Optimizing integrated luminosity of future hadron colliders

Michael Benedikt, Daniel Schulte, and Frank Zimmermann
CERN, 1211 Geneva 23, Switzerland
(Received 1 June 2015; published 13 October 2015)

The integrated luminosity, a key figure of merit for any particle-physics collider, is closely linked to the peak luminosity and to the beam lifetime. The instantaneous peak luminosity of a collider is constrained by a number of boundary conditions, such as the available beam current, the maximum beam-beam tune shift with acceptable beam stability and reasonable luminosity lifetime (i.e., the empirical “beam-beam limit”), or the event pileup in the physics detectors. The beam lifetime at high-luminosity hadron colliders is largely determined by particle burn off in the collisions. In future highest-energy circular colliders synchrotron radiation provides a natural damping mechanism, which can be exploited for maximizing the integrated luminosity. In this article, we derive analytical expressions describing the optimized integrated luminosity, the corresponding optimum store length, and the time evolution of relevant beam parameters, without or with radiation damping, while respecting a fixed maximum value for the total beam-beam tune shift or for the event pileup in the detector. Our results are illustrated by examples for the proton-proton luminosity of the existing Large Hadron Collider (LHC) at its design parameters, of the High-Luminosity Large Hadron Collider (HL-LHC), and of the Future Circular Collider (FCC-hh).

DOI: 10.1103/PhysRevSTAB.18.101002 PACS numbers: 29.20.D-, 29.27.-a, 13.75.Cs, 13.85.-t

I. INTRODUCTION

So far five hadron colliders have been in operation [Intersecting Storage Rings (ISR), Super Proton Synchrotron (S$p\bar{p}$S), Tevatron, Relativistic Heavy Ion Collider (RHIC), and LHC]. Their achievements and primary performance limitations can roughly be summarized as follows [1–3]. The CERN ISR started operation in 1970. A double ring collider, it reached a peak luminosity of $2.2 \times 10^{32}$ cm$^{-2}$ s$^{-1}$ and a maximum beam energy of 31 GeV with coasting beams of 38–50 A current each. The ISR luminosity was limited by space-charge tune shift and spread, coherent beam-beam effects, proton-electron two-stream instabilities, pressure bumps, detector background, and accumulation efficiency [4]. The ISR also produced the first $p\bar{p}$ collisions, and, when operated with bunched beams, it reached a beam-beam tune shift of $\Delta Q = 0.0035$ per interaction point (IP) with 8 crossings [5]. The second hadron collider was the CERN S$p\bar{p}$S operating since 1981 at ten times higher energy than the ISR. The S$p\bar{p}$S discovered the W and Z bosons. Its luminosity was limited by beam-beam interaction, loss of longitudinal Landau damping, number of available antiprotons, hourglass effect, and intrabeam scattering (IBS) [6–8]. The typical beam-beam tune shift was $\Delta Q = 0.005$ at each of three interaction points. The FNAL Tevatron was the first collider based on superconducting magnets. Colliding-beam operation here started in 1987 [9]. Tevatron luminosity was limited by antiproton intensity, beam-beam interaction including long-range effects, luminosity lifetime, number of events per crossing, and IBS. It reached an antiproton beam-beam tune shift of $\Delta Q \approx 0.015$ for each of two collision points [10,11]. The Tevatron discovered the bottom and top quarks. RHIC at BNL, the first heavy-ion collider as well as the first polarized-proton collider, delivers luminosity since 2000. The main limiting factor is intrabeam scattering. Other factors again are beam-beam interaction, luminosity lifetime, and the number of events per crossing [12]. In 2010, 3-D stochastic cooling was successfully implemented in RHIC, reducing the transverse emittances of heavy-ion beams by a factor of 5 [13].

At present, the Large Hadron Collider (LHC) [14,15] is the world’s highest-energy proton collider. Thanks to its elevated energy the LHC is the first machine where radiation damping is expected to dominate over IBS [1]. The scarcity of antiprotons is no longer a problem, as LHC and all future hadron machines will collide protons on protons (or heavy ions against heavy ions) [16,17]. As for the Tevatron, LHC luminosity performance will be limited by beam-beam interaction, luminosity lifetime, and events per crossing.

In addition, for the LHC or any future hadron collider various other phenomena or constraints may be important, such as the electron cloud produced by photoemission or beam-induced multipacting [18–24], beam instabilities [25–28], local beam losses and the efficiency of the collimation system [29–36], and, especially for heavy-ion operation, local magnet quenches induced by the...
collision products [37,38]. None of these potential additional limitations are included in our following analysis.

In 2015 the LHC center-of-mass energy has reached 13 TeV. Extremely preliminary observations indicate that in physics fills at this energy the vertical emittance is approximately constant in time, while the horizontal emittance growth rate is about half as large as it had been at 8 TeV in 2012 [39,40]. The reduced emittance growth may be attributed to the beneficial side effect of increased radiation damping.

The LHC design luminosity is $10^{34}$ cm$^{-2}$ s$^{-1}$ at 14 TeV. By around 2023-24 the LHC will be upgraded to the High-Luminosity LHC (HL-LHC) [41–43]. The HL-LHC should operate at a constant “leveled” luminosity of $5 \times 10^{34}$ cm$^{-2}$ s$^{-1}$, and result in a 10-fold increase of the LHC annual integrated luminosity.

Recently launched in response to a request from the 2013 Update of the European Strategy of Particle Physics [44], the primary focus of the global Future Circular Collider (FCC) study [45] is the design of a 100-TeV hadron collider, called FCC-hh, which could go into operation soon after the end of the HL-LHC program in the mid 2030s.

The development of the FCC-hh can profit from the design studies for previously considered large hadron colliders, such as the ill-fated Superconducting Super Collider (SSC) in Texas [46,47], a Very Large Hadron Collider (VLHC) in Illinois [48], a High Energy LHC (HE-LHC) [48], and the, still ongoing, INFN Eloisatron Project in Italy [49].

The most important figure of merit of any collider, together with its collision energy (and possibly level of beam polarization), is the useful luminosity it produces over time, i.e., its integrated luminosity delivered to one or several particle-physics detectors.

In this article, we discuss the optimization of the integrated luminosity—e.g., through emittance control and judicious choice of store lengths—for both present and future highest-energy proton colliders, the latter profiting from strong synchrotron radiation damping.

Our subsequent discussion is based entirely on analytical calculations. Previous work for future proton-proton colliders employed a slower numerical solution of the underlying differential equations to determine optimized store lengths and average luminosities [50]. Other earlier studies optimized the integrated luminosity of hadron colliders by adopting conditions or assumptions often quite different from those considered here. For example, the integrated luminosity for the LHC luminosity upgrade was first considered in Ref. [51]; the integrated luminosity of the Tevatron was maximized as a function of the antiproton production rate based on a heuristic model for the luminosity decay [52]; and in Ref. [53] the time-dependent luminosity of a generic collider was modeled as an exponential decay.

