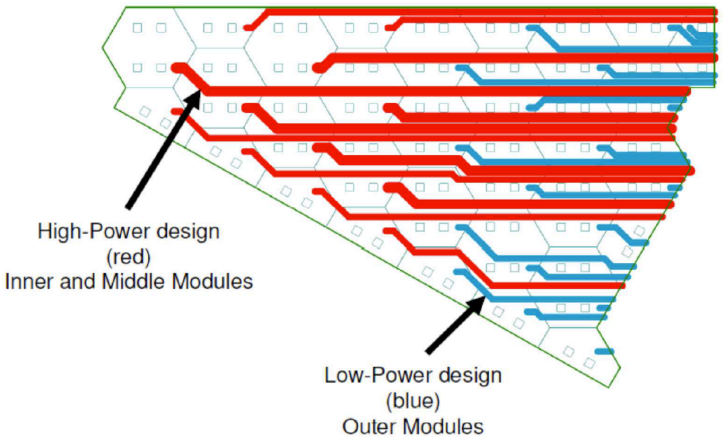


Figure 3.29: Printed circuit board with heavy duty copper tracks used to take the low voltage power to individual modules.

3.5.9 Backing Hadron Calorimeter

As discussed above, the backing hadron calorimeter would be constructed of 5 λ of brass absorber formed in 9 cm-thick brass plates interleaved with scintillator active material. The light signal will be collected from the planes using wavelength-shifting fibers. Twelve planes of scintillator will be used, with each plane split into 20° sections and offset alternately in the z direction to allow strong mechanical connections between the brass layers. This mechanical structure is very similar to the existing HE, but reuse of the current HE material is not proposed due to the difficulty of working with the irradiated absorber and the tight constraints of installation work during LS3.

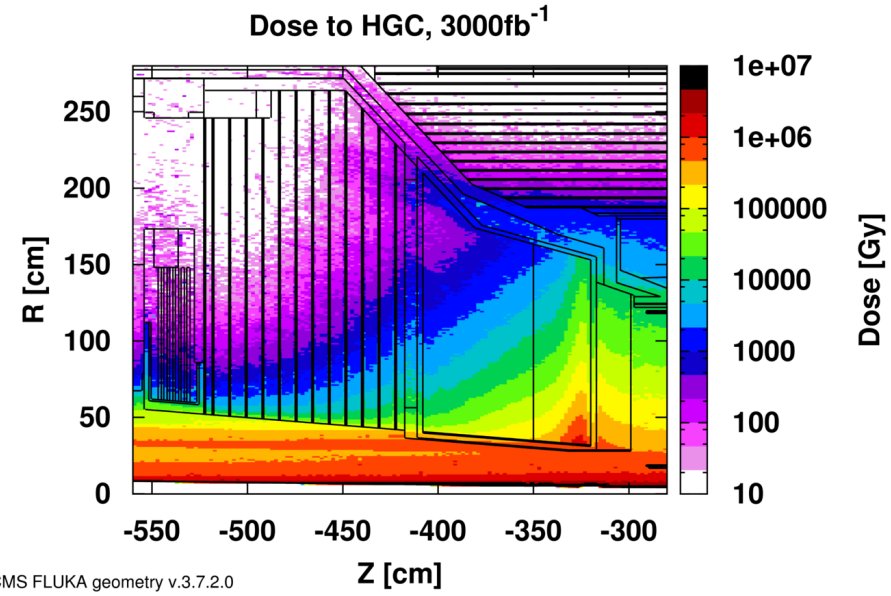


Figure 3.30: Predicted radiation dose in the endcap calorimeters for HL-LHC operation.

The scintillator planes are expected to be constructed using doubly-doped scintillator, as described in Section 3.4.6. The detailed geometry of the individual tiles will depend on the expected local radiation dose. The map of the doses expected in the BH region is presented in Fig. 3.30. The radiation doses range from approximately 5 Mrad to below 1 krad. In the low-dose regions, the traditional sigma-shaped arrangement of the WLS will be used, while in the medium-dose regions, the towers will be structured as set of narrow tiles, each of which is read out by a single WLS fiber running along the length of the finger tile. Figure 3.31 shows a comparison between a sigma tile and a finger tile. This simple concept increases the radiation tolerance of the detector because it significantly shortens the average light path between the particle-scintillator interaction point and the closest WLS fiber.

The front-end electronics for the BH will be heavily based on the HCAL Phase-I upgrade electronics[5]. However, the reconstruction of the endcap and the integration of the HGC services will require changes to the mechanics and structure of the electronics, requiring a reconstruction of the system. The data link of the electronics will be upgraded from the Phase-I bandwidth of 5 Gbps to the standard Phase-II 10 Gbps, allowing for a higher data concentration in the readout and trigger electronics.

3.5.10 Trigger and Off-Detector Electronics

Trigger data will be generated from sums of adjacent channels, using every alternate active plane. The sums will be made with a granularity of 2×2 sensor pads for the both the EE and FH, and sent at the full rate of 40 MHz by the front-end electronics to the services cavern where trigger primitives will be generated. The total number of trigger sums will be about 600k for the EE and 250k for the FH. In the BH, each digitized sample will be transferred to the off-detector electronics, with no pipeline in the front-end.

Trigger primitives, to be sent to the Level-1 calorimeter trigger, will be constructed in the off-detector electronics from the single-plane sums and the BH samples. The calorimeter electronics will be responsible for forming local longitudinal clusters and projective “towers” for use in the calorimeter trigger.

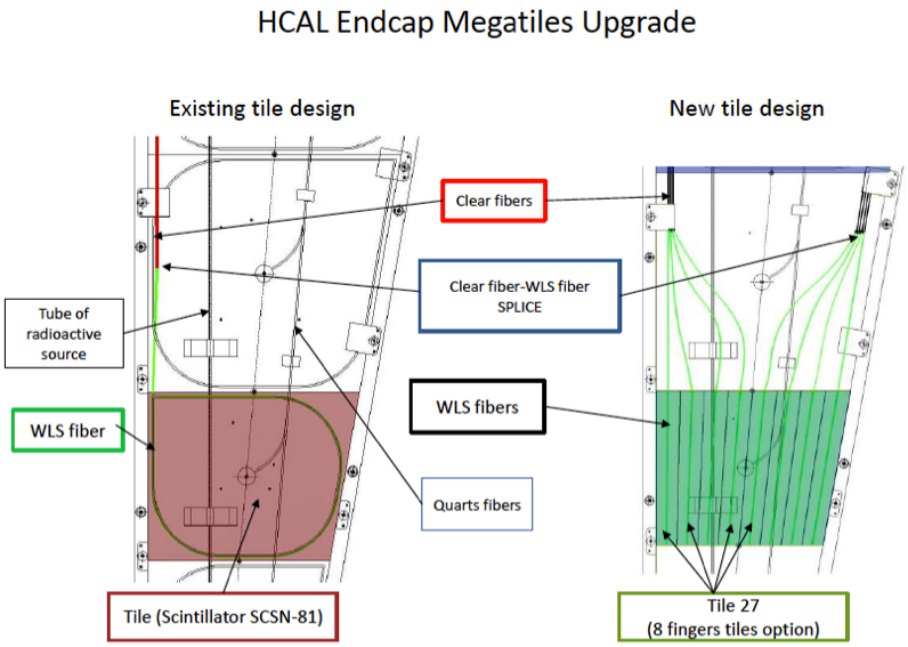


Figure 3.31: Comparison between a plastic scintillator sigma tile (left), currently used in the HE detector, and a finger tile, the proposed design for the new BH, in the low-dose region (right).

Electromagnetic clusters will be formed beginning with energetic seeds and applying a-priori information about the lateral and longitudinal shape of electromagnetic showers. Layer-by-layer pileup and non-clustered-energy corrections can be made in the calorimeter electronics before transmission of the clusters to the trigger electronics. The trigger electronics will be responsible for merging clusters and carrying out isolation and E/H requirements as well integrating tracker information where such information is available.

Preliminary studies of trigger algorithms suggest that, for the electromagnetic trigger, the use of the information from longitudinal granularity and fine lateral granularity will provide effective rate reduction. Comparing rates between the Phase 1 trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGC with a mean of 140 pileup interactions per crossing, the Level-1 single electromagnetic trigger rate is only 1.5 times higher for thresholds up to 20 GeV, and 2.5 times higher for thresholds up to 30 GeV, despite the increase in the instantaneous luminosity by a factor of 3.5. This is shown in Fig. 3.32. The algorithm results in only a 1–2% inefficiency for electrons.

The FH and BH data will be processed in calorimeter electronics of the same design as for the electromagnetic part. Clusters from the hadronic section of the calorimeter will be sent to the calorimeter trigger electronics for use in the electromagnetic and jet algorithms. Clusters will be similar in size to the Phase-I trigger towers ($0.087 \times 0.087 \Delta\eta \times \Delta\phi$) or smaller to allow for precisely-defined jets and isolation regions.

Ongoing exploration of possible jet algorithms shows that the lateral granularity allows the dense core of jets, particularly jets from the hadronic decay of τ leptons, and quark jets from the VBF production of Higgs bosons, to be identified and distinguished from pileup. The fine lateral granularity provides effective rejection of background from pileup, where regions of high activity containing the energy from many overlapping interactions are reconstructed as jets. VBF jets and jets from the hadronic decay of tau leptons can be seeded with more than

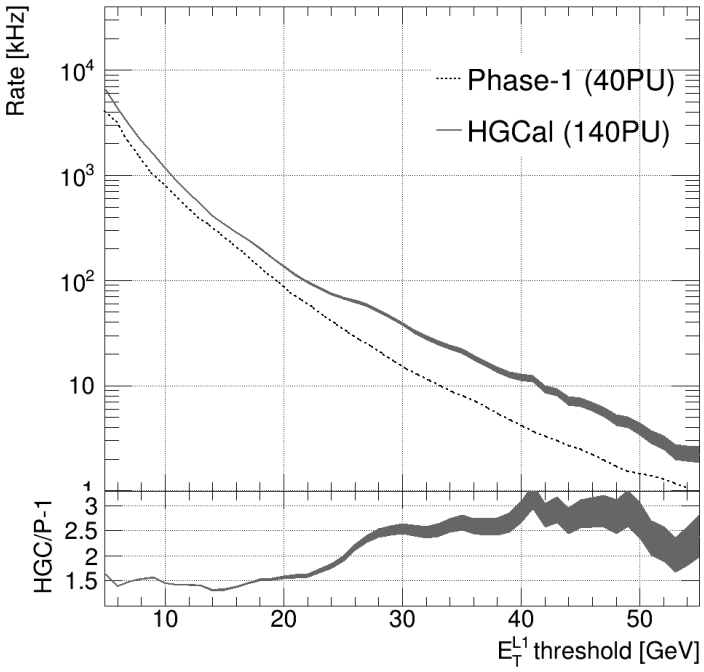


Figure 3.32: Comparison of rates between the Phase 1 electromagnetic object trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGC with a mean of 140 pileup interactions per crossing.

90% efficiency using small clusters in EE and FH. An average of 16 seeds per event, providing regions of interest, are found in events where the mean number of interactions per bunch crossing is 140. Optimization of the size of the η-ϕ region used for reconstructing jet energy has been studied.

A size of $\Delta R = 0.2$ seems to give the best balance between reduction of pileup fluctuations and the minimization of fluctuations due to inadequate jet containment. Cells are included in the energy sum based on a layer-by-layer pileup-dependent threshold, and the resulting jet energy is corrected for pileup and out-of-cone leakage. Figure 3.33 compares the Level-1 single jet trigger rate in the endcaps for the Phase 1 detector at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGC, using the algorithm described, with a mean of 140 pileup interactions per crossing. The rate in the HGC is at most a little more than twice the rate in the Phase 1 endcaps, despite the instantaneous luminosity being 3.5 times higher.

Besides the trigger electronics, several other off-detector electronics systems will be required. While trigger data and the full BH data will be available off-detector for every beam-crossing, the full EE and FH data will be transferred on L1 to DAQ boards which will interface between the calorimeter electronics and the central DAQ system. Robust control and safety systems will also be required to configure, control, and protect the detector.

3.5.11 Calibration and monitoring

We are targeting an electromagnetic energy resolution with a constant term smaller than 1%. The contribution to the constant term in the energy resolution function given by a channel-to-channel intercalibration uncertainty has been studied. When reconstructing electromagnetic showers, it is found that the contribution to the constant term is more than six times smaller

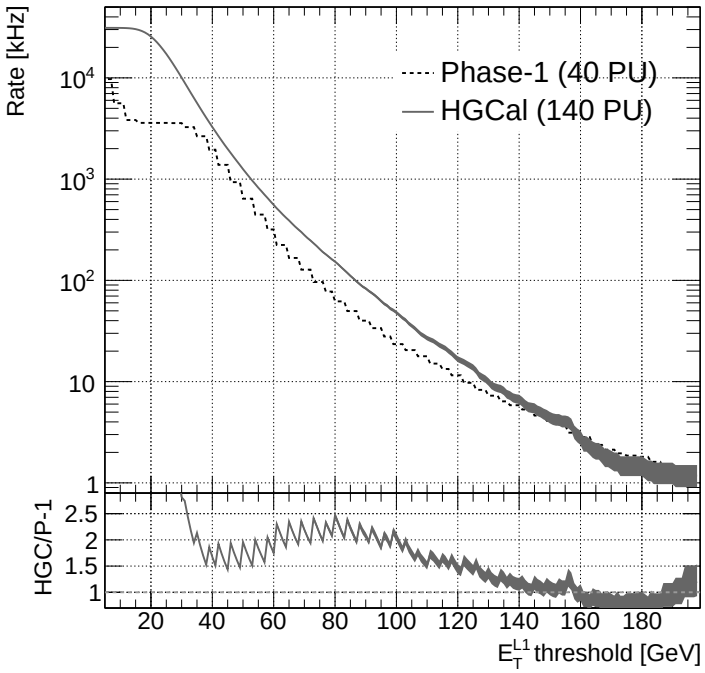


Figure 3.33: Level-1 single jet trigger rates in the endcaps, comparing the Phase 1 trigger running at an instantaneous luminosity resulting in a mean of 40 pileup interactions per bunch crossing, and the HGC with a mean of 140 pileup interactions per crossing.

than the intercalibration uncertainty. We thus target an intercalibration precision of 3% which will contribute less than 0.5% to the constant term.

The sensor intercalibration will be tracked and maintained using the MIP signals in any triggered event, to follow the slow change in charge collection efficiency over the duration of HL-LHC operation. The noise performance of the ToT front-end will enable MIP signals to seen and fitted in almost every sensor cell of the HGC even after exposure to the hadron fluence from an integrated luminosity of 3000 fb^{-1}. In addition, for redundancy, and to fully guarantee the ability to achieve MIP calibration throughout the life of the HGC, dedicated low-capacitance/low-noise cells will be included on each wafer. For these calibration cells a standard hexagonal cell will be divided into seven subcells, a central hexagonal subcell with six neighbours thus providing cells with a signal-to-noise ratio > 5 on each wafer of the HGC, even after the accumulation of the full lifetime hadron fluence.

A simple MIP-tracking algorithm has been implemented in simulated events. Requiring a signal > 0.9 MIP to be seen in the layers before and after the sensor studied enables a clear peak to be seen and fitted for noise levels up to about 0.4 MIP. Figure 3.34 shows the result in sensor cells with noise equivalent to 0.3 MIP located in the region $2.8 < |\eta| < 2.9$. Local isolation, requiring that all sensor cells surrounding the cell under study have a signal < 0.5 MIP, reduces the sensitivity of the fitted peak value to the instantaneous luminosity. About 1.5 M events are needed to reach a statistical precision of 3% for individual cells in the region $2.8 < |\eta| < 2.9$ when the noise is equivalent to 0.3 MIP. The sensors in this region have an active thickness of 100 μm and the simulation is made with a mean of 140 interactions per bunch crossing.

Tightening the MIP-tracking algorithm by requiring corresponding signals in two layers before and two after the sensor under study removes the contribution from instances where there is, in fact, no MIP present (seen on the left of the signal peak in Fig. 3.34) and enables the peak to

be seen and fitted even when the noise level is 0.6 MIP. This noise level corresponds to that in the worst case sensor pads after a fluence corresponding to 3000 fb^{-1}. For this case about 100 times more events are needed to reach a statistical precision of 3%. The largest factor ($\approx$ 15) is due to the demand for signals in the additional layers, and further factors come from the inefficiency of the isolation requirement at high noise levels, and the fitting of a wider signal. A possible way of performing the MIP calibration would be to build histograms, such as shown in Fig. 3.34 (left), in the HLT farm accessing the full Level-1 data sample. Only for the highest noise sensor cells, and only for the highest luminosities, would the low readout thresholds required to perform the calibration be intolerable. Histograms for these cells would be taken (for example) later in a fill after the luminosity had decayed somewhat, allowing the readout thresholds to be lowered.

The electronics chain of each channel will be independently monitored and linearized using a charge injection system based on a chopper circuit and fixed calibration capacitances connected to the front-end input. The additional noise contribution from inclusion of this system is negligible because of the chopper series resistance. The large range of injectable charges (0–10 pC) will allow a finely detailed linearity measurement. Before installation, the channel testing and qualification protocol will include measurement of the charge injection capacitance to $< 1\%$. Knowledge of the charge injection capacitance will allow an independent verification of the MIP calibration at startup, and, when used in conjunction with the MIP calibration, will monitor the charge collection efficiency throughout the lifetime of the detector. Figure 3.34 (right) shows a circuit diagram of the proposed scheme.

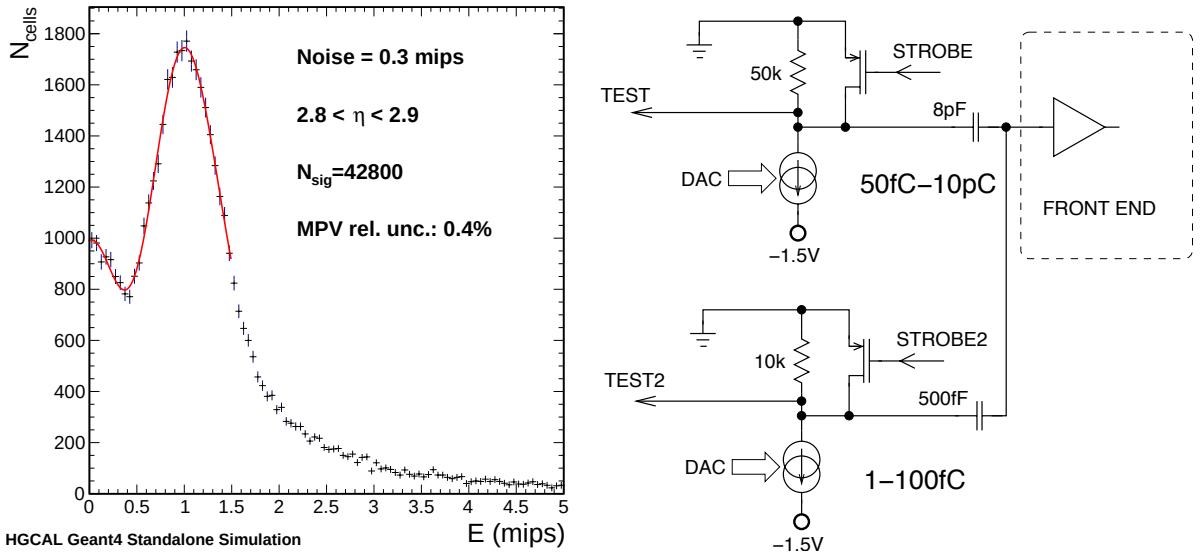


Figure 3.34: (Left) MIP signal peak in sensor cells with noise equivalent to 0.3 MIP, after the simplest MIP tracking and local isolation algorithms have been applied. The sensor cells are located in the region $2.8 < |\eta| < 2.9$ and have an active thickness of 100 μm. In the simulated events the mean number of interactions per bunch crossing is 140. (Right) Electronics calibration circuit diagram. The circuit has two sections with overlapping ranges, one for small (1-100fC) and one for large calibration signals. This avoids switching very low DC levels for the low range, and allows the use of two low resolution (8 bit) DACs instead of one high resolution DAC.

According to a potential supplier the average thickness of all wafers will be contained within the range $\pm 5\,\mu$m of nominal, and within a wafer the diffusion depth of all pads (defining the

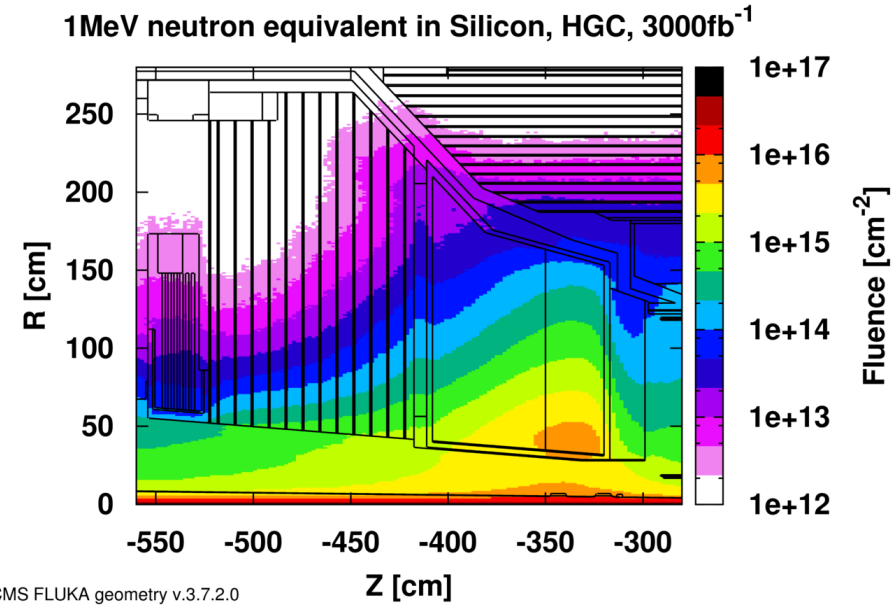


Figure 3.35: Equivalent 1 MeV neutron fluence for Phase-II endcap calorimetry at HL-LHC.

active thickness) will be within $\pm 3\,\mu$m of the average for the wafer. This translates to an effective Gaussian spread of the diffusion depth between the pads in a wafer of 0.6% (1.7%) for wafers with 300 (100) μm active thickness, respectively. As part of the test and qualification protocol for wafers the depletion depth of cells will be characterized by voltage/capacitance measurements with a precision of better than 1%. A number of "longitudinal" towers will be put into test beams to calibrate the responses to electrons and hadrons before startup.

3.5.12 Radiation tolerance

At the HL-LHC the silicon sensors of the HGC will be exposed to hadron fluences ranging from about 2×10^{14} up to about 10^{16} 1 MeV neutron equivalent per cm^2 (neq/cm^2) as shown in Fig. 3.35. These fluences are similar to those in the tracker and pixel volumes for the HL-LHC, and the basic parameters for the HGC sensor design are based on results obtained for the CMS Phase-II Tracker R&D (e.g. Fig. 3.36, showing charge collection as a function of fluence), and further dedicated measurements using neutron fluences up to 1.6×10^{16} n/cm^2.

The main difference between the tracker and the HGC is that whereas in the tracker case the fluence is dominated by charged hadrons, in the case of the HGC it is neutrons that dominate. A dedicated campaign is underway to determine if the performance of the sensors is affected differently by neutrons. This study includes both p-in-n and n-in-p sensors, with active thickness of 300, 200, 100 and 50 μm, exposed to fluences up to the highest to which the HGC will be exposed. First results from neutron irradiation are summarized in Fig. 3.37. At the start of life the collected charge is about 22, 15 and 9 ke for sensors with 300, 200 and 120 μm active thicknesses, respectively. These numbers are calculated using 73 e/μm for the MIP charge deposition in silicon. Based on the measurements shown above, the collected charge after 3000 fb^{-1} is estimated to be in the worst case 10, 6 and 4 ke for the three thicknesses. These numbers are somewhat lower than expected from purely proton irradiation and the reason is being investigated. Our current design specifies an active thickness of 100 μm for the region subjected to the highest flux, but in the light of these measurements we may wish to choose an active thickness of 120 μm. The measured leakage currents for irradiated sensors, at –20° C and at bias voltages of 600 and 800 V, are consistent with expectations, as can be seen in Fig. 3.38

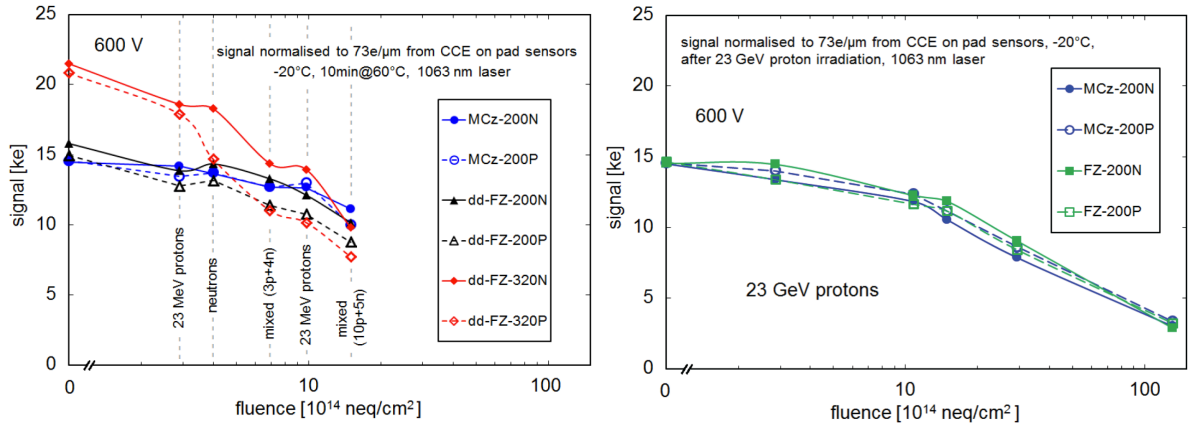


Figure 3.36: Charge collection for silicon sensors versus neutron fluence. The sensors are biased at 600 V after proton beam irradiation, and the fluence of the exposures is expressed in terms of 1 MeV neutron equivalent per cm^2. The plot on the left shows sensors with 300 and 200 μm active-thickness; the plot on the right shows both p-in-n and n-in-p 200 μm sensors and extends to a fluence of 1.4×10^{16} neq/cm^2.

where the leakage currents, as a function of neutron fluence, are compared to the expected α-value of 3.25×10^{-19} [24] for a bias voltage of 800 V. In the figure the leakage currents and the α-value have been scaled from the temperature, –20° C, at which they were measured, to –30° C, the temperature at which the HGC will be operated. A suitable low noise system is being set up with which both the noise and the response to single MIPs (Landau fluctuations) will be measured for the irradiated diodes.

The present design relies on 100 μm active thickness sensors to cover the region of $|\eta| < 3$, where the integrated fluence will be 10^{16} neq/cm^2. For an extended endcap design with coverage up to $|\eta| = 4$, the maximum integrated fluence approaches 10^{17} neq/cm^2. Options under study include the use of either planar sensors with an active thickness down to 50 μm, or of 3D sensors, and finer granularity to maintain adequate signal response.

The 130 nm TSMC CMOS technology which will be used to fabricate the front-end ASICs has been qualified up to 400 MRad and it is expected that good analogue performance can be maintained above the maximum dose of 150 MRad expected in the calorimeter. This technology is also known to be radiation tolerant up to very high fluences (a few $\times 10^{15}$ n/cm^2), although tolerance up to 10^{16} n/cm^2 will have to be demonstrated. The radiation field at the inner edge of the calorimeter falls off very sharply with increasing radius and, if need be, the innermost detector modules could be designed so as to place the ASICs at the outer periphery of these particular modules.

3.5.13 Simulation and performance

The electromagnetic energy resolution has been studied using a standalone GEANT4 v9.6.2 [79] simulation. The range cut is set to be 40 μm for the simulation for electrons, positrons and photons in all silicon volumes and 1 mm elsewhere. With this step size the energy of an electron (photon) that would deposit all its energy in a single step is 70 (1.5) keV.

The stochastic term of the energy resolution function, measured for incident electrons with energies in the range $5 \leq E \leq 500$ GeV is found to vary between 20 and 24%, as shown in Fig. 3.39 (left), somewhat worse than the value of 16.5% found in a test beam by the CALICE collaboration [21], using a prototype having a longitudinal sampling structure very similar to

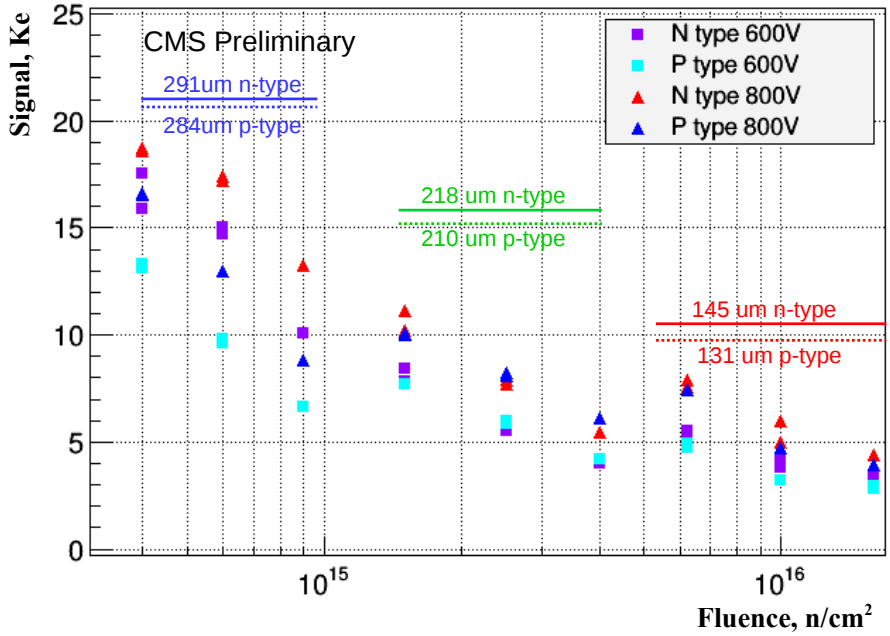


Figure 3.37: Charge collection efficiency for $\approx 300\,\mu$m (leftmost set of points), $\approx 200\,\mu$m (middle set of points), and $\approx 120\,\mu$m silicon sensors (rightmost set of points).

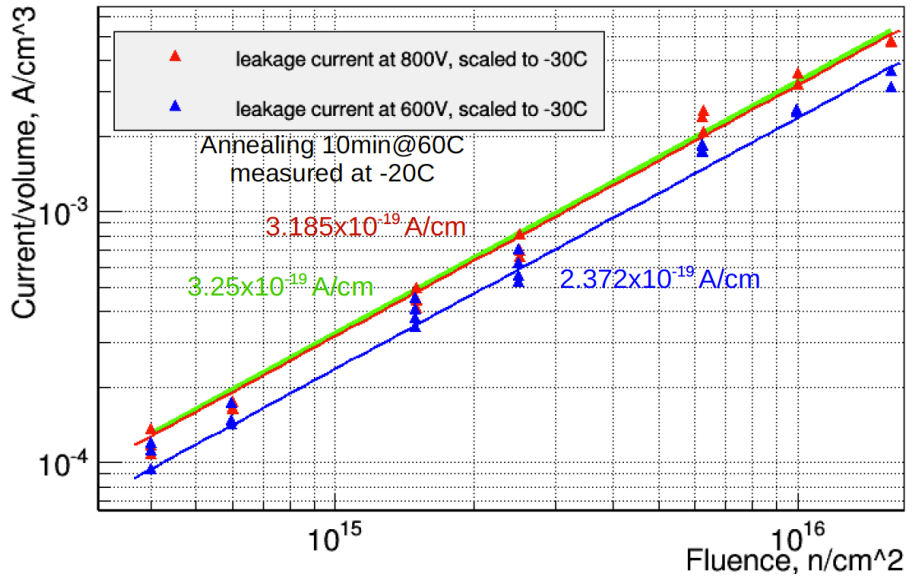


Figure 3.38: Leakage currents, as a function of neutron fluence, measured at –20° C, and scaled to –30° C, the temperature at which the HGC will be operated.

the design described here. As can be seen the stochastic term varies with the active thickness of the silicon sensors, and this largely explains the difference with the CALICE result, for which sensors with an active thickness of 500 μm were used. A further contributing factor is the fact that in the lower η regions of the HGC the angle of incidence of electrons and photons is significantly inclined to the normal (by 25° at $\eta = 1.5$), thus increasing the effective thickness of absorber layers. It has been verified that using a simplified model of the CALICE test beam device, their test beam resolution performance can be reproduced. The constant term is also small, 0.6 or 0.7%.

The Molière radius of the EE can be extracted, i.e. the radius within which 90% of the electromagnetic shower energy is expected to be contained, and is found to be 27.8 mm in simulated electron showers, with the 68% containment radius being 9.7 mm. If the air gap (shown in Fig. 3.25) is increased to 4 mm in the simulation the Molière radius increases to 35 mm but the 68% containment changes by a only a few mm. The benefit of high longitudinal granularity can be seen by looking at the containment per silicon layer. Figure 3.39 (right) shows the 68% and 90% containment radii for each individual silicon layer, showing that for the first half of the layers 90% of the energy deposited by the shower is contained inside a radius significantly less than the Molière radius, and the 68% containment radius is only a few mm.

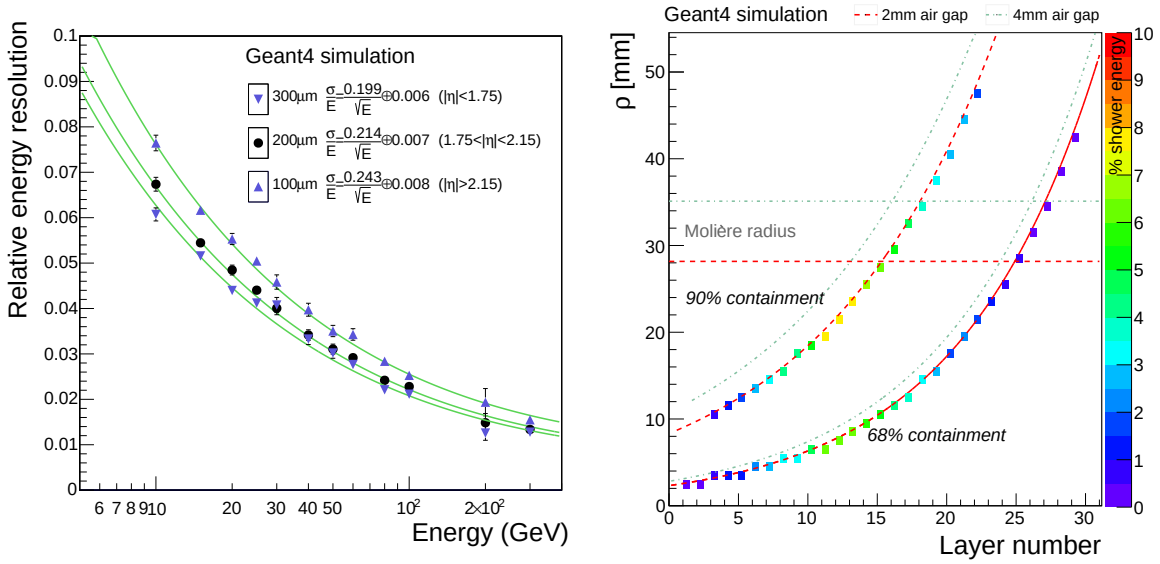


Figure 3.39: (Left) Energy resolution as a function of energy from a standalone simulation of incident electrons on an HCGAL using silicon sensors with an active thicknesses of 300, 200 and 100 μm. (Right) Radii, ρ, containing 68% and 90% of the energy deposited in an individual silicon layer by a shower, as a function of silicon layer. The colour-coded rectangles indicate the fraction of total energy deposited inside the 68% and 90% containment radii of each layer. Fitted parameterizations of the 68 and 95% radii as a function of depth are shown using the design (2 mm) gap clearance and an increased (4 mm) gap.

The occupancy has been studied in events where the number of pileup events is modelled by a Poisson distribution with a mean of 200. The result is shown in Fig. 3.40 for signals above a thresholds of 0.5 MIP, equivalent to an energy of 2.5 MeV, and for signals above a threshold of 5 MIPs. The occupancy for the higher signal threshold is lower by more than a factor of about 4. If the mean number of events per bunch crossing is 140 the occupancies are found to be lower by a corresponding factor.

The ToT front-end results in dead time for sensor cells with signals larger than than 250 fC. The

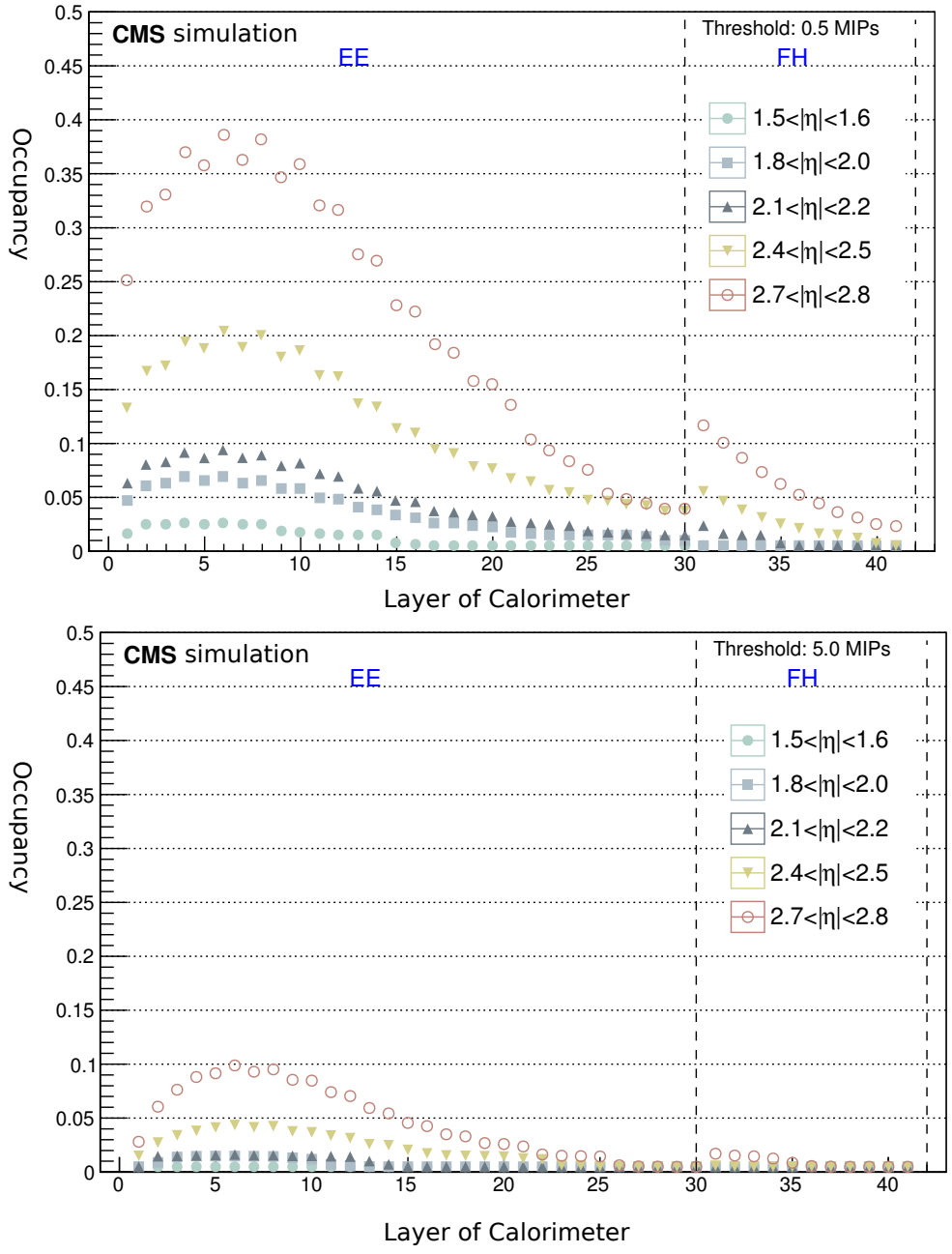


Figure 3.40: Median cell occupancy as a function of layer for signals (top plot) above 0.5 MIP and (bottom plot) above 5 MIP for an instantaneous luminosity such that the mean number of events per bunch crossing is 200.

length of the dead time depends on the signal size, and extends to 270 ns for a cell with 210 pC deposited. The dead time has been investigated using a simulation of a sequence of ten bunch crossings for instantaneous luminosities corresponding to means of 140 and 200 interactions per bunch crossing. The resulting probability for a sensor pad to be busy during any particular bunch crossing, as function of layer, for different η locations, is shown in Fig.3.41. It is found that the probability for a sensor cell to be busy is well below $\times 10^{-4}$ everywhere.

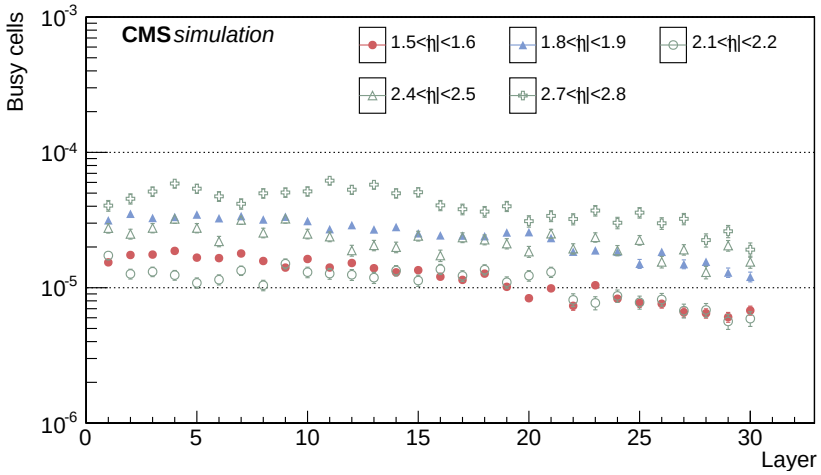


Figure 3.41: Probability for a sensor pad to be busy during any particular bunch crossing, as function of layer, for different η locations, when the instantaneous luminosity is such as to result in a mean of 140 pileup events per bunch crossing.

Showers for which some cells go over threshold are also susceptible to the effect of energy being deposited in cells for which the ToT front-end remains over threshold in subsequent bunch crossings. This effect, together with the effect of busy cells discussed in the previous paragraph, has been studied using photon showers in simulated $H \to \gamma\gamma$ events. The study is performed assuming that the energy in busy cells, and the energy deposited in shower cells which are over threshold, is simply added to the measured shower energy. Figure 3.42 shows the measured energy divided by the true energy, $E_\gamma / E_{\text{true}}$, for photon showers in simulated $H \to \gamma\gamma$ events before inclusion of the additional energy, with the inclusion of the energy from earlier crossings, and with the inclusion of the energy from both earlier and subsequent crossings. It can be seen that the subsequent crossings add more energy than the earlier crossings, but that the overall effect is still negligible.

A further feature of the fine lateral and longitudinal granularity that has been investigated is the determination of the direction of a shower. This is exploited in some of the methods that are currently used to estimate the energy of a shower in the presence of pileup, for example, it is found that an estimate obtained by summing, in each layer, only the cells surrounding the cell through which the axis passes, provides the best estimate of the energy of unconverted photons, and is almost completely insensitive to pileup of an average of 200 events per bunch crossing. The shower pointing precision that is currently achieved is ≈ 4 mrad for $p_T > 40$ GeV photons. Such a precision gives sensitivity to exotic signals with displaced vertices.

The power of a high-granularity calorimeter can only be fully exploited using fully developed particle-flow reconstruction algorithms. Particle-flow, PF, reconstruction makes use of both tracker and calorimeter information to yield much improved energy resolution and performance. A dedicated PF reconstruction package with an excellent performance, designed and tuned around the existing CMS detector, is used in current data taking and analysis by CMS.

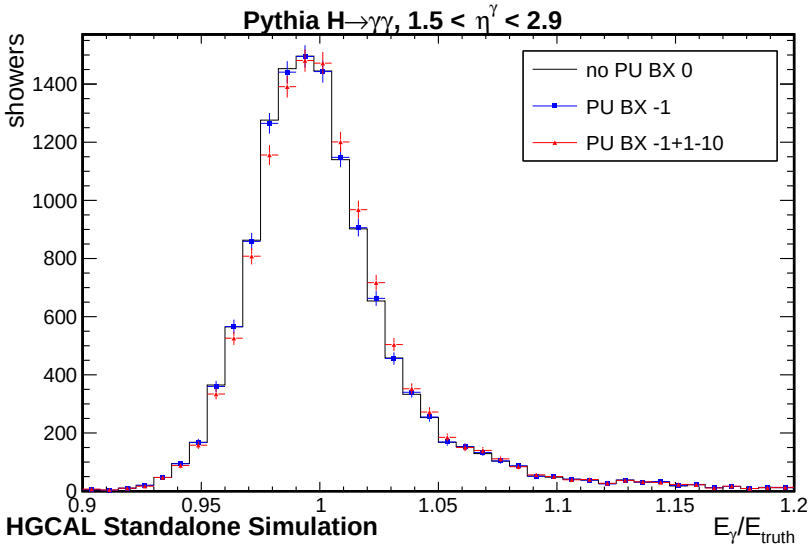


Figure 3.42: Measured energy divided by the true energy, $E_\gamma / E_{\text{true}}$, for photon showers in simulated $H \to \gamma\gamma$ events before inclusion of the additional energy, with the inclusion of the energy from earlier crossings, and with the inclusion of the energy from both earlier and subsequent crossings.

The large community of experts working on high-granularity calorimetry for a Linear Collider have also developed considerable expertise, and produced tools for generic particle-flow reconstruction. One such tool is the PandoraPFA C++ software development kit [99, 100] providing a highly sophisticated PF reconstruction for high-granularity detectors which is flexible and can be incorporated by various user applications and detector configurations. The PandoraPFA package is widely used by almost all ILC/CLIC studies, and by other high energy physics experiments as well (MicroBOONE). A full particle-flow reconstruction, interfacing PandoraPFA, has been integrated into the CMS software. Particle-flow objects are successfully reconstructed, linking tracks to calorimeter clusters composed of several calorimeter hits. Event displays show that complex multiparticle signals, such as hadronic tau decays, are clearly distinguishable from the pileup activity resulting from an average of 140 pileup interactions per bunch crossing.

Table 3.7: Milestones of the R&D, production and installation phases for the ECAL barrel upgrade. The target dates are the technically-driven milestone dates and at least three months of project float exists on each.

Milestone Identifier	Target Date	Description
TP.CA.EB.1	Q2 2016	Specifications set for VFE and FE electronics
TP.CA.EB.2	Q1 2017	First FE prototype in free-flow mode with 5GB/s links
TP.CA.EB.3	Q2 2017	Demonstrate feasibility of EB electronics with DC-DC converters instead of LV regulators
TP.CA.EB.4	Q3 2017	First prototype of the new VFE board with new chips
TP.CA.EB.5	Q4 2017	Technical Design Report
TP.CA.EB.6	Q3 2019	Final FE prototype in free-flow mode with 10GB/s links
TP.CA.EB.7	Q2 2020	Validation of all 3 boards (VFE, LVR, FE) with lab. tests, accelerated ageing and radiation exposure
TP.CA.EB.8	Q2 2020	Validation of off-detector electronics design
TP.CA.EB.9	Q3 2020	Electronics Systems Review

3.6 Project planning and cost estimates

3.6.1 Barrel Calorimeters Project Planning

As discussed above, the EB upgrade will involve operating the detector at 8°C, and replacing the very front-end and off-detector electronics. The steps planned to reach these goals are outlined in Table 3.7. The target dates are the technically-driven milestone dates and at least three months of project float exists on each.

The effort will begin with detailed simulation and emulation studies to understand what type of architecture best suits these needs. These simulations will allow us to define the technical specifications for the front-end chips and boards (see Section 3.4.3) by the middle of 2016 (TP.CA.EB.1). The structure and dimensions of the on-detector electronics is set by the pre-existing cooling blocks to which these cards must be attached.

With the specifications settled, prototypes of the FE, VFE, and LVR boards will be developed. For the FE board, the first prototype will demonstrate the feasibility of the planned control and continuous data read-out scheme at a reduced data rate (TP.CA.EB.2). Subsequent prototypes will follow the development of the CERN GBT chip up to the required data rate (TP.CA.EB.6).

The VFE board will require the development of pre-amplifier and ADC ASICs. This development will require approximately one year, with initial prototype chips submitted through a Multi-Project-Wafer service. The integrated prototype VFE board (TP.CA.EB.4) will be tested on the bench and using the ECAL spare super-module. Up to two cycles of design and MPW submission are possible in the schedule.

The LVR board can be designed once the needed voltages for the VFE and FE chips are known. Once the new DC-DC converters from the CERN micro-electronics group are available, a prototype card will be produced to allow testing with the ECAL spare supermodule for assessment of noise-levels (TP.CA.EB.3).

The results of the initial prototype electronics will allow a firm final design and the preparation of the technical design report (TP.CA.EB.5). Following the initial prototypes, a further round of prototypes is expected for final validation (TP.CA.EB.6). Several integration tests, including beam tests, will be carried out including the use of the spare supermodule (TP.CA.EB.7). Sev-

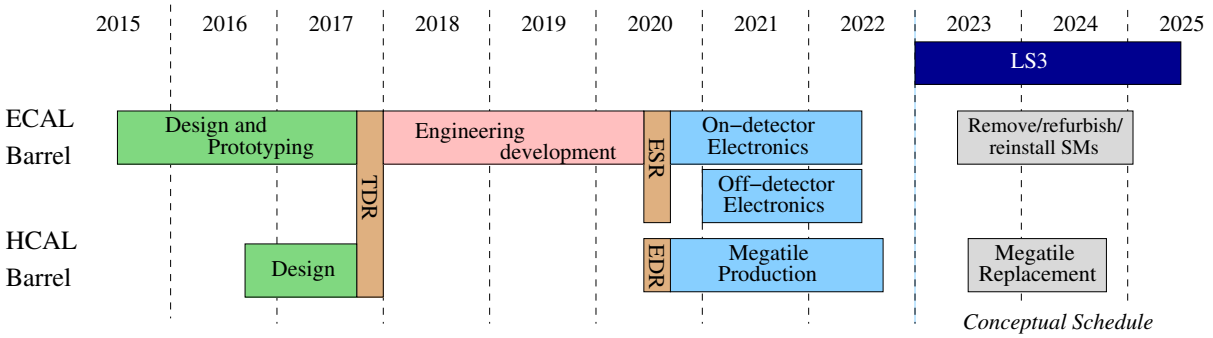


Figure 3.43: Schedule for barrel calorimeter upgrades.

eral small new mechanical components (including the SM patch panel, support structure for cables and fibers), must be designed and will also be tested using the spare SM. These will be further refined and produced in bulk in readiness for the final SM integration.

The off-detector electronics project requires simulation studies throughout the duration of the project, to develop and optimize the low-level data reconstruction that will be needed both for data readout and for triggering. The board design can start later than the on-detector electronics, so that the board production will profit from the latest developments in terms of optical communication and processing power. The off-detector electronics design will be validated using the prototype on-detector electronics and the spare SM (TP.CA.EB.8).

Based on the results of all tests, an Electronic Systems Review (ESR) will be carried out, to establish the readiness for full-scale production of boards (TP.CA.EB.9).

The schedule for the work on the barrel calorimeters is given in Figure 3.43. As discussed in Section 3.4.4, a total of 20 months is required for the removal, rework, and re-installation of the ECAL barrel supermodules. The electronics which will be installed will require two years of production and should be ready for installation six months before the beginning of LS3.

The design and preparation of the integration zone will occur during the electronics production: zone layout, preparation of the SM stands (support frames, cooling, and control electronics), software (control software, databases). The preparation of the area and the preparation of the spare SM will also serve to train the personnel of the integration zone. The preparation and test of the supermodule extraction tool (enfourner) must also be carried out in advance.

The HCAL barrel scintillator production will be combined with the production of megatiles for the endcap calorimeters to enable volume orders and effective use of tooling and production lines.

3.6.2 Endcap Calorimetry Project Planning

The milestones towards the construction of the endcap calorimeter are given in Table 3.8. The Technical Design Report is targetted for the end of 2017 (TP.CA.EC.9) and is informed by design and initial engineering work as well as prototyping, bench and beam tests.

Engineering design, validation, and qualification are necessary before production begins. During this time the schedule will be driven by the development of the front-end ASIC, and all other items are expected to remain in the shadow of this development. The intention is to freeze the specifications and architecture of the ASIC in the final quarter of 2015 (TP.CA.EC.2), and have the final version submitted for production in mid 2018 (TP.CA.EC.10). The plan includes a series of system tests to verify the design prior to launching major tenders, and the qualification of final components, prior to launching production, includes validation in the

Table 3.8: Milestones of the R&D, production and installation phases for the endcap calorimetry upgrade. The target dates are established for the purpose of project planning. Actual dates can be several months later in all cases without affecting technical progress.

Milestone Identifier	Target Date	Description
TP.CA.EC.1	Q4 2015	Receive first HGC-specific sensors for radiation testing and testbeam prototype construction
TP.CA.EC.2	Q4 2015	Define baseline architecture and specifications for on-module electronics
TP.CA.EC.3	Q2 2016	Define baseline module design and specifications
TP.CA.EC.4	Q1–Q4 2016	Prototype beam tests with existing electronics
TP.CA.EC.5	Q4 2016	Define baseline BH megatile and absorber design
TP.CA.EC.6	Q1 2017	Submit V1 of front-end ASIC
TP.CA.EC.7	Q3 2017	First tests of modules with V1 front-end electronics
TP.CA.EC.8	Q3 2017	Define baseline design for cassettes and mechanics
TP.CA.EC.9	Q4 2017	Submit Technical Design Report
TP.CA.EC.10	Q2 2018	Submit V2 of front-end ASIC
TP.CA.EC.11	Q2 2019	System and beam tests of modules with V2 front-end electronics and final modules
TP.CA.EC.12	Q3 2019	Engineering Design Review

context of a full cassette system. A first version of the ASIC will be submitted in early 2017 (TP.CA.EC.6) and available for tests in mid-2017 (TP.CA.EC.7), although it may have reduced functionality as compared to the final ASIC. The second version of the ASIC, a preproduction version having full functionality, would be used for full final bench and beam tests in the first half of 2019 (TP.CA.EC.11).

A comprehensive program of R&D will be carried out to understand sensor, calibration, and system aspects of the detector and to fully benchmark and validate the detector simulation. The program will begin with the availability of the first sensors in late 2015 (TP.CA.EC.1). These sensors will combined with existing readout chips (e.g. the PACE chip developed for the CMS ECAL preshower detector) to allow a series of bench and beam tests in 2016 (TP.CA.EC.4). The sensors will also be used for mechanical and radiation studies leading to the specification of the module design (TP.CA.EC.3). Further engineering will lead to the baseline designs for the cassettes and the detector mechanics (TP.CA.EC.8). For the backing calorimeter, the baseline mechanical design, scintillator design and segmentation, and optical readout scheme will be defined by the end of 2016 (TP.CA.EC.5).

The conceptual schedule for the HGCAL and backing-HE endcap calorimetry option is shown in Fig. 3.44. The production and testing of sensors requires two and a half years, beginning in 2019. Module assembly requires two years. Once sufficient modules are available, the assembly of cassettes will begin. Completed cassettes are inserted into the mechanical structure to form completed endcaps. The endcaps are then sealed and prepared for cold operations. Nine to twelve months testing and commissioning of each complete endcap on the surface are envisaged. The two endcaps will be ready for lowering to the experimental cavern in the third quarter of 2024 in accordance with present CMS planning for the long LHC shutdown LS3.

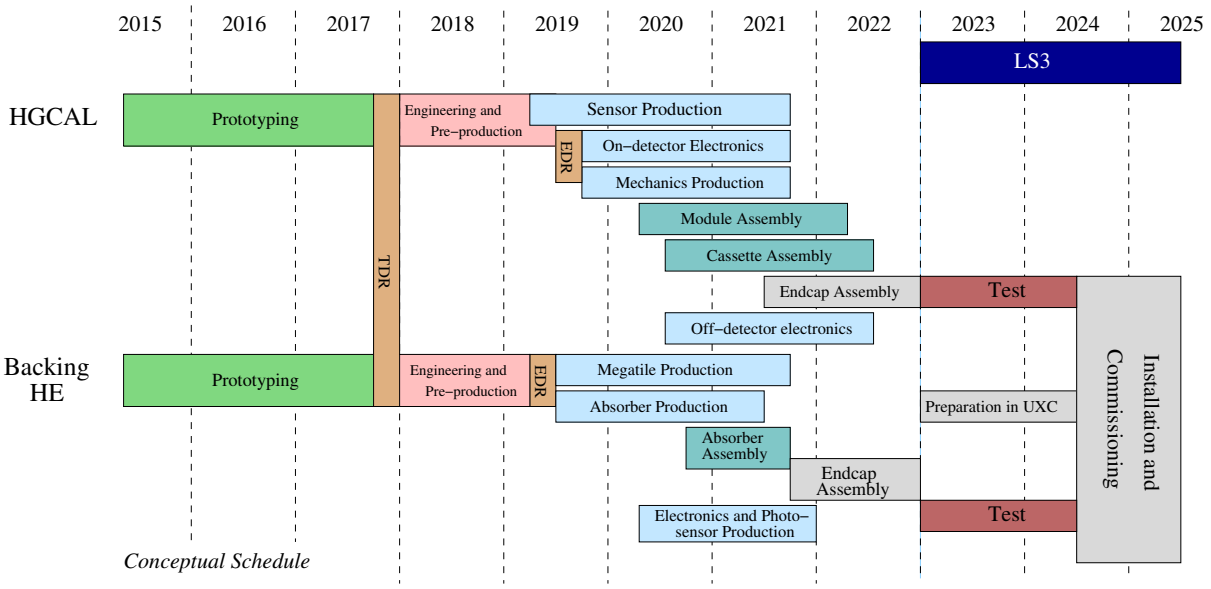


Figure 3.44: Schedule for the HGCAL and backing-HE endcap calorimeter system.

3.6.3 Cost Estimates

3.6.3.1 Barrel Calorimeters

The cost estimates for the barrel calorimeters are given in Table 3.9. The ECAL barrel upgrade front-end upgrade costs are based on the original construction costs updated to 2014. The ASIC costs have been updated using quotes for masks and wafers for the 130 nm technology which is presently targeted for the design. The costs for fibers, cables, and pipework are based on the accurate lengths used for the original construction, updated for current costs.

The HCAL barrel costs are based on the replacement of five layers of scintillator, using quotes obtained recently for the scintillator, wavelength-shifting fiber, and clear fiber. The replacement of the optical decoder in the HB front-end is also included, but the rest of the Phase-I upgrade on-detector electronics is assumed to be reused.

Table 3.9: CORE cost estimate for the ECAL and HCAL barrel calorimeter upgrades.

Estimated CORE cost in MCHF (2014)	
Front end electronics	5.7
Off-detector Services	4.3
Total EB	**10**
Rebuild tiles	1.1
Optical distribution and installation	0.3
Total HB	**1.4**
Total Barrel Calorimeters	**11.4**

3.6.3.2 Endcap Calorimetry

The cost estimate for the high-granularity calorimeter and backing-HE endcap calorimetry system is given in Table 3.10. The module cost is based on vendor quotations for silicon and 130 nm readout ASIC mask and wafer costs. The costs for the mechanical structure are based on the CALICE prototype, while the CO_2 cooling costs are based on the CMS Phase-I Pixel Upgrade

costs. The backing-HE costs are based on the use of plastic scintillator only. It is assumed that the readout cards must be replaced due to required changes in the readout box volume.

Table 3.10: CORE cost estimate for the endcap calorimetry system.

Estimated CORE cost in MCHF (2014)	
Mechanical Structures	5.4
Silicon Modules	28.7
On-detector electronics and services	3.4
BE Electronics and Controls	2.5
Power System	5.0
Cooling System	7.5
Assembly and Installation	1.4
Total HGCAL	**54**
Absorber	6.6
Scintillating Tiles and WLS fibers	1.3
Readout and on-detector electronics	1.2
BE Electronics and Controls	0.3
Power System	0.1
Assembly and Installation	0.3
Total Back HE	**10**
Total Endcap Calorimeters	**64**

Chapter 4

Muon Systems

4.1 Introduction

4.1.1 Goals of the muon upgrade

The muon detector system of CMS consists of three types of muon detectors: Drift Tubes (DT), Cathode Strip Chambers (CSC), and Resistive Plate Chambers (RPC), as shown in Figure 4.1. The muon system was described in detail in the original muon TDR [101]. Most of the system was installed in 2007, while the large chambers in the fourth endcap disk (ME4/2, RE4/2 and RE4/3) were installed in 2014. The DT and CSC provide precise measurement points for muon identification and triggering for $|\eta| < 2.4$. In addition, the RPC provide an extra muon trigger for $|\eta| < 1.6$. The muon detectors are located at four "stations", each separated by a sufficient thickness of iron from the next to isolate electromagnetic showers due to muon bremsstrahlung to a single station. The system provided excellent triggering, muon identification, timing, and momentum measurement during Run-I [102, 103].

The muon system of CMS was designed to operate in the initial LHC environment with luminosity up to $10^{34}/\text{cm}^2\text{s}$. It is not possible to maintain the high level of performance achieved during Run-I in the environment of HL-LHC without an extended program of upgrades to many of the system components. The main factors that need to be taken into account are the higher instantaneous and integrated luminosity, the possible detector degradations during the long time periods before and during Phase-II, and changes to the trigger, most notably the increases in L1 latency and rate described in Chapter 6. In addition, it is planned to add a "Track Trigger", i.e. candidate tracks with p_{T} of at least 2 GeV in the inner silicon tracker that are matched at L1 to muons and extend the inner silicon tracker coverage to $|\eta| > 2.4$ (Chapter 2).

It is important for Phase-II physics to keep the efficiency of the L1 muon triggers high, while maintaining p_{T} thresholds low enough to collect a large fraction of Higgs, top quark, and electroweak bosons for more sophisticated analysis. Specific physics processes where low thresholds are especially important include: Higgs decays through ZZ^* to two or four muons, Higgs decay to $\tau^+\tau^-$ where one τ decays to a muon, and certain SUSY cascade decay chains where mass splittings are small (compressed spectra). Generically, thresholds of about 20 GeV yield high efficiency for electroweak boson and top quark decays to leptons. Additionally, some exotic physics models predict signatures such as muon-like slow particles that may be neutral in the inner tracker, muons that do not point back precisely to the primary vertex, and closely collimated sets of two or more muons; the ability to identify all of these signatures requires good standalone muon trigger capability and enhanced ability to resolve multiple muon tracks. Some of these processes are discussed in Section 10.

There are no Phase-I muon upgrades other than the improvements already made during LS1, namely, the installation of large chambers in the fourth endcap disk, installation of optical links

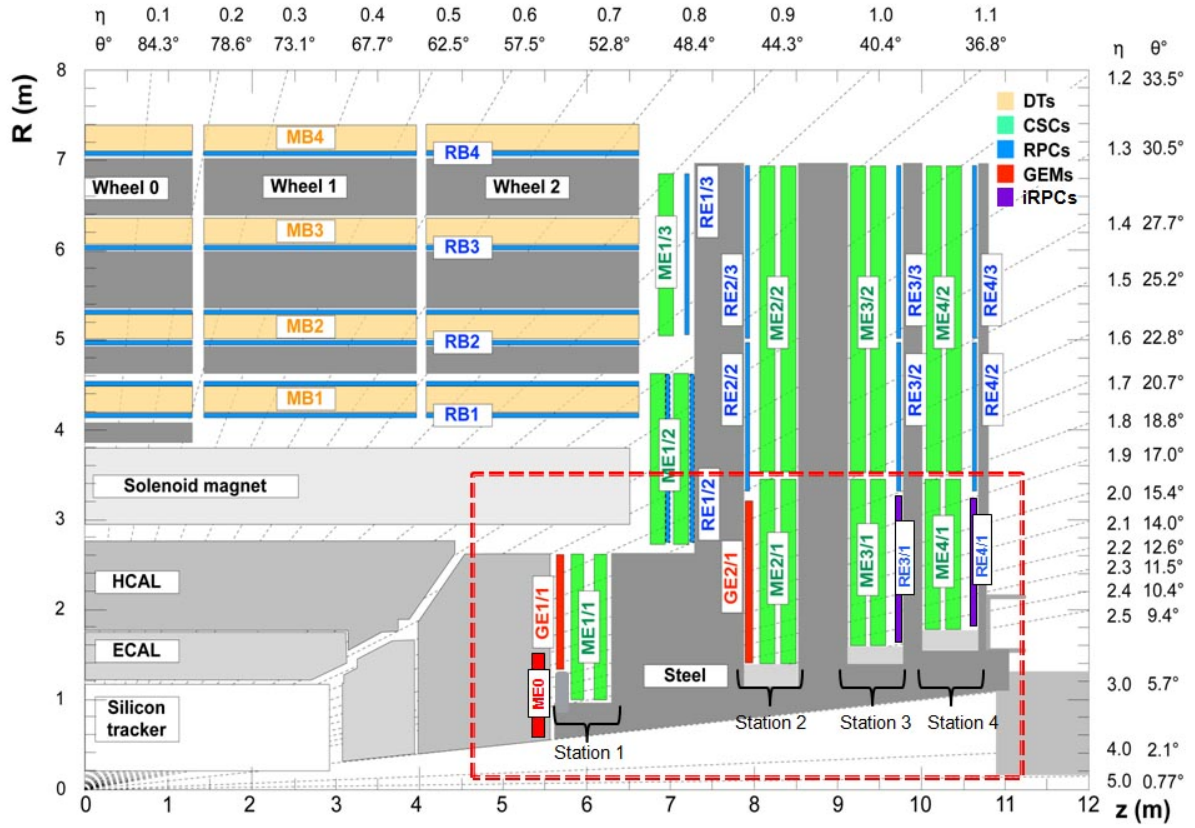


Figure 4.1: A quadrant of the muon system, showing DT chambers (yellow), RPC (light blue), and CSC (green). The locations of new forward muon detectors for Phase-II are contained within the dashed box and indicated in red for GEM stations (ME0, GE1/1, and GE2/1) and dark blue for improved RPC stations (RE3/1 and RE4/1).

that allow relocation of some DT electronics from the collision hall, and installation of improved electronics in the innermost set of CSC chambers (ME1/1).

There are three types of muon upgrades proposed for Phase-II: (i) upgrades of existing muon detectors and associated electronics that ensure their longevity and good performance, (ii) additional muon detectors in the forward region $1.6 < |\eta| < 2.4$ to increase redundancy and enhance the trigger and reconstruction capabilities, and (iii) extension of muon coverage up to $|\eta| = 3$ or more behind the new endcap calorimeter to take advantage of the pixel tracking coverage extension. Overviews of each type of upgrade are presented below, while further details are included in subsequent sections of this chapter.

4.1.2 Upgrade of existing muon detectors

The present muon system is expected to provide excellent performance throughout the HL-LHC program. However, it is known that DT electronics will need replacement due to limited radiation tolerance of some components; this replacement also gives the opportunity to increase the trigger rate capability and performance, and improve maintainability. Additionally, the 108 inner-ring CSC chambers ME2/1, ME3/1, and ME4/1 will need to have their front-end cathode cards replaced, since the combination of increased occupancy plus larger L1 trigger rates and latency in Phase-II will cause their analog pipelines to fill up and lead to unacceptable deadtime.

Since the HL-LHC provides radiation levels and doses well beyond the design expectations of the LHC, new irradiation tests must be performed to confirm that all types of muon detectors will survive the harsher conditions. The past irradiation studies for DTs, CSCs and RPCs with gamma rays need to be repeated at higher doses to better understand the HL-LHC aging effects, such as deposits on anodes and cathodes that will be encountered.

Some modification of existing gas mixtures presently used in RPCs (Freon) and CSCs (CF_4) are desirable to comply with new greenhouse gas rules. Any change of gas mixture must be considered very carefully, as the existing mixtures have proven to yield excellent performance with no sign of systematic aging or other negative effects during LHC operation thus far.

4.1.3 Enhancing the redundancy and trigger in the forward region

For good performance of the muon trigger and offline muon identification in the presence of background, delta rays, bremsstrahlung, etc., it is important to record a sufficient number of muon detector hits on each track. In the Phase-I CMS detector, the number of muon hits is actually least in the most challenging region, i.e. for $1.6 < |\eta| < 2.4$, where the background rates are highest and bending is much reduced. Moreover, the CSC stations are relatively thin, so that the rejection of low-p_{T} muon or punchthrough segments by angle within the stations, as is done in the DT trigger [104], is ineffective. The Phase-II muon upgrade recovers an effective muon system in this region by the addition of new forward muon detectors within spaces that were allocated for RPCs but not instrumented, due to the limited rate capabilities of the previous generation of RPC detectors. The new detectors add additional measurement points and dramatically increase the lever arm of each forward muon station.

The locations of the additional forward muon detectors, specifically GE1/1, GE2/1, RE3/1, and RE4/1 are indicated in the $r - z$ quadrant diagram of Figure 4.1. Since the original design of CMS included RPC detectors in these locations, the space, mounting features, and some services already exist for the addition of the new detectors.

The first two muon stations, where the bending angles are largest and thus momentum determination most effective, will be equipped with GEM detectors denoted GE1/1 and GE2/1 that each contain two high-precision measurement layers. These increase the path length within each station to 28-55 cm, versus $11 - 15$ cm provided by the CSC alone. Preparations for a demonstrator consisting of two GE1/1 chambers is already well advanced and planned for installation during the year-end 2016 technical stop, while the full GE1/1 station could be installed as early as long shutdown 2 (LS2), rather than the nominal Phase-II installation window in LS3.

In the 3rd and 4th muon stations, improved RPC (iRPC) detectors, denoted RE3/1 and RE4/1, will provide much-needed redundancy at modest cost, with benefits to neutron-induced background reduction for the trigger and offline reconstruction. If improved multi-gap RPCs are used, their excellent timing precision of $\sim$ 100 ps can be used for pileup mitigation, and to provide an excellent signature and mass measurement for hypothetical heavy stable charged particles (HSCP).

The technologies of the new detectors easily handle the moderately high particle hit rates $\sim$ 1 kHz/cm^2 expected at maximum HL-LHC luminosity in this region. Moreover, while the CSC detectors have a crossed strip and wire geometry that suffers intrinsically from ghosting in multi-muon events, the new detectors are configured with finer segmentation in η, which allows ghost removal in such events.

4.1.4 Muon pseudorapidity extension

With pixel tracking extension up to close to $|\eta| = 4$, and the replacement of the endcap calorimeters for Phase-II comes the opportunity to extend the muon coverage beyond the present limit of $|\eta| < 2.4$ with the addition of a small but precise muon detector installed at the back of the more compact calorimeters. Since this detector is in front of existing muon endcap station 1, it is denoted ME0, as shown in Figure 4.1. The extended muon coverage should increase the acceptance for all final states with muons, especially multiple-muon states. The additional coverage should also reduce backgrounds where missing transverse energy can be generated or altered by non-identified high-p_T muons.

Since the new endcap calorimeter will be more compact that the current design, approximately 30 cm in the z coordinate is available for this purpose. The GEM technology is easily able to handle the $\sim 10 - 100$ kHz/cm^2 background hit rates that are expected in this region, and six layers are foreseen in order to better reject neutron-induced background hits. The design of the ME0 station will incorporate shielding at the inner radius, but otherwise will extend in $|\eta|$ nearly to the limit of the endcap calorimeter coverage.

4.2 Upgrade of existing muon detectors

4.2.1 Overview of existing muon detector upgrade

During Run-I the CMS muon system operated well inside its design radiation tolerance, and no aging effects were observed. All muon subsystems performed very stably and with high efficiency ($> 95\%$), thus contributing crucially to the Higgs discovery as well as searches and measurements for many other physics signatures comprising muons in the final state. In the future the muon detectors and their associated electronics must operate at much higher luminosities while accumulating the large doses expected from more than 3000 fb^{-1} of acquired data from collisions at the HL-LHC.

4.2.2 Background nature and rates

Background radiation in the muon system arises from low-energy gamma rays and neutrons from p-p collisions, low-momentum primary and secondary muons, punch-through hadrons from the calorimeters, muons and other particles produced in the interaction of the beams with collimators, residual gas and beam-pipe elements. All major background sources affect mostly detectors at the highest pseudorapidity.

In the most forward region ($|\eta| > 1.6$) of CMS, where the neutron flux is highest, the total fluence for neutrons with E>100 keV, after ten years of operation at the LHC design luminosity of 10^{34} cm^{-2}s^{-1}, will be about 6×10^{11} cm^{-2}, and the total ionization dose (TID) $\sim$ 1.8 krad (Figure 4.2). In the barrel region the neutron flux will be more than one order of magnitude lower.

At the HL-LHC neutron fluences are expected to be five times larger than at LHC, that is 3×10^{12} cm^{-2} and $\sim$ 8 krad, over ten years of operation at 5×10^{34} cm^{-2}s^{-1} [101]. Even considering an upgrade of the detector background shielding, such extremely high radiation background will still exceed the design tolerances of several components of the muon system, thus requiring a new assessment of the detector and electronics longevity (radiation tolerance, aging of components, long-term behavior) and performance (rate capability, detection efficiency, pattern recognition, resolution) under accelerated aging tests.

During LHC Run-I, background rates in the muon detectors have been measured at 8 TeV for

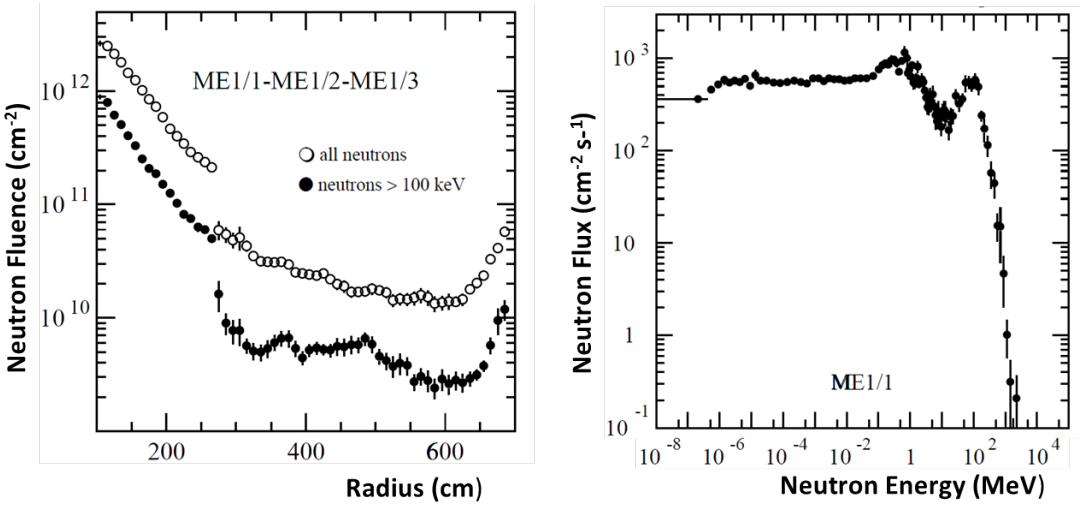


Figure 4.2: (Left) Expected neutron fluence in the ME1 station after ten years of LHC operation at 10^{34} cm^{-2}s^{-1} luminosity shown as a function of the distance from the beam line. The steep variation at R=270 cm is due to the shielding geometry of the first endcap disk. (Right) simulated neutron energy spectrum in the ME1/1 region (R$<$ 260 cm, z=570 cm) during standard LHC operation [101].

several luminosity values up to $\sim 7 \times 10^{33}$ cm^{-2}s^{-1} [105]. In the CSC system, the measured rates extrapolated linearly to the LHC design luminosity of 10^{34} cm^{-2}s^{-1} and 14 TeV center-of-mass energy, will reach 900 Hz/cm^2 at the innermost part of the ME1/1 chambers. Similar measurements have been done with the RPCs, both endcap and barrel, and barrel DT chambers (Table 4.1). In the barrel, the highest rates are in the outermost stations (RB4, MB4), mainly due to diffuse neutron background, and in the innermost stations (RB1, MB1), mainly due to charged particles and punch-through hadrons. On the other hand, in the endcap region the highest rate ($\sim$ 50 Hz/cm^2) is measured in station 2 (RE2). Such rates are however $\sim$ 20 times smaller than the highest measured CSC rates because of the rapidly increasing background rate above $|\eta| = 1.6$, where there are no RPCs.

At the HL-LHC center-of-mass energy of 14 TeV and leveled luminosity of 5×10^{34} cm^{-2}s^{-1}, the background rates are expected to extrapolate linearly from the measured data. The highest estimated background rate (innermost ME1/1 region) is $\sim$ 4.5 kHz/cm^2, and the total collected charge integrated over ten years and 3000 fb^{-1} is $\sim$ 0.3 C/cm on the ME1/1 anode wires. In the barrel, the maximum expected hit rate in the DT is of the order of 25 kHz per drift tube (with an average area of 1000 cm^2). In the RPC, the maximum expected rate is $\sim$ 250 Hz/cm^2 (innermost RE2/2 region) and $\sim$ 2 times smaller in barrel region, well below the maximum tested rate in beam test $\sim$ 1 kHz/cm^2 [106] [107]. However, long-term aging tests have only certified real-size RPCs up to $\sim$ 300 Hz/cm^2. An effort to estimate rates from full simulation of flux and detector sensitivities at HL-LHC is ongoing to confirm predictions from LHC data. Table 4.1 summarizes present estimates and indicates, in comparison, the largest integrated charges to which real-size chambers were certified in past aging tests.

4.2.2.1 Irradiation tests at GIF^{++}

To perform irradiation tests at HL-LHC levels, a new facility, GIF^{++} [108], has been built at CERN and will come into operation by early 2015. At GIF^{++}, detectors and material samples can be irradiated by 662 keV photons emitted by an intense 16.7 TBq ^{137}Cs source with an

Table 4.1: Expected maximum background rates at $\sqrt{s}$=14 TeV and LHC design luminosity of 10^{34} cm^{-2}s^{-1} and at HL-LHC peak luminosity of 5×10^{34} cm^{-2}s^{-1}, and maximum rates and maximum integrated charge reached in past aging tests. Rates are given in Hz/cm^2. No safety factor is included in the estimation.

Detector	Bck. rate LHC	Bck. rate HL-LHC	Max. rate tested	Max int. charge 3000 fb^{-1}	Max int. charge tested
DT	5	25	100	0.15 C/cm	0.48 C/cm
CSC	900	4500	20000	0.3 C/cm	0.35 C/cm
RPC	50	250	300	0.27 C/cm^2	0.05 C/cm^2

average flux rate of $\sim 3 \times 10^7$ Hz/cm^2 at $\sim$ 2 m and $\sim 5 \times 10^6$ Hz/cm^2 at $\sim$ 4 m. The source is combined with a high-momentum particle beam (muon, electron, proton) extracted from the SPS along the H4 beam line. With a momentum of 100 GeV/c and an intensity of 10^4 μ/spill, muons will constitute excellent probes for performance studies in the presence of a very high radiation environment. The large and uniformly irradiated surface will allow testing several large detectors simultaneously to rapidly collect doses comparable to those accumulated after several years of operation at HL-LHC luminosity. All existing and future muon detectors are planning to carry out extensive irradiation campaigns of detectors and electronics, as detailed in the following sections. Since the dose rates given in Table 4.1 are based on extrapolations from present knowledge and assumptions on the features of the background spectrum at the HL-LHC, all subsystems must take into account large safety factors ($\sim$ 3) in testing. Final doses will be collected in one to two years of irradiation at GIF^{++}, allowing the finalization of most radiation tolerance tests. Depending on the results, further investigations or aging mitigation measures will be foreseen.

4.2.3 Drift Tube Chambers for Phase-II

In preparation for the HL-LHC, DT upgrade plans are focused on:

- Aging of the DT chambers: from previous tests we expect only minor degradation. No intervention is foreseen on the chambers themselves, but studies will be done to further reduce the expected degradation, for example, by lowering the HV. The aging and performance of the DT chambers in the HL-LHC environment will be thoroughly tested at the GIF^{++} facility.
- Electronics upgrade. The first level of readout and trigger electronics (so called Minicrate [104]) is a complex system embedded inside the CMS wheels. The survival of this system is not guaranteed in the harsh HL-LHC environment and it is incompatible with increasing the Level 1 trigger acceptance rate beyond 300 kHz. Consequently, the substitution of this electronics is planned for LS3, moving the complex trigger logic to the service cavern. We expect this replacement to allow the implementation of full time (and therefore space) resolution and to provide complete chamber information in the DT trigger system, with improved performance in terms of rate reduction and better matching with the tracker at the Level 1 trigger. Eventually, losses due to dead time in the current trigger electronics will also be eliminated.

In addition, for Run-II a trigger link is being prepared that makes the RPC and HO (equipped with SiPM during LS1) signals available for building muon trigger primitives together with inputs from DT. The benefits of the HO layer for MIP identification, punch-through rejection

and resolution of ambiguities in the barrel muon system are currently being studied with simulations (Muon Track Fast Tag, [109]). Data collected during Run-II will allow the study of the trigger primitive algorithms and the optimization of the HO granularity for L1 muon triggering in Phase-II, in combination with the barrel muon track finder and the Track Trigger systems.

4.2.3.1 Drift Tube chambers and electronics longevity

DT chamber construction is extensively described in [101]. Previous longevity studies [110] are summarized here. Several DT aging effects are considered, and described as follows.

Laboratory tests of the insulation breakdown in a tube indicate that, at the typical voltage of 2 kV/cm, the system lifetime is on the order of 10^8 years, which translates into a 4% probability of failure of one out of the 180000 tubes of the full DT system in twenty years (see page 65 in [101]).

The electronics inside the gas volume of the chamber are the HVB (High Voltage Boards) and the FEBs (Front End Boards), both extremely time consuming to replace. The HVBs are a multilayer FR4 structure hosting passive components only. Failure is highly improbable and not expected if properly operated. Only one HVB out of $\sim$ 11000 has been replaced during LS1. FEBs were tested up to a radiation dose and a neutron flux almost one order of magnitude larger than expected at HL-LHC and no damage was detected. Extensive accelerated aging tests were also performed on 20 FEBs at high temperature (125^{o}C) for 3200 hours and no degradation was observed, translating into a maximal efficiency loss of 2.8% by the end of HL-LHC.

The DT chamber gas mixture is Ar/CO_2 (85/15). Gas leaks, changes in the gas mixture, gas contamination with oxygen, water or even pollutants from outgassing of chamber material can degrade the system behavior. All these aspects are carefully monitored to minimize contamination [104].

Wire aging induced by all the components present inside a tube (glues, O-Rings, plastic parts) was excluded by measurements up to an accumulated charge of 0.48 C/cm that corresponds to three times the expected dose in ten years of HL-LHC [101]. These results and the effects of even higher radiation dosages will be studied at the GIF^{++} facility.

Table 4.2: Estimated maximum loss of efficiency in the DT detector as a consequence of chamber aging at the end of ten years HL-LHC operation (about 3000 fb^{-1}).

Lifetime of the tubes	1%
FEB failure rate	2.8%
HVB failure rate	1.5%
Lower efficiency due to radiation	1%
Total chamber efficiency loss	6.3%

Table 4.2 summarizes the contributions of the various aging effects. In total, and conservatively, we expect less than 6.3% inefficiency due to chamber aging. Studies of the muon reconstruction within several failure scenarios of the DT detector show that such an inefficiency will have a negligible impact throughout HL-LHC. In addition, we are studying a reduction of the strip voltage by 300$\sim$400 V while keeping the wire gain constant, reducing the probability of insulator failure by about one order of magnitude.

4.2.3.2 Drift Tube electronics upgrades

Each DT chamber houses the L1 Trigger and Readout electronics in an aluminum profile, called a Minicrate [104], which hosts the time digitization logic and the complex logic for the L1 trigger primitive generation. The Minicrate is a complex custom assembly with several types of ASICs that cannot be rebuilt today, as well as many interconnecting cables and dense connectors. It cannot be operated without water cooling because of its high power consumption (85 W). It is deeply buried inside the iron yoke and covered by cables and services for other detectors. As a result, it is accessible only when the barrel wheels are open during Long Shutdown periods.

Each one of the 250 Minicrates contains a paired Control Board (CCB) and Server Board (SB), one CCB link board and, on average, six pairs of Read Out (ROB) and Trigger Boards (TRB). During the tests to certify their radiation resistance to ten years of LHC operation, samples of all the Minicrate components and boards were irradiated up to half the doses expected at HL-LHC (a bit more in the case of the TRBs) without any damage. However, for the SB/CCB pack, two components died in a single event, after an irradiation of 10^{11} cm^{-2} protons of 60 MeV energy in tests performed at PSI in November 2003. Tests performed by an independent group also confirm the sensitivity to radiation of one of these components [111].

In addition, operational experience highlighted reliability issues related both to the system complexity and to the limited resilience of the boards (TRBs and ROBs) against power cycling. These issues severely restrict the viability of operating the system up to the end of HL-LHC. Furthermore, the ROB processing speed limits the L1A rate to 300 kHz.

Finally, the high- and low-voltage supplies located in the CMS tower racks are also exposed to radiation. The DT system is one of the largest power supply systems in CMS and the radiation tolerance in HL-LHC of some module types should be verified. A significant decrease in maintenance and operations cost for the power system would be obtained by reducing the power required in the Minicrates.

For these reasons, and because partial rather than full replacement of components in the current Minicrates will require three times longer to install and would compromise the internal connectivity, which might impinge on the length of LS3, it is proposed to replace the full Minicrates.

4.2.3.3 Drift Tube chamber Minicrates upgrade

For simplicity and ease of maintenance, the new Minicrates will only implement the time digitization and control functions. The higher bandwidth needed for data transfer can easily be achieved at low cost using current optical link technologies. This design has the advantage of reducing the power consumption and so requires fewer low voltage power supplies.

The functionalities that will remain close to the DT chamber are: time digitization of the chamber FEB analog signals, control of the FEBs, test pulse generation, monitoring of DT chamber sensors (temperature, pressure, etc.), and slow control services for the RPC and Alignment systems. These will be implemented in a single type of board. Using a similar form factor as the legacy TRB/ROB, there will be 6-7 boards per chamber with 128 channels each (Figure 4.3, left). However a more robust architecture is being studied in which the superlayers are completely decoupled from each other and have independent boards. This new On-Board Electronics for DT (OBEDT) board will receive the chamber signals from each wire and will assign a digital time stamp.

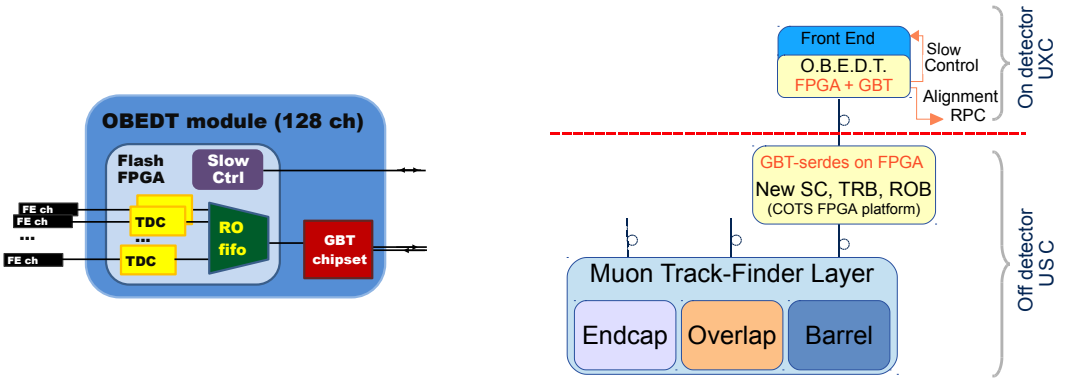


Figure 4.3: Block diagrams of proposed DT electronics. Left: scheme of on-board electronics (OBEDT). Right: overall view of the on-detector and off-detector components of the proposed architecture.

The resolution obtained with the present HPTDC [112] of ∼1 ns is sufficient for the DT detector and can be now obtained using FPGA technology. The logic needed for the TDC implementation has been studied using the last but one generation of Microsemi flash FPGA (130 nm), which is known to be radiation tolerant. It is estimated that using the largest part available, about 100 TDCs of 5-bit range and 1 ns resolution can be accommodated in approximately 10% of the FPGA logic blocks.

Once digitized, hit times will be sent to USC through an optical connection. In order to evaluate the total bandwidth required for a basic OBEDT module, we assume 128 channels per module, a maximum sustained rate of 1 MHz/channel (1 kHz/cm^2), and hit time encoding in 16 bits (TDC count + TTC time reference tag). Then the total bandwidth required for each of the OBEDT modules is about 2 Gb/s, which fits within the capability of a single 3.125 Gb/s CERN GBT link [34].

4.2.3.4 Drift Tube chamber back-end electronics upgrade

The current implementation of the DT electronics [113] not only introduces significant limitations to the maximum readout rate (300 kHz due to ROB processing) but severely constrains the trigger primitive algorithms, most notably because of the low input sampling frequency (80 MHz) and the multiple ghosts due to the highly partitioned architecture. In the proposed upgrade, the trigger primitive generation and the readout event matching will be performed in the underground control room on generic hardware, based on the latest commercial FPGAs. Full chamber information at maximum TDC resolution will be available to improve the DT local trigger primitives. In particular:

- Time resolution of the sampled hits can be improved from the current 12.5 ns to about 3 ns, leading to higher trigger primitive resolution in position, angle, p_T and probably bunch-crossing identification.
- Deadtime can be reduced from 400 ns to about 100 ns which will reduce the loss of hits that are masked by background signals.
- Superlayer granularity for local track fitting will not be limited to groups of nine cells as in the present system. This should imply fewer ghosts and reduced bunch crossing misidentification.
- Improved DT trigger z resolution measurement (from 32 cm to about 1 mm) and the

possibility of early combination of ϕ and θ views will lead to a smaller region for matching with the tracker information.

- The number of trigger track segments that can be identified at the same bunch crossing will not be limited to two per chamber as in the current system.
- The trigger primitive generation may be able to combine information from neighboring chambers, depending on the hardware resources available at the time of construction.

New trigger primitives will have a spatial and angular resolution close to those of present HLT reconstructed segments. Thus, an upper limit for the gain in performance can be estimated considering the present standalone muon reconstruction at HLT. With respect to the current L1, the higher resolution and more sophisticated tracking algorithm will allow the standalone L1 trigger rate to be reduced, and a more efficient and robust matching with the Track Trigger (see Section 6.4).

Studies will be needed to define the algorithm that provides the best improvement at a reasonable implementation cost. Bunch Crossing and Track Identifier (BTI ASIC) and Track Correlator (TRACO ASIC) algorithms used in the present system have been used to estimate the resources that would be needed.

A possible architecture based on an appropriate FPGA card, like a future version of the MP7 [96], based on the Xilinx Virtex-7 FPGA, which is used in present Phase-I upgrades, has been investigated. Under the assumption of an increase in available logic and memory in the FPGA of a factor of ten at the same price as the present MP7, we estimate it will be possible to fit the trigger and readout algorithms for one sector (four or five chambers) on two of such boards, receiving also RPC and HO data. In addition, a final layer for cleaning duplicates and data concentration will be needed, giving a total of 132 such boards. The slow-control links and logic for the DT wheels can be accommodated in five more cards, based on lower-power and lower-cost FPGAs.

A diagram of this system is shown in the right side of Figure 4.3.

4.2.4 Cathode Strip Chambers for Phase-II

In preparation for HL-LHC, the CSC upgrade plans are focused on:

- Aging tests of CSC detectors and electronics beyond the doses expected at the HL-LHC to assess the safety margin and study alternative gas mixtures.
- Replacement of front-end cathode electronics boards on inner chambers in order to handle the increased L1 trigger latency and rate described in Section 6.1.

4.2.4.1 Aging and gas mixture tests of the CSCs

The CSCs were tested for aging effects at the GIF facility in the early 2000's, and found reliable up to the accumulated charge ($0.3 - 0.4$ C/cm of wire length) expected after thirty years of operation at LHC design luminosity [114, 115]. However, ME1/1 chambers, which in CMS cover the forward region ($1.6 < |\eta| < 2.4$) and will be exposed at the HL-LHC to a maximum of $\sim$ 5 kHz/cm^2 at the innermost radius, have not been as thoroughly tested. The ME1/1 chambers were built using different epoxies and with some of their components purchased from different vendors, as compared to the rest of the CSC chambers in CMS. It will therefore be extremely important to re-assess the radiation hardness of all materials and detect any sign of performance degradation at GIF^{++}.

Two fully instrumented CSCs will be tested at GIF^{++}: one ME1/1 and one ME2/1 "control" type, with the goal of ensuring a safety margin of ~ 3 with respect to the expected charge accumulated over the full period (~ 10 years) of HL-LHC operation. This irradiation corresponds to an accumulated charge of 1 C/cm (Table 4.1), which is feasible at GIF^{++}. For example, the dose collected in one year at an expected gamma rate of 100 kHz/cm^2 and a 50% duty factor yields an equivalent charge of 1.5 and 2 C/cm for ME1/1 and ME2/1 chambers, respectively. (The different charge is due to the different wire spacing of ME1/1 and ME2/1 chambers, 2.5 mm and 3.12 mm, respectively.)

Irradiation will consist of both low- and high-dose rate sequences. The low irradiation cycles are intended to simulate the radiation environment expected during runs at the HL-LHC, whereas the high-dose irradiation cycles are for achieving the final TID. During the low irradiation runs we plan to collect hit rate statistics and monitor the chamber parameters: anode currents and counting rate of anode wire signals. During the breaks between high dose exposures we plan to measure various chamber characteristics, such as: dark current from cathode strip planes, anode wire signal noise, the efficiency versus HV (plateau point) for cosmic ray muons, and strip-to-strip resistance. After the irradiations have been completed, we may open the chambers and investigate the physical features that correspond to any aging encountered (e.g. analysis of deposits on anode wires or cathode strips).

If significant aging effects are discovered during these tests, various mitigating actions can then be contemplated, such as lowering the gas gain by reducing the high voltage applied to the anode wires, increasing the gas flow rate, or changing the gas mixture.

We also plan small-scale tests of alternative gas mixtures that could reduce aging effects, and might allow the operation of the CSC without the anti-aging CF_4 gas component, whose usage may become restricted with the advent of new greenhouse gas environmental regulations. If suitable substitute gas mixtures can be identified, they will then be tested with full chamber irradiation as previously described.

4.2.4.2 CSC replacement of cathode front-end electronics

The cathode read-out front-end boards (CFEB) use analog charge storage within custom Switched Capacitor Array (SCA) chips (ASICs) that sample at 50 ns intervals and contain a depth of 96 cells. These boards are used on all CSCs except ME1/1 chambers which, in 2013-2014, have undergone an upgrade of the front- and back-end electronics to improve rate capability and performance at higher luminosity. With the increased L1 trigger latency (up to 12.5 μsec) and rate (up to 750 kHz) that is planned for Phase-II CMS operation, the 108 inner CSC chambers (ME2/1, ME3/1, and ME4/1) will often fill the SCA cells; consequently no additional data can be stored, causing large readout inefficiency. An illustration of the effect is shown in Figure 4.4, which shows 2% data-loss contours in the CFEB electronics due to this effect in the various types of chambers, as extrapolated from Run-I data. Data loss increases dramatically to the right and above the contours; while the CFEBs on all chambers can easily handle the original trigger specifications with latency about 3 μs and 100 kHz rate, it can be seen that with L1 latencies at 12.5 μsec and rates of up to 750 kHz, the ME2/1, ME3/1, and ME4/1 chambers will incur large data losses, while the 360 outer CSC chambers can easily sustain the HL-LHC conditions since for them the data rates are much lower.

The CFEBs can be replaced by boards similar to the "digital" DCFEBs that have already been installed on the ME1/1 chambers. These boards flash-digitize the data continuously and store them in large digital buffers, resulting in zero deadtime and the capability to handle latencies well beyond the current specification. They also send their output data on 3.2 Gb/s optical

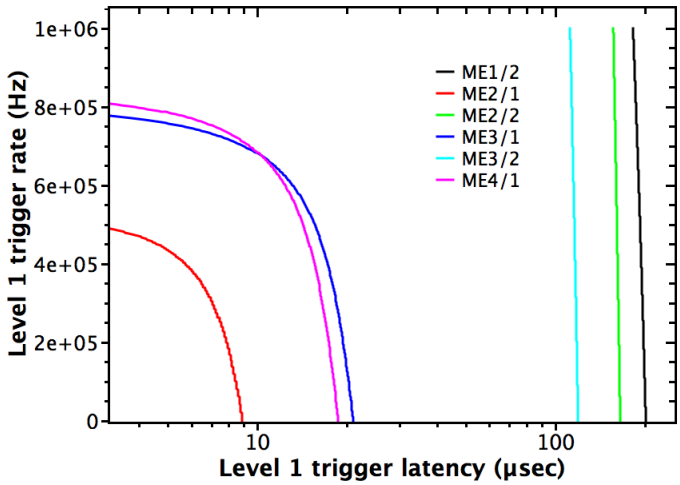


Figure 4.4: Contours of 2% data loss in the CFEB electronics caused by buffer saturation as a function of L1 latency and trigger rate (L1-Accept), for various chamber types. The contours are calculated by scaling from the occupancy of data seen in LHC Run-I.

links, as compared to 1 Gb/s achievable on the copper output cables of the CFEBs; the higher output data capability of the DCFEBs is similarly found to be important to handle the HL-LHC and CMS L1 trigger conditions.

Some additional elements of the CSC data acquisition system, namely the Data Acquisition Motherboards (DMBs) and Detector-Dependant Units (DDUs) that connect to the DCFEBs will need to be replaced as well in order to handle the high data throughput using the fast optical links.

There are important implications of this replacement for the planning of work in the CMS cavern during the long shutdowns. Based on experience with installation of the ME4/2 chambers and refurbishment of ME1/1 chambers during LS1, it is estimated that 4.25 months of access to the YE2 endcap iron disks (for ME2/1 and ME3/1), and 2.2 months of access to the YE3 iron disks for ME4/1 are needed. During LS3, the installation schedule is currently driven by the replacement of the Tracker and other elements of the CMS central detector such as the EB and DT FE electronics. During this work the endcap disks will be moved away from the central detector to the maximum extent, giving no access to the CSCs in stations 2-4. It appears therefore difficult to schedule two such long-lasting activities in the barrel and endcaps within the same shutdown period. Another issue is the availability of the two ASICs needed for the new boards: the "Buckeye" preamp/shaper and a comparator ASIC that is used to localize the muon position with a precision of half a strip for the trigger; 3240 of each are needed (not including spares). There are enough Buckeye spares from the original production to build all of the new boards, but only enough comparator ASIC spares to build 2/3 of the new boards. A new chip production campaign would imply high costs and long lead time. Therefore, early installation of at least some of the new boards during LS2, for example, for ME2/1 and ME3/1, would relieve the installation schedule for LS3, and would also allow recovery of ASICs from the existing CFEB boards in order to build the remainder of the new boards.

Additional testing will be needed to ensure that the CSC electronics can tolerate the increased radiation levels expected. Neutrons (and protons) with energy above 20 MeV are the major source of transient disruptions (SEU = Single Event Upset) or incremental damage to electronics

devices caused by hadronic interactions in the silicon. For the new inner chamber electronics our plan is to measure SEU rates in a high-intensity hadron beam once the final board firmware for Run-II is fully deployed.

Additionally, cumulative effects on electronics devices are permanent incremental damage attributed to total ionizing dose (TID), which degrades the performance of silicon circuits over time. The high-flux gamma irradiation at GIF^{++} will provide a way to study the effect of total ionization dose on CSC electronics and thus complement the earlier measurements made at nuclear reactors. The components of the boards used on the non-ME1/1 chambers were tested in the early 2000s at doses of $5-10$ krad [116], substantially lower than the 30 krad dose to which the components of the new ME1/1 electronics have been tested [117]. In all cases, the exposures were integrated in time ranging from a few minutes to a few hours. Tests of entire boards, for old and new type of electronics, will need to be repeated at GIF^{++}, with higher TID integrated over several months. This will be much closer to the dose ($\sim$ 8 krad) expected during actual operation, and will complement the reactor and cyclotron studies.

4.2.5 RPC detectors for Phase-II

In preparation for HL-LHC, the RPC activities are focused on:

- RPC detector longevity: from Run-I data and previous tests one does not expect any major degradation of detector performance, thus no interventions are foreseen on the chambers themselves. However, the aging of the RPCs up to ten years of HL-LHC environment has to be thoroughly tested and certified at the GIF^{++} facility.
- RPC electronics longevity: new TID and SEU studies will be performed.
- RPC gas mixture: investigation and qualification of a new gas mixture is needed to fulfill new environmental regulations. These tests will be done in conjunction with the aging tests.

4.2.5.1 RPC detector longevity

During Run-I, the RPC system provided stable operation and excellent performance and did not show any aging effects. In the past, extensive long-term tests were carried out at several gamma and neutron facilities certifying the detector performance up to values of dose, charge and fluence close to those expected after ten years of HL-LHC operation. Both full size and small prototype RPCs have been irradiated with photons up to an integrated charge of ~ 0.05 C/cm^2 and ~ 0.4 C/cm^2, respectively [107, 118]. Neutron fluence of $\sim 10^{12}$ n/cm^2 and an equivalent dose of about 45 Gy have been reached [119]. On the other hand, the use of two of the main RPC gas components, $C_2H_2F_4$ and SF_6, is progressively being banned by authorities, due to their large global warming potential (GWP). To certify the detector performance for long-term operation at the HL-LHC a new series of aging and gas studies are needed.

At GIF^{++} we plan to irradiate three spare endcap chambers and two barrel chambers. New mechanics will be designed in order to have fewer different types of chambers: the goal is to have not more than ten types of chambers. The latter will have a partially new design of the service box to separate and better handle gas, cooling and HV connectors. The on-chamber gas distribution, recently found to be the most critical problem in the barrel chambers, will be modified to assure better gas tightness. All these changes will not affect the aging test results. The goal of the GIF^{++} irradiation test is to integrate a 30-year HL-LHC equivalent charge in about one year. The following studies will be performed:

1. The electrode bakelite bulk resistivity will be regularly measured in order to verify its

stability throughout the irradiation period. The bulk resistivity has a great impact on the RPC performance, especially at very high particle rate.

2. RPC current and counting rate will be monitored with and without source. Any variation would represent a change in the detector gas gain or damage of the inner bakelite surface, thus a sign of aging.

3. Chamber rate capability will be periodically studied with cosmic ray muons and, when available, muon beam in the presence of different background rates (by using the moving-source attenuator filters available at GIF^{++}).

4.2.5.2 RPC electronics longevity

The RPC on-detector electronics consists of LV distribution boards, which distribute LV power and slow-control signals, and front-end boards (FEBs,) which amplify and discriminate the signals induced on the strips and transmit them to the RPC Link Board system [120] located in the CMS tower racks.

- During Run-I, very few failures or malfunctioning of FEBs (23 over 6016) have been reported. The FEBs had been previously tested up to a neutron fluence of about 10^{12} n/cm^2 and no damage was observed [121]. In order to qualify the FEB performance at the higher doses and fluence, a new campaign of neutron irradiation testing is planned. Since the FEB electronics is basically analog, SEUs would negligibly increase the spurious noise rate. In addition, by the end of Run-I, about 1% of RPC electronic channels were masked due to a failure in the distribution board caused by discharges in the chamber. A new generation of distribution boards, with stronger protection against discharge, has been already produced and 23 out of 360 barrel distribution boards have been replaced during LS1.
- The RPC Link Board system [120] consists of 1376 Link Boards and 216 Control Boards. Most of them, produced in 2006, have been continuously working since their installation in 2008. Infant mortality has been encountered only during the first year of data taking, mostly due to low-voltage instabilities or frequent system power cycles. The RE4 stations, installed in 2014, were equipped with an improved Link Board system which includes a set of on-board LV regulators. Being located in the CMS tower racks, the link system components are easily accessible and the available spare boards are sufficient to cover the needs up to the end of the HL-LHC running period.
- During Run-I, one SEU-like problem per week was observed at a luminosity of about 5×10^{33}cm^{-2}s^{-1}. If SEU-induced problems become too frequent, a solution would be to increase the rate of the firmware reloading to the flash memories. On the basis of irradiation tests, flash memories should survive to about four times the expected HL-LHC dose [122].

4.2.5.3 RPC gas mixture

The CMS RPC gas mixture is a three-component, non-flammable mixture of 95.2% $C_2H_2F_4$ (Freon R-134a, tetrafluoroethane), 4.5% C_4H_{10} (isobutane) and 0.3% SF_6 (sulfur hexafluoride), humidified at about 40%. The same gas mixture is also used by ATLAS for its RPC system. Recently, because of its ecological impact, the use of fluorine compounds is increasingly restricted by (inter)national regulations following the 1997 Kyoto Protocol, which are aimed at limiting the emission of certain greenhouse gases. R-134a, with its Global Warming Potential (GWP)

of 1430, is one of the hydrofluorocarbons appearing in the Kyoto Protocol gas list as well as SF_6. Possible alternatives are already under study by the industry such as the R-152a, with a GWP of 120, and the most promising ones HFO-1234ze and HFO-1234yf which are non-ozone depleting and with a very low GWP (6 and 4, respectively). However, until the use of these gases is more widespread, the main issue remains a cost $\sim$15 higher than that of R-134a.

The search for a new gas mixture is therefore mandatory for the RPC operation at HL-LHC. R&D tests with alternative gas mixtures have been recently started in two laboratories aimed at exploring a large number of eco-gas candidates. Two different and complementary approaches are being used to identify the best candidate and to cross-validate the results.

- The charge spectra and the intrinsic time resolution properties of new gas mixtures will be studied using a clone of the CMS RPC Gas Gain Monitor system [123] .
- The detector performance (efficiency curve/plateau, cluster size and multiplicity, time resolution) will be assessed with new gas mixtures by using standard CMS front-end electronics. For this R&D two small CMS-like detectors have been already constructed with RE4 spare materials and characterized with cosmic muons using the standard gas mixture.

Once the above R&D provides a good candidate for a new RPC gas mixture, the study will continue at GIF^{++} to include a long-term assessment of the radiation tolerance of the chambers operated with the new gas mixture. In high-radiation environments like those expected at the HL-LHC, several new chemically-reactive impurities could be created inside the RPC gas volume, mainly hydrocarbons, HF, F, Freon-type molecules and other chemical species. They are potentially harmful to the detector materials and the gas system and ultimately they may degrade the detector performance. The RPC current and the flourine concentration will be measured simultaneously as a function of time, source intensity and detector working point.

4.3 GEM detectors GE1/1 and GE2/1

4.3.1 GEM Detectors overview

The GEM detectors GE1/1 and GE2/1 are proposed to instrument the forward region of the CMS muon system with high resolution detectors, each "superchamber" consisting of a pair of triple-GEM (Gas Electron Multiplier) layers that complement the existing CSC (ME1/1 and ME2/1) detectors, providing redundancy, enhanced triggering and reconstruction capabilities.

In the past, GEMs have been demonstrated in several experiments (TOTEM [124], LHCb [125], COMPASS [126]) to be robust and reliable and to have high-rate capability and high longevity. In TOTEM and LHCb, the GEMs have been operating for the last three years, while in COMPASS, the GEMs have been operating for the last 12 years. In TOTEM, triple-GEMs are located at $5.3 < |\eta| < 6.5$ and operate at a hit rate of $\sim 12\,\mathrm{MHz/cm^2}$, with a detector surface $\sim 500\,\mathrm{cm^2}$ with no aging effects due to polymerization nor any change in material properties or performance. In LHCb, the rate is of the order of $1\,\mathrm{MHz/cm^2}$. In COMPASS, the GEMs have been operated with a gas mixture of Ar/CO_2 at a rate of $\sim 2.5\,\mathrm{MHz/cm^2}$ with no degradation of gain, efficiency, energy resolution or time resolution. Both COMPASS and LHCb GEMs have a total active area $\sim$ 7 to 8 $\mathrm{m^2}$.

4.3.2 Construction of GEM detectors

The GEM detector is a micro-pattern gas amplification detector based on a polyimide (Apical) foil located between two conductive layers. The foil is perforated with biconical holes in a

hexagonal pattern and it acts as signal amplifier. With a triple-GEM detector, the gas gain in each GEM foil is typically about 25, thus providing a total gain of 15000, with an operational voltage far away from any critical value. The principle of operation of a GEM and a depiction of the operation of a triple-GEM are shown in Figure 4.5.

Within the CMS GEM R&D, cost-effective production of large areas has been demonstrated and results from beam tests have demonstrated a spatial resolution as low as 100 μm, and a time resolution of 5-8 ns. Efficiencies above 98% have been measured, with rate capabilities of 1 MHz/cm^2. The gas mixtures employed are non-flammable Ar-CO_2 and Ar-CO_2-CF_4. Thus, GEMs have shown themselves to be nearly ideal matches to the requirements of the forward muon system upgrade, as described in further detail below.

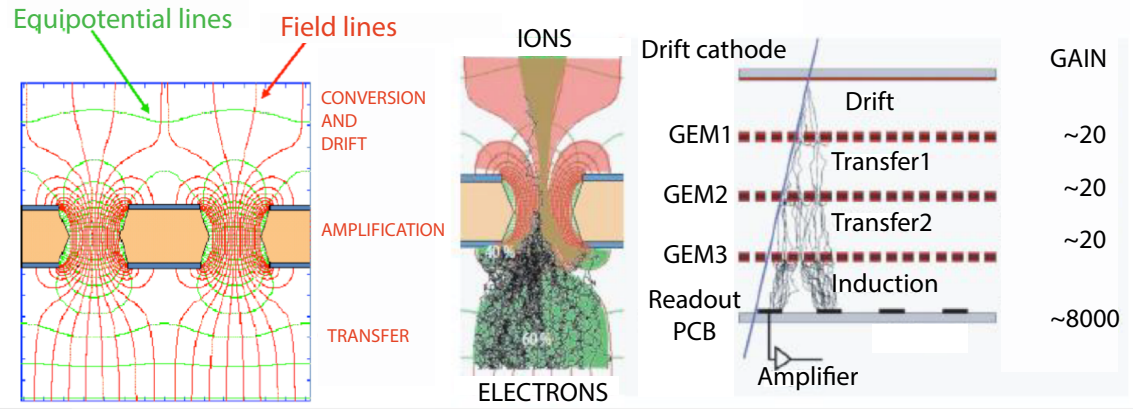


Figure 4.5: Left) Electric fields and equipotentials across the GEM foil. Within the holes the amplifying field ranges from $50 - 80$ kV/cm. Center) A simulation of electron multiplication in the vicinity of the hole in a GEM foil. Right) Arrangement of the triple-GEM detector with three foils, a drift electrode on the top and a readout electrode at the bottom defining drift and induction fields.

Currently used dimensions are: foil thickness of 50 μm, copper thickness of 5 μm, outer hole diameter of 70 μm and hole pitch of 140 μm. The drift gap is 3 mm, while transfer gaps and the induction/collection gap are 2 mm. The typical voltage across each GEM foil is 380 V, with a total across the stack of 4000 V. In the last four years, in the context of the CMS GEM R&D effort (CMS RD10.02), production of large-area GEMs with the single-mask technique has been demonstrated [127].

A novel spacerless and stretching technique has also been developed that allows production of large-area triple-GEM detectors without any gluing: the three GEM foils are mounted into a single stack where they are kept together with a segmented frame, also providing the inter-GEM spacing. Inside the stack frame several nuts are embedded to carry out the foil tensioning. Once the stack is assembled, it is placed on the drift plane over which it is free to slide. A frame is fixed on the drift plane that will close the gas gap and allow simultaneous tensioning of all three foils by means of stretching screws and a dynamometric screwdriver applying a known torque ($\sim$ 1 N/cm). Once the GEM stack has been fully tensioned, the chamber is closed by placing the readout board on top and fixing it with screws. The gas volume is sealed by two O-rings housed in the external frame. This construction technique allows assembly of one full-scale GE1/1 chamber in about two hours, compared to one week for a detector assembled with the conventional gluing technique. Furthermore using mechanically stretched and spacerless technology allows easy reopening of a chamber if necessary.

A total of 10 GE1/1 prototypes have been built at various production sites (CERN, FIT, Gent,

Frascati, Bari, India) during 2010-2014, with steady improvements of the assembly procedure. These prototypes, mostly of length 99 cm and top and bottom widths of 45 cm and 22 cm respectively, have been fully tested on x-ray sources and test beams. A video of the assembly of a GE1/1 prototype according to this technique can be found in [128].

4.3.3 The GE1/1 and GE2/1 detectors

The GE1/1 and GE2/1 detectors are "superchambers", each made of a double layer of trapezoidal triple-GEM chambers. They cover slightly more than 10° in GE1/1 and 20° in GE2/1, overlapping in ϕ just like the corresponding CSC chambers in stations ME1/1 and ME2/1. For both endcaps, 72 superchambers will be needed for GE1/1 and 36 for GE2/1. The two readout planes of the superchambers are separated by 20 mm in z, and the total thickness is 88 mm.

The GE1/1 superchambers alternate in ϕ between long and short versions: the long ones cover an η range of $1.5 - 2.18$ while the short ones cover an η range of $1.6 - 2.18$. The GE2/1 chambers cover an η range of $1.65 - 2.4$. These detectors are about 1.2 m long and 0.8 m wide on the outer side of the trapezoid.

The chambers are segmented in eight rings in r and three (GE1/1) or six (GE2/1) sectors in ϕ; each ϕ sector is in turn segmented into 128 radial strips read out by a single front-end ASIC (the VFAT3 discussed in Sect. 4.3.6). The strip pitch varies from about 0.5 mm to 1.2 mm (w), subtending an angle of 0.45 mrad between strips. The expected spatial resolution with binary readout is $\sim w/\sqrt{12}$, varying from 140 to 350 μm from inner to outer radius. The L1 trigger combines two (GE1/1) or four (GE2/1) strips adjacent in ϕ into a single trigger strip; thus the trigger segmentation is 0.9 mrad (GE1/1) or 1.8 mrad (GE2/1).

In GE1/1, the superchambers are to be installed in a 100 mm gap between a back flange within the YE1 nose and the CSC ME1/1 chambers. The location of these chambers on the "nose" of the first endcap disk of CMS is shown in Figure 4.6 (left), and a closeup of the GE1/1 short and long detectors within the nose is presented in Figure 4.6 (right).

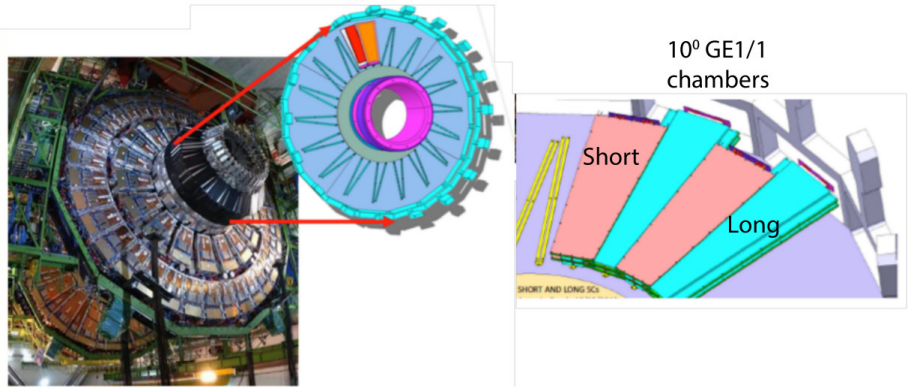


Figure 4.6: Left: location of the GE1/1 detectors on the YE1 nose. Right: detailed view of the overlapping of GE1/1 superchambers.

The state of development of the detectors is such that the GE1/1 station and the associated electronics could be installed in LS2: dummy detectors have been produced to optimize cable and fiber routing, and layout of services, and installation trials have been performed. Installation of a "slice test" consisting of four GE1/1 superchambers is foreseen during the LHC technical stop at the end of 2016.

A prototype detector for GE2/1 is under design and it is foreseen to be constructed by the end of 2014. Services, routing and gas system are also under study.

4.3.4 GEM detector alignment

To ensure the required submillimeter spatial resolution of the GEM chambers, a position monitoring system based on capacitive sensors is planned to be built in the GE1/1 region as early as Run-II. Six sensors (with micrometer resolution), mounted on the long chambers will measure the relative R and ϕ distances between the superchambers. These measurements will allow the geometric reconstruction of the full GE1/1 ring.

The location of the GE1/1 ring in the CMS coordinate system will be performed by track-based alignment methods as a baseline solution. However, a hardware connection to other muon components, in particular to ME1/1, GE2/1 (and eventually ME0) are also considered to validate the track alignment, particularly for the degrees of freedom (weak modes) not accessible in the procedure. We consider the integration of the GEM systems into the existing hardware muon alignment system, extending the current lines of sight from the Modules for the Alignment of the Barrel (MAB) structures to the endcap nose region.

We are also considering equipping the GE2/1 and ME0 chambers with position monitoring devices, similarly to GE1/1, based on the experience gained during Run-II.

4.3.5 GEM test results

4.3.5.1 Performance

Several prototypes have been assembled and tested in the lab with x-rays and with test beams [129] over the last four years. Performance results shown below, unless specifically indicated otherwise, are obtained with a standard triple-GEM detector built with the single-mask technique, using gas mixture of $Ar/CO_2/CF_4$ (45/15/40%), and running at a gas gain of 10^4 (achieved at a total applied voltage of 3.8 kV between the drift cathode and the readout PCB). Figure 4.7 shows results obtained with a muon test beam: (left) shows that the gain rises exponentially with the total applied voltage, and Figure 4.7 (right) shows that the efficiency plateaus well above 95% as the gain exceeds 8000. The gain increases exponentially with HV over two orders of magnitude, and high efficiency is easily reached. Figure 4.8 shows (left) spatial resolution of 270 μm with binary readout in a part of a GE1/1 prototype section that has a strip pitch of 0.9 mm, and (right) time resolution of 5 and 8 ns at plateau for gas mixtures with and without CF_4, respectively. Tests of gain uniformity and rate capability over the surface of a full-scale GE1/1 prototype show a gain uniformity within a few percent, at most a few percent variation with X-ray induced hit rate up to 1 MHz/cm^2, and about a 10-15% gain drop by 2.5 MHz/cm^2 [130, 131].

4.3.5.2 GEM long-term operation and aging studies

New detectors in the CMS muon system are foreseen to operate for 10 years with sufficient longevity margin. While the electron avalanches in GEM detectors occur within the GEM holes, the readout electrodes are well-separated from the detector amplification. This ensures sustaining long-term operation. Including a safety factor of 2, the largest integrated current expected for GE1/1 detectors is of the order of 100 mC/cm^2 for GE1/1 and a few C/cm^2 for the inner radius of the ME0 chambers.

Irradiation tests at the GIF facility at CERN using a Cs^{137} gamma source are ongoing with a GE1/1 prototype detector running at a gas gain of 10^4 with $Ar/CO_2/CF_4$ (45/15/40%). We intend to continue tests to optimize the gas mixture and to study possible CF_4-free gas mixtures at the new Gamma Irradiation Facility (GIF++). The hit rate at the test facility of 1 kHz/cm^2 is similar to that expected at the HL-LHC for GE1/1 and GE2/1, and the accumulated charge

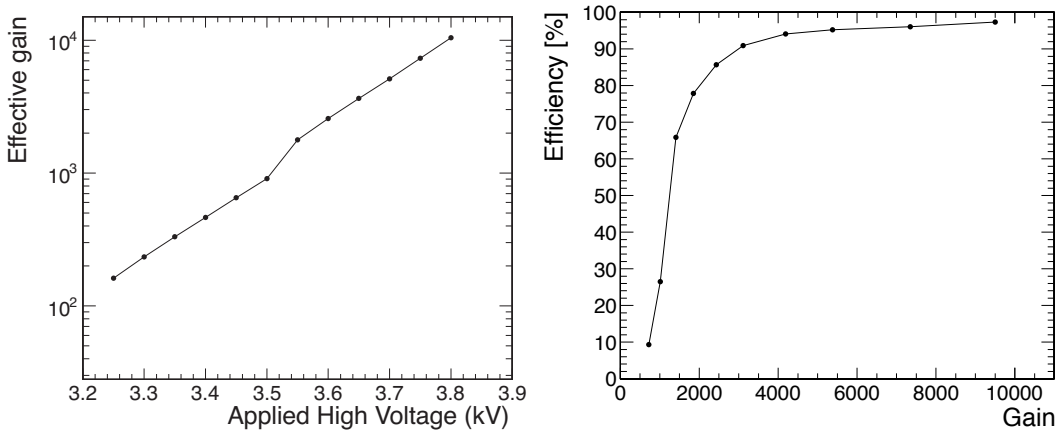


Figure 4.7: Muon test beam measurements. Left: Triple-GEM gas gain versus the total applied HV. Right: efficiency for muon detection versus gas gain.

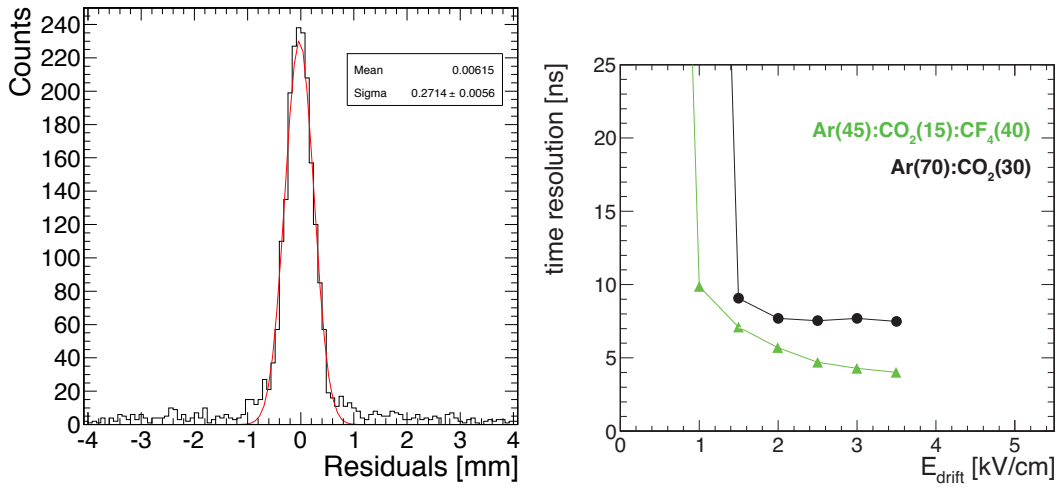


Figure 4.8: Triple-GEM resolutions obtained in a GE1/1 prototype using binary readout. Left: residual distribution with a fitted $\sigma = 270$ μm. (right) Timing resolution as a function of drift field for two representative gas mixtures, Ar/CO_2 with (5 ns) and without (7-8 ns) the addition of CF_4.

thus far is 9 mC/cm^2. No aging effects have been observed to date, and the tests will continue with higher doses at GIF++. Results from previous tests done by LHCb [125] with a 5.9 kV X-ray tube have shown stable operation with no gain drop even at hit rates up to $6 \cdot 10^7$ Hz/cm^2 and an accumulated charge of 4 C/cm^2 using an $Ar/CO_2/CF_4$ 45/15/40% gas mixture.

