SLAC-R-1035

Radiative Bottomonium Spectroscopy at the Υ (2, 3S) Resonances at BaBar

By: Peter M. Lewis

SLAC National Accelerator Laboratory, Stanford University, Stanford CA 94309

Work funded in part by DOE Contract DE-AC02-76SF00515.

SLAC-R-1035

RADIATIVE BOTTOMONIUM SPECTROSCOPY AT THE $\Upsilon(2,3S)$ RESONANCES AT *BABAR*

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Peter M. Lewis
August 2013

SLAC-Report-1035
Work supported by Department of Energy contract DE-AC02-76SF00515

SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Rafe Schindler) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Patricia Burchat)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(David L. Burke)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Stanley J. Brodsky)

Approved for the University Committee on Graduate Studies

Abstract

The compact bound state consisting of a bottom and anti-bottom quark pair interacting via the strong nuclear force is called "bottomonium." A wealth of long-lived bottomonium states can be both experimentally produced and theoretically described, providing a unique tool to probe calculation techniques with experiment. Bottomonia with total angular momentum $J = 1$ and orbital angular momentum $L = 0$ at a variety of radial excitations n – called $\Upsilon(nS)$ – can be produced at electron-positron colliders. The *BABAR* experiment, located at the interaction point of such a collider (the PEP-II storage ring), has observed 122 million $\Upsilon(3S)$ and 100 million $\Upsilon(2S)$ decays. Some of these involve a transition to the bottomonium state $\chi_{bJ}(nP)$ ($L = 1$ and $J = (0, 1, 2)$), emitting a photon, with subsequent transition to a lower $\Upsilon(nS)$, also emitting a photon. The final $\Upsilon(nS)$ can be identified through a decay to two muons. The dependence of the branching fractions and photon energies in this process on the spin state of the intermediate $\chi_{bJ}(nP)$ is a key test of phenomenological models. To this end, this dissertation contains a nearly comprehensive study of these transitions with an emphasis on experimentally optimal discrimination between various models. This focus spurs innovative techniques that complement a large array of physics results, both presented in detail herein.

Acknowledgments

As is traditional, I will include an acknowledgment of those without whom this thesis would not exist; or, conversely, without whom this thesis might have been completed much sooner but who have made the integrated misery significantly less. Somewhat less traditionally, I feel inclined to also thank people, creatures and inanimate objects that in some hard-to-specify way helped me through these seven very long and wonderful years.

First, to my wife Amanda, who has been deeply invested in my physics education from nearly the beginning. She's a lovely and strong woman and the most perfect mate I could ask for. Looking forward to a long life with Manda evokes optimism and hope, two emotions in desperate short supply during a PhD candidacy.

To my daughter Eleanor, who was a little spheroid squirt when we moved here. I don't know anybody who extracts more from each second of living than my little EllaBellaSmellaFella. From scaling a Nicaraguan volcano in brutal mid-day heat to mastering Irish Wedding Dance on the piano to her fierce love of her friends she brings a startling intensity to life.

To my daughter Alexandra, who has only ever known life as the child of a grad student. She is incomparably thoughtful, perpetually-disheveled, cheerful, sweet, brutally-destructive and loving. Some of my favorite moments in the last seven years have involved Ballyhoo's unrestrained love and care and helpfulness.

To my little boy Miles, a little-old-man with big dark eyes and a massive sap streak. In his mind he is a ninja with fancy moves who is a snowy owl riding orcas all day long. Budsto hugs me whenever I want hugs and beats me up when I want to beat up, which is just about perfect.

To my parents, John and Peg, who provided the genes and environment that

made this moment almost inevitable. Without my mother's restless ambition and my father's relentless scientific mind this document wouldn't exist.

To my extra-large extended family: 46 nieces and nephews [and counting] and 28 siblings/siblings-in-law. There are so many wonderful characters, many of whom frequently influence and enliven me. Van, Megret, Chris, Katie and Bonnie – my siblings, unique and wonderful all – provide all the pieces to the template I model my life after. Nick – my "nephew" – is basically me yet entirely the opposite and is one of my favorite people ever.

To Escondido Village, our little utopia on campus. The rich community, wise design and cool grass threatened to entice me to stay forever. I don't have the space to begin to describe what my EV friends have meant to me and my family. Carrie and Matt, Chris and Brittney, Alicia and Nathan, Joni and Aaron, Justin and Amy, Kelsey and Bill; Matt and Silvana; so many more. Without exception, these are the best kind of people, a phenomenal group of life-changing personalities that I love dearly.

To Jared and Kim for the hospitality over the last year. A doctoral candidate makes the worst kind of house guest but you both were perfect hosts and wonderful friends. To Ant Chicki, without whom our stay in Bellingham over the last year would have been flat-out dreadful.

To RASICAM, for taking a 3-year bite out of my thesis, teaching me irrelevant skills and keeping me up late on weekends crimping lugs and popping rivets. Nothing I've done in my career has so well prepared me for experimental science and I derived great satisfaction from the process and especially from the results. I feel extraordinarily lucky to have had the rare experience in graduate school of conceiving, designing, building, coding, fixing, debugging, deploying and operating a self-contained scientific device. I owe a special appreciation to Kevin Reil for adopting RASICAM when I needed to move on and doing it well enough that I rarely wake up in a cold sweat at 2am worried about my robotic infrared baby any more.

To California tiger salamanders [and, by extension, Dr. Alan Launer] for giving me a somewhat random sense of adventure and purpose, at least on cold, rainy winter nights.

To Montebello Open Space Reserve, where I inevitably found myself when I needed

it most.

To Costa Rica, for beckoning me relentlessly with unlimited birds, frogs, snakes and life. Similarly, to Papua New Guinea, whose flag on my desktop keeps me dreaming daily.

A special kind of acknowledgment is due to the various avocations that derailed my focus, sapped my energy, occupied my time and nourished my sanity: birding, field herping, photography, basketball, cycling, backpacking, foreign travel, slacklining, exploring wilderness, more basketball, working out, cooking, really pretty much anything. I owe an especially deep-felt and unironic gratitude for the fields of biogeography and evolutionary ecology for causing a crisis of career well past the point-of-no-return, and to Dr. Bob Drewes who wouldn't let the matter rest. To the 410+ life birds I've seen since moving to Stanford – one for each of my PhD-related troubles, and some just as spectacular.

The authors that kept science alive for me – David Quammen, Stephen Pinker, Scott Weidensaul, Jared Diamond, Richard Dawkins, E. O. Wilson, Adrian Forsyth – and the ones that entertained me and added meaning to life – David Foster Wallace, Edward Abbey, Jonathan Franzen, Kurt Vonnegut – deserve a great deal of gratitude. I was invigorated and inspired by the imposing creative genius of John Lennon, Thom Yorke, Steve Reich, Vijay Iyer, Colin Meloy and the Emerson String Quartet and kept sane by Randall Munroe. But to Carsten Peter and Tim Laman I direct a torrent of resentment for essentially destroying a lifelong dream by being flat out astounding at what they do.

At SLAC are many colleagues and friends who have enriched my life in one way or another: Chris West, who taught me everything I know about particle physics [such as it is] and showed me the best cycling routes in the area; Bryan Fulsom, who provided precious dialog about bottomonium physics and showed me that physics and children can coexist; Peter Kim, an excellent physicist and gracious sounding-board; Howard Rogers, a great friend wielding unmatched experience and practical knowledge; Dave Burke, who actively supported my career; Judy Meo, who was always just a building away when I absolutely needed to get away from my cubicle, and Stephen Sun, whose excellent eclectic conversation ensured that neither of us got a single thing done my first year here. My summer students Brett Sandler, Nick Bern, Trent Benedick and

Chris Sund were rewarding to work with and taught me a great deal.

Bill Dunwoodie, as my de facto thesis adviser over the last year, has been unlimitedly gracious with his time and expertise, and seems to take unnatural pleasure out of doing partial wave expansions, which *somebody* has to do.

Advisers always earn a very unique place in the thesis thankalog; it is rare that anybody so quickly becomes the nexus for all your future success and/or failure. I quickly identified Rafe as the adviser that would allow me to blossom maximally – I felt that the adviser was more important than the experiment and I believe I was right. I'm extremely grateful for his minimalist but sure guiding hand throughout the last seven years.

And last and literally least, to the $\chi_{b0}(1P)$ for being a tiny bump that occupied my life and sometimes sleep off-and-on for a half-decade.

Contents

List of Tables

List of Figures

Contents

Chapter 1

Introduction

The strongly bound $b\bar{b}$ meson – bottomonium – exhibits a rich positronium-like structure and is an ideal laboratory for verifying quantum chromodynamics calculation techniques and fundamental properties [1]. The bottomonium system is largely non-relativistic, allowing relatively simple phenomenological treatments to predict properties of the system with unique success. Precision measurements in the bottomonium system can then be used to inform calculations of QCD in a multitude of contexts.

This dissertation presents a study of a large number of radiative transitions between various states of the bottomonium system using *BABAR* data. The focus is twofold: (1) to provide precision determinations of all radiative branching fractions involved, particularly for the poorly known states and (2) to present experimentally optimal spectroscopic results in the language of the relevant effective theories to avoid the unnecessary loss of precision that occurs commonly during translation from experiment to theory.

A discussion of the relevant physics is presented in this chapter.

1.1 The physics context

To set the proper context for a discussion of the bottomonium system, an overview of the relevant theoretical and experimental physics is provided here, followed by a more detailed description of the physics of the bottomonium system itself in the next section.

1.1.1 The Standard Model

The Standard Model (SM) provides a unified framework for our current understanding of fundamental forces and particles. The fundamental particles include leptons, quarks and gauge bosons. Leptons include electrons and their heavier cousins the muons and taus as well as associated neutrinos. Quarks also come in three generations, with six "flavors": up, down, charm, strange, top and bottom. Interactions between the fundamental particles are seen as arising from the exchange of gauge bosons: photons for electromagnetic, W and Z bosons for weak and gluons for strong interactions. The SM also predicts the existence of the Higgs boson, which couples with many of the other fundamental particles to give them mass.

This dissertation involves bottom quarks interacting with the strong force; the relevant SM theory is called quantum chromodynamics (QCD), described in more detail below with an emphasis on understanding the bottomonium system.

1.1.2 Quantum Chromodynamics

All hadrons (such as nucleons and pions) are composed of quarks and gluons, with gluon exchange mediating the strong interactions. Quarks are massive spin-1/2 fermions that carry one of three "color" charges (called red, blue and green), while antiquarks carry one of the three anticolors (antired, antiblue and antigreen). Gluons are massless spin-1 particles that carry a mixture of one color and one anticolor; of the nine combinatoric color possibilities one is a color singlet (colorless), yielding eight unique gluons. Hadrons in nature are colorless, which shows the utility of the color analogy for strong charge: a combination of three different colors is colorless, as is a color-anticolor pair. Strong interactions conserve color, so an exchange of a gluon between two quarks requires the gluon to carry the color of the initial quark and the anticolor of the final quark, swapping the colors of the two quarks.

This qualitative description of quark-gluon interaction is encapsulated formally in the special unitary group $SU(3)$. Quark color states are represented as vectors in a 3-dimensional complex color space. The eight gluons correspond to 3×3 matrices that are the fundamental representation of $SU(3)$; these can be conceptualized as rotation matrices in color space. These transformations have unit determinant and

are therefore unitary and form the Lie group $SU(3)$. The rules for adding colors fall out of matrix operations within this framework; hadrons arise out of the structure of the symmetry group. The QCD Lagrangian is given by:

$$\mathcal{L} = \sum_q \overline{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab})\psi_{q,b} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu}, \tag{1.1}$$

where repeated indices are summed over, g_s is the QCD coupling constant, $\psi_{q,a}$ are quark-field spinors for a quark with flavor q and mass m_q with a color index $a = 1, 2, 3$ and γ^μ are the Dirac γ-matrices. The gluon field is represented by $\mathcal{A}^C_\mu$ with $C = 1, 2, ...8$ representing the 8 possible color combinations of gluons. The eight 3×3 matrices that are the generators of the $SU(3)$ group, as described above, are denoted t^C_{ab}. $F^A_{\mu\nu}$ is the gluon field strength tensor that is determined by the gluon field $\mathcal{A}^C_\mu$ and the structure constants of the $SU(3)$ group.

Calculation in the bottomonium system from first principles in QCD is extremely complex; the discussion herein focuses on phenomenological and qualitative aspects of QCD that are directly relevant to the physics topic of this dissertation.

Confinement and asymptotic freedom

A distinct feature of QCD is the varying behavior of quark interactions at different interquark scales r. At small r (and at high energies) the strong coupling decreases, a behavior termed asymptotic freedom. As r increases, the energy of the interquark interaction increases, qualitatively understandable as the system resisting isolation of the constituent quarks, which is forbidden. This effect is called confinement. The following cartoon conceptualization describes these two behaviors: two quarks at close distance are interacting weakly, exchanging single gluons. Energy is added to the system, separating the quarks; the interaction becomes stronger, with single-gluon exchange getting replaced with a string of gluon energy density called a color flux tube, analogous to a rubber band. This picture is starkly different from the familiar electromagnetic case in which the electric field energy density diffuses isotropically in space. The flux tube has a roughly constant cross-section and energy density, meaning that the energy stored in it is linear with r. Eventually enough energy is provided to the system that the tube snaps; confinement requires both original quarks to form

new mesons with two new quarks.

This cartoon qualitative description motivates a quantitative phenomenological description of quarkonium ($q\overline{q}$) that tracks the historical development of effective theories used in quarkonium calculations. This discussion not only illustrates the principles of confinement and asymptotic freedom but is directly applicable to bottomonium spectroscopy.

In the nonrelativistic limit and for small r single-gluon exchange dominates and is reminiscent of the Coulomb interaction in electromagetism. The interquark potential derived from this argument from asymptotic freedom can be written analogously:

$$V(r) = -\frac{4}{3}\frac{\alpha(r)}{r}, \tag{1.2}$$

where $\alpha(r)$ is the strong coupling constant. Considering now larger r, confinement introduces a linear term to create what has been called the Cornell potential [7]:

$$V_{Cornell}(r) = -\frac{4}{3}\frac{\alpha(r)}{r} + \sigma r \tag{1.3}$$

where σ is again tuned to experimental data. The Schrödinger equation can be solved with this potential to derive the Υ spectrum well; further potential terms to accommodate spin-dependent interactions are added in Section **??**. Other schemes use different constructions of the interquark potential with usually subtle differences in final form.

Within the Coulomb region of the potential, perturbative calculations are valid; nonperturbative techniques are required in the linear region. The confinement scale Λ_{QCD} delineates these regions. Fig. **??** shows that the bottomonium system spans these two regions in the nonrelativistic limit, allowing a probe of QCD in both regimes. This effective theory formulation and extensions of it are very successful at describing the quarkonia spectra and are radically simpler than first-principle QCD calculations.

1.1.3 Bottomonium in QCD

The large bottom quark mass ($M_{\Upsilon(1S)} \approx 10 M_{proton}$) makes the bottomonium system nearly nonrelativistic with $(v/c)^2 \approx 0.08$. The phenomenological description up to

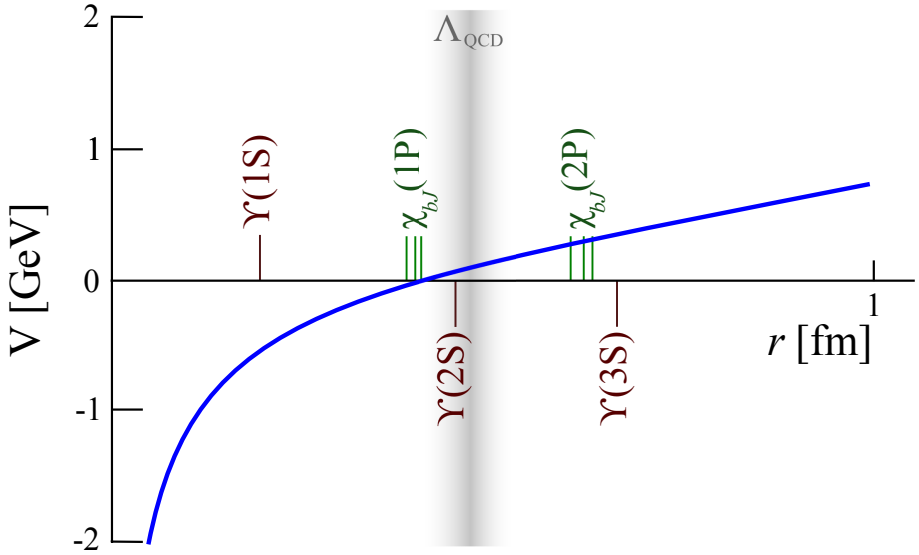


Figure 1.1: A Cornell potential compared to the root-mean-square interquark separation with approximate positions of the relevant bottomonium levels and the confinement scale Λ_{QCD} indicated.

this point has relied on a nonrelatvistic argument; this perspective is extended to include relativistic effects.

The mass spectrum of the bottomonium system consists of many distinct states that correspond to different excitations of radial (n) and rotational (L) modes in $b\bar{b}$ strong-force interactions as well as effects due to spin alignment (Fig. **??**), analogous to positronium structure under EM interactions. We use the standard spectroscopic notation $n^{2S+1}L_J$ where needed, but we adopt other notations to suit the context throughout. Bottomonium decay nominally would be dominated by strong processes. Below the open-flavor threshold ($2M_{B\bar{B}}$) the dominant process $b\bar{b} \rightarrow B\overline{B}$ is kinematically forbidden, opening up a large suite of observable transitions within the bottomonium system. What makes the bottomonium system particularly attractive is that this threshold is above the mass of the $\Upsilon(3S)$ resonance ($3^3\mathrm{0}_1$), allowing precision probes of a large diversity of states in the S-, P- and D-waves. Fig. **??** shows the bottomonium system below the open-flavor threshold, excluding states that are unreachable from the produced $\Upsilon(nS)$.

The bottomonium spectrum is similar to the charmonium spectrum, with analogous rotational and orbital modes and splittings. However, two significant differences exist: the splitting of the $L = 1$ states (χ_{bJ} in the bottomonium case and χ_{cJ} for charmonium) is smaller in the bottomonium system. Additionally, the $2P$ states in the charmonium system are above open-charm threshold and the analogous states in the bottomonium system are above-threshold, allowing bottomonium spectroscopy to probe a much larger array of states. Further orbital states, such as the $1F$, exist but are difficult to reach from the initial Υ states.

This section describes the bottomonium system in more detail. Comprehensive treatments are found at [3], [4], [5] and [6].

Bottomonium properties

Bottomonium with photon quantum numbers $J^{PC} = 1^{--}$ is directly produced via a virtual photon at *BABAR* in the process $e^+e^- \to \gamma^* \to \Upsilon(nS)$. Below the open-flavor threshold the simplest strong decays (to a pair of B mesons) are kinematically forbidden and two-gluon decay is forbidden by C-parity. This leaves three-gluon annihilation as the dominant strong process, but this process is OZI-suppressed leaving a very narrow hadronic width. Consequently, bottomonium states are long-lived and EM processes are competitive.

The bottomonium spectrum exhibits fine splittings due to spin-orbit and tensor couplings [such as between the three $\chi_{bJ}(1P)$ states] and spin-spin hyperfine splittings [such as between the spin-triplet $\Upsilon(1S)$ and singlet $\eta_b(1S)$] that are generally weaker than similar splittings in the charmonium spectrum. The properties of the χ_{bJ} states and associated radiative transitions are of particular interest in this analysis and are thus described in more detail here.

Potential models

The phenomenological strategy of Sec. **??** is extended in this section to include relativistic effects specifically in the bottomonium system. The bottomonium system is the smallest bound state in nature ($\sim 0.2fm$ for the $\Upsilon(1S)$), making the static Cornell potential model conceptually apt. At this level there is no explicit spin dependence

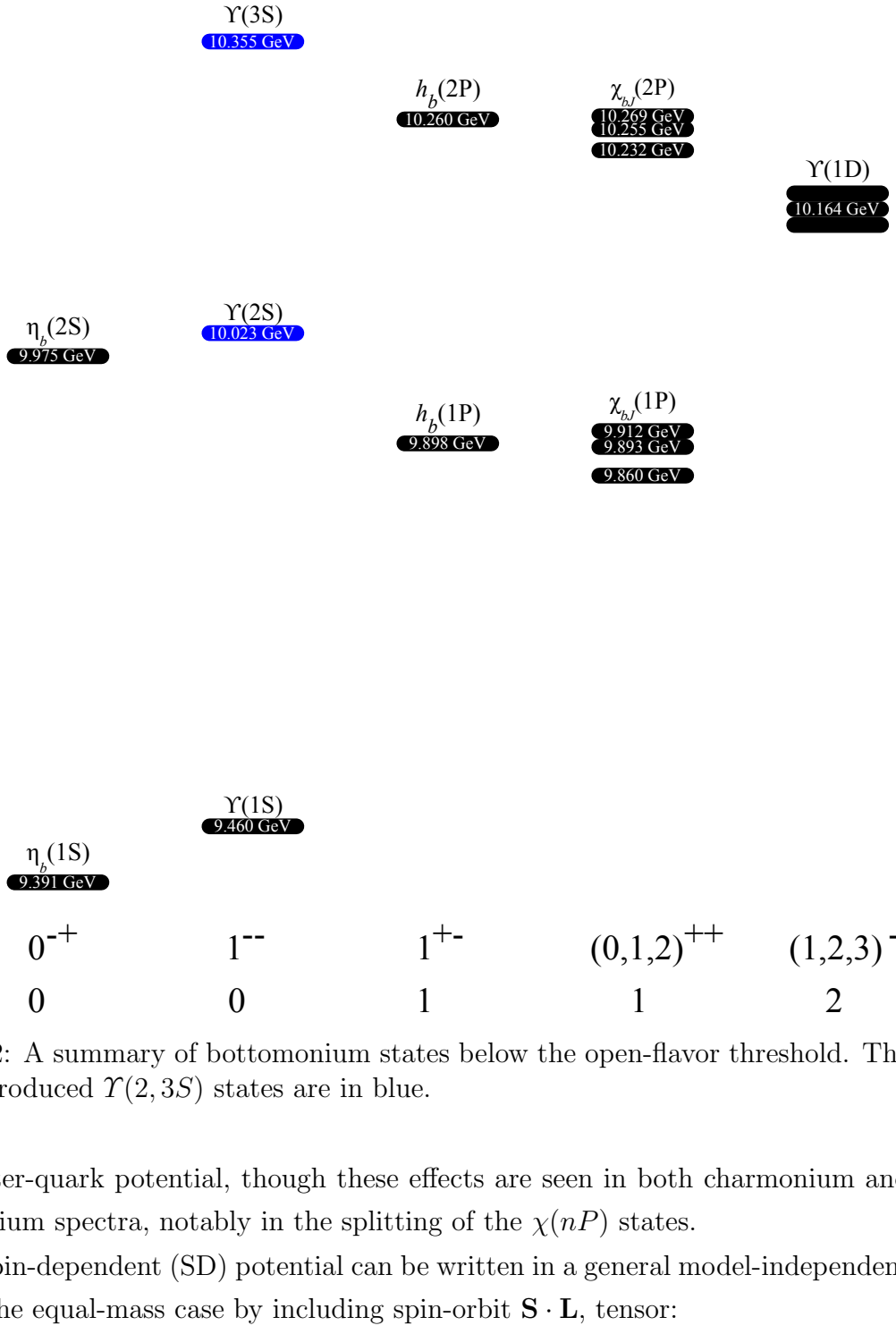


Figure 1.2: A summary of bottomonium states below the open-flavor threshold. The directly produced $\Upsilon(2,3S)$ states are in blue.

on the inter-quark potential, though these effects are seen in both charmonium and bottomonium spectra, notably in the splitting of the $\chi(nP)$ states.

The spin-dependent (SD) potential can be written in a general model-independent sense in the equal-mass case by including spin-orbit $\mathbf{S}\cdot\mathbf{L}$, tensor:

$$S_{12} = 6(\mathbf{S_1}\cdot\hat{\mathbf{r}})(\mathbf{S_2}\cdot\hat{\mathbf{r}}) - 2\mathbf{S_1}\cdot\mathbf{S_2}, \tag{1.4}$$

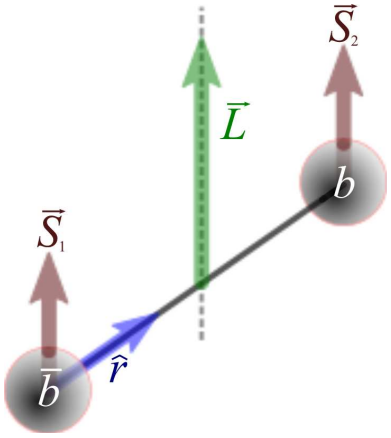


Figure 1.3: A cartoon of the bottomonium system in the χ_{b2} configuration. The black line connecting the b and $\bar{b}$ can be conceptualized as a color flux tube for the higher-L states and as signifying single-gluon exchange for the $L = 0$ states.

and spin-spin $\mathbf{S_1} \cdot \mathbf{S_2}$ terms with associated potentials V_{SO}, V_T and V_{SS} [61]:

$$\begin{aligned} V_{SD}(r) =& \frac{\mathbf{S} \cdot \mathbf{L}}{2m^2} \left[\frac{-dV(r)}{rdr} + 4\frac{dV_{SO}(r)}{rdr} \right] \\ &+ \frac{1}{12m^2} S_{12} V_T(r) \\ &+ \frac{2}{6m^2} \left[2\mathbf{S} \cdot \mathbf{S} - 3 \right] V_{SS}(r). \end{aligned} \tag{1.5}$$

The dependence on the spin-independent potential $V(r)$ is due to Thomas precession, in analogy with the same calculation in electromagnetic interactions. The remaining terms are fully general; the task of understanding and deriving the spin-dependent potentials remains.

An interesting interpretation comes from comparing this formulation to the Breit-Fermi potential that expresses a similar result in the equal-mass case in terms of the

exchange of vector (v) and scalar (s) fields:

$$\begin{aligned} V_{B-F}(r) =& \frac{\mathbf{S}\cdot\mathbf{L}}{2m^2}\left[-\frac{dv(r)+ds(r)}{rdr}+4\frac{dv(r)}{rdr}\right] \\ &+\frac{1}{12m^2}S_{12}\left[\frac{dv(r)}{rdr}-\frac{d^2v(r)}{dr^2}\right] \\ &+\frac{1}{6m^2}\left[2\mathbf{S}\cdot\mathbf{S}-3\right]\nabla^2 v(r). \end{aligned} \tag{1.6}$$

Comparison with Eqn. **??** yields the correspondences:

$$V(r) \longleftrightarrow v(r)+s(r), \tag{1.7}$$

$$V_{SO} \longleftrightarrow v(r), \tag{1.8}$$

$$V_T \longleftrightarrow \frac{dv(r)}{rdr}-\frac{d^2v(r)}{dr^2}, \tag{1.9}$$

$$V_{SS} \longleftrightarrow \nabla^2 v(r). \tag{1.10}$$

The physical picture this evokes is instructive: vector interactions correspond to single-gluon exchange in the Coulomb region, with scalar interactions describing flux tube interactions in the confinement regime. The scalar term only appears associated with the spin independent potential $V(r)$ – the spin-dependent potentials are all functions of vector exchange in the perturbative region. In the scalar region we can picture the flux tube rotating with the quark pair, generating no frame-dependent relativistic effects.

This picture can inform a discussion of the two $\chi_{bJ}(nP)$ states in the bottomonium system; the two systems are identical but for their energy, which pushes the interactions of the $2P$ state higher into the scalar confinement region of the interquark potential. Since the spin-dependent potentials depend only on the vector field term we would expect the mass splittings of the $2P$ levels to be smaller than those of the $1P$ state, and the balance of spin-orbit and tensor corrections will change. Precision spectroscopy of the $\chi_{bJ}(nP)$ states can then probe assumptions about the nature of the potential in this sensitive transition region regardless of whether the scalar and vector exchange interpretation is strictly correct.

The spin-spin term $\nabla^2 v(r)$ will clearly approach zero quickly as the size of the

system expands beyond the Coulomb well. Indeed we will see that only the S-wave bottomonia are small enough for the hyperfine spin-spin splittings to be significant.

The matrix elements of the spin-orbit, tensor and spin-spin terms can be parametrized as:

$$a = \frac{1}{2m^2}\left\langle -\frac{ds}{rdr} + 3\frac{dv}{rdr}\right\rangle, \tag{1.11}$$

$$b = \frac{1}{12m^2}\left\langle \frac{dv}{rdr} - \frac{d^2}{dr^2}\right\rangle, \tag{1.12}$$

$$c = \frac{1}{6m^2}\left\langle \nabla^2 v(r)\right\rangle. \tag{1.13}$$

Using these expectation values we write the masses of the triplet $\chi_{bJ}(nP)$ states in relation to the spin-weighted center of gravity $\overline{M_{nP}} = \frac{1}{9}(M_{nP(0)} + 3M_{nP(1)} + 5M_{nP(2)})$:

$$M_{nP(J)} = \overline{M_{nP}} + a\left\langle \mathbf{L}\cdot\mathbf{S}\right\rangle + b\left\langle S_{12}\right\rangle + c\left\langle \mathbf{S_1}\cdot\mathbf{S_2}\right\rangle \tag{1.14}$$

The matrix elements for the operators are given by [47]:

$$\left\langle \mathbf{L}\cdot\mathbf{S}\right\rangle = \frac{1}{2}\left[J(J+1) - L(L+1) - S(S+1)\right] \tag{1.15}$$

$$\left\langle S_{ij}\right\rangle = -\frac{\left[12\left\langle \mathbf{L}\cdot\mathbf{S}\right\rangle^2 + 6\left\langle \mathbf{L}\cdot\mathbf{S}\right\rangle - 4S(S+1)L(L+1)\right]}{(2L-1)(2L+3)} \tag{1.16}$$

$$\left\langle \mathbf{S_1}\cdot\mathbf{S_2}\right\rangle = \frac{1}{2}\left[S(S+1) - \frac{3}{2}\right] \tag{1.17}$$

For the triplet P-wave states $\{\chi_{b0}, \chi_{b1}, \chi_{b2}\}$ these expectation values become $\left\langle \mathbf{L}\cdot\mathbf{S}\right\rangle = \{-2, -1, 1\}$, $\left\langle S_{ij}\right\rangle = \{-4, 2, -2/5\}$ and $\left\langle \mathbf{S_1}\cdot\mathbf{S_2}\right\rangle = \{1/4, 1/4, 1/4\}$. The spin-spin term does not distinguish between the three spin states; splittings from the center of gravity are independent of c. This term is therefore dropped as we focus attention on the $\chi_{bJ}(nP)$ states.

The masses of the $\chi_{bJ}(nP)$ states become:

$$M_{nP(2)} = \overline{M_{nP}} + a - 2b/5, \tag{1.18}$$

$$M_{nP(1)} = \overline{M_{nP}} - a + 2b, \tag{1.19}$$

$$M_{nP(0)} = \overline{M_{nP}} - 2a - 4b, \tag{1.20}$$

allowing a measurement of the splittings to constrain a and b (see Fig. **??**).

We propose to optimize the determination of a and b experimentally by casting these parameters in terms of line splittings $\delta_0 = M_{nP(1)} - M_{nP(0)}$ and $\delta_2 = M_{nP(1)} - M_{nP(2)}$ thus bypassing any systematic errors associated with absolute mass measurements:

$$a = \frac{1}{6}\left(\delta_0 - \frac{5}{2}\delta_2\right), \tag{1.21}$$

$$b = \frac{5}{72}\left(2\delta_0 + \delta_2\right). \tag{1.22}$$

This formulation is model-independent; direct discrimination between models is possible by comparing predictions for these splitting parameters with experimentally determined values.

Additionally, the ratio R_χ is a convenient parameter to probe theoretical predictions in a way that is sensitive to the underlying models [2] and has been used to provide a cancellation of experimental mass determination systematic uncertainties:

$$R_\chi = \frac{2a - \frac{12}{5}b}{a + 6b} = -\frac{\delta_2}{\delta_0}. \tag{1.23}$$

Traditionally, R_χ has been the translation guide that interfaces experimental splitting measurements with theory, since experimental determinations of this quantity have avoided systematic errors associated with determining the absolute line energy in radiative transitions. However, the direct systematics-minimal determination of a and b in Eqn. **??** is preferable to the combined value R_χ, though comparison with previous experimental and theoretical results is still interesting, primarily because previous measurements of these ratios have not been in universal agreement. A persistent question is whether $R_{\chi(2P)} > R_{\chi(1P)}$, which is contrary to most models but consistent with many of the experimental determinations.

We conclude that experimentally optimized determinations of a, b and R_χ are a valuable contribution for those engaged in effective theory calculations and present results accordingly.

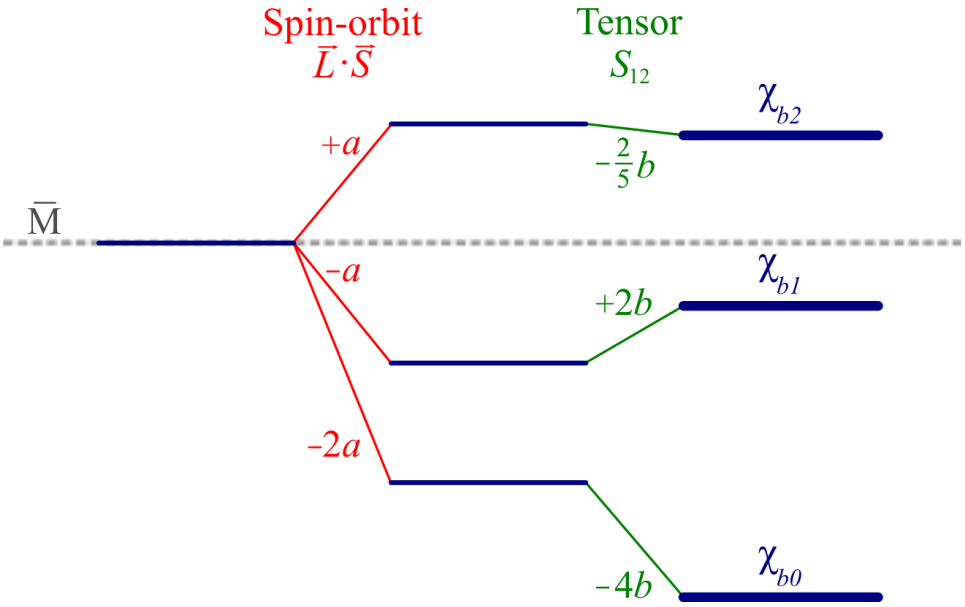


Figure 1.4: A demonstration of the effect of spin-orbit and tensor interactions on the three χ_{bJ} states in terms of the splitting parameters a and b. The gray dashed line is the spin-weighted center of gravity $\overline{M}$ of the system.

Lattice NRQCD

Phenomena in bottomonium are generally categorized into three widely spaced energy scales: hard (the mass m of the quarks), soft (relative quark-antiquark momentum $p \approx mv$, $v \ll 1$) and ultrasoft (binding energy $E \approx mv^2$), all of which are assumed to be larger than the characteristic scale of QCD Λ_{QCD}. The mass of the bottom quark m_b is very large when compared to Λ_{QCD}, allowing perturbative techniques to be applicable in that regime; the lower scales are generally inaccessible to pertubative calculations. The hierarchy $m \gg p \gg E$ with $m \gg \Lambda_{QCD}$ has been exploited to simplify calculations in bottomonium that are dramatically complicated by the presence of these scales simultaneously. This process, described below as it relates to bottomonium, involves the formulation of so-called Effective Field Theories (EFT).

An EFT describes the observables in a low energy region by integrating out the degrees of freedom associated with the higher scales scales. This process gives equivalent results to fundamental QCD with much simpler calculations. Nonrelativistic QCD [NRQCD] is a widely successful EFT [44] that integrates out modes of energy

and momentum from QCD, yielding an expansion in orders of $1/m_b$.

The NRQCD approach has been particularly useful for heavy quarkonium calculations on the lattice. In the nonperturbative regime where the strong coupling constant α_s approaches unity – in bottomonium this essentially corresponds to all states but the $\Upsilon(1S)$ – the simplest way to solve first-principles QCD is to discretize spacetime on an Euclidian grid and describe gauge fields as closed loops on the grid [43], greatly diminishing the computational burden of QCD in the continuum. This approach is referred to as lattice QCD and utilizes numerical methods for calculating path integrals on the lattice. Lattice calculations in NRQCD in the bottomonium system verify the general form of the interquark potential from first principles and accurately predict the bottomonium levels. Simulations of the interquark potentials generally show that the short range vector behavior interpretation of the interquark potential from the previous section to be largely correct and confirms the short-range nature of the spin-dependent forces.

The NRQCD Hamiltonian used in these lattice calculations in terms of chromoelectric $\widetilde{\mathbf{E}}$ and chromomagnetic $\widetilde{\mathbf{B}}$ fields [44] is given by:

$$H = H_0 + \delta H; \tag{1.24}$$

$$H_0 = -\frac{\Delta^{(2)}}{2m_b}, \tag{1.25}$$

$$\begin{aligned}\delta H = &- c_1\frac{(\Delta^{(2)})^2}{8(m_b)^3} + c_2\frac{i}{8(m_b)^2}\left(\nabla\cdot\widetilde{\mathbf{E}} - \widetilde{\mathbf{E}}\cdot\nabla\right)\\ &- c_3\frac{1}{8(m_b)^2}\sigma\cdot\left(\widetilde{\nabla}\times\widetilde{\mathbf{E}} - \widetilde{\mathbf{E}}\times\widetilde{\nabla}\right) - c_4\frac{1}{2m_b}\sigma\cdot\widetilde{\mathbf{B}}\\ &- c_5\frac{\nabla^{(4)}}{24m_b} - c_6\frac{(\nabla^{(2)})^2}{16n(m_b)^2},\end{aligned} \tag{1.26}$$

where ∇ is the symmetric lattice derivative, m_b is the bare bottom quark mass and the $\Delta^{(i)}$ terms are required for discretization. Crucially, the leading-order spin-dependent terms are proportional to the coefficients c_3 and c_4 and correspond to spin-orbit and tensor terms. Analogously to the procedure in Section**??**, the triplet χ_{bJ} splittings can be used to tune these two parameters, which end up being simply related to a and b despite arising from a very different framework. A precision measurement of the

leading-order spin-dependent terms in the NRQCD Hamiltonian benefits all lattice calculations in the same framework, extending the impact of this measurement well beyond bottomonium.

The framing of the c_3 and c_4 tuning from experiment within the lattice community is somewhat different than that of the procedure for a and b described in the previous section. Using the expectation values of the previous section we construct mass combinations that isolate the spin-orbit and tensor terms by canceling expectation values of the relevant operators. For example,

$$-2M_{nP(0)} - 3M_{nP(1)} + 5M_{nP(2)} \tag{1.27}$$

is proportional to c_3 and the combination

$$2M_{nP(0)} - 3M_{nP(1)} + M_{nP(2)} \tag{1.28}$$

is proportional to c_4^2 and independent of c_3 [43, 44]. Generally these mass combinations have been calculated from world average masses and no experimental results have presented optimal measurements. However, we note that these combinations are simply related to the parameters a and b so that an improvement in the precision of the mass combination is possible by avoiding absolute mass measurements. Specifically, the first combination is equal to $12a$ and the second to $-\frac{72}{5}b$.

Although not a completely distinct result from the potential formulation of a and b in the previous section, presenting the splitting parameters in the language of lattice NRQCD with experimentally optimal determination could improve the accuracy of a large array of lattice calculations, not just constrained to bottomonium physics.

E1 matrix elements

The potential model lends itself to a description of electromagnetic [EM] transitions in the bottomonium system using the multipole expansion in the Schrödinger formalism analogous to atomic transitions. Electric dipole [E1] followed by magnetic dipole [M1] are the lowest-order transitions. Furthermore, photon wavelengths for radiative transitions within the bottomonium system are in the range $1.5 - 15fm$, much larger than the inter-quark spacing of $\sim 0.2fm$, thus dipole transitions dominate.

The transition width for E1 transitions with initial state $n_i L_i J_i$ to a final state $n_f L_f J_f$ can be calculated in the potential model using:

$$\Gamma_{E1} = \frac{4}{3} e_b^2 \alpha C_{if} (2J_f + 1) E_\gamma^3 |\langle n_f L_f | r | n_i L_i \rangle|^2, \tag{1.29}$$

where C_{if} is a statistical factor that depends on the initial- and final-state quantum numbers [equal to 1/9 for transitions between S and P states], e_b is the charge of the b quark, E_γ is the photon energy and r is the inter-quark separation. The E1 matrix element $|\langle n_f L_f | r | n_i L_i \rangle|$ depends on the potential model, so direct measurement of the radiative branching fraction serves as model validation.

Some model-independent observations about the E1 transition width are instructive. The factor E_γ^3 strongly favors the higher-energy transitions, such as $\chi_{bJ}(2P) \to \gamma\Upsilon(1S)$. However, the E1 matrix element $|\langle 1S | r | 2P \rangle|$ is expected to be small due to the mismatch in the number of nodes in the initial and final radial wave functions. The phase-space factor $2J_f + 1$ favors $\Upsilon(nS) \to \gamma\chi_{b2}(mP)$ over the similar cascades with $J = 0$ or 1.

Of particular interest, nonrelativistic models do not discriminate between different spin-states in the E1 transition matrix elements. The implication is that a measurement of the spin-dependence of the matrix elements provides direct insight into the nature of spin forces in strong interactions in the context of models with relativistic corrections. As an example, consider the transition $\Upsilon(2S) \to \gamma\chi_{bJ}(1P)$. The ratio of E1 widths of two distinct spin states $J = 0$ to $J = 1$ simplifies to:

$$\frac{\Gamma_{E1}(J=0)}{\Gamma_{E1}(J=1)} = \frac{E^3_{\gamma_{2S\to 1P(J=0)}}}{3E^3_{\gamma_{2S\to 1P(J=1)}}} \frac{|\langle 1P(J=0)|r|2S\rangle|^2}{|\langle 1P(J=1)|r|2S\rangle|^2} \tag{1.30}$$

This ratio is also simplest to measure experimentally, therefore this presents a clear avenue for probing the strong force in relativistic spin interactions.

The E1 transition $\Upsilon(nS) \to \gamma\chi_{bJ}(mP)$

The $\Upsilon(nS)$ states are spin-triplet states which reach only the $\chi_{bJ}(mP)$ states via E1 transitions with large branching fractions [$\mathcal{B}(\Upsilon(2S) \to \gamma\chi_{bJ}(1P)) = 0.18 \pm 0.01$ and $\mathcal{B}(\Upsilon(3S) \to \gamma\chi_{bJ}(2P)) = 0.27 \pm 0.04$]. The photons produced in $\Upsilon(3S) \to \gamma\chi_{bJ}(2P)$

and $\Upsilon(2S) \to \gamma\chi_{bJ}(1P)$ are of low-enough energy [$80 - 160\,\mathrm{MeV}$] that the spin-dependent splitting is resolvable in *BABAR*'s calorimeter. The $J = 2$ and 1 peaks are unfortunately unresolvable in the high-energy $\Upsilon(3S) \to \gamma\chi_{bJ}(1P)$ transition, with a chance of observing the outlying $J = 0$ peak.

As noted above, relativistic potential models generally predict a spin dependence for the radiative branching fractions in this transition [8]. A significant enhancement of the $J = 2$ rate relative to the $J = 1$ rate as well as a suppression of the $J = 0$ rate is expected.

Selection rules

For any fermion/anti-fermion pair, $P = (-1)^{L+1}$ and $C = (-1)^{L+S}$. E1 transitions can only couple states with different parity, whereas M1 transitions are forbidden between these states. Conservation of C and P therefore dictate that $\Delta S = 0, \Delta L = \pm 1$ for E1 transitions and $\Delta S = \pm 1, \Delta L = 0$ for M1. Thus the suppressed M1 transitions are required to reach the spin-singlet η_b states from the $\Upsilon(nS)$ but cannot reach the spin-triplet states χ_{bJ} and $\Upsilon(1D)$.

Fig. **??** shows the states reachable through E1 and M1 transitions from the directly produced $\Upsilon(2,3S)$ states. In this analysis we focus exclusively on the E1 transitions.

The E1 transition $\chi_{bJ}(nP) \to \gamma\Upsilon(nS)$

The χ_{bJ} states, with quantum numbers $J^{PC} = (0,1,2)^{++}$, have different hadronic widths due primarily to selection rules. Color-charge conservation forbids one-gluon decay: a single colored gluon (with $J^P = 1^-$) cannot couple the colorless $b\bar{b}$ initial state with the colorless hardronic decay products. This leaves annihilation through two gluons as the dominant strong process, with annihilation rate proportional to α_s^2. However, this decay route is forbidden for the χ_{b1} due to a generalization of Yang's theorem preventing massive spin-1 particles from decaying to two identical massless spin-1 particles [9]. The annihilation rate for the remaining strongest strong process is proportional to α_s^3, thus the EM width for the $J = 1$ peak is expected to be significantly larger than the other two.

In the exclusive two-photon cascade $\Upsilon \to \gamma\chi_{bJ}; \chi_{bJ} \to \gamma\Upsilon$ (including any legal combination of $\Upsilon(nS)$ and $\chi_{bJ}(mP)$ states), the combined branching fraction for the

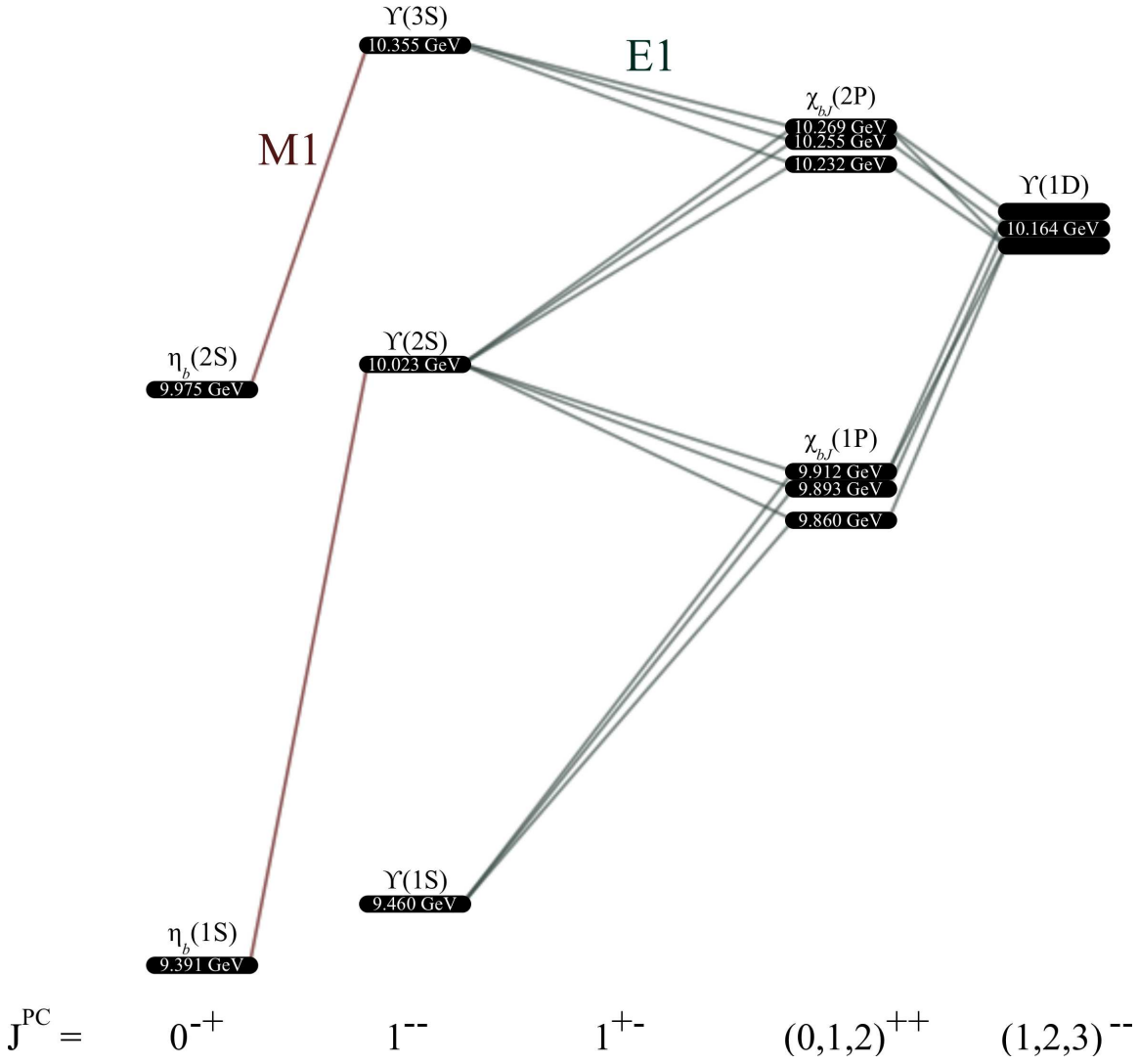


Figure 1.5: Bottomoniuim states reachable through EM transitions from the directly produced $\Upsilon(2,3S)$ states.

$J = 0$ channel is drastically suppressed, making a precision measurement of that quantity a top priority in bottomonium spectroscopy.

1.1.4 Previous bottomonium results

Bottomonium was discovered experimentally by the E288 collaboration at Fermilab in 1977 [3]. In subsequent years most of the states below the open-flavor threshold

have been observed, with a flurry of new activity in the last 5 years. Most recently the singlet ground state $\eta_b(1S)$[28] was observed for the first time at *BABAR*. The CLEO experiment released the first observation of one of the three $\Upsilon(1D)$ states in 2004[25] and became the first experiment to see the transition $\chi_{b0}(1P) \to \gamma\Upsilon(1S)$[21]. Belle saw both the $h_b(1P)$ and $h_b(2P)$[26] – the spin-singlet $L = 1$ states – as well as the $\eta_b(2S)$ – spin-singlets with $L = 0$ – in 2012 [27]. At the time of this writing, only the $\eta_b(3S)$ and two of the three $\Upsilon(1D)$ states have yet to be definitively seen below the open-flavor threshold.

Despite experimental evidence for a large fraction of the relevant bottomonium states, significant work remains to be done. Some of the states – for example the $\chi_{b0}(nP)$ – are poorly known even as new results continue to come in.

Bottomonium at *BABAR*

From 1999 to 2007 the PEP-II collider ran at energies above the open-flavor threshold, mostly at the $\Upsilon(4S)$ resonance. For the last four months of operation prior to decommissioning in April 2008 ["Run 7"] PEP-II recorded an integrated luminosity of $30.22\,\text{fb}^{-1}$ at the $\Upsilon(3S)$ and $14.45\,\text{fb}^{-1}$ at the $\Upsilon(2S)$ resonances corresponding to $122M$ and $100M$ events, respectively. As of writing these collections constitute the largest $\Upsilon(3S)$ and second-largest $\Upsilon(2S)$ collections in the world [Table **??**]. Run 7 data has yielded many bottomonium results, perhaps the most prominent being the discovery of the bottomonium ground state $\eta_b(1S)$ in M1 transitions from the $\Upsilon(3S)$ and first evidence for the $h_b(1P)$ in hadronic transitions. Most relevant for this analysis are results of an inclusive search of Run 7 data using photons converted to electron pairs in the detector material [59]. This approach provides a considerable improvement in photon resolution. However, the conversion rate is low, particularly for lower-energy photons such as those from $\Upsilon(2S) \to \gamma\chi_{bJ}(1P)$, so the technique is best suited for probing the $\chi_{bJ}(nP) \to \Upsilon(mS)$ and $\Upsilon(3S) \to \chi_{bJ}(1P)$ transitions.

1.2 Analysis scope

Our goal is to achieve world-best measurements of E1 branching fractions in the bottomonium system and experimentally optimized fine splitting parameters for the

$\chi_{bJ}(nP)$ states. Whereas previous results have achieved high precision in the higher-energy transitions using the inclusive spectrum or converted photons we target the low-energy transitions in an exclusive reconstruction with a $\gamma\gamma\ell^+\ell^-$ final state. The exclusive mode provides a very clean signal albeit with substantially reduced statistics. The archetype exclusive process for this analysis is $\Upsilon(2S) \to \gamma\chi_{bJ}(1P); \chi_{bJ}(1P) \to \gamma\Upsilon(1S); \Upsilon(1S) \to \ell^+\ell^-$. It is in this exclusive mode that CLEO first observed the transition $\chi_{b0}(1P) \to \Upsilon(1S)$ in 2011 [21].

[Before proceeding we introduce some notation in order to simplify discussion of these exclusive processes. For the cascade discussed above we use the notation $2S \to 1P \to 1S$. Each $\to$ represents one E1 transition with photons labeled $\gamma_{2S\to1P}$ and $\gamma_{1P\to1S}$. It is implied that the final state in the chain decays to two leptons; as it turns out these will always be μ pairs. The first-leg photons will often be referred to as "soft" and the second-leg photons as "hard" when the context is unambiguous.]

The EMC provides reasonable separation of the three χ_{bJ} peaks in the low-energy regime [below $\sim 200\,\mathrm{MeV}$] but separation is impossible for the higher-energy transitions. Therefore this analysis focuses on these four cascades:

- $2S \to 1P \to 1S$: Fit $\gamma_{2S\to1P}$ spectrum.
- $3S \to 2P \to 1S$: Fit $\gamma_{3S\to2P}$ spectrum.
- $3S \to 2P \to 2S$: Fit $\gamma_{3S\to2P}$ spectrum.
- ($3S \to 2P \to 1D \to 1P \to 1S$: Fit combined $\gamma_{3S\to2P}$ and $\gamma_{2P\to1D}$ spectrum.)

The final bullet turns out to be a fruitless analysis, covered only in Appendix A. The $3S \to 1P \to 1S$ cascade is excluded because both photons in the cascade are at an energy range where the EMC resolution prevents disentanglement of the three overlapping peaks.

In order to allow cancellation of reconstruction and cut systematic errors we measure the ratios of product branching fractions:

$$\mathscr{F}^{J/J'}_{nS\to mP\to kS} \equiv \frac{\mathcal{B}_{nS\to mP(J)} \times \mathcal{B}_{mP(J)\to kS}}{\mathcal{B}_{nS\to mP(J')} \times \mathcal{B}_{mP(J')\to kS}}, \tag{1.31}$$

This quantity can be determined to high precision, complementing less precise determinations of the individual branching fractions. The final results will include the following ratios: $\mathscr{F}^{0/1}_{2S\to1P\to1S}$, $\mathscr{F}^{2/1}_{2S\to1P\to1S}$, $\mathscr{F}^{0/1}_{3S\to2P\to1S}$, $\mathscr{F}^{2/1}_{3S\to2P\to1S}$, $\mathscr{F}^{0/1}_{3S\to2P\to2S}$, $\mathscr{F}^{2/1}_{3S\to2P\to2S}$, and relevant ratios for the $\Upsilon(1D)$ cascades. Derivation of single-leg or ratio branching fractions are possible using results from other analyses.

Although this analysis is not ideally suited for measuring the spin-orbit mass splittings of the nP states due to low statistics in the exclusive mode, we will report these measurements. Once again the strategy is to bypass large systematic effects; in this case the systematics involved in extracting absolute line energy from a fit to the EMC line shape are minimized by ignoring the absolute energy scale of the photon lines and only extracting splitting energies.

1.2.1 Available data

In addition to data taken at the $\Upsilon(2,3S)$ resonances, MC simulations of the main signal and background modes are used to generate a model spectrum and probe backgrounds and efficiencies. This section contains a summary of the datasets used.

Data

During *BABAR*'s final run (Run 7), extended operation at the $\Upsilon(3S)$ and $\Upsilon(2S)$ resonances produced what was then the largest collection of data taken at these two bottomonium states in the world. Table **??** shows the amount of data available for this analysis after the conclusion of Run 7 as well as the comparable collections utilized by both the CLEO and Belle collaborations as of cessation of operations for all three experiments.

The $(2,3)S\to(1,2)P$ transition produces signal photons $\gamma_{(2,3)S\to(1,2)P}$ in the low energy range $86-163\,\mathrm{MeV}$, a region with large numbers of "beam" photons – that is, photons with origins in the beam pipe and not associated with physics events. This large low-energy background presents a significant challenge for extracting peak yields in the $\gamma_{(2,3)S\to(1,2)P}$ spectrum. This challenge is particularly acute for the experiments at asymmetric colliders, namely Belle and *BABAR*. CLEO, operating at a symmetric collider, sees much less low-energy neutral background and also has better calorimeter

resolution, improving peak separation. *BABAR*'s key advantage over CLEO is an order of magnitude more data; the advantage over Belle is an order of magnitude more data in $\Upsilon(3S)$ and the temporary fact that Belle has not completed a similar analysis on the $\Upsilon(2S)$ dataset.