Numerous past studies addressed the optimization of hadron-collider peak luminosity, e.g., by shaping the longitudinal bunch profile [54] and increasing the crossing angle [54]; on the contrary by introducing crab cavities [55–59], i.e., radiofrequency cavities supporting a transversely deflecting mode, which can be employed in order to tilt the head and tail of a bunch in opposite direction so that either the collision becomes effectively head-on while the bunch centroids still cross at a nonzero angle, or so that they create an intentional bunch tilt at the collision point for the purpose of luminosity leveling in the so-called “crab-kissing scheme” [60]; by operating with flat beams and using special sextupole magnets to introduce a “crab waist” [61–63]; or by various beam-beam compensation schemes [64–76]. Also the parameter optimization of synchrotron-radiation dominated hadron colliders was discussed on various occasions [1,49,77,78], as were models for higher-level machine complexity and cost [79–82].

Of course, the integrated luminosity is strongly (but not uniquely) correlated with the peak luminosity as hinted at in Ref. [83]. The maximum possible peak luminosity of a collider is in practice limited by several factors, in particular, (i) by the maximum value for the total available beam current, which may, e.g., be determined by the injector complex or by heat load from synchrotron radiation; (ii) by an empirical maximum acceptable value for the total beam-beam tune shift, above which poor beam lifetime, emittance growth, or beam instabilities are encountered; and (iii) by a maximum value for the number of pileup events per bunch crossing, i.e., for the number of inelastic scattering events per bunch collision observed by the particle-physics detectors above which the quality of the data collected for physics analyses rapidly degrades.

II. BEAM-BEAM TUNE SHIFT, PEAK LUMINOSITY, PILEUP, AND BURN OFF

For the Future Circular Collider (FCC), in addition to a limited total beam current, the second restriction may be set by the beam-beam tune shift, since the pileup limit could, in principle, be overcome, e.g., with advanced detectors or, more simply, by reducing the bunch spacing, and increasing the number of bunches, while reducing the transverse emittance in proportion to the bunch intensity.

The LHC was designed for a maximum total beam-beam tune shift of 0.01. However, in actual operation during 2011–12 the total LHC beam-beam tune shift has been more than twice this value, and in dedicated beam-beam studies values of up to, and above, 0.03 have been attained without any noticeable effect on the beam quality or on the luminosity lifetime [84–86]. A beam-beam limit above 0.03 also is consistent with the results of strong-strong and weak-strong computer simulations of the LHC beam-beam collisions [87]. Mimicking the LHC design, the FCC baseline assumes a total tune shift of 0.01 for the two main experiments, but, given the actual LHC experience, in the examples presented later we will also consider a higher beam-beam limit of 0.03.
The HL-LHC relies on a constant instantaneous luminosity, not exceeding $5 \times 10^{34} \text{ cm}^{-2} \text{s}^{-1}$, which corresponds to approximately 140 events on average per bunch crossing for operation with 25 ns bunch spacing (and, more precisely, with about 2750 bunches per beam, see Table I). This is achieved through challenging luminosity leveling techniques, for instance via a gradual reduction of the beta function at the interaction point, $\beta^*$, taken to be equal in the two transverse planes, and for the two colliding beams.

We consider the collision of round Gaussian beams with equal horizontal and vertical geometric rms emittances, $\epsilon \equiv \epsilon_x = \epsilon_y$, and equal beta functions at the collision point, $\beta^* \equiv \beta_x^* = \beta_y^*$. Further we assume the presence of crab cavities around each collision point (i.e., no luminosity loss from a finite crossing angle), and we also neglect the hourglass effect (which is permitted for $\beta^*$ values larger than the rms bunch length). Under these assumptions—which, for example, are well fulfilled for the parameters of the HL-LHC—the luminosity $L$ at a collision point is given by

$$L = \frac{\nu_{\text{rev}} n_b N_b^2}{4\pi \beta^* \epsilon},$$

where $f_{\text{rev}}$ denotes the revolution frequency, $N_b$ the bunch population (the number of protons per bunch), $n_b$ the number of bunches per beam, and $\epsilon$ the geometric rms emittance.

We also suppose that there are $n_{\text{IP}}$ collision points around the ring circumference, and, for simplicity, that the luminosity is equal for all of them. Under the same assumptions as made for the luminosity, the total beam-beam parameter (or the approximate beam-beam tune shift) becomes

$$\Delta Q_{\text{tot}} = r_p n_b \frac{N_b}{4\pi \gamma^* \epsilon},$$

where $r_p$ designates the “classical proton radius” (about $1.5 \times 10^{-18}$ m), and $\gamma$ the relativistic Lorentz factor.

The number of inelastic scattering events per bunch collision $\mu$ (approximately equal to the event pileup in the particle physics detectors) can be calculated from the formula

$$\mu \equiv \sigma_{\text{inel}} \frac{L}{n_b f_{\text{rev}}} = \sigma_{\text{inel}} \frac{N_b^2}{4\pi \beta^* \epsilon}.$$
where $\sigma_{\text{inel}}$ refers to the (inelastic) cross section for any event “seen” by the particle-physics experiment.

If we take the dominant source of beam loss to be due to the burn-off in collision (a good assumption for machines like the LHC and HL-LHC), the rate of change of the bunch intensity is proportional to the instantaneous luminosity as

$$\frac{dN_b}{dt} = -\sigma_{\text{tot}} n_{\text{IP}} \frac{L}{n_b} = -\frac{\sigma_{\text{tot}} f_{\text{rev}} n_{\text{IP}} N_b^2}{4\pi\beta'} e \equiv -K N_b^2,$$  \hspace{1cm} (4)

where $\sigma_{\text{tot}}$ denotes the total cross section (about 100 mbarn at 14 TeV center of mass energy; comprising inelastic and elastic components), and $L$ the luminosity at each IP (assumed to be the same), and, for later use, we have introduced the parameter $K$:

$$K \equiv \frac{\sigma_{\text{tot}} f_{\text{rev}} n_{\text{IP}}}{4\pi\beta'}.$$ \hspace{1cm} (5)

The character of the solution of Eq. (4) differs according to the constraints and assumptions imposed. In particular it depends on the behavior of the emittance.

The energy-dependent total and inelastic cross sections, $\sigma_{\text{tot}}$ and $\sigma_{\text{inel}}$, can be estimated from the scaling laws [50,88–90],

$$\sigma_{\text{tot}}[\text{mbarn}] \approx 42.1 s^{-0.467} - 32.19 s^{-0.540} + 35.83$$
$$+ 0.315 \ln^2 \left( \frac{s}{34.0} \right),$$ \hspace{1cm} (6)

and [91,92]

$$\sigma_{\text{inel}}[\text{mbarn}] \approx \sigma_{\text{tot}} - 11.7 + 1.59 \ln s - 0.134 \ln^2 s,$$ \hspace{1cm} (7)

where $s$ designates the square of the center-of-mass energy in units of GeV$^2$. At 14 TeV c.m. the total and inelastic cross sections are then expected to be 112 and 83 mbarn, respectively [88]. At 100 TeV they will increase to 156 and 110 mbarn.