Neutron backgrounds of the order of 10^{14}/year $\cdot$ cm^2 are expected and several such irradiation tests have been performed in the framework of the RD51 collaboration. Detailed discharge studies have been done with GEM detectors exposed to neutrons and alpha particles, showing no degradation [132]. Nevertheless further detailed cyclotron tests are planned.

4.3.6 Electronics system overview

A block diagram of the main system components in the GEM signal/control electronics is shown in Figure 4.9.

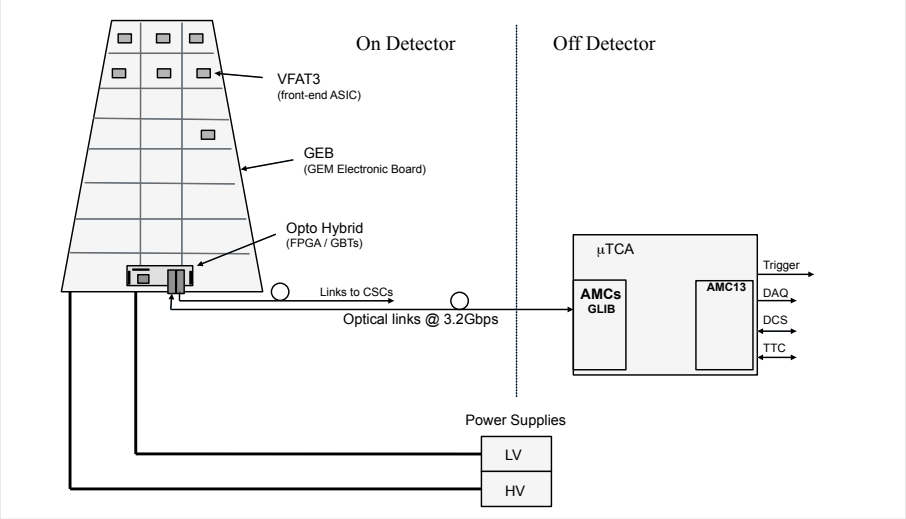


Figure 4.9: A diagram of the GEM electronics readout system for CMS.

The diagram illustrates the main components for the readout of a single GEM chamber and is divided into two main regions, namely on-detector and off-detector. Visible in the on-detector part is the segmentation of the GEM chamber. The 128 strips from each sector are connected to the inputs of the front-end ASIC (VFAT3) via a connector on the GEM readout PCB. The VFAT3 itself is mounted on a hybrid which plugs into the GEM readout PCB connector. The control, readout and power to/from the VFAT3 hybrid is delivered via electrical signals (E-links) running through a large flat PCB known as the GEM Electronics Board (GEB). An Opto-hybrid board also plugs into the GEB which contains the GigaBit Transceiver (GBT) chip set, an FPGA and the optical receivers and transmitters to provide the link to the off-detector electronics region.

There are two optical paths to the Opto-hybrid. The first is bidirectional and runs between the μTCA crates located in the counting room and the Opto-hybrid. This path is used for sending set-up and control signals to the front-end chips. The return path is used for VFAT3 tracking data packets and return slow-control data. The second path is unidirectional and takes the VFAT3 fixed latency trigger data from the GEM system to the CSC system.

4.3.6.1 The VFAT3 front-end ASIC

The VFAT3 is the next iteration front-end ASIC for GEM triggering and readout, currently under development following the previous development of VFAT2 [133]. The basic features of

the VFAT3 are summarized here:

- 128 channels:
- positive and negative polarity for charge collection;
- trigger path with granularity of two channels;
- full data packets readout at 1 MHz;
- latency up to 25 μs;
- time resolution of less than 7.5 ns (with detector);
- integrated calibration and monitoring functions;
- interface to and from the GBT at 320 Mbps;
- Robustness to ionizing radiation and against single-event upsets (SEU)

The block diagram for the VFAT3 ASIC functions is shown in Figure 4.10.

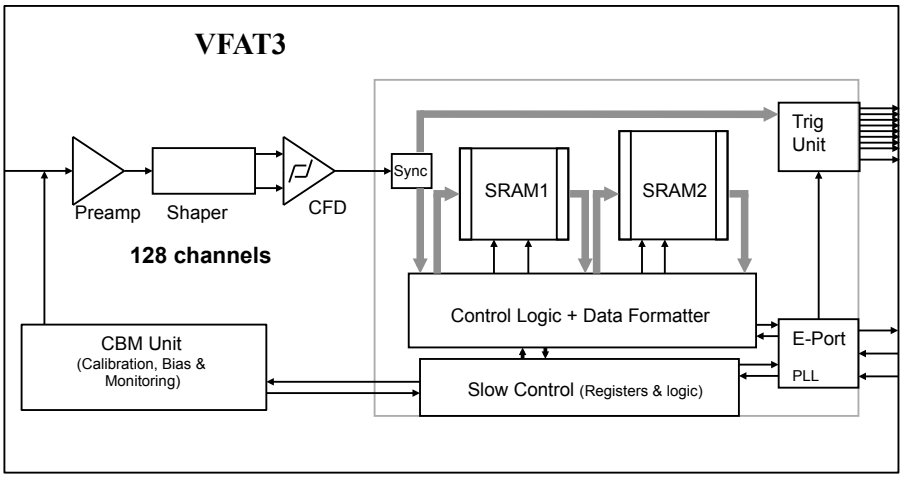


Figure 4.10: VFAT3 block diagram.

Each VFAT3 channel contains a charge sensitive preamplifier and shaper, followed by a constant-fraction discriminator. Following the discriminator is a Sync unit which synchronizes the comparator result with the 40 MHz clock. The data then split into two paths, one with a fixed latency for trigger signals, and the second with variable latency for tracking data. All communication with the VFAT3 occurs through the E-port. This includes slow-control commands and response as well as fast trigger commands, clock and calibration signals. The chip is highly programmable to offer maximum flexibility. Programmable options include front-end features such as gain and shaping times as well as options related to data packet content and size to optimize bandwidth usage at high rates.

4.3.7 GEM off-detector electronics

The off-detector electronics provide the interfaces from the detector and its VFAT3 front-end electronics to the CMS DAQ, TTC and Trigger systems. The design foreseen for the CMS GEM off-detector electronics is based on FPGAs and Multi-GBit/s links that adhere to the μTCA standard. The off-detector system is based on the preferred CMS μTCA crate, the VadaTech VT892, which supports twelve AMC (Advanced Mezzanine Card) boards and two μTCA Carrier Hub (MCH) slots. The MCH1 slot houses a commercial MCH module, used for gigabit Ethernet communication and IPMI control. The MCH2 slot houses the AMC13, a custom AMC developed by Boston University to interface the μTCA crates to the CMS data acquisition system and to provide the TTC signals downlink.

The AMC cards that equip the μTCA crates will most probably be the MP7 [96] (Master Processor) board or a future evolution. The current MP7 is based on the Xilinx Virtex-7 FPGA and Avago MiniPOD optical modules, and can provide 144 serial optical links capable of operating above 10 Gb/s. In total, 16 MP7 boards are needed to read-out the full GE1/1 and GE2/1 system.

While the trigger and tracking data are sent from the detectors to the μTCA electronics, the trigger data are also sent directly to the CSC Trigger Mother Board (TMB) located in the CMS peripheral crates. There the GEM trigger data are combined with the CSC data to improve the Level-1 trigger efficiency of the CSC system at the earliest stage of the CSC trigger processing. This trigger data path to the CSC TMB will use optical fibers running at 3.2 Gb/s, which are already installed in the case of GE1/1, located along the CSC detectors inside CMS.

4.4 Forward RPC detectors RE3/1 and RE4/1

4.4.1 Forward RPC detector overview

This section discusses the instrumentation of the Muon System in the RE3/1 and RE4/1 regions with an improved version of RPCs, referred to as iRPCs, capable of handling the particle rates foreseen during the HL-LHC phase. The new chambers would complement the existing ME3/1 and ME4/1 stations, currently instrumented with CSCs only, providing improvements to the L1 muon trigger. Moreover, in one of the options under study, these chambers could provide an improved time resolution down to better than 100 ps, which may be exploited, for instance, for pileup mitigation.

The proposed RE3/1 and RE4/1 stations, schematically shown in Figure 4.1, will cover the region $1.8 < |\eta| < 2.4$. Each of the 72 chambers, foreseen for both endcaps, will span a 20 degree sector. Simulation studies to determine the optimal readout segmentation of these chambers are still ongoing; as a starting point a geometry is considered with five η partitions, each equipped with 192 strips, with a pitch ranging from 0.30 to 0.62 cm (to be compared with 1.30 and 3.93 cm respectively in the present endcap RPCs). The spatial resolution will be improved proportionally to the strip width, i.e. by a factor 4-6 compared to the present system.

4.4.2 Forward RPC detector requirements

Improvements in rate capability with respect to the present RPC detectors are needed. The first results provided by FLUKA simulations, and the comparison with data from the CSCs, provide an estimate of the rate around 700 Hz/cm^2 in the hottest points of RE3/1 and RE4/1 (see Figure 4.11 and [134]). Taking into account an additional safety factor, a rate capability of around 2 kHz/cm^2 should be required for the new chambers, while standard CMS double-gap RPCs have been tested to reach around 1 kHz/cm^2 (see, for instance, [107] and [106]).

The rate capability of RPCs (see [135]) can be improved in various ways (all of which are or will be investigated in the framework of the muon upgrade *R&D* activities, as will be shown in the following section):

1. Reducing the electrode resistivity ρ, which has the direct effect of reducing the recovery time needed for the electrodes to be charged up again after a discharge in the gas gap; this increases the rate capability proportionally.

2. Reducing the average charge generated in the avalanches, and transferring part of the signal amplification from the gas to the front-end electronics. Reducing the overall charge

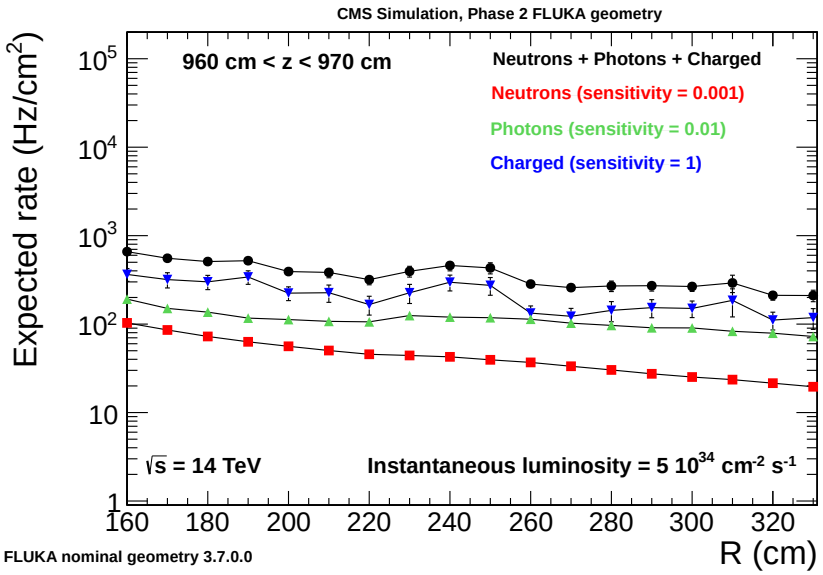


Figure 4.11: Simulated particle hit rate as a function of radius at the RE3/1 station, assuming a 0.001 sensitivity to neutrons, 0.01 to photons and 1 to ionizing particles.

path in the gap results in a reduced voltage drop on the electrode plates, and a reduced period of inefficiency; it also reduces aging processes.

3. Changing the detector configuration, which includes many possible options, like changing the electrode thickness or their number (multigap). For a given induced signal (i.e. detector efficiency), these operations increase the ratio of the induced signal to the moving charge in the gap, with beneficial effects similar to those previously described.

In addition, aging effects during the full HL-LHC program need to be accurately considered, as discussed in Section 4.2.2 for existing detectors. For instance, at a rate of 0.5 kHz/cm^2 and an average charge per avalanche of 20 pC, the integrated charge in the detector will reach about 0.6 C/cm^2, a significantly higher value than tested so far. (see Table 4.1). Moreover, an adequate safety factor of three will be considered.

4.4.3 Forward RPC technologies under study

To satisfy the requirements mentioned in the previous section, all major factors in potential improvements are considered: electrode resistivity, higher-amplification electronics, and detector configuration. The final technique adopted may combine the benefits of the different approaches.

The resistivity of the high-pressure laminates (HPL) made with Bakelite resin used in the present CMS RPC detectors is $1 - 5 \times 10^{10}$ Ω cm. Recent production for RE4 chambers had resistivities near the low end of this range. By adjusting the proportions of phenol-impregnated paper and melamine resin in the production process, a range of $0.5 - 1 \times 10^{10}\Omega$cm might be achievable. Prototype detectors made using this lower-resistivity HPL will soon be produced and tested.

Low-resistivity electrodes can also be produced using semiconductive silicate glass (see [136]) that can now reach a resistivity around $10^{10}\Omega$cm, which is significantly lower than for standard borosilicate (float) glass (which is around $10^{13--14}\Omega$cm) and essentially comparable to

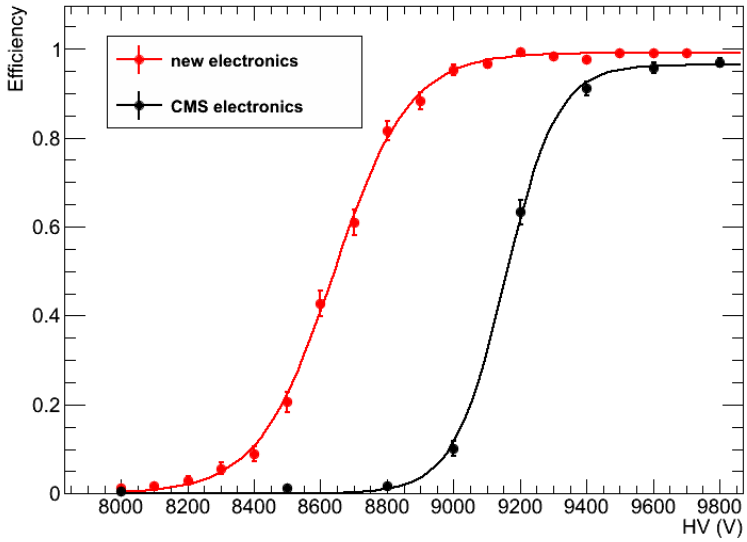


Figure 4.12: Efficiency plateaus for the cases of the standard CMS RPC (black points) and new ATLAS prototype (red points) electronics. The shift in the beginning of the plateau toward the left indicates that in this case a lower charge is crossing the RPC gap. This can help to improve the rate capability of the RPC.

the value for HPL. Compared to HPL, glass electrodes can be produced with smoother surfaces, avoiding the oiling process during the gas-gap production. They also do not require humidification of the gas mixture during chamber operation. However, while HPL sheets can be produced in large sizes, this new glass is now only produced in small (e.g. 30×30 cm^2) pieces. The glass pieces must be glued together or combined mechanically to cover the typical areas needed for a RE3/1 or RE4/1 chamber. Other, new types of promising low-resistivity materials are currently being developed, such as vanadium-based glass with tunable resistivity, Si_3N_4/SiC ceramics [137], and glass or poly-ether-ether-ketone (PEEK) doped with carbon nanotubes (CNT), could also be considered for further R&D.

One version of higher-amplification electronics has been developed in the framework of the muon system upgrade of the ATLAS experiment [138]. Tests comparing the performance of CMS chambers equipped with the standard CMS electronics and an early prototype of this new electronics have been performed with cosmic rays; a shift of about 460 V in the efficiency curves is obtained (see Figure 4.12); the average charge corresponding to these voltages for 90% efficiency is reduced from 20 pC to about 3 pC, a factor of 7 lower. This is a strong hint at an improved rate capability that will be tested in realistic conditions at the upcoming GIF++ facility.

The changes to the RPC detector configuration that are considered to enhance the rate capability are thinner electrodes (from 2 mm at present to 1.5 or even 1 mm) with double-gap HPL; and multi-gap configurations using HPL and glass with thin gaps and electrodes. Thinner electrodes increase the signal induced on the readout strips and, subsequently, improve the rate capability. Multi-gaps divide the charge necessary to induce a signal higher than the electronics threshold in more than one gap, and enhance the rate capability as well.

The principle of a multi-gap HPL is shown in Figure 4.13 (left). The standard double-gap configuration used in CMS has been modified by adding an additional gap on both sides of

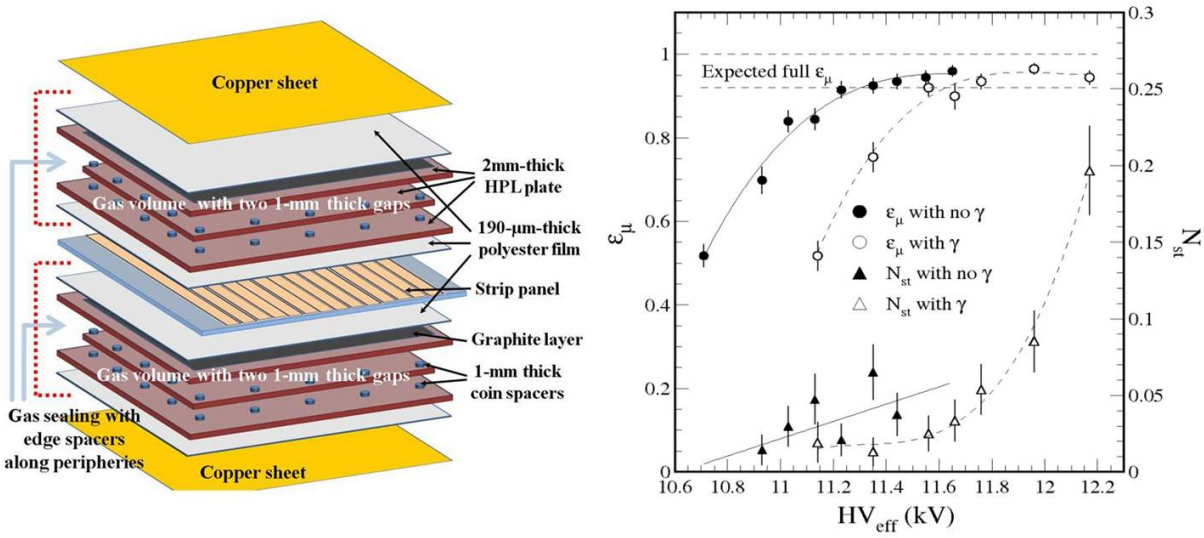


Figure 4.13: Left: view of a multigap HPL RPC prototype. Right: efficiency plateau (circles) and streamer probability (triangles) vs. operating voltage with (open symbols) and without (solid symbols) a 3 kHz/cm^2 γ irradiation.

the readout strips. The thickness of each of the four gaps is 1 mm. In this particular case, the material used for the eletrodes is identical to the one already used for CMS/RPC chambers (same resistivity), and the readout electronics as well. Efficiency for cosmic rays muon vs. operating voltage in the two cases of no irradiation and irradiation with a 150 mCi ^{137}Cs γ-ray source providing a hit rate around 3 kHz/cm^2 on the whole chamber is shown in Figure 4.13 (right). The chamber reaches full efficiency and has a plateau several hundreds volts wide in both cases where the probability of a streamer is acceptably low ($<$ 10%), satisfying the requirement on rate capability. Several other results are described in detail in [139].

Single and Multi-gap glass RPCs (GRPCs) are described in [136]. Tests performed with a localized beam irradiating small areas of the chambers have shown promising results, with a rate capability exceeding 10 kHz/cm^2 and a time resolution better than 100 ps for a multi-gap configuration. This time resolution was obtained reducing the gap thickness from 400 microns to 200 microns and using four gaps. Tests to measure the performance in conditions of uniform irradiation are foreseen at GIF++ in the next year. Moreover, simulation studies to quantify the benefits that detectors with an improved time resolution in RE3/1 and RE4/1 would bring to the CMS potential for physics are being actively pursued.

4.4.4 Forward RPC on-chamber and off-chamber electronics

Benefiting from synergies across different experiments and groups, the amount of R&D needed for iRPC on-detector electronics should be limited. Actually either the present CMS/RPC electronics, or electronics already developed for other applications could be used.

Some of the detector protoypes described in the previous section have already been tested with the CMS/RPC electronics, which is described in details in [140]; this choice would have the added value not to require any modifications to the off-detector electronics, since the integration with the rest of the CMS DAQ system would take place *via* the link boards system already in use.

As already pointed out, significant improvement could also come from the electronics de-

scribed in [138]. Essentially, this is characterized by a higher amplification and lower input threshold, but requires improved shielding, grounding and HV filters to operate effectively. It should be interfaced with the CMS Link Board system [120] through a custom card, which is currently under development.

Finally, if multi-gap, impreoved-time-resolution RPCs are used, then the front-end electronics originally designed for time-of-flight measurements with silicon photomultipliers (SiPMs) could be exploited [141]. This electronics is based on the PETIROC 16-channel ASIC and has already been used for testing glass RPCs. The interface of this electronics to the CMS DAQ chain is under investigation.

4.4.5 Forward RPC R&D

A final choice of the technical solution will be taken after the tests foreseen at the GIF++ in 2015. All different prototypes previously described will be tested together in the same setup. Primary goals of the tests will be to assess the rate capability of the chambers and their stability after integrating a charge corresponding to 3000 fb^{-1}, i.e. around 1-1.5 C/cm^2. The margin of safety obtained for the different designs will be part of the technical choice criteria.

Specific aging tests on prototypes built with materials or electrodes not already used in the past, such as glass, are of particular importance to ensure the radiation hardness of the new chambers and to detect any possible aging of the materials in contact with the gas mixture. Similar procedures and infrastructures as described in Section 4.2.2 will be used for these tests. The parameters to be measured and monitored will include the electrode (HPL or glass) resistivity, the current drawn, the chamber efficiency and time resolution at low and high particle rate, and the sensitivity to neutrons and photons. Special attention will be given to the production of HF in the gas mixture as a function of rate, since it is a primary source of detector aging.

4.5 Very forward muon GEM detector ME0

The ME0 station will be installed in a space of ~ 30 cm freed behind the new endcap calorimeters as shown in Figure 4.14. The coverage is foreseen to extend up to $\eta = 3$, or more, depending on the boundary defined by these calorimeters. The configuration described here extends up to a value of 3, taking into account the shielding that will be required at the inner radius. The lower η boundary of ME0 is affected by substantial mechanical infrastructure such as bolts and plates that must be located near the periphery of the endcap "nose" to support the several hundred tonnes of calorimeter cantilevered off of the YE1 disk. A value of 2.0 is presently considered. The detector is expected to supply good muon ID in offline analysis. As demonstrated in Section 4.3, GEM detectors are able to handle the $\sim 10 - 100$ kHz/cm^2 rates and the integrated doses of 5 C/cm^2 expected in this region during Phase-II. A six-layer device is foreseen to allow proper rejection of neutron background (see Section 4.7.4). The design of ME0 incorporates neutron and photon shielding using, for example, borated polyethylene and lead. The detector will be made of units of 20° wedges with inner and outer radii of 300 and 1495 mm.

Triple-GEM detectors with their demonstrated performance, as described in Section 4.3, are expected to fulfill the rate and longevity requirements in this region. This will be validate by extending the tests ongoing at the GIF, GIF++ and neutron facilities for the GEM chambers foreseen in GE1/1 and GE2/2.

The scheme of the six detectors in a 20° wedge is shown in Figure 4.15. Each wedge will also contain a sliding rail for assembly on the support structure. The services will run along the back of the wedge onto the service ducts.

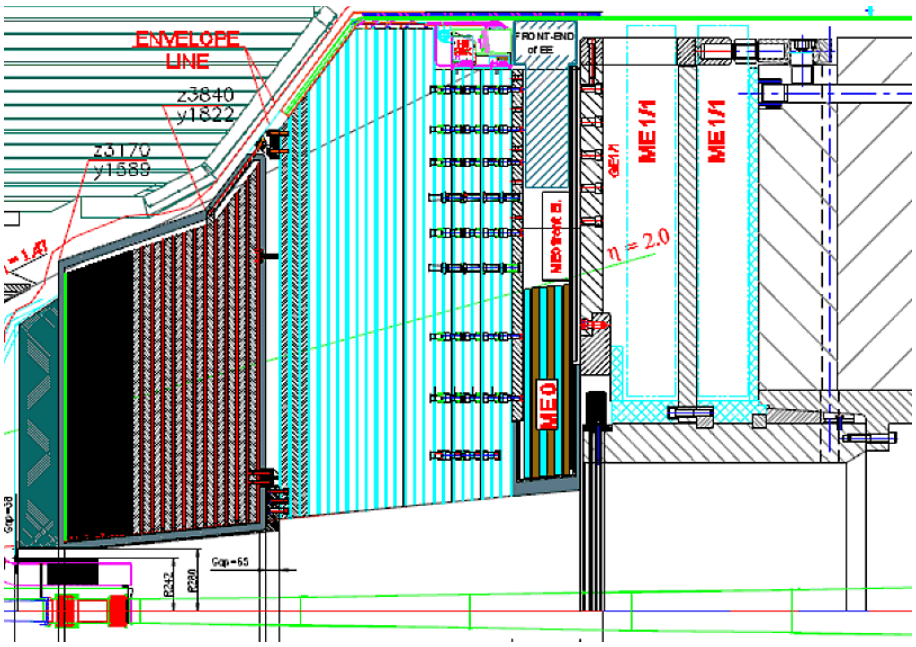


Figure 4.14: A cross-section of the endcap region including ME0 placed behind the endcap calorimeters at $523 < z < 554$ cm and extending from $2.0 < \eta < 3.0$. Space behind the calorimeter at larger radii will be used to accommodate mechanical elements and endcap calorimeter services.

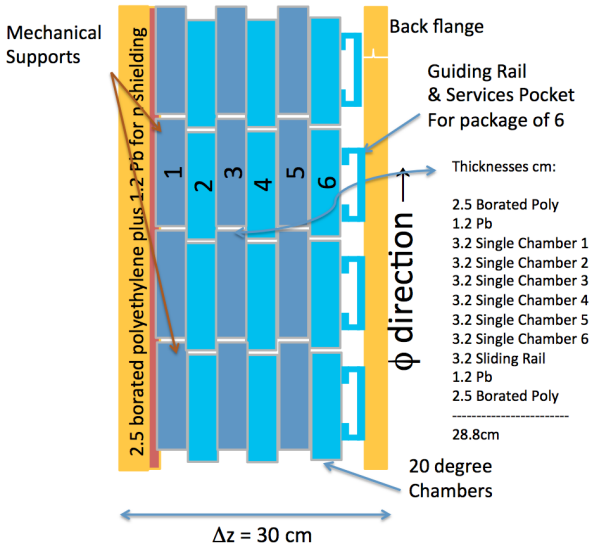


Figure 4.15: A preliminary layout of approximate 100° of a 6-layer ME0 layout in ϕ-z view, using 20° triple-GEM chambers in the 30 cm space made available behind the new Endcap Calorimeters.

The electronics for this station will be based on the VFAT3 architecture described in Section 4.3.6. As in GE1/1, the nominal segmentation of each ME0 chamber layer is 8 rings in η and 3 sectors in ϕ; each ϕ sector is in turn segmented into 128 radial strips read out by the VFAT3 ASIC chip. The strip pitch varies from about 0.27 mm to 1.34 mm and the angle subtended between strips is 0.9 mrad. Each layer will contain 3072 readout channels. The LV power cable package contains 6 cables for the VFATs, each carrying 6 A at 3.3 V; and 6 cables for the optical hybrids, each carrying 7 A at 3.3 V. The total LV power per endcap will be 4.7 kW.

4.6 Infrastructure requirements

The mechanical and installation aspects, as well as the required gas, cooling, LV, HV systems and the cabling work required for the Phase-II muon upgrades are described as follows:

For the additional forward detectors GE1/1, GE2/1, RE3/1, and RE4/1, space envelopes were already defined in the original detector design. In general, these detectors are allowed a maximum thickness of 100 mm, which generally limits their design to two detection layers, and a staggered overlapping geometry. For ME0 the allocated space will depend on the new design of the forward calorimeter region. For the time being, we assume a space envelope along z of 300 mm, which is compatible with a six-layer ME0 design.

In the GE1/1 area, which is especially congested due to the endcap calorimetry and the ME1/1 services, the power cables, gas and cooling pipes were already installed for potential RE1/1 chambers and thus can be used for the GEM chambers. Mock installation trials performed in May 2014 have shown that completion of the installation (addition of optical readout and trigger fibers, data and control cables) should not pose any particular problems. Power and infrastructure services foreseen for RE1/1 are perfectly adequate for the GE1/1 requirements.

For the other proposed forward muon detectors, all the services will need to be integrated with the present systems and installed, i.e. gas and cooling water pipes, LV and HV cables, and optical fibers. This will probably require temporary removal of some of the ring 2 and 3 RPC chambers for proper routing of cables and services from the disk periphery to the detectors.

For the muon rapidity extension, the ME0 detector installation will be part of the process of the complete rebuild of the forward calorimeter on the YE1 endcap disk, and services will need to be provided in careful coordination with the planning for the calorimeter; this planning has already begun.

The power/cooling requirements of the new muon detectors are approximately 120 W for each 10° two-layer GE1/1 super-chamber, 240 W for each 20^o GE2/1 super-chamber, 260 W for each 10° 6-layer ME0 unit, and 20 W for each RPC chamber. Overall, the heat dissipated on each YE1 disk by the additional GEM detectors will not exceed 20 kW, which is well within the specifications of the present cooling system. As originally planned for the RE1/1 detectors, the cooling circuits which are now feeding the HE Readout Boxes (RBX) will as well supply cooling water to the GE1/1 chambers. The $\sim$ 500 W cooling capacity available on each of the 18 circuits is adequate for the 240 W plus the $\sim$ 100 W power dissipated by 2 GE1/1 superchambers and one RBX, connected in series, even taking into account an upgrade of the HE RBXs. Cooling for the GE2/1, RE3/1 and RE4/1 is rather straightforward, with connections directly off the existing endcap cooling manifolds which provide enough flow and cooling capacity to take up the extra dissipated heat. Since the water flow provided by the present cooling system seems adequate, no additional pump needs to be added to the circuit, and probably only a larger system heat-exchanger will be required.

For the electronics upgrade of the DT system, the new power requirements will be reduced by about a factor of 2 in comparison with the present electronics, and the current gas and cooling infrastructure is considered adequate.

For the gas system new mixer racks located in the surface gas room must be foreseen for the new muon detectors (GEM and RPC), depending on the employed gas mixture. A new recirculation pump, gas mixer and purifier for GE1/1 is already planned and will be constructed by 2016. It will supply the 360 l GE1/1 volume at a gas flow of 180 l/h. There is already a plan to install the $\sim$ 230 m transfer lines from the surface gas hut to the underground service cavern (USC). A sufficient number of spare channels is available in the existing RPC distribution racks located in the experimental cavern (UXC) to supply gas to the independent GEM/RPC detector channels, although some modifications will be needed. New distribution racks are planned for ME0 and RE4/1.

Off-chamber, some amount of rack space on the endcap towers will be needed for LV power supplies, RPC Link boards and optical patch panels. Additional HV cables will need to be installed between the HV power supply located in USC and the detectors if the present RPC HV system architecture is maintained. In the service cavern a few μTCA crates will suffice for the GEM data acquisition electronics and a few rack computers will suffice for the slow controls of the new detectors.

Overall, the infrastructure requirements of the proposed muon upgrades are modest and can easily be accommodated.

4.7 Performance of the Upgraded Muon System

In this section we discuss the limitations of the existing muon system, and demonstrate how the upgrades successfully address the high-luminosity challenges to maintain the high physics performance achieved in Run-I.

In the context of the new Track Trigger introduced for Phase-II (see Chapters 2 and 6) it is important to provide efficient muon identification to minimize the rate of fake associations with the few hundred tracks produced at each beam crossing. This will be best achieved in order to minimize the trigger thresholds if a relatively low rate and good precision standalone muon trigger can be maintained for low transverse momentum muons.

In addition, a performant standalone muon trigger will ensure redundancy with the track trigger and allow recovering inefficiencies for signals where higher thresholds can be tolerated. One example is the single-muon signature from a W-prime particle decay. Another, is the broad range of physics signatures where non-prompt muons or muon-like particles are produced far away from the beam line. In such cases, the track trigger may become inefficient and a standalone muon trigger will be needed.

While muon triggering and reconstruction are expected to continue to work well under HL-LHC conditions in the region up to $|\eta| \sim 1.6$, substantial degradations are expected due to the lack of redundancy in the forward region ($1.6 < |\eta| < 2.4$), unless additional muon detectors are added. This is discussed in Section 4.7.1.

The muon trigger is undoubtedly the most difficult in this forward region where the bending angle in the magnetic field is small and the background is large. The addition of fine-grained GEM chambers in the two first stations will extend the measurement lever arm by more than a factor two, providing substantially better transverse momentum assignment. The new RPC

chambers in the two next stations will complete the system to improve the background rejection. The muon trigger performance with these upgrades is discussed in Section 4.7.2, while the benefit in the offline muon reconstruction is described in Section 4.7.3.

The performance of the very forward ($|\eta| > 2.4$) extention for muon tagging is discussed in Section 4.7.4.

4.7.1 Forward muon redundancy

While it is expected that theCSCs will survive the entire HL-LHC operation, it is difficult to anticipate the degradation of the system over a period of operation that will exceed 25 years. For instance, over the course of Run-I, up to 10% of CSCs in station ME1/1 encountered operational problems that were only possible to solve during LS1. It is therefore a primary goal of the upgrade to provide redundancy for the current system in the forward region.

A qualitative indication that additional chambers are also needed in the forward region to overcome performance issues at high luminosity is presented in Figure 4.16 (left). It shows the average number of ϕ-measuring muon hits associated with a global muon track, as a function of η for the Phase-I muon detector (black histogram). While in the barrel muon detector, with $\eta < 0.8$, muon tracks are associated with an average of 25 hits, this number decreases to only 18 in the forward region with $\eta > 1.6$. In the same figure, the colored curves illustrates the rate of one of the major backgrounds, the neutron flux through the first station of the muon system. It undergoes a drastic increase (note the logarithmic scale on the right) with increasing $|\eta|$, and other backgrounds such as the flux of low-p_T muons also exhibit a similar trend. The muon detector redundancy is therefore currently the least in the region where the backgrounds are the largest. With two additional hits in each GEM detector and one per iRPC, the number of hits will become similar over the full coverage of the muon system.

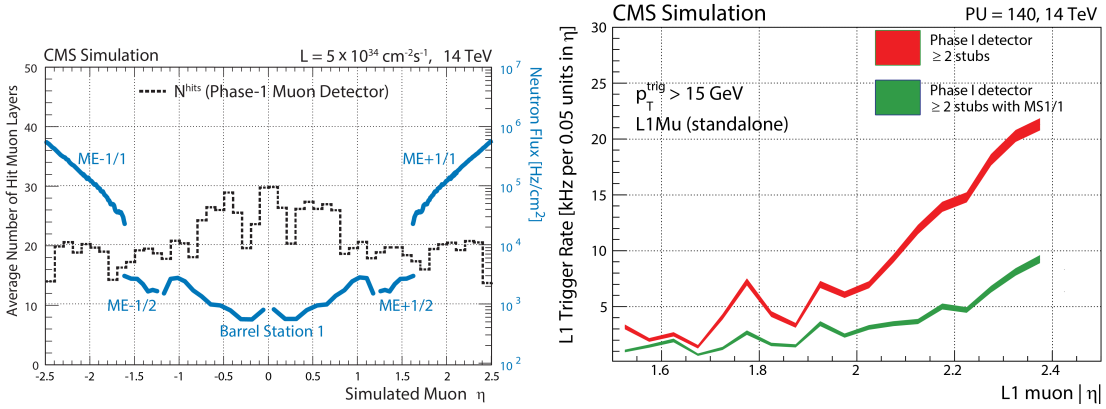


Figure 4.16: (Left) the histogram shows the average number of ϕ-measuring muon layers with reconstructed hits that are attached to a standalone muon track, for simulated muons from $Z \to \mu\mu$ as a function of η. It is compared to the flux of neutrons in Hz/cm^2 shown as colored curves (note the log scale on the right), which are the dominant cause of background hits, for the muon station first crossed by a muon with a given η. (Right) Standalone muon trigger rate as a function of pseudeorapidity for the current Phase-I muon detector under loose (red) and tight (green) trigger conditions.

In Figure 4.17, the fast deterioration of the muon trigger efficiency with even a moderate fraction of non-triggering CSC chambers is presented. Details of the simulation are as follows: in the Phase-II case, the trigger requires hits in two or more stations including hits in ME1/1, in

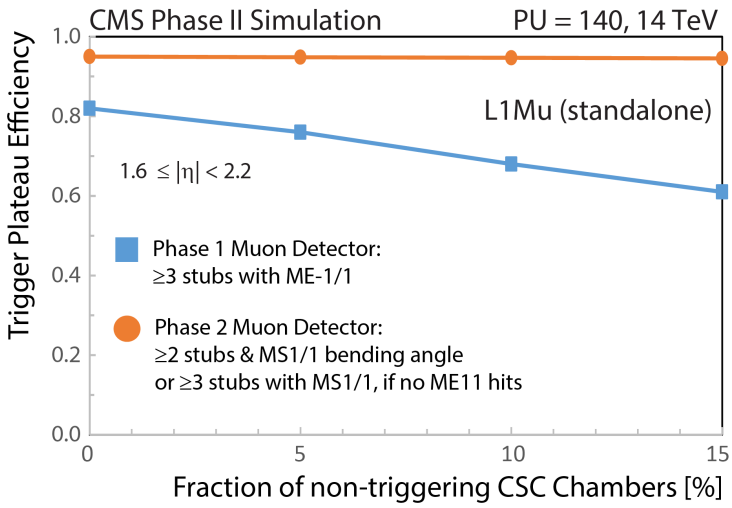


Figure 4.17: Single-muon trigger efficiency at the plateau in p_T as a function of the fraction of non-triggering CSC chambers, in Phase-I and Phase-II.

which case a bending angle cut is applied. If hits are not reconstructed in ME1/1 and the bending angle becomes unmeasurable, the trigger requires hits in three or more stations including GE1/1. In the Phase-I configuration, the trigger algorithm used is more restrictive due to an otherwise unacceptably high L1 trigger rate. Despite this, the rate remains substantially higher than for the Phase-II case and it can be seen that even without loss in the M1/1, the efficiency would be significantly lower without the upgrade.

Degradation of the CSCs will also affect the ability to find muons in the offline data analysis. This is illustrated in Figure 4.18. The top plot shows that the degradation of muon reconstruction and identification (ID) efficiency is small in the central region, up to $|\eta| = 1.6$, while it will significantly drop at higher $|\eta|$ values without upgrades. In the bottom plot, it can be seen that the efficiency is mostly recovered over the full detector coverage with the new chambers (2023 geometry). In the simulation, two stations containing segments in the case of CSC or hits in the case of GEM or iRPC are required in the muon reconstruction.

4.7.2 Performance of the forward muon trigger

Figure 4.16 (right) shows the fast growth of the muon trigger rate in the forward region of the current muon system as a function of $|\eta|$, for a p_T threshold at 15 GeV. The rate is driven by muon momentum mis-measurements associated with the tails in the p_T resolution of the muon trigger. The CSC trigger measures muon p_T using the positions of "stubs" (short segments consisting of nearby hits on several layers) reconstructed in muon stations that the track crosses: if a soft muon undergoes a substantial scattering in the material of the absorber, it can be reconstructed as a high-p_T candidate. The same figure illustrates that of the four muon stations, the first one is of special importance: the presence of reconstructed hits in this station greatly improves the muon momentum resolution and can therefore significantly reduce the trigger rates.

4.7.2.1 Muon trigger performance after LS2

With the existing system, even in beam conditions expected in Phase-I after LS2, a luminosity of $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ and a PU of 50, maintaining a muon trigger efficiency above 95% with the current thresholds will become impossible. In the region $|\eta| < 1.6$, the upgraded muon

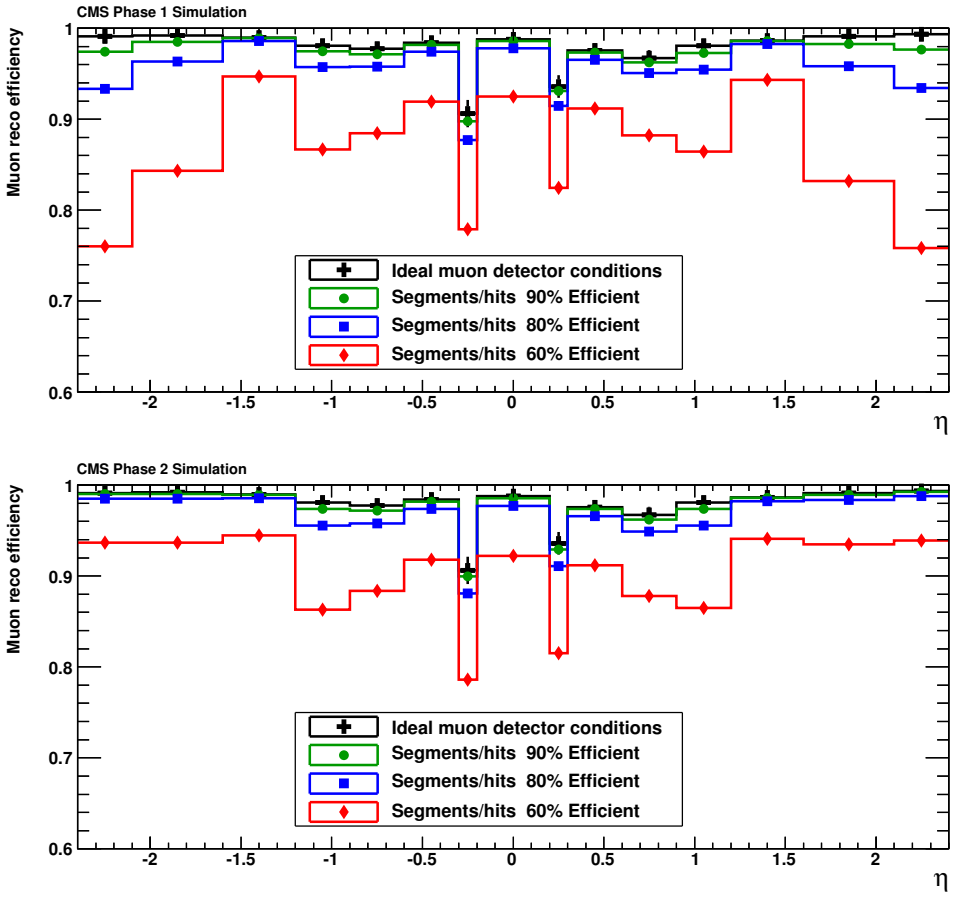


Figure 4.18: Muon reconstruction efficiency versus η for various uniform degradations of segment (CSC) and hit (GEM, iRPC) efficiencies. The top plot (2015 Geometry) illustrates the situation without the new forward muon detectors, while they are included in the bottom plot (2023 Geometry). In both cases, the black line represents the situation in Run-I.

track finder will use the information from both the CSC and RPC detectors for a combined momentum fit that will allow good control of the rate while keeping efficiency high. Due to the lack of redundancy, it will not be possible to do this in the region of $|\eta| > 1.6$. The blue curve in Figure 4.19 shows that instead an efficient trigger (with a plateau efficiency of 94%) would require a too-high-bandwidth allocation of > 20 kHz out of the total L1 trigger acceptance rate of 100 kHz. Tightening the muon selection criteria will quickly deteriorate the trigger efficiency, and any malfunctioning in the first station will triple the trigger rate in the affected region.

Early installation of the GE1/1 station in LS2 will address most of the causes of the trigger performance degradation. In addition to implementing the redundancy necessary for efficient selection of muon stubs in the most critical region, it will allow much improved measurement of the muon bending angle in the first station, using the lever arm formed by the GEMs and the CSCs. This will greatly reduce the tails in the transverse momentum resolution, and therefore the trigger rate as illustrated in Figure 4.19. Furthermore, the overall trigger efficiency will be improved with the additional hits provided by the new station. A plateau efficiency as high as 96% is expected even with a very preliminary version of the GEM-CSC integrated reconstruction algorithm. With this upgrade, CMS will be able to maintain a single-muon trigger threshold of $p_{\mathrm{T}} \sim 15$ GeV, ensuring full efficiency for offline reconstruction of muons with $p_{\mathrm{T}} > 20$ GeV.

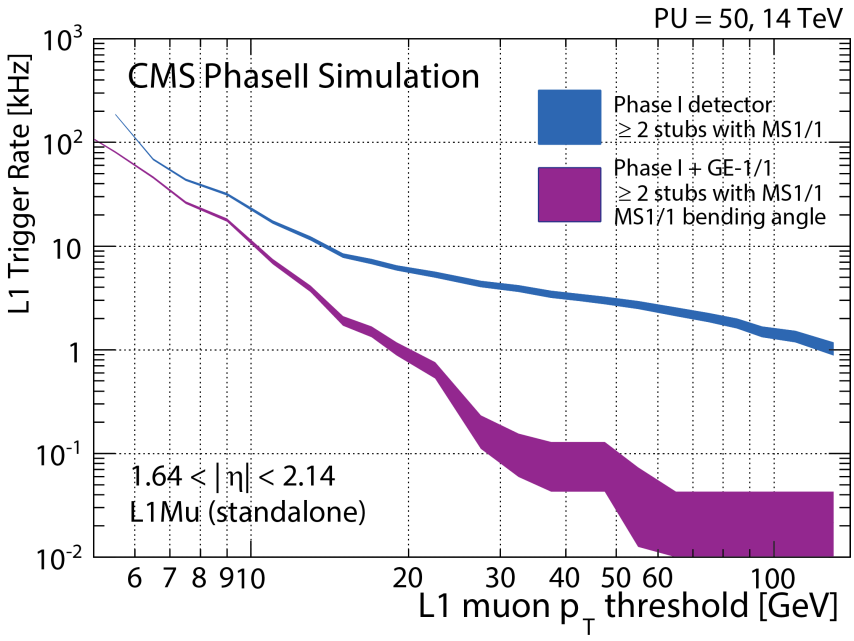


Figure 4.19: L1 muon trigger rate at a luminosity of $2 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ as a function of p_T threshold. For the Phase-I system, 2 or more stubs, one of which is in the ME1/1 station are required. With the addition of GE1/1, the bending angle between the two stations can be used and the trigger rate is greatly reduced.

4.7.2.2 Muon trigger performance at HL-LHC

With the further increase in instantaneous luminosity in Phase-II, the addition of GE1/1 will not be sufficient to maintain a high efficiency and low rate general purpose muon trigger. Apart from the expected increase in the trigger rate from higher frequency of collisions, the degradation in stub reconstruction in stations 2, 3 and 4 will increase the fraction of muon candidates with a low number of reconstructed stubs. The plots in Fig. 4.20 shows that at 140 PU, even for a perfectly working existing system (red curves), the typical stub reconstruction efficiency drops below 90% with significant dips due to the high-voltage spacers inside the CSCs. The reduction in the average number of reconstructed stubs on a track in turn increases the frequency of muon p_T mismeasurements, which inflates the trigger rate and flattens the rate curve. The same figure shows that the installation of stations GE2/1 and RE3/1 (the RE4/1 case is very similar to RE3/1) restores the local-reconstruction (stub) efficiency. Much like with GE1/1 and ME1/1, pairing GE2/1 and ME2/1 allows an online measurement of the bending angle, which can be used to improve the momentum resolution in the trigger. Overall, the deployment of GE2/1, RE3/1 and RE4/1, and the availability of the bending angle measurements in the first two stations provide adequate range of discriminating variables for designing and optimizing an efficient trigger with low thresholds and low trigger rate for Phase-II.

Deployment of the tracking trigger in LS3 will allow an ultra-high purity and low-rate trigger targeting prompt muons by matching standalone muon candidates with the Tracker tracks. The excellent momentum resolution of the Tracker eliminates the flattening of trigger-rate curve owing to mismeasured low-p_T muons and yields a very sharp turn-on of the trigger efficiency. Using tracking isolation, which is less sensitive to PU than calorimeter isolation, and combining objects targeting exclusive final states allows very high purity and low trigger rates. The new combined trigger objects, referred to as L1TkMu, use track-trigger tracks extrapolated to the muon station planes and matched with L1 standalone muon candidates. A L1TkMu object re-

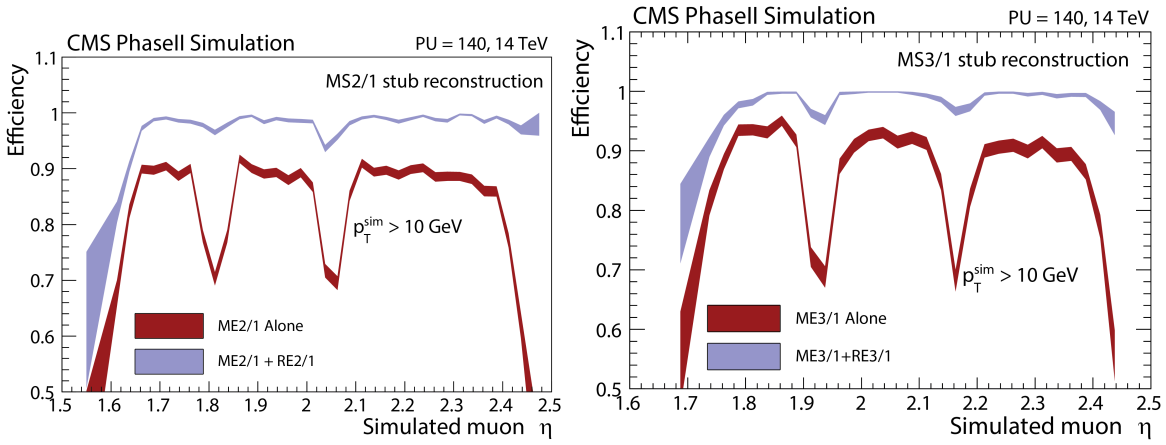


Figure 4.20: (left): The local-trigger primitive ("stub") reconstruction efficiency in station 2 as a function of eta for the present Phase-I detector (red curve) and with the addition of GE2/1 (blue curve). (right) Similar plot for station 3 with the addition of RE3/1 chambers.

quires a standalone muon track with hits in two or more stations, optionally including the first muon station. The latter condition improves the purity of the trigger by rejecting cases where a sufficiently high-p_T pion is matched to a soft secondary muon produced in the interaction of a pion in the material of the calorimeter or the absorber. To minimize accidental matches, the parameters of the Tracker and standalone muon tracks can be compared. Eventually, the L1TKMu object is assigned the momentum measured in the inner tracker. More details about the track trigger can be found in Chapter 6. The efficiency of this combined trigger is illustrated in Fig. 4.21, where the rates achieved in the region $1.2 < |\eta| < 2.4$ are plotted as a function of the transverse momentum threshold for configurations with and without the addition of the forward muon upgrade. It should be noted that the trigger rate can be further reduced using isolation requirements to achieve even lower thresholds. Examples of physics signatures that could greatly benefit from such a trigger are: $H \to \tau\tau \to \mu\tau_{had} + X$ where the muons produced in the three-body decay of tau leptons are usually soft; or possible SUSY production of light gauginos where muons produced in the decay $\chi^+ \to \mu\chi^0 + X$ are also typically quite soft.

Initial studies of the effect of strip pitch on the reduction of the L1 trigger rate have been performed, using a muon $p_T > 20$ GeV threshold and tuning the bending angle cut to reach 98% efficiency for 30 GeV muons. With the nominal 2-to-1 ganging of pairs of GEM strips in the trigger in GE1/1, the single-muon trigger rate reduction factor is 5.6 ± 0.9 compared to the Phase-I trigger, while with a 4-to-1 trigger ganging of GEM strips the factor drops slightly to 4.9 ± 0.8; for ganging of strips beyond this the trigger rate reduction factor drops quickly. For the nominal 4-strip trigger pads in GE2/1, the reduction factor is 2.4 ± 0.2, dropping to 2.0 ± 0.2 for 8-strip trigger pads. In both cases, increased trigger segmentation beyond the nominal values does not significantly improve the trigger performance.

4.7.3 Enhancement of the offline reconstruction of forward muons

In the following, we present the performance of the muon reconstruction with the Phase-II upgrades, distinguishing the nominal coverage acceptance ($1.6 < |\eta| < 2.4$), and the extension in the $2.0 < |\eta| < 3.0$ range. In the former region, the improvements come from the installation of the double-layered GEM chambers in front of ME1/1 and ME2/1 (GE1/1, GE2/1) and from the two additional iRPC chambers in the external station (RE3/1, RE4/1). In the the second region, the improvements come from muon identification with the new 6-layered triple-GEM

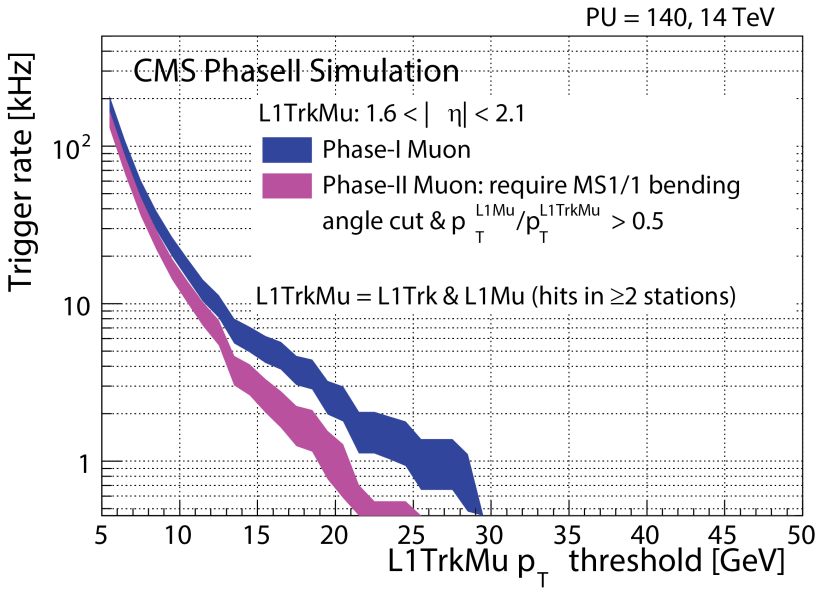


Figure 4.21: Rate of the L1TkMu trigger as a function of the p_T threshold for $1.2 < \eta < 2.4$ and 140 PU. A significant improvement is obtained with the requirement of reconstructed hits in the first muon station that eliminates soft muons produced by interactions of high p_T pions in the calorimeter.

station (ME0) behind the new endcap calorimeters.

Since an early installation of the GE1/1 chambers during LS2 is highly desirable (see Section 4.7.2.1), this configuration has been studied and is referred to as "2019 scenario". The "2023 scenario" includes all the new Phase-II detectors (GE1/1, GE2/1, RE3/1 and RE4/1, and ME0). The current detector geometry without upgrades is referred to as "no upgrade".

The analysis uses an ideal detector with no miscalibration and misalignment. We have included in our simulation a 98% detector efficiency, based on the test beam results. The beam-induced background is included in the simulation for all the new detectors: hit rates predicted using FLUKA, including the detector sensitivities, are used to generate additional spurious signals that are added when simulating the electronics response. The probability of these additional hits are weighted to the luminosity to properly represent the contribution of these backgrounds at a given instantaneous luminosity.

The performance of the reconstruction algorithms has been tested using samples of single muons generated with different values of p_T and flat distributions in η and ϕ. The performance on single particles is evaluated from samples of single muons uniformly distributed in ϕ and η, overlaid with an average of 140 PU events. Hits from all new stations are added in the final fit of the muon tracks. Here we report results on the standalone muon performance, while the performance of the global muon reconstruction and identification, including the track information, is reported in Chapter 9.

4.7.3.1 Enhancement of performance in the nominal muon coverage

In CMS, the excellent momentum resolution of muons is mainly the result of the precise resolution of the tracker system. For high-momentum muons, though, the muon system contributes substantially to the overall momentum measurement [102], since curvature resolution scales as the square of the lever arm and the muon system provides a much larger lever arm than

the tracker alone. However, there is a poorer resolution in the forward region compared to the barrel, caused by magnetic field inhomogeneity, by showering effects at high momenta and by the lower number of measured points with respect to the rest of the CMS muon system. Additional detectors in the muon system, with good spatial resolution, will improve the limited p_T resolution in the endcap region and provide higher efficiency in a high-background environment.

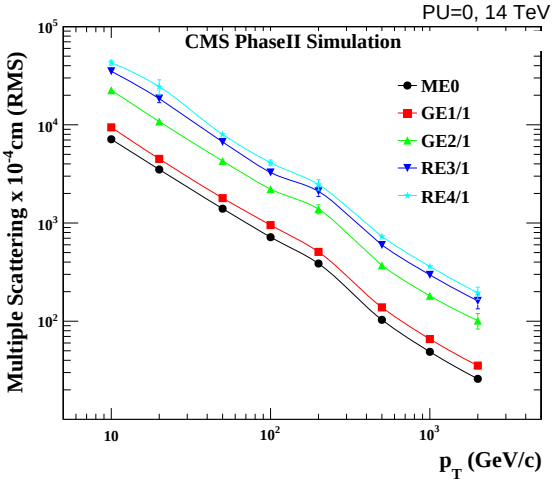


Figure 4.22: The RMS of the multiple scattering displacement as a function of muon p_T for the different forward muon stations. All of the electromagnetic processes such as bremsstrahlung and magnetic field effect are included in the simulation.

The spatial resolution lower limit, set by the multiple scattering experienced by muons in traversing the CMS material, has been estimated from simulation, assuming the current geometry scenario. The results are displayed in Fig. 4.22, which shows the RMS displacements caused by multiple scattering as a function of p_T in the different endcap stations at a fixed η and ϕ ($(|\eta|, \phi) = (2.0, 0.0)$). The RMS displacement for $p_T = 200$ GeV is about 500 μm at the surface of GE1/1, about 1 mm at GE2/1, and reaches 3-4 mm at the fourth station (RE4/1). These values have been used in setting the various detector granularities as given in Sections 4.3.3 and 4.5.

4.7.3.2 Standalone Muon reconstruction redundancy

While the present barrel muon system has enough redundancy to maintain its performance at the HL-LHC, the lack of redundancy of the system in the forward region will become an issue if any CSC chambers become inoperative. We have studied how the performance of the q/p_T resolution, of the charge mis-identification and of the reconstruction efficiency will be affected in the three configurations mentioned above, considering the case in which one segment in CSC is lost. The q/p_T resolution is defined as the Gaussian width of the q/p_T residual distribution:

$$\frac{\delta(\frac{q}{p_T})}{\frac{q}{p_T}} = \frac{q^{Rec}/p_T^{Rec} - q^{Sim}/p_T^{Sim}}{q^{Sim}/p_T^{Sim}}, \tag{4.1}$$

where q is the charge and p_T^{Sim} and p_T^{Rec} are the simulated and reconstructed transverse momenta, respectively, plotted in the range [-6,+6]. Sigma of the q/p_T residual distribution is obtained by fitting the distribution to the mean$\pm$RMS. Figure 4.23 (left) shows the recover, obtained thanks to the additional stations (squared markers), of the RMS of the q/p_T resolution

in the case of failure of ME1/1 station with respect to the no upgrade case (cross markers). The overall gain obtained with the upgraded system (circle markers) with respect to no upgrade case (triangle markers) is also shown in the same Figure and in Figure 4.24 (middle).

Figure 4.23 (right) shows that the standalone muon performance can be restored in case of failure of either ME1/1 or ME2/1 or ME3/1 or ME4/1 station adding new detector (GE1/1, GE2/1, RE3/1, RE4/1 respectively). The effect is reported for each station as improvement of the RMS of q/p_{T} residual distribution adding new detector with respect to the case of non-operational ME station. As expected the major recovery (up to 60%) is obtained by adding GE1/1 (squared filled markers) and GE2/1 (triangles filled markers), due to the high spatial resolution, comparable with the CSC one. The overall improvement, up to 35%, with the upgraded system with respect to no upgrade case is shown in Figure 4.23 (right, black dots).

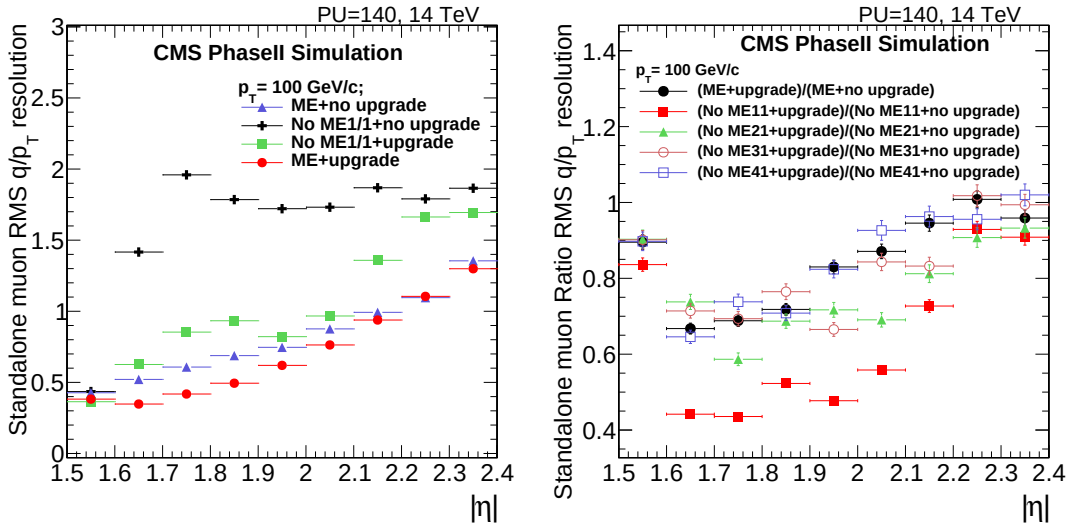


Figure 4.23: On the left Restoration of RMS of q/p_{T} distribution as a function of the simulated eta for $p_{\mathrm{T}} = 100$ GeV at 140 PU by adding the upgrade detectors in case of ME1/1 failure in the 2023 scenario at PU = 140. On right the gain obtained by adding new detectors to each failure configuration is shown as ratio of RMS of q/p_{T} resolution.

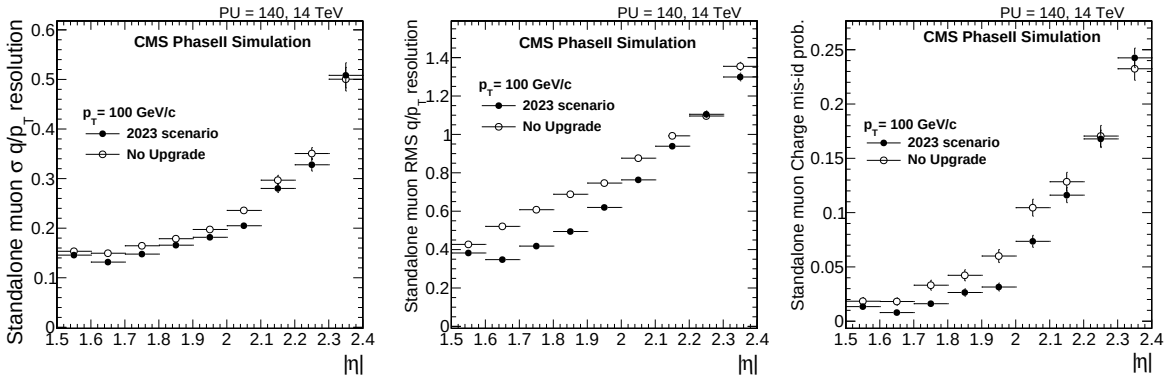


Figure 4.24: Improvements with added redundancy of the upgrade forward muon detectors, for standalone muons, as a function of the simulated $|\eta|$ for $p_{\mathrm{T}} = 100$ GeV at 140 PU, for the "2023 scenario" (filled markers) and "no upgrade" (empty markers) : sigma (left), RMS (middle) of the q/p_{T} residual distribution and the charge mis-identification probabilities (right).

Finally, in the "2023 scenario" the additional hits from new detectors in track fitting will im-

prove the overall the muon track resolution and reduce the charge mis-identification probability, which in case of "no upgrade" scenario degrade with 140 PU. In figure 4.24, the sigma (left) and the RMS (middle) of the q/p_{T} distribution, and the charge mis-identification probability (right) are shown as a function of $|\eta|$ for a standalone muon of $p_{\mathrm{T}} = 100$ GeV with (circles markers) and without (triangle markers) upgrade. Adding the new detectors in the high $|\eta|$ region provides an improvement up to 10% in the q/p_{T} resolution (right Figure), a much better control of the tails of the q/p_{T} distribution (middle Figure), where the majority of the badly reconstructed tracks are (as also shown in Figure 4.23 (middle and right)) and a large reduction, up to $\sim$50%, of the charge mis-identification probability.

4.7.4 Enhancement of far forward muon coverage

This section discusses the expected performance with the muon tagging in the region covered by ME0, i.e. $2.0 < |\eta| < 3.0$. This station has a slight overlap with ME1/1 and GE1/1 in order to provide good muon chamber relative alignment using high-p_{T} muon tracks.

For an initial estimate, the expected hit rates are used to estimate the probability of reconstructing a local muon segment from random coincidence of neutron-induced hits. Assuming that ME0 strips have length of $l = 10$ cm and that the coincidence of hits in four out of six layers of ME0 within a region of width $d = 1$ cm will yield a reconstructed local muon segment, the probability of finding a muon segment in the search area is found to be only $10^{-5} - 10^{-6}$, at a luminosity of 5×10^{34} cm^{-2}s^{-1}. This very small probability needs to be multiplied by the number of tracker tracks consistent with the vertex associated with the triggered event and found in the pixel extension to be consistent in curvature with a high-p_{T} track; the number is physics-channel dependent. For events with an already-identified vertex, there may only be a few such tracks, but in a few channels, for example a single-lepton-plus-photon channel, one may need to consider all primary vertices. Even if in the latter case 100 candidate tracks meet the criteria, the probability of a random match will be only $10^{-3} - 10^{-4}$. Verification of the good muon tagging performance will be performed as soon as the simulation software for the reconstruction with the ME0 station becomes available.

4.8 Project planning and cost estimates

4.8.1 Organization

Currently, each of the CMS muon sub-detectors has its own project manager and project organization. The sub-detector PMs report to a common Muon Institution Board. The Phase 2 muon upgrade activities are embedded in the sub-detector organizations and steered by an upgrade coordinator reporting to the PM. The muon detector project is undergoing reorganization under a muon management board chaired by a single project manager. In the new muon organization, the Phase 2 muon R&D will be centrally coordinated, while the major responsibility for construction will continue to be taken by the muon subdetectors. The new organization will improve cross-detector upgrade planning and decision making.

4.8.2 Schedule

Major steps planned during the R&D phase of the muon upgrades are listed in Table 4.3. The Muon Technical Design Report (MTDR) is planned to be completed at the end of 2017. The milestones shown are planned to cover the developments that need to be done up to the Engineering Design Reviews (EDR)that will launch production. The target dates are established for

the purpose of project planning. Actual dates can be several months later in all cases without affecting technical progress.

In preparation for the MTDR (milestone TP.MU.CM.3), all current or proposed muon detector technologies must show the ability to withstand the rates and doses expected from the HL-LHC (milestone TP.MU.CM.1). When final prototypes are available, including new RPC and GEM technology chambers and new electronics for all muon detectors, they must be similarly qualified (TP.MU.CM.4). The GIF++ facility is envisioned for most of these tests, and infrastructure such as chamber supports, electronics, and gas systems are needed. Accumulating HL-LHC doses at GIF++ will take 9-12 months of irradiation. Also before the MTDR, the situation with respect to muon technologies that use greenhouse gases (CSC, RPC, and GEM) must be explored and reviewed (milestone TP.MU.CM.2). The eco-gas R&D efforts include small-scale gas systems, small chambers, and supplies of various gases.

In preparation for the MTDR, the DT minicrate electronics replacement project will develop a first version trigger algorithm (TP.MU.DT.1) and a detailed system design both for the UXC electronics and for the USC electronics (TP.MU.DT.2). Items needing R&D include the UXC system (Section 4.2.3.3) that will replace present Minicrates, that is: new Minicrate mechanics, TDC implementation (first in ProASIC FPGAs and later in improved FPGAs that will be available at the time of production), slow controls, optical links and the final OBEDT board with all previous items included. Also, extensive R&D is needed for the USC part to develop the trigger algorithms in the appropriate hardware platform (Section 4.2.3.4), and proper test systems will need to be developed to validate the overall design. All of these elements will be required for completion of the prototype R&D effort (TP.MU.DT.3), and the validation of the entire system in situ at Point 5 during LS2 (TP.MU.DT.4).

The replacement of cathode electronics on the forward (inner ring) of CSC chambers (Section 4.2.4.2) requires additional simulation work to verify that the DCFEBs that were used for ME1/1 installation during LS1 are appropriate to HL-LHC for the other stations, and engineering to prepare to produce more of the Comparator ASIC chips that were originally produced in the year 1999 (TP.MU.CS.1). Since the optical link data throughput for DMB and DDU boards will be insufficient, the back-end CSC electronics will need to be redesigned for these chambers (TP.MU.CS.2), and demonstrated using prototype boards (TP.MU.CS.3).

The VFAT3 front-end ASIC (Section 4.3.6) is required for readout of all GEM detectors (TP.MU.GE.1). Installation of the GE1/1 slice test in the YE1 nose at the end of 2016 (TP.MU.GE.2) is necessary to ensure good integration of hardware and software into CMS early on, and to gain valuable operational experience in the high-rate environment. Construction of GE2/1 (TP.MU.GE.3) and ME0 (TP.MU.GE.4) chamber prototypes will enable performance verification (TP.MU.GE.5) before the MTDR. Notable challenges are the large size of foils for GE2/1 and the thin multi-layer design foreseen for ME0. In addition to the VFAT3, there are various other GEM electronics such as the GEB, OH boards and the HV system, that will need to be prototyped (TP.MU.GE.6). The preparation of the entire GE1/1 system (CERN-LHCC-2015-001, CMS-TDR-15-00) for installation (TP.MU.GE.7) is not strictly an R&D milestone, but is included in the list as an important marker of the progress of the GEM developments. Development of the design of chambers will continue for a while beyond the MTDR for GE2/1 (TP.MU.GE.8) and ME0 (TP.MU.GE.9) before the EDR.

The integration design (geometry, mounting, services) for the additional RPC stations, RE3/1 and RE4/1 (Section 4.4) will be defined (TP.MU.RP.1) before the MTDR. In order to enable the choice of improved RPC technology before the submission of the MTDR (TP.MU.RP.4), there must be fully-qualified chamber prototypes (TP.MU.RP.2) as well as the corresponding elec-

Table 4.3: Planned next steps for the Muon upgrade projects.

Project	Milestone Identifier	Target date	Description
Common muon	TP.MU.CM.1	Q2 2017	Technology verification of longevity from irradiation tests, etc. for DT, CSC, RPC, and GEM detectors
	TP.MU.CM.2	Q3 2017	Review of proposed eco-gas mixtures for RPC, CSC, GEM
	TP.MU.CM.3	Q4 2017	Muon TDR submission
	TP.MU.CM.4	Q3 2019	Validation of radiation tolerance with final prototypes of all chambers
DT	TP.MU.DT.1	Q1 2017	First version of the DT trigger algorithm for USC electronics
	TP.MU.DT.2	Q2 2017	DT minicrate electronics system design
	TP.MU.DT.3	Q4 2018	Completion of prototyping for DT minicrate electronics
	TP.MU.DT.4	Q1 2019	System validation at P5 during LS2
CSC	TP.MU.CS.1	Q2 2016	CSC inner chamber front-end electronics design
	TP.MU.CS.2	Q2 2017	CSC inner chamber back-end electronics design
	TP.MU.CS.3	Q4 2018	Completion of prototyping for CSC inner chamber electronics
GEM	TP.MU.GE.1	Q4 2016	GEM front-end VFAT3 electronics design
	TP.MU.GE.2	Q1 2017	GE1/1 in-situ slice test installed
	TP.MU.GE.3	Q2 2017	Construction of GE2/1 chamber prototype
	TP.MU.GE.4	Q2 2017	Construction of ME0 chamber prototype
	TP.MU.GE.5	Q3 2017	Performance verification of GE2/1, ME0 prototypes with muon beam
	TP.MU.GE.6	Q4 2017	Completion of prototyping for GE2/1, ME0 electronics
	TP.MU.GE.7	Q1 2018	GE1/1 ready for installation
	TP.MU.GE.8	Q2 2018	Completion of design and prototyping for GE2/1 chambers
	TP.MU.GE.9	Q4 2018	Completion of design and prototyping for ME0 chambers
RPC	TP.MU.RP.1	Q2 2016	Integration design of RE3/1 and RE4/1 chambers
	TP.MU.RP.2	Q2 2017	Qualification of prototypes for improved RPCs
	TP.MU.RP.3	Q2 2017	Prototype of improved RPC front-end electronics
	TP.MU.RP.4	Q3 2017	Choice of improved RPC technology
	TP.MU.RPC.5	Q3 2018	Completion of RE3/1 and RE4/1 chamber design and integration

tronics design (Section 4.4.4, milestone TP.MU.RP.3). Following the MTDR, work will continue to complete the design and integration (TP.MU.RP.5) before the EDR.

The overall conceptual schedule for the Phase-II muon upgrades is shown in Fig. 4.25. While for the GEM and RPC projects, EDRs will be held to launch construction of the new chambers, the DT and CSC upgrade construction will be authorized through Electronics Systems Reviews (ESRs). These reviews may be combined or separate depending on the final schedules for each of the components of the muon upgrade.

4.8.3 Cost estimates

In the case of the DT minicrate replacement, the functionality of the current system is well understood, and R&D is well underway to re-engineer this system to meet the requirements. The costs presented represent a straightforward extrapolation of known electronics technology. In the case of the CFEB replacement, the Digital CFEB (DCFEB) and associated off-chamber

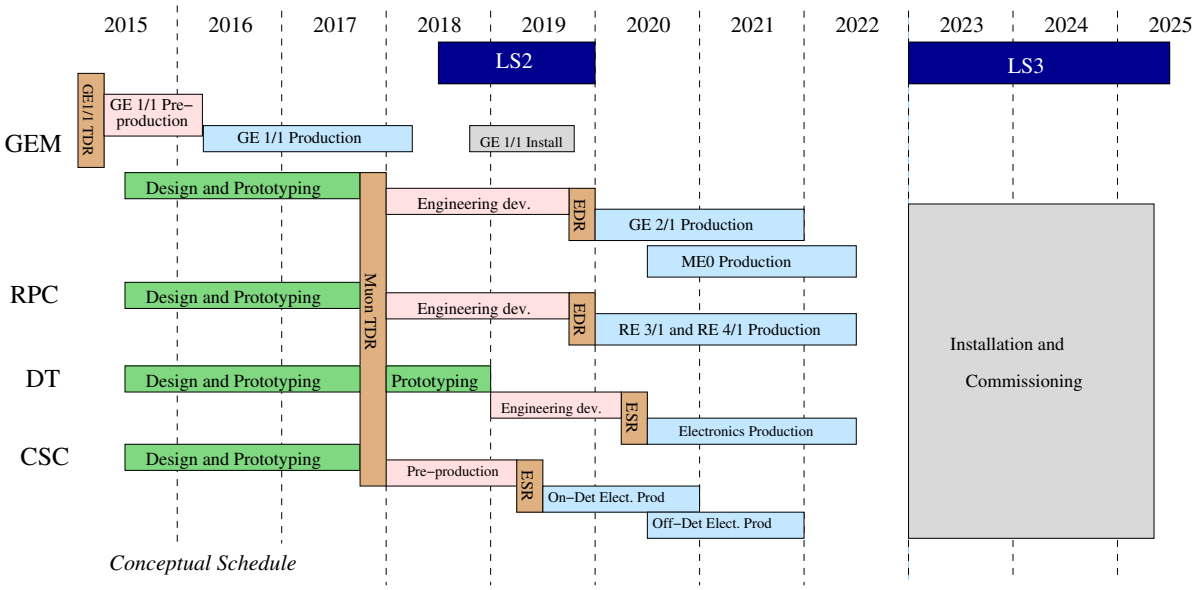


Figure 4.25: Conceptual schedule for the Muon portions of the Phase-II upgrade.

boards already installed during LS1 on the ME1/1 chambers serve as accurate models for the cost estimation.

The new GE1/1 forward muon station is based on GEM technology that is by now very well understood, having gone through extensive prototyping in CMU as well as operational experience in other high-rate experiments. Electronics and infrastructure dominate the total cost, not the detectors themselves. The costs have already been well established, since this station has been proposed for installation during LS2.

The cost estimate for ME0, built with GEM technology, is also a straight-forward extrapolation of the closely reviewed estimate for GE1/1. Some aspects will need detailed engineering before the muon TDR is written, since this station lies within a completely new endcap that imposes significant constraints on the design coming from: the mechanical support of the very heavy endcap "nose"; and the depth available behind the calorimeter, and space available for electronics, cabling and services.

The cost estimates for additional forward muon stations (GE2/1, RE3/1, and RE4/1) are based on straightforward extrapolations. In the case of GE2/1, the cost estimate is extrapolated from the closely reviewed GE1/1. In the case of the RE3/1 and RE4/1 stations, the cost estimate is extrapolated from similar RPC detectors built recently, such as the RE4/2 chambers that have been installed during LS1. The situation with respect to engineering and installation services is favorable, since these detectors were foreseen (as RPC detectors) in the original design of CMU, and space was allocated for their mounting.

A summary of the present cost estimates is shown in Table 4.4.

Table 4.4: CORE Cost estimates for the Phase-II Upgrades of the Muon System.

Estimated CORE cost in MCHF (2014)	
Minicrate electronics	3.3
Trigger electronics	2.1
Opto-links	0.7
Total DT electronics	**6.1**
Electronics boards	3.1
Low voltage system	0.2
Opto-links	0.4
Total CSC electronics	**3.7**
Chamber assembly	2.5
Electronics	3.2
Power system	1.2
Services	0.7
Installation	0.3
Total GE11 and GE21	**7.9**
Chamber assembly	1.2
Electronics	0.3
Power system	0.5
Services	0.2
Installation	0.2
Total RE31 and RE41	**2.3**
Chamber assembly	1.2
Electronics	2.0
Power system	0.9
Services	0.3
Installation	0.2
Total ME0 Muon extension	**4.5**
Total Muons	**25**

Chapter 5

Beam Radiation Instrumentation and Luminosity

This chapter describes the instrumentation needed to measure and give feedback on the HL-LHC beam conditions, measure the radiation products in the CMS experimental cavern, and measure with precision the delivered luminosity, whenever there is beam in the LHC machine.

In order to ensure reliable luminosity and beam background monitoring for HL-LHC, all the presently installed Phase-I beam monitoring devices that are located in high radiation environments close to the beam pipe will need to be completely replaced due to radiation damage. The extreme radiation levels expected will require higher radiation tolerance technologies and detectors adapted mechanically to the future HL-LHC integration constraints in the tracker volume and forward region. In the case where radiation hard sensors and electronics are not yet identified to withstand the full integrated luminosity expected for HL-LHC, the integration considerations of the detectors should allow for occasional replacement of the sensitive detector elements, during an extended technical stop or long shutdown (LS).

Installed Phase-I detector systems not degraded by radiation effects will still require some maintenance so that the present functionality of the systems is maintained for HL-LHC operation.

The collective functionality of the beam monitoring systems for CMS should satisfy the following:

- Active protection of sensitive parts of the CMS detector from dangerous beam loss events whenever there is beam in the machine.
- Online monitoring and realtime feedback to the LHC machine on the machine-induced-background in the CMS detector, and a notification of poor beam conditions whenever there is beam in the machine.
- Online and passive monitoring of the radiation environment of the CMS cavern.
- Online measurement of the bunch-by-bunch luminosity delivered to CMS and real-time feedback to the LHC machine.
- Precision measurements of the luminosity recorded by CMS to determine cross-sections.
- Online measurement of the relative timing between the two beams, to centre the longitudinal collision region prior to collisions, and a measurement of the time between the beams and the CMS reference clock for correct adjustment of the phase of the CMS clock distributed centrally to all sub-detectors.
- Provide L1 trigger inputs of the bunch fill pattern for each beam and of the machine

induced background candidates.

Sub-detectors to address these beam monitoring functionalities for CMS during the Phase-I running period are listed in Table 5.1, and a similar set of instruments are needed for HL-LHC operating conditions. The functionalities of the instrumentation for the HL-LHC environment will be further described in this chapter, identifying upgrades and R&D needed for the existing infrastructure.

Table 5.1: Overview of presently installed CMS beam monitoring systems. Z is the longitudinal distance from the CMS interaction point (IP5) and R is the radial distance from the centre of the LHC beam pipe and X is the horizontal coordinate.

Subsystem (Sensor type)	Location	Sampling Time	Function	# Channels
BCM1L (Polycrystalline Diamond)	Pixel Volume Z=±1.8 m , R=5.0 cm	40 μs	Protection & Beam-Background Monitoring	8
BCM2L inner (Polycrystalline Diamond)	CASTOR Table Z=±14.4 m , R∼5.0 cm	40 μs	Protection & Beam-Background Monitoring	8
BCM2L outer (Polycrystalline Diamond)	CASTOR Table Z=±14.4 m , R∼28.0 cm	40 μs	Beam-Background Monitoring	16
Ionization Chamber	CASTOR Table Z=±14.55 m , X∼28.0 cm	40 μs	Beam-Background Monitoring	2
Passive dosimetry (TLD)	On Detector & Experimental Cavern	∼ months	Dose Monitoring	>200
Medipix (Hybrid Silicon Pixel detector)	Experimental Cavern & Service Cavern Z = 15 m, X = 12 m	1 minute	Dose & Mixed-Radiation Monitoring	2
HF Neutron RadMon (Polyethylene & Ionization chamber)	Behind HF Z=±13 m , R∼1.4 m	1 second	Neutron Flux Monitoring	16
BHM Phase-I upgrade (Synthetic Fused Quartz)	Rotating shielding Z=±20.6 m , R=1.8 m	6 ns	Beam Background Monitoring	40
BCM1F Phase-I upgrade (Single-crystalline Diamond)	Pixel Volume Z=±1.8 m , R=5.0 cm	6 ns	Beam Background & Luminosity Monitoring	48
Pixel Luminosity Telescope (Pixelated Silicon)	Pixel Volume Z=±1.8 m , R=5.0 cm	25 ns	Luminosity Monitoring (Online and Offline)	48
HF Luminosity Readout	HF Z=±1∼m, R=13 –140 cm	25 ns	Luminosity Monitoring (Online and Offline)	24
BPTX (Button Beam Pickup)	Upstream of IP5 Z=±175 m	200 ps	Longitudinal Position & & Timing Monitoring	2

5.1 Active Protection

The Beam Condition Monitor (BCML) detectors are linked to the abort system of the LHC and used to protect the CMS silicon tracker from catastrophic beam loss events. The BCML detectors measure the current created by ionization in a polycrystalline (pCVD) diamond sensor. This signal current is proportional to the ionizing particle flux through the active detector material thus providing a relevant measurement for the radiation load in detectors at that location.

The presently installed BCML detectors [142] are read out using the LHC Beam Loss Monitor (BLM) electronics [143–145], with the ability to dump the beam within 2 to 3 LHC orbits, on the detection of a signal above a given threshold, for a particular integration time. A short duration (40 μs) threshold protects the silicon tracker from potentially damaging amounts of energy deposition in the electronics, and a longer-term ($\sim$83.3 s) integration threshold protects against high beam background conditions that can result in problematic conditions for data taking and reconstruction as well as an increased dose-rate to the inner detector region.

The small sensor size ($10 \times 10 \times 0.4\,\mathrm{mm}^3$), relative radiation hardness [146, 147] and ability to operate without the need for services such as cooling, make this a robust detector system, especially for locations where low material budget is important. These BCML systems are powered independently of CMS and are in full operational mode before the LHC accelerator can be filled and an interlock is in place to ensure that this condition is satisfied. The BCML systems are connected to an uninterruptible power system, in order to permit LHC to run even if the local CMS power distribution is not functional for some reason.

To enhance the acceptance coverage of potential poor beam background conditions and for operational redundancy, there are two BCML detector systems installed in CMS as shown in Table 5.1. Two LHC-type ionisation chambers used for beam loss measurements in the LHC are installed close to the BCM2L location and are connected to the same electronics. These are used for monitoring and performance studies of the BCML systems.

5.1.1 Limitations of BCML for HL-LHC based on Run-I performance

The performance of the BCML detectors to date has been reliable and useful monitoring of the beam conditions has been provided to CMS and to the LHC [148, 149]. The excellent performance of the LHC vacuum and collimation systems [150], and the early detection of beam loss events measured by the LHC beam loss system, has resulted in a good quality beam delivered to CMS thus far. No LHC beam dump has yet occurred, due to the onset of a dangerous beam loss event as measured by the BCML in CMS. On various occasions beam loss events have occurred and were measured by the BCML system, showing its sensitivity towards these events, though none of these events have been significant enough to trigger a beam abort.

With the $29\,\mathrm{fb}^{-1}$ of delivered luminosity thus far in LHC Run-I, the detectors have accumulated about $1.0 \times 10^{14}\,\mathrm{cm}^{-2}$ (BCM1) and $2.8 \times 10^{14}\,\mathrm{cm}^{-2}$ (BCM2 inner) 24 GeV proton equivalent ($\mathrm{p_{eq}}$) fluence [151], as estimated with FLUKA Displacement Per Atom (DPA) scoring. The choice of the DPA scaling to the damage potential of a 24 GeV proton, is to compare measurements done in a mixed-radiation environment, like CMS, with proton test-beam irradiations of diamond sensors typically done with protons at this energy, by the RD42 collaboration.

Despite the relative radiation hardness of the pCVD sensor material, serious radiation damage and loss in signal efficiency has already occurred [152]. As the trend of the data from Fig. 5.1 suggests, the sensitivity of the currently installed sensors, operated at about $0.5\,\mathrm{V}/\mu\mathrm{m}$, will be further reduced after additional exposure to radiation. A replacement of detectors will be necessary before the time that the signal efficiency falls below the noise level of the frontend electronics, as indicated by the red-dashed level in Fig. 5.1. Based on a data–driven extrapolation using a parametric fit, this level will be reached after about $120\,\mathrm{fb}^{-1}$ in the case of BCM2L inner. Then a replacement of the sensors will be necessary and/or a change in the operating conditions. The sources of this signal degradation are under study and a theoretical model to justify the parameters of the empirical fit is not yet fully developed. This result suggest that for HL-LHC conditions, the baseline strategy requires access to both the BCM1L and BCM2L detector locations to replace the sensors periodically.

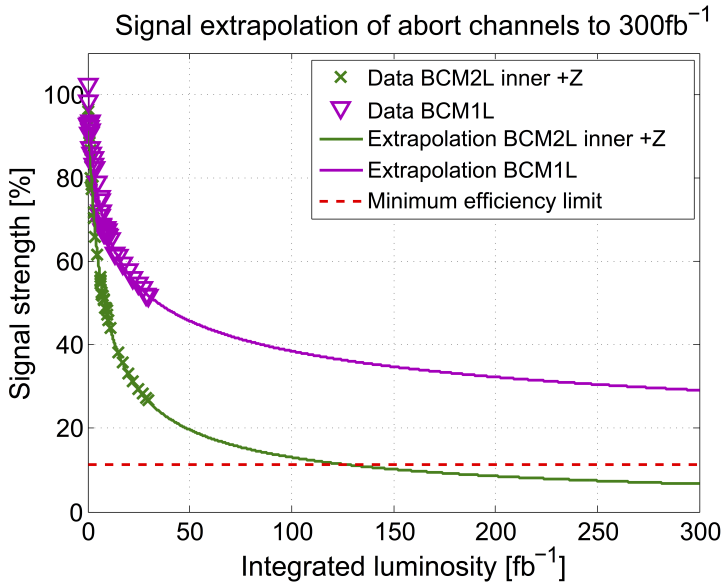


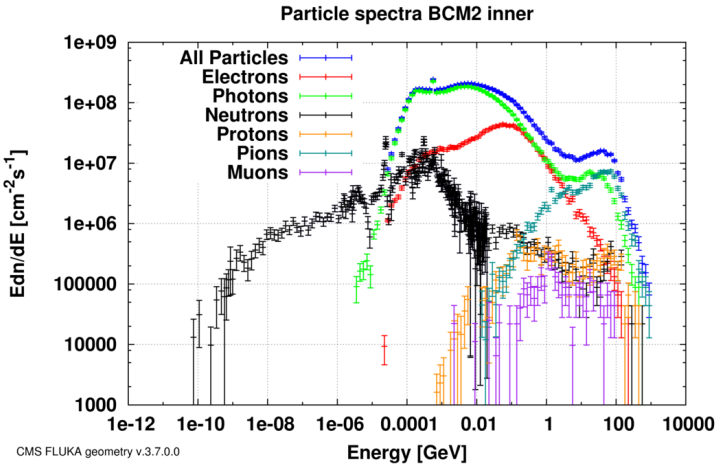
Figure 5.1: Signal strength measurements for the BCML detectors, operated at 0.5 V/ μm during LHC Run-I, and an extrapolation using a parametric fit to an integrated Luminosity of 300 fb^{-1}. The minimum signal efficiency compatible with the threshold settings in the backend electronics is indicated by the red dotted line.

5.1.2 Upgrades of BCML for HL-LHC

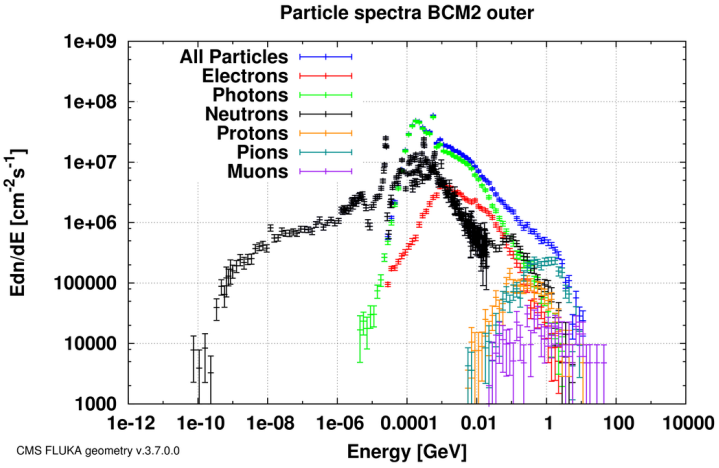
As for LHC operation, CMS will require redundant, reliable and radiation hard BCML detectors for its protection against catastrophic beam losses in the HL-LHC beam era. It is mandatory to operate the detector independent of services of cooling or gas, and the detector should be designed to be robust against changes in external environmental conditions such as temperature, humidity and magnetic field. For maximum sensitivity to beam losses, the detectors should be located close to the beam pipe. The radiation environment for potential locations of the BCML detectors for HL-LHC is shown as a function of the energy of the particle in Fig. 5.2 and summarised in Table 5.2. The sensor material should be able to withstand an average hadronic fluence of up to 6.2 $\times$ 10^{15} cm^{-2} (averaged over the region R $=$ 3.45 $-$ 9.60 cm), for the full operational period of 3000 fb^{-1} of integrated luminosity. This imposes stringent requirements on the radiation hard sensor technologies. It is advisable to design a detector system that is relatively easily accessible and to replace the sensors on a few occasions during the HL-LHC running period, thus lowering the radiation tolerance requirements.

The technology choices for the sensor material are still under study. The pCVD diamond sensors remain attractive since they can be used in locations where material budget is important and are robust against changes in external environmental conditions. Owing to their insufficient radiation hardness, as presently measured, the mechanics of a diamond-based BCML detector should be designed to be easily retractable and produced from light materials to minimise radiation induced activation, so that an extended technical stop can be exploited for a replacement of the sensors. R&D on improving the quality of pCVD sensors, for increased charge collection efficiency, improved uniformity to minimise the probability of erratic currents will be of benefit for the potential performance of future BCML-type detectors.

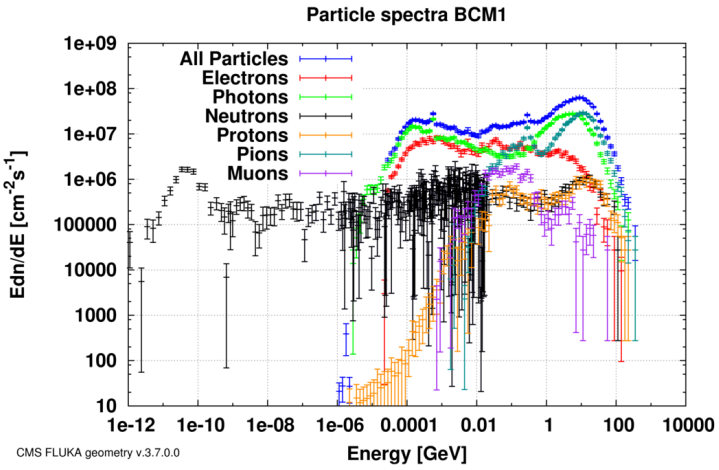
In locations where there is less of a constraint on material budget, small gas-sealed ionization chambers that require no active gas flow services, as used in the LHC beam loss system and



(a) BCM2L inner



(b) BCM2L outer



(c) BCM1L

Figure 5.2: Particle energy spectra at the possible future BCML locations for HL-LHC at 7.0 TeV beam energy, grouped per detector location, with regions listed in Table 5.2. The radiation in (a) is for BCM2L inner, (b) is for BCM2L outer and (c) is for BCM1L location, compatible with HL-LHC integration constraints. The data is normalized to the flux at nominal luminosity for HL-LHC of $5 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The energy spectra is plotted for different particle types above their transport thresholds, except for the *All Particles* where the neutron contribution is included only above 20 MeV.

Table 5.2: The average radiation fluence after 3000 fb^{-1} for potential locations for future BCML detectors as estimated from FLUKA simulations for HL-LHC conditions. This summarises the fluence shown in differential form, see Fig. 5.2. The regions for the fluence scorings are the following: BCM1L (R = 3.45-9.60 cm, Z = 183.34-190.20 cm); BCM2L Inner (R = 7.45-12.50 cm, Z = 1433.0-1439.0 cm) and BCM2L Outer (R = 25.0-33.0 cm, Z = 1433.0-1439.0 cm) and corresponds to the *average* fluence for each region.

	Particle Fluence [cm^{-2}] (for 3000 fb^{-1})		
	BCM1L	BCM2L Inner	BCM2L Outer
Charged particles	9.5×10^{15}	1.6×10^{16}	9.0×10^{14}
Hadrons	6.2×10^{15}	5.0×10^{15}	2.8×10^{15}

installed in for Run II, in Table 5.1, will also be considered.

Generic R&D of other sensor technologies will be followed, such as relatively low cost, optical grade, single-crystal sapphire sensors. Although Sapphire sensors have a smaller charge collection efficiency that pCVD sensors, they have been demonstrated for use as beam-loss measurements at the FLASH accelerator, and are planned to be installed at FLASH-II and XFEL [153]. In order to study their feasibility for operation in CMS for Phase-II, a prototype Sapphire sensor was installed in the BCM2L system during LS1 for additional radiation studies. Performance of the sapphire sensors during Run-II will be quantified by comparisons with pCVD diamond sensors in the same radiation environment. These prototype tests will be important for evaluation of sapphire as a viable detector material candidate for the BCML Phase-II upgrade and operation.

For BCM1L, the upgraded mechanics should be designed by taking into account constraints imposed by the Phase-II pixel tracker upgrade, since the volume will be shared for mechanics and services. In the case of BCM2L, an increase in the beam aperture for HL-LHC, implies a modification of the CT2 beam pipe geometry with one of larger radius. This imposes a modification to the BCM2L mechanical support structure.

Given the criticality of this device for CMS safety it is proposed to continue to use a very simple readout system, such as signal current, with no active components in the radiation environment. The BCML electronics should remain compatible with LHC technical standards and exploit developments of the LHC Beam Loss Monitor (BLM) readout for HL-LHC by the CERN Beam Instrumentation Group (BE/BI), on which the system will continue to be based for both the backend readout and abort functionality.

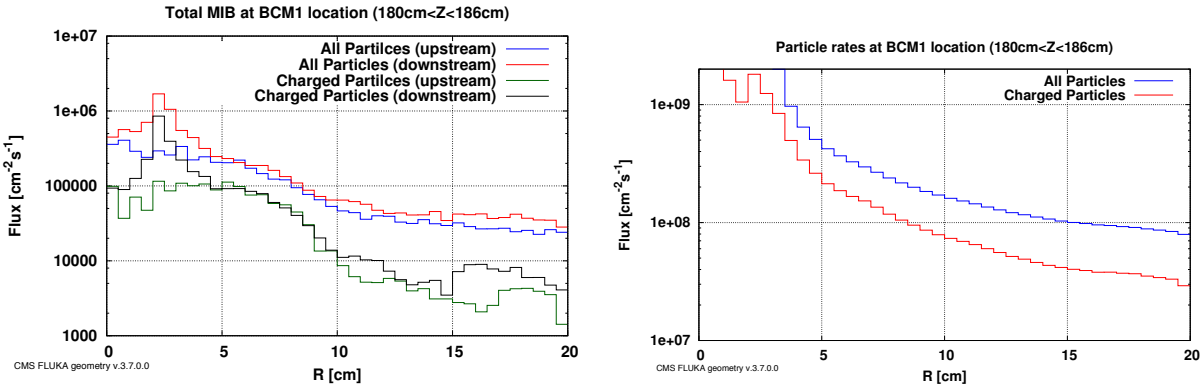
A higher signal current per channel is achievable by both increasing the active detector volume and increasing the applied electric field to the sensor. Both of these techniques will be investigated for the Phase-II detector. Work is ongoing in the LHC Beam Instrumentation group to upgrade also the frontend electronics to cope with higher radiation levels in the LHC. A reduction in noise in this frontend electronics could allow the possibility to use detectors with lower signals. The ASIC will have a larger dynamic range and will be able to accept signals of both polarities, which could be interesting for applications that foresee a potential alternating polarity HV. A realization of a frontend design with an alternating polarity HV (with about 0.1-10 Hz switching frequency) would require some R&D on the frontend electronics and services.

5.2 Machine Induced Background monitoring

Machine induced background (MIB) particles are particles entering the CMS cavern from the LHC tunnel, travelling with the beam. They are usually created in particle showers originating from beam particles interacting with residual gas atoms or the collimation system. The CMS Tracking performance, for both the Muon Systems and the Tracker, can be degraded by high MIB particle rates. Detector systems have therefore been implemented for diagnostics at both large and small radius, to be sensitive to MIB. The measurements are required to distinguish, with nanosecond timing precision, MIB particles from the the flux of prompt collision products and activation products that arrive out-of-time.

The measurements of MIB particle rates are used to identify safe-operating conditions, as defined by sensitive CMS sub-detectors. MIB particles orientated parallel to the beam line can traverse many elements of the CMS pixel detector giving rise to high occupancy in specific topologies, resulting in potentially high deadtimes for the pixel readout, and hence must be minimised. These measurements are important as MIB can be the precursor to dangerously high beam-losses. CMS may take internal actions based on these measurements, for example change the high voltage for the silicon pixel detector. CMS also sends these measurements to LHC where they are used to identify and/or diagnose degraded beam conditions and re-optimize LHC where possible; for this purpose realtime measurements with latencies on the order of seconds are required.

The simulated HL-LHC MIB particle flux arriving at the location of the LHC/CMS machine interface is shown in Fig. A.3 , as explained in Appendix A.1.7. As an example, the charged particles MIB flux after transport through CMS, is plotted at the BCM1 location ($Z = \pm 1.8\,\mathrm{m}$) in Fig 5.3a. This is the most sensitive location to monitor MIB in the pixel volume.



(a) MIB products

(b) Collision products

Figure 5.3: a) The machine induced background (MIB) flux, as simulated for a nominal HL-LHC machine, at the two BCM1 locations ($Z = \pm 1.8\,\mathrm{m}$) after transport of simulated MIB particles through the CMS detector and beampipe elements. b) Particle flux from collision products as a function of radius at the BCM1 location for an instantaneous luminosity of $5 \times 10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, as estimated using FLUKA.

The MIB particle flux for a "nominal" HL-LHC is typically many orders of magnitude below the rates of particles coming directly from collision products, as illustrated by comparing Fig. 5.3a with Fig. 5.3b. The MIB particle flux is also typically less than the flux of albedo particles. Arrival time discrimination and directionality aid in distinguishing MIB particles from collision and albedo products

- Timing: MIB particles pass through CMS almost at the same time as the proton bunches. Collision products pass through the detector together with outgoing bunches. At specific locations (Z = 1.875 m , 5.625 m , 9.375 m , 13.125 m , 16.875 m , 20.625 m) so called "golden locations", the outgoing collision products and the incoming MIB particles are maximally separated in time. For a bunch spacing of 25 ns, incoming MIB particles will arrive at z locations with flight times of $N \times 6.25$ ns, for N = 1, 3, 5..., maximally out of phase with collision products and in time with the incoming bunches. A detector with good time resolution is required to discriminate between particles arriving at 12.5 ns intervals and it can help discriminate also against asynchronous albedo particles.
- Directionality: MIB particles enter into CMS along the beam axis, moving in the same direction as the incoming beam. Collision products instead move in the opposite direction, while albedo particles do not have a preferential direction. A detector with a strong directional response can help discriminate between MIB, outgoing collision and albedo products.

There are two dedicated detector systems to measure MIB particles for CMS, namely the Beam Halo Monitor (BHM) and the Upgraded Fast Beams Conditions Monitor (BCM1F). Both these systems have been installed during LS1 and will be commissioned during the start of Run-II. The expected performance of these systems, described below, is based primarily on test beam and laboratory measurements, and some considerations from the previously installed BCM1F detector that was operational between 2008 and 2013 [154]. The Phase-I upgrade of the BCM1F will also be used for online luminosity, as explained in Section 5.4.1.1. The systems are equipped with electronics to provide an efficient MIB tag to the L1 trigger, to flag the presence of beam background events for considerations in the physics analysis.

5.2.1 MIB measurement at the entrance of the CMS cavern

The Beam Halo Monitor (BHM) [155] is a detector system using Čerenkov light signals produced in synthetic fused quartz. It has been deployed on the Rotating Shielding ($Z = \pm 20.6$ m; $R =$ 1.8 m), see Fig. 5.4 (left). At this location the detectors are sensitive to muon MIB arising from interactions in collimators in the long straight sections. The BHM system makes use of both timing and directionality to identify background, owing to its fast signal (3.0 ns FWHM) and a signal amplitude response that depends on the direction of the incoming particle, as shown in Fig. 5.4 (right). Each detector unit is orientated such that the incoming beam background from the tunnel will produce a large signal relative to collision products.

5.2.1.1 The BHM system and considerations for HL-LHC

The BHM system should be maintained to be operational in HL-LHC. This system is in principle already optimized for operation in Phase-II and beyond, although the need for some modifications may become apparent with the experience gained during Run-II. The expected radiation dose at BHM location is 0.02 Gy per fb^{-1} and all the primary components, i.e. radiator and photodetector, have been tested for irradiation up to 1 kGy, and thus are expected to be operational for a 3000 fb^{-1} run. The electronics of this system is located in the service cavern and is hence not expected to suffer from any deterioration of performance due to radiation damage whereas the photodetector may need replacement due to ageing.

There is no design at this stage for a BHM-type detector system that could exploit the directionality of Čerenkov light and be installed at lower radius for enhanced sensitivity to MIB. Potential directions to be investigated include the use of quartz fibers appropriately aligned

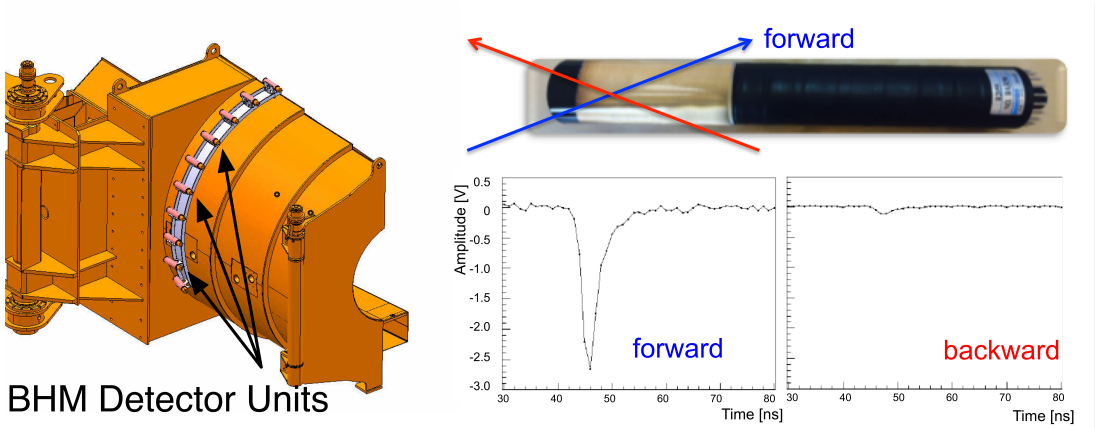


Figure 5.4: (left) The BHM system installed around CMS rotating shielding (1/4 of installation shown) during LS1. (right) The amplitude as a function of time, show the very different responses to forward and backward going particles.

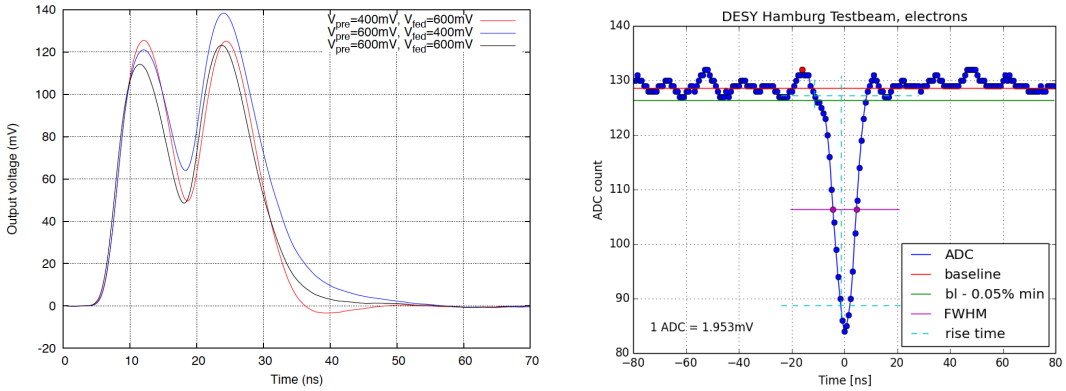
around the beam pipe to generate and capture Čerenkov light produced by MIB particles originating from the tunnel direction. Both amplitude and timing information would be combined to discriminate light produced by collision products versus MIB. For such a detector concept, the light transport and detection should be considered taking into account the magnetic field and radiation environment in the location of the photodetector. Photodetectors based on multi-anode PMTs, fine mesh PMTs or silicon photomultipliers (depending on the radiation environment) could be considered. These technologies offer high temporal and spatial resolution, both of which could be exploited for a fast and directional MIB measurement closer to the CMS detector.

5.2.2 MIB measurement close to the tracking volume of CMS

The BCM1F single-crystalline (sCVD) diamond detector is located close to the CMS pixel detector and makes use of timing characteristics to identify MIB particles by exploiting the "golden location" of 1.8 m. This system has demonstrated its crucial importance in 2011 when a non-conformity in the RF finger assembly of the vacuum chamber at Z = -18.3 m from the IP gave rise to very high deadtimes in the CMS pixel barrel frontend, and for which the BCM1F measurement gave a clear signal and diagnostic [148]. The Phase-I upgrade of the BCM1F detector installed during the LHC LS1, is especially designed for optimal performance in the LHC beam conditions up to a luminosity of $2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$. It is instrumented with 24 sCVD diamond sensors, split, metallized with two pads, each of area $0.125\,\mathrm{cm^2}$ per channel and a fast radiation hard preamplifier in 130 nm technology [156] with a FWHM of less than 10 ns and a high double-pulse resolution capability, as illustrated in Fig. 5.5a [157].

5.2.2.1 The BCM1F system and considerations for HL-LHC

Diamond detectors of sCVD type have been shown to develop radiation induced deformations of the electric field in the sensor, leading to a reduced charge collection efficiency as a function of absorbed dose. The charge collection efficiency can depend on the rate and affects the efficiency for detecting single minimum ionizing particles [151]. In Run-I the sensors were operated at 300 V ($0.6\,V/\mu m$), whereas for the Phase-I upgrade, all BCM1F components have been qualified to operate at 1000 V. This higher voltage capability for Run-II will be used and studied as a potential means for mitigating polarization effects. The efficiency of the sCVD diamond sensor material as a function of fluence at these operating parameters will be quantified



(a) (b)

Figure 5.5: a) The frontend ASIC response to two MIPs (3 fC) delayed by 12.5 ns, as would be the expected maximum timing separation between a particle produced from a collision and a particle originating from MIB in the same bunch crossing. b) Test beam of signal measured from a single electron, traversing the sCVD diamond sensor, amplified and shaped using the BCM1F Phase-I upgrade readout. The signal was digitized with an 8 bit ADC sampling at 1.25 GS/s.

in Run-II. It is expected that the sensors will require replacement for Phase-II operation. The expected integrated hadronic fluence is about $1.0 \times 10^{16}\,\mathrm{cm}^{-2}$ for the full Phase-II operation. It is expected that the sensors will probably require replacement periodically. The optical readout components of the BCM1F data chain will also exhibit radiation damage which will contribute to the loss in signal efficiency [158]. This will necessitate the upgrade and/or replacement of these components for Phase-II operation.

The Phase-II BCM1F detector granularity could be increased by a factor of ~ 3 to compensate for the higher occupancy at a luminosity of $5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, in order to remain efficient in identifying MIB signals.

Due to modifications in the tracker volume for the proposed Phase-II forward tracker extension, there is only a small volume available for the Phase-II BCM1F detector at the "golden location" of ± 1.8 m, which is needed for discrimination of MIB based on arrival time. The Phase-II BCM1F should be implemented as a very low mass, compact detector in order to avoid deterioration of the tracking performance, adding a certain amount of complexity to the design.

The backend electronics of this detector system will be demanding, due to the requirement of both dead-time free monitoring in high data rates and sub-bunch spacing timing resolution. The peak finding algorithms to detect single MIPs in the backend electronics FPGAs will need to be specially adapted to HL-LHC occupancies and timing requirements. The electronics for Phase-II is expected to adopt similar considerations as in Phase-I [159] which are based on μTCA based electronics standards. The BCM1F electronics makes use of deadtime free, trigger-less readout. This expected hit rate for the Phase-I detector system is about 300 MHz for a pileup of about 50. The scaling assumption for the Phase-II detector system and electronics is to maintain a hit rate of about 300 MHz for a nominal beam pileup of 140. Overlapping peaks, as shown in Fig. 5.5a with a separation of 12.5 ns must be still recognised for HL-LHC using fast sampling ADC's processed by peak finding algorithms implemented in onboard FPGAs.

The Phase-I BCM1F detector only makes use of timing to suppress hits from collision products.

A new detector concept for Phase-II could investigate methods to exploit the directionality of incoming MIB particles compared to outgoing collision products for additional discriminating power to identify MIB. This requires R&D, as discussed in Section 5.2.1.1.

The implementation of a BCM2F system (i.e. a fast, MIP sensitive detector in a similar location as the BCM2L systems), will also be investigated. This could offer a redundancy in the low radius machine induced background measurement, or an alternate location to the present BCM1F location at ± 1.8 m. The present BCM2L location offers the advantage of more frequent access possibilities, but it is not a "golden location", and hence is less efficient in identifying MIB particles from collision products based on timing. Identifying easily accessible "golden locations", where one could potentially install an additional low radius Phase-II MIB detector system, still requires study. This will be done taking into consideration the CMS forward integration constraints. A longitudinal position of more than 7.3 m from the IP, is interesting as this would allow for a clear measurement of the MIB associated to the first incoming bunch in a bunch train, prior to the arrival of outgoing collision products.

The choices between the MIB particle detector upgrade options will be made subsequent to further LHC and BRIL MIB simulation studies, Run-II experience and a better understanding of CMS integration constraints at other "golden locations". For all forward or low radius locations, radiation hardness will be of paramount importance and will certainly drive the final detector design and R&D activities.

5.3 Radiation Monitoring

The radiation fields of various electromagnetic and hadronic particles, of various energy regimes, in the CMS detector and cavern should be monitored during both LHC operation and after beam dump periods. Estimating the online residual ambient dose rates helps to optimise access periods, minimising the dose to personnel. The measured fields are used to benchmark radiation simulations and optimise the shielding design of the cavern.

Systems to measure the fluence of particles in the CMS cavern and detectors with a time scales between months and nanoseconds are important for feedback on the radiation environment and for a measure of accumulated dose. Passive and active dosimetry is installed throughout the CMS cavern to monitor the dose to equipment and to compare measurements with expectation from radiation simulation models. During HL-LHC operation, it will be important to monitor in particular the neutron fluence in specific energy ranges to understand possible sources of radiation damage to electronics, of both instantaneous and cumulative nature. Single event upsets (SEU), for example a memory bit flip, can result in a temporary functional failure of electronics due to both a low and high energy hardron fluence and thermal neutrons. Cumulative effects, due to an increased total ionising dose or displacement damage to the atomic lattice of materials used in the sensors and electronics can result in a degradation of performance over time.

5.3.1 Radiation field monitoring for the CMS detector and Experimental Cavern

Neutron monitoring can be implemented using the Neutron RADMON [160] detectors based on a thermal neutron detector with a moderating shell, optimised for the expected radiation environment and neutron energies for HL-LHC. This detector has good sensitivity to the evaporation part of the neutron spectrum (0.5 -15 MeV). Neutron RADMON can work in magnetic field and in the neutron fluence range from 0.01 up to 10^5 $cm^{-2}s^{-1}$. The stationing of Neutron RADMON in the CMS cavern, monitors online the main component of the scattered radiation

field. This is used to estimate the accumulated dose to the electronics in the cavern. The Neutron RADMON sensitivity function can be tuned by the choice of the moderator structure and material. Deploying monitors with different moderators allows an estimation of the neutron spectrum, at the monitor locations.

Neutron RADMONs will be installed in critical points in the CMS cavern, for example at the entrance of corridors to the underground experimental cavern and near the racks with electronics. Monitoring of the radiation field around the CMS detector and the Forward Shielding allows a crosscheck with radiation field simulations.

Mixed radiation field detectors can be built with hybrid pixelated sensors. Such detectors are able discriminate between major particle types. Particle track classification makes it possible to separate particle types comparing a characteristic pattern for a cluster of pixels [161]. To measure neutrons, the sensors are in addition coated with a very thin conversion layer. The conversion reactions differ for thermal, intermediate and fast neutrons and thus the coatings have to be adapted to the required environment [162]. Development of such detectors using Medipix3 [163] are being considered for HL-LHC. Time-resolved measurements of the mixed radiation field also can be performed with this technology, taking advantage of the Timepix CMOS ASIC [164], which supports additional modes of operation: Time-of-Arrival and Time-over-Threshold. Gated measurements can also be implemented, which could offer an interesting handle on studying the radiation environment during particular time intervals. A prototype system based on Medipix3 will be installed in CMS in 2016. Experience in operating and exploiting this detector system will be important for optimisation for HL-LHC requirements and for determining whether modifications in radiation hardness of the readout electronics are required for Phase-II.

5.4 Luminosity Measurement

Luminosity is a key parameter for an LHC experiment since a component of the physics program consists of precision cross-section measurements that necessitate an accurate luminosity determination. A sound understanding of the luminosity measurement and its uncertainties is therefore critical, especially for the study of physics processes where the luminosity is the dominant error contribution to the measurement.

In addition to the requirement for physics normalisation, an online luminosity measurement is needed for the tuning of the machine parameters to optimise the running conditions for CMS, for example in the establishment of target pileup values. Latencies in the calculation of the online luminosity of order one second are desirable since they are used in the LHC machine luminosity optimisation beam tuning algorithms, together with a small statistical uncertainties of order of 1%. Owing to the need for an accurate accounting of the dose delivered to the CMS detector an online luminosity measurement needs to be operational whenever there is beam in the machine. A redundant online luminosity measurement system is prudent in order to guarantee feedback to the LHC, even in the case when some CMS detector services are non-functional, and continuous accounting of the luminosity should one device fail. The use of more than one luminosity measurement technique also provides cross checks for an enhanced evaluation of the systematic error.

The measured rates in the luminosity detectors are calibrated to an absolute luminosity scale using the Van der Meer (VDM) scan technique [165] at low pileup, where the transverse size of the beam is increased, compared to standard operation, for a better measurement resolution of the transverse beam overlap and to reduce collective beam-beam effects that can cause

time dependent systematic effects in the measurement [166]. The calibration value calculated during the low luminosity VDM scan is then extrapolated to high luminosity operation, at a significantly higher luminosity than the conditions at which the scan was taken. The luminosity detector should therefore exhibit a linear response to particle rates over a wide range of luminosity operating conditions. Since VDM scans require dedicated beam conditions, calibrations are performed only a few times per year. This puts also a stringent requirement on the intrinsic long term stability of the internal calibration of the luminosity detectors themselves, and necessitate that the detector stability is monitored continuously and the stability of each luminometer (offline and online methods) is compared against another.

In Run-I, the HF tower occupancy method was used to perform high statistics bunch-by-bunch measurements using a dedicated readout that was available in real time and available independent of the CMS central DAQ. A second method, based on offline pixel cluster counting, was used to provide highly linear and stable luminosity values for offline analysis [167–169]. The BCM1F system was used as a backup relative luminometer for the rare occasions when HF measurements were unavailable [170].

Long term efficiency variations are also monitored by making use of processes with clean experimental signatures such as $Z \to \mu\mu$ [171], detailed comparisons of the rates for these processes and of the rates in the various luminometers are used to identify and correct for efficiency and similar effects and to estimate the final systematic uncertainty in the luminosity measurements.

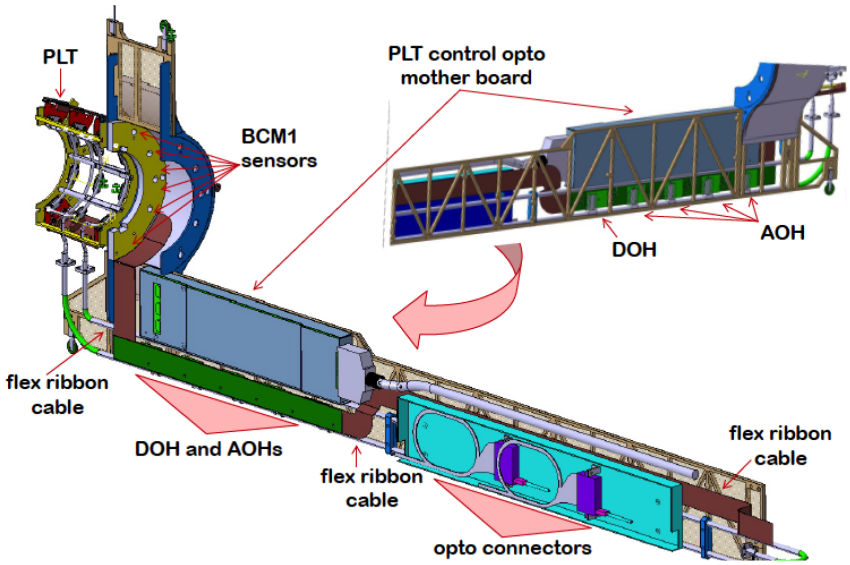


Figure 5.6: Drawings of the PLT and BCM1 Phase-I mechanics, 1/4 of the detector system is shown, installed at a Z location of ± 1.8 m from the IP, inside the pixel support tube. The inner radius of the central part of the detector is at about 43 mm, extending in radius of about 76 mm.

5.4.1 Online luminosity

5.4.1.1 Considerations for CMS Online Luminosity Systems for HL-LHC

During LS1, in the context of the CMS Phase-I upgrades, two precision online luminometers based on different technologies, to supplement the HF measurement, have been installed on both ends of CMS to provide the experiment with a redundant online luminosity system. The first, namely Pixel Luminosity Telescope (PLT) [172], is a dedicated set of 16 telescopes,

comprising of 3 planes. Each telescope is equipped with standard CMS Silicon Pixel sensors (100×150 μm^2 pixel area) equipped with a 40 MHz Fast-Or readout to give track counting on a bunch by bunch basis, inside the pixel tracker volume. The PLT will also provide a quasi-online measurement of the beam spot during VDM scans and measurement of the MIB particles, by using a prompt reconstruction of tracks originating from direct collision products and parallel background tracks. The second system is the upgraded BCM1F system [158], capable of identifying signals with about 12 ns separation. The PLT and BCM1 are installed on a combined carbon fiber carriage mechanics, as shown in Fig. 5.6, inside the pixel support tube.

The system already used prior to LS1 is the HF. During Run-II, for the HF luminosity algorithms will be quantified for high rate operation.