Experiment	$\Upsilon(2S)$	$\Upsilon(3S)$
BABAR	$98.6M$	$122M$
Belle	$158M$	$11M$
CLEO	$9M$	$6M$

Monte Carlo

A summary of the available relevant Monte Carlo (MC) collections is shown in Table **??**. All other MC modes contribute negligible background in the exclusive reconstruction. The most important use of these modes is in the assembly of the "Representative MC Ensemble" used as a full simulation of the signal spectrum. The scales given are used to adjust the number of events in each mode to match the expected branching fractions for each mode in order to create the ensemble. The signal MC [given the internal identifiers "SP modes" 9244, 9245 and 9246] are full simulations of the three exclusive signal cascades without mixing from any other process. Consequently, the signal cascades in the $\Upsilon(2S)$ generic [SP mode 9016] collection are manually removed to allow this mode to be used as an uncontaminated continuum background source.

SP #	Mode	# MC	# Run 7	Scale
9016	Generic $\Upsilon(2S)$ decays:	$332M$	$98.6M$	0.30
	$\Upsilon(2S) \to \pi^0\pi^0\Upsilon(1S)$		$0.2M$	
	$\Upsilon(2S) \to \mu\mu$		$1.9M$	
	All other $\Upsilon(2S)$ generic		$96.5M$	
9244	Full $\chi_{b0}(1P)$ cascade	$87,000$	$1.2K$	0.014
9245	Full $\chi_{b1}(1P)$ cascade	$709,000$	$57K$	0.080
9246	Full $\chi_{b2}(1P)$ cascade	$493,000$	$33K$	0.068
3981	$\mu\mu(\gamma_{FSR})$	$16.7M$	$15.5M$	0.93

Table 1.1: A summary of the relevant MC modes used to construct the representative sample. Column 3 shows the number of events included in the MC collection. Column 4 contains the number expected in Run 7 based on expected decay rates and cross-sections given *BABAR*'s recorded luminosity at the $\Upsilon(2S)$ resonance using branching fractions from the PDG [53]. Column 5 shows the factor required to scale each MC sample to the size expected in data.

Chapter 2

The *BABAR* detector

This chapter summarizes the detector hardware utilized by *BABAR*. A wise reader would refer to the full detector descriptions in [48].

2.1 Physics motivation

The impetus for the *BABAR* experiment was a study of $CP-$violating processes in the decay of B mesons. Consequently the design of the PEP-II $B-$Factory and *BABAR* detector itself have been optimized for this physics program. However, the high luminosity provided by PEP-II to *BABAR* throughout its almost-10-year operation and stringent detector performance demands of the $CP-$violation program provide opportunities for a wide array of physics topics, including bottomonium transition studies.

BABAR's aim is to generate an enormous number of entangled $B^0 - \overline{B^0}$ pairs in decays of the $\Upsilon(4S)$ resonance. Each meson in the entangled pair can oscillate between the meson and the anti-meson states provided that the overall entangled state contains one meson and one anti-meson. However, each meson is free to decay; once one does, the coherence is lost and the remaining B^0 or $\overline{B^0}$ is free to oscillate until its own decay. The asymmetries in the time-dependent behavior of B^0 versus $\overline{B^0}$ mesons can be used to measure $CP-$violation parameters and thus probe the Standard Model. From an experimental perspective the focus is on correct flavor tagging of each meson produced and reconstruction of the meson decay vertices. These requirements flavor

the entire instrumental design as will be shown in the following material.

2.2 PEP-II

The PEP-II B-Factory is an asymmetric e^+e^- collider located at SLAC National Laboratory. To suit its original physics objective as a factory for B mesons, PEP-II ran at a center of mass energy $\sqrt{s} = 10.58\,\text{GeV}$ with beam energies of 9.0 GeV for elections and 3.1 GeV for positrons for the majority of its run, corresponding to the $\Upsilon(4S)$ resonance. The purpose of the asymmetric design is to provide a Lorentz boost to the produced B mesons in the lab frame to the extent that the decay vertices are separable from the interaction point. This feature is not utilized in bottomonium programs and indeed contributes to a higher level of background than a symmetric collider operating at the same center of mass energy given the nonlinear behavior of background production processes.

For data taken at lower resonances the energy of the positron beam remains unchanged while the energy of the electron beam is reduced to 8.61 GeV and 8.07 GeV for the $\Upsilon(3S)$ and $\Upsilon(2S)$ resonances, respectively.

2.3 Detector systems

The *BABAR* detector, pictured in Figs. **??** and **??**, consists of five principle subdetectors. The purpose of these subdetectors is summarized here, while technical details appear in the following sections:

Silicon Vertex Tracker (SVT) measures angles and positions of charged particles in the very inner layer of the detector.

Drift Chamber (DCH) measures momentum for charged particles as well as dE/dx for particle identification. The DCH and SVT provide a comprehensive particle tracking system.

Detector of Internally Reflected Cherenkov light (DIRC) provides particle identification targeted primarily at K/π discrimination.

Electromagnetic Calorimeter (EMC) measures time and energy of photons; also provides dE/dx information for track ID.

Instrumented Flux Return (IFR) identifies muons and neutral hadrons.

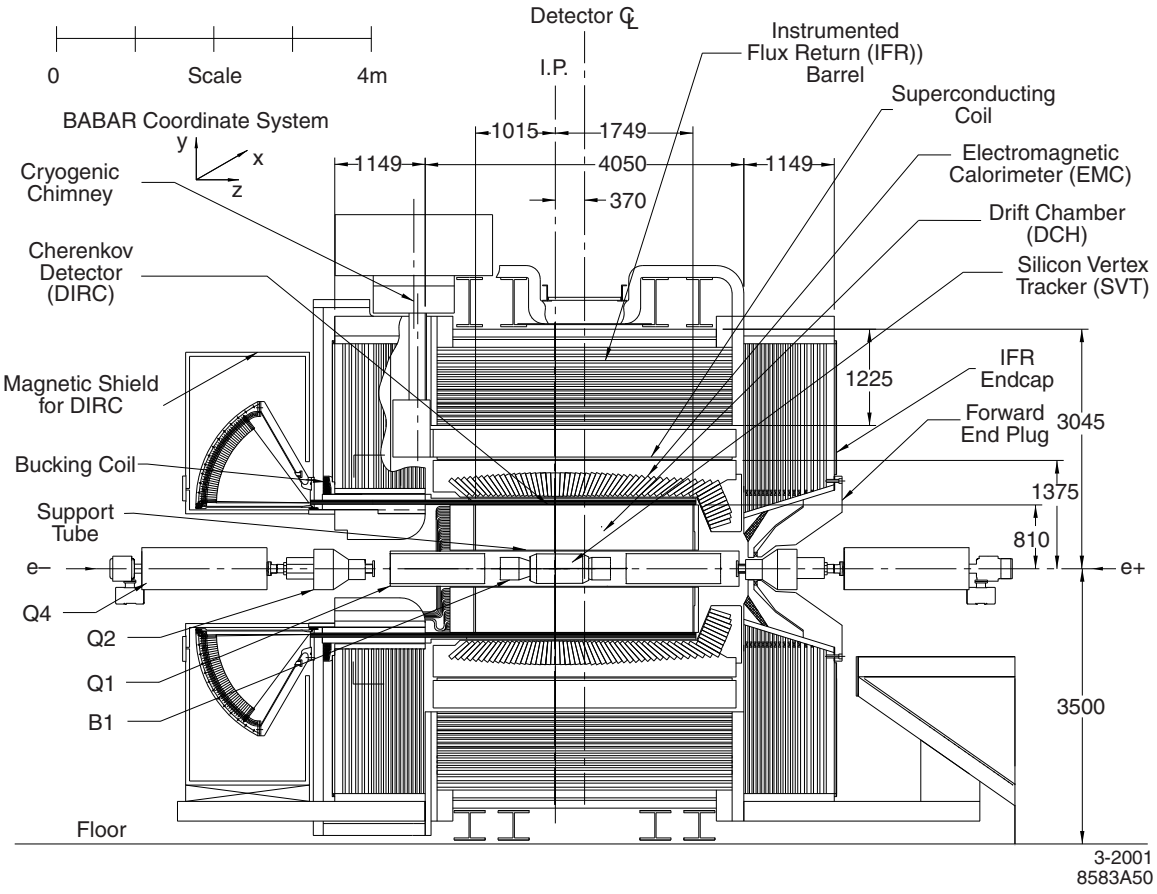


Figure 2.1: Longitudinal section of the *BABAR* detector.

2.3.1 Silicon Vertex Tracker

Boosted B mesons decay roughly $< \beta\gamma c\tau_{B^0} > \sim 260\mu$m from the interaction point (IP); high-precision tracking of charged particles as close to the IP as possible is critical in order to consistently resolve this separation. To that end the SVT (Fig. **??** and Fig. **??**) has three layers of double-sided silicon strip detectors immediately outside the beam pipe. The strips of one side are oriented perpendicular to the strips on the

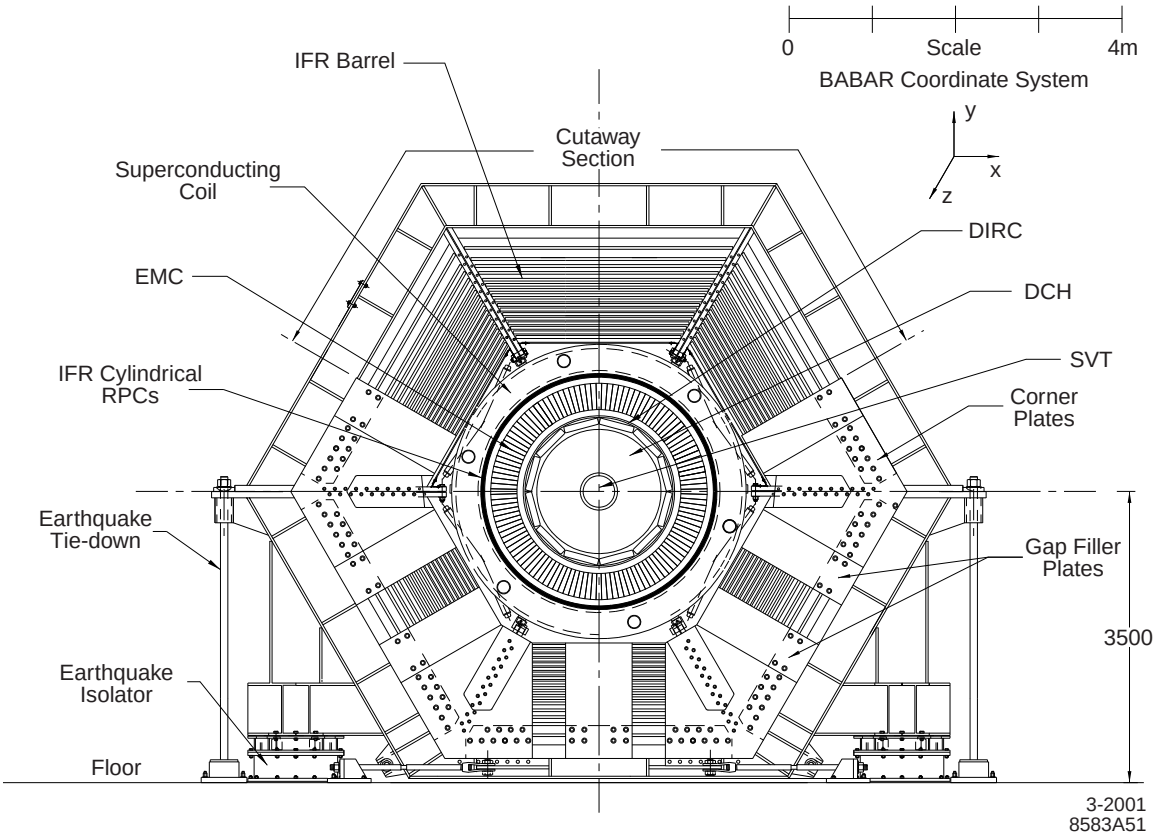


Figure 2.2: End view of the *BABAR* detector.

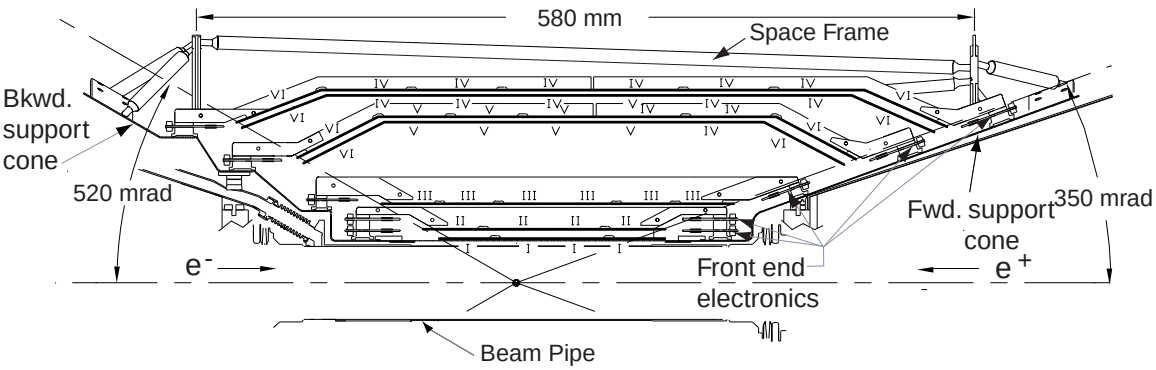


Figure 2.3: Longitudinal section of the SVT.

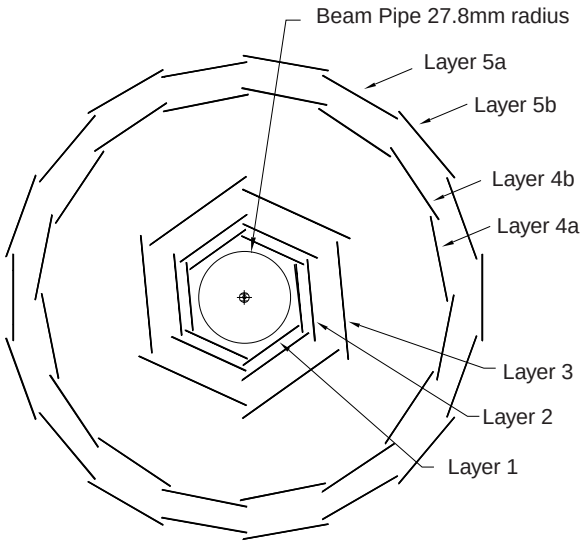


Figure 2.4: SVT transverse section, showing the placement of the five detector layers.

other side to provide z-position information. Two more layers are placed further from the IP in order to obtain a long lever arm for track angle measurements.

The DCH (see the following section) cannot adequately measure low-or high-momentum tracks due to a cutoff in acceptance below about 120 MeV and an unmeasureably small sagitta for minimally curved tracks. The SVT can see particles with low-enough momenta and measures track curvature precisely enough to complement the DCH in these respects.

A major design consideration of the SVT is to minimize the radiation length X_0 of material in the detector. Multiple-scattering in detector material degrades the momentum resolution for charged tracks. Electron pairs produced in photon conversions within the detector material can lose significant unrecovered energy by bremsstrahlung, thus degrading the photon reconstruction. Additionally these conversion events suffer lower efficiency due to the electrons being bent out of the detector by the magnetic field. At normal incidence the SVT is less than $0.04X_0$ thick.

The aggressive design and placement of the SVT is optimized for reconstruction of $B-$decay vertices and performance in this respect is not crucial in bottomonium physics programs. However, for the sake of completeness we note that the highest premium is placed on z-axis resolution, which is about 80μm, sufficient to identify $B-$decay vertices. The $x-y$ resolution is sufficient to resolve the decay vertices of

light charm mesons and is around 100 MeV. It is possible to determine the resolution of the SVT by using cosmic ray tracks that pass through the IP – considering the incoming and outgoing tracks as two separate particles, the differences in the measured parameters for the two tracks gives the parameter resolutions: 23μm, 29μm, 0.32$mrad$ and 0.53×10^{-3} for the distance of closes approach, the azimuthal angle and z of the track at the closest approach and the dip angle (the angle between the track momentum and the z-axis).

2.3.2 Drift Chamber

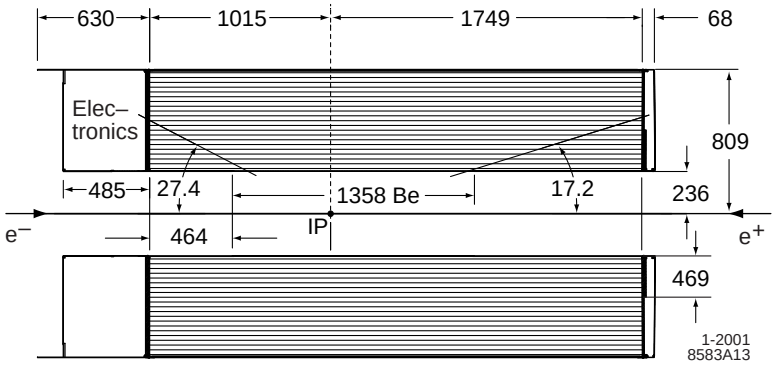


Figure 2.5: Longitudinal section of the DCH.

The DCH, shown in Fig. **??**, is composed of 40 layers of roughly hexagonal drift cells with low-mass wires at each node filled with an 80:20 helium:isobutane gas mixture. The principle purpose of the DCH is to measure the momentum and angles of charged particles with high precision and efficiency. Transiting particles leave up to 40 ionization loss measurements, with longitudinal information coming from 24 of the layers being placed at a small angle with respect to the z-axis.

Each cell (Fig. **??**), approximately 11.9mm×19.0mm in radial×azimuthal dimensions, is composed of a single sense wire at a positive high electrical potential surrounded by six field wires at ground potential. This configuration provides the electric potential required to draw ionized gas molecules to the sense wires while maintaining a roughly circular distribution of isochrone lines (contours of equal drift time) in the majority of the detector volume (Fig. **??**). A total of 28,768 high-tension wires in 7,104 channels provide tracking for this detector which is second-closest to the IP.

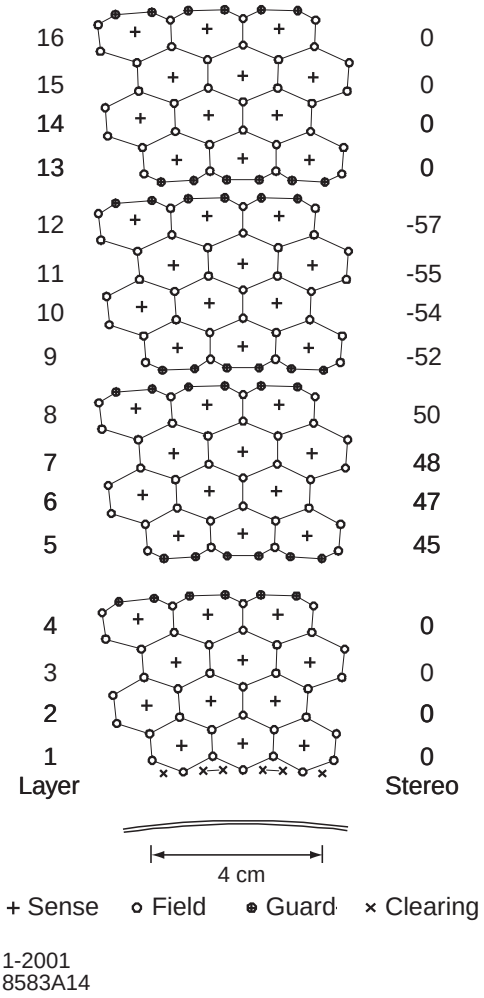


Figure 2.6: Transverse section of drift cells for the four innermost layers. The lines between the field wires are not physical.

The dE/dx measurements provided by the DCH (Fig. **??**) assist in particle ID; specifically, the signature of background protons is distinct and K/π separation is reliable up to about 800 MeV.

At normal incidence the DCH is $0.0108X_0$ thick, a crucially small number given the inner position of the detector. The transverse momentum resolution σ_{p_t} is determined with cosmic ray events as described in the previous section, with:

$$\frac{\sigma_{p_t}}{p_t} = (0.13 \pm 0.01)\% \times \left(\frac{p_t}{1\,\mathrm{GeV}}\right) + (0.45 \pm 0.03)\% \tag{2.1}$$

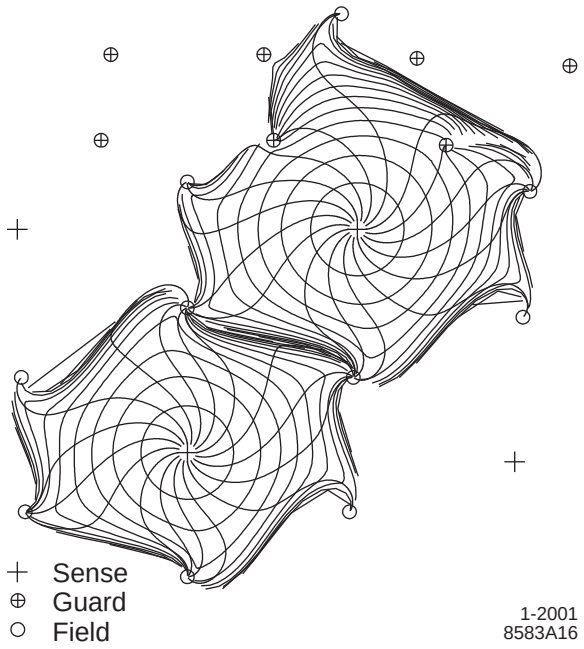


Figure 2.7: $100ns$ isochrones for two layers of the DCH. The isochrones are circular near the sense wires but are complicated in shape near the field wires.

The first term comes from measurement accuracy and the second term is caused by multiple Coulomb scattering, which dominates the momentum error except for the highest momentum tracks.

2.3.3 Detector of Internally Reflected Cherenkov light

To satisfy CP-violation physics program requirements *BABAR* needs to be able to tag the flavor of one of the B mesons using kaons up to 2 GeV. As shown in Fig. **??**, the DCH is unable to provide K/π separation at energies above 700 MeV. DIRC is a first-of-its-kind detector for particle ID that provides a $\sim 4\sigma$ or greater K/π discrimination from the Cherenkov threshold up to 4.2 GeV.

When a particle with high momentum enters a material with an elevated index of refraction n the local speed of light c/n may decrease sufficiently that the particle is now traveling faster than the local speed of light. The excess energy is radiated as a cone of light called Cherenkov radiation. The cone is comprised of photons at a characteristic angle θ_C with respect to the flight direction, where $cos\theta_C = 1/(nv/c)$;

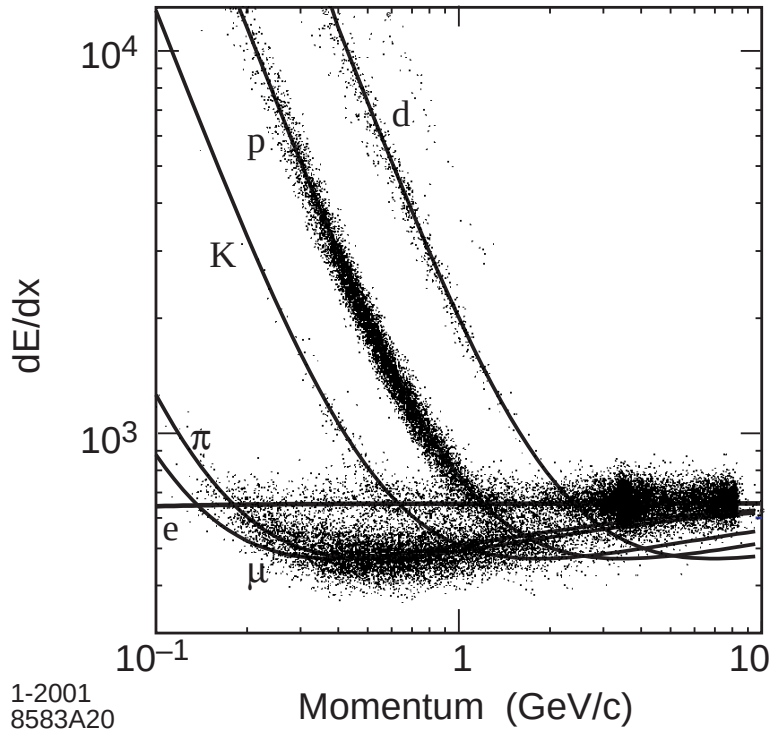


Figure 2.8: Measurements of dE/dx in the DCH as a function of track momentum compared to Bethe-Block predictions for various source particles.

a measurement of the cone angle is a measurement of the track velocity.

DIRC (Fig.**??**) is composed of 144 4.9m-long, 17mm-thick and 35mm-wide fused silica bars arrayed around the IP in the configuration shown in Fig. **??**. The bars have parallel sides, preserving all internal reflection angles. This configuration is essentially a light pipe that shepherds the Cherenkov cone photons to the large water tank – matched in refractive index to minimize internal reflection losses – which is covered on the rear face with photomultiplier [PMT] tubes. The Cherenkov cones are visible as rings in the PMT array. The timing of the arrival of the photons is also used to discriminate against photons from other backgrounds, mostly low-energy machine sources.

Fig. **??** shows the measured Cherenkov angle and photon arrival time for single muons in $\mu^+\mu^-$ events. Photons outisde of $\pm$8ns of the expected arrival time Δt_γ are removed to reject the majority of the machine background photons in the nominal 300ns trigger window. The unique design of DIRC has been extremely useful

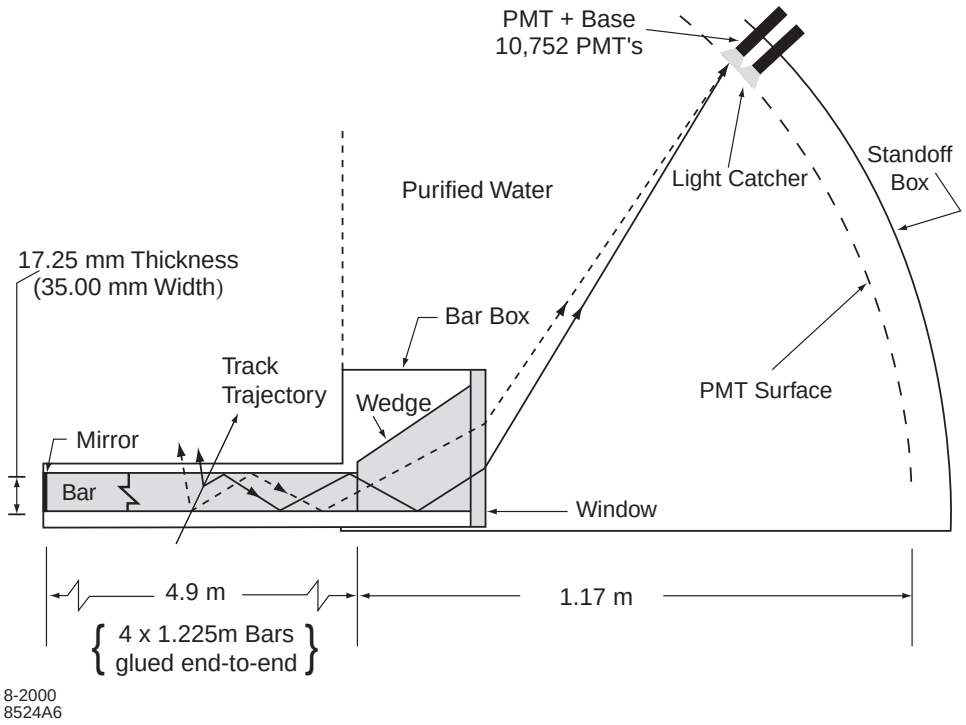


Figure 2.9: Schematic of the DIRC system.

in K/π discrimination (see Fig. **??**) – however, this ability is of minimal utility in bottomonium analyses.

2.3.4 Electromagnetic Calorimeter

The physics topic of this dissertation depends intimately on the performance and characteristsics of the EMC. An in-depth discussion of the EMC electronics pipeline with a focus on cluster timing is found in Appendix C, complementing the more general discussion here.

The EMC measures electromagnetic showers from neutral particles and charged tracks with excellent efficiency as well as energy and angular momentum resolution. Photons from 20 MeV to 9 GeV are almost fully recovered in the EMC, while dE/dx measurements of tracks assist in particle ID.

A finely segmented array of $6,580$ trapezoidal thallium-doped cesium-iodide crystals are arranged in a projective geometry covering 15.8° to 141.8° in polar angle with

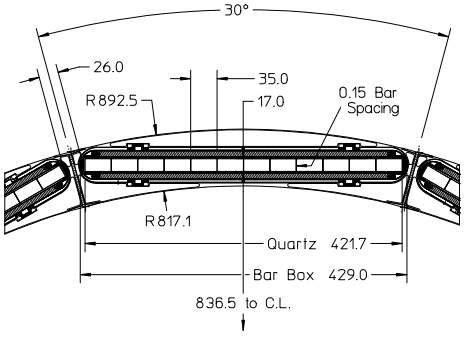


Figure 2.10: Transverse section of the DIRC bar arrangement with dimensions in mm.

full azimuthal coverage (see Fig. **??**). The crystals are arranged in the barrel region in 48 rings with 120 identical crystals each and with the remaining 820 crystals in the forward endcap region. The crystal length is chosen so that typical EM showers are almost entirely contained within the detector material; crystals in the forward region are slightly longer to accomodate a higher expected photon energy distribution due to the boost. To minimize preshowering the crystals are supported from the back, with less than $0.3 - 0.6X_0$ of material in front of the crystal faces for all but the inner three rings of endcap crystals which are shadowed by the SVT and magnet structure.

Each crystal has a unique response which can also degrade over time due to exposure to beam radiation, particularly the inner end of the crystal. Since the degradation is strongest in the inside end, energy calibrations at two different scales are required to target the inner more degraded areas with low-energy, short-penetrating photons and

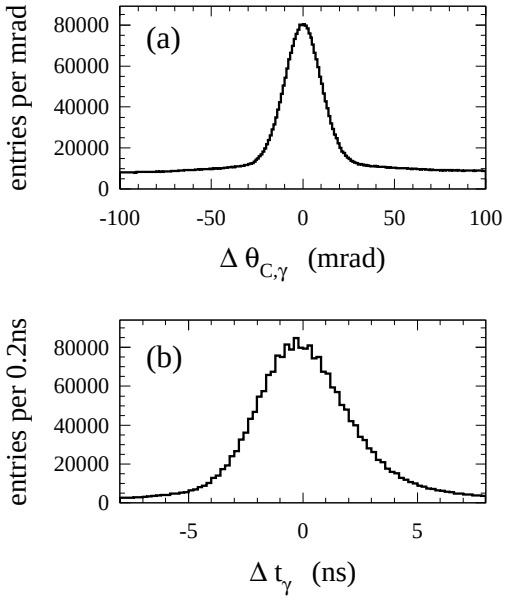


Figure 2.11: The difference between the measured and expected Cherenkov angle (a) and photon arrival time (b) for single muons in $\mu^+\mu^-$ events.

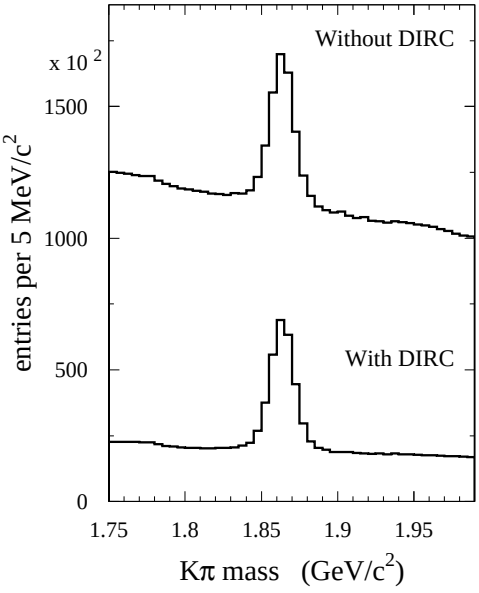


Figure 2.12: Invariant $K\pi$ mass spectrum with and without the DIRC system for K/π discrimination in the D^0 mass peak window.

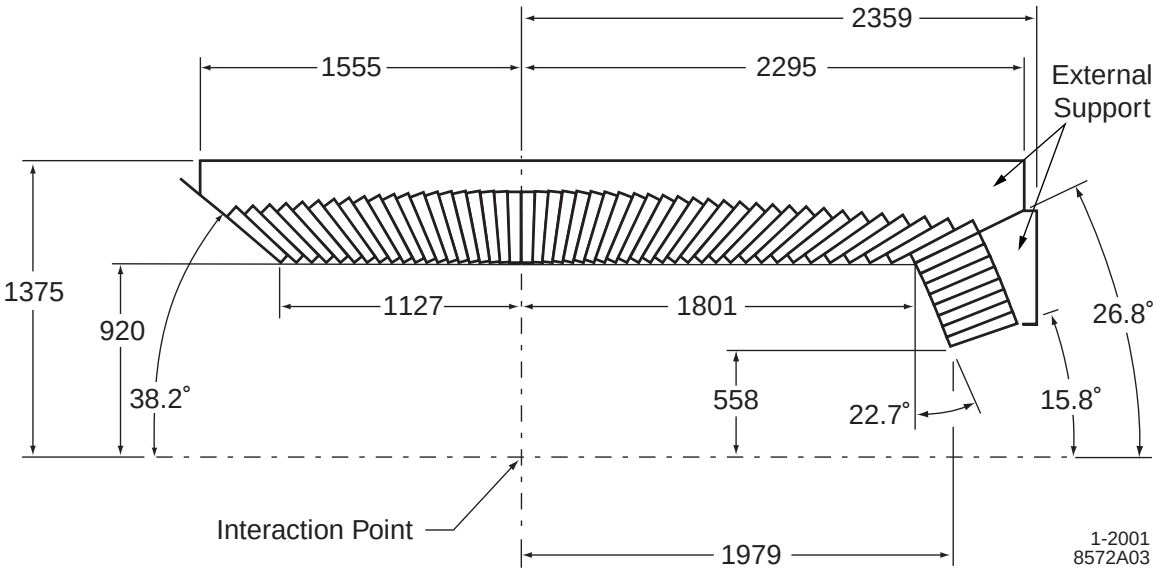


Figure 2.13: Longitudinal section of the top portion of the EMC. The forward endcap region is to the right.

reach the back of the crystals with high-energy deep-penetrating photons. The following energy calibration is performed regularly (for a timing calibration performed in conjunction with this analysis see Appendix C): a flux of low-energy neutrons is used weekly to irradiate Fluorinert, which consequently produces 6.13 MeV photons used in the inner region absolute energy calibration. The Flourinert is pumped through pipes immediately in front of the crystals. For the high energy calibration, copious Bhabha scattering events (taken from a sample of rejected Bhabha events during physics runs) are isolated and a relationship between the electron polar angle and energy is exploited to determine the crystal energy error. The sum

$$\chi^2 = \sum_k \left[\frac{\sum_i c_i \epsilon_i^k - E_{dep}^k(\theta, \phi)}{\sigma^k} \right]^2 \tag{2.2}$$

is minimized where the sum runs over the single crystal energies ϵ_i^k of all clusters k with energy resolution σ_i and the deposited cluster energy $E_{dep}^k(\theta, \phi)$ is determined from MC. A logarithmic interpolation is used to fill the gap between the high- and low-energy calibrations.

Even perfectly calibrated crystals will experience energy loss due to absorbtion and

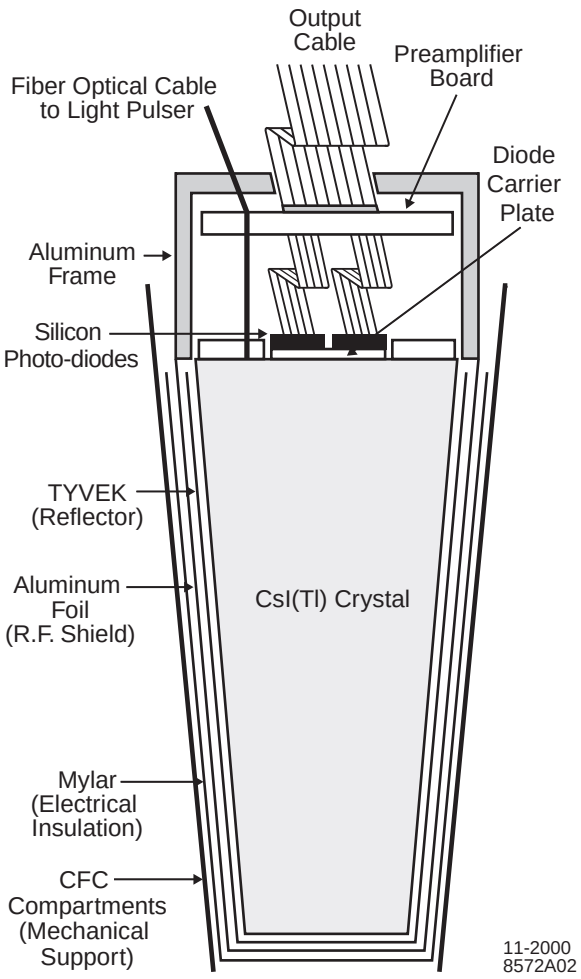


Figure 2.14: Schematic of a single EMC crystal with readout electronics mounted on the rear face. The mechanical support (CFC) is shown opening at the front to minimize preshowering.

shower leakage. Cluster energy corrections depend mostly on the energy deposited and polar angle. Below 800 MeV these corrections are derived from π^0 decays, fitting coefficients for correction terms proportional to $\ln E$ and $\cos\theta$ to correctly return the π^0 mass. At higher energies $\mu\mu\gamma$ events are used to constrain the cluster energy for photons over 1 GeV.

The energy resolution of the EMC is thus parameterized as:

$$\frac{\sigma_E}{E} = \frac{(2.30 \pm 0.03 \pm 0.3)\%}{\sqrt[4]{E(\,\mathrm{GeV})}} \oplus (1.35 \pm 0.08 \pm 0.2)\%, \tag{2.3}$$

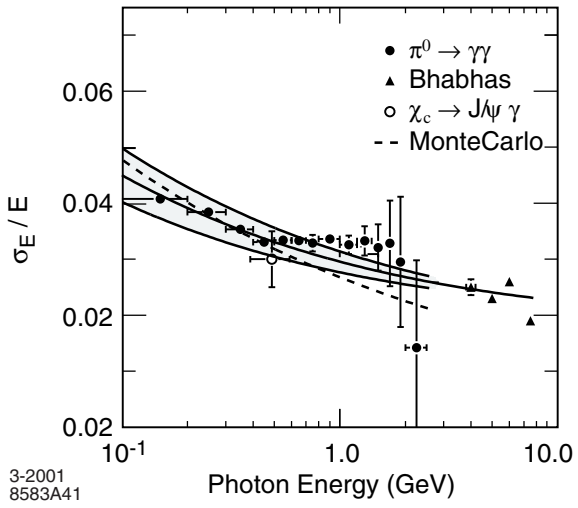


Figure 2.15: Energy resolution of the EMC determined from a variety of photon sources with the parameterization of Eqn. **??** shown as a solid curve with an error range.

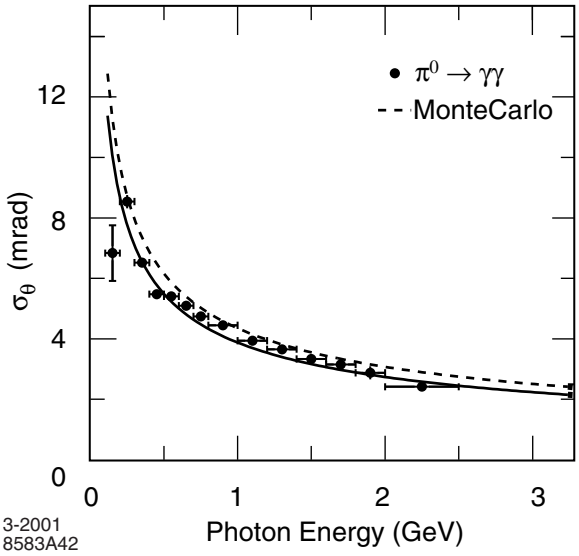


Figure 2.16: Angular resolution of the EMC determined from photons from π^0 decays with the parameterization of Eqn. **??** shown as a solid curve.

where $E$is the photon energy with RMS error σ_E in GeV. This parameterization is shown in Fig. **??**. This parameterization is poorly constrained for photon energies relevant to this analysis, which will be a significant topic in later sections of this

dissertation.

With a projective geometry the angular resolution of the EMC is determined almost entirely by the angular granularity of the detector and the Moliére radius of CsI(Tl), which is the radius in which 90% of the energy of a shower is contained. Symmetric π^0 and η decays are used to parameterize the angular resolution:

$$\sigma_\theta = \sigma_\phi = \left[\frac{4.16 \pm 0.04}{\sqrt{E(\,\mathrm{GeV})}} \oplus 0.00 \pm 0.04\right] mrad, \tag{2.4}$$

shown in Fig. **??**.

2.3.5 Instrumented Flux Return

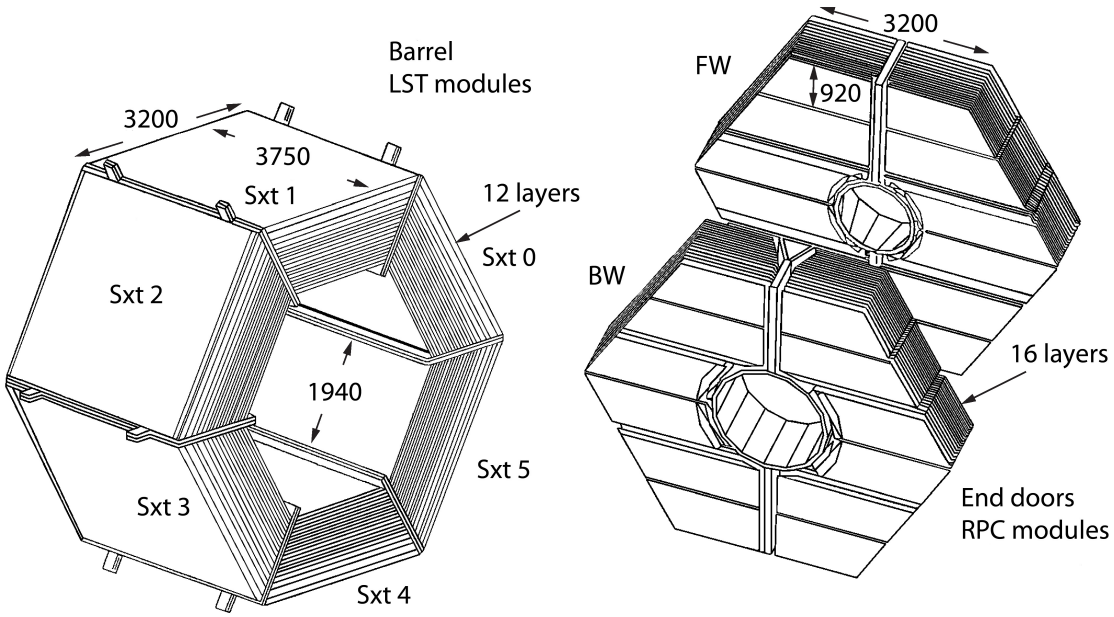


Figure 2.17: Layout of the IFR with forward (FW) and backward (BW) end doors with dimensions in mm.

The bottomonium analysis presented herein requires identification of $\mu\mu\gamma\gamma$ final states. The EMC details in the last section are relevant to the two photons; the IFR is the instrument used for muon identification.

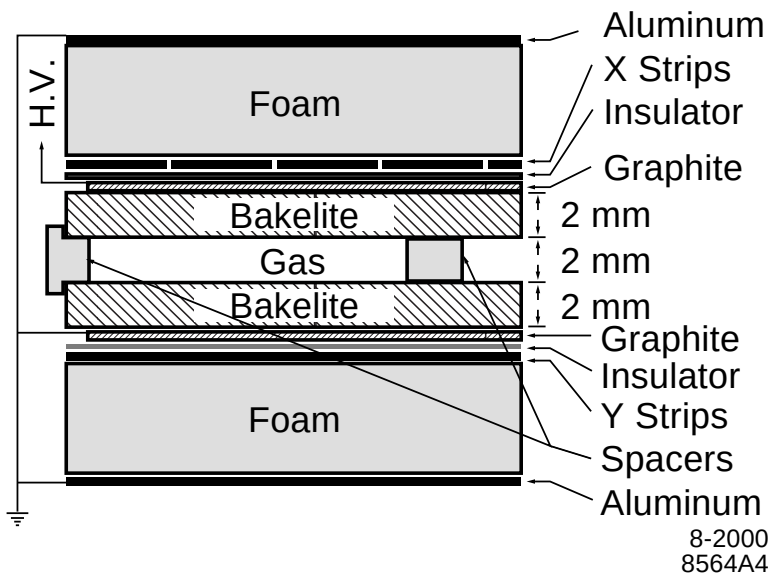


Figure 2.18: Schematic cross-section of a planar RPC.

The IFR (see Fig. **??**) is used to identify muons and neutral hadrons and is located immediately outside the cryostat housing the $1.5T$ superconducting magnet coil. Prior to 2006, the space between steel flux return segments was filled with single-gap resistive plate chambers (RPCs), shown schematically in Fig. **??**, with 19 layers in the barrel and 18 in the endcaps. This system experienced rapidly declining performance (Fig. **??**) and was replaced with a higher-functioning configuration which is described here in more detail.

In the forward encap region the upgraded IFR consists of over 200 upgraded RPCs in four "doors", plus 12 layers of Limited Streamer Tubes (LSTs) in the barrel (Fig. **??**) region, which consists of six sections in a hexagonal pattern. LSTs consist of layers of gas-filled cells of square cross-section with a single wire at high voltage in the center. A plane is mounted below the tube with strips perpendicular to the cell axis to gather induced charge. The wires acts as the readout channels, with charge determining the ϕ coordinate and induced charge on the plane giving a z-coordinate. Multiple layers give r-coordinate information, allowing three-dimensional reconstruction to be performed with a simple and robust device.

Pion misindentification and muon efficiency are both improved in relation to the

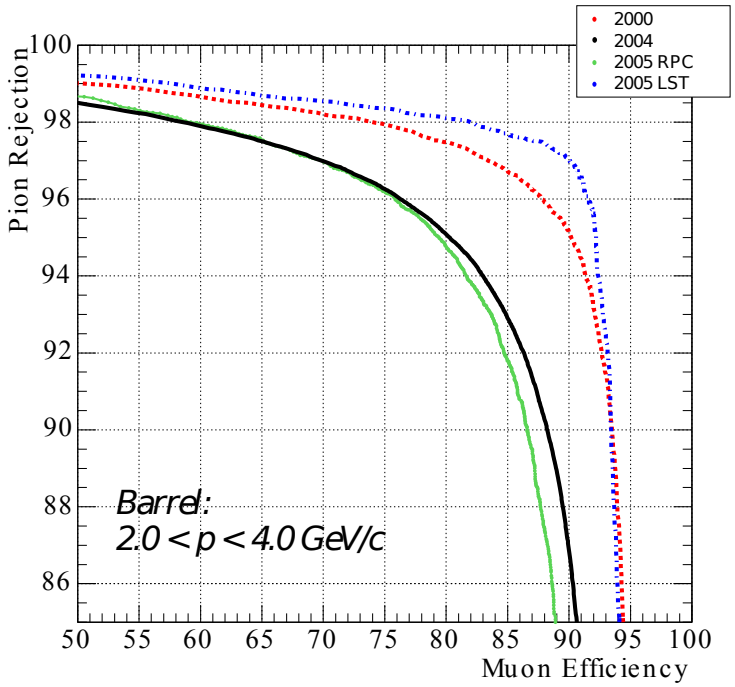


Figure 2.19: Pion rejection rate versus muon efficiency for the IFR system before the 2006 upgrade. Note specifically the decrease in both pion rejection and muon efficiency between 2000 and 2005. The blue line represents a small number of LSTs installed in 2004 as part of a phased installation.

original RPC design (Fig. **??**) with better longevity. The muon identification in the analyses presented in this document was achieved exclusively with the new IFR system.

2.4 Trigger

Information from the various detector subcomponents combine to produce a highly efficient trigger, with a total trigger efficiency greater than 0.99 for $B\overline{B}$ events and greater than 0.95 for continuum events. The two stages of triggers are discussed here; the first, Level 1, is a hardware trigger. Level 3 is a software trigger that uses real-time partial reconstruction of physics events (a contingency for a Level 2 trigger was never activated). The trigger system was designed to accomodate up to ten times the

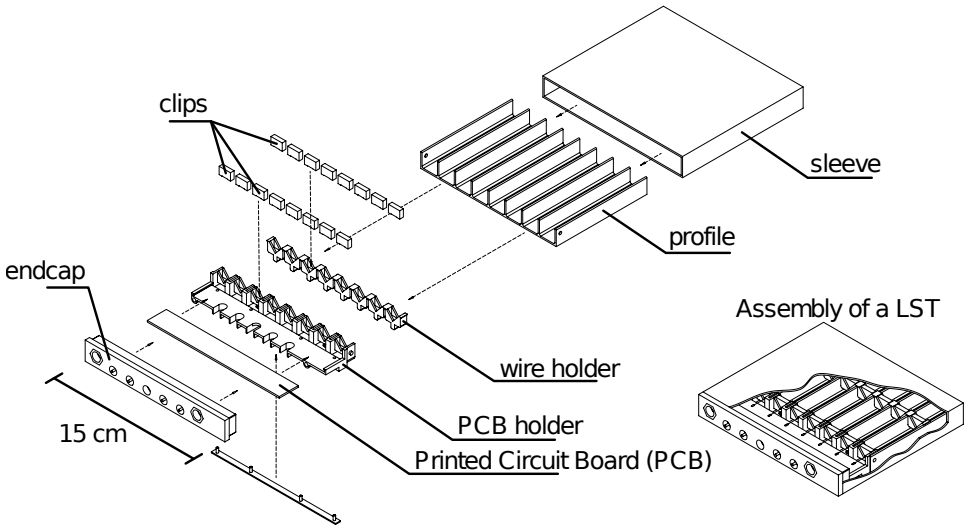


Figure 2.20: Exploded view of the LST modules used in IFR upgrades.

projected PEP-II background rates.

2.4.1 Level 1 Trigger

The hardware trigger, Level 1 (L1), has an output rate of roughly $1kHz$ during normal operation. The latency between e^+e^- collision and L1 trigger is fixed in a window of $11 - 12\mu m$. The L1 hardware trigger is based on high-transverse-momentum DCH tracks, EMC showers and IFR tracks (primarily for diagnostic purposes), which all output summary data on particle position and energy called primitives. The EMC primitives come from a sectioning of the crystals into 40 strips in the ϕ and 6 strips in the θ dimensions. DCH primitives are derived from track transverse momentum, z position and number of layers crossed.

A centralized trigger engine evaluates eight primitives and sends a L1 accept signal to begin readout of electronics buffers if the correct conditions for the global L1 trigger logic are met. The various subcomponent buffers are designed to store a backlog of multiple events to accomodate temporary high L1 trigger rates. Upon an L1 accept signal the full detector readouts are sent to the L3 trigger.

2.4.2 Level 3 Trigger

Of the roughly $1kHz$ output rate of the L1 trigger, the L3 passes about 90Hz to disk. This software trigger relies on partial event reconstruction, primarily to discriminate against physics events that are not interesting. The L3 trigger is again based on DCH and EMC information.

For the L3 DCH trigger, one track with $p_T > 600\,\text{MeV}$ or two loose tracks are required. Tracks are reconstructed with a fast algorithm to extrapolate to the interaction region. The single track must pass within 1.0cm of the IP in the transverse plane and within $7.0cm$ of the IP on the z-axis. The loose tracks are required to satisfy $p_T > 250\,\text{MeV}$ and vertex constraints of 1.5cm and 10cm for transverse and longitudinal directions, respectively.

The L3 EMC trigger evaluates clusters with energy over 20 MeV. A loose timing selection is also imposed on the clusters, as described in Appendix C. Either two EMC clusters with $E_{CM} > 350\,\text{MeV}$ or four total clusters are required with an even mass greater than 1.5 GeV. No association between tracks and clusters are made in the trigger logic.

L3 is highly efficient for the high-multiplicity hadronic events that typify CP-violation physics analysis, with typical efficiencies over 0.999.

2.5 Datasets

Table **??** summarizes the data accumulated by *BABAR* during its run. For the analysis presented herein the entire $\Upsilon(2S)$ and $\Upsilon(3S)$ collections are used; the brief analysis contained in Appendix B uses $\sim 233\text{fb}^{-1}$ of $\Upsilon(4S)$ data.

Resonance	$\mathscr{L}_{\rm on}(\mathrm{fb}^{-1})$	$\mathscr{L}_{\rm off}(\mathrm{fb}^{-1})$
$\Upsilon(4S)$	$424.18 \pm 0.04 \pm 1.82$ (0.43)	$43.92 \pm 0.01 \pm 0.19$ (0.43)
$\Upsilon(3S)$	$27.96 \pm 0.03 \pm 0.16$ (0.58)	$2.623 \pm 0.008 \pm 0.017$ (0.72)
$\Upsilon(2S)$	$13.60 \pm 0.02 \pm 0.09$ (0.68)	$1.419 \pm 0.006 \pm 0.011$ (0.88)

Table 2.1: Integrated luminosities for on- and off-resonance data samples recorded at *BABAR*. The full $\Upsilon(2S)$ and $\Upsilon(3S)$ datasets are used in this analysis. The first uncertainty is statistical, the second systematic, and the relative uncertainty in percent is given in the parentheses.

Chapter 3

$2S \to 1P \to 1S$

Analysis of the $2S \to 1P \to 1S$ cascade (Fig. **??**) is pursued first and used as a template for the very-similar $3S \to 2P \to 2S$ and $3S \to 2P \to 1S$ cascades. Consequently most of the significant details of reconstruction, backgrounds, selections, fitting and systematics for all three two-photon analyses are found exclusively in this chapter.

The singular focus of this analysis is to fit the $\gamma_{2S\to1P}$ spectrum and extract yields from the three peaks corresponding to cascades through the $\chi_{bJ}(1P)$ states with $J = 0, 1, 2$.

3.1 $2S \to 1P \to 1S$ event selection and reconstruction

it Appropriate events include $\gamma\gamma\ell^+\ell^-$ final state particle candidates with certain requirements. For each event, candidate decay chains are reconstructed in the signal cascade hypotheses based on the detected final state particles. The statistical parsimony[1] of each event under this hypothesis is used to select the best candidate and

[1] In ordinary English *parsimony* refers to a measure of the simplicity of an explaination; implicitly, the simplest explaination is considered the most correct. In the context of this document, explainations take the form of interpretations of statistical correspondences between a physical cascade hypotheses and observational signatures. Specifically, if the correspondence is poor, the simplest *statistical* interpretation – and the one considered correct – is that the physical model is incorrect, rather than the result being due to statistical variation of quantities.

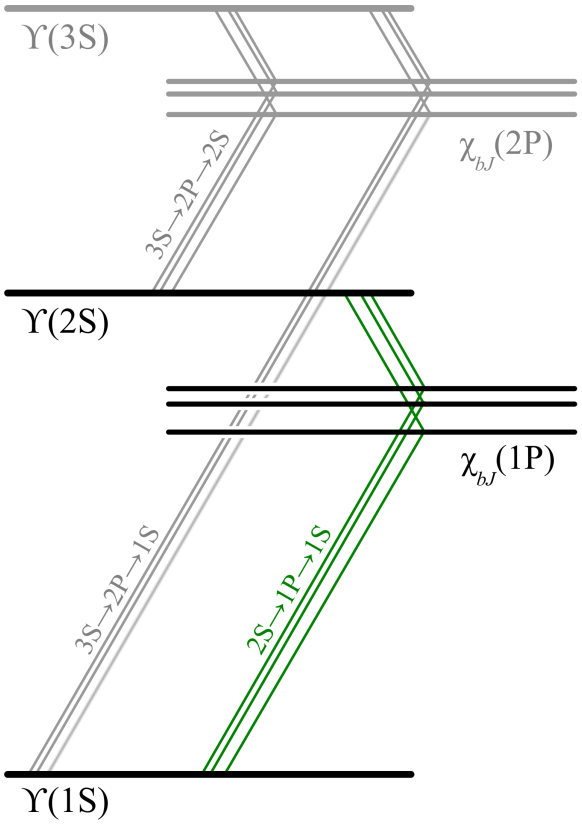


Figure 3.1: The decay mode evaluated in this chapter.

accept or reject the entire event. The calculation of the parsimony metric and selection of the candidate final state particles is discussed in this section.