### III. Luminosity Decay and Optimum Fill Length with Constant Emittance (LHC)

For the nominal LHC, we may assume that the transverse damping due to synchrotron radiation roughly balances the blow up from intrabeam scattering [1,39], as well as other smaller contributions, such as the effect of residual-gas scattering or the emittance growth due to the core depletion in collision [93,94], and that, as a result, the transverse emittances $e \equiv e_{x,y}$ stay roughly constant during a physics fill.

In this case the solution of (4) is an algebraic decay of the beam intensity

$$N_b(t) = \frac{N_0}{1 + t/\tau_b},$$ \hspace{1cm} (8)

where $N_0 \equiv N_b(0)$ refers to the initial intensity and

$$\tau_b \equiv e \frac{4\pi\beta' e}{KN_0} = \frac{N_0 N_b}{L(0) \sigma_{\text{tot}} n_{\text{IP}}},$$ \hspace{1cm} (9)

the (initial) beam lifetime.

Inserting (8) into (1) we obtain the luminosity evolution

$$L(t) = L(0) \frac{1}{(1 + t/\tau_b)^2},$$ \hspace{1cm} (10)

and the integrated luminosity (per IP) after time $t$ becomes

$$\Sigma(t) = \int_{t'=0}^{t} L(t') dt' = \tau_b L(0) \frac{t/\tau_b}{1 + t/\tau_b}. \hspace{1cm} (11)$$

Suppose for an instant that the transverse emittances were not constant, but increased linearly in time with a rise time $\tau_e$, i.e.,

$$\epsilon_{x,y}(t) \approx \epsilon_0 \left( 1 + \frac{t}{\tau_e} \right).$$ \hspace{1cm} (12)

Then, also assuming $\tau_e \gg \tau_b$, Eq. (8) remained approximately valid, Eq. (10) transformed into

$$L_e(t) \approx L(0) \frac{1}{(1 + t/\tau_b)^2 (1 + t/\tau_e)},$$ \hspace{1cm} (13)

and the integrated luminosity (per IP) after time $t$ became

$$\Sigma_e(t) \approx \int_{t'=0}^{t} L_e(t') dt'$$
$$= \tau_b L(0) \left[ \frac{\tau_e t}{(\tau_e - \tau_b)(\tau_e + t)} + \frac{\tau_b \tau_e}{(\tau_b - \tau_e)^2} \ln \frac{\tau_b (\tau_e + t)}{\tau_e (\tau_b + t)} \right].$$ \hspace{1cm} (14)

In the limit $\tau_e \to \infty$ this reduces to formula (11).

Next, going back to the case of constant emittance, described by (11) and representing the LHC design baseline, we wish to optimize the physics fill time, i.e., the run time $t_r$ which maximizes the average luminosity. To this end we introduce the (mean) “turnaround time” $t_{\text{tr}}$, i.e., the average time between the end of one physics fill at time $t_r$ and the start of the luminosity production in the following fill, at (on average) the time $t_r + t_{\text{tr}}$.

The turnaround time $t_{\text{tr}}$ typically comprises the time periods spent on ramp down, injection setup, injection, acceleration, collimation setup, and $\beta''$ squeeze, until it is safe for operation to restart the data taking. The average turnaround time also includes time spent on resolving any technical problems, waiting periods due to nonavailability
of the injectors, etc. Relevant experience with the turn-around times at Tevatron, RHIC, and the lepton-hadron collider HERA was compiled in Ref. [95]. Limits on the LHC turn-around time are discussed in Ref. [96]. For the FCC we assume that its turn-around time will be similar to the one of the LHC (i.e., about 5 hours on average in the first years of operation).

The average luminosity depends on the length of a physics run, \( t_r \), as

\[
L_{\text{ave}}(t_r) = \frac{\Sigma(t_r)}{t_r + t_a} = L(0) \frac{t_r}{(1 + t_r/\tau_b)(t_r + t_a)}. \tag{15}
\]

It is maximized if the run time equals the geometric mean of the initial beam lifetime and the turn-around time:

\[
t_{r, \text{opt}} = \sqrt{\tau_b t_a}. \tag{16}
\]

For the optimum run time the average luminosity becomes

\[
L_{\text{ave}} = L(0) \frac{\tau_b}{(\tau_b + \sqrt{t_a})^2}. \tag{17}
\]

We note that \( \tau_b \) depends on \( L(0) \) and on the total beam intensity, according to (9).

In this situation, both the beam-beam tune shift and the event pileup are maximum at the start of a physics fill and then decrease, like the bunch intensity and the luminosity, respectively. These solutions were presented, for example, in Ref. [97].

To give a numerical example, for the nominal LHC parameters of Table I and a turn-around time of 5 h, typical for the year 2012 [96], the optimum run time equals 14.2 h and the average luminosity (17) becomes

\[ L_{\text{ave, LHC}} \approx 5.5 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}. \]

The integrated annual luminosity can simplistically be estimated by multiplying the total time scheduled for physics production \( T_{\text{tot}} \), the machine availability \( A \) (time without hardware failures divided by total time scheduled), and the average luminosity over the time periods without any hardware failure, as

\[
\Sigma(1 \text{ year}) = \int_{\text{year}} L(t) \, dt = T_{\text{tot}} A L_{\text{ave}}. \tag{18}
\]

Considering an LHC availability of 71%, i.e., the value achieved in the year 2012 [96], and assuming that \( T_{\text{tot}} = 160 \) days are scheduled for physics per year, we can use (18) to estimate the integrated luminosity for the nominal LHC as 55 fb\(^{-1}\) per year.

Defining the machine efficiency, \( \eta \), as the time spent in physics divided by the total allocated calendar time, the latter may be related to the availability \( A \), and to the average run time \( t_r \), so as to estimate the minimum needed efficiency as

\[
\eta \approx A \frac{t_r}{t_r + t_a}. \tag{19}
\]

In 2012 the average run time of the LHC was about 6 h, much lower than the optimum run time 14.2 h. Inserting the actual availability of \( A \approx 71\% \) and the actual average run time of 5.5 h, (19) yields \( \eta \approx 37\% \), which matches the value of \( \eta \) reported from the LHC 4-TeV operation in 2012 [98].

IV. LUMINOSITY AND OPTIMUM FILL LENGTH WITH LEVELING (HL-LHC)

The detector technology sets limits on the total number of events per crossing (e.g., for calorimetry), as well as on the longitudinal event line density (for tracking of the primary vertices) and, possibly, also on the number of events per unit time during the collision. The nominal LHC parameters correspond to a peak pileup of about 20 events per crossing. During the 2012 LHC run the average pileup already was about 21 for ATLAS and CMS, and the maximum pileup in physics runs about 40 (i.e., twice the design). Higher peak values, close to 80 events per crossing, were reached in dedicated machine studies with a few bunches. The upgraded ATLAS and CMS detectors of the HL-LHC are expected to cope with an average pileup of \( \mu_{\text{tot}} = 140 \) events per bunch crossing, and tails up to 200.