The performance of BCM1F will also be quantified for high pileup conditions in Run-II. The BCM1F backend electronics allows for sophisticated signal processing to identify multiple pulses in consecutive and also the same bunch crossing and resources for cross-channel coincidence logic. These algorithms will be necessary, since the single sensor occupancy for detecting a single hit above threshold in a simple level discriminator will require non-linear corrections at higher pileup. As for HF, an upper limit in terms of pileup for all BCM1F algorithms for luminosity is still to be quantified, but it is unlikely to be an optimally performant detector for HL-LHC, with its present sensor dimensions. It is proposed to reduce the BCM1F sensor size by a factor of ~ 3 for Phase-II operation, to have the performance of future detector at a pileup of 140, comparable with the present detector at a pileup of 50. Radiation damage is expected to the BCM1F sensors and analogue optical readout, which will be monitored and quantified during Run-II, and require replacement for Phase-II.

The longevity of the silicon PLT detector is about 500 fb^{-1}, after which significant loss in charge collection efficiency will occur, requiring a replacement of the sensors. The PLT detector in its present form, is not a candidate for HL-LHC operation and would need to be replaced.

The extrapolation from VDM intensities to LHC operating intensities will be further extended by a factor ten in going to HL-LHC compared to Run-I/Run-II operation, thus placing even more stringent requirements on the linearity of this extrapolation. Significant contributions to the systematic errors on the eventual online and offline luminosity arise from non-linear effects due to detector pileup/saturation, activation of the detector or surrounding materials, late arriving radiation and the suchlike. These are strong motivations to ensure that CMS has more than one type of technology used in its luminometers and in particular the implementation of both a calorimetric and a tracking type luminometer will remain our baseline

5.4.1.2 Proposal for Online Luminosity System Upgrade

During HL-LHC, at least one luminosity measurement system is required to be fully operational whenever there is beam in the machine; to account for all delivered luminosity, including periods of LHC machine development. This will constrain at least one system to be fully operational in all beam modes, and designed to take data independent of the status of the CMS DAQ. One online luminosity system should also be robust against interruptions to CMS services, like cooling and power. The online luminosity measurement system must also be able to feed the LHC luminosity-levelling algorithms with stable and well behaved measurements at a frequency of about 1 Hz.

We expect to continue the use of the online HF readout as one of the online luminosity measurements. Although the current system exhibits substantial non-linearities and is subject to modest calibration drifts in time, experience has shown that it may be possible to derive cor-

rections that result in acceptable accuracy for online operation. While the tower-occupancy method will eventually saturate, early Monte Carlo studies indicated that the E_T-sum method may offer the possibility to run at higher pileup. Further studies in Run-II are required to quantify the systematic uncertainties of the non-linearities that may be expected, and the use of this type of luminosity measurement in Phase-II. Depending on the final adopted upgrade options for other forward Calorimeters from CMS, it may be interesting and cost-effective to implement luminosity measuring front ends for these calorimeters.

The physical volume currently used by the Phase-I upgraded BCM1F and PLT detectors, is proposed to be incorporated and used as an active Phase-II forward tracking volume. Some integration considerations in the redesign of the mechanics, as shown in Fig. 5.6, and services will therefore be mandatory for these detectors for HL-LHC. Considerations for an upgraded BCM1F detector at $\pm$ 1.8 m is discussed in Section 5.2.2.1.

All systems must be designed to allow relatively straightforward access and/or replacement as a way to compensate for damage at very high integrated luminosities.

Given the successes of the pixel cluster counting methods for offline luminosity measurement in Run-I and having good reason to expect that a cooled si-PLT will provide a stable per-bunch online luminosity measurement for Run-II, a strong candidate for Phase-II online luminometry would be to exploit this type of pixel cluster measurement and/or PLT-like algorithms using the Phase-II hardware possibilities.

We propose to take advantage of radiation-hard sensor R&D and mechanical synergies with the forward pixel upgrade with the implementation of 3 special forward pixel-planes, though with partial phi coverage, in addition to the most forward pixel disks. These planes should be "on" before stable-beams. They would be read out independently of the standard pixel tracker. These planes should be triggered either by a dedicated set of random luminosity triggers O(100 kHz) range, sufficient for a good statistical sample, or alternately/additionally the readout chip could be implemented with a free-running 40 MHz readout using a FAST-OR, as in the case of the PLT. Both these options have the advantage of complete independence from the standard tracker. Our baseline costing will be built entirely on standard FPIX modules, extensions to that baseline may make use of different detector technologies subsequent to successful R&D.

The FAST-OR option, giving the abilty to do some tracking at 40MHz in the same way as with the current PLT would be a valuable addition to the future pixel front end chip. This offers a triggerless ability to get deadtimeless luminosity from every bunch and every crossing. If incorporated at an early stage as an option in the design it should cost very little yet offer a real advantage in the luminosity measurement.

The choices of luminometer upgrades will be made subsequent to initial Run-II experience that will allow benchmarking of the Phase-I detectors performance and extrapolations of this performance to the conditions expected in HL-LHC. Design feasibility studies are required to further quantify the expected performance, operational advantages, design challenges and costs of these upgrades.

5.4.2 Offline Luminosity

On the assumption that the mean hit rate in the Phase-II pixel detector is up to 2 GHz cm^{-2} for a pileup of 140 with a pixel size of 50 $\times$ 50 μm^2 at the lowest radius position, the fraction of pixels hit per bunch crossing is about 0.1%. Given this low occupancy, the probability of overlapping hits is negligible and the offline pixel cluster counting algorithm as used in Run-I,

can be applied, as the occupancy conditions are comparable to the present LHC situation. The proposed online systems described in Section 5.4.1.2 can also yield the possibility to measure pixel hit and cluster rates at 40 MHz potentially to augment the offline measurement of pixel counting.

5.5 Online Timing and Longitudinal Interaction Point

Located at $\pm 175\,m$ from the CMS interaction point (IP5), the LHC is equipped with 4 button beam position monitor pickup electrodes (2 in the horizontal and 2 in the vertical planes). The signals from these four electrodes are combined to be insensitive to the beam position and used as a reference timing signal for each beam. These reference are used as a measurement of the Z location of the IP, by the difference in arrival times between beam 1 and beam 2 and also as a measurement of the phase of the beams at CMS with respect to the CMS distributed clock. The signals from this detector can also be exploited to measure the bunch fill pattern, the bunch intensity and mean bunch length, providing important independent measurements to CMS on the beam conditions. The system has also a requirement to deliver to the CMS L1 trigger the beam presence, for each beam.

The functionality of the presently installed system and backend electronics has to be maintained for HL-LHC, and in the case of the input signals to the L1 trigger, adapted to the requirements of the upgraded L1 trigger receiver hardware.

5.6 Project planning and cost estimates

5.6.1 Project Organization

The Phase-II BRIL upgrade presented in this document has been developed by the BRIL Project team and will be incorporated into the BRIL project.

A top-level management team composed of the BRIL Project Manager (PM), that is proposed by the BRIL Institution Board, and his or her Deputy (DPM) oversees the maintenance and operation of the present systems and the preparation of the upgrades. The PM and DPM are supported by a Level-2 management team in the area of the Run Coordination and Operations, Technical Field Coordination, Detector Performance Studies and Radiation Simulation. The BRIL Project Management reports to the BRIL Institution Board. The BRIL Resource Board oversees the BRIL budgets and reports to the BRIL Institution Board. The BRIL Editorial Board and Conference Committee oversee the preparation of BRIL-specific papers and participation in conferences.

The BRIL project management will appoint, from the R&D team, a Level-2 Upgrade Coordinator in 2016 with responsibility of overseeing a robust R&D program needed to match the BRIL milestones including the preparation of the Phase-II BRIL Technical Design Report for 2020.

5.6.2 Schedule and Next Steps

The installation of the BRIL detectors for Phase-II is expected in late 2024. Sufficient contingency in the schedule is reserved for thorough testing of the assembled detectors prior to the installation in CMS and subsequent commissioning in order to insure optimal performance at the restart of accelerator operation.

A view of the conceptual project schedule is shown in Fig. 5.7.

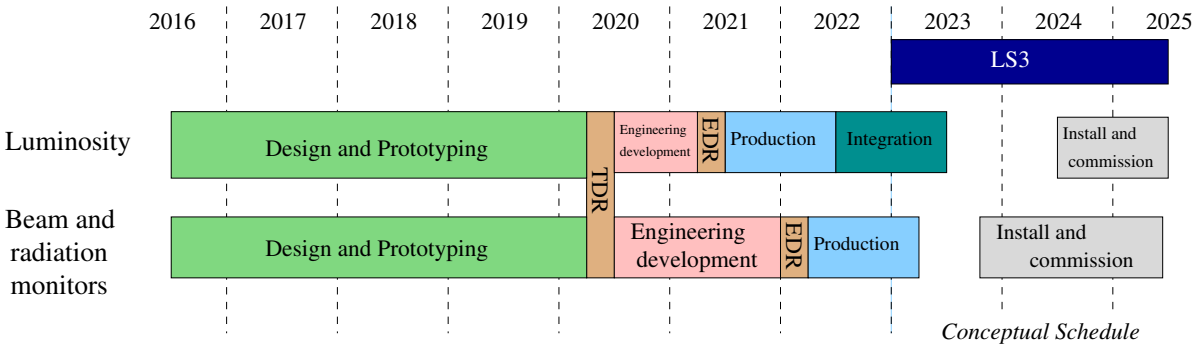


Figure 5.7: Conceptual schedule for the BRIL Phase-II detector upgrades.

Table 5.3: Planned next steps for the upgrades of the BRIL project sub-systems. The target dates are established for the purpose of project planning. Actual dates can be several months later in all cases without affecting technical progress.

Sub-system	Milestone Identifier	Target date	Description
Luminosity	TP.BI.LD.1	Q2 2017	Concept for luminosity detector
	TP.BI.LD.2	Q4 2019	Demonstration of concept
Machine - Induced Background	TP.BI.MB.1	Q4 2017	Choice of sensor technology
	TP.BI.MB.2	Q1 2018	Specification of frontend ASIC
	TP.BI.MB.3	Q4 2019	Demonstration of module
Beam Monitoring for Protection	TP.BI.BP.1	Q1 2019	Choice of sensor technology
	TP.BI.BP.2	Q1 2019	Design of detector modules and mechanics
	TP.BI.BP.3	Q3 2019	Demonstration of module
Radiation Monitoring	TP.BI.RM.1	Q1 2019	Concept for Radiation Monitor technology
Common	TP.BI.CM.1	Q2 2020	Submission of Technical Design Report

The main milestones for the years 2015 – 2020 are listed in Table 5.3 for the various BRIL subsystems, and show the preparatory steps for the BRIL Technical Design Report. The date for the submission of the Technical Design Report depends on both the anticipated R&D time needed to optimise the detector technology choices and the foreseen time estimated for production, assembly and testing prior to installation in LS3. The milestones shown are planned to cover the developments that need to be done before the Engineering Design Reviews (EDR) that launch production.

For the luminosity detector, time is foreseen to gain experience in exploiting the CMS PLT for luminosity measurements in LHC Run II, and in understanding its performance at high rates. This operational experience, together with considerations for the Phase-II Pixel and Tracker detector upgrade, will be used in developing a conceptual design for the Phase-II luminosity detector (TP.BI.LD.1). It is planned to identify and exploit wherever possible common developments with the Phase-II Pixel and Outer Tracker detector upgrade for the sensors, frontend readout system and electronics. Adaptations in the readout, triggering and data processing will be necessary for the luminosity measurement. The demonstration of the detector concept should be made (TP.BI.LD.2) prior to the submission of the Technical Design Report (TP.BI.CM.1).

For the detector optimised for MIB measurements, time is foreseen to gain experience in exploiting the present BCM1F Phase-I Upgrade for beam background measurements in high rates. R&D is foreseen in radiation hard sensors and radiation hard readout ASICs with excellent time resolution. Synergies in frontend ASIC design with other CMS upgrade projects and with projects outside of CMS will be explored. The optimal sensor choice (TP.BI.MB.1) and frontend electronics specification (TP.BI.MB.2) form the basis for the conceptual design of the detector. Prior to the submission of the Technical Design Report (TP.BI.CM.1), a functioning detector module will be demonstrated (TP.BI.MB.3).

The BCML detectors, optimised to protect CMS from dangerous beam losses, are located in regions of extremely high fluence, close to the beam pipe. R&D is foreseen in radiation hard sensors which are compatible with the backend electronics readout used in the HL-LHC machine protection system. The technology for the sensors will be proposed (TP.BI.BP.1) and supporting mechanics that is compatible with both the Phase-II beam pipe and the mechanics of the Phase-II Pixel detector will be designed (TP.BI.BP.2). A detector module will be demonstrated (TP.BI.BP.3) prior to the submission of the Technical Design Report.

The concept for detectors to monitor the neutron and mixed-field radiation in the CMS cavern will be proposed (TP.BI.RM.1), based on experience with the HF Neutron RADMON and Medipix/Timepix sensors installed for the BRIL Phase-I Upgrade and their potential adaptation to the needs of HL-LHC cavern environment.

5.6.3 Cost estimates

The active protection system has been costed based on considerations on the expected lifetime of the pCVD diamond of 6×10^{15} 24 GeV proton equivalent fluence. Included in the cost estimate is the installation of 2 full sets of BCM1L sensors and of 5 sets of BCM2L sensors in the lifetime of HL-LHC; the difference being due to the different radiation environments in these two locations.

The BCM1F system, like the BCM1L, will need 2 replacements during HL-LHC. Diamond replacement costs for the BCML and BCMF systems are based on current industrial quotations. The frontend costs are based on recent experience with the Phase-I BCM1F Upgrade. The backend costs are based on recent experience with μTCA systems.

Neutron and mixed radiation detector cost estimates are based on the HF Neutron RADMON experience and on the expected upgrades of the Medipix/Timepix readout chips for hybrid pixelated detectors.

It is assumed for the cost estimate that the pixel based luminometer makes use of standard Phase-II Pixel components with a dedicated Luminosity trigger/backend. The costing of this is based on the same methods used in the Phase-II Pixel upgrade. The proposed luminometer represents approximately 3% of additional coverage compared the the total Phase-II pixel surface and this is used to estimate the frontend costs. There are in addition other costs that don't scale linearly with detector coverage, due to the quantization of mechanics, power supplies, services and backend electronics boards; all these are needed in order to have an independent powering system and DAQ for the luminosity measurement functionality. One full replacement of the luminosity frontend over the HL-LHC lifetime is costed.

A summary of the present cost estimates is shown in Table 5.4.

Table 5.4: Cost estimates for the Phase-II Upgrades of the BRIL systems.

Estimated CORE cost in MCHF (2014)	
Beam Monitoring for Protection	0.7
Machine-Induced Background	1.1
Radiation Monitoring	0.3
Luminosity Measurement	1.9
Total Beam Monitors and Luminosity	**4.0**

Chapter 6

Trigger: Level 1 and HLT

6.1 Introduction

As is presented in the physics part of this TP (Chapter 10), the central mission of the CMS HL-LHC physics program will include precise measurement of the properties of the recently discovered Higgs boson and continued searches for New Physics beyond the Standard Model. Higher luminosity will extend the discovery mass reach, allow more sensitive searches for signatures for new physics, and enable studies of any newly found particles and their interactions.

Therefore, a key goal of the Phase-II upgrade will be to maintain the overall physics acceptance under the challenging HL-LHC conditions. CMS must retain its capability to efficiently trigger events originating from low-mass physics processes (e.g. Higgs production at 125 GeV) and for performing precision measurements of low to medium p_{T} physics objects, leptons (e, μ, τ), photons, jets (including b tagging), and Missing Transverse Energy, as well as for the reconstruction of complex event topologies to identify low to medium p_{T} Ws, top quarks, VBF jets, etc.

To achieve this, a number of major Trigger upgrades are foreseen during the third Long Shutdown (LS3). Besides the addition of a Tracker Trigger at L1, which is an integral part of the design of the new silicon tracker, CMS also plans an upgrade to the detector readout, L1 trigger and HLT systems, which would allow up to 750 kHz L1 read-out (compared to the current 100 kHz), up to 12.5 μsec L1 Latency (compared to the current 4 μsec) and up to 7.5 kHz permanent event storage rate (compared to the current 0.5 to 1 kHz).

These upgrades are motivated by dedicated L1 menu studies, which are explained in detail in Section 6.5. Using L1 trigger algorithms designed for the Phase-I detector, based on the Level-1 TDR[7], we estimate that, for beam conditions of 140 PU, at least 1500 kHz of L1 acceptance rate would be required to maintain the same physics acceptance as outlined for Phase-I. For an environment of 200 PU the same L1 menu would require almost 4000 kHz. Such bandwidth are beyond the technical feasibility of the CMS upgrades. However, these studies also demonstrate that adding tracking information to L1 trigger objects substantially reduces these rates to about 260(500) kHz for the same beam conditions of 140(200) PU. These estimates do not include any uncertainties such as simulation imperfections, limitation in releasing the L1 track based triggers in hardware as well as the ability of the individual subsystems to provide the required L1 rate. The present estimates are also only representative of past operation experience and studies are ongoing to estimate if key signals accessible only at the HL-LHC would benefit form larger acceptance rates. We estimate that accounting for these uncertainties requires a margin of 50% in the total L1 acceptance rate. Therefore, we expect that the addition of a Track Trigger at L1 in conjunction with a L1 bandwidth increase up to 750 kHz will enable CMS to maintain its Phase-I physics acceptance up to 200 PU, with sufficient margin to adapt to experimental

and physics uncertainties.

The L1 latency increase up to 12.5 μsec will provide sufficient time for the L1 Track Trigger, which requires hardware track reconstruction and matching with the muon system and calorimeter information.

With the L1 acceptance rates proposed above, and assuming that about a 1/100 event selection is maintained from L1 to HLT output, a permanent event storage of 5(7.5) kHz rate will be required to operate at 140(200) PU.

In the following sections of this chapter details about the Phase-II Trigger upgrades and how they help to meet the physics requirements of CMS at the HL-LHC are provided and discussed. This chapter is structured to provide an outline of a first version of the Phase-II Trigger Architecture in Section 6.2. Section 6.4 provide details about design and performance of the foreseen Track Trigger at L1. A first outline of a L1 menu for different assumptions of L1 bandwidth and trigger object performance is provided in Section 6.5 and the corresponding HLT and DAQ requirements are discussed in Section 6.6. An overview of a first version of the hardware design of the Phase-II Trigger is given in Section 6.7. The chapter is completed with an overall cost estimate of the Phase-II Trigger project in Section 6.8.

6.2 Trigger Architecture

6.2.1 Incorporation of Tracking at L1

As described in Sect. 2.3.2.5 the tracker will provide L1 tracks, built from the outer tracker information, for each bunch crossing. These tracks, parameterized by five helix track parameters, will be used in the L1 trigger to form the trigger decision by combining the tracks with information delivered by the muon and calorimeter trigger system to:

- better identify charged leptons (e, μ, τ);
- improve the p_{T} determination of charged leptons, in particular that of muon candidates;
- determine the isolation of leptons and photons with respect to the neighboring tracks;
- determine the "vertex" of charged leptons and jet objects;
- determine an event primary vertex and the transverse missing energy carried by L1Tracks that come from this vertex.

With this, the Global Trigger will have several new handles to increase its selectivity. It will be possible to improve the selection of isolated leptons and photons by imposing a tracker-based isolation requirement - the latter being more robust with respect to pileup than isolation criteria that rely solely on calorimeter information. Multi-object triggers can also be made more robust by imposing that the objects come from the same interaction, exploiting the z-vertex of the objects. Hadronic triggers that require that the scalar or vector sum of the transverse energies of jets (HT or MHT) be above a given threshold will also be made more robust by restricting these sums to the jets that come from a common vertex. More details are given in Section 6.4.

To achieve sufficient performance the proposed properties for the L1 tracking are the following:

- Highest possible efficiency, over the full η range, for isolated high p_{T} tracks which are used to form lepton (e, μ, and τ) triggers.
- A p_{T} threshold of $\sim$2 GeV, to ensure high efficiency for tracks in jets, for vertex identification and for track based isolation.

- A good p_T resolution, primarily for low transverse momentum muons at threshold.
- A vertex z position resolution of about 1 mm, to allow efficient selection of tracks that come from the same vertex position within a window of 5 to 10 mm.
- Track quality criteria (eg. number of stubs used in track fit and the χ^2) for use in calculating inclusive quantities, such as track based missing transverse momentum.

The studies of the L1 track formation, as discussed in chapter 2, indicate that performance at the level described above should be feasible. In an average minimum bias event at $\sqrt{s} = 14$ TeV, about 30 charged particles are produced within the acceptance of the tracker, $|\eta| < 2.4$. However, only about 2.7% of these particles have $p_T > 2\,\text{GeV}/c$. A bunch crossing with $\langle \text{PU} \rangle = 140$ will therefore contains about 115 tracks have $p_T > 2$ GeV/c. To account for fluctuations and possibly higher average PU we assume that up to 300 tracks per bunch crossing in the trigger. Retaining the full precision from the track finding and track fitting, will require ~ 100 bits corresponding to a bandwidth of about 1200 Gbit/s to transfer the track data from the track finding to the down stream processing. It has to be noted that the data volume could be reduced for trigger algorithms that only requires high p_T tracks, for example for $p_T > 5\,\text{GeV}$ the number of tracks decreases by a factor of 5.

In addition to the track trigger, based on the outer tracker information, we are conducting simulation studies to evaluate possible benefit of using the pixel detector. This could further improve the trigger objects association to a same vertex and the electron efficiency, and also allow reconstructing secondary vertices for b-tagging identification. Although this would significantly complicate the readout of the pixel system, a possible implementation could consist in reading out, at few hundreds kHz, small regions of the detector defined by the the matching of the L1 track and calorimeter information.

6.2.2 Impact of trigger upgrades on the sub-detector electronics

Provision of tracking information for the L1 trigger will require an increase in L1 Latency for the tracking trigger calculations, the combination with the calorimeter and muon trigger information and the increased algorithm complexity. The L1 track trigger information is estimated to be available for processing by the trigger system approximately 5 μsec after the interaction occurs. The tracks are then processed to find the primary vertex, and to be associated with this vertex. Then, they are matched with objects found by the calorimeter trigger and fitted with tracks found by the muon trigger. These associations and fits are used to calculate track-correlated L1 object and their characteristics. Finally, the tracks are used to calculate the isolation of the found muon and calorimeter objects. These five steps correspond to the regional processing stage, each is estimated to take at least 0.5 μsec, yielding 2.5 μsec of processing time, finishing at least 7.5 μsec after the interaction occurred. The next stage of processing, or global processing, consists of global sums, kinematic calculations, correlations between trigger objects, and trigger decision logic (including trigger rules), which are also estimated to take another 1 μsec of processing time, finishing at least 8.5 μsec after the interaction occurred. The propagation of the L1 decision then takes another 1 μsec, arriving at the front ends at least 9.5 μsec after the interaction occurred. This estimate needs a safety factor of at least 30%, the amount by which the original CMS trigger exceeded its target latency. Since a significant part of this safety factor could be absorbed by the uncertainty in the L1 track trigger latency alone, we set the minimum design latency for the Phase-II L1 trigger to 12.5 μsec. It has to be noted that this value is only a limit for the present outer tracker electronics design, contrary to the other sub-detectors. It allows minimizing power consumptions in the front-end electronics but it could be increased in case a higher latency would be technically required or reveal to be beneficial for performance

of the trigger.

The overall impact of the new trigger specifications on the modification of the sub-detector electronics is summarized in the following sections. Mostly, the modifications of the on-detector electronics allows at the same time to eliminate latency and L1 acceptance rate limitations. The specific upgrades involved by the higher rates and data volumes on the DAQ and on the online and offline computing are discussed in chapters 7 and 8.

6.2.2.1 Tracker

The Tracker front-end chips store the full information collected at each bunch crossing in pipeline buffers , which is retrieved and read out upon reception of a L1 trigger signal. In addition, the Phase-II Tracker will also provide tracking information at each bunch crossing for the Level-1 trigger decision. As described in the Tracking section, The Tracker front-end electronics selects signals of particles with momentum above a threshold of approximately 2 GeV, and a Level-1 track finding layer in the back-end electronics reconstructs the tracks and calculates the helix parameters.

6.2.2.2 ECAL

The present ECAL electronics, as described in section 3.2 has a maximum latency of 6.4 μs and approximately 150 kHz of level 1 acceptance rate. To ensure sufficient latency for the new trigger, the ECAL electronics will need to be upgraded. Since the EE and ES will be entirely replaced, this will require modification of the EB electronics only. The concept for these new electronics is to read out single crystal information at 40 MHz, moving all trigger primitive generation and buffers off-detector. This will allow the system to operate without on detector L1 acceptance rate limitation and the off-detector data buffers will provide sufficient latency for the new trigger. In addition, this upgrade will provide the highest crystal granularity information at L1 for better precision in the association to the tracker information, and therefore improved identification and isolation of the electromagnetic objects. Eventually, improved suppression of Hadron-induced anomalous signals in the EB APDs will also be built in the new front-end electronics.

6.2.2.3 HCAL

The HCAL Phase-I upgrade foresees that by the end of LS2 the entire HCAL back-end electronics will be μTCA-based. For the existing design of the μTCA Phase-I hardware, we can make the following considerations about the HL-LHC L1 trigger latency and rate. HCAL performs continuous (40 MHz) readout from the detector front-end to the back-end, where the data is inserted in a pipeline awaiting for the trigger decision. A latency of 12.5μs and beyond is within the resources of the Phase-I electronics. The L1 acceptance rate, under reasonable assumptions for the number of readout samples and the zero-suppression level, is limited to about 250 kHz by the available bandwidth of the links toward central DAQ. The foreseen rate of 750 kHz could be achieved by a redesign and replacement of the back-end cards hosting such links. However, since EE and HE will be replaced for Phase-II and the front- and back-end of EB will be redone, a more-sensible scenario foresees that the entire new endcap calorimeter and both EB and HB would have newly designed back-end electronics systems. In this scenario, the existing Phase-I μTCA HCAL back-end electronics would be fully devoted to HO and HF. With the increased availability of μTCA components for HF and HO, achieving a 750 kHz rate will become possible.

6.2.2.4 Muons

In the DT subdetector, the electronics attached to the chambers inside the Minicrates, the Readout Boards (ROBs) will require modifications in order to be able to sustain rates beyond 300 kHz, irrespective of the L1 trigger latency. There are a total of 1500 Readout Boards (ROBs) in the DT system which are based on the CERNs ASIC HPTDC (High Performance Time to Digital Converter) that digitize the time of arrival of the chamber signals, store them in memories and perform a matching when the trigger acceptance signal arrives. ROB boards contain 4 HPTDCs which share a common bus to send out the information to the next level by means of a 240 Mbps copper link. HPTDCs buffer sizes should be sufficient to ensure a latency well beyond 12.5 μsec; however, the readout protocol is a bottleneck for increasing the trigger rate. The allowed limit varies significantly with the number of hits per event which depend mostly on the background distribution over the detector. Present extrapolations give approximately 4-10 hits per event per ROB, imposing trigger rate limits at 300 kHz, requiring modifications to stand higher trigger rates. With the above modifications, there is no concern if latencies and/or L1 rates would need further increase.

For the RPCs, the present system is not capable of working beyond 200 kHz L1 rate, but unlike other systems, all the limitations are restricted to the off-detector electronics in USC. Since most of these electronics will be replaced already during Phase-I Trigger upgrade, it is expected that the readout performance will significantly improve, removing critical bottlenecks. The final acceptable rate will also depend on how the interface to the DAQ system is performed which may also change by the time of Phase-II, so in principle the modifications if any, required, should be relatively small.

The present CSC readout and trigger system can operate at a maximum of 400 kHz acceptance rate and 6 μs latency without a significant loss in either readout or trigger efficiency. Beyond this, the main limitation comes from the Cathode Front-End Board (CFEB) placed on the CSC chambers located in the innermost endcap rings (η >1.55). The CFEB stores data in an analog Switched Capacitor Array (SCA) containing 96 buffers. When a CSC trigger primitive consisting of a 3-layer coincidence is found, 8 of the SCA buffers are used up until the trigger acceptance signal is received. At high rate and long latency it can happen that all 96 buffers are used up and no further information can be stored in the CFEB. An additional limitation comes from the copper cables used for output transmission of CFEB data at 80 MHz. As discussed in detail in Section 4.2.4.2, the replacement of the CFEBs on 108 out of a total of 468 CSCs with digital CFEBs, similar in design to those installed during LS1 on ME1/1 chambers, and their associated optical fiber data chain and off-chamber electronics, would overcome the L1 rate and latency limitations.

6.3 Improvements to the Existing L1 Trigger Systems

This section describes Phase-II improvements to the L1 Calorimeter, Muon and Global Trigger Systems beyond the Phase-I upgrade. The addition of the L1 tracking trigger is covered in the next section. The L1 calorimeter and muon trigger improvements are motivated by the requirements for combination with the new tracking trigger and the addition of new information to the trigger, either from adding new detectors, such as the additional Phase-II end-cap muon chambers, or required replacement of electronic systems to accommodate the Phase-II trigger architecture, such as the Phase-II EB readout that provides crystal-level energies to the trigger. The Global Trigger improvements are motivated by the need to fully exploit the new and considerably increased information provided in the trigger objects, including more precise

position and momentum resolution as well as such characteristics as track matching quality, calorimeter shower shape and number of tracking and muon hits.

For the L1 trigger, full exploitation of the tracking trigger requires good resolution for matching calorimeter and muon trigger information with the tracking information. This helps to minimize the "fake" rate from matching unassociated calorimeter and muon trigger objects with tracks and optimizes the matching of tracks with calorimeter or muon trigger information. In addition, finer resolution calorimeter and muon trigger information enable these systems by themselves to improve performance in the face of increasing pileup, providing better "stand-alone" performance. For some specific signals, "stand-alone" calorimeter or muon triggers could increase the efficiency over that of track and muon or track and calorimeter triggers. In addition, improved prescaled "stand-alone" calorimeter and muon triggers will reduce the trigger rate required to measure the efficiency of triggers that combine trigger tracks with muon or calorimeter trigger information. Finally, the finer resolution information from the calorimeter and muon triggers, the new information from the tracking trigger and the refined and increased information contained in the resulting trigger objects enables more sophisticated and effective topologically-based global trigger calculations such as invariant mass and also application of more more sophisticated and effective criteria on combinations of trigger objects such as their isolation.

6.3.1 Calorimeter Trigger

The purpose of the calorimeter trigger system is to identify electron, photon, jet and τ candidates, based on information from the ECAL and HCAL detectors across the entire range of CMS coverage in η. In addition, the calorimeter trigger computes global energy sums (Total E_{T}, $E_{\mathrm{T}}^{\mathrm{miss}}$, H_{T}), and is capable of providing a real-time estimate of bunch-by-bunch background levels as an input to pileup-dependent trigger selections. These tasks remain essentially unchanged in the HL-LHC era. However, the overall selectivity of the L1 trigger will be improved by combining the output of the calorimeter trigger with information from other sub-detectors before the final decision stage.

Precision measurements of Higgs properties and other electroweak phenomena are central goals of the HL-LHC physics programme and can only be attained by efficiently accepting events that have photons or leptons with low transverse momentum and loose identification requirements. The lower luminosity of the Phase-I upgrade running conditions will allow accepting such electroweak scale phenomena with good efficiency with the coarse geometrical granularity of calorimeter trigger towers. However at the high luminosity of the HL-LHC, with an average of 140 pileup interactions per bunch crossing, it is important to reduce fake matches due to track occupancy and further reject spikes by exploiting the full geometrical resolution of the ECAL barrel (whose single crystal size is approximately equal to the Molière radius in $PbWO_4$, $R_M = 22$ mm) at the trigger level. This will provide improved performance of the L1 calorimeter trigger itself, and better position resolution of the calorimeter objects for matching with tracks, to mitigate the effect of combinatorial background.

Another important goal of the Phase-II L1 trigger is to produce objects as close as possible to those used in the online selection and in the physics analyses. Already during Run-I, adjustments to the HLT algorithms were made to match the offline analysis algorithms wherever possible, resulting in significantly improved efficiencies for physics signals. However, the performance was limited by the coarse granularity of the calorimeter-tower information available. With the Phase-II upgrade, candidate particle objects, such as electrons, photons, and taus, with position, energy, and isolation characteristics close to those of the final particles will provide

sharper turn-on efficiency curves and better acceptance for electroweak scale physics. Such improvements are critical, for example, in efficiently selecting Higgs decays to offshell Z bosons, which can have multiple leptons in the final state that often have low transverse momenta and/or are not well isolated. Access to individual crystal information in combination with tracks at the L1 trigger will not only impact electroweak and Higgs measurements, but will also benefit searches for new physics that often involve low p_{T} leptons in the cascade decays of possible heavy new particles. In such cases, the L1 trigger rate can be kept sufficiently low by adjusting the fine-grained isolation requirements for a variety of trigger objects (photons, electrons, taus, etc).

The summary of the key improvements to the Phase-II Calorimeter Trigger system is as follows:

- Substantially enhanced input data will be available, following the upgrade of the on-detector and off-detector electronics of the barrel ECAL and HCAL. In particular, the digitized response of every crystal of the barrel ECAL will be available to the trigger (instead of the present 5 x 5 crystal sums), providing energy measurement at a spatial resolution of 0.0175×0.0175 in (η, ϕ), as opposed to the current input to the trigger consisting of trigger towers with granularity 0.0875×0.0875.
- Increased bandwidth into the trigger will allow better energy resolution and dynamic range for trigger primitives, along with additional information such as sub-BX timing measurements for trigger primitives derived from calorimeter enhancements providing high resolution timing.
- An increased latency of up to 4μs will be available for stand-alone calorimeter trigger processing, compared to less than 2μs in the pre-LS3 system. This will allow the implementation of object-finding algorithms with far greater selectivity than today, making full use of the improved input data.
- The object-level output of the calorimeter trigger will be combined with tracking information in order to reduce fake rates for electron objects, provide enhanced isolation criteria for electrons and photons, and form correlations in z-vertex position for all object categories. Both stand-alone and tracking-enhanced calorimeter trigger objects will be available to the Global Trigger, allowing an optimal trade-off between efficiency and rate across the whole E_{T} range.
- Detailed information from the calorimeters will be available for use in muon isolation calculations.

The basic processing steps in the calorimeter trigger chain are shown in Fig. 6.1, and consist of:

- Suppression or masking of subdetector-dependent artifacts, for example neutron-induced ECAL noise
- Time reconstruction of energy deposits, conversion to transverse energy, and application of energy calibration constants
- An optional zero suppression step before transmission from the subdetector electronics to the main L1 trigger
- Clustering of ECAL crystals into units suitable for first-pass object identification
- e/γ and τ identification, including fine-grained position estimation for identified objects, using the full ECAL barrel granularity
- Application of calorimeter-based isolation criteria to e/γ and τ objects
- Identification of jets, either via multiple fixed-sized window algorithms, or an iterative clustering algorithm

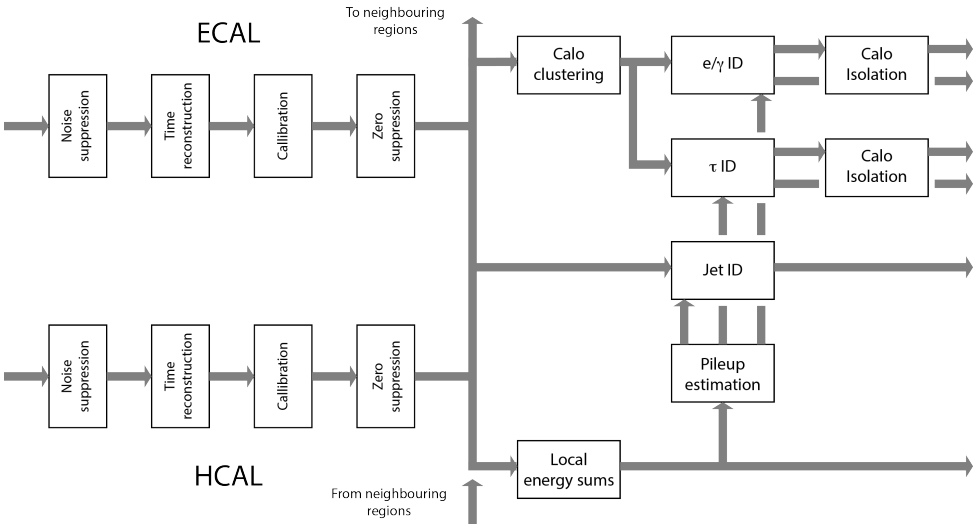


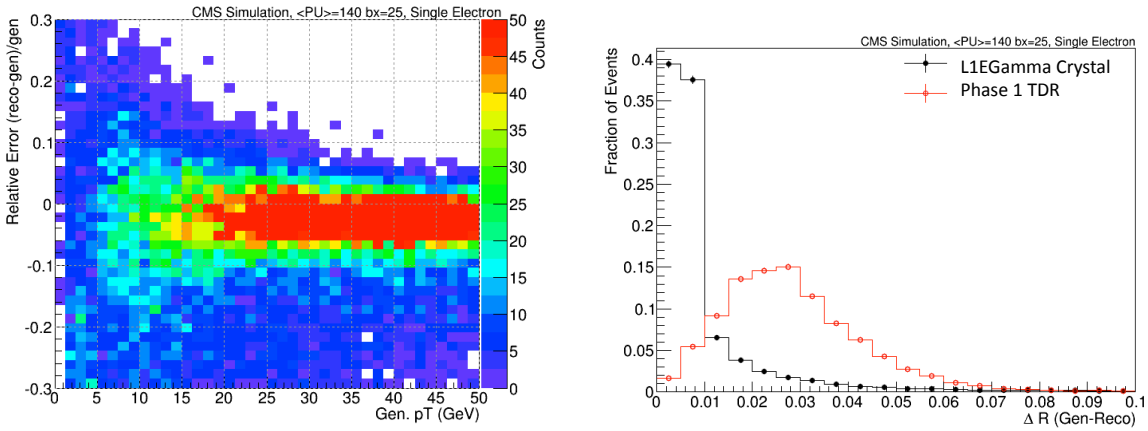
Figure 6.1: Processing steps in the L1 calorimeter trigger chain. The mapping of processing onto hardware modules is flexible.

- Estimation of pileup background on a bunch-by-bunch basis, allowing pileup-dependent object identification and isolation criteria to be applied

The calorimeter trigger will be incrementally upgraded in the period between LS1 and LS2, with new μTCA-based hardware providing significantly higher processing capacity than in the pre-LS1 era. However, is it envisaged that the HL-LHC trigger will be based on entirely new state-of-the-art hardware, and make use of an architecture that brings together calorimeter, muon and track trigger information at the earliest possible stage. Since the implementation will make use of flexible generic modules, there are several possible options for the mapping of calorimeter trigger processing steps onto hardware. For instance, it may be desirable to co-locate some of the earliest trigger processing functions with the sub-detector readout hardware. An example dataflow architecture is presented later in this Chapter.

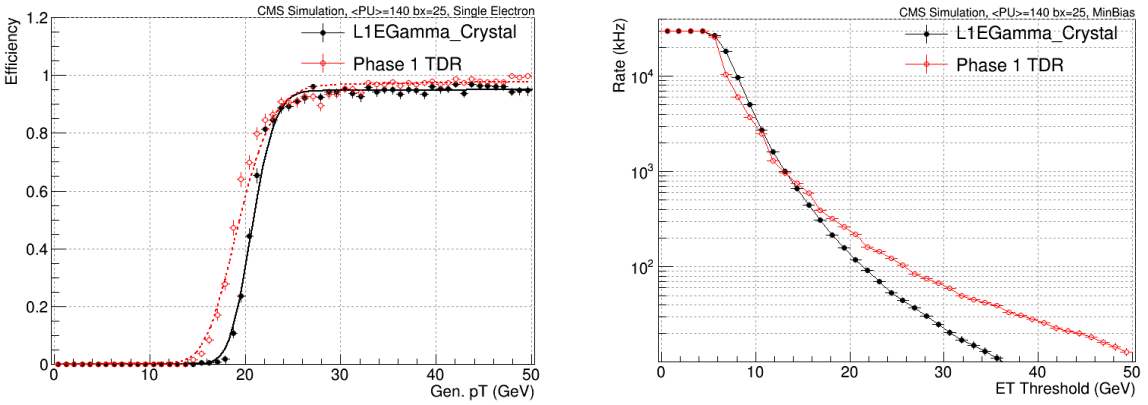
6.3.1.1 Calorimeter isolation and position with full crystal granularity

The energy resolution performance that can be reached using the full ECAL barrel granularity is demonstrated in Fig. 6.2(a), where the reconstructed E_{T} is compared to the generated E_{T} for a single electron sample as a function of the generated electron E_{T}. Figure 6.2(b) shows the corresponding geometrical resolution, for the difference in R with respect to the generated electron, when measured using the crystal information and for the trigger described in the Phase-I Trigger Upgrade TDR. Significantly improved resolution is obtained using the full ECAL barrel granularity compared with the case when only the much coarser calorimeter-tower granularity is used. Figure 6.3(a) shows the electron trigger efficiency with respect to the E_{T} of the generated electron for a threshold of 20 GeV. By exploiting the full ECAL barrel granularity, a sharper turn-on is obtained compared to Phase-I. Finally, Fig. 6.3(b) shows the trigger rate for electron/photon objects as a function of the object E_{T} threshold for a sample of background events, and compares again the two cases. For a E_{T} threshold of 20 (30) GeV, exploiting the full ECAL barrel granularity provides a lower trigger rate by a factor of two (three). These initial performance results demonstrate the ability to preserve good acceptance for low p_{T} and loosely isolated objects in high luminosity running conditions at the HL-LHC.



(a) (b)

Figure 6.2: Performance of electron reconstruction algorithms using the full ECAL barrel granularity: (a) The difference between reconstructed and generated electron E_T as a function of the generated E_T; (b) the difference between reconstructed and generated electron R, compared with the coarser calorimeter-tower based algorithm from the Phase-I Trigger Upgrade TDR.



(a) (b)

Figure 6.3: Performance of electron reconstruction algorithms using the full ECAL barrel granularity, compared with the coarser calorimeter-tower based algorithm from the Phase-I Trigger Upgrade TDR: (a) efficiency as a function of generated electron E_T; (b) rate for background events as a function of the trigger object E_T threshold.

6.3.2 Muon Trigger

Based on the experience of Run-I operations, the purity of the candidates selected using the L1 muon trigger system is generally high. However, the trigger rate for the p_T thresholds required to maintain high efficiency for offline muons originating from electroweak scale physics ($p_T \sim$ 20 GeV) remained significant. There are two main causes of the high rate, both related to accuracy of the standalone momentum measurement in the muon system. The modest core momentum resolution requires L1 thresholds substantially (typically 5-7 GeV) lower than the corresponding thresholds used in the offline, creating a substantial overhead in the rate. The second reason is the non-negligible tails of momentum resolution, which leads to a significant additional contribution due to soft muons being reconstructed as high momentum candidates by the L1 algorithms. The latter effect also causes "flattening of the trigger curve above L1

$p_T \sim 20$ GeV with the reconstructed candidates dominated by such large mis-measurements. The Phase-I trigger upgrades will reduce the tails of momentum measurement by increasing the average number of spatial points readily available for a good momentum measurement. This will be achieved by combining trigger primitives from all muon systems, e.g. DT and RPC or CSC and RPC, in single track reconstruction and momentum fit. Since there is no redundancy to the CSCs in the region of $|\eta| > 1.6$, these improvements will only be possible if a GEM1/1 station is installed during LS2.

With the complete upgrade of the muon systems in Phase-II, the availability of full-resolution DT $|\eta|$ information at the first level of the trigger will significantly improve track matching in the longitudinal plane. The z-resolution of single-station track segments is presently reduced from 0.5 mm achievable with all available DT information, to a coarse 32 cm at the DT trigger primitive level. Improvements to DT trigger primitives are discussed in more detail in Section 4.2.3.4. In the endcap region, combining all the new GEM and iRPC stations and CSC information both at the primitive-building and muon-reconstruction stage will provide significant improvements to both "stand-alone" muon momentum measurement, and muon-tracker matching performance.

The baseline design of the muon trigger for Phase-II generally follows that of the Phase-I design. Trigger primitives from the muon detectors will be transmitted via fast optical links to the local track reconstruction boards. The electronics boards will be executing the most detailed local tracking algorithms that can be programmed into their FPGAs. Local stand-alone muon candidates will be matched to L1 tracker-trigger tracks to generate high quality global L1 muon candidates. One difference in local reconstruction algorithm design is that the local stand-alone muon algorithms will have to focus on efficiency and position resolution besides momentum resolution, to optimize the matching with the tracks from the tracking trigger, minimizing the fake matches at high pileup.

An evaluation of the resources needed to pre-process the full stream of DT detector data and to provide track segments suitable for Level-1 trigger track-finding is presented in the Section 4.2.3.4.

In the present plans, the data sizes of the other muon trigger primitives (CSC, RPC) are not expected to change. In the baseline solution, the Phase-I muon track trigger, consisting of three subsystems: barrel, endcap and overlap, with 12 FPGA-based boards each, could be reused, or replaced by a smaller scale, state-of-the-art system.

Access to high precision track p_T measurement, possible with the new Level-1 tracking trigger, provides an opportunity for designing a high efficiency and purity muon trigger with sufficiently low p_T thresholds as required by physics. The weaknesses of the current muon Level-1 trigger system stem primarily from the tails in momentum resolution, which lead to low p_T muons (mostly from b-decays, punch through, and decays in flight) occasionally being reconstructed as high momentum tracks. While such mis-reconstructions happen rarely, the production rate for soft muons is very high, leading to a large contribution to the rate of muon trigger. This effect is most pronounced in the forward region where the particle rates are high, magnetic field is weak and redundancy is the lowest. With no improvements, maintaining the standalone muon trigger rate within reasonable bounds would only possible at the cost of a substancial reduction in efficiency.

The planned upgrades of the muon system coupled with the new tracking trigger allow preservation of efficient muon triggering, as in Phase-I, while also significantly improving the standalone muon trigger by itself. With the addition of the GEM-based detectors in the first two

stations and iRPC detectors in the farther two stations of the forward region, the reconstructed muon tracks are guaranteed to have enough hits for a muon p_T measurement allowing improved resolution and reduced tails in the momentum measurement. Second, the much increased lever arm formed by the pairs of GEM and CSC chambers in the first two stations and the high spatial precision of the two detectors will enable an online measurement of muon candidates' bending angle, which reduces the tails of the Level-1 momentum resolution. Measurement of the bending angle using chamber pairs in the same station serves as an effective second measurement of muon transverse momentum, complementary to the measurement by the muon Track Finder, as it is performed within a single station and does not suffer from the effects of multiple scattering in the absorber material.

The deployment of the tracking trigger will allow designing low threshold, high efficiency and high purity muon trigger paths, which will maintain and possibly extend the CMS reach to include signatures with soft muons, including low threshold multi-muon trigger paths and track-based isolation requirements. As it was mentioned earlier, a number of new physics scenarios predict final states containing muons originating from decays of sufficiently long living particles. Because of a significant impact parameter and/or absence of hits in the inner tracker layers, such muons may not be reconstructed by the track trigger. Maintaining standalone muon triggering capabilities (even if with somewhat higher thresholds) will therefore remain important both for maintaining physics sensitivity to displaced signatures as well as providing an important calibration for the paths based on matching tracking and muon trigger primitives targeting prompt muon signatures.

6.3.3 Global Trigger

The Global Trigger (GT) and the Global Muon Trigger (GMT) will be replaced for the Phase-I upgrade between LS1 and LS2. The custom-built VME modules are being replaced by μTCA boards. The design is fully modular and allows future increases in performance for the Phase-II upgrade.

To allow for an increasing number of parallel L1 trigger path calculations ("algorithms") in the future, the design foresees that multiple boards with identical hardware but different firmware can work in parallel. All inputs will be split and provided to all processor boards and their verdict will be combined in "OR" and forwarded to the Trigger Control and Distribution System (TCDS). Most inputs are supplied on optical fibers while the output (basically one bit at 40 MHz) is copper to save latency. Ancillary inputs for "external conditions" (such as from beam and beam halo counters) are received on copper lines via μTCA input boards and forwarded to the GT processor boards via the μTCA backplane.

The Phase-II upgrade will result in an increase of inputs to GT. Muon and calorimeter objects will be matched to tracker tracks in an intermediate hardware level upstream of GT. This will provide GT with an increased number of candidate objects (with and without matching or fitting to tracker tracks) of higher than present precision, requiring additional processing power and bandwidth for appropriate processing. For the time being we assume that going from Phase-I to Phase-II upgrade, GT will receive roughly twice the number of objects and twice the number of bits per object.

The current hardware platform for the Phase-I GT upgrade would have sufficient input resources and bandwidth to support the Phase-II upgrade. Additional processing power can be provided by using more GT boards in parallel. However, as the processor board is generic, GT could also use another type of board using a more powerful FPGA with higher input bandwidth.

Furthermore, to make use of the track trigger, it is foreseen that GT can calculate objects for specific vertices. At present, and up to LS3, all quantities are overall values for a given bunch crossing. Therefore, until LS3 only one value of quantities such as missing transverse energy can be calculated, while the use of the tracker trigger will allow calculating this and other values for each vertex individually. This will make it possible to construct at L1 missing transverse energy objects with energies not associated with the primary vertex removed and also provide correlations between objects identified with the primary vertex.

6.4 Track Trigger

This section presents results of performance studies of triggers which include L1 tracking. The algorithms presented in this section are used in the trigger menu studies shown in Sect. 6.5. In the studies presented here the performance of the L1 tracking is as described in Sect. 2.3.2.5. The L1 tracking algorithm used in these studies, tracklet based track finding, incorporates the effects of detector resolution, fakes from combinatorics, but assumes a fully efficient detector and sufficient bandwidth to transfer all stubs to the L1 track-finding processor. The calorimeter L1 objects are obtained from algorithms similar to those described in the L1 TDR, but adapted to the tighter pile-up conditions. The simulation of L1 muon candidates uses the currently available algorithm for the Phase-I detector, which is close to what was used in Run-I.

6.4.1 Muons

The current stand-alone muon trigger system occasionally assigns a high transverse momentum to low p_{T} muons. This results in a flattening of the muon trigger rates when the threshold increases beyond $\mathcal{O}(30\,\mathrm{GeV})$, and increasing the trigger threshold to higher p_{T} values is no longer effective in reducing the rate. The Phase-I upgrades will improve the stand-alone p_{T} determination compared to what was available at Run-I, but this will not be enough to cope with the high rates expected at Phase-II. The introduction of the track trigger will provide a major improvement to muon triggering due to the excellent momentum resolution of the inner tracking. For the new track+muon trigger, the required offset between L1 thresholds and the desired offline minimum p_{T} will become negligible, which allows a large reduction in the rate. The track trigger will also eliminate any residual tails in momentum resolution of the muon trigger.

6.4.1.1 Matching muon trigger primitives to L1 tracks

For the central pseudorapidity range of $|\eta| < 1.1$, an "outside-in" matching algorithm has been studied, which starts from L1 candidates in the muon system and extrapolates their trajectory towards the tracker. Only the two innermost stations are used, which measure the bending angle ϕ_B in the station and the pseudorapidity of the muon candidate. Predefined windows, parameterised by these measurements, are then used to extrapolate the trajectory of the muon candidate towards the beam line, and L1 tracks that have a ϕ_0 and θ_0 within the window are matched to the muon candidate. In case several matches are found, the track that is closest in (η, ϕ) to that measured in the DT chamber is selected, and the track p_{T} defines the muon transverse momentum[1].

[1] An alternative algorithm, in which tracker stubs (instead of L1 tracks at the origin) are matched to the DT candidate, in each layer of the tracker, is also under consideration.

6.4.1.2 Matching L1 tracks to L1 muons

An "inside-out" matching algorithm has also been implemented, in which the parameters of L1 tracks are used to extrapolate the trajectory of the corresponding charged particles to the muon stations. A matching is then performed with L1 muon candidates delivered by the muon trigger system, using windows in pseudo-rapidity and azimuth which have been determined taking into account multiple scattering and the η and ϕ granularity of the L1 muon system. When a L1 track passes the matching cuts, it defines the kinematics of the muon candidate. This algorithm has been implemented in the full η range of the muon system.

6.4.1.3 Performance

In the central pseudorapidity range, the two algorithms can provide similar rates and efficiencies. Figure 6.4 (top) shows, for two ranges in pseudorapidity, the efficiency of a single muon trigger when the stand-alone muon trigger is used and when it is associated with L1 tracks, as a function of the muon generated transverse momentum. The latter provides a very good efficiency and, thanks to the much improved p_{T} determination provided by the L1 track, a much sharper turn-on. This sharper turn-on translates into the large rate reduction shown in Fig. 6.4 (bottom). For a threshold of about 20 GeV, the rate of a single muon trigger can be reduced by a factor of $\mathcal{O}(10)$, as compared to the standalone muon trigger configuration used in Run-I. While the standalone muon trigger configuration used here does not include the improvements expected from the planned muon system upgrades, the large reduction in the rate is indicative of the strong enhancement of muon triggering capabilities brought by the track trigger.

6.4.2 Electrons

Electrons are identified by matching L1 tracks to tower based L1 e/γ objects found by the calorimeter trigger. The matching is performed in η and ϕ where the calorimeter η is corrected for the vertex position, z_0, of the track and the ϕ position is corrected for the bending of the track in the magnetic field. Two selections are considered: a high E_{T} selection optimized for electrons over $\mathcal{O}(20)$ GeV where tracks are required to have a p_{T} greater than 10 GeV, and a selection for lower energy electrons where looser cuts are applied in order to retain a reasonable efficiency. The intention is that the high energy selection will be used for single electron triggers, while the lower energy electrons will be used for triggers that combine an electron with some other object. The left plot of Fig. 6.5 shows the efficiency for the selection, while the right plot shows the rate as a function of the E_{T} threshold. For a threshold of 20 GeV, the rate of a single electron trigger can be reduced by a factor of about 5 when L1 e/γ objects are matched to tracks.

The principle challenge with electrons is to obtain a high efficiency for finding the L1 track, given the fact that many electrons interact with the material in the detector and radiate away a large fraction of the original energy. As shown in Fig. 2.25 the efficiency for reconstructing electrons is significantly lower than that of both muons and pions. The efficiency for electron reconstruction has been optimized by considering looser matching windows in the tracking algorithm as well as only using the innermost hits, which are less affected by bremsstrahlung, to determine the p_{T} of the electron track. In addition to track matching, an algorithm that matches stubs directly to the L1 e/γ objects has been considered. This algorithm has a similar performance to the L1 track based algorithm; however the latter will be much simpler to implement since it will not require access to all tracker stubs, but instead only to the much smaller set of L1 tracks.

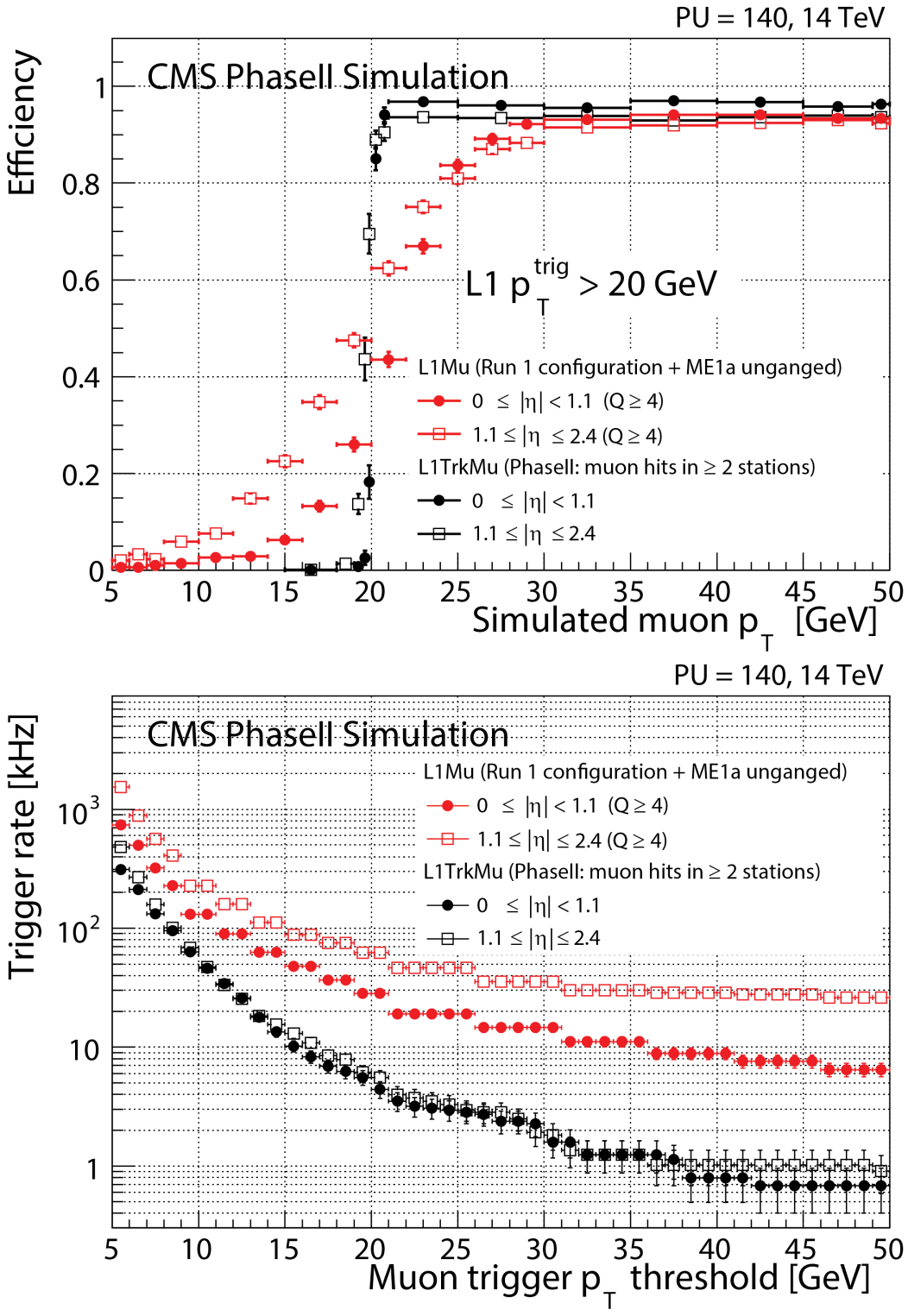


Figure 6.4: Top: Efficiencies for a single muon trigger with 20 GeV threshold as a function of the generated transverse momentum of the muon, for stand-alone L1 muons (red symbols) and for muons that are matched to L1 tracks (black symbols). Two ranges in pseudo-rapidity are shown. Bottom: Rates of single muon triggers as a function of the p_T threshold. For triggers based on stand-alone L1 muons, the quality cut ($Q \geq 4$) that was used during Run-I is applied.

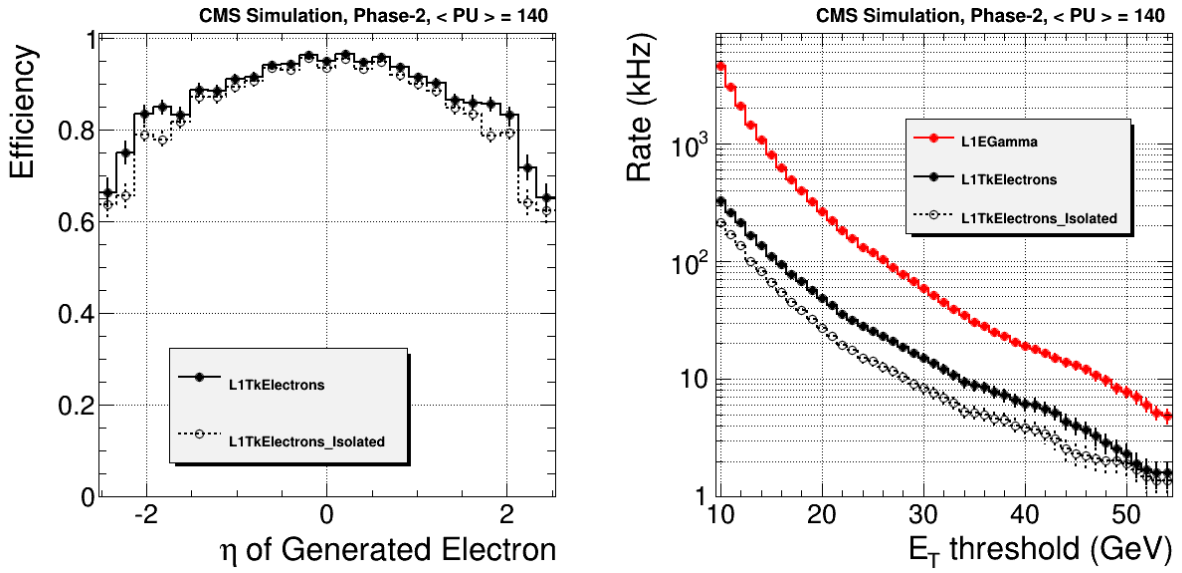


Figure 6.5: Left: Efficiency vs. η of the track-matching selection optimised for high E_{T} electrons, obtained on a sample of electrons with $\langle E_{\mathrm{T}} \rangle = 35\,\mathrm{GeV}$ in which a L1 e/γ candidate of $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ has been found. Right: rate of a single electron trigger as a function of the threshold: based on L1 e/γ objects (red dots); on such objects that pass the "high E_{T}" track-matching (black dots); on objects that pass in addition the tracker-based isolation (black open squares) described in 6.4.3.

6.4.3 L1 Track Based Isolation for electrons and muons

In offline analyses and in the High Level Trigger, an isolation requirement with respect to tracks is an efficient handle to increase the purity of lepton samples. Track-based isolation is also intrinsically more robust with respect to pileup than a calorimeter-based isolation. Examples of isolation requirements with respect to L1 tracks are considered here, and their effect on single electron and single muon triggers is presented. The usage of track-based isolation in the identification of tau leptons at L1 will be discussed in Section 6.4.5.

We start from muon or electron candidates delivered by the algorithms described previously (see Sections 6.4.1 and 6.4.2). A relative isolation variable is defined as $Isol = \Sigma p_{T,L1track}/p_{T,l}$ where $p_{T,l}$ is the transverse momentum of the lepton (given by the calorimeter for electron candidates, and by the tracker for muons) and the sum in the numerator runs over L1 tracks that:

- have a transverse momentum $p_{T,L1track}$ above a minimal threshold $p_{T,min}$;
- satisfy loose quality requirements;
- are in a cone around the lepton track: $\Delta R < \Delta R_{max}$, where ΔR denotes the distance in azimuth-pseudorapidity between the considered L1 track and the L1 track matched to the lepton candidate. The optimisation of this parameter leads one to consider a cone radius $\Delta R_{max} \sim 0.2 - 0.3$, consistent with the observed limited extension of the lepton "footprint". For electrons, the isolation cone is restricted to $\Delta R_{min} < \Delta R < \Delta R_{max}$, in order to avoid tracks due to Bremsstrahlung photons that convert;
- have a z_0 consistent with that of the L1 track matched to the lepton, z_{lepton} : $|z_0 - z_{lepton}| < \Delta z_{max}$.

The L1 track matched to the lepton is excluded from the isolation sum. The performance of the isolation requirement is quantified by the fraction of "background" and "signal" events for

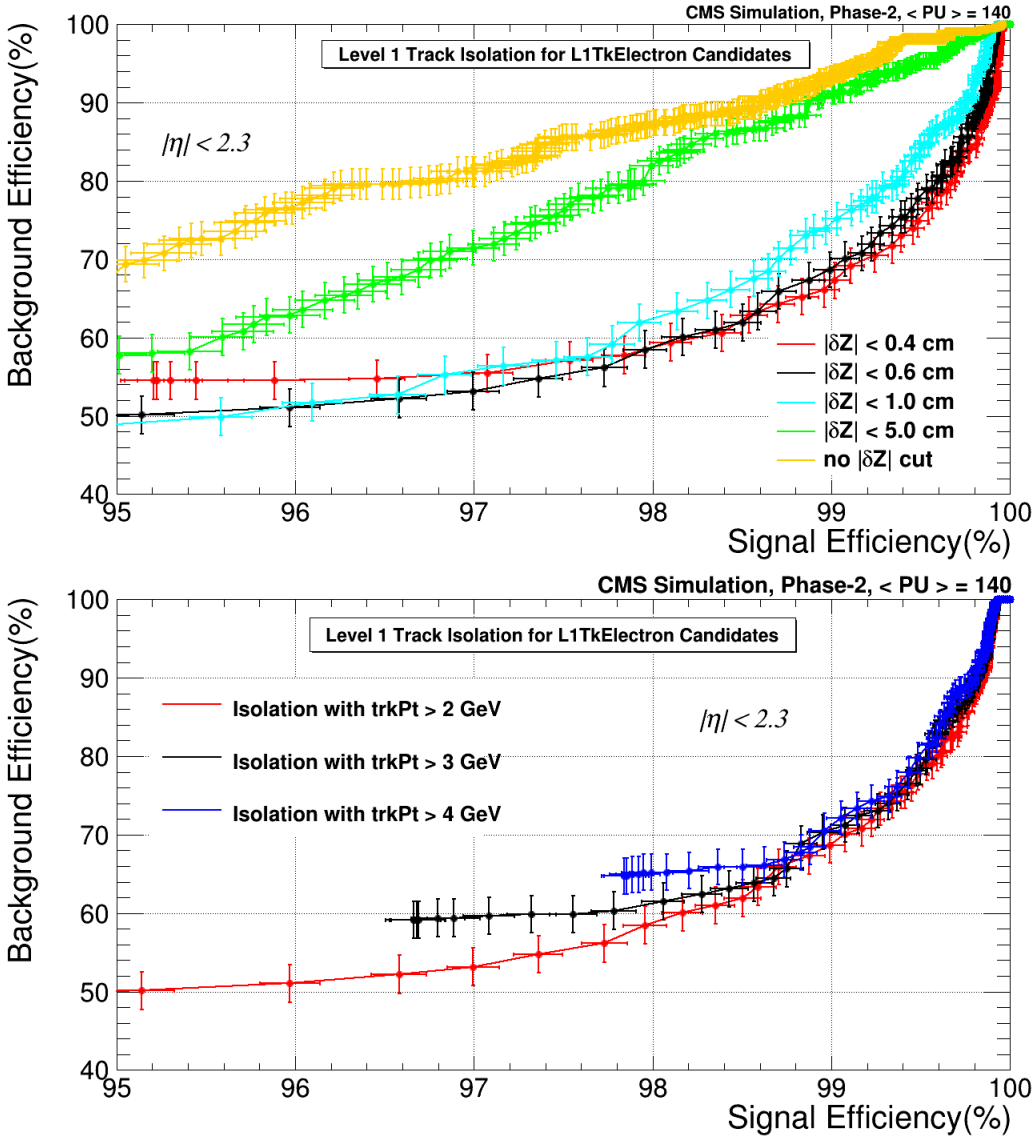


Figure 6.6: Performance of a L1 track-based relative isolation cut applied to L1TkElectrons of $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ in minimum bias events (background) and in $W \to \nu e$ events (signal). In the top plot, different values for the parameter Δz_{max} are considered. The bottom plot shows how the performance degrades when the minimal p_{T}, $p_{T,min}$, down to which the L1 tracks are available, is increased from 2 to 4 GeV.

which *Isol* is below a given cut.

Figure 6.6 (top) shows the fraction of minimum bias events in which a "L1TkElectron" (i.e., a L1 e/γ object satisfying the matching requirement described in 6.4.2) with E_{T} above 20 GeV would survive an isolation cut, as a function of the same fraction in $W \to \nu e$ events. The different curves correspond to different values of Δz_{max}, and each point in the curves corresponds to a different upper cut on the relative isolation variable. The best performance is obtained for Δz_{max} of about half a centimeter, consistent with the $\mathcal{O}(1\,\mathrm{mm})$ resolution for the z_0 measurement of those tracks, for which a rate reduction factor of about 2 can be achieved while maintaining more than 95% efficiency on the signal. Figure 6.6 (bottom) also shows that increasing the minimal threshold $p_{T,min}$ from 2 GeV to 3 GeV only leads to a small deterioration of the performance of the isolation requirement. Similar observations have been made for tracker-based isolation of muon candidates.

6.4.4 Isolated photons

Similar to what was shown in Section 6.4.3 for charged leptons, an isolation requirement with respect to L1 tracks can help identify (isolated) photons. An important difference compared to the case of electrons is that the position of the z vertex is a-priori unknown, while, for electrons, the L1 track matched to the L1 e/γ candidate provides a reference z_0. An isolation variable for photons can be defined similarly to what is done for electrons, except that all tracks in the annulus $\Delta R_{min} < \Delta R < \Delta R_{max}$ enter in the isolation sum, irrespective of their z_0 position. Moreover, the ΔR distance is now calculated using the η and ϕ of the calorimeter L1 e/γ object (hence, as seen from the center of the detector), such that the footprint of genuine photons extends to larger distances than that of electrons. Figure 6.7 (left) shows the fraction of minimum bias events in which a L1 e/γ candidate with E_{T} above 20 GeV would survive an isolation cut, as a function of the same fraction in single photon events that have the kinematic properties of photons from a $H \to \gamma\gamma$ decay. Several curves are shown, corresponding to different values of ΔR_{min}. The importance of excluding the tracks from converted photons via a large enough value for ΔR_{min} is clearly visible. It can also be seen that a loose (tight) isolation cut can reduce the background rate by a factor of about 2 (3), while retaining 95% (90%) of the photons considered[2]. This tracker-based isolation requirement performs considerably better than a calorimeter-based isolation making use of tower-level inputs. While the latter also retains 90% of photons coming from a $H \to \gamma\gamma$ decay, the corresponding rates are larger. This is illustrated in Fig. 6.7 (right), which shows rates of a double-photon trigger using inclusive L1 e/γ objects, "Iso e/γ" objects isolated in the calorimeter, and L1 e/γ objects that pass the loose tracker-based isolation cut. For thresholds of about 18 and 10 GeV on the leading and subleading legs, the rate can be reduced by a factor of > 6, more than twice better than what is achieved with the calorimeter-based isolation.

6.4.5 Taus

The identification of tau leptons that decay hadronically, using the information delivered to the L1 calorimeter trigger system, is a very challenging task. Two different approaches have been investigated to improve this selection using the L1 tracks. The first approach uses tau candidates as selected by the calorimeter trigger ("CaloTaus"): tracks are matched to the CaloTau and a track based isolation requirement is applied. In the second approach, we start with L1

[2]When comparing with the results shown for electrons in Fig. 6.6, one should keep in mind that photons from $H \to \gamma\gamma$ have a considerably larger E_{T} than electrons from $W \to e\nu$, and are more central. For single photons with the same $< E_{\mathrm{T}} >$ and $< |\eta| >$ as electrons from W decays, one recovers the performance shown by the "no δz cut" of Fig. 6.6 (top).

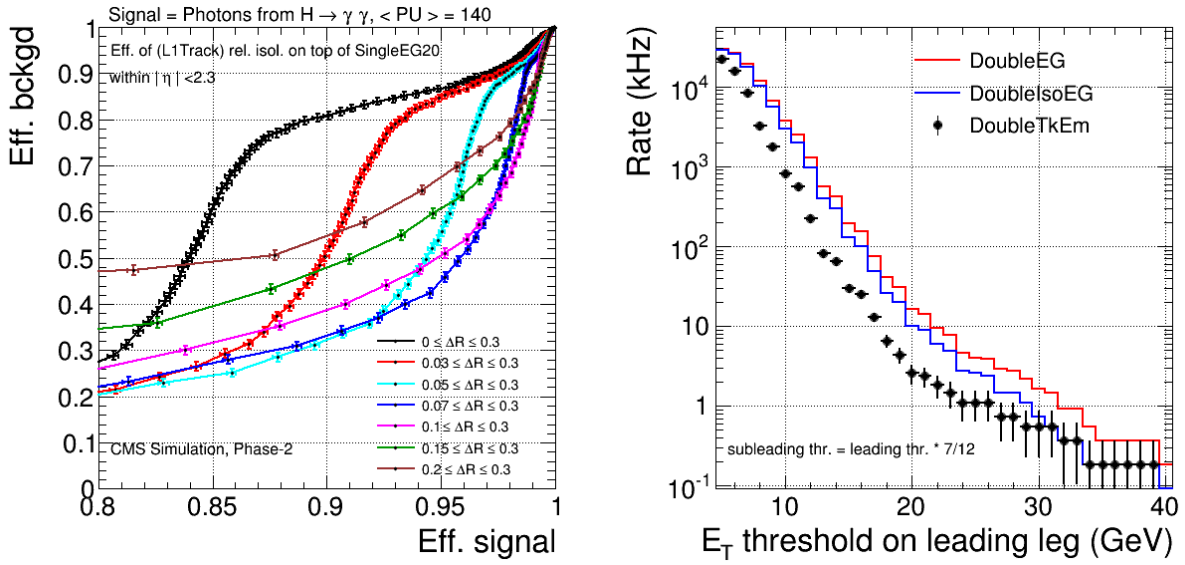


Figure 6.7: Left: Performance of a L1 track-based relative isolation cut applied to L1 e/γ candidates of $E_{\mathrm{T}} > 20\,\mathrm{GeV}$ in minimum bias events (background) and in single photons coming from a $H \to \gamma\gamma$ decay (signal). Right: rates of a di-photon trigger based on inclusive L1 e/γ objects (red histogram) , on L1 e/γ objects that pass the calorimeter-based isolation requirement (blue histogram), and on L1 e/γ objects that are isolated with respect to L1 tracks (closed dots).

tracks and add crystal level L1 e/γ objects. The two approaches are described in more detail below together with their expected performance.

For the "CaloTau + Tracks" algorithm, the CaloTau objects that are within a distance $\Delta R < 0.15$ of a good quality track with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ are selected in a first step. An isolation cone is then opened around the direction of this matched track (would several tracks be found, the one that is closest to the CaloTau is retained), defined by $\Delta R_{min} < \Delta R < 0.3$ with $\Delta R_{min} = Min(0.15, 3.5\,\mathrm{GeV}/E_T)$ where E_T denotes the transverse energy of the CaloTau. The CaloTau is accepted as a tau candidate if no track is found with a z_0 within 1 cm of that of the matched track and within this isolation cone. The calorimetric E_T defines the transverse energy of the tau candidate.

For the "Tracks + e/γ" algorithm, we start by finding candidate leading tracks. A leading track is required to have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ and no higher p_{T} track must be found within a cone of $\Delta R < 0.5$ around its direction. To suppress fake candidates from misreconstructed tracks, only tracks with 5 stubs or more and with a good χ^2 are used. Starting from the highest p_{T} tracks, additional tracks are combined with the leading track if they come from the same vertex ($\Delta z < 0.8\,\mathrm{cm}$) and if the invariant mass of the combined tracks is less than the τ mass. After having added the tracks, we add crystal-level L1 e/γ objects in the same way: objects with $E_{\mathrm{T}} > 5\,\mathrm{GeV}$, starting with the highest E_{T}, are added as long as the mass of the tau candidate remains below the τ mass. Finally, a track based relative isolation is calculated from the additional tracks, within an isolation cone of size 0.5, that come from a vertex consistent with that of the leading track, $\Delta z < 0.6\,\mathrm{cm}$. The relative isolation is required to be less than 0.15.

The performances of the CaloTau, CaloTau + Tracks, and Tracks + e/γ algorithms are compared in Fig. 6.8. The upper plots show the trigger rate as a function of the signal efficiency for VBF $H \to \tau\tau$, for a single τ selection on the left and a double τ selection on the right, when the trigger threshold is varied. For the double τ selection, we use the same threshold for both τ candidates, and, when tracks are used to select both taus, the two candidates should be consistent

with coming from the same vertex within 1 cm. The signal efficiency is measured with respect to events where both taus from the Higgs decay are within $|\eta| < 2.3$, decay hadronically, and lead to a visible transverse energy larger than 20 GeV. The lower plots show the background rate for a given p_T threshold. The Tracks + e/γ algorithm and the CaloTau + Tracks algorithm show comparable performances; the menu studies described in Sect. 6.5 make use of the former. For the single τ selection, we can maintain a 50 kHz trigger rate while having an efficiency of 45% for the VBF $H \to \tau\tau$ selection. In Fig. 6.9 we show the turn on curves obtained for the different τ selectors.

6.4.6 Event Vertexing

The L1 tracks reconstructed in an event can be used to determine the primary vertex. This determination is a pre-requesite for calculating a tracker-based missing transverse energy, as detailed below in Section 6.4.8. It can also be used for multijet or hadronic triggers presented in 6.4.7.

A fast reconstruction of the primary vertex can be made by histogramming the z_0 of all L1 tracks that fullfill quality requirements and have a transverse momentum above a threshold $p_{T,min}$. A peak is expected to show up in this histogram, at the position of the event vertex. This simple determination can be improved significantly by:

- weighting the histogram with the p_T of the tracks;
- maximizing the total scalar p_T carried by L1 tracks in three consecutive z bins of the histogram, instead of picking up the bin with the largest Σp_T. This allows to account for events where the tracks coming from the main interaction spill over two bins of the histogram. The primary vertex is then defined as the weighted average of the center of the z bins.

Figure. 6.10 (left) shows the performance of this vertexing algorithm as obtained on $t\bar{t}$ events (inclusive decays). It can be seen that:

- a resolution of better than 1 mm can be achieved (Gaussian core)
- the resolution does not degrade significantly if the $p_{T,min}$ threshold is increased from 2 GeV to 5 GeV.

In $t\bar{t}$ events, the efficiency to reconstruct the primary vertex within 5 mm of the true vertex is $\sim 97\%$. Restricting it to be within 1 mm of the generated vertex, the efficiency is about 90% for $p_{T,min} = 2$ GeV, and remains at $\sim 86\%$ with $p_{T,min} = 5$ GeV.

The performance is of course much worse for events with a low multiplicity of high p_T tracks. To ensure that the reconstructed vertex is reliable, a minimal threshold on the sum of the p_T of tracks attached to the vertex can be imposed, as shown in Fig. 6.10 (right).

6.4.7 Jets, HT, and MHT

The rate for multi-object triggers increases drastically with the pileup due to contributions from many different proton-proton interactions in a given bunch crossing. One of the most sensitive triggers to high pileup is the HT trigger, which just sums the transverse energy of the identified jets, or the MHT trigger which makes a vectorial sum of the p_T of these jets (resulting in an estimate of the missing transverse energy in the event). Similarly, the multi-jet triggers are also very sensitive to pileup. We can significantly reduce this pileup dependence by requiring that the different jets are consistent with originating from the same vertex.

L1 calorimeter jets are matched to L1 tracks to determine the z position of the jet vertex. First all

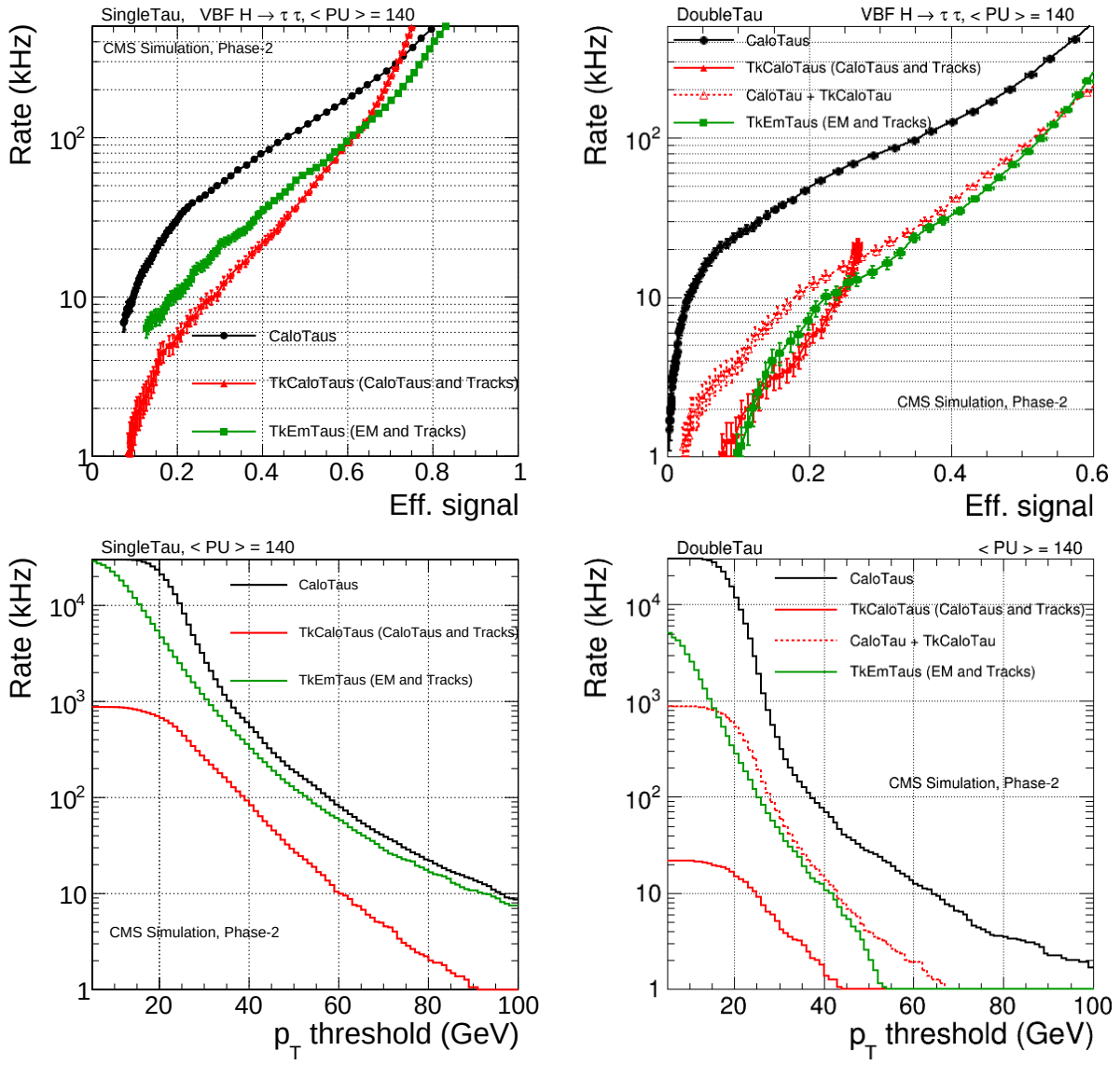


Figure 6.8: Comparison of the performance of the CaloTau, CaloTau + Tracks, and Tracks + e/γ algorithms. The upper left plot shows the trigger rate vs. VBF $H \rightarrow \tau\tau$ signal efficiency for a single τ selection. The upper right plot similarly shows the trigger rate vs. signal efficiency for a di-τ selection; when tracks are used to select both taus, the two candidates should be consistent with coming from the same vertex within 1 cm. The lower plots show the trigger rate vs. p_{T} threshold.

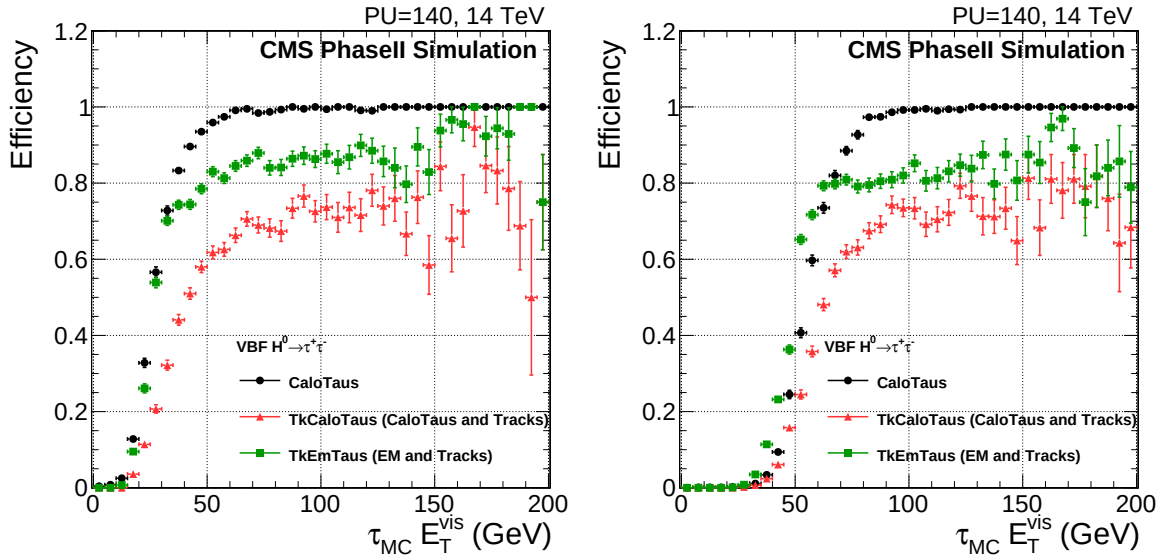


Figure 6.9: These plots show the turn on curves for the different τ selectors as function of the visible E_T of the τ. The left plot is for a threshold of 25 GeV and the right plot for a 50 GeV threshold.

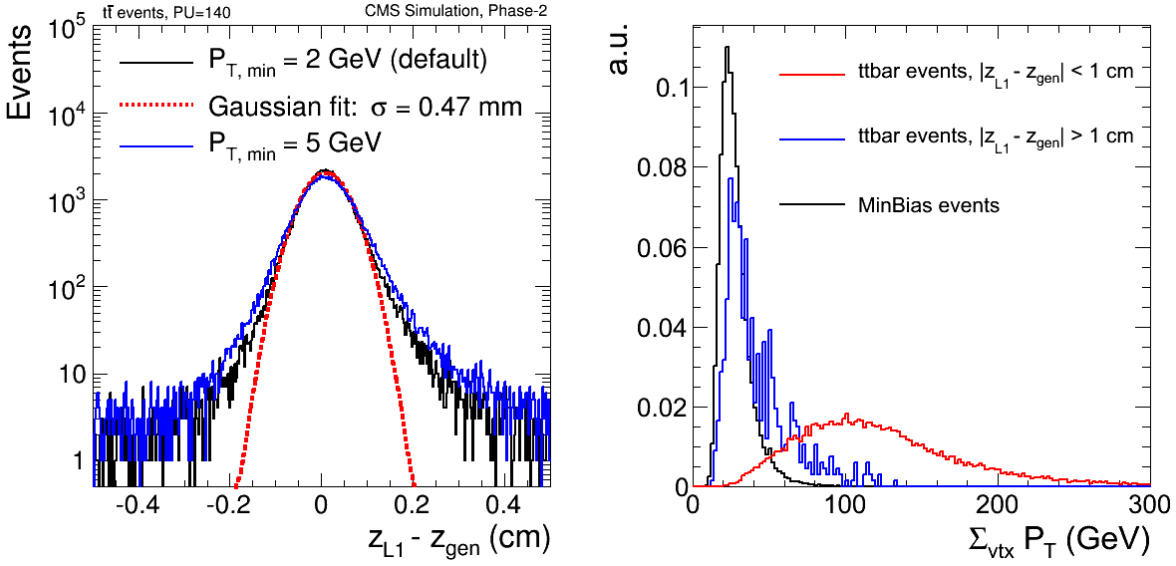


Figure 6.10: Left: Resolution of the reconstructed primary vertex in $t\bar{t}$ events. Right: Distribution of the sum of the transverse momenta of the L1 tracks attached to the reconstructed vertex: in $t\bar{t}$ events, when the reconstructed vertex falls within 1 cm of the generated vertex; in $t\bar{t}$ events, when the reconstructed vertex differs significantly from the generated vertex; in minimum bias events.

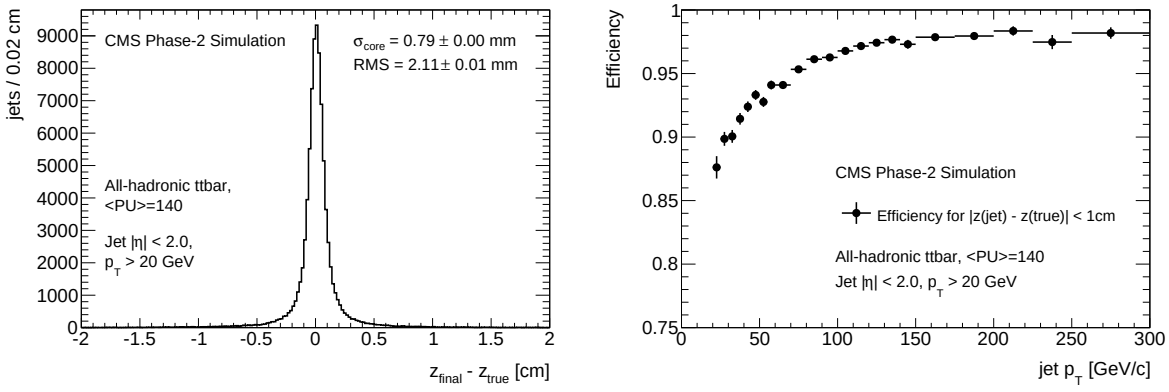


Figure 6.11: Left: Resolution for jet vertex position based on L1 track match. Right: Efficiency for reconstructing the jet vertex within 1 cm of the true vertex, as a function of the jet p_{T}.

tracks in an $\eta - \phi$ cone around the jet are selected. A first estimate for the z vertex is obtained from a p_{T} weighted average of the tracks. Outliers are removed in two passes before the final vertex position is calculated. The resolution obtained on the z position is about 1 mm as shown in Fig. 6.11 (left). The efficiency for finding the vertex position of the jet to better than 1 cm accuracy is about 95% for jets with $p_{\mathrm{T}} > 70$ GeV (see Fig. 6.11, right).

Having the jet vertex position, we can use this when forming triggers as illustrated in Fig. 6.12. The performance of a MHT trigger is assessed using events corresponding to stop pair production, where each stop decays into a top and a neutralino. The amount of missing transverse energy in the events depends on the mass difference between the stop and the neutralino. Two scenarios have been considered, which are close to the expected discovery reach of HL-LHC, and which lead to events with an average generator-level MET of ~ 200 GeV and ~ 300 GeV, respectively. Figure 6.12 shows the performance of a MHT trigger for these two example scenarios. In each case, events with a generator-level missing transverse energy larger than 100 GeV are used to define the signal. The blue and magenta open squares show the rate and efficiency of a MHT trigger that uses only the calorimeter information. Two jet algorithms have been used as input to the MHT calculation; they differ in particular in the amount of "pileup subtraction" that they achieve, one being more conservative than the other. The blue and magenta closed dots show how the performance of these MHT triggers is improved when MHT is calculated only from jets that come from the same vertex[3], within 1 cm. The vertex constraint can reduce the rates by a factor of typically 5 to 10, a larger reduction being obtained when using the jet algorithm that is less efficient in subtracting the pileup, as expected. After the vertex constraint, both algorithms lead to comparable rates. For a signal with a "medium" amount of missing transverse energy (Fig. 6.12 right), the vertex constraint allows for a MHT trigger that is $\sim 95\%$ efficient, with an acceptable rate of a few tens of kHz.

6.4.8 Tracking based MET

The reconstruction of the primary vertex makes it possible to compute the missing transverse momentum carried by all L1 tracks attached to the vertex. Only tracks that have a transverse momentum above a minimal threshold $p_{T,min}$ and that have a z_0 consistent with the primary vertex within ~ 1 cm are considered in the vectorial sum. The tracks are also required to pass quality cuts, which have been designed to limit the contribution of fake tracks that are often

[3]The leading jet is used here to set the z-vertex reference. Using the event vertex instead, determined as explained in Section 6.4.6, gives very similar results.

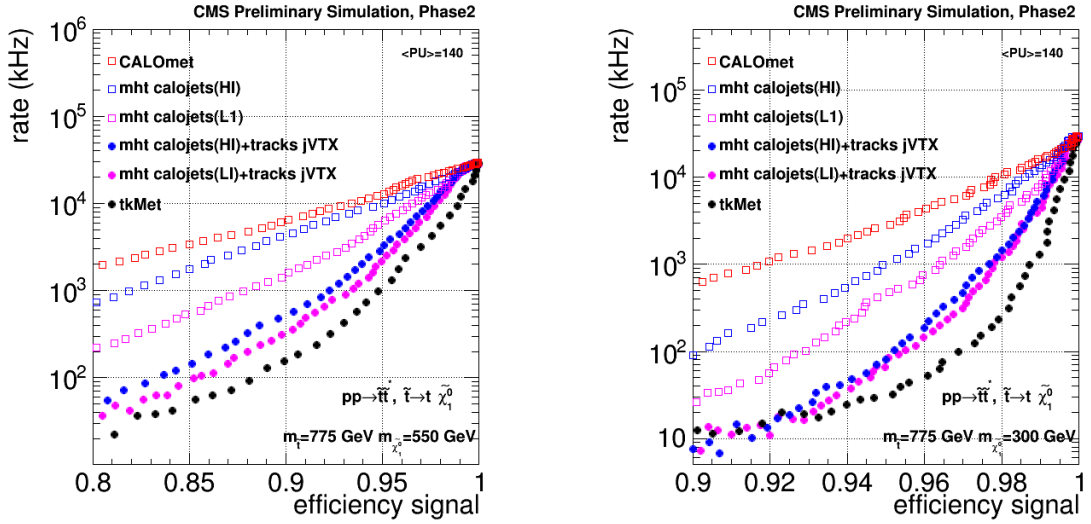


Figure 6.12: The rate as a function of the signal efficiency for inclusive missing transverse energy triggers, in three supersymmetric scenarios. The open symbols show the performance of triggers that do not make use of L1 tracking information: the red squares correspond to the calorimetric MET trigger, while the blue and magenta squares correspond to triggers on MHT, the missing transverse momentum reconstructed from jets. The blue and magenta closed dots show the performance of MHT triggers when the jets are constrainted to originate from the same vertex. The black closed dots correspond to a trigger based on TrkMET.

mismeasured. The resulting "track-based missing transverse momentum", or "TrkMET", is much more robust with respect to pileup than the calorimetric MET.

The performance of a TrkMET trigger is assessed on the stop-pair production events used previously. The black dots in Fig. 6.12 show the signal efficiency and the rates of a trigger based on TrkMET, while the red dots correspond to a cut on the calorimetric MET. For the two scenarios considered, a cut on TrkMET allows one to reduce the rates by one to two orders of magnitude, compared to a cut on the calorimetric MET. An inclusive TrkMET trigger that would ensure a signal efficiency of 80% to 85% in the very challenging "low MET" scenario (Fig. 6.12 left) appears to be within reach, with a rate of a few tens of kHz. In contrast, an inclusive calorimetric MET trigger that would efficiently select events from the "low MET" scenario is completely out of reach.

6.4.9 Multi-object triggers

Many triggers can be built that require several objects, for example two muons, or a muon and a jet. Generalizing the multi-jet triggers discussed in 6.4.7, the corresponding rates can be significantly reduced by demanding that these objects all come from the same vertex. For example, a z-vertex consistency requirement of 1 cm retains $\sim$ 97% of $Z \to \mu\mu$ events, while it reduces the rate of a low threshold dimuon trigger by a factor of $\sim$ 3. All di-object triggers considered in Section 6.5 below make use of this 1 cm vertex consistency requirement, except a few of them for which one leg is not required to be confirmed by the tracker.

6.5 Level-1 Menu at HL-LHC

In this section we explore the Level-1 trigger performance using a simplified menu similar to the one developed for Phase-I. It includes 20 major trigger paths that capture key physics signals and covers approximately 70% of the total Level-1 rate that would be needed for a full menu. The remaining 30% corresponds to trigger algorithms that have specific physics targets, triggers with higher-$|\eta|$ objects, diagnostic and prescaled triggers, etc. The trigger rates are estimated assuming operation at beam condition of 140 PU. The trigger algorithms used for these studies are those described in Section 6.4. Optimization of these algorithms represents an ongoing effort. In particular, the standard muon algorithms do not yet include all the upgrades featured in this document and therefore additional improvement in muon rates can be expected. In addition, algorithms using Level-1 tracking also have potential for further improvements. The rates presented in this section are therefore an indication of general trends, not final or optimal predictions.

6.5.1 Simplified Level-1 Menu Thresholds and Rates

The method used for these studies starts from the thresholds established for the Phase-I operation, as described in the Level-1 TDR[7]. It first measures the individual trigger rates for each algorithm. After measuring the rates, the thresholds are adjusted for any algorithm with an individual rate that is greater than 20% of the total rate or any algorithm with a low rate due to algorithmic improvements. As in the Level-1 TDR, the thresholds quoted have been scaled from online values to offline values. The scaling is chosen such that the trigger object is 95% (85% for taus) of the plateau efficiency for an offline cut at the threshold. This allows for a simpler comparison between algorithms that have different resolutions and thus different turn-on behavior. These threshold values can be seen in Table 6.1. For any given event, several trigger algorithms may be satisfied. Therefore, we must account for overlaps (correlations) between different algorithms when estimating the total rate of our menu. When we perform this procedure using trigger algorithms without any Level-1 tracking information, we find a total Level-1 trigger rate of approximately 1500 kHz for 140 pileup and almost 4000 kHz for 200 pileup events, which is well beyond the bandwidth being considered in this proposal.

Level-1 Tracking information can provide important rejection of backgrounds in the trigger. Our sample menu shown in Table 6.1 uses Level-1 tracks in the formation of the trigger objects. The details of these algorithms, including efficiency, are discussed in Section 6.4. Single lepton triggers include tracking requirements. Isolated electrons use tracking isolation rather than traditional calorimeter isolation. In the Phase-I triggers, photons and electrons are selected by the same e/γ triggers. With the addition of Level-1 tracking, these paths need to be separated and so we have introduced single γ and double γ triggers. Dilepton triggers make use of tracking information for the primary leg and in some cases on the secondary leg. When both legs of the dilepton trigger have a Level-1 track, we also impose a $\Delta z \leq 1$ cm requirement. The hadronic triggers also use Level-1 tracks. Multijet triggers require that all jets that contribute to the trigger decision be consistent with coming from the same z-vertex position ($\Delta z \leq 1$ cm). Likewise, the energy sum quantities (H_{T} and $H_{\mathrm{T}}^{\mathrm{miss}}$) use collections of jets that are consistent with the same z-vertex position. The rate of each trigger algorithm is shown in Table 6.1. At the bottom of the table, we give the total rate for our small sample menu and an estimate of the total rate required for a complete physics menu based on the estimate that our sample menu represents 70% of a full menu. The total Level-1 menu rate is approximately 260 kHz, not including any margin to cover uncertainties in extrapolations to higher energy and luminosity, actual detector performance, detector readout electronics, and the hardware realization of the algorithms. It has also to be noted that if Table 6.1 illustrates the power of the Level-1 tracking

when made in conjunction with an upgrade to the total Level-1 bandwidth, it is not an optimized Level-1 menu. More capabilities are being explored, such as invariant mass triggers, standalone missing P_T triggers constructed from tracks, standalone muon triggers, etc. Some of these are discussed in Section 6.4 and it is likely that with future refinement of the physics case at the HL-LHC, the present Level-1 table will be complemented with some of these triggers or even new ones, depending on the physics results obtained during Phase-I.

Table 6.1: Level-1 Menu using algorithms that include track trigger capabilites. The beam conditions are $\sqrt{s} = 14$ TeV, and $L = 5.6 \times 10^{34}$ cm^{-2}s^{-1} with a bunch spacing of 25 ns and pileup of 140. The rate for each algorithm is given along with the total rate, which accounts for overlaps between algorithms. For algorithms that depend on more than one object, the thresholds are listed in the order corresponding to the algorithm name. Objects which use Level-1 tracking are indicated with "(tk)". From Run-I data, we estimated that our sample menu of 20 triggers accounts for approximately 70% of the total Level-1 rate. The last line of the table presents the total estimated rate when we scale for the remaining 30%. No additional safety factor for uncertainties in our extrapolations has been applied.

$L = 5.6 \times 10^{34}$ cm^{-2}s^{-1} $\langle PU \rangle = 140$	Level-1 Trigger with L1 Tracks	
Trigger Algorithm	Rate [kHz]	Offline Threshold(s) [GeV]
Single Mu (tk)	14	18
Double Mu (tk)	1.1	14 10
ele (iso tk) + Mu (tk)	0.7	19 10.5
Single Ele (tk)	16	31
Single iso Ele (tk)	13	27
Single γ (tk isol)	31	31
ele (iso tk) + e/γ	11	22 16
Double γ (tk isol)	17	22 16
Single Tau (tk)	13	88
Tau (tk) + Tau	32	56 56
ele (iso tk) + Tau	7.4	19 50
Tau (tk) + Mu (tk)	5.4	45 14
Single Jet	42	173
Double Jet (tk)	26	2@136
Quad Jet (tk)	12	4@72
Single ele (tk) + Jet (tk)	15	23 66
Single Mu (tk) + Jet (tk)	8.8	16 66
Single ele (tk) + $H_{\mathrm{T}}^{\mathrm{miss}}$ (tk)	10	23 95
Single Mu (tk) + $H_{\mathrm{T}}^{\mathrm{miss}}$ (tk)	2.7	16 95
H_{T} (tk)	13	350
Rate for above Triggers	180	
Est. Total Level-1 Menu Rate	**260**	

6.5.2 Thresholds versus Total Level-1 Bandwidth

In addition to the menu presented above, we have explored the thresholds that would be required to accommodate a similar menu with a total bandwidth varying from 100 kHz to 1 MHz. Our procedure can briefly be described as follows (for more information see [7]). For each algorithm in the menu, we allocate an initial fraction of the total bandwidth according to the fractions used in Table 6.1. These fractions are used along with the the total bandwidth being considered to assign a rate to each algorithm. We then determine the threshold(s) that give the assigned rate. The initial thresholds for each algorithm are then used to determine the total Level-1 trigger rate, while properly accounting for overlaps between trigger algorithms. This process could lead to a total Level-1 rate that is substantially larger than or less than the maximum bandwidth being considered. We follow an iterative procedure that scales the allocated bandwidths either up or down. After each scaling, we determine new algorithm thresholds and then re-evaluate the rate of the full menu. We iterate until we converge on the target total Level-1 bandwidth. For the iterative procedure, we apply a simple democratic scaling of the assigned bandwidths. We perform this study both without and with the Level-1 tracking to show the performance of the latter to reduce the thresholds as a function of the total bandwidth.

Figure 6.13 shows the evolution of the trigger thresholds versus total Level-1 bandwidth for single lepton triggers. The thresholds given on the y-axis represent the primary offline thresholds described above. The x-axis gives the estimated total Level-1 bandwidth required for operating with these thresholds. The hatched area represent variation on the threshold values including both the statistical uncertainty from our samples and assuming a 20% variation on the rate at any given threshold. The statistical uncertainty tends to be largest at small total Level-1 bandwidth values because the individual thresholds must be raised significantly and pushed out on the tails of distributions where statistics become small. In general, the thresholds fall with increasing bandwidth. The thresholds fall sharply at the lower total bandwidths and then begin to plateau. In all cases, the presence of Level-1 tracks leads to lower thresholds. Figure 6.14 shows the situation for a set of dilepton triggers. Finally, Figures 6.15 and 6.16 show the hadronic triggers and lepton cross triggers. It should be noted that Level-1 tracking provides no extra rejection for the Single Jet trigger and therefore the thresholds with and without tracking are identical. In the case of the Double and Quad Jets, the Level-1 tracking provides rejection by demanding that jets that form the trigger are all consistent with coming from the same z-vertex position.

6.5.3 Bandwidth Requirements

The studies presented above used Monte Carlo samples generated with 140 pileup per event. We investigated the impact of running our trigger menu at a pileup of 200 events, which corresponds to a luminosity of 8×10^{34} cm^{-2}s^{-1}. When using the menu given in Table 6.1, the rate for a full Level-1 menu will increase to 500 kHz. The total rate increases by a factor of 1.9, slightly faster than a linear scaling with luminosity. In addition, we must account for simulation uncertainties, possible limitations in realizing the Level-1 track based triggers in hardware as well as the ability of the individual subsystems to provide the required Level-1 rate. For this we estimate that a safety factor of 1.5 is required leading to a bandwidth of about 750 kHZ to offer operation capability up to beam conditions of 200 PU. As the menu and physics thresholds presented in this section represent typical algorithms and rate allocations from current experience, we can expect that further ongoing physics studies and data from the upcoming run will influence the algorithms and thresholds used and may necessitate adjustment in the allocated rates associated with specific physics channels.

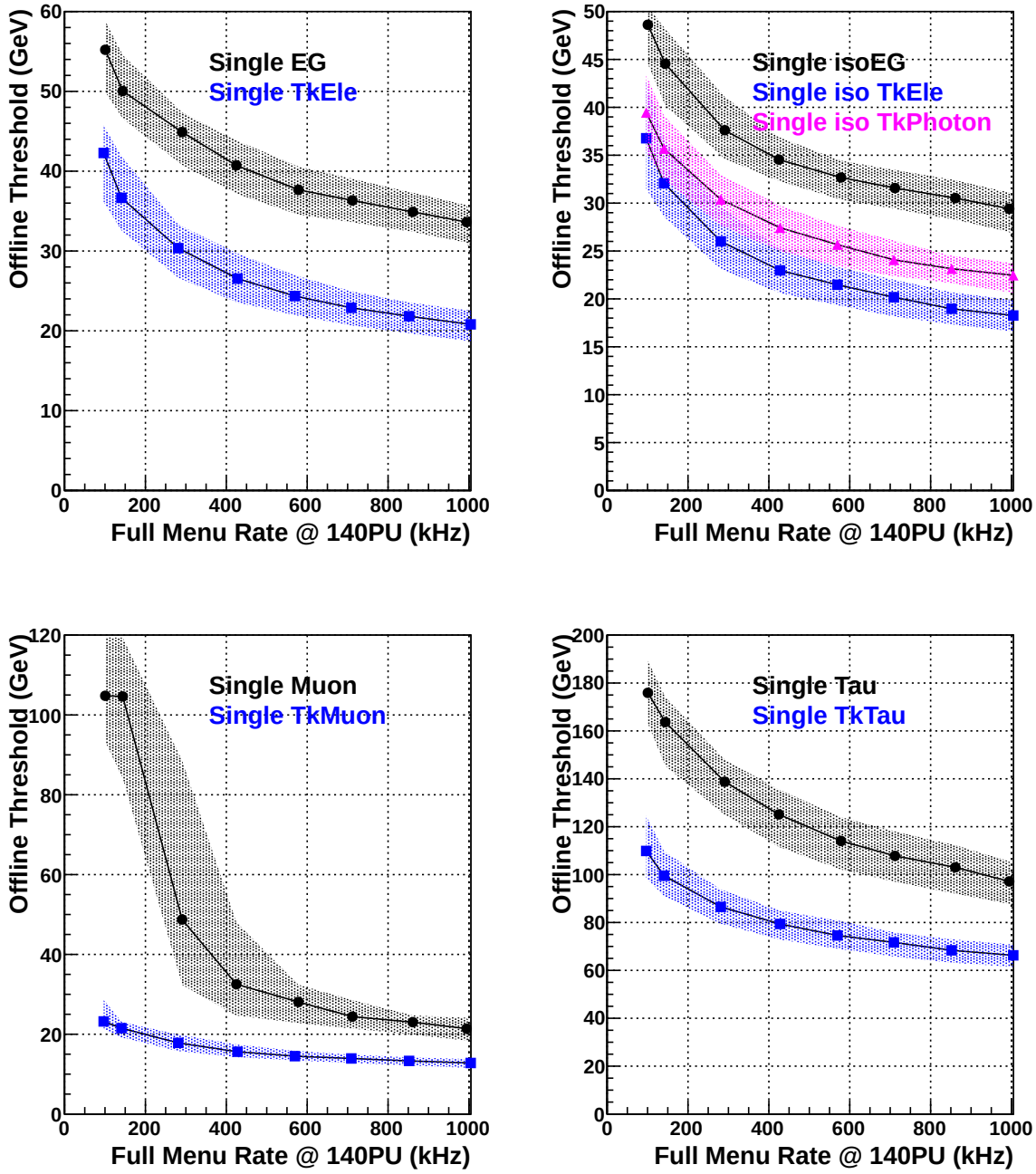


Figure 6.13: Thresholds versus bandwidth for single lepton triggers. (a) Single e/γ (b) Single Isolated e/γ (c) Single Mu (d) Single Tau. The width of the band represents the uncertainty on our predicted thresholds. The x-axis gives the estimated total Level-1 trigger rate at 140 pileup for both the tracking and non-tracking algorithms.

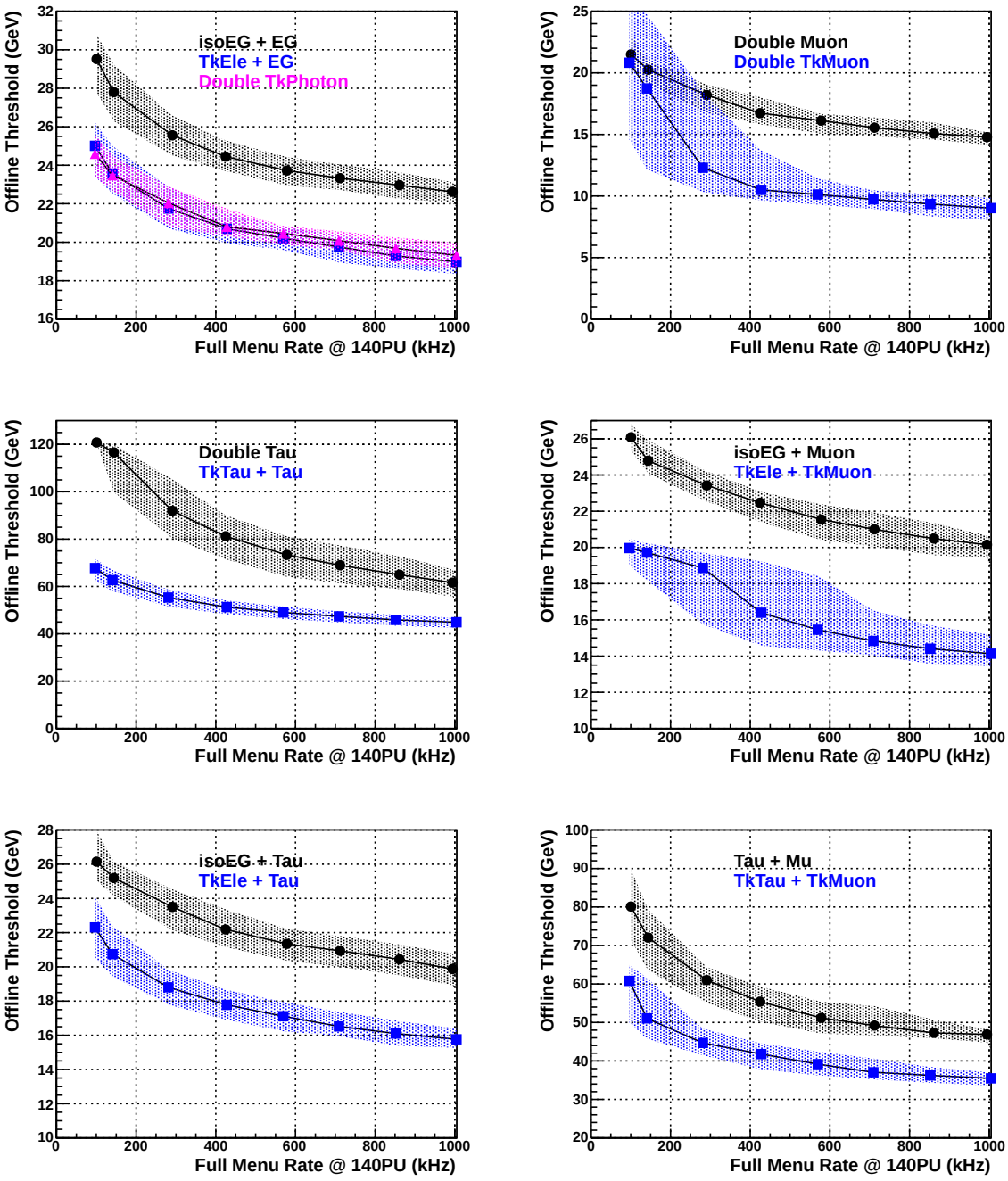


Figure 6.14: Thresholds versus bandwidth for several dilepton triggers. (a) Double e/γ (b) Double Mu (c) Double Tau (d) EG + Mu (e) EG + Tau (f) Mu + Tau The x-axis gives the estimated total Level-1 trigger rate at 140 pileup for both the tracking and non-tracking algorithms.

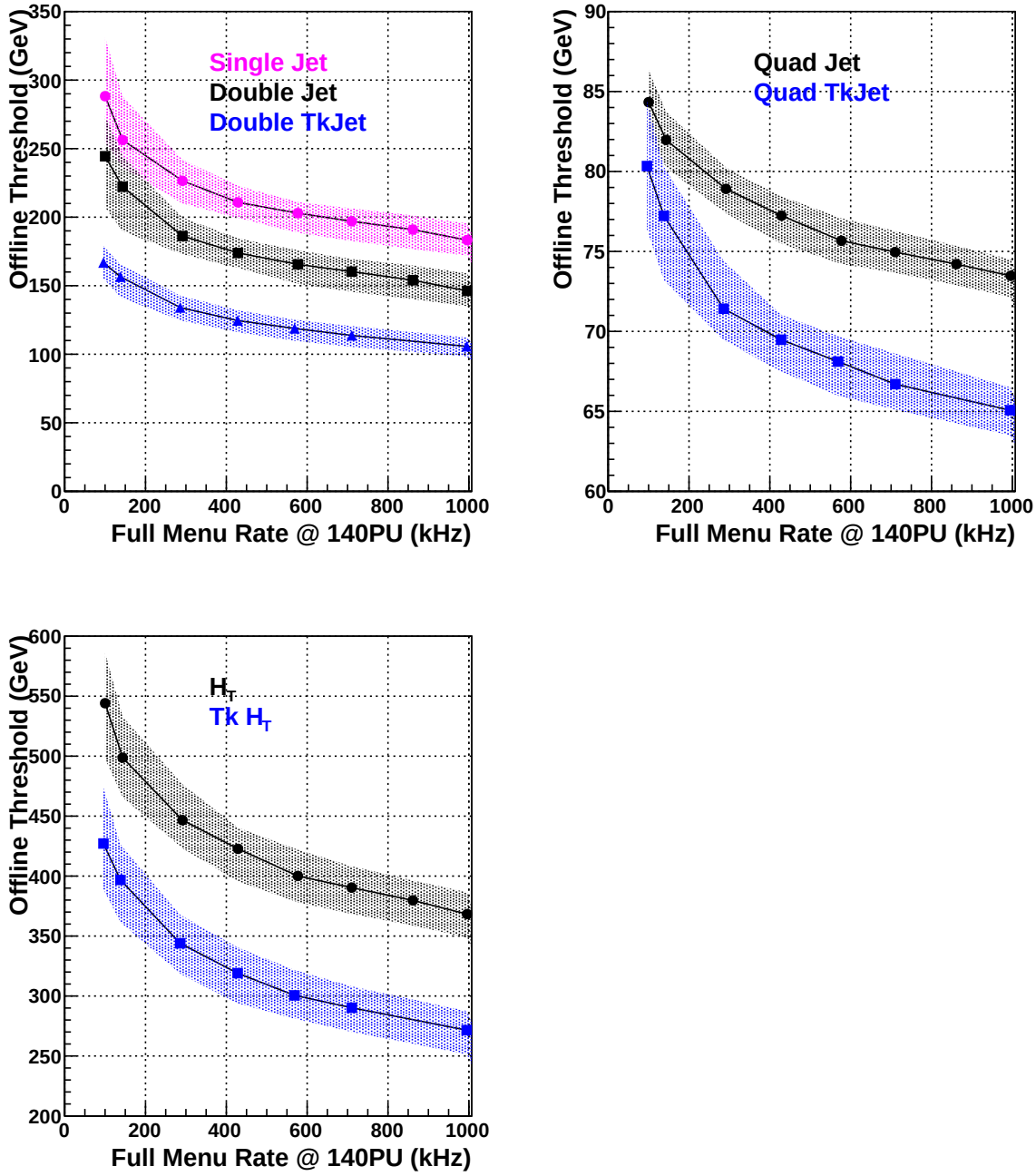


Figure 6.15: Thresholds versus bandwidth for hadronic triggers. (a) Single Jet and Double Jet (b) Quad Jet (c) H_{T}. Note: In the case of the Single jet, Level-1 tracking does not provide any additional rejection power and so only a single curve is displayed. The x-axis gives the estimated total Level-1 trigger rate at 140 pileup for both the tracking and non-tracking algorithms.

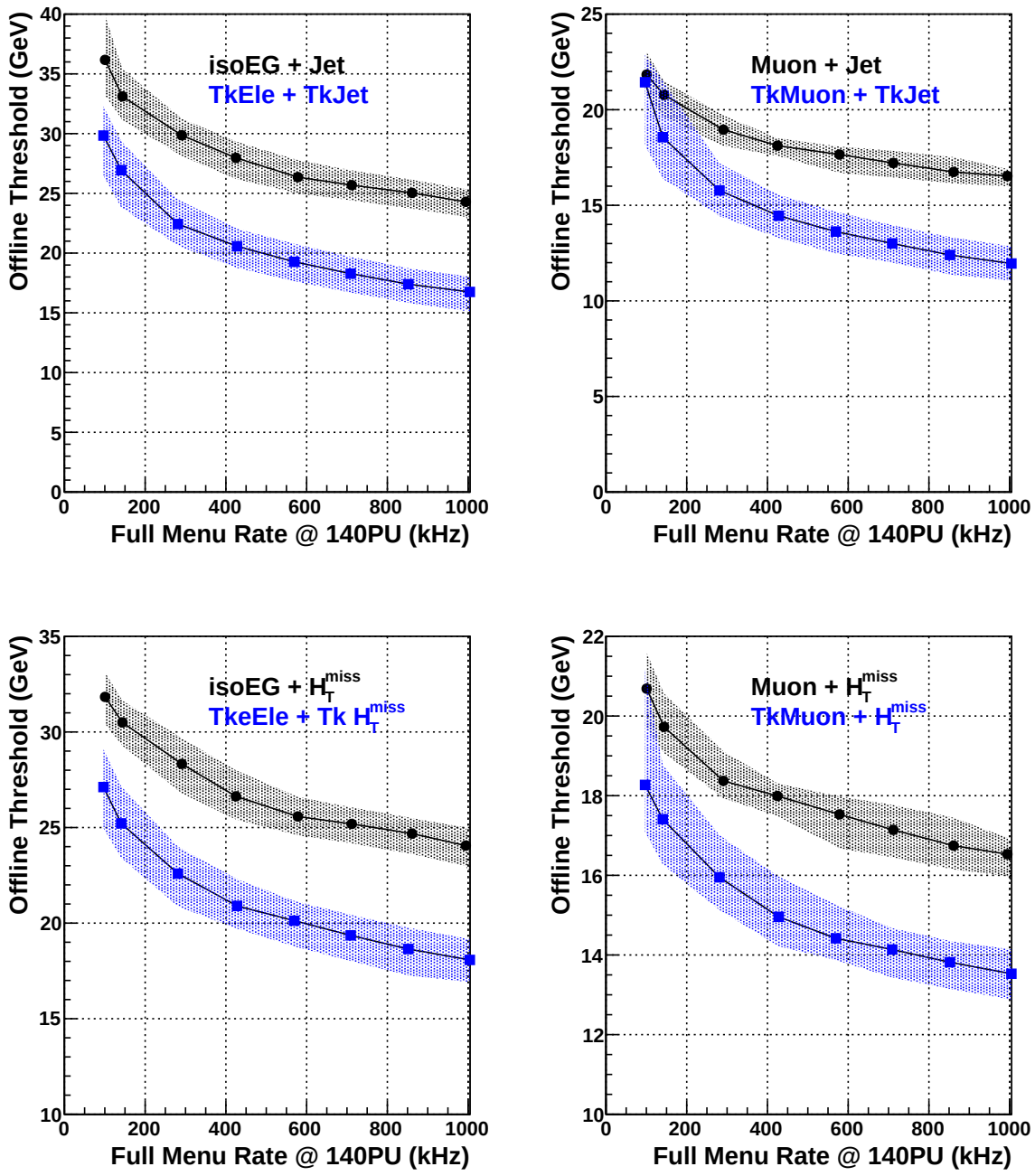


Figure 6.16: Thresholds versus bandwidth for lepton cross-triggers. (a) Single e/γ + Jet (b) Single Mu + Jet (c) Single e/γ + $H_{\mathrm{T}}^{\mathrm{miss}}$ (d) Single Mu + $H_{\mathrm{T}}^{\mathrm{miss}}$. The x-axis gives the estimated total Level-1 trigger rate at 140 pileup for both the tracking and non-tracking algorithms.

6.6 HLT and DAQ

The High level Trigger of CMS in Phase-II will have to process an input rate of 500 kHz to 750 kHz, a factor 5 to 7.5 higher than in the present design. On the other hand, it is proposed to increase the average output rate to the permanent storage by the same amount, to 5 kHz for 140 pileup events and 7.5 kHz for operation at 200 pileup events. These increases would approximately match the expected increase in luminosity and cross section of the bulk of electro-weak scale processes at the HL-LHC compared to the 2012 data taking. The implication at the HLT is to maintain a rejection factor of a 100 as in the current system and with similar purity. Provided that the upgraded detectors will have the expected performance this reduction is certainly achievable by HLT algorithms similar to those being used now.

The main challenge for HLT will be the increased processing power needed to process the input stream, which we estimate to be about 24 times that of the 2012 run for 140 PU. This figure comes from the increased L1 input rate, the increased event complexity due to the higher pileup conditions and the higher center-of-mass energy. As measured during the 2012 run, up to a pileup of 63, a linear increase with PU can be assumed leading to an increase by a factor 4 (at 140 PU). A further factor will come from the increased physics processes cross sections at the higher center-of-mass energy. This will require tighter selection criteria and this effect has been accounted to contribute a factor 1.5 in the processing needs. A further factor 0.8 is applied to take into account the expected improvement in the CPU time from using the tracks found with $p_{\mathrm{T}} > 3\,\mathrm{GeV}$ by the L1 track trigger. For operation at 200 PU the overall processing power need will increase from 24 to about 52 compared to 2012.