3.1.1 Final state particle selection

Cascade reconstruction requires identification of the four final state particles $\gamma\gamma\ell^+\ell^-$. Leptonic universality implies that the leptonic decay rates are independent of lepton flavor with branching fractions around 2.5%. However, the e^+e^- mode suffers higher backgrounds (from Bhabha scattering) and substantially attenuated signal due to an online Bhabha veto. The inclusion of the e^-e^- mode promises little improvement in the measurement precision and thus the $\mu\mu$ mode is used for all further cascades in this analysis.

The *BABAR* standard list `muBDTLoose` provides μ candidates. This list is based on

a decision tree algorithm which serves as a muon identification selector; the details of this algorithm are beyond the scope of this documentbut are described thoroughly in Ref. [55]. Muon selection efficiency is tuned to 80% while mis-ID rate is enhanced compared to cut-based lists with the same muon efficiency.

Photon candidates come from the standard list `GoodPhotonLoose`, consisting of single calorimeter bumps not matched with any track with a minimum lab energy of 30 MeV and with a zero-mass hypothesis. The lateral moment of the associated cluster is required to be below 0.8. Photon energies in this document are center-of-mass energies unless stated otherwise.

3.1.2 Cascade reconstruction

Two μ candidates are combined to create an $\Upsilon(1S)$ candidate. Two photon candidates are added to reconstruct the $\Upsilon(2S)$. The number of candidate cascades identified in each event depends on the number of candidate final state particles; the count is often larger than one. The multiplicity is combinatoric in photon candidate number.

At this level of event selection no requirements are placed on final state particles besides those required for particle ID; this list includes all sets of $\mu^+\mu^-\gamma\gamma$ candidates in all events observed while running at the $\Upsilon(2S)$ resonance.

3.1.3 Cascade fitting

The two photon candidates span the energy gap between two very-well-known narrow resonances, the $\Upsilon(3S)$ and $\Upsilon(2S)$. These two resonances have natural widths well below the detector resolution, allowing the constraint of the of the two Υ masses to improve photon resolution. The $\pi^0\pi^0$ background needs four photons to span the energy gap; this constraint also improves discrimination between signal and $\pi^0\pi^0$ processes based on kinematic parsimony. Fixing the location and energy of the interaction point further constrains the fits.

Cascade fitting is accomplished using *BABAR*'s `TreeFitter`. This fitting algorithm utilizes a χ^2 fitting method to adapt a set of parameters to a set of data points. The actual algorithm uses a Kalman filter and is beyond the scope of this document. A benefit of this fitting method is that it can treat the whole cascade simultaneously,

SP mode #	Mode	ϵ_{reco}
9244	$\chi_{b0}(1P)$ cascade	0.210 ± 0.002
9245	$\chi_{b1}(1P)$ cascade	0.2236 ± 0.0006
9246	$\chi_{b2}(1P)$ cascade	0.2198 ± 0.0007

Table 3.1: Reconstruction efficiencies for the three signal MC modes. Errors are solely statistical; a large unknown systematic error is hidden.

maximizing and measuring the parsimoniousness of the entire process. The kinematic fit returns a fit probability (derived from the χ^2 value and number of degrees of freedom in the fit) which is a powerful fit parsimony metric given the constraints imposed and the kinematic uniqueness of the signal mode.

3.1.4 Best candidate selection

The "winning" cascade from each event is chosen based on the kinematic fit probability. The cascade candidate with the highest fit probability under the signal hypothesis is chosen. Events with no signal fit probability above zero are discarded. $E_{\gamma_{1P\to 1S}}$ is required to be in the range $300 - 480\,\mathrm{MeV}$.

3.1.5 MC reconstruction efficiencies

Table **??** shows reconstruction efficiencies in signal MC. The statistically significant J-dependent biases are discussed in more detail in Section **??**.

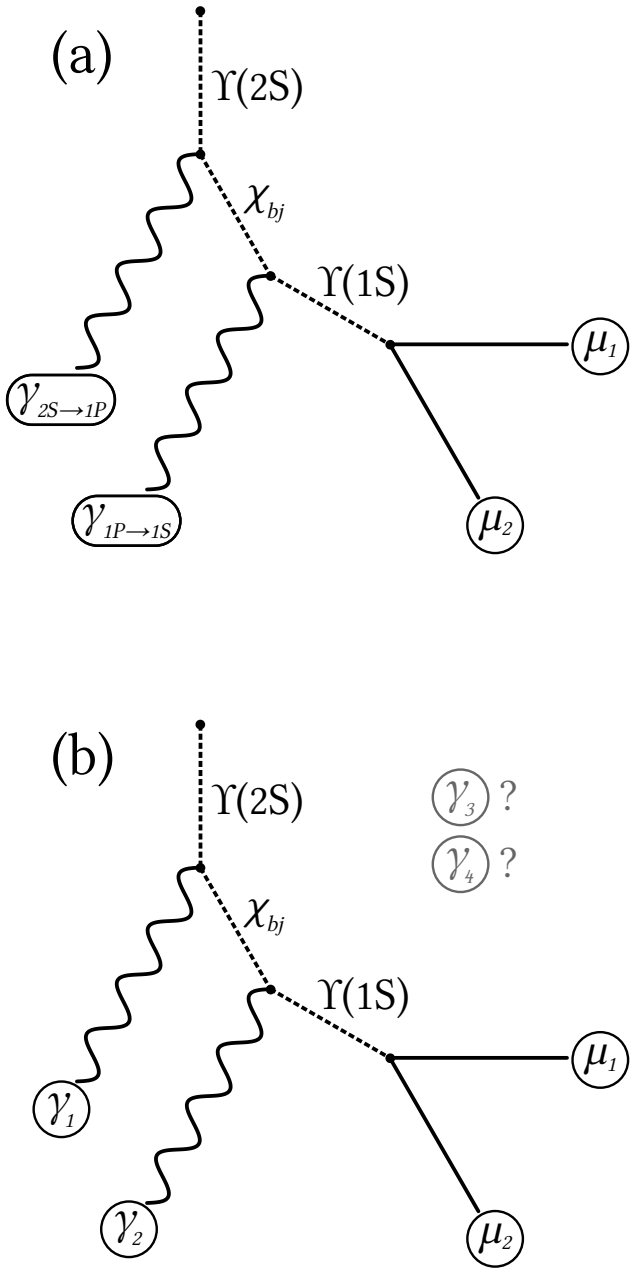


Figure 3.2: Event diagrams for signal (a) and $\pi^0\pi^0$ (b) cascades in the signal reconstruction hypotheses. Final state particles are represented by solid lines, reconstructed composite particles by dashed lines and non-reconstructed intermediate states by dotted lines. The final state particle labels in circles are those reconstructed by the given cascade hypothesis in the given process in a typical case [sources for the candidate photons in practice will vary]. Particles disconnected from the diagram and grayed represent final state physics particles not reconstructed. The signal cascade has final state particles $\gamma_{2S\to1P}$, $\gamma_{1P\to1S}$, μ_1 and μ_2 while the $\pi^0\pi^0$ cascade includes γ_1, γ_2, γ_3, γ_4, μ_1 and μ_2. The kinematic parsimony of the fit to the background $\pi^0\pi^0$ cascade is clearly very poor.

3.2 $2S \to 1P \to 1S$ background

The choice of performing analysis in the exclusive mode is designed to provide clean event selection. The $\chi_{b0}(1P)$ peak in the $\gamma_{2S\to1P}$ spectrum is expected to be very small, therefore background suppression is still important. Discussion of the relevant background modes and mitigation is presented in this section.

3.2.1 Backgrounds in MC

Candidate cascades must have the correct final state particles with a plausible kinematic match to survive initial selection. However, the ubiquity of low-energy photons in the EMC – mostly "beam" photons originating in interactions between the beam and stray gases in the beampipe or the beampipe itself – provides ample opportunities for fake signal in the $\gamma_{2S\to1P}$ acceptance window. Beam photons decrease in abundance to insignificant levels at energies above roughly 200 MeV, meaning that a physics source for the $\gamma_{2P\to1S}$ candidate is required as well as a source for the two μ candidates.

A description of the likely background sources is provided here.

Background source: $\mu\mu(\gamma)$

SP mode 3981 models the direct μ pair production with associated ISR and FSR photons at the $\Upsilon(2S)$ resonance. A collection of $16.7M$ simulated $\mu\mu(\gamma)$ events is available, slightly larger than the expected number in Run 7.

Fig. **??** shows the $\mu\mu$ invariant mass plot for all μ pairs in the $\mu\mu(\gamma)$ MC sample of $e^+e^- \to \mu^+\mu^-$ at the $\Upsilon(2S)$ resonance. The long tail extending to the left of the main peak at the $\Upsilon(2S)$ mass is due to initial- or final-state radiation, a spectrum of which is shown in Fig. **??** along with beam photons from the same sample. The probability of a "stray" photon in μ production processes with energies in the $\gamma_{1P\to1S}$ window is small, though the high cross-section for this process and abundance of beam photons in the $\gamma_{2P\to1S}$ window guarantees a significant number of these events will fake the correct kinematics and be mistaken for the signal cascade.

In addition to these $\mu\mu(\gamma)$ events, direct decay of produced $\Upsilon(2S)$'s to a μ pair will appear in generic $\Upsilon(2S)$ decay MC. This background should be indistinguishable

from the $\mu\mu(\gamma)$ source described.

Background source: $\pi^0\pi^0$

SP mode 9016 simulates generic $\Upsilon(2S)$ decays with roughly three times the total number generated during Run 7. In these generic decays, the process $\Upsilon(2S) \to \pi^0\pi^0\Upsilon(1S)$; $\Upsilon(1S) \to \mu\mu$ is expected to be the main background mode. In this case μ selection is indistinguishable from the signal cascade and there are four final-state photons associated with the physics event. Fig. **??** and Fig. **??** show the source of photon candidates for $\pi^0\pi^0$ in Generic MC, similar to Fig. **??** except in this case the events are required to reconstruct as the signal cascade in order to show the correct source of the reconstructed signal photons. In this case the presence of a photon in the $\gamma_{1P\to1S}$ acceptance window – -which is almost always a daughter of one of the π^0's – -makes it unlikely that another π^0 daughter will satisfy the kinematic requirements and land within the $\gamma_{2S\to1P}$ window.

The main implication of the presented plots is that the $\gamma_{1P\to1S}$ background is largely composed of beam photons regardless of the background process. This observation has significant implications for background rejection.

3.2.2 Representative MC ensemble

SP modes 9244, 9245 and 9246 provide simulated decay chains for the signal cascades through χ_{b0}, χ_{b1} and χ_{b2}. When combined correctly with the $\Upsilon(2S)$ generic and $\mu\mu(\gamma)$ MC modes a model $\gamma_{2S\to1P}$ spectrum can be constructed which is useful for tuning selections. The scales for each mode are taken from the PDG [53] and a calculation of $e^+e^- \to \mu^+\mu^-$ cross-sections at *BABAR* [30] and are shown in Table **??**. This group of scaled representative MC samples is referred to as the "MC ensemble" throughout the remainder of this document.

Using the scales presented in Table **??**, the $\gamma_{2S\to1P}$ and $\gamma_{1P\to1S}$ spectra are plotted using the MC ensemble in Fig. **??**.

SP #	Mode	# MC	# Run 7	Scale	# reconstructed	Scaled #
9016	Generic $\Upsilon(2S)$ decays:	$332M$	$98.6M$	0.30	$4,941$	$1.5K$
	$\Upsilon(2S) \to \pi^0\pi^0\Upsilon(1S)$		$0.2M$		$2,290$	$0.69K$
	$\Upsilon(2S) \to \mu\mu$		$1.9M$		$2,603$	$0.78K$
	All other generic		$96.5M$		Negligible	
9244	Full χ_{b0} cascade	$87,000$	$1.2K$	0.014	$18,111$	$0.25K$
9245	Full χ_{b1} cascade	$709,000$	$57K$	0.080	$146,586$	$12K$
9246	Full χ_{b2} cascade	$493,000$	$33K$	0.068	$107,213$	$7.3K$
3981	$\mu\mu(\gamma)$	$16.7M$	$15.5M$	0.93	$9,649$	$9.0K$
998	uds				Negligible	
1005	$c\bar{c}$				Negligible	
3429	$\tau\tau$				Negligible	

Table 3.2: A summary of the relevant MC modes used to construct the MC ensemble. Column 3 shows the number of events included in the MC collection. Column 4 contains the number expected in Run 7 based on expected decay rates and cross-sections given *BABAR*'s recorded luminosity at the $\Upsilon(2S)$ resonance using branching fractions from the PDG [53]. Column 5 shows the factor required to scale each MC sample to the size expected in data. Column 6 contains the number of events from the scaled MC sources that reconstruct as the signal and pass the initial fit probability and selection requirements. The final column shows the number of events in the scaled MC sample that constitutes the representative MC ensemble.

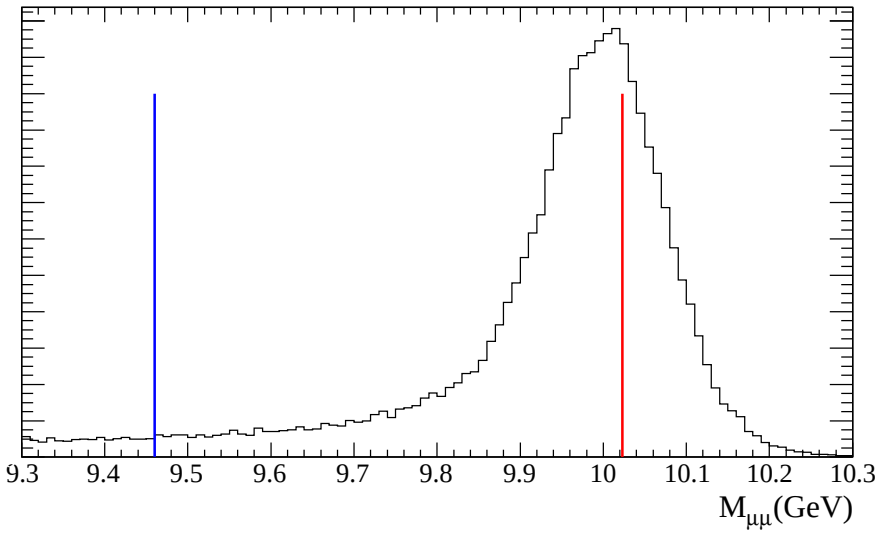


Figure 3.3: Di-μ invariant mass in a subset of the $\mu\mu(\gamma)$ MC. The mass of the $\Upsilon(2S)$ is marked with a red line and the $\Upsilon(1S)$ with blue. The long tail extending well below the nominally-produced $\Upsilon(2S)$ peak is due primarily to initial- and final-state radiation losses. Entries near the $\Upsilon(1S)$ mass are responsible for faked $\Upsilon(1S) \to \mu\mu$ decays.

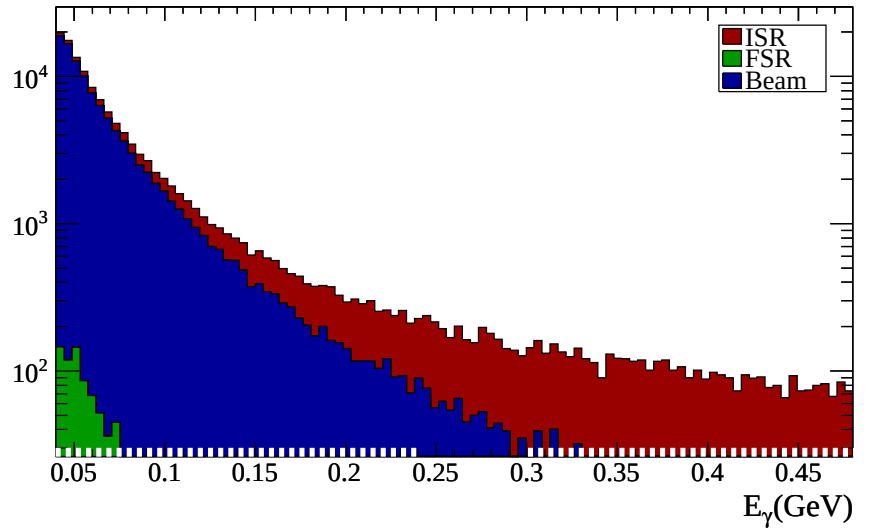


Figure 3.4: Photon energy and origin for all photon candidates seen in the same sample of MC $\mu\mu(\gamma)$ events as in Fig. **??**. The hatched regions correspond to the acceptance windows for the $\gamma_{2S\to1P}$ and $\gamma_{1P\to1S}$ candidates. It is evident that $\gamma_{2S\to1P}$ candidates for selected $\mu\mu(\gamma)$ events come mostly from ubiquitous beam photons and $\gamma_{1P\to1S}$ candidates mostly come from ISR.

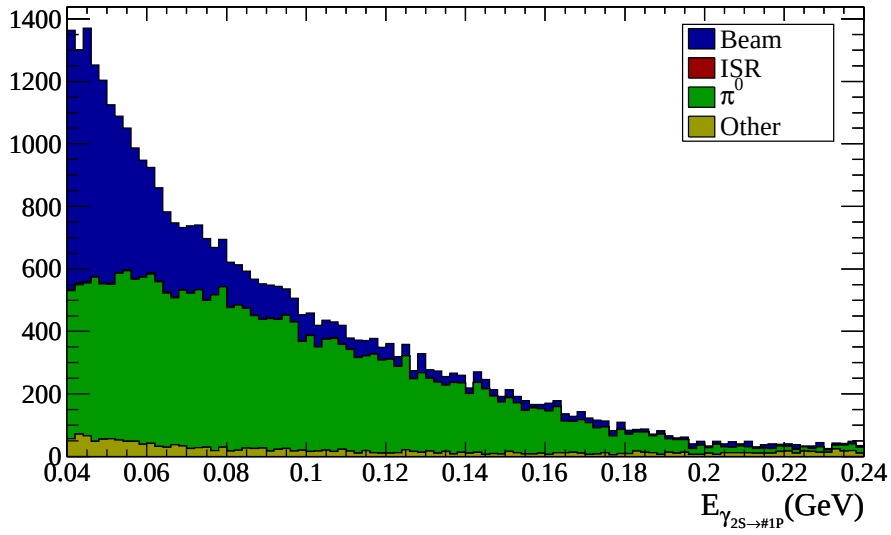


Figure 3.5: MC mothers for $\gamma_{2S\to1P}$ candidates in reconstructed Generic $\Upsilon(2S)$ decays with initial selections.

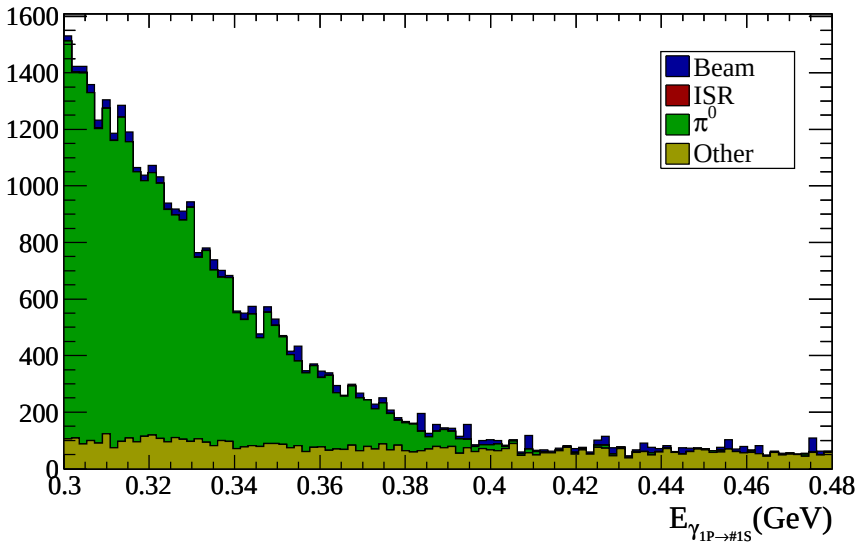


Figure 3.6: MC mothers for $\gamma_{1P\to1S}$ candidates in reconstructed Generic $\Upsilon(2S)$ decays with initial selections.

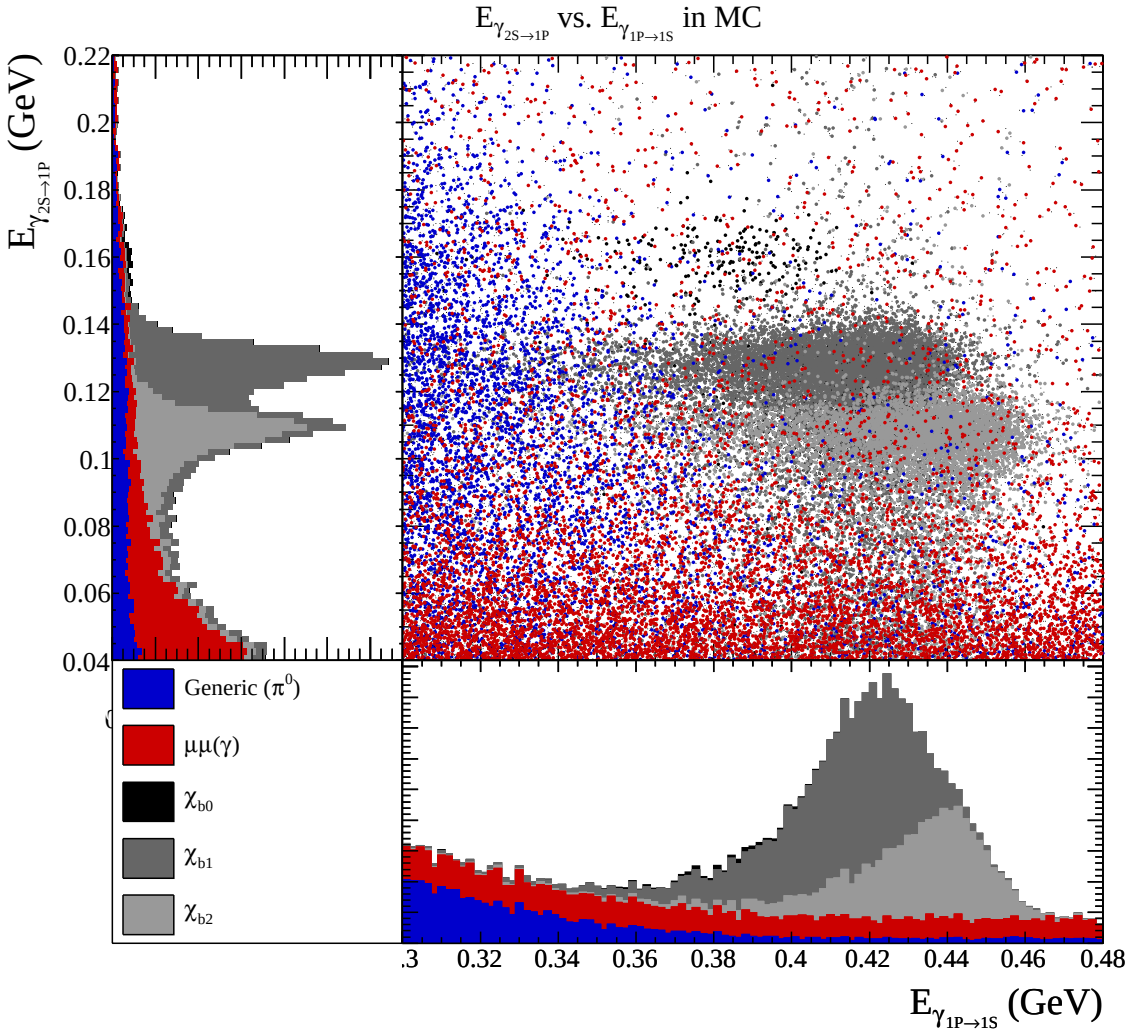


Figure 3.7: Scatterplot of $E_{\gamma_{2S\to1P}}$ vs $E_{\gamma_{1P\to1S}}$ in ensemble MC, showing the distribution of the main background sources. Very few background events fake both photon energies simultaneously.

3.3 $2S \to 1P \to 1S$ cuts

With low backgrounds in the exclusive mode it's quite likely that the measurement error will be dominated not by background effects but by fitting difficulties arising from overlapping signal peaks. Thus the most useful cut performance metric is fit significance using the MC ensemble.

A particular difficulty arises from the presence of the long tails on the signal peaks as shown in Fig. **??**. Around 8% of the photons incident on the EMC start showering before entering the detector, leaving unrecovered energy outside of the calorimeter. An artificially decreased cluster energy is recorded for those photon candidates. When the signal and background contributions are fit to extract total yields, there is a very real danger that the tail and background contributions could be conflated. Minimizing the background contribution and tightening the tails are the topics of this section.

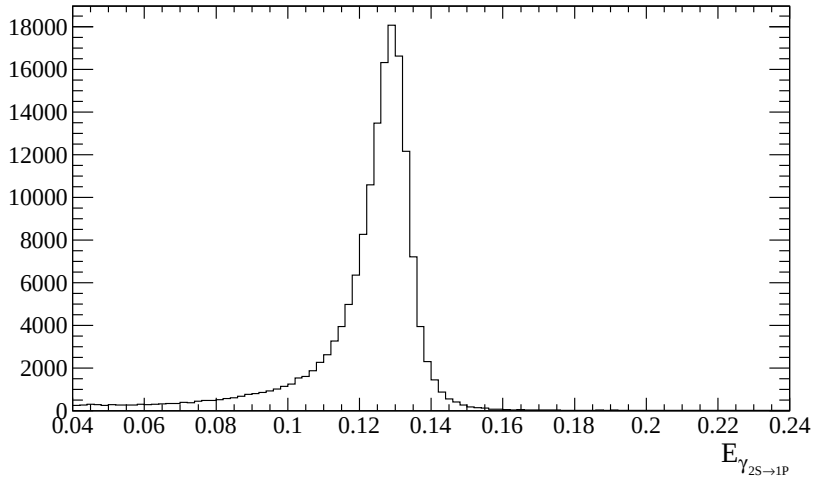


Figure 3.8: Signal MC $\gamma_{2S\to1P}$ line for $J = 1$ using initial selections. The large low-energy tail is due to unrecovered shower energy outside of the EMC.

3.3.1 Signal fit probability cut

As described in Section **??**, the cascade fitting process returns an overall fit probability for the entire cascade as a kinematic parsimony metric. This value is a particularly effective background discriminant given the fixed masses of the $\Upsilon(2S)$ and $\Upsilon(1S)$.

Using the MC ensemble, Fig. **??** shows the distribution in signal fit probability for the main candidate sources. Whereas the background sources are fairly uniform in fit probability – consistent with the idea that they achieve high fit probability only by "accident" – the signal fit probability peaks strongly at high fit probabilities. A cut on the signal fit probability is therefore useful as a means of rejecting background with minimal impact on the signal.

Fit probability and tails

In addition to background subtraction, other benefits arise from a cut on the signal fit probability. Fig. **??** shows the relationship between signal fit probability and candidate CM energy for all reconstructed $\gamma_{2S\to1P}$ candidates in the χ_{b1} MC sample. The fit probability disfavors events where $E_{\gamma_{2S\to1P}}$ and $E_{\gamma_{1P\to1S}}$ combined are not large enough to account for the energy spacing between the $\Upsilon(2S)$ and $\Upsilon(1S)$ states. The tails represent unrecovered energy, meaning that the further a candidate's reconstructed energy is from the nominal signal peak energy the lower its fit probability will be. Accordingly, a cut on the fit probability will also remove tail events from the signal peak disproportionately to the other signal events. Removing signal events in the tail may increase signal significance by removing the ambiguity as to the source of the events in the low-energy region of the spectrum as well as minimize the risk of mis-counting tail events that "leak" outside the signal window. Additionally, for the critical χ_{b0} peak, tightening the tail will decrease the error in relative peak yields caused by the χ_{b1} peak picking up tail entries from the χ_{b0}.

For these reasons, the optimal fit probability cut is likely to more aggressive than that suggested by simply maximizing the signal-to-background ratio.

3.3.2 Timing selection

As a subset of this analysis, an offline correction of the EMC timing is presented in Appendix C. See the appendix also for a detailed description of the timing significance variable S and other timing-related topics. In this case the timing difference significance S is defined between the two signal photons:

$$S = \frac{|t_{2S\to 1P} - t_{1P\to 1S}|}{\sqrt{\sigma^2_{t_{2S\to 1P}} + \sigma^2_{t_{1P\to 1S}}}}. \tag{3.1}$$

Regardless of the background source, the $\gamma_{2S\to 1P}$ candidate is usually out-of-time and the $\gamma_{1P\to 1S}$ candidate is almost always in-time. A cut on the maximum value of S, called here S_{max}, selects events where the two photons arrive nearly simultaneously. This method avoids the uncertainty of comparing both photon times with a global reference time. In principle the two photon times could be compared to the cluster times for the muons but the large clusters of the in-time $\gamma_{1P\to 1S}$ candidates provide better timing determination.

The fit applied to the signal cascade makes no use of timing information. The fit probability cut has no power to discriminate against photon candidates that just happen to have the right energy to make a "good" kinematic fit. The fact that this isn't particularly rare is a testament to the huge number of beam photons detected at low energies. The timing significance cut, on the other hand, is designed specifically to discriminate against these out-of-time candidates but has no power to reject candidates based on kinematics. For these reasons, the timing cut and signal fit probability cuts are complementary and combine for powerful low-energy photon background rejection. Fig. **??** illustrates the orthogonality and complementarity of these two cuts.

3.3.3 Cut choices

Ideally the choice of cuts would be tailored to maximize the significance of the χ_{b0} peak. However, the peak significance is limited largely by line overlap which is not modeled well in MC. Significance is also limited by the continuum background in the region of the peak, but the relative sizes of signal and background in this region are not known a priori. Neither of these effects depend strongly on the exact choices of timing and fit probability cuts, so values are chosen based on scrutiny of Fig. **??**. A purer selection could have been achieved with a combination of the two variables [to form an arc on the plot] but without reliable MC timing information the two variables need to remain uncorrelated.

3.3.4 Cut efficiencies

To measure any J-dependent efficiency biases we evaluate the number of events passing the various selections in signal MC. The actual quantities of interest in the product branching fraction ratio measurements are the efficiency ratios, seen in Table **??**. A slight but statistically significant bias is seen in the reconstruction efficiency ratio, with the $J = 1$ cascade reconstructing with a higher efficiency than the $J = 0$ and $J = 2$ cascades.

The most likely explanation for a J-dependent bias in detection or reconstruction efficiencies is due to different angular distributions of the final state particles [detector acceptance is not uniform]. The angular distributions in this type of decay have been calculated in model-independent formulations [34] and helicity amplitudes from these calculates are explicitly used in event generation using the `EvtGen` package. The MC detection and reconstruction efficiencies thus are expected to be largely correct.

The efficiency ratios for all signal fit probability cuts are very consistent; these biases are presumably due to the fitting algorithm assigning improper widths to the three peaks. The variance of the efficiency ratios across all fit probabilities tested is used as the error on this quantity. The width assumed by the fitting algorithm does not necessarily match MC or data – in principle the corrections on MC will not be the same as those on data. However, this effect is tiny compared to the statistical uncertainty in the reconstruction efficiencies.

Table **??** shows the corresponding efficiencies on the background MC samples. Fig. **??** shows the signal spectrum with these cuts as well as the data spectrum, confirming that additional selection criteria will result in minimal gains in signal significance.

Selection	$\epsilon_{J=1}/\epsilon_{J=0}$	$\epsilon_{J=1}/\epsilon_{J=2}$	$\epsilon_{J=2}/\epsilon_{J=0}$
Reco. and initial selections	1.0637 ± 0.009	1.0170 ± 0.004	1.046 ± 0.009
Signal fit probability	0.9999 ± 0.0007	0.9961 ± 0.0007	1.002 ± 0.001
Total	1.0621 ± 0.0092	1.0131 ± 0.0045	1.0485 ± 0.0093

Table 3.3: Efficiency ratios for the three signal modes from MC. The timing cut cannot discriminate between two physics events, therefore it cannot contribute systematic error to the efficiency ratio. Efficiency ratios were computed for a wide range of fit probability cuts with very little variation; the results here are general for all selections on this variable and the variance is folded into the total ratio error, which is dominated by statistics. For a discussion of the source of the line-dependent efficiencies see the text.

Selection	$\epsilon_{Generic}$	$\epsilon_{\mu\mu(\gamma)}$
Reco. and initial selections	2.8×10^{-5}	7.0×10^{-4}
Signal fit prob. $> 10^{-5}$	0.099	0.172
$S < 2.0$	< 0.92	< 0.92
Cuts	< 0.09	< 0.16
Cuts + reconstruction	$< 10^{-6}$	$< 10^{-4}$

Table 3.4: Efficiencies for the main background sources for reconstruction and cuts. Efficiencies are taken from MC with the exception of the timing cut. The timing cut efficiency was taken from Appendix C, with the expectation that the non-signal modes will have substantially lower efficiencies than the in-time efficiency of 0.92 quoted: these modes will have some admixture of in- and out-of-time photon candidates with some averaged efficiency somewhere between the in- and out-of-time efficiencies.

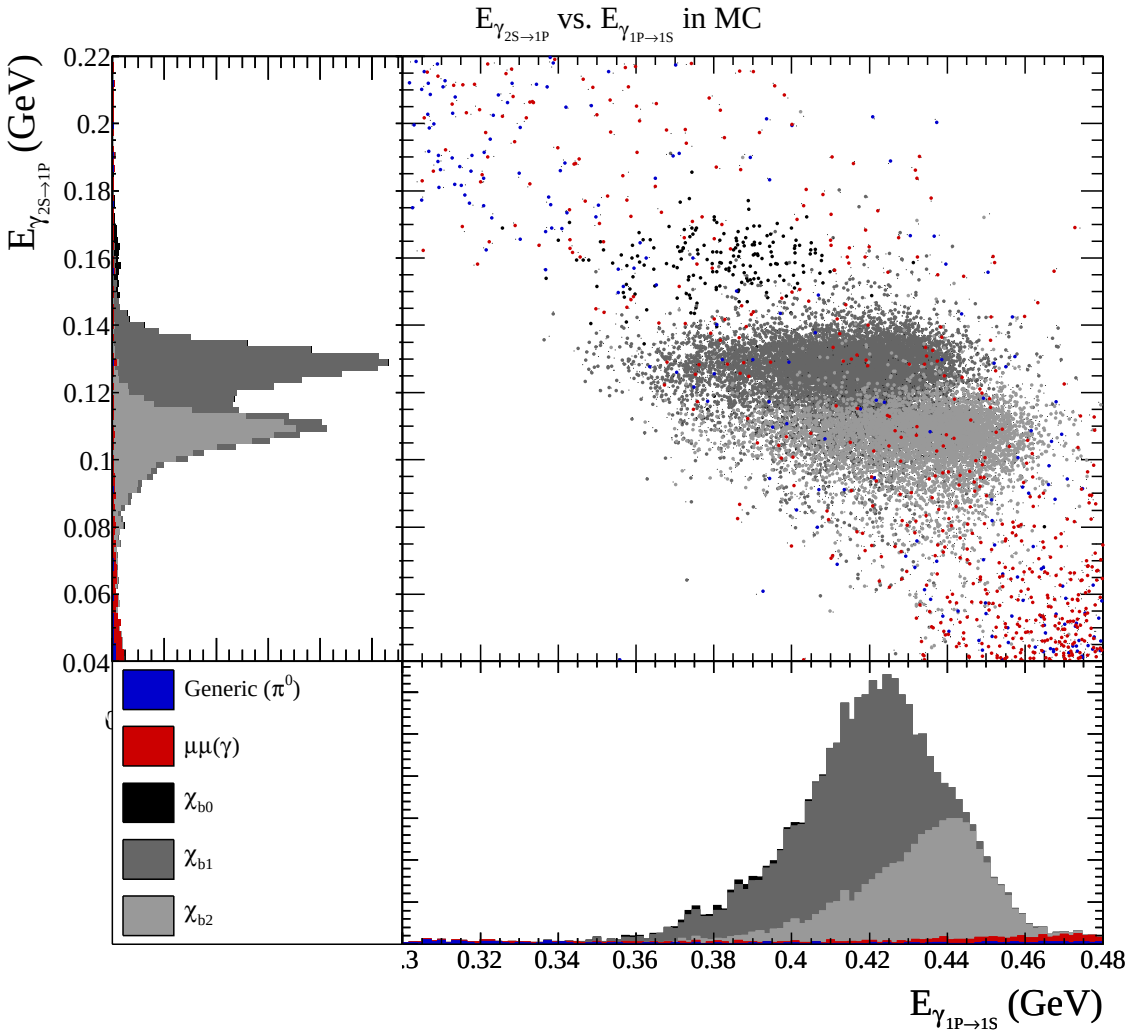


Figure 3.9: Scatterplot of $E_{\gamma_{2S\to1P}}$ vs. $E_{\gamma_{1P\to1S}}$ in ensemble MC only showing events with a signal cascade fit probability greater than 10^{-5}. Compare to Fig. **??**, with only probability 0 events excluded. The selection is most adept at removing entries where $E_{2S\to1P} + E_{1P\to1S}$ departs substantially from $M_{\Upsilon(2S)} - M_{\Upsilon(1S)}$. A line drawn through the three signal peaks in the scatter plot defines the equivalency of these two values, and the bulk of the remaining background after the fit probability selection is "accidentally" near this line.

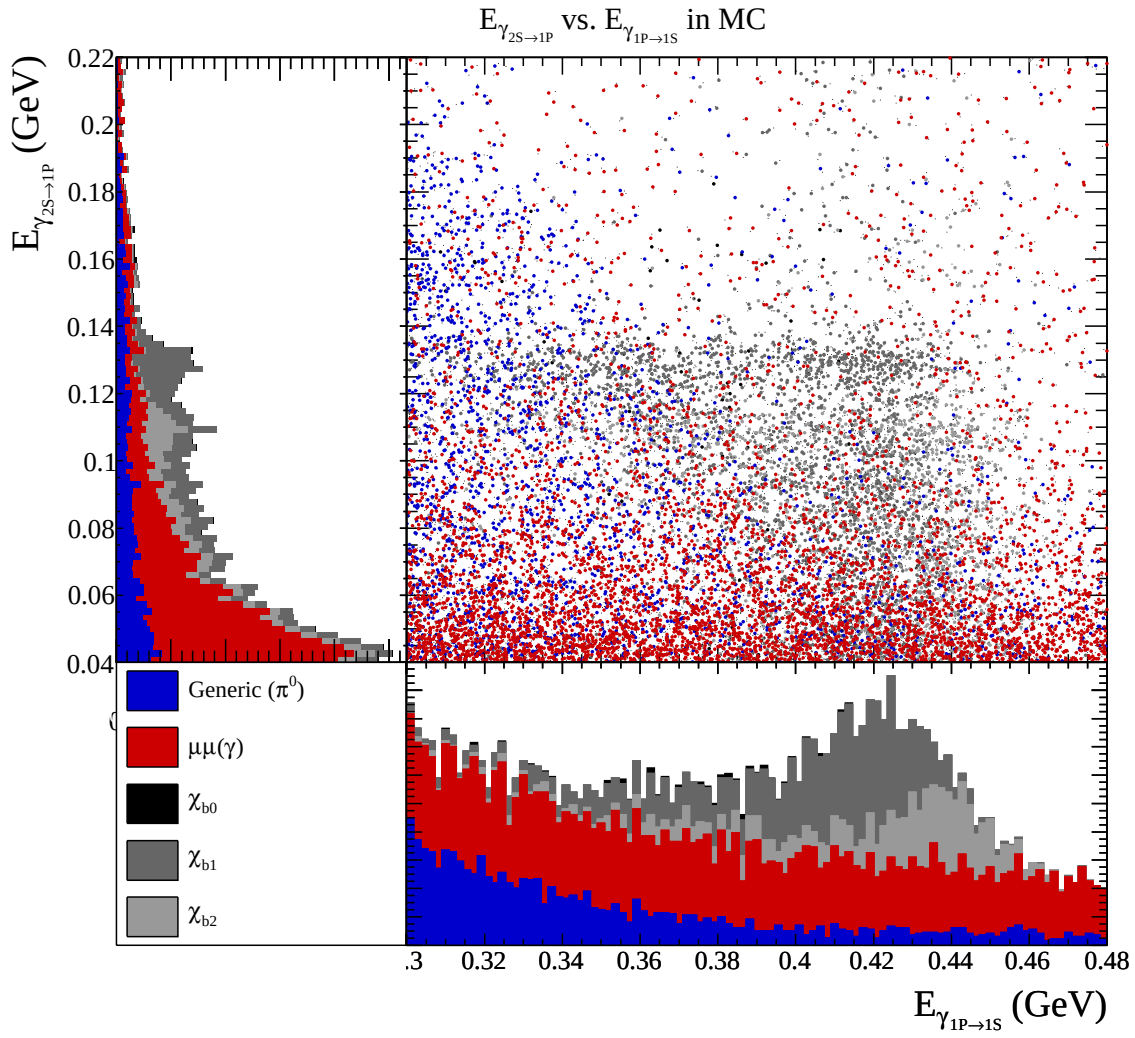


Figure 3.10: Difference between Fig. **??** and Fig. **??**, showing only entries with a signal cascade fit probability less than 10^{-5}. The removed entries in the signal modes are almost entirely tail.

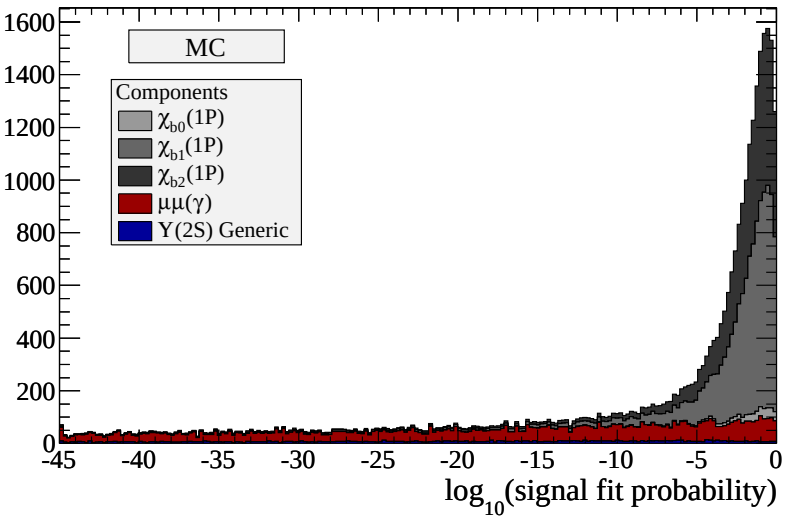


Figure 3.11: Histogram of the log signal fit probability for the MC ensemble.

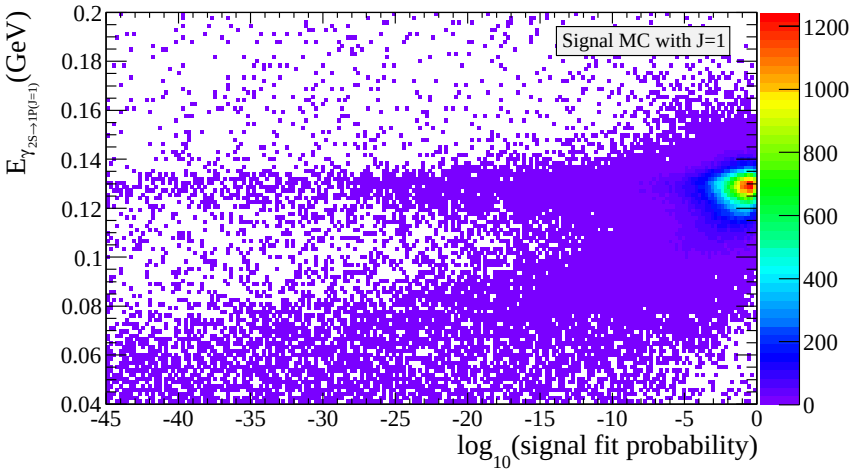


Figure 3.12: A two-dimensional histogram of the reconstructed photon energy vs. log signal fit probability for the $J = 1$ signal MC. The large "wing" extending diagonally down from the main peak constitutes the tail of the energy spectrum.

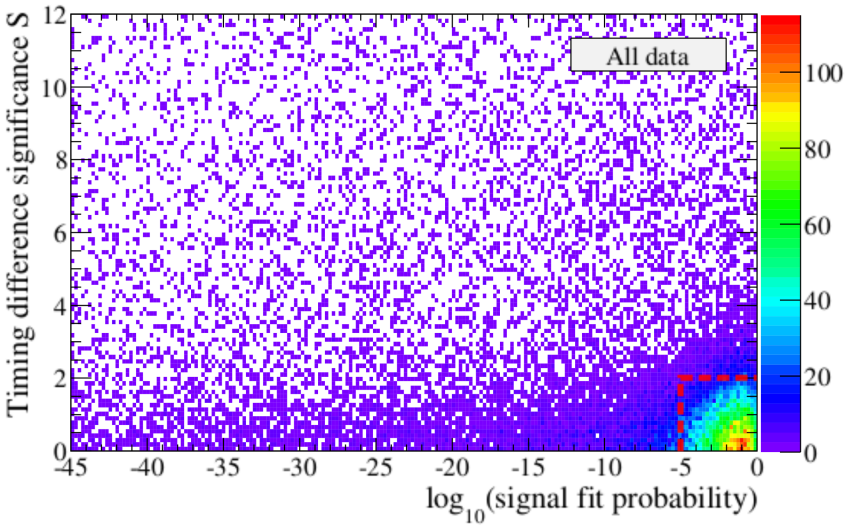


Figure 3.13: A two-dimensional histogram of the timing difference significance S between $\gamma_{2S\to1P}$ and $\gamma_{1P\to1S}$ plotted against the log signal fit probability using the full data sample. The peak in the lower right-hand corner consists of in-time and high-probability events – namely, signal. The faint tail running left from this peak is presumably tail entries and residual $\pi^0\pi^0$ background. Outside of these two features the distribution is uniform, confirming that the fit probability and timing cuts are complementary. The dashed red box shows the choices for cuts on these two variables.

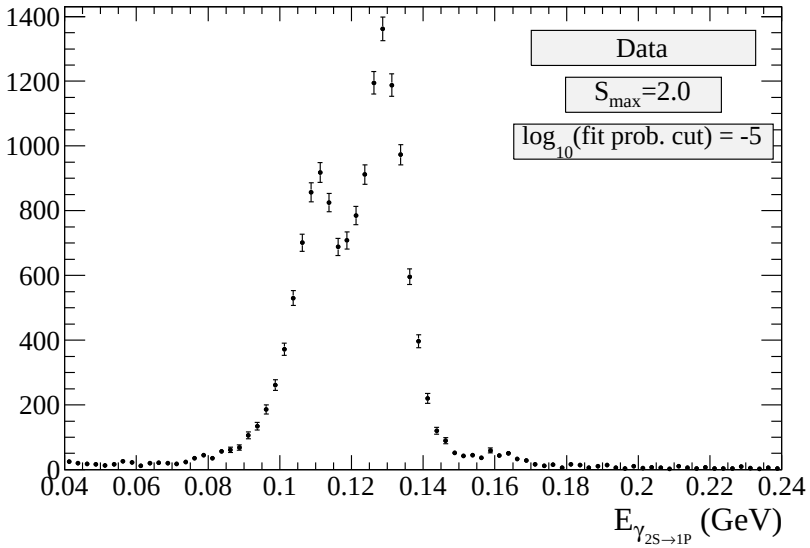


Figure 3.14: The $\gamma_{2S\to1P\to1S}$ spectrum with final selections on Run 7 data.

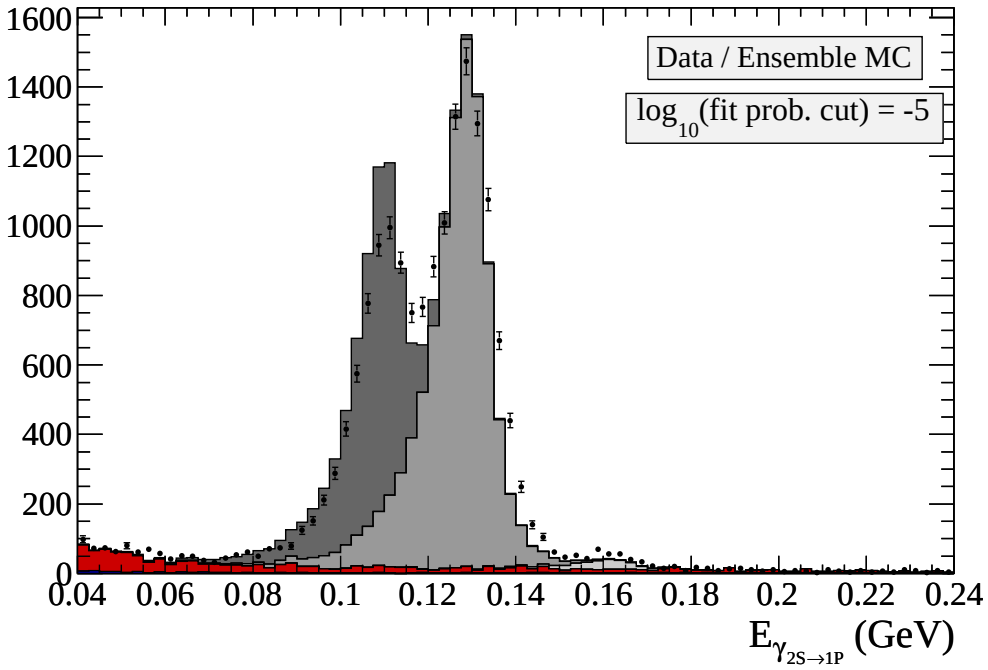


Figure 3.15: MC ensemble compared to signal spectrum. The simulated $\gamma_{2S\to1P}$ spectrum uses the scales and MC modes shown in Table **??** with the final fit probability cut but no timing selection due to the inadequacy of the MC timing information.

3.4 $2S \to 1P \to 1S$ fitting

The primary task in extracting precise yields from the spectrum lies in finding a fitting procedure that is accurate and robust. The line shape in the $\gamma_{2S\to1P}$ spectrum is not trivial and depends on the choice of signal cascade fit probability selection. For a given fit probability selection the best-fit parameters for the line shape are extracted from MC. These parameters are used to build the model for the data fit.

3.4.1 Fitting procedure

Spectrum fitting utilizes an extended unbinned maximum-likelihood method implemented in the `ROOT`-based fitting package `RooFit`. The "extended" label signifies that the total yield of the fit PDF is fixed to the spectrum integral with Poisson error. This constraint removes a degree of freedom from the fit and makes extraction of individual component yields accurate.

3.4.2 Lineshape parametrization

Monochromatic lineshapes in the EMC are determined by detector dynamics. The width of the distribution is entirely a factor of detector resolution, which is itself a function of line energy. In addition to the width, the low-energy tails contribute to a complex shape dependent on energy and reconstruction.

For calorimeter lineshapes with low-energy tails the so-called "Crystal Ball" function is commonly used as a model. This probability density function [PDF] has a power-law low-energy side and a Gaussian peak and high-energy side. Requiring the function and its first derivative to be continuous everywhere this parametrization is typically flexible enough to accommodate most lineshapes. However, at the low photon energies of this analysis and with the fairly high statistics of this spectrum a Crystal Ball parametrization of the lineshape using signal MC performs poorly; the MC lineshape shows substantial deviation from Gaussian shape on the high-energy side.

Extending the flexibility of the low-energy side of the Crystal Ball function to the high-energy side, a "Double-sided Crystal Ball" function performs well on the MC

lineshape:

$$f_{DCB}(E;\sigma,\mu,n_1,\alpha_1,n_2,\alpha_2) = N \cdot \begin{cases} A_1 \cdot (B_1 - \xi)^{-n_1}, & \text{for } \xi \leqslant -\alpha_1 \\ \exp(-\frac{\xi^2}{2}), & \text{for } \alpha_1 < \xi < \alpha_2 \\ A_2 \cdot (B_2 + \xi)^{-n_2}, & \text{for } \xi \geqslant \alpha_2 \end{cases} \tag{3.2}$$

where

$$A_i = \left(\frac{n_i}{\alpha_i}\right)^{n_i} \cdot \exp\left(-\frac{\alpha_i^2}{2}\right), \tag{3.3}$$

$$B_i = \frac{n_i}{\alpha_i} - \alpha_i, and \tag{3.4}$$

$$\xi = \frac{E-\mu}{\sigma}. \tag{3.5}$$

An example of a Double-sided Crystal Ball function with parameters labeled is shown in Fig. **??**. This is a function with a Gaussian core and width σ which transitions at points α_1 and α_2 [in units of σ] on either side into power-law curves with coefficients given by n_1 and n_2. The function and its derivative are required to be continuous everywhere, requirements which lead to the forms for A_i and B_i as defined. This model allows a large degree of flexibility in the tail regions of the distribution which is essential for minimizing the errors in the measured peak yields. N is the normalization constant which will be the main result in the fit.

3.4.3 MC lineshape

Monte Carlo attempts to model the processes that contribute to the complex line shapes in the EMC. A single line can be isolated in signal MC to refine the fit with the hope that the model developed there can be used successfully on data. The χ_{b1} line has the greatest statistics in signal MC so this peak is used to determine line shape. The line shape is significantly impacted by the choice of fit probability cut [Fig. **??**]. Fig. **??** shows the best double-sided Crystal Ball fit for the fit probability cut used in this analysis. The timing cut has no effect on line shape.

The parameters extracted from the fit to the MC line can be informative but are not expected to match the actual data lineshape perfectly, as the processes that lead to the complex lineshape are difficult to model precisely. The EMC resolution as a

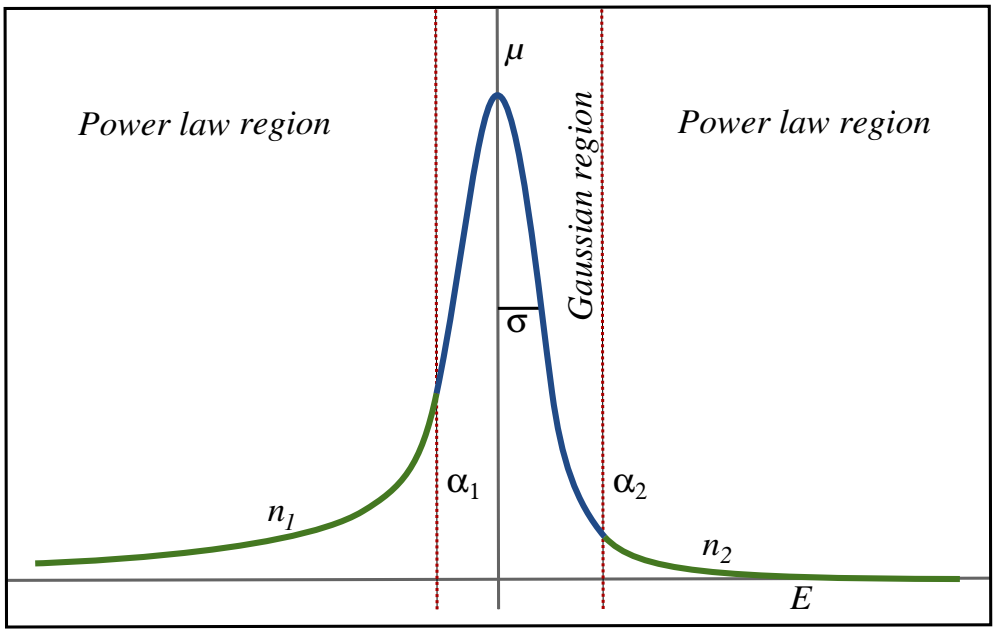


Figure 3.16: General Double-sided Crystal Ball shape with parameters.

function of line energy has been parametrized [33] as

$$\frac{\sigma_E}{E} = \frac{(2.30 \pm 0.03 \pm 0.3)\%}{\sqrt[4]{E(GeV)}} \oplus (1.35 \pm 0.08 \pm 0.2)\%. \tag{3.6}$$

The parametrization is fit to data from multiple sources as shown in Fig. **??**. There are no data points low enough to directly constrain the parametrization in the energy region of interest in this analysis, $80 - 160\,\mathrm{MeV}$. Extrapolation down to these energies cannot approach the precision required to fit line resolutions from this parametrization, though it will still be useful for constraining the slope of the resolution in photon energy.