The HL-LHC machine is designed so as to achieve a “virtual” peak luminosity \( \hat{L} \) considerably higher than the maximum value imposed by the acceptable event pile. During operation the peak luminosity will be controlled and reduced (“luminosity leveling”) in order to sustain the operational luminosity, and the associated event pileup, at a constant level over a significant length of time [41–43].

This luminosity leveling during a physics store can be accomplished in a number of ways [41–43,97]:

(i) dynamic \( \beta^* \) squeeze, (ii) crossing angle variation, (iii) changes in the crab rf voltage, including the elegant “crab kissing” scheme [60], (iv) dynamic bunch-length reduction, or (v) controlled variation of the transverse distance between the two colliding beams. In each case a different collision-related beam parameter is varied as a function of time during the store, so as to lower the actual luminosity compared with the maximum achievable value at the beginning of the physics collisions, and, conversely, to increase the specific luminosity later during the fill. As a result the luminosity remains roughly constant, “leveled,” during the data taking of the particle-physics experiments, thereby avoiding periods with extremely high or extremely low luminosity. Depending on the parameter being varied the beam-beam tune shift may increase, decrease, or remain unaffected as a result of the leveling. Also the longitudinal extent of
the luminous region is affected differently by the various leveling methods.

Due to proton consumption in the collisions, the bunch intensity, $N_b$, decays according to (4), with $L$ equaling the leveled luminosity $L_{lev}$, namely

$$\frac{dN_b}{dt} = -\sigma_{tot} n_{IP} \frac{L_{lev}}{n_b}.$$  \hspace{1cm} (20)

As a consequence, the beam intensity decreases linearly in time and the effective initial beam lifetime is

$$\tau_{eff} = \frac{N_0 n_b}{\sigma_{tot} L_{lev}},$$  \hspace{1cm} (21)

where $N_0 = n_b(0)$. Next, introducing the ratio of virtual peak luminosity and leveled luminosity, $k = \hat{L}/L_{lev}$, we can express the maximum leveling time as

$$t_{lev} = \tau_{eff} \left(1 - \frac{1}{\sqrt{k}}\right) \equiv \tau_{eff} F,$$  \hspace{1cm} (22)

where $F \equiv (1 - 1/\sqrt{k})$ designates the ratio of leveling time and effective beam lifetime. For the general case, where the physics run is extended beyond the end of the leveling period by a certain decay time $t_{dec}$ (see Fig. 1), the time-averaged luminosity becomes [99]

$$L_{ave} = L_{lev} \frac{t_{lev} + t_{dec} \tau_{eff} / \left(t_{dec} + \tau_{eff}\right)}{t_{dec} + t_{lev} + t_a},$$  \hspace{1cm} (23)

with $t_a$ denoting the average turnaround time, i.e., the time between the end of one physics run and the start of the next one (time needed for magnet ramp down, injection, acceleration, $\beta^*$ squeeze, collimation setup, etc.).

In this case the average luminosity assumes a maximum value, $L_{ave, opt}$ if $t_{dec}$ equals the optimum decay time [43,99]:

$$t_{dec,opt} = \tau_{eff} \frac{1 + F}{1 + F + \sqrt{F^2 + (1 + F)t_a / \tau_{eff}}}.$$  \hspace{1cm} (24)

The larger the turnaround time $t_a$ is compared with the effective lifetime $\tau_{eff}$, the longer is the optimum decay time. The resulting optimum total length of a single run becomes

$$t_{r,opt} = t_{lev} + t_{dec,opt}.$$  \hspace{1cm} (25)

The variation of the beam-beam tune shift during a physics store depends on the leveling scheme [100]. In case of $\beta^*$ variation, the beam-beam tune shift is maximized at the beginning, but then decreases during the store. Conversely, when leveling via the bunch length, crossing angle, or crab voltage the tune shift is minimized at the beginning, but then increases, by a factor of 2 to 3. When leveling with the transverse offset the beam-beam tune shift changes sign and its modulus can increase even more strongly during the leveling process, by up to an order of magnitude [100].

For the HL-LHC target parameters in Table I, we find $t_{dec,opt} \approx 3.34$ h. If we require $L_{tot}$ to be 250 fb$^{-1}$, and again consider $T_{tot} = 160$ days (per year), we can use (18) to deduce the minimum availability required to be about 52%, which may be compared with the actual LHC availability of 71% in the year 2012 [96]. Assuming this value (52%) for the minimum machine availability $A$ required, together with an optimum HL-LHC run time $t_r$ of 11.7 h, and an average turnaround time of 5 hours, from (19) the corresponding minimum needed efficiency $\eta$ becomes 36.4%, which is comparable to the efficiency of about 36.5% obtained during LHC 4-TeV operation in 2012 [98], which, therefore, appears an achievable target. More refined estimates of integrated luminosities or necessary efficiencies might be obtained by considering a realistic random run-time distribution of (prematurely aborted) physics stores.

Figure 1 illustrates the HL-LHC luminosity time evolution for a single fill with and without leveling (here, all the curves shown are extended beyond the respective optimum run times). The curve for the nominal LHC is also included for comparison. Figure 2 displays the luminosity evolution with and without leveling over several successive fills, where the fill length for either case has been optimized for maximum luminosity, assuming a turnaround time of 5 hours. Without leveling the optimized fill length would be less than half the fill length of the leveled case, implying a significant increase in the fraction of time spent for turnaround without any luminosity. The solid lines indicate the time-averaged luminosities with and without leveling. The difference is only some 25% while the peak luminosity differs by 300%.
Finally, Fig. 3 shows the integrated luminosity as a function of time. It is about \(4 \text{ fb}^{-1}\) per day. With an availability as low as 40% and considering 160 days scheduled for physics per calendar year, the luminosity delivered per year will exceed \(250 \text{ fb}^{-1}\).

V. OPTIMIZING LUMINOSITY WITH RADIATION DAMPING (FCC-HH)

A. Natural emittance shrinkage

Damping due to synchrotron radiation is significant at future hadron colliders, both transversely and even more longitudinally. However, in the following we assume that the bunch length and the longitudinal emittance are kept constant via continuous controlled excitation with "pink noise," which counteracts the natural emittance shrinkage \([1,50]\). Already nowadays, such automated continuous longitudinal excitation is routinely applied during the energy ramp of the LHC \([101]\). At the LHC, the control of the longitudinal beam profile avoids the harmful consequences of shorter bunches, such as unacceptable heating of kicker ferrites \([102]\) or beam instabilities due to loss of Landau damping, observed during LHC commissioning as well as in dedicated beam studies \([103]\).