While the previous estimates are based on the current computing performance, improvements can be expected from emerging technologies using many-cores and parallel computing. To this extent, CMS is adapting its framework to achieve event-level parallelism already at the beginning of Run-II, and algorithm-level parallelism in the following years. Work has also started to evaluate sub-algorithm level parallelism, needed to use GPUs or other many-core processors. The cost/benefits of this approach will have to take into account the need of rewriting part of the reconstruction code with different languages, and the need to keep the HLT code as close as possible to what is used in the offline reconstruction. Finally, improvement in the CPU time per event can be expected from code optimizations, and from using the result of the L1 tracking step as a seed for the HLT tracking, speeding up its first steps. Given all the above, we expect that the HLT processing needs for Phase-II will be technically achievable, this is further discussed in Chapter 7 and 8.

6.7 Hardware Design

For the HL-LHC, the L1 trigger data will combine information from tracking, calorimeter and muon triggers at a regional level with finer granularity than presently employed. After this regional correlation stage, the physics objects will be transmitted to the Global Trigger. The important new feature is that some of the tracking, isolation, and other regional trigger functions will be performed in a new hardware layer composed of regional cross-detector trigger crates.

This design requires cards with sufficient number of links with appropriate bandwidth to carry the necessary data between the calorimeter, muon and tracker trigger systems and with sufficient FPGA logic resources to implement the necessary logic tests for e, γ, μ, τ jet and MET object identification and association with tracker trigger primitive information. Sufficient inter-card backplane bandwidth will be needed to ensure efficient coverage across boundaries.

The Level-1 trigger will continue to be based upon use of high speed FPGAs with large numbers of high speed links, optical fiber links running at or in excess of 10 Gbps and the latest telecommunications backplanes and infrastructure to provide extensive interconnectivity and support for the hardware. Recent industry trends indicate that 40 Gbps or even 100 Gbps data links could be commercially available over the relevant timescale.

A number of hardware realizations are under study for use to achieve these requirements. These include μTCA as presently used in the CMS Phase-I Trigger Upgrade, ATCA as deployed by ATLAS and PCI-express designs which include exploitation of computer infrastructure. For purposes of this proposal and its costing, μTCA is used as the model.

The design will incorporate and be compatible with the existing Phase-I trigger hardware and will continue to utilize this hardware where possible. However, where greater granularity is required or direct incorporation of tracking information, these parts will be rebuilt.

6.8 Project planning and cost estimates

6.8.1 Organization

The Phase-II trigger upgrade presented in this document has been developed by the Trigger Performance and Strategy Working Group of the CMS upgrade organization. Following this technical proposal, the Phase-II trigger upgrade will become part of the central CMS Trigger project.

6.8.2 Schedule

The installation of the Phase-II trigger system is expected to start in 2024. Sufficient contingency in the schedule is reserved for a thorough commissioning of the electronics prior to the installation in CMS in order to ensure optimal performance at the restart of accelerator operation. The schedule of R&D is driven by the need to test the trigger electronics modules with early prototypes of new detector back-end electronics and where relevant new detector modules in test beams with full electronics chains in order to understand the detailed performance of the Phase-II trigger system components. In addition, extensive system tests are required to validate this high bandwidth interconnected system.

Major steps planned for the years $2016-2020$ are listed in Table 6.2. The target dates are established for the purpose of project planning. Actual dates can be several months later in all cases without affecting technical progress.

The trigger R&D activities include efforts in algorithms as well as in hardware. To allow the development of the calorimeter, muon, and tracking trigger electronics to proceed, the planned algorithms and the necessary trigger primitives as well as link counts and formats will need to be defined. This is expected to proceed in two stages with an initial definition (TP.L1.1) followed by a baseline definition one year later (TP.L1.3). The development of the hardware to realize the trigger will begin with the demonstration of key implementation technologies such as 40 Gb data links, which have a general applicability across the Phase-II upgrade (TP.L1.2). The completion of a detailed emulator for the baseline trigger (TP.L1.4) is also a critical R&D milestone which will inform the final implementation of the trigger hardware.

The trigger hardware has four development streams which will be pursued in parallel: Calorimeter Trigger Processor, Track Correlator Processor, Muon Trigger Processor and Global Trigger Processor. After the completion of the detailed emulator and the definition of hardware technology baseline (TP.L1.5), full function prototypes will be developed which are the

Table 6.2: Planned next steps for the Trigger upgrades.

Milestone Identifier	Target date	Description
TP.L1.1	Q2 2016	Initial definition of trigger algorithms, primitive objects, and inter-layer objects
TP.L1.2	Q4 2016	Initial demonstration of key implementation technologies
TP.L1.3	Q2 2017	Baseline definition of trigger algorithms, primitive objects and interchange requirements with subdetectors
TP.L1.4	Q4 2017	Software emulator demonstrates implementation of core phase 2 trigger menu with baseline objects
TP.L1.5	Q1 2018	Definition of hardware technology implementation baseline
TP.L1.6	Q4 2018	Full-function prototypes produced which allow local comparison with emulator
TP.L1.7	Q4 2019	Demonstrator system shows integration and scaling, global/full-chain comparison with emulator
TP.L1.8	Q1 2020	Phase 2 Trigger TDR

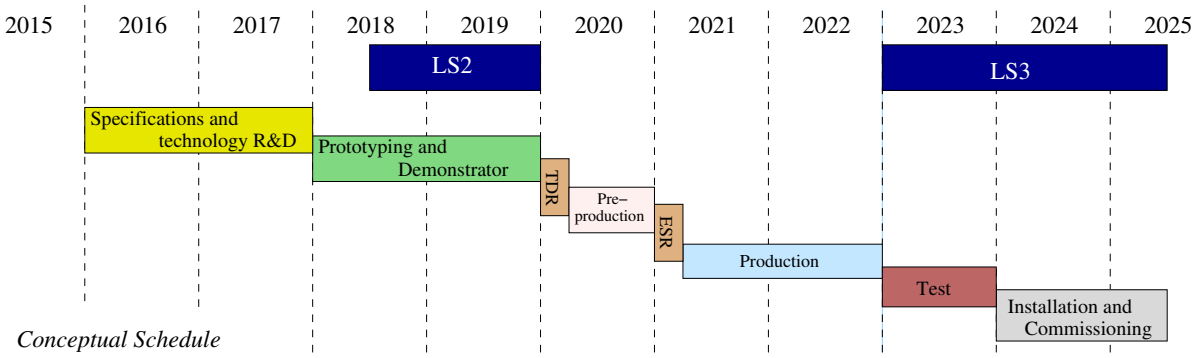


Figure 6.17: Conceptual schedule for the Trigger Phase-II upgrade.

first boards which have sufficient channels, processing capability and bandwidth optical links to meet the requirements of the final boards (TP.L1.6). These boards will cover only a portion of the trigger processing logic, however, and only local comparisons will be possible between the hardware behavior and the emulator. End-to-end comparisons over a slice of the detector will be possible with the demonstrator systems (TP.L1.7) which include multiple full-capability prototype boards. The goal will be to demonstrate a system, with infrastructure and testing environment, capable of being connected to its front end detector for test-beam validation. The Trigger Technical Design Report is planned to be completed in early 2020 (TP.L1.8) so that the TDR can be based on the results from the Trigger Demonstrators.

The overall conceptual schedule for the Phase-II trigger upgrades is shown in Fig. 6.17. The milestones in Table 6.2 lead to the Trigger TDR. After the TDR is a preproduction and final qualification period, followed by the Electronic Systems Review which will authorize construction. Installation will take place early enough in LS3 to allow for internal trigger system testing as well as commissioning of connections to the trigger primitive generation systems of the detectors.

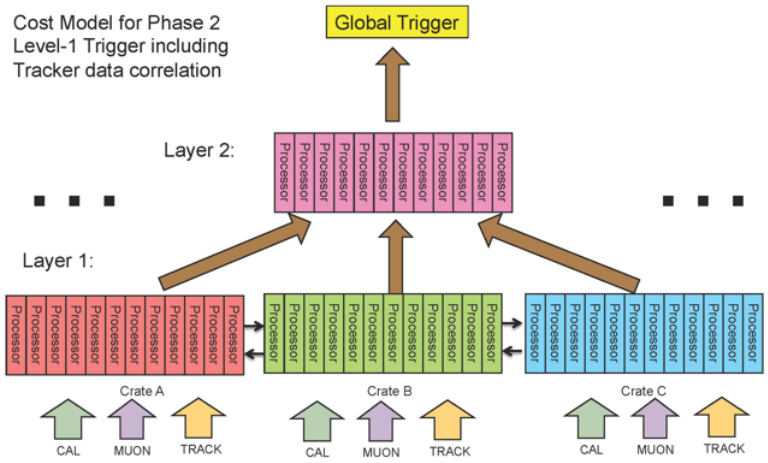


Figure 6.18: Example of the model structure used in estimating the cost of the Phase-II Level-1 Trigger, showing the incorporation of Level-1 tracking information.

6.8.3 Cost estimates

The cost of the track trigger is mostly contained within the Tracker estimate, with the L1 Trigger performing the correlation of the produced track with muon and calorimeter trigger information. For purposes of costing, we use a model of architecture where the input trigger data is pre-processed and characterized in a first layer, with the trigger algorithms executed in a second layer, which provides the output that is sent to the global trigger as is shown in Figure 6.18. For purposes of cost only and not design, the number of boards is calculated from the data bandwidth expected from simulation and the I/O bandwidth of the MP7 trigger processing board developed for the Phase-I upgrade. The cost estimate is based on the present cost of this board.

The calorimeter trigger will need to process the data at higher granularity in order to be optimally matched with the track trigger information. As this is very similar to the current calorimeter trigger, the cost estimate is essentially scaled to the higher number of channels involved.

The endcap muon trigger covering $|\eta|$ from 1.6 to 2.5 will need to be rebuilt to incorporate additional chambers in the endcap and to provide input to the tracking correlator. The cost is based on a similar Phase-I upgrade in this region. In addition, the modifications of the existing muon triggers covering the barrel and overlap regions to provide the input to the tracking correlator are costed.

The new Global Trigger will require an increase of capability once tracking information is introduced (at least a factor of 2 in the information contained in the found objects since they have much more precise coordinates from tracking, tracking isolation quantities), much more complicated algorithms, such as invariant mass cuts (enabled by the precise coordinates) and finally, an increase in the number of objects (at least a factor of 2 since the tracker will introduce a whole new series of objects found first by the tracker). The cost estimate is based on a scaling of the Phase-I upgrade Global Trigger according to the number of input channels.

The total cost of the L1 trigger upgrade system is shown in Table 6.3. The costs of the TCDS and HLT upgrades are discussed in Chapter 7.

Table 6.3: Estimated CORE cost for the L1 Trigger Upgrade.

Estimated CORE cost in MCHF (2014)	
Track Correlation Trigger	1.6
Calorimeter Trigger	3.5
Muon Track Finder (Barrel and Overlap)	0.7
Muon Track Finder (Endcap)	0.9
Global Trigger	0.6
Total Trigger	**7.3**

Chapter 7

Data Acquisition and Trigger Control

7.1 Introduction

The computing and networking equipments in the Data Acquisition system (DAQ) and High Level Trigger (HLT) are replaced every 4-5 years, due to end of life. LS3 is anticipated to be such a period, in which the new equipments will accommodate the additional resources required for the Phase-II sub-detector and L1-trigger upgrades.

The main parameters of the DAQ/HLT systems for Phase-II, in comparsion with the current system, are summarized in Table 7.1. These values are peak values at the start of the LHC fill, when luminosity is the highest. Clearly, the DAQ/HLT system is required to handle the peak values, whereas the offline computing needs to handle the average conditions. The HL-LHC is anticipated to operate with luminosity levelling, which will give roughly a constant luminosity throughout the fill. Two assumptions are considered in the following, operation at 140 and 200 PU with a L1 trigger rate of respectively 500 kHz and 750 kHz.

The capability to operate with the Phase-II upgrade up to a L1 acceptance rate of 750 kHz, will enable significantly more flexibility in online event selection to maximize the physics reach. We currently assume the same rejection factor of roughly a 100 at HLT as at present (section 6.6). This will lead to an HLT output rate of about 5(7.5) kHz in case of operation at 140(200) PU. These requirements are not expected to raise technical issues either for online or offline computing by the time of Phase-II. However, we continue to work towards minimizing computing needs by improving the reconstruction, both with new selection algorithms and with code development for multi-core processors and co-processors (Chapter
refchap:swcomp). With this work, the selection performance achievable at the HLT may improve.

The DAQ system for Run-I was designed for an average event size of 1 MB. The actual event

Table 7.1: DAQ/HLT system parameters.

	LHC Run-I	LHC Phase-I upgr.	HL-LHC Phase-II upgr.	
Energy	7-8 TeV	13 TeV	13 TeV	
Peak Pile Up (Av./crossing)	35	50	140	200
Level-1 accept rate (maximum)	100 kHz	100 kHz	500 kHz	750 kHz
Event size (design value)	1 MB	1.5 MB	4.5 MB	5.0 MB
HLT accept rate	1 kHz	1 kHz	5 kHz	7.5 kHz
HLT computing power	0.21 MHS06	0.42 MHS06	5.0 MHS06	11 MHS06
Storage throughput (design value)	2 GB/s	3 GB/s	27 GB/s	42 GB/s

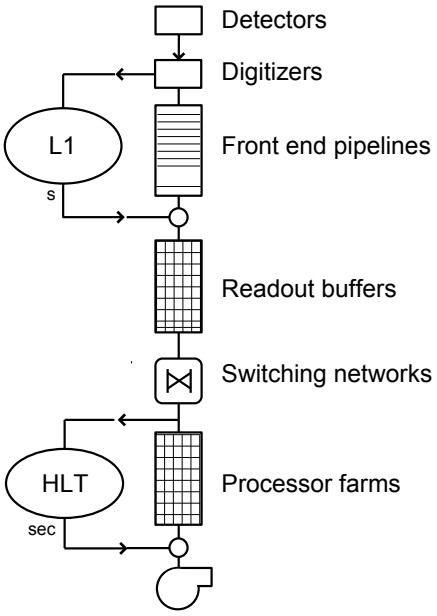


Figure 7.1: Two-level DAQ architecture with L1 trigger and HLT.

size at the highest PU was ≈0.8 MB and an approximate linear behaviour as function of PU was observed, with an offset of ≈0.3 MB. The event size was dominated by the tracking detectors, where the tracker strips and pixel contributed more than 60%.

For the Phase-I upgrade an increase in the event size is expected due to:

- Operation at higher (in-time) PU and at 25 ns bunch spacing.
- New Pixel detector with more layers.
- New readout for the Phase-I L1 trigger upgrade.
- New sensors and higher segmentation for HCAL.

The Phase-II upgrade results in entirely new sub-detectors and/or new readout systems. A first estimate of the event size amounts to roughly 4.1-4.5 MB in the range of PU 140-200 (see Section 7.2, below). The design value for the event size to be handled by the DAQ is assigned accordingly, taking into account a small margin (10%).

The HLT computing power required at 140 PU and an L1 rate of 500 kHz has been estimated to be 24 times that of the 2012 run (see Section 6.6). It is expected to scale with PU, and hence would amount to about 11 MHS06 in case of operation at 200 PU and 750 kHz L1 rate.

The local storage of the DAQ system for Run-I was designed for an effective throughput of 2 GB/s. This effective throughput is required for the storage of the main physics stream at the HLT output rate, as well as calibration and data quality monitoring streams. For Phase-II, the required throughput is determined by the event size and an L1 accept rate up to 7.5 kHz, whereas the rate of calibration and data quality monitoring streams is assumed to be similar as for Run-I and 2.

A system overview of the two-level DAQ architecture detector is shown in Figure 7.1. The proposed Phase-II DAQ/HLT architecture is the same as for the currently implemented system, with a single level hardware trigger (L1) and a second level software software trigger (HLT) on commercial processors. The difference with the current system is the increase of L1 decision latency from 3.4 μs to 12.5 μs, and the increase of the maximum L1 rate from 100 kHz to 750 kHz.

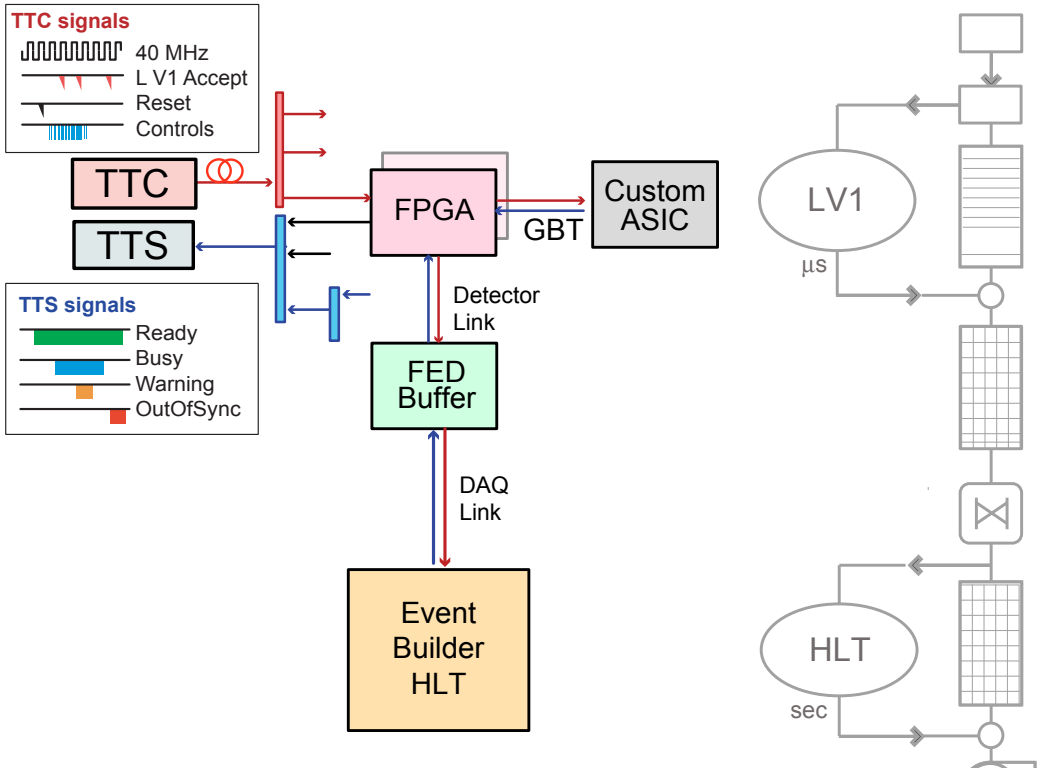


Figure 7.2: System overview of the sub-detector read-out electronics, DAQ and HLT. The TTC and TTS signals could be transmitted using PON technology (see Section 7.3).

The first level is a clock driven synchronous system operating at the 40 MHz LHC clock. The second level is an asynchronous event driven system operating at the time scale of a second. The increase in L1 decision latency has no implications for the DAQ/HLT system, due to the buffering in the front-end pipelines and readout-buffers before the switching networks.

The HLT farm represents a sizeable computing resource for CMS, which is not used for data taking when the LHC has no physics beam (more than 50% of the year). CMS developed a system, based on cloud provisioning techniques, to use the HLT farm for offline processing during the idle periods. This system has been developed in 2013 and is in production since early 2014. As the HLT computing resources for Phase-II represent a sizeable increase with respect to Run-I and 2, we foresee to continue to use the HLT resources for offline computing during idle periods.

7.2 Sub-detector DAQ system

A schematic of the dataflow for a generic Phase-II sub-detector is shown in Fig 7.2. The following functional components can be distinguished:

- On-detector front-end electronics, which for most sub-detectors is based on custom ASICs.
- Off-detector electronics based on FPGAs. For the communication between the detector front-ends and back-ends a new bi-directional standard is being developed which will carry trigger and timing signals in addition to readout data. This link system is an LHC-wide development and consists of on-detector and off-detector

Table 7.2: Parameters of the DAQ system of the Phase-II sub-detectors, and a provisional assignment of DAQ links to the Event Builder. For the forward calorimeter both options (Shaslik and HGCal) are listed, but only one is taken into account for the Event Builder design parameters.

sub-detector	#links on- to off-detector	type	use	data reduction	estimated event size	# DAQ links (100 Gbps)
TK-outer	15 k	GBT (3.2 Gbps)	DAQ + Trig 20%+80%	on-det.	0.5-0.6 MB	100
TK-pixel	2 k	GBT (6.4 Gbps)	DAQ only	on-det.	0.7-1.0 MB	200
ECAL-barrel	12 k	GBT (3.2 Gbps)	streaming	off-det.	1.2 MB	200
HCAL			streaming	off-det.	0.2 MB	40
HB	1296	GBT (3.2 Gbps)				
HF	864	GBT (3.2 Gbps)				
HO	768	GOL (1.0 Gbps)				
Forward-calo						
option Shaslik	12 k	GBT (3.2 Gbps)	streaming	off-det	1.2 MB	200
option HGCal	6 k	10 Gbps	DAQ only	on-det ?	1 MB	200
Muons DT	5760	GBT (3.2 Gbps)	streaming	off-det.	0.1 MB	20
Muons CSC	1080	GBT (3.2 Gbps)	DAQ + Trig 50%+50%	off-det.	0.1 MB	20
Trigger	N/A	N/A	N/A	N/A	0.2 MB	20
Event Builder					≈4.2-4.6 MB	≈ 800

GigaBit Transceiver (GBT) elements connected by a Versatile Link (VL) (see [34] and [35]).

- The FED (Front End Driver)-buffer is the other functional block of the off-detector electronics, which will typically concentrate the data from several GBT receiving elements and perform data reduction, and, depending on the sub-detector, transmit data to the trigger system (not shown).

All functional components discussed so far operate synchronously with the 40 MHz LHC clock. The FED buffer also provides the uniform (across sub-detectors) interface to the event builder, termed DAQ-link. This DAQ-link could be implemented with the FPGA on the FED and the associated Serializer/Deserializer (SerDes) and high speed serial links connecting the FED directly to a commercial switching network, or (as in the present system) a point-to-point connection from the FED to another physical module acting as interface to the switching network. From the DAQ-link onwards, the system is asynchronous and event driven.

Table 7.2 summarizes the parameters of the various sub-detector DAQ systems. It also includes an initial estimate of the expected event size. The event size is dominated by the calorimeters and due to the high occupancy and sampling over many BXs only a small variation (order 10%) in event size between PU of 140 and 200 is expected. This estimate will be refined at the time of the TDR of the sub-detectors when the electronics, data reduction capabilities and expected occupancies are specified.

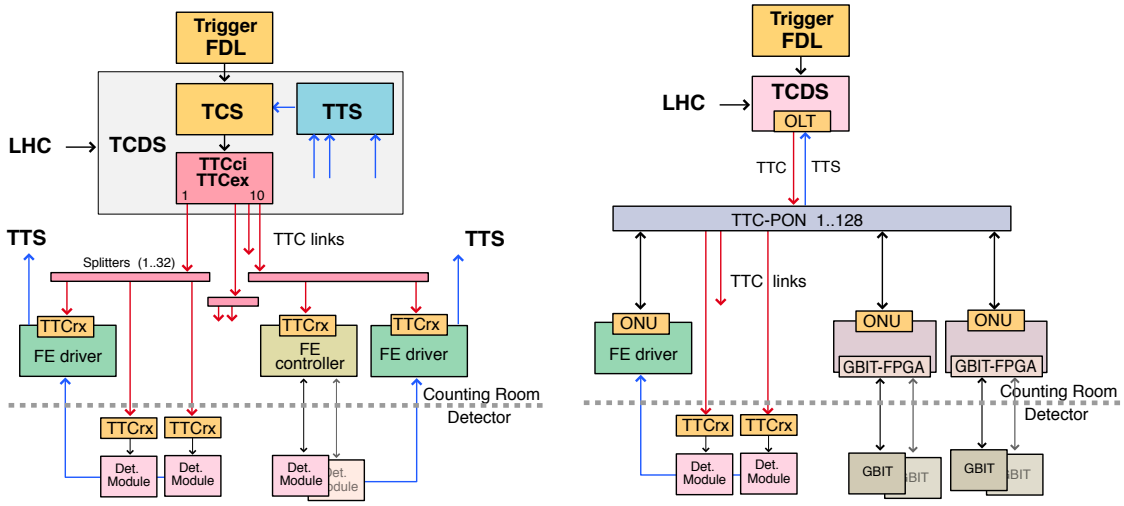


Figure 7.3: TCDS system. Left shows the current system based on the custom (80 Mbps) TTC technology. Right shows the Phase-II system based on PON technology. The TTC and TTS signals are distributed and collected via an OLT (Optical Line Terminal), passive optical splitters (with ratio 1..128) and a number of ONUs (Optical Network Unit).

7.3 Trigger control and distribution system

The trigger control and distribution encompasses the following sub-systems:

- TTC (Trigger, Timing and Control) system. This system is uni-directional and uses optical links to distribute the LHC clock, the L1-trigger accept and various control signals. The TTC system was an LHC-wide development (see [173]).
- TTS (Trigger Throttling System) collects and merges the readout status (such as BUSY) and can throttle the trigger to avoid loss of synchronisation.
- TCS (Trigger Control System), which provides the high level control of the TTC and TTS and interfaces with the trigger FDL (Final Decision Logic). The FDL determines for each LHC BX whether it is a L1 Accept.

For Run-I the TCS was a module integrated in the global trigger hardware (see [174]). The TTC system was based on TTCci and TTCex modules and signals are distributed at 80 Mbps to TTCrx ASICs in the off-detector FE (Front End) driver and FE controller modules and to TTCrx ASICs on the on-detector electronics (see [175]). The TTS is a dedicated system built for CMS with a hierarchy of merging modules connected to the FE driver modules via unidirectional electrical links (see [176]).

The TCDS (Trigger Control and Distribution System), operational for Run-II, combines the functions of the TTC system, TTS and TCS with added functionality compared to Run-I. The new TCDS is implemented in MicroTCA (see [177]) technology, and replaces the TCS in the global trigger and all TTCci and TTCex modules. The TTC links to the front ends are unchanged. The protocol and signal speed of the TTC signals are the same as for Run-I. The TCDS interfaces to the existing TTS system for the legacy FEDs and provides the TTS for the new MicroTCA based FEDs introduced for sub-detectors involved in the Phase-I upgrade.

For the Phase-II upgrade, the bi-directional data links between detector front-ends and back-ends will have a bandwidth of 5-10 Gbps (GBT versatile links), and transmit data readout as well as timing and trigger signals (see Fig 7.2).

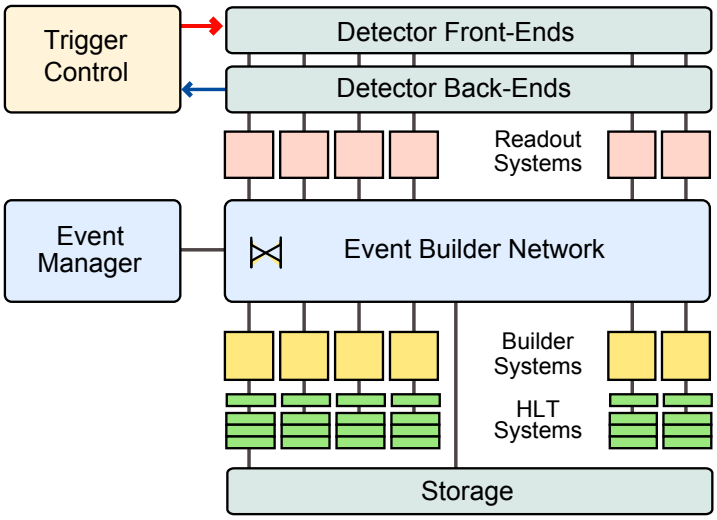


Figure 7.4: Schematic overview of DAQ system.

The TCDS is foreseen to be upgraded for Phase-II to a single bi-directional system with increased bandwidth compared to the current system. The bi-directional links will transmit downstream (TTC) clock, trigger and control and upstream (TTS) status, such as BUSY. The higher bandwidth will allow to send data with each individual bunch crossing (BX), which can be used to steer the on- and off-detector readout on certain trigger types, or to steer the event building.

The upgraded TCDS could be implemented with the standard Passive Optical Network (PON) technology, with extra requirements on clock recovery and fixed latency signals added (see [178]). A possible architecture is shown in Fig. 7.3 on the right side. The system consists of two levels, one between Central TTC system (PON based) to sub-detector off-detector FPGA, and the other between sub-detector off-detector FPGA and on-detector (e.g. using GBT/VL).

The current demonstrator model for the new PON system using 10G-EPON (10 Gbps Ethernet PON) components (see [179]) has a downstream line rate of 11.2 Gbps or $\approx$ 100 times that of the current TTC system with about the same transmission latency ($\approx$100 ns, excluding cables). This will allow transmitting 216 bits of data per BX. The upstream line rate is 2.8 Gbps, where the actual payload and latency depends on the splitting ratio of the PON system.

7.4 Event Building, HLT and Storage

A schematic of the DAQ system is shown in Fig. 7.4. Read-out systems linked to the back-end electronics of the sub-detectors collect and buffer event fragments at the L1 rate. The read-out systems are connected via an event builder network to builder systems, which collect all event fragments belonging to the same L1 trigger and transmit the fully-built events to the nodes of the filter farm running the HLT. Hence, the HLT has access to the full detector information and is only limited by the availabe CPU resources and quality of online calibration. The event building process is steered by the Event Manager. Events selected by the HLT are collected and temporarily saved on a storage system, before transmission to the CMS offline facility.

Given the long lead time to Phase-II, it is premature to present details of a possible implementation that will be highly dependent on technology evolution. Depending on the details of the switching technology, experience has shown that event building can be achieved with an ef-

fective throughput 25% - 100% of the bandwidth of the switching network. The DAQ/HLT will be re-implemented taking advantage of progress in computing and networking industry in the next decade. The DAQ requirement in Table 7.1 correspond to an effective throughput of 30 Tbps at 200 PU for the event building network. Commercial networking equipment with a switching capacity of 50 Tbps and network interfaces operating at 40 Gbps (Ethernet) or 56 Gbps (Infiniband 4xFDR) are available today. The new DAQ system for Run-II is already based on these networking technologies, albeit with a smaller switching capacity (see Section 7.5.2, below). Hence, it appears feasible to anticipate a DAQ system in 2023 with the required throughput. As an example, a system with 800 data sources based on 100 Gbps links, and assuming ≈38% bandwidth efficiency, would provide the required throughput (see Table 7.2). This assigment includes sufficient reserve in the bandwidth of the DAQ links and switch ports to the event building network to accomodate that the data sources can not be perfectly balanced in practice and that commercial network links have discrete transmission speeds.

7.5 Areas of R&D

During the next decade, CMS will track the technology evolution in FPGAs, networking and computing to assure a solid technical and economical solution for the regular replacement due to obsolescence of the DAQ/HLT equipment up to Phase-II. All those hardware technologies follow the trend of Moore's law (the observation that the number of transistors in a dense integrated circuit doubles approximately every two years). Hence, a very significant performance increase over the decade between LS1 and LS3, is expected. Furthermore, there is also a rapid evolution of software technologies in general and experience has shown that evolutions in hardware technologies often necessitate upgrade of software as well in order to fully exploit the full potential of underlying hardware.

7.5.1 FPGA and Links and SerDes

FPGA and the associated links and SerDes are used for the connection between the back-end electronics and the central DAQ and the interfacing to standard commercial IT equipment. High speed I/O standards (such as Ethernet, XAUI, SATA, PCIe, Fibre Channel, Infiniband) have evolved from 1 Gbps in 2000, 2.5 Gbps in 2005 to 10 Gbps in 2010 per lane. Several lanes are typically combined to obtained a higher bandwidth data stream, for example Infiniband 4xQDR combines four 10 Gbps lanes with 8/10b encoding to obtain an effective 32 Gbps data rate link. FPGA and associated links and SerDes show a similar trend in recent years. The current generation of FPGAs features several 10 Gbps links.

R&D includes the evaluation of new products with higher I/O bandwidth, the study of custom protocols to transmit event data, the feasibility of implementing standard protocols in an FPGA in order to interface directly to standard commercial switching networks and computing nodes (both Ethernet and HPC interconnects).

7.5.2 Networking for event building and event distribution

Networking technology is relevant for event building and event distribution and collection from HLT. The two main technologies of interest to us are Ethernet and HPC fabric interconnects. In the case of Ethernet there is a trend in the market towards two directions: expensive carrier class routers with large buffers and many ports; or network of relatively small switches with small buffers and limited features. In the case of HPC fabric interconnect multiple standard or propriety products are competing on the market.

The CMS DAQ system in Run-I ([180] and [181]) employed Ethernet (multiple 1 Gbps links) and the HPC interconnect Myrinet (2.5 Gbps). The new DAQ system for Run-II [182] took advantage of the progress in networking technology. For data concentration 10/40 Gbps Ethernet is used, as well as an implementation of a reduced TCP/IP in FPGA for a reliable transport between custom electronics and commercial computing hardware. The HPC Interconnect Infiniband (4xFDR 56 Gbps) is used for the event builder network. The evolution of network technology will continue and evaluation of new products is foreseen.

One of the main objectives is to study the throughput that can be achieved for the event building traffic. The event building traffic does not resemble the typical traffic in a data center and is also not typical for a HPC cluster which is mainly concerned with low latency. A further area of study investigates the effective use of the bi-directional links in network, considering that event building traffic is essentially uni-directional.

7.5.3 Processor nodes for DAQ

The event builder consists of processor nodes connected to switching networks. These processor nodes perform the input/output and buffering of event data and the control of this process. CMS used various generations of general purpose (x86) servers for this. The latest generations are dual-socket based servers featuring the Xeon processor line, which has now 40 PCIe-Gen3 lanes and up to 8 physical cores per socket and a NUMA (Non Uniform Memory Access) architecture. There is a trend towards integration of components in to the same Xeon die (CPU itself, memory controllers, PCI interface, GPU), which might lead to a further incorporation of the network interface in the future. Experience has shown that extensive effort in software development and testing is required to fully exploit the full potential of the underlying hardware. R&D includes study of I/O performance, NUMA, multi-core, use of co-processors for partial event building and effective utilization of very high bandwidth network interfaces.

7.5.4 Processor nodes for HLT

The HLT code is common with the offline software and typically uses optimized algorithms providing sufficient precision with a small execution time. General purpose servers based on dual-CPU (x86) configurations were the main platform for both HLT and offline processing in the past decade. Improvement gains in processing power have come mainly from an increased number of cores, a trend that appears continuing. However, there are also developments towards specialized co-processors and vectorization, which might provide more economic solutions. The CMS software and computing project is engaged in re-designing the offline software in order to fully exploit many-core architectures and support alternative processor architectures. This involves significant changes to the framework to support execution of algorithms in parallel threads, and to the algorithms themselves to implement finer grain parallelization and vectorization. Specialized hardware and architectures are likely to be deployed first in controlled environments, such as the HLT farm, where the hardware can be controlled and specified. R&D includes the integration of HLT processor nodes based on alternative architectures.

7.5.5 Storage

The DAQ system for Run-II changed to an architecture where the HLT processing is entirely file based. This allows the DAQ and HLT systems to be independent and to use the HLT software in the same way as for the offline processing. The builder systems (see Fig. 7.4) record fully built events on a file system based on ram-disk technology as temporary storage, which is made accessible to the HLT nodes connected to a particular builder unit with 1/10/40 Gb Ethernet

via NFS network file system. Hierarchical collection of HLT accepted events and monitoring meta-data are stored in a global file system. The requirements for the temporary storage file system of each builder system is a throughput of ≈3 GB/s for simultaneous write and read and a size of ≈256 GB. These requirements were met with ram disk technology. However, SSD technology is improving both in terms of I/O throughput, as well as endurance.

The DAQ storage system in Run-I was able to record files containing events selected by the HLT with a throughput of about 2 GB/s. It was implemented with 16 dedicated computer nodes operating in parallel connected to SAN (Storage Area Network) devices. For Run-II a more standard solution was chosen based on a cluster file system server providing a global file system accessible by the DAQ nodes. The requirements on storage for Phase-II are a sustained throughput of ≈40 GB/s and a size of about 2 PB/s. This is an order of magnitude increase compared to the current system. Systems meeting these requirements are commercially available today, but expensive. Also here, price-to-performance is expected to evolve significantly in the next decade.

7.5.6 Software

R&D includes areas of network technologies, protocol stacks, web technologies, databases, cluster file systems, expert systems and online failure prediction methods, and data analytics tools.

7.6 Project planning and cost estimate

7.6.1 Project

The Phase-II DAQ upgrade is handled by the DAQ project. The DAQ project is responsible for the development, operation, maintenance and upgrade of the DAQ system and the HLT computing infrastructure. At present, the DAQ project constitutes together with the L1-trigger project the TriDAS system area within CMS. The HLT algorithms and operation are the responsibility of the Trigger coordination area in CMS.

The requirements for the DAQ/HLT for Phase-II appear feasible from the technical point of view and are likely obtainable within a reasonable budget. The DAQ/HLT system needs full connectivity to all sub-detector back-end electronics electronics from the outset. However, the size of the event builder and/or HLT farm can be adapted easily to a configuration appropriate to CMS wide requirements with a short lead time.

Several small scale demonstrator systems will be constructed focussing on each of the areas outlined in section 7.5 and in addition a medium sized demonstrator system to adress integration and scaling. This work will start during Run-II and continue up to LS3. Technology tracking and evaluation with demonstrator systems will be an ongoing activity of the DAQ project in the next decade in order to deliver a cost-effective implementation, both for the regular 5 year equipment replacement cycle as well as the major Phase-II upgrade.

A Technical Design Report is foreseen during the second half of 2020 about four years before initial deployment, starting in 2024. The TCDS and DAQ systems will be procured and fabricated in 2022-2023 while the HLT systems will be procured in 2025 to benefit maximally from pricing trends.

7.6.2 Cost estimate

A preliminary estimate for the CORE costs is presented in Table 7.3.

Table 7.3: Estimated CORE Cost for TTC/DAQ/HLT systems, for a 200 PU operating condition.

Estimated CORE cost in MCHF (2014)	
TTC system	1
DAQ read out, network and storage	5
HLT computing nodes	11
Total	**17**

The cost of the TTC system is based on the costs of the Run-I system and its upgrade for Run-II.

The DAQ read-out, network and storage include:

- The read-out of the sub-detector back-end electronics.
- The event building processor nodes and networks.
- The event distribution to the HLT farm nodes.
- The collection and storage of accepted events.

The estimate is based on the costs of those systems for Run-I and Run-II and assumes a factor of $\approx$20 performance improvement over 10 years at fixed costs, as was observed in the recent past.

The costs of the HLT resources are estimated using the actual costs incurred in 2013 extrapolated to 2025, assuming a factor 1.25 performance/cost improvement per year, as has been established for dual-CPU (x86) general purpose servers (see [183]). The costs for HLT compute nodes scales with the computing power required (HS06) which would amount here to 11 MCHF estimated for the case of 200 PU. This is based on an expected costs of 1 CHF/HS06 in 2025, which has a substantial uncertainty. Those servers were the main platform for both HLT and offline processing in the past decade. These gains in processing power have come mainly from an increased number of cores, a trend that appears continuing. However, alternative architectures may provide more economic solutions.

Chapter 8

Software and Computing

8.1 Introduction

As described in chapter 6 and 7, offline software and computing systems will be expected to process, archive, and serve for analysis 5 to 7.5 kHz of complex events, and to create a corresponding simulation sample during the Phase-II period. In this Section we estimate the required resources, based on the current processing and analysis model, and we describe the elements of an R&D program to adapt to changes in computing technology. With a successful completion of this R&D program, we expect that CMS will be able to carry out the HL-LHC physics program within a similar growth of the computing resources as currently planned.

The effective growth in CPU power for a fixed cost is currently 25% a year as explained in the WLCG Computing Evolution document [183]. This corresponds to a doubling every three years. In the three doubling cycles between the beginning of Run-II and the beginning of Phase-II we would expect an increase in processing of a factor of about eight assuming a flat spending profile.

The historical increase in LHC computing and disk storage is shown in Figure 8.1. The total CPU capacity has increased at a rate of 360 kHS06 per year and the disk at 34 PB a year on average. The roughly linear increase is the combination of three separate periods. The years 2008 until 2010 were the procurement ramp for LHC as the scale of the available system was tested and commissioned. The period from 2010 to 2013 corresponds to the first LHC run, where the computing and storage increased at a rate defined by the incoming data to process and analyze. The resources needed to accommodate the higher HLT output rate and event complexity expected in Run-II are those foreseen in 2015. The linear growth was related to availability of funds, and not only driven by technology improvements and a flat spending profile. It has to be noted that in addition to the uncertainty in projected resources, it is difficult to predict the evolution of computing capabilities on the time scale of Phase-II.

This chapter discusses the activities that are ongoing to adopt current new technology trends and how those are expected to evolve toward Phase-II. Achieving a system, which is both deploy-able with the resources and support expected and capable of handling the trigger rate and event complexity expected, appears possible, but it will be challenging and will require a sizable amount of developments.

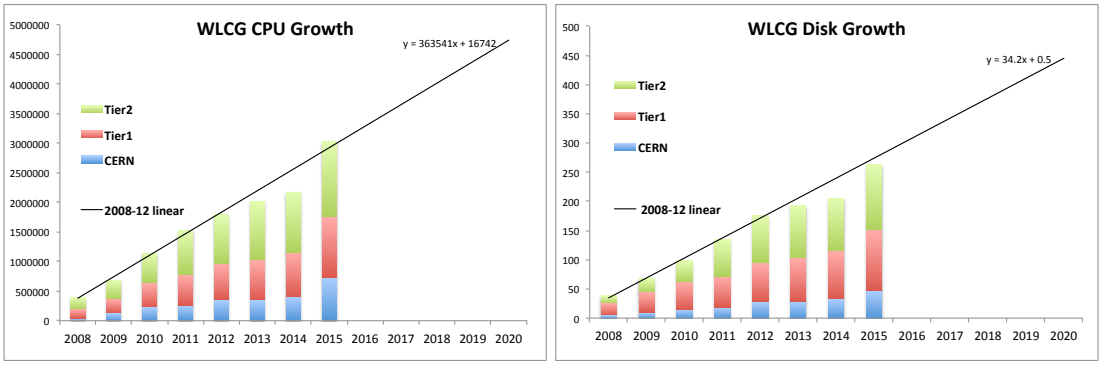


Figure 8.1: Shows the CPU and disk growth through the first 7 years of the program (Expected growth for 2015 shown).

8.2 Event processing and simulation resource needs for Phase-II

8.2.1 Event reconstruction

There are four primary drivers in the resource usage for event reconstruction: output rate from the CMS HLT, event complexity (pileup), choices in the acceptable level of physics algorithm performance (for instance the requirements on the minimum track p_{T} to be reconstructed), and characteristics of the detector itself (e.g., channel count and complexity of the detector geometries). In this section we summarize the computing resources needed assuming no significant changes to our current processing model, which includes fully reconstructing all events saved by the HLT into a standard event data format.

To derive resource needs for the Phase-II CMS detector upgrades, we assume that the algorithm performance for physics objects must remain unchanged compared to Run-II. We use a standard $t\bar{t}$ simulated event sample to estimate resources, and quote results as ratios to the current Run-II resource projections, which include all simulation and reconstruction application developments for Run-II to date. In each comparison, we consider the current detector with Run-II beam conditions (25 pileup events per crossing and 25 ns bunch spacing), the Phase-I detector with Phase-I beam conditions (50 pileup events per crossing), and the Phase-II detector with the Phase-II beam conditions (with either 140 or 200 pileup events per crossing). The Phase-II detector considered includes the tracker and muon upgrades. All performance numbers are derived from event samples created by the CMS Geant4 based detector simulation application, and event digitization and reconstruction run on those samples.

Table 8.1 summarizes the increase in CPU requirements and compressed size per event after the CMS reconstruction. While the Run-II reconstruction application is currently under development, our expectation is that the current software performance will not change noticeably due to this development. The event size of the analysis object data (AOD) format also varies depending on the detector and pileup conditions. Contrary to the situation for CPU needs, the size per event of the AOD primarily used by CMS analysis users for Run-II is less certain. CMS has recently developed a substantially smaller AOD format for Run-II, potentially only 10% of the Run-I AOD size. As the change to the mini-AOD within the analysis groups is still ongoing, we consider the Run-I AOD as the baseline for resource estimates for Phase-II. However, the ratio between the current situation and the Phase-II are not expected to differ significantly between the Run-I and Run-II AOD format. Therefore, we use the Run-I AOD format as the basis for our disk resource requirements for analysis data.

As anticipated, we observe large increases in resource requirements for both CPU and storage

Table 8.1: Summary of reconstruction and AOD resource requirements per event relative to those in Run-II.

Detector	Pile-up (Ave./crossing)	Reconstruction time (Ratio to Run-II)	AOD size (Ratio to Run-II)
Phase-I	50	4	1.4
Phase-II	140	20	3.7
Phase-II	200	45	5.4

Table 8.2: Summary of detector simulation and digitization resource requirements per event relative to those of Run-II.

Detector	Pileup (Ave./crossing)	Detector simulation (Ratio to Run-II)	Digitization (Ratio to Run-II)
Phase-I	50	1	4
Phase-II	140	1	9
Phase-II	200	1	13

resources due to the high pileup in the Phase-II scenario. An important caveat in the CPU resource estimates in Table 8.1 is that the reconstruction software used for the projections presented here is largely unchanged from that developed for the Run-II pileup conditions. CMS has undertaken work to improve both the CPU performance of the reconstruction algorithms, particularly those used in tracking and the global event reconstruction, for both Run-II and Phase-II pileup conditions. Work on rewriting algorithms specifically to improve technical performance characteristics and the use of new C++ compilers, including both integration of C++11 standards and advanced compilation options, continues to yield considerable improvements in performance. However, work on Phase-II conditions specifically has been largely limited to the few algorithms where particularly poor scaling behavior was observed. We anticipate that further work on the performance improvements specifically on the Phase-II conditions will reduce the reconstruction time by about a factor 2 while maintaining the current physics performance.

8.2.2 Event simulation

Table 8.2 summarizes CPU requirements for the detector simulation and digitization for simulated event samples. Contrary to the event reconstruction, the detector simulation resource requirements do not change with detector conditions. This is due to the choice made in CMS to simulate pileup events in the digitization algorithms by overlaying events. Similar to the developments on-going in the reconstruction algorithm work, the CMS simulation is under considerable development to reduce the per event resource requirements. Within CMS we have implemented novel sampling techniques for shower development simulations and other purely technical improvements to gain approximately a factor of two in CPU needs. These improvements are already part of the CMS Run-II simulation application. In addition to this, considerable research and development has started within the Geant4 collaboration towards a redesign of Geant4 with considerably better technical performance [184].

Due to the pileup simulation, the digitization of the Geant4 output to create the simulated raw data scales linearly with the product of the average number of pileup events per bunch crossing and the number of bunch crossings that must be simulated. CMS currently simulates bunch crossings before and after the one of the simulated hard interaction, 16 in total, in order to fully

Table 8.3: Estimated CPU processing needs as ratios of those in Run-II. The ratios include both the increase in HLT output rate and processing time.

	HLT output	Data	Simulation			
Detector	rate (kHz)	Reco.	Detector sim.	Digi.	Reco.	Total
Phase-I	1	4	1	3.5	4	3
Phase-II (140)	5	100	5	47	100	65
Phase-II (200)	7.5	340	7.5	100	340	200

capture the calorimeter electronics response. The results in Table 8.2 assume the simulation of 16 bunch crossings will continue to be required. The software has already been redesigned in order to be able to scale to as many pileup events as necessary within an essentially fixed memory budget. The CPU time within the digitization is dominated by the detector specific digitization algorithms and thus can depend significantly on the final Phase-II configuration of the CMS detector.

8.2.3 Reconstruction and simulation resource needs

Table 8.3 summarizes the total CPU requirements for reconstruction and simulation assuming no substantial changes to the CMS analysis model. We also assume that the accelerator operation time is the same as in Run-II and that the size of the simulated samples continues to increase according to the HLT output rate. These assumptions imply that 5 to 7.5 times more events per year for both real data and simulation samples will be processed offline in Phase-II. The total column is a weighted average of the processing steps based on how frequently each step is run. Assuming a factor of eight growth in computing resources and a factor of two improvement in pure technical changes to algorithm performance, this leaves a deficit of a factor 4 to 12 to be gained from progress in computing capabilities, for operation between 140 PU and 200 PU.

The increase in total AOD on disk is roughly a factor 20(40) for 140(200) PU, as derived from Table 8.1 and including the increase in HLT output rate. We have not explicitly included the CPU requirements for analysis (user) jobs. These currently consume a substantial and roughly constant fraction of the CMS CPU resources. However, because they are not standard application configurations, their scaling into Phase-II is more difficult to estimate. We have therefore assumed that these resources will scale in a manner similar to that of the reconstruction application CPU requirements and AOD output sizes.

8.3 Computing Technology Evolution

Computing for the LHC is largely based on ideas from the MONARC distributed computing models [185]. Over the nearly 15 years since the original ideas were developed there have been refinements and substantial increases in scale, but many of the initial ideas about how resources and data would be accessed are still visible. The strict hierarchy of site capabilities proposed has slowly evolved to allow more flexibility. Sites can now share work-flows and activities initially reserved for the host lab computing, such as reconstruction workflows carried out at Tier-2 centers.

We expect this evolution to continue with increased heterogeneity in the type of resources and architectures used, who provides them and with what expectations of services, and how data is served to them. To improve computing performance, CMS will both need to adapt to these new resources and to investigate use of new technologies.

8.3.1 Processing

During the entire LHC program up to now, nearly all calculations have been performed on a generation of processor based on the x86 architecture. With the primary Linux platform, the x86 architecture has been the workhorse of the LHC computing and, while the Worldwide LHC Computing grid is highly distributed, it has allowed proper compatibility of various hardware and operating systems. For upgrades, the LHC computing will need to evolve to use alternative and more performant computing architectures and platforms following the focus of industry developments. CMS is carefully tracking these new trends.

Recent industry investments are in areas of GPUs, specialized co-processors and low power applications for mobile computing. All of these represent challenges and potential gains. Using GPUs introduces significant diversity to the system, complicates the programming, and changes the approaches used in scientific calculation, but can increase performance by orders of magnitude for specific calculations. Co-processors have similar potential improvement gains, increase in the diversity and complexity of the system, and additional programming challenges. Low power mobile platforms are most interesting when they are combined into massively parallel, specialized system where a single device may have the same number of cores as a remote computing center may have today. These systems would be used rather as a super-computer than as a batch farm, which will require the field to grow expertise in this much more interconnected computing environment.

Specialized hardware and architectures are likely to be deployed first in extremely well controlled environments, like trigger farms and other dedicated centers where the hardware can be controlled and specified. The next step is likely to implement schedulable dedicated specialized systems to permit large-scale calculations, achieving similar goals as in a super computer center request. Large scale clusters of specialized hardware owned by the experiment are likely to arrive at a later stage, and only if they can replace a complete class of computing resources and improve performance at a reduced cost.

A complementary trend is the evolution of computing as a service as in "cloud" solutions. Currently commercial offerings, academic resources and opportunistic resources are all being offered through cloud provisioning techniques. While commercial solutions are still more expensive than well used dedicated resources, there is a steady decrease in costs. Academic clouds function largely like other academic clusters but the cloud environment expects the user to build up more of the services. Opportunistic computing is an interesting area with a growing number of resources with under utilized systems available, particularly at night, being offered for applications that can make effective use of limited or predictable duration. CMS has seen a variety of places proposing cloud-provisioning tools as the primary interface to use the computing. Up to now no site has chosen to contract with a commercial cloud provider to meet the obligations of an experiment, but this too will come if the costs continue to drop and the experiments can make easy access of the resources through the provisioning tools. Small-scale and expensive clusters, without sufficient history of computing, will likely be the first to be outsourced.

By the start of Phase-II, it can be expected that most of the resources will be supplied with a variety of dynamic cloud infrastructures. Already the CERN and the HLT computing resources are operated in this way for about 20% of the CMS computing and there have been several similar proposals from other sites. The evolution to cloud-provisioned resources also encourages the adoption of specialized high-capability dedicated hardware. As specialized hardware is expensive, it may be applicable only to part of the computing requirements, based on a proper scheduling of accesses to be cost effective.

8.3.2 Storage

The disk space needs in CMS scale with the number of events collected (HLT output rate) and their complexity (event size). In 2015 CMS will deal with approximately 10B new events combining data and simulation, with an analysis stream unit of 3PB so that 10 copies/versions could be stored on Tier-2s centers. Considering that multiple copies of the same raw data are required, processed with different software releases, and that the data from multiple years may be required for some analyses, it is clear that many replicas must be hosted and the current model preferentially places jobs where the data is physically located.

In Phase-II, the data volume will increase by a factor of 5 to 7.5 in number of events, with respectively a factor 4 to 6 increase in the event size. Currently the effective growth in disk space at fixed cost is 20% per year as explained in the WLCG Computing Evolution document [183], which means that disk space only doubles every 3.8 years. The factor of 3 or more potential deficiency in disk storage as compared to 2015 indicates that CMS will need to develop techniques that need fewer replicas to accomplish the same processing and analysis goals.

Additionally, while disk capacities have grown, IO and search times have been nearly constant. For active accessed storage it is expected that storage will evolve to solid state devices. These have low search times and excellent IO performance, but also a higher price and smaller capacity.

A technique to reduce the number of data replicas is to serve multiple CPU resources from a single storage source. CMS has invested significant efforts in Run-I optimizing the ability of the processing and analysis application to read data over high latency links while maintaining reasonable CPU efficiency. This has led to the successful deployment of the CMS data federation based on XRootd [186]. In general data federations are intended to serve a portion of the applications, while the majority of the data is served from local storage. Dynamic replication and clean up, as well as predictive placement and network awareness, are needed to enable a data federation to become the primary source of storage access for the majority of the CPU resources.

In an environment where hardware resources are specialized and scheduled it will be important to queue a large volume of data and feed the specialized systems that can quickly process and generate output into a local cache. From a data management and data transfer perspective this is very similar to how clusters are currently operating with datasets transferred in advance. In a cloud environment the concept of local data starts to lose its relevance. Cloud storage can be categorized as size and I/O capacity, therefore, provided the bandwidth is sufficient, storage and processors may not need to be physically close.

To serve a cloud provisioned system the data management system begins to look like a content delivery network (CDN). The CMS data federation being developed is a good first step to a more advanced system. Dynamic replication and clean up as well as predictive placement and network awareness are all needed to enable the CDN to become the primary source of the bulk, non-specialized computing. Moving to a CDN for data management with no expectation from the application for data locality simplifies the use of cloud resources, of opportunistic resources, and of local user controlled systems. It preserves the deterministic pre-placement for specialized systems where the hardware is expected to be too fast to be served over wide area access, or so specialized that the loss of data access could not be tolerated.

Development in commercial computing has focused on delivering content either through CDNs or through peer-to-peer systems. In the next decade computing processing for CMS will evolve to be less deterministic, with more emphasis on cloud provisioned resources, opportunistic

computing, local computing, and volunteer computing. The data management system will also evolve to be much less deterministic to allow efficient use of new and diverse resources.

8.3.3 Networking

CDN systems have the potential to introduce enormous flexibility in data access for a variety of computing systems, but is demanding on the network capacity. CDN systems for video distribution are among the largest users of the networks to residential homes, and one can reasonably expect these distribution systems to be appropriate for some of the largest users of research and education networks. To optimize I/O, currently 50-100kB/s per core is needed for reconstruction and more for analysis. A 10k core processing farm, which in ten years will likely be the average for a Tier-2 center could be served for reconstruction with a 10-20Gb/s connection. Sustaining 10,000 analysis jobs will require a Tier-2 to have a 100Gb/s link if the primary data source is remote. Sites providing infrastructure to serve data to multiple sites will require multiple 100Gb/s export links within a decade. As part of a comprehensive data management system intermediate caches may be automatically populated and used by local systems to enable local access.

Networking is one of the areas with the fastest evolution, with global traffic growing at a rate of 30% per year [183]. Already in 2014 the largest Tier-2 centers are moving to 100Gb/s networks and the first transatlantic networks should have been upgraded by the end of Run-II. Networking enjoyed an explosive rate of growth with the introduction of optical networks. There is no equivalent disruptive technology change on the horizon on the time scale of Phase-II.

8.4 Areas of R&D in software and computing

The resources of the offline processing system of CMS is the limiting factor in the ability to collect data. Sections 6 and 7 show that the trigger and DAQ systems could allow collecting data at a higher rate. In order to push back the limit, offline and computing teams work continuously to reduce the amount of resources needed to perform each of the required tasks. Through code and operations improvements, over the years, CMS has reduced the amount of processing, disk storage, and even archival storage needed for each event. This process involves dedicated R&D programs to continue investigating new technologies and techniques.

These developments are described below. As the resources are limited and the challenges are common, they are carried out in collaboration with the other LHC experiments, through forums like the ECFA workshops, the on-going "concurrency forum" run by the CERN SFT group and new software collaborations currently being defined.

8.4.1 Application software

It is difficult to predict precise directions of computing technologies on the timescale of Phase-II operations. Therefore, we focus CMS software related R&D work on adapting to generic trends in technology rather than on exploiting specific new products available on the market today. This work is critical for our ability to continue making effective use of the CPU processing capability of future architectures as a big focus of computing technology development is now performance per Watt rather than pushing towards ever more powerful processing cores [187]. We will discuss work to develop and exploit multi-threaded algorithms, I/O improvements and heterogeneous computing systems.

A major focus of the R&D is on the software environment for low-power computing architectures, such as Intel Atom or ARM processors, and co-processors, such as Intel Xeon Phi or

graphics processing units (GPUs). The motivation of this work is not to support these platforms for production in the short term, but rather to gain experience in using them on the longer time scale perspective of Phase-II. Short term R&D work will focus on both making needed modifications to CMS simulation and reconstruction to run on these platforms, and small-scale capability demonstrators with an eye towards how the development can then be used in a full-scale CMS reconstruction or simulation application. Evaluating and working to improve actual performance per Watt as well as memory usage of CMS software is particularly interesting for the low-power platform studies. This work has already started to access novel hardware for development and testing, through direct collaborations with hardware vendors and through the CERN-OpenLab structure. It will also evaluate potential benefits or constraints from having computing resources that are considerably more heterogeneous than those used by CMS today.

The primary challenge in this area is to evolve to a parallel application that can take advantage of large numbers of processing resources in a single process. CMS has already developed and deployed a multi-threaded capable event-processing framework [188] capable of making efficient use of a small number of cores. Running a multi-threaded application both reduces the burden on the workflow management system and dramatically improves the memory use characteristics of our application, which is particularly important for most future processor technologies with changes towards more numerous computing cores per CPU. Similarly to the trend towards larger number rather than more powerful processing cores, the DRAM memory capacity and access speed have increased less quickly than the number of processing cores [189]. This becomes an important limitation for continuing the current CMS, and more broadly high-energy physics, processing mode of running one job per core, each using about 2 GB of memory.

To ease its adoption, this framework can run a mix of single-threaded modules (legacy code) and multi-threaded capable algorithms (individual algorithms are known as "modules" in the CMS software framework). Current work is focused on the changes needed to the offline reconstruction algorithms to migrate the remaining legacy modules in order to use this framework efficiently. As Figure 8.2 illustrates, a high fraction of the application must be fully able to use the multi-threaded capability efficiently [190]. If only 10% of the application is not capable of running in parallel, the overall usage of the resources drops to less than 50% before 20 cores are used. Currently CMS can reach 75% resource utilization on 4 cores after only a few months of adapting its reconstruction software to run in a multi-threaded environment, and is tracking well with the expectations of Amdahl's law.

Current work in CMS focuses on using a handful of threads. This number of concurrent threads is sufficient to process an entire luminosity section of data in a single job with an acceptable latency and to fully exploit processing resources while staying within the memory budget per core both in offline and in the HLT. Towards Phase-II, the new R&D will focus on transitioning from a multi-threaded application to a many-threaded one. Independent of computing technology trends, the need for this can be illustrated by scaling up the trigger rate (x 5-7.5) and processing time (x 20-45) while maintaining the requirement that the fundamental blocks of data (luminosity sections) must be processed in one application within the same time period as today. This R&D focuses on the implementation of fine-grained parallelism by identifying sets of work within algorithm modules that can be processed in parallel (e.g., track fitting). In general, some algorithm redesign is needed so that the requirement of a threaded application can be considered from the design phase. One important aspect of this effort is the development of sufficient programming expertise within the experiment to develop and maintain these algorithms through the Phase-II data taking period. Currently used programming techniques

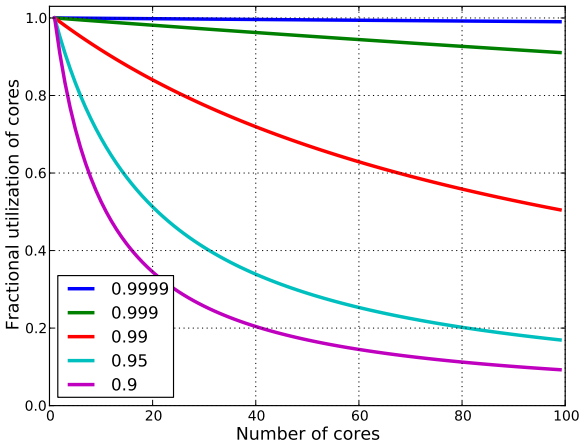


Figure 8.2: Best case CPU resource utilization as a function of the number of computing cores available for different fractions of the application that is fully multi-threaded capable.

within high-energy physics will be inadequate for the development of the critical algorithms in the Phase-II reconstruction application.

Analysis applications make up another substantial component of the computing resources used by CMS. Unlike standard production, there are a large number of processing patterns that change as physics-object identification algorithms are updated. However, a large amount of information is available from any user-submitted analysis job. CMS will reassess physics analysis usage patterns in 2015 as the new AOD format is deployed across CMS Tier-2 sites. On the timescale of Phase-II, it is relevant to evaluate the benefits from alternative file formats, compression and other means of improving the CPU and I/O throughput of real analysis applications. These applications will also naturally benefit from improvements in the reconstruction algorithms towards large-scale use of the CMS multithreaded framework.

CMS has established a discussion forum for evaluating potential research directions, new technologies, and to set up the software development environment to facilitate collaborations. An important component of this collaboration is with software development experts outside of CMS including the broader scientific computing community.

8.4.2 Computing

In a very similar way as application development is focusing on making more efficient use of new architectures and technology, computing development is focusing on improving the overall system efficiency. The triggering system in CMS is capable of writing far more events than computing could afford to process or analyze. Typically the triggering system is limited by the number of events it can reasonably reject and not the number of events it can accept. For Phase-II, the CMS computing organization is considering a number of ways to reduce the scale of the resources required for event processing, to improve the efficiency for processing and analysis of the collected events, and to increase the quantity of computing resources available.

8.4.2.1 The Active Dataset

Events collected in CMS are largely treated in the same way. Datasets are collected, processed, archived, reprocessed, and analyzed completely. A triggering decision takes 100 ms and then each event is treated as having the same physics potential and interest as any other event from

collection through curation. Nearly all events are background and most are later rejected by analysis selections. Some amount of skimming is done for analysis selection, but all events are reprocessed and re-skimmed when needed.

Collecting and archiving events is not expensive; the estimate of the tape cost for a year's worth of Phase-II data at 5 kHz is 400 kCHF from the LHC evolution document [183]. The cost is driven by the resources to process, reprocess, serve, and analyze the events. Currently every event collected in CMS is processed on average 2.5 times and hosted as AOD 4-6 times on disk. The number of events accessed by analysis jobs is harder to calculate, but the average CMS analysis job was four hours in 2012 and there were 57 million jobs completed in total. At a reading rate of 1 Hz, every Run-I simulation and data event was read more than 40 times in 2012.

If CMS were to reduce the actively processed and accessed sample as the events were better understood, it could decrease the scale of the computing resource needs by significant factors. After initial processing an event might be classified as known background and placed in a distribution, but not continuously reprocessed or accessed for analysis. The longer events are analyzed and processed the better understood they are and more efficient choices can be made on their use. The concept of the active evolving dataset takes into account the longer time available to make reasonable decisions on how to treat data. This relatively simple concept is employed already in CMS and previous generations of experiment, but if applied more systematically across the whole experiment could reduce the amount of computing needed.

8.4.2.2 Opportunistic Computing

CMS enjoyed sufficient resource provisioning for computing during Run-I. There was sufficient contingency in the system to handle correcting unexpected problems, and nearly all motivated requests by the experiment could be accommodated. Already in Run-II CMS has very little contingency and difficult choices will need to be made. Opportunistic computing, defined here as any additional resource that is available without an explicit cost, is now being exploited and will continue to be an important contribution for Phase-II. In Run-II the target is to provide 10% of the processing needed by CMS in the form of scavenging opportunistic computing.

The first step was to exploit resources that CMS owns but does not use efficiently all the time. The Higher Level Trigger farm (HLT) was commissioned for offline processing in 2013. The next steps are to aggregate resources that look similar to existing CMS facilities, but are underutilized or could be temporarily assigned to CMS. Data can then be moved to the opportunistic cluster using the data federation. A portion of the 2012 parked data was processed using external computing that could be pledged to CMS for a period of time with no cost.

In order to maintain even 10% opportunistic resources looking forward to Phase-II, CMS will need to expand into architectures and configurations that were not supported in the past. Work is needed in the application, the storage and the data management systems to support a much more diverse set of computing centers moving forward. CMS will need to expand into facilities that have not traditionally supported LHC physics: leadership class super computers, donations of commercial cloud allocations, volunteer computing, and new facilities that will be developed.

8.4.2.3 Improved Techniques

In addition to the two relatively non-controversial ideas discussed above, there are more speculative development activities being organized in the CMS Computing project. A significant change in the overall computing landscape has been the development of "Big Data" tools and

techniques, many of which have not been fully exploited in high energy physics. CMS in Phase-II will produce tens of petabytes of analysis data and simulation each year. The current techniques that are used for reducing the data and selecting events rely on many passes made by individual users. During the next few years the CMS computing project will look at streamlining some of the analysis access to apply techniques developed for "Big Data" tools to improve the scalability of the I/O and decrease the computing needed to reduce datasets.

8.5 Outlook

Assuming only technology improvements and maintaining existing techniques, offline software and computing areas would miss about a factor 4(12) of the resources needed to accommodate the challenging conditions expected in Phase-II at 140(200) PU.

CMS is facing a much more complex computing program for HL-LHC due to both detector changes as well as increased trigger rates and event complexity. We anticipate the need to evolve to be ready to effectively use future commodity computing technologies in order to meet the offline processing needs of the central processing and user data analysis. Internally, CMS has organized a weekly software and computing focused R+D meeting for discussing these issues, work in progress on potential solutions and to foster collaboration across CMS institutions. CMS members are also actively participating in the new HEP Software Foundation (HSF), and are part of the HSF startup team. This has already proven to be a valuable mechanism for collaboration with other LHC experiments. CMS members have also formed active collaborations with CERN (e.g., with CERN TechLab), industry and outside computer science groups.

Areas of initial focus in CMS R+D include:

- Adapting critical components of the reconstruction to heterogeneous, many-core computing resources. Current work in this area includes expanding the set of platforms on which CMSSW is supported, and to prototype new approaches to algorithms, in particular for tracking, redesigned to better utilize a many-core processing capability.
- Evaluation of resource scheduling as well as cost performance trade-offs of using specialized centers for dedicated workflows. We are developing a computing model simulation to evaluate potential mechanisms to optimize data access and to guide how the data federation now in place for Run 2 should evolve.
- Extending the CMSSW event processing framework for heterogenous, many-core computing resources. Initial steps in this direction are to extend the CMSSW multi-threaded framework to support additional fine-grained parallelism mechanisms and to reduce the impact of synchronization points in the data (i.e. luminosity-section boundaries)
- Evaluation of the user data analysis model for HL-LHC. We are considering how alternative data reduction and selection techniques (e.g., techniques from the "big data" community or other research areas) could impact I/O intensive workflows as well as manage the very large data samples that users will need to
- Developing and/or integrating tools and infrastructure components: These include power monitoring profiling tools, and prototyping new technologies such as ways to use containers CMSSW development and distribution.

Chapter 9

Overall Detector Performance

The CMS detector is designed to reconstruct all physics objects with high efficiency, good resolution and low fake rate. These objects are the light quarks (u,d,s, and c) and gluons, manifested as jets; the heavy quarks, bottom and top; the charged leptons, electrons, muons and taus; and photons. Neutrinos are not detected directly but their presence is indicated by missing transverse energy. Since new high mass objects are expected to decay into these physics objects or into other weakly interacting particles whose presence is, like the neutrinos, signified by missing transverse energy, the study of combinations of these objects provides CMS with the opportunity to discover new physics. The purpose of the Phase-II CMS upgrade is to preserve the object quality that will be achieved by CMS right after the Phase-I upgrade is completed, but at the significantly higher luminosity with the correspondingly higher pileup of the HL-LHC.

The CMS strategy for object reconstruction was the basis of its original design. However, during the first running periods in 2011 and 2012, once the performance of the detector was established, many new ideas were developed and led to refinements in the analysis strategy. These developments are expected to be the basis of the analysis in the foreseeable future. New ideas will continue to improve performance beyond what can achieved today.

The approach currently employed is known as "Global Event Description (GED)". Its goal is to provide a complete and accurate description of each hard-scattering event by reconstructing all of its physics objects in a unique manner. This implies making all corrections, including corrections for pileup, so that data analysts can work with high level characteristics of the objects, such as the energy-momentum vectors, without concerning themselves with the details of their alignment, calibration, and reconstruction. The actual method by which the GED is implemented is called "Particle Flow (PF)". The basic principle of this technique is to use information from all available detectors to build a mutually exclusive list of particle candidates that can be used as physics objects themselves or as "seeds" for the construction of higher level (more complex) objects such as hadronic taus, jets, and missing transverse energy.

An important example of the use of PF in CMS is in the reconstruction of hadronic jets. Typically, about 55% of the jet energy is in charged particles; 30% in photons, mostly from π^0s; and 15% in neutral hadrons, mainly from neutrons and neutral kaons. "Conventional" jet reconstruction uses information from the electromagnetic and hadronic calorimeters to identify and reconstruct the jets. However, since the tracking system has better energy resolution for most charged hadrons than the calorimeter, PF uses tracking information, together with the information from the calorimeters, to produce the best measurement of the jet momentum vectors. More information on the PF approach is given in [191].

The use of PF elevates the tracking system to a new importance because it is now involved in the identification and reconstruction of every physics object. In fact, the tracker is an enabling

detector for whole new kinds of analyses, for example, the analysis of merged or "boosted" jets, whose charged component provides important information needed to correctly identify the object and determine its characteristics. The tracker, because of its granularity and its ability to associate tracks correctly to closely-spaced vertices, provides the best information about pileup. For all these reasons, tracking plays a central role in the CMS Phase-II upgrade.

In the following section, the PF approach to the reconstruction of several physics objects is described briefly. In subsequent sections, simulations are used to demonstrate the performance of the CMS detector after the Phase-II upgrade. Most of the results in this chapter are obtained using a full, detailed detector simulation based on GEANT4. The following conditions and detector configurations are compared:

- the Phase-I detector operated at 50 PU to benchmark the required performance goals for Phase-II;
- the Phase-I detector operated at 140 PU with modeling for the radiation damage after exposure to 1000 fb^{-1} of integrated luminosity, for the outer tracker and hadron and electromagnetic calorimeters, to identify the key areas of performance degradation that must be addressed by the upgrades (no degradation is included to the muon detectors); and
- the Phase-II detector operated at 140 PU to evaluate performance reach for the upgrades put forward in this Technical Proposal.

At present, the Phase-II detector simulation includes the upgraded outer tracker, muon systems, and calorimetry. The pixel detector upgrade, however, is still being designed so the simulation contains the Phase-I pixel detector in the barrel and an extended version of the current pixel detector in the forward detector to provide tracking at higher η. The improvements under consideration have smaller pixels in both the barrel and the forward direction and the additional detector stations in the forward direction. The primary and secondary vertex reconstruction and identification performance will certainly be better than what is achieved in the studies presented here.

Maintaining object performance, meaning reconstruction efficiency and resolution, fake rates, and absolute energy/momentum scales, is critical to achieving the full potential of the HL-LHC. A particular analysis may involve several physics objects to reconstruct the signal and reject backgrounds. The figure of merit for CMS is the 'effective efficiency' $\times$ 'integrated luminosity'. The effective efficiency is the total efficiency for all the objects needed in the analysis, diluted by the effect of background. Background depends on object resolution on the signal, which in turn determines how many background events fall into the signal region. Fake rate also increases the background. Vetoes may be needed to reject backgrounds and the effectiveness of these vetoes also depends on object efficiency and fake rate. If typical object performance falls by even 10-15%, the sensitivity to a particular complex final state can easily be degraded by 50% of more, effectively negating 5 years of HL-LHC running. Therefore, the goal of the design of the Phase-II upgrade is for the new detector at pileup of 140 to equal or exceed the performance of the Phase-I detector at pileup of 50.

The challenge to the tracking system is illustrated in Figure 9.1, which shows an event display of the vertex region for a simulated top-pair event with an additional 140 interactions. The challenge to the calorimeters from pileup is even greater since they lack the high granularity and pointing capability of the tracker.

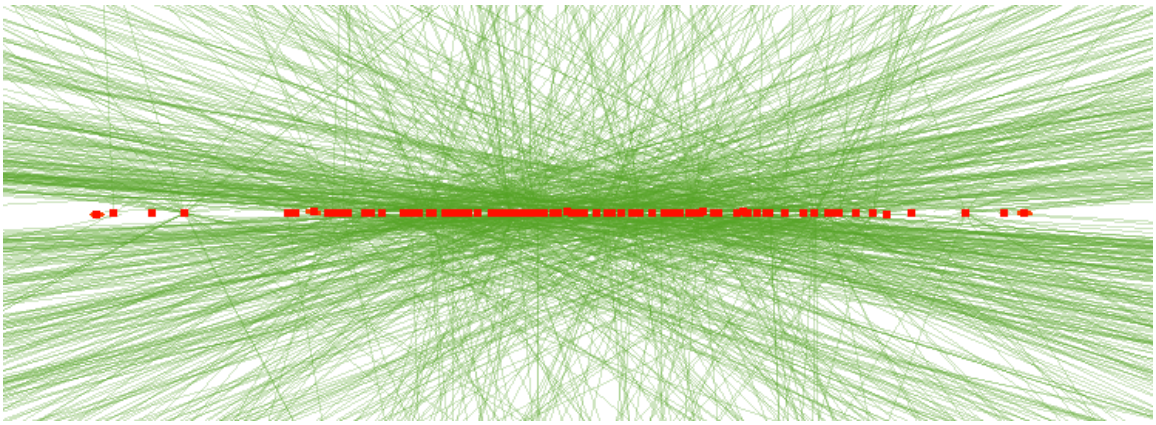


Figure 9.1: An event display showing reconstructed tracks and vertices of a simulated top-pair event with additional 140 interactions overlaid for the Phase-II detector.

9.1 The Particle Flow Algorithm

The Particle Flow algorithm starts with the results of reconstruction performed locally using individual CMS subdetectors. Identification of concentrations of deposited energy, or "clusters", is performed in each of the calorimeters. Track reconstruction is carried out using clusters from the silicon strip tracker and the silicon pixel system. Each muon system produces muon track segments that can be linked into stand-alone muon tracks. Clusters of various kinds and muon track segments along with tracking information are the inputs to the Particle Flow algorithm.

9.1.1 Iterative Track Reconstruction

In order to provide efficient tracking with low fake rate for all physics objects , including taus, needed for the particle flow technique, a new iterative tracking algorithm was developed. The strategy is the following: first, tracks with very pure seeds, for example those having well-defined hits in three layers of the pixel detector, exceeding the requirement of a minimum number of hits in the silicon strip tracker, and satisfying tight vertex constraints, are reconstructed. The charged particle objects obtained in this pass are reconstructed with moderate efficiency and with very low fake rate. Each subsequent iteration proceeds by removing hits associated with tracks found in the previous pass from the pool of available hits and then finding additional tracks by loosening the quality requirements to find a new group of tracks with high efficiency while carefully limiting the fake rate. With this technique, charged particles with as few as three hits and transverse momentum as small as 300 MeV can eventually be reconstructed with acceptable efficiency and low fake rate. Tracks originating from secondary vertices, including those from the nuclear interaction of a hadron or a photon conversion in the tracker material or the decay of neutral hadrons, such as K^0s or Λ^0s, are found by specific algorithms using hits that remain after the first steps of the reconstruction.

9.1.2 Muon Particle Flow Reconstruction

Muons are reconstructed by utilizing information from the tracker and the muon system to create "tracker muons" seeded by inner tracks and requiring matches to muon segments or "global muons" requiring a global track in both subdetectors. Tracks that are identified as muons are flagged in the next steps of the reconstruction. Special criteria may be applied to them to eliminate muons from decay in flight of hadrons or from punchthrough.

9.1.3 Electron and Photon Particle Flow Reconstruction

The PF algorithm starts from calorimeter clusters created out of locally reconstructed calorimeter hits (RecHits). The clustering algorithms implemented for the proposed subdetectors are described in Chapter 3.

The PF clusters from the electromagnetic calorimeter are grouped together by an algorithm that dynamically expands its search space in the bending plane to form superclusters that are used to seed electron and photon reconstruction, taking into account the shower shape due to bremsstrahlung and conversions induced by the material in the CMS tracker. Superclusters become seeds for electron tracks by back-propagating the trajectories from the calorimeter to the tracker creating outside-in tracker seeds for ECAL driven electrons. Tracker seeds can also be created from existing tracks by extrapolating the track tangents in the tracker layers and associating PF clusters that are compatible with the tangent position on the ECAL surface. Those tracks seed tracker-driven electrons, improving the performance of electron reconstruction at low transverse momenta.

After seeds are created, track reconstruction is run with a Gaussian Sum Filter (GSF) algorithm to account for the electron energy loss. All tracks, including electron candidates, are then propagated to the calorimeter surface with a simple helix algorithm.

The next step is reconstruction of electrons and photons by analyzing cluster and track information. Pre-identified electrons and photons are selected and all the linked elements, clusters and tracks, are removed from the pool of components available for the next steps of the PF reconstruction.

9.1.4 Neutral and Charged Hadron Particle Flow Reconstruction

The next step in PF reconstruction is the linking of the individual elements. Tracks, ECAL and HCAL clusters are linked based on their distance on the calorimeter surfaces and PF blocks are formed for all objects that are linked between them.

The blocks are processed by the PF algorithm to create particles. Tracks linked to calorimeter clusters yield charged hadron candidates. In case of energy excess in the calorimeter, compatible with the calorimeter resolution, a neutral candidate is also created by the corresponding cluster. Inside the tracker boundary, an ECAL (HCAL) excess not consistent with a track yields a PF photon (a neutral hadron). On the other hand, outside the tracker coverage most deposits are usually due to charged particles. Therefore neutral hadrons are created by combined ECAL and HCAL deposits. In this Technical Proposal, where the tracker is extended up to $\eta = 3.8$, the boundary is set at $\eta = 3.0$. Beyond this value tracks need to be matched to the forward calorimeter, HF. The coarse granularity of the forward calorimeter implies usually a full jet reconstructed as a PF hit therefore preference is given to neutral hadrons in this range. Below this range, the tracker is followed by the upgraded endcap calorimeters, which have the very high position granularity crucial for PF reconstruction. In the case of non-isolated muons that pass tight muon ID requirements, additional particles are created by any calorimeter excess on top of the muon MIP deposits. The possibility of muon brem is also taken into account by absorbing the calorimeter energy of the muon. A muon that is isolated is identified as a good muon if it passes loose identification criteria. The separate requirements of identification for isolated and non-isolated muons in PF improves performance since it optimally rejects punch through and decays in flight inside jets, while it accepts prompt muons with high efficiency.

9.1.5 Isolation

The requirement that an object is isolated in the event means that the energy flow or p_{T} in its vicinity, exclusive of the object itself, is below a certain threshold. Leptons or photons coming from electroweak interactions are usually isolated, whereas fake leptons and photons and leptons and photons coming from QCD sources, such as jets, are often accompanied by charged or neutral particles. Isolation is usually applied by adding up the p_{T} of extra particles in a cone in $\eta - \phi$ space around the object in question: $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. The radii of isolation cones typically range from 0.2 to 0.5, depending on the level of efficiency and background rejection needed for a particular analysis. To be classified as isolated, the extra p_{T} in the cone must be less than a fraction of the p_{T} of the object itself. Typical analyses use fractions of 10-20%.

Particles from vertices other than the Primary Vertex (PV) can fall into the isolation cone, raising the threshold used to declare the object isolated, thereby reducing the efficiency with which isolated particles can be identified. For charged tracks, it is possible to exclude from the isolation calculation tracks coming from pileup vertices. For neutral deposits in the isolation cone, it is not possible to determine which are from the PV and which are from pileup interactions. Various methods are applied to remove the neutral energy on a statistical basis.

Isolation is discussed in more detail in the sections on the different physics objects.

9.1.6 Particle Flow Reconstruction of Higher Level Objects

The resulting collection of particle candidates produced by the PF algorithm is used in further steps of object reconstruction and analysis. Jets are produced by running special jet clustering algorithms, such as the anti-k_T algorithm, on the PF candidate collection. Missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, is defined as the negative of the transverse vectorial sum of the PF candidates. Isolated leptons and photons are identified directly by PF candidates, avoiding double counting of deposits of tracks in the calorimeters, and in case of boosted decays allowing automatic subtraction of nearby particles from the isolation cones. In addition, b-jets form a special case of jets that have properties characteristic of b-quark decays, such as secondary vertices. W and Z bosons are reconstructed from isolated single leptons and isolated lepton pairs, respectively, and also from pairs of jets. Hadronic taus are built by combining charged hadrons and PF photons to reconstruct the intermediate tau decay modes containing a_1 and ρ mesons. Top quarks are constructed from isolated leptons and jets, with a special role for b-jets, which are present in nearly every top decay. Higgs bosons decay into $\gamma\gamma$, W^+W^-, ZZ^*'s, $b\bar{b}$ quarks, $\tau^+\tau^-$ and other modes, that may be detectable at the HL-LHC, such as $\mu^+\mu^-$. The efficiency and fake rate of these higher level objects, and of new objects that will be sought at the HL-LHC, depends critically on the quality of the more basic objects from which they are constructed and, in many cases, on isolation.

9.2 Vertex Reconstruction Performance

The HL-LHC will achieve its high luminosity by increasing the number of interactions per bunch crossing. The number of crossings will remain fixed at its Run-II value of about 2808, with 25 ns bunch spacing. The luminosity increase will come from lowering the emittance of the beams by improving the whole LHC injection chain, focusing the beams more tightly by replacing the low-β quadrupoles that were built for the original LHC with quadrupoles that have a higher gradient and a larger aperture, and practically eliminating the reduction of luminosity from crossing angle effects by use of Crab Cavities. Moreover, the luminosity will be "levelled", that is maintained at a relatively constant rate throughout the store, either by

adjusting the overlap of the beams transversely at the IR, adjusting the transverse size of the beams so they are larger at the beginning of the store and are reduced as the individual beam currents diminish during the store, or adjusting the crossing angle with the Crab Cavities. The goal is to level the luminosity at $\sim 5 \times 10^{34} \mathrm{s}^{-1} \mathrm{cm}^{-2}$. "Luminosity levelling" reduces the peak pileup that the detectors must handle for a given integrated luminosity.

The operating conditions expected for the HL-LHC will have a profound impact on the detector performance. The IR is expected to have a longitudinal extent of about $\sigma_z \approx 5$cm. The RMS of the beams transversely will be $\sigma_x \approx \sigma_y \approx 7\mu$m. Given the HL-LHC design values for the β^*, the emittance, and the crossing angle, and extrapolating the cross section from 8 TeV to 14 TeV, giving $\sigma_{inel} = 81 \pm 3$mb, the average number of collisions per crossing, referred to as μ, is 135 ± 5. In Run-I of the LHC, the individual bunch crossings exhibited a spread of $\pm 8\%$. With this in mind, CMS has established a "baseline" design pileup of 140 interactions per crossing. This number determines the occupancy of the various detectors and the density of vertices along the z-axis.

Typically, there is only one hard scattering per triggered bunch crossing. This is called the "primary vertex" (PV). The correct identification of the PV in the presence of $\sim$140 other vertices is an important goal and a difficult challenge for the charged particle tracking system. In addition to identifying the PV, it is important to determine as accurately as possible the set of all tracks that are associated with it. In CMS, the set of reconstructed tracks that do not come from the PV is used by special algorithms such as "Charged Hadron Subtraction (CHS)" or "PileUp Per Proton Interaction (PUPPI)"[192] to remove the effects of pileup. Fortunately, a complete reconstruction of all pileup vertices is not required to do this. In fact, some pileup interactions produce only a few low p_{T} tracks and are essentially not reconstructible by the standard CMS tracking program. For some specific studies, special algorithms are used to find tracks at very low p_{T}, but this is not the usual practice in high luminosity, high pileup running. Typically, because of this and because of vertex merging, an average of 90 vertices are reconstructed out of 140 actual inelastic interactions in a typical beam crossing.

With an average of 140 interactions occurring in an interaction region whose RMS length is only 5 cm, there will often be crossings in which tracks from nearby vertices merge and result in the wrong identification of the primary vertex or tracks from nearby vertices are merged with the true PV, causing a shift in its apparent z-position. The correct reconstruction of the primary vertex is also an important step in establishing the presence of secondary vertices, such as those from b quark decays, in the hard scattering event. Reconstruction of states containing b-jets is important for B physics, top physics, Higgs physics, and BSM searches. Finally, a quantity, such as the total number of reconstructed vertices, that has a one-to-one relationship to the total number of pileup vertices, can provide a useful check on whether pileup is properly accounted for by showing that the physics result is independent of it.

As explained above, even at low pileup, a fraction of the vertices do not have sufficient tracks to be reconstructed. Figure 9.2a gives the efficiency for reconstructing and matching all generated vertices (including the pileup vertices) vs pileup. There are two matching criteria: the z separation between reconstrucuted and generated vertex must be <200 μm (the resolution of all vertices is about 70 μm and for signal vertices is about 20 μm as there is a strong dependence on the momentum and number of tracks); and at least 10% of the tracks in the reconstructed vertex must originate from the same generated vertex. The upgraded Phase-II detector reconstructs about 60-70% of the primary vertices, somewhat better than the other two configurations. The slight decline for the Phase-II detector as the pileup increases is the expected behavior since, at very high pileup, more vertices will be close together and will be merged by the vertex finding

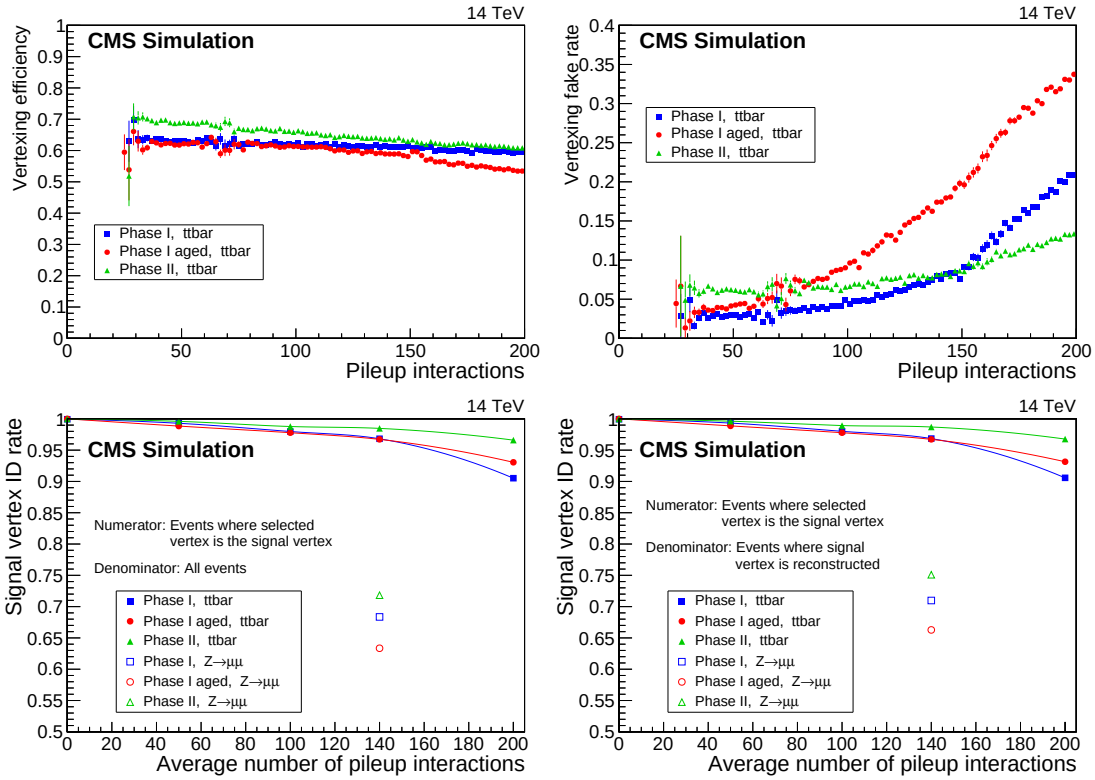


Figure 9.2: For the three detector conditions: (upper left) The efficiency for reconstructing and matching all generated vertices (signal and pileup) vs pileup; (upper right) the fake rate vs the number of of pileup interactions; (lower left) the ratio of the number of events in which the vertex identified as the PV by the new algorithm is matched to the generated signal vertex to the total number of events, plotted vs. pileup; and (lower right) the ratio of the number of events in which the signal vertex is matched to the number of events in which it is reconstructed, plotted vs. pileup

algorithm.

Figure 9.2b shows the fake rate vs the number of of pileup interactions. The fake rate is defined as the fraction of reconstructed vertices in which there is no matching generated vertex, (with the matching as defined for the efficiency). Vertices are also counted as fake when two reconstructed vertices are matched to the same simulated vertex (duplicates). The upgraded detector has by far the lowest rate of incorrect vertex identifications when the pileup becomes very high. The aged Phase-I detector has the worst performance at high pileup. The condition with the best tracking, the undamaged Phase-II detector, shows the best behavior while the aged Phase-I detector shows the worst behavior.

In the Run-I analysis, for the primary vertex reconstruction, vertices are ordered by decreasing value of ΣP_T^2 of their associated tracks. The PV is usually defined as the reconstructed vertex with the largest value of this quantity. However, this method does not work well for the high pileup of Phase-II. A new vertex reconstruction algorithm, which has been developed for high pileup by the CMS b-tagging group, is used in this analysis.

Figure 9.2c shows the ratio of events in which the vertex identified as the PV by the new algorithm is, in fact, matched to the generated signal vertex to the total number of events, plotted vs. pileup for $t - \bar{t}$ events and for $Z \to \mu^+\mu^-$ events. Figure 9.2d shows the ratio of events in

which the signal vertex is matched to the events in which it is reconstructed, which is a type of vertex-finding efficiency, plotted vs. pileup. The PV efficiency for the softer Z events is lower than for the $t - \bar{t}$ events but in both cases, the upgraded Phase-II detector performs better than the other configurations.

Figure 9.3 shows the difference in the longitudinal position, z, of the reconstructed primary vertex and the generated vertex in crossings containing a $t\bar{t}$ event accompanied by a mean of 140 additional pileup events. The left hand side compares this distribution for the no-pileup to 140 pileup case for the aged Phase-I detector. The right hand side compares the no-pileup to 140 pileup case for the Phase-II detector.

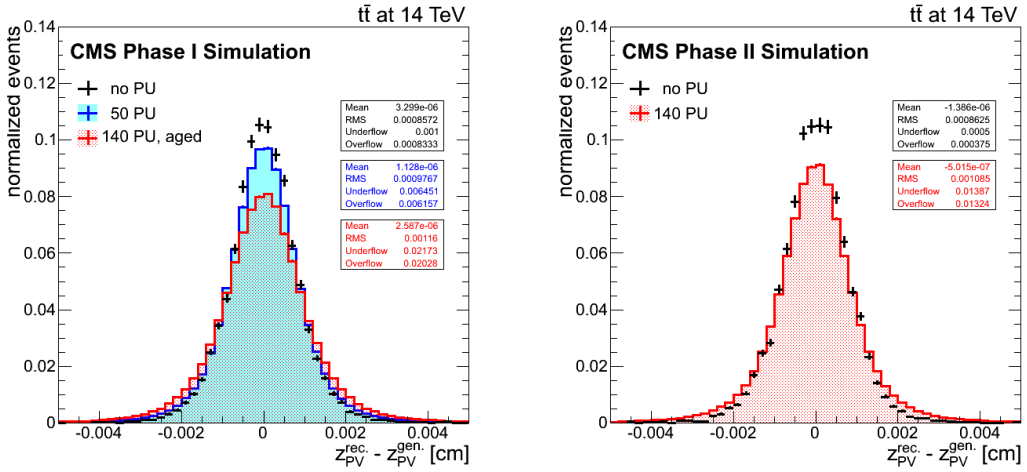


Figure 9.3: a) The difference of the z of the reconstructed primary vertex and the generated primary vertex for $t\bar{t}$ events for the no pileup case and the aged Phase-I detector with mean pileup of 140, and b) The difference of the z of the reconstructed primary vertex and the generated primary vertex for $t\bar{t}$ events for the no pileup case and the Phase-II detector with mean pileup of 140.

In general, the Phase-II detector does an excellent job of finding the primary vertex. The ability to identify tracks that are not from the primary vertex will be seen as the key to defeating pileup from charged tracks.

9.3 Jet and Missing Transverse Energy Performance

9.3.1 Introduction

Jets are the experimental signatures of the production of quarks and gluons in high-energy processes. The nature of the QCD interaction forbids color-charged particles, such as quarks and gluons, to exist and travel in free form and instead forces these particles to fragment into a large array of hadrons and other composite particles that are individually of neutral color. The energy of the initial color-charged particle is spread among all its fragments, which are usually contained in a cone in a well defined location of the detector. The collection of the resulting particles is called a jet. Jets are among the most fundamental observables at hadron colliders and are a fundamental tool in searches for new physics. Neutrinos, as well as hypothetical weakly interacting particles, cannot be directly reconstructed by the CMS detector but their presence can be inferred from a momentum imbalance in the plane transverse to the beam

direction. Usually the magnitude of this imbalance is referred to as missing transverse energy and is denoted by $\not{E}_{\mathrm{T}}$. The need to correctly reconstruct jets and $\not{E}_{\mathrm{T}}$ in the presence of high pileup is one of the crucial design requirements of the Phase-II detector.

The performance of jet and missing transverse energy reconstruction of the three combinations of pileup and detector configuration are studied using the full simulation of the CMS detector, which is based on GEANT4 [193]. Samples of simulated QCD events are generated with PYTHIA6.4.22 [89], tune Z2STAR (CTEQ6L PDF). For each detector configuration we create an additional sample that is identical at generator level but has no added pileup. These samples are used for the calibration of the jets. For $\not{E}_{\mathrm{T}}$ studies, $Z/\gamma^* \to \mu^+\mu^-$ Drell-Yan samples are used with the same generator and detector simulation.

9.3.2 Jet Reconstruction and Pileup Mitigation

Since the HL-LHC is expected to operate with 140 pileup interactions, it is necessary to employ pileup mitigation techniques to maintain jet and $\not{E}_{\mathrm{T}}$ object performance. CMS reconstructs jets using three methods that produce three types of jets called PF, PFchs, and PUPPI jets, respectively. Each approach starts with the output of the particle-flow algorithm [194, 195] described above. These PF candidates are reconstructed quantities that could have originated either from a pileup interaction, from the primary interaction that we are interested in, or from an overlap of energies from particles from the pileup as well as the primary interaction. The better we can distinguish between these cases the better we can mitigate the effects of pileup in a given event.

The PF, PFchs, and PUPPI jets differ in that they perform different operations on these PF candidates before passing them to the clustering algorithm. Jets are reconstructed using the anti-k_T clustering algorithm [196] with a cone size parameter $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$ in (η, ϕ) space, implemented using the *FastJet* package [197, 198].

The performance of reconstructed jets is defined by comparison with the particle-level simulation jets. Particle-level jets are reconstructed by clustering the four-momentum vectors of all stable ($c\tau > 1$ cm) particles generated in the simulation with exception of the neutrinos, which are excluded from the clustering. Each reconstructed jet is spatially matched in the (η, ϕ) space with a particle-level jet by requiring $\Delta R < 0.2$.

The PF, PFchs, and PUPPI jets are described very briefly in the following sections.

9.3.2.1 Particle Flow (PF) Jets

PF jets are simply jets in which the four-momentum vectors of the particle flow candidates are passed directly to the clustering algorithm without any kind of filtering or manipulation. PF jets have been one of the baseline jet benchmarks at CMS for the Run-I period. Figure 9.4(left) (extracted from [199]) shows the average transverse energy deposited in the CMS detector by a single pile up interaction at $\sqrt{s} = 8$ TeV in terms of particle flow candidates. The distribution is expected to remain about the same at 14 TeV. The center and right plots show the expected impact of pileup on jets at a center-of-mass energy of $\sqrt{s} = 14$ TeV .

On average the extra p_{T} added to a PF jet, or "offset p_{T}", from a pileup interaction is composed of 65% from charged hadrons, 25% from photons, and 10% from neutral hadrons. Simulation and data from Run-I show a linear relationship of the energy deposited in the jet cone with the number of interactions in the event. At mean pileup of 140, about 45 GeV is added to a typical jet in the barrel as shown in Figure 9.4. This varies only by about 10% as the jet energy varies from 20 GeV to 1 TeV. In the more forward direction, the offset can be as high as 70 GeV per jet.

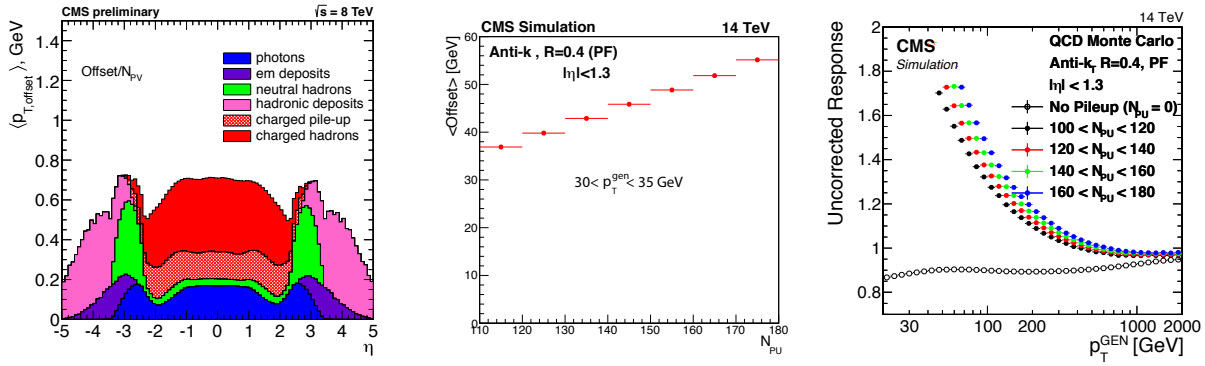


Figure 9.4: (left) Breakdown of the average offset p_{T}, in terms of the PF candidates, as a function of η, for events with one PU interaction; (center) Average transverse energy added to a PF jet as a function of the number of pileup vertices for $30< p_{\mathrm{T}} <35$ GeV; (right) Response of uncorrected PF jets as a function of $p_{\mathrm{T}}^{\mathrm{gen}}$ in the Phase-II sample.

9.3.2.2 Charged Hadron Subtraction (CHS)

Charged Hadron Subtraction (CHS) is the first line of defense against pileup. The CHS algorithm effectively removes charged particles from the list of PF candidates if their tracks can be directly traced to an interaction vertex that is not the defined primary vertex. A track is considered to be "directly traced" if the track's z position is near a pileup vertex within the vertex resolution. Because the vertex efficiency is well below 100%, there are many tracks that are not associated with any given reconstructed vertex and which are, therefore, kept in consideration during clustering. Other alternatives, such as keeping only tracks that point to the primary vertex within some distance, were found in simulation to be overaggressive as they removed tracks that did belong to the primary vertex but had slightly inferior track reconstruction quality. In the current detector, there is no directional information for a given photon or neutral hadron that enables it to be associated with a particular interaction vertex so this method does not help reduce the neutral component of the pileup. "PFchs Jets" are jets that are created after the clustering of the PF-candidates, where charged hadron candidates originating from pileup interactions have been removed. Most of the physics results from Run-I data used the CHS method for pileup mitigation.

9.3.2.3 Pile-Up Per Particle Identification (PUPPI)

For each particle in the event the PUPPI algorithm [192] computes the likelihood that the particle comes from the primary vertex. The algorithm then applies a weight to that particle based on the computed likelihood and passes all the weighted candidates to the clustering algorithm. The resulting clustered jets are called PUPPI jets.

The weight for charged PF candidates that originate from the primary vertex is 1 and is 0 for those from the other vertices. This is equivalent to the CHS method described above. For the neutrals, including charged particles outside the tracker acceptance, the weights are based on the probability of being from a non-pileup vertex determined in the following manner.

First, for each particle i in the event, a variable α_i is computed for all particles in a cone $R_{min} \leq \Delta R_{ij} \leq R_0$ of its vicinity. Typically $R_{min} = 0.02$, and $R_0 = 0.3$. The computation of α_i depends on whether the particle i has tracking information (equation 9.1) as in the central region of the detector or whether is has not (equation 9.2) as in the very far forward region, $3 < |\eta| < 5$, which is covered only by the Forward Hadron Calorimeter (HF) in the Phase-I detector. The

difference is that the sum in the first case runs over the charged particles of the primary vertex and in the second case over all particles.

$$\alpha_i^C = log\left[\sum_{j\in \mathrm{CH,LV}} \xi_{ij} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0)\right] \tag{9.1}$$

$$\alpha_i^F = log\left[\sum_{j\in \mathrm{event}} \xi_{ij} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0)\right] \tag{9.2}$$

$$\text{with } \xi_{ij} = \frac{p_{T_j}^2}{\Delta R_{ij}^2}. \tag{9.3}$$

The variable α distinguishes between collinear and soft diffuse structures in the neighborhood of the particle. The collinear structure is indicative of particles originating from the primary interaction.

In each event, the mean of α_i for charged particles not from the primary vertex (i.e. from pileup) is computed. This distribution is considered to be representative of the neutral particles also coming from the pileup vertices. The weight of the neutral particles is derived by transforming the corresponding α_i into a χ^2 probability defined as:

$$\chi_i^2 = \frac{(\alpha_i - \hat{\alpha}_{PU})^2}{\sigma_{PU}^2}. \tag{9.4}$$

Here, $\hat{\alpha}_{PU}$ is the average weight over all pileup particles and σ_{PU} is the square root of its variance. Using the weight associated with each particle to compute its collinearity with surrounding particles results in a strong suppression of the soft and non-collinear pileup particles. This results in improved jet resolution compared to PF and PFchs jets across the entire p_T and η range of CMS. PUPPI performs better at pileup mitigation than CHS and has recently become the preferred method for use in jet reconstruction. In the following, PUPPI jets will be used to demonstrate the capabilities of the Phase-II detector.

9.3.3 Performance of Jets

Several corrections must be made to the raw jet energy and p_T to produce quantities that have the offsets from pileup removed and are corrected for imperfections in the detector and the reconstruction technique. Only then do the derived jet quantities correctly represent the properties of the particle-level jets. These corrections must be performed separately for each algorithm for each combination of pileup and detector configurations in order to allow for a meaningful comparison of jet and $\not\!\!E_\mathrm{T}$ performance.

9.3.3.1 Raw Response

The jet energy response is defined as the mean of the reconstructed jet p_T over the corresponding particle-level jet p_T. The addition of a constant pileup energy dramatically changes the response behaviour over the range of $p_\mathrm{T}^\mathrm{gen}$. At low $p_\mathrm{T}^\mathrm{gen}$ values, where the jet energy is comparable to that of the pileup depositions, the response increases sharply with decreasing $p_\mathrm{T}^\mathrm{gen}$, while at large $p_\mathrm{T}^\mathrm{gen}$ values, where the jet energy is many times larger than the pileup deposition, the effect of pileup is unnoticeable. In this region, the intrinsic response of the detector is clearly seen since the uncorrected jet energy response approaches values that are only a few percent below unity, as shown in Figure 9.4(right).

9.3.3.2 Jet Energy Calibrations

Corrections for jets are derived for each jet type and sample following reference [200]. The corrections are applied sequentially so that the next correction stage is applied on top of the jet already corrected by previous stages. The first correction is for pileup energy depositions, effectively subtracting the average energy deposited in the jet by pileup, bringing the jet to the energy it would have if the sample had no pileup. Second, the jet energy is corrected for relative η variations in the detector response and finally there are corrections for the absolute scale.

The average amount of energy due to pileup, or offset energy, is parametrized as a function of the median energy density estimator ρ as defined in [201]. The offset is computed by comparing the same events in the samples generated with and without pileup energy deposition. Since these samples are identical at generator level the only difference in their reconstructed jets arise from the lack of pile up. Events in one sample are matched to the corresponding event in the other sample, and jets in an event are matched to the corresponding jets in the other event. The pileup corrections bring the average energy deposition of reconstructed jets to the value they have in the no-pileup sample with an accuracy of better than 3%. These corrections are derived separately for PF and PFchs and PUPPI jets and for the three different detector configuration samples. After pileup corrections we apply p_{T}- and η-dependent corrections. These are derived independently for each jet type and for each simulation sample and they are of the order of 10%. The fully corrected jet energy response is well within 2% in the region of $|\eta| < 3.0$ for all jet types and in all samples.

9.3.3.3 Corrected Resolution

The performance of jets in the different detector configurations is quantified by the response resolution of fully-corrected jets. In a given $p_{\mathrm{T}}^{\mathrm{gen}}$ and η region, a Gaussian function is fitted to the distribution of the corrected response of all jets in that region. The jet resolution is taken as the width of the fitted Gaussian. The resolution is then further divided by the average jet response in that $p_{\mathrm{T}}^{\mathrm{gen}}$ and η region to correct for any residual under/over corrections.

In the central region $|\eta| < 1.3$, the resolution degrades for lower values of $p_{\mathrm{T}}^{\mathrm{gen}}$ as expected from the more important role of the pileup energy deposition on those softer jets. At larger $p_{\mathrm{T}}^{\mathrm{gen}}$ values beyond 1 TeV the resolution is about 5% and is the same for all jet types. The resolution significantly degrades in going from the Phase-I 50 PU sample to the Phase-I aged 140 PU sample. This is the combined effect of the larger pileup contribution as well as the aging of the detector in the latter sample. In comparison, the Phase-II 140 PU sample, which in this region has only tracker upgrades with respect to the Phase-I samples, has only marginal resolution improvement for PF jets, which don't take full advantage of this new tracker for pileup rejection. The tracker upgrades are exploited better in the PFchs jets, which are able to regain about 17% of the resolution over the Phase-II samples at $p_{\mathrm{T}}^{\mathrm{gen}}$ =30 GeV. Furthermore, the PUPPI jets are able to regain even more for a 20% resolution improvement at the same $p_{\mathrm{T}}^{\mathrm{gen}}$ value.

The Phase-I detector at 140 PU provides a markedly poorer resolution for all jet types, especially for the low $p_{\mathrm{T}}^{\mathrm{gen}}$ range. The performance of low p_{T} jets has large impact on the $\not{E}_{\mathrm{T}}$ resolution. Comparing the resolution in the more central region to the more forward (endcap) regions, a more significant degradation appears in the endcap region for the Phase-I 140 pileup aged samples, clearly demonstrating the effects of radiation damage in the endcap region of the Phase-I detector. In this detector region the Phase-II 140 PU sample contains an upgraded electromagnetic calorimeter plus a rebuilt hadronic calorimeter, both of which have a direct

effect on the jet energy measurements.

The physical structure of the detector affects directly the jet resolution in different η regions. The plots in Figure 9.5 shows the resolution as a function of η for jets in the region $30 < p_{\mathrm{T}} < 150$ GeV. The Phase I aged 140 PU sample is degraded particularly in the region of $2.5 < |\eta| < 3$ in which it presents two large spikes in resolution. The extension of the tracker in the Phase-II detector together with the new calorimeters significantly reduces those spikes bringing them in line with the resolution seen in the more central part of the detector. This is seen in all jet types and is another performance gain achievable with the upgraded detector of Phase-II.

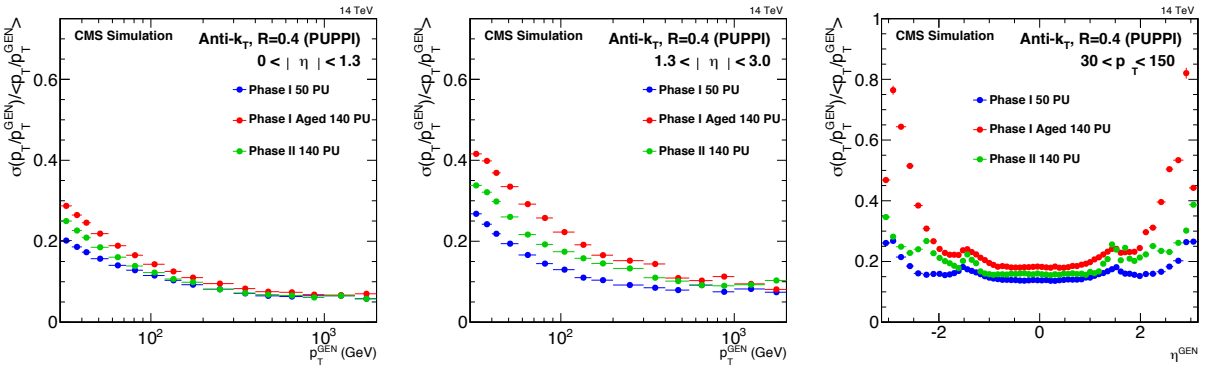


Figure 9.5: Jet response resolution of corrected PUPPI jets as a function of jet p_{T} for jets (left) in $|\eta| < 1.3$ and (middle) in $1.3 < |\eta| < 3.0$; (right) Jet response resolution as a function of η in the range $30 < p_{\mathrm{T}}^{\mathrm{gen}} < 150$ GeV.

9.3.3.4 Pileup-Jet Misidentification

The increase in the number of pileup events results in larger number of jets that can mimic or "fake" hard-scatter jets from the primary interaction. The number of these "pileup jets" depends on the p_{T} as well as on the details of the jet algorithm. We quantify the ratio of fake jets as the average number of reconstructed jets in a given p_{T} and η bin to the average of the number of reconstructed jets that can be matched to a particle-level jet in the same sample. The Figures 9.6 shows this ratio as a function of p_{T} for PUPPI jets. The fake rates more than doubles for the Phase-I aged detector at pileup of 140. The fake rates for the upgraded Phase-II detector are restored to the low values of the Phase-I unaged detector at pileup of 50.

9.3.4 Performance of Missing E_T

Missing transverse energy, commonly known as "MET" and written symbolically as $\not{E}_{\mathrm{T}}$ is defined as the imbalance in the transverse momentum of all visible particles in the final state. $\not{E}_{\mathrm{T}}$ is calculated as the negative vectorial sum of transverse momentum of all particles that interact via the electromagnetic or strong forces with the detector. PF $\not{E}_{\mathrm{T}}$ is used commonly within CMS by summing over all PF particles in the event. Although pileup interactions are not expected to have significant amounts of $\not{E}_{\mathrm{T}}$, it has been observed that pileup interactions have a considerable effect on the MET resolution [202]. By further exploiting the pileup mitigation technique of PUPPI algorithm, PUPPI $\not{E}_{\mathrm{T}}$ is calculated using PF candidates whose transverse momenta are weighted by the PUPPI algorithm as described in Section 9.3.2.3.

The magnitude of $\not{E}_{\mathrm{T}}$ can be under-measured for a variety of reasons, like minimum energy threshold of the calorimeter clustering, inefficiency in track reconstruction, non-linear response of the detector etc. These biases can be compensated by correcting the p_{T} of jets to the particle level p_{T} using the jet energy correction as described in Section 9.3.3.2.

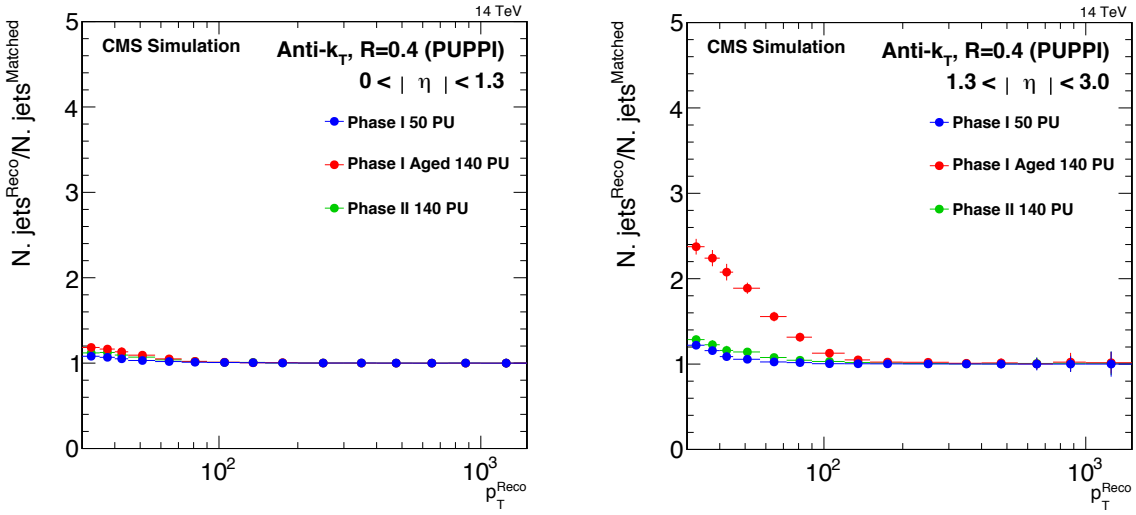


Figure 9.6: Ratio of the number of reconstructed PUPPI jets to the number of reconstructed jets matched to particle level jets for different p_T bins.

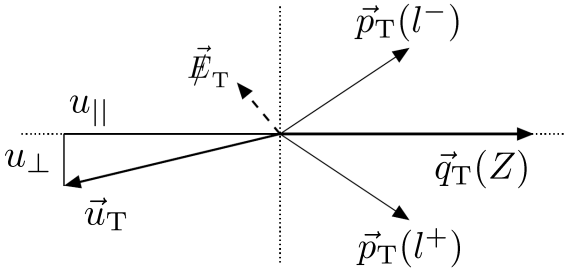


Figure 9.7: Illustration of $Z \to \mu^+\mu^-$ event kinematics in the transverse plane. The vector $\vec{u}_T$ denotes the vectorial sum of all particles reconstructed in the event except for the two muons from the Z decay.

The performance of $\not{E}_T$ is studied using samples of events where a Z boson is produced, reconstructed, and identified in the $Z \to \mu^+\mu^-$ decay. While there is no genuine $\not{E}_T$ in these samples, the PF objects, exclusive of the dimuon, are used to reconstruct the p_T of the dimuon, which serves as a proxy for $\not{E}_T$ and is very well measured by the tracking system.

By comparing the well measured vector boson with the hadronic recoil system, we can measure the scale and resolution of $\not{E}_T$. As illustrated in Figure 9.7, the reconstructed vector boson momentum in the transverse plane is $\vec{q}_T$, and transverse momentum of the hadronic recoil, defined as the vectorial sum of the transverse momenta of the other objects besides the two muons from the Z boson decay is $\vec{u}_T$. Momentum conservation in the transverse plane requires $\vec{q}_T + \vec{u}_T + \vec{\not{E}}_T = 0$. The hadronic recoil can be projected onto the direction of $\vec{q_T}$, yielding the parallel ($u_{||}$) and perpendicular ($u_\perp$) components.

The value of $u_{||}$ is typically negative, since the hadronic system is usually pointing in the opposite direction from $\hat{q}_T$. We refer to $< -u_{||} > /q_T$ as the response of the $\not{E}_T$. Figure 9.8 shows the response of the component parallel to the Z boson. The PUPPI $\not{E}_T$ tends to reach unity more slowly than the PF $\not{E}_T$. The PUPPI $\not{E}_T$ of Phase I 140PU with aging reaches unity much more slowly, comparing to the Phase-I 50PU and Phase-II 140PU.

We refer to the width of the $u_{||}$ and $u_\perp$ distributions as the $\not{E}_T$ resolution. It is taken as the σ

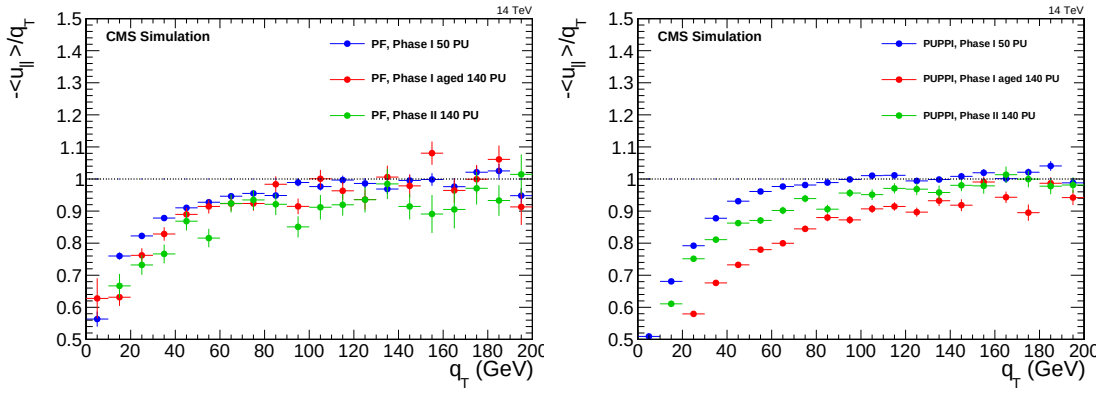


Figure 9.8: (left) PF and (right) PUPPI response curves of the hadronic recoil component parallel to Z boson as a function of Z boson q_{T}, measured in $Z \to \mu^+\mu^-$ events in Phase-I 50PU no aging (blue point) , Phase-I 140PU with aging (red point), and Phase-II 140PU (green point) samples.

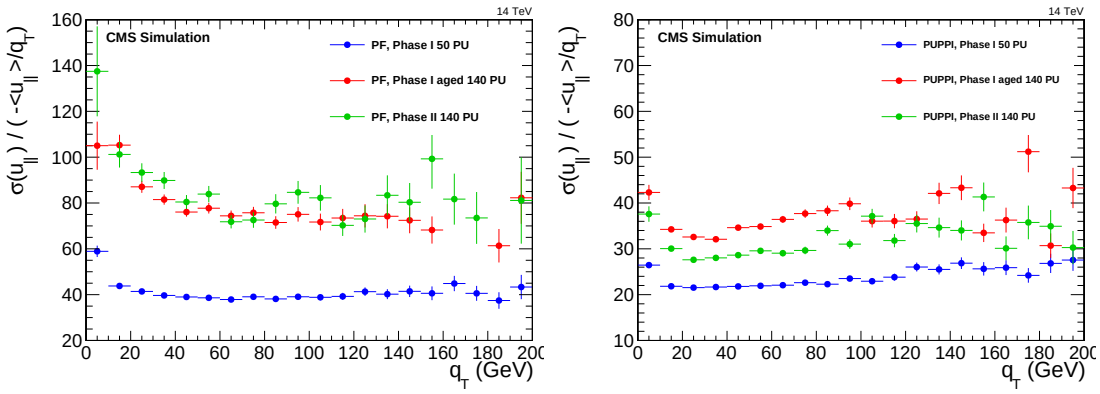


Figure 9.9: (left) PF and (right) PUPPI $\not\!\!E_{\mathrm{T}}$ resolution curves for the parallel component of hadronic recoil to Z boson, measured in $Z \to \mu^+\mu^-$ events in Phase-I 50PU no aging (blue point), Phase-I 140PU with aging (red point), and Phase-II 140PU (green point) samples.

of a Gaussian function fitted to the $u_{\parallel}$ and $u_{\perp}$ distributions, which have been corrected for $\not\!\!E_{\mathrm{T}}$ response. Figure 9.9 shows the resolution of the parallel components of PF and PUPPI hadronic recoil. Because of the smearing from pileup, we don't observe a strong q_{T} dependent for the resolution curves as observed in [202]. The Phase-II detector has similar PF $\not\!\!E_{\mathrm{T}}$ resolution as the Phase-I with aging at 140PU, because of the large pileup smearing effect. With the PUPPI pileup mitigation, the Phase-II detector with 140 pileup has $\not\!\!E_{\mathrm{T}}$ resolution that is a factor of two better than is achieved with PF $\not\!\!E_{\mathrm{T}}$.

Particle flow and pileup mitigation rely heavily on tracking. In the Phase-I detector, there is no tracking to help in the forward direction, $|\eta| > 2.5$. The Phase-II upgrade addresses this problem by extending tracking to $|\eta| \sim 4$. The effectiveness of this forward tracking is shown in Figure 9.10, which compares the distribution of $\not\!\!E_{\mathrm{T}}$ in Drell-Yan events for the Phase-I detector with the identical detector except that it has the tracking extended in η. Since this sample should have relatively little $\not\!\!E_{\mathrm{T}}$, the tracking extension has clearly reduced the false $\not\!\!E_{\mathrm{T}}$ by a large amount.

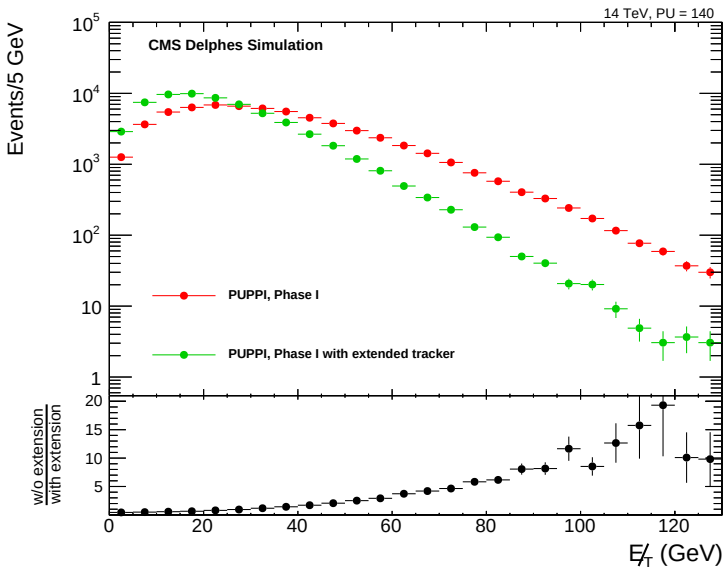


Figure 9.10: The $\not{E}_{\mathrm{T}}$ distribution for the Phase-I detector with and without the high-η tracking extension; and (lower) the ratio of the number of events observed without and with the tracking extension vs $\not{E}_{\mathrm{T}}$.