An alternative MC ensemble is generated with "corrected" resolution. This correction convolves a Gaussian with the MC resolution to better match MC with data at higher energies. The correction in this energy region ends up giving widths that are as much too wide as the original uncorrected MC widths are too narrow. However, this corrected MC sample is useful for testing the generality of the fitting procedure over a range of resolutions.

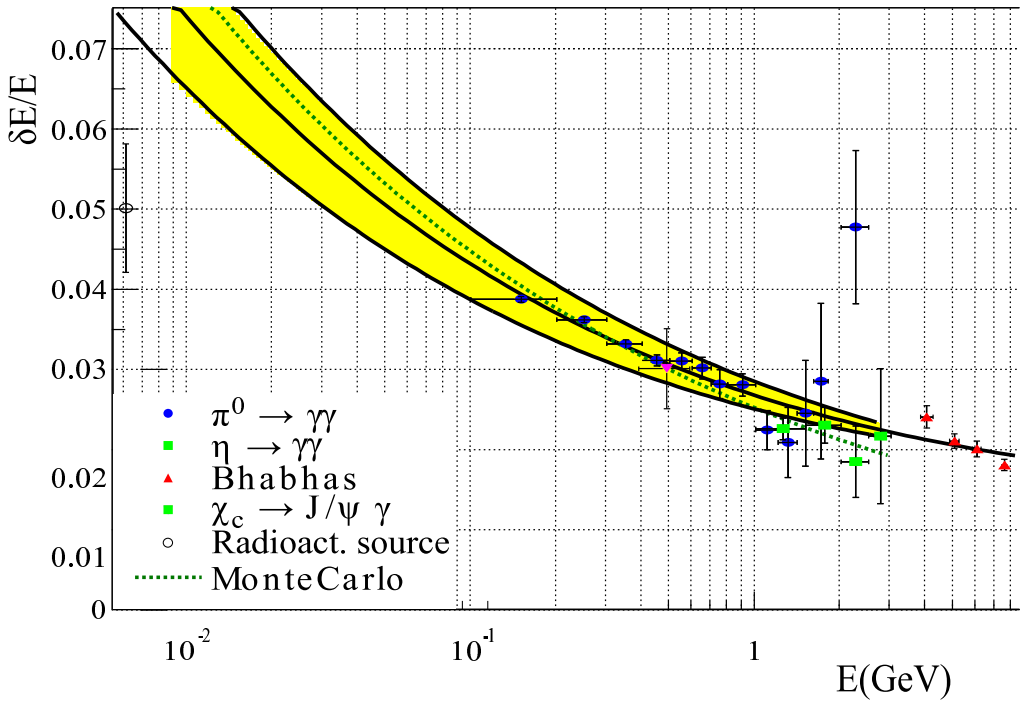


Figure 3.17: Parametrization of the EMC fractional resolution as a function of photon energy E. The energy range used in this analysis is below the data used to constrain the parametrization; although the form of the parametrization will be helpful, the exact resolution function will not be reliable in this low-energy extrapolation. Note that the MC resolution is broader than the parametrization in the low-energy region –this turns out to be inaccurate.

3.4.4 MC background fit

The background contribution is minimal and smooth and is well-described in MC by the summation of two functions, one a decaying exponential with parameter λ and the other a line with slope a.

Fig. **??** demonstrates a fit to the $\Upsilon(2S)$ and $\mu\mu(\gamma)$ samples from the ensemble MC using these two functions. The fitting window is deliberately allowed to be much larger than the expected width of the signal peaks so that the two relevant parameters – λ and a – can be well-determined in the edges where there is no signal remaining.

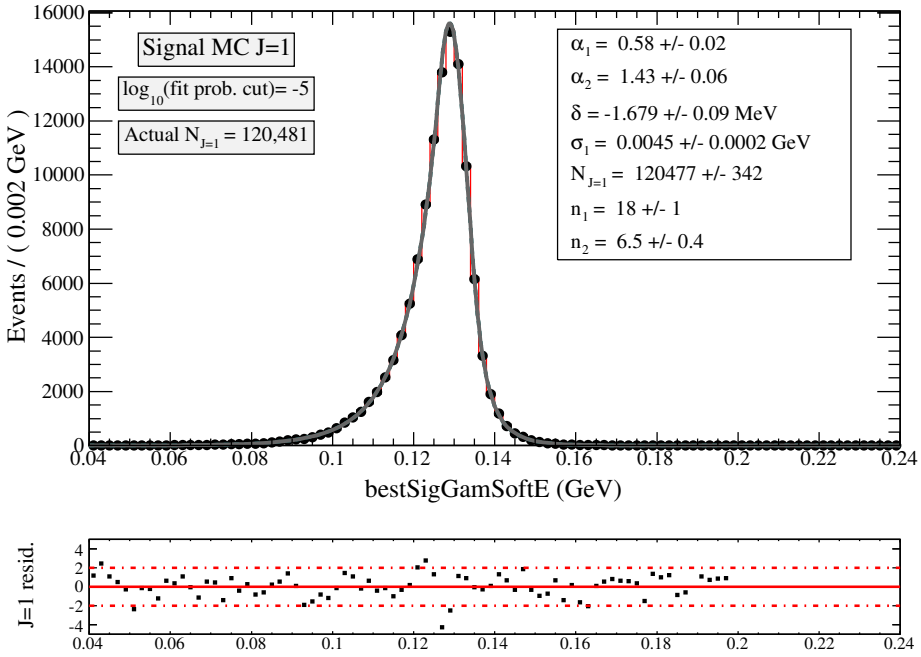


Figure 3.18: Best-fit double-sided Crystal Ball solution for χ_{b1} signal MC with signal fit probability $> 10^{-5}$. The MC spectrum is shown as black dots with error bars (mostly hidden due to small size in the full signal MC collection); the fit is the solid gray line. All floated parameters are shown in the legend, with δ referring to the offset of the PDF's mean from the nominal mean value of 129.64 MeV and all other parameters as shown in Eqn. **??**. The plot on the bottom shows normalized bin-by-bin fit residuals as a visual measure of the goodness-of-fit. Note that this data sample is roughly 12 times the size of the actual data peak, meaning that the residuals are artificially poor in this case. Critical fit regions are the "shoulders" around $80 - 100$ MeV and $140 - 160$ MeV. The fit performs very well in these areas. The integral of the spectrum is a fixed parameter, so the fact that the nominal yield $N_{J=1}$ is equal [within Poisson error] to the fit value is not significant.

3.4.5 Fitting the MC ensemble

The spectrum from the MC ensemble is the primary tool for validating the full spectrum fit. However, to ensure full generality of results – in order to expect acceptable fit behavior on data given acceptable behavior on the MC ensemble – we must be careful about how much MC-specific information is assumed in the fit. As an extreme example, the high-statistics background and signal MC modes can be fit independently

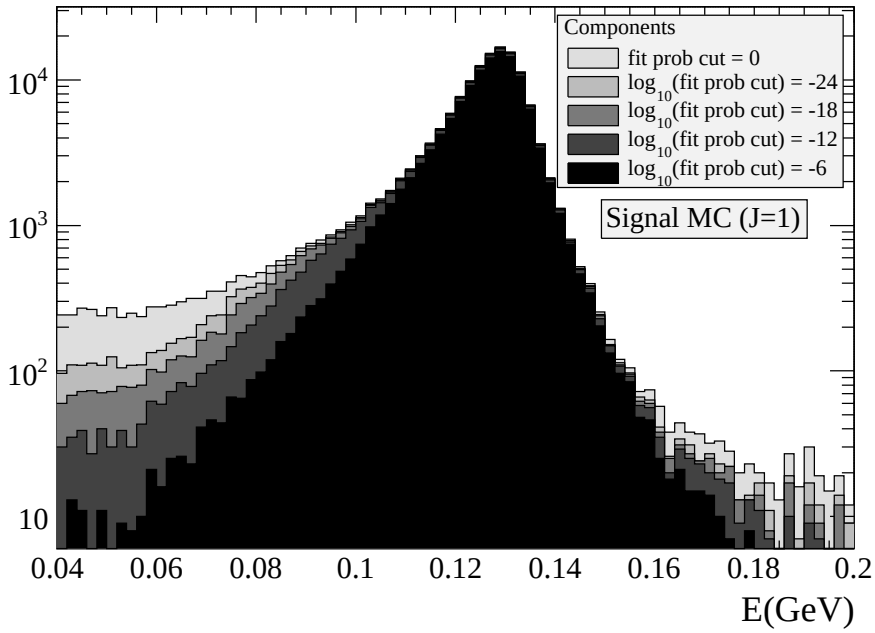


Figure 3.19: Distribution of the $J = 1$ line for an array of fit probability cuts, illustrating the effect of the fit probability cut on the tail shape.

to extremely high precision and all PDF parameters from those fits can be applied to the ensemble MC spectrum. That fit will have extremely high accuracy and precision but will completely lack generality and is thus not useful. However, using some information derived from MC is needed to constrain the fit to avoid systematic biases and allow fit convergence. This section contains a description of those choices as well as tests for consistency for the MC ensemble fit, summarized in Table **??**.

Background constraints

Signal extraction has extremely low sensitivity to the small beam [exponential] component in the data fit, therefore the exponential parameter λ can safely be fixed to the value found on the fit to the MC background sources in Fig. **??**.

The linear component comes almost exclusively from the $\pi^0\pi^0$ background, which we can expect to be modeled well in MC. The continuum background shape has extremely low sensitivity to the exact EMC lineshape, so we fix the linear slope a to that derived from the MC ensemble. The relative ratio of these two sources, $frac$ is

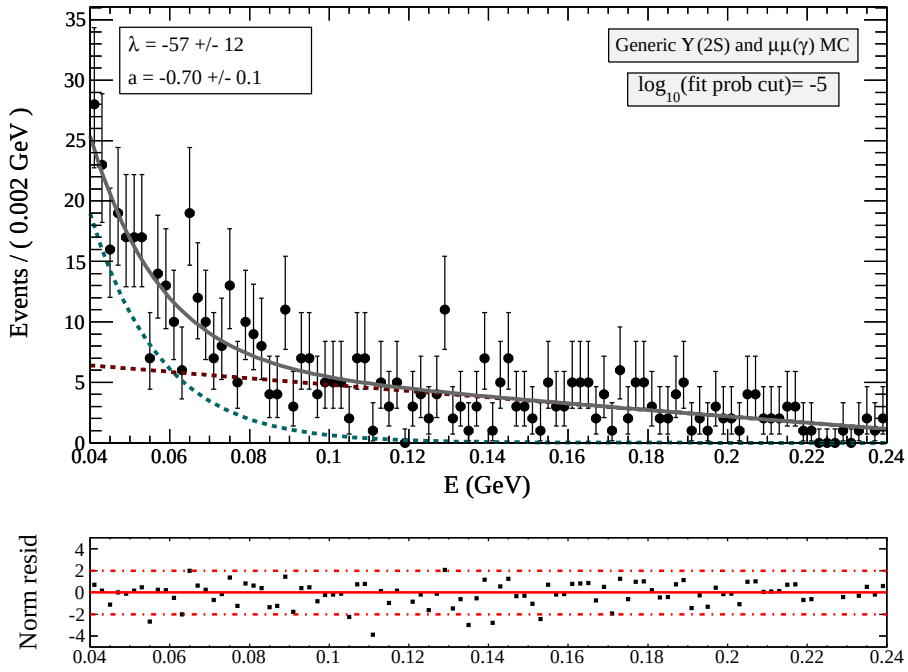


Figure 3.20: Fit to the background components of the MC ensemble using maximum statistics while maintaining expected ratios from the two sources. The ratio between the exponential and linear contributions is allowed to float.

allowed to float to accommodate the expected differences between the beam yield in MC and data.

Lineshape constraints

Perhaps the singular challenge of this analysis is to accurately and precisely fit the overlapping signal peaks in data without biasing the results and without a reliable simulated lineshape. A tendency for the fit to conflate $J = 2$ and $J = 1$ peak entries plagued many versions of this fit, creating an unacceptable systematic bias. The procedure presented here utilizes exactly as much MC lineshape information as can be trusted to constrain the lineshapes.

Given the large degree of overlap between the three signal peaks, accurate signal extraction depends largely on how well the peak parameters can be constrained while maintaining generality. In MC the lineshape at these energies is changing rapidly enough that poor performance is achieved by constraining all three peaks to have the

exact same shape. However, the peak parameters cannot be directly fixed from MC.

Given that the absolute value of all five shape parameters in each peak cannot be fixed to MC, it follows that they must float. However, the fit cannot converge with all 15 signal parameters independently floating. To resolve this dilemma we make the following assumptions:

Linearity Lineshape parameters vary linearly with either the resolution σ or energy $E_{\gamma_{2S\to 1P}}$

MC slopes The slope of the lineshape parameters in resolution or energy is reliable in MC

Power tail The power-law values $n_{1,2}$ can be fixed from signal MC

The first assumption relies on the fact that the three signal peaks are very close in energy with no clear discontinuity in lineshape. The second assumption may be less obvious; it's not immediately clear why the MC width σ would be wrong but the slope $d\sigma/dE$ would be reliable, for example. Consider that the MC resolution is constrained to the correct EMC resolution at higher energies and extrapolated to these lower energies. A small deviation in the slope of the extrapolation would create a large error in the absolute resolution at these energies. The slope will in general always be more reliable than the absolute value.

The third assumption stems from the observation that the fits are extremely insensitive to exact $n_{1,2}$ value. Fixing the three peak tail parameters to values obtained from MC helps constrain the other shape parameters without loss of generality or accuracy.

These assumptions have substantial systematic errors implications which will be investigated in detail using toy studies.

Fig. **??** shows parameters derived from fits to both uncorrected and corrected signal MC with $n_{1,2}$ fixed. It appears that the linear assumption is correct, and the slopes derived in these fits are used to fix the values of σ and $\alpha_{1,2}$ for the $J = 0, 2$ peaks based on the fitted values for $J = 1$.

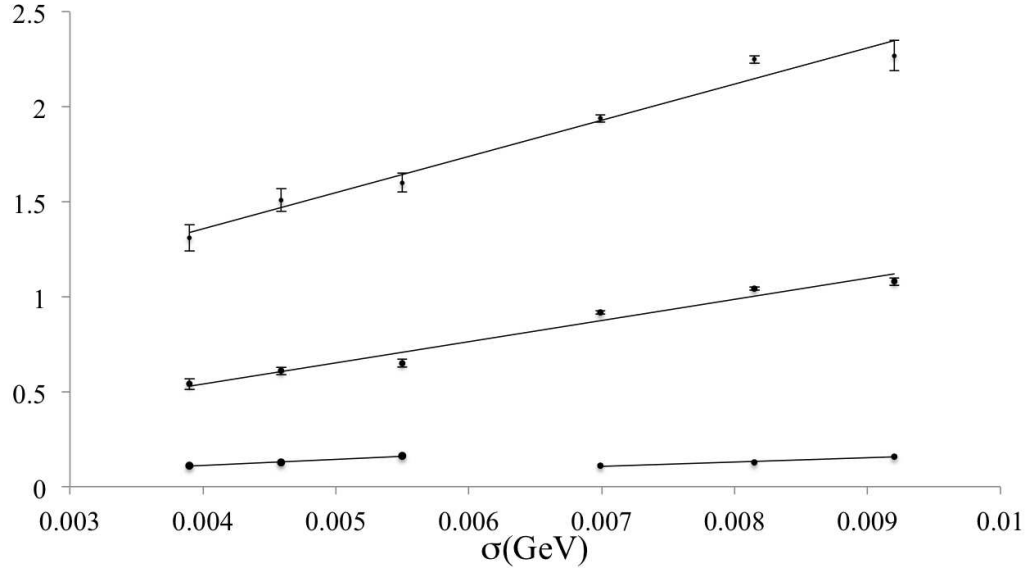


Figure 3.21: Linear fits to the lineshape parameters $\alpha_{1,2}$ (top and middle) and line energy E (bottom) as functions of the fitted width σ in signal MC. The power-law tail parameters $n_{1,2}$ have been fixed to signal MC. The triad of data points on the left correspond to the uncorrected MC sample and the triad to the right to the corrected. The data resolution lies somewhere between these two sets of points. The slopes extracted from these fits are used to constrain the slope of these parameters in the spectrum fit.

MC ensemble fits

Fig. **??** shows the fit to the full ensemble with floating parameters displayed on the plot with zoomed view in Fig. **??** and uncorrected fit in Fig. **??**. Detailed parameter information is displayed in Table **??**.

In principle the correct approach is to float the desired relative parameters N_J/N_1 and $\delta_0 = \mu_0 - \mu_1$. With certain correlations significant differences in value and errors may be obtained. We recast the fitting functions in terms of these relative parameters and refit using `MINOS` to obtain correct asymmetric errors. No significant differences in solutions compared to the original formulation is observed; the independent parameter formulation is maintained to simplify systematic studies.

Parameter	Description	Fixed?	Value
λ	Background exponential parameter	Yes	-67
a_1	Background linear slope	Yes	-0.59
$frac$	Ratio of exponential to linear yields	NO	0.03 ± 0.02
μ_0	Gaussian mean energy for $J = 0$	NO	$161 \pm 1\,\text{MeV}$
μ_1	Gaussian mean energy for $J = 1$	NO	$128.8 \pm 0.1\,\text{MeV}$
μ_2	Gaussian mean energy for $J = 2$	NO	$109.8 \pm 0.2\,\text{MeV}$
$\sigma(J = 1)$	Gaussian width for $J = 1$	NO	$6.2 \pm 0.2\,\text{MeV}$
$\sigma(J = 0)$	Gaussian width for $J = 0$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_0 - \mu_1)$
$\sigma(J = 2)$	Gaussian width for $J = 2$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_2 - \mu_1)$
$\alpha_1(J = 1)$	Low-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	0.89 ± 0.04
$\alpha_1(J = 0)$	Low-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_1(J = 1) + 111.6 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_1(J = 2)$	Low-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_1(J = 1) + 111.6 \cdot (\sigma(J = 2) - \sigma(J = 1))$
$\alpha_2(J = 1)$	High-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	$1.7 \pm 0.1\sigma$
$\alpha_2(J = 0)$	High-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_2(J = 1) + 190.7 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_2(J = 2)$	High-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_2(J = 1) + 190.7 \cdot (\sigma(J = 2) - \sigma(J = 1))$
n_1	Low-end power-law coefficient [shared]	Yes	15
n_2	High-end power-law coefficient [shared]	Yes	5.5
$N_{J=0}$	Yield of $J = 0$ peak	NO	249 ± 31
$N_{J=1}$	Yield of $J = 1$ peak	NO	$9,207 \pm 126$
$N_{J=2}$	Yield of $J = 2$ peak	NO	$5,000 \pm 106$
N_{bkg}	Yield of background	NO	840 ± 54
N	[Implicit parameter – total yield constraint]	Yes	$15,136$

Table 3.5: Summary of parameters from the full data fit in Fig. **??**. Fixed parameters come from high-statistics signal MC.

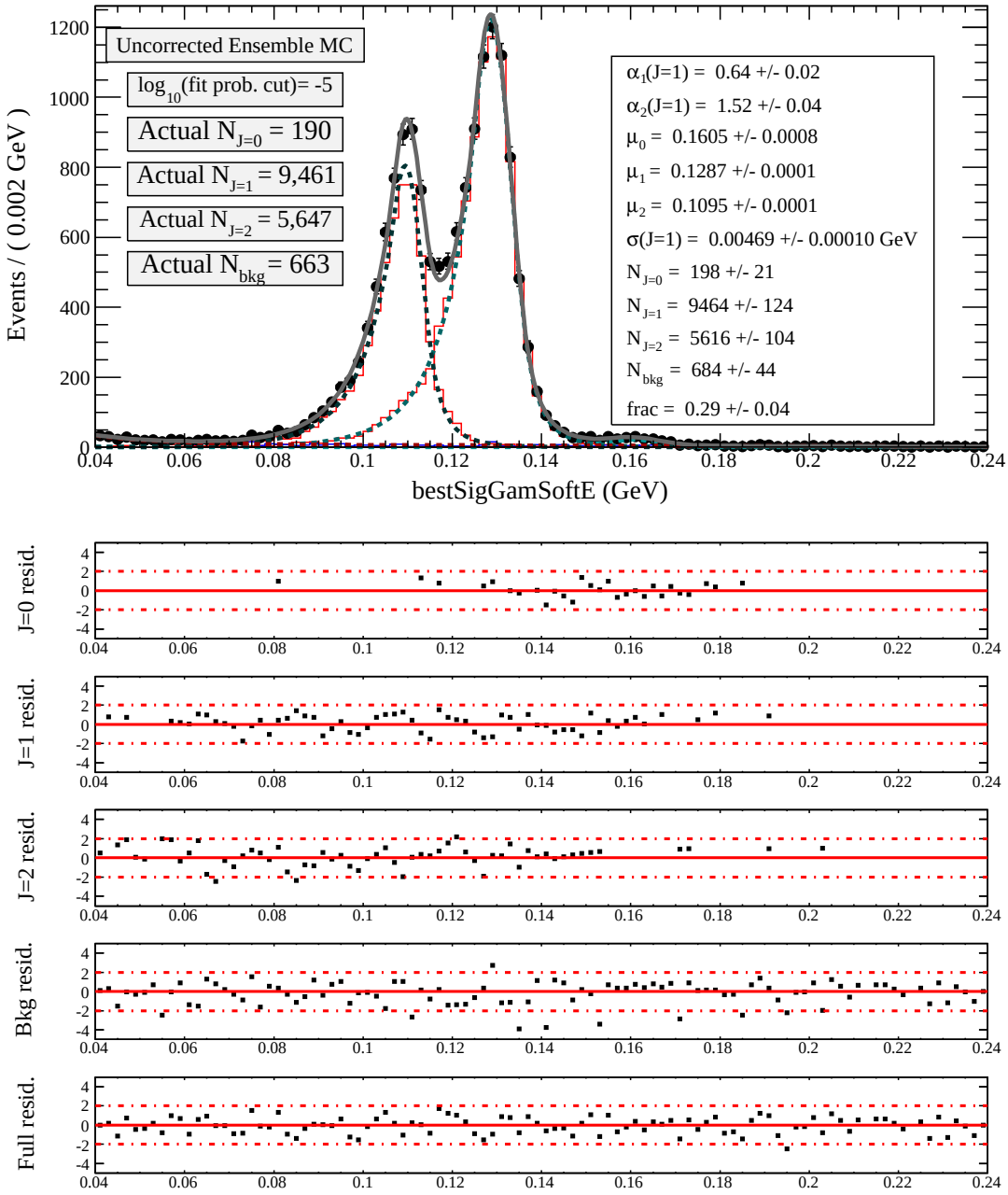


Figure 3.22: Fit to the full uncorrected MC ensemble with component-by-component normalized residuals. The residuals demonstrate that the component yields are not accurate by "accident" but that the lineshapes properly converge on the correct values without being directly fixed from MC.

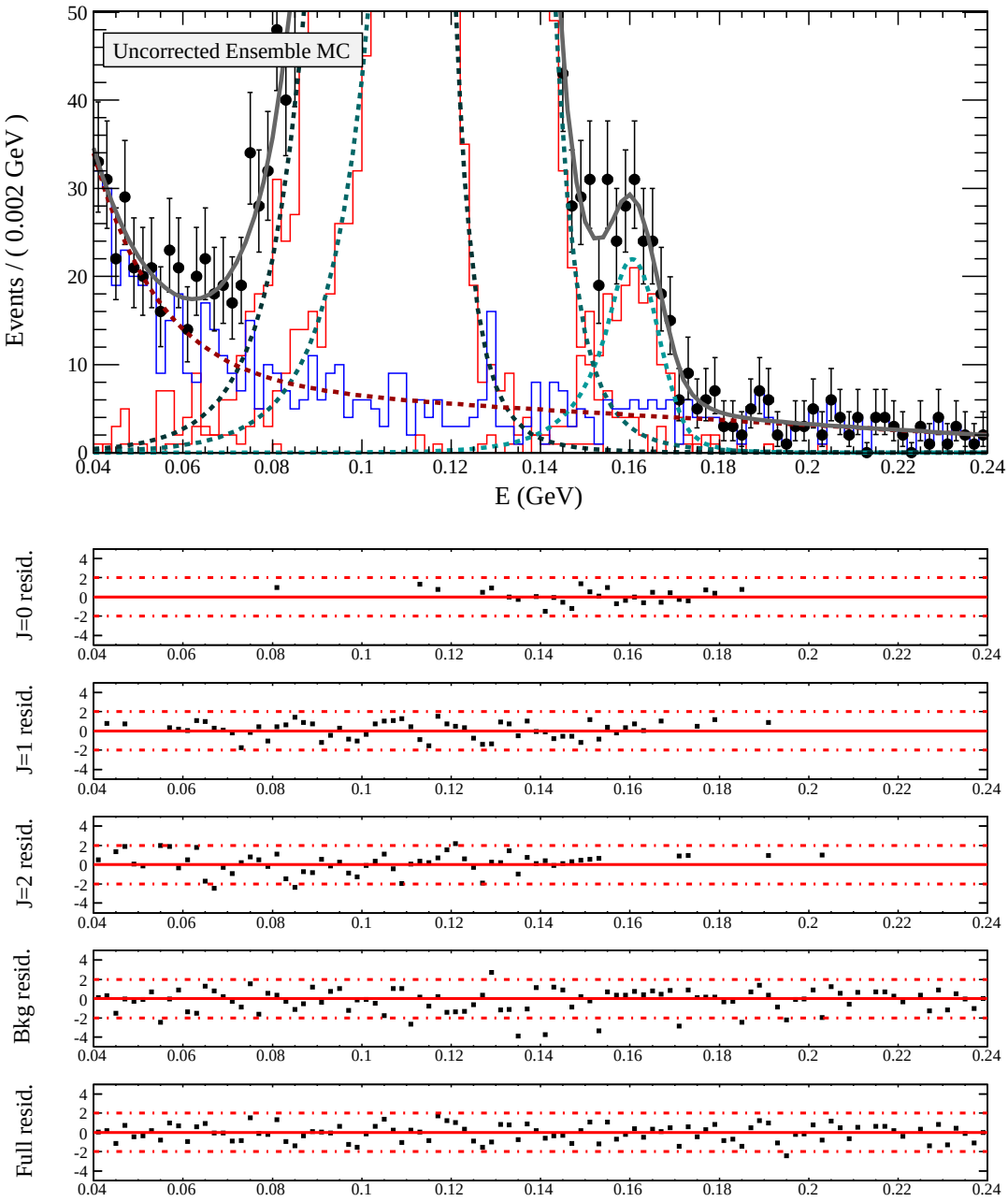


Figure 3.23: Zoomed view of Fig. **??**. Background MC histogram is in blue, signal MC are the red histograms. The dashed lines correspond to the fitted components.

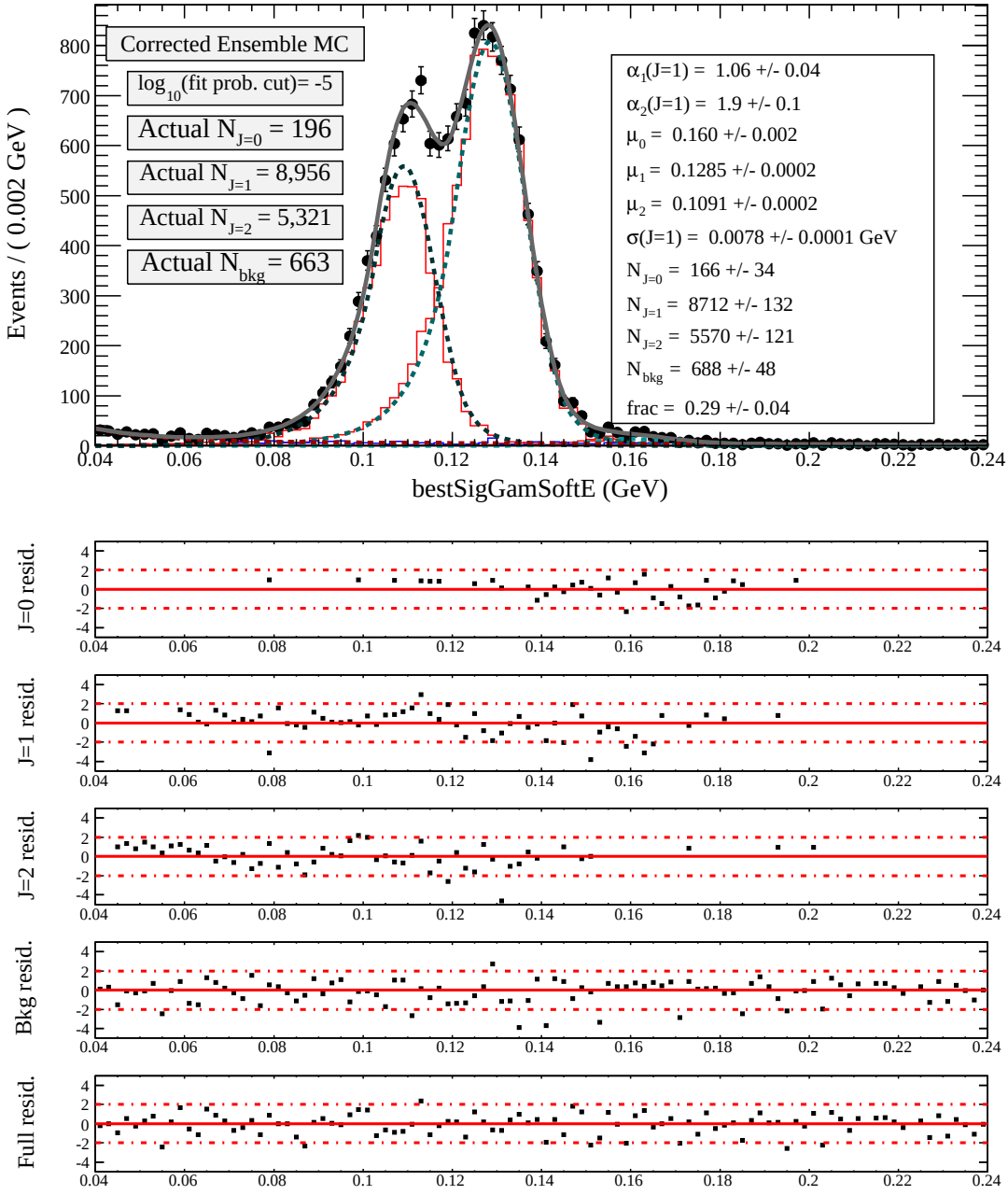


Figure 3.24: Fit to the corrected MC ensemble, verifying that the fit is general in the range of peak widths that covers the EMC resolution.

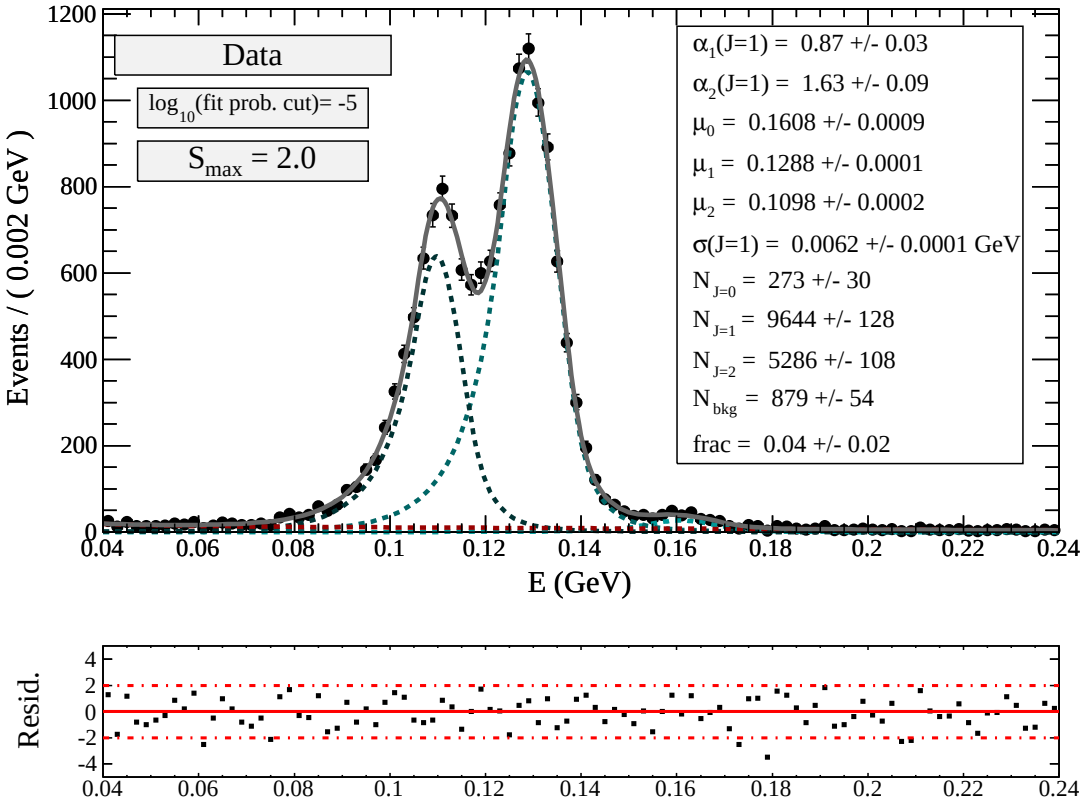


Figure 3.25: Fit to data with final selections.

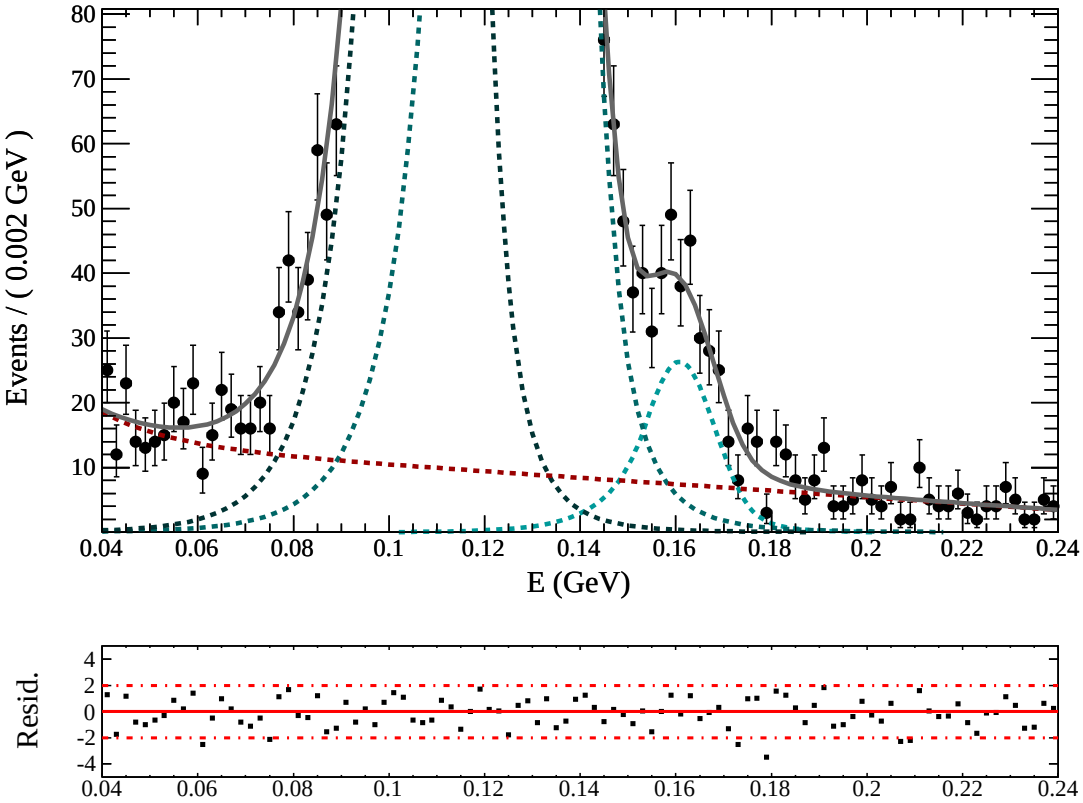


Figure 3.26: The final data fit zoomed to show the $J = 0$ peak.

3.5 $2S \to 1P \to 1S$ systematics

The product branching ratios are derived from measured quantities:

$$\mathscr{F}^{J/J'}_{nS\to mP\to kS} = \frac{N_J}{N_{J'}} \cdot \frac{\epsilon_{J'}}{\epsilon_J}, \tag{3.7}$$

where N_J is the yield of the J peak in the $\gamma_{nS\to mP\to kS}$ spectrum and ϵ_J is the signal efficiency – a combination of detection, reconstruction and selection efficiencies. The systematic uncertainties in the efficiencies are unknown and difficult to measure; the ratio allows them to cancel given that the reconstruction and selections are identical for each of the three signal cascades. The dominant source of error in the ratio is the fit error on the peak yields which are not cancellable.

The mass splittings are derived trivially from the measured Gaussian means $\mu_{0,1,2}$:

$$\Delta M^{1P}_{J-J'} = \pm(\mu J - \mu J'), \tag{3.8}$$

where uncertainty in the correspondence between the measured means and the line energy drop out in the difference and the sign depends on the transitions being evaluated.

The fitting scheme presented here relies on a few very specific assumptions about the nature of the data. So-called "toy" studies allow a direct measurement of the uncertainty introduced by these assumptions. Additionally, toy studies are required to validate the performance of the fit on a more general class of spectra.

3.5.1 Traditional toy MC

The traditional implementation of toy studies utilizes the fit to the complete MC spectrum. Re-simulating the spectrum from physics models through detector response is extraordinarily expensive. Instead, the converged fit PDF – which is found to characterize the spectrum well as shown in Fig. **??** – is used in a Monte Carlo simulation to generate new spectra with normal statistical variance. In this case all the parameters are fixed to the converged values in the MC ensemble fit but the resultant spectrum varies from the original with expected statistical fluctuation. Generating these "toy" spectra is very cheap and a large collection of spectra can be fit in the method of

the MC ensemble to check if the fitting scheme is robust under ordinary variation in spectrum shape.

Since the generating PDF is known – including individual yields for each parameter – a collection of a large number of fits to toy spectra can describe how well the fit converges on the "correct" component yields. Philosophically, of course, this assumes that the generating PDFs are themselves correct representations of the signal shapes and yields, though in the case of the MC ensemble we have already verified the accordance between the signal MC peaks and their corresponding fit components.

The most important result from each fit to a toy spectrum is the "pull" of the fit, defined as:

$$pull = \frac{N_{fit} - N_{generated}}{\sigma_{fit}}, \tag{3.9}$$

meaning the pull is the multiple of the fit error away from the "correct" yield for each component. For a perfect fit with accurate errors the pull distribution then would be centered on 0 with a Gaussian distribution of width 1.

Fig. **??** shows results for toy studies using $2,000$ generated spectra. Note that the pulls are consistent with 0 and the fit errors are very consistent. The width of the pull distributions for the $J = 1$ and $J = 2$ peaks are both significantly lower than the expected 1 – this effect appears to be a result of `RooFit` overestimating the errors involved in separating these two peaks. In any event, a misestimation of fit errors will become irrelevant shortly.

The naïve conclusion would be that toy MC validates the robustness and acuracy of the fit. However, Fig. **??** shows the discordance between the signal MC and data lineshapes in the EMC. To be precise, all that the toy MC has proven is that the MC ensemble fit is accurate on spectra the have the same shape as the MC ensemble. But extrapolating this conclusion to the data spectrum would be inappropriate in light of the apparent lineshape differences, therefore toy studies are required using spectra derived from the data.

3.5.2 Data toys

As an attempt to evaluate whether the ensemble MC toy results hold on the data spectrum, 2,000 toys were generated using a fit to the data spectrum as the generating

PDF.

Fig. **??** shows the result of fits to 2,000 toy spectra using the fit to the data spectrum as the generating PDF. Results are similar to the MC toys, suggesting that the fit also behaves well with data-like lineshapes.

3.5.3 Aristotle is unhappy

In his writings, Aristotle identified many fundamental logical fallacies, two of which are:

Ignoratio elenchi This fallacy can be simply stated "proving the irrelevant." In endeavoring to prove A, a proof of B – no matter how convincing – means nothing [32].

Petitio principii Also known as "begging the question," this fallacy is committed when A is assumed without proof in the process of proving A [31].

Each of these fallacies have been violated by the toy studies of the previous two sections:

MC toys These toy studies prove that the fitting scheme is consistent and accurate on MC-derived spectra. This does nothing to show that the fit is consistent and accurate on the data spectrum. *Ignoratio elenchi.*

Data toys These toy studies prove that the fitting scheme is consistent and accurate on data-derived spectra. However, there is no guarantee that the fit to data is accurate. A bad case of *Petitio principii.*

The implications are not merely philosophical. In the process of this analysis we have tried many fitting schemes which show good quality fits with correct toy pulls but disagree with each other by up to 2σ. A more general set of toy studies is required.

3.5.4 Jitter MC toys

Without a method for generating reliable MC spectra, the next best solution is to generate toys which cover the whole class of spectra which give fit results similar to

those seen with the data spectrum. This is a brute force method but should remove the ambiguity about what the toy results say, with the added benefit of being a direct measurement of systematic errors and uncertainties in the fitting process.

To this end a new set of toy spectra is produced using jittered shape and yield parameters, called here "jitter MC." Table **??** shows the jitter range for each generating PDF parameter – the parameters are jittered over the full range with a flat distribution. Note that the $J = 0$ and $J = 2$ peaks are decoupled from each other; the systematic error from the assumption that the shape parameters are linear is thus measured and folded into the combined systematic errors.

Jitter MC generation

The parameter space is very sparsely covered by any reasonable number of toys if all 21 parameters are allowed to jitter blindly. Some common-sense constraints on the generating PDF's are used without losing generality in order to thoroughly cover the required parameter space. Other parameters are jittered by $\pm 3\sigma$ around a nominal value. Following is a detailed description of these choices by parameter type:

Background The background shape parameter λ is fixed to MC with minimal effect. The linear slope a_1 is jittered around the value extracted from a sideband fit to the data spectrum. The ratio of the yields of these two sources $frac = N_{linear}/N_{exponential}$ jitters similarly.

Gaussian means μ_J The central locations of the three signal peaks jitter around the values extracted from the data fit.

Gaussian widths $\sigma(J)$ The width of the central $J = 1$ peak is jittered centered on the data fit value with a range ± 3 times the data fit error for this parameter. Each other peak width is independently extrapolated from $\sigma(J = 1)$ using the nominal parameter slope derived from signal MC ± 5 times the error on the slope of that fit. To confirm that the parameter fit error is meaningful, this range is compared to an internal parametrization of the EMC resolution [an extrapolation from higher energies] to good agreement [33].

Low-energy transition points $\alpha_1(J)$ For $J = 1$, the low-energy transition point jitters around the value extracted from the data fit. $\alpha_1(J = 0, 2)$ are extrapolated from this value analogous to the method used for the Gaussian widths using slopes and slope errors extracted from the MC parameter fit.

High-energy transition points $\alpha_2(J)$ Identical to the low-energy transition points with the correct slope and slope error.

Tail powers $n_{1,2}$ Jittered within the full range of meaningful values independently.

Component yields N_X Jittered $\pm 3\sigma$ around values extracted from the data fit.

Jitter MC selection

The shape of a simulated signal peak is uniquely determined by the 5 shape parameters. The jitter MC contain no assumptions about line shape that is derived from MC. This generality comes at the price of precision. However, two levels of selections on the fits purify the sample of generated spectra to only include a fairly small class of spectra to which the data spectrum is certainly a member:

Toy selection: fit convergence Only fits that converge with high-quality covariance matrices are included. This is a requirement met by the fit to the data spectrum so no generality is lost. About 10% of the generated spectra fail this selection, which helps purify the generated spectra sample.

Toy selection: fitted line shape parameters Table **??** summarizes selections to fitted parameters. The choices are made to purify the toys while maintaining generality.

The jitter MC results are presented in Fig. **??** for the few jitter MC toys that survive these selections out of $50,000$. The fitting systematic bias and error for the component yields can be read directly off the pull distribution:

$$N = [N_{fit} - \mu_{pull} \cdot \sigma_{N_{fit}}] \pm \sigma_{pull} \cdot \sigma_{N_{fit}}, \tag{3.10}$$

where N_{fit} refers to the fitted yield on the data spectrum, $\sigma_{N_{fit}}$ is the error on that yield, μ_{pull} is the mean of the best-fit Gaussian for the N pull distribution and σ_{pull}

Parameter	Selection	Low value	High value
$\sigma(J=1)$	Data fit $\pm 3\sigma_\sigma$	5.9 MeV	6.5 MeV
$\mu 0$	Data fit $\pm 3\sigma_{\mu 0}$	159.0 MeV	162.6 MeV
$\mu 1$	Data fit $\pm 3\sigma_{\mu 1}$	128.6 MeV	129.0 MeV
$\mu 2$	Data fit $\pm 3\sigma_{\mu 2}$	109.4 MeV	110.2 MeV
N_0	Data fit $\pm 3\sigma_{N_0}$	213	333
N_1	Data fit $\pm 3\sigma_{N_1}$	9,388	9,900
N_2	Data fit $\pm 3\sigma_{N_2}$	5,070	5,502
N_{bkg}	Data fit $\pm 3\sigma_{N_{bkg}}$	771	987

Table 3.6: Summary of selections made on fitted parameters from the $3S \to 2P \to 2S$ jitter toys.

is the width of that Gaussian in units of σ_{fit}. Jitter results for the Gaussian means $\mu_{0,1,2}$ are shown in Fig. **??** using the same subsample of successful toys as above.

This correction has the effect of encapsulating all fit systematics into one measurement, though some precision is sacrificed compared to using an accurate model. In principle this strategy is an overestimation of systematic errors, which is just as inaccurate as under-estimating systematic errors – in combination with other results, these results will be under-weighted. However, the overestimation is made small by the toy selection; furthermore, tighter toy selections do not decrease the spread of the parameter pull distributions. Additionally the overestimation can only be a fraction of the difference between the correction and no correction, which is itself always a fraction of 1σ. Therefore the overestimation is considered negligible.

Parameter	Nominal value source	Nom. value	Param. σ	Jitter range
λ	Fit to MC ensemble background	-67		[Not jittered]
a_1	Fit to data sidebands	-0.66		Nominal $\pm 3\sigma_{a_1}$
$frac$	Fit to data	0.04	0.02	Nominal $\pm 3\sigma_{frac}$
μ_0	Fit to data	160.8 MeV	0.9	Nominal $\pm 3\sigma_{\mu_0}$
μ_1	Fit to data	128.8 MeV	0.1	Nominal $\pm 3\sigma_{\mu_1}$
μ_2	Fit to data	109.8 MeV	0.2	Nominal $\pm 3\sigma_{\mu_2}$
$[m_\sigma]$	Signal MC and EMC res. param.			0.01 – 0.08, about $\pm 5\sigma$
$\sigma(J=1)$	Fit to data	6.2 MeV	0.1 MeV	Nominal $\pm 3\sigma_{\sigma(J=1)}$
$\sigma(J=0)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_0 - \mu_1)$
$\sigma(J=2)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_2 - \mu_1)$
$[m_{\alpha_1}]$	MC parameter fit	111.6	9.9	
$\alpha_1(J=1)$	Fit to data	0.87	0.03	Nominal $\pm 3\sigma_{\alpha_1}$
$\alpha_1(J=0)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_1(J=2)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=2) - \sigma(J=1))$
$[m_{\alpha_2}]$	MC parameter fit	190.7	15.6	
$\alpha_2(J=1)$	Fit to data	1.63	0.09	Nominal $\pm 3\sigma_{\alpha_2}$
$\alpha_2(J=0)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_2(J=2)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=2) - \sigma(J=1))$
n_1	Manual			5.0 – 100.0
n_2	Manual			5.0 – 100.0
$N_{J=0}$	Fit to data	273	30	Nominal $\pm 3\sigma_{N_{J=0}}$
$N_{J=1}$	Fit to data	9,644	128	Nominal $\pm 3\sigma_{N_{J=1}}$
$N_{J=2}$	Fit to data	5,286	108	Nominal $\pm 3\sigma_{N_{J=2}}$
N_{bkg}	Fit to data	879	54	Nominal $\pm 3\sigma_{N_{bkg}}$

Table 3.7: Summary of parameters used to generate the jitter MC spectra. See the text for details.

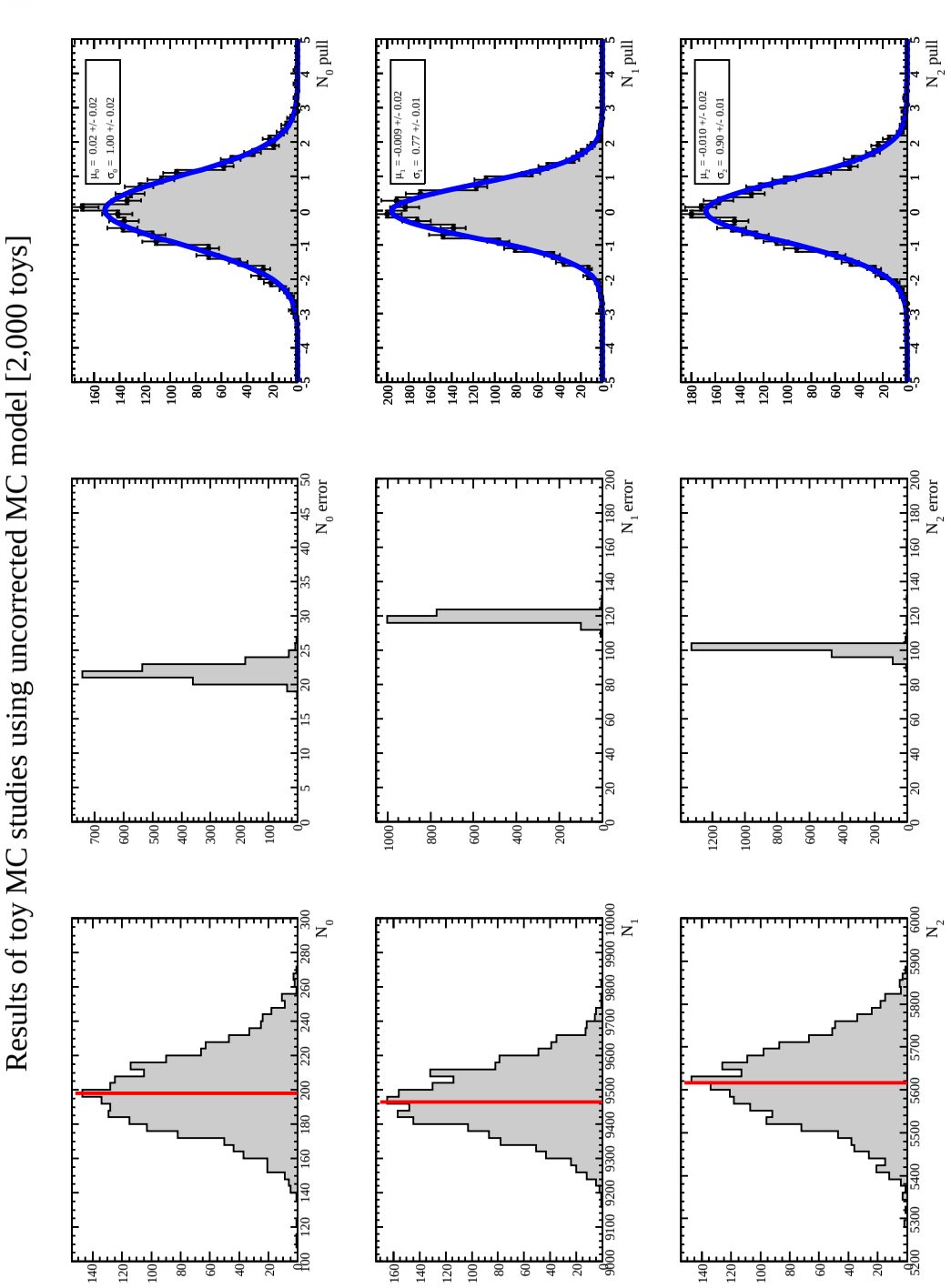


Figure 3.27: Fitted N_0, N_1 and N_2 yields from fits to 2,000 simulations of the $\gamma_{2S\to 1P}$ spectrum using the fit to the MC model as the generating PDF. The absolute value of the fit error and pulls are also plotted for each yield. Red lines indicate the "correct," – i.e. generating PDF yield – for each component.

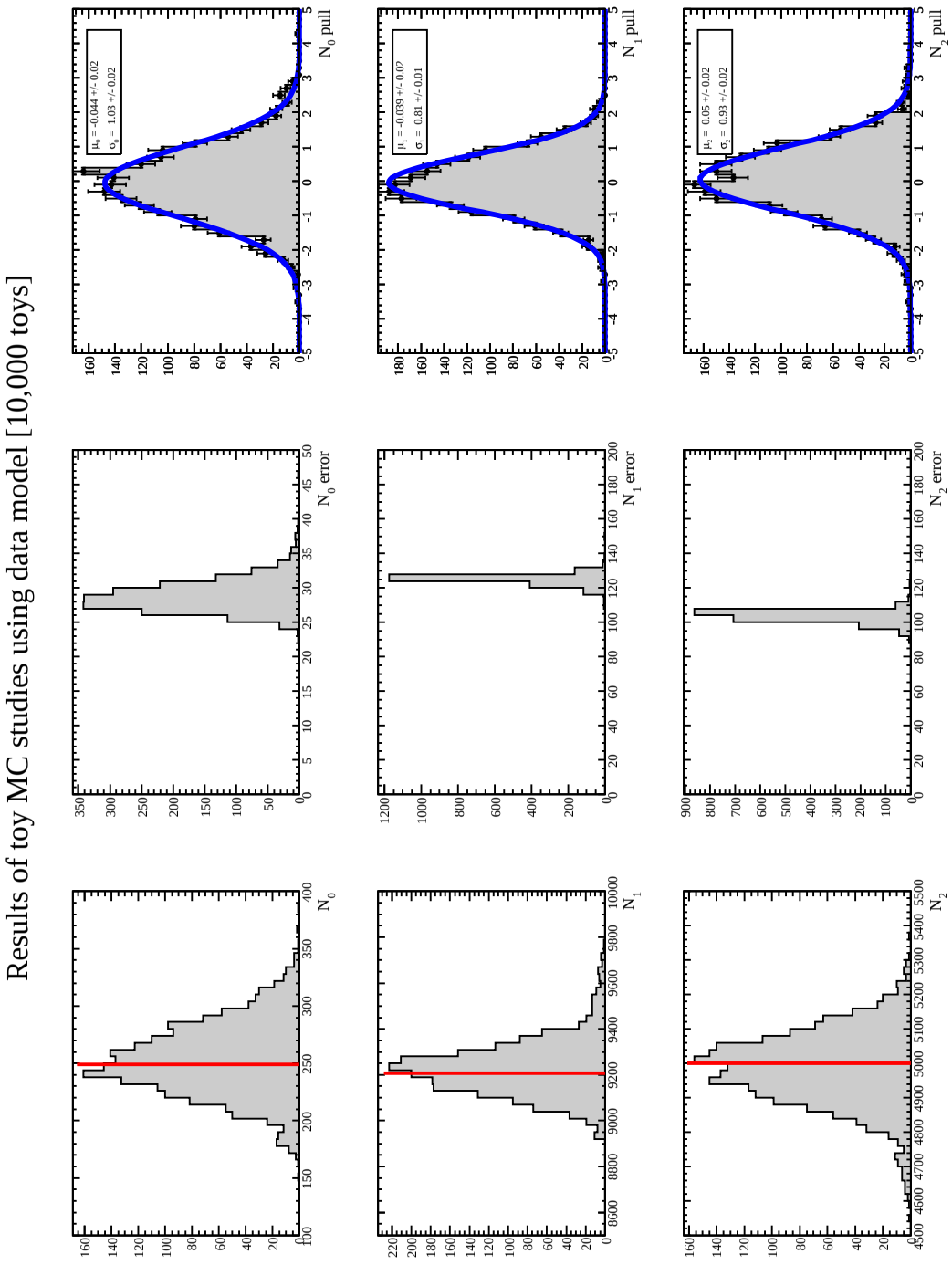


Figure 3.28: Toy study results using a fit to the data spectrum as the generating PDF. See the text for caveats.