Transversely we can better profit from the radiation damping. In the first period of a physics run, after the start of the collisions at time 0, both transverse emittances of either beam are allowed to shrink under the influence of the naturally occurring synchrotron radiation damping, as

\[
e(t) = e_0 \exp \left( -\frac{t}{\tau} \right) + e_{eq} \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right],
\]

where the parameter \(\tau\) signifies the transverse emittance damping time, and \(e_{eq}\) denotes the equilibrium emittance. In the horizontal plane, \(e_{eq}\) arises from quantum excitation in conjunction with the nonzero design dispersion. Its value is given by \(e_{eq} \approx C_q \theta_{hc}^2 F\), where \(\theta_{hc}\) denotes the bending angle per half cell (0.0165 rad for the LHC, 0.001 rad for the FCC-hh), \(F\) is a numerical factor of order 1 which depends on the optics and the filling factor \([104]\) (\(F \approx 3.1\) for a FODO cell with 90 degree phase advance and 80% dipole-magnet filling factor in the arcs), and \(C_q = 55/(32\sqrt{3})hc/(m_pc^2) \approx 2.09 \times 10^{-16} \text{ m}\) designates the quantum constant for protons. For the hadron colliders which we consider in this paper (LHC, HL-LHC and FCC-hh) the equilibrium emittances from synchrotron radiation are two or three orders magnitude smaller than the emittances optimized for use in operation (see Tables I and II). Reasons for the small equilibrium emittances include the combination of large proton mass, low photon energies (with critical energies ranging from a few tens of eV for the LHC to a few keV for the FCC-hh), the low dispersion and the small bending angle per
half cell (thanks to the large ring circumferences). We may, therefore, apply the approximation $\epsilon_{eq} \ll \epsilon_0$ and, over the time period of interest (length of a fill), also $\epsilon_{eq} \ll \epsilon(t)$, so that the emittance evolution of (26) becomes

$$e(t) = \epsilon_0 \exp \left( -\frac{t}{\tau} \right). \quad (27)$$

Inserting (27) into (4) the time dependent bunch intensity then fulfills

$$\frac{dN_b}{dt} = -K \frac{N_b^2}{\epsilon_0} e^{t/\tau}. \quad (28)$$

The integration of (28) can be performed by separation of variables

$$\int_{N_0}^{N_b(t)} \frac{1}{N_b^2} dN_b = - \int_0^t \frac{1}{\epsilon_0} e^{t/\tau} dt, \quad (29)$$

with the result

$$N_b(t) = \frac{1}{\epsilon_0} \left( e^{t/\tau} - 1 \right) + \frac{1}{N_0}, \quad (30)$$

where we have abbreviated the initial bunch intensity as $N_0 = N_b(0)$.

A similar set of equations would describe the intensity evolution for heavy-ion collisions at RHIC, where instead of synchrotron radiation a stochastic-cooling system shrinks all three beam emittances during a physics store. However, unlike the radiation damping times, the stochastic cooling rates depend on the bunch intensity and on the instantaneous emittances. For this reason, the final formulas derived in this section are not directly applicable to RHIC.

**B. Running at the beam-beam limit**

The beam-beam limit in hadron colliders has been the subject of much debate [105–115] and still is an active research area. For example, recent simulation studies explore the ultimate limit for the LHC [87]. A large body of actual experience from the Tevatron, RHIC, and LHC is reported in the literature, e.g., [116–121]. In general no well-defined limit exists. However, it is reasonable to assume that for each hadron collider there exists a, possibly different, maximum value of the beam-beam tune shift which should not be exceeded, since otherwise emittance growth, intensity loss rate, or detector background would become unacceptably high. In addition, we may expect that, with further increasing strength of radiation damping, the beam-beam limit in highest-energy hadron colliders will start to resemble the beam-beam limit.
in electron-positron ring colliders, for which above a certain value of \( \Delta Q_{\text{tot}} \) one of the transverse emittances blows up in proportion to the beam intensity [122]. We, therefore, assume that the emittances may continue to shrink until a maximum acceptable value for the total beam-beam tune shift \( \Delta Q_{\text{max}} \) (the “beam-beam limit”) is reached at a time \( t_1 \). By definition, this time \( t_1 \) follows from the tune shift limit as

\[
\Delta Q_{\text{max}} = \Delta Q(t_1) = n_{\text{IP}} \frac{r_p N_b(t_1)}{4\pi\gamma\epsilon_0} \frac{e^{\epsilon_1/\tau}}{e^{\epsilon_1/\tau} - 1} + \frac{1}{N_0},
\]

where

\[
\Lambda = n_{\text{IP}} \frac{r_p}{4\pi\gamma\epsilon_0}.
\]

From (31) we deduce the time limit for period 1:

\[
t_1 = -\tau \ln \left[ \frac{\Delta Q_{\text{max}} - \frac{\Lambda}{N_0}}{\frac{\Lambda}{N_0} - \frac{\Lambda}{\epsilon_0}} \right].
\]

Now the integrated luminosity for period 1 becomes

\[
\Sigma_1 = \int_0^{t_1} L(t) dt = \frac{f_{\text{rev}} n_b}{4\pi\beta^*} \int_0^{t_1} N_b^2 \frac{e^t}{e^{t}} dt = \frac{f_{\text{rev}} n_b N_0^2}{4\pi\beta^*\epsilon_0} \left[ \frac{1}{B(e^{t}/\tau - 1) + 1} \right]^2 dt.
\]

With the additional, dimensionless parameter

\[
B = \frac{K\tau N_0}{\epsilon_0},
\]

the integral \( \Sigma_1 \) of (34) can be solved as

\[
\Sigma_1 = \frac{f_{\text{rev}} n_b N_0^2}{4\pi\beta^*\epsilon_0} \frac{\tau}{B} \left( 1 - \frac{1}{1 - B + Be^{\epsilon_1/\tau}} \right).
\]

Inserting (33) into (36) yields

\[
\Sigma_1 = \frac{N_0 n_b}{n_{\text{IP}}\sigma_{\text{tot}}} \left[ 1 - \frac{1}{1 + \frac{f_{\text{rev}}\sigma_{\text{tot}}}{4\pi\epsilon_0} \left( \frac{N_0 n_p r_p e^{-\epsilon_1/\tau} B Q_{\text{max}}}{f_{\text{rev}}\sigma_{\text{tot}} B Q_{\text{max}} - \beta^* r_p} \right) \right].
\]

which may be simplified to

\[
\Sigma_1 = n_{\text{IP}} \frac{n_p \gamma_
u}{r_p B Q_{\text{max}}} f_{\text{rev}} \gamma \left( \frac{\Delta Q_{\text{max}} - \Delta Q_0}{1 - \frac{f_{\text{rev}}\gamma\sigma_{\text{tot}}}{r_p B Q_{\text{max}}}} \right)
\]

where

\[
\Delta Q_0 = n_p \frac{r_p N_0}{4\pi\gamma\epsilon_0}
\]

denotes the initial tune shift.

To give an example, considering the FCC baseline parameters [123,124] as a starting point at \( t = 0 \), and a beam-beam limit of \( \Delta Q_{\text{max}} = 0.03 \), we find \( t_1 = 1.2 \text{ h} \), and \( \Sigma_1 = 0.4 \text{ fb}^{-1} \).