9.3.5 Summary of Jet and Missing E_T Performance

This study shows that there would be significant degradation of the jet resolution if the detector is not upgraded and that there will be a significant recovery of that resolution if the Phase-II upgrades are carried out. In particular, for PUPPI jets the recovery will bring the resolution almost to that achieved in Phase-I with 50 PU running conditions in the regions where both calorimeter and tracker are being upgraded. At the same time Figure 9.5 shows that resolution spikes in the region of $2.5 < |\eta| < 3.0$ observed in the Phase-I detector configuration are not present in the Phase-II detector. Furthermore, the number of fake jets produced by pileup can and does complicate the analysis of an event, especially at low jet p_{T}. The Phase-II detector also improves the response and resolution of missing transverse momentum, compared to Phase-I 140PU. The PUPPI $\not{E}_{\mathrm{T}}$ of Phase-II detector can reach a similar resolution as the PF $\not{E}_{\mathrm{T}}$ of the Phase-I detector at 50PU scenario. Finally, a study comparing the $\not{E}_{\mathrm{T}}$ performance of a Phase-I detector with a tracking extension to higher η demonstrates a signficant improvement, which is the basis of including such an extension in the design of the Phase-II upgrade.

9.4 B-Tagging Performance

Many physics processes lead to the presence of b quarks in their final states. The ability to discriminate the jets arising from the hadronization of these quarks (b-jets) from the ones coming from the light partons is an important ingredient for the measurement of the properties of standard model particles decaying to b quarks, for example the top quark and the Higgs boson, as well as for the search for new physics predicted by theories beyond the standard model.

9.4.1 b-jet Identification Algorithm

During Run-I, the first data taking period of the LHC (2010-2012), the CMS Collaboration has developed several algorithms to identify (*tag*) b jets [203], based on the distinctive signature of heavy hadron decays such as displaced tracks, secondary vertices, and the presence of soft leptons in the jet cone. Many of the analyses that needed b-jet identification used the Combined

Secondary Vertex (CSV) tagger, that combines the information of reconstructed secondary vertices with the information from displaced tracks. During and after Run-I a more performant version of the CSV tagger was developed. This new algorithm (CSVv2) is used to illustrate the performance of b-jet identification for the Phase-I and Phase-II detector configurations.

The identification of the event vertex of the hard scattering, the "primary vertex" (PV), is a key element for the reconstruction and identification of b jets. The trajectories of the charged particles in the CMS inner tracker are reconstructed using an iterative tracking procedure with a Kalman filter [204]. These charged particle tracks are then clustered based on the z coordinate of their closest approach to the beam line. The vertex position is estimated with an adaptive vertex fit [205] using a collection of tracks compatible with originating from the same interaction. During Run-I operations, with an average of 21 additional pp collisions per bunch crossing, the event PV was identified as the reconstructed vertex with the largest $\sum p_{\mathrm{T}}^2$ of its associated tracks. At the very high luminosity design for the LHC during the Phase-II detector operations, an average of about 140 pileup (PU) interactions are expected, leading to about 80 reconstructed primary collision vertices in each event. When the number of additional interactions in the same bunch crossing increases, the chosen PV is not always identified correctly as the vertex corresponding to the hard interaction. Since the performance of b-jet identification techniques is sensitive to the choice of the primary vertex, a more robust method has been developed. This new technique consists in replacing the individual tracks contributing to the $\sum p_{\mathrm{T}}^2$ by jets, obtained by clustering the tracks originating from the same vertex using the anti-K_t jet clustering algorithm with a distance parameter of 0.4 (AK4) [196]. For each primary vertex, the $\sum p_{\mathrm{T}}^2$ is computed by using these jets as well as the remaining single tracks (including identified leptons) and the missing transverse momentum at the PV in order to account for neutral particles. Different weights can be applied to the jets, leptons and missing transverse momentum. The PV with the largest $\sum p_{\mathrm{T}}^2$ obtained in this way is then chosen as the one corresponding to the hard scattering and the jets associated to it are used to quantify the performance of the b-tagging algorithm.

Jets are obtained using the methods described above. Jets with a transverse momentum $p_T > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$ or $|\eta| < 3.0$ are selected to study the performance of the b-jet identification algorithm after the Phase-II upgrade. A relaxed requirement on the pseudorapidity of the jets enables the study of the possibility for b-jet identification with the Phase-II inner tracker that has an extended coverage in pseudorapidity with respect to Phase-I.

Of all tracks only a subset of well-reconstructed displaced tracks is used as input for the b-jet identification algorithms. The tracks must have a transverse momentum of at least 1 GeV. To ensure a high-quality fitted track, the χ^2 of the fit divided by the number of degrees of freedom is required to be less than 5. At least eight hits in the inner tracker must be associated with the track, of which at least two hits are required in the pixel system, since track measurements in the innermost layers provide most of the discriminating power. One of the most discriminating variables used for b-jet identification is the impact parameter. A loose selection on the track impact parameter is used to further increase the fraction of well-reconstructed tracks and to reduce the contamination by decay products of long-lived particles, e.g. neutral kaons. The transverse and longitudinal impact parameters are defined as the transverse and longitudinal distance to the primary vertex at the point of closest approach in the plane transverse to the beam axis. Their absolute values must be smaller than 0.2 cm and 17 cm, respectively. Tracks are associated to jets in a cone $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2} < 0.3$, where the jet axis (with pseudorapidity η and azimuthal angle Φ) is defined by the primary vertex and the direction of the jet momentum. The distance of a track to the jet axis is defined as the distance between the point of closest approach of the track to the axis. In order to reject tracks from PU vertices this

quantity is required to be less than 0.07 cm. The distance between the primary vertex and the point of closest approach must be less than 5 cm.

Secondary vertices are reconstructed independently from jets with the inclusive vertex finder (IVF). A collection of tracks with a large impact parameter with respect to the primary vertex is used to cluster candidate secondary vertices using track distances and angles. After this coarse clustering, these vertices are reconstructed with the adaptive vertex fitter. Vertices with a large fraction of shared tracks are merged afterwards. The next step consists of an arbitration of tracks between the primary vertex and the secondary vertex to optimize the assignment of tracks to vertices and is followed by another vertex fit and merging step.

The CSVv2 algorithm follows the same approach as CSV [203], except that the secondary vertices obtained with the inclusive vertex finder are used. The flight direction of the secondary vertex has to be within a cone $\Delta R < 0.3$ around the jet direction. When one or more secondary vertices are reconstructed, the secondary vertex information is combined with the track-based lifetime information. When no secondary vertex is found, tracks with a signed impact parameter significance [203] exceeding 2 are clustered in a so-called "pseudo vertex" allowing for the computation of a subset of secondary-vertex based quantities without performing an actual vertex fit. When even this is not possible, a "no vertex" category is defined and only track-based lifetime information is used. The variables are combined into a single discriminator using a neural network to distinguish between jets from b quarks and those from charm or light quarks and gluons.

Monte Carlo (MC) simulated samples of $\mathrm{t\bar{t}}$ events have been generated with MADGRAPH [206] which was interfaced to PYTHIA 6 [89] for parton showering and with the Z2* tune [207]. In order to study the tagging efficiencies as a function of the jet p_{T}, multi-jet events were generated with PYTHIA 8 [208] using the 4C tune [207] and a $\hat{p}_T$ range of 15 to 3000 GeV for the hard-scattered partons. The MC samples have been produced with the full detector simulation based on GEANT4 [79], applying PU simulation, event reconstruction and object identification software.

Tagging efficiencies are evaluated for the three standard combinations of conditions. For Phase-II the pixel detector in the simulation is currently the same as in Phase-I, implementing the new disks in the forward regions but not incorporating expected resolution improvements. Similarly the track reconstruction is expected to be improved, in particular in the forward region with pseudorapidity up to 3.0. For each of the detector conditions, the CSVv2 algorithm was retrained.

The efficiency for choosing the right primary vertex can be estimated in the simulation by comparing its reconstructed z position to that of the generated hard interaction. In $\mathrm{t\bar{t}}$ events without PU the efficiency is found to be 100% and decreases to 99% for Phase-I with 50 PU, 96% in Phase-I-aged conditions with 140 PU and 97% in Phase-II conditions with 140 PU. The efficiency of choosing the correct PV depends on the event kinematics. This is illustrated in Fig. 9.11 where the efficiency of choosing the correct PV is shown as a function of the transverse momentum of the highest p_{T} (leading) jet in multi-jet events. Three pseudorapidity ranges are considered for the leading jet: $|\eta|$ within a range of 0–1.8, 1.8–2.4 or 2.4–3.0. The efficiency of choosing the correct PV decreases for events with only low p_{T} jets, in particular in the forward regions. For $|\eta| < 1.8$, it is larger than 95% for events with a leading jet p_{T} exceeding 50 (90, 70) GeV in the Phase-I (Phase-I-aged, Phase-II) conditions. Note that the Run 1 algorithm for PV sorting (shown in Fig. 9.11 for the Phase-I-aged conditions) would be significantly less efficient. The pixel-detector extension in Phase-II provides an improved efficiency for the PV choice in the forward region $2.4 < |\eta| < 3.0$ compared to the Phase-I-aged detector configuration.

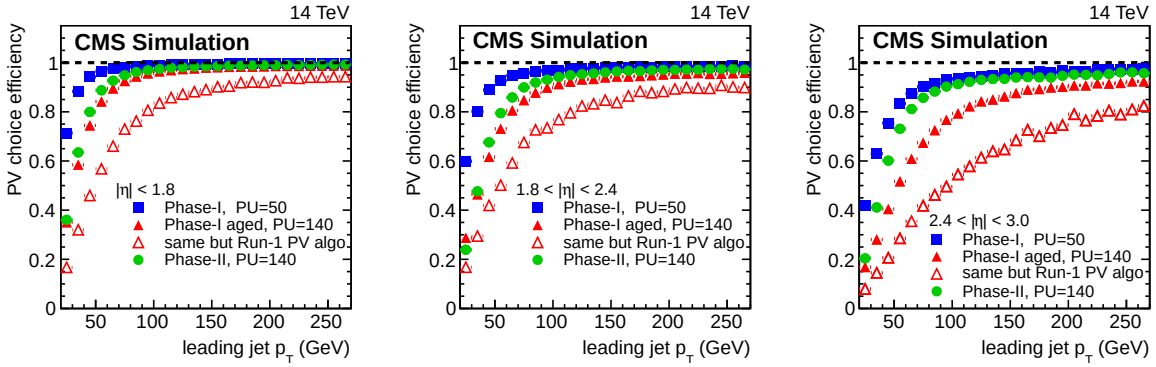


Figure 9.11: Efficiency of choosing the correct PV of the hard interaction as a function of the leading jet p_T in the $|\eta|$ range 0–1.8 (left), 1.8–2.4 (middle) or 2.4–3.0 (right). Simulated multi-jet events are used for three detector scenarios; Phase-I detector with <PU>= 50 (blue squares), Phase-I-aged detector with <PU>= 140 (red triangles) where, as an illustration, the new PV sorting algorithm (full symbol) has a significantly higher efficiency than the Run 1 algorithm (open symbol); Phase-II detector with <PU>= 140 (green circles). The pixel-detector extension in Phase-II provides an improved efficiency for the PV choice in the forward region $2.4 < |\eta| < 3.0$ (right) compared to the Phase-I-aged condition.

9.4.2 Performance of b-jet Identification

The performance of the b-tagging algorithm is characterized by examining the probability to misidentify (mis-tag) a light-parton jet for a given efficiency to identify b jets. The goal is to achieve high efficiencies for b-tagging while keeping the probability for mis-tagging as small as possible.

In the simulation, the jet flavour is defined as a prompt b if only one weakly decaying b hadron is found within a cone $\Delta R < 0.3$ around the jet axis. It is defined as a b from gluon splitting if another weakly decaying b hadron is found within $\Delta R < 0.4$. If no b hadron is found, a prompt c (or c from gluon splitting) is defined by considering the presence of weakly decaying c hadron(s). Otherwise the reconstructed jet is defined as u, d, s quark or gluon (udsg) flavour if such a generated parton is found within $\Delta R < 0.3$. In addition, for all of these flavors, a generated jet with p_T exceeding 8 GeV is required to be found within $\Delta R < 0.25$ of the reconstructed jet, where the generated jet is obtained by clustering the generated particles with the AK4 algorithm [196]). In this way, reconstructed jets originating from pileup collisions are removed when determining the b-tagging performance. In order to ease the comparison of performance between $t\bar{t}$ events and multi-jet events, b and c jets from gluon splitting are not considered. In $t\bar{t}$ events, jets matched to a charged lepton (e, μ or τ) from W decay are also rejected.

The b-tagging performance of the CSVv2 tagger is shown in Figs. 9.12 for jets within $|\eta| < 2.4$ for different levels of expected pileup in Phase-I and Phase-II conditions in simulated $t\bar{t}$ events. The ranges $|\eta| < 1.8$ and $1.8 < |\eta| < 2.4$, respectively, are shown in 9.13. For the Phase-I-aged scenario, the tracks are required to have at least 6 associated hits (instead of 8), which compensates partially the reduced hit reconstruction efficiency in the silicon strip tracker after aging. For the Phase-I detector with 1000 fb^{-1} aging and <PU>= 140 condition, the b-tagging performance is significantly degraded compared to Phase-I at <PU>= 50 or Phase-II at <PU>= 140. As an illustration for jets with $p_T > 30$ GeV and $|\eta| < 2.4$, the b-jet identification efficiency is 0.50, 0.37, 0.46 (0.74, 0.60, 0.72) for the Phase-I, Phase-I-aged, Phase-II conditions, respectively,

if the mis-identification probability for udsg jets is 0.001 (0.01). The performance in Phase-II, without aging, is close to that achievable in Phase-I, especially in the central region, despite the larger number of pileup collisions.

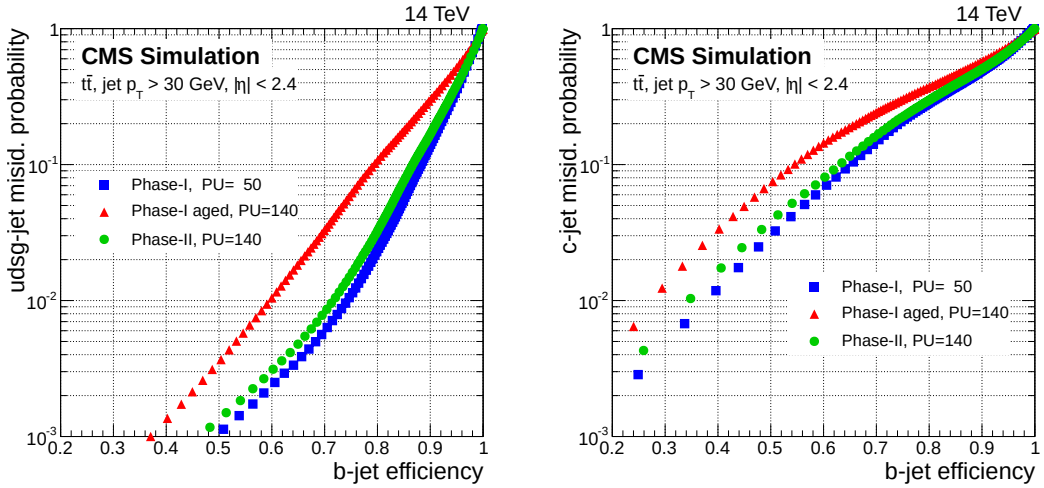


Figure 9.12: Characterization of the b-tagging performance, expressed as mis-identification probability for (left) udsg-jet and (right) c-jet as a function of b-jet tagging efficiency for jets with $p_T > 30$ GeV and $|\eta| < 2.4$. Simulated $t\bar{t}$ events are used in Phase-I conditions with <PU>= 50 (blue) or <PU>= 140 with aging (red), and for Phase-II condition with <PU>= 140 (green). The b-tagging performance for the Phase-I aged detector scenario at high pileup is significantly degraded compared to Phase-I without aging at medium pileup. However the performance in Phase-II largely compensates that performance loss, despite the higher number of pileup collisions.

One of the main upgrades of the CMS detector for Phase-II is the extension of the inner pixel detector in the forward region. This will offer the unprecedented possibility to study b-tagging of jets with pseudorapidity $|\eta|$ larger than 2.4. The b-tagging performance in Phase-II is thus compared between several pseudorapidity ranges in Fig. 9.14. For a light-parton mis-identification probability of 0.01, the b-jet tagging efficiency is of about 0.74, 0.58, 0.41 in the $|\eta|$ ranges 0–1.8, 1.8–2.4, 2.4–3.0, respectively. Although the performance degrades with increased pseudorapidity, the extended pixel coverage with the Phase-II detector upgrade will allow b-tagging capabilities up to $|\eta| \sim 3.0$. Obviously, improvements are expected when optimizing the track reconstruction and b-tagging efficiency in this region.

Thresholds can be set on the b-tagging discriminator value in order to define operating points corresponding to a given mis-identification probability for light-parton jets. The operating points are derived for the various detector conditions and pseudorapidity ranges. Efficiency curves are derived for b jets and taken as input to the "Delphes" fast simulation used for various studies of Higgs boson properties and sensitivities to new physics with the Phase-II detector.

Efficiencies for b-jet and c-jet identification for the Phase-I, Phase-I-aged and Phase-II scenarios are presented in Fig. 9.15 as a function of the jet p_T, for a constant mis-identification probability of 0.001 for udsg-jets over the full p_T and η range. Three $|\eta|$ ranges are considered for the b and c jets; 0–1.8 and 1.8–2.4 where the three detector scenarios can be compared and 2.4–3.0 where b-jet identification is only possible in the case of the Phase-II scenario due to the extended pixel detector. With the Phase-I detector the aging will reduce the b-jet identification efficiency by

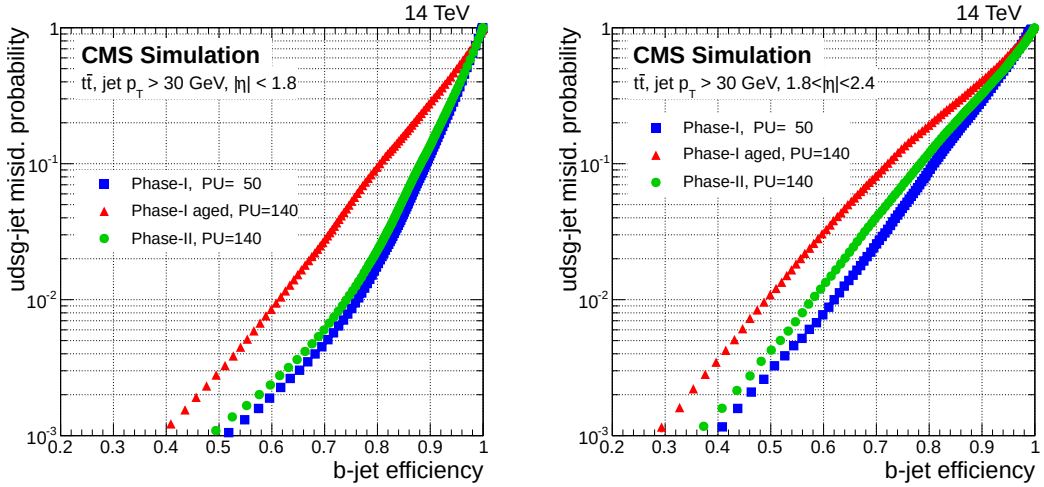


Figure 9.13: Same as Fig. 9.12 in the pseudorapidity region $|\eta| < 1.8$ (left) and $1.8 < |\eta| < 2.4$ (right).

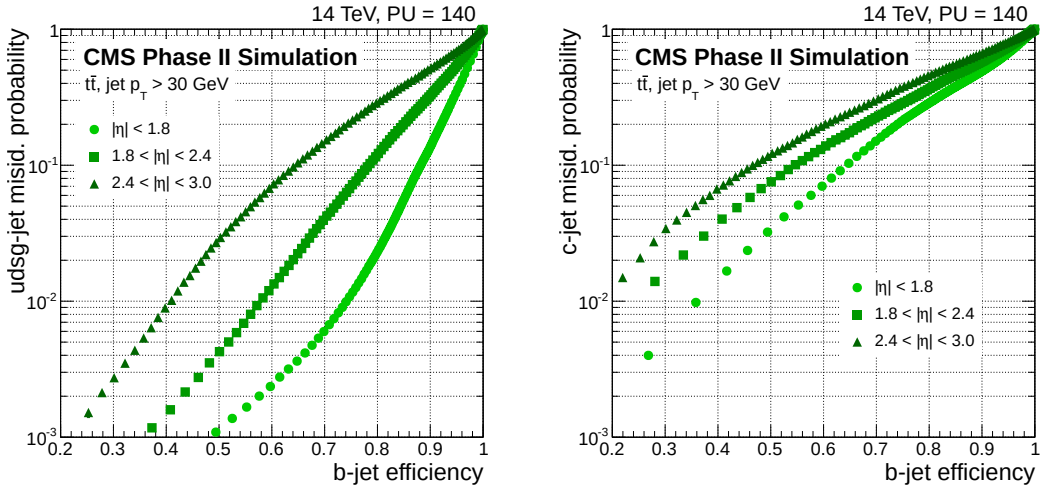


Figure 9.14: Performance in Phase-II detector scenario with $t\bar{t}$ events at <PU>= 140 for different pseudorapidity ranges. The performance is expressed as mis-identification probability for (left) udsg-jets and (right) c-jets as a function of b-jet tagging efficiency. Although the performance degrades with increased pseudorapidity, the extended pixel coverage with the Phase-II detector upgrade allows b-tagging capabilities up to $|\eta| \sim 3.0$.

an absolute amount of more than 10% for jet p_T smaller than 150 GeV, despite a tuning of the track selection for the number of hits.

For $|\eta| < 2.4$, efficiencies in Phase-II with <PU>= 140 are close to those in Phase-I with <PU>= 50 without aging, and significantly better than Phase-I with <PU>= 140 and aging. This is observed in the full p_T and η range, except at low p_T where displaced tracks from pileup interactions can increase the mis-identification of light udsg-jets, and thus decrease the b-jet identification efficiency for the fixed mis-identification probability. The effect is more pronounced at large $|\eta|$, as expected. The efficiency to choose the right primary vertex is smaller for events with low p_T leading jets, contributing also somewhat to the relative decrease in b-jet tagging efficiency at high pileup.

As already noticed for the b-tagging performance in simulated $t\bar{t}$ events, the Phase-II detector allows a basic discrimination between b-, c- and udsg-jets for multi-jet events in the region $2.4 < |\eta| < 3.0$, which is provided by the pixel-detector extension. It is expected that the b-tagging performance in this region will improve with dedicated optimizations.

A similar conclusion is obtained when examining Fig. 9.16 where the mis-identification probability for udsg-jets is presented for a fixed b-jet tagging efficiency of 0.5 for the full jet p_T and η range. The aging of the Phase-I detector will increase the mis-identification probability by a factor of two to ten. On the other hand the Phase-II detector is able to reduce the mis-identification probability significantly, resulting in a performance that is close to that of the Phase-I detector without aging and in an environment with a lower number of pileup collision.

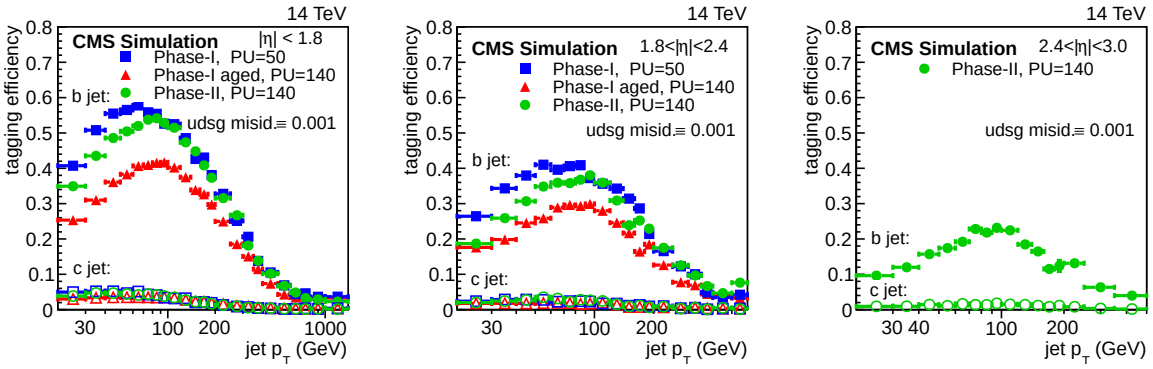


Figure 9.15: Parametrization of the tagging efficiencies for prompt b-jets (filled symbols) and prompt c-jets (open symbols), corresponding to a fixed mis-identification probability of 0.001 for udsg-jets for any p_T and η. The tagging efficiencies are shown as a function of the jet p_T in the $|\eta|$ ranges: (left) 0–1.8, (middle) 1.8–2.4, (right) 2.4–3.0. Simulated multi-jet events are used in three standard combinations of detector, pileup, and aging. The b- and c-jet tagging efficiencies are significantly reduced in Phase-I-aged conditions. The efficiencies in Phase-II conditions at high pileup are found to be close to those from Phase-I at medium pileup for jets with a p_T larger than ~ 100 GeV. The pixel-detector extension in Phase-II also allows b-tagging in the pseudorapidity range 2.4–3.0.

9.4.3 Summary of b-tagging Performance

We have investigated the b-tagging performance in the very challenging conditions expected at the LHC for the proposed Phase-I and Phase-II upgrades of the CMS detector after the 2022 shutdown. The performance of the b-tagging algorithms for the Phase-I and Phase-II detector

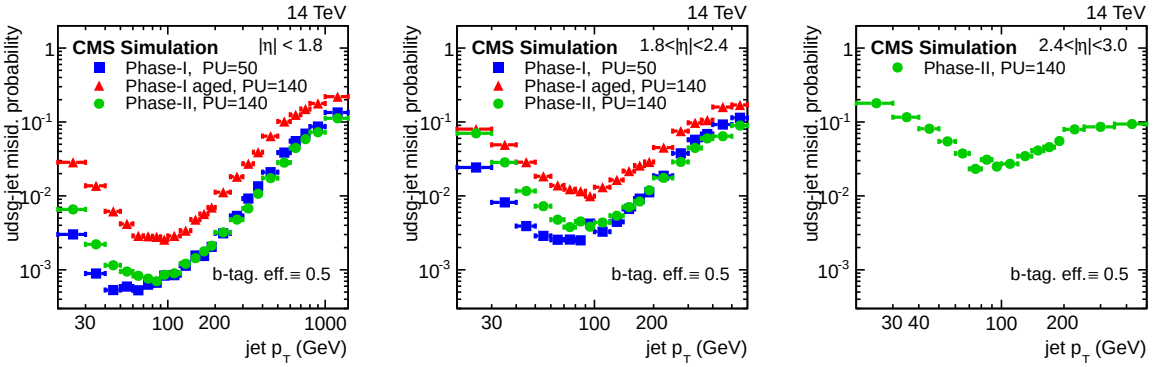


Figure 9.16: Parametrization of the mis-identification probability of udsg-jets, corresponding to a fixed b-jet tagging efficiency of 0.5 for any p_T and η. The mis-identification probability is shown as a function of the jet p_T in the $|\eta|$ ranges: (left) 0–1.8, (middle) 1.8–2.4, (right) 2.4–3.0. Simulated multi-jet events are used in three standard combinations of detector, pileup, and aging. The mis-tagging increases significantly in Phase-I-aged conditions. In Phase-II conditions at high pileup, the mis-tagging is found to be close to that from Phase-I at medium pileup for jet with a p_T larger than $\sim$ 100 GeV. The pixel-detector extension in Phase-II also allows b-tagging in the pseudorapidity range 2.4–3.0.

are presented as curves of b-jet identification efficiency versus the mis-identification probability of udsg- and c-jets for different pileup conditions. The performance is also studied as a function of the jet p_T and the pseudorapidity. For Phase-I, the b-tagging performance degrades as the pileup increases and the detector ages with large accumulated luminosity, with an absolute loss in b-jet tagging efficiency of 13% (14%) at udsg-jet mis-identification probability of 0.001 (0.01) on average in simulated $t\bar{t}$ events. The gain in absolute b-jet tagging efficiency for the upgraded Phase-II detector compared to the aged Phase-I detector simulation is 9% (12%) for a 0.001 (0.01) light udsg-jet mis-identification probability. In addition, compared to the Phase-I detector, during Phase-II the b-tagging capability is extended to a larger pseudorapidity of $|\eta| \sim 3.0$, due to the larger planned coverage of the Phase-II pixel tracker. Since the actual Phase-II pixel detector will be an improved version of the one currently in the simulation, and the reconstruction software, which is currently an extension of the Phase-I software, will also be optimized for the final detector, the performance shown here should be viewed as a conservative estimate.

9.5 Electron and Photon Performance

9.5.1 Introduction

A significant degradation of the electron and photon performance is expected at the HL-LHC from the aging of the Phase-I detector and from the high pileup. This motivates the Phase-II upgrade of the electromagnetic calorimeter. The barrel ECAL will undergo a significant electronics upgrade to replace aging components, while at the same time making the revisions needed to solve problems encountered in normal operation, such as APD spikes, and to accomodate the higher Level 1 trigger latency and data rates. In the endcap region, the radiation damage will degrade the performance of Phase-I calorimeters severely that they must be completely replaced. After extensive R&D, three options were proposed for the replacement. After 18 months of study, CMS chose a tungsen-silicon electromagnetic calorimeter with very high

lateral and longitudinal segmentation; followed by a brass-silicon hadronic calorimeter of approximately 3.5λ, which also provides high lateral and longitudinal segmentation, and extends past shower maximum; and finally a "backing hadron" calorimeter section with coarse lateral and longitudinal segmentation, which is based on Phase-I HCAL design with plastic tiles read out by wavelength-shifting fibers with the light detected by SiPMs to capture the tail of the hardon shower. This ensemble of calorimeters, referred to collectively as the "high Granularity Calorimeter" or "HGCAL", is described in Chapter 3.

The high longitudinal segmentation of electromagnetic secction of HGCAL will mitigate the pileup effect as well as improve further the identification of electron and photon showers. The new electromagnetic calorimeter will provide extended angular coverage for the precision measurement of electromagnetic objects out to $|\eta| = 3$

The overall electron and photon performance of CMS is also improved because of the new tracker. The new tracker design will lead to a strong reduction of the material budget in front of the calorimeter, with an integrated thickness of the tracker material reduced by more than a factor 2 up to $|\eta| = 1.5$. This will allow a further improvement of the performance of the reconstruction and selection of electrons and photons, from the reduction of bremsstrahlung emission and photon conversions.

In the followoing sections, the performance of the electron and photon reconstruction and selection with the Phase-II upgrade version of the CMS detector are presented, and compared to those corresponding to the Phase-I upgrade version of the CMS detector, with aging and pileup conditions corresponding to either those expected at the time of the Phase-I upgrade or those expected for the Phase-II LHC running.

9.5.2 Electron and Photon Reconstruction

Clusters are formed grouping crystals in the ECAL barrel or cells in the HGCAL endcaps starting from a seed and aggregating crystals or cells in transverse directions following the decreasing energy profile expected for electromagnetic showers. In the HGCAL endcaps, clusters built in each layer are then linked across layers to form a single electromagnetic cluster exploiting the known transverse properties of electromagnetic showers, choosing the best link according to minimal angle and distance betwen clusters [209]. In the HGCAL endcaps, clusters are further pre-identified as electromagnetic clusters based on a set of few robust variables involving the longitudinal and transverse profile as measured in HGCAL (see 9.5.4). The corresponding cells are assiged to these clusters and removed from the list of clusters available for subsequent reconstruction steps. The remaining unclusterized cells are used to build hadron clusters. The shower average position and axes are reconstructed using a Principal Component Analysis (PCA), using a log-weighting of the cell energies to unfold the bias arizing from the non-uniform transverse profile.

Clusters of clusters are then built by grouping electromagnetic clusters in order to recover the bremsstrahlung energy spread along the ϕ direction due to the magnetic field. These "superclusters" are used to seed tracks by looking for matching track seeds in the innermost part of the tracker. Tracks are reconstructed from the selected track seeds using a Gaussian Sum Filter (GSF) that optimaly deals with the highly non Gaussian distribution of energy loss by electrons [210].

Electron candidates are formed from the association of a supercluster in the calorimeter and a GSF track reconstructed in the tracker. Loose cuts are imposed on the position matching. In the HGCAL endcaps the matching makes use of the seed cluster position (leading cluster in the

supercluster) rather than the supercluster to avoid the bias arising from the very high PU. A higher transverse energy (E_T) threshold of $E_T > 9$ GeV is also required for the supercluster. A loose cut is applied on the hadronic energy measured behind the electromagnetic supercluster in the hadron calorimeter (HCAL) or in the hadronic section of the HGCAL for the endcaps. In the HGCAL, the cone size used is small in order to mitigate the PU contribution that otherwise constitutes the main source of efficiency loss. The reconstruction working point is such that it maximizes the efficiency, ie the fake rate is the maximum affordable within the CPU and data volume constraints. More details can be found in [211].

Photon candidates are built from superclusters imposing a fiducial cut in η that excludes the transition region between the ECAL barrel and the beginning of the endcaps ($|\eta| < 1.4442$ and $1.566 < |\eta| < 2.5$), where the mass from services (cables and cooling pipes) degrades the resolution badly.

The electron reconstruction efficiencies for the three benchmark configurations are presented in Fig. 9.17. The efficiency is defined as the number of reconstructed electrons matched within $\Delta R(\eta, \phi) < 0.15$ to generator-level electrons coming from prompt $Z \rightarrow$ ee, divided by the total number of generator-level electrons. The reconstruction efficiency is presented as a function of the number of PU interactions per bunch crossing, separately for the ECAL barrel and for the HGCAL endcaps. The comparison is shown for $|\eta|$ <2.5 as only the Phase-II detector allows for electron and photon reconstruction beyond that value. Electron candidates are required to have $E_T > 10$ GeV.

The performance is significantly degraded for the aged Phase-I detector in HL-LHC conditions, in particular in the endcaps with an efficiency loss of up to $\sim$ 20%. With the Phase-II upgrade detector, the electron reconstruction efficiency performance is fully recovered, with the corresponding fake rate kept under control.

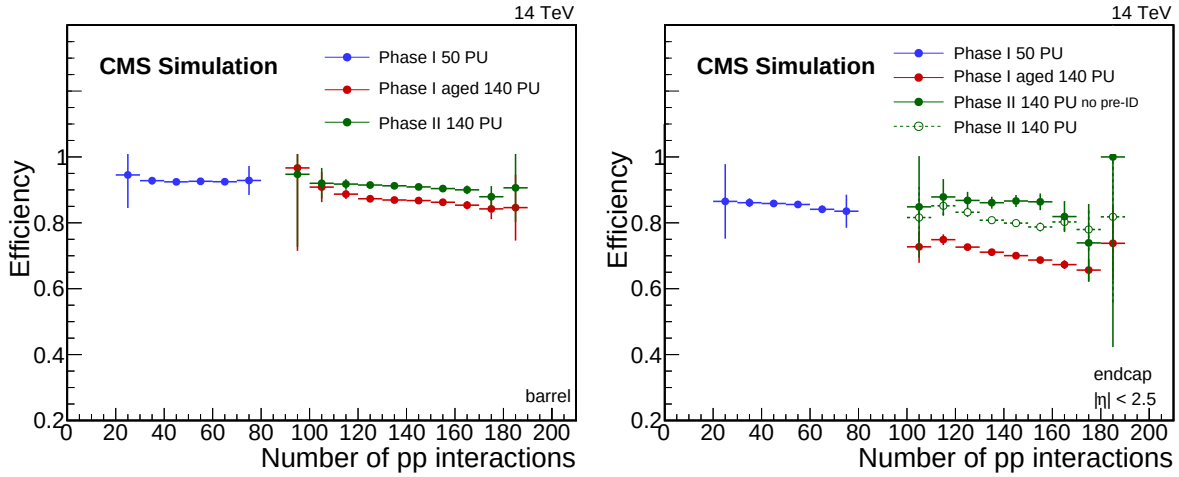


Figure 9.17: Electron reconstruction efficiency in the ECAL barrel (left) and in the HGCAL endcaps (right) as a function of the number of pileup interactions per crossing. The efficiency is shown for three benchmark configurations and for simulated Drell-Yan events.

9.5.3 Energy Measurement

The raw sum of energy deposits in reconstructed superclusters needs to be corrected for the variation of shower containment in the clustered crystals or cells and for the energy loss of electrons and converted photons that arises before reaching the calorimeter. The Phase-II up-

grade benefits from the significantly reduced amount of tracker material before the calorimeter, compared to Phase-I. Other effects such that the proximity of showers to inter-crystal gaps in the barrel or ϕ-sector boundaries in HGCAL needs to be taken into account.

The best performance is obtained using a multivariate regression technique, taking as input in the case of photons the η coordinates of the supercluster as well as a number of shower shape observables. For barrel photons the ϕ coordinate of the supercluster is also used. Finally, observables related to the amount of pileup in the event are included to correct for the pileup effect. More details on the energy estimation and corrections can be found in Ref. [212].

The energy resolution for photons in the ECAL barrel as a function of p_T is presented in Fig. 9.18 for the three benchmark configurations. The photon candidates are from a simulated sample of $H \to \gamma\gamma$ events. A strong degradation of the resolution, of up to a factor ~ 2 at $p_T \sim 20$ GeV, is visible for the Phase-I aged scenario. The resolution loss is almost entirely recovered with the Phase-II detector, with a resolution performance of $\sim 1.3\%$ for a photon p_T of ~ 60 GeV typical of $H \to \gamma\gamma$ decays.

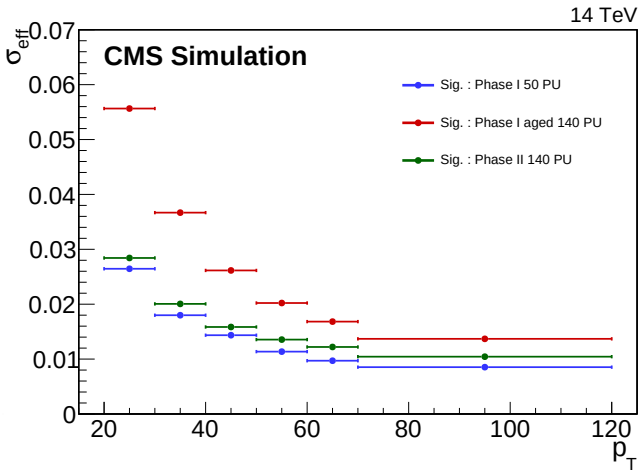


Figure 9.18: Energy resolution for photons in the ECAL barrel as a function of the photon p_T in simulated $H \to \gamma\gamma$ events.

9.5.4 Electron Identification

Electron identification is based on observables that characterise the quality of the track-cluster matching in position and momentum, as well as from pure ECAL observables (shower shape) and pure tracker observables (bremsstrahlung fraction seen by the tracker, number of hits and χ^2 of the track). The best performance is obtained from a multivariate Boosted Decision Tree (BDT) that combines all information taking into account the correlations between the variables for the signal and for the background. Different working points are used depending on the physics analysis, from the highest possible efficiency that is suitable for H$\to$ ZZ $\to$ 4e to the highest purity that is suitable for H$\to$ WW $\to$ 2e2ν. More details on the electron identification algorithms can be found in [211].

In the endcaps, the HGCAL for the Phase-II upgrade provides fine longitudinal shower shape information that is additionaly used. It also helps to mitigate the PU, and, together with the transverse segmentation, to maintain high discriminating power of the shower shape variables despite the non-projective nature of the detector. Example input variables are presented in Fig. 9.19. Signal distributions are from simulated Drell-Yan events and background distribu-

tions are from a sample of simulated dijet events with $p_T > 15$ GeV. Distributions are shown for reconstructed electron candidates with $E_T > 10$ GeV.

The layer structure is visible in the distribution of the shower start position along the z-axis (Fig. 9.19, top-left) and background electrons tend to have electromagnetic showers starting later in HGCAL, as expected. The shower length compatibility (Fig. 9.19, top-right) is defined as the measured minus expected shower length, divided by the expected shower length sigma. Expectations for the measured energy are obtained from a shower parametrization adapted to the HGCAL medium [213]. The measured length is reconstructed as the distance between the shower average from PCA and the shower start position. This purely longitudinal variable also shows a significant discriminating power. The shower transverse width along η, $\sigma_{\eta\eta}$ (Fig. 9.19, bottom-left), is a powerful shower shape variable. In building this variable, the fine longitudinal and transverse granularity is used to evaluate the distance of each contributing cell to the reconstructed shower axis, thereby restoring projectivity and maintaining a very high resolution for the signal distribution. In order to mitigate the PU effect, the fine granularity is further used to restrict contributing cells to those at a distance $< 1.5\mathrm{R_M}$, where $\mathrm{R_M}$ is the Moliere radius, to the reconstructed shower axis. Finally, the H/E distribution is presented (Fig. 9.19 bottom-right), where H is the energy of all reconstructed hadron clusters behind the electron cluster within $\Delta R(\eta, \phi) < 0.05$. The small cone size used allows the mitigation of the PU effect on this variable, as can be seen from the small remaining tail at very high H/E values in the signal distribution.

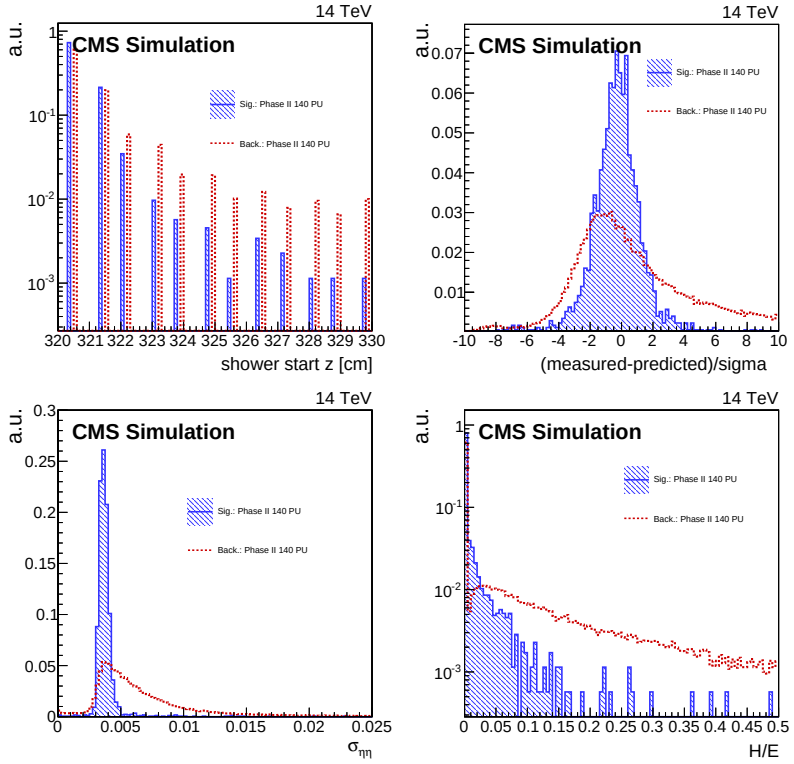


Figure 9.19: Example electron identification variables in HGCAL endcaps.

The performance for electron identification for the three benchmark configurations is assessed using a BDT trained separately for $10 < E_T < 20$ GeV and for $E_T > 20$ GeV with a sample of simulated Drell-Yan events for the signal and a sample of simulated dijet events with $p_T > 15$ GeV for the background. The above shower shape variables as well as more involved vari-

ables based on the comparison of the measured shower longitudinal and transverse profiles are used for the HGCAL Phase-II configuration, in addition to pure tracker and standard track-cluster matching observables. The results are presented in Fig. 9.21. The efficiency is defined as the number of reconstructed electrons passing some BDT cut divided by the number of reconstructed electrons in the sample. The background efficiency is defined as the number of reconstructed electrons matched to a simulated jet from the hard interaction passing the BDT cut, divided by the number of simulated jets. Figure 9.21 (left) presents the background efficiency as a function of the signal efficiency for the three benchmark configurations, for electron candidates with $10 < E_T < 20$ GeV. The signal and background efficiencies are then presented in Fig. 9.21 (right) as a function of the candidate E_T for a cut on the BDT output corresponding to an integrated efficiency of approximately 95% that is suitable for H$\rightarrow$ ZZ $\rightarrow$ 4e.

Finally, the overall selection efficiency and fake rate after the reconstruction and identification steps is presented in Fig. 9.20 as a function of E_T. A 95% efficiency electron identification working point is chosen for the HGCAL sample as is suitable for H$\rightarrow$ ZZ $\rightarrow$ 4e. The working point for the other samples is set so to get an integrated reconstruction times identification efficiency as close as possible to the one obtained for the HGCAL scenario.

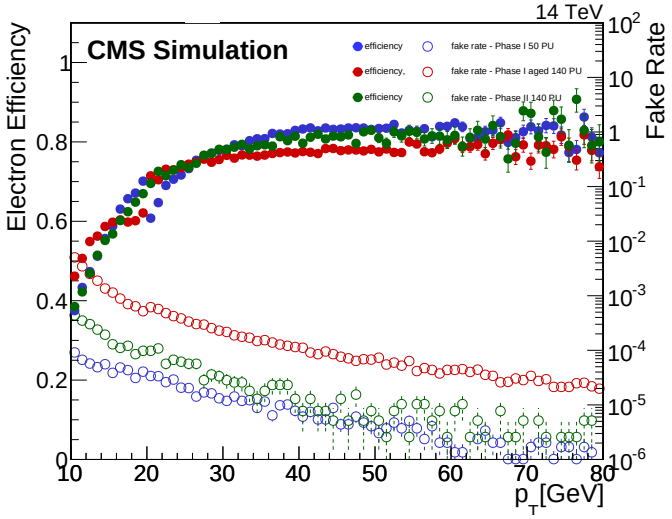


Figure 9.20: Electron overall selection efficiency and fake rate in the endcaps in bins of E_T for the three benchmark scenario.

9.5.5 Photon Identification

To compare the photon selection performance for the three benchmark configurations, samples of simulated events with a photon and a jet are used. The signal is defined as reconstructed photons matched to the simulated photon from the primary vertex within $\Delta R(\eta, \phi) < 0.1$. The background is defined as all other reconstructed photons separated from the signal photon by a distance $\Delta R(\eta, \phi) > 1$. The background therefore includes fake photons from the jet and from pileup. Particles with $p_T > 30$ GeV and $1.6 < |\eta| < 2.5$ (Phase-I samples) or $1.6 < |\eta| < 2.9$ (Phase-II sample) are considered. The identification is achieved using a simple BDT for each sample. The BDTs are trained on samples where the signal and background distributions were re-weighted so as to be flat in η and p_T. All samples are trained in the range of $1.6 < |\eta| < 2.5$ and $30 < p_T < 1000$ GeV. The Phase-I BDTs use as inputs H/E (cone size 0.15), $\sigma_{\eta\eta}$, as defined in Ref. [212], η and p_T. The HGCAL BDT uses the variables developed specificaly for HGCAL and

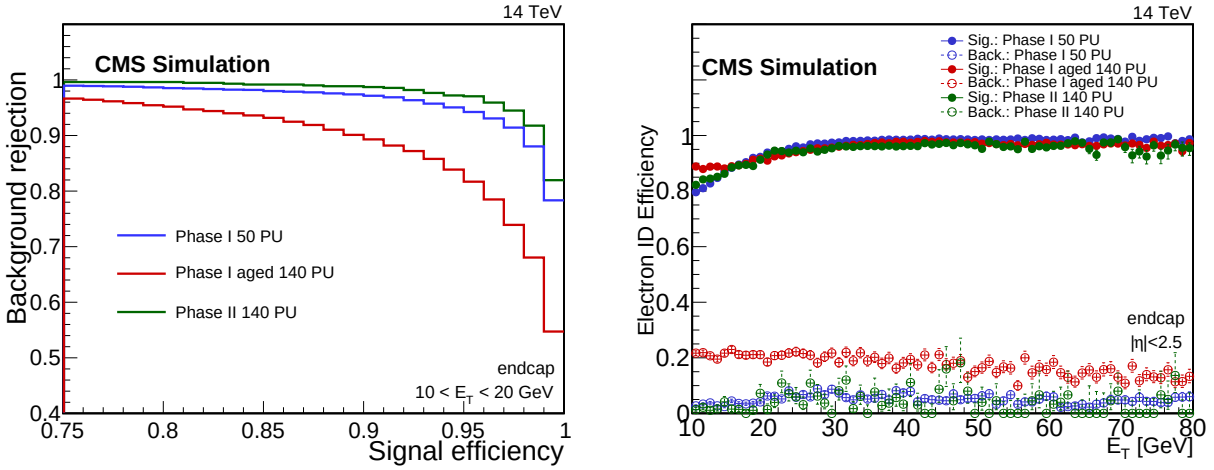


Figure 9.21: Electron identification performance in the endcaps for the three benchmark scenario: (left) background vs signal efficiency; (right) electron selection and background efficiencies as a function of E_T for a cut on the BDT ouptut corresponding to an integrated efficiency of approximately 95% for all samples.

described in 9.5.4: H/E (with a cone size of 0.15), $\sigma_{\eta\eta}$, the shower length compatibility (adapted to the photon case), η and p_T. The efficiency is defined as the number of reconstructed photons passing some BDT cut divided by the total number of simulated signal photons in the sample. The fake rate is defined as the total number of particles incorrectly reconstructed as photons passing the BDT cut, divided by the number of events in the sample. An example working point for each sample is chosen such that the average efficiency across the η range 1.6 to 2.5 is approximately 85%. Figure 9.5.4 then presents the breakdown of efficiency and fake rate per sample in bins of $|\eta|$ of 0.1 (left) or in bins of p_T of 10 GeV (right) for that working point. The plot on the left also shows the performance of the HGCAL for $|\eta|$ up to 2.9.

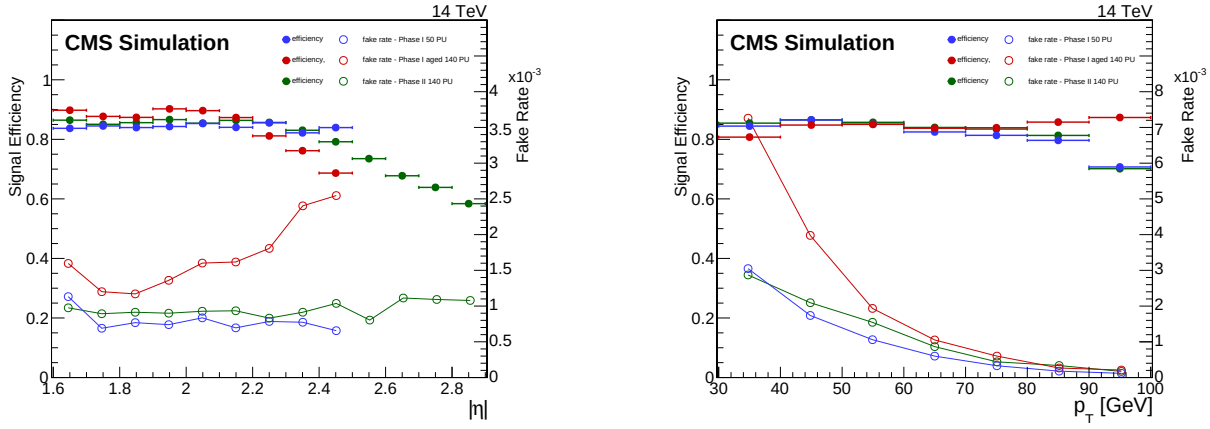


Figure 9.22: Photon selection efficiency and fake rate in the endcaps in bins of $|\eta|$ (left) and p_T (right) for the three benchmark samples. The working point is chosen such that the average efficiency across the η range 1.6 to 2.5 is approximately 85% for all samples.

9.5.6 Summary of Electron and Photon Performance

The performance of electron and photon reconstruction and selection for the Phase-II upgrade of the CMS detector have been presented. The Phase-I endcap electromagnetic calorimeter

would degrade signficantly by $1000\,\text{fb}^{-1}$ and would cease to be useful at all soon thereafter. The upgraded Phase-II detector recovers the performance of the Phase-I detector even though the pileup is considerably higher and it is, of course, designed to survive and maintain high preformance for the entire HL-LHC period.

9.6 Muon Performance

9.6.1 Introduction

For the measurement of muons, the single most important aspect is the 3.8 T solenoidal field, with inner diameter of 6 m and length of 13 m. The magnetic flux generated by the strong central field is large enough to saturate the steel in the return yoke. Muon reconstruction is performed using the all-silicon inner tracker at the center of the detector and up to four stations of gas-ionization muon detectors installed outside the solenoid and sandwiched between the layers of the steel return yoke. This provides two measurements of the trajectories of each muon, resulting in a good level of redundancy and excellent momentum resolution, specified to vary from one to few percent at $100\,\text{GeV}/c$, depending on $|\eta|$, without making stringent demands on spatial resolution and the alignment of muon chambers. The favorable length to radius ratio allows efficient muon measurement up to pseudorapidity of $|\eta| < 2.4$. The strong magnetic field also enables an efficient first-level trigger with an acceptable rate.

The inner tracker is composed of a pixel detector and a silicon strip tracker, it measures charged-particle trajectories in the pseudorapidity range $|\eta| < 2.5$. The muon system covers the pseudorapidity range $|\eta| < 2.4$ and performs three main tasks: triggering on muons, identifying muons, and improving the momentum measurement and charge determination of high-p_T muons.

While all the muon detectors will undergo improvements in LS3 to maintain their performance throughout the HL-LHC era, the main effort will be to preserve the performance of the endcap muon system, increasing the coverage of detector layers for the $|\eta| > 1.6$ region, since that system experiences the highest particle fluxes and suffers the most from problems with pileup. A second goal is to extend muon detection coverage to higher η.

The simulations shown in this section are based on Drell–Yan dimuon production. For isolation studies, QCD multijet samples are also used.

The main challenge for the muon system at the HL-LHC is pileup. Aging of the muon detectors from radiation damage is not the major concern at the startup of Phase-II. On the other hand, some of the electronics of the Drift Tube system will degrade and some of the CSC electronics will suffer from high deadtimes. These electronics systems will be upgraded. However, the simulation of performance of the Phase-I muon detector at $1000\,\text{fb}^{-1}$ presented in this section does not include these effects. Consequently, the performance of the aged Phase-I detector is to be considered somewhat optimistic, when compared to the performance of the Phase-II detector.

9.6.2 Muon Reconstruction and Identification

In the standard CMS reconstruction for pp collisions [214, 215], tracks from muons are reconstructed both in the inner tracker (*tracker track*) and in the muon system (*standalone-muon track*). Based on these objects, two reconstruction approaches are used:

- *Global Muon reconstruction (outside-in).* For each standalone-muon track, a matching tracker track is found by comparing parameters of the two tracks propagated onto a common surface. A *global-muon track* is then fitted combining hits from the tracker track and standalone-muon track, using the Kalman-filter technique [216]. At large transverse momenta, $p_{\mathrm{T}} \gtrsim 200\,\mathrm{GeV}/c$, the global-muon fit (as well as a set of dedicated $high - p_T$ refits combining information from tracker and muon chambers with different logics) can improve the momentum resolution compared to the tracker-only fit [215, 217].
- *Tracker Muon reconstruction (inside-out).* In this approach, all tracker tracks with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}/c$ and total momentum $p > 2.5\,\mathrm{GeV}/c$ are considered as possible muon candidates and are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple Coulomb scattering in the detector material. If at least one muon segment matches the extrapolated track, the corresponding tracker track qualifies as a Tracker Muon.

Tracker Muon reconstruction is more efficient than the Global Muon reconstruction at low momenta, $p \lesssim 5\,\mathrm{GeV}/c$, because it requires only a single muon segment in the muon system, whereas Global Muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station and typically requires segments in at least two muon stations.

Owing to the high efficiency of the tracker-track reconstruction [218] and the very high efficiency for reconstructing segments in the muon system, about 99% of muons produced in pp collisions within the geometrical acceptance of the muon system (and having sufficiently high momentum) are reconstructed either as a Global Muon or a Tracker Muon, and very often as both. Candidates found both by the Global Muon and the Tracker Muon approaches that share the same tracker track are merged into a single candidate.

The combination of different approaches provides robust and efficient muon reconstruction. Physics analyses can set the desired balance between identification efficiency and purity by applying further selections based on various muon identification variables. The performance of the following two basic muon identification algorithms is presented:

- *Loose Muon selection.* The candidate is required to be reconstructed either by the Tracker Muon or the Global Muon algorithm. In addition, the candidate must satisfy the selection criteria of the PF Muon algorithm [219].
- *Tight Muon selection.* For this selection, the candidate must be reconstructed both as a Global Muon and a Tracker Muon. It also needs to satisfy the PF muon algorithm selection. The Global-Muon track fit must have a $\chi^2/d.o.f.$ less than 10 and must include measurements in more than five inner-tracker layers, including at least one hit in the first 3 (2) layers in the barrel (endcap) region, and at least one muon chamber hit. In addition, the Tracker Muon algorithm must find matching segments in at least two muon stations. Finally, the candidate must have a transverse impact parameter $|d_{\mathrm{xy}}| < 2$ mm and a longitudinal impact parameter $|d_{\mathrm{z}}| < 5$ mm with respect to the primary vertex in the event. The primary vertex identification algorithm described in section 9.2, based on the square-p_T sum of all associated tracks, has been found to be inefficient for $Z/\gamma^* \rightarrow \mu\mu$ events in Phase-II pileup conditions. Hence, the reconstructed vertex closest to the simulated primary vertex is used for the calculation of the impact parameters. This allows us to estimate the muon performance independently of the vertex identification efficiency, under the assumption that specific, more efficient, vertexing algorithms may be developed at analysis level. With

this selection, the rate of muons from decays in flight is significantly reduced, at the price of a few percent loss in efficiency for prompt muons, such as those from W and Z decays. The Tight Muon selection is used in many physics analyses in CMS.

Muon reconstruction and identification efficiency is calculated as the fraction of simulated muons that are associated to a reconstructed muon passing a given identification selection criteria. The association between simulated and reconstructed muons is performed by matching the simulated signals on each detector layer to the corresponding reconstructed hits, used to fit the muon track.

Figures 9.23 and 9.24 show the identification efficiency of Loose and Tight Muons respectively. Results are provided as a function of p_{T} and $|\eta|$ of the simulated muons and with respect to the number of simulated pileup interactions. The upgraded detector allows for the identification of muons at high pileup with the same efficiency as the Phase-I detector with lower pileup. The aged detector shows a drop in efficiency of approximately 15%, dominated by the tracking inefficiency. As discussed in 9.6.1, the efficiency reduction for the high-pileup scenario of Phase-I would be greater if the effect of electronics aging of the muon spectrometer were included.

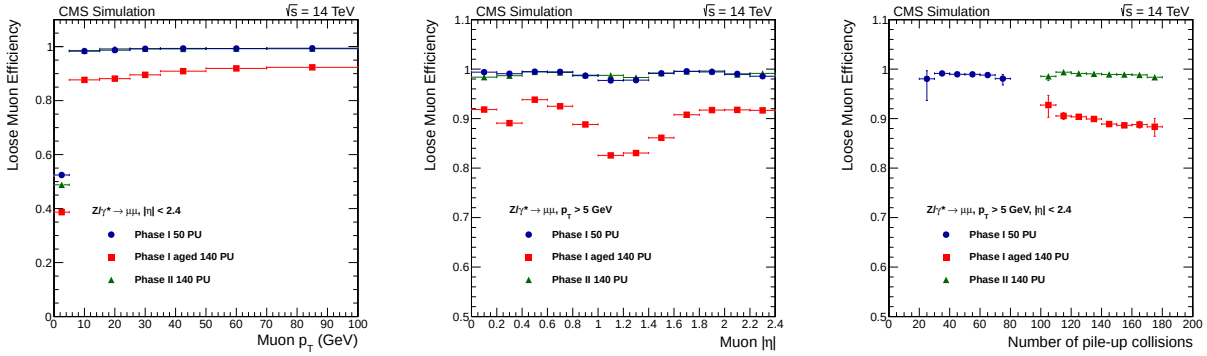


Figure 9.23: Loose Muon reconstruction and identification efficiency in DY events as a function of the simulated muon p_{T}, $|\eta|$, and number of simulated pileup interactions for three detector and pileup scenarios.

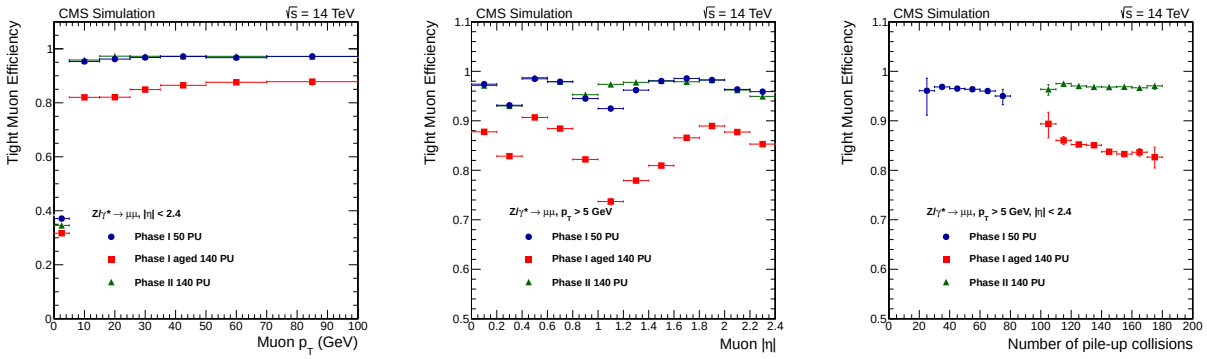


Figure 9.24: Tight Muon reconstruction and identification efficiency in DY events as a function of the simulated muon p_{T}, $|\eta|$, and number of simulated pileup interactions for three detector and pileup scenarios.

9.6.3 Background Muon Multiplicity

Background muons are defined as all the reconstructed and identified muon candidates that are not matched to a simulated muon from the main "hard" interaction—in the following

results, Drell–Yan dimuon production is considered as signal. Background muons originate from various sources: muons from heavy-flavor and light-flavor hadron decays; fake muons from hadron punch-through; random matching of tracker tracks with standalone tracks (or segments) in the muon spectrometer; and duplicates of the same muon.

Figure 9.25 shows the average background-muon multiplicity per event (left) and the average background-muon rate (right) for the Loose Muon identification. The background rate is defined as the ratio between the reconstructed and identified muons that are not matched to a simulated muon and the total number of reconstructed and identified muons. Results are presented as a function of the transverse momentum of the reconstructed muons. The same three detector and pileup scenarios used for efficiency studies are considered. The background multiplicity is maximum at low p_T, where there is the highest pileup contribution. The background rate depends on the multiplicity and on the kinematics of the signal sample. For dimuons from Drell–Yan production, it has a minimum around $p_T \simeq 40\,\text{GeV}/c$, where the number of muons from Z decays—hence the denominator of the background rate—is largest. Above $p_T \simeq 15\,\text{GeV}/c$, the Phase-I aged detector has the largest background multiplicity level, due to the high rate of badly reconstructed tracks. Below that value of p_T, the Phase-II detector background is larger. This is due to the higher Phase-II tracking efficiency, which also applies to the reconstruction of tracks from pileup interactions. Similar conclusions were obtained when comparing the background multiplicity of Tight Muons in different scenarios. In that case, however, the level of background muons is further reduced with respect to the observation for Loose Muons. A reduction factor between one and two orders of magnitude is found. As an example, the background muon multiplicity for Tight Muons of p_T ¡ 5 GeV is below 5×10^{-2} for all considered scenarios.

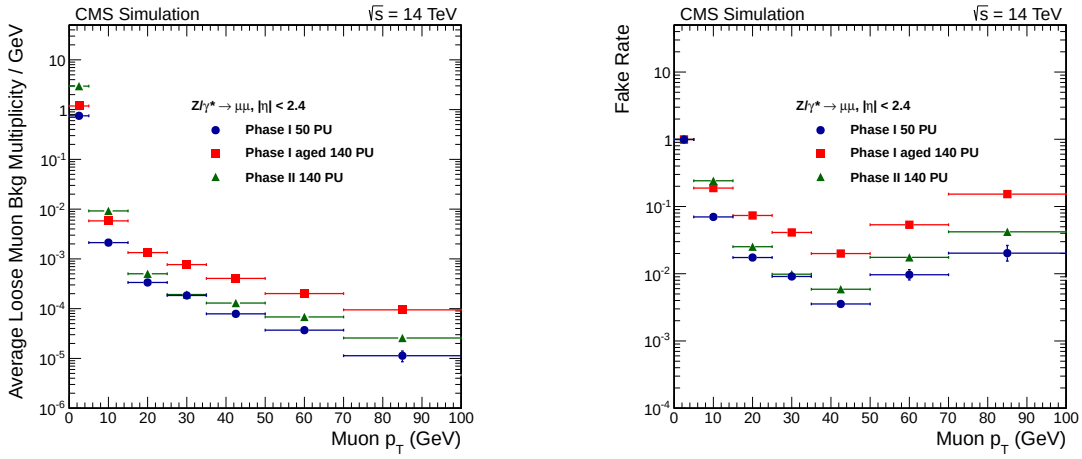


Figure 9.25: Average background-muon multiplicity (left) and rate (right) for the Loose Muon identification as a function of muon p_T.

9.6.4 Muon Resolution

The transverse momentum resolution is defined as $\sigma(p_T)/p_T$, where $\sigma(p_T)$ is the width of a Gaussian fit to the distribution of the reconstructed muon transverse momentum with respect to their true values. The result is not significantly affected by the muon identification criteria. Figure 9.26 shows the muon p_T resolution estimated in a simulated Drell–Yan sample for p_T ranges from 5 to 10 GeV/c (left) and larger than 40 GeV/c (right) as a function of the muon pseudorapidity. The improvement in resolution due to the reduction of material and better spatial measurements of the Phase-II tracker is visible.

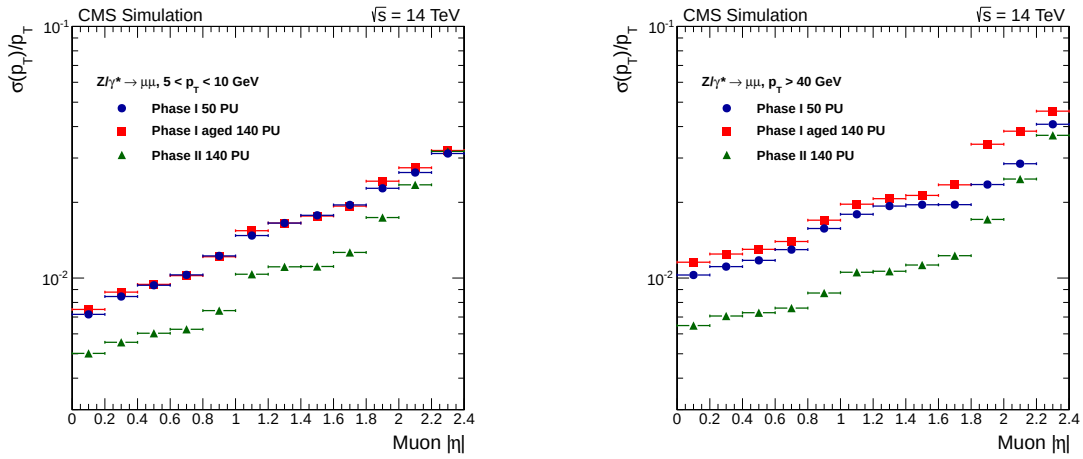


Figure 9.26: Muon p_{T} resolution estimated in a simulated Drell–Yan sample for p_{T} ranges from 5 to 10 GeV/c (left) and larger than 40 GeV/c (right) as a function of the muon $|\eta|$.

9.6.5 Extension of η Coverage: ME0

A new muon detector, ME0, will be located in space that will become available when the endcap calorimetry is replaced in LS3. This device covers the range $2.0 < |\eta| < 2.8$. ME0 chambers consist of 6-layer GEM detectors. Segments are built by fitting groups of three or more hits falling in box of $\Delta\eta \times \Delta\phi = 0.02 \times 0.05$ rad. By requiring hits in a minimum of three layers to define a segment, backgrounds from neutrons can be effectively eliminated. *ME0 Muon* objects are formed by matching tracker tracks to ME0 segments in local x and y position, as well as in global direction in the bending plane, ϕ. If a charged track matches a muon segment within $\Delta\phi = 0.15$ rad, we have a *Tight* ME0 Muon. If a charged track matches to within $\Delta\phi = 0.5$ rad, then we have a *Loose* ME0 Muon. Efficiencies and background multiplicities are computed with definitions similar to those used in Sec. 9.6.2 and 9.6.3. The reconstruction efficiency of ME0 Muons is shown as a function of $|\eta|$ in Fig. 9.27. The efficiency and background yield as a function of p_{T} are shown in Fig. 9.28. The new ME0 detector will provide efficient muon identification with reasonably low backgrounds that will add an additional measurement for muons up to $|\eta| < 2.4$ and will extend the CMS muon coverage up to $|\eta| = 2.8$.

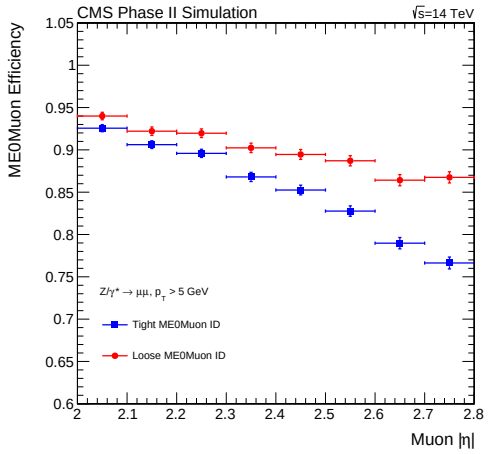


Figure 9.27: Efficiency for ME0 Muons as a function of $|\eta|$.

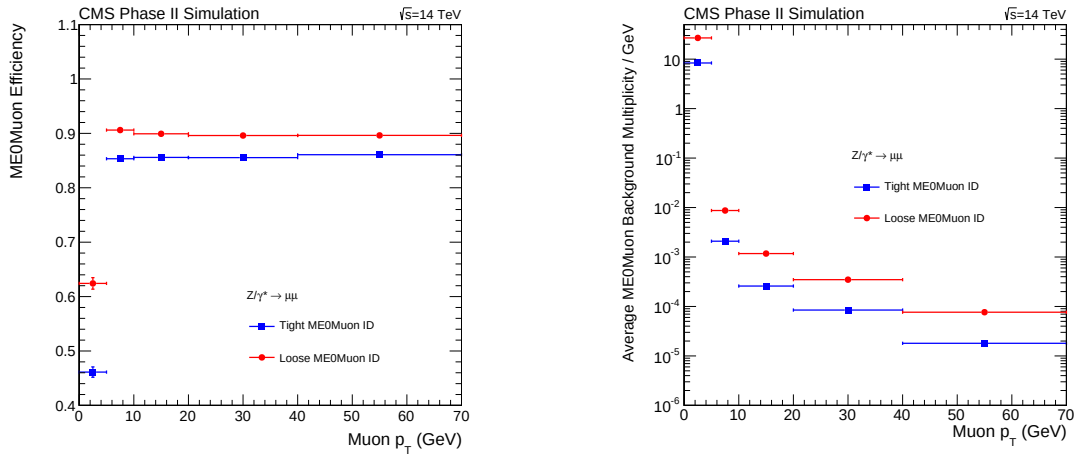


Figure 9.28: Efficiency and background for ME0 Muons as a function of p_T.

9.6.6 Summary of Muon Performance

The performance of muons is compared for different detector configurations and pileup scenarios considering two different identification criteria. For both, the muon reconstruction and identification efficiency is above 95% with the Phase-I detector with a 50 pileup events scenario. It falls by approximately 10% when we consider 140 pileup events and the effect of tracker aging. In the high pileup scenario, the Phase-II detector fully recovers the efficiency of Phase-I without the high pileup events and aging effect. The multiplicity of background muons with p_T ¿ 15 GeV that pass the looser identification criteria considered for the study is below 10^{-3} for the Phase-I detector in the low pileup scenario. The performance is similar for the Phase-II detector at 140 pileup, whereas it is roughly 5-10 times larger for the aged Phase-I detector in the same pileup conditions.

Finally, a dedicated reconstruction was developed to identify muons up to $|\eta| = 2.8$, exploiting a new set of chambers (ME0) included in the Phase-II detector configuration. Despite the reduced coverage in terms of muon detector layers for $|\eta|$ ¿ 2.4, good performance is achieved. By tuning the matching criteria between the tracker tracks and the segments reconstructed in the ME0 chambers, it is possible to reach an average identification efficiency of 85-90%, while keeping the background contamination reasonably low.

9.7 Tau Performance

9.7.1 τ Reconstruction Overview

The *tau lepton* (τ^-) is the heaviest known charged lepton with the mass of 1.777 GeV [220]. It belongs to the third generation of standard model fermions. Like the muon, it decays via the weak interaction. However, because of its large mass, it decays with a much shorter mean lifetime of 2.9×10^{-13} s and it can decay both to hadrons and to lighter leptons. In fact, because of the larger phase space, the decays to hadrons are dominant (about 2/3 of all tau decays). The detailed listing of the tau lepton branching fractions is given in Table 9.1.

One role of taus at the LHC is to act as probes of the Yukawa coupling of the Higgs boson. Because of the sizable $H \rightarrow \tau^+\tau^-$ branching fraction and small irreducible background, it is the most sensitive channel for the measurement of one of the Higgs-fermion couplings [221, 222] and a good tool for the measurement of the Higgs boson polarization [223]. Several extensions of the standard model predict abundant production of tau leptons, so they also serve as a probe

Table 9.1: Dominant intermediate resonance of the τ^-. Data from ref. [220].

final state	branching fraction	intermediate resonance
$e^-\overline{\nu}_e\nu_\tau$	17.83 ± 0.04 %	–
$\mu^-\overline{\nu}_\mu\nu_\tau$	17.41 ± 0.04 %	–
$\pi^-\nu_\tau$	10.83 ± 0.06 %	–
$\pi^-\pi^0\nu_\tau$	25.52 ± 0.09 %	$\rho(770)$
$\pi^-2\pi^0\nu_\tau$	9.30 ± 0.11 %	$a_1(1260)$
$\pi^+2\pi^-\nu_\tau$	8.99 ± 0.06 %	$a_1(1260)$
$\pi^+\pi^02\pi^-\nu_\tau$	2.70 ± 0.08 %	–

to beyond-standard-model physics.

Because the majority of the tau decays are hadronic, CMS employs a dedicated procedure to reconstruct tau leptons from the light hadrons inside jets. The details of this algorithm can be found in [224–226]. The visible products of hadronic tau decays ("tau candidates") are reconstructed from the particle candidates that have been created by the *ParticleFlow* (PF) algorithm [227] and clustered into jets using the anti-k_T [196] algorithm. In order for a jet to be used as an input for the tau reconstruction, it must have $p_T > 14$ GeV. The tau candidates are created by the following sequence of steps:

1. Based on the compatibility (closest distance in z) with the highest-p_T track of a jet, a vertex is assigned to that jet.
2. π^0candidates are constructed by combining PF electron and photon candidates from the input jet in "strips" (wider in ϕ than in η), to take into account possible broadening of calorimeter depositions from photon conversions.
3. Combinations of 1-2 charged particles from the input jet with 0-2 strips and combinations of 3 charged particles without any strips are created. All signal candidates must be compatible with the selected vertex.
4. Only the candidates that are kinematically compatible with the hadronic tau decays given in lines 3-6 of Table 9.1 in terms of reconstructed mass and collimation of the candidates are selected.
5. If there is more than one tau candidate for a given input jet, the one with the highest p_T is selected. In case several candidates have similar p_T, the one which is the most isolated (calculated from all particles compatible with the vertex in a cone of 0.5 around the tau direction) is selected.

The vast majority of tau candidates constructed in this way are fake: e.g. every input jet with at least one reconstructed charged pion compatible with the jet vertex is regarded as a tau candidate. Therefore applying the *tau identification* is a necessary step for every analysis using hadronically decaying taus. This means that the tau candidate needs to be isolated in order to pass the discrimination against QCD jets. Electrons can be misreconstructed as jets, and so they can fake hadronic tau decays as well. Therefore additional discrimination against electrons is required. During Run-I, CMS also applied additional discrimination against muons. However, because the efficiency for identifying a muon by "normal" techniques is more than 99% efficient for hadronic tau decays and the misidentification rate is around 10^{-4}, extra muon discrimination is unnecessary and has been eliminated from the analysis for this Technical Proposal.

9.7.2 Tau Isolation Performance

This section shows the expected performance of hadronic tau reconstruction and identification based on a simulation of the state of the CMS detector after Phase-I and Phase-II upgrades. The three standard scenarios are considered: the "nominal Phase-I conditions" with expected 50 pileup interactions per event; the "aging" scenario where the damage caused to the full CMS detector except the pixel tracker by 1000 fb^{-1} of collected data is modeled together with 140 pileup interactions every event; and the "nominal Phase-II conditions" with 140 pileup interactions per event.

The products of hadronic tau decays are typically more isolated than tau candidates misreconstructed from hadrons clustered in a quark or gluon jet. Therefore the sum of momenta of particles reconstructed in the vicinity of the tau candidate is a very important handle to decrease the tau misidentification rate. The performance study presented in this section focuses on a case where only charged particles are being used for the isolation p_{T} sum calculations. The reason is that this quantity is intrinsically more stable against pileup because the charged particles associated with pileup vertices can be ignored in the p_{T} sum calculation. The contribution of neutral particles originating from the same interaction as the tau candidate is much harder to estimate and will be the subject of future studies. Nevertheless, during Run-I CMS established that the charged isolation is the dominant part of the discrimination against jets and neutral isolation brings only a modest performance improvement.

The tau performance estimates are based on the algorithm described in the previous section that has been developed and commissioned for LHC Run-I conditions [226]. This algorithm has not yet been fully optimized for the post-LS2 running conditions, so results presented in this section can be considered as conservative lower estimates of the tau performance after LS2. However, the working points for Phase-II and aging scenarios have been changed to yield the same efficiency as nominal Phase-I conditions.

The tau candidates considered for the isolation efficiency calculation must pass the following criteria:

- generated visible (i.e. without neutrinos) $p_{\text{T}} > 20\,\text{GeV}$,
- generated visible $|\eta| < 2.3$,
- matched to a reconstructed tau candidate within $\Delta R < 0.5$ that passes:
  - $p_{\text{T}} > 20$ GeV,
  - $|\eta| < 2.3$.

Additionally, to pass a selection "isolation efficiency numerator", the reconstructed tau candidates matched to examined visible decay products of generated hadronic taus must pass the isolation discriminator. This means that the absolute p_{T} sum of all tracks that are within the cone of $\Delta R < 0.5$ around the reconstructed τ direction *and* are likely to come from the same interaction as the τ itself (i.e. they pass selection based on transverse and longitudinal impact parameters) is smaller than a given value. For the nominal Phase-I scenario, the maximal allowed p_{T} sum was 2 GeV (corresponding to the high efficiency working point of Run-I). In the scenarios with 140 interactions per event, there are more tracks reconstructed around the hadronic tau decay products that pass the criteria of coming from the tau vertex. This means that the isolation sums are larger and fewer tau candidates pass the selection. Therefore the maximal allowed p_{T} sum was increased to 3.5 GeV in Phase-II scenario and to 4.0 GeV in Phase-I with 140 interactions to match the nominal Phase-I efficiency. The efficiency of the charged isolation applied on generated taus in $Z \to \tau^+\tau^-$ decays in the nominal Phase-I and two HL-LHC

scenarios is shown on figure 9.29.

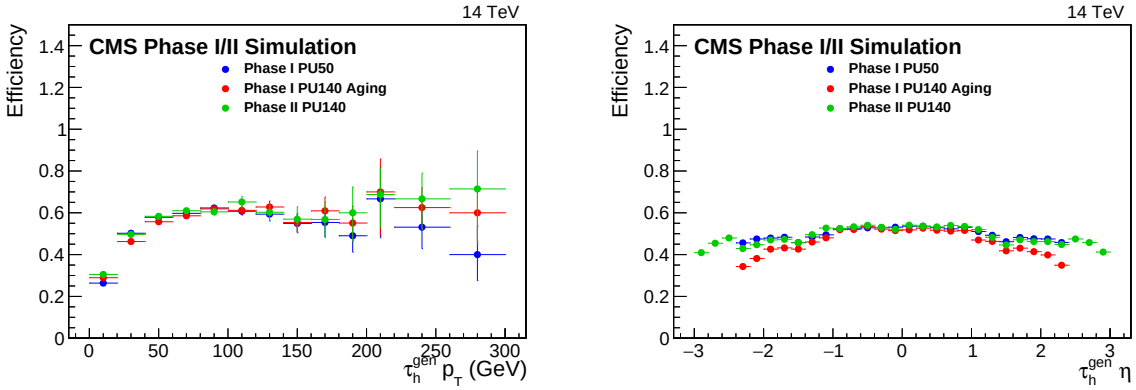


Figure 9.29: Efficiency of the selection of true taus with $p_{T\ vis} > 20$ GeV matched to a reconstructed tau candidate with $p_T > 20$ GeV using the charged isolation as a function of true visible tau p_T (left) and η (right). The three upgrade scenarios described in the text are compared. The isolation selection is tuned to yield similar efficiencies in all scenarios. The cut on $|\eta| < 2.3$ is not applied in the right-hand plot.

The jet misidentification rate quantifies the probability that a true jet will be reconstructed and identified as a hadronic decay of the tau lepton. The following criteria must be met by a jet to be considered for the misidentification rate calculation:

- generated jet with $p_T > 20$ GeV and $|\eta| < 2.3$ is within $\Delta R < 0.5$ of reconstructed jets,
- distance along z-axis between primary vertex and the jet constituents is smaller than 0.2.

In order for a reconstructed jet to be considered as misidentified tau candidate, it must meet the following criteria:

- the jet is matched to a reconstructed tau candidate with $p_T > 20$ GeV and $|\eta| < 2.3$,
- the tau candidate passes the charged isolation selection.

The probability for a generated jet to be reconstructed as a tau candidate and pass the charged isolation selections described in previous paragraph is shown in figure 9.30.

Figures 9.29 and 9.30 show that scenarios with 140 interactions per event have more jets misidentified as hadronic taus for the same selection efficiency. Nevertheless, the performance of the Phase-II detector is much better than the one Phase-I detector in HL-LHC conditions and only slightly worse than nominal Phase-I performance.

The efficiency of the tau ID (figure 9.29) after the Phase-I upgrade ("50 PU" scenario) is actually comparable to the Run-I performance for a similar working point (50-60%). The increase in the number of jets passing the tau identification in scenarios with 140 interactions is caused mostly by looser isolation requirements. However, if we kept the isolation selection same in all the scenarios, the misidentification rate would be similar for Phase-II and nominal Phase-I detector, while it would be increased for the Phase-I detector with aging and 140 interactions. The performance of the tau identification when the same isolation selection is used in all three scenarios is shown in figures 9.31 and 9.32.

The selection efficiency decrease is caused by the larger number of tracks reconstructed in the

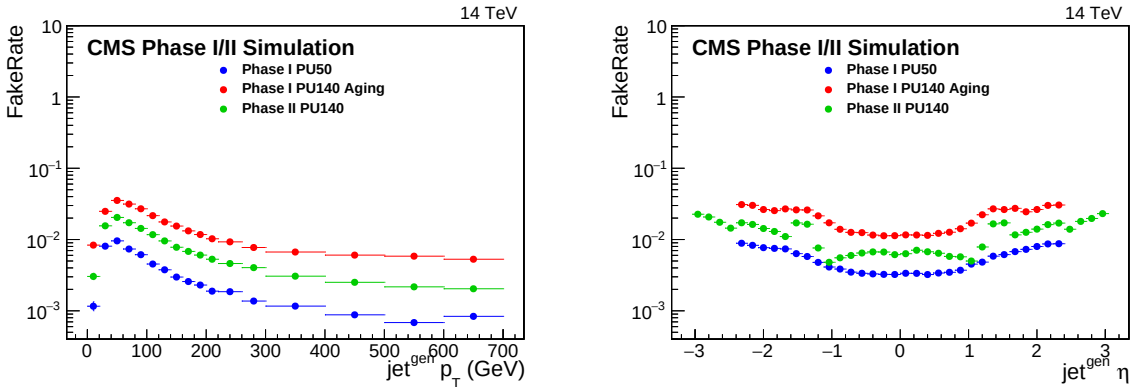


Figure 9.30: Probability of a true jet to be reconstructed as a hadronic tau and pass the charged isolation selection as a function of true jet p_{T} (left) and η (right). The three upgrade scenarios described in the text are compared. The isolation selection is tuned to yield similar efficiencies in all scenarios. The cut on $|\eta| < 2.3$ is not applied in the right-hand plot.

tau candidate vicinity, as discussed previously. The misidentification rate increase in the aging scenario is on the other hand caused by (slightly) decreased tracking efficiency due to radiation damage applied to the tracker, so the sum of the transverse momenta of the charged jet constituents around the fake tau candidate is slightly smaller and thus the number of jets passing is higher in the Phase-I aging scenario. However, the Phase-II tracker upgrade recovers the lost tracking efficiency and sets the fraction of jets misidentifed as hadronic tau decays back to its "initial" Phase-I performance.

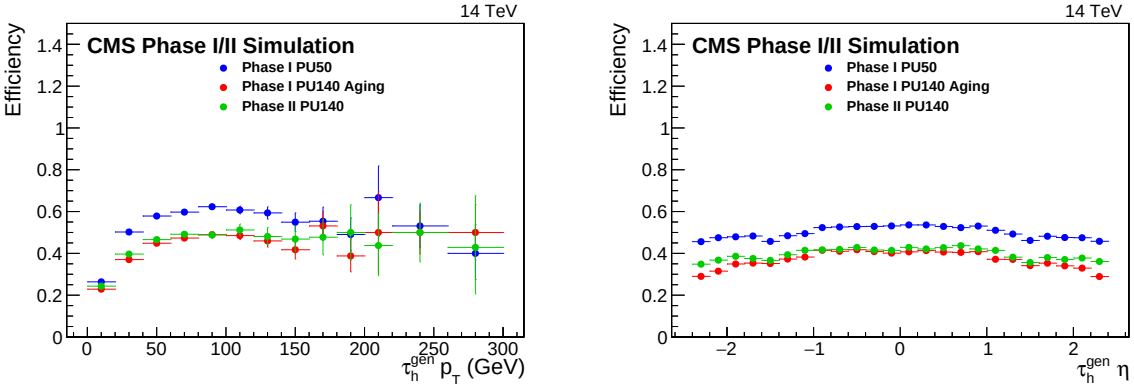


Figure 9.31: Efficiency of the selection of the true taus with $p_{\mathrm{T\ vis}} > 20$ GeV matched to a reconstructed tau candidate with $p_{\mathrm{T}} > 20$ GeV using the charged isolation as a function of true visible tau p_{T} (left) and η (right). The three upgrade scenarios described in the text are compared. The isolation selection was the same for all three scenarios.

9.7.3 Electron Rejection Performance

An important part of the tau identification is the rejection of electrons that are reconstructed as hadronic tau candidates. The reason is that electrons can look similar to 1-prong tau decays. On top of this, bremsstrahlung can resemble the π^0decay. The tau identification algorithm therefore uses several observables that have different distributions for electrons and hadronic tau decay products to reject tau-like electrons. Such variables are e.g. HCAL and ECAL energy fractions of the reconstructed tau or its constituents and the fraction of reconstructed tau energy

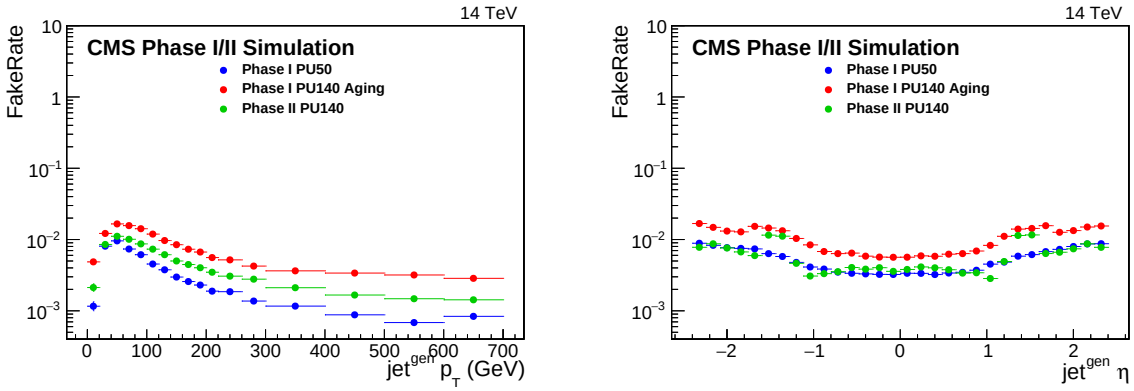


Figure 9.32: Probability of a true jet to be reconstructed as a hadronic tau and pass the charged isolation selection as a function of true jet p_{T} (left) and η (right). The three upgrade scenarios described in the text are compared. The isolation selection was the same for all three scenarios. The cut on $|\eta| < 2.3$ is not applied in the right-hand plot.

carried by photons. However, for the Phase-II calorimeter endcap, a new set of observables is chosen due to its novel design. The calorimeter and tracker based observables are then used as an input to a multivariate discrimination based on boosted decision trees. The discriminator is (re)trained for all three scenarios to cover the difference in running conditions and detector performance. In all cases, it is trained separately for the barrel and endcap due to the different instrumentation there.

Figure 9.33 shows the fraction of electrons, from simulated $Z \to ee$ decays, reconstructed and identified as hadronic taus as well as the fraction of generated hadronic taus passing the electron rejection discriminator. The performance is degraded in the scenarios with 140 interactions. The Phase-II detector significantly decrease the amount of electrons misidentified as hadronic tau decays (compared to the aged Phase-I detector with 140 interactions).

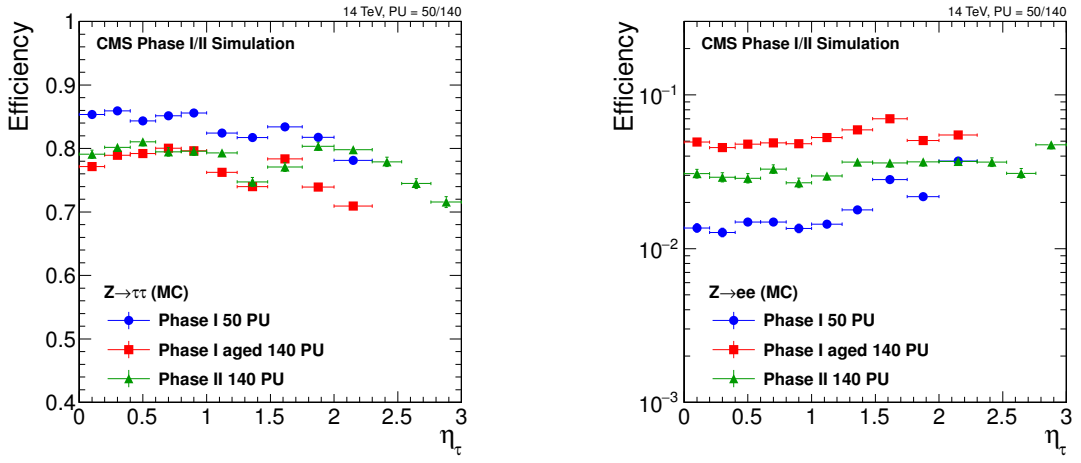


Figure 9.33: Efficiency of the selection of a true taus with $p_{\mathrm{T\ vis}} > 20$ GeV matched to a reconstructed tau candidate with $p_{\mathrm{T}} > 20$ GeV using the tight working point of BDT-based electron rejection as a function of reconstructed tau η (left). The probability for a true electron to pass the same algorithm as a function of reconstructed tau η (right). Three upgrade scenarios described in the text are compared.

9.7.4 Summary of tau Performance

The tau reconstruction and identification have been studied in the context of the detector and conditions expected after LS2. Although the pileup is considerably larger than in Run-I, the nominal tau ID performance is actually similar to the Run-I performance. However, the high-luminosity running leads to decreased tau ID performance due to a decreased selection efficiency in the high-density environment. The decreased tracking performance leads to an increase of the number of jets misidentified as hadronic taus. This effect is however mostly fixed by the Phase-II tracker upgrade. The extended tracker also allows the reconstruction of hadronic tau decays with pseudorapidity between 2.5 and 3. For the same efficiency, the amount of misidentified jets increases by a factor of 2, mostly due to the looser isolation selection. For the same misidentification rate, the efficiency loss is about 20% but further optimizations may recover some of the performance loss.

More detailed studies are needed to address the issue of light lepton misidentification, but all the results so far indicate that hadronic taus can be successfully used even during the HL-LHC run to perform the precision measurement of the Higgs boson coupling to tau leptons.

9.8 Summary of Physics Object Performance

The reconstruction of the physics objects of the SM with high efficiency, good resolution, low backgrounds, and low fake rates is the key to discovering new physics, either by direct observation of new particles or by detection of departures from SM behavior from precision measurements. The excellent performance of the original CMS detector resulted in the discovery of the Higgs boson with only $25\,\text{fb}^{-1}$ of integrated luminosity. The Phase-I upgrade will ensure that this high level of performance does not degrade too badly up to an integrated luminosity of 300-500 fb^{-1} and for pileup as high as 50. However,the performance of the Phase-I detector at the HL-LHC would be seriously degraded because of the higher pileup of 140-200 and the damage from radiation doses of $\sim 300\,\text{fb}^{-1}$ per year and it would not be functioning at all midway through the planned ten years of operation. The simulations presented in this chapter demonstrate that the proposed Phase-II upgrade will restore the object performance to roughly the same level as the Phase-I detector before it has experienced any degradation from radiation. The upgrade will allow CMS to realize the full benefits of the HL-LHC upgrade and will ensure that it discovers any new physics that might be present at mass scales of up to a few TeV, as shown in the next chapter.

Chapter 10

Exploring the High Luminosity LHC Physics Program

The standard model (SM) of particle physics is a framework developed for the description of fundamental particles and their interactions. It is one of the greatest triumphs of science, representing more than a century of experimental discoveries and theoretical breakthroughs. Nevertheless, we know that the SM is an incomplete theory. There are open fundamental questions and physics phenomena in nature that the SM does not adequately explain.

Cosmological and astrophysical observations show that the SM only describes about 15% of the total matter present in the universe. The remaining 85% should be dark matter, a hypothetical kind of matter that cannot be seen with telescopes. The SM does not supply any fundamental particles that are good dark matter candidates. This mystery has stirred worldwide interest. Another fundamental physical phenomenon not explained by the SM is that the Universe is mostly made out of matter. The SM predicts that matter and antimatter should have been created in almost equal amounts. In addition, the SM has some features, which point to a lack of understanding. The hierarchy problem is the most important of these features. The SM introduces particle masses through spontaneous symmetry breaking caused by the Higgs field. Within the SM, the mass of the Higgs boson gets very large quantum corrections due to the presence of virtual particles. These corrections are much larger than the actual mass of the Higgs boson. As a result the bare mass parameter of the Higgs boson in the SM must be fine-tuned to cancel these quantum corrections. The level of fine-tuning needed in the SM is considered to be unnatural.

New physics that would contribute particles and phenomena is expected to be found at the electroweak (EW) scale. One example of a very powerful extension of the SM that the CMS collaboration hopes to uncover is supersymmetry (SUSY), which predicts a whole new spectrum of fundamental particles that each partner with a particle in the SM. The SUSY partners bring an important balance to the overall description of nature at the most fundamental level that solves the hierarchy problem. In SUSY models the Higgs sector is modified such that the Higgs boson properties may deviate from the SM expectation and additional Higgs bosons are predicted. Moreover, in many SUSY models the lightest SUSY particle is predicted to be weakly interacting, stable, and as such an excellent candidate for dark matter. There are many other models that also provide tantalizing new ways to go beyond the SM. We do not know which, if any, of these models is the right one.

In most extensions of the SM, new particles can be expected at the TeV scale. To date, no such particles have been observed. This implies that their mass is above the current level of sensitivity, or that their cross sections are lower than expected, or that their experimental signatures are very difficult to observe.

Performance/ Physics	Higgs VBF H→ττ	Higgs H→μμ	Higgs H→ZZ→4l	Higgs HH→bbγγ	Higgs HH→bbττ	SMP VBS	SUSY VH(bb) +MET	EXO A_{fb} (Z')	EXO Dark Matter	EXO HCP	BPH $B_{s,d}$→μμ
Tracker											
Performance		mass resolution	mass resolution	b-tagging	b-tagging						mass resolution
Extensions	forward jets / MET		acceptance		MET resolution	forward jets	MET resolution	acceptance	acceptance		
Trigger											
Bandwidth	acceptance				acceptance						
Track Trigger	background rejection				background rejection						background rejection
Calorimeter											
ECAL	forward jets / MET		acceptance	acceptance	MET resolution	forward jets	MET resolution	acceptance	acceptance		
HCAL	forward jets / MET				MET resolution	forward jets	MET resolution				
Muons											
Extension			acceptance					acceptance	acceptance		

Figure 10.1: Overview of measurements enabled by Phase-II detector upgrades.

The HL-LHC physics program will build on the experience acquired, and the results obtained, in the first phase of the LHC operation. Independent of potential discoveries in this first period, the physics program will continue the quest to answer fundamental questions in particle physics. The ten times larger dataset of the HL-LHC will open new opportunities and transform the field.

The strategy is therefore to perform precision measurements of SM phenomena including the study of the Higgs boson and other processes sensitive to new physics, of BSM phenomena in case of a discovery, and to search directly for new physics. At the HL-LHC the tested parameter space is increased significantly in mass reach and in coupling strength. To explore the data delivered by the HL-LHC fully, it is crucial to maintain excellent quality of the detector performance throughout the lifetime of the experiment.

A number of examples of measurements are addressed below and a summary is given in the table shown in Figure 10.1. The selected measurements are enabled by the proposed upgrades of the CMS detector, namely the upgrade of the tracking system with an extension of the pseudorapidity coverage, the increase trigger bandwidth and the capabilities of the track trigger, the upgrades to the barrel and the replacement of the endcap calorimeter, and the extension of the muon system in pseudorapidity coverage. We begin the discussion with a study of the Higgs sector and vector boson scattering before we move to searches for SUSY and more exotic new physics possibilities. The last channel discussed is the study of dimuon events from B_s and B_d decays, where the improved tracking resolution and the availability of the track trigger enable significant progress.

10.1 Exploitation of the Higgs boson

In 2012 the LHC experiments ATLAS and CMS observed a new particle consistent with the properties of a Higgs boson [228–230] with a mass of about 125 GeV. This discovery strongly suggests the validity of the Higgs mechanism for electroweak symmetry breaking and mass generation for the fundamental particles.

Within the SM, once its mass is known, all other properties of the Higgs boson can be predicted with high accuracy. Presently, using the Run-I data of the LHC, a combined mass determination of the ATLAS and CMS experiments gives a value of $m_H = 125.09 \pm 0.21(stat.) \pm 0.11(syst.)$ GeV [231], i.e. a measurement with a precision of 0.2%. Many other properties, in particular the couplings of the fundamental particles to the Higgs boson, are far less well measured, typically to a precision of the order of 20%. Recently, a novel method to determine the total width of the Higgs boson resonance was explored, using off-shell Higgs boson production with decay to Z-boson pairs, leading to an upper limit of about 5.4 times the expected value of the width in the SM, that is predicted to be around 4 MeV [232].

The measurements of Higgs boson properties made with the Run-I data are already impressive, but a number of important measurements cannot be done yet. The most prominent one is the measurement of the Higgs boson trilinear self-coupling, that will directly probe the shape of the Higgs field potential, testing whether it has the typical "mexican hat" shape or a more complex structure. This measurement can be performed using events containing a Higgs boson pair. Its cross section is however a factor 1000 smaller than single Higgs boson production, hence its study requires large data samples.

High integrated luminosity is also needed for precision measurements of many of the properties of the Higgs boson, such as its couplings to other particles, a detailed CP analysis of the Higgs boson, determination of its total width, and measurements of rare decay channels, to list some aspects of the program. The Higgs boson properties are expected to be sensitive to new physics. Given its role in the SM the Higgs boson could also be a portal for interactions with Dark Matter. Studies [233, 234] show that it is mandatory to measure the properties of the Higgs particle as precisely as possible, at least to the level of a few %, in order to access and discriminate between a broad class of new physics scenarios.

CMS has previously reported projections on the expected sensitivity of several Higgs boson decay channels and in particular on signal strengths and Higgs boson coupling modifiers [9]. These are based on the measurements that have been made public in 2013, extrapolated to the larger datasets of 300 and 3000 fb^{-1}, and for a center-of-mass energy of 14 TeV, by scaling the signal and background yields accordingly. These extrapolations assumed that the CMS upgrades will provide the same level of detector and trigger performance as in the 2012 data taking period. The estimated precision on the measurements for modified couplings for a SM-like Higgs boson is shown in Figure 10.2. Two scenarios are used to extrapolate uncertainties, namely scenario 1 where all systematic uncertainties are left unchanged and scenario 2 where the theoretical uncertainties are reduced by a factor of two and all other systematic uncertainties are reduced by the square root of the integrated luminosity. The two scenarios are expected to bracket a realistic extrapolation. The precision is increased by more than a factor of two with the HL-LHC dataset.

Complementary to Higgs boson precision measurements are searches for possible other family members of a more extended Higgs sector, both at higher and lower masses than 125 GeV.

The LHC is, for the foreseeable future, the only machine capable of studying the Higgs particle. The HL-LHC is designed to deliver a factor 10 more luminosity than what is anticipated from

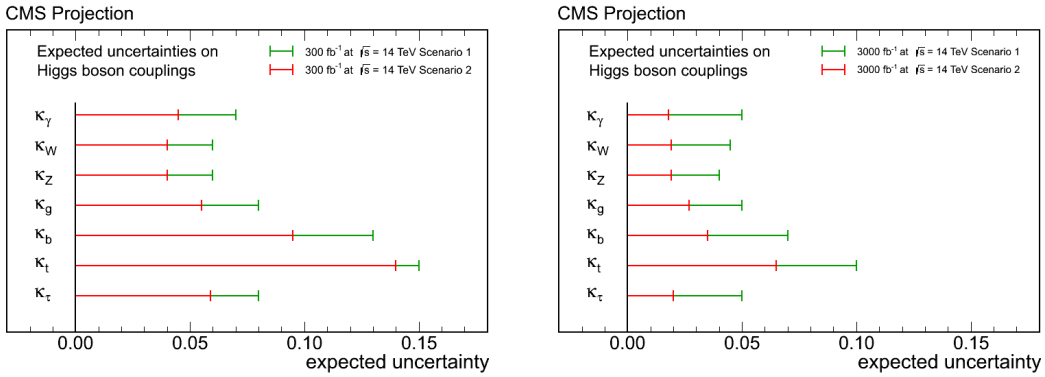


Figure 10.2: Estimated precision on the measurements for modified couplings for a SM-like Higgs boson [9]. The projections assume a center-of-mass energy of 14 TeV and a dataset with integrated luminosity of 300 and 3000 fb^{-1}. The projections are obtained with two uncertainty scenarios as described in the text.

the LHC Run-II and III. It is a unique opportunity for in-depth Higgs boson studies if the experiments can fully benefit from the higher data rates. The key element will be to have a similar, or possibly even better, sensitivity for Higgs boson detection with upgraded detectors, that are capable of dealing with the higher pileup at the very high luminosity operation of the HL-LHC. For the Higgs boson studies in this document we assume an average of 140 pileup events per bunch crossing. Hence, it is important to verify the experimental capabilities in detail, by studying a few key processes, as well as exploring the potential for new channels such as the Higgs boson pair production.

10.1.1 $H \to ZZ^* \to 4\ell$ analysis

The Higgs boson decay into two Z bosons, each decaying into two charged leptons, electrons or muons, is the golden channel in the study of the Higgs boson. Electrons and muons can be measured very accurately, with high efficiency, and excellent energy and momentum resolution. The complete final state of the Higgs boson decay can be reconstructed which leads to a signal of high purity, measured as a peak over a smooth background distribution. The four lepton events allow for a detailed CP analysis of the Higgs particle by measuring angular distributions, such as the angle between the ZZ decay planes and the decay angles in these planes, which contain information of the CP properties of the parent boson. The very accurate determination of the production rate of $H \to ZZ$ bosons is crucial for searches for deviations from the SM in the study of Higgs boson couplings. As mentioned before, the analysis of the full four lepton mass spectrum contains information of the total width of the Higgs boson.

The HL-LHC will produce about sixteen thousand Higgs boson events per experiment in the golden channel through the gluon-fusion production process and about 1400 through vector boson fusion. Selecting these events with the largest possible acceptance is crucial to the Higgs physics program. Excellent electron and muon reconstruction at low transverse momentum and a large rapidity coverage are key. Since four leptons need to be reconstructed in this final state, any single object inefficiency is potentiated. The Run-I analysis uses electrons with a p_T requirement as low as 7 GeV and muons with a p_T as low as 5 GeV with a pseudorapidity coverage of up to $|\eta| < 2.5$.

Figure 10.3 shows the efficiency times acceptance for the Higgs boson signal as a function of the selection criteria following the selection applied to the Run-I data. A significant increase in

the event selection efficiency can be observed with the Phase-II upgraded detector with respect to the configuration simulating the aged Phase-I detector. In terms of efficiency the upgraded detector shows the same performance as the Phase-I detector with lower pileup.

The four lepton mass distributions for a pileup of 140 events are shown Figure 10.4 normalized to an integrated luminosity of 3000 fb^{-1}. The signal is shown together with the irreducible $ZZ \rightarrow 4\ell$ background. The aged Phase-I detector and the Phase-II detector scenarios are considered. The increased signal yields and the improved mass resolution obtained with the Phase-II upgraded detector are clearly visible.

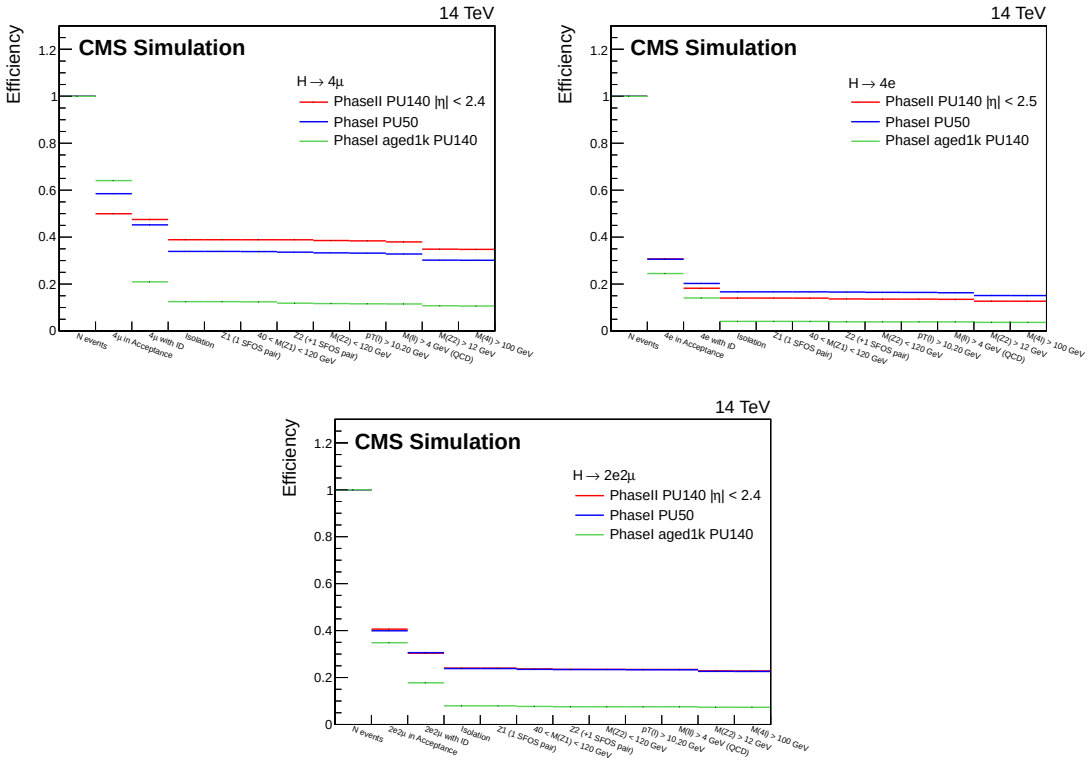


Figure 10.3: The cut flow table for the full analysis chain is shown for the Phase-I detector with pileup of 50 (blue), the aged Phase-I detector with pileup 140 (green) and the Phase-II detector with pileup of 140 (red) for the signal sample $H \rightarrow ZZ^* \rightarrow 4\ell$. A significant increase in the selection efficiency after full selection can be observed with the Phase-II detector.

In order to assess the further increase in signal yields made possible by an extension of the pseudorapidity coverage of the muon detector, we have parametrized the expected performances of the Phase-II detector up to $|\eta| = 4.0$, using the Delphes fast simulation [22]. We have processed the $H \rightarrow ZZ \rightarrow 4\mu$ signal final state through this simulation and the cut flow table results are shown on Figure 10.5 for $|\eta| \leq 2.4$, 3.0, 3.5 and 4.0. A significant increase in signal efficiency extending the muon detector coverage to larger η values is visible: the Phase-II detector with a muon coverage of $|\eta| \leq 3.0$ shows a 20% larger acceptance with respect to a coverage of $|\eta| \leq 2.4$.

10.1.2 $H \rightarrow \mu\mu$ analysis

The coupling of the Higgs boson to fermions has been established during the Run-I of the LHC. First evidence has been reported by CMS with measurements of third-generation fermions,

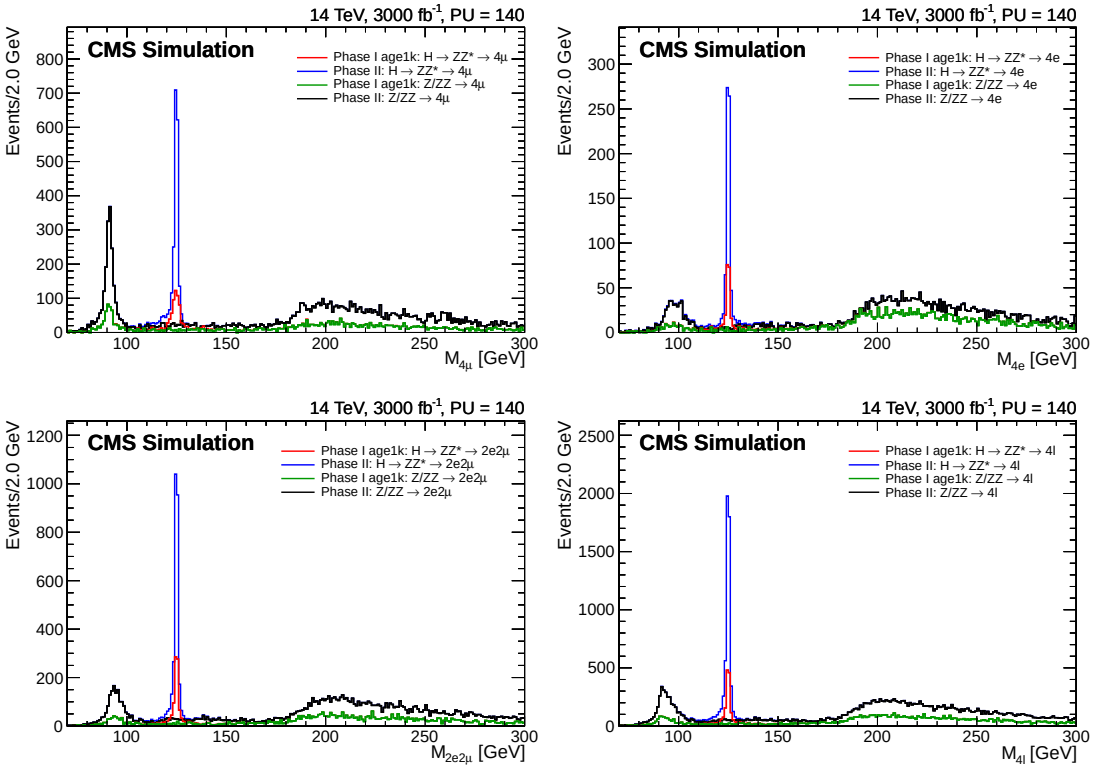


Figure 10.4: Four lepton mass distributions obtained with 3000 fb^{-1} for the signal sample, $H \rightarrow ZZ \rightarrow 4\ell$, and for the irreducible $ZZ \rightarrow 4\ell$ background. Both processes have been simulated with the aged Phase-I detector with pileup of 140 and the Phase-II detector with pileup of 140. The bottom right plot shows the sum of all 4ℓ final states.

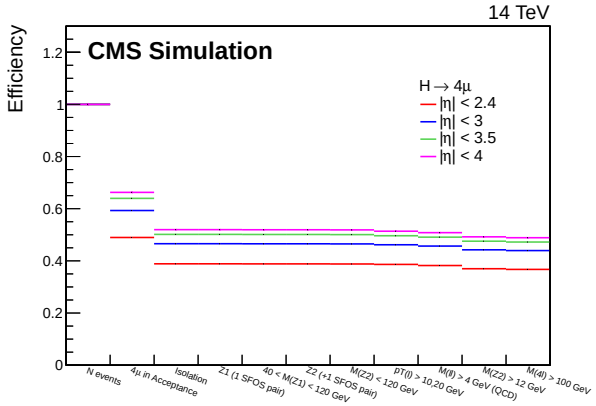


Figure 10.5: The cut flow table for the full analysis chain is shown for the Phase-II detector with pileup of 140, parametrized with Delphes, for the signal sample $H \rightarrow ZZ^* \rightarrow 4\mu$ for various hypotheses of the muon detector coverage.

namely through studies of the decay to b-quarks and tau-leptons [221]. The HL-LHC will give unique access to Higgs boson couplings to second-generation fermions. Measurements of the couplings of the Higgs boson to the second generation are more challenging, as a result of their smaller values and hence smaller experimental rates. The most promising channel is the search

for the decay of a Higgs boson in two muons with a branching fraction of 2.2×10^{-4} expected in the SM. The channel has been studied using the Run-I data reaching a sensitivity of 6 times the SM prediction [235]. Projections of the analysis to the HL-LHC era show that the coupling of the Higgs boson to muons can be measurement with a precision of about 8%, still limited by the statistical uncertainty.

In $H \to \mu^+\mu^-$ events the kinematics of the Higgs boson can be fully reconstructed. The signal will consist of a small bump over a large di-muon background from Drell-Yan events, hence an excellent di-muon mass resolution is crucial. This puts constraints on the required performance of the new tracking system. Special attention will be given to the study of the vector-boson fusion channel that exhibits experimental advantages compared to the gluon fusion one, due to the reduced backgrounds and lower theoretical uncertainties in the Higgs production cross section.

Figure 10.6 shows a fit to the signal di-muon mass distribution for Higgs boson events simulated with the Phase-I and Phase-II detectors. The distributions are normalized to both the acceptance of each di-muon category and selection efficiency of events in each di-muon category.

Due to the reduction of material and better spatial measurements of the upgraded Phase-II tracking detector, the mass resolution is 40 % better and the efficiency to reconstruct the muon pair is 20 % larger with respect to an aged Phase-I detector. The measurement of the Higgs boson coupling to muons is expected to improve with the square-root of the improvement in resolution and efficiency. Based on previous projections, an uncertainty in the Higgs boson coupling to muons of about 5 % is expected.

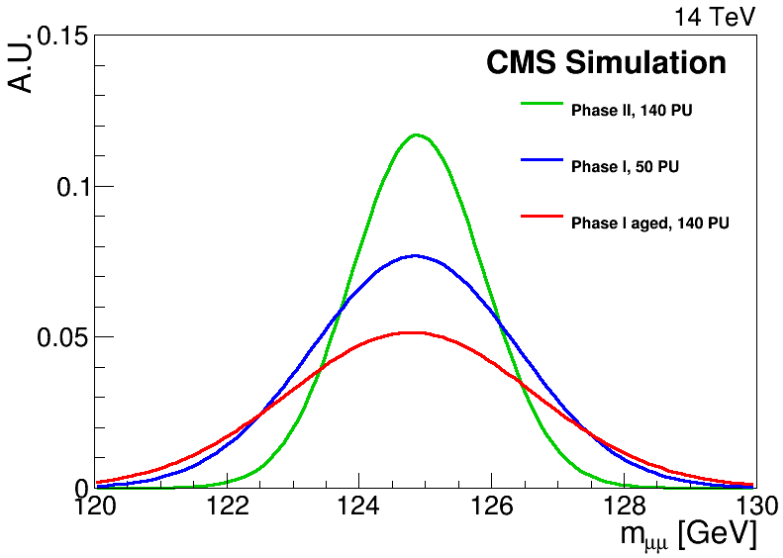


Figure 10.6: Di-muon mass distributions for Higgs boson events simulated with the Phase-I (nominal and aged) and Phase-II detectors. The distributions are normalized to take the relative selection efficiency of different detectors into account.

10.1.3 $H \to \tau\tau$ analysis

Projections of the Run-I $H \to \tau\tau$ analysis show that the coupling modification of the Higgs boson to tau leptons with respect to the SM expectation can be measured with a precision of 2–5 %. Modifications of the Higgs boson couplings to fermions of this scale or larger are expected

from some BSM Higgs models, particularly those predicting multiple Higgs doublets. This makes the $H \to \tau\tau$ channel an excellent probe for new physics.

The $H \to \tau\tau$ measurements rely strongly on the performance of almost all aspects of the CMS detector. The list of objects used in the analysis ranges from electrons, muons, hadronic taus, jets, b-tagged jets, to missing transverse energy. High efficiencies and low misidentification rates are crucial to control challenging backgrounds and to explore the $H \to \tau\tau$ decay in full. In this section we discuss the capability to match jets with the event primary vertex, missing transverse energy resolution, and trigger requirements.

To improve the signal to background ratio, events are categorized based on the number of reconstructed jets and the VBF production signature is explored. Additional jets from pileup collisions dilute the analysis performance unless they are identified and rejected. This can be achieved by matching the charged constituents of reconstructed jets to the primary vertex. Figure 10.7 shows the efficiency for associating a jet from the VBF process to the primary vertex as a function of the VBF tagging jet pseudorapidity. The tagging jet pseudorapidity reaches its maximum at $|\eta| = 3$. Starting with a generated VBF quark with $p_T > 30$ GeV matched to a jet, this is the efficiency for that jet to have $\beta > 0.1$, where β is defined as the p_T-weighted fraction of tracks within the jet cone from the primary vertex. The benefit from the extended coverage of the tracking detector is clearly visible.

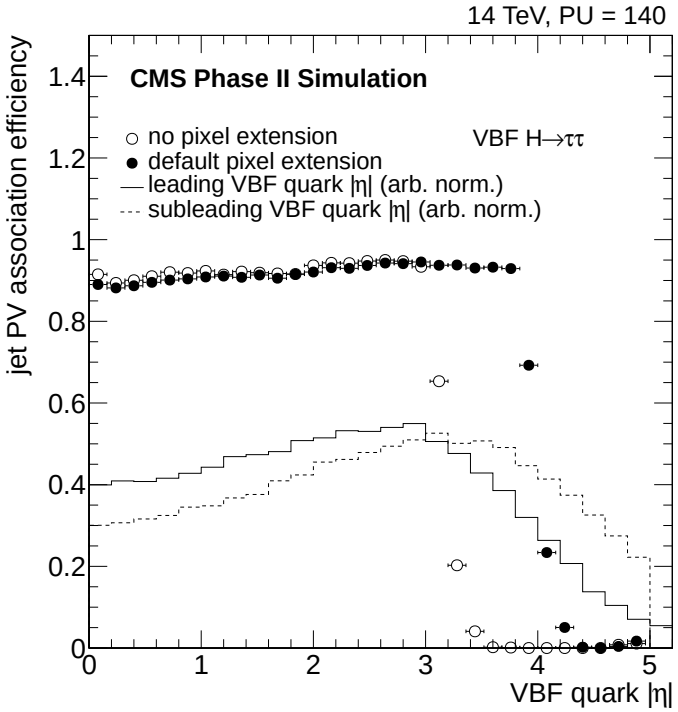


Figure 10.7: Efficiency for associating a jet from the VBF process to the primary vertex. The pseudorapidity distributions for the leading and sub-leading VBF quarks is shown in the same plot.

A likelihood-based mass reconstruction technique allows to discriminate the Higgs boson signal from the challenging Drell-Yan background. The performance of the analysis scales roughly linearly with the di-tau mass resolution. Figure 10.8 shows the di-tau mass resolution versus the missing transverse energy resolution normalized to the Run-I performance [222]. The de-

pendence of the di-tau mass resolution on the missing transverse energy resolution for the SV Fit algorithm is indicated by the black points. The resolution of the di-tau mass reconstructed from the visible tau decay is shown with a red line. The visible mass would be used in the analysis if the missing energy resolution is about 3 times the current resolution. The missing energy resolution for various detector configuration is indicated with colored lines. Comparing the orange line and the turquoise line, which represent the PUPPI missing transverse energy algorithm for the Phase-I and Phase-II detector, a significant improvement in resolution can be seen [192]. The main driver for this improvement stems from the extension of the tracking system to larger pseudorapidity. Future improvements in the reconstruction will further improve the missing transverse energy resolution. The purple line indicates the resolution using PF candidates. The resolution for PF MET is dominated by pileup, illustrating the need for pileup mitigation techniques. The green line shows a projection of the current performance to large pileup and agrees well with the performance estimated for the Phase-II detector. The agreement between the green and orange line underlines that the extended tracking enables efficient pileup mitigation for missing transverse energy reconstruction.