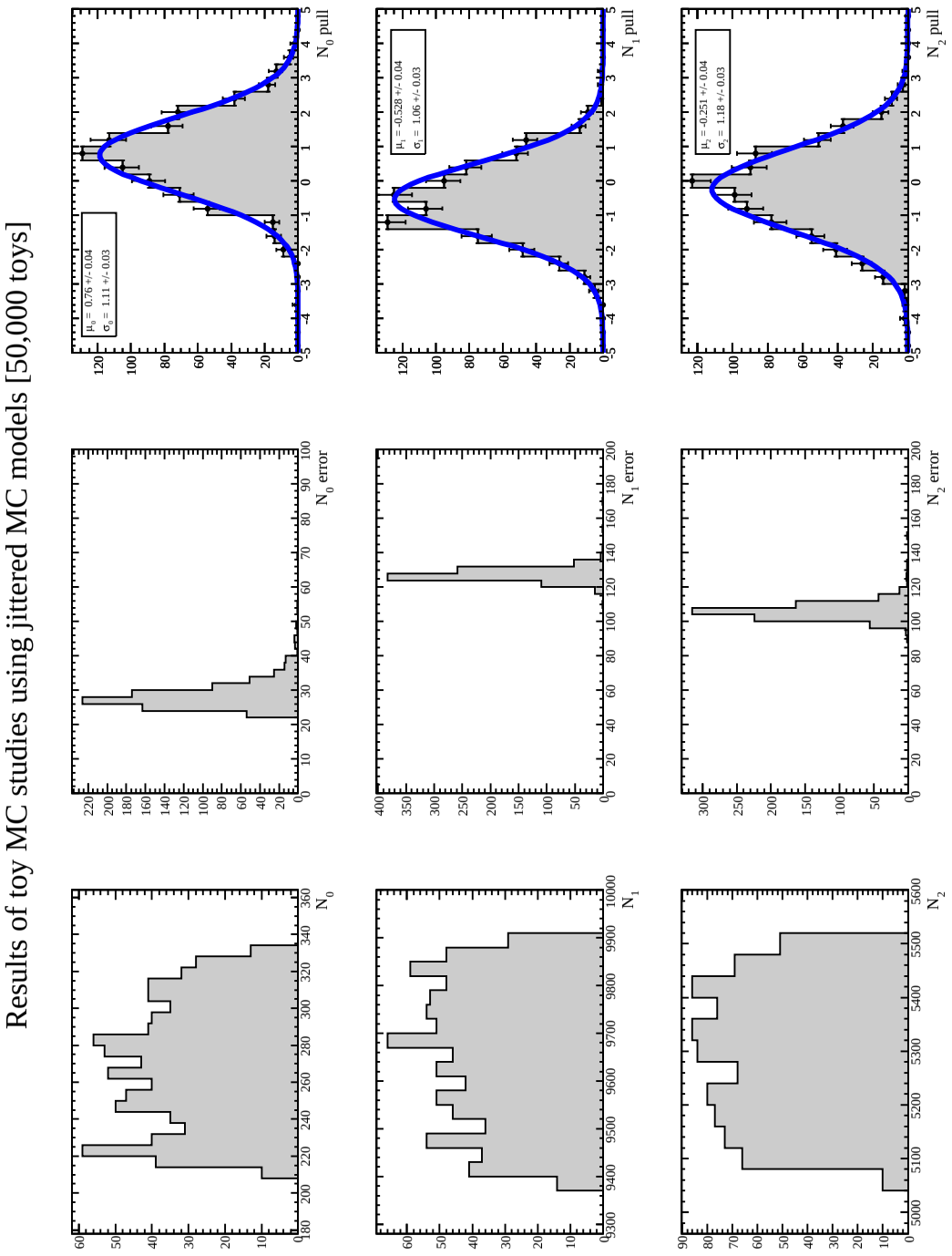


Figure 3.29: The results of 50,000 fits to the jitter MC spectra for the yield parameters $N_{0,1,2}$.

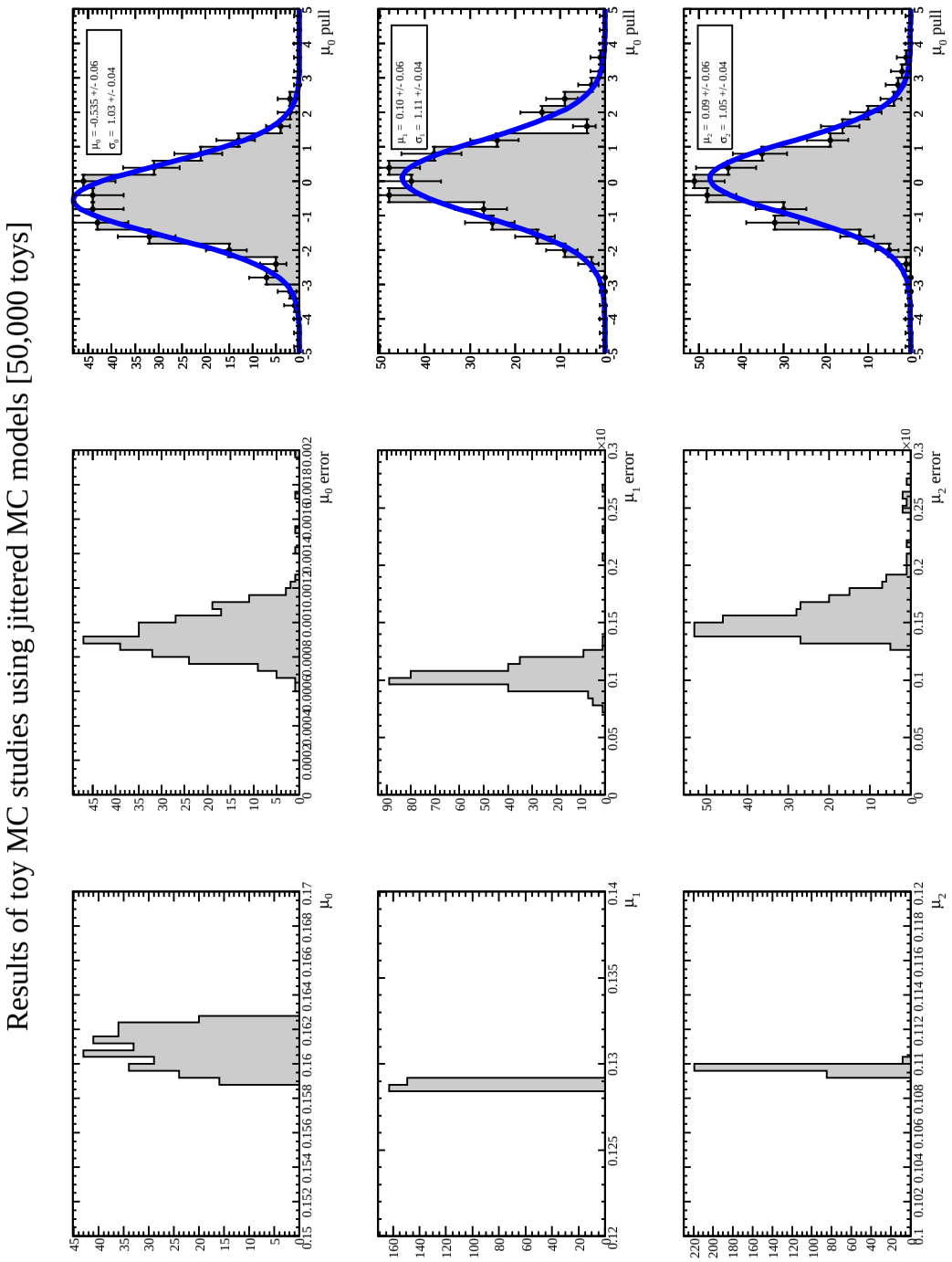


Figure 3.30: The results of 50,000 fits to the jitter MC spectra for the Gaussian mean parameters $\mu_{0,1,2}$.

3.6 $2S \to 1P \to 1S$ results

We calculate the spin-dependent branching fraction ratios from the corrected yields and efficiencies from Table **??** following Eqn. **??**:

$$\mathscr{F}^{0/1}_{2S\to 1P\to 1S} = \frac{296 \pm 33}{9576 \pm 136} \times (1.062 \pm 0.009) = (3.28 \pm 0.37)\%, \tag{3.11}$$

$$\mathscr{F}^{2/1}_{2S\to 1P\to 1S} = \frac{5259 \pm 127}{9576 \pm 136} \times (1.013 \pm 0.004) = (55.6 \pm 1.6)\%, \tag{3.12}$$

The mass splittings are measured to be:

$$\Delta M^{1P}_{1-0} = (161.28 \pm 0.93 - 128.79 \pm 0.11)\,\mathrm{MeV} = 32.49 \pm 0.93\,\mathrm{MeV}, \tag{3.13}$$

$$\Delta M^{1P}_{2-1} = (128.79 \pm 0.11 - 109.78 \pm 0.21)\,\mathrm{MeV} = 19.01 \pm 0.24\,\mathrm{MeV}. \tag{3.14}$$

Comparable results derived from individual branching fractions as reported in the PDG are:

$$\mathscr{F}^{0/1}_{2S\to 1P\to 1S} = (2.89 \pm 0.69)\%, \tag{3.15}$$

$$\mathscr{F}^{2/1}_{2S\to 1P\to 1S} = (58.4 \pm 6.9)\%, \tag{3.16}$$

$$\Delta M^{1P}_{1-0} = 33.34 \pm 0.66\,\mathrm{MeV}, \tag{3.17}$$

$$\Delta M^{1P}_{2-1} = 19.43 \pm 0.57\,\mathrm{MeV}. \tag{3.18}$$

See Chapter 6 for a discussion of the results.

Chapter 4

$3S \to 2P \to 2S$

Analysis of the $3S \to 2P \to 2S$ cascade (Fig. **??**) is very similar to the $2S \to 1P \to 1S$ case. Details given in this chapter are therefore restricted to instances where there is a notable difference between the two modes; otherwsise the procedures of this analysis are identical to the $2S \to 1P \to 1S$ analysis.

4.1 $3S \to 2P \to 2S$ event selection and reconstruction

Reconstruction and event selection proceeds analogously to the $2S \to 1P \to 1S$ analysis, with a small variation in best event selection to be explained during the cuts discussion. The MC reconstruction efficiencies, found in Table **??**, are quite similar to those seen in the previous analysis. This result is consistent with the idea that the J-dependence is due to angular distribution effects, which should be identical here to the $2S$ case.

The soft and hard photons are defined in candidate selection according to $40 < E_{3S\to 2P} < 160\,\mathrm{MeV}$ and $160 < E_{2P\to 2S} < 280\,\mathrm{MeV}$. The low energy of the "hard" photon candidates is the defining characteristic of this analysis mode.

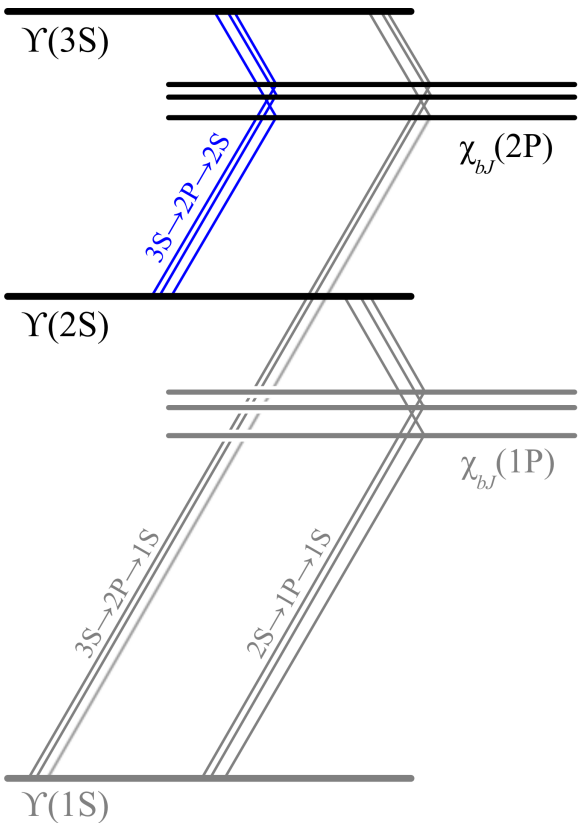


Figure 4.1: The decay mode evaluated in this chapter.

SP mode #	Mode	ϵ_{reco}
11502	$\chi_{b0}(2P)$ cascade	0.200 ± 0.005
11501	$\chi_{b1}(2P)$ cascade	0.221 ± 0.004
11500	$\chi_{b2}(2P)$ cascade	0.215 ± 0.005

Table 4.1: Reconstruction efficiencies for the three signal MC modes for the $3S \to 2P \to 2S$ analysis. Errors are solely statistical; a large unknown systematic error is hidden.

4.2 $3S \to 2P \to 2S$ background

Backgrounds in the $3S \to 2P \to 2S$ mode are similar to those for the $2S \to 1P \to 1S$ case, with the same significant modes contributing. No significant differences are seen qualitatively with the $\mu\mu(\gamma)$ background except for the expectation that this background will be greater in this case given the proximity of the $\Upsilon(3S)$ and $\Upsilon(2S)$ resonances. The lower photon energies imply a larger beam contribution.

4.2.1 Backgrounds in MC

Figs. **??** and **??** show the MC mothers of $\gamma_{3S\to 2P}$ and $\gamma_{2P\to 2S}$ in generic $\Upsilon(3S)$ decays reconstructed as the $3S \to 2P \to 2S$, excluding the signal mode. Some differences in the background contributions relative to the $2S \to 1P \to 1S$ case are evident, described here.

Background source: FSR

A new category of photon background becomes important, particularly for the high-energy photons. These photons in MC come from the `PHOTOS` package which models QED Bremsstrahlung photons in particle decays. Although in MC these photons are associated with $\Upsilon(nS)$ mothers, they can be understood as equivalent to the FSR photons seen in $\mu\mu(\gamma)$. These are physics photons and are thus in-time.

The process $\Upsilon(3S) \to \gamma_{FSR}\mu\mu$ with the $\gamma_{2P\to 2S}$ coming from γ_{FSR} and the $\gamma_{3S\to 2P}$ candidate from beam is indistinguishable from $\mu\mu(\gamma)$ events and is equally susceptible to timing discrimination.

Background source: $\pi^0\pi^0$

The process $\Upsilon(3S) \to \pi^0\pi^0\Upsilon(2S)$ is suppressed relative to the signal cascade by a factor of 6 compared to the $\Upsilon(2S)$ case. This effect is seen clearly in Figs. **??** and **??** in comparison to Figs. **??** and **??**. The nominal fit probability cut as employed in the $2S$ analysis will likely render this background source negligible.

Background source: beam

Beam background appears to be a significant source of background for both photon candidates. The lower energy of the $\gamma_{2P\to 2S}$ photons compared to $\gamma_{1P\to 1S}$ allows many beam photons to be accepted; cuts in the $2S$ analysis assumed that the $\gamma_{1P\to 1S}$ candidate was in-time regardless of source, but this will not be universally true in this analysis for γ_{hard} candidates.

4.2.2 $3S \to 2P \to 2S$ MC ensemble

Table **??** shows the scaling and MC samples used to create the MC ensemble for $3S \to 2P \to 2S$. Using these scalings the 2D scatterplot $E_{\gamma_{3S\to 2P}}$ vs. $E_{\gamma_{2P\to 2S}}$ is shown in Fig. **??** before selections and in Fig. **??** after a nominal fit probability cut, the same used in the $2S \to 1P \to 1S$ analysis. It is clear that the $\mu\mu(\gamma)$ background is more troublesome in this mode.

The $J = 0$ peak in the MC ensemble is radically more prominent visually than in the data spectrum using PDG branching fractions. In order to better approximate the obscurity of this peak in the MC ensemble, we manually decrease the scale factor to a value of 0.002 from a nominal 0.01. Final results will confirm that the PDG branching fractions are anomalously large for this mode.

SP #	Mode	# MC	# Run 7	Scale	# reconstructed	Scaled #
8739	Generic $\Upsilon(3S)$ decays:	$450.5M$	$122M$	0.27	$11,604$	$3,133$
	$\Upsilon(3S) \to \pi^0\pi^0\Upsilon(2S)$				648	175
	$\Upsilon(3S) \to \mu\mu$				$10,811$	$2,919$
	All other generic				145	Negligible
11502	Full $\chi_{b0}(2P)$ cascade	$621k$	$6.4k$	0.002^*	$124,095$	248
11501	Full $\chi_{b1}(2P)$ cascade	$717k$	$59k$	0.082	$158,498$	$12,997$
11500	Full $\chi_{b2}(2P)$ cascade	$616k$	$33k$	0.053	$132,290$	$7,011$
3981	$\mu\mu(\gamma)$	$39.2M$	$32M$	0.82	$36,949$	$29,929$

Table 4.2: See Table **??** for discussion. The signal cascade modes include both $\mu\mu$ and ee decay modes of the $\Upsilon(2S)$, thus the reported MC count is half the total $\Upsilon(3S)$ decays in those modes. Entries denoted by $*$ are scaled manually as described in the text.

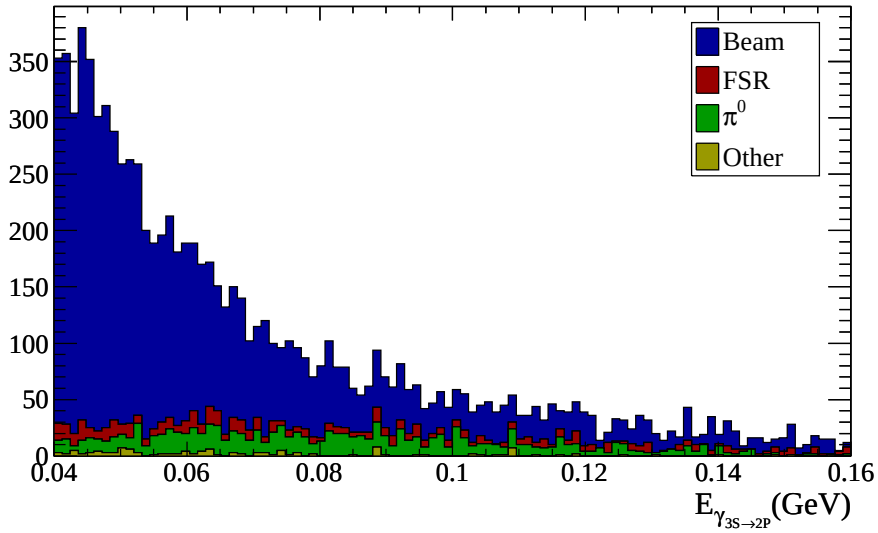


Figure 4.2: Mothers of $\gamma_{3S\to 2P}$ candidates in generic $\Upsilon(3S)$ MC for reconstructed $3S \to 2P \to 2S$ cascades. Cascade candidates are required to have a signal fit probability above zero but multiple candidates are allowed in each event. An arbitrary subset of the full generic $\Upsilon(3S)$ dataset is used.

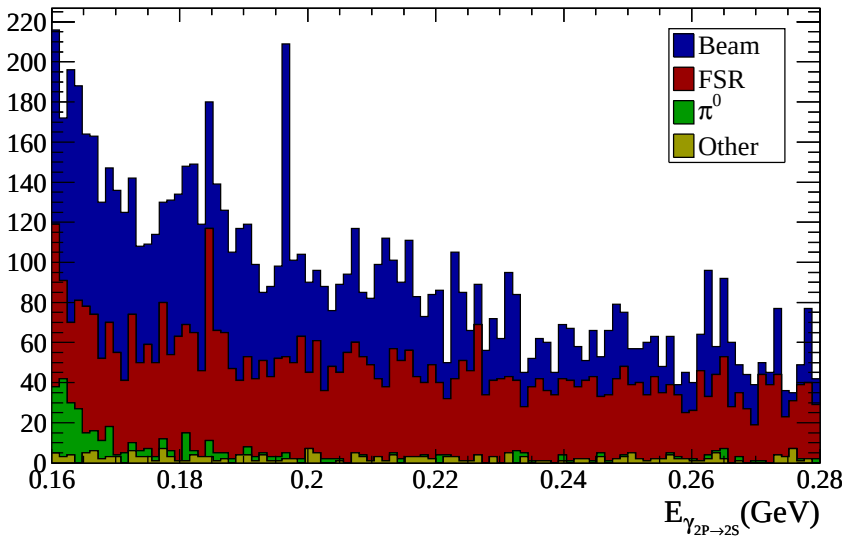


Figure 4.3: Mothers of $\gamma_{2P\to 2S}$ candidates in generic $\Upsilon(3S)$ MC for reconstructed $3S \to 2P \to 2S$ cascades.

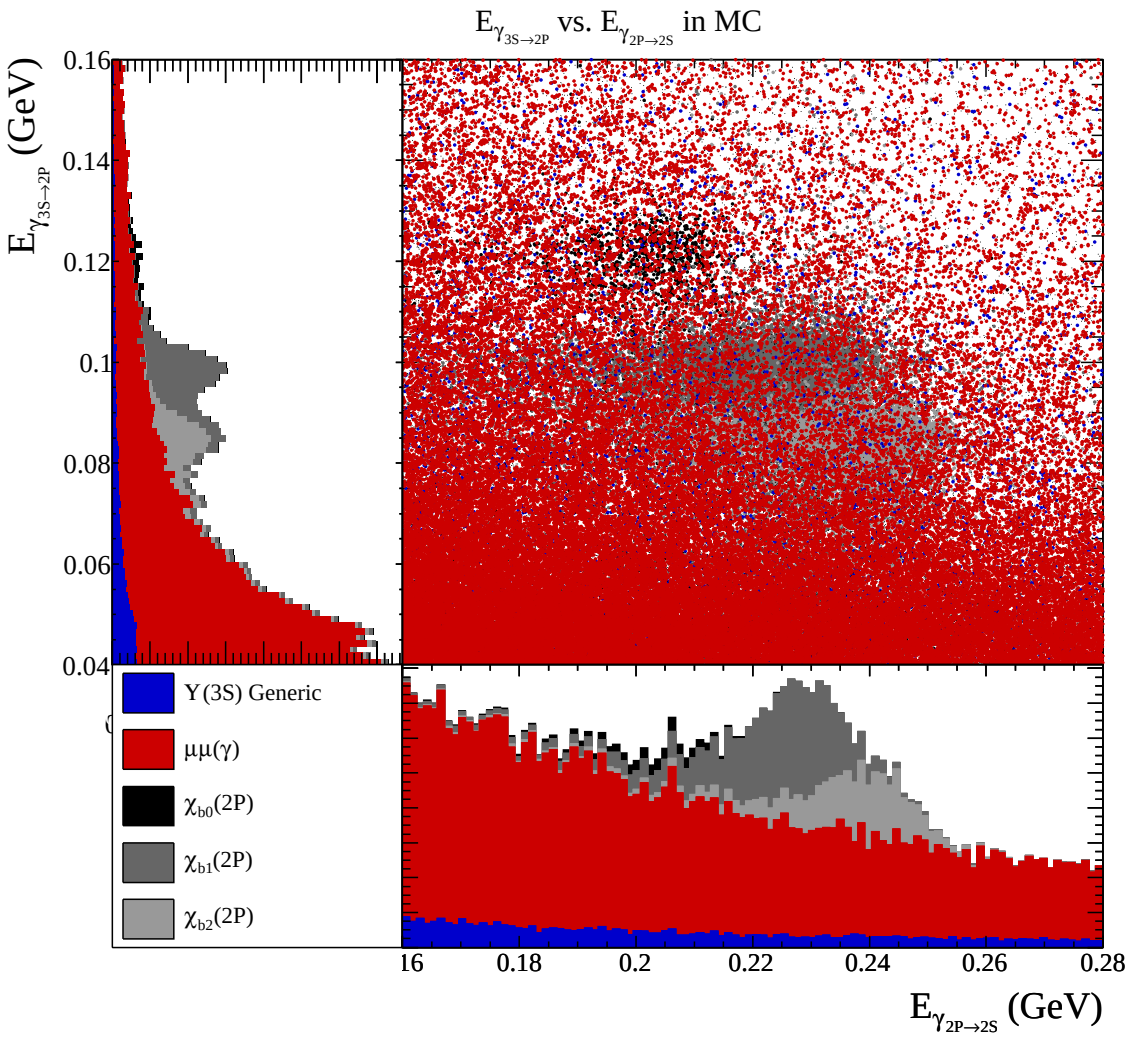


Figure 4.4: Scatterplot of $E_{\gamma_{3S\to 2P}}$ vs $E_{\gamma_{2P\to 2S}}$ in ensemble MC, showing the distribution of the main background sources.

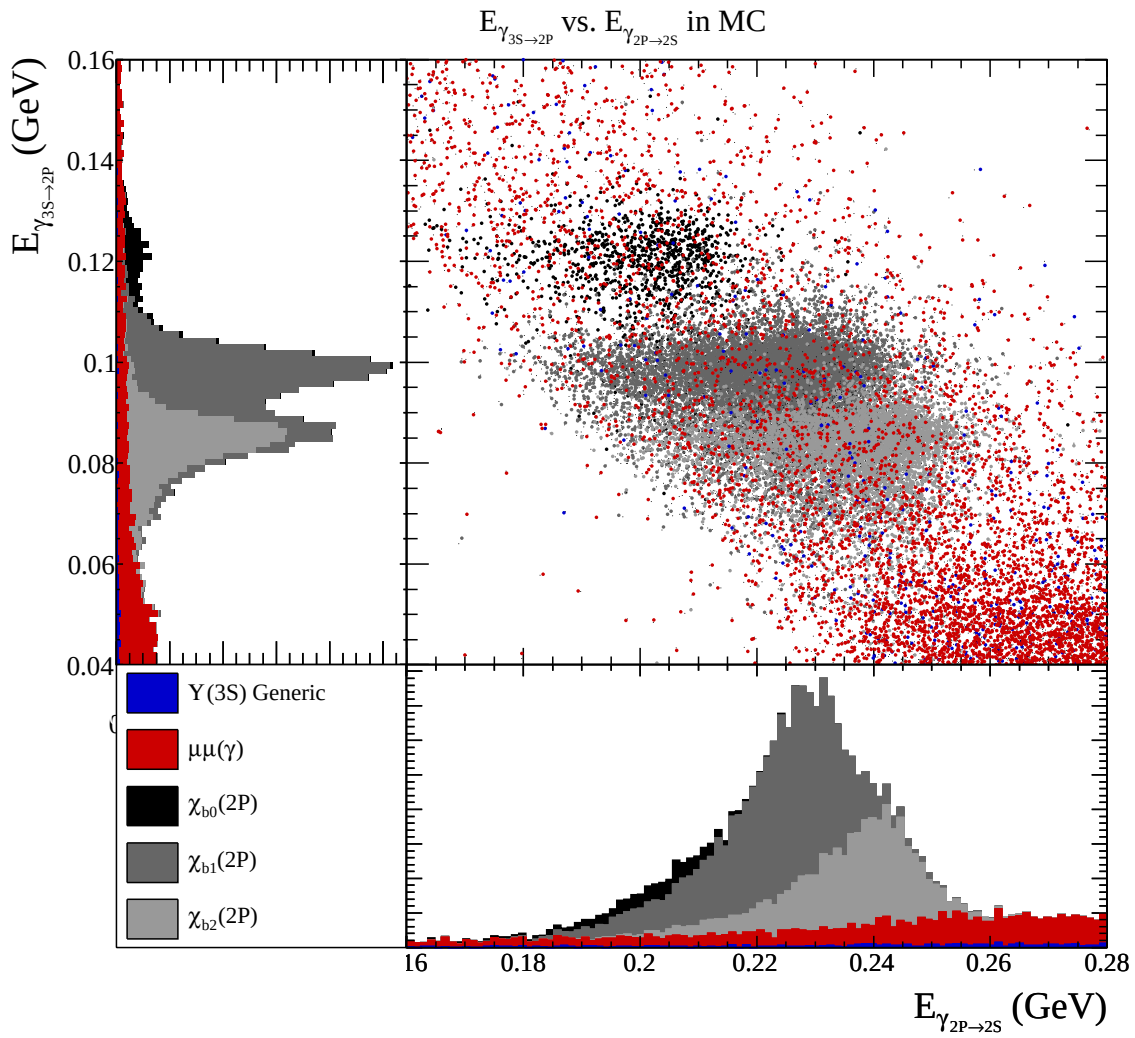


Figure 4.5: Scatterplot of $E_{\gamma_{3S\to 2P}}$ vs $E_{\gamma_{2P\to 2S}}$ in ensemble MC including only candidates with signal fit probability greater than 10^{-5}.

4.3 $3S \to 2P \to 2S$ cuts

Background rejection in this analysis targets the same main sources of background – $\mu\mu(\gamma)$ and $\Upsilon(3S) \to \pi^0\pi^0\Upsilon(2S)$ – as in the $2S \to 1P \to 1S$ analysis with the same tools. An adjustment of the selection strategy is needed to accomodate the significantly higher $\mu\mu(\gamma)$ contribution, described below.

4.3.1 Timing cut

With a large portion of the background for both photon candidates coming from beam background, the $\gamma_{2P\to 2S}$ candidate can no longer be used as an in-time benchmark for timing. Comparing the timing between two out-of-time photons is still useful; there is no loss in signal efficiency and only a modest loss in background rejection. Both μ candidates leave clusters in the EMC which are guaranteed to be in-time, so in principle these can be used as an in-time benchmark. However, the energy deposited [$\sim 200\,\mathrm{MeV}$] by muons in the MEC is largely independent of muon energy and the clusters include much fewer individual crystals than a similar-energy photon [see Appendix C]. This means that the μ cluster times have an associated error that is much larger than comparing to a signal photon. Comparing the timing of the two signal photons is still preferable.

In the $\Upsilon(2S)$ analysis, the timing cut provided a very small improvement in peak discrimination due to the already-low backgrounds. Given the significantly larger $\mu\mu(\gamma)$ background in the $3S \to 2P \to 2S$ mode the timing selection becomes much more important. In the $2S$ case, no significant difference was seen between selecting on the timing difference significance before or after choosing the best cascade candidate. In order to reduce uncertainties in the effects of the timing cut in the MC ensemble we chose to select for timing after best cascade selection. A different approach is used in this analysis, presented below.

Improved cascade candidate selection

Each event has, in general, many cascade candidates. To choose the "best" cascade candidate we required only that the two photon candidates fit within the energy

Selection	$\epsilon_{J=1}/\epsilon_{J=0}$	$\epsilon_{J=1}/\epsilon_{J=2}$	$\epsilon_{J=2}/\epsilon_{J=0}$
Reco. and initial selections	1.106 ± 0.005	1.029 ± 0.004	1.075 ± 0.004
Signal fit probability	0.9576 ± 0.0013	0.9601 ± 0.0006	0.997 ± 0.001
Total	1.0593 ± 0.0047	0.9883 ± 0.0041	1.0719 ± 0.0048

Table 4.3: Efficiency ratios for the three $3S \to 2P \to 2S$ signal modes from MC.

acceptance windows and we selected the cascade candidate with the highest fit probability that met that condition. This was the "best" candidate. Timing selection occurred after best candidate selection in the $2S \to 1P \to 1S$ analysis. Now we perform this timing selection on every pair of photon candidates within each event; a pair must be acceptably in-time before being considered as a best cascade candidate. Fig. **??** shows the effect of this change using $S_{max} = 2.0$.

In order to simulate this effect in the MC ensemble, before selection we reject pairs of photons both from in-time sources at a rate of 0.92 and with at least one photon from out-of-time sources at a rate of 0.40. The resulting MC ensemble is shown in Fig. **??** compared to the data spectrum with the new best candidate selection and fit probability cut.

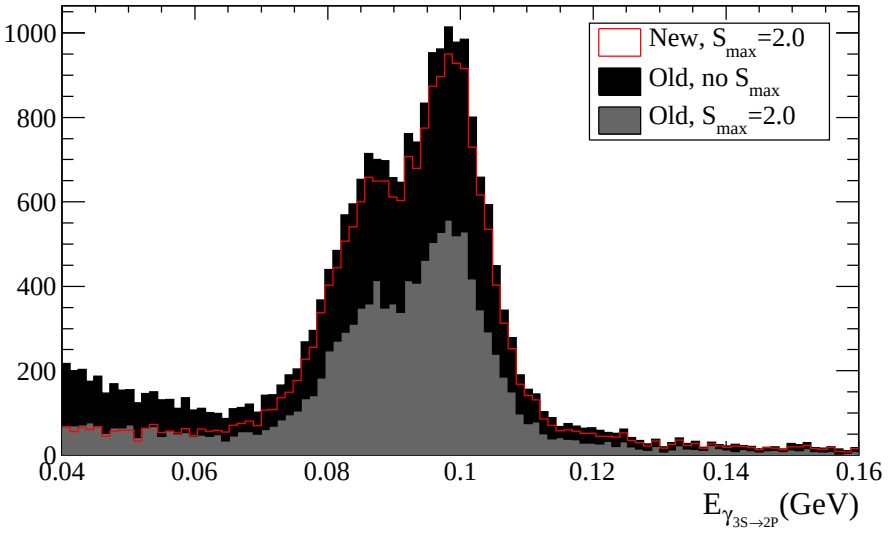


Figure 4.6: Comparison of the old and new best cascade selection on the data spectrum with cascade fit probability greater than 10^{-5}. The new strategy – rejecting cascades from best candidate consideration if the timing difference significance between the two signal photons exceeds S_{max} – has greatly improved signal efficiency compared to the old method.

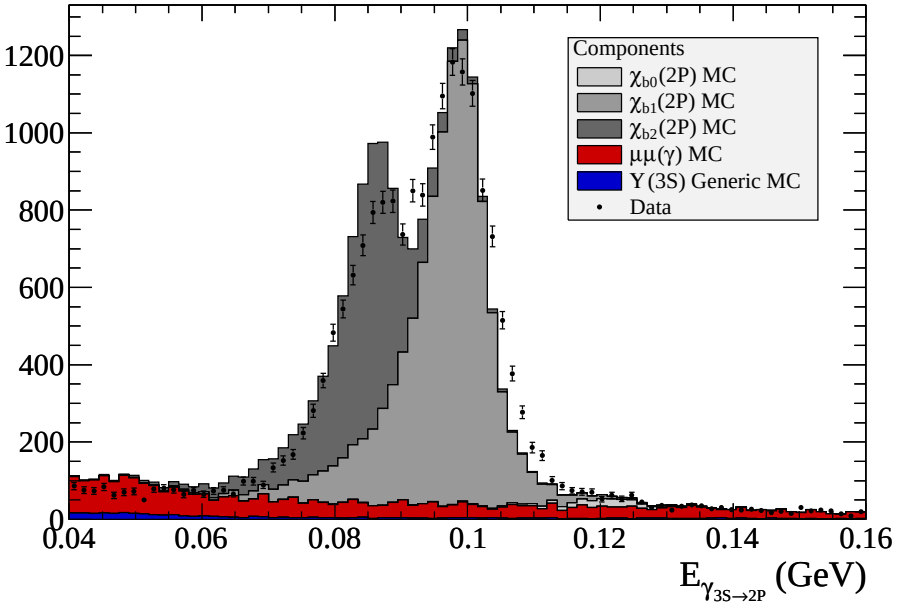


Figure 4.7: Comparison of the data spectrum to the ensemble MC spectrum with corrections for timing as described in the text. Large differences between the lineshapes and beam photon yields between data and MC are evident.

4.4 $3S \to 2P \to 2S$ fitting

Fitting of the $E_{3S\to 2P}$ spectrum follows the same general procedure as in the previous analysis. New lineshape and background parameterizations are required.

4.4.1 Lineshape parameterization

With identical reconstruction and selections, the lineshape trends as measured in the previous analysis can be expected to be largely similar. However, differences in lineshape can arise from two sources: [1] a shift in the line energies and [2] differences in signal fit probability selection due to inconsistent widths between the cascade fitter and the $2P \to 2S$ peaks. In order to obtain maximum precision we re-calculate these parameters for this cascade; this will also allow a second test of the idea that the parameter slopes are independent of shape. Fig. **??** shows the linear trends of the lineshapes in signal MC. The fitted slopes differ from the $2S \to 1P \to 1S$ case by less than one linear fit standard deviation, further supporting the hypothesis that these slopes are lineshape-independent.

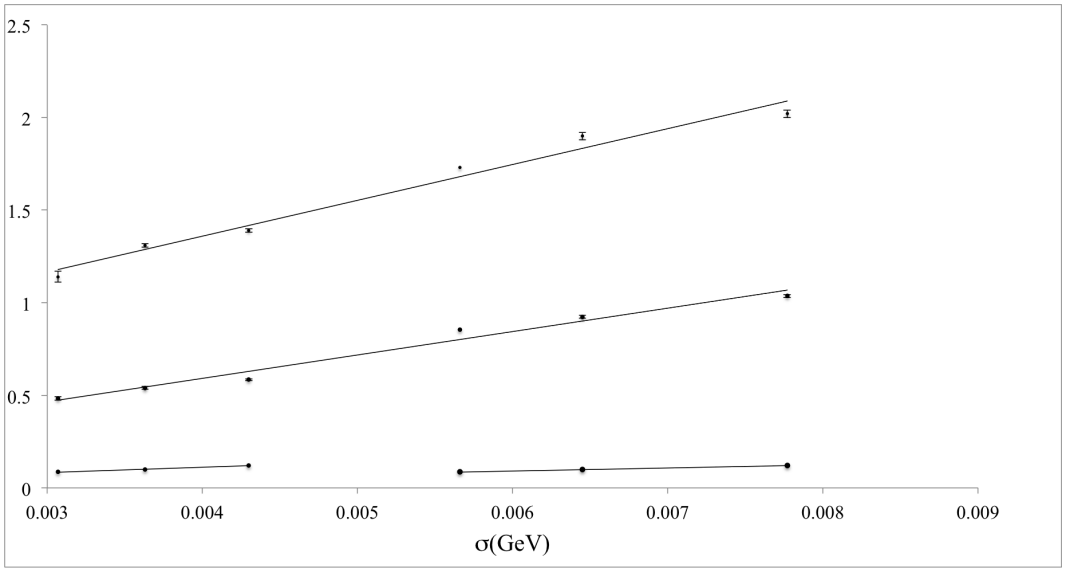


Figure 4.8: Linear fits to the lineshape parameters in $3S \to 2P \to 2S$ cascades. Compare to Fig. **??**.

4.4.2 MC background fit

Linear and exponential parameters are extracted from a fit to the background components of the MC ensemble as shown in Fig. **??**.

4.4.3 Data fit

The final data fit for the $3S \to 2P \to 2S$ analysis is shown in Fig. **??**. Although the $J = 0$ "bump" is not obvious in this view, the zoomed view [Fig. **??**] shows a clear departure from double Crystal Ball shape on the right shoulder of the $J = 1$ peak. Jitter MC studies are required to determine the true significance of this bump and whether the null hypothesis $[N_0 = 0]$ can be excluded.

Using lineshape trends and background parameters derived from the MC background fit in Fig. **??** and the lineshape parameterization in Fig. **??**, the fitted parameter constraints are shown in Table **??**. Although the parameter ranges and slopes have changed the strategy is identical to the fit to the $2S \to 1P$ spectrum in the previous analysis.

Parameter	Description	Fixed?	Value
λ	Background exponential parameter	Yes	-57
a_1	Background linear slope	Yes	-0.265
$frac$	Ratio of exponential to linear yields	NO	0.21 ± 0.03
μ_0	Gaussian mean energy for $J = 0$	NO	$121 \pm 2\,\text{MeV}$
μ_1	Gaussian mean energy for $J = 1$	NO	$98.4 \pm 0.1\,\text{MeV}$
μ_2	Gaussian mean energy for $J = 2$	NO	$85.1 \pm 0.2\,\text{MeV}$
$\sigma(J = 1)$	Gaussian width for $J = 1$	NO	$5.47 \pm 0.10\,\text{MeV}$
$\sigma(J = 0)$	Gaussian width for $J = 0$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_0 - \mu_1)$
$\sigma(J = 2)$	Gaussian width for $J = 2$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_2 - \mu_1)$
$\alpha_1(J = 1)$	Low-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	0.96 ± 0.04
$\alpha_1(J = 0)$	Low-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_1(J = 1) + 123 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_1(J = 2)$	Low-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_1(J = 1) + 123 \cdot (\sigma(J = 2) - \sigma(J = 1))$
$\alpha_2(J = 1)$	High-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	$1.6 \pm 0.1\sigma$
$\alpha_2(J = 0)$	High-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_2(J = 1) + 192 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_2(J = 2)$	High-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_2(J = 1) + 192 \cdot (\sigma(J = 2) - \sigma(J = 1))$
n_1	Low-end power-law coefficient [shared]	Yes	21
n_2	High-end power-law coefficient [shared]	Yes	5.3
$N_{J=0}$	Yield of $J = 0$ peak	NO	189 ± 55
$N_{J=1}$	Yield of $J = 1$ peak	NO	$11,380 \pm 181$
$N_{J=2}$	Yield of $J = 2$ peak	NO	$5,446 \pm 145$
N_{bkg}	Yield of background	NO	$2,769 \pm 124$
N	[Implicit parameter–total yield constraint]	Yes	$19,784$

Table 4.4: Summary of parameters from the full data fit in Fig. **??**. Fixed parameters come from high-statistics signal MC.

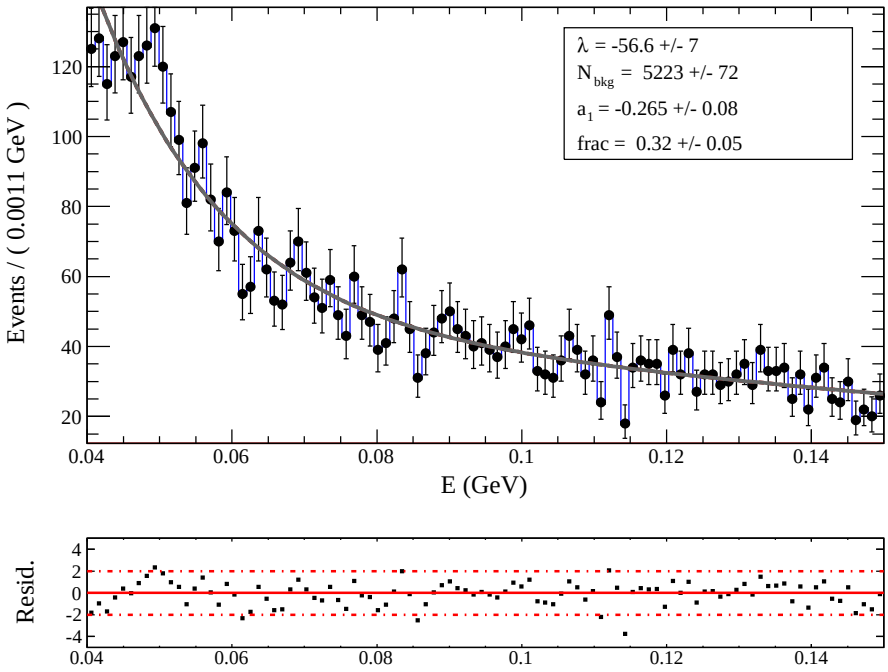


Figure 4.9: Fit to the background components of the MC ensemble in reconstructed $3S \to 2P \to 2S$ cascades using only $\Upsilon(3S)$ generic decays. The ratio between the exponential and linear contributions is allowed to float. Below 50 MeV in MC the smooth background seems to be disrupted; this has a negligible effect on the final data fit.

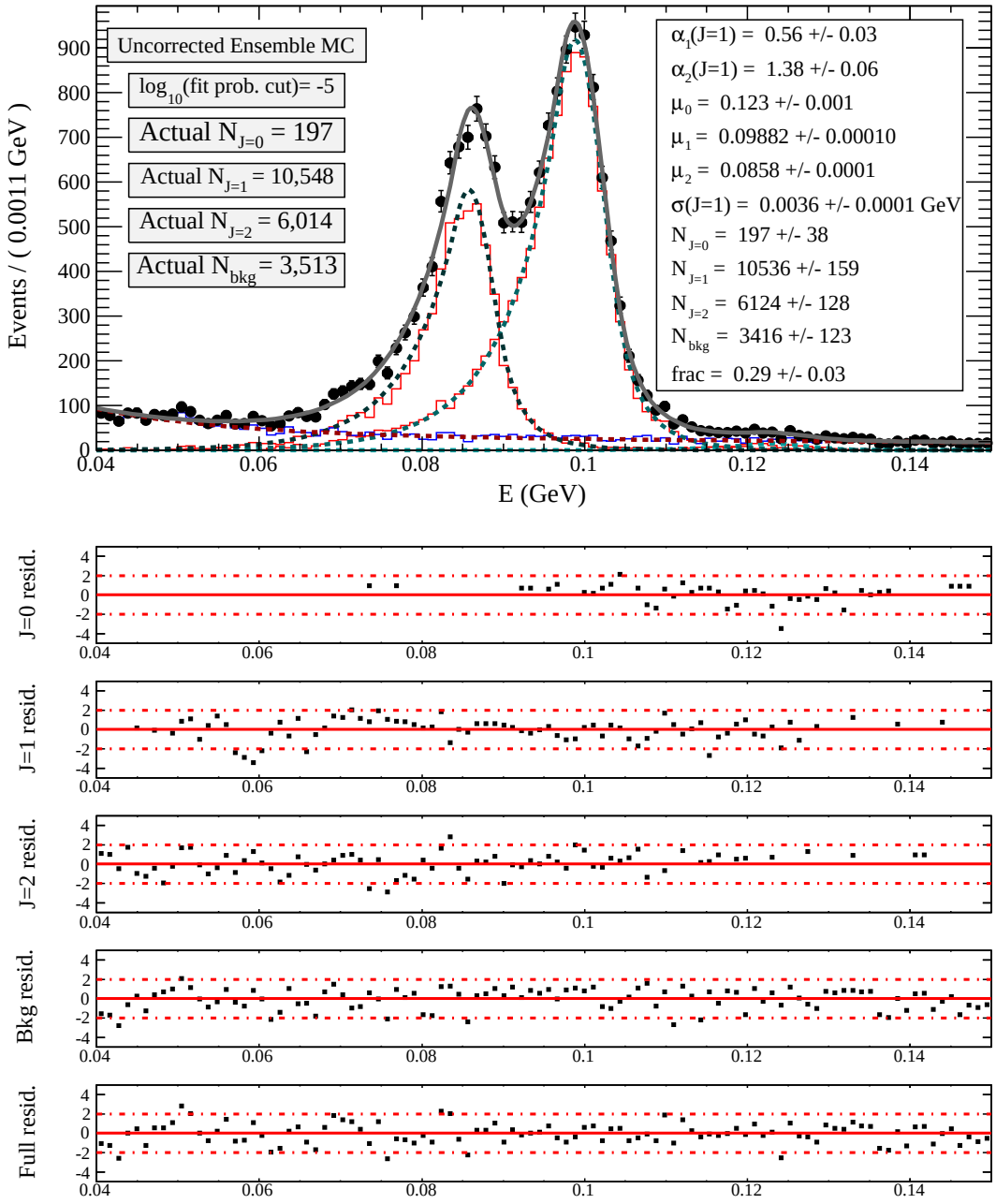


Figure 4.10: Fit to the full $3S \to 2P \to 2S$ uncorrected MC ensemble with component-by-component normalized residuals.

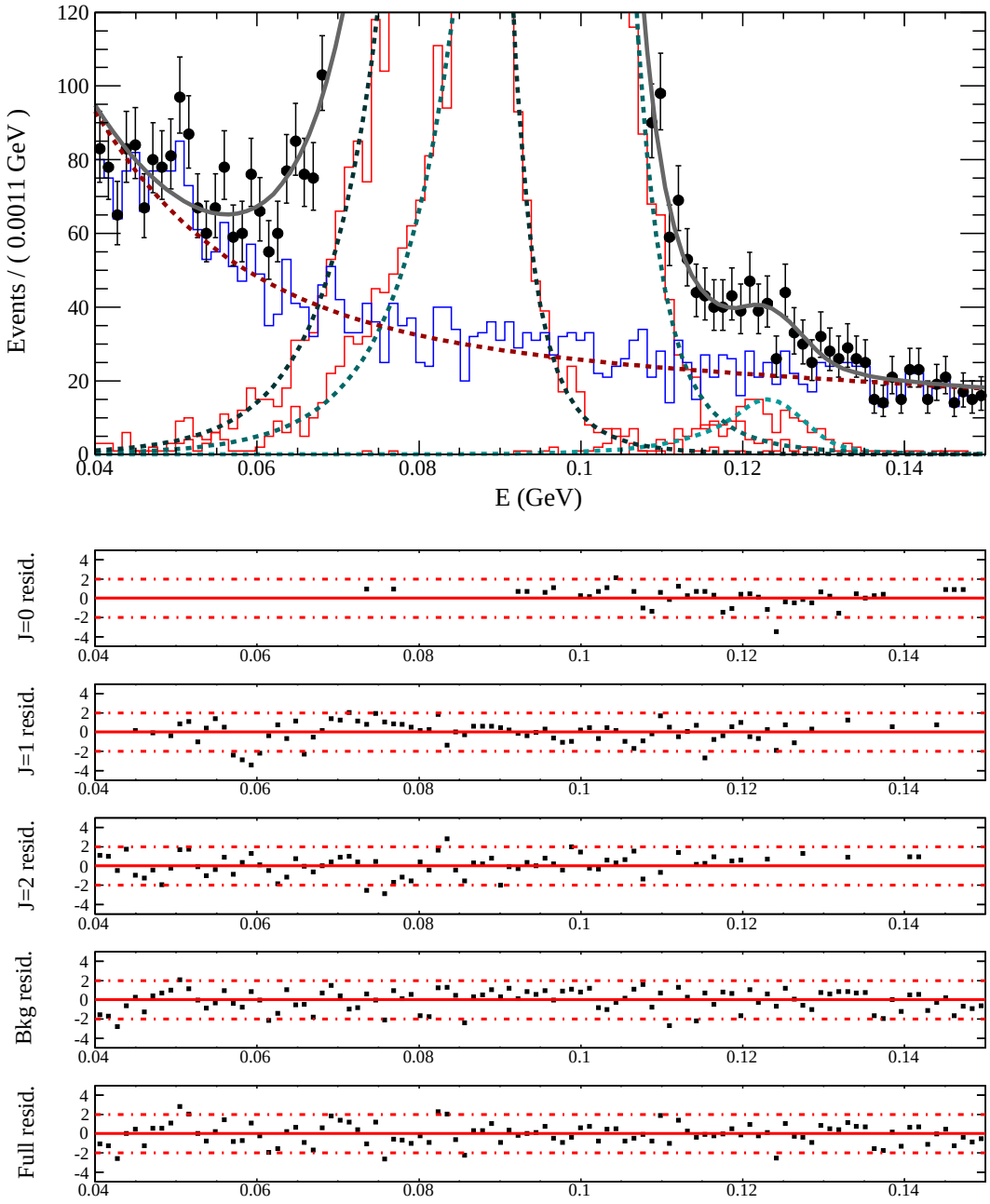


Figure 4.11: Zoomed view of Fig. **??**. Background MC histogram is in blue, signal MC are the red histograms. The dashed lines correspond to the fitted components.

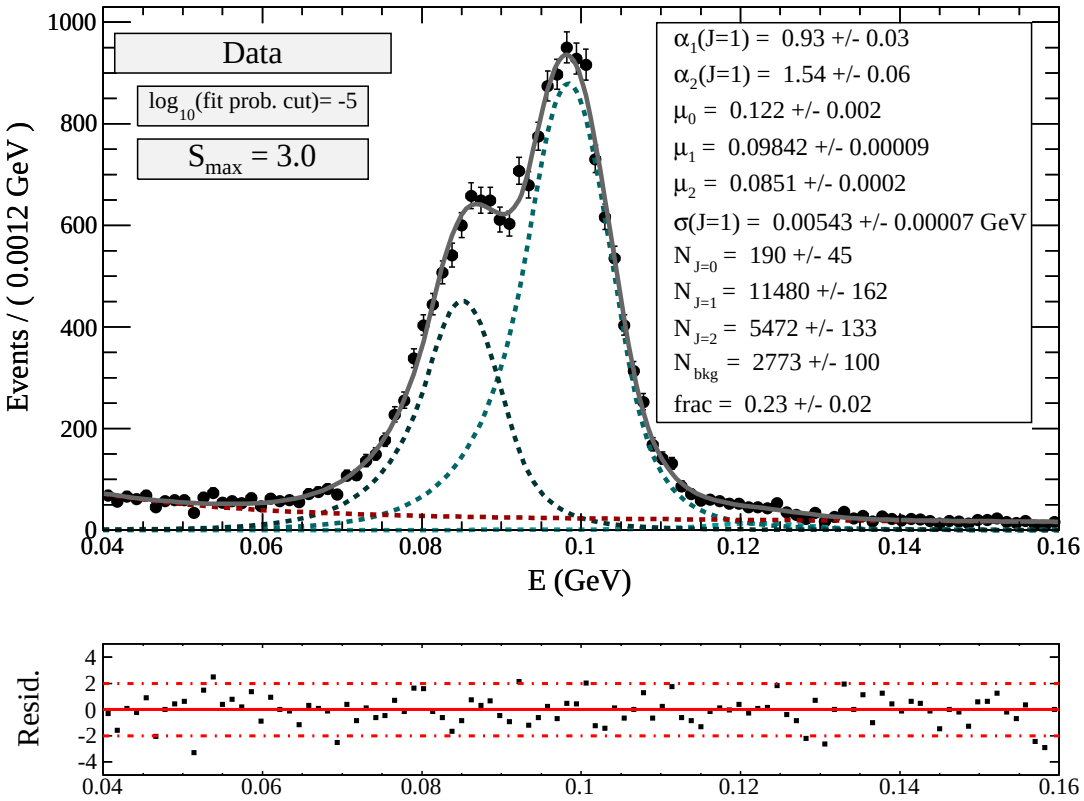


Figure 4.12: Fit to $3S \to 2P \to 2S$ data with final selections.

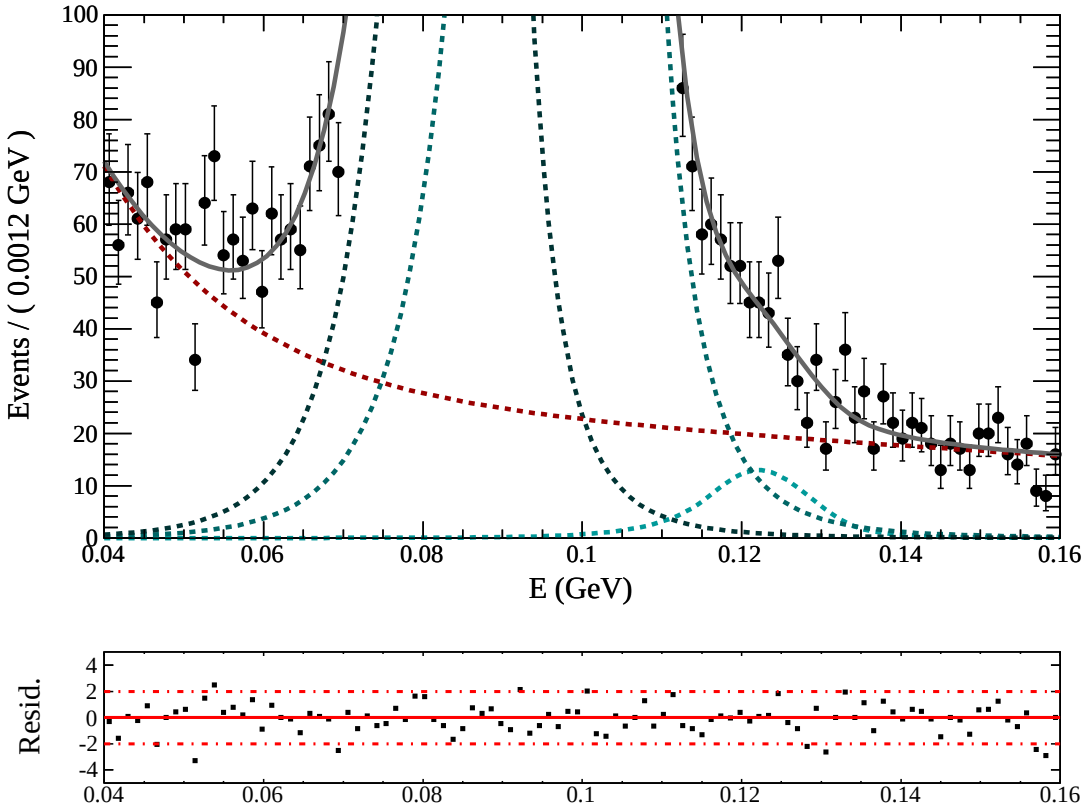


Figure 4.13: The final $3S \to 2P \to 2S$ data fit zoomed to show the $J = 0$ peak.