From time \( t_1 \) onward the collisions continue with constant tune shift.

\[
\Delta Q_{\text{max}} = n_{\text{IP}} \frac{r_p N_b(t)}{4\pi\gamma\epsilon_0} \quad \text{for } t > t_1.
\]

This can be achieved by controlling the emittance \( \epsilon(t) \) through transverse noise excitation.

From (4) the change in intensity is

\[
\frac{dN_b}{dt} = -\sigma_{\text{tot}} n_{\text{IP}} L n_b = -\sigma_{\text{tot}} f_{\text{rev}} \gamma \Delta Q_{\text{max}} \frac{\beta^* r_p}{B} N_b = -X N_b,
\]

where we have introduced

\[
X = \frac{\sigma_{\text{tot}} f_{\text{rev}} \gamma \Delta Q_{\text{max}}}{\beta^* r_p}.
\]

The solution is an exponential decay for both the bunch intensity and the emittance with time constant

\[
\tau_2 = \frac{1}{X} = \frac{\beta^* r_p}{\sigma_{\text{tot}} f_{\text{rev}} \gamma \Delta Q_{\text{max}}},
\]

that is

\[
\epsilon(t) = \epsilon_0 e^{t_1/\tau} e^{-(t-t_1)/\tau_2}
\]

and

\[
N(t) = \frac{1}{K\tau_0 (e^{t_1/\tau} - 1) + \frac{1}{N_0}} e^{-(t-t_1)/\tau_2}.
\]

The time dependent luminosity then becomes

\[
L(t) = \frac{f_{\text{rev}}\gamma n_b \Delta Q_{\text{max}}}{\beta^* r_p n_{\text{IP}}} N(t)
\]
\[ L(t) = \frac{f_{rev}n_h\Delta Q_{\text{max}}}{\beta^* r_p n_{\text{ip}}} \frac{1}{\epsilon_0} \left( e^{t_1/\tau_2} - 1 + \frac{1}{n_0} e^{-x(t_1)/\tau_2} \right). \]  

The integrated luminosity at time \( t \) \((t > t_1)\) is

\[ \Sigma(t) = \Sigma_1 + \Sigma_2(t) = \Sigma_1 + \int_{t_1}^{t} L(t)dt \]

\[ = n_0 n_{h2} \Delta Q_{\text{max}} \left[ \frac{f_{rev} \tau_2}{\beta^* r_p n_{\text{ip}}} \left( 1 - \frac{\Delta Q_0}{\Delta Q_{\text{max}}} \right) \right] \]

\[ + f_{rev} \tau_2 \frac{\Delta Q_0}{\epsilon_0} \left( e^{t_1/\tau_2} - 1 + \frac{1}{n_0} e^{-x(t_1)/\tau_2} \right). \]  

Finally, as before in the case of constant emittance, we want to maximize the average luminosity. Introducing again the “turnaround time” \( t_{ta} \), i.e., the average time between the end of one physics fill at time \( t_1 \) and the start of the luminosity production in the following fill, at (on average) the time \( t_1 + t_{ta} \), the average luminosity depends on the run time \( t_r \) as

\[ L_{\text{ave}}(t_r) = \frac{\Sigma_1 + \Sigma_2(t_r)}{t_r + t_a}. \]  

The optimum run time \( t_{r,\text{opt}} \) is found by maximizing \( L_{\text{ave}} \). We define the auxiliary parameters

\[ C \equiv f_{rev} \tau_2 \frac{1 - \frac{\Delta Q_0}{\Delta Q_{\text{max}}}}{1 - \frac{f_{rev} \tau_2}{\beta^* r_p n_{\text{ip}}} \Delta Q_0} \]  

\[ D \equiv -f_{rev} \tau_2 \frac{\epsilon_0}{\epsilon_0} \left( e^{t_1/\tau_2} - 1 + \frac{1}{n_0} e^{-x(t_1)/\tau_2} \right). \]  

From

\[ \frac{d}{dt_r} \left( \frac{C + De^{t_1/\tau_2}}{t_r + t_a} \right)_{t_r = t_{r,\text{opt}}} = -\left( \frac{C + De^{t_1/\tau_2}}{(t_{r,\text{opt}} + t_a)^2} \right) \frac{D e^{-t_{r,\text{opt}}/\tau_2}}{\tau_2(t_{r,\text{opt}} + t_a)} = 0 \]

the optimum run time follows as

\[ -\left( \frac{t_{r,\text{opt}} + t_a}{\tau_2} + 1 \right) e^{-t_{r,\text{opt}}/\tau_1} - t_{r,\text{opt}} - 1 = C D e^{-t_{r,\text{opt}}/\tau_2} \]  

or, solving for \( t_{r,\text{opt}} \), as

\[ t_{r,\text{opt}} = -\tau_2 - t_a - \tau_2 W_1 \left[ \frac{C e^{-t_a/\tau_2}}{D} \right] \]

where \( W_1(x) \) denotes the more negative branch of the multivalued Lambert \( W \) function (defined over the interval \(-1/e < x < 0\)). [The Lambert \( W \) function \([125]\) is the inverse of the function \( f(z) = z e^z \), or \( z = f^{-1}(z e^z) \equiv W(z e^z) \), where \( z \) is any complex number.]

C. Example: FCC-hh scenarios at the beam-beam limit

We now illustrate our results with some example calculations for proton-proton physics at the Future Circular Collider (FCC).

The key physics goals of the FCC are the complete exploration of the Higgs boson and a significant extension, via direct and indirect probes, of the search for physics phenomena beyond the standard model \([126]\). The baseline FCC-hh integrated-luminosity goal of \( 3 \) ab\(^{-1}\) translates into a discovery reach of about 32 TeV for standard model-like couplings. Raising the luminosity by a factor of 10 increases the discovery reach only by about 20% in energy. The higher luminosity leads to much increased event rates, and better statistics, at low masses, and would, for example, allow measuring the Higgs self-coupling to better than 5%. Synthesizing the discussions from several theory workshops, an ultimate integrated luminosity goal of 10–20 ab\(^{-1}\) for the FCC-hh seems well justified \([126]\).

The FCC-hh luminosity can be increased in a number of ways. First, the IP beta function may be reduced. An advanced interaction-region (IR) optics is already being developed, which can reach \( \beta^*_y = 30 \) cm \([127]\], yielding almost a factor 4 gain in peak luminosity, compared with the baseline (and “phase 1”) value of \( \beta^*_y = 1.1 \) m. Second, the beam-beam limit of \( \Delta Q_{\text{tot}} = 0.01 \) assumed in the baseline, appears conservative as the LHC and the Tevatron have routinely been running with two times larger values, and as more than three times higher tune shifts have been obtained in LHC beam experiments without any noticeable impact on beam lifetime or emittance growth \([84]\). Much stronger radiation damping at the FCC-hh (transverse emittance damping time of 1 h) might further boost the achievable beam-beam tune shift if the effect of the radiation damping is similar to the one found at lepton colliders \([128]\). In addition, head-on beam-beam compensation by electron lenses, recently demonstrated at RHIC \([129]\), is likely to support even higher tune shifts. For all the above reasons we consider the possibility of a total beam-beam tune shift as high as \( \Delta Q_{\text{tot}} = 0.03 \) (sum of two IPs). Third, we assume that the initial turnaround time \( t_a \) (the period from the end of one physics fill to the start of the next physics collisions) can be reduced from 5 h in the baseline to 4 h, after a couple of years of beam operation.