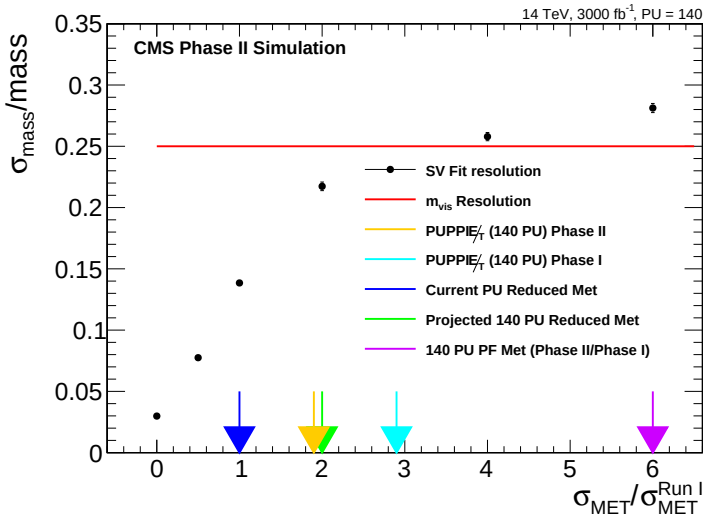


Figure 10.8: Di-tau mass resolution versus missing transverse energy resolution normalized to the Run-I performance. The dependence of the di-tau mass resolution on the missing transverse energy resolution for the SV Fit algorithm is indicated with the black points. The resolution of the mass reconstructed from the visible tau decay is shown with a red line. The missing energy resolution for various detector configuration relative to the Run-I performance is indicated with colored lines.

The leptons and pions from tau decays in Higgs boson events as well as VBF tagging jets have a typical p_T of 20 to 50 GeV. Excellent trigger capabilities are therefore crucial to fully explore the VBF and $H \to \tau\tau$ program at the HL-LHC.

Tracking information at the Level-1 trigger level will be essential to measurements in the doubly hadronic $H \to \tau\tau$ final state. The current hadronic tau triggers use only the available calorimeter information to discriminate between hadronic tau decays and jets. However, adding track information at the trigger level will allow dedicated triggers to more efficiently reject misidentified jets, which have substantially different structures from hadronic tau decays.

The Run-I $H \to \tau\tau$ analysis used a combination of double hadronic tau and double hadronic tau plus the one jet triggers for the fully hadronic final state, with a selection-level p_T threshold of 45 GeV in the $|\eta| < 2.1$ region. The L1 trigger menu including the track trigger lists a single

tau trigger with an offline threshold at 88 GeV and a di-tau trigger with an offline threshold at 56 GeV for each object with $|\eta| < 2.4$. The equivalent trackless trigger thresholds for the same rate are 140 GeV and 90 GeV, respectively.

A combination of the single and double tau triggers in the scenario with track trigger increases the absolute trigger acceptance by a factor of 1.7 compared to applying the Run-I analysis p_T threshold on in the same $|\eta|$ region, largely due to the addition of the single tau trigger which was not available in Run-I. Compared to the combination of tau triggers without the tracking capability, the track-aware scenario increases the absolute acceptance by 5.5 times, a huge gain from the addition of tracking information at the Level-1 trigger level. In summary, a precision Higgs boson program with taus will be very difficult without an upgrade of the trigger system including the track trigger.

10.1.4 Higgs boson pair production

Studies of Higgs boson pair production at the HL-LHC will provide insight on Higgs boson trilinear coupling [236]. This measurement would directly probe the Higgs field potential since the self-coupling is related to the third derivative of the Higgs potential at its minimum. The process is also sensitive to other BSM effects, as new physics can modify the rate of Higgs boson pair production. The dominant Higgs boson pair production mode at LHC is through gluon fusion. Figure 10.9 shows the dominant Feynman diagrams. Di-Higgs events can be produced via a box diagram and through the Higgs boson self-coupling contribution. The two processes interfere destructively and the cross section is near minimum for the SM. It should be noted that the cross section increases by a factor of two if the Higgs boson self coupling is zero. In fact, the Higgs boson pair production cross section is about 1000 times smaller than single Higgs boson production cross section.

The final states are generated with MADGRAPH [237] interfaced with PYTHIA6 [89] for parton showering and fragmentation. The sample is normalized to the NNLO cross section of 40.7 fb [238].

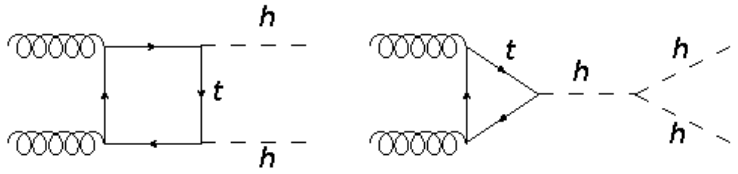


Figure 10.9: Feynman diagrams contributing to gluon fusion Higgs boson pair production.

Studies are performed of di-Higgs production and decay into for $bb\gamma\gamma$, bbWW, where the W boson decays leptonically, and $bb\tau\tau$. Its crucial that the Phase-II detector can cope with the challenging environment of HL-LHC, as pileup mitigation, b-tagging, tau-tagging, photon identification efficiencies, and mass resolutions are instrumental to perform these measurements. Triggers are assumed to be 100% efficient in these studies. Delphes fast simulation [22] is used for bbWW results. The parameterized performance of the Phase-II detector in Delphes is taken from the corresponding GEANT-based [79] full simulation samples. The $bb\gamma\gamma$ analysis uses MC truth-level information with smearing functions to model the performance of the Phase-II detector. A combination of the two approaches mentioned above is used for the $bb\tau\tau$ final state.

10.1.4.1 $bb\gamma\gamma$ final state

The signal events of interest contain two high-p_T photons and two high p_T jets originating from b quarks. Only 320 such events are expected to be produced per experiment at HL-LHC with 3 ab^{-1}. The backgrounds can be broadly categorized into resonant backgrounds which contain a Higgs boson decaying to two photons, and non-resonant backgrounds, which do not contain Higgs boson decays. The main resonant backgrounds are ZH, where a Higgs boson is produced in association with a Z boson, which subsequently decays to two b-jets, $t\bar{t}H$, where a Higgs boson is produced in association with a top quark-antiquark pair, and $b\bar{b}H$, where a Higgs boson is produced in association with a b quark-antiquark pair. The non-resonant QCD backgrounds considered include production of $b\bar{b}\gamma\gamma$, production of $jj\gamma\gamma$ with light jets mis-tagged as b-jets, production of $b\bar{b}j\gamma$ and $b\bar{b}jj$ with one or two jets mis-identified as photons, respectively, and production of four jets dominated by mis-tagged charm jets. The non-resonant background processes have cross sections that are several orders of magnitude larger than the resonant backgrounds but are suppressed by low rates of mis-tagged jets and mis-identified photons expected for the Phase-II detector. Finally, $t\bar{t}(\gamma)$ enters as a background when both top quarks decay semi-leptonically to produce electrons and the electrons are subsequently mis-identified as photons.

Events containing two photons with p_T greater than 25 GeV and $|\eta| < 2.5$, and two b-tagged jets with p_T greater than 30 GeV and $|\eta| < 2.4$ are selected. In addition, one of the two photons is required to have $p_T > 40$ GeV. Due to the larger background coming from jets being mis-identified as photons in the endcap region of the detector, the events are split into categories, one with both photons in the barrel and one with at least one photon in the endcap. To suppress $t\bar{t}H$ background it is required that no electrons or muons are reconstructed and that the number of jets with $|\eta| < 2.5$ is less than four. Further kinematic requirements are applied to suppress the background events. It is required that the ΔR between the two photons and the ΔR between the two b-tagged jets is less than 2.0 and the minimum of the ΔR between photons and b jets is greater than 1.5.

Process / Selection Stage	HH	ZH	$t\bar{t}H$	$b\bar{b}H$	$\gamma\gamma$+jets	γ+jets	jets	$t\bar{t}(\gamma)$
Object Selection & Fit Mass Window	23.8	30.5	184	6.5	3721	1619	287	597
Kinematic Selection	13.4	15.1	3.4	2.1	192	98	20	22
Mass Windows	9.0	3.4	1.6	0.8	13.0	6.3	1.1	1.2

Table 10.1: The expected event yields of the signal and background processes for 3000 fb^{-1} of integrated luminosity are shown at various stages of the cut-based selection for the both photons in the barrel region. Mass window cuts are 120 GeV to 130 GeV for $M_{\gamma\gamma}$ and 105 GeV to 145 GeV for M_{bb}. A large fit mass window, 100 GeV to 150 GeV for $M_{\gamma\gamma}$ and 70 GeV to 200 GeV for M_{bb}, is used for the likelihood fit analysis. The statistical uncertainties on the yields are of the order of percent or smaller.

A two-dimensional maximum likelihood fit of the di-photon, $M_{\gamma\gamma}$, and di-b-jet, M_{bb}, mass distributions is performed to extract the signal yield and uncertainty. The expected event yields for signal and backgrounds for various stages of event selection when the two photons are in the barrel region is summarized in Table 10.1. Figure 10.10 shows the di-photon mass distribution for signal and background processes together with the result of a pseudo-experiment. The median expected uncertainty in the signal yield is about 67%. Figure 10.11 shows the evolution of this result as the b-tagging and photon identification efficiencies change. From this figure one can directly derive the impact on the analysis from improvements on the detector,

the reconstruction, or the analysis strategy.

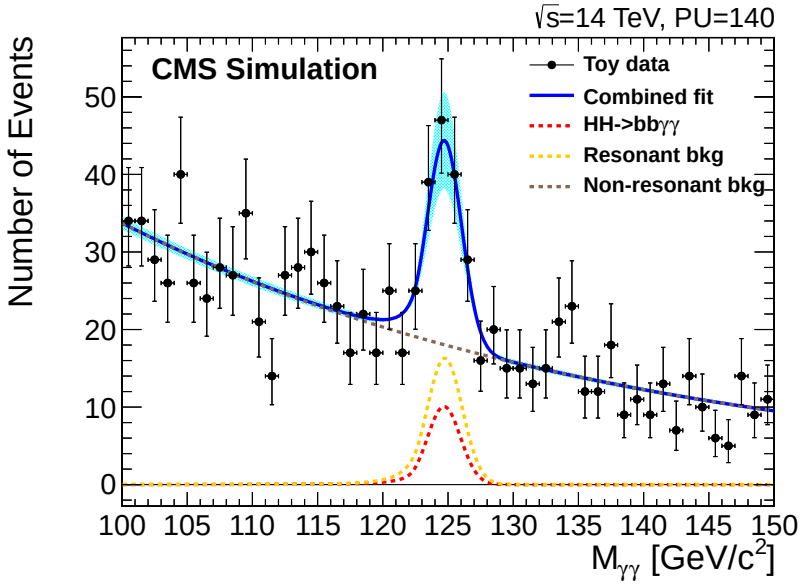


Figure 10.10: Di-photon mass distribution for the estimated signal and background contributions. The data points show the result of a pseudo-experiment.

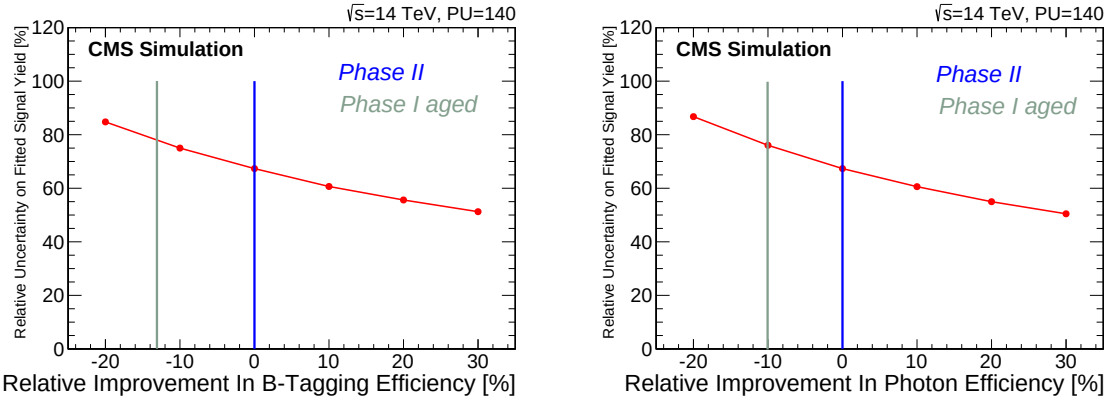


Figure 10.11: The median expected relative uncertainty on the signal yield as a function of the relative change of b-tagging (left) and photon identification (right) efficiencies.

10.1.4.2 $bb\tau\tau$ final state

The $\tau_\mu\tau_h$, and $\tau_h\tau_h$ di-tau final states, where τ_h denotes hadronic tau decays, and τ_μ denotes tau decays to muons, are studied. About 9000 $bb\tau\tau$ di-Higgs events per experiment are expected at HL-LHC with 3 ab^{-1}. However, the $t\bar{t}$ background with fully leptonic decays to taus is overwhelming. Another source of large background is Drell-Yan production of a Z boson decaying into a pair of tau leptons produced in association with jets, where light jets are mis-tagged as b-jets. The important single Higgs boson backgrounds are ZH, where the Higgs boson is produced in association with a Z boson, and $t\bar{t}H$, where the Higgs boson is produced in association with a top quark-antiquark pair. The remaining backgrounds considered are single top and $t\bar{t}$ produced in association with a vector boson, and di-boson processes. The QCD multi-jet

background is negligible in the signal region, as verified by studying the LHC data available at $\sqrt{s} = 8\,\text{TeV}$.

Events are selected containing two b-tagged jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$, and two taus with $p_T > 60\,\text{GeV}$, or $p_T > 90\,\text{GeV}$ for the leading tau and $p_T > 45\,\text{GeV}$ for sub-leading tau, and $|\eta| < 2.1$ for the $\tau_h\tau_h$ di-tau final state, $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$ for the τ_h and $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$ for the τ_μ in $\tau_\mu\tau_h$ di-tau final states. To further reduce background events with light jets mimicking hadronic tau decays it is required that jets originating from hadronic tau decays contain an isolated track. Selections are applied on the di-tau mass, $M_{\tau\tau}$, and the di-b-jet mass, M_{bb}, mass distributions to identify Higgs boson decays to tau and b pairs, respectively. The requirement for m_{bb} is $90\,\text{GeV} < m_{bb} < 130\,\text{GeV}$, and $110\,\text{GeV} < m_{\tau\tau} < 140\,\text{GeV}$ for $m_{\tau\tau}$. A likelihood-based mass reconstruction technique (SVFIT) is used to reconstruct the di-tau mass.

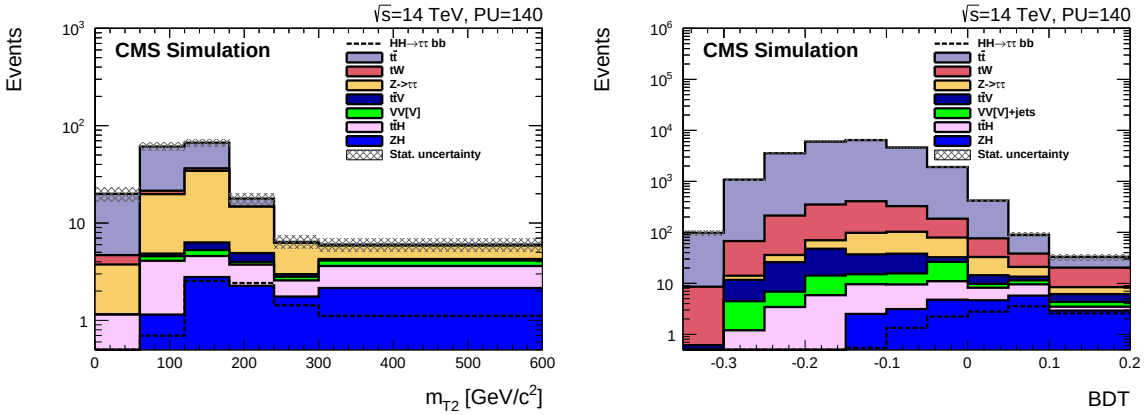


Figure 10.12: m_{T2} (left) and BDT score (right) distributions in $\tau_h\tau_h$ and $\tau_\mu\tau_h$ channels, respectively. The yields are the expected SM contributions.

A kinematic bounding variable, m_{T2}, is introduced to further discriminate the dominant $t\bar{t}$ background from the di-Higgs signal [239]. By construction, m_{T2} is bounded above by the top quark mass for $t\bar{t}$ background events while it is unbounded for di-Higgs signal events. For the $\tau_\mu\tau_h$ di-tau final states a BDT discriminant was trained to further exploit the boosted kinematics of di-Higgs production. The input variables are the masses, transverse momenta, and ΔR distances of the di-tau, di-b-jet, and di-Higgs systems. The m_{T2} variable is also included in the training. Figure 10.12 shows the distributions of the BDT discriminant for $\tau_\mu\tau_h$ channel on the left and m_{T2} distribution for the $\tau_h\tau_h$ di-tau final state on the right. The m_{T2} distribution is used to extract the signal in the $\tau_h\tau_h$ di-tau final state and the BDT discriminant for the $\tau_\mu\tau_h$ di-tau final state.

The expected significance for di-Higgs boson production is 0.5, and 0.7 standard deviations, for $\tau_\mu\tau_h$, and $\tau_h\tau_h$ di-tau final states, respectively. For the combination 0.9 standard deviations are expected. The resulting expected uncertainty in the signal strength is approximately 105%. Theoretical uncertainties in the Higgs boson production are included in this result. Renormalization and factorization scale uncertainties in the di-Higgs signal production are 20% for NNLO calculation. The PDF uncertainty is 9%. The systematic uncertainty on luminosity is taken to be 2.6%. Scale uncertainties on jets, tau leptons, and missing energy are also included.

The performance of the trigger system is crucial to achieve the result described above, in particular the capability to trigger on charged particles at Level-1. For the $\tau_h\tau_h$ final state, the di-tau trigger has an offline threshold of 56 GeV on both tau legs, and single tau trigger threshold is 88 GeV for the Level-1 sample menu described in this document. These thresholds are signifi-

cantly higher without the track trigger, 95 GeV on both tau legs for di-tau trigger and 138 GeV for single tau trigger. Considering these less performant thresholds, the signal and background yields are reduced by about a factor of two. For the $\tau_\mu\tau_h$ final state the situation is similar. The single-muon trigger threshold is 18 GeV with track trigger and 50 GeV without track trigger. The thresholds for muon-tau trigger legs are significantly higher as well. Again, the signal and background yields are reduced by a factor of two by requiring 50 GeV cut on the p_T of the muon and hadronic tau. Thus, in both final states the effect on the sensitivity of this analysis is significant. The overall sensitivity is reduced by 40%, the equivalent of using only half of 3000 fb^{-1}.

10.1.4.3 $bbWW$ final state

About 1500 fully leptonic signal events per experiment are expected at the HL-LHC, where the leptons are either muons or electrons. The dominant background process is the $t\bar{t}$ production and fully leptonic decay. Other backgrounds have negligible contribution in comparison to $t\bar{t}$ and only the dominant $t\bar{t}$ background is studied. Selected events are required to have two b tagged jets with $p_T > 30$ GeV and two opposite-sign leptons with muon $p_T > 20$ GeV, electron $p_T > 25$ GeV, and all objects with $|\eta| < 2.5$. Additional requirements that remove reduce the background include a requirement on the di-lepton mass, $M_{ll} < 85$ GeV, di-bjet mass, 60 GeV $< M_{bb} <$ 160 GeV, ΔR between the two leptons, $\Delta R_{ll} < 2$, ΔR between the two b-tagged jets, $\Delta R_{bb} < 3.1$, and $\Delta\phi$ between the di-b-jet and the di-lepton systems, $\Delta\phi_{bb,ll} > 1.7$.

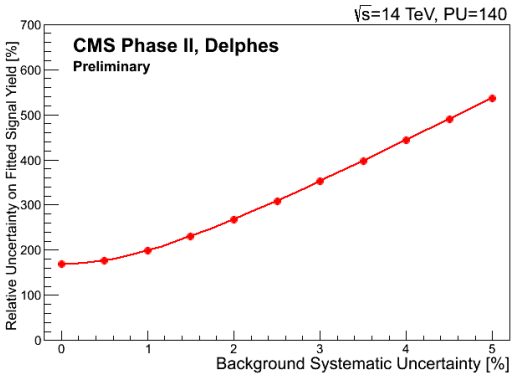


Figure 10.13: The average expected relative uncertainty on the signal yield as a function of the background systematic uncertainty.

A neural network (NN) discriminator based on the kinematic properties of the event is trained to further reduce the background. The NN takes into account the correlations among the input variables used for training. The signal selection is obtained by applying a threshold on the NN discriminator leading to 3875 background events and 37.1 signal events. Figure 10.13 shows the expected uncertainty in the di-Higgs signal yield as a function of the background uncertainty from 0% to 5%. The results suggest a promising contribution of this final state when combined with the other final states at the HL-LHC.

10.1.4.4 Conclusion

The measurement of Higgs boson pair production at the HL-LHC directly probes the Higgs boson trilinear coupling. The studies presented here show that an observation of the process is possible in combination of multiple channels and using results from ATLAS and CMS. Combining the studies of $bb\gamma\gamma$ and $bb\tau\tau$ final states, the expected significance for di-Higgs boson

production is 1.9 standard deviations. The resulting expected uncertainty on the signal yield is approximately 54%. These studies are a snapshot of the understanding today and we expect significant improvements in future studies of di-H signatures.

10.2 Exploring vector boson scattering

Precision measurements in the electroweak sector, and studies of processes that have sensitivity to the electroweak symmetry breaking, are particularly interesting. In this section we discuss examples of massive vector boson scattering (VBS), which are expected to be very sensitive to any new physics in this sector, and are currently unobserved.

One cornerstone of the SM is the electroweak symmetry breaking mechanism, where vector bosons acquire mass through their coupling to the Higgs field. Along with that, the scattering amplitudes of longitudinally polarized W or Z vector bosons ($V_L V_L \rightarrow V_L V_L$) are unitarized through the presence of the Higgs boson. In fact, when acquiring mass, the vector bosons gain their longitudinal degree of freedom with respect to the massless formulation of the theory, by coupling to the Higgs boson. The presence of these additional vertices introduces the new terms in the scattering cross section calculation that prevent the final result to be divergent at high energies. If the SM is an effective theory of a more general one with an additional strongly-coupled sector, the unitarization of VBS is only partially operated by the Higgs boson, and the full regularization needs to happen via new physics, which intervenes at a large energy scale. In a large data sample VBS may be used to determine that scale, and possibly resolve different models that can be nearly degenerate at the Higgs resonance, while showing different behaviors at higher energies.

This section presents studies of the potential of the CMS experiment at the HL-LHC in this context. The VBS signal will be accessible when two quarks from the beams emit vector bosons, which in turn interact with each other. The two quarks get deflected from the beam direction and originate jets of hadrons in the detector, called tag jets, which clearly identify this category of events. In this way the HL-LHC is used as a vector boson collider, providing access to purely electroweak processes in a high energy regime. Any new resonances in the VBS invariant mass spectrum, or a deviation of data from the SM expectations, would be a ground-breaking result, direct evidence of the scale at which new physics enters into play.

The small cross section of this electroweak process, and the large background due to the production of vector boson pairs and two hadronic jets via strong interactions (irreducible background), make this analysis very challenging. Besides, in the harsh experimental conditions of the HL-LHC, an average of about 140 events will overlap with each other at each beam collision (pileup). In this high-occupancy environment, background contamination can also arise from cases where hadronic jets are by mistake identified as leptons (j-l misidentification rate), or jets from different events overlap and mimic the features of tagging jets. Therefore, the new central tracking system is beneficial in reducing the j-l misidentification rate and its extension to $|\eta| = 4$, together with a new radiation resistant and highly granular forward calorimeter, reduces the contamination of jets from pileup events and allows to trigger on these events with higher efficiency.

For this study, only fully leptonic decays of the vector bosons are selected, each decaying into electrons or muons, in order to avoid the ambiguities due to the separation of the tag jets from the vector boson decay products. In particular, projections are presented for same-sign WW scattering and WZ scattering, which are among the final states that are most sensitive to new physics. The sensitivity to the SM processes is determined through the measurement of EW VBS cross sections and of its longitudinal component $V_L V_L \rightarrow V_L V_L$. The sensitivity to partial unitarization scenarios is studied as an indicator for the determination of new strongly-coupled sectors in the theory. Events are generated where the Higgs boson does not contribute to the unitarization of VBS and this alternative hypothesis is compared to the SM case. New physics resulting in enhanced VBS cross sections at high energy could be due to high-mass scalar reso-

nances, such as Higgs bosons in theories with extended Higgs sectors, or new high mass vector bosons, such as those expected in grand unified theories with larger gauge groups. Such additional states would alter the strong interference between the components of VBS processes and potentially lead to strong enhancements of the cross section at high energy. We describe these types of effects using the EFT field theory formalism, where in the parameterization FS-type operators correspond to the effects of scalar particles and FT-type operators correspond to the effect of transversely polarized vector bosons. While dimension six operators are well constrained by other measurements, the VBS process is modified by dimension eight operators through anomalous quartic gauge couplings. Therefore, we simulate events with non-zero contribution for each of the nine possible operators in the EFT expansion, in a range comparable to the current sensitivity. The coefficients for the additional terms in the Lagrangian are scanned independently, and for the FS0-FS1 case also a simultaneous scan is performed, since the combination FS0-FS1 provides a direct measurement of new physics that only modifies the SM quartic gauge coupling.

All these benchmarks are tested for the possible future upgrade configurations of the CMS apparatus, and in the case of the aged current detector.

The signal and irreducible background have been generated at leading order with the PHANTOM [240] and MADGRAPH [206] generators, while the detector response has been simulated parametrically with DELPHES [22]. The aging effect on the detector components is determined in the detailed simulations and parametrized in the present study through its effect on event selection efficiencies and on systematics uncertainties related to the reconstruction and identification physics objects. Detector effects specific to these analyses, such as the j-l misidentification rate have been evaluated with the detailed simulation of the CMS detector, accounting for the specific conditions of each upgrade design. Anomalous couplings are studied in the frame of the EFT formalism [241] implemented in the MADGRAPH generator, while non-unitarized scenarios are described by the PHANTOM program. Final-state leptons are considered if they are found in the detector with $|\eta| < 2.4$ and $p_T > 20$ GeV, and jets have to satisfy $|\eta| < 4.7$ and $p_T > 30$ GeV.

The performance of the analysis has been evaluated, for each final state, after applying rectangular selections on m_{jj} and $\Delta\eta_{jj}$ exploiting the typical VBS signature. Since the signal is an EW process, little color flow is expected in the event. Therefore, events are required to have H_T <125 GeV (150 GeV if the tracker extends up to $|\eta| = 4$), where H_T is the energy flow measured with jets composed only by tracks. This quantity is insensitive to pileup, since tracks are required to originate from the primary vertex of each event, and its effectiveness depends on the extension of the tracker detector. A minimal requirement is also put on the measured missing transverse energy, for both the WW (40 GeV) and WZ (30 GeV) studies.

A wide set of variables has been investigated for each of the benchmarks, including the kinematic properties of each single object and the angles among them. The analysis performance is determined with a template fit on the most sensitive ones, along one or two dimensions. When meaningful, the analysis is split according to the charge and flavour of the final state leptons. Figure 10.14, left, shows the $\Delta\phi$ between the two final state charged leptons for the WW scattering, after the VBS selections. The signal is represented by the red line, stacked over the backgrounds. The dashed purple line shows the different expectation in the extreme case when the Higgs boson does not play any role in the WW scattering unitarization (no-Higgs scenario). The difference of the Higgsless scenario and the SM one is used as signal on top of the SM itself, as an indicator of the analysis sensitivity to models where the Higgs boson performs a partial unitarization of VBS. The distributions on the right-hand side show the expected dif-

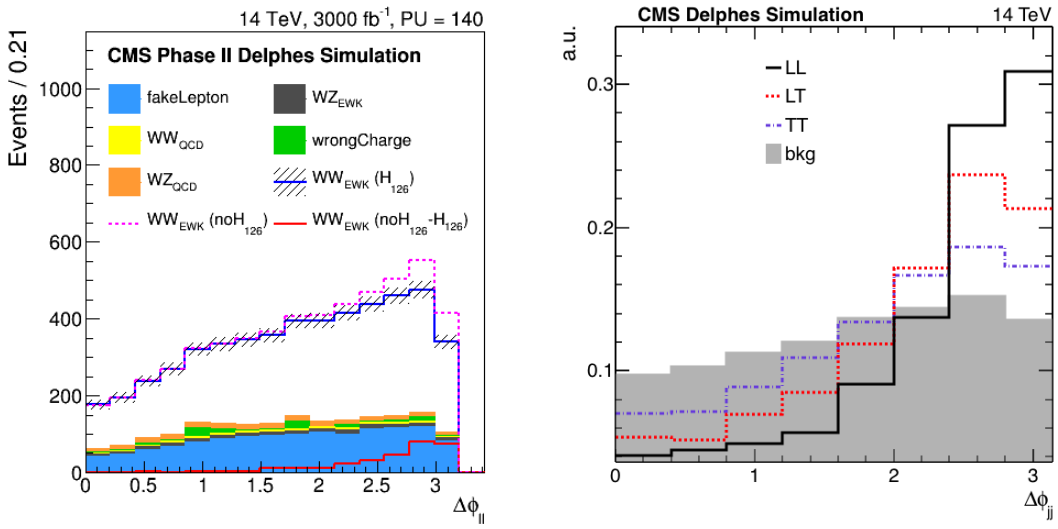


Figure 10.14: Left, the $\Delta\phi$ between the two final state charged leptons for the same-sign WW scattering, after the VBS selections, for positive muons in the non-aged Phase-I scenario. Right, an example of the expected differences for polarized scatterings in the WZ analysis.

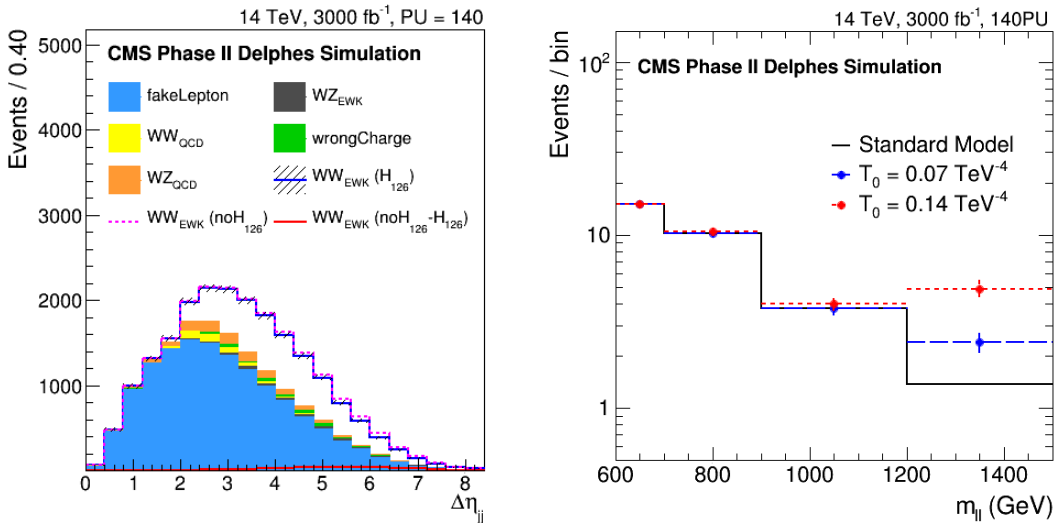


Figure 10.15: Left, the differences in shape of $\Delta\eta_{jj}$ between the two final state jets for the WW scattering, before the VBS selections, for signal and background. Right, the $m_{\ell\ell}$ distribution at the end of the analysis chain, still for the WW scattering, showing the expectation for the SM case and for a signal hypothesis with enhanced anomalous couplings.

ferences for polarized scatterings in the WZ analysis. In this case the $V_L V_L \to V_L V_L$ signal is searched for on top of the background and the other scattering components. Figure 10.15, on the left, shows the $\Delta\eta_{jj}$ between the two final state jets for the WW scattering, before the VBS selections. On the right, the $m_{\ell\ell}$ distribution is shown at the end of the analysis chain, still for WW scattering, showing the expectation for the SM case and for a signal hypothesis with enhanced anomalous couplings.

10.2.1 The same-sign WW scattering

Besides the two tag jets, the final state of this channel is composed of the two same-sign charged leptons from the W bosons decay (either electrons or muons) and missing transverse energy due to the undetected neutrinos. The irreducible background is well kept under control, as clearly visible in Figure 10.14, left. On the other hand, the $t\bar{t} \rightarrow WbWb$ process, where one W boson decays leptonically and the other one decays hadronically generates an important source of contamination when one of the jets (either from the hadronic W boson decay or from b quarks) is wrongly identified as a charged lepton. Therefore, additional selections are put in place: only two well-identified and isolated same-sign leptons, with $p_T > 20$ GeV, have to be reconstructed, together with two non-b-tagged jets with $p_T > 30$ GeV. No additional loose leptons can be found in the event, and $m_{\ell\ell}$ cannot have a value close to the Z boson mass. We also require $m_{\ell\ell} >$12 GeV and $\Delta\eta_{\ell\ell} < 2$. The actual values of the VBS thresholds have been optimized for each benchmark independently, and are of the order of 600–800 GeV for m_{jj} and 2.5 for $\Delta\eta_{jj}$. Events are discarded if a soft muon is found inside a jet, to reduce the $t\bar{t}$ background. The two leading leptons are required to be within the tag jets along the η direction and the di-jet and di-lepton systems shall be closer than 6 units of ΔR. The b-tagging efficiency depends in principle on the tracker extension, but given the complexity of the algorithms, to be cautious it is considered non zero only up to $|\eta| = 2.5$. After these requirements, the expected number of events after 3000 fb^{-1} of integrated luminosity are reported in Table 10.2 for the various possible upgrade scenarios considered. Besides the $t\bar{t}$, also the production of a W boson

Scenario	WW			WZ	
	Signal	I-bkg	R-bkg	Signal	I-bkg
Phase-I 50 PU	2958	387	1389	986	1364
Phase-II 140 PU	3155	374	1662	1100	1513
Phase-I aged 140 PU	2210	347	1662	746	1148

Table 10.2: The expected number of events, after the analysis selections, for same-sign WW and WZ final states. Signal, irreducible background (I-bkg) and reducible one (R-bkg) are reported. Each line corresponds to a possible scenario for the CMS operations.

with jets is accounted for in the reducible backgrounds. Since the j-l misidentification rate is determined through simulation only, results are reported as a function of a global scale factor multiplied to the misidentification rate itself, which has been in turn derived from a detailed description of the CMS detector in similar conditions. Events with leptonically decaying WZ boson pairs and jets, when a lepton from the Z is not detected, are also considered.

After 3000 fb^{-1} of integrated luminosity the total cross section of the EWK production of same-sign W bosons plus two jets will be determined with a precision at the order of 5%, after accounting for the main expected sources of uncertainty.

Figure 10.16 shows the expected significance for the discovery of the longitudinal W scattering, obtained from a two-dimensional template fit for the $(\Delta\phi_{jj}, p_T(\ell_1))$ variables, and as a function of a global scale factor applied to the j-l misidentification rate, where ℓ_1 denotes the lepton with the highest p_T one in each event. Figure 10.17 instead shows the expected 95% CL exclusion limit for the Higgsless scenario, obtained from a two-dimensional template fit in the $(R, m_{\ell\ell})$ variables, where R is defined as in [242]. The result is reported in terms of the strength modifier μ of the difference with respect to the SM case, as a function of a global scale factor applied to the j-l misidentification rate. If the 95% confidence level exclusion is reached for μ=1, the analysis would be sensitive to the case when the Higgs boson does not participate to the WW scattering unitarization. If the exclusion is attained for values of μ smaller than one, the analysis will be sensitive to scenarios of partial unitarization as well. As can be seen from these

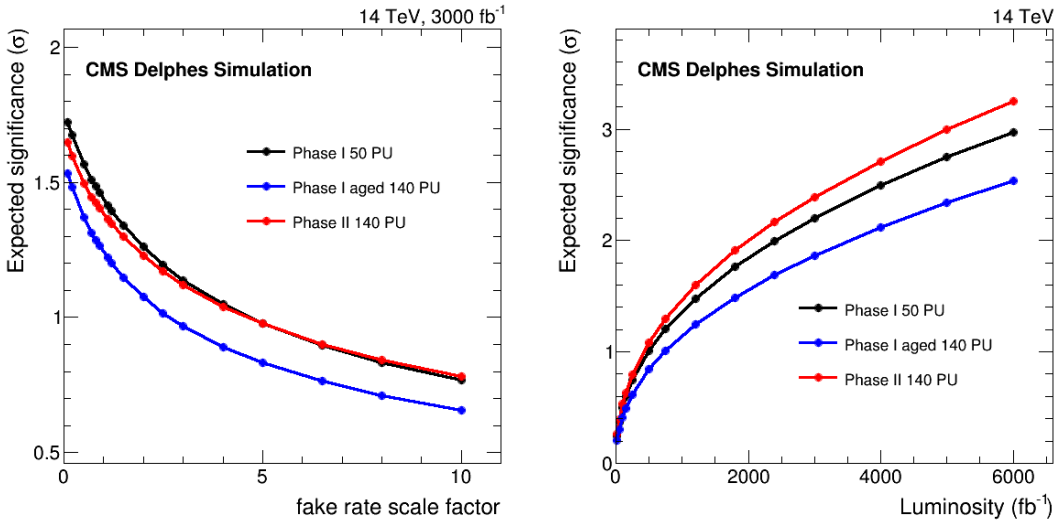


Figure 10.16: The expected significance for the discovery of the longitudinal same-sign WW scattering after 3000 fb^{-1} of integrated luminosity as a function of the data/MC j-l misidentification rate scale factor (left), and as a function of the integrated luminosity when the j-l misidentification rate scale factor is set to one (right).

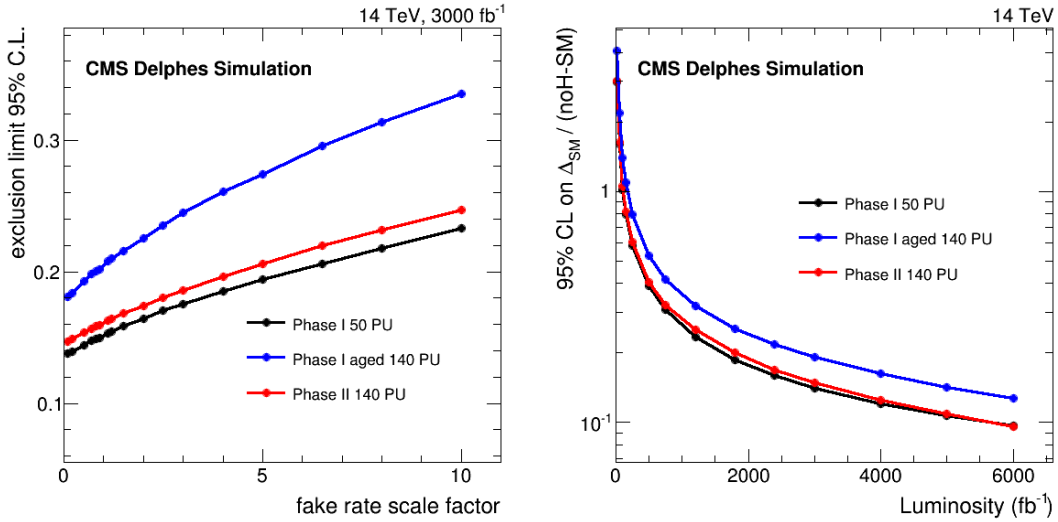


Figure 10.17: The expected 95% CL exclusion limits for the Higgsless scenario after 3000 fb^{-1} of integrated luminosity and as a function of the data/MC j-l misidentification rate scale factor (left), and as a function of the integrated luminosity when the j-l misidentification rate scale factor is set to one (right). The limit is expressed as deviation from the SM divided by the difference of the Higgsless case from the SM itself.

projections, the upgraded CMS detector is more sensitive than the aged version of the current one and recovers the same performances one would get with the same luminosity, acquired during data conditions of the LHC Run-I.

Table 10.3 shows the expected 95% CL limits on the coefficients for BSM higher-order operators in the EFT Lagrangian in the various scenarios considered. Since the deviations from the SM happen in high energy tails of the distributions, and the results approach the systematic limit already, the difference between the scenarios is typically at the order of 10%.

	Phase-I	Phase-II	Phase-I aged
S_0	1.06	1.07	1.17
S_1	3.51	3.55	3.87
M_0	0.78	0.75	0.82
M_1	1.10	1.06	1.14
M_6	1.56	1.49	1.63
M_7	1.37	1.32	1.45
T_0	0.067	0.077	0.083
T_1	0.036	0.033	0.036
T_2	0.119	0.111	0.119

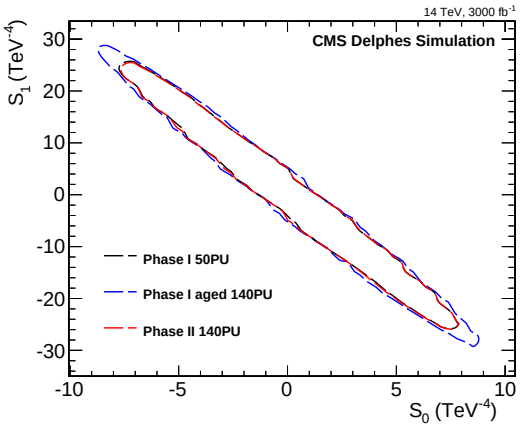


Table 10.3: Expected 95% CL limits on the coefficients for BSM higher-order (dimension eight) operators in the EFT Lagrangian (left) and 2D contour for the S_0 and S_1 parameters (right).

10.2.2 WZ scattering

This channel is identified by searching for three charged leptons in the final state, where two of them have opposite sign, same flavour and an invariant mass compatible with the one of a Z boson. The undetected neutrino generates missing transverse energy, and the longitudinal component of its momentum can be determined by requiring it to be produced, together with the charged lepton, in the decay of a W boson. The background due to Drell–Yan events plus jets, when one jet is wrongly identified as a charged lepton, is not significant in this case, while the irreducible component is more important than in the same-sign WW scattering. The high-granularity forward calorimeter technology and the tracker extension play a crucial role in identifying and removing pileup jets. Besides the VBS selections of $m_{jj} > 600$ GeV and $\Delta\eta_{jj} > 4.0$, events should have a third reconstructed and identified lepton, with $|\eta| < 2.4$ and $p_T > 20$ GeV. Same-flavour, opposite-sign lepton pairs should not have a mass consistent with the Z boson mass within 6 GeV and have at least 20 GeV of mass. After these conservative requirements, the expected number of events after 3000 fb^{-1} of integrated luminosity are reported in Table 10.2 for the various possible upgrade scenarios considered.

Figure 10.18, left, shows the expected significance on the total cross section of the EWK production of WZ boson pairs plus two jets as a function of the integrated luminosity. Values are determined with a two-dimensional fit in the $(p_T(jj), \Delta\eta_{\ell^\pm\ell^\pm})$ variables, and results are reported for the different upgrade scenarios. The $\Delta\eta_{\ell^\pm\ell^\pm}$ represents the pseudo-rapidity difference between the two same-sign leptons identified in each event. Figure 10.18, right shows the expected significance for the discovery of the longitudinal WZ scattering, obtained with a two-dimensional fit for the $(\Delta\phi_{jj}, p_T(\ell_1))$ variables. As can be seen from these projections, the upgraded CMS detector is more sensitive than the aged version of the current one. The highest expected exclusion power for the Higgsless scenario with respect to the SM has been obtained fitting $(p_T(\ell_3), \Delta\eta_{jj})$, yielding a 95% confidence level exclusion of a strength modifier of about 0.5 in the Phase-I case, with 140 pileup events.

10.2.3 Combined WW and WZ analysis

Results from the WW and WZ analyses have then been combined, to determine the overall performance for the longitudinal scattering observation and the search for deviations from the SM due to partial unitarization schemes. Table 10.4 shows the relevant results of the combination, assuming a scale factor of one for the j-l misidentification rate.

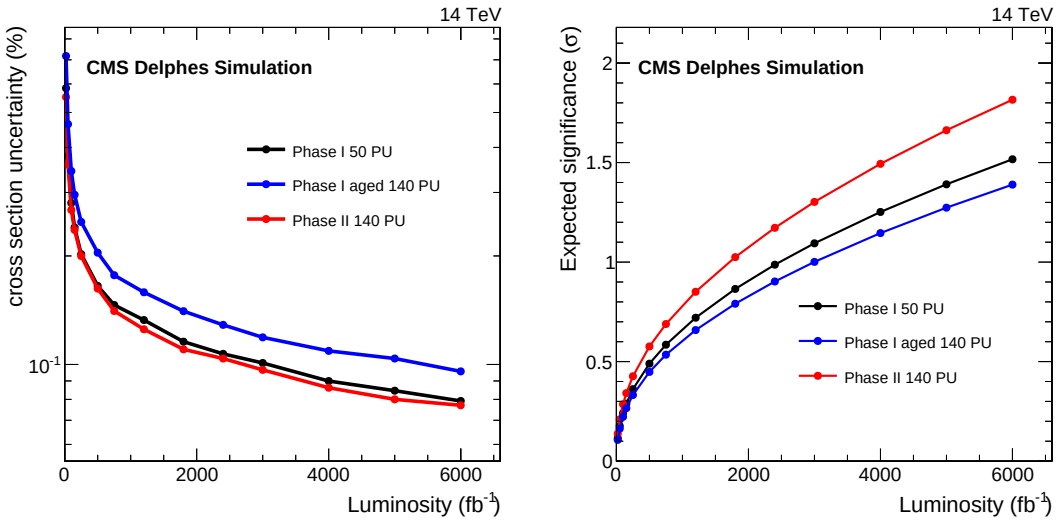


Figure 10.18: The expected cross section uncertainty of the EW WZ scattering (left), and the one for the longitudinal component of the scattering in the same final state (right) as a function of the integrated luminosity.

3000 fb^{-1}, 14 TeV	Phase-I	Phase-II	Phase-I aged
Higgsless 95% CL μ exclusion	0.14	0.14	0.20
V_L V_L scattering significance	2.50	2.75	2.14

Table 10.4: Results of the combination of WW and WZ analyses, assuming a scale factor of 1 for the j-l misidentification rate, for the longitudinal scattering observation and the search of deviations from the SM due to partial unitarization schemes.

In conclusion, these preliminary studies show that the upgraded CMS detector will recover the performances of the current one and in some cases grant an improvement in the physics performances necessary for the verification of the EWSB in a model-independent way with respect to the detailed description of the low-mass Higgs resonance.

10.3 Supersymmetry searches and measurements

The search for supersymmetry (SUSY) is a major goal of the LHC physics program. The range of SUSY scenarios is broad, and both high-luminosity data samples and the full set of CMS detector capabilities will be needed to provide sensitivity to the signatures, cross sections, and decay branching fractions of interest. If evidence for one or more new particles is discovered, an extensive program of measurements will be required to determine whether they are indeed SUSY partners of SM particles and to address even more challenging issues, such as discerning the mechanism of SUSY breaking. Because of the complexity of the SUSY spectrum and the associated decay processes, such a program would likely extend for many years, as was the case for studies of the standard-model hadron spectrum. This section presents results from a set of studies [243, 244] that address key questions on the physics potential of the anticipated SUSY program, assuming integrated luminosities from $300\,\text{fb}^{-1}$ (LHC Run 2+3) to $3000\,\text{fb}^{-1}$ (HL-LHC).

The motivations for supersymmetry are strong, in spite of the absence of specific evidence for this framework. First, SUSY provides a candidate particle, the lightest supersymmetric particle (LSP), that may account for all or part of the astrophysical dark matter. For example, in the minimal supersymmetric standard model (MSSM), the lightest neutralino, designated $\widetilde{\chi}_1^0$, is a superposition of the spin-1/2 superpartners of the neutral gauge and Higgs bosons, and it can in principle satisfy the constraints from indirect dark matter observations. Second, if a SUSY particle spectrum is present, the three running gauge couplings of the standard model interactions can converge at a common high energy, a requirement for gauge unification. Finally, the discovery of a Higgs boson by the CMS and ATLAS experiments has given new urgency to the gauge hierarchy problem. Assuming that the Higgs boson is a fundamental scalar particle, its mass is extremely sensitive to short distance quantum corrections. Without some kind of new physics that compensates for these effects, the Higgs mass would be pulled up to the Planck scale, barring a coincidental near-perfect cancellation of the bare Higgs mass parameter and the enormous shift induced by quantum corrections. A broad class of SUSY scenarios, known as natural models [245], can stabilize the Higgs boson mass through additional contributions involving diagrams with the scalar superpartner, $\widetilde{\text{t}}$, of the top quark (stop), as well as other SUSY particles.

With the discovery of a Higgs boson, the gauge hierarchy problem is no longer a hypothetical issue, and many searches in the current and future SUSY program are guided by expectations for natural SUSY models. In such models, the top squark, bottom squark, gluino, and higgsino are constrained to be relatively light, while the other superpartner masses are essentially unconstrained. Although the parameter space of natural SUSY models is only loosely defined, the high energy and high integrated luminosity of the full LHC program are generally regarded as sufficient to probe most of the relevant model space.

While natural SUSY models are perhaps the best motivated of all scenarios, they are by no means the only models of interest. The studies presented here consider several possibilities to provide a sense of the challenges posed by a range of models.

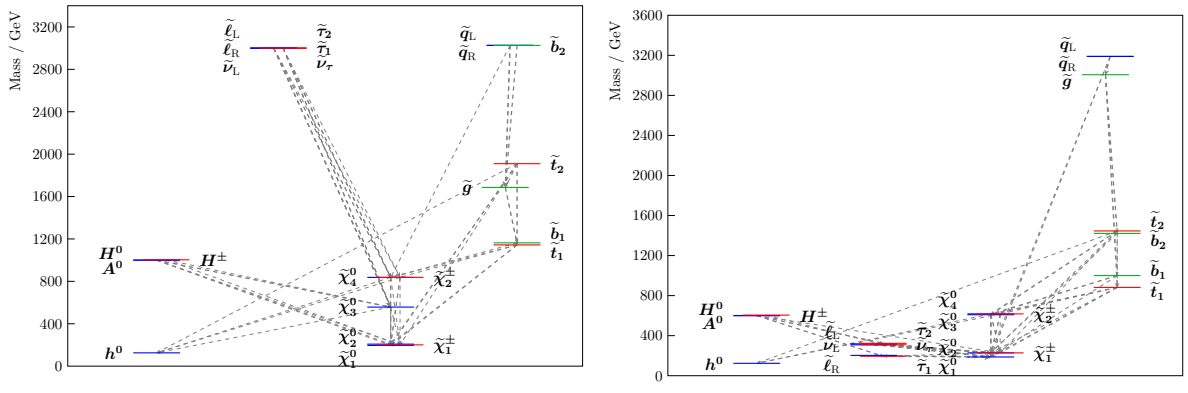
This discussion of SUSY searches is divided into five sections. Section 10.3.1 starts with a discussion of the different types of models that are used in the interpretation of SUSY searches. An overview of CMS Run 1 results is then presented, followed by projections for similar searches to 14 TeV and to higher luminosities. Detector aging effects are then discussed in the context of a particular SUSY search. All of the results presented in Section 10.3.1 are interpreted using simplified model spectra (SMS). Section 10.3.2 presents an overview of search strategies used in a new set of 14 TeV studies, which are based on nine separate analysis strategies. These

are applied to five full-spectrum benchmark SUSY models, which include three natural SUSY scenarios, as well as stau and stop coannihilation scenarios. The features of these models, and the selection requirements of the nine analyses are briefly described. Section 10.3.3 presents the results obtained from the nine analyses. We consider not only the discovery sensitivity, but also how, in the event of a discovery, the pattern of signals and the associated kinematic distributions can provide many clues to understanding the nature of the underlying particle spectrum. From these studies, it is clear that the full HL-LHC data sample will provide critical information, even if discoveries are made much earlier. Section 10.3.4 summarizes the main results and conclusions of these studies.

10.3.1 SUSY models, searches with Run 1 data, and simplified-model projections to higher energies

10.3.1.1 SUSY models used in the interpretation of searches

This section compares full-spectrum SUSY models, which are used in the new studies presented in later sections, with simplified-models, which have been used in many of the interpretations of CMS Run 1 data. Figure 10.19 shows the mass spectra for two of the SUSY models, NM3 and STC. The left-most column shows the particles that make up the Higgs sector, with the neutral, CP-even Higgs boson fixed at the observed mass, $m_H = 125\,\text{GeV}$. (All five models share this feature, but other aspects of their Higgs sectors can vary from model to model.) SUSY particles with electroweak interactions (only) are shown in the two middle columns. These are the scalar leptons and scalar neutrinos (sleptons and sneutrinos) and an electroweak sector consisting of the charginos and neutralinos, which are spin-1/2 fermions. The SUSY partners in the electroweak sector, sometimes referred to as ewkinos, are superpositions of higgsinos and gauginos of the same charge. The far-right-hand column shows the strongly interacting sector, with the gluino and the squarks. Because quarks are spin-1/2 objects, they each have two SUSY partners, corresponding to the L- and R-handed chiral projections, which have different gauge quantum numbers. Thus, $\widetilde{q}_L$ and $\widetilde{q}_R$ are distinct particles. Mixing is expected to be significant in the third generation of squarks, leading to mass eigenstates designated $\widetilde{t}_1$, $\widetilde{t}_2$, $\widetilde{b}_1$, and $\widetilde{b}_2$.



(a) NM3 (b) STC

Figure 10.19: Examples of SUSY full-spectrum models: (a) the natural SUSY model NM3 and (b) the stau coannihilation model STC, which are among the five full-spectrum scenarios used in the studies presented here. In NM3, the masses of the $\widetilde{g}$, $\widetilde{t}_1$, $\widetilde{t}_2$, and $\widetilde{b}_1$ are all below 2 TeV. The $\widetilde{\chi}^0_1$ is higgsino-like. In the STC model, the gluino is much heavier than the top squarks, and the slepton sector is light, with the $\widetilde{\tau}$ nearly degenerate with the $\widetilde{\chi}^0_1$. The lines between different states indicate transitions with branching fractions greater than 5%.

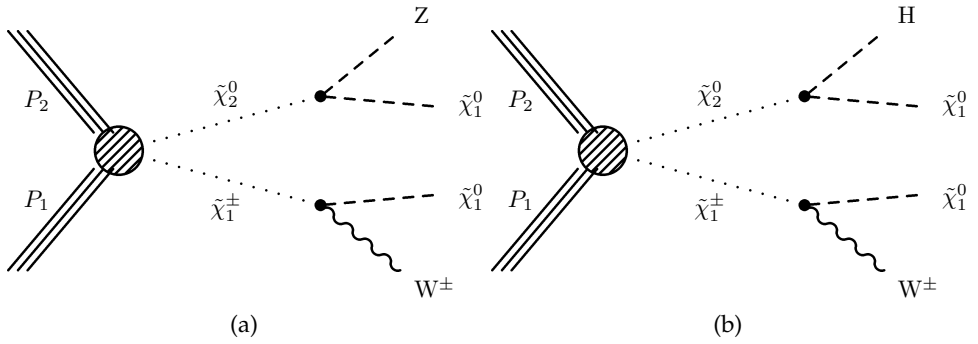


Figure 10.20: Simplified SUSY models for chargino-neutralino production leading to (a) $W^\pm Z + \not{E}_T$ and (b) $W^\pm H + \not{E}_T$ final states.

As we have noted, natural SUSY models are characterized by a small set of relatively light particles. This set consists of $\widetilde{t}_1$, $\widetilde{t}_2$, $\widetilde{b}_1$ (which in this case is $\widetilde{b}_L$), the gluino, and the higgsinos. The masses of the first- and second-generation squarks are not constrained to be light; the sleptons and the electroweak sector, apart from the higgsinos, are also not constrained. While the natural SUSY framework provides important experimental guidance, we will see that the phenomenology can vary substantially across these models.

The interpretation of the results of a SUSY search can be performed either with full-spectrum models, such as the NM3 and STC models shown in Fig. 10.19, or with simplified model spectra (SMS) [246, 247]. Simplified models limit the number of relevant SUSY particles for a particular signature to a minimal number, often just two. For example, a simplified model with just a single top squark, $\widetilde{t}$, and a lighter neutralino, $\widetilde{\chi}_1^0$, is sufficient to describe the scenario of top squark pair production with $\widetilde{t}_1 \to t\widetilde{\chi}_1^0$. Figure 10.20 shows production and decay diagrams associated with two simplified models relevant to the production of a chargino-neutralino pair ($\widetilde{\chi}_1^\pm \widetilde{\chi}_2^0$) pair, with different $\widetilde{\chi}_2^0$ decays, (a) $\widetilde{\chi}_2^0 \to Z\widetilde{\chi}_1^0$ and (b) $\widetilde{\chi}_2^0 \to H\widetilde{\chi}_1^0$.

Simplified models have been extremely important in helping to provide a systematic and general structure to guide the SUSY search program. In addition, the small number of mass parameters needed to describe such models greatly facilitates scans over the relevant range of the model space. It is important, however, to be aware of the nature of the simplifying assumptions and the limitations that they introduce. For example, by omitting certain particles from an SMS, the branching fraction for the searched-for signature can be artificially enhanced, since competing decay modes are eliminated. This would not affect a measurement (or upper limit) on the product of the production cross section and the branching fractions for the full decay chains, which are usually reported as part of the results. However, the inferred lower limit on the SUSY particle masses could be misleadingly high, because the calculation of these limits requires one to explicitly assume some values for the relevant branching fractions (usually, but not always, taken to be 100%). Interpretations based on both approaches, simplified-spectrum models and full-spectrum models, are discussed here. In the case of conclusions based on simplified models, we have tried to note the key assumptions.

10.3.1.2 Summary of CMS Run 1 SUSY results and sensitivity projections with simplified models

The CMS Run 1 SUSY exclusion limits [248] follow a pattern that is expected from the production cross sections, with strong interaction processes providing sensitivity to the highest mass scales. For a given SUSY particle mass, the largest cross section is associated with gluino

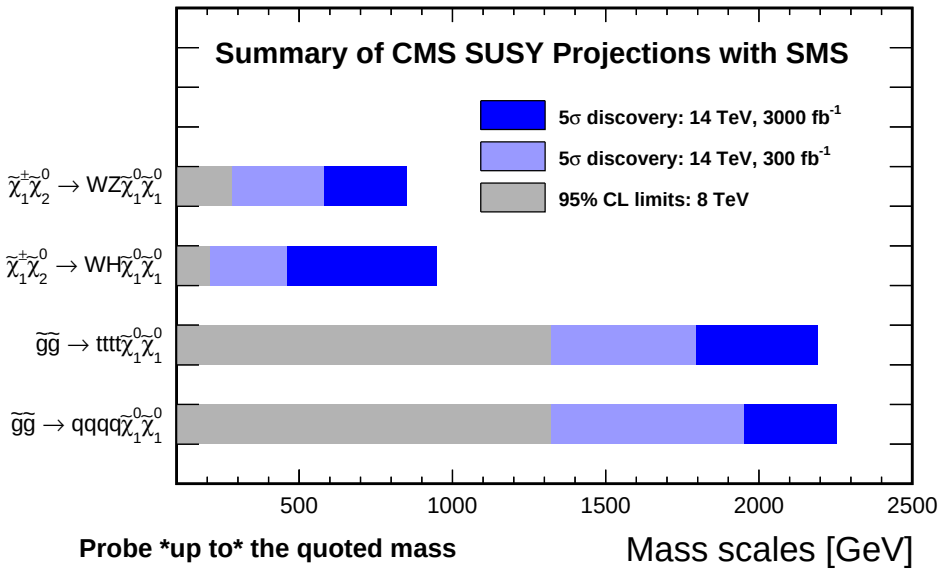


Figure 10.21: Mass reach of SUSY searches from selected 8 TeV results (masses excluded at 95% CL) and from projections for 14 TeV running at high luminosities (highest masses for 5σ observation). Simplified models are used for the interpretations in each case. The processes listed are, from top to bottom, the direct electroweak production of $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0$ pairs, using the $W^{\pm}\widetilde{\chi}_1^0 Z\widetilde{\chi}_1^0$ and $W^{\pm}\widetilde{\chi}_1^0 H\widetilde{\chi}_1^0$ final states; gluino pair production with $\widetilde{g} \to t\bar{t}\widetilde{\chi}_1^0$; and gluino pair production with $\widetilde{g} \to q\bar{q}\widetilde{\chi}_1^0$. For the latter two processes, the gluino decay can proceed via either on- or off-shell squarks. For the case $\widetilde{g} \to q\bar{q}\widetilde{\chi}_1^0$, eight intermediate-state squarks (assumed to be degenerate in mass) contribute (the L and R partners of each of the first- and second-generation quarks). The searches for final states reached via strong interaction processes in general have sensitivity to higher masses, but in most of natural SUSY models, the Higgsinos have the lowest expected masses.

pair production, and Figure 10.21 shows that gluinos are excluded [249] up to a mass of about 1.3 TeV, assuming an SMS decay scenario with $\widetilde{g} \to t\bar{t}\widetilde{\chi}_1^0$. This process takes place via $\widetilde{g} \to \widetilde{t}\bar{t}$ (charge conjugate final states implied), where the $\widetilde{t}$ can be produced either on or off mass shell. The limit applies only for $m(\widetilde{\chi}_1^0) < 500\,\text{GeV}$, because higher neutralino masses lead to insufficient $\not{E}_T$ to suppress the $t\bar{t}$ background. Limits have also been obtained for the case in which $\widetilde{g} \to b\bar{b}\widetilde{\chi}_1^0$ and for gluino decays to the sum of first- and second-generation squark-antiquark pairs, assuming squark mass degeneracy [250, 251].

Searches for the direct pair production of $\widetilde{t}$ quarks lead to mass limits of $m(\widetilde{t}) > 750\,\text{GeV}$ assuming that the decay proceeds via $\widetilde{t} \to t\widetilde{\chi}_1^0$ and that the mass difference $m(\widetilde{t}) - m(\widetilde{\chi}_1^0)$ is significantly larger than $m(t)$. Searches for pair production of first and second generation squarks, (assumed to be mass degenerate) yield the common mass limit $m(\widetilde{q}) > 900\,\text{GeV}$ in an SMS scenario with $\widetilde{q} \to q\widetilde{\chi}_1^0$ and $m(\widetilde{\chi}_1^0) < 300\,\text{GeV}$ [252].

Even though the cross sections for electroweak processes are much smaller, the signatures can be extremely distinctive, as in the case of a trilepton final state. Furthermore, in most natural SUSY scenarios, the Higgsinos expected to be the lightest SUSY partners. Searches [252, 253] for $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0 \to W^{\pm}Z + \not{E}_T$ imply $m(\widetilde{\chi}_1^{\pm}) > 275\,\text{GeV}$ assuming a 100% branching fraction and that $m(\widetilde{\chi}_1^{\pm}) = m(\widetilde{\chi}_2^0)$ with $m(\widetilde{\chi}_1^0) < 50\,\text{GeV}$. Alternatively, searches for $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0 \to W^{\pm}H + \not{E}_T$ exclude $m(\widetilde{\chi}_1^{\pm})$ in the range 140 GeV–220 GeV, assuming 100% branching fraction and that $m(\widetilde{\chi}_1^{\pm}) = m(\widetilde{\chi}_2^0)$ with $\widetilde{\chi}_1^0$ nearly massless. By combining these results, one can exclude $m(\widetilde{\chi}_1^{\pm})$ in the

range 140 GeV–220 GeV, again assuming that $m(\widetilde{\chi}_1^{\pm}) = m(\widetilde{\chi}_2^0)$ and that $\widetilde{\chi}_1^0$ is nearly massless. In all of these interpretations of the search results, the $\widetilde{\chi}_1^{\pm}$ and $\widetilde{\chi}_2^0$ are assumed to be Wino-like, which is relevant for determination of the expected production cross section.

Figure 10.21 also gives projected discovery sensitivities at 14 TeV with high integrated luminosities for several representative SUSY signatures. For each process, the mass that is discoverable at the 5σ level is shown for an integrated luminosity of $300\,\text{fb}^{-1}$ and $3000\,\text{fb}^{-1}$. For the gluino, the current 8 TeV 95% lower limit is at 1.3 TeV; the discovery reach at 14 TeV with $300\,\text{fb}^{-1}$ extends to 1.9 TeV and the reach at 14 TeV with $3000\,\text{fb}^{-1}$ is about 2.2 TeV. For some analyses, a 5σ discovery significance can nominally be achieved by $300\,\text{fb}^{-1}$, while in other cases, more data are required. These are highly model-dependent statements, however. The studies developed for this document consider several full-spectrum scenarios and also provide a picture of how the interplay among the different search results can be used to unravel the nature of excess yields in an highly complex model space.

10.3.1.3 Detector performance issues and effects of radiation damage

Virtually all aspects of the detector performance are relevant to SUSY searches, including jet reconstruction, b-jet tagging, lepton identification, photon identification, and especially the measurement of the missing transverse momentum vector, whose magnitude is denoted by $\not\!\!E_{\text{T}}$. The full range of detector capabilities is required to suppress the SM background, which typically includes events arising from $\text{t}\bar{\text{t}}$, W + jets, Z + jets, and QCD multijet production. Large values of $\not\!\!E_{\text{T}}$ indicate the production of a high-transverse-momentum particle that has only weak couplings, such as a neutralino in SUSY or a neutrino in the SM. In the context of natural SUSY models, b-jet tagging plays a critical role in searches for third-generation squarks, produced either directly or in gluino decays.

Leptons that are isolated from hadronic jet activity (which is typically measured within a cone around the lepton momentum direction using both tracking and calorimetry) are signatures of the electroweak decay of heavy particles, such as W- or Z-bosons. The measurement of lepton isolation can be affected by extra particles in the event that are produced in proton-proton collisions other than the one generating the lepton. These collisions can involve protons from either the same pair of in-time beam bunches (in-time pileup), or protons from the previous or subsequent beam bunches (out-of-time pileup). This is a key experimental issue that becomes more severe at high luminosity and 25 ns operation.

The degradation of the detector performance as the result of radiation damage will affect most observables relevant to SUSY searches, such as b-jet tagging performance and the $\not\!\!E_{\text{T}}$ resolution. To explore how such degradation can potentially affect a SUSY search, we have evaluated these effects for the case of electroweak production of a chargino-neutralino pair $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0$, with $\widetilde{\chi}_1^{\pm} \to \text{W}^{\pm}\widetilde{\chi}_1^0$ and $\widetilde{\chi}_2^0 \to \text{H}\widetilde{\chi}_1^0$ (see Fig. 10.20(b)). The $\widetilde{\chi}_1^0$ is unobserved, so the final state is $\text{W}^{\pm}\text{H} + \not\!\!E_{\text{T}}$, with the W decaying leptonically and $\text{H} \to \text{b}\bar{\text{b}}$.

The effect of the degradation in the trigger efficiency, lepton identification, b-tagging efficiency, and $\not\!\!E_{\text{T}}$ resolution have been estimated, and the key results are summarized in Figure 10.22. With an integrated luminosity of $300\,\text{fb}^{-1}$ (50 pileup interactions), the existing detector is expected to have 5σ discovery sensitivity in a small range of $m(\widetilde{\chi}_2^0) = m(\widetilde{\chi}_1^{\pm})$ around 450 GeV. With the upgraded detector, the 5σ discovery reach with $3000\,\text{fb}^{-1}$ is vastly larger, up to $m(\widetilde{\chi}_2^0) = m(\widetilde{\chi}_1^{\pm}) \simeq 950\,\text{GeV}$. (A mean of 140 pileup internations is assumed.) Furthermore, the allowed range of $m(\widetilde{\chi}_1^0)$ extends to much higher values. However, if the existing (and therefore degraded) detector is used instead of the upgraded detector, the gain in discovery reach, shown with $1000\,\text{fb}^{-1}$, is significantly reduced. This key search is particularly sensitive

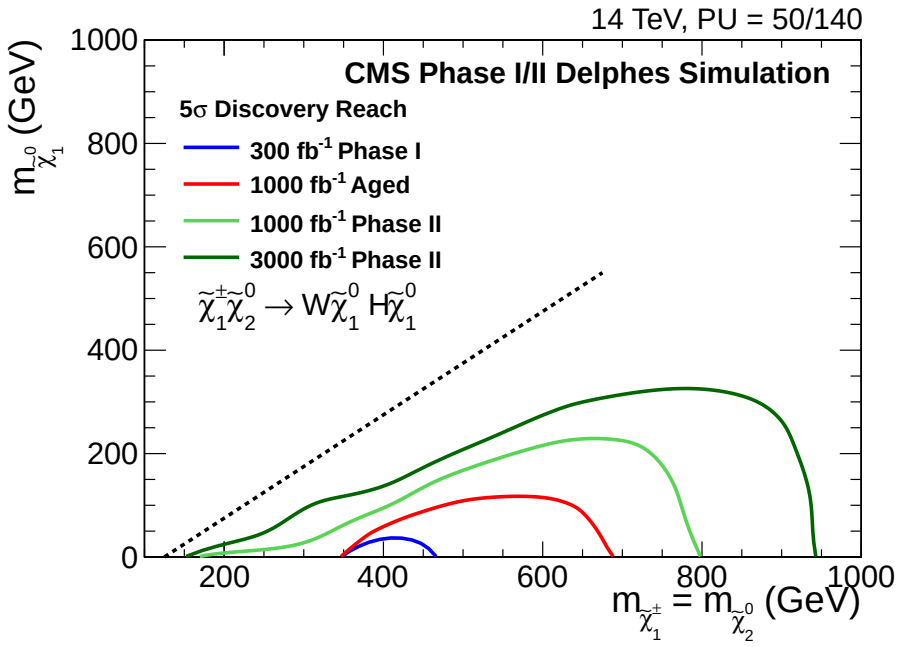


Figure 10.22: Effect of detector aging on the sensitivity for the SUSY search for $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ in the $W^\pm H + \not{E}_T$ final states. The estimated effects of degradation in the detector performance for b-quark tagging, lepton-identification, $\not{E}_T$ resolution, and trigger efficiency have been applied.

to detector performance and provides a demonstration of how important the detector upgrade can be.

10.3.2 Overview of SUSY search strategies and full-spectrum models

The searches presented here span a range of final states ranging from a single high-p_T jet recoiling against $\not{E}_T$ to signatures with three leptons and multiple b-tagged jets. The full SUSY search program involves many more signatures, but these nine studies allow us to address many of the key issues.

The scope and results of these studies are summarized in Figure 10.23, which lists the discovery significances obtained when the analyses are applied to the five full-spectrum benchmark models. This section describes the basic selection requirements for each of these searches and then discusses the main features of the models. The results of these studies are presented in Section 10.3.3.

The details of these searches, including a full description of the selection criteria and tables showing the reduction of backgrounds as the selection requirements are applied, are presented in a separate document [244]. Here we focus on the main topological aspects of the signatures, as well as the key kinematic variables used in the analyses.

- **All-hadronic final state with b-jets and the H_T and H_T^{miss} variables:** ≥ 3 jets, ≥ 2 b-tagged jets, no leptons, QCD suppression with $\Delta\phi(\text{jet}, H_T^{miss})$, $H_T > 2.5$ TeV, $H_T^{miss} > 1.3$ TeV.
- **All-hadronic final state with b-jets and the M_{T2} variable:** ≥ 8 jets, ≥ 3 b-tagged jets, no leptons, $H_T > 2$ TeV, M_{T2} thresholds $500 - 800$ GeV.
- **Direct bottom squark production in the dijet channel:** 2 b-tagged jets with $p_T > 300$ GeV and > 200 GeV, no additional jets, no leptons, $\Delta\phi(\text{jet}_1, \text{jet}_2) < 2.5$, $H_T >$

Analysis	Luminosity (fb^{-1})	Model				
		NM1	NM2	NM3	STC	STOC
all-hadronic (H_T-H_T^{miss}) search	300					
	3000					
all-hadronic (M_{T2}) search	300					
	3000					
all-hadronic $\tilde{b}_1$ search	300					
	3000					
1-lepton $\tilde{t}_1$ search	300					
	3000					
monojet $\tilde{t}_1$ search	300					
	3000					
$m_{\ell^+\ell^-}$ kinematic edge	300					
	3000					
multilepton + b-tag search	300					
	3000					
multilepton search	300					
	3000					
ewkino WH search	300					
	3000					

$< 3\sigma$ | $3 - 5\sigma$ | $> 5\sigma$

Figure 10.23: Overview over the SUSY search analyses and their application to the different full-spectrum models.

750 GeV, $\not{E}_T > 450$ GeV. $M_T(b_{1,2}, \not{E}_T)$ thresholds $500 - 900$ GeV.

- **Direct top squark production in the single-lepton channel:** 1 isolated lepton (e or μ), ≥ 5 jets, 1 or 2 b-tagged jets, centrality cut, $\not{E}_T > 400\ (800)$ GeV, $M_T > 260$ GeV, $M_{T2}^W > 260$ GeV.
- **Compressed SUSY spectra, including top squark production with $\tilde{t} \to c\tilde{\chi}_1^0$, in the monojet-like final state:** $p_T(\text{jet}_1) > 900$ GeV, $\Delta\phi(\text{jet}_1, \text{jet}_2) < 1.8$, veto events with third jet with $p_T > 100$ GeV, no leptons, $\not{E}_T > 600$ GeV.
- **Neutralino-slepton cascade signature using the kinematic edge in the dilepton mass ($m_{\ell^+\ell^-}$) distribution:** 2 opposite-sign, same-flavor isolated leptons (e or μ), ≥ 6 jets, ≥ 1 b-tagged jet, $\not{E}_T > 450$ GeV, $H_T > 1250$ GeV.
- **Electroweak production of chargino-neutralino pair in the $W^{\pm}H$ + E_T final state:** 1 lepton (e or μ), 2 b-tagged jets, $90 < m(b\bar{b}) < 150$ GeV, $M_{CT} > 160$ GeV, $M_T > 100$ GeV, $\not{E}_T$ thresholds 200 GeV to 500 GeV.
- **Gluino pair production in the trilepton + b-jets final state:** ≥ 3 leptons (e or μ), b-tagged jets: 2–3 bin and ≥ 4 bin, $\not{E}_T > 500$ GeV.
- **Electroweak production of chargino-neutralino pair in the $W^{\pm}Z$ + E_T final state:** ≥ 3 leptons (e or μ), On-Z and above-Z mass regions in $m_{\ell^+\ell^-}$, veto events with b-tagged jets, multiple signal regions in M_T vs. $\not{E}_T$.

While each search in this list is given a name that reflects both the nominal target process and the experimental signature, in many cases the signature is relevant to multiple SUSY processes.

In general, the lepton p_T requirements are in the range 15 GeV to 40 GeV, while jet p_T requirements are in the range 30 GeV to 60 GeV. An exception is the search for bottom-squark pair production, where the b-tagged jets are required to satisfy much higher threshold requirements, as noted above.

The studies described here were performed using simulated event samples based on two detector configurations. The 300 fb^{-1} samples were generated with a simulation of the Phase I detector with 50 pileup interactions [5–7], while for the HL-LHC, a Phase II baseline detector configuration with a pileup of 140 were used.

Monte Carlo simulations samples were produced with these configurations based on GEANT detector simulation, from which the object efficiencies and resolutions were determined and implemented in the DELPHES 3.0.10 fast simulation program [254], which was used for both signal and background production. DELPHES is able to include pileup interactions from inelastic proton-proton interactions simulated with PYTHIA6 [89]. Using MADGRAPH5 [237], samples of 10 to 100 million events per background process were produced. The events include up to four extra partons from initial- and final-state radiation, matched to PYTHIA6 for fragmentation and hadronization. The background cross sections were normalized to next-to-leading-order (NLO) calculations, which is based on the work in preparation for the Snowmass summer study 2013 and discussed in more detail in Refs. [255–257].

The systematic uncertainties assigned in these studies are based partly on those achieved in current 8 TeV analyses. However, the uncertainties are adjusted to reflect their dominant contributions and how these are expected to evolve as the selection procedures are adjusted for higher energy running and higher integrated luminosities. A fuller discussion of these assumptions is presented in Ref. [244].

The following sections describe the main properties of the full-spectrum SUSY models. All of the models are constructed such that their calculated dark matter abundances are at or below the observed value from WMAP. Two of the models, STC and STOC, involve dark matter coannihilation scenarios in which a second SUSY particle is nearly degenerate in mass with the $\widetilde{\chi}_1^0$, which helps to keep the dark matter abundance sufficiently low. More complete information, including the mass spectra, cross sections, and branching fractions, is given in Ref. [244].

10.3.2.1 The natural SUSY models (NM1, NM2, NM3)

The natural SUSY models NM1, NM2, and NM3 typify some of the most important scenarios for future searches. The strongly interacting sectors of NM1 and NM2 are nearly identical to that of NM3 (shown in Figure 10.19). In each model, $m(\widetilde{g}) \approx 1.7\,\text{TeV}$, $m(\widetilde{t}_1) \approx 1.1\,\text{TeV}$, $m(\widetilde{t}_2) \approx 1.9\,\text{TeV}$, and $m(\widetilde{b}_1) \approx 1.2\,\text{TeV}$, whereas $\widetilde{b}_2$ and the first and second generation squarks are much heavier, with masses around 3 TeV.

In each of these models, the gluino pair-production cross section at $\sqrt{s} = 14\,\text{TeV}$ is $\sigma(\widetilde{g}\widetilde{g}) = 5.4$ fb, the largest of any strong process. The cross sections for direct pair production of $\widetilde{t}_1\widetilde{t}_1^*$ pairs are in the range $\sigma(\widetilde{t}_1\widetilde{t}_1^*) \approx 3 - 4$ fb, while $\sigma(\widetilde{b}_1\widetilde{b}_1^*) \approx 3$ fb. Gluino-squark pair production is significant, $\sigma(\widetilde{g}\widetilde{q}) \approx 2$ fb in spite of the fact that the first and second generation quarks have large masses, $m(\widetilde{q}) \approx 3\,\text{TeV}$.

Because gluinos decay strongly and the daughter squarks have similar masses across these models, the gluino branching fractions are also nearly identical in NM1, NM2, and NM3. The decays $\widetilde{g} \to \widetilde{t}_1\bar{t}$ and $\widetilde{g} \to \widetilde{b}_1\bar{b}$ dominate, with branching fractions of 60% and 40%, respectively (conjugate modes are implied). The subsequent decays of $\widetilde{t}_1$ and $\widetilde{b}_1$ lead to final states with four b-quarks, highlighting the importance of b-jet tagging. (The STOC model is constructed to provide an interesting alternative, in which $\widetilde{t}_1 \to c\widetilde{\chi}_1^0$ with nearly 100% branching fraction.)

The electroweak and leptonic SUSY sectors of the natural models differ, however, with dramatic consequences for the $\widetilde{t}_1$ and $\widetilde{b}_1$ decay branching fractions, as well as for the decay patterns

of colorless particles. These effects arise from different patterns of masses in the electroweak and lepton sectors, as well as from differences in the gauge and higgs content of the neutralinos and charginos. For example, the decay $\widetilde{t}_1 \to t\widetilde{\chi}_1^0$ almost never occurs in NM1 and NM2, but this mode has a 40% branching fraction in NM3. These differences can be traced to the electroweak sector. For example, in NM1 the LSP has mass $m(\widetilde{\chi}_1^0) \approx 420\,\text{GeV}$ and is Higgsino-like, but in NM2 and NM3, it has mass $m(\widetilde{\chi}_1^0) \approx 200\,\text{GeV}$ and is Bino-like.

Specific features of NM1, NM2, and NM3 that arise from their different electroweak sectors include the following:

- In NM1, the sleptons are light, with $m(\widetilde{\ell}_L) \approx m(\widetilde{\tau}_1) \approx 430\,\text{GeV}$, and furthermore $m(\widetilde{\chi}_1^0) < m(\widetilde{\ell}_L) < m(\widetilde{\chi}_2^0)$. (In NM2 and NM3, the sleptons are effectively decoupled.) This pattern of neutralino and slepton masses leads leads to the $m_{\ell^+\ell^-}$ dilepton edge signature originating from the cascade process $\widetilde{\chi}_2^0 \to \widetilde{\ell}_L^{\pm}\ell^{\mp}$, $\widetilde{\ell}_L^{\pm} \to \ell^{\pm}\widetilde{\chi}_1^0$. (The direct decays $\widetilde{\chi}_2^0 \to Z/H\widetilde{\chi}_1^0$ are suppressed by a small mass splitting between the two neutralinos.)
- In NM2, both $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^{\pm}$ are Wino-like, with masses around 530 GeV, leading to a large $\widetilde{\chi}_2^0\widetilde{\chi}_1^{\pm}$ pair production cross section. For the $\widetilde{\chi}_2^0$, the decay $\widetilde{\chi}_2^0 \to H\widetilde{\chi}_1^0$ is dominant, while $\widetilde{\chi}_1^{\pm} \to W^{\pm}\widetilde{\chi}_1^0$ has a 100% branching fraction. These decays provide a powerful signature that can be used to search for $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0$ production.
- In, NM3 the LSP $\widetilde{\chi}_1^0$ is higgsino-like with a mass of around 200 GeV, and the dominant $\widetilde{t}_1$ decays are $\widetilde{t}_1 \to t\widetilde{\chi}_1^0$ and $\widetilde{t}_1 \to t\widetilde{\chi}_2^0$. This leads to a spectacular signature for gluino pair production with four top quarks and large $\not{E}_T$.

These consequences of these production and decay patterns for searches are discussed in Sec. 10.3.3.1 and more fully in Ref. [244].

10.3.2.2 Stau coannihilation model (STC)

In the stau coannihilation model [258], all of the sleptons and sneutrinos are light, and the $\widetilde{\tau}_1$ and $\widetilde{\chi}_1^0$ masses are nearly degenerate, with $m(\widetilde{\tau}_1) = 194\,\text{GeV}$ and $m(\widetilde{\chi}_1^0) = 187\,\text{GeV}$). The mass degeneracy of the $\widetilde{\tau}_1$ and the $\widetilde{\chi}_1^0$ allows efficient co-annihilation of dark matter to lower the predicted relic density to its observed value.

The gluino is heavy ($m(\widetilde{g}) \approx 3\,\text{TeV}$), suppressing its production; this is the only benchmark model in which the gluino production cross section is effectively negligible. However, the $\widetilde{t}_1$ and $\widetilde{b}_1$ are relatively light, with $m(\widetilde{t}_1) \approx 880\,\text{GeV}$ and $m(\widetilde{b}_1) \approx 1\,\text{TeV}$). These masses lead to significant direct pair-production cross sections: $\sigma(\widetilde{t}_1\widetilde{t}_1^*) \approx 19$ fb and $\sigma(\widetilde{b}_1\widetilde{b}_1^*) \approx 8.3$ fb. (Antiparticles of SUSY partners are denoted here with an asterisk, which does not refer to an off-shell particle.) The decay $\widetilde{b}_1 \to b\widetilde{\chi}_1^0$ has a large branching fraction ($\approx 70\%$) and provides a key signature in an all-hadronic search. As a consequence of $\widetilde{t}_1$ decays to the low-mass electroweak sector, the single-lepton $\widetilde{t}_1$ search and the trilepton +b-jet search both provide good sensitivity.

The STC model lies in the parameter space that yielded the highest likelihood in fits to all pre-LHC experimental data within the constrained MSSM (cMSSM) [258]. These fits preferred scenarios with a small mass difference, about 10 GeV, between the $\widetilde{\tau}$-NLSP and the $\widetilde{\chi}_1^0$ as LSP. Within the context of the cMSSM, this region is ruled out by LHC searches based on the strongly interacting sector, which in the cMSSM is coupled to the electroweak sector by constraints arising from GUT-scale mass unification. However, without these cMSSM-specific constraints, the part of the spectrum that is closely coupled in the fit to electroweak and flavor precision observables and to dark matter is not in conflict with existing LHC results.

10.3.2.3 Stop coannhilation model (STOC)

The stop coannihilation [259, 260] model STOC is also formulated in the cMSSM parameter space. In this model, $m(\widetilde{t}_1) \approx 400\,\text{GeV}$ is very low, and the top squark is nearly degenerate with the $\widetilde{\chi}_1^0$, which is bino-like. As a consequence, the direct top squark pair production cross section is enormous, $\sigma(\widetilde{t}\widetilde{t}^*) \approx 2.1$ pb. The top squark decays are effectively invisible, however, because they proceed via the loop process $\widetilde{t} \to c\widetilde{\chi}_1^0$, in which the daughter charm jet is extremely soft due to the small mass splitting between $\widetilde{t}_1$ and $\widetilde{\chi}_1^0$. Nevertheless, if the $\widetilde{t}_1$-pair system is boosted against a hard jet from initial-state radiation, the process is experimentally accessible in the single jet + $\not\!\!E_T$ signature, as in monojet searches [261].

In the STOC model, the gluino is the second lightest strongly interacting particle, with $m(\widetilde{g}) \approx 2.1\,\text{TeV}$. The gluino pair-production cross section is $\sigma(\widetilde{g}\widetilde{g}) \approx 0.5$ fb, about one-tenth that in NM1–NM3. The squark-gluino cross section is comparable, $\sigma(\widetilde{q}\widetilde{g}) \approx 0.3$ fb, and leads to a significant event yield in the all-hadronic H_T–H_T^{miss} search.

Although the gluino pair-production cross section is large enough to be experimentally visible, the decay signatures are quite distinct from those in the natural SUSY models. This may be surprising, given that the gluino decays with 50% branching fractions each to $\widetilde{t}_1\bar{t}$ and $\widetilde{b}_1\bar{b}$, which are comparable to those in the natural models. As we have seen, however, $\widetilde{t}_1$ decays are essentially invisible in STOC, so gluino pair production with $\widetilde{g} \to \widetilde{t}\bar{t}$ simply yields two top quarks plus large $\not\!\!E_T$. In summary, for STOC, the discoveries of signals in the monojet and $t\bar{t}$ +$\not\!\!E_T$ searches, combined with the absence of new physics in the $4b + \not\!\!E_T$ signature, could suggest a scenario of this type.

10.3.3 Results from sensitivity studies based on full-spectrum models

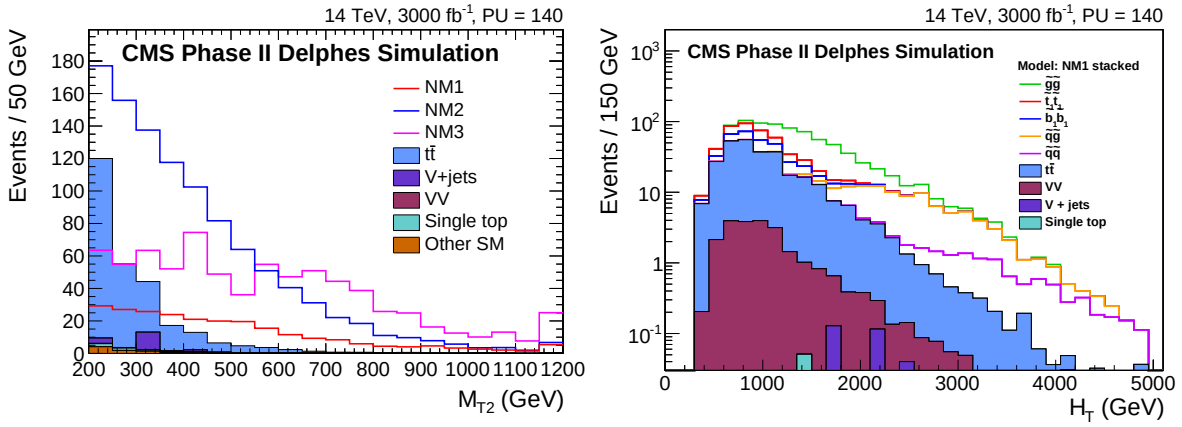
Figure 10.23 summarizes the sensitivities of the nine search methods to the benchmark models studied in each case. The results are color coded according to the level of signal significance observed using simulated data samples of 300 fb^{-1} and 3000 fb^{-1}. Because none of the signatures involves a mass peak, determining the exact origin of an excess will be a major challenge, and the pattern of results across different search channels will provide key information. Kinematic distributions will also be valuable for characterizing the processes responsible for any signals.

10.3.3.1 Results for the SUSY models NM1, NM2, and NM3

Figure 10.23 shows that there are both strong similarities, as well as some distinctive features, in the pattern of expected search results for NM1, NM2, and NM3. As discussed in Section 10.3.2.1, these models have nearly identical strongly interacting sectors, including relatively light $\widetilde{g}$, $\widetilde{t}_1$, $\widetilde{t}_2$, and $\widetilde{b}_1$. However, these models differ significantly in the properties of their electroweak and leptonic sectors.

The all-hadronic M_{T2} analysis with a requirement of $\geq$ 3b-tags has strong discovery potential for these models. The M_{T2} variable (also referred to as stransverse mass) is described in the references [262, 263]. The signals arise primarily from gluino pair production, with $\widetilde{g} \to \widetilde{t}\bar{t}$ and $\widetilde{g} \to \widetilde{b}\bar{b}$, as discussed in Sec. 10.3.2.1. Figure 10.24(a) shows the distribution of the M_{T2} for each of the three natural SUSY models and for the SM backgrounds. The dominant SM background arises from $t\bar{t}$ production, but these events are concentrated at lower values of M_{T2}. Significant signals are observed in the higher M_{T2} regions for each of the scenarios, although the yields vary substantially over the natural models. The slopes of the distributions are also different and reflect the masses and decay patterns.

The resulting sensitivity for NM2 and NM3 already exceeds 5σ at 300 fb^{-1}, while that for NM1



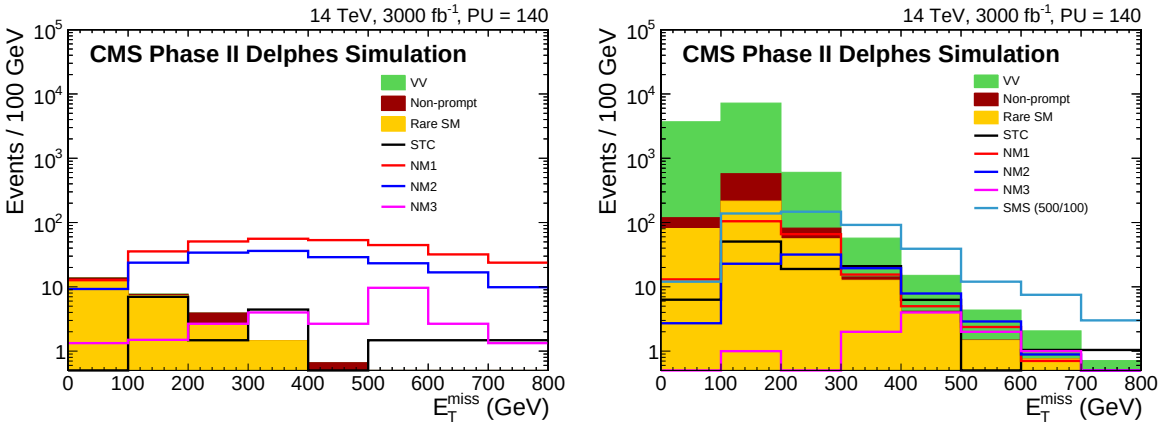
(a) (b)

Figure 10.24: Distributions from two searches with sensitivity to the NM1, NM2, and NM3 models. (a) All-hadronic search with M_{T2}: the distribution of M_{T2} after the full selection (including ≥ 3 b-tags) except the M_{T2} requirement itself. (b) Single-lepton search: distribution of H_T after the selection requirements are applied to all other variables. The contributions of the SM backgrounds are shown as stacked histograms in both (a) an (b), as they are elsewhere in this paper. In (a), the SUSY signal contributions from the different models are shown overlaid, but in (b) they are shown stacked because the histograms represent different processes within a single SUSY scenario, NM1.

is lower, requiring roughly 2000 fb^{-1} to reach this significance. The lower sensitivity to NM1 in the all-hadronic search arises from the properties of its electroweak sector, which produces more leptons. Because events with observed leptons are vetoed in an all-hadronic analysis, the average efficiency for events in this model is lower than that for NM2 and NM3. This feature illustrates the complementarity of different search signatures. The mass splittings in a given model also have strong effects on the sensitivity. In NM3, the lower $\widetilde{\chi}_2^0$ and $\widetilde{\chi}_1^\pm$ masses lead to a harder M_{T2} distribution. This distinctive feature boosts the discovery sensitivity for NM3.

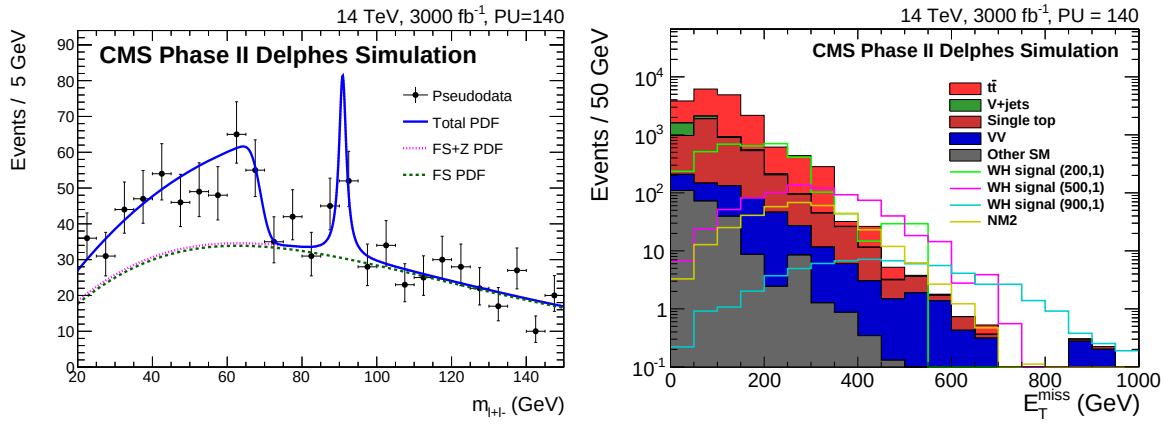
The single-lepton search also has strong sensitivity to each of the natural SUSY models. Figure 10.24(b) shows the distribution of H_T, the scalar sum of the jet p_T values, for the single-lepton search. Rather than showing the signal distributions for each of the natural SUSY models, only NM1 is shown, but the contributions from the various SUSY production processes within NM1 are displayed. For most of the H_T region, up to around 2 TeV, the signal yield is dominated by gluino-pair production, but at high H_T, above 2.5 TeV, gluino-squark production becomes prominent. The flavor of the produced squark is inherited from that of the incoming quark from the proton, so $\widetilde{u}$ and $\widetilde{d}$ dominate. It is remarkable that, with a 3000 fb^{-1} sample, this contribution is observable even though the masses of the $\widetilde{u}$- and $\widetilde{d}$-squarks are 3 TeV. In fact, the presence of contributions from multiple SUSY decay chains often complicates the extraction of a signal from a particular processes of interest. This search is nominally designed for direct $\widetilde{t}$ pair production, but Fig. 10.24(b) shows that other SUSY processes would dominate the observed yields in this scenario.

The trilepton + b-tag search also has good sensitivity to NM1, NM2, and NM3. The events are separated into two bins with b-tag multiplicity of either 2–3 or ≥ 4. The higher multiplicity bin is well suited to searching for gluino pair production in which each gluino decays into either a $\widetilde{t}\bar{t}$ or $\widetilde{b}\bar{b}$ pair. Figure 10.25(a) shows the distribution of $\not{E}_T$ in the trilepton + $\geq$ 4b event sample, which is well-suited for NM1 and NM2. In NM3, the b-jets are softer due to smaller mass



(a) (b)

Figure 10.25: Trilepton searches with and without b-jet tagging. (a) Distribution of $\not{E}_{\mathrm{T}}$ for the trileptons + b-jets search in the bin with≥ 4 b-tagged jets. (b) Distribution of $\not{E}_{\mathrm{T}}$ in the trilepton search for $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_1^0$, with $\widetilde{\chi}_1^{\pm} \to \mathrm{W}^{\pm}\widetilde{\chi}_1^0$ and $\widetilde{\chi}_2^0 \to \mathrm{Z}\widetilde{\chi}_1^0$. In this search, no b-tagging requirements are used. The selected events satisfy $200 < M_T < 400\,\mathrm{GeV}$ and have an $\ell^+\ell^-$ pair that reconstructs to the Z-boson mass.



(a) (b)

Figure 10.26: (a) Distribution of the dilepton invariant mass $m_{\ell^+\ell^-}$ in NM1, showing the distinctive kinematic edge associated with the decay chain $\widetilde{\chi}^{\pm} \to \widetilde{\ell}_L^{\pm}\ell^{\mp}$; $\widetilde{\ell}_L^{\pm} \to \ell^{\pm}\widetilde{\chi}_1^0$. (b) Search for $\widetilde{\chi}_2^0\widetilde{\chi}_1^{\pm} \to \mathrm{WH} + \not{E}_{\mathrm{T}}$: distribution of $\not{E}_{\mathrm{T}}$.

splittings, and the 2–3 b-jet multiplicity bin provides better sensitivity.

Figure 10.25(b) shows the distribution of $\not{E}_{\mathrm{T}}$ from a trilepton search in which events with b-jets are vetoed. This analysis is designed for processes such as $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0$, with $\widetilde{\chi}_1^{\pm} \to \mathrm{W}^{\pm}\widetilde{\chi}_1^0$ and $\widetilde{\chi}_2^0 \to \mathrm{Z}\widetilde{\chi}_1^0$, where the signature consists of leptons and $\not{E}_{\mathrm{T}}$ only. For generality, separate signal regions are defined for events with $\ell^+\ell^-$ pairs reconstructed on and off the Z-boson mass peak. Figure 10.23 shows that the trilepton analysis without b-jet tagging provides $\geq 3\sigma$ sensitivity for NM1, NM2, and STC at 3000 fb^{-1}.

Figure 10.23 also shows that the kinematic edge search is sensitive only to NM1, while the WH search is sensitive only to NM2. Thus the combination of the probes of the electroweak sector provide powerful information that distinguishes among NM1, NM2, and NM3. The $\widetilde{\chi}_2^0$

plays a key role in both of these signatures. The dilepton edge analysis targets a distinctive signature in the $m_{\ell^+\ell^-}$ distribution. This signature arises in NM1 because the slepton mass $m(\widetilde{\ell}_L) \approx 430\,\text{GeV}$ is below $m(\widetilde{\chi}_2^0) \approx 520\,\text{GeV}$ and above $m(\widetilde{\chi}_1^0) \approx 420\,\text{GeV}$, leading to the neutralino-slepton cascade decay chain $\widetilde{\chi}_2^0 \to \widetilde{\ell}_L^{\pm}\ell^{\mp}$, $\widetilde{\ell}_L^{\pm} \to \ell^{\pm}\widetilde{\chi}_1^0$. Figure 10.26(a) shows the distribution of $m_{\ell^+\ell^-}$ in the dilepton-edge analysis of NM1 for the signal sample (opposite-sign, same-flavor leptons), where a prominent triangle-shaped signal excess is evident, with its distinctive edge at the upper end. The location of the edge is determined by the $\widetilde{\chi}_{1,2}^0$ and $\widetilde{\ell}_{\text{L}}$ masses through the relation $m_{\text{edge}} = \sqrt{(m^2_{\widetilde{\chi}_2^0} - m^2_{\widetilde{\ell}})(m^2_{\widetilde{\ell}} - m^2_{\widetilde{\chi}_1^0})}/m_{\widetilde{\ell}}$. Most of the background can be estimated from an $\text{e}^{\pm}\mu^{\mp}$ control sample. Sensitivity to this critical feature of NM1 requires the full high-luminosity LHC data sample.

The NM2 model produces distinctive signals that arise from $\widetilde{\chi}_1^{\pm}\widetilde{\chi}_2^0$ production. In this model, $B(\widetilde{\chi}_1^{\pm} \to \text{W}^{\pm}\widetilde{\chi}_1^0) \approx 100\%$, $B(\widetilde{\chi}_2^0 \to \text{H}\widetilde{\chi}_1^0) \approx 90\%$, and $B(\widetilde{\chi}_2^0 \to \text{Z}\widetilde{\chi}_1^0) \approx 10\%$, leading to low-multiplicity signatures with leptons, $\not\!\!E_{\text{T}}$, and, in the case of $\text{H} \to \text{b}\bar{\text{b}}$, a pair of b jets with an invariant mass $m(\text{b}\bar{\text{b}}) \approx m_H$. The origin of these signals is the large splitting between $m(\widetilde{\chi}_2^0) \approx m(\widetilde{\chi}_1^{\pm}) \approx 530\,\text{GeV}$ and $m(\widetilde{\chi}_1^0) \approx 200\,\text{GeV}$, allowing $\widetilde{\chi}_1^{\pm} \to \text{W}^{\pm}\widetilde{\chi}_1^0$ and $\widetilde{\chi}_1^0 \to \text{H}\widetilde{\chi}_1^0$. Figure 10.26(b) shows the distribution of $\not\!\!E_{\text{T}}$ in the search using $\text{W}^{\pm} \to \ell^{\pm}\nu$ and $\text{H} \to \text{b}\bar{\text{b}}$. With $3000\,\text{fb}^{-1}$, this search is expected to observe a signal for NM2 with 3–5σ significance, as shown in Figure 10.23.

For processes such as $\widetilde{\chi}_2^0\widetilde{\chi}_1^{\pm}$, where there are plausible scenarios in which only a small number of particles play a role, it is useful to interpret the results more generically using a simplified model spectrum. The relevant simplified model processes have already been shown in Fig. 10.20(b). Besides the SM background and the signal for NM2, Fig. 10.26(b) shows the signal distributions for three simplified model scenarios. The models involve three SUSY particles but are characterized by only two parameters: $m(\widetilde{\chi}_2^0) = m(\widetilde{\chi}_1^{\pm})$ and $m(\widetilde{\chi}_1^0)$. The production cross section decreases with increasing $m(\widetilde{\chi}_2^0) = m(\widetilde{\chi}_1^{\pm})$, but for $m(\widetilde{\chi}_1^0)$ fixed essentially at zero, the extent of the high $\not\!\!E_{\text{T}}$ tail increases. Because there are only two independent model parameters, one can scan over the parameter space and present the results as a region in the plane of $m(\widetilde{\chi}_1^0)$ vs. $m(\widetilde{\chi}_2^0) = m(\widetilde{\chi}_1^{\pm})$. Figure 10.27 shows the boundaries of the 5σ discovery regions for data samples of $300\,\text{fb}^{-1}$ and $3000\,\text{fb}^{-1}$ under three different assumptions: $\mathcal{B}(\widetilde{\chi}_2^0 \to \text{Z}\widetilde{\chi}_1^0) = 100\%$, $\mathcal{B}(\widetilde{\chi}_2^0 \to \text{H}\widetilde{\chi}_1^0) = 100\%$, and 50% for each branching fraction. (These scenarios do not span the possible space in a full-spectrum model scenario. Both the $\widetilde{\chi}_1^{\pm}$ and the $\widetilde{\chi}_2^0$ can in principle decay via additional modes.) The exclusion region from the 2012 analysis is also shown. The new data will greatly expand the discovery region for this key signal process.

10.3.3.2 Results for the stau coannihlation model STC

As discussed in Sec. 10.3.2.2, the stau coannihilation model is characterized by a light slepton sector, with the $\widetilde{\tau}_1$ degenerate in mass with the $\widetilde{\chi}_1^0$; a relatively light third-generation squark sector; and a heavy, essentially decoupled gluino. As a consequence, the STC model has a very different profile of strong-production cross sections from the natural models, where gluino pair production plays a major role. In STC, the dominant strong-interaction processes are $\widetilde{\text{t}}_1\widetilde{\text{t}}_1^*$ and $\widetilde{\text{b}}_1\widetilde{\text{b}}_1^*$ pair production, with cross sections of $19\,\text{fb}^{-1}$ and $8.3\,\text{fb}^{-1}$, respectively. Figure 10.23 shows that the all-hadronic search with H_T and $H_{\text{T}}^{\text{miss}}$, the all-hadronic $\widetilde{\text{b}}_1$ search, the single-lepton search, the trilepton search, and the trileptons + b-tag search all provide sensitivity to this scenario.

Figure 10.28 shows distributions of the boost-corrected contransverse mass [264, 265] M_{CT} from the the all-hadronic search for $\widetilde{\text{b}}_1\widetilde{\text{b}}_1^*$ pair production, which focuses on the signature with 2b-

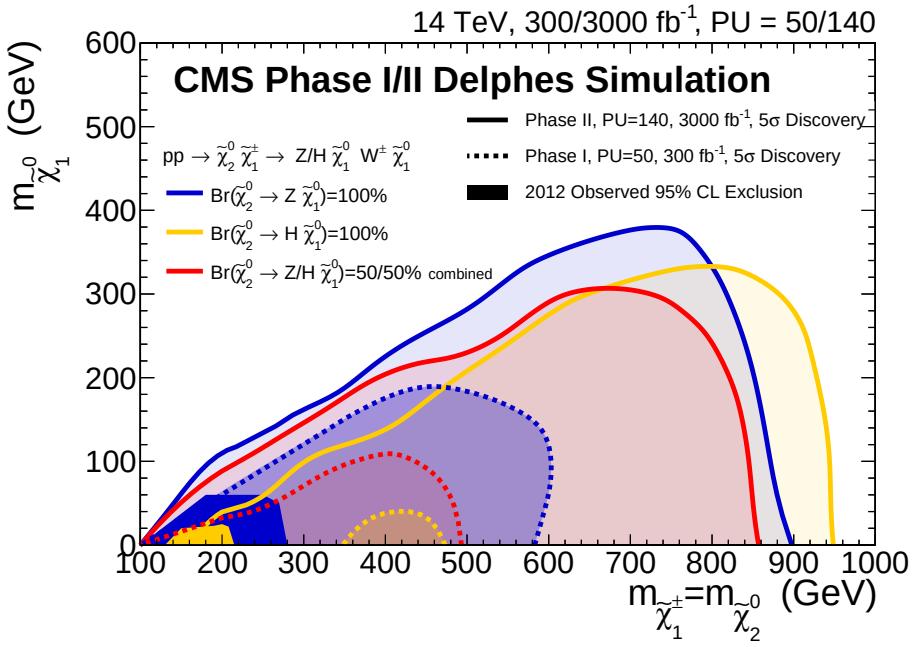
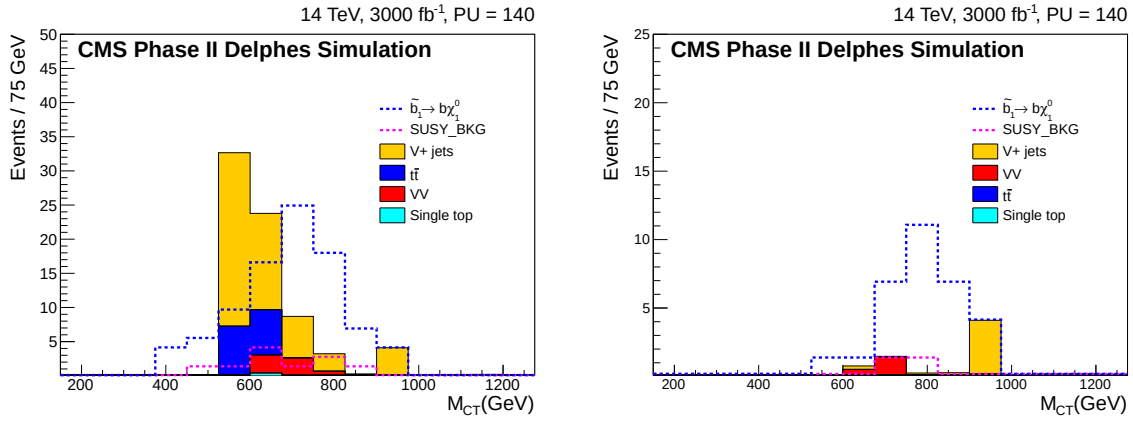


Figure 10.27: Search for $\widetilde{\chi}_2^0\widetilde{\chi}_1^\pm$ production in the $\mathrm{W}^\pm\mathrm{Z} + \not\!\!E_\mathrm{T}$ and $\mathrm{W}^\pm\mathrm{H} + \not\!\!E_\mathrm{T}$ final state. The excluded regions are shown in the simplified model parameter space of $m(\widetilde{\chi}_1^0)$ vs. $m(\widetilde{\chi}_1^\pm) = m(\widetilde{\chi}_2^0)$ for various assumptions. In such plots, the mass of the produced particle (or particles) is generally shown on the x-axis, while the mass of the LSP is shown on the y-axis. As a consequence, the excluded region is bounded by the decreasing production cross section on the right, but by the decreasing $\not\!\!E_\mathrm{T}$ as one approaches the diagonal.



(a) (b)

Figure 10.28: All-hadronic $\widetilde{\mathrm{b}}_1$ search. Distributions of the M_{CT} variable for (a) $M_T > 750\,\mathrm{GeV}$ and (b) $M_T > 950\,\mathrm{GeV}$. The endpoint of the M_{CT} distribution is a function of $m(\widetilde{\mathrm{b}}_1)$ and $m(\widetilde{\chi}_1^0)$.

tagged jets + $\not\!\!E_\mathrm{T}$. This variable is designed for the analysis of events in which two heavy particles decay into a jet + $\not\!\!E_\mathrm{T}$. The M_{CT} distribution for signal events has an endpoint at the value $M_{\mathrm{CT}}^{\mathrm{max}} \approx (m^2(\widetilde{\mathrm{b}}_1) - m^2(\widetilde{\chi}_1^0))/m(\widetilde{\mathrm{b}}_1)$ [264, 265]. Figure 10.28(a) shows the M_{CT} distribution for $M_T > 750\,\mathrm{GeV}$, while Fig. 10.28(b) shows the distribution for $M_T > 950\,\mathrm{GeV}$. The tighter M_T requirement provides additional suppression of the SM background, but the position of the endpoint is essentially unchanged.

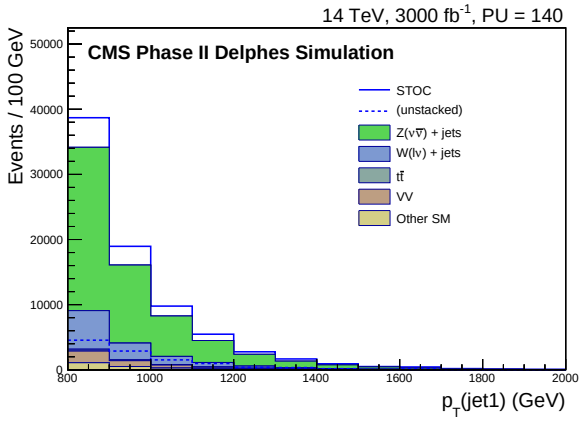
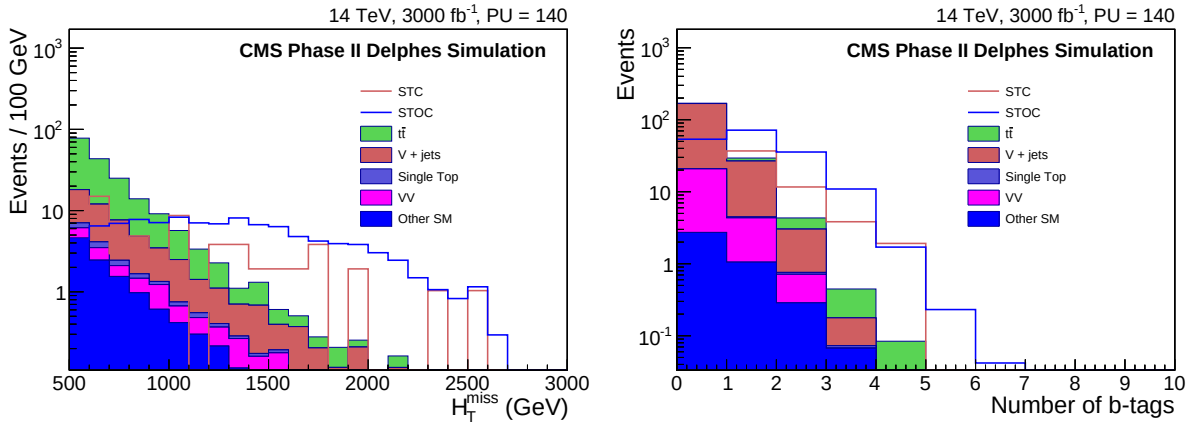


Figure 10.29: Monojet-like search in STOC: distribution of the p_T of the leading jet in the event.



(a) (b)

Figure 10.30: All-hadronic H_T–H_T^{miss} search. (a) Distribution of H_T^{miss} for the SM backgrounds and the STC and STOC signal contributions. (b) Distribution of the number of b-tagged jets after all other selection requirements.

Figure 10.25(a) would seem to indicate that the trileptons + b-jets search does not have useful sensitivity to the STC model. However, the events in this figure are restricted to the $N(\mathrm{b}) \geq 4$ multiplicity bin. For the $N(\mathrm{b})$=2–3 bin, the search reaches 5σ sensitivity at an integrated luminosity around 500 fb^{-1}. The absence of a signal in the four b-tagged jet bin indicates that the gluino is heavy, so that gluino pair production (with $\tilde{\mathrm{g}} \to \tilde{\mathrm{t}}\bar{\mathrm{t}}$) is suppressed.

10.3.3.3 Results for the stop coannihilation model (STOC)

The key features of the top-squark coannihilation (STOC) model were described in Section 10.3.2.3. Figure 10.23 shows that the all-hadronic search with H_T and H_T^{miss}, the single-lepton search, and the monojet search all provide sensitivity to this scenario.

As noted in Sec. 10.3.2.3, the very low top squark mass in STOC leads to a large cross section for direct top-squark production. Because the $\tilde{\mathrm{t}}$ and $\tilde{\chi}_1^0$ are nearly mass degenerate, however, the $\tilde{\mathrm{t}}_1$ is effectively invisible, and the process is not accessible experimentally without the presence of an additional jet from initial-state radiation. Figure 10.29 shows the distribution of the leading jet p_T from the analysis of the monojet-like final state, where an excess above the SM background is evident. The analysis must determine the very large, irreducible background

contribution from $Z(\rightarrow \nu\bar{\nu})$ + 1 jet. The significance of this signal approaches 5σ with 3000 fb^{-1}.

The observation of a large signal cross section in the monojet-like channel is indicative of the pair production of a strongly interacting particle that is nearly degenerate in mass with the LSP. Other analyses would help to understand this result. Figure 10.30 shows that the all-hadronic H_T-H_T^{miss} search would also observe a signal (arising from gluino pair production, with $\tilde{g} \rightarrow \tilde{t}\bar{t}$), but the excess would be associated with two b-tagged jets, not four b-tagged jets, as is normally the case with gluino pair production followed by $\tilde{g} \rightarrow \tilde{t}_1\bar{t}$ or $\tilde{g} \rightarrow \tilde{b}_1\bar{b}$. Since $m_{\tilde{b}}$ is large, the search for the $b\bar{b} + \not{E}_T$ final state would not observe an excess, but the single-lepton + $\geq$ 2b-tagged jets + $\not{E}_T$ analysis would observe one from gluino pair production with $\tilde{g} \rightarrow \tilde{t}\bar{t}$. Together, these observations would suggest the production of a light $\tilde{t}_1$ nearly degenerate with the LSP.

10.3.4 Conclusions

The discovery of a Higgs boson at 125 GeV has given new urgency to the question of how the electroweak scale is stabilized against short-distance quantum corrections that must inevitably arise in the SM. Addressing this question is one of the major goals of the LHC program for the next decade and beyond. The solution can in principle be either accidental or natural. Supersymmetry provides one, but by no means the only, approach to resolving the gauge hierarchy problem.

If a SUSY spectrum of any kind emerges from the LHC program, we will have finally broken through to beyond-the-SM physics in accelerator experiments. The studies presented here show that a broad range of searches will provide a way to discover and characterize what may be a very complex spectrum. In the absence of reconstructed mass peaks, the pattern of such results provides essential information for characterizing the different sectors of such a spectrum. Natural SUSY models are expected to produce signals in channels with gluinos, third-generation squarks, and higgsinos. Because the decays of the third-generation squarks would be sensitive to the arrangement of the electroweak and slepton sectors, the observed pattern of signals and their individual kinematic distributions will provide crucial information for identifying the underlying spectrum.

Several conclusions arise from these studies.

- The breadth of the SUSY search program implies that all of the capabilities of the CMS detector are required.
- For some of the measurements, the performance of the degraded detector will severely compromise the program.
- For natural SUSY models, which are designed to address the gauge hierarchy problem, discovery in one or more of the favorable search signatures could occur within the first 300 fb^{-1} of data taking. Sensitivity to other key signatures, however, can require 33000 fb^{-1}.
- Because a given search channel provides only limited information, interpreting any observed excesses in terms of a specific particle spectrum will be a major challenge.
- To map out the properties of a particle spectrum, it is essential to have the full pattern of results that will be obtained at the highest integrated luminosities.

The program of SUSY measurements will provide critical information needed to address the issues of dark matter, the unification of forces, and the gauge hierarchy program. In particular, the processes and signatures that high luminosity LHC running will allow us to probe are crucial for excluding or discovering natural SUSY models. And beyond even those major issues,

the SUSY program will allow us to address the even deeper question of whether nature utilizes supersymmetry in the fundmanental laws of physics.

10.4 Exotica searches and measurements

Alternative new physics scenarios must be explored in the absence of signals of supersymmetry in data. All these possible extensions are conventionally classified under the label of "Exotica". Some of them are driven by the same naturalness or hierarchy problems that motivate SUSY, like the search for effects from the presence of extra dimensions. Other extensions are driven by experimental indications of deviations from the SM, like the search for dark matter, or the search for new particles at the TeV scale related with the mechanism that gives masses to light neutrinos. Finally, Exotica searches also cover generic deviations that are motivated by completions of the SM postulated at very high energy scales that are not directly accessible by the LHC, but with possible implications at the TeV scale. Examples are the search for new gauge bosons, leptoquarks, technicolor, or new contact interactions.

Although the sensitivity to new physics scales increases more significantly with an increase in the center-of-mass energy of the collisions, the effect of an increase in integrated luminosity from 300 fb^{-1} to 3000 fb^{-1} is also substantial, as illustrated in Fig. 10.31. Besides a quantifiable increase in the mass reach, the improved statistical sensitivity can help to establish an observation in a region where only a tiny indication of an excess was previously present. On should note that many of the proposed extensions of the SM are just benchmark models, with branching fractions or simplifications that may be optimistic at the end of the day. In addition, larger data sets allows for a better understanding of many backgrounds in the search.

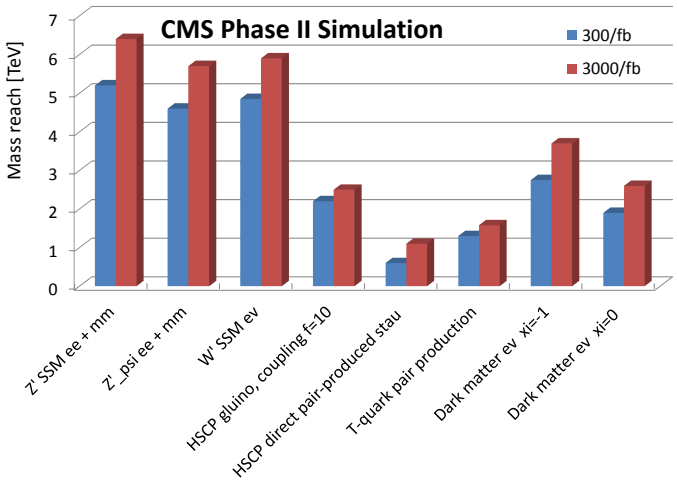


Figure 10.31: Projected performance of selected searches for new physics. Except for dark matter, the 5σ discovery reach in terms of particle mass is shown. For dark matter the EFT cut-off scale, Λ, is the relevant parameter which can be interpreted as a mediator mass in the limit where EFT is applicable.

In order to profit from the order of magnitude increase in the collected data set, a good performing detector is required. In case of an already existing signal, enhanced acceptance capabilities in the forward region will help to better disentangle new physics models. Improved granularity and detector resolution will also be required for the very high instantaneous luminosities expected in Phase-II. Finally, there are some exotic searches that demand special trigger and detection capabilities, like being sensitive to highly ionizing particles, displaced vertices or very low-momentum tracks. We evaluate the relevance of the proposed modifications to the CMS detector using several examples:

- the search for new massive neutral narrow resonances,
- the search for Dark Matter in final states containing single jets or W bosons accompanied by missing energy,
- heavy stable charged particles and displaced signatures,
- triggering very soft leptons.

The search for high-mass narrow resonances at high luminosity is illustrated with a study that shows the capability to identify the parity properties of of a new spin-1 resonance (Z') using studies of the forward-backward asymmetry in dielectron final states. We studied the impact of an increased acceptance and an improved charge determination in the forward regions of the detector. The dark matter analysis illustrates the capabilities of the detector to provide a precise measurement of the missing transverse energy in a harsh environment, with more than hundred simultaneous events on top of the hard scattering interaction. The search for heavy stable charged particles and displaced vertex signatures targets scenarios that demand unique detector capabilities. Finally, very compressed supersymmetric spectra become unusually exotic scenarios. They can be identified by the presence of soft leptons in intermediate decays, which require a very performing tracking system at the lowest trigger level.

10.4.1 Properties of a dielectron resonance

Searches for dilepton resonances at the highest accessible masses require excellent detector performance in terms of resolution and coverage. For instance, CMS charge identification capabilities are expected to improve for the Phase-II design, particularly in the forward regions. This has important implications for the measurement of the forward-backward asymmetry of a new dilepton resonance, which requires an unambiguous determination of the direction of the negatively charged lepton. Fig. 10.32-left shows the charge misidentification probability for the case of a 3.6 TeV dielectron resonance as a function of the lepton pseudorapidity. A significant degradation is observed for the case an aged detector scenario, leading to big acceptance losses in regions of maximal sensitivity. The plot on the right of the same figure illustrates the typical benefits of an extended acceptance. About 10-20% of the selected events have at least one electron in the endcap region, the effect being more significant at lower invariant masses.