4.5 $3S \to 2P \to 2S$ systematics

Toy studies using jitter MC derived from the data fit are used to calculate systematic errors and corrections to the fit yields as before. The jitter ranges of the PDF parameters used to generate the toy spectra are shown in Table **??**. The results of the fits to the $50,000$ jitter toys are shown in Fig. **??** for the yield parameters and in Fig. **??** for the means.

Selection of toy fits is more stringent than in the $2S \to 1P \to 1S$ jitter toy studies due to the fact that the errors of the three fitted yields in the data spectrum are on the high side of the peaks extracted from the toys. The set of included jitter spectra can be narrowed to the subset that fit to similar yield precision as the data spectrum by excluding low-error fits. These fits are presumably associated with a sub-class of spectra that have better peak separation than the data fit, therefore generality is not compromised. Peak separation is determined by a combination of peak mean, width and shape parameters and cannot easily be selected for explicitly. Table **??** shows the selections made to the set of toy fits which were identified as successful fits with high-quality covariance matrices. The yield errors are chosen to be within ± 2 times the square root of the data fit yield errors; this is an arbitrary choice.

Jitter corrections are largely similar to those calculated for the $2S \to 1P \to 1S$ analysis. This consistency is an indication that the measured yield biases are inherent to the fitting strategy and confirms that the yield corrections are "real."

Parameter	Selection	Low value	High value
$\alpha_1(J=1)$	Data fit $\pm 2\sigma_{\alpha_1}$	0.87	0.99
$\alpha_2(J=1)$	Data fit $\pm 2\sigma_{\alpha_2}$	1.42	1.66
$\sigma(J=1)$	Data fit $\pm 2\sigma_{\sigma}$	5.29 MeV	5.57 MeV
N_0	Data fit $\pm 2\sigma_{N_0}$	100	280
N_1	Data fit $\pm 2\sigma_{N_1}$	11,156	11,804
N_2	Data fit $\pm 2\sigma_{N_2}$	5,206	5.738
N_{bkg}	Data fit $\pm 2\sigma_{N_{bkg}}$	2,573	2,973
σ_{N_0}	Data fit $\pm 2\sqrt{\sigma N_0}$	32	58
σ_{N_1}	Data fit $\pm 2\sqrt{\sigma N_1}$	136	187
σ_{N_2}	Data fit $\pm 2\sqrt{\sigma N_2}$	110	156

Table 4.5: Summary of selections made on fitted parameters from the $3S \to 2P \to 2S$ jitter toys.

Parameter	Nominal value source	Nom. value	Param. σ	Jitter range
λ	Fit to MC ensemble background	-57		[Not jittered]
a_1	Fit to data sidebands	-0.46		Nominal $\pm 3\sigma_{a_1}$
$frac$	Fit to data	0.21	0.03	Nominal $\pm 3\sigma_{frac}$
μ_0	Fit to data	122 MeV	2	Nominal $\pm 3\sigma_{\mu_0}$
μ_1	Fit to data	98.42 MeV	0.09	Nominal $\pm 3\sigma_{\mu_1}$
μ_2	Fit to data	85.1	0.2	Nominal $\pm 3\sigma_{\mu_2}$
$[m_\sigma]$	Signal MC and EMC res. param.			0.01–0.08, about $\pm 5\sigma$
$\sigma(J=1)$	Fit to data	5.43 MeV	0.07 MeV	Nominal $\pm 3\sigma_{\sigma(J=1)}$
$\sigma(J=0)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_0 - \mu_1)$
$\sigma(J=2)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_2 - \mu_1)$
$[m_{\alpha_1}]$	MC parameter fit	123.5	9.2	
$\alpha_1(J=1)$	Fit to data	0.93	0.03	Nominal $\pm 3\sigma_{\alpha_1}$
$\alpha_1(J=0)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_1(J=2)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=2) - \sigma(J=1))$
$[m_{\alpha_2}]$	MC parameter fit	191.8	17.1	
$\alpha_2(J=1)$	Fit to data	1.54	0.06	Nominal $\pm 3\sigma_{\alpha_2}$
$\alpha_2(J=0)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_2(J=2)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=2) - \sigma(J=1))$
n_1	Manual			5.0–100.0
n_2	Manual			5.0–100.0
$N_{J=0}$	Fit to data	190	45	Nominal $\pm 3\sigma_{N_{J=0}}$
$N_{J=1}$	Fit to data	11,480	162	Nominal $\pm 3\sigma_{N_{J=1}}$
$N_{J=2}$	Fit to data	5,472	133	Nominal $\pm 3\sigma_{N_{J=2}}$
N_{bkg}	Fit to data	2,773	100	Nominal $\pm 3\sigma_{N_{bkg}}$

Table 4.6: Summary of parameters used to generate the $3S \to 2P \to 2S$ Jitter MC spectra.

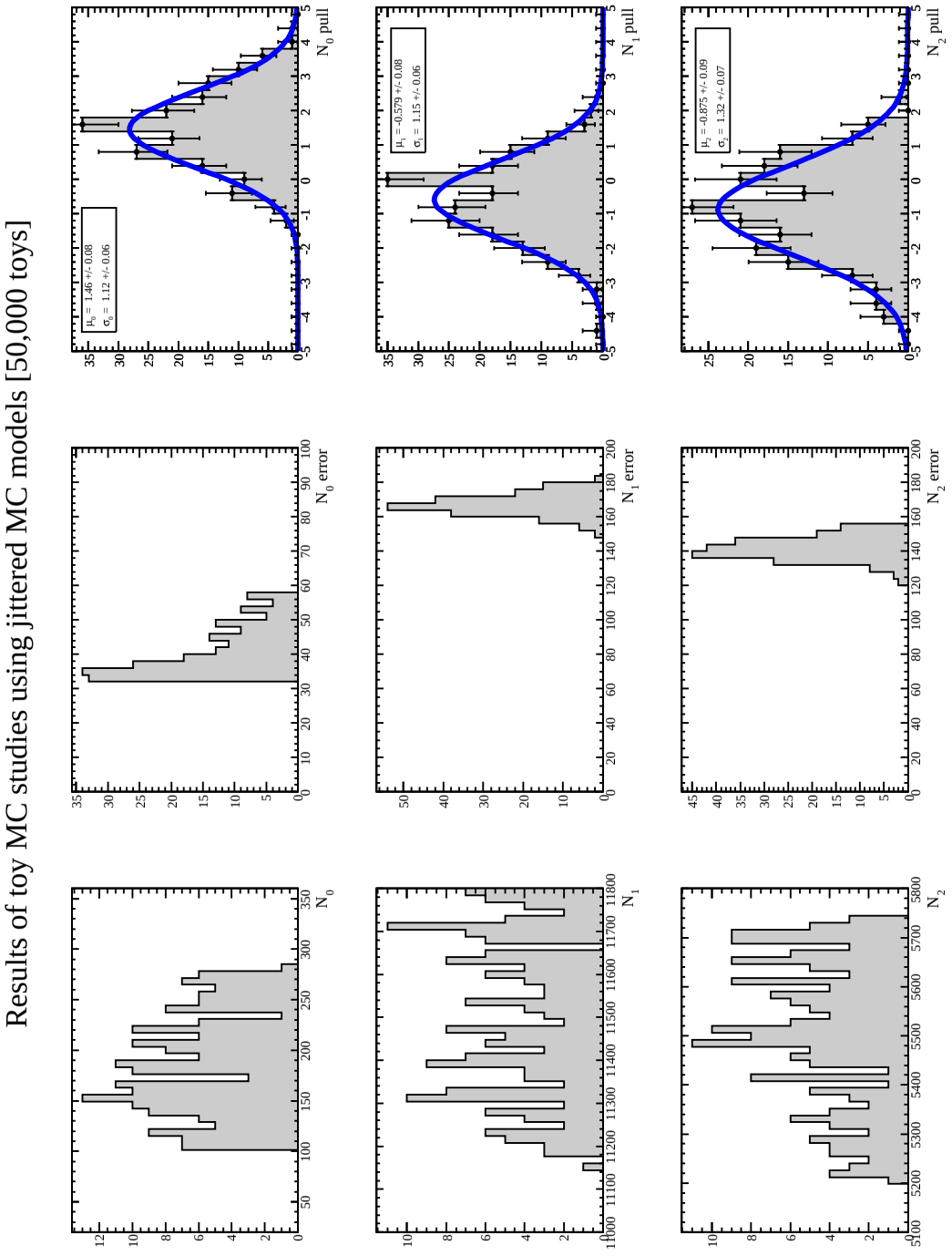


Figure 4.14: The results of 50,000 fits to the $3S \to 2P \to 2S$ jitter MC spectra.

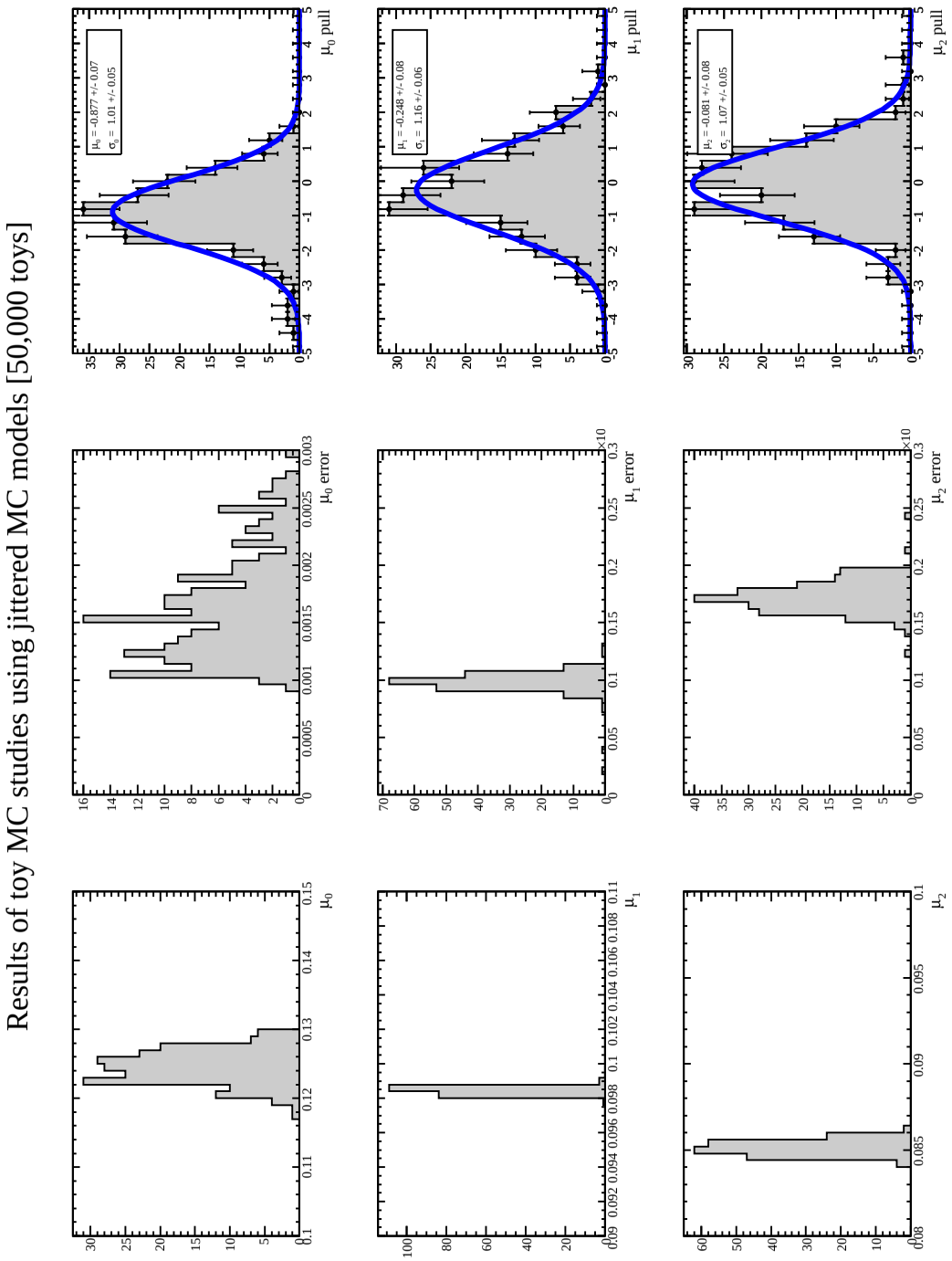


Figure 4.15: The results of 50,000 fits to the jitter MC spectra for the Gaussian mean parameters $\mu_{0,1,2}$.

4.6 $3S \to 2P \to 2S$ results

Using the yield corrections obtained using the jitter MC, we present the product branching fraction ratio results from the $3S \to 2P \to 2S$ analysis:

$$\mathscr{F}^{0/1}_{3S\to 2P\to 2S} = \frac{256 \pm 50}{11386 \pm 186} \times (1.059 \pm 0.005) = (2.38 \pm 0.47)\%, \tag{4.1}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 2S} = \frac{5356 \pm 176}{11384 \pm 186} \times (0.988 \pm 0.004) = (47.0 \pm 1.7)\%, \tag{4.2}$$

The mass splittings are measured to be:

$$\Delta M^{2P}_{1-0} = (121.0 \pm 2.0 - 98.4 \pm 0.1)\,\text{MeV} = 23.7 \pm 2.0\,\text{MeV}, \tag{4.3}$$

$$\Delta M^{2P}_{2-1} = (98.4 \pm 0.1 - 85.1 \pm 0.2)\,\text{MeV} = 13.31 \pm 0.24\,\text{MeV}. \tag{4.4}$$

Comparable results derived from individual branching fractions as reported in the PDG are:

$$\mathscr{F}^{0/1}_{3S\to 2P\to 2S} = (10.8 \pm 5.3)\%, \tag{4.5}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 2S} = (55 \pm 17)\%, \tag{4.6}$$

$$\mathscr{F}^{0/2}_{3S\to 2P\to 2S} = (19 \pm 11)\%. \tag{4.7}$$

$$\Delta M^{2P}_{1-0} = 22.96 \pm 0.84\,\text{MeV}, \tag{4.8}$$

$$\Delta M^{2P}_{2-1} = 13.19 \pm 0.77\,\text{MeV}. \tag{4.9}$$

See Chapter 6 for a discussion of the results.

Chapter 5

$3S \to 2P \to 1S$

Analysis of the $3S \to 2P \to 1S$ cascade (Fig. **??**) is very similar to the $3S \to 2P \to 2S$ case. Discussion here proceeds where significant differences between the two analyses is evident.

5.1 $3S \to 2P \to 1S$ event selection and reconstruction

Reconstruction and event selection proceeds identically to the $3S \to 2P \to 2S$ analysis, including the "new" timing-based best cascade candidate selection. Table **??** shows the signal MC reconstruction efficiencies.

The soft and hard photons are defined in candidate selection according to $40 < E_{3S\to 2P} < 160\,\mathrm{MeV}$ and $620 < E_{2P\to 1S} < 820\,\mathrm{MeV}$. The soft photon covers the same transition as in the $3S \to 2P \to 2S$ analysis. However, the much-higher energy of the hard photons in this analysis change both the background sources and soft photon lineshapes after signal fit probability selection relative to that analysis.

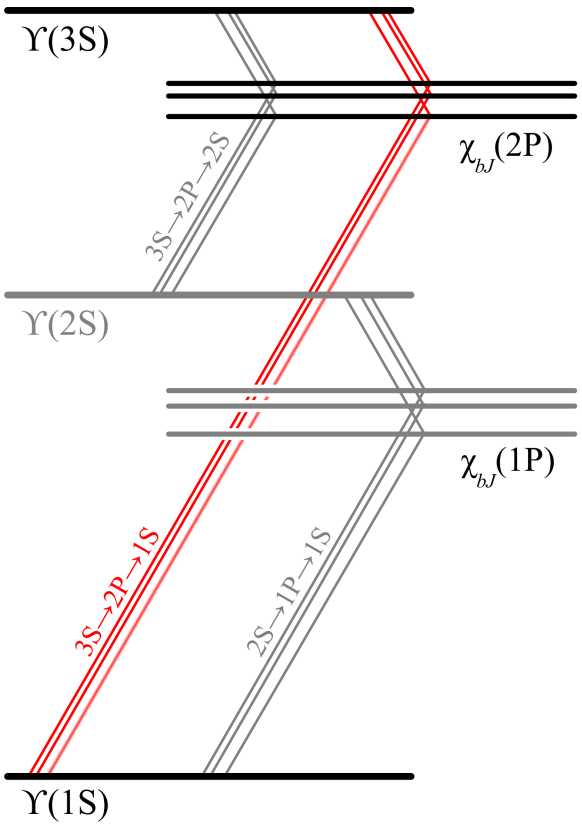


Figure 5.1: The decay mode evaluated in this chapter.

SP mode #	Mode	ϵ_{reco}
8816	$\chi_{b0}(2P)$ cascade	0.227 ± 0.009
8820	$\chi_{b1}(2P)$ cascade	0.239 ± 0.008
8824	$\chi_{b2}(2P)$ cascade	0.228 ± 0.008

Table 5.1: Reconstruction efficiencies for the three $3S \to 2P \to 1S$ signal MC modes.

5.2 $3S \to 2P \to 1S$ background

.The $3S \to 2P \to 1S$ cascade shares a first leg with the already-described $3S \to 2P \to 2S$ cascade. However, the very different energy ranges of $\gamma_{2P\to 2S}$ and $\gamma_{2P\to 1S}$ — $160 - 280\,\mathrm{MeV}$ vs. $600 - 900\,\mathrm{MeV}$, respectively — change the balance of the important background modes.

5.2.1 Backgrounds in MC

Figs. **??** and **??** show the MC mothers of $\gamma_{3S\to 2P}$ and $\gamma_{2P\to 1S}$ in generic $\Upsilon(3S)$ decays through the $\Upsilon(1S)$. Some differences in the background contributions relative to the $3S \to 2P \to 2S$ case are evident, most notably the absence of beam photons in the $2P \to 1S$ transition and large contribution from FSR.

Background source: $\pi^0\pi^0$

There cannot be any contribution from $\Upsilon(3S) \to \pi^0\pi^0\Upsilon(2S)$ since the $\Upsilon(2S)$ lies below the χ_{bJ} states; no source for the $\gamma_{2P\to 1S}$ candidate can exist kinematically. The transition $\Upsilon(3S) \to \pi^0\pi^0\Upsilon(1S)$ is the only $\pi^0\pi^0$ mode relevant, but the huge disparity between the energies of the two cascade photons makes reconstruction of these events extremely unlikely.

Background source: FSR

With no beam photons in the $\gamma_{2P\to 1S}$ energy range, all of the photon candidates in the generic sample are in-time and almost all are from QED bremsstrahlung, as modeled by `PHOTOS`. The $\gamma_{3S\to 2P}$ candidates in these events are almost entirely out-of-time, making timing selection the only reliable way of rejecting events with FSR candidates, whether they are from $\Upsilon \to \gamma_{FSR}\mu\mu$ decays or $\mu\mu(\gamma)$ events.

Recall that the $\mu\mu$ pair are required to come from the interaction point in the `TreeFitter` kinematic fit of the cascade. The high energy of the FSR photon in $\mu\mu(\gamma)$ events alters the kinematics of the $\mu\mu$ pair to a degree that the pair no longer appear to come from the interaction point; the fit probability cut is expected to be significantly more effective at removing $\mu\mu(\gamma)$ background in this analysis than in the previous two. Comparison of Figs. **??** and **??** show this to be true in MC.

5.2.2 $3S \to 2P \to 1S$ MC ensemble

As before, we construct an MC Ensemble to model the signal spectrum. The only potential implication for signal extraction is in obtaining the correct form for the background shape. Table **??** shows the relevant scaling factors; note the expected extremely low $\mu\mu(\gamma)$ contribution.

Fig. 5.6 shows the data spectrum compared to the MC ensemble. The background is evidently not well-modeled in MC. Although it's not entirely clear what the source of the discrepancy is, it is likely caused by poor modeling of the FSR kinematics in `PHOTOS`. Note that the $\chi_{bJ}(2P) \to \Upsilon(1S)$ branching fractions are the poorest-known of all those used in these analyses, so the yield differences between MC and data are not surprising.

SP #	Mode	# MC	# Run 7	Scale	# reconstructed	Scaled #
8739	Generic $\Upsilon(3S)$ decays:	$450.5M$	$122M$	0.27	$3,284$	887
	$\Upsilon(3S) \to \pi^0\pi^0\Upsilon(2S)$				183	49
	$\Upsilon(3S) \to \mu\mu$				$2,953$	797
	$\Upsilon(3S) \to \eta\Upsilon(1S)$				102	28
	All other generic				46	Negligible
11502	Full $\chi_{b0}(2P)$ cascade	$162k$	$1.6k$	0.010	$36,270$	363
11501	Full $\chi_{b1}(2P)$ cascade	$162k$	$35k$	0.216	$38,193$	$8,250$
11500	Full $\chi_{b2}(2P)$ cascade	$162k$	$28k$	0.171	$36,581$	$6,255$
3981	$\mu\mu(\gamma)$	$39.2M$	$32M$	0.82	8	Negligible

Table 5.2: The constitution of the $3S \to 2P \to 1S$ MC ensemble. See **??** for discussion.

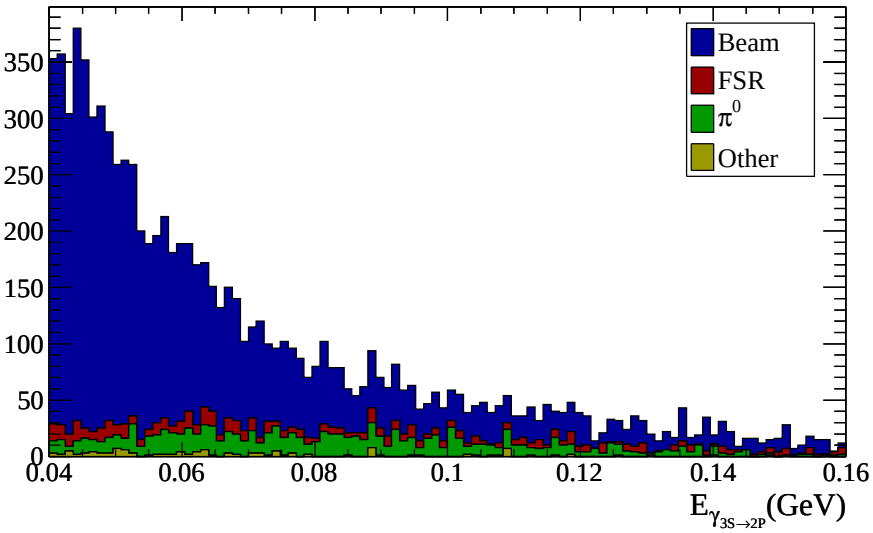


Figure 5.2: Mothers of $\gamma_{3S\to 2P}$ candidates in generic $\Upsilon(3S)$ MC. Cascade candidates are required to have a signal fit probability above zero but multiple candidates are allowed in each event.

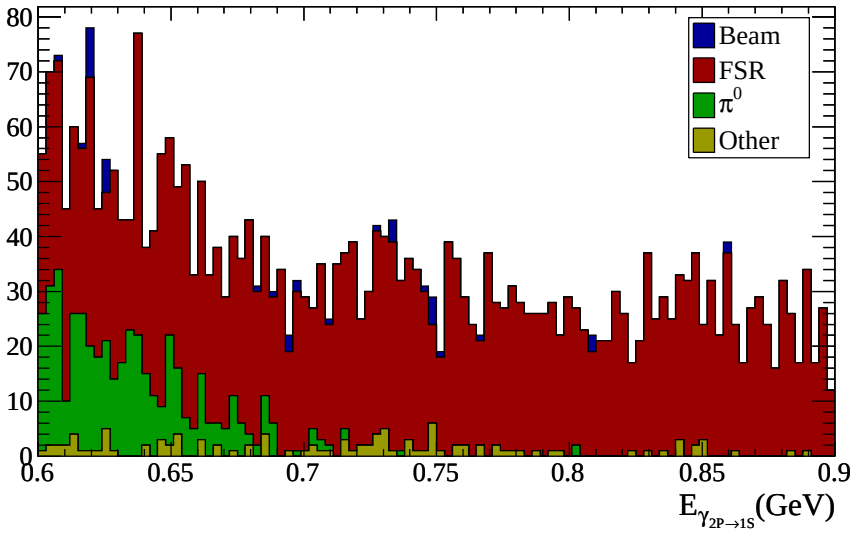


Figure 5.3: Mothers of $\gamma_{2P\to 1S}$ candidates in generic $\Upsilon(3S)$ MC.

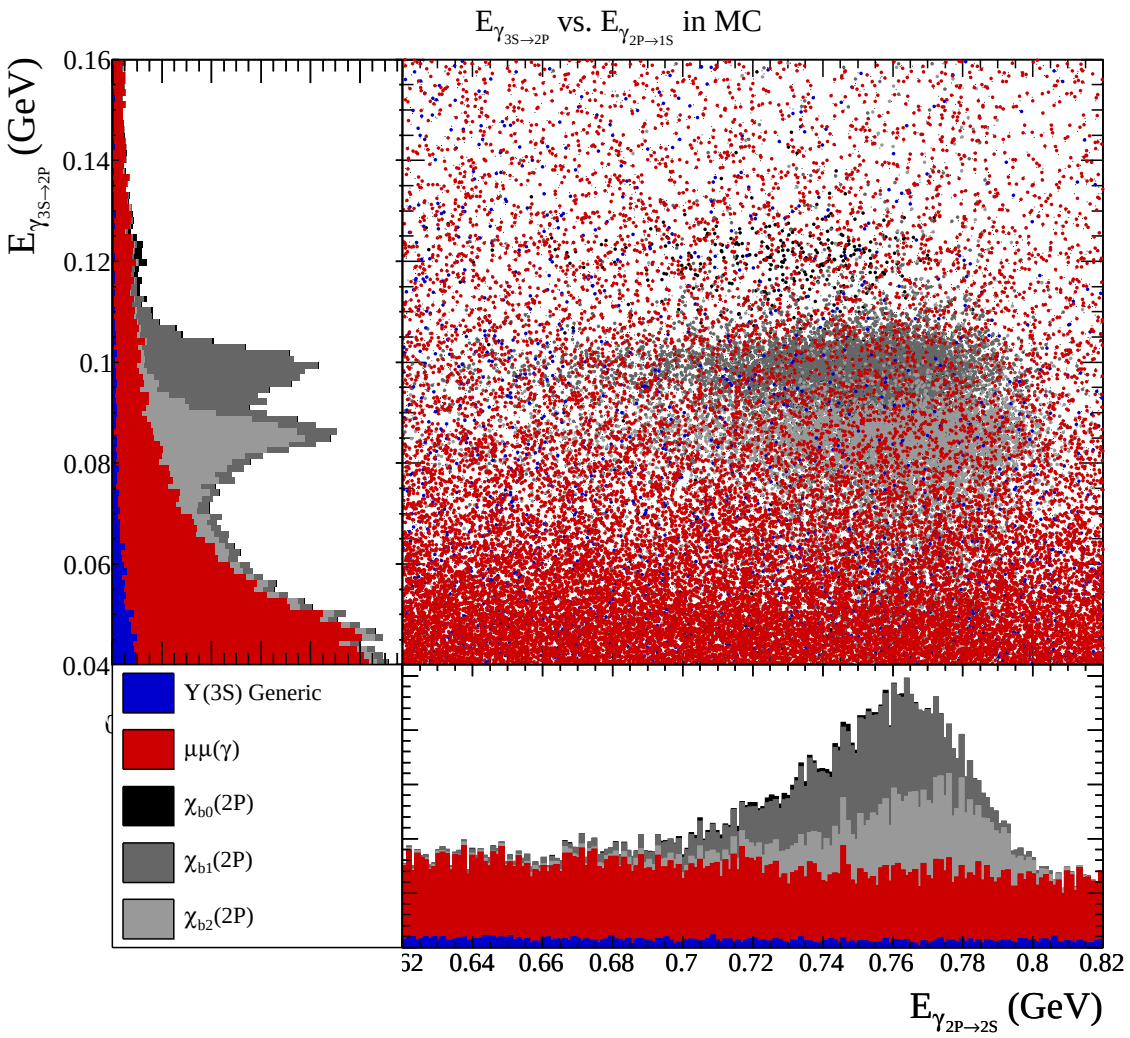


Figure 5.4: Scatterplot of $E_{\gamma_{3S\to 2P}}$ vs $E_{\gamma_{2P\to 2S}}$ in $3S \to 2P \to 1S$ ensemble MC, showing the distribution of the main background sources.

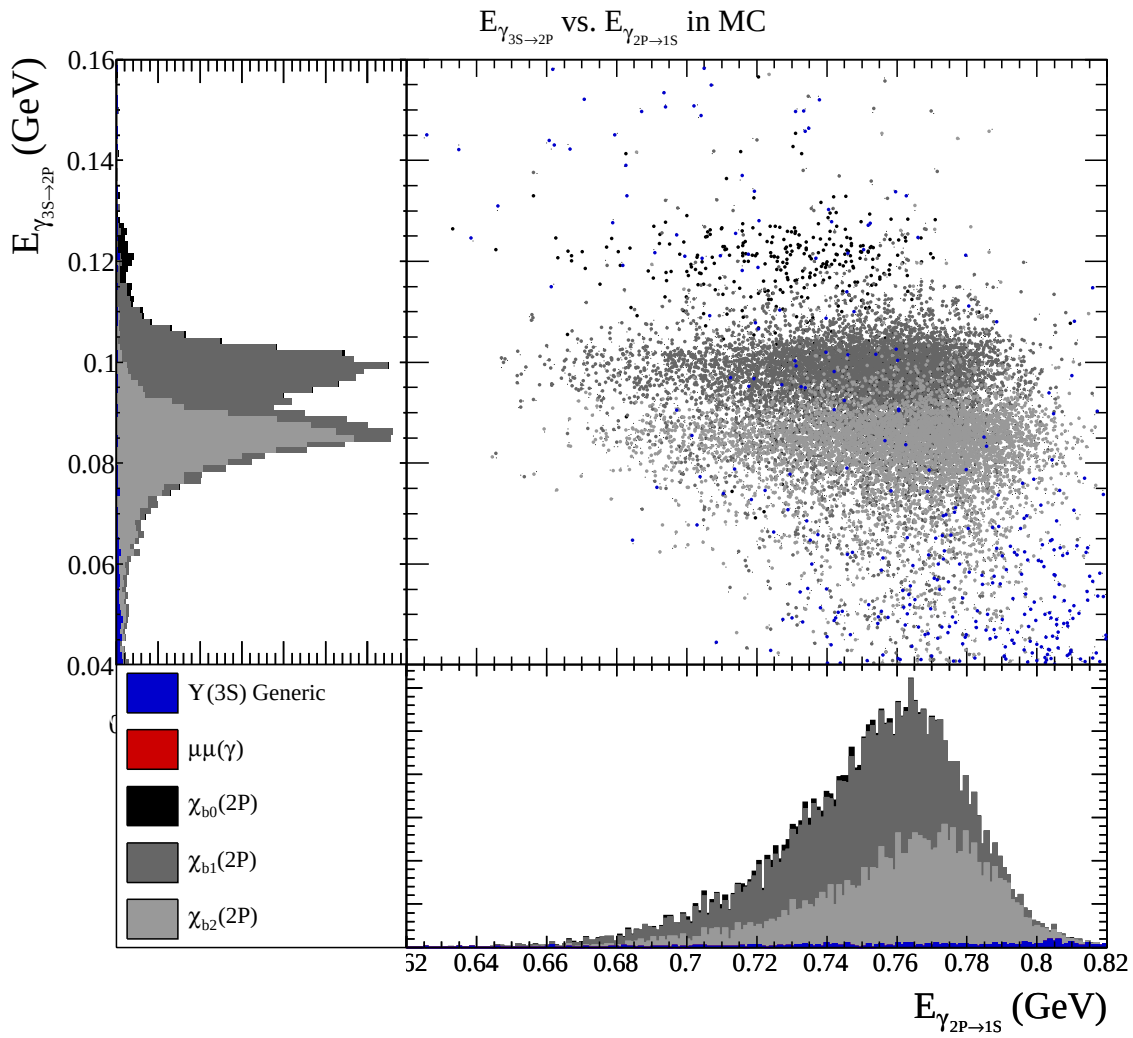


Figure 5.5: Scatterplot of $E_{\gamma_{3S\to 2P}}$ vs $E_{\gamma_{2P\to 2S}}$ in $3S \to 2P \to 1S$ ensemble MC including only candidates with signal fit probability greater than 10^{-5}.

Selection	$\epsilon_{J=1}/\epsilon_{J=0}$	$\epsilon_{J=1}/\epsilon_{J=2}$	$\epsilon_{J=2}/\epsilon_{J=0}$
Reco. and initial selections	1.052 ± 0.008	1.045 ± 0.004	1.006 ± 0.008
Signal fit probability	1.0068 ± 0.0002	0.982 ± 0.002	1.025 ± 0.002
Total	1.059 ± 0.009	1.026 ± 0.009	1.031 ± 0.009

Table 5.3: Efficiency ratios for the three signal modes from MC in $3S \to 2P \to 1S$.

5.3 $3S \to 2P \to 1S$ cuts

Selections are similar to those used in the $3S \to 2P \to 2S$ analysis. The fit probability is required to be above 10^{-5} as usual. However, an adjustment in the timing selection due to the large differences between the energies of the second-leg photons gives best cut results.

5.3.1 Timing cut

Calculation of the in-and out-of-time efficiencies in timing significance selection in Appendix C utilized a proxy mode tailored to the $3S \to 1P \to 1S$ cascade. Inaccuracies in calculating the cluster time error are likely to arise in analyses with much higher-energy photons such as this one. These inaccuracies will change the exact form of the photon efficiency curves. Although a value of $S_{max} = 2.0$ worked well for both $2S \to 1P \to 1S$ and $3S \to 2P \to 2S$, we find that $S_{max} = 3.0$ is more suitable for this analysis [see Fig. **??**]. The calculated cluster timing error for high energy photons is apparently lower than it should be for consistency.

Optimizing the S_{max} selection allows for better signal efficiency and similar background rejection while maintaining complete generality of the spectrum. As in the previous analysis, the timing selection is performed prior to selecting the best cascade candidate in an event.

Using the new timing selection, the MC ensemble compared to the data spectrum is shown in Fig. **??**.

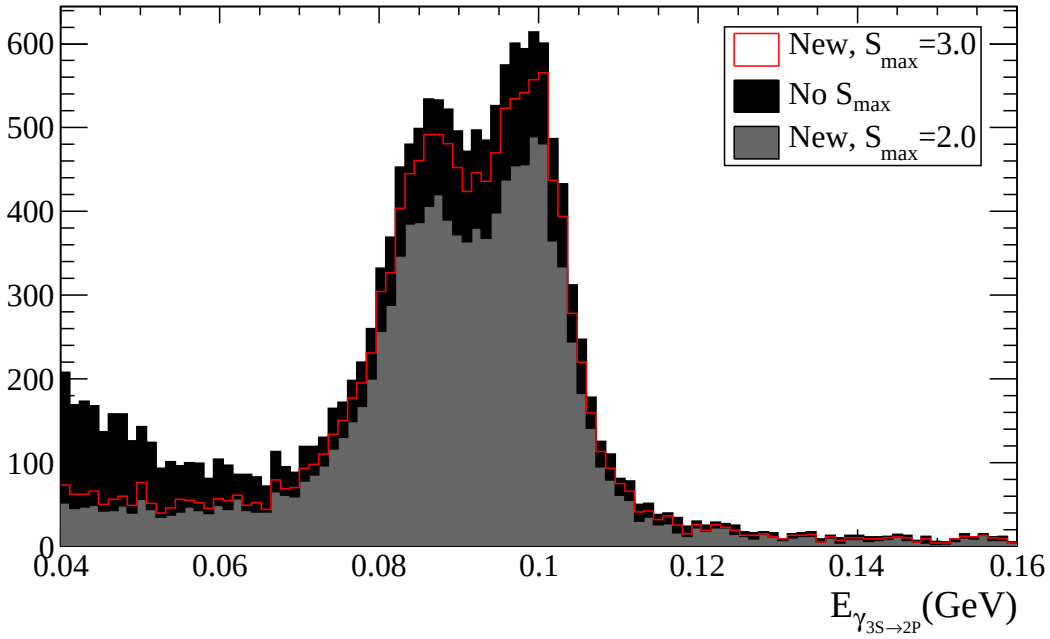


Figure 5.6: Comparison of the $\gamma_{3S\to 2P}$ spectrum in $3S \to 2P \to 1S$ data with no S_{max} selection and selections of $S_{max} = 2.0$ and 3.0, with the "new" timing cut referring to selection on S_{max} before best cascade candidate choice as described in the text. Best signal/background discrimination is achieved with the $S_{max} = 3.0$ selection. Signal fit probability is required to be above 10^{-5} for each histogram.

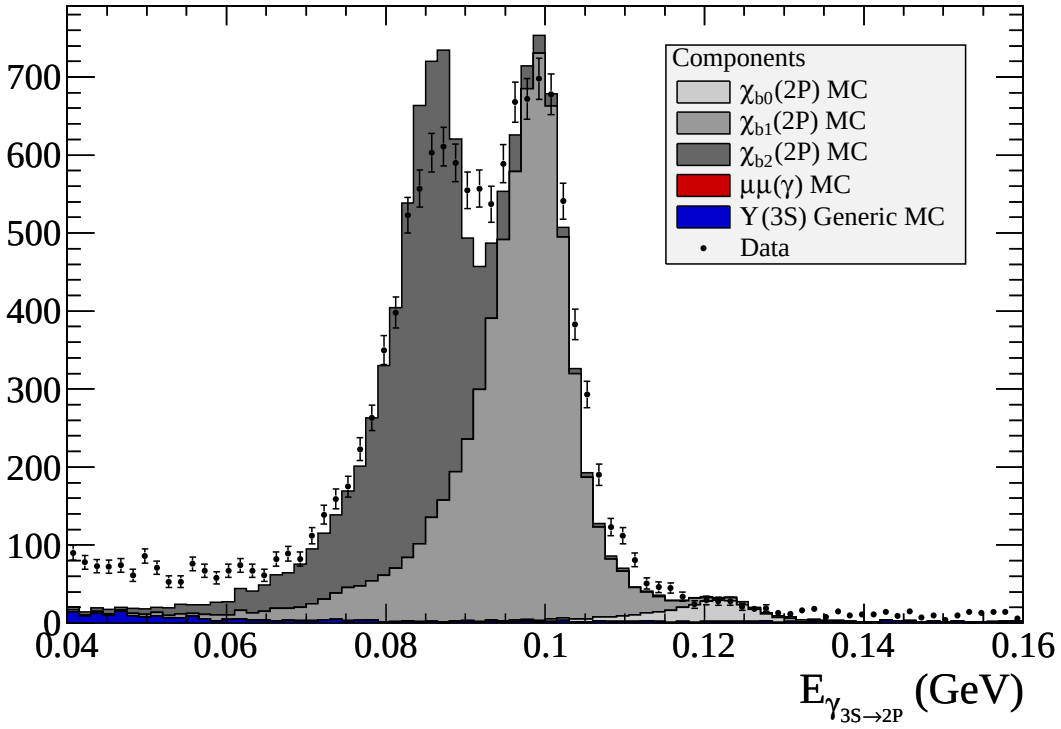


Figure 5.7: Comparison of the $3S \to 2P \to 1S$ data spectrum with the MC ensemble.

5.4 $3S \to 2P \to 1S$ fitting

Fitting proceeds according to the model established by the $2S \to 1P \to 1S$ analysis, with minor differences discussed below.

5.4.1 Lineshape parametrization

Although the $E_{\gamma_{3S\to 2P}}$ spectrum includes signal photons from exactly the same transition as in the $3S \to 2P \to 1S$ analysis, the lineshapes are not necessarily identical; the much broader high-energy photons in the $3S \to 2P \to 1S$ cascade change the nature of the signal fit probability selection's effect on lineshape. Lineshape trend parametrization follows the same procedure here as before, with results shown in Fig. **??**. The extracted lineshape slopes are nearly identical to the $3S \to 2P \to 2S$ case but with significantly different offsets. Once again this result demonstrates the robustness of the slope extraction method to changes in energy, resolution and lineshape.

5.4.2 MC background fit

Linear and exponential parameters are extracted from a fit to the background components of the MC ensemble as shown in Fig. **??** as before.

5.4.3 Data fit

The data fit proceeds as before, with parameters and constraints shown in Table **??**. Fit results are shown in Figs. **??** and **??**. In this analysis the $J = 0$ peak is not clear visually; the 2.0σ signal may or may not exist. Jitter MC is required to determine if this result is inconsistent with no signal.

Parameter	Description	Fixed?	Value
λ	Background exponential parameter	Yes	-71
a_1	Background linear slope	Yes	0.0
$frac$	Ratio of exponential to linear yields	NO	0.48 ± 0.03
μ_0	Gaussian mean energy for $J = 0$	NO	$120 \pm 3\,\text{MeV}$
μ_1	Gaussian mean energy for $J = 1$	NO	$98.4 \pm 0.1\,\text{MeV}$
μ_2	Gaussian mean energy for $J = 2$	NO	$85.6 \pm 0.2\,\text{MeV}$
$\sigma(J = 1)$	Gaussian width for $J = 1$	NO	$5.1 \pm 0.1\,\text{MeV}$
$\sigma(J = 0)$	Gaussian width for $J = 0$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_0 - \mu_1)$
$\sigma(J = 2)$	Gaussian width for $J = 2$	Yes	$\sigma(J = 1) + 0.04 \cdot (\mu_2 - \mu_1)$
$\alpha_1(J = 1)$	Low-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	0.86 ± 0.03
$\alpha_1(J = 0)$	Low-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_1(J = 1) + 127 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_1(J = 2)$	Low-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_1(J = 1) + 127 \cdot (\sigma(J = 2) - \sigma(J = 1))$
$\alpha_2(J = 1)$	High-side transition for $J = 1$ (in multiples of $\sigma(J = 1)$)	NO	$1.55 \pm 0.09\sigma$
$\alpha_2(J = 0)$	High-side transition for $J = 0$ (in multiples of $\sigma(J = 0)$)	Yes	$\alpha_2(J = 1) + 200 \cdot (\sigma(J = 0) - \sigma(J = 1))$
$\alpha_2(J = 2)$	High-side transition for $J = 2$ (in multiples of $\sigma(J = 2)$)	Yes	$\alpha_2(J = 1) + 200 \cdot (\sigma(J = 2) - \sigma(J = 1))$
n_1	Low-end power-law coefficient [shared]	Yes	4.6
n_2	High-end power-law coefficient [shared]	Yes	4.4
$N_{J=0}$	Yield of $J = 0$ peak	NO	74 ± 37
$N_{J=1}$	Yield of $J = 1$ peak	NO	$7,077 \pm 141$
$N_{J=2}$	Yield of $J = 2$ peak	NO	$4,695 \pm 129$
N_{bkg}	Yield of background	NO	$1,555 \pm 74$
N	[Implicit parameter–total yield constraint]	Yes	13,401

Table 5.4: Summary of parameters from the full data fit in Fig. **??**. Fixed parameters come from high-statistics signal MC.

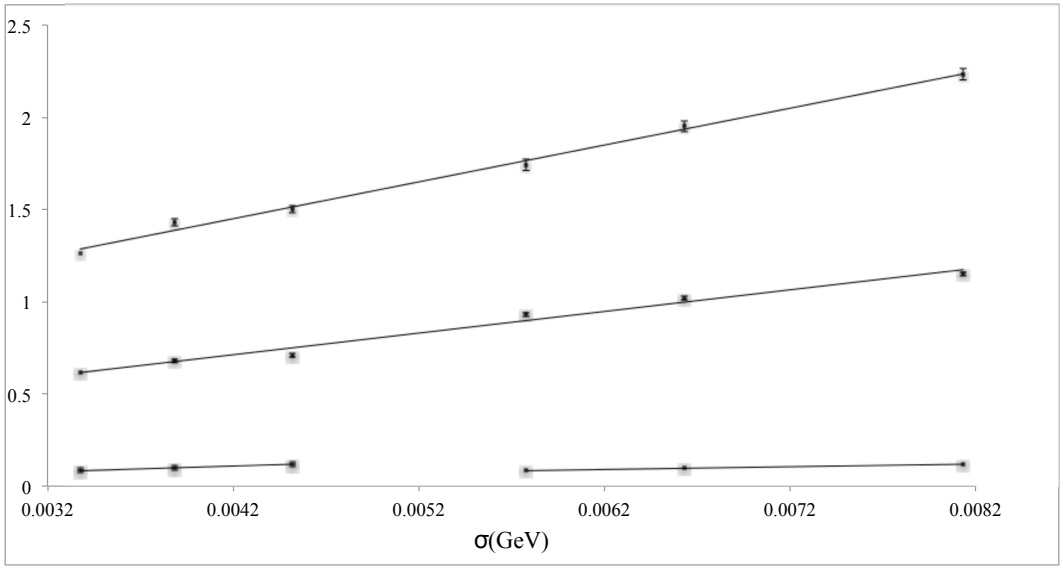


Figure 5.8: Linear fits to the lineshape parameters $\alpha_{1,2}$ (top and middle) and line energy E (bottom) as functions of the fitted width σ in signal MC for $3S \to 2P \to 1S$. Compare to Fig. **??**. Extracted α_1 and α_2 slopes are nearly identical to the $3S \to 2P \to 2S$ case, but with significantly different offsets; this is the expected behavior.

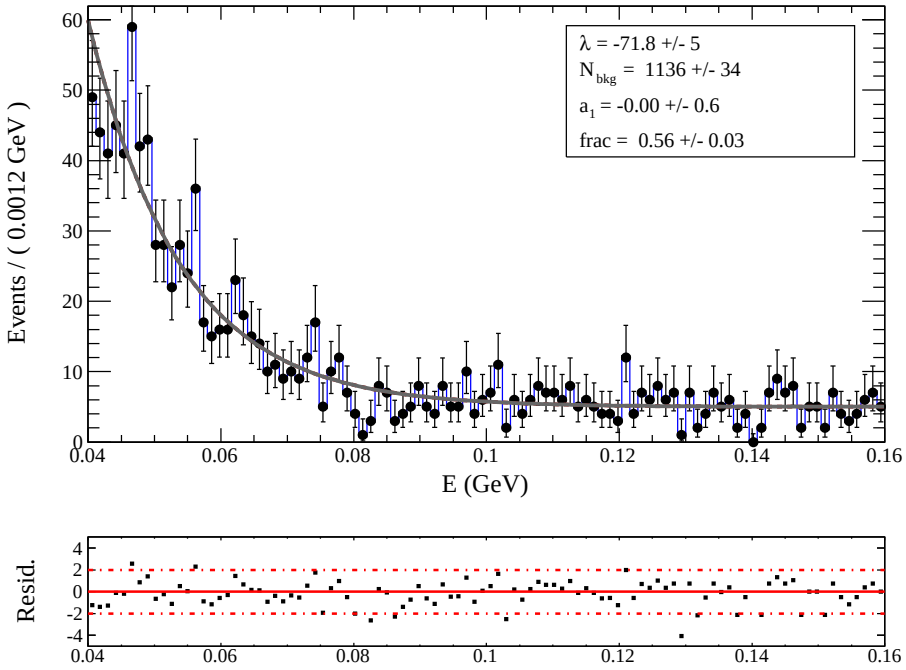


Figure 5.9: Fit to the background components of the $3S \to 2P \to 1S$ MC ensemble, in this case solely $\Upsilon(3S)$ generic decays in the full MC statistics.

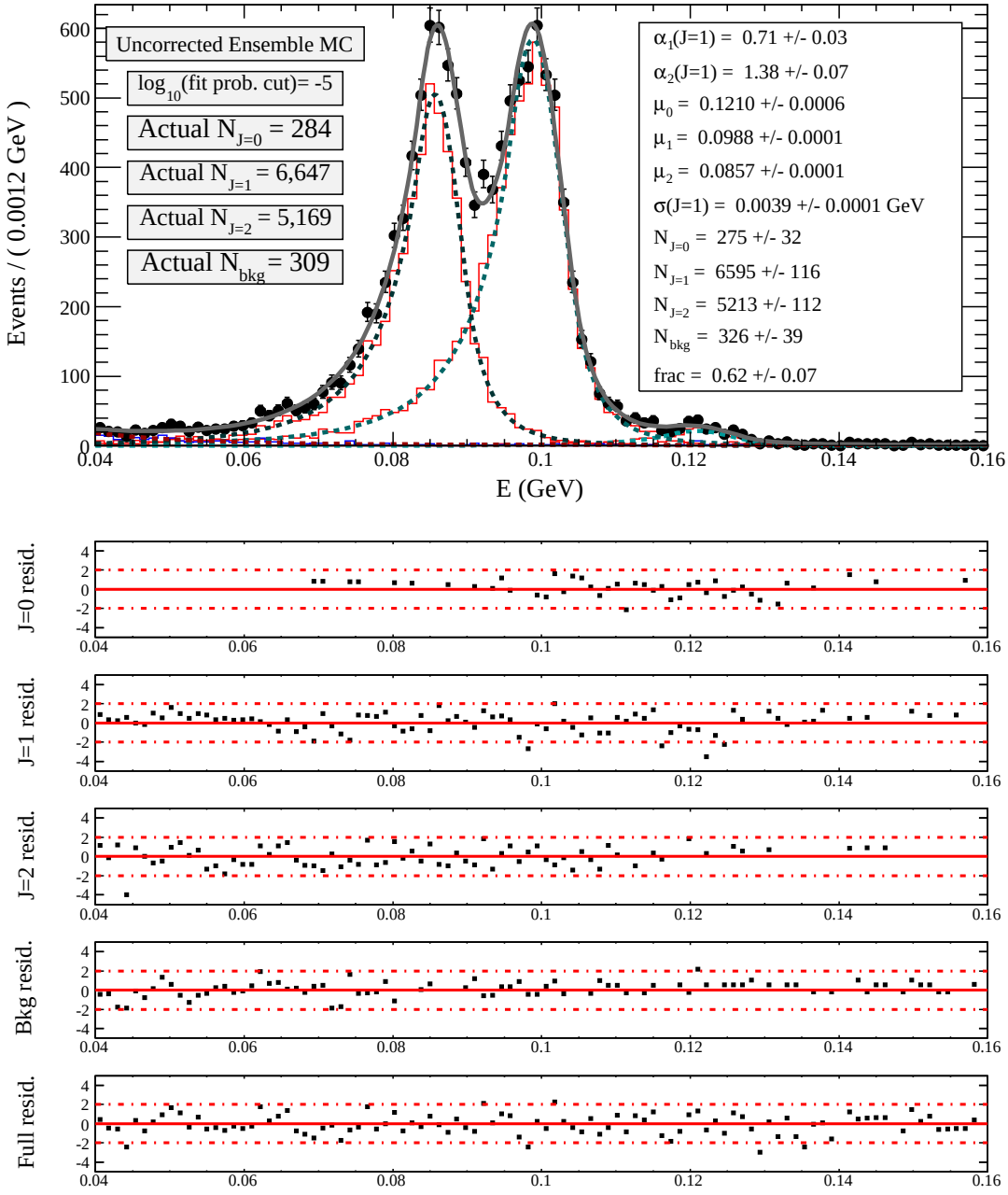


Figure 5.10: Fit to the full $3S \to 2P \to 1S$ uncorrected MC ensemble with component-by-component normalized residuals.

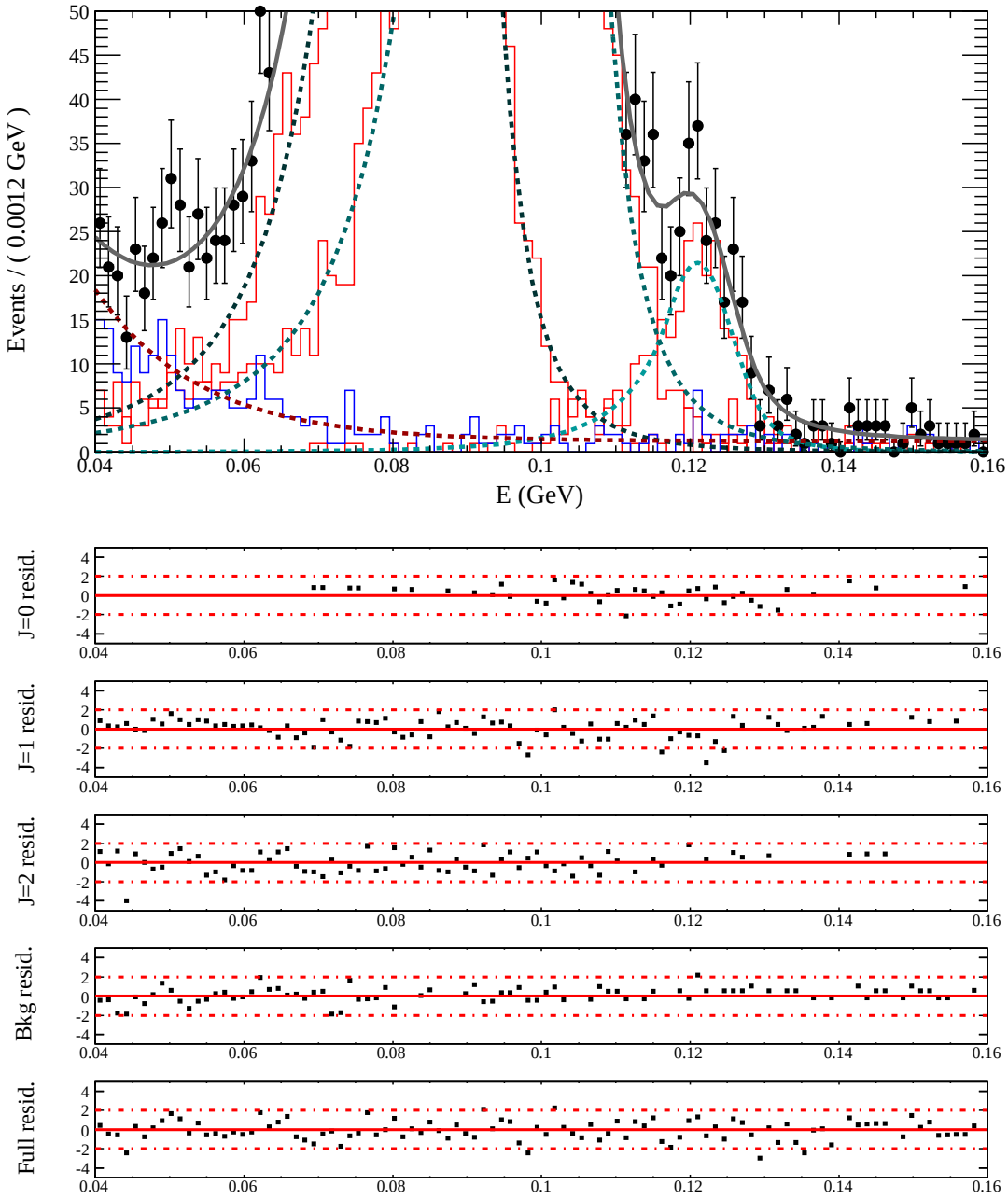


Figure 5.11: Zoomed view of Fig. **??**. Background MC histogram is in blue, signal MC are the red histograms. The dashed lines correspond to the fitted components.