Based on the above considerations we envisage two operational phases of the FCC-hh \([130]\). “Phase 1” corresponds to the baseline with a peak luminosity of \( 5 \times 10^{34} \) cm\(^{-2}\) s\(^{-1}\) and an average luminosity production of 250 fb\(^{-1}\) per year. “Phase 2” achieves about a factor 6 higher peak luminosity of \( 3 \times 10^{35} \) cm\(^{-2}\) s\(^{-1}\).
and produces more than 1000 fb\(^{-1}\) per year. Depending on the final operation schedule, an overall integrated luminosity of a few tens of ab\(^{-1}\) can be expected over a period of 20–30 years.

Figure 4 presents the evolution of instantaneous luminosity, integrated luminosity, bunch intensity, emittance, total beam-beam tune shift, and event pileup as a function of time during 24 h, for FCC-hh phases 1 and 2 from Table II.

![Fig. 4. Instantaneous luminosity, integrated luminosity, bunch intensity, emittance, total beam-beam tune shift, and event pileup as a function of time during 24 h, for FCC-hh phases 1 and 2 from Table II.](image)

and are taken into account. Other additional phenomena like gas scattering, Touschek effect, intrabeam scattering, and space charge are insignificant for the 50-TeV beams, in the scenarios considered.

A few key parameters of FCC-hh phases 1 and 2 are shown in Table II. The integrated luminosity values are obtained assuming that on average 180 days per year are scheduled for physics operation (after accounting for shut-downs, maintenance, machine developments, etc.). In the physics period the availability is taken to be 70%.

Figure 5 (left picture) illustrates the effect of radiation damping for phase 2 operation. The nominal emittance damping time is 1 h at 50 TeV. For a shorter damping time the beam-beam limit would be reached earlier at a higher peak luminosity (with more unspent protons), while for a longer damping time of 1.5 h (e.g., corresponding to a lower beam energy of \(\sim 44\) TeV) the luminosity would be
almost constant during the physics fill, as a result of “natural leveling” thanks to the synchrotron radiation.

The right picture of Fig. 5 illustrates the importance of the turnaround time for the integrated luminosity. Especially for phase 2, where protons are burnt quickly and runs are short, the fastest possible turnaround time would be desired. The total turnaround time comprises the ramp-down/ramp-up times of the FCC-hh main ring, as well as the filling time at injection, which depends on the injector complex.

D. Running at the pileup limit

We now consider the possibility that, as the emittances are shrinking due to synchrotron radiation, the first limit encountered is set by the maximum acceptable pileup rather than by the beam-beam tune shift.

This actually is the baseline situation for the HL-LHC, discussed earlier. Already in regular LHC operation with 50 ns bunch spacing and proton bunch intensities of $\geq 1.5 \times 10^{11}$ the maximum pileup acceptable for the ATLAS and CMS detectors would be exceeded at 13 TeV beam energy, for which reason the nominal bunch spacing of 25 ns has been chosen for LHC Run No. 2 (2015–2018). Whether this will also be the case for the FCC-hh depends on its achievable peak luminosity, as well as on the bunch spacing (25 ns or 5 ns), the detector technology then available, and, not the least, the nature of the relevant particle-physics studies.

Denoting the maximum pileup by $\mu_{\text{max}}$, the corresponding maximum luminosity depends on the number of bunches per beam, $n_b$, and on the revolution frequency, $f_{\text{rev}}$, as

$$L_{\text{max}} = f_{\text{rev}} n_b \frac{\mu_{\text{max}}}{\sigma_{\text{inel}}}.$$  \hspace{1cm} (55)

The maximum pileup $\mu_{\text{max}}$ is reached at time $t_2$ defined by

$$\mu_{\text{max}} = \sigma_{\text{inel}} \frac{N^2_h(t_2)}{4\pi\beta^2 e(t_2)} = \frac{\sigma_{\text{inel}} N_0}{4\pi\beta^2 e_0} \left[ \frac{N_0}{B(e^{t_2/\tau} - 1) + 1} \right]^2 e^{t_2/\tau}.$$  \hspace{1cm} (56)

Defining

$$Z = \frac{\sigma_{\text{inel}}}{\mu_{\text{max}} 4\pi\beta^2 e_0}$$  \hspace{1cm} (57)

and using $B$ from (35), we can solve (56) for $t_2$:

$$t_2 = \tau \ln \left[ 1 - \frac{1}{B} + \frac{ZN^2_0}{2B^2} - \frac{N_0}{B} \sqrt{Z(1 - \frac{1}{B} + \frac{ZN^2_0}{4B^2})} \right],$$  \hspace{1cm} (58)

with the implicit assumption that the pileup limit is reached prior to the beam-beam limit.

The integrated luminosity from $t = 0$ to $t = t_2$ is given by $\Sigma_1$ of Eq. (36) with $t_1$ replaced by $t_2$:

$$\Sigma_{1,h} = \int_0^{t_2} L(T)dt = \frac{f_{\text{rev}} n_b N^2_0}{4\pi\beta^2 e_0} \frac{\tau}{B} \left( 1 - \frac{1}{1 - B e^{t_2/\tau}} \right)$$

$$= \frac{f_{\text{rev}} n_b N^2_0}{4\pi\beta^2 e_0} \frac{\tau}{B} \left[ 1 - \frac{1}{\frac{ZN^2_0}{2B} - N_0 \sqrt{Z(1 - \frac{1}{B} + \frac{ZN^2_0}{4B^2})}} \right].$$  \hspace{1cm} (59)

At time $t = t_2$ the bunch intensity, emittance and beam-beam tune shift are

$$N_h(t_2) = N_0 \frac{1}{1 + B(e^{t_2/\tau} - 1)}$$  \hspace{1cm} (60)

$$e(t_2) = e_0 \exp \left( -\frac{t_2}{\tau} \right).$$  \hspace{1cm} (61)
\[ \Delta Q(t_2) = \Delta Q_0 \frac{e^{\beta_3/\tau}}{1 + B(e^{\beta_3/\tau} - 1)}. \] (62)

For times \( t > t_2 \) the emittances need to be controlled so as to keep the luminosity constant (it needs to be “leveled” as for the HL-LHC, but in this case through a control of the naturally shrinking emittance). According to (4) and (3) the further rate of intensity decrease is described by

\[ \frac{dN_b}{dt} = -\frac{\sigma_{\text{tot}} f_{\text{rev}} n_{W}}{\sigma_{\text{inel}}} \mu_{\text{max}} = -K \frac{4\pi \beta^3 \mu_{\text{max}}}{\sigma_{\text{inel}}} = -\frac{K}{Z\varepsilon_0}. \] (63)