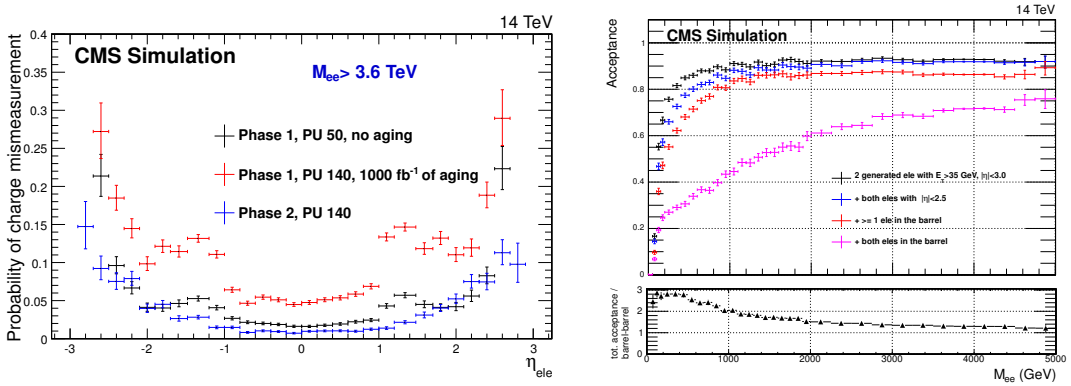


Figure 10.32: Left: Electron charge misidentification probability. Right: Acceptance for different detection scenarios as a function of the dielectron invariant mass.

One of the cleanest ways to observe physics beyond the SM at the LHC would be an alteration of the dilepton mass spectrum at high masses. This could be a narrow resonance, as predicted for instance by GUT Z' theories [266] or the Randall Sundrum (RS) model [267] with a small

wraped extra dimension. While many models considered would manifest themselves as a sharp resonance at very high dilepton invariant masses, its physical width could be relatively large, and the exact nature of such a resonance could only be revealed with high luminosity by studying the differential angular distributions of the final state leptons. These distributions are sensitive to the production mechanism and to the spin-parity properties of the new resonance. A spin 0 particle can be produced through gg or $q\bar{q}$ fusion and yields a flat $\cos\theta$ distribution. In case of a colorless spin 1 (which can only be produced through $q\bar{q}$ fusion) or spin 2 (produced either via gg or $q\bar{q}$ fusion) resonance, the coefficients of the terms proportional to an even power of $\cos\theta$ in $d\sigma/d\cos\theta$ are fixed for a given spin and production mode. A number of signal models were studied for spin states of 0, 1, or 2. With the help of a likelihood ratio potential signal hypotheses can be separated with sufficient statistics. As illustrated in Fig. 10.33, with an accumulated statistics of 300 fb^{-1} the two cases of two spin-1 resonance shown in the legend cannot be separated, while the 3000 fb^{-1} of Phase-II do allow to separate spin-1 hypotheses with more than three sigma. Also spin-0 can be separated while spin-2 would need even more luminosity.

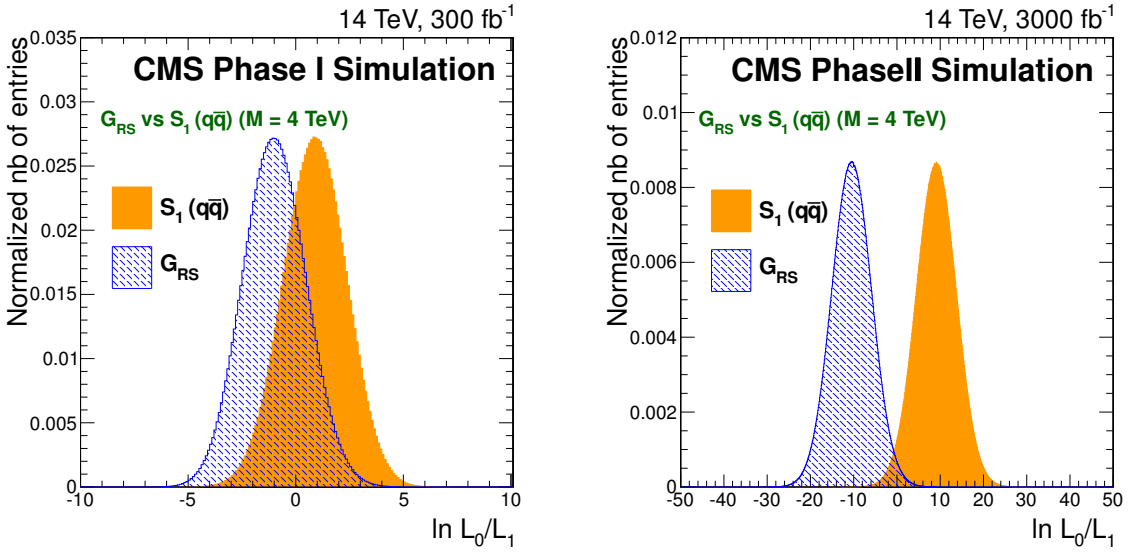


Figure 10.33: Likelihood ratio between two hypotheses (RS graviton, G_{RS}, and a spin 1, S_1, resonance) for a signal corresponding to a G_{RS} at M = 4 TeV and for an integrated luminosity of 300 fb^{-1} and 3000 fb^{-1}.

Another way of searching for new physics is a precise measurement of the forward-backward asymmetry,A_{FB}, of the SM Drell-Yan spectrum rather than selecting a model in particular. Here we study the precision that can be reached on the full dielectron invariant mass spectrum including lower masses (between 500 GeV and 2 TeV) where the effects of the extended acceptance and improved charge identification (see Fig. 10.32) could become sizable. The forward-backward asymmetry measurement is particularly sensitive to detector performance, as it requires the most accurate identification of the lepton charges and directions in the resonance decay and of the direction of the incoming quark (as opposed to the antiquark). The $\cos\theta$ distribution in the center of mass system of the hard process is experimentally approximated by the measured Collins-Soper angle $\cos\theta_{CS_{meas}}$, which is an adequate choice, minimally sensitive to higher order QCD corrections. The lepton charge assignment plays a critical role, and only events with oppositely charged electrons are accepted in the selection. As discussed before, the suggested modifications for Phase-II will help to reduce the charge misidentification probabil-

ity, particularly in the endcap regions. Let us note that the asymmetry is maximal in the very forward regions (see Fig.10.34), which are the detector components undergoing substantial upgrades for the HL-LHC. The coverage will also increase from $|\eta|$ <2.5 in Phase-I to $|\eta|$ <3.0 in Phase-II, which has a sizable impact on the acceptance for low-mass resonances. The statistical precision that could be reached on the Drell-Yan A_{FB} measurement after having collected the equivalent of 3000 fb^{-1} of data is represented in Fig.10.34. In the first bins, the uncertainty is around 40% larger for the Phase-I aged detector with respect to the Phase-II detector. As expected, the relative improvement decreases when the dielectron mass increases.

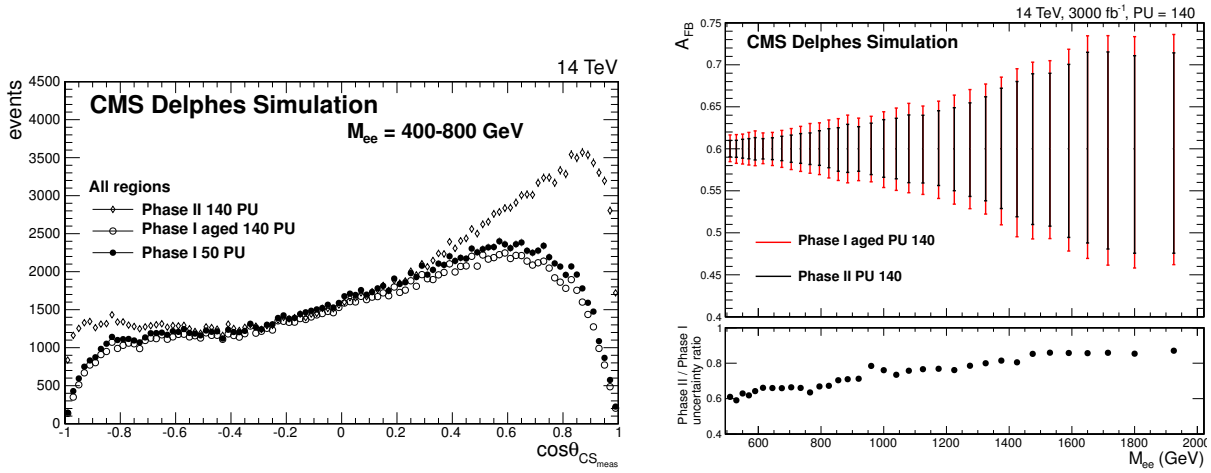


Figure 10.34: Left: Distribution of $\cos\theta_{CS_{meas}}$ for low mass resonances for the three detection configurations. Right: Precision of A_{FB} measurement for Drell-Yan events as a function of the dielectron invariant mass for the Phase-I aged and the Phase-II scenarios.

10.4.2 Dark matter

Astrophysical evidence for Dark Matter (DM) is one of the few existing hints for physics beyond the SM. The LHC can study many different types of possible interactions between quarks and DM. The LHC operating in high luminosity mode will deeply probe search regions with very low DM masses ($\lesssim$ 10 GeV) and regions of phase-space where direct-detection experiments are blind because of their irreducible background from neutrino-nucleus scattering.

To illustrate the physics sensitivity in Phase-II, we consider the case where a W boson recoils against a pair of DM particles, which, following the W boson decay, can yield an event containing an electron plus missing transverse energy. Events containing a single electron with p_T >100 GeV (following the Run-I single-electron trigger threshold) are selected if the ratio of the electron p_T and missing transverse energy is in the range 0.4< p_T/E_T^{miss} <1.5 and the angle $\Delta\phi$ between both objects is larger than 2.5. There is an implicit M_T cut due to the lepton p_T requirement. The discriminating variable is the transverse mass, M_T, calculated from the electron E_T and missing transverse energy which is caused by the neutrino.

Fig. 10.35 and Table 10.5 show the consequences of an aging Phase-I detector in final states with one electron and missing transverse energy: the losses in efficiency become unacceptable for lepton momenta up to about 1 TeV. For DM this is particularly relevant since low masses always play a role as illustrated with the M_T spectrum in Fig. 10.35-right. The DELPHES reconstruction efficiency has been compared to full simulation samples using a number of simulated W′ signals, which cover a comparable M_T range, demonstrating good agreement to DELPHES.The average signal efficiency, which affects the part of the M_T spectrum above the

ξ	Phase-I	Phase-I"aged"	Phase-II
-1	59.8%	31.8%	59.3%
0	57.5%	30.4%	56.8%
+1	16.0%	5.7%	16.2%

Table 10.5: Signal efficiencies determined for an EFT DM sample (electron channel) with M_{DM}=10 GeV determined by DELPHES simulation for the three detector scenarios.

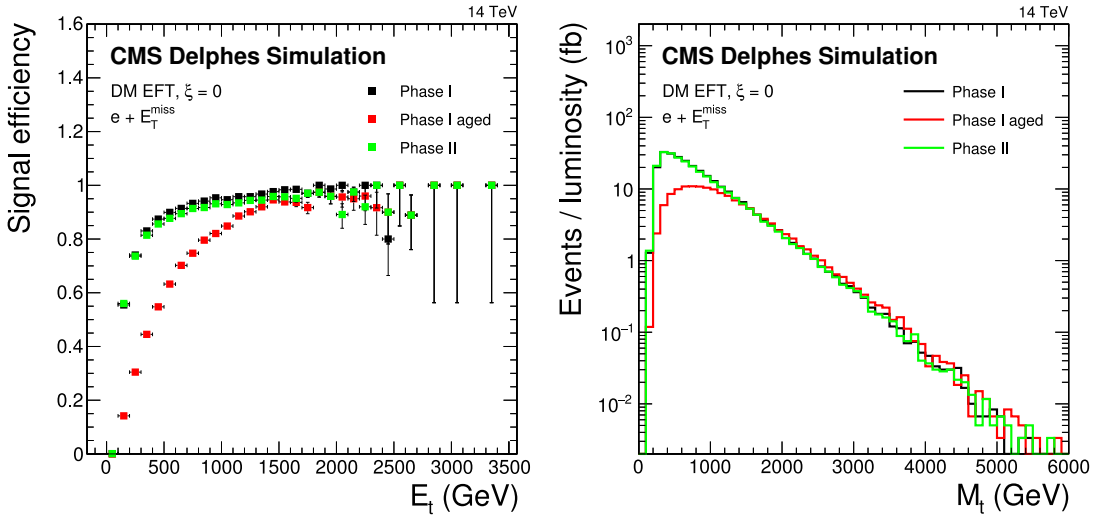


Figure 10.35: Signal efficiency (left) and transverse mass spectrum for an electron and missing transverse energy final state for the three detector scenarios. The signal efficiency is shown as M_T of the charged lepton, while the M_T spectrum is shown for the discriminating variable of M_T.

previously mentioned threshold, drops by roughly half for the Phase-I aged detector. Performances of the Phase-I and Phase-II detectors are comparable. A DM signal would create a small excess of events over the SM expectation distributed nearly equally in M_T unlike e.g. a W′ signal which rather yields a Jacobian peak at high masses. The main background, around 60%, is due to off-shell W boson production requiring a good understanding of higher order corrections. Other contributing backgrounds are DY, diboson (WZ, ZZ, WW) and $t\bar{t}$ events where all but one lepton escape detection. The analysis follows the strategy from Run 1 [268] without further optimization for HL-LHC conditions as the signal sensitivities, shown in Table 10.5 for the electron channel studied with DELPHES, are quite comparable to the ones observed in Run 1.

As a special feature, this channel is sensitive to different couplings of DM to up- and down-type quarks - parametrized by ξ - and may provide an explanation for inconsistent results from DD experiments. The production can be modeled with an effective field approach (EFT) where no specific assumptions apply to the mediator other than it being very heavy (labeled Λ), or with a simplified model where the mediator mass and its couplings (and therefore its width) are specified. For high mediator masses as will be probed at the HL-LHC, both models yield comparable description and sensitivity[269]. The strong dependence of the M_T spectrum and the signal efficiency on the value of ξ is due to the significantly different M_T shape.

Discovery would reach EFT scales (or mediator masses) of Λ >3.7 for ξ=-1 and Λ >2.6 for ξ=+1

(see Fig. 10.31). Fig. 10.36 illustrates the situation for the monolepton channel in terms of 95% C.L. exclusion limits using interpretations in the framework of a simplified model. The mediator is modeled as a Z′-like particle with either vector or axial-vector coupling, a fixed minimal width of $\Gamma_{med} = M_{med}/8\pi$ and the mass range shown in Fig. 10.36. The monojet channel, which is often used for comparison (see ref [270] for projection) relies on an initial state radiation jet for tagging the event, which, along with missing transverse energy due to the produced DM-pair, leads to a final state of jet + E_T^{miss}. This channel is expected to profit from the improved jet performance discussed in Chapter 9. Given its kinematics it cannot be sensitive to interference but its sensitivity is comparable to the monolepton channel with constructive interference (ξ=-1). The projected performance of the monojet channel is based on Run-I performance projected to larger accumulated statistics [270]. The monolepton sensitivity in Fig. 10.36 is again based on DELPHES simulation and illustrates the situation for the maximum sensitivity corresponding to constructive interference (ξ=-1) and the narrow width indicated in the legend. The hadronic W boson channel should reach even further but has not be studied in this framework.

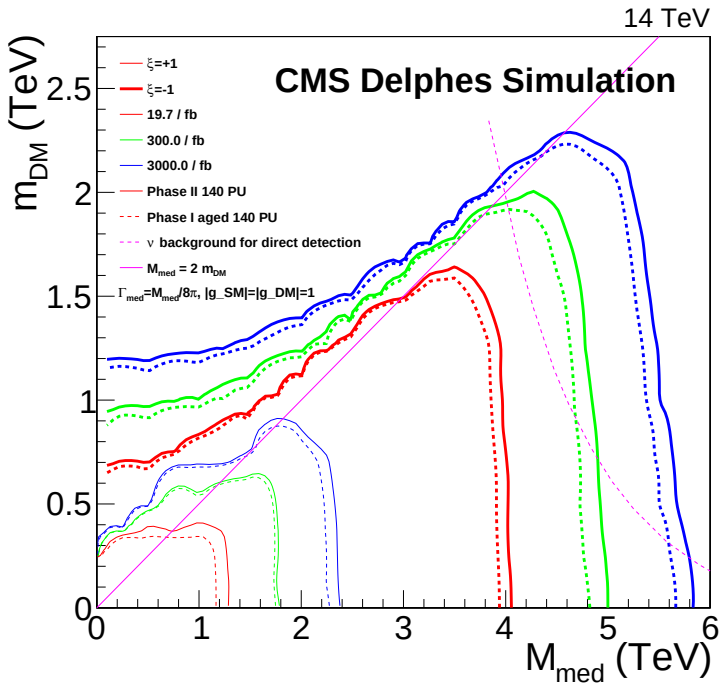


Figure 10.36: Dark matter reach of the monolepton channel as a function of the DM mass and mediator mass for the two extreme cases of $\xi = \pm 1$.

10.4.3 Exotic signatures: highly ionizing particles and displaced vertices

A large number of new physics searches rely on very exotic signatures, such as anomalous ionization depositions, slow moving particles, (dis)appearing tracks, and secondary vertices significantly displaced with respect to the primary interaction vertex. Many of these signatures become quite natural in new physics scenarios where some of the new massive particles become stable or long-lived. Searching for them demands specific detector capabilities, special triggers or adapted reconstruction algorithms, which implies that usually they are uncovered by other searches using standard objects.

One example is the production of heavy stable charged particles, such as supersymmetric taus (staus), moving with moderate velocities β <1 (in our example β=0.8). Such particles are likely

to be missed by standard SUSY searches, due to the absence of special time-of-flight or ionization treatments in usual trigger and reconstruction algorithms. One signature facilitating their detection is an anomalous dE/dx deposition in the pixel and tracker layers, as expected from charged particles in the non-ultrarrelativistic regime. This search serves several theoretical models [271]. A second option is a specialized analysis of the time evolution (TOF) of the signal deposits in the detector. Fig. 10.37-left shows the projected cross section limits for 3000 fb^{-1} for pair-produced staus in comparison with the Phase-I (300 fb^{-1}) sensitivity for a) a pessimistic scenario with missing dE/dx information (only time-of-flight information, blue curve) and b) a scenario where four pixel plus three tracker layers provide the necessary dE/dx information.

A variety of theoretical models predict the existence of long-lived particles that decay within the CMS detector volume to final states that include leptons, photons, or jets. If the cross section for such processes is small, or if the lifetime of the long-lived particles is large, such that only a small fraction decay within the detector, then discovery may only be made at HL-LHC. Exotic signatures may derive from the decay a resonance, such as a (non-SM) Higgs boson, that can decay to a pair of neutral, long-lived bosons X, each decay to a pair of leptons. Such a signature is typical of 'hidden valley' models. The experimental search simply requires the presence of two leptons that are consistent with originating from a displaced secondary vertex. A variant of this model, in which the long-lived particle is highly Lorentz boosted, and so decays to leptons that are almost collinear. Other signal models yielding displaced leptons, which have already been studied in CMS and for which similar considerations would apply, include R-parity violating, long-lived neutralinos, each decaying to dileptons and a neutrino; or long-lived third-generation squarks, each decaying to a quark and a lepton.These searches are usually background-free, and so lead to limits on the signal cross section (for which no reliable theoretical predictions exist), that scale in inverse proportion to the luminosity times signal acceptance.

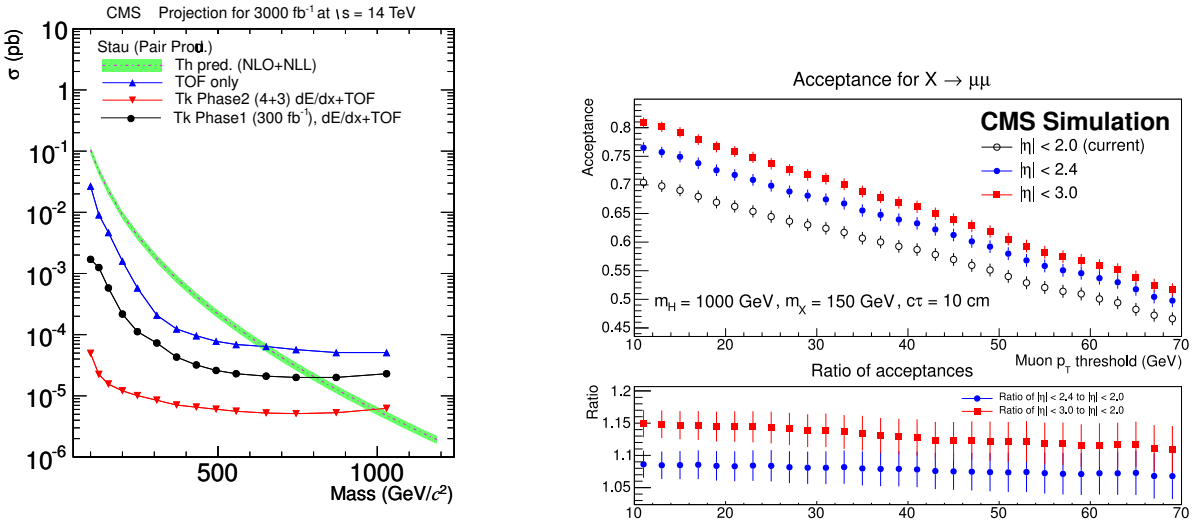


Figure 10.37: Left: Projected cross section reach in terms of 5 sigma discovery for 3000 fb^{-1} for pair-produced staus with the following detection scenarios: a) TOF-only (blue), b) 4+3 dE/dx+TOF (red) in comparison to Phase-I (black). Right: Acceptance for $X \to \mu^-\mu^+$ as a function of the requirement on the minimum allowed lepton p_T, for the $H \to XX$ model. The results are shown for three different requirements on the maximum permitted lepton rapidity.

There are no reliable theoretical predictions for the lifetime of the X boson. If the X bosons

decay before leaving the tracker volume, the displaced leptons can be reconstructed in the tracker provided that they pass through a sufficient number of tracker layers, which typically means that they must be produced at a transverse distance from the beam-line of $L_{xy} < 50$ cm. If the X bosons decay further out than this, it may still be possible to reconstruct displaced muons using stand-alone reconstruction in the muon chambers, though this possibility has not yet been studied in detail for HL-LHC. Triggering on displaced leptons is also non-trivial, with conventional lepton triggers expected to become inefficient for leptons that have a transverse impact parameter exceeding a few millimetres.

Figure 10.37-right shows the predicted 'acceptance' of the model in which long-lived X bosons decay to dileptons. The *acceptance* is defined as the fraction of $X \to \mu^+\mu^-$ decays in which the X boson decays at a transverse distance from the beam-line of $L_{xy} < 50$ cm and both leptons pass kinematic requirements on their rapidity ($|\eta|$) and p_{T}. High Higgs boson masses yield leptons with such high p_{T} that any difficulty in triggering on and reconstructing them can only be due to their non-prompt nature. Lower masses presents the additional challenge that the lepton p_{T} are lower, demanding that the trigger p_{T} thresholds must also be kept low. In both cases, triggering is only possible without a beamspot or vertex constraint. The acceptance is estimated with a generator-level study using events simulated with PYTHIA, in which the mean lifetime $c\tau$ of the X boson is chosen to correspond to a mean L_{xy} of 20 cm. The analysis is able to reconstruct X bosons that decay at L_{xy} values between a few hundred microns (such that the leptons are well separated from the beam-line) and approximately 50 cm (beyond which displaced track reconstruction is not possible).

Reconstruction requires a dedicated algorithm (no vertex constraint), usually limited to the tracker (no pixel detector) due to the large displacement. While in Run-I the analysis acceptance had to be limited to $\eta \leq 2.0$ to avoid the high track-misidentification rate in the forward region, it can be extended up to $\eta \leq 3.0$ in Phase-II. The gain in acceptance is shown in Fig. 10.37-right for the muon channel. Similar arguments apply to the electron channel, where in addition the improved charge identification capability of the Phase-II detector (see Fig. 10.32) will be beneficial.

10.4.4 Triggering with soft leptons

Some exotic scenarios involve the production of particles in models with highly compressed mass spectra. Typical examples are natural supersymmetric models where the lowest neutralino and chargino mass states are quasi-degenerate Higgsinos, with small mass splittings $m_{\chi_2^0} - m_{\chi_1^0}$, $m_{\chi_1^\pm} - m_{\chi_1^0}$, of order O(5-10 GeV) [272] for a wide range of $M_1, M_2 > |\mu|$ parameter space. Splittings of this size makes those events invisible to standard multi-lepton searches. A search strategy outlined in ref. [272] suggests to exploit the channel $pp \to j + E_T^{miss} + 2\ell$ where the jet is from initial state radiation, missing transverse energy from the stable, neutral SUSY particle (LSP) and the leptons from the decay of the next-to-LSP electroweakino via gauge bosons (W or Z).

Studies were performed for the scenario $M_1 = M_2 = 500$ GeV and $\mu = 110$ GeV and 150 GeV. The resulting final state yields a relatively hard jet, missing transverse energy and a very soft accompanying lepton with $p_{\mathrm{T}} \geq 1$ GeV.

Based on the L1-track-trigger menu from Chapter 3, a standard analysis with single object triggers would have the following offline p_T thresholds: 18 (31) GeV for a single muon (electron) trigger and 173 GeV for a single-jet trigger. These thresholds would result in efficiencies of 7% (muons), 4% (electrons), and 20% (jets), which do not allow to significantly increase the sensitivity beyond the LEP limit of $\sim$103 GeV [272].

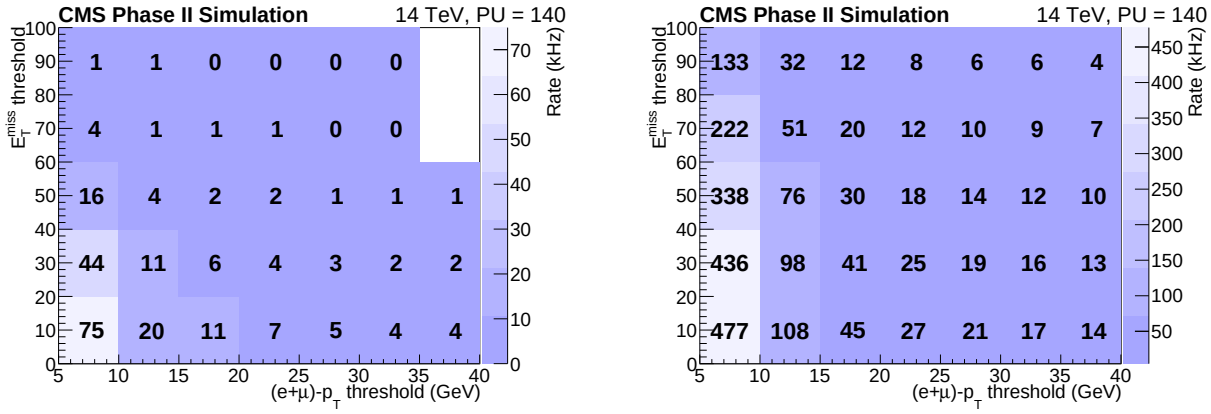


Figure 10.38: Rates with Phase-II track-trigger (left) and Phase-I trigger (right) for the combination of electron and muon channels with the same E_T^{miss} requirement.

A L1 cross-trigger based on three objects, namely a lepton, a jet, and missing transverse energy, was developed by exploiting the track trigger from Section 6.4 for the soft lepton. For this study it is assumed that such an exclusive $\ell + j + E_T^{miss}$ trigger can use a bandwidth of ≈10 kHz for a single channel, either with electrons and muons, and ≈ 20 kHz for the combination of both. One has to find a good compromise of the trigger rate and signal efficiency. For each trigger rate two thresholds, for the lepton and missing transverse energy, can be tuned. It appears that the lepton p_{T} is most sensible and should be kept as low as possible, certainly below 10 GeV. Only the Phase-II track trigger provides the needed sharp turn-on and purity. A trigger rate around 10 kHz can be met with the following thresholds: the lowest possible lepton p_{T} of 5 GeV being combined with E_T^{miss} ≥20 GeV in the muon and E_T^{miss} ≥40 GeV in the electron channel, yielding signal efficiencies of 25% (15%) for muons (electrons). In the electron channel the rates are slightly higher due to misreconstructed leptons.

For the combination of electron and muon channels $\ell + j + E_T^{miss}$, Fig. 10.38-left shows the trigger rates for the Phase-II detector as a function of the lepton p_{T} and the missing transverse energy thresholds.

The impact of the upgraded Phase-II tracker can be seen when comparing it to a Phase-I trigger in Fig. 10.38-right which are substantially higher. The much lower rates of the combined Phase-II $\ell + j + E_T^{miss}$ cross trigger allow to chose lower thresholds for each of the objects resulting in a higher signal efficiency. To achieve a comparable rate with Phase-I, the thresholds would have to be lepton p_{T} ≥20 GeV and E_T^{miss} ≥60 GeV.

10.5 B physics

The excellent B physics performance of the current CMS detector shows that it is possible to produce competitive results in this field also for a non-dedicated hadron collider experiment. The LHC luminosity upgrade will open new possibilities for B physics measurements, especially the study of rare processes like the $B^0_{(s)} \to \mu^+\mu^-$ decays. On the other hand, the high instantaneous luminosity and increased pileup will create demanding conditions for the trigger and offline reconstruction of B production and decay.

In this section, the B-physics potential of the upgraded CMS detector is illustrated by presenting a study of the $B^0 \to \mu^+\mu^-$ and $B_s \to \mu^+\mu^-$ decays. This work is based on the $\mathcal{B}(B_s \to \mu^+\mu^-)$ measurement published by the CMS collaboration with the LHC Run-I data [273], and focuses on the implementation of a prototype L1-trigger algorithm and on the estimate of the final analysis sensitivity.

The only way to build a L1 trigger for the B^0 and B_s signal at the HL-LHC is exploiting the track trigger architecture of the upgraded CMS detector (see Section 6.4.1), as the non-upgraded L1 trigger would bring unsustainable event rates at the low thresholds needed for the measurement, due to the inferior resolution and high pileup contamination.

Events are triggered by requiring two opposite-charge muons, reconstructed matching the track trigger track in the inner silicon detector with an outer track in the muon chambers. Each muon is required to have transverse momentum $p_{\mathrm{T}}^{\mu} > 3\,\mathrm{GeV}$ and pseudorapidity $|\eta^{\mu}| < 2$. In addition, the two muons are required to originate from the same primary vertex within 1 cm.

The Level 1 B candidate is built from the two muons and is required to have $p_{\mathrm{T}}^{\mu\mu} > 4\,\mathrm{GeV}$, $|\eta^{\mu\mu}| < 2$, and $3.9 < M^{\mu\mu} < 6.9\,\mathrm{GeV}$.

The invariant mass distributions of the Level 1 B candidates for $B^0 \to \mu^+\mu^-$ and $B_s \to \mu^+\mu^-$ signal events are shown in Fig. 10.39 for an integrated luminosity of 3000 fb^{-1}. The invariant mass resolution at L1 is about 70 MeV for both resonances.

The rate of the L1 trigger on Minimum Bias events is a few hundreds Hz. This rate is in addition to the dimuon L1 trigger rate shown in Table 6.1, and is equal to only a small fraction of it.

The offline signal reconstruction is performed by requiring two opposite-charge muons, each having $p_{\mathrm{T}}^{\mu} > 4\,\mathrm{GeV}$. The B candidate is built from the two muons and is required to have $p_{\mathrm{T}}^{\mu\mu} > 5\,\mathrm{GeV}$ and $4.9 < M^{\mu\mu} < 5.9\,\mathrm{GeV}$.

Three different scenarios have been considered: the 2019 Phase-I detector with no aging and with an aging corresponding to 1000 fb^{-1} of integrated luminosity, and the Phase-II upgraded detector. For the first two scenarios, both barrel and endcaps are taken into account ($|\eta^{\mu}| < 2.5$), while for the Phase-II scenario only barrel is considered ($|\eta^{\mu}| < 1.4$). All the resolutions do not show significant changes between the two aging conditions of the 2019 detector, thus only the latter is used in the following.

The invariant mass resolutions measured in the barrel, extracted from the widths of Gaussian fits to the mass peaks found with the full detector simulation, are 42 MeV for the Phase-I scenario and 28 MeV for the Phase-II scenario. The Phase-I resolutions are comparable to those of the Run-I measurement, while the Phase-II gives an improvement of a factor 1.5. In general, no significant differences between the resolutions of the two peaks are seen.

Figure 10.40 shows toy Monte Carlo projections of the B^0 and B_s analysis results for the Phase-I and Phase-II scenarios. The left plot corresponds to barrel events and a total integrated lu-

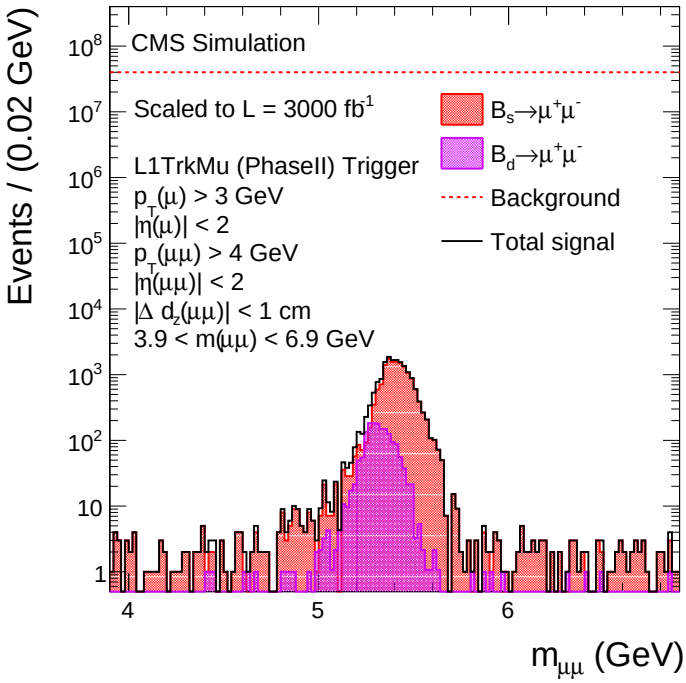


Figure 10.39: The dimuon invariant mass distributions at Level 1 trigger for $B^0 \to \mu^+\mu^-$ (purple) and $B_s \to \mu^+\mu^-$ (red) events for a total integrated luminosity of 3000 fb^{-1}. The red dashed line shows the expected background level.

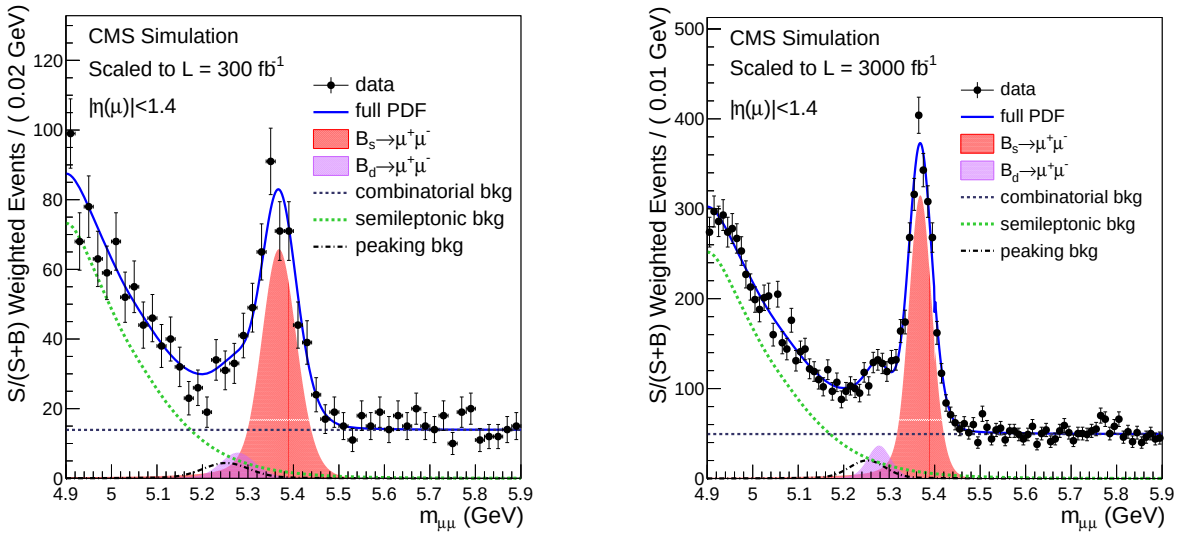


Figure 10.40: Projections of the mass fits to 300 fb^{-1} (left) and 3000 fb^{-1} (right) of integrated luminosity, respectively assuming the expected performances of Phase-I and Phase-II CMS detector. Both plots are for barrel only.

minosity of 300 fb^{-1}, while the right plot corresponds to barrel events and a total integrated luminosity of 3000 fb^{-1}. In both cases, the $\sigma \times \mathcal{B}$ predicted by the SM is assumed for B^0 and B_s. The background models are taken from the $\mathcal{B}(B_s \to \mu^+\mu^-)$ measurement published by CMS with the LHC Run-I data. Conservatively, a loss of efficiency equal to 35% for the signal and 30% for the backgrounds is assumed for the Phase-II projections with respect to the Run-I results. These results show that while in the 2019 scenario the B^0 peak is covered by the long resolution tail of the B_s resonance, in the 2023 scenario the two peaks can be resolved due to the improved invariant mass resolution.

The analysis performance is estimated for the Phase-I and Phase-II scenarios under the assumptions outlined above for a total integrated luminosity of 300 fb^{-1} and 3000 fb^{-1} respectively. The sensitivity for the observation of the $B^0 \to \mu^+\mu^-$ decay is expected to be $\approx 6.8\,\sigma$ ($\approx 2.2\,\sigma$), while the branching fractions $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_s \to \mu^+\mu^-)$ can be measured with a precision of 18% and 11% (48% and 13%) respectively for the Phase-II (Phase-I) scenario. Their ratio $\frac{\mathcal{B}(B^0\to\mu^+\mu^-)}{\mathcal{B}(B_s\to\mu^+\mu^-)}$ can be measured with a 21% (50%) uncertainty for the Phase-II (Phase-I) scenario.

Chapter 11

Infrastructure and Logistics of Work in USC

11.1 Introduction

This section describes the part of the intended CMS Phase-II upgrade programme covering common systems, infrastructure and CMS facilities on CERN sites, as well as common tasks such as technical coordination and central technical support for the overall Phase-II upgrade programme, which will be carried out between June 2015 and June 2025.

The individual particle and energy detection sub-systems deployed in the CMS experiment, as well as the data acquisition, trigger and offline systems, are the responsibility of sub-collaborations of institutes (in some cases including CERN), which are responsible for their construction, maintenance and upgrade. The magnet, experiment infrastructure, laboratory testing and storage facilities, overall technical coordination and central technical support needed to satisfactorily assemble and maintain the detector sub-systems and to allow them to function together as a coherent scientific apparatus, are provided either by CERN as the host laboratory or through CMS collaboration resources. Over the next decade, which is planned to include two operation periods and two long-shutdown periods, this provision will, as previously, cover "operation", "maintenance" and "upgrade", often conducted simultaneously. The common systems, infrastructure, facilities or projects are under the direct responsibility or oversight of CMS Technical Coordination. Resources for items or activities defined as CMS "common systems", "infrastructure", "common tasks" or "common facilities" come from several sources:

- Provision for host laboratory support or other selected general support made by CERN.
- CMS Maintenance and Operation "A" budgets (M & O A) to which all member institutes, including CERN, contribute.
- CMS Phase-I or Phase-II upgrade common project budgets (to which, again, all member institutes, including CERN, contribute).
- Detector specific M & O "B" or upgrade budgets of subsystems which may chose to pay for system-specific, centrally coordinated and/or executed work (M & O B is contributed by member institutes of a sub-collaboration and administered by the subsystem concerned).

The attribution of common items or activities to particular categories is subject to interpretation. The guidelines, as currently interpreted by CMS Technical Coordination, are briefly described in Appendix A. In many cases, the resources for a single activity or work-stream are logically split between the different categories. Technical Coordination, in unison with Resource Coordination, is responsible for coordinating the flow of resources to execute the common activities in all categories as a coherent, unified programme.

After a brief reference to the predicted evolution of the radiation environment induced by particle collisions, which has substantial consequences for the scope and timing of common Phase-II upgrades, the remainder of this chapter describes common items or tasks which are considered to be entirely, or partly, classifiable as Phase-II upgrade. Where known, the overlap with other categories will be noted.

11.2 Radiation Environment

The increased particle fluence, integrated absorbed dose and activation levels induced by collision products as LHC luminosity increases will present one of the biggest challenges to be addressed by the Phase 2 upgrade programme.

11.2.1 Monitoring equipment

Increased fluences and integrated rates directly affect some common equipment. Examples are position-monitoring sensors, cameras etc, which will have to be replaced more frequently, shielded better, or substituted with more radiation-tolerant equipment. The basic, online, beam-induced background monitors (BCM1L, BCM2 and BHM) and the fast timing monitors (BPTX), which are a common responsibility, though operated within the BRIL project, will have to be upgraded to deal with higher rates. Similarly, the active and passive area dosimetry systems will need to be adapted to a higher dynamic range.

11.2.2 Radiation Simulations

In order to produce and maintain global maps of the predicted fluences and integrated doses resulting from LHC operation, along with the evolution of activation levels and residual ambient equivalent dose rates during shutdowns, the common project will provide support (under the executive coordination of the BRIL project) to the CERN staff in HSE and CMS charged with conducting and benchmarking radiation simulations. The simulation team has to work closely with common and sub-detector projects in the optimization of the detailed Phase-II design and to adapt simulation results as the luminosity predictions and LHC operation schedule evolve. The implications of fluences and integrated absorbed dose for the design and operation of specific sub-detectors, are assumed to be handled by the sub-projects themselves. Oversight to maintain coherence amongst the many simulation tasks for central and sub-system use is assigned to the BRIL project.

Preliminary simulated maps of the residual ambient equivalent dose rate, H*(10) anticipated after 1 week cooling at the start of LS2, LS3 and LS4 are shown in Fig. 11.1 and Fig. 11.2.

The simulations were performed with Monte Carlo radiation transport code FLUKA[13, 14]. A CMS FLUKA model that represents the Run 2 configuration was used for all simulations including those for LS4. The maps in Fig. 11.1 and Fig. 11.2 represent estimates of dose rates for the closed configuration and demonstrate the effect of luminosity increase, rather than changes in material composition. A software tool, which is used with the FLUKA code, has recently been developed by the BRIL team to determine the dose rates for CMS open scenarios and this will be used for future simulations of access configurations. Dose rates after the first week of LS2 may be a factor of around 5 higher than those seen in the first week of LS1, due to luminosity increase and the fortuitously long cooling period due to the end-of-year lab closure and subsequent heavy ion running period which preceded LS1. The 20–25% difference between LS2 and LS3 reflects the relatively small change in the instantaneous luminosities foreseen to be typical just before each shutdown. For LS4 onwards, however, an increase by as much as

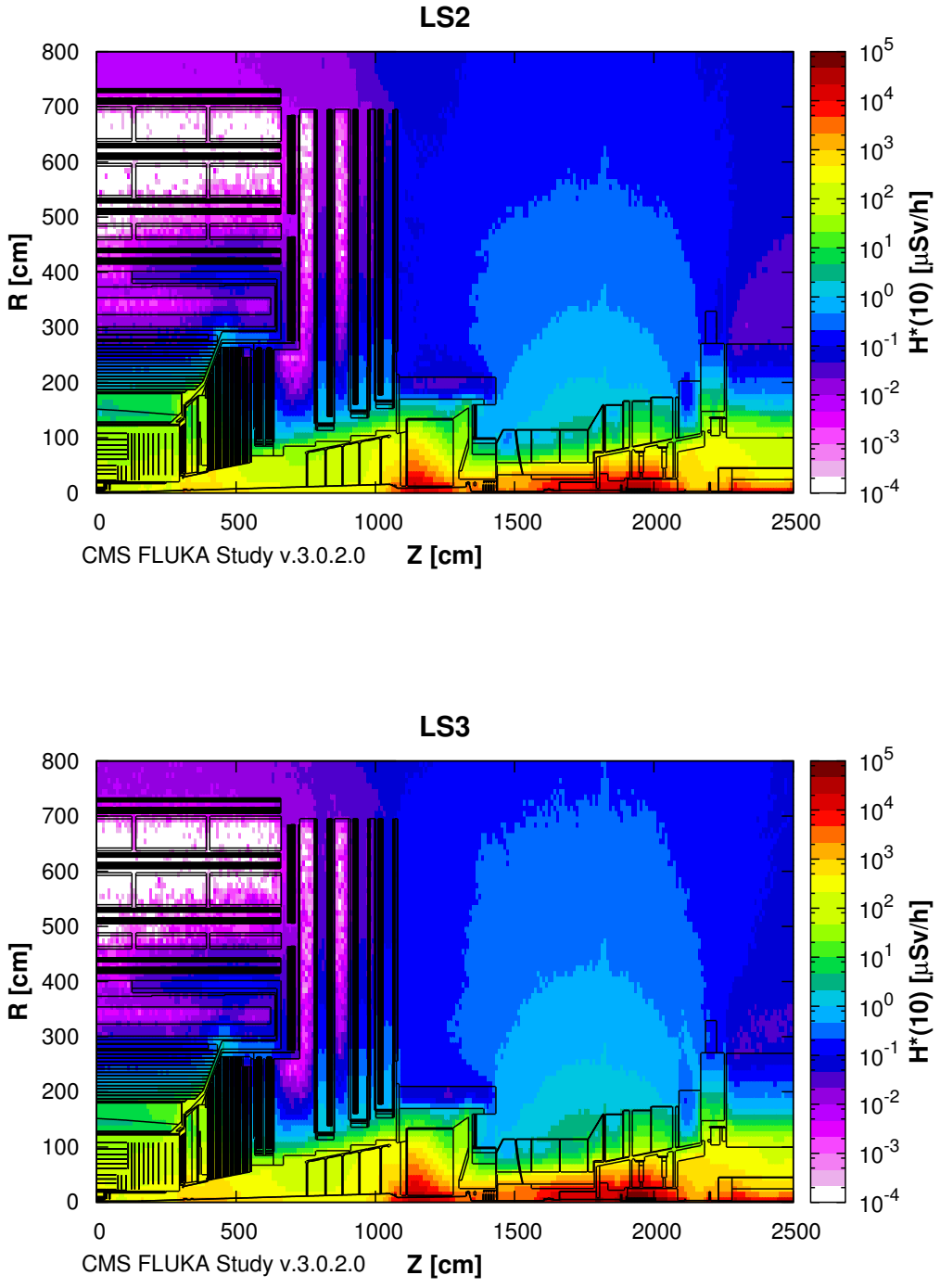


Figure 11.1: Simulated equivalent dose-rate H*(10) map for one quadrant of CMS after 1 week cooling in LS2 (top) and LS3 (bottom). A geometry model representing a Run 2 configuration was used for all simulations. (Plots courtesy of CERN DGS/RP).

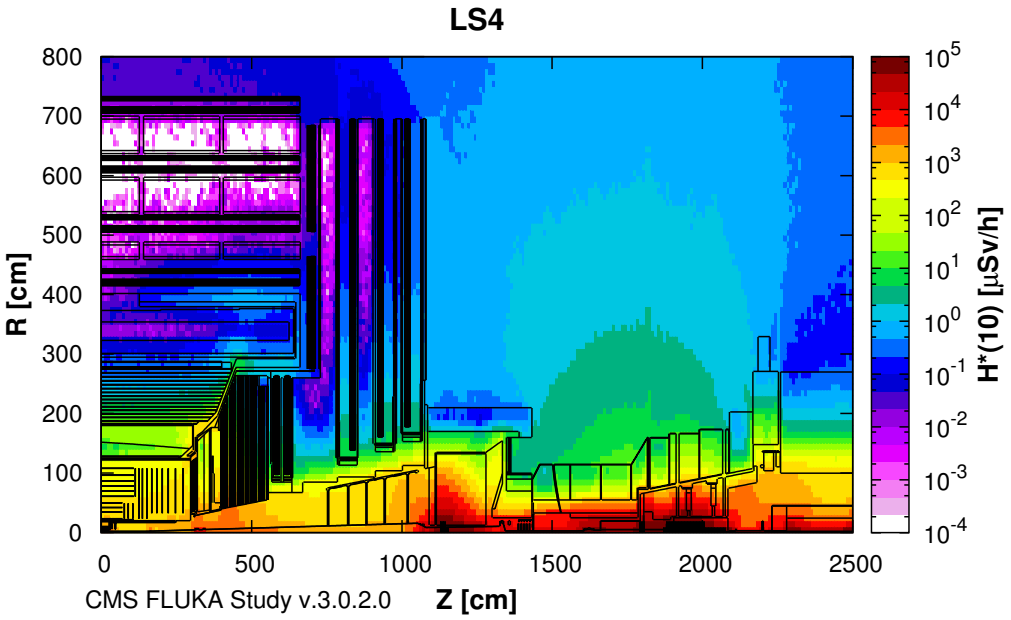


Figure 11.2: Simulated equivalent dose-rate H*(10) map for one quadrant of CMS after 1 week cooling in LS4. A geometry model representing a Run 2 configuration was used for all simulations. (Plots courtesy of CERN DGS/RP).

another factor 5 is anticipated. The effect of "cooling time" (used in the radioprotection context to mean: time to allow radioactive decay of short-lifetime isotopes resulting from activation) is demonstrated in Fig 11.3 and Fig 11.4, where predicted dose rates in the radial range 0-300 cm are plotted against z for different times following the end of high luminosity p-p operation, for LS2, LS3 and LS4.

The maxima due to the vacuum bellows and flanges plus the ECAL face at around 3.5m, the HF face at 11.5 m, and the TAS at 18-20m, stand out clearly. A cooling time of 3-4 months produces a (position-dependent) factor 3–5 reduction in dose rate in LS2, somewhat less in LS3, but substantially less in LS4, where the reduction factor is only 2-3, due to the build-up of long-lived isotopes.

11.2.3 Applicable Radiation Safety Policy

The radiation safety policy being applied by CMS, in line with the general objectives of ALARA dose targets (As Low As Reasonably Achievable) and fully consistent with CERNs radiation safety policy, demands a design for installation, operation and maintenance of CMS Phase-II where the anticipated annual individual doses received by personnel are comparable with the typical annual doses due to cosmic radiation and the local natural environment in the Geneva area. Thus the objective for the maximum annual individual dose for LS1 activities, CERN-wide was 3mSv, a target expected to be maintained, or reduced to 2mSv, for LS2 and LS3. In addition every effort has to be made to keep the dose to each individual involved in a particular work-package/work-stream $< 100\mu$Sv and the collective dose to the team involved $< 500\mu$Sv. Besides avoiding a lengthy internal scrutiny process, respecting these limits will help avoid

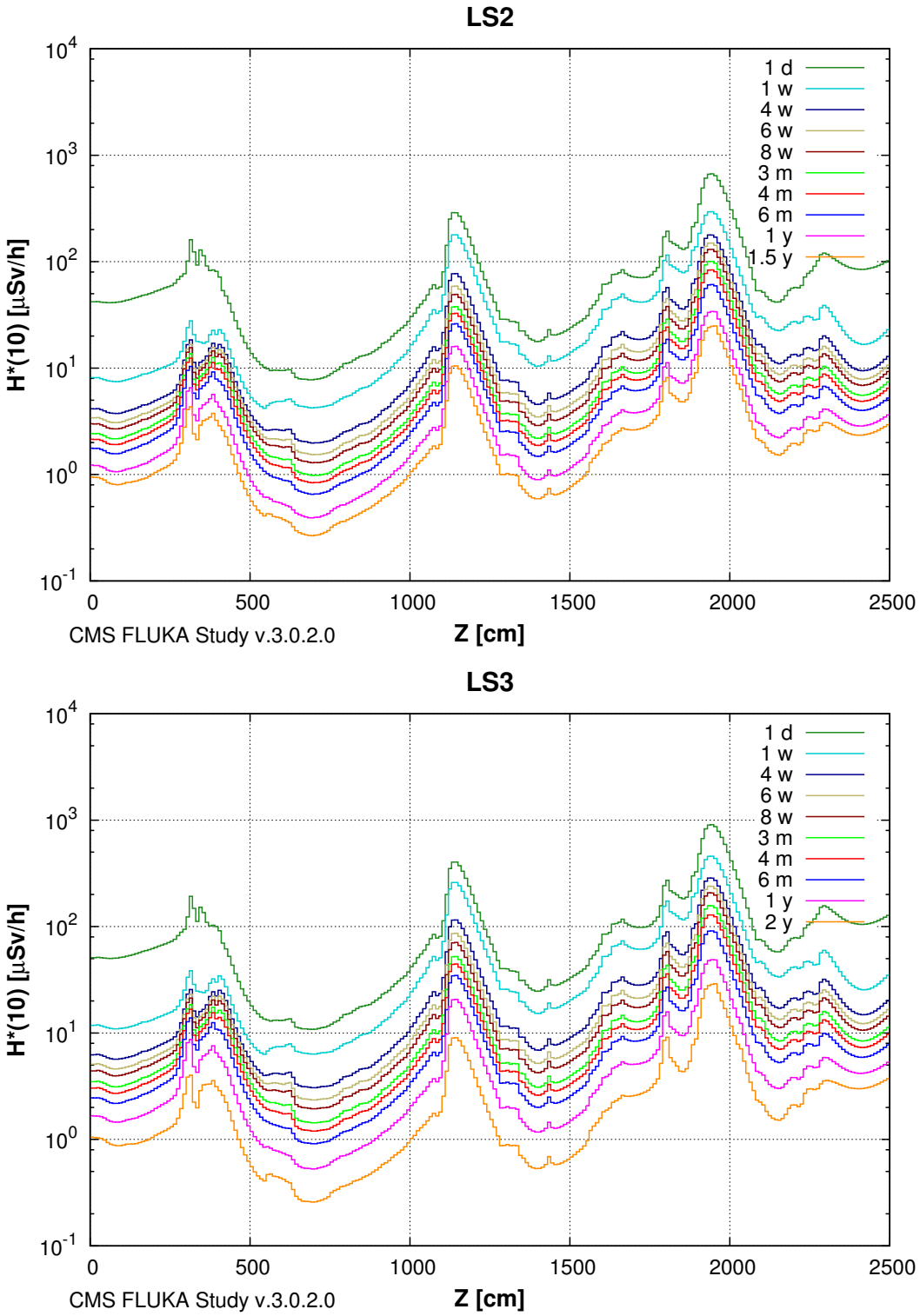


Figure 11.3: Simulated dose-rate H*(10) vs z in the range 0–300 cm, after times ranging from 1s to 1.5 years after the end of high luminosity p-p operation for LS2 (top) and LS3 (bottom). A geometry model representing a Run 2 configuration was used for all simulations (plots courtesy of CERN DGS/RP).

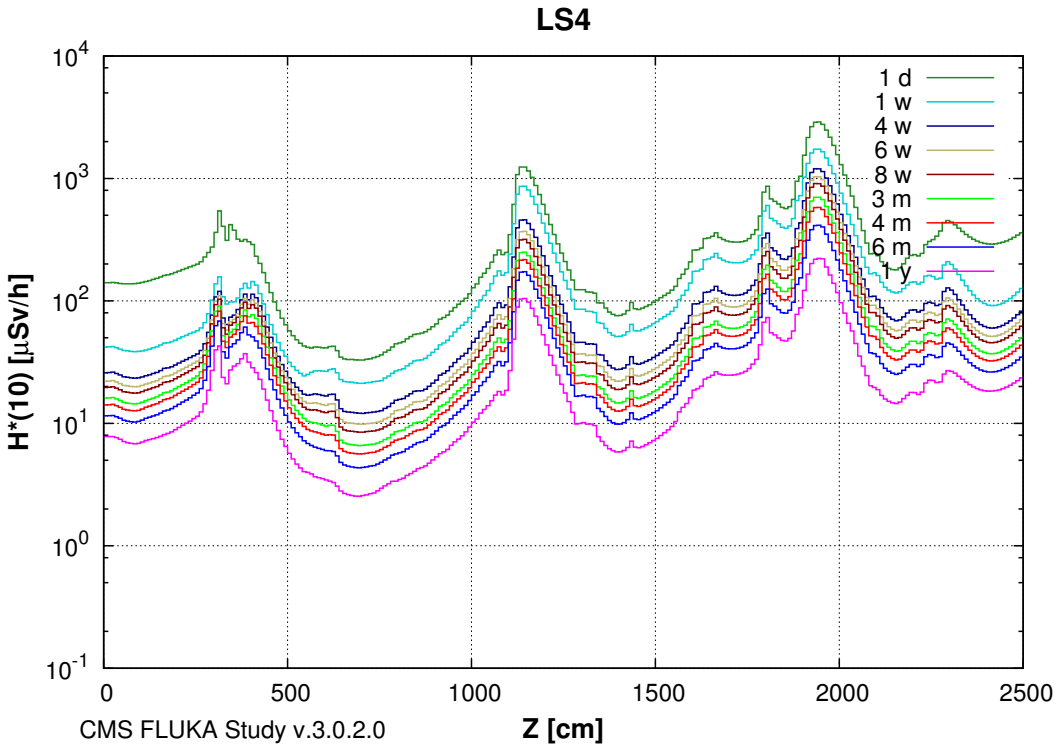


Figure 11.4: Simulated dose-rate H*(10) vs z in the range 0–300 cm, after times ranging from 1s to 1.5 years after the end of high luminosity p-p operation for LS4. A geometry model representing a Run 2 configuration was used for all simulations (plots courtesy of CERN DGS/RP).

dose limitations on the work of the few, key, multi-talented field personnel on whom the efficiency and smooth execution of the planned work programme chiefly depend. The main strategies that can be applied to minimise doses to personnel due to residual activation are: choice of materials, cooling time (allowing short-lived isotopes to decay), shielding and distancing workers from activated areas (remote handling plus work-site discipline). These strategies are mirrored in several foundation features of the common project. These are:

1. Programme to replace the steel beam vacuum chambers in the 3m-18m range by aluminium alloy chambers (described in Section 11.6) which will reduce overall activation one month into a shutdown by a factor 5 to 15.

2. Decision to make the above change in LS2, when longer cooling times (up to 4 months) can be accommodated, thus to take advantage of lower activation at LS3 start.

3. On-going implementation of temporary shielding systems integrated into maintenance configurations.

4. Construction of detailed models of critical zones for use in rehearsing procedures.

5. Investment in remote monitoring and handling development.

11.3 Phase 2 Upgrade Program

11.3.1 Assumptions

In the planning and program development, the following assumptions are used:

1. The year-end technical stop (YETS) in 2016-17 will be extended to 19 weeks to allow installation of the CMS Phase-I pixel tracker upgrade.

2. Long shutdown 2 (LS2) will last at least 18 months (period with no beams through P5), starting at the earliest in July 2018, while Long Shutdown 3 (LS3) will last 30 months starting at the earliest in January 2023.

3. 10 years of physics operation are planned after LS3, interleaved with shutdowns LS4, LS5. [giving $\geq$ additional 2500 fb^{-1} in integrated luminosity (LHC capability 3500 fb^{-1}), at peak luminosity $5 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ (LHC capability $7.5 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$)].

4. The Phase-I upgrade common project programme, within the limits of the funding so far received for it, covered common upgrade activities during Long Shutdown 1 (LS1) and will cover common activities supporting the approved pixel, HCAL and Trigger subsystem upgrades in the period (including year-end technical stops) up to the beginning of LS2, under the scheduling assumption that only these Phase-I upgrade activities, approved by 2014, are taking place (alongside operation and routine maintenance). For the present document it is assumed that full funding of the Phase-I common project will eventually be achieved, which will support these Phase-I upgrades up their completion during LS2.

5. The common project task for the CMS Phase-II upgrade programme commences in June 2015, and continues until Jun 2025. [As will be seen below, an early start to Phase-II work, with a substantial work-load in LS2, is an essential feature of the current CMS baseline planning, aimed at completing the proposed Phase-II programme by the end of LS3.].

11.3.2 Scope of the Phase 2 Common Project

The common project will, as for construction and Phase 1 upgrade, cover:

1. Integration and installation activities necessary to ensure safe assembly and operation of the Phase-II detector.

2. Modifications to common systems such as the solenoid magnet, beampipe, and services infrastructure, not specific to any individual sub-detector or to the DAQ, Trigger, Control or Offline Systems, necessary either to accommodate Phase-II detector upgrades or to allow CMS systems to continue effective physics operation at HL-LHC, well beyond the performance specification and predicted lifetime of the original experiment design after its Phase-I upgrades.

11.3.3 Main modifications considered

1. Solenoid magnet: cryogenics, powering and control system consolidation and improvement required to prolong the operating lifetime until at least 2035. Following the consolidation carried out in LS1, no major re-build or revision of the cryogenics or power systems is envisaged, rather a series of measures designed to ensure the magnet can operate

reliably at full field for a further two decades. Chief amongst those already identified is the installation of cooled freewheel thyristors to provide for ride-through of power converter electrical or cooling faults. There will also be a requirement for simulations and measurements of the central or fringe fields and of parameters (eg Residual Resistance Ratio) which are indicators of magnet lifetime, along with studies of the effectiveness of particular operating regimes in prolonging that lifetime.

2. Beampipe, beampipe support and beam vacuum system: replacement of all outer steel sections of the beamipe (from $\pm$3m to $\pm$18m) by equivalent aluminium alloy sections to comply with LHC aperture changes, reduce activation levels to which maintenance crews will be exposed, and accommodate possible changes in the eta limits of forward tracking and calorimetry.

3. Crane system: addition of a second carriage to the underground crane in experimental cavern UXC55, largely to allow a greater degree of parallelism in the LS2 and LS3 work plans.

4. Services supplying the experiment: replacement or upgrade of elements of the gas, water cooling, electrical power, dry gas and HVAC systems. These will be needed either because the end-of (maintainable) lifetime is expected to be reached, or to accommodate upgraded detector systems with modified requirements.

5. Common facilities: provision of suitable laboratory, assembly, workshop and storage space at CERN needed to execute the Phase-II upgrade programme in a timely, efficient and safe manner. This includes an extension of the surface assembly building at LHC Point 5, refurbishment of already allocated assembly and laboratory buildings on CERN Meyrin and Prevessin sites, improved or enhanced test facilities, including test-beam zones, cosmic ray test stands, irradiation facilities etc, plus the update of the Electronics Test and Integration Facility on the Prevessin site.

6. Moving systems, tooling and access equipment: revision of the major element winching system by a hydraulic system will be required to achieve higher precision and reduce, through automated control, the number of skilled personnel needed for operation. Massive new tooling will be necessary for Phase-II detector installation, particularly the removal and replacement of the endcap calorimetry elements. Simultaneous access to many different parts of the detector will be needed to satisfy schedule constraints, which may also drive re-design of some of the platforms and access devices used heretofore.

7. Safety and radioprotection systems: adaptation of the Detector Safety System will be needed to accommodate the Phase-II detector, along with replacement of CMS specific, safety-related equipment which has reached the end of its maintainable lifetime. Substantial changes to CMS specific safety systems may be needed to comply with applicable new legislation (CERN or host state). Safety studies will be needed for authorisation of Phase-II upgrade work along with training, or training equipment, specific to Phase-II detector assembly, testing, installation and subsequent operation, including remote monitoring and diagnostic equipment, full scale models, dedicated handling gear, radioprotection shielding and CMS specific safety courses.

11.3.4 Main common project activities foreseen

1. Execution (including safety oversight) of the common systems and facilities changes outlined above.

2. Overall planning of the technical stops and shutdowns in the period 2015–2025 and specific project planning for the common items. Specific planning support to individual subsystems may be provided by arrangement.

3. Engineering and Electronics/Electrical Integration to ensure coherence of all Phase-II elements.

4. Specific engineering design and simulation support for heavy engineering aspects of the Phase-II upgrade.

5. Common installation and technical support for upgrade activities during shutdowns, covering transport teams (crane drivers and riggers, transport between sites) cabling in common ducts and chains, provision of access platforms and scaffolding, provision of experienced technical support to assist in sub-detector removal and (re-) installation in the cavern environment and, lastly, provision of workshop technical personnel beyond those normally needed for maintenance.

11.4 Schedule

With some notable exceptions, most changes (sub-detectors and common systems) can only be made during long LHC shutdowns or extended year-end technical stops, whose assumed timeframe is described above.

The 30-month beam-to-beam duration set as the target for LS3 presents a major planning and logistical challenge and is only viable provided substantial preparatory work and some parts of the detector upgrade are completed during LS2 and those year-end-technical-stops before LS3 start. Upgrade work will define the LS3 critical path, with the load of maintenance and repair work (estimated from LS1 experience to be 40% of all tasks) running concurrently, but, as far as feasible, parasitically. Besides opening and closing of the experiment, five activities, of the many currently anticipated, will determine the underground logistics, and critical path, namely:

- Pixel and strip tracker replacement
- Revision of Electromagnetic Calorimeter barrel (EB) modules with entirely new on-board electronics, involving dismounting all 36 modules to surface, refitting in parallel batches at re-work/test stations and re-mounting the modules inside the HCAL.
- Replacement of the endcap calorimetery and first muon layer structures forming the "nose" mounted on the first endcap yoke.
- Electronics replacement for all barrel muon stations and part, or all, of the inner ring of the endcap muon system.
- Infrastructure changes such as the TAS and beampipe replacement, the installation of a second crane bridge, the replacement of the main elevator, and the replacement of other elements of importance for safety.

The first 3 items must necessarily be tackled in the same shutdown, as discussed in the next section.

11.4.1 Assumptions of the planning

1. Crane-sharing in the experimental cavern has been proven in LS1 to often limit the possibilities for simultaneous detector work in the typically 2-4 working zones which the CMS

modular structure can make available. The installation of a second crane bridge (carrying a 5t or 7t hook) is proposed and assumed to be executed in LS2.

2. The tracker and electromagnetic calorimeter changes are constrained by the exceptionally large, common un-cabling and re-cabling overheads to happen in the same shutdown. Radiation damage and trigger rate projections (based on estimated LHC performance) further constrain changes 1 and 3 to be done in LS3. Barrel muon electronics obsolescence replacement requires a change which cannot be ready before LS3 but, to maintain reliability, should not be delayed further.

3. The installation of new fast triggering detectors in the endcap muon system (GE1/1) is assumed be done in LS2.

4. The revision of electronics in one or more of the inner ring muon endcap stations (ME2/1, 3/1, 4/1) is assumed to be completed in LS2. For the purposes of the initial plan presented here, this leads to a considerable saving in time and surface resources in LS3.

5. The type of infrastructure work indicated above, particularly involving activated elements requiring substantial "cooling times", or affecting underground access conditions, is particularly disruptive to a coherent and efficient planning of LS3. A key element of the provisional Phase 2 plan is therefore to pre-empt as many of these types of activity as possible before LS3 start. Thus replacement of the outer beampipes (change from steel to aluminium) and the installation of a second crane in the experimental cavern are assumed to happen in LS2; similarly the replacement of elevators is assumed to happen during LHC Run2 (preferable) or LS2.

6. The logistics plan assumes additional laboratory space at Point 5 (in building SX5) and the correspondingly enhanced services (gas, cooling, power etc) needed for detector testing and re-commissioning are available 6 months after the end of LS2. This means the extended surface facility (see Section 11.10.2) needs to be completed by the end LS2, to allow the workshop areas to be moved there from SX5 as soon as the LS2 workload is completed.

7. Substantial active storage facilities are needed to house and maintain in good order the ME 1/2 and ME1/3 Cathode Strip Chambers, the outer 2 rings of muon detectors mounted on the first endcap disk, which must be removed during LS3 in order to replace the services exiting the endcap calorimetry radially across the face of this disk. These facilities are assumed to be provided by a temporary hard-shell building, installed adjacent to SX5 for the duration of LS3 and thus within easily extendable reach of the services (detector gas, climate control) available therein. A further soft shell, weather proof storage area may also be needed as a staging area (as was done in LS1).

8. Since EB super-module revision is the dominant influence on shutdown length, the baseline plan assumes that two "enfourneurs" (supermodule removal/re-installation tooling systems) are available and that refurbishment/test stations are available at P5 to handle 6 supermodules simultaneously.

9. Recalling the construction philosophy of CMS, it is proposed to remove the calorimetry "nose" from the YE1 yoke disk at each end in one piece, using customized heavy lift equipment, which would then, in each case, subsequently refit a new "nose", preassembled in the surface facilities at Point 5. [NB. Rebuilding, or re-loading, the absorber of the endcap hadron calorimetry in situ is extremely time-consuming and blocks the logistic arrangement for long periods, prolonging LS3 over more than 3 calendar years].

10. There is assumed to be no conflict of resources, or other disruption to CMS activity, underground or on the surface at P5, due to excavation or other preparatory work associated with the refit of the accelerator for HL-LHC.

11.4.2 Provisional schedule

An overview Gannt chart, based on a detailed week by week analysis, is shown in Fig. 11.5.

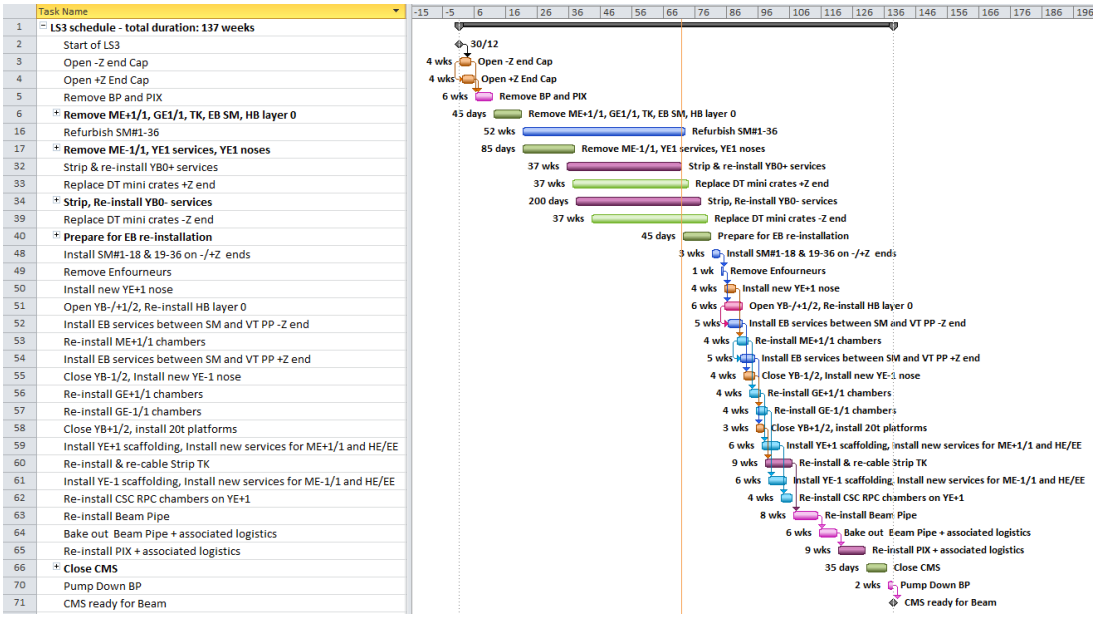


Figure 11.5: Overview of the LS3 planning.

Under the assumptions listed above, and employing the same scheduling principles used successfully in LS1, this provisional planning foresees Phase 2 upgrades being completed by the end of an LS3 of duration 137 weeks, still somewhat longer than the target duration of 30 calendar months. The gains to be achieved by fully exploiting two cranes, two "enfourneurs", etc., and by further optimising the timing of the removal/replacement of the two endcap nose assemblies, are still to be fully assessed.

11.5 Magnet and Cryogenics

Maintaining the performance of the CMS magnet at 3.8T for an operating lifetime of 30 years, much longer than anticipated, is imperative to the success of Phase-II. Following the consolidation carried out in LS1, no major re-build or revision of the cryogenics or power systems is envisaged, rather a series of measures designed to ensure the existing magnet system can operate reliably at full field for a further two decades. The changes in Phase-I focussed on the cryogenics, introducing redundancy in the compressor system to eliminate the risk of very long down-time during physics operation, and powering redundantly from the machine and general services networks to reduce cryo-plant shutdowns (and consequent slow discharges) due to power glitches. Phase-II focuses on eliminating slow discharges due to transient power converter faults, on revising the control system (LS2), including auto-restart of the "hot-spare" compressors and on simulation and measurement to better understand the ageing processes and how to mitigate them,

Table 11.1: History of slow discharge cycles and fast discharges of the CMS magnet.

Year	Number	On request	Unexpected	FD at nominal field
2006	12	9	3	5
2008	12	9	3	1
2009	10	5	5	1
2010	14	7	7	0
2011	7	4	3	2
2012	9	2	7	1
2013	2	1	1	0
2015	2	1 (2014)	1	0
Total	68	38	30	10

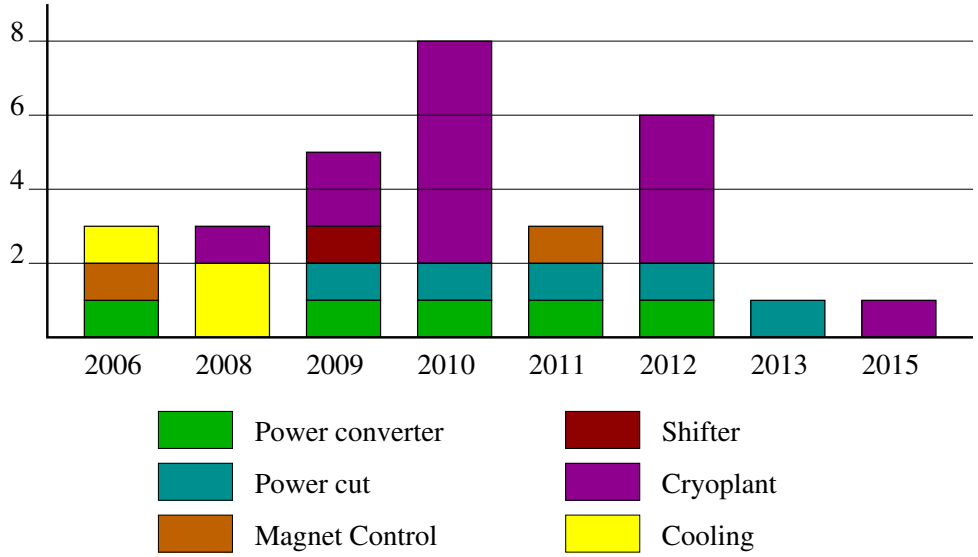


Figure 11.6: Main identified causes for magnet unexpected discharges over the past years.

Ageing effects are expected from mechanical and thermal cycling of the magnet; thus planned on-off cycles must be minimised and unplanned cycles eliminated as far as possible. In general the objective is to keep the magnet cold except during very long shutdowns, avoid any accidental fast-discharge, reduce to an absolute minimum the incidence of slow discharges (usually due to electrical or cooling faults affecting the power converter) and, through liaison with the beams department, minimise the number of cycles enforced by the requirements of machine commissioning and operation. The number of cycles in each year since magnet start-up in 2006 is shown in Table 11.1. The original design assumed around 200 cycles would be needed in the magnet lifetime. A design target of 5 slow cycles per year should be aimed at from now on to give confidence that the magnet will be fully capable in 2036 (30 year lifetime). Fast discharges (FD) are to be avoided.

The main source of ageing studied so far is the work-hardening of the pure aluminium stabiliser around the superconducting cable due to mechanical cycles of the magnet (transit between 2T and 3.8T and back). This is expressed as a reduction in the Residual Resistance Ratio (RRR). At some point, originally estimated at around 200 cycles, the conductor system may not be able to safely absorb the current necessary for 3.8T. The predicted evolution of RRR has been simulated for 3.8T as shown in Fig. 11.7, along with the original 4.0T simulation. Measurements

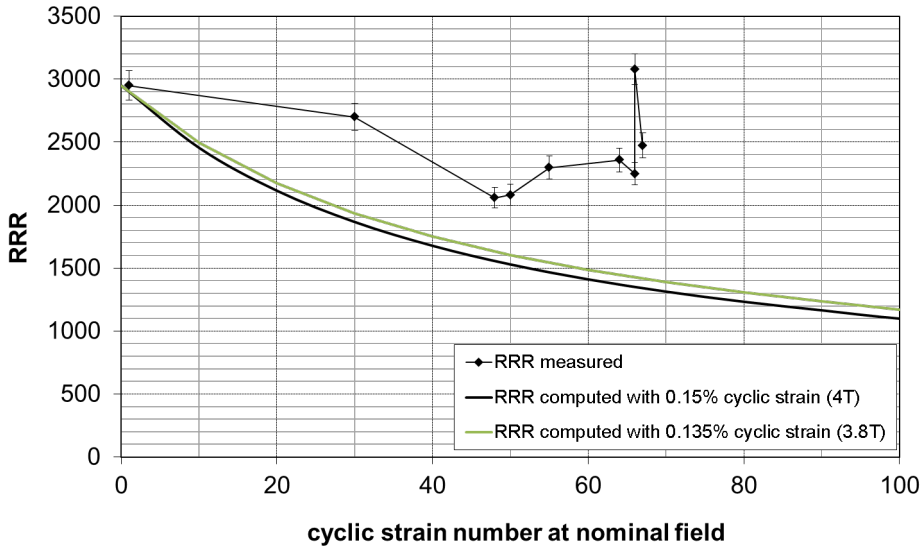


Figure 11.7: Residual resistance ratio vs number of strain cycles.

of RRR, taken after fixed numbers of cycles, are superimposed. The model approximates the cyclic strain as if it applied uniformly across the conductor, which is not in fact the case.

Warm-up to room temperature is known to provoke recovery of the measured RRR value, at least in the shortterm. However the understanding of this mechanism and its permanence are poor. The measurement taken after the magnet was cooled down in 2014 indicated a large recovery took place despite the reverse thermal cycle from room temperature back to 4K. However, as can be seen in Fig 11.7, the effect of this recovery mostly disappeared after the subsequent magnetic cycle, leaving the RRR about 10% higher than the last value measured before LS1.

The existing freewheel thyristors in the magnet power converter (see Fig. 11.8) are air-cooled and can tolerate the full current for only a few minutes at most to protect the converter against current transients. After close consultation with CERN TE department, CMS proposes to replace these by water-cooled freewheel thyristors, with an independent UPS and a water cooling system with diesel power back-up. In case of a transient power converter fault (electrical or cooling), the main breakers stay closed and the solenoid current decays with a time-constant of about 39 hours through the busbars and the freewheel thyristors. A dedicated control system then re-engages the power converter when ready. This modification would give immunity to most power converter electrical faults and a majority of cooling faults (see Fig. 11.8).

A demonstrator system will be built and installed on the CMS M1 3T test magnet in the H2 beamline at CERN, during YETS 2016-17, in order to prove the concept before implementing it in CMS during LS2. The final thyristor system, with dedicated busbars and water cooling circuits, will be integrated underground in USC5-S3, outside the existing power converter cabinets. Other magnet-specific consolidation and preventive maintenance activities, planned for LS2 and LS3, are:

1. Replacement of the cryoplant lubricant separator in SH5 to diminish the risk of breox pollution in the cold box (better done in LS2 than LS3)

2. Replacement of primary vacuum pumps and valves (LS3).

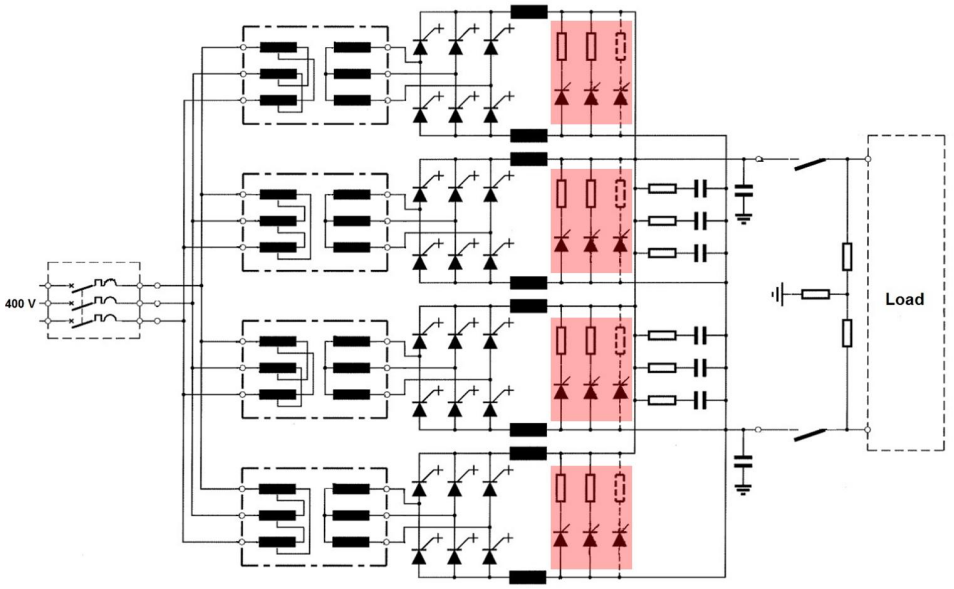


Figure 11.8: Magnet power converter schematic, with the existing freewheel thyristors highlighted in red. The load is the CMS magnet with the dump resistor in parallel.

3. Replacement of electrical contacts in the switch breakers (LS2/LS3).
4. Replacement of corroded copper flexible power connections (under monitoring, part to be done in LS2, balance in LS3).
5. Upgrade of the magnet safety system (planned in LS2), to be tested on the demonstrator implemented in 2015 on the CMS M1 3T test magnet in the H2 beamline at CERN.
6. Upgrade of the magnet control system by replacement of the PLCs (LS3).

11.6 Beam pipe

Figure 11.9 shows the $\pm$23m extent of the beam-pipe either side of CMS, which is within the CMS experimental cavern. The design of the central, endcap, HF and CT2 chambers, including the bellows around 3m and 16m and the pumping station at 13.4 m, are heavily influenced by CMS requirements (detector performance, backgrounds, activation), but must still satisfy accelerator requirements for aperture, vacuum stability, impedance etc. It should be noted that the 18m "VAX" pumping stations and the TAS (Target Absorber Secondary) absorbers at 18-21 m, including the vacuum chambers within them and the "FIN" shielding around them, are wholly the responsibility of the LHC machine, but must still satisfy integration constraints within the experiment forward shielding. In the Phase-I upgrade, changes to the beam-pipe were physics performance motivated and consisted of reducing the radius of the central chamber to the minimum feasible so as to bring the pixel tracker first layer as close as possible to the interaction point.

For Phase-II, there are 3 motivations for further change:

1. The increase in the aperture of the final focus quadrupoles in LS3, to prepare for HL-LHC optics, will require replacing the TAS, FWD, and CT2 chambers to match.

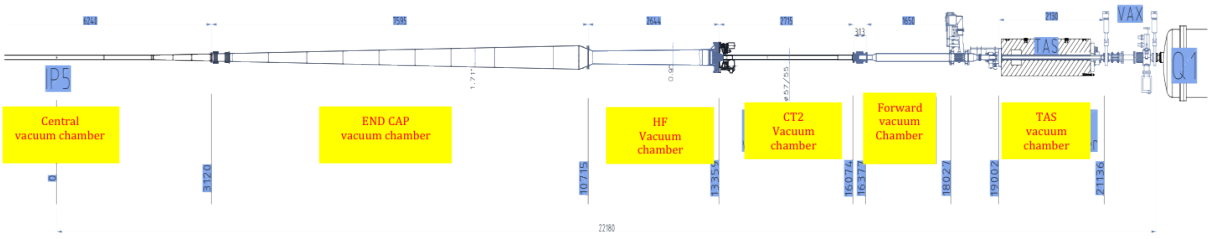


Figure 11.9: Layout of CMS beam-pipe from interaction point to first quadrupole. All dimensions are in millimeters.

2. The conical endcap chamber design has proven very effective in minimising activation and backgrounds, allowing the use of stainless steel at nominal LHC luminosity. However, reducing dose to personnel during shutdown activity is a high priority once higher luminosity is contemplated. The capability of opening CMS with the beam-pipe in place will be retained for Phase-II and therefore substantial work-streams in close proximity to the vacuum chambers and flanges can be expected. A system of RP shielding (partly constructed and tested during Phase-I upgrade) is envisaged for rapid interventions which exploit leaving the beam-pipe in place, but, in general, all vacuum chambers within $\pm$18m will always be removed during Long Shutdowns to minimise mechanical risk, logistic complexity and irradiation of personnel. Given the very busy schedule of LS3, and the possible absence of the serendipitous extended cooling time, such as preceded LS1, the dose rates encountered by beam-pipe personnel during the first interventions of opening the experiment and removing the vacuum chambers, are estimated, for an unchanged chamber material, to be more than an order of magnitude higher than at the same stage of LS1. To minimise the doses involved in manipulating the beampipe or working near it (ALARA principle), it is proposed to substitute all the existing stainless steel chambers and vacuum flanges by new ones built from Al 2219 alloy. FLUKA simulations, shown in Fig. 11.10, predict a typical reduction factor of at least 5 in contact dose rate in the period between 1 week and 1 month into shutdown when first interventions occur.

   It can be inferred from simulations that a cooling time (hiatus in activity to allow short-lived isotope decay) of at least 6 months at the outset of LS3 before removing beam-pipe elements would be needed to produce the same reduction in dose-rate as that gained from changing the beam-pipe material. Within the constraints of the LS3 work-plan, there is no way to avoid cooling time transferring directly on to the critical path, unless there is a prolonged period of low luminosity or heavy ion operation before this shutdown.

   The change in vacuum chamber material will also bring benefits in reducing RF heating caused by the conical endcap chamber geometry (again, by a factor 4-5) and in reducing machine-induced background in the tracking and muon systems (yet to be quantitatively evaluated).

   The replacement of all stainless steel chambers and flanges in LS2 (where significant cooling time can be afforded) is thus very strongly recommended.

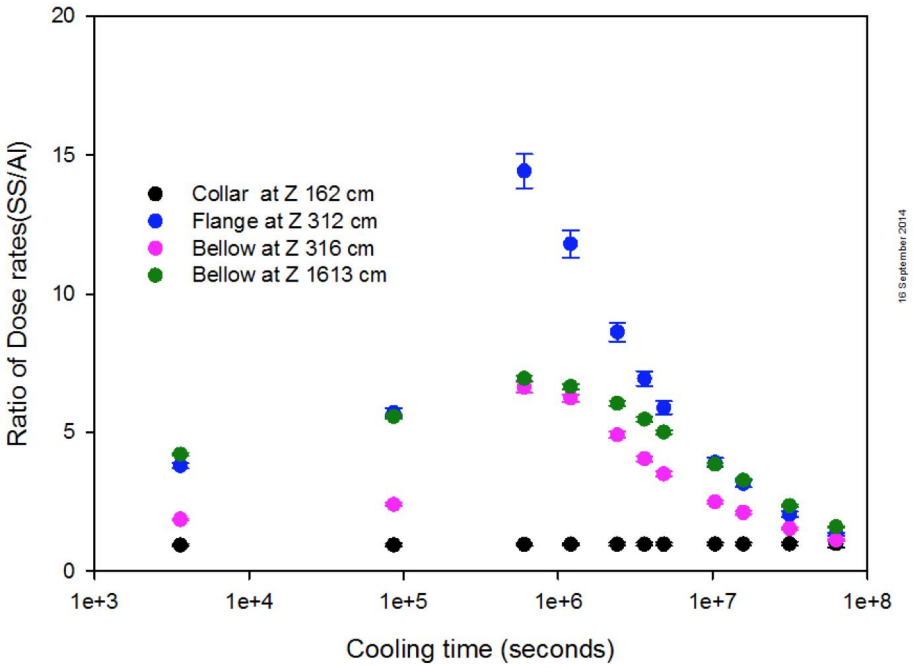


Figure 11.10: Ratio of contact dose rates vs time at the outset of LS3 for various z locations along the beam vacuum chambers, where high levels of activation are expected due to concentration of material. Phase-I (stainless steel) and Phase-II (Al 2219) vacuum chambers are compared. The Ti-Zr collar at z = 1620mm, unchanged between Phases 1 and 2, is included as a control.

3. The desire to examine extending the range of forward calorimetry to higher $|\eta|$ in order (for instance) to give homogeneous coverage of the region of interest for Vector Boson Fusion processes, automatically implies the need to study a beam-pipe design with a lower endcap cone angle (ie increased η of the pipe wall) and a re-designed ion pump station at 13.4m, in order to maintain radial clearances between the vacuum chambers and pumps and the innermost edge of the endcap calorimeter, when opening the experiment.

Presently the HL-LHC parameter list envisages changing only the inner diameter of the copper TAS absorber (1.8m long, outer diameter 250mm) from 34mm to 54mm to accommodate the increased quadrupole aperture. The unchanged outer diameter is assumed as input to the CMS Phase-II common project, in that the existing FIN shielding surrounding the TAS is taken to be unchanged. Recent studies are pessimistic about being able to replace the TAS in LS2, since the inner aperture of the existing quadrupoles must be protected until they are replaced. No feasible design could be found based on the concept of a core section removable for Run 3, and so it is likely that TAS replacement will have to be scheduled late in LS3, just before beampipe re-installation, to allow adequate cooling.

The imperative of replacing the steel vacuum chambers (including CT2) with higher diameter chambers in LS2 implies no re-installation of the existing TOTEM T2 or CASTOR detectors after LS2. The CT2 chamber (currently 55mm in diameter, compared with the existing TAS diameter of 34mm) will be replaced by a chamber of at least 65mm in diameter (compared with the new TAS diameter of 54mm).

A decision on the shape (η limit) of the vacuum chamber and design of the local pump assemblies will have to be taken by late 2015. Design details and costings have so far been worked out with TE-VSC for the existing $\eta = 4.9$ cone angle, retaining the existing triple ion-pump

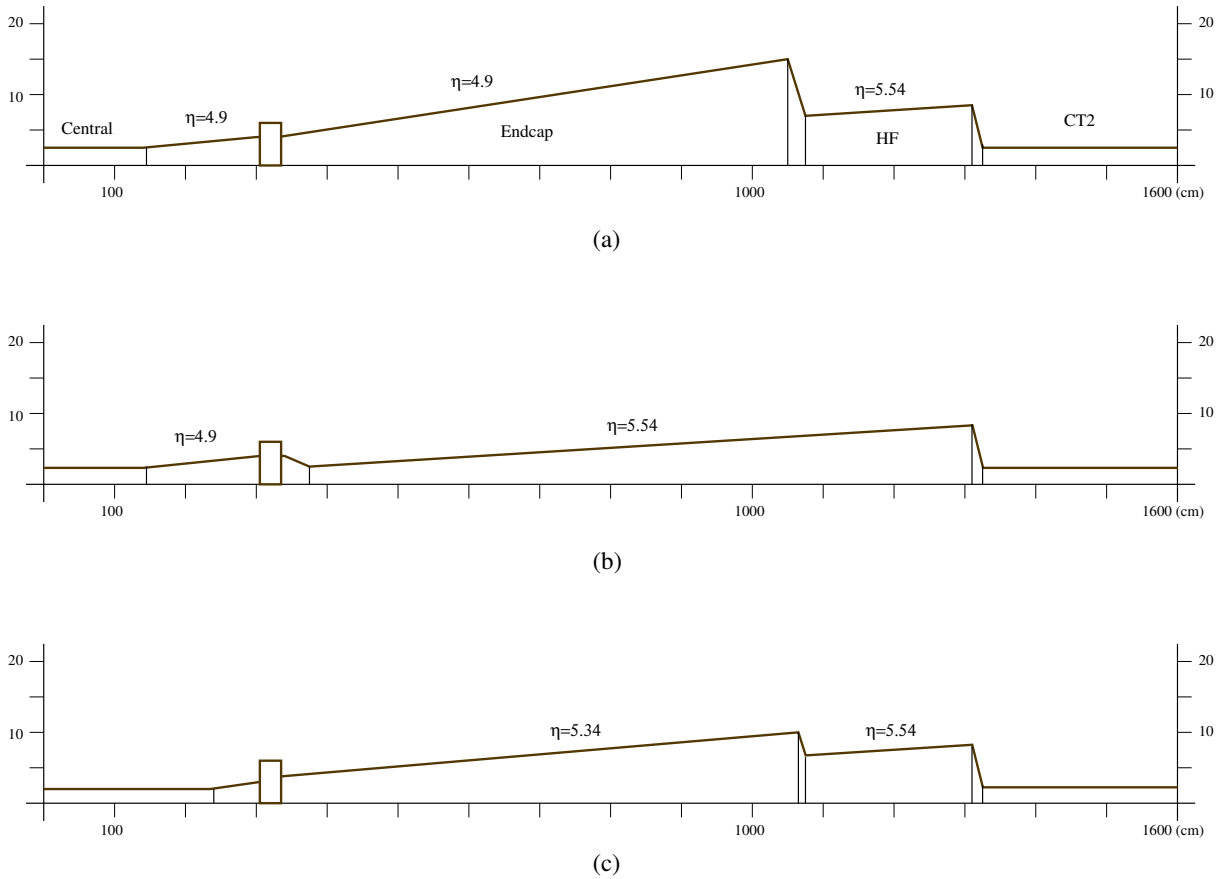


Figure 11.11: Phase-II geometry options, all incorporating the existing central beryllium chamber, with Al2219 for all other chambers. (a) existing geometry apart from CT2 and forward pipes to accommodate increased TAS aperture. (b) reduced endcap cone angle, allowing inner radius of endcap calorimetry to be reduced by 50mm, interfacing with existing central pipe through bellows at 3.1m. (c) as for (b), but maintaining separate endcap and HF pipes, allowing the same reduction in endcap calorimetry inner radius, but giving lower vacuum impedance and a cleaner matching to a (modified) central chamber to avoid provoking RF heating.

station at 13.4m. (The existing geometry is illustrated in Fig. 11.11(a).

Higher cone angles are possible in principle (as shown in Fig. 11.11(b,c), up to a pragmatic limit of $\eta = 5.54$ (the cone angle of the HF pipe), where considerations of mechanical and vacuum stability, mechanical integration, plus tight alignment constraints on the cone, are expected to become limitations. An endcap/HF pipe at $\eta \approx 5.54$ would allow the endcap detector structures to be extended inwards from $\eta = 3.0$ to $\eta = 3.8$, although the limit of the active detector would be somewhat less than this due to the need for mechanical and environmental housings, shielding and alignment system components. This change would require re-design of the vacuum chamber mechanical supports, the vacuum system (notably the introduction of an annular ion pump instead of the existing triple ion pump at 13.5m) and the RP shielding systems. Mechanical, vacuum and dynamic stability remain to be established by calculation. It is likely (though not yet examined in detail) that a change in the cone angle might necessitate modification of the central pipe, which becomes conical, currently along $\eta = 4.9$ within the beryllium section (from 1450mm outwards). This option is sketched in Fig. 11.11(c). The beryllium and aluminium endcone sections of of the central chamber would have to be cut off and

replaced with a section in the new geometry, an operation which can probably only be done by the manufacturer in the USA. This operation is possible within the LS3 planning, but not without risk.

The installed CMS Phase-I central chamber, of inner diameter 43.4mm, is not within the shadow of the 54mm HL-LHC TAS. Both ATLAS and CMS have argued that the parallel beams shadowing condition (ie that all parts of the beampipe must be in the shadow of the TAS for parallel beams) should not apply for HL-LHC, where no high $\beta*$ operation is foreseen. Should this be agreed, the existing Phase-I central chamber can be retained (or modified for higher cone angle). A new central chamber with inner diameter as low as 38mm could be considered (based on the latest refinements of the aperture criteria applied in the boundary conditions of the new ALICE central chamber), but only if strongly justified on physics performance grounds. If the shadowing condition is applied, the natural fall-back would be to use the 58mm inner-diameter central beryllium chamber used in Run 1 (again with modifications for increased cone angle if necessary). The Run 1 beampipe central chamber should therefore be retained in good condition.

Whatever the final choice, it is essential to maintain the policy of having three examples of each vacuum chamber, with two installed and one retained as a spare. In the case of the central chamber, a hybrid with only the central 1m section in beryllium is acceptable and arrangements to share spare components with other experiments may be possible. All other new chambers will be manufactured in triplicate.

11.7 UXC55 crane system

CMS major elements were pre-assembled on the surface and lowered as a series of 15 heavy loads (350t to 2000t) using a customized, temporary, fixed-position gantry crane installed over the PX56 shaft and SX5 building, employing strand-jacking technology. Further underground assembly was designed to involve loads $<$ 20t. A key reason to have only one crane carriage (20t and 5t hooks) in the underground experimental cavern was the importance of being able to move loads up to 20t from one extreme end of the cavern to the other without having to transfer between carriages.

Experience from shutdown work 2009-2014 suggests that the original concept of a single light (20t) bridge crane in UXC 55, although valid during construction, may be less applicable for the future, given the actual LHC operating regime.

CMS is examining possible improvements in collaboration with the CERN Engineering Department (EN-HE group) based on the following concept (see Fig. 11.12):

- In long shutdowns, where the critical path sequence for detector work is often determined by crane availability, there would be considerable advantages in being able to execute crane-dependent maintenance of upgrade work packages simultaneously in two of the many (typically 4) locations made available by the modular construction of CMS (eg + z and z endcaps simultaneously). In general an additional carriage carrying a 5t hook is adequate for such work, although 7t or 10t would improve design flexibility if it does not cause range restrictions or require large support infrastructure modifications. An interlock system would be needed to prevent collisions of the two carriage.
- The need for a single carriage to be able to transfer at least a 20t load from one end of the cavern to the other remains. It has been something of a struggle to design and

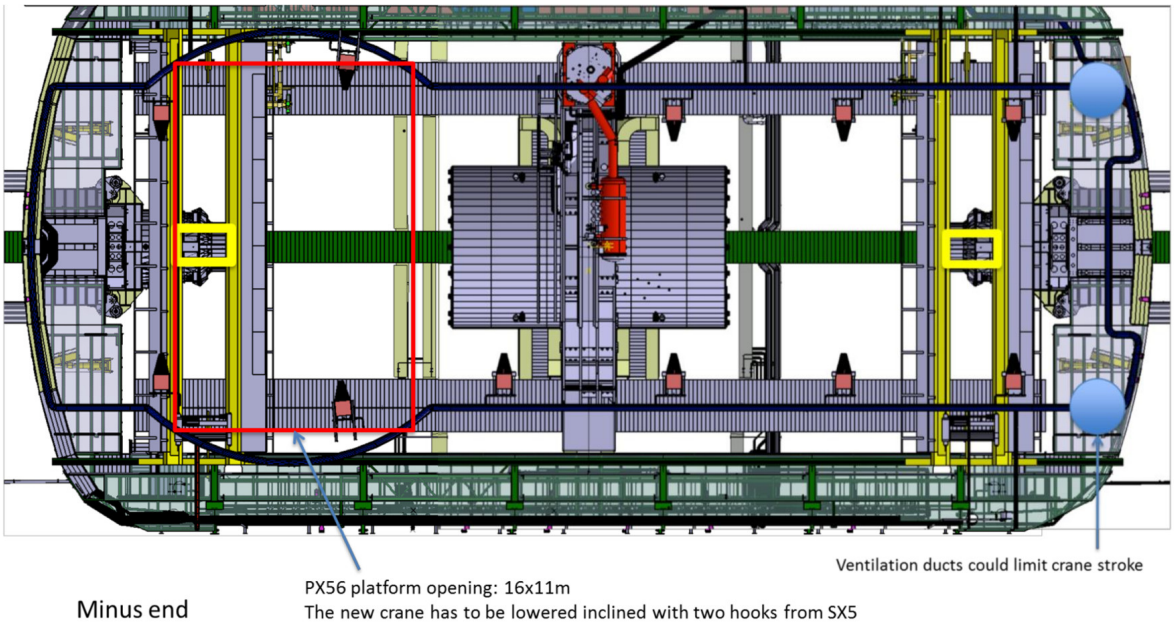


Figure 11.12: Example of engineering studies for the installation of a second crane bridge in UXC.

build within the 20t limit per object. If an increase to 25t or 30t load limit can be achieved at little extra cost, this would be valuable, but is not the highest priority.

- Personnel access to the detector from a suspended cage is quite frequently needed. A nacelle-like arm integrated into one of the crane carriages would be a useful feature, leading to safer working and fewer special authorization requests.

11.8 Service Infrastructure

11.8.1 Detector Cooling

Both mixed and chilled water cooling systems supplying CMS requirements at P5 are currently loaded near to the limit of their capacity. The present system (illustrated in Fig. 11.13) delivers to the detector a total cooling power of 0.2 MW as chilled water at 5-6 degrees (primary water cooling for Tracker fluorocarbon system) and a further 0.9 MW as mixed water at 13 deg C. There is very little margin in either system. The chilled water capacity to SCX (HLT farm, control room, network switches) was upgraded to 1MW during Phase 1 and is considered sufficient for all future contingencies. The mixed water cooling of the rack system delivers 2.6 MW. It is likely that future upgrades may require somewhat more capacity (10-20%), but may also challenge the individual cooling capacity (about 12kW max) in each rack.

For the purposes of this proposal, it is assumed that (as in construction and Phase 1 upgrade), water-cooling systems for CMS specific facilities at CERN are a central CMS responsibility. Water-cooling systems for detectors are assumed to be a central responsibility up to local on-board distribution manifolds or connections (defined for each sub-system). Other detector-specific cooling systems which do not use water (eg fluorocarbon or 2-phase CO_2) are presently assumed to be the responsibility of the subsystem concerned, although under the overall oversight of Technical Coordination, through the cooling coordinator.

A preliminary analysis of Phase 2 detector requirements suggests that the main change needed will be additional chilled water. Proposals under discussion to lower the operating temperature of ECAL and HCAL electronics are also taken into account, since this automatically requires a shift from mixed water to chilled water.

The proposed changes to chilled water capacity are summarised in Table 11.2.

Table 11.2: Proposed changes to chilled water capacity.

Area	Power Increase (kW) Phase-I	Phase-II	Comment
HVAC	500	500	Existing chiller
Tracker and pixel primary	200	400	New chiller
ECAL+HCAL	0	600	New chiller
Endcap calorimetry	0	500	New chiller

The conclusion is that a further 1.5 MW in chilled water capacity is needed. CERN Engineering Department is studying whether this can be achieved within the capacity of the existing cooling towers. In this document it is assumed that no change to the towers is needed.

This change to chilled water would involve replacing the heat exchanger, keeping the pump-station and existing pipework, but would require an extensive programme of insulation. The upgraded installation will include pipework to allow routing of 150kW chilled water capacity to the surface assembly building (SX5) to supply the test stations for ECAL barrel supermodule refurbishment. For HCAL, a moved to chilled water would require decoupling from the yoke circuit and a substantial amount of new insulated pipework. Details of this task have not yet been assessed. Removing the ECAL (417kW needed) from its current mixed water cooling is considered likely to cover any additional requirements which could be placed on the mixed water circuit (eg from the rack circuit).

For reasons of required availability during LS3, the changes required have to be done during LS2 or Run 3 (pre-LS3). Based on initial discussions with the CERN engineering department, it is believed that existing spares stocks and routine maintenance can cover the effects of ageing on all major mechanical components of the cooling system over the next 2 decades. However, the CMS specific cooling control systems will become un-maintainably obsolete and so a provision must be made to replace them.

11.8.2 Electrical power system

The installed electrical power at P5 totals about 13.5 MW for a peak demand of about 9.0 MW, used mostly for the cooling (5.5 MW) and ventilation systems. The CMS detector consumes on average about 2.5 MW (1 MW for the detector and 1.5 MW for the cryogenics and magnet). A considerable margin is therefore available to cover increased detector power consumption, but this has to be seen against the substantial increases in electrical consumption by services (eg cooling), which are also envisaged. For the purposes of this document, it is assumed that CERN (EN-EL) as host laboratory, will take care of providing the electrical power required to run the experiments accepted to operate at HL-LHC, along with their ancillary systems. The Phase 2 common project (apart from changes to the magnet power system described above) is therefore only concerned with CMS-specific changes.

These mostly apply to the CMS Detector Power System and to the computing facilities on the first floor of SCX5, which are fed from the 18kV machine power network, as illustrated in Fig. 11.14.

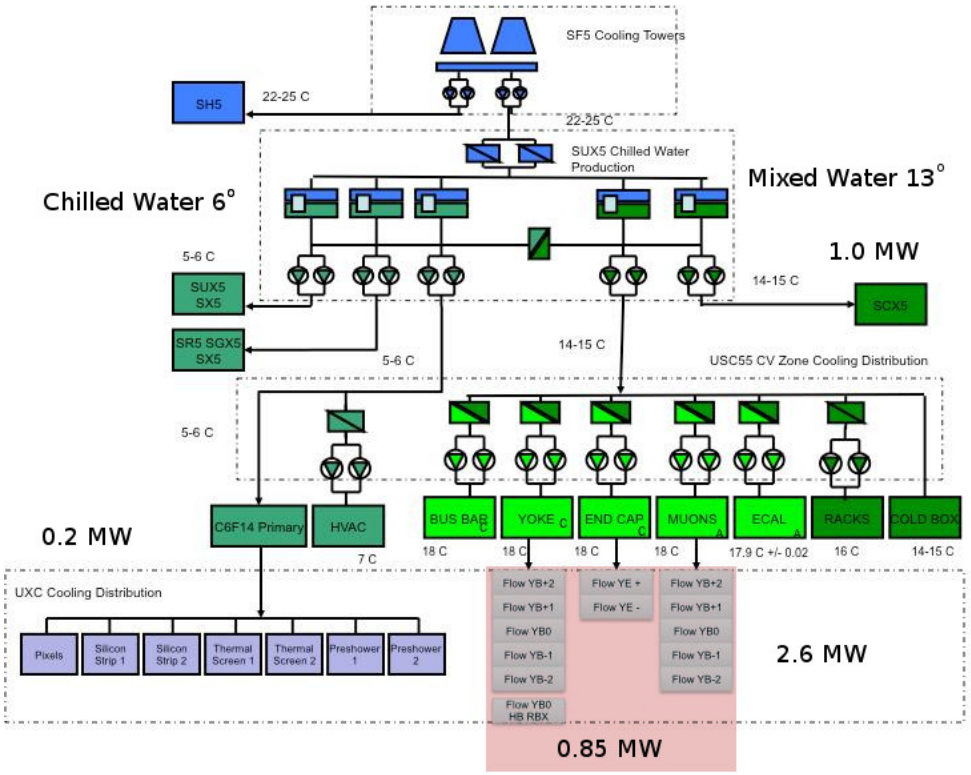


Figure 11.13: Current Layout of Cooling Infrastructure at Point 5. The power available for CMS detector-related cooling is indicated.

The Detector Power System supplies filtered power to the sensitive electronics, power systems and processors on-detector and in the service cavern. It provides for 400kV or 18kV glitch ride-through, and also for the opportunity of a graceful shutdown, via battery backedup power, in the case of extended power failure. After Phase 1 modifications, this system consists of 4 x 500kVA UPS units, providing $n+1$ redundancy. The CMS detector currently draws 1200kW, which allows for the loss of one 500kVA UPS unit while still maintaining 20% headroom in power capacity. Further increases in detector power consumption are expected as a result of radiation damage. To accommodate these increases, as well as new Phase 2 detectors, it is proposed to expand the system by adding one more 500kVA unit (brown, dotted, see Fig. 11.14).

Following the experience of Run 1, in particular after analysing CMS recovery time and induced faults after 18kV power failures, it is considered essential to provide the CMS computing facilities with the same level of power quality protection as accorded to the CMS detector. Currently, most of the CMS computing facility at Point 5 is directly exposed to disturbances on the local electrical network. The proposed extension of the UPS system (shown in green in Fig. 11.14) will require a basis of two more 500kVA units and an additional one for $n+1$ redundancy. This separate system will provide immunity from electrical disturbances and allow for a graceful shutdown of the SCX5 CMS computing facility, which includes key network switches and the high level trigger (HLT) farm, also doubling as a major data-processing centre in beam-free periods.

Both UPS expansions will be carried out as a single project under Technical Coordination oversight, but a substantial fraction of the cost will be provided by the DAQ sub-system, whose facility will thereby be largely protected against unnecessary power cycles and the attrition induced by power incidents.

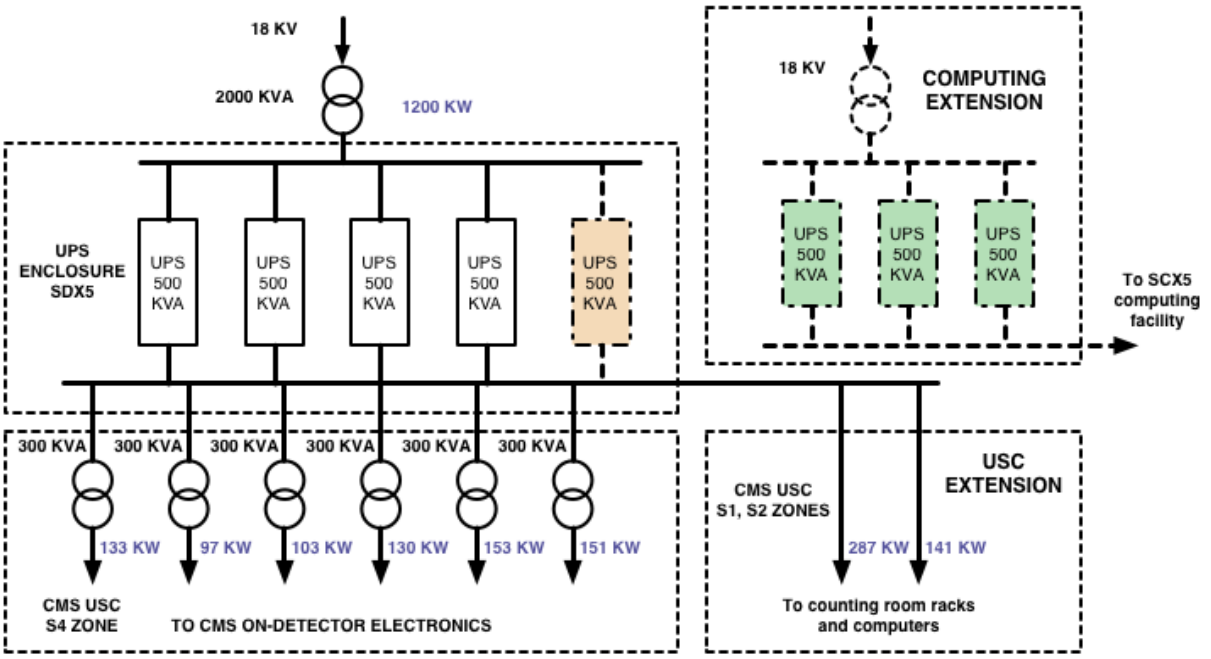


Figure 11.14: Detector power distribution for CMS (post-LS1), showing (dotted) the proposed Phase 2 extensions.

Implementation of the full upgraded system in its current location in the SDX5 enclosure is incompatible with the long-term interests of CMS for the following reasons:

- Expansion of this enclosure to accommodate four new 500kVA UPS units and additional electrical switchgear will cut off access through SDX5 between the control room and the new SL53 offices and meeting rooms (see Fig. 11.16).
- Recent incidents have highlighted the desirability on safety grounds of relocating the battery banks, currently in the SUX5 (cooling and ventilation) building. Recommendations from a panel, including HSE representatives, are awaited, but are expected to recommend housing large battery banks in dedicated buildings.

The assumption made here is that safety issues and HSE recommendations will require an eventual relocation of the CMS Detector Power System as part of the evolution of a Phase 2 upgrade. Implementation of the Computing Extension should be executed in concert with the long-term CMS strategy, and to that end, the enclosure (barn) containing the UPS units and battery banks should be located away from the current locations in SDX5 and SUX5 and not close to any ventilation intakes.

Figure 11.16 shows a potential location for the barn, (next to the cooling towers, building 3565). It would only require the environmental stability consistent with power system equipment. Substantial new cabling installation (in existing tunnels and trenches), would be required.

The implementation would take place in 3 stages. The first, to be executed as soon as possible (YETS 2015 preferably) would involve moving the 80kVA control room UPS alongside the existing battery banks in building 3582 (SUX) and implementing fire mitigation procedures. In stage 2, for completion in 2017, the UPS barn would be built and would initially house only the Computing Extension UPS units and battery banks. In the final stage, the new, Phase 2 CMS Detector Power System would also be installed in the barn and made operational during LS3, considering that, by then, the current system will have reached the end of its operational

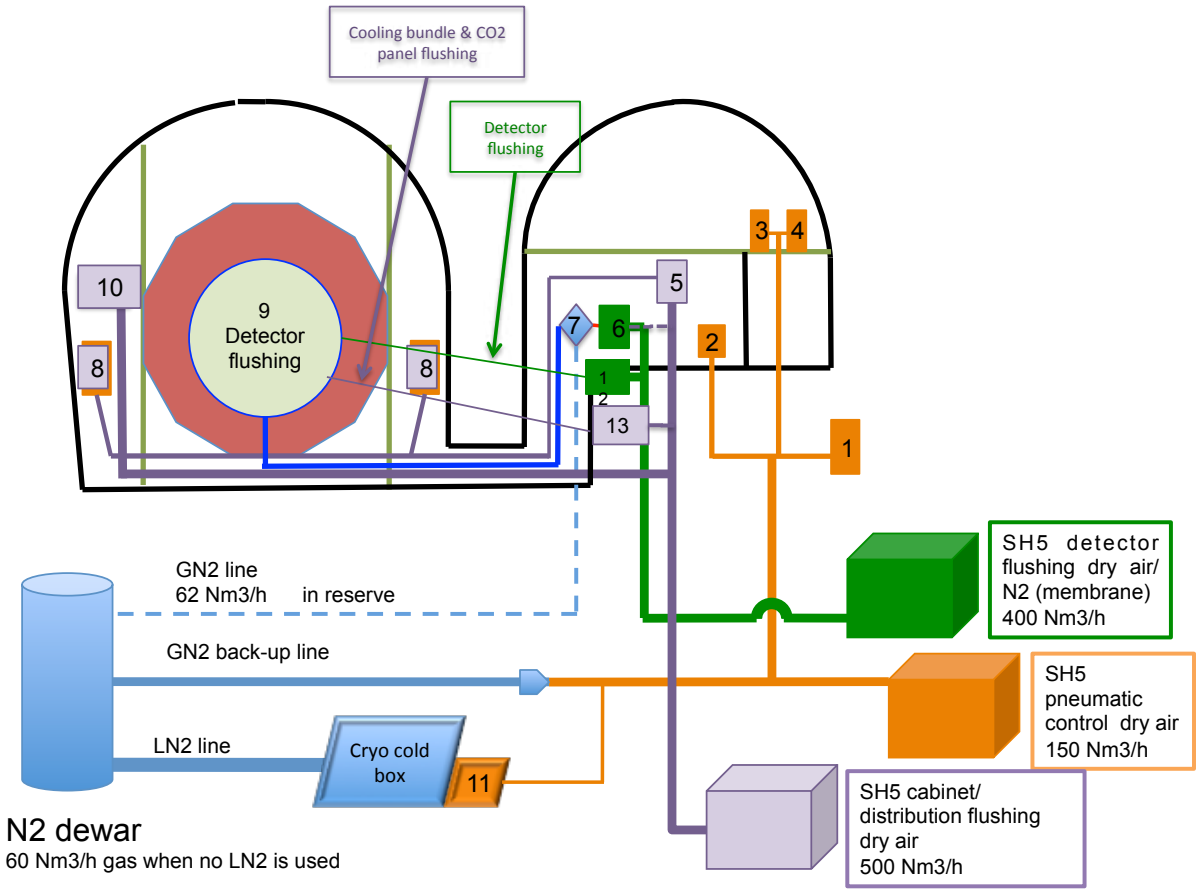


Figure 11.15: Layout of the Point 5 dry gas system after LS1 modifications: green: air/N_2 detector environment control, purple: dry air flushing, orange: pneumatic control.

lifetime.

11.8.3 Dry Gas system

As part of the Phase 1 upgrade, major changes to the dry gas system have been made to accommodate the intention of operating detectors at lower temperatures (notably the Tracker, now operating at $-15°C$). A new dry air/nitrogen plant (based on a membrane technique of locally extracting Nitrogen from air) now provides dedicated detector environment control (humidity and fire prevention) with diesel backup. A new dry air flushing plant, recognised during LS1 as needed to maintain a dry environment in the coolant distribution systems of detectors which operate cold (such as Tracker) has just been delivered by EN department and commissioned. Operation is imminent, as soon as the associated distribution pipework and cabinets are complete. The original dry air system will then be dedicated to pneumatic control and the use of nitrogen evaporated from liquid will be eliminated, except as an emergency backup. The layout of the system after these LS1 modifications is shown in Fig. 11.15.

For Phase 2, the requirement for dry gas will increase, to an extent dependent on detector choices for operating conditions. For instance, the new silicon-based endcap calorimeter (operating well below $0°C$) will incur the same sort of dry gas (air or N_2) requirements for detector flushing as the Phase 2 Tracker and will also likely require dry air flushing for the plants and manifolds of the CO_2 cooling system foreseen.

Although existing calorimeters may also in future wish to operate at lower than ambient temperature (but above 0°C), there is substantial margin in the dry air flushing capacity and the capacity released from the change in the Tracker coolant from fluorocarbon to CO_2 is likely to cover developing needs after LS3. It is therefore assumed that the main change to the dry gas system is likely to be another membrane plant covering new detector environment control needs.

11.8.4 Detector Gas systems

Apart from the addition (at sub-system expense) of one (or two) new systems serving the GE1/1 and GE2/1 endcap GEM detectors and the ME0 muon layer behind the Phase 2 calorimetry nose, no major upgrade of the CMS gas systems is believed to be necessary on performance grounds for the HL-LHC era. However, on the timescale of a decade, several items of common gas infrastructure will reach the end of their serviceable lifetime and need to be replaced. Notable amongst these are gas purity analysers, programmable logic controllers and the entire network using pressurized nitrogen for pneumatic control. Additionally, there are consequences of the CERN initiative to reduce the annual impact of greenhouse gases, released to atmosphere by experiments, by 80% before the end of LS3. This may require a substantial effort to recuperate (though not necessarily to re-use) the vented gas from gaseous detectors (notably from RPCs the CMS CSC system is already recuperating and re-using exhaust gas). In case alternative gas mixtures with a lower greenhouse impact are found to be viable, several major features of the gas mixing and distribution system may have to be changed for affected sub-detectors. Possible examples are mass flow controllers, flow meters, analysers, purification systems and pumps. At this moment, the resources needed can only be estimated very approximately.

11.9 Safety and Safety Systems

While the new Phase 1 detectors will present few new safety challenges and require only a limited amount of new safety infrastructure (smoke and leak detection, fire suppression etc), the re-equipment of the existing Detector Safety System (replacing un-maintainably obsolescent parts) and CMS-specific fire suppression systems (CO_2 bottles, water mist etc) and personnel safety items for a further decade of operation (plus the corresponding long shutdowns) will require substantial investment of resources. In addition, the radiation environment to which workers are exposed and the activation levels of equipment near the beamline, will increase substantially. (depending on cooling times arising from the LHC schedule). In the worst, and most likely, case, the dose rate encountered by personnel opening the detector at the beginning of LS3 will exceed that encountered at the beginning of LS1 by an order of magnitude. A further factor of 3-5 could be expected in the case of an emergency intervention necessary during Run 4 or later (assuming HL-LHC luminosity). It is therefore imperative to complete and augment the removable radioprotection shielding for personnel, begun and partially tested during LS1, which shields workers from activated areas, with the aim of reducing the effective dose rate by factor 10). In order to apply the ALARA principle effectively, it will also be necessary to rehearse certain tasks where a large individual or collective dose is liable to be accumulated. This will require accurate full-scale models to be built of certain areas where high activation or extended working time are anticipated.

CMS-specific simulations to map fluence, dose rate and activation levels through out the cavern will be needed both for the design of the beampipe and shielding and to plan interventions during technical stops or shutdowns. Subdetector specific simulations are assumed to be re-

sourced by other contributors to the BRIL project.

Short interventions in UXC during operations are strongly discouraged and will be increasingly complex in an activation environment worsening with improved LHC performance. It is therefore foreseen to engage in a substantial programme to augment remote monitoring and diagnostic capabilities inside the experimental cavern. (For example, the existing camera system is insufficiently radiation tolerant). The magnet fringe field provides a very unusual challenge to such diagnostics. In addition to and as a complement to, radiological shielding, hand-tooling designed to allow the operator some distance margin from activated surfaces will have to be designed or procured. In all of these developments, where applicable, CMS will work in collaboration with ATLAS and the LHC technical groups.

The increased volume of potentially activated material will require an increase in the size of the buffer zones where material is placed for radio-protection screening after removal from the experimental cavern. Similarly, additional cleaning precautions will have to be taken during shutdown activity to minimise the un-necessary exposure of waste materials to activation and to avoid the possibility (still considered quite slight in LS3) of contamination by activated dust.

11.10 Surface Assembly Buildings, Workshops, Laboratories, and Storage space

11.10.1 Overview

The broad philosophy behind CMS surface facilities remains the same as that agreed with DG-HSE, detailed to the CERN site committee in 2010 and largely implemented before or during LS1. Maintenance of detector elements exposed to LHC beam is preferably carried out at P5, where dedicated facilities for detector maintenance (operational support centre, OSC) have been created in the SX5 building (3585) formerly used for surface assembly of the detector. New or replacement elements are assembled in dedicated laboratory areas on the Meyrin site (Building 186 : silicon tracking systems and near-beam detectors, Building 27: calorimetry) and on the Prevessin site (Building 904: gaseous detectors for muon detection and endcap calorimeter pre-assembly, Building 892: barrel calorimeter module testing).

To meet the space requirements of the Phase-II programme, it is important first to obtain occupancy of the full original space allocated to CMS in 904 following the transfer of the CMS facilities in ISR-4 to HSE-RP. The only major area still to be made available is the laboratory 904-R-002, needed by the muon systems by the end of 2015. (currently this area is still occupied by TE department, awaiting new space promised to them). This transfer will free up areas of building 904 to prepare for calorimeter pre-assembly.