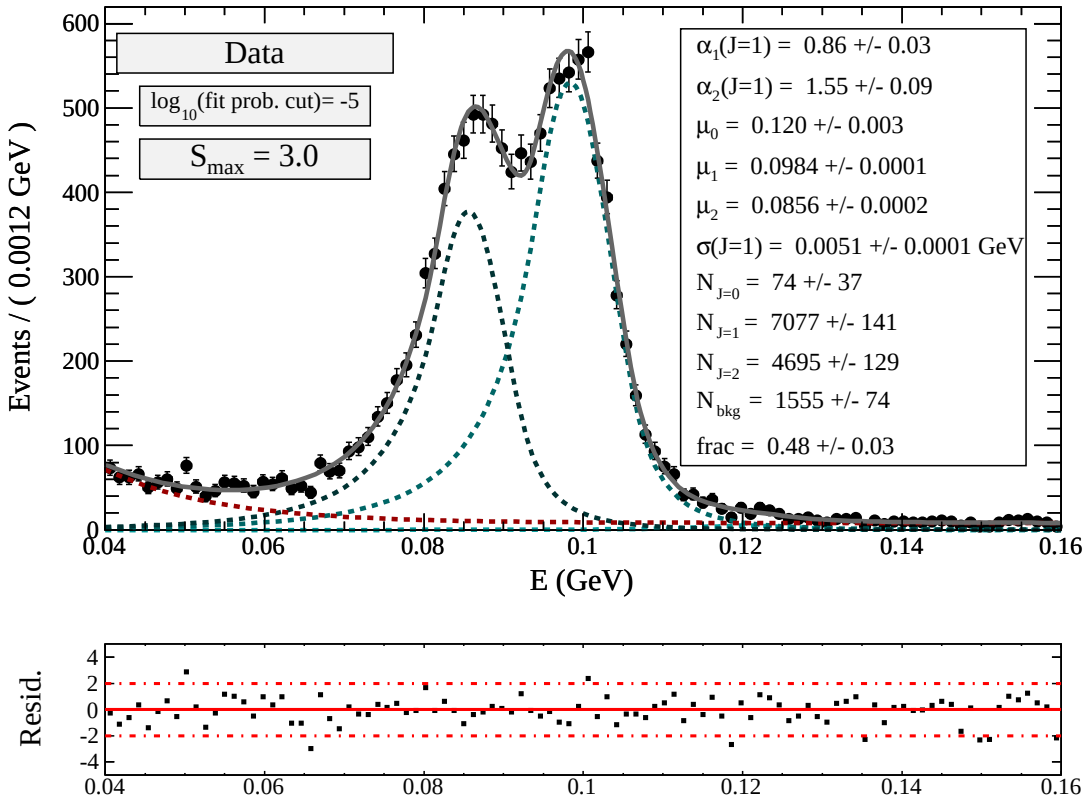


Figure 5.12: Fit to $3S \to 2P \to 1S$ data with final selections.

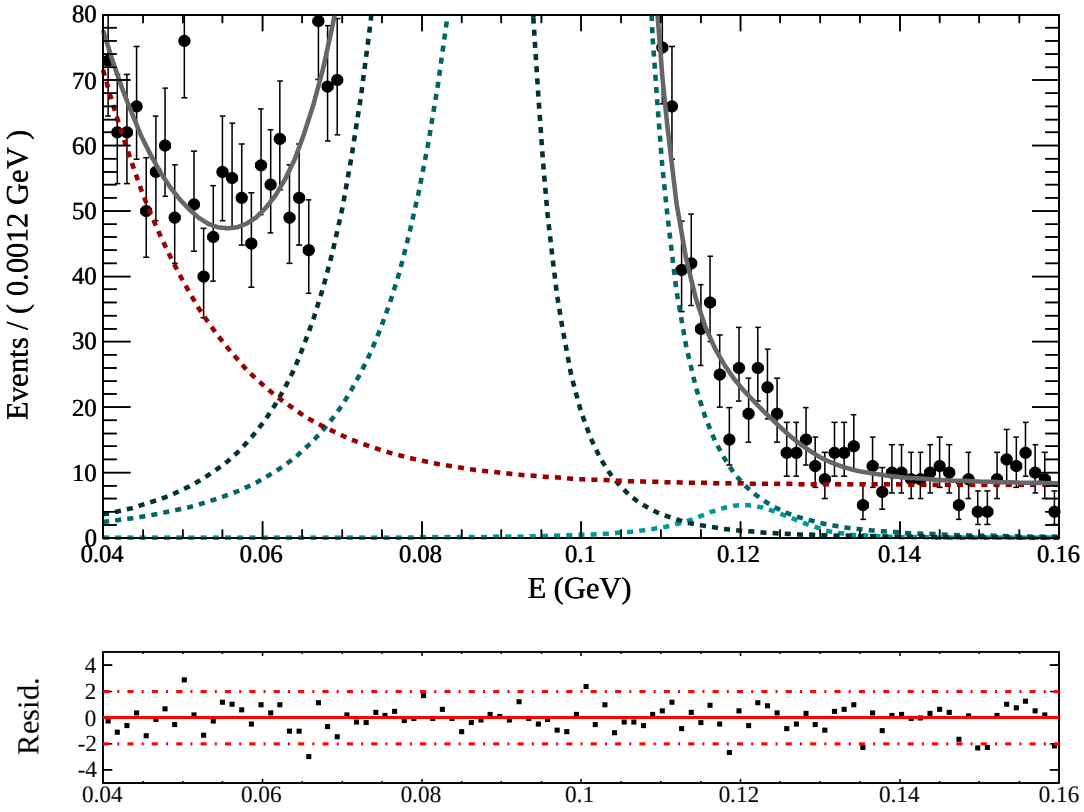


Figure 5.13: Zoomed view of Fig. ??.

5.5 $3S \to 2P \to 1S$ systematics

Toy studies proceed as in the previous analyses. With such low pre-systematics significance in the $\chi_{b0}(2P)$ peak, the jitter MC provides an opportunity to determine whether zero yield can be ruled out for this peak.

Table **??** shows the usual jitter parameter choices for the $3S \to 2P \to 1S$ toy spectra, with post-fit selections shown in Table **??**. Note that the N_0 jitter range includes negative and null values; the null hypothesis is thus explicitly tested in the jitter MC toy studies.

Results for the $3S \to 2P \to 1S$ jitter MC toys are shown in Fig. **??** for the yield parameters and Fig. **??** for the means. Pulls are consistent with the previous two analyses. Critically, significance of a non-null N_0 yield is actually enhanced with the correction factors derived from the toy fits.

Parameter	Nominal value source	Nom. value	Param. σ	Jitter range
λ	Fit to MC ensemble background	-71		[Not jittered]
a_1	Fit to data sidebands	0.0		Nominal $\pm 3\sigma_{a_1}$
$frac$	Fit to data	0.48	0.03	Nominal $\pm 3\sigma_{frac}$
μ_0	Fit to data	120 MeV	3	Nominal $\pm 3\sigma_{\mu_0}$
μ_1	Fit to data	98.4 MeV	0.1	Nominal $\pm 3\sigma_{\mu_1}$
μ_2	Fit to data	85.6	0.2	Nominal $\pm 3\sigma_{\mu_2}$
$[m_\sigma]$	Signal MC and EMC res. param.			0.01–0.08, about $\pm 5\sigma$
$\sigma(J=1)$	Fit to data	5.1 MeV	0.1 MeV	Nominal $\pm 3\sigma_{\sigma(J=1)}$
$\sigma(J=0)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_0 - \mu_1)$
$\sigma(J=2)$				$\sigma(J=1) + (m_\sigma \pm 5\sigma_{m_\sigma}) \cdot (\mu_2 - \mu_1)$
$[m_{\alpha_1}]$	MC parameter fit	127.6	10.3	
$\alpha_1(J=1)$	Fit to data	0.86	0.03	Nominal $\pm 3\sigma_{\alpha_1}$
$\alpha_1(J=0)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_1(J=2)$				$\alpha_1(J=1) + (m_{\alpha_1} \pm 5\sigma_{m_{\alpha_1}}) \cdot (\sigma(J=2) - \sigma(J=1))$
$[m_{\alpha_2}]$	MC parameter fit	200.4	12.7	
$\alpha_2(J=1)$	Fit to data	1.55	0.09	Nominal $\pm 3\sigma_{\alpha_2}$
$\alpha_2(J=0)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=0) - \sigma(J=1))$
$\alpha_2(J=2)$				$\alpha_2(J=1) + (m_{\alpha_2} \pm 5\sigma_{m_{\alpha_2}}) \cdot (\sigma(J=2) - \sigma(J=1))$
n_1	Manual			2.0–50.0
n_2	Manual			2.0–50.0
$N_{J=0}$	Fit to data	74	37	Nominal $\pm 3\sigma_{N_{J=0}}$
$N_{J=1}$	Fit to data	7,077	141	Nominal $\pm 3\sigma_{N_{J=1}}$
$N_{J=2}$	Fit to data	4,695	129	Nominal $\pm 3\sigma_{N_{J=2}}$
N_{bkg}	Fit to data	1,555	174	Nominal $\pm 3\sigma_{N_{bkg}}$

Table 5.5: Summary of parameters used to generate the $3S \to 2P \to 2S$ jitter MC spectra in $3S \to 2P \to 1S$ cascades.

Parameter	Selection	Low value	High value
$\alpha_1(J=1)$	Data fit $\pm 3\sigma_{\alpha_1}$	0.77	0.95
$\alpha_2(J=1)$	Data fit $\pm 3\sigma_{\alpha_2}$	1.28	1.82
$\sigma(J=1)$	Data fit $\pm 3\sigma_{\sigma}$	4.8 MeV	5.4 MeV
N_0	Data fit $\pm 3\sigma_{N_0}$	1	185
N_1	Data fit $\pm 3\sigma_{N_1}$	6,654	7,500
N_2	Data fit $\pm 3\sigma_{N_2}$	4,308	5,082
N_{bkg}	Data fit $\pm 3\sigma_{N_{bkg}}$	1,222	1,777
σ_{N_0}	Data fit $\pm 3\sqrt{\sigma_{N_0}}$	18	55
σ_{N_1}	Data fit $\pm 3\sqrt{\sigma_{N_1}}$	105	177
σ_{N_2}	Data fit $\pm 3\sqrt{\sigma_{N_2}}$	95	163

Table 5.6: Summary of selections made on fitted parameters from the $3S \to 2P \to 1S$ jitter MC toys. Some differences in selections compared to the $2S$ analysis are used to accommodate a higher level of statistics.

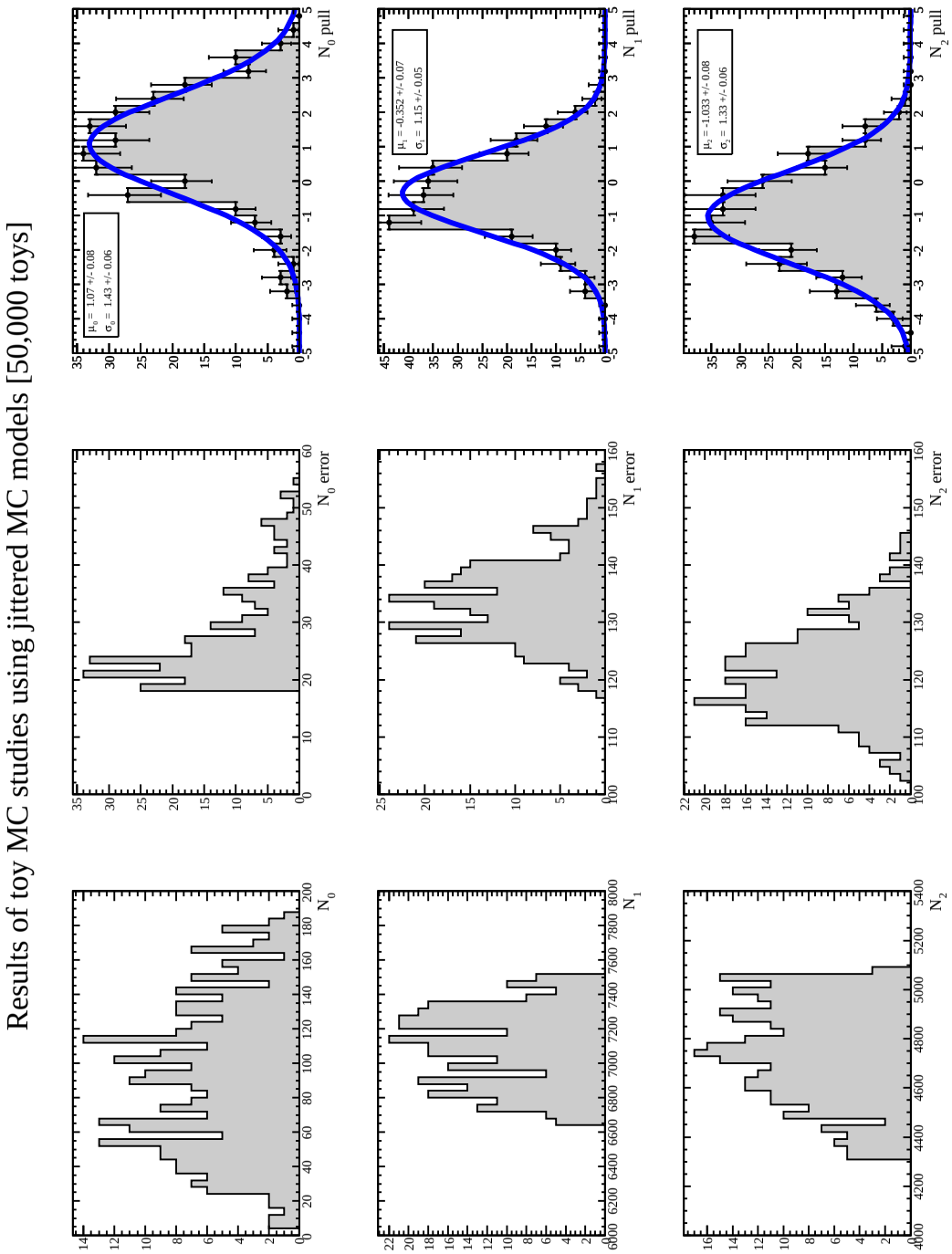


Figure 5.14: The results of 50,000 fits to the $3S \to 2P \to 1S$ jitter MC spectra for the yield parameters $N_{0,1,2}$.

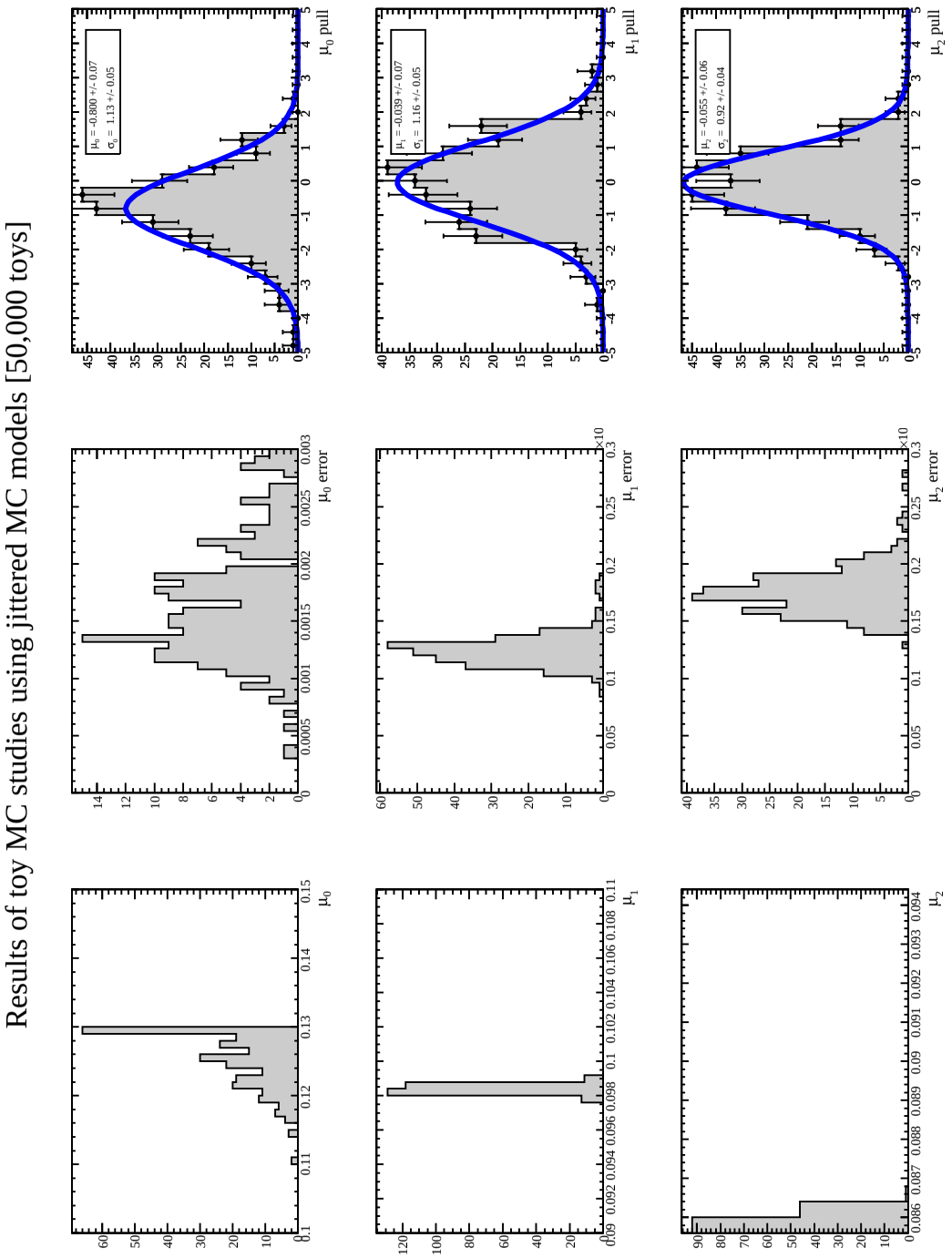


Figure 5.15: The results of 50,000 fits to the jitter MC spectra for the Gaussian mean parameters $\mu_{0,1,2}$.

5.6 $3S \to 2P \to 1S$ results

Using the yield corrections obtained using the jitter MC, we present the product branching fraction ratio results from the $3S \to 2P \to 1S$ analysis:

$$\mathscr{F}^{0/1}_{3S\to 2P\to 1S} = \frac{114 \pm 53}{7027 \pm 162} \times (1.059 \pm 0.009) = (1.71 \pm 0.80)\%, \tag{5.1}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 1S} = \frac{4562 \pm 172}{7027 \pm 162} \times (1.027 \pm 0.009) = (66.6 \pm 3.0)\%, \tag{5.2}$$

The mass splittings are measured to be:

$$\Delta M^{2P}_{1-0} = (122.4 \pm 3.4 - 98.40 \pm 0.12)\,\text{MeV} = 24.0 \pm 3.4\,\text{MeV}, \tag{5.3}$$

$$\Delta M^{2P}_{2-1} = (98.40 \pm 0.12 - 85.61 \pm 0.20)\,\text{MeV} = 12.79 \pm 0.23\,\text{MeV}. \tag{5.4}$$

Comparable results derived from individual branching fractions as reported in the PDG are:

$$\mathscr{F}^{0/1}_{3S\to 2P\to 1S} = (10.8 \pm 5.3)\%, \tag{5.5}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 1S} = (55 \pm 17)\%, \tag{5.6}$$

$$\Delta M^{2P}_{1-0} = 22.96 \pm 0.84\,\text{MeV}, \tag{5.7}$$

$$\Delta M^{2P}_{2-1} = 13.19 \pm 0.77\,\text{MeV}. \tag{5.8}$$

See Chapter 6 for a discussion of the results.

Chapter 6

Discussion

The analyses discussed in this dissertation have followed a general philosophy of providing systematics-minimal determinations of bottomonium spectroscopy results. These results fall into three categories: (1) primary results, such as the product branching ratios; (2) secondary results, directly obtained from the primary results in language targeted at particular relevant theories and (3) derived results, using quantities from the PDG and other analyses. These results are summarized here with a discussion of the relevance and context of the findings.

6.1 Summary of results

Primary, secondary and derived results are given here with some context. A full discussion of the implications of these results are given in the next section.

6.1.1 Primary results

The product branching ratios have been measured:

$$\mathscr{F}^{J/1}_{nS\to P(J)\to mS} = \frac{\mathcal{B}(nS\to P(J))\mathcal{B}(P(J)\to mS)}{\mathcal{B}(nS\to P(1))\mathcal{B}(P(1)\to mS)} \tag{6.1}$$

with the following values:

$$\mathscr{F}^{0/1}_{2S\to 1P\to 1S} = (3.28\pm 0.37)\%, \tag{6.2}$$

$$\mathscr{F}^{2/1}_{2S\to 1P\to 1S} = (55.6 \pm 1.6)\%, \tag{6.3}$$

$$\mathscr{F}^{0/1}_{3S\to 2P\to 2S} = (2.31 \pm 0.56)\%, \tag{6.4}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 2S} = (46.9 \pm 2.0)\%, \tag{6.5}$$

$$\mathscr{F}^{0/1}_{3S\to 2P\to 1S} = (1.71 \pm 0.80)\%, \tag{6.6}$$

$$\mathscr{F}^{2/1}_{3S\to 2P\to 1S} = (66.6 \pm 3.0)\%. \tag{6.7}$$

The measured mass splittings are given by $\Delta M^{nP}_{J-J'} = M^{nP}_{J} - M^{nP}_{J'}$:

$$\Delta M^{1P}_{1-0} = 32.49 \pm 0.93\,\mathrm{MeV}, \tag{6.8}$$

$$\Delta M^{1P}_{2-1} = 19.01 \pm 0.24\,\mathrm{MeV}, \tag{6.9}$$

$$\Delta M^{2P}_{1-0} = 23.8 \pm 1.7\,\mathrm{MeV}, \tag{6.10}$$

$$\Delta M^{2P}_{2-1} = 13.01 \pm 0.23\,\mathrm{MeV}, \tag{6.11}$$

where in the nP case we have combined the splittings from the two $3S$ modes with a variance-weighted mean:

$$\overline{\delta} = \frac{\frac{\delta_1}{\sigma_1^2} + \frac{\delta_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}, \tag{6.12}$$

where δ_i refers to the measured value of $\Delta M^{2P}_{J-J'}$ for the two $3S$ analyses and σ_i are the associated errors. The variance of the weighted mean is given by:

$$\sigma^2 = \left[\frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}\right]\left[\frac{(\delta_1 - \overline{\delta})^2}{\sigma_1^2} + \frac{(\delta_2 - \overline{\delta})^2}{\sigma_2^2}\right], \tag{6.13}$$

where the χ^2 term is dropped if it is less than 1. The final splittings are simply $\Delta M^{2P}_{J-J'} = \delta \pm \sigma$.

6.1.2 Secondary results

The secondary results consist of the fine structure parameters discussed in Section **??**. These results are directly derived from the line splittings from the primary results and are summarized in Table **??**.

The mass combinations used to isolate the leading-order spin-dependent coefficients c_3 and c_4^2 in the NRQCD Hamiltonian, as discussed in Section **??** are found to be:

$$-2\delta_0 + 5\delta_2 = 12a \tag{6.14}$$

for the c_3 term, for which we measure a value of (160.0±2.2) MeV and (112.7±3.8) MeV for the $1P$ and $2P$ triplets, respectively. The tensor term becomes

$$2\delta_0 - \delta_2 = -\frac{72}{5}b \tag{6.15}$$

for which we measure (-46.0 ± 1.9) MeV and (-34.5 ± 3.5) MeV for $1P$ and $2P$, respectively.

Table 6.1: Comparison of fine splitting parameters in the $1P$ and $2P$ systems. Results are compared to the world averages as of 1993 in column 4 and two more recent measurements where these parameters were explicitly reported. See [2] for a comparison of a large number of theoretical predictions.

Parameter	nP	This analysis	[2]	[15]	[14]
a(MeV)	$1P$	13.34 ± 0.18	14.2 ± 0.8		
	$2P$	9.40 ± 0.31	9.4 ± 0.2		
b(MeV)	$1P$	3.19 ± 0.13	3.0 ± 0.3		
	$2P$	2.39 ± 0.25	2.3 ± 0.1		
R_χ	$1P$	0.585 ± 0.018	0.65 ± 0.03	0.574 ± 0.012	0.54 ± 0.03
	$2P$	0.549 ± 0.042	0.58 ± 0.01	0.584 ± 0.014	

6.1.3 Derived Results: E1 matrix elements

In the nonrelativistic limit the $E1$ matrix elements of Eqn. **??** only depend on the radial and orbital quantum numbers nL; ratios of matrix elements for two transitions that differ only in spin are thus convenient probes of relativistic corrections. We use the derived branching ratios from the next section to calculate matrix element ratios

with the $J = 1$ element in the denominator:

$$\begin{aligned}\frac{|\langle f_J|r|i_J\rangle|^2}{|\langle f_1|r|i_1\rangle|^2} &= \frac{3}{2J_f+1}\frac{E_1^3}{E_J^3}\frac{\mathcal{B}(i_J \to f_J)}{\mathcal{B}(i_1 \to f_1)} \\ &= \frac{3}{2J+1}\left(1 \pm \frac{\delta_J}{E_1}\right)^{-3}\frac{\mathcal{B}(i_J \to f_J)}{\mathcal{B}(i_1 \to f_1)}\end{aligned} \tag{6.16}$$

where in the last line we have recast the absolute photon energies in reference to the $J = 1$ energy and the measured mass splittings to minimize uncertainties. The sign comes from the definition of δ_J so that it is negative (positive) for $S \to P$ transitions involving the 0/1 (2/1) ratio and the opposite for the $P \to S$ transitions.

The calculation of the matrix element ratios requires branching ratios which we do not explicitly measure. Instead we use the best available branching fractions as shown in Tables **??**, **??**, **??**. Similarly for the absolute line energies we calculate the spin-weighted center of energy for the $1P$ and $2P$ states $\overline{\mu^{mP}}$ and the spin-weighted center of gravity $\overline{M^{mP}}$ from PDG [53] masses similarly. We derive all line energies using $E_J = \overline{M^{mP}} \pm (\overline{\mu^{mP}} - \mu_J)$ where the sign depends on the transition. The derived line energies are shown in Table **??**, although only the $J = 1$ values are used in the matrix element ratio calculations, shown in Table **??**.

6.1.4 Derived results: branching fractions

Tables **??**, **??** and **??** at the end of this document show the derived branching fractions using results from this analysis and the best available product and bare branching fractions. The strategy employed in this analysis aims specifically at reaching the $1P(0) \to 1S$, $2P(0) \to 2S$ and $2P(0) \to 1S$ transitions, which we have seen with significances of 8.9, 4.1 and 2.1σ with derived branching fractions calculated to 6.1, 3.4 and 2.1σ significances, all world-best. Our derived branching fractions are largely consistent with previous results with one minor exception: the modest tensions in Table **??** are discussed in more detail in Section **??**.

Table 6.2: Photon energies for transitions evaluated in this analysis, calculated by correcting the spin-weighted average of Gaussian means μ_J to the correct average as calculated from PDG masses [53]. Only $J = 1$ values and mass splittings are used for the matrix element ratio calculations.

Transition	E_γ(MeV)
2S$\rightarrow$ 1P(0)	162.8 ± 1.0
2S$\rightarrow$ 1P(1)	130.34 ± 0.45
2S$\rightarrow$ 1P(2)	111.33 ± 0.49
3S$\rightarrow$ 2P(0)	123.4 ± 1.8
3S$\rightarrow$ 2P(1)	99.61 ± 0.64
3S$\rightarrow$ 2P(2)	86.59 ± 0.66
1P(0)$\rightarrow$ 1S	400.1 ± 1.0
1P(1)$\rightarrow$ 1S	432.62 ± 0.45
1P(2)$\rightarrow$ 1S	451.63 ± 0.49
2P(0)$\rightarrow$ 2S	208.6 ± 1.8
2P(1)$\rightarrow$ 2S	232.44 ± 0.51
2P(2)$\rightarrow$ 2S	245.35 ± 0.53
2P(0)$\rightarrow$ 1S	771.5 ± 1.8
2P(1)$\rightarrow$ 1S	795.29 ± 0.48
2P(2)$\rightarrow$ 1S	808.31 ± 0.50

6.2 Context of the Results

Presented herein is an array of primary, secondary and derived spectroscopic results comprising a nearly comprehensive study of electric dipole transitions between the $(1,2,3)S$ and $(1,2)P$ bottomonium states. These results include best-yet evidence for the $\chi_{b0}(2P) \to (1,2)S$ transitions, competitive determination of splitting parameters and a calculation of spin-dependent matrix element ratios and branching fractions.

The PDG values for $\mathcal{B}(2P(0) \to 2S)$ and $\mathcal{B}(2P(0) \to 1S$ both come from a single fit to the combined exclusive spectrum analyzed by the CUSB-II Collaboration [19] with a $J = 0$ peak yield of 17 ± 7 events shared between the two modes. There is some evidence from CLEO [23] and *BABAR* [59] that $\mathcal{B}(2P(0) \to 2S)$ is somewhat high; we agree, with our derived value for this branching fraction 1.3σ below the value used by the PDG. We find a lower value with a smaller significance (0.8σ) also in the $2P(0) \to 1S$ transition derived from the same fit. The larger tensions in Table **??** all stem from the use of the PDG value to derive other quantities; we see similar tensions in Table **??** if we replace the product branching fractions from CLEO [23] with the $2P(0) \to 1S$ results from the PDG/CUSB-II. As a semi-qualitative parsimony argument this suggests that the branching fractions for the $3S$ modes as reported in the PDG are modestly overstated for both $2P(0) \to 2S$ and $2P(0) \to 1S$.

We see $\sim 1\sigma$-level spin-dependence in the matrix element ratios shown in Table **??** for several ratios. These results are competitive with previous analyses, particularly for the 2/1 ratios.

The splitting parameter measurements in Table **??** are competitive and largely consistent with previous results. Our value for the splitting ratio $R_{\chi(1P)}$ supports more recent CLEO results [15, 14] over the 1993 world-average [2] shown, with a value almost 2σ below the then-average. Our measurements of the NRQCD parameters c_3 and c_4 are competitive with world-averages commonly used, with a distinct improvement in the determination of c_4.

6.3 Impact

In the spirit of learning to communicate better between experiment and theory this analysis has pioneered a basic analysis technique that provides experimentally optimal

fine splitting parameters for the relevant potential model and lattice calculations. The strategy re-casts all splitting parameters in terms of measured line splittings and not absolute masses, avoiding the large systematic uncertainties involved in absolute mass determination. The technique is conceptually and mechanically extremely simple but has not been used previously.

The exclusive strategy of these analyses is far from ideal for fine structure measurements – given the large χ_{b0} hadronic widths, inclusive analyses looking at the transition $3S \to 2P$ and $2S \to 1P$ should provide a much more precise determination of the line splittings. This analysis could use much of the machinery from the analyses presented here and would provide fine structure parameters well beyond those that have been used to tune and verify a wide range of effective theories in QCD physics.

Inclusive analyses are better suited for the mass splitting and $3S \to 2P$ matrix element measurements, and we suggest such an analysis using the relative-energy techniques we have presented to target high-precision measurements of these quantities with explicit determinations of a, b, c_3 and c_4.

The systematics-minimal product branching ratio strategy that comprises the major contribution from this analysis is also rather unique. It is our intention that these high-precision results will generate the world-best determination of a large number of bare and product branching fractions continually as new smaller-scale results come in. Consequently, the matrix elements derived in this analysis should also continue to generate new high-precision values.

Table 6.3: Matrix element ratios derived using results from this analysis and the best available product and bare branching fractions from Tables **??**, **??** and **??**. Results in column 4 are the world averages as of 1993; subsequent explicit measurements are shown in the final column.

$\frac{n_iL_i(J_i) \to n_fL_f(J_f)}{n_i'L_i'(J_i') \to n_f'L_f'(J_f')}$	$\frac{\lvert\langle n_fL_f(J_f)\lvert r\rvert n_iL_i(J_i)\rangle\rvert^2}{\lvert\langle n_f'L_f'(J_f')\lvert r\rvert n_i'L_i'(J_i')\rangle\rvert^2}$	σ from unity	[2]	[23, 14]
$\frac{3S \to 2P(0)}{3S \to 2P(1)}$	0.74 ± 0.19	1.3	0.74 ± 0.06	
$\frac{3S \to 2P(2)}{3S \to 2P(1)}$	0.95 ± 0.16	0.3	1.17 ± 0.04	
$\frac{2S \to 1P(0)}{2S \to 1P(1)}$	0.85 ± 0.13	1.2	0.95 ± 0.16	0.75 ± 0.28
$\frac{2S \to 1P(2)}{2S \to 1P(1)}$	0.998 ± 0.085	0.0	0.92 ± 0.11	1.02 ± 0.11
$\frac{2P(0) \to 1S}{2P(0) \to 2S} \Big/ \frac{2P(1) \to 1S}{2P(1) \to 2S}$	0.59 ± 0.35	1.2	0.37 ± 0.3	
$\frac{2P(2) \to 1S}{2P(2) \to 2S} \Big/ \frac{2P(1) \to 1S}{2P(1) \to 2S}$	1.59 ± 0.43	1.4	1.33 ± 0.26	1.21 ± 0.06

Table 6.4: Summary of derived results from the $2S \to 1P \to 1S$ analysis. *Italicized* entries are directly reported by the stated sources; all others are derived from compounding values within that analysis. Other analysis results are from PDG averages except those marked with * which come from [21] with greater significance than derived quantities from the PDG. The significance of the derived quantities for this analysis and the comparison values are shown in columns 3 and 5; **bold** entries are those from which the highest-significance value is derived in this analysis. Tensions between the two sources are displayed in the last column.

Measurement	*BABAR* 2013 (10^{-2})	(σ)	Other (10^{-2})	(σ)	Tension (σ)
$\mathcal{B}(2S \to 1P(0))$	4.2 ± 1.0	4.2	*3.8 ± 0.4*	9.5	0.4
$\mathcal{B}(2S \to 1P(1))$	7.06 ± 0.65	10.9	*6.9 ± 0.4*	17.3	0.2
$\mathcal{B}(2S \to 1P(2))$	6.64 ± 0.59	11.3	*7.15 ± 0.35*	20.4	0.7
$\mathcal{B}(1P(0) \to 1S)$	1.97 ± 0.32	**6.1**	*1.76 ± 0.35*	5.0	0.4
$\mathcal{B}(1P(1) \to 1S)$	34.7 ± 3.0	11.4	*33.9 ± 2.2*	15.4	0.2
$\mathcal{B}(1P(2) \to 1S)$	17.7 ± 1.4	12.6	*19.1 ± 1.2*	15.9	0.7
$\mathcal{B}(2S \to 1P(0))/\mathcal{B}(2S \to 1P(1))$	63 ± 15	4.2	55.1 ± 6.6	8.3	0.5
$\mathcal{B}(2S \to 1P(2))/\mathcal{B}(2S \to 1P(1))$	98.7 ± 9.4	10.5	104 ± 7.9	13.2	0.4
$\mathcal{B}(1P(0) \to 1S)/\mathcal{B}(1P(1) \to 1S)$	5.96 ± 0.98	**6.1**	5.2 ± 1.1	4.8	0.5
$\mathcal{B}(1P(2) \to 1S)/\mathcal{B}(1P(1) \to 1S)$	53.7 ± 4.4	**12.3**	56.3 ± 5.1	11.1	0.4
$\mathcal{B}(2S \to 1P(0)) \cdot \mathcal{B}(1P(0) \to 1S)$	0.0748 ± 0.0094	**7.9**	$*0.066 \pm 0.011$	5.8	0.6
$\mathcal{B}(2S \to 1P(1)) \cdot \mathcal{B}(1P(1) \to 1S)$	2.39 ± 0.16	15.3	$*2.28 \pm 0.13$	17.9	0.6
$\mathcal{B}(2S \to 1P(2)) \cdot \mathcal{B}(1P(2) \to 1S)$	1.268 ± 0.080	15.9	$*1.333 \pm 0.078$	17.0	0.6
$\mathscr{F}^{0/1}(2S \to 1P \to 1S)$	*3.28 ± 0.37*	8.9	2.89 ± 0.52	5.5	0.6
$\mathscr{F}^{2/1}(2S \to 1P \to 1S)$	*55.6 ± 1.6*	34.8	58.3 ± 4.7	12.3	0.5

Table 6.5: Summary of derived results from the $3S \to 2P \to 2S$ analysis. See the caption of Table **??** for details. * [23].

Measurement	*BABAR* 2013 (10^{-2})	(σ)	Other (10^{-2})	(σ)	Tension (σ)
$\mathcal{B}(3S \to 2P(0))$	4.2 ± 2.1	2.0	*5.90 ± 0.60*	9.8	0.8
$\mathcal{B}(3S \to 2P(1))$	11.3 ± 2.4	4.7	*12.6 ± 1.2*	10.5	0.5
$\mathcal{B}(3S \to 2P(2))$	11.4 ± 2.5	4.5	*13.1 ± 1.6*	8.2	0.6
$\mathcal{B}(2P(0) \to 2S)$	1.84 ± 0.54	**3.4**	**2.6 ± 1.1*	2.5	0.6
$\mathcal{B}(2P(1) \to 2S)$	37.2 ± 6.8	5.4	**41.5 ± 6.0*	6.9	0.5
$\mathcal{B}(2P(2) \to 2S)$	17.5 ± 4.8	3.7	**19.3 ± 3.3*	5.9	0.3
$\mathcal{B}(3S \to 2P(0))/\mathcal{B}(3S \to 2P(1))$	37 ± 18	2.0	46.8 ± 6.5	7.2	0.5
$\mathcal{B}(3S \to 2P(2))/\mathcal{B}(3S \to 2P(1))$	101 ± 23	4.4	104 ± 16	6.5	0.1
$\mathcal{B}(2P(0) \to 2S)/\mathcal{B}(2P(1) \to 2S)$	4.9 ± 1.4	**3.6**	**6.2 ± 2.7*	2.3	0.4
$\mathcal{B}(2P(2) \to 2S)/\mathcal{B}(2P(1) \to 2S)$	45.1 ± 7.2	**6.2**	**47 ± 10*	4.5	0.1
$\mathcal{B}(3S \to 2P(0)) \cdot \mathcal{B}(2P(0) \to 2S)$	0.108 ± 0.0030	**3.6**	**0.140 ± 0.054*	2.6	0.5
$\mathcal{B}(3S \to 2P(1)) \cdot \mathcal{B}(2P(1) \to 2S)$	4.69 ± 0.74	6.4	**4.69 ± 0.64*	7.4	0.0
$\mathcal{B}(3S \to 2P(2)) \cdot \mathcal{B}(2P(2) \to 2S)$	2.2 ± 0.31	**7.0**	**2.20 ± 0.33*	6.6	0.0
$\mathscr{F}^{0/1}(3S \to 2P \to 2S)$	*3.31 ± 0.56*	8.9	*3.0 ± 1.2	2.5	0.5
$\mathscr{F}^{2/1}(3S \to 2P \to 2S)$	*46.9 ± 2.0*	34.8	*46.9 ± 9.5	4.9	0.0

Table 6.6: Summary of derived results from the $3S \to 2P \to 1S$ analysis. See the caption of Table **??** for details. * [23], † [19].

Measurement	*BABAR* 2013 (10^{-2})	(σ)	Other (10^{-2})	(σ)	Tension (σ)
$\mathcal{B}(3S \to 2P(0))$	2.5 ± 2.0	1.2	*5.90 ± 0.60*	9.8	1.6
$\mathcal{B}(3S \to 2P(1))$	10.2 ± 1.5	6.9	*12.6 ± 1.2*	10.5	1.2
$\mathcal{B}(3S \to 2P(2))$	12.5 ± 1.9	6.5	*13.1 ± 1.6*	8.2	0.3
$\mathcal{B}(2P(0) \to 1S)$	0.38 ± 0.18	**2.1**	†*0.90 ± 0.60*	1.5	0.8
$\mathcal{B}(2P(1) \to 1S)$	9.4 ± 1.4	6.6	**11.60 ± 0.99*	11.8	1.3
$\mathcal{B}(2P(2) \to 1S)$	6.9 ± 1.3	5.2	**7.00 ± 0.89*	7.8	0.0
$\mathcal{B}(3S \to 2P(0))/\mathcal{B}(3S \to 2P(1))$	22 ± 18	1.2	46.8 ± 6.5	7.2	1.3
$\mathcal{B}(3S \to 2P(2))/\mathcal{B}(3S \to 2P(1))$	110 ± 18	6.3	104 ± 16	6.5	0.3
$\mathcal{B}(2P(0) \to 1S)/\mathcal{B}(2P(1) \to 1S)$	3.7 ± 1.8	**2.0**	*†*7.8 ± 5.2*	1.5	0.7
$\mathcal{B}(2P(2) \to 1S)/\mathcal{B}(2P(1) \to 1S)$	64 ± 10	6.2	**60 ± 9.3*	6.5	0.3
$\mathcal{B}(3S \to 2P(0)) \cdot \mathcal{B}(2P(0) \to 1S)$	0.022 ± 0.011	**2.1**	†*0.050 ± 0.041*	1.2	0.6
$\mathcal{B}(3S \to 2P(1)) \cdot \mathcal{B}(2P(1) \to 1S)$	1.19 ± 0.14	8.5	**1.310 ± 0.094*	13.9	0.7
$\mathcal{B}(3S \to 2P(2)) \cdot \mathcal{B}(2P(2) \to 1S)$	0.872 ± 0.074	**11.8**	**0.790 ± 0.086*	9.2	0.7
$\mathscr{F}^{0/1}(3S \to 2P \to 1S)$	*1.71 ± 0.80*	2.1	*3.8 ± 3.2	1.2	0.6
$\mathscr{F}^{2/1}(3S \to 2P \to 1S)$	*66.6 ± 3.0*	22.2	*60.3 ± 7.9	7.7	0.7

Appendix A

The $\Upsilon(1D)$ spectrum

The most interesting remaining exclusive analysis in the bottomonium system is the $3S \to 2P \to 1D \to 1P \to 1S$ cascade. The $J = 1$ and $J = 3$ states have not been observed. This analysis is potentially extremely challenging: the soft photons in the transitions $3S \to 2P$ and $2P \to 1D$ have very low energy, making beam background a distinct challenge. The splittings are quite narrow, which makes disentangling the three peaks also difficult. The product branching fraction of the four-step cascade is small, yielding low statistics. Additionally, the signal spectra include up to a dozen overlapping peaks with unknown means and yields. However, the timing selection used throughout this document provides a uniquely good tool for evaluating this spectrum at *BABAR* so we attempt to gain some insight into the $\Upsilon(1D)$ system using an analysis procedure much like those previously discussed.

A.1 $1D$ predictions

At least two phenomenology papers[63][64] predict yields, energies and backgrounds for this exact analysis. Predictions from [64] are shown in Table **??** for line energies and yields. Given the success of similar calculations for the nP states these predictions are expected to be fairly accurate in line energies and yields.

Following the predictions in Table **??** we expect four main peaks in the narrow energy range $87 - 111$ MeV, with each peak corresponding to two or more superimposed peaks. All other peaks are at higher energies with no hope of disentangling them,

though in principle fits to the high energy peaks could be used to make constraints on the relative yields of the signal region photons. The signal spectrum will combine photons from the first two transitions ($\gamma_{3S\to 2P}$ and $\gamma_{2P\to 1D}$) in the energy window $40 - 140\,\mathrm{MeV}$. The degree of contamination from $3S \to 2P \to 2S \to 1P \to 1S$ peaking background will depend on the success of the rejection procedure for this background.

A.2 $1D$ reconstruction

Reconstruction proceeds in three parallel hypotheses, allowing independent evaluation of the signal and main background modes. The three modes are the signal ($3S \to 2P \to 1D \to 1P \to 1S$), "$\pi^0\pi^0$" ($\Upsilon(3S) \to \pi^0\pi^0\Upsilon(1S)$) and "2S" ($3S \to 2P \to 2S \to 1P \to 1S$). Details of reconstruction for each of these three hypotheses are given here.

A.2.1 The signal hypothesis

Reconstruction of the signal mode proceeds in analogy to reconstruction used on the other two-photon modes. The masses of the $\Upsilon(1S)$ and $\Upsilon(3S)$ candidates are fixed (as well as the beam energy and location) and the fit probability of each set of $\gamma\gamma\gamma\gamma\mu^+\mu^-$ candidates in each event under the signal hypothesis is calculated.

Selection of the best signal candidate in each event also proceeds as before, with both soft photons $3S \to 2P$ and $2P \to 1D$ required to have timing within 2.0σ of the hard $1P \to 1S$ photon and all four photons are required to fall within appropriate energy windows. The highest-probability cascade candidate is then chosen in each event provided the probability exceeds 10^{-5}.

A.2.2 The $\pi^0\pi^0$ hypothesis

A significant source of four-photon background with a true $\Upsilon(1S)$ is the cascade $\Upsilon(3S) \to \pi^0\pi^0\Upsilon(1S)$, with a branching fraction of around 2%. The four photons in this cascade add to the $\Upsilon(1S)$ mass to properly reconstruct the $\Upsilon(3S)$ mass, making

Transition		Energy (MeV)	Yield
$3S \to$	$2P(0)$	123	7
	$2P(1)$	100	**643**
	$2P(2)$	87	**266**
$2P(0) \to$	$1D(1)$	81	7
$2P(1) \to$	$1D(1)$	106	**93**
	$1D(2)$	99	**551**
$2P(2) \to$	$1D(1)$	119	2
	$1D(2)$	112	**73**
	$1D(3)$	107	**190**
$3S \to$	$2P(0)$	123	24
	$2P(1)$	100	587
	$2P(2)$	87	324
$2S \to$	$1P(0)$	162	27
	$1P(1)$	131	619
	$1P(2)$	110	290
(Mult.)		87	$266 + \epsilon_{2S}324$
(Mult.)		100	$1194 + \epsilon_{2S}587$
(Mult.)		107	283
(Mult.)		111	$73 + \epsilon_{2S}290$
$2S \to$	$1P(1)$	131	$\epsilon_{2S}619$

Table A.1: Summary of the relevant predictions from a phenomenological treatment of the $3S \to 2P \to 1D \to 1P \to 1S$ analysis [64], ignoring all transitions above 170 MeV for the signal (top) and background $3S \to 2P \to 2S \to 1P \to 1S$ processes (middle) assuming a reconstruction efficiency of 0.2. The bottom section contains combined peaks (within ± 1 MeV) for the significant yields (**bold** entries) assuming the $2S$ background mode is rejected with an efficiency of ϵ_{2S}. A fitting scheme would need to distinguish four peaks in the narrow range $87 - 111$ MeV with no way to disentangle the multiple modes contributing to each peak.

this background source resistant to selections on the signal fit probability, unlike in the previous analyses.

We reconstruct the $\pi^0\pi^0$ cascade in parallel with the signal cascade, with $\Upsilon(1S)$ and $\Upsilon(3S)$ masses fixed. For each event we find the cascade candidate with the highest fit probability under the $\pi^0\pi^0$ hypothesis. If either of the π^0 candidates in this "best" $\pi^0\pi^0$ cascade has a mass in the range $127 - 143\,\mathrm{MeV}$ the event is rejected.

The $\pi^0\pi^0$ background should be smooth in any of the signal regions; any stray $\pi^0\pi^0$ events not rejected in this manner contribute to the continuum background and do not bias the peak yields.

A.2.3 The $3S \to 2P \to 2S \to 1P \to 1S$ hypothesis

As previously discussed, the four-photon cascade through the $\Upsilon(2S)$ is particularly problematic since it contributes peaks in the signal region and is similar in total yield to the signal mode. The small kinematic difference between this background cascade and the signal cascade can be exploited to reject most or all of these events.

For each event the $3S \to 2P \to 2S \to 1P \to 1S$ cascade is reconstructed with all three Υ masses fixed. In principle the cascade fit probability can be used to discriminate against these events while preserving signal-mode events. However, of the selected "best" signal mode candidates exactly zero of them have a non-zero fit probability under the $2S$ hypothesis; this may be due to effective selections.

A.3 $1D$ fitting

Given the large number of peaks in the signal region it is essential to identify all usable parameter constraints. Consider the four significant background peaks associated with the $3S \to 2P \to 2S \to 1P \to 1S$ cascade; the total yield of each step is constant, so we can write (ignoring the $J = 0$ peak):

$$N_{3S\to 2P(1)} + N_{3S\to 2P(2)} = N_{2S\to 1P(1)} + N_{2S\to 1P(2)}. \tag{A.1}$$

Additionally, the relative yields $N_{3S\to 2P(2)}/N_{3S\to 2P(1)}$ and $N_{2S\to 1P(2)}/N_{2S\to 1P(1)}$ can be fixed to those obtained for the $3S \to 2P \to 2S$ and $2S \to 1P \to 1S$ analyses with

the three peak means fixed to the solutions from those fits. A single yield parameter can then describe the entire $2S$ spectrum; choosing $N_{2S\to1P(2)}$ for this we write:

$$N_{2S\to1P(2)} = N_{2S\to1P(1)}\frac{5286}{9644}, \tag{A.2}$$

$$N_{3S\to2P(1)} = N_{2S\to1P(1)}\frac{1+\frac{5286}{9644}}{1+\frac{5446}{11380}}, \tag{A.3}$$

$$N_{3S\to2P(2)} = N_{2S\to1P(1)}\frac{1+\frac{5286}{9644}}{1+\frac{11380}{5446}}, \tag{A.4}$$

with $\mu_{2S\to1P(1)} = 128.0\,\text{MeV}$, $\mu_{2S\to1P(2)} = 108.0\,\text{MeV}$, $\mu_{3S\to2P(1)} = 98.4\,\text{MeV}$, and $\mu_{3S\to2P(2)} = 85.1\,\text{MeV}$.

For the signal contribution we use four peaks which correspond to:

- **1:** $3S \to 2P(2)$
- **2:** $[3S \to 2P(1)] + [2P(1) \to 1D(2)]$
- **3:** $[2P(1) \to 1D(1)] + [2P(2) \to 1D(3)]$
- **4:** $2P(2) \to 1D(2)$

Despite some components of these peaks having fix-able means we float these four in the non-overlapping ranges μ_1: $(83, 87)\,\text{MeV}$, μ_2: $(96, 100)\,\text{MeV}$, μ_3: $(103, 108)\,\text{MeV}$ and μ_4: $(108, 113)\,\text{MeV}$. These assumptions may not be correct, but are an interesting test of the phenomenological predictions.

All peaks share all shape parameters which are taken from the similar-energy $3S \to 2P(1)$ solution in the main analysis. These parameters are not expected to be exactly right given the different fit probability selection and number of photons in this analysis, but a good estimation of the line shape is expected given the poor signal mode statistics.

A.4 $1D$ results

Fig. **??** shows the combined spectrum of $\gamma_{3S\to2P}$ and $\gamma_{2P\to1D}$ with a signal fit probability selection of 10^{-8}, the $\pi^0\pi^0$ selection described above and $S_{max} = 2.0$.

Assuming for the moment that the $2S$ cascade contribution has the correct yield (which is not clear), the only peaks with any significance are peaks 1 and 2. Peak 1 is completely uninteresting – it is entirely due to the $3S \to 2P(2)$ transition which has been described in the previous analyses. Peak 2 contains an undisentangleable combination of $3S \to 2P(1)$ (not interesting) and $2P(1) \to 1D(2)$. The second component of peak 2 corresponds to the only $\Upsilon(1D)$ state yet observed. The unobserved $1D$ states are both contained in a single peak – Peak 3 – which has zero yield.

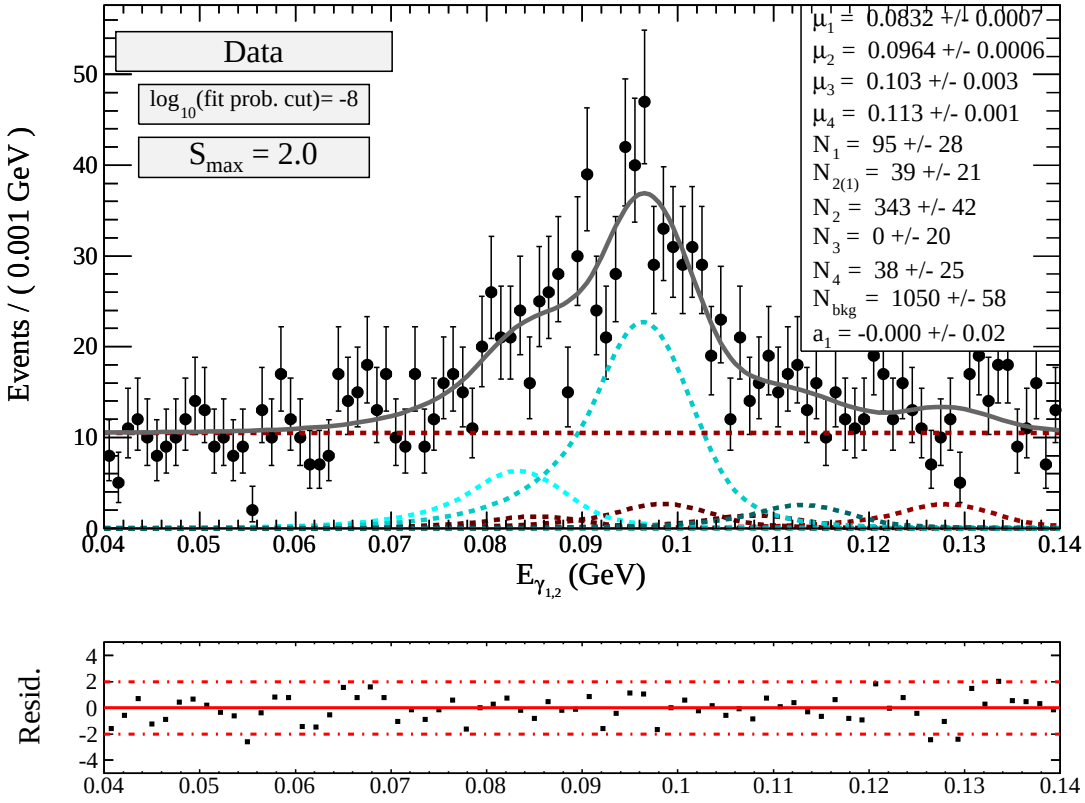


Figure A.1: Combined spectrum of $\gamma_{3S\to 2P}$ and $\gamma_{2P\to 1D}$ candidates as described in the text with fit. The red curves are the fitted $3S \to 2P \to 2S \to 1P \to 1S$ peaks and the blue peaks are the extracted signal peaks with definitions as described in the text.

A.5 $1D$ discussion

With an aggressive timing selection, π^0 veto and apparently little $2S$ background, an over-constrained fit fails to find any evidence of the unobserved $\Upsilon(1D)$ states. Some work could be done to further purify the $\Upsilon(1D)$ spectrum and perhaps constrain the fits further. However, this exercise is likely to be futile with *BABAR* data given the detector resolution, statistics and fitting challenges described here.

Appendix B

An attempt at an observation of the $\chi_{bJ}(3P)$

The radiative bottomonium sprectroscopy presented in this document is only possible below the open-flavor threshold where EM processes are competitive. However, above the threshold *BABAR* has collected 424.2fb^{-1} of data at the $\Upsilon(4S)$ resonance, a large enough sample that the soft transition $\Upsilon(4S) \to \chi_{bJ}(3P)$ may be visible despite the overwhelming hadronic width of the $\Upsilon(4S)$. A search for this transition is presented in this appendix.

The triplet $3P$ states were first observed [but not differentiated] by the ATLAS experiment in 2012 [62] in hard transitions to $\Upsilon(1,2S)$ with energies suggesting $\sim$ 50 MeV photons in the soft transition. In this analysis ATLAS benefitted from a high production rate but suffered from poor resolution that prevents disentanglement of the three $\chi_{bJ}(3P)$ states. The transition energy suggested by ATLAS is in a region where beam photons are extremely copious; if it is possible to see this transition at *BABAR* the timing selection will be required.

B.1 $3P$ data

We use Runs 5 and 6, representing roughly 1/2 of the total data taken at the $\Upsilon(4S)$ resonance from `AllEvents-Run5-R24-v10` and`AllEvents-Run6-R24-v10` with `UpsilonFilter` which only processes 4-prong events. Candidate pions come from

`piCombinedLoose` and otherwise all other reconstruction details are identical to the previous analyses.

B.2 $3P$ event selection and reconstruction

At the $\Upsilon(4S)$ resonance the $S \to P$ branching fraction is roughly two orders of magnitude suppressed relative to the $\Upsilon(3S)$ case. However, μ pair production with associated ISR and FSR photons is a continuum process and is not suppressed in the same way. This suggests that the $\mu^+\mu^-\gamma\gamma$ mode will be overwhelmed with $\mu^+\mu^-(\gamma)$ background. We instead look for $\mu^+\mu^-\pi^+\pi^-\gamma\gamma$ final states by reconstructing the following cascades, where $nS \xrightarrow{\pi\pi} mS$ denotes the transition $\Upsilon(nS) \to \pi^+\pi^-\Upsilon(mS)$ and the final state is implied to decay via two muons:

- $4S \to 3P \to 3S \xrightarrow{\pi\pi} 2S$
- $4S \to 3P \to 3S \xrightarrow{\pi\pi} 1S$
- $4S \to 3P \to 2S \xrightarrow{\pi\pi} 1S$

Combining these three modes we expect a clean spectrum with largely beam photon sources but likely with poor signal-mode statistics.