Introducing the parameter

\[ W \equiv \frac{K}{Z\varepsilon_0} = \frac{\sigma_{\text{tot}} f_{\text{rev}} n_{W}}{Z\varepsilon_0} \frac{1}{\sigma_{\text{inel}}} \mu_{\text{max}} \] (64)

solving (63) the intensity evolution from time \( t_2 \) onward becomes

\[ N_b(t) = N_b(t_2) - (t - t_2)W. \] (65)

According to (56) during this period the emittance shrinks quadratically with time, namely

\[ e(t) = e(t_2) \left[ 1 - W \frac{(t - t_2)}{N_b(t_2)} \right]^2, \] (66)

and the beam-beam tune shift increases as

\[ \Delta Q_{\text{tot}}(t) = n_{W} \frac{r_p}{4\pi\gamma} \frac{N_b(t_2)}{e(t_2)} \frac{1}{1 - W \frac{t-t_2}{N_b(t_2)}}. \] (67)

The time interval of constant luminosity ends if one reaches either a limit of the leveling mechanism, such as the minimum value of \( \beta^* \) or zero effective crossing angle, or the maximum acceptable beam-beam tune shift.

Consider the latter case. That is the luminosity after an initial increase, due to radiation damping, is first restricted by the pileup limit from the particle-physics detector, and later in a third phase, starting at time \( t_3 \) as the beam current, emittance and luminosity decreases, confined by the beam-beam limit.

The time \( t_3 \) follows from the equation

\[ \Delta Q_{\text{tot}}(t_3) \equiv \Delta Q_{\text{max}}, \] (68)

or, using (67),

\[ t_3 = t_2 + \frac{N_b(t_2)}{W} \left[ 1 - \frac{n_{W} r_p N_b(t_2)}{4\pi\gamma e(t_2) \Delta Q_{\text{max}}} \right]. \] (69)

Since the luminosity over this time interval is constant, the integrated luminosity from \( t = t_2 \) to \( t = t_3 \) is simply given by

\[ \Sigma_{2,b} = \int_{t_2}^{t_3} L(t)dt = (t_3 - t_2)L_{\text{max}}. \] (70)

At time \( t = t_3 \) the bunch intensity, emittance, and beam-beam tune shift are

\[ N_b(t_3) = N_0 \frac{1}{1 + B(e^{\beta_3/\tau} - 1)} - (t_3 - t_2)W, \] (71)

\[ e(t_3) = e_0 \exp \left[ -\frac{t_2}{\tau} \right] \left[ 1 - W \frac{(t_3 - t_2)}{N_b(t_2)} \right]^2, \] (72)

\[ \Delta Q(t_3) = \Delta Q_{\text{max}}. \] (73)

From \( t_3 \) onward, operating at the beam-beam limit, we can apply the considerations and results from Eqs. (40)–(48). The intensity and emittance decrease exponentially, with time constant \( \tau_2 = 1/X \) of Eq. (43).

The instantaneous luminosity for \( t > t_3 \) is given by an equation of the form (47):

\[ L(t) = \frac{f_{\text{rev}} n_b N_b(t_3)^2}{4\pi\gamma e(t_3)} e^{-\frac{t-t_3}{\tau_2}}. \] (74)

The integrated luminosity at time \( t (t > t_3) \) is

\[ \Sigma_b(t) \equiv \Sigma_{1,b} + \Sigma_{2,b} + \Sigma_{3,b} = \Sigma_{1,b} + \Sigma_{2,b} + \int_{t_3}^{t} L(t)dt. \] (75)

Similar to the discussion following Eq. (49) the average luminosity depends on the run time \( t_r \) as

\[ L_{\text{ave}}(t_r) = \Sigma_{1,b} + \Sigma_{2,b} + \Sigma_{3,b}(t_r), \] (76)

and the optimum run time \( t_{r,\text{opt},b} \) is again found by maximizing \( L_{\text{ave}} \).

The solution can again be expressed through the Lambert function \( W_1 \) as in Eq. (54), namely,

\[ t_{r,\text{opt},b} = -\tau_2 - t_a - \tau_2 W_1 \left[ \frac{C_b e^{-\tau_2/\tau_2} t_{\tau}}{D_b} \right], \] (77)

where, this time, the auxiliary parameters are given by...
Once more we consider the FCC-hh phase 2 from Table II, but we now assume that the maximum pileup limit is 700. In this case, while the emittances are first naturally shrinking the pileup limit of $\mu_{\text{max}} = 700$ is reached already after $t_2 = 12$ minutes and then the beam-beam limit of $\Delta Q_{\text{max}} = 0.03$ at time $t_3 = 2.7$ h (for comparison without the pileup limit the beam-beam limit of 0.03 is reached at time $t_2 = 1.7$ h). The optimum run times for the two cases are $t_{r,\text{opt}} = 3.5$ h without pileup limit, and $t_{r,\text{opt}} = 4.1$ h with the pileup limit of 700, respectively.

Figure 6 shows the luminosity, integrated luminosity, bunch intensity, emittance, total beam-beam tune shift, and event pileup as a function of time, comparing FCC-hh runs with and without a limit $\mu_{\text{max}} = 700$.

E. Example: FCC-hh scenario with pileup and beam-beam limit

Once more we consider the FCC-hh phase 2 from Table II, but we now assume that the maximum pileup limit is 700. In this case, while the emittances are first naturally shrinking the pileup limit of $\mu_{\text{max}} = 700$ is reached already after $t_2 = 12$ minutes and then the beam-beam limit of $\Delta Q_{\text{max}} = 0.03$ at time $t_3 = 2.7$ h (for comparison without the pileup limit the beam-beam limit of 0.03 is reached at time $t_2 = 1.7$ h). The optimum run times for the two cases are $t_{r,\text{opt}} = 3.5$ h without pileup limit, and $t_{r,\text{opt}} = 4.1$ h with the pileup limit of 700, respectively.

Figure 6 shows the luminosity, integrated luminosity, bunch intensity, emittance, total beam-beam tune shift, and event pileup as a function of time, comparing FCC-hh runs with and without pileup limit.

VI. CONCLUSIONS

We have derived analytical expressions for the optimized integrated luminosities and for the optimum store lengths at high-energy hadron colliders limited either by event pileup or by the beam-beam tune shift, in the presence or absence of strong radiation damping.

The analytical solutions were illustrated with examples from the LHC design, for the planned HL-LHC and for the
proposed FCC-hh. Our results indicate that, for all three colliders, the target values of their integrated luminosity are well within reach.

**ACKNOWLEDGMENTS**

Informative discussions with M. Hostettler and G. Papotti on the emittance evolution during LHC physics stores have been much appreciated. We also thank G. Arduini and M. Giovannozzi for a careful reading of the manuscript. This work was supported in parts by the European Commission under the Capacities 7th Framework Programme project EuCARD-2, Grant Agreement No. 312453, and the HORIZON 2020 project EuroCirCol, Grant Agreement No. 654305.


[87] K. Ohmi and F. Zimmermann, Beam-beam limit in hadron colliders (to be published).