11.10.2 SX5 and the SX5 extension

The major new working space on CERN sites required by CMS for Phase-II is at P5, to allow for on-site detector refurbishment and pre-assembly on the scale assumed in the baseline planning. In an evolution of the concept first presented in 2010, the new construction now proposed is a multi-purpose extension of area 1000m^2 (approximate dimensions 22 x 46m x 14m) attached to surface assembly building SX5 (3585), shown in plan view in Fig. 11.16. This allows easy transfer of material between the buildings, through an existing external roller door in SX5, along with relatively straightforward extension of services to the extension. It also avoids known areas of unstable ground on the site, (for instance along the axis of SX5 to the east) and will

avoid perturbing the site roads during construction. The transfer of electro-mechanical and infrastructure workshops from SX5 into the extension, will make way for enlargement of the detector laboratories of the Operational Support Centre (OSC), which are equipped for work on activated detectors if needed.

Particular upgrade tasks currently envisaged for the enlarged OSC are the upgrade (electronics replacement) and testing of the 36 barrel ECAL crystal super-modules and the final assembly and testing, before lowering, of the new endcap nose calorimeters. The ECAL activity drives the critical path of LS3 planning, but the existing Operational Support Centre in SX5, designed to meet LS1 requirements, has no provision for work on the crystal calorimeters. Sufficient space will be available in the extended OSC to refurbish the ECAL barrel in parallel batches of 6 super-modules simultaneously, whilst, at the same time, conducting final assembly and testing of the Phase-II nose calorimeters.

In this scheme, it is hoped to minimize the infrastructure costs of the SX5 extension, which will not require laboratory facilities and services. The main services provision in the extension would be for the workshops. There is an added advantage in properly separating the workshops, presently the cause of much unwanted dirt and noise in the laboratory areas. The workshops would occupy 400m^2 of the extension. The remainder would be dedicated to logistics areas and storage space for sensitive tooling and spares, currently encumbering the workshops inside SX5, stored outside, or stored in barracks or tents with inadequate climate control and security.

A self-propelling gantry crane, of the type shown in Fig. 11.17, running on floor mounted rails and with a capacity about 20t, would cover the length of the SX5 extension building from an external access door up to the SX5 interface door. The SX5 extension building shell can therefore be lightweight. Transfer between buildings would be achieved via a trolley or a dedicated short traverse, again with a capacity of 20t. HVAC requirements can be similar to SX5, although more efficient modern systems such as those installed in the new SL53 office and visitor reception building will be investigated.

The preliminary resourcing assumption is that a CERN department (GS) will take care of constructing the extension. Part or all of the crane and the service infrastructure (power, HVAC, compressed air etc, installed by EN dept) will be attributable to the Phase-II common project. The possibility of a turn-key project will also be investigated.

11.10.3 Workshop machines

The CMS central workshop at Point 5 is crucial to effective detector and infrastructure work on the site and has also played a very significant role during LS1 in supporting sub-detector work, especially where needs for tooling or repair arise at short notice. Maintaining this capability is key to the common project planning for the next decade. Along with other project specific workshops at CERN, many of the machines are old (in our case inherited from LEP experiment site workshops) and increasingly difficult to keep maintained and in conformity with applicable regulations. Providing a workshop for the 2020s will require (besides the building described above) a steady investment in equipment over the next 10 years. The proper source of resources for this has still to be formally identified, but first indications are that this is most naturally categorized as a Phase-II upgrade.

11.10.4 Control room

The CMS control room in SCX5 was improvised from an intended rack room due to limits on construction funds, but has nevertheless functioned adequately. Changes are needed be-

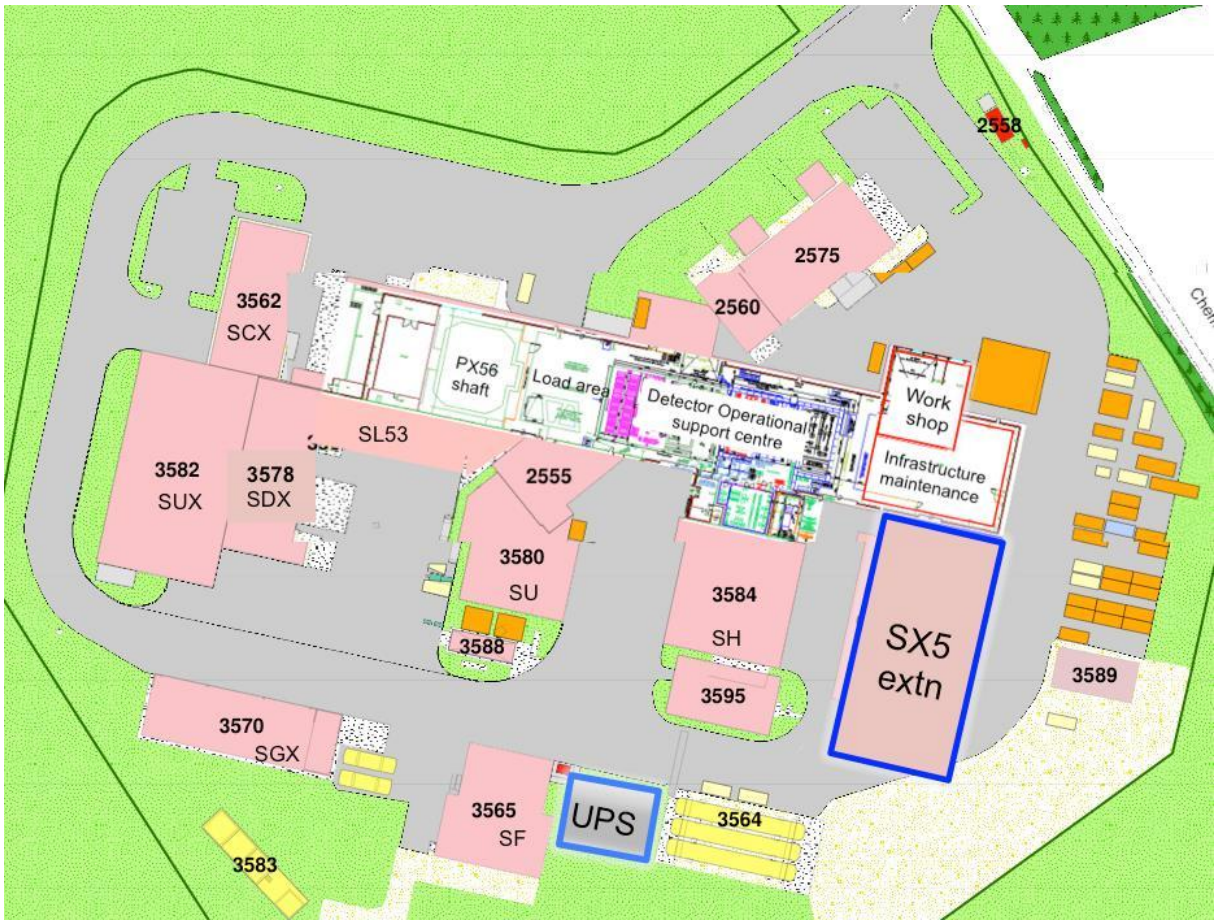


Figure 11.16: View of P5, Cessy site, showing disposition inside SX5, outline of proposed SX5 extension and a possible location for the UPS barn.

cause of the pressure to find space for additional detector workstations, the incorporation of the magnet and cryo control stations (previously separate) and the large visitor stream at CMS during both operation and maintenance. In addition, the presence of the 80kVA UPS (supporting control room functions) within the structural walls of the control room itself is identified as a hazard. The intention is to remove the UPS Rack room, currently in a partitioned off area within the control room volume, initially to a location alongside the existing UPS battery bank and eventually to the dedicated UPS barn. Control room functions will then be expanded into the liberated area, allowing space for a separate visitor observation corridor through the back of the revised control room. Although some CERN funding has already been provided and some contribution can be expected from DAQ funds (for consolidation of DAQ functions), charges to the Phase-II common project may arise.

11.11 Engineering and electronic integration and support

To maintain the general mission of mechanical and electronic integration over the next decade, it is first assumed that existing the existing core of CERN staff, plus staff paid from Collaboration funds, is maintained. Additional work to ensure coherence of Phase-II elements is estimated to require one additional staff mechanical engineer, one additional staff electrical project engineer and two additional designers over the next 10 years. Provision should also be made for the participation of one or more visiting associate engineers almost continuously

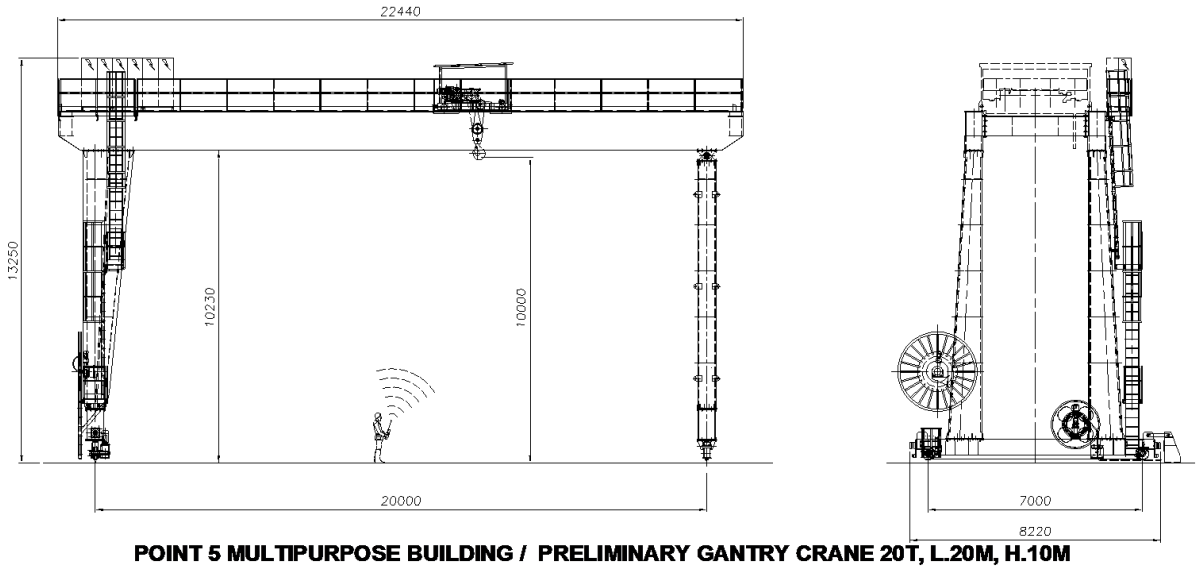


Figure 11.17: Point 5 Multipurpose Building / Preliminary Gantry Crane 20T, L.20M, H.10M.

throughout the Phase-II project. Experience also shows that oversight of host lab projects executed for CMS is necessary at a level about 10% of the cost. Support for the CERN-CMS heavy engineering team in the form of CMS-funded fellows or technical students will also be needed to deliver the support structure design, as well as mechanical and thermal simulation work, currently assumed to be provided to the Phase-II upgrade subsystem projects. As was done for CMS construction, specific agreements will have to negotiated made for in-kind institute contributions to integration work, as well as to specific engineering design and simulation support for heavy engineering aspects.

11.12 Technical support

The needs for commonly managed technical manpower for LS2 and LS3 can be scaled from 2009 and from LS1. In general, services work (water, power, control, leak detection etc) is assumed to be a central responsibility up to an on-detector patch panel agreed between each subsystem and Technical Coordination. Commonly managed work can go further at subsystem request and cost. The resources needed for YB0 stripping/re-cabling can be scaled from the central resources of 60,000 man-hours expended on cabling YB0 underground in 2007. Resources needed for stripping and re-cabling both YE1s are assumed to be about the same as required for YB0. All materials (cableways, pipes, cables, insulation), along with as much as is clearly factorisable of installation costs, will be assigned to the corresponding sub-detector budgets. Non-water cooling systems are included in common integration but design and installation up to the chilled water exchanger is a sub-system responsibility.

11.13 Opening/closing systems and tooling

Large scale hydraulic, pneumatic and winching systems and jacks are needed to open and close the shielding, forward calorimeters and the magnet yoke, and to change the logistic configuration of the major elements. Experience from LS1 is that the strand-jacking system used to move yoke elements supported on air pads is very labour intensive, requiring skilled labour which is

hard to re-inforce for shutdown periods. In addition, the length of the existing barrel transport beams (multiple air-pad structures used to move any one of the outer barrel wheels) leads to great difficulties in establishing optimum scenarios where both endcap disks and yoke wheels are accessible. It is planned to build shorter transport beams to more quickly access a range of logistic scenarios assumed in the LS3 planning. In addition, successful tests of long-stroke hydraulic jacks in LS1 indicate that these could be the basis of a more automated opening system, which would substitute for the strand-jacks in most maintenance operations and enable configuration changes to be mastered within the skilled team which can reasonably be maintained by CERN, re-inforced with some less skilled hired manpower (field support units) when needed.

The complete revision of the detectors forming the nose attached to the first endcap yoke disk would be very time and space consuming if done underground. The baseline planning assumes that each nose in turn is detached from its supporting disk as a single 300t element, manoeuvred via a cradle and transfer system under either the PX56 shaft (-z end) or through the TX 54 transfer gate to the PM54 shaft (+z end), at which point a heavy-lift "crawler" crane lifts the cradled nose up the shaft to the surface through temporary openings in the roofs of the SX5 building (-z end) or SDX building (+z end). The crawler crane would load the nose on a trailer for the short trip to the SX5 building where the same crane would discharge it into the extended RP workshop of the OSC, where any dis-assembly needed before disposal, or re-use, of particular elements would be done. It is assumed that the upgraded nose structures, comprising the combined calorimetry and additional muon detectors, would undergo final assembly and testing in the extended OSC. The crawler crane cradle, and other necessary rotation jigs and transport systems, would then be used in the inverse of the removal process, to sequentially install both new noses. The intensive use of the heavy-lift crane for short periods (in the current planning, one dedicated to removal of both noses and the second dedicated to replacement of both noses) minimises the disruption to both the surface and underground logistics and allows the planning to focus on the critical path driver, which is the removal, refurbishment and replacement of the 36 electromagnetic barrel super-modules.

Sketches illustrating the major steps in the nose removal/replacement sequence form Fig. 11.18.

The proposed replacement endcap calorimeters, being more granular devices, partly operating at sub-zero temperature, inevitably require more services, particularly power, cooling and dry air. This integration of these new services encounters obstacles within the nose system itself, which will not be discussed here, and also in the already excessive (80%) occupancy of the two YE1 cable chains (-x and +x sides), which serve all detectors on all disks in an endcap and have a combined capacity of $2 \times 3200\,\mathrm{cm}^2$. The proposed solution is to install new cable chains underneath the nearside X3 balcony of the UXC. These would provide an additional capacity of $640\,\mathrm{cm}^2$, which would be dedicated to CO_2 cooling and dry-air provision. The principle and location of this new cable chain is illustrated in Fig. 11.19.

11.14 Cost estimates

11.14.1 Categorization of Common Systems and Tasks

In this document, it is assumed that the general guidelines for the categorization of work as host lab, M & O A or upgrade common project are as follows:

(a) Work will generally be classified as a host laboratory responsibility if:

1. It is listed as such under the General Conditions for Experiments

2. In the case of assembly buildings, surface facilities or experimental caverns, it is generic in the sense of being (approximately) independent of the detailed specifications of the installed experiment.

3. A CERN department is taking responsibility for it as a consolidation or upgrade activity

It should be noted that the provisions in the General Conditions are basic and restricted to providing an overall framework complying with basic personnel protection requirements and satisfying very basic infrastructure and service needs. In particular, they do not cover equipping laboratory or experimental areas with infrastructure or services adapted to supplying detectors (specialised power, cooling, compressed air, network etc). Site-specific safety features, including those linked to the efficiency of operating or maintaining an experiment or assembly area, or protecting the experimental apparatus, are also generally not included. Nevertheless, CERN departments have generally made contributions beyond the basic minimum and reasonable precedents exist for all the attributions assumed herein.

(b) It is assumed (after informal guidance from the RRB scrutiny group) that work will generally be classifiable as Maintenance and Operation if it covers operation, repair or routine replacement of already installed equipment needed to operate at the ultimate LHC luminosity of $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ up to an integrated luminosity of around $500\,\mathrm{fb}^{-1}$ (the performance limits of the CMS detector after Phase 1 upgrade). This includes replacement of equipment that fails, or becomes inoperable, or un-maintainable, before the end of its design lifetime consistent with the above LHC performance targets.

(c) It is assumed, by inference from the above, that work will generally be classified as Phase 2 upgrade if it is a necessary part of equipping the experiment to deliver physics performance consistent with effective exploitation of the HL-LHC programme ([$\geq$ additional $2500\,\mathrm{fb}^{-1}$ (HL-LHC capability $3500\,\mathrm{fb}^{-1}$), luminosity $5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$(HL-LHC capability $7.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$)], while maintaining technical reliability and data-taking efficiency at the current high levels.

11.14.2 Cost estimates

A summary of the present CORE cost estimates for the infrastructure upgrades, the common systems and support, and the installation work is shown in Table 11.3.

Table 11.3: CORE Cost estimates for the Phase-II Infrastructure upgrades, common systems and support, and installation.

	Estimated CORE cost in MCHF (2014)
Magnet power and cryogenics	1.6
Beampipe	1.7
Infrastructure	6.1
Test Facilities	0.4
Surface facilities	0.3
Safety Systems	2.3
Electronics Integration	0.9
Engineering Integration	3.2
Technical support	8.9
Total Infrastructure, Common Systems and Support, and Installation	**25**

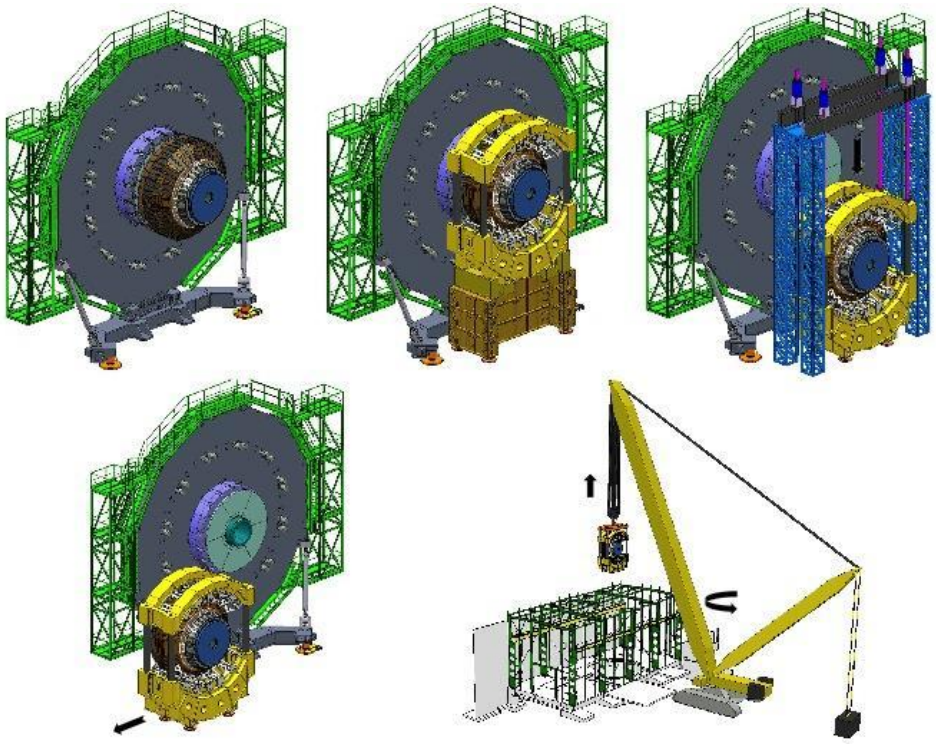


Figure 11.18: Major steps in the nose removal/replacement sequence.

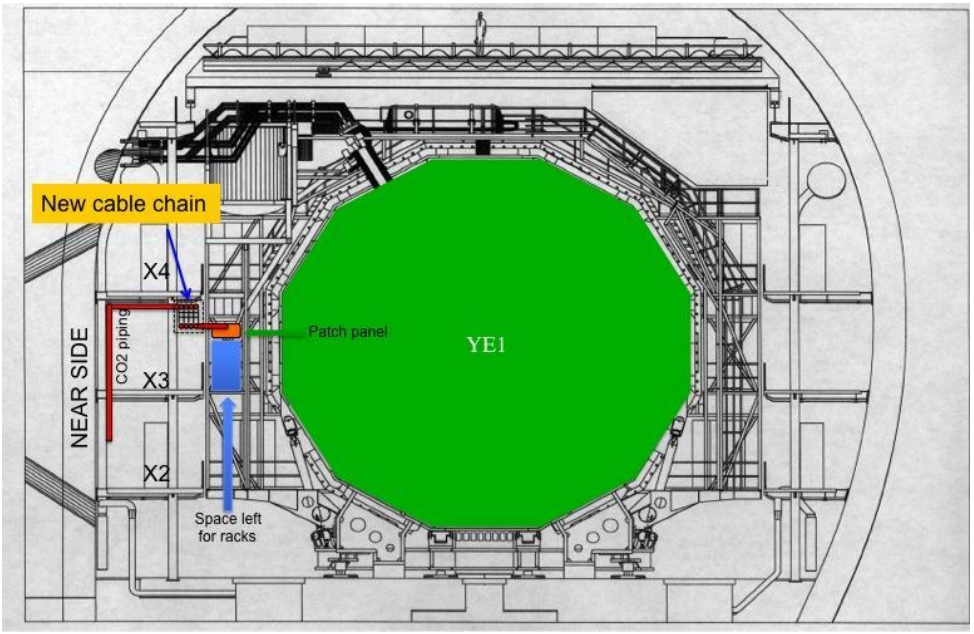


Figure 11.19: Proposed new cable chains serving YE1 disk from under near-side balcony of UXC.

Chapter 12

Project Cost and Schedule

12.1 Estimated Cost

The estimated costs for the upgrades are expressed as CORE cost. CORE costs were reported for the construction of CMS and for the current Phase-I upgrades. They represent the material replacement value of the installed equipment, in terms of the (*M*&*S*) (materials and services) for the production phase of the project. They include engineering costs incurred during production at a vendor or contractor, production fabrication and construction costs, QA and system testing costs during the assembly process, and transportation costs, integration and installation including costs associated with technical labour supplied at CERN for these purposes. They do not include *R*&*D* and prototype costs associated with developing the design, the cost of infrastructure and facilities at CMS institutions, or institution personnel costs. In addition to the systems installed in the experiment, the estimates include costs for test stands needed for commissioning, beam studies and firmware development. Spare parts to cover production losses and to ensure successful commissioning are included in the CORE estimates, while spares to support long term maintenance and operation (*M*&*O*) are not. In many cases, in order to deliver assembled or produced components of the upgrade with this CORE-value, additional costs are borne by institutions and funding agencies, including (*R*&*D*), design engineering and other institution labor. Costs are reported here in 2014 CHF, with no correction for inflation to future years. No contingency is included in the estimates. In most cases cost estimates for the dominant costs are based on information provided by vendors. Other costs are typically based on the original construction of the experiment or the current Phase-I upgrades. Where the design concepts are not sufficiently advanced, for example the L1 Trigger, the estimates are based on material costs and a number of units scaled from the present detector, with a cost per unit based on Phase-I.

A mid-level cost breakdown for each element of the upgrade scope is provided in the preceding sections. Table 12.1 below provides a summary for the entire upgrade, with a total CORE-cost estimate of 265 MCHF.

12.2 Project Timeline

An outline of the project timeline is illustrated in Figure 12.1. The tracking system, endcap calorimetry and track-trigger upgrades require significant *R*&*D*. The timeline for the projects anticipate about 2 years for *R*&*D* to finalize the designs ahead of the TDRs, and a further 1-2 years for prototyping to develop production grade components. The construction is then planned for 4-5 years with installation during LS3. Detailed schedules will be presented in the TDRs which are anticipated from early 2017 to late 2017 except for the Trigger and DAQ upgrades. In order to maximize the benefit from advances in computing technology and cost

Table 12.1: Summary of CORE costs for the CMS Phase-II Upgrade.

CORE cost estimate	**MCHF (2014)**	**Further Details in Table**
Pixel Detector	23	2.5
Outer Tracker	89	2.5
Tracking System	**112**	
EB electronics	10	3.9
HB scintillators	1	3.9
Endcap HGC+BHE	64	3.10
Calorimeters	**75**	
DT and CSC electronics	10	4.4
Muon stations:GE11,GE21, RP31 and RP41	10	4.4
Muon extension ME0	5	4.4
Muon Systems	**25**	
Beam Monitors and Luminosity	**4**	5.4
L1 Trigger	7	6.3
HLT	11	7.3
DAQ	6	7.3
Trigger and DAQ	**24**	
Infrastructure, Systems and Support, Installation	**25**	11.3
Total	**265**	

reduction, it is expected that these TDRs and project plans will be finalized nearer to LS3, in 2020.

The TDRs will include project schedules and milestones to allow the necessary management oversight. The new detectors and unprecedented high PU conditions require significant development and tuning of reconstruction algorithms. While initial results have been represented in this document, the work will continue in tandem with the *R&D* to optimize the design and to prepare CMS to commission the upgraded detector.

12.3 Cost Profile

The estimated profile for CORE expenditures is illustrated in Figure 12.2. Expenditures on construction will start in 2017 with a ramp-up through 2018 and a broad distribution through 2024. With further development of designs for the TDRs, the cost estimate and profile for each project will be updated.

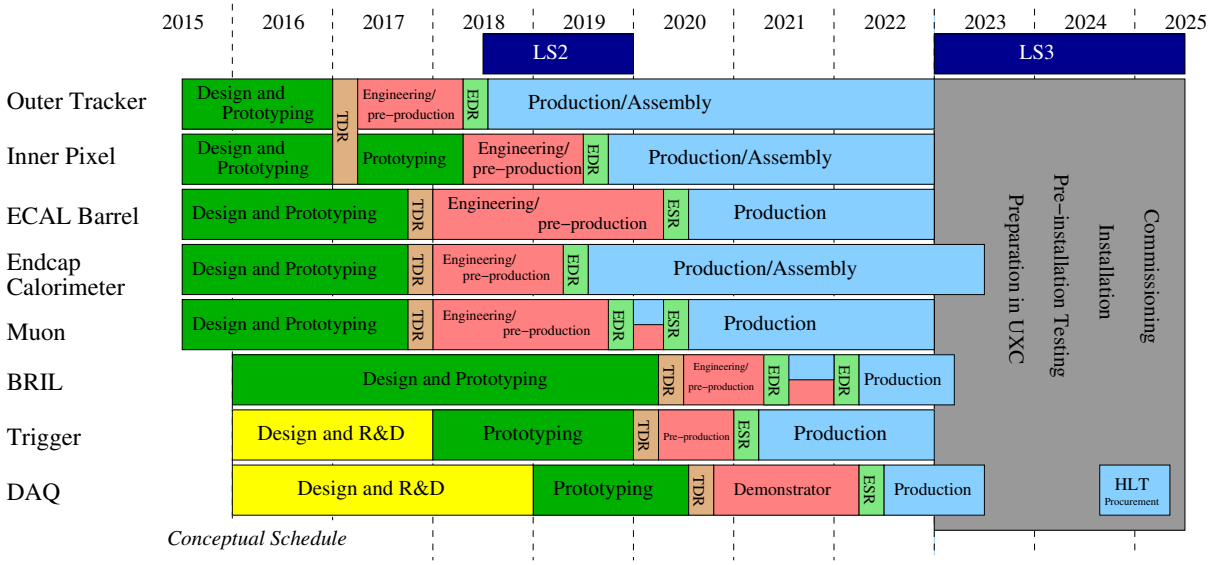


Figure 12.1: Outline of the Phase-II Timeline. Each project will include a detailed schedule in the respective TDR.

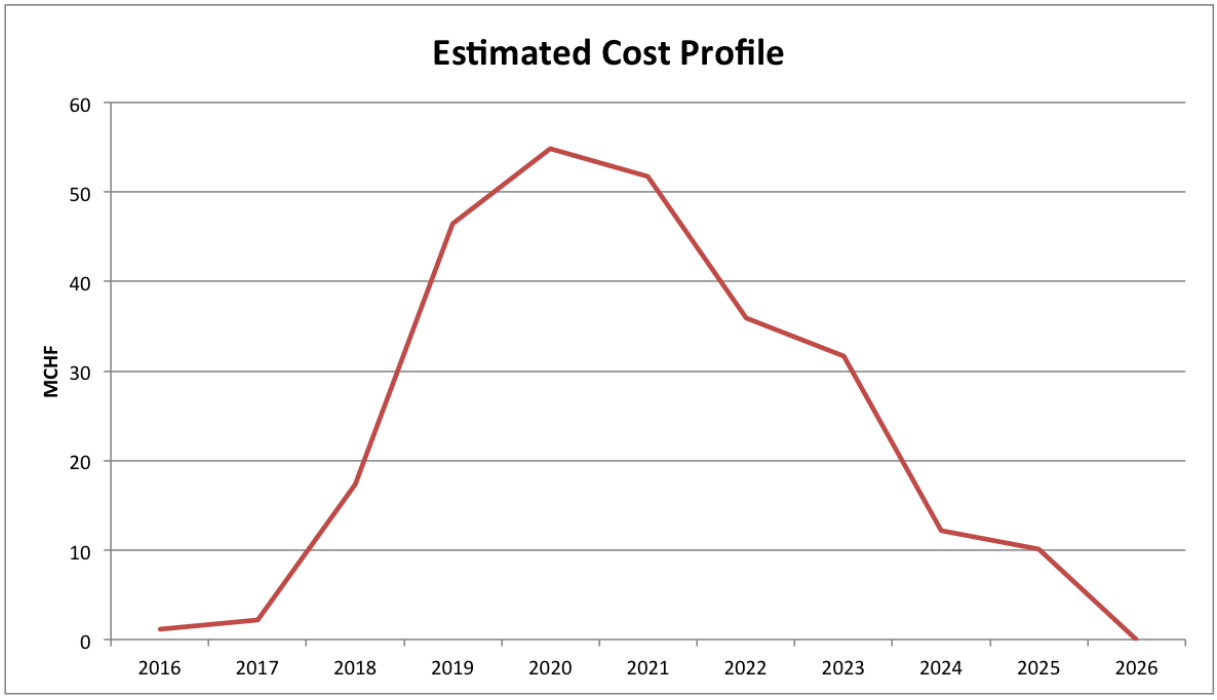


Figure 12.2: Anticipated CORE spending profile with the installation occurring during LS3 in 2023-2025.

Appendix A

Radiation simulation tools, models and parameters

A.1 FLUKA and MARS codes

BRIL radiation simulations are performed with the FLUKA 2011.2b.6 Monte-Carlo program [13], [14] and the MARS'109 code [12]. In FLUKA, single events are simulated. For most CMS simulations, a proton-proton collision is used as primary event, and all particles are transported until they decay or fall under a predefined energy cut-off. However, typically FLUKA is not used for the purpose of recording single events or for studying the tracing of single particles. The output is usually averaged over all simulated primary events and normalized per primary event. In FLUKA all secondary particles will be generated and in MARS (an inclusive code) some fixed number of secondary particles is generated in one step, with weights according to averaged multiplicities of such particles. MARS simulations are therefore typically faster. For both codes, the output is typically a set of averaged values, e.g. of fluences or doses. However, FLUKA has many more user defined output formats, specific features for activation studies and a more extensive low energy neutron group library. Whereas MARS has been used for several specific studies, most of the results presented in this technical proposal are based on FLUKA calculations. A fuller description of the FLUKA output options, models and simulation assumptions is therefore presented in this Appendix. The MARS CMS model contains many of the same features as the FLUKA model. However, some of the regions are represented in less detail in order to maintain the advantage of faster simulation times.

A.1.1 FLUKA output options

A concept that is used in many of the FLUKA user-defined outputs is the so called track length estimation, where the path length of a particle, passing through a specified volume, is recorded. The track length is normalized to the detector volume, resulting in a fluence value per unit area. This method is more suitable than e.g. counting particles, as it properly weights particles that have a steep angle, only scrape the detector volume, decay inside the volume or are created in the middle of the volume.

Many different user defined outputs, or so called scorings, are available in FLUKA. Here the ones that are used for plots presented in the report are described:

- **USRBIN:** The volume of interest is overlaid with a binning mesh that can be Cartesian, in cylindrical coordinates, or as a special case bound to a geometry region. This scoring is used to obtain particle fluences by calculating the track length density, or doses using the energy deposition. The obtained 3D distributions can be projected to two dimensions and shown, e.g., as a 2D flux map.

- **USRTRACK:** The output is the fluence as a function of the energy. This estimator has to be assigned to a certain geometry region and cannot be scored at an arbitrary location.
- **USRBDX:** The fluence or the current through an interface between two geometry regions are given. For the fluence calculation the particles are weighted with the cosine of their direction with respect to the normal of the boundary, for the current no weighting is done. Double differential distributions in energy and crossing angle are given. This estimator has to be assigned to particular geometry regions and cannot be scored at an arbitrary location.

Scorings are configured for different particle types or groups of particle types, e.g. charged hadrons, all neutral, neutrons or electrons, hereafter referred to as particle type. Using the energy deposited in the material, quantities like energy deposition, absorbed dose and NIEL are determined. Using the track length scoring technique, with the use of additional predefined functions where necessary, quantities such as fluence and ambient dose equivalent are determined.

A.1.2 FLUKA and MARS geometry construction

A FLUKA or MARS geometry is created by combining regions of basic geometric shapes. These shapes can be finite objects such as spheres, boxes or cones or infinite elements such as planes or infinite cylinders. The complexity of the geometry that can be created is limited by the number of objects or elements that are reasonably used. A large number increases the computing time. Hence the geometry models are a simplification of the reality, composed of the objects and elements and the material within them. It is, however, important that the correct material composition and the material densities are used to account for the correct total mass. Only then calculations of particle shower cascades become reliable. For activation studies trace elements that have the potential to get highly activated should be taken into account. Cut-off and transport parameters are configured per geometry region.

A.1.3 FLUKA models of the CMS Detector and Cavern

For the TP baseline model (v.3.7.0.0) key components are taken over from the Phase-I Run-II detector model [18]. These are described in the following subsection. Further subsections cover specific changes to form the TP baseline Phase-II detector and upgrade options for the forward calorimeters. The key components of the TP baseline model are shown in Fig A.1(A-C). The Phase-I (Run-II) model comprises a more detailed geometry, than the one used in simulations presented in various CMS subdetector TDRs [83, 101].

A.1.3.1 Key components from the Phase-I model used for the TP baseline model

The electromagnetic calorimeters are represented by a volume filled with $PbWO_4$ at the location of the crystals. The electronics volumes are filled with an appropriate average material. The preshower detector is composed of layers of silicon, absorbers, cooling screens and neutron moderators. The support cone is also included.

The hadronic calorimeter is implemented in layers of brass and scintillator, shaped in cylinders and discs. The HB uses brass with a 70 % copper and 30 % zinc mixture, interleaved with a plastic scintillator. The HE has a detailed brass description, which is composed of 75 % copper, 24.5 % zinc and various trace elements. The brass is interleaved with layers of air, scintillator and reflective material.

The vacuum tank is implemented and the CMS magnet is integrated as an aluminum tube.

The yoke is included according to the available technical drawings. The DT muon chambers in between are modeled as cylinders. They consist of an average honeycomb material, approximated by low-density aluminum. The DT detectors consist of aluminum layers with gas in between. The CSCs are composed of alternating layers of gas and Lucite with G10 plates in between. The RPCs are composed three layers of bakelite with gas in between. The end cap muon shielding is composed of borated polyethylene and lead with antimony according to these production drawings.

The HF detector active area is represented by a cylindrical volume, with a shield around it. The HF is modeled as an average material with a density of 7.51 g cm^{-3}, composed mainly of: 97 % Fe, 1 % Si and 1 % O. The area behind HF (where the fibers are located in reality) is filled with air.

The forward shielding is modeled in detail, including the cracks between collar shielding and rotating shielding and the alignment holes. A model for the blockhouse containing the FIN, TAS and Q1 is included. The cavern is taken into account according to design drawings, however displacement of elements by the magnetic field is not considered. Uncertainties of the particle fluxes in the cavern are driven by imperfections of the forward shielding model.

The cavern model includes the concrete wall (with an approximated shape), the cavern floor, the YB feet, the HF risers and the electronics racks on the balconies. Other bulky elements in the cavern are not modeled yet, but will be included in future models. For example the YE feet, the electronics and cables at X0, the electronics towers on top of CMS, and the balcony support structures are currently not implemented.

Models of TOTEM T1 and T2 are present in the Phase-I (Run-II) geometry, however they are removed for the TP baseline geometry.

A.1.3.2 The TP baseline model

The specific changes to the Phase-I model made to form the TP baseline are listed below:

A.1.3.2.1 Pixel Detector and Tracker No detailed information is available yet for the Phase-II Pixel detector. The model of the pixel barrel detector for Phase-II, as shown in A.1 (C) (bottom left) is based on the model and material budget for Phase-I (Run-II). It consists of detection layers comprising of silicon sensors, averaged material of support tubes, electronics and cables. The forward disks are added as layers of silicon at the locations proposed for Phase-II. Cables and cooling lines are also included.

The Tracker geometry is implemented based on the current proposal for Phase-II. Each tracker layer comprises of silicon sensors supplemented by an average material accounting for electronics, support structure, cooling and cables, as shown in Fig. A.1 (C).

A.1.3.2.2 Tracker neutron shielding The preshower detector has been removed. At its location, a neutron shielding has been added, as shown in Fig. A.1 (D), to protect the Tracker from neutron albedo[1], created inside the Endcap calorimeters. As shielding material borated polyethylene is used with a boron concentration of 5 %. The thickness of the shielding is 18 cm, to fill the whole preshower volume. An 8 mm thick support cone has been included under the shielding and the calorimeter. The lower radius of the shielding is 28.8 cm and it covers the calorimeter up to a radius of about 130 cm.

[1] Evaporation neutrons emitted isotropically from the spallation reactions in the calorimeter

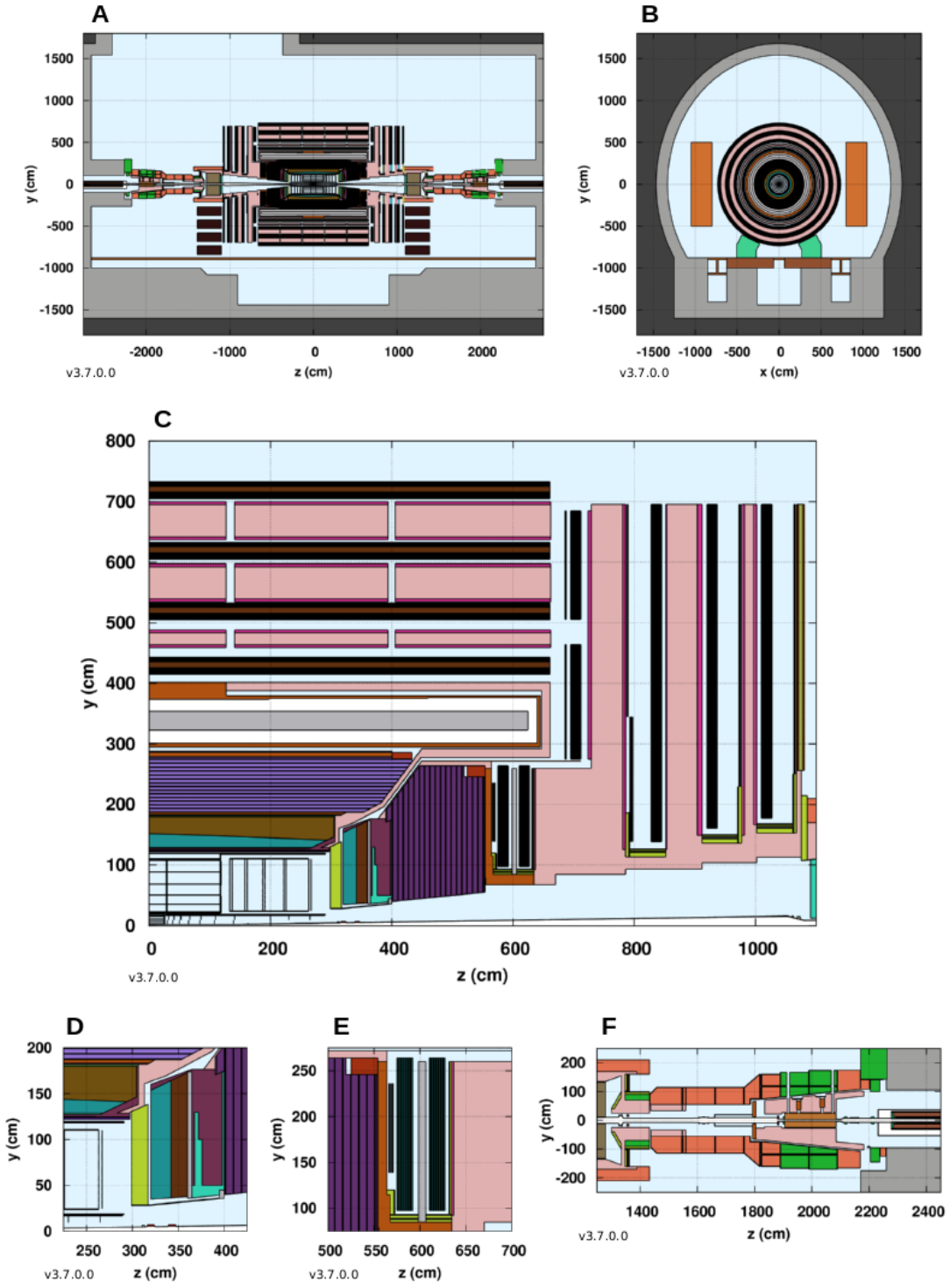


Figure A.1: The CMS FLUKA model used for simulations of the HL-LHC; TP baseline model. **A**: A vertical slice at X = 0 through the CMS detector and cavern. The shaft, cavern walls and floor are clearly visible. **B**: A vertical slice in the X-Y plane at Z = 0. The racks on the balconies are shown in orange on each side of the detector, and the YB feet in light green. The detector layers are modeled as cylinders. **C**: A vertical slice in the X-Z plane, showing the barrel and the Endcap. **D**: A slice in the Y-Z plane, showing the borated polyethylene shielding in the Endcap region in light green in front of the Phase-I Endcap Calorimeter. **E**: The new GEM detectors. **F**: Rotating shielding, TAS and blockhouse region, including part of the new inner Triplet magnets.

A.1.3.2.3 GEM detectors GEM detectors will replace the RPCs at the ME1/1 and ME2/1 locations. Two layers of triple GEM detectors are implemented at their expected location as shown in Fig. A.1 (E). One layer of the GEM detector consists of 7 mm of gas, with copper covered Kapton layers inside, and an aluminum housing with a total of 8 mm aluminum per layer.

A.1.3.2.4 Beam pipe and TAS The CMS beam pipe has been changed in material from stainless steel to aluminum (AL2219), since in aluminum fewer particle showers and lower activation are created. To compensate for the lower stability, the thickness has been increased by 50 %. As a requirement of the LHC to facilitate HL-LHC beam optics, the radius of the CASTOR-Totem 2 (CT2) beam pipe, the TAS and the beam pipe in the Triplet magnets have to be increased. The inner radius of the CT2 beam pipe has been increased to 7.25 cm. All flanges and bellows are increased in radius accordingly. The inner radius of the TAS has been increased to 2.85 cm. The geometry model of the Q1 magnet of the Triplet is more detailed than the present Phase-I model. A model of the HL-LHC magnets has been provided by the FLUKA team. The beam pipe elements between the TAS and the Triplet magnets are not defined yet. Hence, the beam pipe has been increased in radius to match the beam pipe of the Triplet ($R_{inner} = 6.9\,cm$) magnets and the radial position of the beam pipe elements have been scaled accordingly. The CT2 beam pipe and TAS region is shown in Fig. A.1 (F).

A.1.3.2.5 Endcap Calorimetry The FLUKA geometry of the endcap calorimeters are shown in Fig A.2. The model (v.3.7.2.0), as shown Fig. A.2 (left), includes:

- Two volumes representing the electromagnetic and the hadronic sections of the silicon-tungsten portion of the calorimeter. Different average materials, based on the latest detector configuration, are assigned to the volumes.
- A thermal screen of 3 cm thickness surrounds the HGC detector.
- The support cone for the tracker shielding is adapted to fit the outline of the calorimeters.
- The model of the HE detector is identical to that in Phase-I, except that it is shortened in length.

A.1.3.2.6 Volume behind the endcap calorimetry The volume behind the endcap calorimeters consists of:

- Support wedges, which are used to mount the Endcap calorimeters on the YE1 wheel.
- The ME0 detector, consisting of 6 layers of GEM detectors.
- Shielding below ME0 consisting of a layer of borated polyethylene and a layer of lead with 4 % antimony.
- ME0 stainless steel insertion rails.
- A volume filled with an average material accounting for ME0 electronics and cabling.

A.1.4 Simulation parameters

The primary proton–proton collision is simulated at 7 TeV energy per proton. The protons collide under a crossing angle of 590 μrad. The events are randomly distributed within 5 cm along the Z axis.

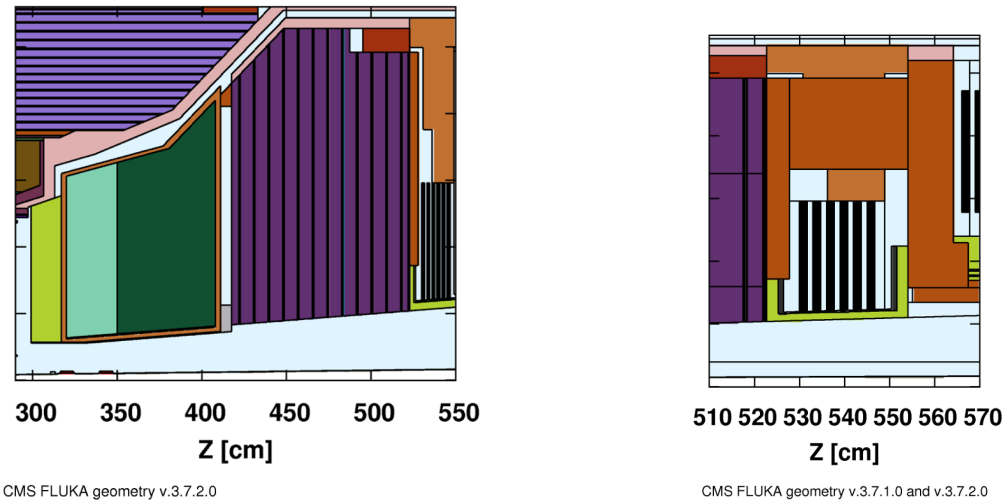


Figure A.2: (Left) The FLUKA model of the endcap calorimetry. (Right) The FLUKA representation of the ME0 region.

Particles are transported until they decay or until their energy falls below their transport cut-off. In this case, their remaining energy is deposited on the spot. All hadrons and muons have an energy cut-off of 1 keV, except neutrons, which are transported down to an energy of 0.01 meV. Electrons and photons are assigned different cutoffs in different materials. In most materials 5 keV for photons and 30 keV for electrons are applied. In heavy materials, the cut-offs are increased. In the beam pipe 10 keV for photons and 100 keV for electrons and in the calorimeters 30 keV for photons and 300 keV for electrons, respectively, are used. In the TAS the cut-offs are 300 keV for photons, and 3 MeV for electrons.

A.1.5 Scoring grids

All the results of the particle fluences are scored in an R-Φ-Z grid. Depending on the location, the cell sizes vary. The main scoring cell sizes are:

- A grid, covering the full cavern volume, with 10 cm grid, uniform the in R and in Z coordinates and 16 sectors uniformly positioned in Φ. The particle types scored are: All particles, charged particles, neutral hadrons, charged hadrons, neutrons, protons, charged pions, all hadrons with $E > 20$ MeV, electrons + positrons, muons, photons, doses, NIEL and 1 MeV neutron equivalent in silicon.
- A grid covering the central sub-detectors including HB and HE. The grid size is 1 cm in R, 2 cm in Z and there is no segmentation in Φ. The particle types scored are the same as for the full cavern scoring.
- A grid covering the Pixel detector volume with a grid size of 1 mm in R for the range $R < 20$ cm and 1 cm in Z up to the tracker bulk head. There is no segmentation in Φ. The particle types scored are: Charged particles, energy dose and 1 MeV neutron equivalent in silicon.

For specific sub-detector studies, various other scoring options have been simulated. The details of these scorings are given in the respective chapters.

A.1.6 Radiation simulations uncertainties

The statistical uncertainties quoted are the error of the mean values, which are obtained by averaging over many simulation cycles. The simulations are performed with a sufficiently large number of primary events such that the statistical uncertainties are negligible, except in certain regions of very low particle fluence. There are two main contributions to the systematic uncertainties for all results: uncertainties due to the event generator (which is described in the Introduction Chapter, Section 1.4.1), and uncertainties due to the simplification of the CMS FLUKA model in terms of volumes and materials. The uncertainties vary from region to region and are expected to be lower in the central detector, where the particle flux is less dependent on the details of the detector structure, but higher where particles originate from long secondary cascades, being relatively strongly influenced by the geometry model. An additional systematic error is introduced by the normalization of the quantities. The inelastic collision cross section at $\sqrt{s}$ = 14 TeV is unknown. It is predicted by various event generators. Being the only event generator tuned with LHC data, the results of EPOSLHC is used, which predicts 80 mb inelastic cross section [16].. The value of the inelastic cross section can be updated, once a measurement has been performed during Run-II operation.

A.1.7 Simulations of sources of Machine Induced Background

The sources of the machine induced background (MIB) particles from the HL-LHC beam are simulated based on nominal HL-LHC beam optics, beam parameters and vacuum conditions. The type and kinematic parameters of the particles entering the CMS cavern at the interface plane (Z = 22.6 m from the IP) are written to a file. These simulations were performed by the LHC FLUKA team (CERN/EN/STI) [274] and the result files are provided to CMS. In the second step these particles are transported into the CMS cavern using the CMS FLUKA geometry.

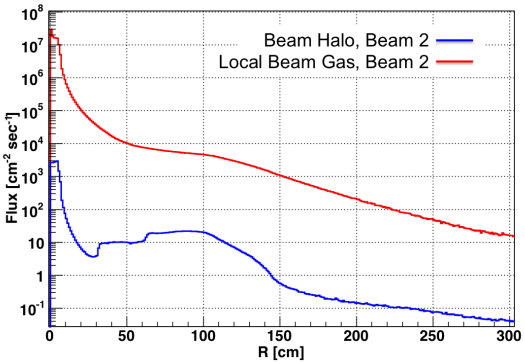


Figure A.3: The flux of all particles entering the CMS/LHC interface plane of Z = 22.6 m from the IP as a function of the radial distance from the beam axis. red: beam gas interactions. blue: beam halo originating from interactions on the TCTs in IR5 B2.

Beam gas interactions are interactions of the LHC beam particles with residual gas in the LHC vacuum chambers. The residual gas consists mainly of hydrogen, carbon or oxygen atoms. Interactions with all three nuclei are separately simulated at random locations up to 140 m upstream from the CMS IP, where the absorber of neutral particles, the Target Absorber Neutral (TAN) is positioned. Since the vacuum pressure is not constant over the long straight section, a weighting factor is applied to the secondary particles of each primary proton-nucleus interaction, depending on the vacuum pressure at the location of the initial interaction. Preliminary HL-LHC pressure profiles for each nucleus were made available by the LHC vacuum group

in a start-up situation, the expected worst case, and also in a conditioned situation, after the vacuum is stabilized. The difference in the particle fluences of the two situations is expected to be roughly a factor of 10. In Fig. A.3 (red line) the particle flux expected from beam gas interactions is shown as a function of the radial distance from the beam axis for $Z = 22.6$ m from the IP. Only beam 2 is simulated at present for a conditioned machine.

The sources of beam halo interactions with the collimation system are estimated using SixTrack [275] simulations. Almost all protons lost in the collimation system hit the primary collimators at point 7, but a small fraction of protons escape this system and hit the tertiary collimator (TCT) located about 150 m upstream of CMS. The simulated number of proton hits at the TCT per proton hitting the primary collimators is 3.5×10^{-5}. With SixTrack, a map of the locations, where primary protons hit the TCT, is obtained. This is used as input for the FLUKA simulation of the long straight section. In order to normalize the particle rates, a hit rate at the primary collimators is needed. A beam lifetime has to be assumed to calculate this quantity. The HL-LHC beam lifetime is not yet known and any number between 10 h and 200 h may be reasonably assumed. During Run-I operation the beam lifetime due to collimator hits was of the order of 80 h [274] and hence this number is used to normalize the beam halo rates presented here from beam 2. The predicted flux is shown in Fig. A.3 (blue line). Fig. A.3 also shows that the dominant source of MIB are the beam gas interactions.

Appendix B

Glossary of Special Terms and Acronyms

- 2S modules = modules composed of two superimposed strips sensors of approximately 10 x 10 cm^2, mounted with the strips parallel to each other.
- 3D sensor = sensor where the electrodes penetrate as columns the silicon bulk.
- μFDL = Micro Final Decision Logic.
- μGT = Micro Global Trigger.
- μGTL = Micro Global Trigger Logic board.
- μHAL = Micro Hardware Abstraction Layer.
- μHTR = μTCA Hadron Calorimeter Trigger and Readout Card.
- μTCA = Micro Telecommunications Computing Architecture.
- Al-CF = Carbon Fibre Reinforced Aluminum.
- ALARA = "As low as reasonably achievable" principle for radiation and environmental safety
- AM = Associative Memories.
- AMC = Advanced Mezzanine Card (from the ATCA specification).
- AMC13 = A μTCA data concentration and clock distribution card specifically designed for the CMS experiment.
- ANSYS = Engineering Simulation Software.
- APD = Avalanche Photodiode.
- ASIC = Application-Specific Integrated Circuit.
- ATCA = Advanced TeleCommunications Architecture.
- B-Go = Fast control signals distributed via the TTC system.
- B-HG = Back hadronic section of HGC.
- BC0 = Bunch Crossing Zero.
- BcN = Bunch crossing Number.
- BCR = Bunch Counter Reset.
- BDT = Boosted Decision Tree, is a multivariate analysis method.
- BEE = Back End Electronics
- BER = Bit Error Rate.
- BTI = Bunch and Track Identifier.
- BTIM = BTI Module.
- BX = Bunch crossing.

- BXN = Bunch Crossing Number.
- C4 bump bonding = industrialized flip-chip bonding.
- CALICE = Calorimeter for ILC. A collaboration to investigate calorimetry for the ILC and CLIC.
- CBC = CMS Binary Chip.
- CC = Charge Collection.
- CCE = Charge Collection Efficiency.
- CCITT-CRC = Standard cyclical redundancy check algorithm developed by the International Telecommunications Union.
- CCM = Clock and Control Module.
- cDAQ = central Data Acquisition.
- CFEB = Cathode Front-End Board, part of the CSC system.
- CFRP = Carbon Fibre Reinforced Polymer.
- CIC = Concentrator IC.
- CLCT = Cathode Local Charged Track (cathode view muon stub), part of the CSC system.
- CLCT/TMB = A board containing both CLCT and TMB functions.
- CLIC = Compact Linear Collider.
- CMC = Common Mezzanine Card.
- CMOS = Complementary Metal-Oxide-Semiconductor.
- CMSSW = Compact Muon Solenoid Software, is the CMS experiment software package.
- COTS = Components Off The Shelf.
- CRC = Cyclical-redundancy check, a widely-used family of algorithms for identifying data corruption.
- CSC = Cathode Strip Chamber.
- CTE = coefficient of thermal expansion.
- CTP7 = Calorimeter Trigger Processor 7 card, featuring the Xilinx FPGA Virtex-7 chip.
- DAQ = Data AcQuisition.
- DAQ2 = Upgrade to DAQ system during LS1.
- DAQMB = Data Acquisition MotherBoard, part of the CSC system.
- DB = Database.
- DC = Direct Current.
- DC-DC = Devices to convert DC voltage from one level to another.
- DCC = Data Concentrator Card.
- DCS = Detector Control System.
- dd-FZ = Silicon float-zone material with a deep diffusion zone at the back.
- DDR = Double Data Rate, the transfer of data on a synchronous digital link at both the rising and falling edges of a clock.
- DDU = Device Dependent Unit.

- DPM = Dual Port Memory.
- DQM = Data Quality Monitoring, a CMS system that monitor quality of recorded data.
- DSP = Digital Signal processing.
- DT = Drift Tubes.
- DTBS = Drift Tube Barrel Sorter.
- DTBX = Drift Tubes with Bunch Crossing identification capability.
- DTC = Data, Trigger & Control Board
- DTCCB = Drift Tubes Chamber Control Board.
- DTCI = Drift Tubes Control Interface.
- DTCM = Drift Tubes Master Control.
- DTTF = Drift Tube Track Finder.
- DTWCB = Drift Tubes Wheel Control Card.
- DTWS = Drift Tube Wedge Sorter.
- e-link = Electrical link.
- E-HG = Electromagnetic section of HGC.
- ECAL FG Veto = A bit to veto non electromagnetic showers from ECAL fine grain crystal data.
- EE = Endcap Electromagnetic Calorimeter
- EEPROM = Electrically-Erasable Programmable Read-Only Memory.
- EG = Electron-Gamma, normally used in the context of a trigger algorithm that will select events with electron or gamma candidates.
- EIC = Electron Isolation Card.
- EID Card = Electron Identification Card.
- EISO ASIC = Electron Isolation ASIC.
- EJ309 = Liquid Scintillator made by Eljen Technologies, Inc
- EM = Electromagnetic
- ES = Preshower Detector currently deployed in CMS
- ET = Scalar sum of transverse energy components over the calorimeter systems.
- ETmiss = 2-vector sum of transverse energy over the calorimeter systems.
- ETTF = Eta Track Finder.
- EU = Extrapolation Unit.
- EVB = Event Builder.
- EvN = Event Number.
- Ex,Ey = Components of ETmiss.
- FBT = First Best Track.
- FDL = Final Decision Logic.
- FE = Front-End.
- FEA = Finite Element Analysis.
- FEB = Front End Board.

- FEC = Forward Error Correction *or* Front End Chip (an ASIC).
- FED = Front End Driver.
- FEE = Front-End Electronics.
- FG = Bit characterising the fine grain profile of energy within the trigger tower.
- FIFO = First-In First-Out logic device that can be used to store and retrieve data.
- FIN = Fixed Iron Nose.
- FINOR = Final OR.
- FIR = Finite-Input Response Filter.
- FJet = (or FwdJet) Forward Jet, normally used in the context of a trigger algorithm that will select events with forward jet candidates.
- FMM = Fast Merging Module.
- FPGA = Field-Programmable Gate Array.
- FPPA = Floating Point PreAmplifier.
- FSB = Final Sorter Board.
- FSM = Finite State Machine.
- FSU = Final Selection Unit - sorting element of CSC Sector Processor.
- FTE = Full Time Equivalent, is a unit to measure workload of a task.
- FTP = Foiled Twisted Pair, is a type of wiring in which two conductors of a single circuit are twisted together for the purposes of cancelling out electromagnetic interference.
- FTT = Fake Track Tagger.
- FZ = Float-Zone silicon material grown via the float-zone process.
- GB = Ghost Buster.
- GbE = Gigabit Ethernet, is a term describing various technologies for transmitting Ethernet frames at a rate of a gigabit per second.
- GBSCA = Slow control chip for the GBT system.
- GBT = Gigabit Transceiver Project developed at CERN, source of the GBTX and associated chips.
- GBTX = Gigabit Transceiver ASIC developed at CERN.
- GCT = Global Calorimeter Trigger.
- GEM = Gas Electron Multiplier, is a type of gaseous ionization detector used in nuclear and particle physics.
- GIF = Gamma Irradiation Facility: area with high-intensity radioactive gamma source.
- GLIB = μTCA card developed by the CERN microelectronics group for testing and operating with the GBT link.
- GMT = Global Muon Trigger.
- GPS = Global Positioning System.
- GT = Level 1 Global Trigger.
- GTFE = Global Trigger Front End.
- GTL = Global Trigger Logic board.
- HEP = High Energy Physics.

- GTX = Is a type of high speed electronics link.
- H-HG = Front hadronic section of HGC.
- H/E = Ratio of energy deposits between HCAL and ECAL trigger towers.
- H4 = H4 Beam line at CERN
- HB = Barrel portion of HCAL covering pseudorapidity less than 1.2.
- HCAL = Hadronic CALorimeter.
- HCAL FG Bit = ”Fine-Grain” bit to indicate if HCAL tower energy is consistent with a MIP.
- HDL = Hardware Description Language.
- HE = Endcap portion of HCAL covering pseudorapidity between 1.2 and 3.
- HF = Very forward portion of HCAL covering pseudorapidity between 3 and 5.
- HGC = High Granularity Calorimeter
- HI = Heavy Ions, at the LHC refers to collisions between lead ions.
- HL-LHC = High Luminosity Large Hadron Collider, is an upgrade planned for the current LHC machine around 2020.
- HLT = High Level Trigger, a collection of software trigger algorithms.
- HO = Outer Barrel Hadronic Calorimeter
- HO = HCAL Outer.
- HPC = High Performance Computing.
- HPK = Hamamatsu Photonics K.K.
- HSE = CERN’s Occupational Health & Safety and Environmental Protection Unit.
- HT = Magnitude of the vectorial sum of transverse energy of jets (hadronic).
- HTM = Missing Total transverse energy of jets, normally used in the context of a trigger algorithm that will select events with a minimum amount of this variable.
- HTR = HCAL Trigger and Readout Card.
- HTRG = BTI High Level Trigger.
- HTT = Total transverse energy of jets (hadronic), normally used in the context of a trigger algorithm that will select events with a minimum amount of this variable.
- HTTP = Hypertext Transfer Protocol.
- HV = High Voltage.
- HV-CMOS = CMOS with high voltage capaility.
- IC = Integrated Circuit.
- I/O = Input/Output.
- I2C = Inter-Integrated Circuit chip-to-chip communications protocol.
- IB = Institution Board.
- ID = Identification, algorithm/method to identify particle candidate objects.
- ieta = integer index for defining detector volumes as a function of pseudorapidity.
- ILC = International Linear Collider
- IM = Input Module.
- IP = Impact Parameter, minimum distance between a trajectory and a selected vertex.

- IPBus = is a protocol to control and communicate with Ethernet-attached xTCA hardware.
- iphi = integer index for defining detector volumes as a function of azimuthal angle.
- IPMI = Intelligent Platform Management Interface, is a standardised computer system interface used by system administrators.
- ISA = Industrial Standard Architecture PC communication/control bus.
- ISAJET = A program to simulate high energy particle interactions.
- ISO = Isolation bit.
- isoEG = Isolated Electron-Gamma, normally used in the context of a trigger algorithm that will select an isolated electron or gamma candidate.
- isoMuon = Isolated Moun, normally used in the context of a trigger algorithm that will select an isolated moun gamma candidate.
- isoTau = Isolated Tau, normally used in the context of a trigger algorithm that will select an isolated tau gamma candidate.
- J2 = Green emission fluorescent dye similar to Y11
- JSC = Jet Summary Card.
- JTAG = Joint Test Action Group; test and diagnostic bus standard by IEEE1149.1.
- L = (or LUMI) Luminosity, depending on units may mean Instantaneous Luminosity or Integrated Luminosity.
- L1 = Level-1 trigger.
- L1A = Level-1 Accept.
- LAN = Local Area Network.
- LB = Link Board.
- LC = Common optical fiber connector standard.
- LCT = Local Charged Track, or muon stub, part of the CSC system.
- LDEMUX = Link Demultiplexer (an FPGA or anASIC).
- LFSR = Linear Feedback Shift Register.
- LHC = Large Hadron Collider.
- LHCb = LHC Beauty Physics Experiment
- LHCC = Large Hadron Collider Experiments Committee.
- LINX = Link Test Board.
- LMUX = Link Multiplexer (an FPGA).
- LOC = Lines Of Code.
- LpGBT = Low-power version of the Gigabit transceiver.
- LPM = Local Partition Managers.
- LS1 = Long Shutdown 1, first LHC long shutdown from beginning 2013 to end of 2014.
- LS2 = Long Shutdown 2, second LHC long shutdown scheduled for around 2018.
- LS3 = LHC Long Shutdown 3
- LS3 = Long Shutdown 3, third LHC long shutdown scheduled for around 2022.
- LSP = Lightest Supersymmetric Particle, is the generic name given to the lightest of

the additional hypothetical particles found in supersymmetric models.

- LTRG = BTI Low Level Trigger.
- LTS = Low Trigger Suppression.
- LUT = Look-Up Table (memory).
- L1 tracks = Trigger primitives of the Tracker.
- LINAC = Linear Accelerator.
- LV = Low Voltage.
- LVDS = Low Voltage Differential Signaling, a specification for differential digital logic.
- LVPECL = Low Voltage Positive ECL.
- LVTTL = Low Voltage TTL.
- LYSO = Lutetium Yttrium Silicate Crystal
- LYSO(Ce) = LYSO Crystal which is Cerium Activated
- MAD = Multiple Amplifier and Discriminator.
- MaPSA = Macro-Pixel-Sub-Assembly.
- MB1..MB4 = Muon Barrel Stations.
- MC = Mini Crate or Monte Carlo.
- MCH = μTCA Control Hub.
- MCP = Microchannel plate photosensor or photomultiplier.
- MCz = Magnetic Czochralski silicon material grown via the Magnetic Czochralski process.
- ME = Muon Endcap.
- MET = Missing Transverse Energy.
- minbias = Soft proton-proton collision.
- MIP = Minimum Ionizing Particle (muon).
- MLB = Master Link Board.
- MMC = Mezzanine Management Controller, part of the μtca specification.
- MPA = Macro Pixel ASIC.
- MP7 = Master Processor 7 card, featuring the Xilinx FPGA Virtex-7 chip.
- MPC = Muon Port Card, part of the CSC system.
- mPU = MicroProcessor Unit.
- MRB = Master Readout Board.
- MS = Muon Sorter of CSC trigger.
- MSL = Mask and Sort Logic.
- MSSM = Minimal SUSY Standard Model - a simplified
- mSUGRA = Minimal Supergravity - a constrained model of SUSY.
- MTest = Meson Laboratory Test Beam Facility at Fermilab
- MTF7 = Muon Track Finder 7 card, featuring the Xilinx FPGA Virtex-7 chip.
- MTP = Optical fiber connector standard allowing up to 12 fibers in a single connector.

- MTP-LC = An optical cable with an MTP connector on one end and many LC connectors on the other.
- Muon MIP Bit = A bit to characterise if a 4x4 trigger region is consistent with a MIP.
- NFS = Network File System, is a distributed file system protocol.
- ngCCM = New generation Clock and Control Module (part of the FEE).
- ngFEC = New generation Front-End Controller (part of the BE electronics).
- n-in-n sensor = Sensor with n-bulk and n-implant electrodes (pixel/strips).
- n-in-p sensor = Sensor with p-bulk and n-implant electrodes (pixel/strips).
- O2O = Is a software to simplify the propagation of configuration online.
- OCS = Optical Communication System.
- ODU = Optical Decoder Unit.
- OM3 = Is a type of optical fiber mostly used for communication over short distances.
- oRM = Optical Receiver Mezzanines.
- OrN = Orbit Number.
- oRSC = Optical Regional Summary Card
- oRX = Optical Receiver Boards.
- oSLB = Optical Synchronization and Link Boards.
- P&R = Place & Route.
- PAC = Pattern Comparator (RPC trigger ASIC).
- PACT = PAttern Comparator Trigger.
- PC = Personal Computer.
- PCB = Printed Circuit Board, is used to mechanically support and electrically connect electronic components.
- PCI = Personal Computer Interface Bus.
- PCIe = Peripheral Component Interconnect Express, is a high-speed serial computer expansion bus standard.
- PDE = Photon-detection efficiency defined as the product of geometrical acceptance of an SiPM and the silicon quantum efficiency.
- PF = Particle flow. A reconstruction method that combines information from different detectors.
- PHIAU = Phi Assignment Unit.
- PHITRB128 = 128 channels Trigger Board in longitudinal CMS plane.
- PHITRB32 = 32 channels Trigger Board in longitudinal CMS plane.
- PHTF = Phi Track Finder.
- pileup = Overlapping of multiple soft interactions during a single LHC beam crossing.
- p-in-n sensor = Sensor with n-bulk and p-implant electrodes (pixel/strips).
- PLD = Programmable Logic Device, term used by Altera Corp., similar to FPGA.
- PLL = Phase Locked Loop.
- PM = Power Module, rack mounted power supply module.
- PMC = PCI Mezzanine Card.

- PMT = Photomultiplier Tubes.
- PON = Passive Optical Network.
- Priority Encoder = Logic that selects for output the highest value of a set of input values.
- PS = Proton Synchrotron at CERN
- PSB = Pipelined Synchronizing Buffer.
- pseudorapidity (η) = $-\ln\tan\frac{\theta}{2}$, where θ is the angle of the particle momentum with respect to the anti-clockwise beam direction.
- PS modules = Modules composed of two sensors superimposed on each other, one segmented in strips, and the other in macro-pixels of size 100 μm x 1.5mm.
- p-stop or p-spray = Technology to isolate pixels or strips.
- PSU = Power Supply Unit.
- p_{T} = Transverse momentum of a physics object.
- p_{T} module = Module intrinsically capable to identify particle above a chosen p_{T} value.
- PU = Pileup, average quantity of particle collisions per bunch crossing.
- PYTHIA = A program to simulate high energy particle interactions.
- QA = Quality Assurance.
- qAU = Quality Assignment Unit.
- QCD = Quantum Chromodynamics - a theory that describes strong interactions.
- QGP = Quark-Gluon Plasma, is a phase of quantum chromodynamics where quarks and gluons are asymptotically free.
- QIE = Charge (Q) Integrator and Encoder, a series of ASICs.
- Q1 = The quadruple magnet closest to the CMS interaction point.
- R&D = Research and Development.
- RAM = Random Access Memory.
- RAW = Raw Data, Data tier format that contains the information read directly from the detector.
- RB = Readout Board.
- RBX = Readout Box, housing the HB and HE FEE.
- RCMS = Run Control and Monitoring System.
- RCS = Run Control System.
- RCT = Regional Calorimeter Trigger.
- RD50/RD53 = CERN RD50/RD53 collaboration.
- RDPM = Readout Dual Port Memory.
- RECO = Reconstructed Data, Data tier format that contains the reconstructed physics objects information from the CMS recorded data.
- RF = Radio-Frequency.
- RLDEMUX = Special LDEMUX on a RB (an FPGA).
- RLDRAM = Reduced-latency Dynamic random access memory, is a type of dynamic random access memory.

- RO = Readout.
- ROC = Read-Out Chip.
- ROOT = Is an object-oriented program and library developed by CERN.
- ROP = Readout Processor.
- ROSE100 = ECAL Readout and Trigger Board.
- RPC = Resistive Plate Chamber.
- RPM = Red Hat Package Manager, is a software package management system.
- RSS = Rich Site Summary.
- RT = Regional Trigger.
- RUI = Readout Unit Interface.
- Rx = Optical Receiver.
- S/N = Signal=to-noise ratio.
- SAS = Serial Attached SCSI, is a point-to-point serial protocol that is used to move data to and from computer storage devices such as hard drives and tape drives.
- SATA = Serial Advanced Technology Attachment, is a computer bus interface for connecting host bus adapters to mass storage devices such as hard disk drives and optical drives.
- SB = Server Board.
- SBT = Second Best Track.
- SC = Sector Collector, part of the DT system.
- SCADA = Supervisory Control And Data Acquisition.
- SCB = Sector Collector Board, part of the DT system.
- SCSI = Small Computer Systems Interconnect.
- SCU = Server and Control Unit.
- SDRAM = Synchronous dynamic random access memory, is a type of dynamic random access memory.
- SEE = Single Event Effect, a generic effect of radiation on electronics.
- SEL = Single Event Latchup.
- SerDes = Serializer/Deserializer.
- SET = Single Event Transient.
- SEU = Single Event Upset.
- SFP = Small Form-factor Pluggable standard for optical and other transceivers which connect to a standard-defined socket.
- SFP+ = Extension of the SFP standard to support up to 10 Gbps data rates.
- SiPM = Silicon Photomulitplier.
- SL = SuperLayer.
- SLB = Synchronisation and Link Board of the CMS calorimeter trigger system, also: Slave Link Board.
- SLINK = is a CERN specification for an easy-to-use FIFO-like data-link.
- SLVS = Scalable Low Voltage Signaling.
- SM = Standard Model of Particle Physics, is a theory concerning the electromagnetic,

weak, and strong nuclear interactions.

- SOAP = Simple Object Access Protocol, is a protocol specification for exchanging structured information.
- SP = Sector Processor.
- SR = Sector Receiver.
- SRAM = Static Random Access Memory: memory that does not need refresh cycles.
- SRB = Slave Link Board.
- SRIO = Serial RapidIO, is a interconnect technology for interconnecting chips on a circuit board.
- SSA = Short Strip ASCI.
- SSTL = Stub Series Terminated Logic.
- STP = Shielded Twisted Pair.
- stub = Pair of hits in the two sensors of a p_{T} module compatible with particles above the chosen threshold.
- SU = Synchronization Unit.
- SUSY = SUper SYmmetry - an as yet unobserved symmetry between fermions and bosons.
- SW = Software.
- TA = Track Assembler.
- TAN = Target Absorber Neutral.
- TAS = Target Absorber Secondaries.
- TAU = Track Assembler Unit - element of CSC Sector Processor.
- Tau Veto = A bit to indicate if energy in a 4x4 region is not consistent with a narrow shower.
- TB = Trigger Board.
- TB2S = Tracker Barrel with 2S modules.
- TBPS = Tracker Barrel with PS modules.
- TC = Trigger Crate.
- TCC = Trigger Concentrator Card.
- TCDS = Trigger Control and Distribution System.
- TCM = Trigger Concentrator Module.
- TCP/IP = Transmission Control Protocol over Internet Protocol.
- TCS = Trigger Control System.
- TDC = Trigger Data Concentrator card, also: Time-to-digital converter.
- TDR = Technical Design Report.
- TEDD = Tracker End-Cap Double Disk.
- TEP = Technology Evaluation Platform.
- TF = Track Finder.
- THETATRB = Trigger Board in transverse CMS plane.
- TID = Total Ionizing Dose.
- TIM = Timing Module.

- tkLayout = Standalone software package to support detector design effort.
- TMB = Trigger MotherBoard: performs anode-cathode coincidence, part of the CSC system.
- TME = Trigger Menu Editor.
- TMR = Triple Modular Redundancy, a technique for avoiding errors in electronics circuits.
- TMT = Time-Multiplexed Trigger, trigger design that processes events in parallel rather than sequentially.
- TPG = Trigger Primitive Generator.
- TPM = Trigger Processor Module, also: Track Pipeline and Multiplexer.
- TRACO = Track Correlator.
- TRC = Trigger Readout Crate.
- TriDAS = Trigger and DAQ.
- TRIG = Trigger.
- TS = Trigger Server, also: Track Segment, also: Tigger Supervisor.
- TSϕ = Trigger Server in longitudinal CMS plane.
- TSM = Track Sorter Master.
- TSMD = Trigger Sorter Master - Data.
- TSMS = Track Sorter Master Sorter.
- TSq = Trigger Server in transverse CMS plane.
- TSS = Track Sorter Slave, also: Trigger Sorter Slave.
- TST = Tracker Support Tube.
- TST = Trigger Server Theta.
- TSV = Through Silicon Via.
- TTC = Trigger Timing and Control, a system for distribution of clocking and control.
- TTCci = TTC CMS interface.
- TTCcpm = TTC central partition manager.
- TTCex = TTC Encoder and Transmitter.
- TTClpm = TTC local partition manager.
- TTCrx = TTC Receiver ASIC.
- TTCvi = TTC VME Interface.
- TTL = Transistor-transistor logic, a common type of digital logic signaling.
- TTS = Trigger Throttle System.
- Tx = Optical Transmitter.
- UDP/IP = User Datagram Protocol using the Internet Protocol.
- uHTR = See μHTR
- USC = Underground Service Cavern, where the CMS counting room is located.
- UXC = Underground Experimental Cavern, where the CMS detector is located.
- VBF = Vector Boson Fusion.
- VCSEL = Vertical Cavity Surface Emitting Laser.

- VHDL = Very-high-speed integrated circuits Hardware Description Language.
- VL = Versatile Link.
- VME = Electronics and mechanics standard for crates, buses and application boards.
- VME = Versa Module Eurocard bus, is an ANSI/IEEE computer hardware bus standard.
- VTRX = Versatile Link Transmitter/Receiver, optical transceiver developed by the CERN Versatile Link project.
- VTTx = Versatile Link Dual Transmitter, optical transmitter developed by the CERN Versatile Link project which combines two transmission channels in a single SFP module.
- WBM = Web Based Monitoring.
- WLS = Wavelength shifter.
- WS = Wedge Sorter.
- WTE = Warning Test Enable.
- XDAQ = Cross DAQ, is a data acquisition software framework
- XMAS = XDAQ Monitoring and Alarm System.
- YETS = Year-End Technical Stop, is a relatively brief stop of the LHC, typically less than three months in length, during the winter holidays.
- ZCD = Zero Crossing Discriminator.

References

[1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[3] CMS Collaboration, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs", *Phys. Lett. B* **736** (2014) 64, doi:10.1016/j.physletb.2014.06.077, arXiv:1405.3455.

[4] CMS Collaboration, "Technical Proposal for the Upgrade of the CMS Detector through 2020", Technical Report CERN-LHCC-2011-006, CMS-UG-TP-1, LHCC-P-004, 2011.

[5] CMS Collaboration, "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", Technical Report CERN-LHCC-2012-015, CMS-TDR-010, 2012.

[6] CMS Collaboration, "CMS Technical Design Report for the Pixel Detector Upgrade", CMS Technical Design Report CERN-LHCC-2012-016, CMS-TDR-11, 2012.

[7] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CMS Technical Design Report CERN-LHCC-2013-011, CMS-TDR-12, 2013.

[8] CMS Collaboration, "Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model", Technical Report CMS-PAS-HIG-14-009, CERN, Geneva, 2014.

[9] CMS Collaboration, "Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process", arXiv:1307.7135.

[10] S. Malik et al., "Interplay and Characterization of Dark Matter Searches at Colliders and in Direct Detection Experiments", arXiv:1409.4075.

[11] L. Carpenter et al., "Mono-Higgs: a new collider probe of dark matter", *Phys. Rev. D* **89** (2014) 075017, doi:10.1103/PhysRevD.89.075017, arXiv:1312.2592.

[12] I. Azhgirey and V. Talanov, "Proc. of XVIII Workshop on the Charged Particle Accelerators", Technical Report CMS-NOTE-2005-010, 2000.

[13] G. Battistoni et al., "The FLUKA code: Description and benchmarking", *AIP Conf. Proc.* **896** (2007), no. SLAC-REPRINT-2007-184, 31.

[14] A. Ferrari, P. R. Sala, A. Fassò, and J. Ranft, "FLUKA: A multi-particle transport code (program version 2005)". CERN, Geneva, 2005.

[15] S. Roesler, R. Engle, and J. Ranft, "The Monte Carlo Event Generator DMPJET III", Technical Report SLAC-PUB-8740, 2000.

[16] T. Pierog et al., "EPOS LHC : test of collective hadronization with LHC data", Technical Report arXiv:1306.0121, Jun, 2013. Comments: 16 pages, 34 figures.

[17] I. Bergstrom and H. Vincke, "Residual dose rates in CMS during long shutdowns following the LHC operation until LS5", Technical Report CERN-RP-2014-052-REPORTS-TN, CERN, 2014.

[18] A. Dabrowski et al., "Development of the CMS nominal FLUKA Model; versions 2.0.0.0 and 3.0.0.0", Technical Report AAA-AA-AA-0002, CERN, 2014.

[19] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.

[20] CMS and ATLAS Collaborations, "Expected pileup values at the HL-LHC", Technical Report CMS-NOTE-2013-003.

[21] C. Adloff et al., "Response of the CALICE Si-W electromagnetic calorimeter physics prototype to electrons", *Nucl.Instrum.Meth.* **A608** (2009) 372, doi:10.1016/j.nima.2009.07.026.

[22] J. de Favereau et al., "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", arXiv:1307.6346.

[23] M. Schwarz, "CMS pixel simulations", *Nucl. Instrum. Meth.* **A511** (2003) 88.

[24] M. Moll, "Radiation Damage in Silicon Particle Detectors - Microscopic Defects and Macroscopic Properties",.

[25] A. Ferrari, P. Sala, A. Fassò, and J. Ranft, "FLUKA: A multi-particle transport code (program version 2005)", Technical Report CMS-2005-010, 2005.

[26] CMS Collaboration, "The CMS Particle Flow Algorithm", arXiv:1401.8155.

[27] C. C. Foudas, A. Rose, J. Jones, and G. Hall, "A Study for a Tracking Trigger at First Level for CMS at SLHC", arXiv:0510227.

[28] ATLAS Collaboration, "ATLAS Insertable B-Layer Technical Design Report",.

[29] G. Aglieri et al., "Monolithic active pixel sensor development for the upgrade of the ALICE inner tracking system", *JINST* **8** (2013) C12041, doi:10.1088/1748-0221/8/12/C12041.

[30] M. Bubna et al., "Testbeam and Laboratory Characterization of CMS 3D Pixel Sensors", arXiv:1402.6384.

[31] The RD53 collaboration.

[32] S. Bonacini et al., "Characterization of a commercial 65 nm CMOS technology for SLHC applications", *JINST* **7** (2012) P01015, doi:10.1088/1748-0221/7/01/P01015.

[33] E. Conti, J. Christiansen, P. Placidi, and S. Marconi, "Pixel chip architecture optimization based on a simplified statistical and analytical model", *JINST* **9** (2014) C03011, doi:10.1088/1748-0221/9/03/C03011.

[34] P. Moreira, J. Christiansen, and K. Wyllie, "The GBT manual", technical report, CERN, 2013.

[35] P. Farthouat and M. Hansen, eds., "ACES 2014 - Fourth Common ATLAS CMS Electronics Workshop for LHC Upgrades". (2014).

[36] G. L. G.-L. Casse, A. A. Affolder, P. Allport, and M. Wormald, "Measurements of charge collection efficiency with microstrip detectors made on various substrates after irradiations with neutrons and protons with different energies", in *Proceedings of Science, PoS(VERTEX 2008)*, p. 036. 2008.

[37] CMS Tracker Sensor Collaboration, "Campaign to identify the future CMS tracker baseline", *Nucl.Instrum.Meth.* **A658** (2011) 30, doi:10.1016/j.nima.2011.05.028.

[38] CMS Tracker Collaboration, "Characterisation of silicon sensor materials and designs for the CMS Tracker Upgrade", *PoS* **Vertex2012** (2013) 016.

[39] CMS Tracker Collaboration, "Planar silicon sensors for the CMS tracker upgrade", *Nucl.Instrum.Meth.* **A732** (2013) 113, doi:10.1016/j.nima.2013.07.058.

[40] CMS Tracker Collaboration, G. Steinbruck, "Towards Radiation Hard Sensor Materials for the CMS Tracker Upgrade", in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, p. 1828. IEEE, 2012.

[41] CMS Tracker Collaboration, T. Pohlsen, "Radiation hard silicon sensors for the CMS tracker upgrade", in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE*, p. 1. IEEE, 2013.

[42] A. Junkes, "Influence of radiation induced defect clusters on silicon particle detectors",.

[43] A. Nurnberg and T. Schneider, "Lorentz angle measurements as part of the sensor R&D for the CMS Tracker upgrade", *JINST* **8** (2013) C01001.

[44] CMS Tracker Collaboration, "Silicon sensor developments for the CMS tracker upgrade", *JINST* **7** (2012) C01110, doi:10.1088/1748-0221/7/01/C01110.

[45] CMS Tracker Collaboration, A. Dierlamm, "Planar sensors for future vertex and tracking detectors", in *Vertex 2013 22nd International Workshop on Vertex Detectors*. 2013.

[46] CMS Tracker Collaboration, M. Printz, "Radiation Hard Sensor Materials for the CMSTracker Upgrade - Charge Collection of Different Bulk Polarities", in *9th International "Hiroshima" Symposium on the Development and Application of Semiconductor Tracking Detectors*. 2013.

[47] CMS Tracker Collaboration, "Analysis of testbeam data of irradiated silicon prototype sensors for the CMS tracker upgrade", *Nucl.Instrum.Meth.* **A730** (2013) 195, doi:10.1016/j.nima.2013.03.018.

[48] T. Mäenpää, "Performance of different silicon materials for the upgraded CMS tracker", technical report, RD13, 2013. CMS CR-2013/236.

[49] M. Bernard-Schwarz, "Measurements and Irradiation Analysis of Silicon Structures for the CMS Upgrade",.

[50] CMS Tracker Collaboration, T. Bergauer, "CMS Process control using dedicated test structures and radiation", in *Workshop on Quality Issues in Current and Future Silicon Detectors*. 2011.

[51] M. Dragicevic, "The New Silicon Strip Detectors for the CMS Tracker Upgrade",.

[52] CMS Tracker Collaboration, "Silicon sensor developments for the CMS tracker upgrade", *PoS* **RD11** (2011) 020.

[53] M. Huhtinen, "Simulation of non-ionising energy loss and defect formation in silicon", *Nucl.Instrum.Meth.* **A491** (2002) 194, doi:10.1016/S0168-9002(02)01227-5.

[54] L. Snoj, G. Zeronik, and A. Trkov, "Computational analysis of irradiation facilities at the JSI TRIGA reactor", *Applied Radiation and Isotopes* **70** (2011) 483, doi:10.1016/j.apradiso.2011.11.042.

[55] CMS Tracker Collaboration, R. Dalal et al., "Comparison of Radiation Hardness Properties of p+n- and n+p- Si Strip Sensors Using Simulation Approaches", in *23rd RD50 Workshop*. 2013.

[56] F. Morsani et al., "The AMchip: A VLSI associative memory for track finding", *NIM* **23** (1992) 446.

[57] M. Raymond et al., "The CMS binary chip for microstrip tracker readout at the SLHC", *JINST* **7** (2012) C01033, doi:10.1088/1748-0221/7/01/C01033.

[58] D. Braga et al., "CBC2: A microstrip readout ASIC with coincidence logic for trigger primitives at HL-LHC", *JINST* **7** (2012) C10003, doi:10.1088/1748-0221/7/10/C10003.

[59] G. Blanchot et al., "Hybrid circuit prototypes for the CMS Tracker upgrade front-end electronics", *JINST* **8** (2013) C12033, doi:10.1088/1748-0221/8/12/C12033.

[60] CMS Tracker Collaboration, M. Pesaresi et al., "Phase II tracker upgrade: beam test with prototype 2S modules", in *CMS Tracker week, January 2014*. 2014.

[61] P. Vichoudis et al., "First results with the Gigabit Link Interface Board (GLIB)", *JINST* **6** (2011) C12060, doi:10.1088/1748-0221/6/12/C12060.

[62] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.

[63] CMS Collaboration, "tkLayout: a design tool for innovative silicon tracking detectors", *JINST* **9** (2014) C03054, doi:10.1088/1748-0221/9/03/C03054.

[64] CDF Collaboration, "Performance of the CDF online silicon vertex tracker", *IEEE Trans.Nucl.Sci.* **49** (2002) 1177, doi:10.1109/TNS.2002.1039633.

[65] C. Cheshkov, "Fast Hough-transform track reconstruction for the ALICE TPC", *Nucl.Instrum.Meth.* **A566** (2006) 35, doi:10.1016/j.nima.2006.05.036.

[66] L. Ristori, "An artificial retina for fast track finding", *Nucl.Instrum.Meth.* **A453** (2000) 425, doi:10.1016/S0168-9002(00)00676-8.

[67] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", arXiv:1405.6569.

[68] CMS Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at s = 7 TeV", *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.

[69] A. Ronzhin et al., "Development of a new fast shower maximum detector based on microchannel plates photomultipliers (MCP-PMT) as an active element", *Nucl.Instrum.Meth.* **A759** (2014), no. 0, 65, doi:http://dx.doi.org/10.1016/j.nima.2014.05.039.

[70] L. Brianza et al., "Response of microchannel plates to single particles and to electromagnetic showers", arXiv:1504.02728.

[71] R. De Oliveira, M. Maggi, and A. Sharma, "A novel fast timing micropattern gaseous detector: FTM", arXiv:1503.05330.

[72] S. N. White, "R&D for a Dedicated Fast Timing Layer in the CMS Endcap Upgrade", *Acta Phys. Pol. B Proc. Suppl.* **7** (2014) 743, doi:10.5506/APhysPolBSupp.7.743, arXiv:1409.1165.

[73] S. White, "Experimental Challenges of the European Strategy for Particle Physics", arXiv:1309.7985.

[74] P. Adzic et al., "Radiation hardness qualification of PbWO4 scintillation crystals for the CMS Electromagnetic Calorimeter", *JINST* **5:P03010** (2010) doi:10.1088/1748-0221/5/03/P03010.

[75] M. Huhtinen et al., "High-energy proton induced damage in PbWO-4 calorimeter crystals", *Nucl.Instrum.Meth.* **A545** (2005) 63, doi:10.1016/j.nima.2005.01.304.

[76] E. Auffray et al., "Experimental study of the lead tungstate scintillator proton-induced damage and recovery", *IEEE Trans. Nucl. Sci.* **59 n.5** (2012) 2219.

[77] P. Lecomte, D. Luckey, F. Nessi-Tedaldi, and F. Pauss, "High-energy proton induced damage study of scintillation light output from PbWO-4 calorimeter crystals", *Nucl.Instrum.Meth.* **A564** (2006) 164, doi:10.1016/j.nima.2006.04.043.

[78] CMS ECAL Collaboration, "Test beam evaluation of CMS ECAL Endcap Supercrystals constructed from hadron-damaged crystals". http://cms.cern.ch/iCMS/jsp/openfile.jsp?type=DN&year=2013&files=DN2013_007.pdf, 2013. DN-2013/007.

[79] GEANT4 Collaboration, "GEANT4: A Simulation toolkit", *Nucl.Instrum.Meth.* **A506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[80] F. Gentit, "Litrani: A General purpose Monte Carlo program simulating light propagation in isotropic or anisotropic media", *Nucl.Instrum.Meth.* **A486** (2002) 35, doi:10.1016/S0168-9002(02)00671-X.

[81] Z. Antunovic et al., "Radiation hard avalanche photodiodes for the CMS detector", *Nucl.Instrum.Meth.* **A537** (2005) 379, doi:10.1016/j.nima.2004.08.047.

[82] S. Baccaro et al., "Radiation damage effect on avalanche photodiodes", *Nucl.Instrum.Meth.* **A426** (1999) 206, doi:10.1016/S0168-9002(98)01493-4.

[83] CMS Collaboration, "CMS: The electromagnetic calorimeter. Technical design report",.

[84] K. Bell et al., "Vacuum phototriodes for the CMS electromagnetic calorimeter endcaps", *IEEE Transactions on Nuclear Science* **51 Issue 5** (2004) 2284, doi:10.1109/TNS.2004.836053.

[85] CMS Collaboration, "CMS: The Hadron Calorimeter Project. Technical design report",.

[86] V. Hagopian and I. Daly *AIP Conf. Proc* **450** (1998) 53, doi:10.1063.1.56958.

[87] CMS-HCAL Collaboration, "Design, Performance, and Calibration of CMS Forward Calorimeter Wedjes", *EPJ* **C53** (2008) 139.

[88] K. Cankocak et al., "Radiation hardness measurements of high OH- content quartz fibres irradiated with 24GeV protons up to 1.25Grad", *Nucl.Instrum.Meth.* **A585** (2008) 20.

[89] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[90] I. Dafinei, E. Auffray, P. Lecoq, and M. Schneegans, "Scintillator and phosphor materials symposium, San Francisco, USA 1994", *Mat. Res. Soc. Symp. Proc.* **348** (1994) 99.

[91] P. Lecoq et al., "Lead tungstate (PbWO-4) scintillators for LHC EM calorimetry", *Nucl.Instrum.Meth* **A365 Issues 2-3** (1995) 291, doi:10.1016/0168-9002(95)00589-7.

[92] P. A. Semenov et al., "First study of radiation hardness of lead tungstate crystals at low temperatures", *Nucl. Instr. Meth.* **A582** (2007) 575, doi:10.1016/j.nima.2007.08.178.

[93] CMS Collaboration, "Time reconstruction and performance of the CMS electromagnetic calorimeter", *JINST* **5** (2010) T03011, doi:10.1088/1748-0221/5/03/T03011.

[94] K. Hansler, "A 0.13-mu-m CMOS technology: Its radiation hardness and its application in high energy physics experiments",.

[95] L. Feld et al., "DC-DC powering for the CMS pixel upgrade", *Nucl.Instrum.Meth.* **A732** (2013) 493, doi:10.1016/j.nima.2013.06.053.

[96] K. Compton et al., "The MP7 and CTP-6: multi-hundred Gbps processing boards for calorimeter trigger upgrades at CMS", *JINST* **7:C12024** (2012) doi:doi:10.1088/1748-0221/7/12/C12024.

[97] ILC Collaboration, "ILC Reference Design Report Volume 4 - Detectors", technical report, 2007. arXiv:0712.2356.

[98] A. Akindinov et al., "Performance of the ALICE Time-Of-Flight detector at the LHC", *Eur. Phys. J. Plus* **128** (2013) 44, doi:10.1140/epjp/i2013-13044-x.

[99] M. Thomson, "Particle Flow Calorimetry and the PandoraPFA Algorithm", *Nucl. Instrum. Meth.* **A611** (2009) 25, doi:10.1016/j.nima.2009.09.009, arXiv:0907.3577.

[100] J. Marshall, A. Münnich, and M. Thomson, "Performance of Particle Flow Calorimetry at CLIC", *Nucl. Instrum. Meth.* **A700** (2013) 153, doi:10.1016/j.nima.2012.10.038, arXiv:1209.4039.

[101] CMS Collaboration, "CMS The Muon Project Technical Design Report",.

[102] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s)=7 TeV at the LHC", (2013, SLACcitation =) P11002, doi:10.1088/1748-0221/8/18/P11002.

[103] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at sqrt(s)=7 TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002.

[104] CMS Collaboration, "CMS Technical Design Report for the Trigger and Data Acquisition project, Volume I",.

[105] S. Costantini et al., "Radiation Background in the Muon System", Technical Report AN-2011-316, 2011.

[106] CMS Collaboration, "Beam test result of a large real-size RPC for the CMS/LHC experiment", *Nucl. Instrum. Meth.* **A546** (2000) 23.

[107] CMS Collaboration, "Study of long-term performance of CMS RPC under irradiation at the CERN GIF", *Nucl. Instrum. Meth.* **A533** (2004) 102.

[108] M. Capeans et al., "Proposal to the CERN SPSC: A GIF++ Gamma Irradiation Facility at the SPS H4 Beam Line", Technical Report SPSC-2009-029 / SPSC-P-339, CERN, 2009.

[109] Y. Erdogan et al., "A possible extension of the barrel muon trigger for the CMS experiment at HL-LHC: Muon Track fast Tag (MTT)", Technical Report DN-14-008, 2014.

[110] E. Conti and F. Gasparini, "Test of the wire ageing induced by radiation for the CMS barrel muon chambers", *NIM* **465A 472-481** (2001).

[111] M. M. et al., "Software Evaluation and Irradiation Test of the SBS Technologies Model 620 PCI-to-VME Bus Adapter.", *Proceedings of the 9th Workshop on Electronics for LHC Experiments. Pp. 389-391*.

[112] C. F. Bedoya et al., "Electronics for the CMS muon drift tube chambers: the read-out minicrate", *Nuclear Science Symposium Conference Record* **2** (2004) 1309.

[113] CMS Collaboration, "Bunched beam test of the CMS drift tubes local muon trigger", *Nucl. Instrum. Meth.* **534** (2004) 441.

[114] CMS Collaboration, "Aging tests of full scale CMS cathode strip chambers", *Nucl. Instrum. Meth.* **A515** (2003) 226.

[115] CMS Collaboration, "Aging tests of CMS muon chamber prototypes", *Nucl. Instrum. Meth.* **A488** (2002) 240.

[116] CMS Collaboration, "Results of radiation tests of the cathode front-end boards for the CMS endcap muon chambers", *Nucl. Instrum. Meth.* **A471** (2001) 340.

[117] CMS Collaboration, "Radiation Testing of Electronics for the CMS Endcap Muon System", *Nucl. Instrum. Meth.* **A698** (2013) 242.

[118] CMS Collaboration, "Quantitative aging study with intense irradiation tests for the CMS forward RPCs", *Nucl. Instrum. Meth.* **A602** (2009) 771.

[119] CMS Collaboration, "Neutron irradiation of RPCs for the CMS experiment", *Nucl. Instrum. Meth.* **A508** (2003) 120.

[120] K. Banzuzi et al., "Link board for CMS RPC – production issues", *Nucl. Instrum. Meth.* **A587** (2008) 110.

[121] CMS Collaboration, "Neutron-induced Single Event Upset on the RPC front-end chips for the CMS experiment", *Nucl. Instrum. Meth.* **A484** (2002) 494.

[122] CMS Collaboration, "Radiation tests of CMS RPC muon trigger electronic components", *Nucl. Instrum. Meth.* **A538** (2005) 708.

[123] CMS Collaboration, "Operational Experience of the Gas Gain Monitoring system of the CMS RPC muon detectors", *Nucl. Instrum. Meth.* **A617** (2010) 146.

[124] G. Antchev et al., "TOTEM Detector at LHC", *Nucl.Instrum.Meth.* **A617** (2010).

[125] A. Augusto Alves Jr et al., "The LHCb Detector at the LHC", *JINST* **S08005** (2008).

[126] E. Ageev et al., "(COMPASS) A new measurement of the Collins and Sivers asymmetries on a transversely polarised deuteron target", *Nucl. Phys.* **B 765, 31** (2007).

[127] D. Abbaneo et al., "Characterization of GEM Detectors for Application in the CMS Muon Detection System", *Nucl. Sci. Symp. Conf. Rec.* **1416-1422** (2010).

[128] GEM-CMS Collaboration, "Movie about GE1/1 assembly", (2014).

[129] D. A. *et al.*, "Test Beam Results of the GE1/1 Prototype for CMS High-η Muon System Future Upgrade", *IEEE Nucl.Sci.Symp.Conf.Rec.* **1806-1810** (2011).

[130] A. Sharma, "GEM technology choices for CMS". `https://indico.cern.ch/event/315330/contribution/3/material/slides/0.pptx`, 2014.

[131] J. Merlin, "GEM gain calibration". `https://indico.cern.ch/event/300480/session/6/contribution/86/material/slides/0.pdf`, 2014.

[132] S. Bachman et al., "Discharge studies and prevention in the gas electron multiplier (GEM)", *Nucl. Instrum. Meth.* **A479** (2002).

[133] P. Aspell et al., "VFAT2: A front-end system on chip providing fast trigger information, digitized data storage and formatting for the charge sensitive readout of multi-channel silicon and gas particle detectors", in *Proceedings of TWEPP Prague, Czech Republic, 3-7 September 2007*. 2007.

[134] S. Costantini et al., "RPC radiation background at the CMS experiment", *in print on JINST* (2014).

[135] M. Abbrescia, "The dynamic behaviour of Resistive Plate Chambers", *Nucl. Instrum. Meth.* **A533** (2004) 7.

[136] Y. Haddad et al., "High rate resistive plate chamber for LHC detector upgrades", *Nucl. Instrum. Meth.* **A718** (2013) 424.

[137] A. Laso Garcia et al., "Extreme high-rate capable timing resistive plate chambers with ceramic electrodes", *JINST* **7 P10012** (2012).

[138] R. Cardarelli et al., "Performance of RPCs and diamond detectors using a new very fast low noise preamplifier", *JINST* **8 P01003** (2013).

[139] K. Lee et al., "Rate capability study for a four-gap phenolic RPC with a CS-137 source", *in print on JINST* (2014).

[140] M. Abbrescia et al., "New developments on front-end electronics for the CMS Resistive Plate Chambers", *Nucl. Instrum. Meth.* **A456** (2000) 143.

[141] J. Fleury et al., "Petiroc and Citiroc: front-end ASICs for SiPM read-out and ToF applications", in *JINST*, volume 9 C01049. 2013.

[142] S. Müller, W. Boer, and T. Müller, "The Beam Condition Monitor 2 and the Radiation Environment of the CMS Detector at the LHC". PhD thesis, KIT Karlsruhe., Karlsruhe, 2011.

[143] B. Dehning et al., "LHC beam loss monitor system design", *Beam Instrumentation Workshop 2002: Tenth Workshop, AIP Conference Proceedings* **648** (2002) 229.

[144] J. Emery et al., "Functional and Linearity test system for the LHC Beam Loss Monitoring data acquisition card", *12th Workshop on Electronics For LHC and Future Experiments* (2007), no. CERN-AB-2007-063, 447.

[145] C. Zamantzas, C. Da Vi?, and B. Dehning, "The Real-Time Data Analysis and Decision System for Particle Flux Detection in the LHC Accelerator at CERN.". PhD thesis, Brunel U., Brunel, 2006. Presented on 28 Feb 2006.

[146] W. de Boer et al., "Radiation Hardness of Diamond and Silicon sensors compared.", *Phys. Status Solidi 204 (2007) pp.3009* **204** (May, 2007) 3009.

[147] RD42 Collaboration, "Review of the development of diamond radiation sensors", *Nucl.Instrum.Meth.* **A434** (1999), no. 1, 131.

[148] A. Dabrowski, "The performance of the Beam Conditions and Radiation Monitoring System of CMS", *Proceedings Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Valencia* (2011) 489, doi:10.1109/NSSMIC.2011.6153979.

[149] M. Guthoff, "Design and Experiences with the Beam Condition Monitor as protection system in the CMS Experiment of the LHC.", *Conf. Proc.* **C11-05-16.4** (May, 2011) TUPD42.

[150] R. Bruce et al., "Simulations and Measurements of Cleaning with 100 MJ Beams in the LHC", *4th International Particle Accelerator Conference* (May, 2013) 52.

[151] M. Guthoff, "Radiation Damage to the diamond-based Beam Condition Monitor of the CMS Detector at the LHC". PhD thesis, KIT Karlsruhe, 2014.

[152] M. Guthoff et al., "Radiation damage in the diamond based beam condition monitors of the CMS experiment at the Large Hadron Collider (LHC) at CERN", *Nucl.Instrum.Meth.* **A730** (2013) 168, doi:10.1016/j.nima.2013.05.041.

[153] A. Ignatenko et al., "Beam Halo Monitor for FLASH and the European XFEL", *Proceedings of IPAC-2012, New Orleans, 2012.* **C1205201** (2012) 816.

[154] A. Bell et al., "Fast Beam Conditions Monitor BCM1F for the CMS Experiment", *Nucl.Instrum.Meth.* **A614** (2010) 433, doi:10.1016/j.nima.2009.12.056, arXiv:0911.2480.

[155] S. Orfanelli et al., "Design of a Novel Cherenkov Detector System for Machine Induced Background Monitoring in the CMS Cavern", *Proceedings, 2nd International Beam Instrumentation Conference, Oxford UK* (2013) 33.

[156] D. Przyborowski, "Development of Front End Electronics for the Beam Condition Monitor for CMS", Technical Report EDMS Id: 1404356, CERN, 2013.

[157] J. Leonard et al., "Fast Beam Conditions Monitor for CMS: Performance and Upgrade", *Nucl.Instrum.Meth.* (2014) doi:http://dx.doi.org/10.1016/j.nima.2014.05.008.

[158] J. L. Leonard, "Upgraded Fast Beam Conditions Monitor for CMS online luminosity measurement", in *Proceedings of the 3rd International Conference on Technology and Instrumentation in Particle Physics*, DPG 2014, Mainz(Germany). March, 2014.

[159] A. Zagozdzinska et al., "The Fast Beam Condition Monitor BCM1F Backend Electronics upgraded, MicroTCA based architecture", *Proceedings submitted, XXXIV-th IEEE-SPIE Joint Symposium on Photonics, Web Engineering, Electronics for Astronomy and High Energy Physics Experiments, Wilga 2014* (2014).

[160] A. Ershov, A. Demianov, A. Gribushin, and A. Kaminsky, "HF RADMON status and upgrade plans",.

[161] T. Holy et al., "Pattern recognition of tracks induced by individual quanta of ionizing radiation in Medipix2 silicon detector", *Nucl.Instrum.Meth.* **A591** (2008), no. 1, 287.

[162] D. Pfeiffer et al., "Design, Implementation and First Measurements with the Medipix Neutron Camera in CMS", *J. Instrum.* **6** (May, 2011) P08005.

[163] The Medipix collaboration http://medipix.web.cern.ch/medipix.

[164] V. Gromov et al., "Development and applications of the Timepix3 readout chip", *PoS* **VERTEX2011** (2011) 046.

[165] S. van der Meer, "Calibration of the effective beam height in the ISR", Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.

[166] S. M. White, H. Burkhardt, and P. Puzo, "Determination of the Absolute Luminosity at the LHC". PhD thesis, Orsay, Universit Paris-Sud 11, 2010.

[167] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[168] CMS Collaboration, "Luminosity Calibration for the 2013 Proton-Lead and Proton-Proton Data Taking", Technical Report CMS-PAS-LUM-13-002, CERN, Geneva, 2014.

[169] J. L. Leonard, "Luminosity measurement at CMS", in *Proceedings of the 3rd International Conference on Technology and Instrumentation in Particle Physics*. 2014.

[170] CMS Collaboration, "Measurements of luminosity and normalised beam-induced background using the CMS Fast Beam Condition Monitor", *PoS* **ICHEP2012** (2013) 526.

[171] CMS Collaboration, "Luminosity Monitoring and Determination by Z Decays to Muons in CMS", Technical Report CMS-PAS-LUM-13-003, CERN, Geneva, 2014.

[172] R. Hall-Wilton et al., "Results from a Beam Test of a Prototype PLT Diamond Pixel Telescope", Technical Report CMS-NOTE-2009-022. 1, CERN, Geneva, Aug, 2009.

[173] B. G. Taylor, "Optical timing, trigger and control distribution for LHC detectors", *IEEE Trans.Nucl.Sci.* **41** (1994) 1294.

[174] A. Taurok et al., "The central trigger control system of the CMS experiment at CERN", *JINST* **6** (2011) P03004.

[175] J. Troska et al., "Implementation of the timing, trigger and control system of the CMS experiment", *IEEE Trans.Nucl.Sci.* **53** (2006) 834.

[176] R. Arcidiacono et al., "The final prototype of the Fast Merging Module (FMM) for readout status processing in CMS DAQ", in *10th Workshop on Electronics for LHC Experiments, Boston, United states, 13-17 Sept 2004*. 2004.

[177] PICMG Collaboration, "MicroTCA specifications", technical report.

[178] I. Papakonstantinou et al., "A Fully Bidirectional Optical Network With Latency Monitoring Capability for the Distribution of Timing-Trigger and Control Signals in High-Energy Physics Experiments", *IEEE Trans.Nucl.Sci.* **58** (2011) 1628.

[179] S. Baron, "TTC upgrade plans, see https://indico.cern.ch/event/287628/other-view?view=standard", in *ACES 2014 - Fourth Common ATLAS CMS Electronics Workshop for LHC Upgrade*. 2014.

[180] CMS Collaboration, "CMS Technical Design Report for the Trigger and Data Acquisition project, Volume II", Technical Report CERN-LHCC-2002-026. CMS-TDR-6.2, 2002.

[181] G. Bauer et al., "Operational experience with the CMS Data Acquisition System", in *Proceedings of CHEP 2012*. 2012.

[182] T. Bawej et al., "The new CMS DAQ system for run-2 of the LHC", in *Real Time 2014 conference, submitted to IEEE Trans.Nucl.Sci.* 2014.

[183] I. Bird et al., "Update of the Computing Models of the WLCG and the LHC Experiments", Technical Report CERN-LHCC-2014-014. LCG-TDR-002, CERN, Geneva, Apr, 2014.

[184] J. Apostolakis et al., "The path toward HEP High Performance Computing", volume Conf. Ser. 513, p. 052006. 2014.

[185] M. Aderholz, "Models of Networked Analysis at Regional Centres for LHC Experiments (MONARC) - Phase 2 Report", *CERN/LCB* **2000-001**.

[186] K. Bloom, "CMS Use of a Data Federation", in *Proceedings of CHEP 2013*. 2013.

[187] S. H. Fuller and L. I. Millet, "The Future of Computing Performance: Game Over or Next Level?", *Computer* **44** (2011).

[188] C. D. Jones and E. Sexton-Kennedy, "Stitched Together Transitioning CMS to a Hierarchical Threaded Framework", volume Conf. Ser. 513, p. 022034. 2014.

[189] M. V. Wilkes, "The Memory Wall and the CMOS End-Point", *ACM Computer Architecture News* **23** (1995), no. 4,.

[190] G. M. Amdahl, "Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities", in *AFIPS Conference Proceedings*. 1967.

[191] CMS Collaboration, "Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and $\not\!\!E_{\mathrm{T}}$", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[192] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup Per Particle Identification", *JHEP* **1410** (2014) 59, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[193] S. Agostinelli et al., "Geant4a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250, doi:http://dx.doi.org/10.1016/S0168-9002(03)01368-8.

[194] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",.

[195] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV",.

[196] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[197] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[198] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the k_t jet-finder", *Phys. Lett. B* **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.

[199] CMS Collaboration, "Jet Energy Corrections and Uncertainties. Detector Performance Plots for 2012.",.

[200] C. Collaboration, "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS",.

[201] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[202] CMS Collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector", CMS Physics Analysis Summary CMS-PAS-JME-12-002, 2013.

[203] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013.

[204] CMS Collaboration, "Description and performance of CMS track reconstruction", CMS Physics Analysis Summary CMS-PAS-TRK-11-001, 2011.

[205] W. Waltenberger, R. Frühwirth, and P. Vanlaer, "Adaptive Vertex Fitting", *J. Phys. G* **34** (2007) N343, doi:doi:10.1088/0954-3899/34/12/N01.

[206] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[207] CMS Collaboration, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9, 2.76$, and 7 TeV", *JHEP* **04** (2013) 072, doi:10.1007/JHEP04(2013)072, arXiv:1302.2394.

[208] T. Sjöstrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Comm.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:hep-ph/0710.3820.

[209] OPAL Collaboration, "Precision Luminosity for Z0 Lineshape Measurements with a Silicon-Tungsten Calorimeter", *Eur. Phys. J. C* **14** (2000) 373, doi:10.1007/s100520000353, arXiv:9910066.

[210] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at LHC", *CMS Note* **2006/001** (2005).

[211] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector at $\sqrt{s}$=8 TeV", (2014). arXiv:1014.XXXX. To be submitted.

[212] CMS Collaboration, "Performance of photon reconstruction and selection in the CMS detector at $\sqrt{s}$=8 TeV", (2014). arXiv:1014.XXXX. To be submitted.

[213] CMS Collaboration, "The Fast Simulation of the CMS experiment", *Journal of Phyiscs: Conference Series* **513** (2014) 022012, doi:10.1088/1742-6596/513/2/022012.

[214] CMS Collaboration, "Performance of CMS muon reconstruction in cosmic-ray events", *JINST* **5** (2010) T03022, doi:10.1088/1748-0221/5/03/T03022, arXiv:0911.4994.

[215] CMS Collaboration, "CMS Physics Technical Design Report, volume I: Detector performance and software", TDR CERN-LHCC-2006-001, CMS-TDR-008-1, 2006.

[216] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A* **262** (1987) 444, doi:10.1016/0168-9002(87)90887-4.

[217] The CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s}$ = 7 TeV", *Journal of Instrumentation* **7** (October, 2012) 2P, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[218] CMS Collaboration, "Measurement of Tracking Efficiency", CMS Physics Analysis Summary CMS-PAS-TRK-10-002, 2010.

[219] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-003, 2010.

[220] Particle Data Group, J. Beringer et al., "Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[221] CMS Collaboration, "Evidence for the direct decay of the 125 GeV Higgs boson to fermions", *Nature Phys.* **10** (2014) 557, doi:10.1038/nphys3005, arXiv:1401.6527.

[222] C. Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons", *JHEP* **1405** (2014) 104, doi:10.1007/JHEP05(2014)104, arXiv:1401.5041.

[223] S. Berge, W. Bernreuther, and H. Spiesberger, "Higgs CP properties using the τ decay modes at the ILC", *Phys. Lett. B* **727** (2013) 488, doi:10.1016/j.physletb.2013.11.006, arXiv:1308.2674.

[224] CMS Collaboration, "Performance of tau-lepton reconstruction and identification in CMS", *JINST* **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001, arXiv:1109.6034.

[225] CMS Collaboration, "Study of tau reconstruction algorithms using pp collisions data collected at $\sqrt{s}$ = 7 TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-004, 2010.

[226] CMS Collaboration, "TauID Performance in Run 1", CMS Physics Analysis Summary CMS-PAS-TAU-14-001, 2014. in preparation.

[227] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", CMS Physics Analysis Summary CMS-PAS-JME-10-002, 2010.

[228] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[229] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV", *JHEP* **06** (2013) 081, arXiv:1303.4571.

[230] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[231] ATLAS, CMS Collaboration, "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments", arXiv:1503.07589.

[232] CMS Collaboration, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs", *Phys. Lett. B* **736** (2014) 64, doi:10.1016/j.physletb.2014.06.077, arXiv:1405.3455.

[233] R. S. Gupta, H. Rzehak, and J. D. Wells, "How well do we need to measure Higgs boson couplings?", *Phys. Rev. D* **86** (2012) 095001, doi:10.1103/PhysRevD.86.095001, arXiv:1206.3560.

[234] TLEP Design Study Working Group Collaboration, "First Look at the Physics Case of TLEP", *JHEP* **1401** (2014) 164, doi:10.1007/JHEP01(2014)164, arXiv:1308.6176.

[235] CMS Collaboration, "Search for a standard model-like Higgs boson in the $^+$ and e^+e decay channels at the LHC", *Phys. Lett. B* **744** (2015) 184, doi:10.1016/j.physletb.2015.03.048, arXiv:1410.6679.

[236] J. Baglio et al., "The measurement of the Higgs self-coupling at the LHC: theoretical status", *JHEP* **1304** (2013) 151, doi:10.1007/JHEP04(2013)151, arXiv:1212.5581.

[237] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[238] J. Grigo, K. Melnikov, and M. Steinhauser, "Virtual corrections to Higgs boson pair production in the large top quark mass limit", *Nucl. Phys. B* **888** (2014) 17, doi:10.1016/j.nuclphysb.2014.09.003, arXiv:1408.2422.

[239] C. E. Alan J. Barr, Matthew J. Dolan and M. Spannowsky, "Di-Higgs final states augMT2ed-selecting hh events at the high luminosity LHC", *Phys. Lett. B* **728** (2014) 308, doi:10.1016/j.physletb.2013.12.011.

[240] A. Ballestrero et al., "PHANTOM: A Monte Carlo event generator for six parton final states at high energy colliders", *Comput.Phys.Commun.* **180** (2009) 401, doi:10.1016/j.cpc.2008.10.005, arXiv:0801.3359.

[241] C. Degrande et al., "Effective Field Theory: A Modern Approach to Anomalous Couplings", *Annals Phys.* **335** (2013) 21, doi:10.1016/j.aop.2013.04.016, arXiv:1205.4231.

[242] K. Doroba et al., "The $W_L W_L$ Scattering at the LHC: Improving the Selection Criteria", *Phys. Rev. D* **86** (2012) 036011, doi:10.1103/PhysRevD.86.036011, arXiv:1201.2768.

[243] CMS Collaboration, "Study of the Discovery Reach in Searches for Supersymmetry at CMS with 3000 fb^{-1}", CMS Physics Analysis Summary CMS-PAS-FTR-13-014, 2013.

[244] CMS Collaboration, "Supersymmetry discovery potential in future LHC and HL-LHC running with the CMS detector", CMS Physics Analysis Summary CMS-PAS-SUS-14-012, 2014.

[245] M. Papucci, J. T. Ruderman, and A. Weiler, "Natural SUSY endures", *JHEP* **1209** (2012) 035, doi:10.1007/JHEP09(2012)035, arXiv:1110.6926.

[246] J. Alwall, P. Schuster, and N. Toro, "Simplified models for a first characterization of new physics at the LHC", *Phys. Rev.* **D 79** (2009) 075020, doi:10.1103/PhysRevD.79.075020, arXiv:0810.3921.

[247] D. Alves et al., "Simplified models for LHC new physics searches", *J. Phys. G* **39** (2012) 105005, doi:10.1088/0954-3899/39/10/105005, arXiv:1105.2838.

[248] CMS Collaboration, "Interpretation of Searches for Supersymmetry with simplified Models", *Phys. Rev. D* **88** (2013), no. 5, 052017, doi:10.1103/PhysRevD.88.052017, arXiv:1301.2175.

[249] CMS Collaboration, "Search for supersymmetry in pp collisions at in events with a single lepton, large jet multiplicity, and multiple b jets", *Phys. Lett. B* **733** (2014), no. 0, 328, doi:http://dx.doi.org/10.1016/j.physletb.2014.04.023.

[250] CMS Collaboration, "Exclusion limits on gluino and top-squark pair production in natural SUSY scenarios with inclusive razor and exclusive single-lepton searches at 8 TeV.", CMS Physics Analysis Summary CMS-PAS-SUS-14-011, 2014.

[251] CMS Collaboration, "Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s}$ = 8 TeV", *Eur. Phys. J. C* **73** (2013) 2677, doi:10.1140/epjc/s10052-013-2677-2, arXiv:1308.1586.

[252] CMS Collaboration, "Search for supersymmetry in hadronic final states using M_{T2} with the CMS detector at $\sqrt{s} = 8$ TeV", CMS Physics Analysis Summary CMS-PAS-SUS-13-019, 2013.

[253] CMS Collaboration, "Searches for electroweak production of charginos, neutralinos, and sleptons decaying to leptons and W, Z, and Higgs bosons in pp collisions at 8 TeV", *Eur. Phys. J.* **C 74** (2014), no. 9, 3036, doi:10.1140/epjc/s10052-014-3036-7, arXiv:1405.7570.

[254] DELPHES 3 Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **1402** (2014) 057, doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.

[255] J. Anderson et al., "Snowmass Energy Frontier Simulations", arXiv:1309.1057.

[256] A. Avetisyan et al., "Methods and Results for Standard Model Event Generation at $\sqrt{s}$ = 14 TeV, 33 TeV and 100 TeV Proton Colliders (A Snowmass Whitepaper)", arXiv:1308.1636.

[257] A. Avetisyan et al., "Snowmass Energy Frontier Simulations using the Open Science Grid (A Snowmass 2013 whitepaper)", arXiv:1308.0843.

[258] O. Buchmueller et al., "Likelihood functions for supersymmetric observables in frequentist analyses of the CMSSM and NUHM1", *Eur. Phys. J. C* **64** (2009) 391, doi:10.1140/epjc/s10052-009-1159-z, arXiv:0907.5568.

[259] J. R. Ellis, K. A. Olive, and Y. Santoso, "Calculations of neutralino stop coannihilation in the CMSSM", *Astropart. Phys.* **18** (2003) 395, doi:10.1016/S0927-6505(02)00151-2, arXiv:hep-ph/0112113.

[260] J. Ellis, K. A. Olive, and J. Zheng, "The Extent of the Stop Coannihilation Strip", arXiv:1404.5571.

[261] CMS Collaboration, "Search for top squarks decaying to a charm quark and a neutralino in events with a jet and missing transverse momentum", CMS Physics Analysis Summary CMS-PAS-SUS-13-009, 2013.

[262] A. Barr, C. Lester, and P. Stephens, "A variable for measuring masses at hadron colliders when missing energy is expected; m_{T2}: the truth behind the glamour", *J. Phys. G* **29** (2003) 2343, doi:10.1088/0954-3899/29/10/304, arXiv:arXiv:hep-ph/0304226.

[263] M. Burns, K. Kong, K. T. Matchev, and M. Park, "Using subsystem M_{T2} for complete mass determinations in decay chains with missing energy at hadron colliders", *JHEP* **0903** (2009) 143, doi:10.1088/1126-6708/2009/03/143, arXiv:arXiv:0810.5576.

[264] D. Tovey, "On measuring the masses of pair-produced semi-invisibly decaying particles at hadron colliders", *JHEP* **0804** (2008) 034, doi:10.1088/1126-6708/2008/04/034, arXiv:arXiv:0802.2879.

[265] G. Polesello and D. R. Tovey, "Supersymmetric particle mass measurement with the boost-corrected contransverse mass", *JHEP* **1003** (2010) 030, doi:10.1007/JHEP03(2010)030, arXiv:0910.0174.

[266] P. Langacker, "The Physics of Heavy Z' Gauge Bosons", *Rev. Mod. Phys.* **81** (2009) 1199, doi:10.1103/RevModPhys.81.1199, arXiv:0801.1345.

[267] L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[268] CMS Collaboration, "Search for physics beyond the standard model in final states with a lepton and missing transverse energy in proton-proton collisions at $\sqrt{s}$ = 8 TeV", arXiv:1408.2745.

[269] O. Buchmueller, M. J. Dolan, and C. McCabe, "Beyond Effective Field Theory for Dark Matter Searches at the LHC", *JHEP* **1401** (2014) 025, doi:10.1007/JHEP01(2014)025, arXiv:1308.6799.

[270] O. Buchmueller, M. J. Dolan, S. A. Malik, and C. McCabe, "Characterising dark matter searches at colliders and direct detection experiments: Vector mediators", *JHEP* **1501** (2015) 037, doi:10.1007/JHEP01(2015)037, arXiv:1407.8257.

[271] C. Collaboration, "Searches for long-lived charged particles in pp collisions at $\sqrt{s}$=7 and 8 TeV", arXiv:1305.0491.

[272] Z. Han, G. D. Kribs, A. Martin, and A. Menon, "Hunting Quasi-Degenerate Higgsinos", *Phys. Rev. D* **89** (2014) 075007, doi:10.1103/PhysRevD.89.075007, arXiv:1401.1235.

[273] CMS Collaboration, "Measurement of the B(s) to mu+ mu- branching fraction and search for B0 to mu+ mu- with the CMS Experiment", *Phys. Rev. Lett.* **111** (2013) 101804, doi:10.1103/PhysRevLett.111.101804, arXiv:1307.5025.

[274] R. Bruce et al., "Sources of machine-induced background in the ATLAS and CMS detectors at the CERN Large Hadron Collider", *Nucl.Inst.Meth.A* **729** (2013) 825.

[275] F. Schmidt, "SixTrack. Users Reference Manual", Technical Report CERN/SL/94-56-AP, CERN, 1994.