Reconstruction is analogous to the previous analyses, with fixed masses for all three of the mS states involved in each mode to minimize $\pi^0\pi^0$ background. Best event selection is also identical with $S_{max} = 2.0$.

B.3 $3P$ results

Fig. **??** shows the spectrum from $4S \to 3P$ candidates in the three combined radiative/hadronic cascade modes. No hint of a signal near the ATLAS mass of 50 MeV is observed and we conclude that no $3P$ signal is observable at *BABAR*.

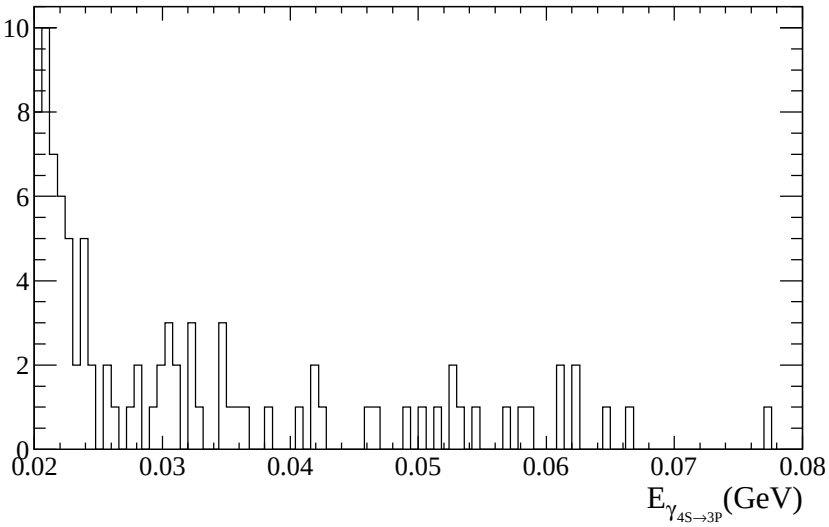


Figure B.1: Soft photon candidates from the $4S \to 3P$ transition in events as described in the text. No possibility for observation is seen.

Appendix C

EMC Timing

An improvement of the EMC timing for discrimination against out-of-time background is discussed.

C.1 Timing introduction

A certain class of analyses can benefit substantially from utilizing the timing information provided by *BABAR*'s Electromagnetic Calorimeter [EMC]. Specifically, photon candidates coming from machine background arrive "out of time" – they are not correlated in time with other candidates in the event or with the trigger. All clusters associated with final-state particle candidates in physics events will occur "in time" – arrival times are essentially identical for all clusters associated with the same event and the offset from the trigger is fixed. In particular, analyses with photons below $\sim 100\,\mathrm{MeV}$ can see tremendous amounts of out-of-time background. However, the EMC timing is uncalibrated and poorly understood and thus has not been used as a standard background discriminant in *BABAR* analyses. A comprehensive offline EMC timing calibration is presented herein.

C.2 EMC timing pipeline

A brief discussion of the source of the timing information and its properties is presented here. A more extensive description can be found in [48] and [42].

C.2.1 A digi is born

Energy deposited in one of the 6580 CsI crystals comprising the EMC produces a light pulse that is detected by a photodiode mounted to the rear of the crystal. This signal passes through a preamp and shaping circuit and is then digitized at a sample rate of $269ns$ and read onto a circular buffer. Upon arrival of the trigger, 64 bins are read from the circular buffer and the signal peak is expected between bins 21 and 29 of the sampled waveform. The "moment time" t is calculated in this window:

$$t = \frac{\sum_{i=21}^{29} E_i t_i}{\sum_{i=21}^{29} E_i} \tag{C.1}$$

The event time t_0 identifies the energy-weighted mean time of the samples above 30 MeV within the buffer, not limited to the narrow peak window:

$$t_0 = \left. \frac{\sum_{i} E_i t_i}{\sum_{i} E_i} \right|_{E_i > 30MeV} \tag{C.2}$$

The event time and the moment time differ substantially only if the peak does not fall within the expected bins. As an L3-level online out-of-time crystal-level background filter, the feature extraction algorithm imposes a $|t - t_0| < 120ns$ requirement on the digitized waveform. For waveforms that pass this cut, the moment time t is multiplied by $4ns$ and bundled with the peak crystal energy [together referred to as "digi"] and associated with other digis to constitute clusters which are associated with candidate particles.

C.2.2 Properties of the moment time

During Run 1, cluster time was determined using a parabolic fit of the waveform peak. Beginning in Run 2 waveform timing has been determined solely using the moment time method, optimized for computing power and bandwidth limitations and not offline out-of-time background rejection. Calculating the time in this manner

introduces some complicating features:

- The moment time pulls towards the center of the window.
- The moment time pulls more strongly towards the center of the window for peaks that are further from the center.
- In-time peaks may not be perfectly centered in the window.

The net result of these effects is that the out-of-time background peaks nearly coincident with the in-time signal and measured time is strongly non-linear in "real" time ["nearly coincident" because in practice the window is not perfectly centered on the in-time peak]. However, for a single crystal the in-time peak is sharper than the out-of-time distribution, meaning that additional out-of-time discrimination may be possible offline.

Before cutting on the digi or cluster times, it is important to characterize the timing information returned by the EMC in reality. While the single-digi time is fairly straightforward, it becomes clear that use of multiple digis to calculate a cluster time is fraught. These issues are explored thoroughly in the remaining sections.

C.3 Digi and cluster timing

C.3.1 Sources of timing error

As discussed above, all EMC clusters associated with final state particles in a single physics event will arrive essentially simultaneously at the detector. The cluster times are made up of one digi time for each crystal represented in a cluster and these underlying digi times should be similarly identical. The data show that this is not the case, and optimal use of the EMC timing information will require careful study to mitigate the sources of the timing inconsistency.

Fig. **??** shows the distribution of raw digi times for a small sample of `CalorNeutral` [single-bump EMC clusters not associated with a charged track] candidates in Run 7 with $E_{cluster} > 200\,\mathrm{MeV}$, representing almost entirely in-time clusters. The RMS error of $53ns$ arises from the following possible components:

- σ_{cryst}: The cumulative effects of various crystal timing offsets throughout the detector.
- σ_{trig}: The random event-to-event offset of the zero reference time.
- σ_{peak}: The fundamental uncertainty of the time measurement.
- $\sigma_{E_{digi}}$: Smearing of digi times due to a systematic digi energy bias in the digi time.

These variances cannot simply be added in quadrature since σ_{cryst} and σ_{trig} are functions of the number of digis being used and σ_{peak} is plausibly a function of E_{digi}. Measuring the relative importance of these variance contributions is essential to maximizing the utility of the timing information.

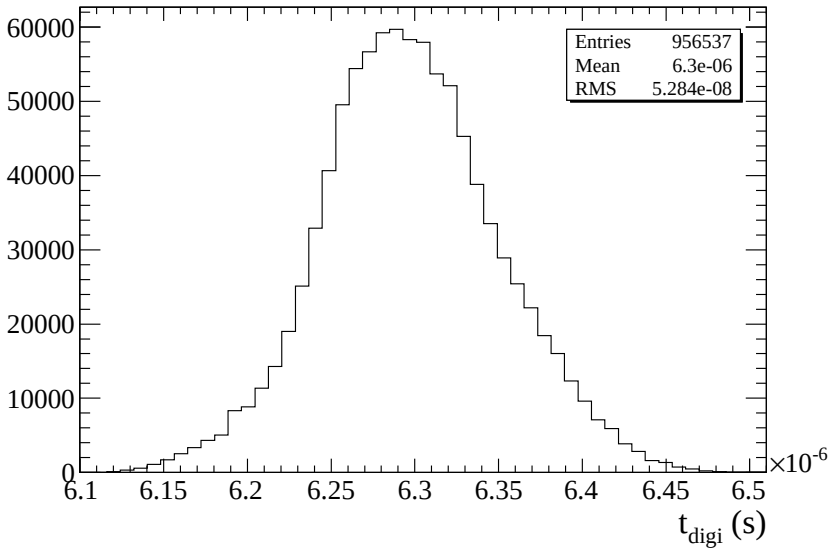


Figure C.1: Digi timing distribution for `CalorNeutral` candidates in Run7 with $E_{cluster} > 200\,\text{MeV}$. The time axis is measured in putative s but in practice this time is not linear in real time and the offset is not meaningful.

C.3.2 Digi time/energy relationships

In principle, both the digi time t_{digi} and the digi time error $\sigma_{t_{digi}}$ can depend on the digi energy E_{digi}. The following two sections investigate whether the energy dependence

of these quantities is substantial enough to require correction of the digi times.

$\overline{t_{digi}}$ vs. E_{digi}

Fig. ?? shows $\overline{t_{digi}}$ vs. E_{digi} for `CalorNeutral` candidates with $E_{cluster} > 200\,\mathrm{MeV}$ for a fixed crystal. Within the limit of the trigger jitter there is no detectable digi time bias as a function of digi energy.

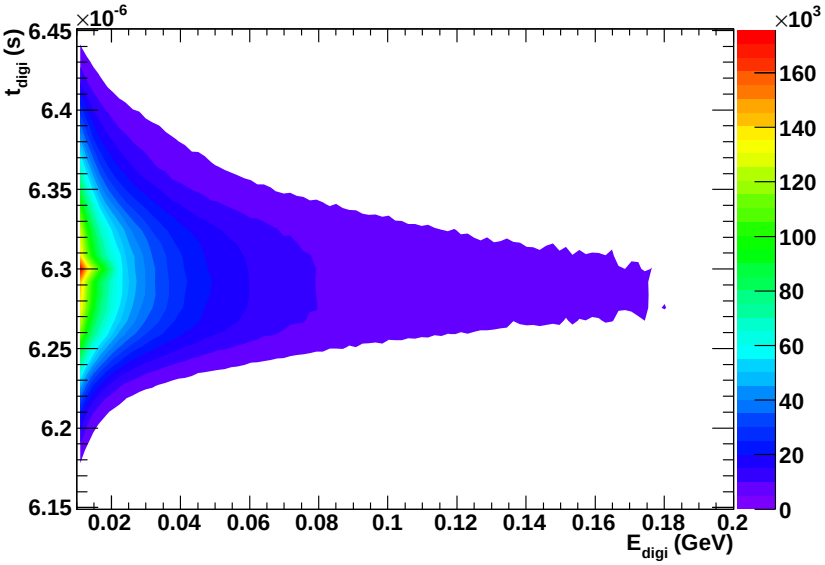


Figure C.2: Smoothed contour plot of t_{digi} vs. E_{digi} for digis with ($\phi_{digi} = 61, \theta_{digi} = 22$) belonging to a sample of `CalorNeutral` candidates with $E_{cluster} > 200\,\mathrm{MeV}$. If a systematic shift of digi times occurred as a result of increasing digi energy the thrust of the distribution would not be entirely horizontal. There is no evidence that the digi energy influences the mean expected digi time.

$\sigma_{t_{digi}}$ vs. E_{digi}

In constructing cluster times, the offline reconstruction uses a simple digi energy weighting, where the index i labels each digi in a cluster:

$$t_{cluster} = \frac{\sum_i E_i t_i}{\sum_i E_i} \tag{C.3}$$

This weighting is an implicit assertion that $\sigma^2_{t_{digi}} = 1/E_{digi}$ under the assumption that the variance in the digi times within a single cluster is basically Gaussian. The use of a precise maximum likelihood estimator may improve timing discrimination.

To determine the exact form of the $\sigma_{t_{digi}}$ vs. E_{digi} relationship, two fixed crystals are chosen [in this case crystals with the ϕ and θ indicies of $(\phi_1, \theta_1) = (60, 22)$ and $(\phi_2, \theta_2) = (61, 22)$]. These crystals are chosen to be adjacent in order to provide the maximum number of intra-event multiple-crystal hits. Using a fixed crystal pair eliminates crystal-to-crystal offsets and demanding that both digis come from the same cluster removes trigger jitter. Fig. **??** shows the distribution of the timing difference variance $\sigma_{\Delta t_{digi}}$ for a sample of unfiltered `CalorNeutral` candidates in Run 7. The energy range is binned into 20 unequal ranges and the middle of each range is taken as the energy value for that range. Both digi energies are required to be within the same energy range so that the form of the timing difference error $\sigma_{\Delta t_{digi}}$ can be assumed to be that of the single-digi timing error $\sigma_{t_{digi}}$. The power law fit gives an exponent value of -0.734 ± 0.005. Apparently the traditional weighting method did not give enough weight to higher-energy digis.

For combining single digi times throughout the remainder of this analysis, this form of the weighted sum will be used [in this and all future sums the digi index is suppressed; the sums are to be interpreted as being over all relevant digis or digi pairs]:

$$\bar{t} = \frac{\sum E^{1.47} t}{\sum E^{1.47}} \tag{C.4}$$

For weighting the time difference between two crystals with digi energies E_1 and E_2, introduce $\varepsilon_{1,2}$ for notational simplicity:

$$\varepsilon_{1,2} = 1/(\sigma^2_{t_1} + \sigma^2_{t_2}) = \frac{E_1^{1.47} E_2^{1.47}}{E_1^{1.47} + E_2^{1.47}} \tag{C.5}$$

The two-digi weighted time difference becomes:

$$\overline{\Delta t} = \frac{\sum \varepsilon_{1,2} [t_2 - t_1]}{\sum \varepsilon_{1,2}} \tag{C.6}$$

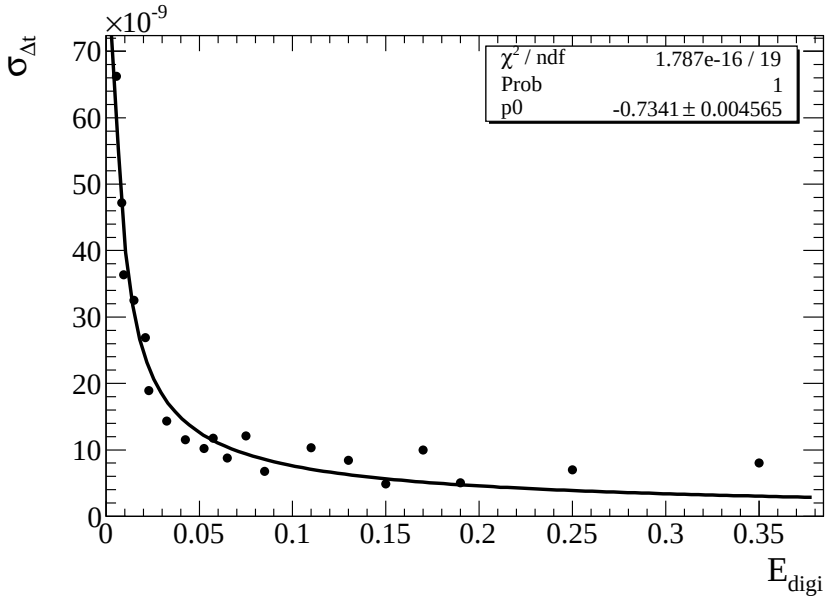


Figure C.3: Best fit power law curve for $\sigma_{\Delta t_{digi}}$ vs E_{digi}. The model equation is $A \cdot E_{digi}^{B}$, with A being irrelevant to the weights. The least-squares fit gives $B = -0.734 \pm 0.005$. Since E_{digi} is close to the same for both digis compared in the Δt measurement, the single-digi variance $\sigma_{t_{digi}}$ has the same power-law solution as the two-digi comparison. The maximum likelihood estimator becomes $1/\sigma_{t_{digi}}^2 = E^{1.47}$. This fit hypothesis is better in the low-digi energy region, but the overwhelming majority of the digis in a cluster are in the lower energies. The full fitted form of the single-digi variance is $\sigma_{t_{digi}}^2 = 1.05 ns^2 / E_{digi}^{1.47}$.

Peak measurement error

The moment time method for determining digi time is by its nature more accurate for more strongly peaking signals than signals with less contrast within the expected peak window. High-energy peaks thus yield better digi timing information than low-energy peaks. Improved digi timing determination cannot possibly exceed the precision of the peak error, so quantification of this uncertainty is needed in order to gauge the effectiveness of the corrections.

The time difference variance $\sigma_{\Delta t_{digi}}$ as a function of digi energy E_{digi} is shown in Fig. **??**. This time difference variance is simply related to the single-digi variance which comes entirely from peak uncertainty: $\sigma_{peak}(E_{digi}) = (1/\sqrt{2})\sigma_{\Delta t_{digi}}(E_{digi})$. σ_{peak} is as good as $\sim 7ns$ for very high-energy digis but averages to $\sim 20ns$ when all digis

are included.

Trigger jitter

Trigger jitter refers to the inconsistency of the definition of the time origin $t = 0$ in the digi waveform digitization stemming from effects in the trigger pipeline. All digis in all clusters associated with a single trigger have the same origin definition, but digis separated by events have different definitions. An empirical description of the effect that trigger jitter has on digi and cluster timing is pursued, ignoring for practical purposes its underlying source.

Looking at the energy bin $28\,\mathrm{MeV} < E_{digi} < 30\,\mathrm{MeV}$ in Fig. **??**, $\sigma_{peak} \approx 10ns$. Again taking a fixed crystal pair but now requiring that the two digi hits span events and fall within the same narrow energy range, the resultant distribution [Fig. **??**] represents the sum $\sqrt{\sigma_{peak}^2 + \sigma_{trig}^2} = 52ns$. Using the estimate for σ_{peak} above, it appears that $\sigma_{trig} \approx 50ns$. In this case, σ_{trig} is somewhat larger than the value that contributes to Fig. **??** because the latter data contains multiple digis in each event and the former requires each digi pair to straddle events. However, it appears that trigger jitter is the major component of the single digi timing variance.

σ_{cryst}

Crystals used in the EMC were sourced from various manufacturers and the front-end electronics contain shaping circuits of two different specifications. Fig. **??**(b) shows the average digi time [for $E_{digi} > 20\,\mathrm{MeV}$] in each of the EMC crystals in a clean in-time $E_{cluster} > 200\,\mathrm{MeV}$ photon sample. The large block of substantially higher-offset crystals spanning the barrel in θ for $39 \leq \phi \leq 68$ corresponds to a batch of shaping circuits with shaping times $\sim 48ns$ longer than those associated with the other crystals, and is responsible for the asymmetry in Fig. **??**. Horizontal bands in θ correlate to crystals sourced from different manufacturers; Fig. **??**(a) shows a color-coded map of the source of crystals in the barrel for comparison. In addition there are crystal-to-crystal inconsistencies within each region due to random crystal or electronic differences. No calibration or correction for these constant crystal-to-crystal time offsets is used in the cluster timing determination in reconstruction, leading to severely compromised timing information.

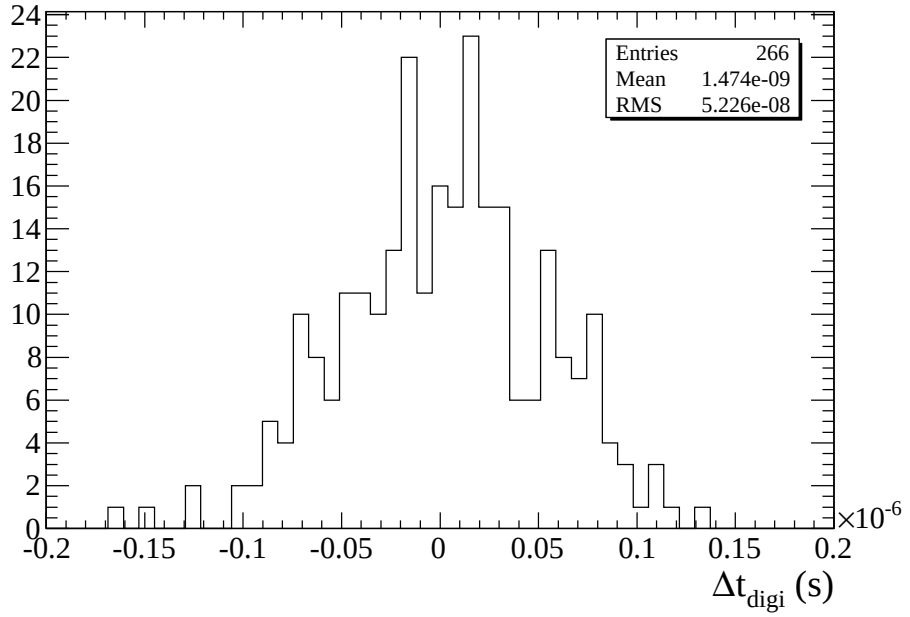


Figure C.4: The digi timing difference for adjacent crystals where the two digis chosen always occur in different events. Trigger jitter overwhelms peak measurement error for all but the lowest-energy digis.

In principle, Fig. **??** contains all the information necessary to make a crystal-by-crystal correction at the digi level. However, the precision of this correction is limited by trigger jitter. A similar correction using intra-event data should produce much more precise and accurate corrections.

The variance of these offset values is not a reasonable estimation of σ_{cryst} because the number of digi hits per crystal is strongly biased toward the high-shaping-time region. Plotting the time difference between two digis in the same event with both digi energies required to be greater than 50 MeV gives a distribution with an RMS error of $28ns$. Assuming that $\sigma_{peak} \approx 10ns$, this gives $\sigma_{cryst} \approx 26ns$. It is clear that crystal offsets are the most significant source of timing error for intra-event timing comparisons. Conveniently, these offsets are constant and correctable, and the remainder of this document addresses the optimal techniques for correcting the crystal offsets.

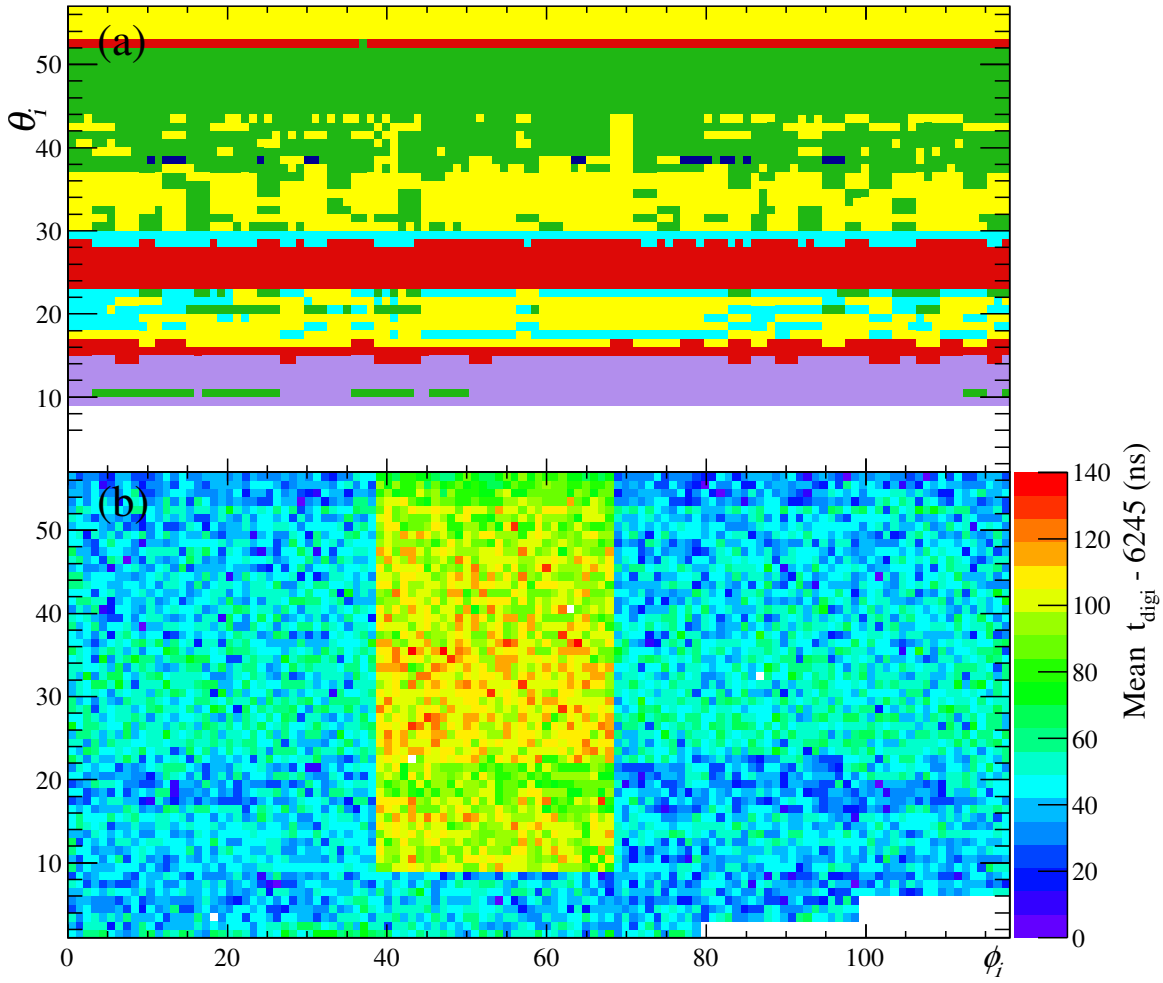


Figure C.5: (a) Color-coded distribution of crystal manufacturers in the barrel region. (b) Average t_{digi} (with a fixed offset subtracted for plotting purposes) in ns for a small sample of `CalorNeutral` events with $E_{digi} > 20\,\text{MeV}$ and $E_{cluster} > 200\,\text{MeV}$, giving almost exclusively in-time data. The horizontal bands correlate to the bands in crystal manufacturers in (a) and the vertical block corresponds to longer shaping time circuits in the front-end electronics. The four blank entries in (b) correspond to permanently dead crystals.

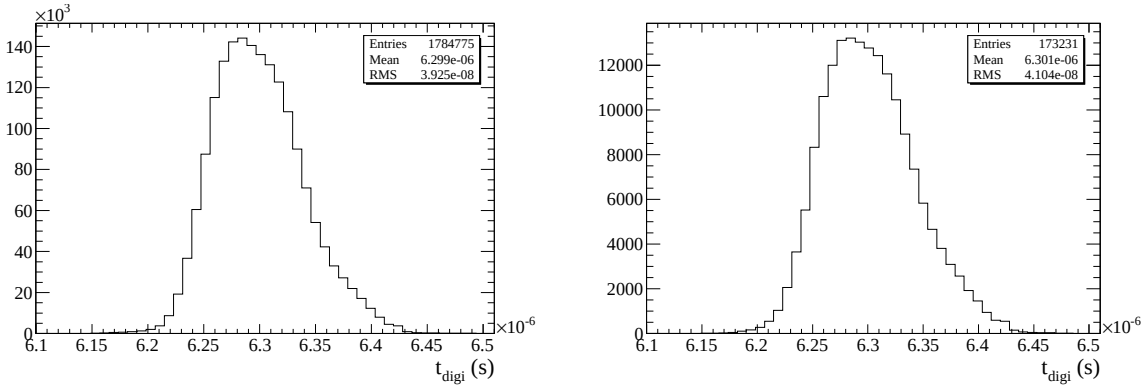
C.4 Crystal corrections

As demonstrated in the previous section, trigger jitter dominates the digi timing uncertainty. Precision determination of the crystal offsets must then use intra-event information. A method is presented here for combining the statistical power of large

datasets with the relative precision of intra-event timing comparisons.

C.4.1 Generating the crystal corrections

The crystal corrections are assumed to constitute constant offsets unique to each crystal. The corrections require a "pure" in-time source of digi data. Two extremely copious sources satisfy this requirement: high energy photons and all muons. Figs. **??** and **??** show t_{digi} distributions for qualifying digis comparing high-energy `CalorNeutral` clusters and muons from `muBDTLoose`. A "qualifying" digi is one with $E_{digi} > 20\,\text{MeV}$ and for the case of the high-energy neutral clusters the digi must be associated with a cluster with $E_{cluster} > 200\,\text{MeV}$. For reference, the digi timing distribution for `CalorNeutral` candidates with no cut on $E_{cluster}$ and with the same digi energy cut has an RMS error of $\sim 55ns$.



(a) t_{digi} distribution for all digis associated with `CalorNeutral` clusters with $E_{cluster} > 200\,\text{MeV}$ and $E_{digi} > 20\,\text{MeV}$.

(b) t_{digi} distribution for all digis associated with `muBDTLoose` clusters with with $E_{digi} > 20\,\text{MeV}$.

Muons are essentially minimum-ionizing particles in the EMC and leave a very consistent $\sim 200\,\text{MeV}$ in the calorimeter regardless of muon momentum for normally incident tracks. This sharp energy peak is only visible in cluster energy or in the digi energy in clusters with only one contributing digi. Intra-event corrections require the use of multiple digis per cluster and with the 200 MeV being shared over several crystals the timing resolution benefit of such a copious source of high-energy digis vanishes.

Additionally, the high-energy neutral clusters have two large advantages over the μ data: [1] more than an order of magnitude more qualifying digis and [2] a factor of ~ 4 increase in the number of qualifying digis per cluster (14.2 vs. 3.8). The number of qualifying digi pairs in a cluster increases combinatorically with cluster size, so there is a great benefit to using large clusters. A comparison of the t_{digi} distributions from muons and high energy photons (see Figs. **??** and **??**) demonstrates that the timing signatures of the two sources are indistinguishable except for the difference in statistics, implying that the high-energy photon sample is comparably clean as the muon sample. Accordingly, the high-energy photon sample is used to generate the timing events.

Seed data

With a suitable selection of in-time digi pairs identified, the process of generating the individual crystal offsets proceeds as follows.

Ntuples are generated from all `CalorNeutral` candidates with $E_{cluster} > 200\,\mathrm{MeV}$ in Run 7 data. Each entry in the ntuple consists of E_{digi}, t_{digi}, ϕ and θ for two digis that have $E_{digi} > 20\,\mathrm{MeV}$ and occur in the same event. Each even-ϕ and even-θ crystal is designated a "seed" crystal. For a seed crystal s, digi pairs where one digi occurs at the seed crystal are identified. For each non-seed crystal i with crystal indicies (ϕ_i, θ_i), the contribution of multiple digi pair timing differences is averaged in a weighted sum using the weights from Eqn. **??**. The numerator N_s and denominator D_s are left separate to facilitate combination with other seeds later, and the total variance squared S_s is also saved. The summation indicies have been suppressed; the summations run over all qualifying digi pairs:

$$N_s(\phi_i, \theta_i) = \sum \varepsilon_{s,i} \left[t_i - t_s \right] \tag{C.7}$$

$$D_s(\phi_i, \theta_i) = \sum \varepsilon_{s,i} \tag{C.8}$$

$$S_s(\phi_i, \theta_i) = \sum \frac{1}{\varepsilon_{s,i}} \tag{C.9}$$

Combining seeds

Each seed now contains data to correct each crystal in reference to the seed, but the seeds need to be corrected in reference to each other. First, for two neighboring seed crystals $s1$ and $s2$, crystals (ϕ_i, θ_i) are identified in which both seeds have nonzero values $N_{s1}(\phi_i, \theta_i)$ and $N_{s2}(\phi_i, \theta_i)$. Each non-seed crystal provides components to the weighted seed-to-seed offsets labeled N_{s2-s1}, D_{s2-s1} and S_{s2-s1} in analogy to Eqn.'s **??**, **??** and **??**:

$$N_{s2-s1} = \sum \left[\frac{1}{S_{s1}(\phi_i, \theta_i) + S_{s2}(\phi_i, \theta_i)}\right] \left[\frac{N_{s2}(\phi_i, \theta_i)}{D_{s2}(\phi_i, \theta_i)} - \frac{N_{s1}(\phi_i, \theta_i)}{D_{s1}(\phi_i, \theta_i)}\right] \tag{C.10}$$

$$D_{s2-s1} = \sum \left[\frac{1}{S_{s1}(\phi_i, \theta_i) + S_{s2}(\phi_i, \theta_i)}\right] \tag{C.11}$$

Recognizing that the seed-to-seed time offset N_{s2-s1}/D_{s2-s1} is a constant, the numerator for the combined seed crystal offsets [designated $s1 + s2$] is:

$$N_{s1+s2}(\phi_i, \theta_i) = N_{s1}(\phi_i, \theta_i) + \left[N_{s2}(\phi_i, \theta_i) - D_{s2}(\phi_i, \theta_i)\left(\frac{N_{s2-s1}}{D_{s2-s1}}\right)\right] \tag{C.12}$$

$$D_{s1+s2}(\phi_i, \theta_i) = D_{s1} + D_{s2} \tag{C.13}$$

Combining seed crystals requires some care; the seed-to-seed time offset N_{s2-s1}/D_{s2-s1} is best known for neighboring seeds. Although combining all seeds will invariably choose one seed as the baseline time reference, attempting to directly offset each seed individually with respect to a single reference seed will give results that decrease in accuracy as the distance between the reference crystal and the crystal being offset increases. To eliminate this effect, seed combination proceeds in the following manner:

- Neighboring seeds in ϕ are combined using Eqns. **??** and **??**. These seeds are called 2^{nd}-generation seeds and the total number of seeds is cut in half, from 1650 to 825.

- Neighboring 2^{nd}-generation seeds in θ are combined, leaving 412 3^{rd}-generation seeds and 1 2^{nd}-generation seed.

- Neighboring 3^{rd}-generation seeds in ϕ are combined.
- Etc.

The process continues until the 11^{th}-generation seed covers every seed crystal and contains the components $N_{s1+s2+...}$ and $D_{s1+s2+...}$ to correct each crystal to the time of the crystal at $(\phi, \theta) = (0, 2)$. These corrections $\tau(\phi, \theta)$ are given by:

$$\tau(\phi, \theta) = \frac{N_{s1+s2+...}(\phi, \theta)}{D_{s1+s2+...}(\phi, \theta)} \tag{C.14}$$

A look-up table of the corrections is loaded at run-time to correct all digi times and compute a new cluster time from the weighted sum of all cluster digis i:

$$t_{cluster} = \frac{\sum E_i^{1.47} \left[t_i - \tau(\phi_i, \theta_i) \right]}{\sum E_i^{1.47}} \tag{C.15}$$

The cluster timing variance is calculated from the weighted mean error using the full form of the fitted maximum likelihood estimator from Fig. **??**:

$$\sigma_{t_{cluster}} = \frac{1}{\sum 9.5 \times 10^{18} E_i^{1.47}} \tag{C.16}$$

For large clusters a difference-of-squares approach can be used to measure rather than predict the cluster timing variance, but this method breaks down with small clusters due to the granular nature of the digi times. For clusters from low-energy photons or particularly muons there is a substantial likelihood that all crystals associated with the cluster will report the exact same time, resulting in variances that are artificially zero.

No cluster timing error is calculated in the standard reconstruction.

C.4.2 The corrections

As a preliminary check to confirm that the corrections successfully remove crystal-to-crystal offsets in data, the data from Fig. **??**(b) are plotted in Fig. **??** with the calculated offsets subtracted. Although using the average digi time from Fig. **??** as the corrections would have successfully removed the gross crystal offset features, Fig.

?? shows the value of the calculated corrections over the raw digi average times. The main features and most of the variance shown in this figure are found in the average digi time data but not the calculated offsets.

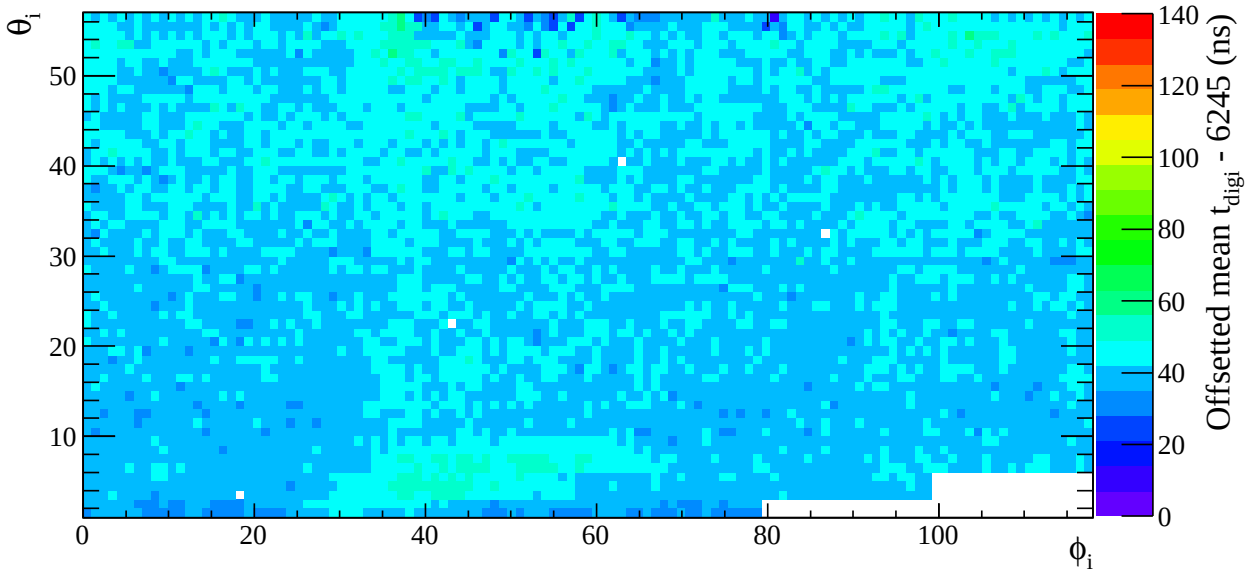


Figure C.6: Average digi time from Fig. **??**(b) with calculated offsets subtracted in the same scale. Using these imprecise inter-trigger data to test the precise intra-trigger offsets at most proves that most of the gross features in the uncorrected plot are removed and that variance is significantly reduced. Features in the corrected plot represent biases in the inter-trigger average digi time and not in the corrections themselves.

The corrections successfully remove the crystal-to-crystal offsets to a very high precision, but the effect that this has on the out-of-time discrimination of EMC candidates requires careful consideration and is addressed in the following section.

C.5 Proxy studies

The Monte Carlo EMC timing information does not include the correct crystal-to-crystal offsets or shaping time information, therefore it cannot be used as a reliable modeling tool for the performance of timing cuts. A well-understood proxy peak in Run 7 data is used to measure in-time and out-of-time efficiencies for various cuts on

the timing difference significance $S_{i,j}$ between clusters i and j:

$$S_{i,j} = \frac{|t_j - t_i|}{\sqrt{\sigma_{t_i}^2 + \sigma_{t_j}^2}} \tag{C.17}$$

This timing difference significance is the primary intra-event metric for timing separation between two clusters. In the remainder of this appendix S_{max} will represent the maximum value of S allowed by the timing cut and the indicies will be dropped.

C.5.1 The proxy mode

To extract in-time and out-of-time efficiencies for various S cuts a proxy peak is needed. This is an in-time peak that is well-understood and sitting above exclusively out-of-time background. The peak should not be a signal peak for the analysis the cut is used in but should be a good model for it. An evaluation of the timing selection using the $2S \to 1P \to 1S$ signal mode is presented here.

In the $2S \to 1P \to 1S$ cascade the energy of $\gamma_{soft}(\gamma_{2S\to 1P})$ is low enough that machine background is a dominant source of background. Although the precise in-time and out-of-time efficiencies calculated in this section are only applicable to this bottomonium analysis, the results can also stand as validation of the timing corrections.

Crystal offsets are calculated independently for different runs, and here only Run 7 is treated. The relevant bottomonium transitions are:

- $\Upsilon(2S) \to \chi_{bJ}\gamma_{soft}$
- $\chi_{bJ} \to \Upsilon(1S)\gamma_{hard}$
- $\Upsilon(1S) \to \mu^+\mu^-$

The background consists almost entirely of out-of-time photons reconstructed under the γ_{soft} hypothesis with the other final state candidates coming from various sources. The proxy peak comes from a closely related cascade: $\Upsilon(2S) \to \pi^0\pi^0\Upsilon(1S); \Upsilon(1S) \to \mu^+\mu^-$, where one of the π^0 candidates – referred to as the "proxy" – decays to γ_{soft} and one in-time photon [with energy over 200 MeV] and the other π^0 candidate – the "spare" – is required to have the correct mass.

Out of the $\sim$ 100 million $\Upsilon(2S)$'s detected in Run 7, $\sim 215,000$ are expected to decay via this exclusive decay mode. Furthermore, requiring that one of the two π^0's in the transition from the $\Upsilon(2S)$ carry the overwhelming majority of the transition energy reduces the sample size drastically. On the other hand, this yields a perfect proxy peak: the invariant mass distribution of the proxy is an in-time peak sitting above exclusively out-of-time background and the effect of the timing cut on the relative size of these two contributions should model the signal exactly.

C.5.2 The proxy peak

The purpose of the proxy study is to quantify the in-time and out-of-time efficiencies for various timing significance cuts on a single photon in the γ_{soft} energy range [though the same techniques can be used in any relevant energy range]. The di-γ invariant mass distribution of the daughters of the proxy π^0 candidate models the in-time and out-of-time efficiency given that only one of the daughter photon candidates – the one labeled γ_{soft} – is likely to be out-of-time.

Given an acceptable rejection of signal events, there is one remaining source of in-time background in the invariant mass distribution of the proxy π^0. Incorrect labeling of the four final-state photons will add a non-peaking contribution that mimics the non-peaking out-of-time distribution while both photons are in-time. A suitable rejection of mis-labeled reconstructions is critical. Following is a discussion of the generation and analysis of the proxy peak.

Proxy reconstruction

For each $\Upsilon(2S)$ cascade, final state muons from `muBDTLoose` and photons from `goodPhotonLoose` are combined under both the proxy and signal hypotheses. Each cascade is fit using `TreeFitter` with `Beam` and `Energy` constraints. Cascades with a fit probability of 0 are discarded and cascades are labeled proxy if the fit probability under that hypothesis exceeds that of the signal hypothesis. Daughter π^0 candidates in identified proxy cascades that have one daughter photon in the γ_{soft} window and the other one over 200 MeV are retained along with the spare π^0 from the same cascade, while all others are discarded. Fig. **??** shows the energy spectrum of the photon reconstructed as

γ_{soft} in the proxy hypothesis in the γ_{soft} energy window, confirming that the proxy identification procedure rejected essentially all photons from the signal cascade.

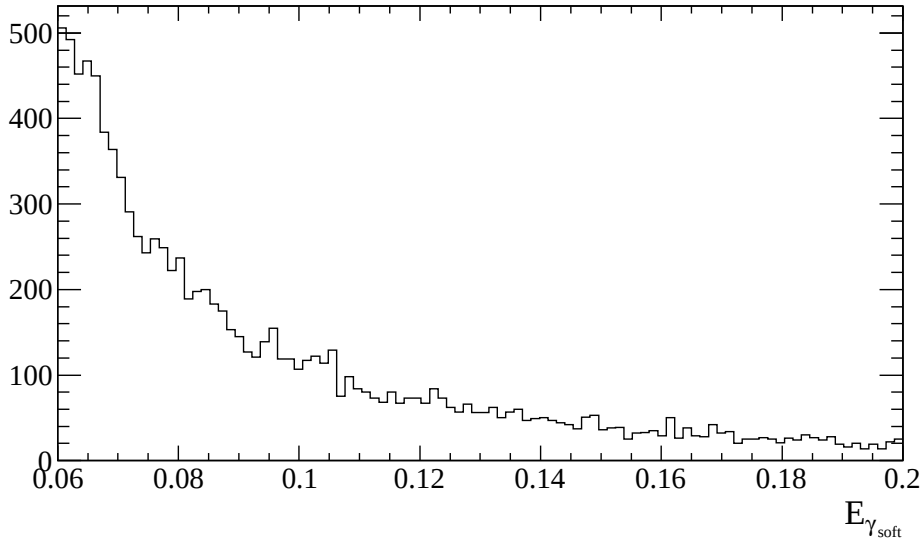


Figure C.7: Photon energy spectrum for the proxy π^0 daughter labeled γ_{soft}. All three peaks from the $\Upsilon(2S) \to \gamma_{soft}\chi_{bJ}$ transitions occur in this window; the absence of any recognizable peaks indicates that the proxy selection includes very few signal cascades.

For each pair of proxy and spare π^0 candidates identified, the invariant mass of the γ_{soft} candidate combined with each spare π^0 daughter is calculated, and the same for the γ_{hard} candidate. Events where at least one of these four invariant masses falls within the range $100 - 155\,\mathrm{MeV}$ are rejected, completely removing the only potential source of in-time background: mis-labeled π^0 daughters. Some out-of-time background is also rejected with this cut, but the absolute scale of the out-of-time background is not important provided that it is large enough to fit accurately. Fig. **??** shows the invariant mass distribution for each pair of final state photons except those matching the cascade hypothesis. The peak centered on $135\,\mathrm{MeV}$ comes from mis-labeled π^0 daughters and the rejection window is shown.

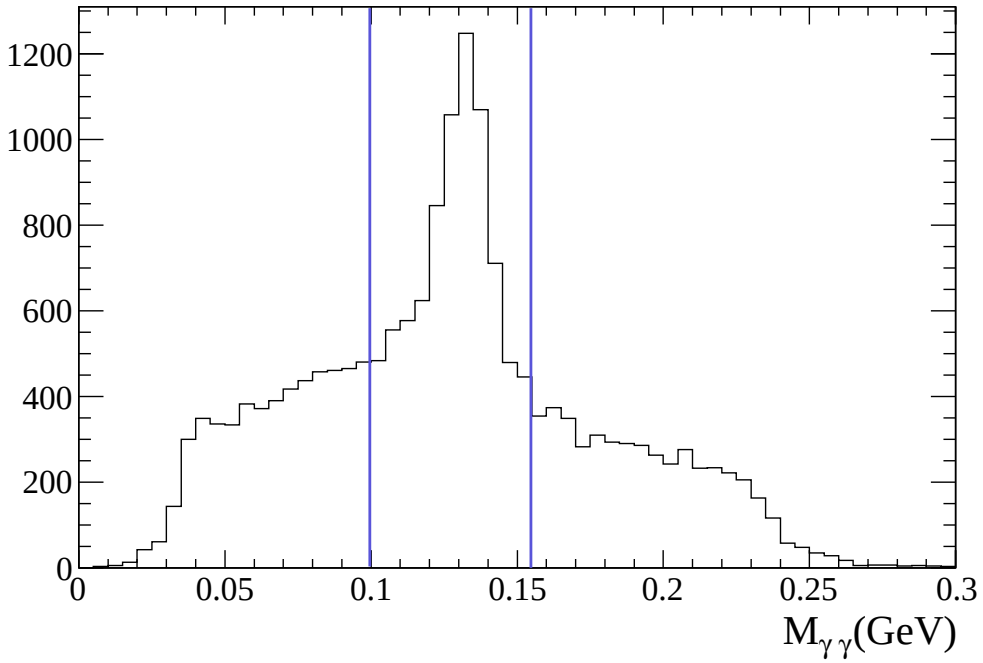


Figure C.8: Invariant mass distribution of final state photon pairs except the assigned proxy hypothesis. The region between the blue lines is rejected to remove the peak in the π^0 region which is the result of incorrect candidate labeling. The "background" in this plot is π^0 candidates with one or more out-of-time photon candidates or in-time candidates with the correct reconstruction.

Proxy peak fitting

The uncut proxy π^0 mass plot is shown in Fig. **??** with a 2^{nd}-order polynomial measuring the out-of-time background yield Y_{out} and a Novosibirsk function centered on the nominal π^0 mass measuring the in-time yield Y_{in}. Fitting proceeds via a binned maximum-likelihood fit.

Efficiencies/objective function

The cut requires that the timing significance S be less than the cut value S_{max}. An objective function $f(S_{max}) = \pi(S_{max}) \cdot \epsilon_{in}(S_{max})$ quantifies the quality of the cut, where $\pi(S_{max}) = Y_{in}(S_{max})/[Y_{in}(S_{max}) + Y_{out}(S_{max})]$ is the signal purity and $\epsilon(S_{max}) = Y_{in}(S_{max})/Y_{in}(S = \infty)$ is the signal efficiency. Fig. **??** shows both $f(S_{max})$

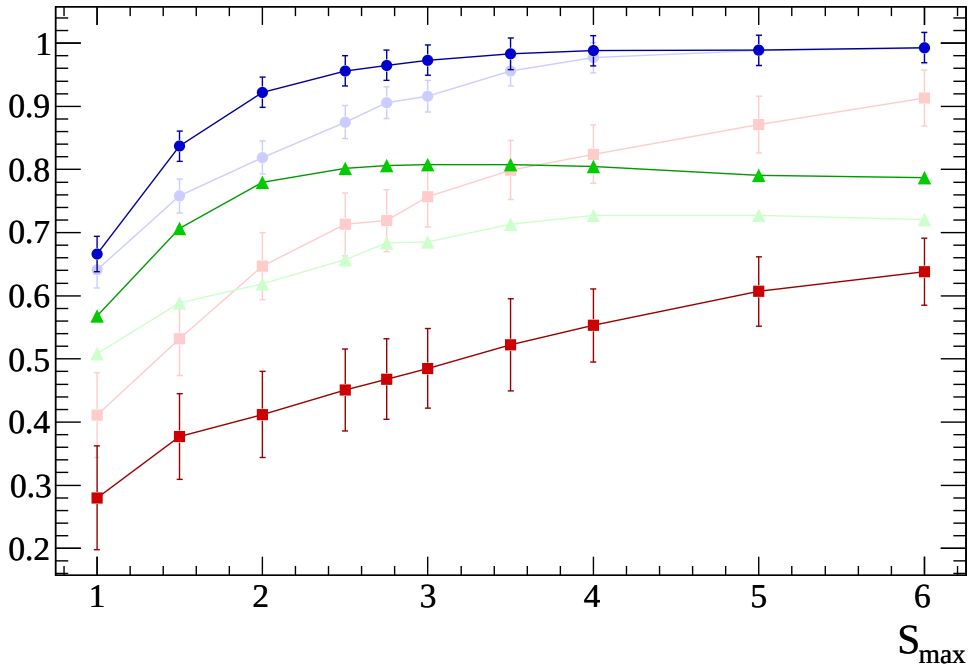


Figure C.9: In-time efficiency ϵ_{in} [blue dots], out-of-time efficiency ϵ_{out} [red squares] and the objective function f [green triangles] as functions of the timing significance cut S_{max}. The faint curves represent the uncorrected timing data and the dark curves are for the corrected timing information. Both in-time retention and out-of-time rejection are significantly enhanced with the corrections. The peak of the objective function occurs around $S_{max} = 3.0$, where the in-time efficiency is still consistent with $\epsilon_{in} = 1.00$, although the optimal cut depends strongly on the relative sizes of the in-time and out-of-time contributions.

and $\epsilon(S_{max})$ for an array of S_{max} values extracted from the fit using uncorrected and corrected timing information. The corrected timing shows a radically improved out-of-time discrimination power for low-energy photons. In this particular energy range, out-of-time background rejection of 66% [$\epsilon_{out} < 0.44$] is possible while maintaining a signal efficiency of $\epsilon_{in} > 0.95$.

The proxy sample is very pure before cuts, so optimization on the objective function in the proxy studies will not necessarily yield the best cut for a given analysis. The increase in signal significance will be much more dramatic with a lower signal-to-background ratio before cuts. Additionally, the background efficiency numbers have

a fairly large uncertainty due to the small yield in the fits, which is not a critical problem for this analysis.

Nominally the in-time efficiency is equal to the error function $\epsilon_{in}(S_{max}) = erf(S_{max})$, which would allow simple optimal timing cut and signal efficiency calculations on data. However, in practice the timing errors are not well-modeled by a Gaussian. Parameterizing the error function by scaling the width fits the in-time efficiency fairly well but no precision gain over fitting the proxy peak is observed.

C.5.3 4-cluster timing

In the signal cascade there are three in-time clusters: γ_{hard} and the two muon candidates. Combining these three times results in no significant improvements in background rejection, plausibly due to the very small clusters and low energy associated with muon tracks and the relatively large clusters and high energy associated with γ_{hard} showers.

The power of the EMC timing information is not limited to out-of-time discrimination. For example, the proxy mode is itself the main source of fake γ_{hard} candidates and has a unique timing signature: four in-time neutral clusters. Rejection of background modes which contribute extra in-time clusters to the event is therefore possible with creative uses of the timing significance information.

C.6 Conclusions

The digi-based timing corrections substantially improve machine background rejection using the EMC's digi timing information. A wide variety of timing-related cuts can be invoked to remove out-of-time background in analyses with low-energy EMC clusters while compromising very little signal. The timing cut is used between γ_{soft} and γ_{hard} candidates in every analysis presented in this dissertation, including the appendices.

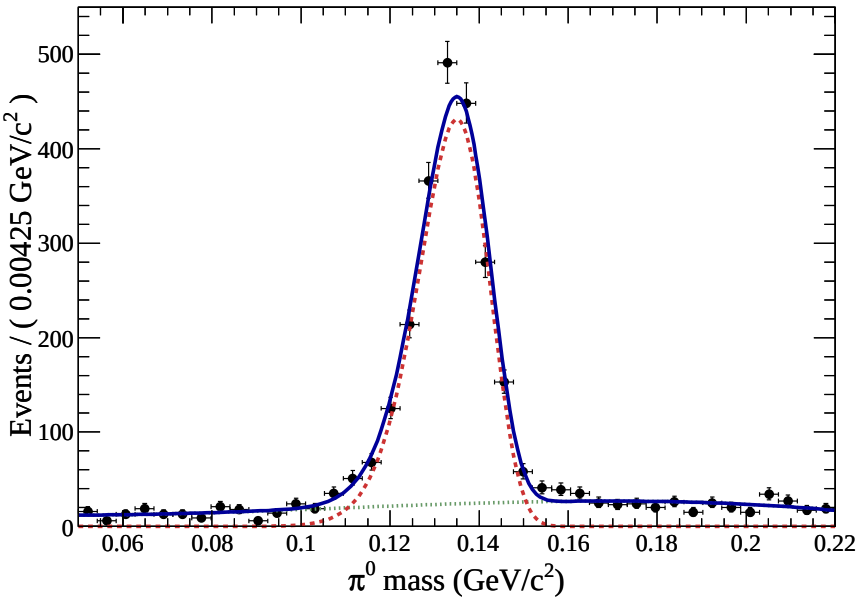


Figure C.10: Candidate proxy π^0 mass distribution with no timing cuts. The dotted green line is the best-fit 2^{nd}-order polynomial for the out-of-time background and the dotted red line shows the fitted Novosibirsk function modeling the proxy π^0 mass peak.

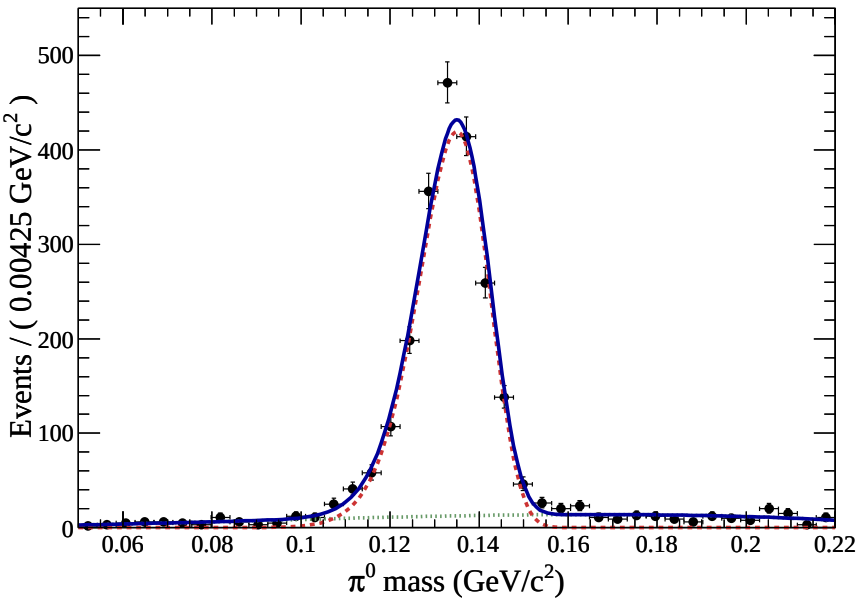


Figure C.11: Candidate proxy π^0 mass distribution as in Fig. **??** with a timing significance cut of $S_{max} = 3.0$.

Bibliography

[1] For comprehensive reviews see E. Eichten *et al.*, Rev. Mod. Phys. **80**, 1161 (2008) and N. Brambilla *et al.*, Eur. Phys. J. C **71**, 1 (2010); and the references therein. See also [2].

[2] D. Besson and T. Skwarnicki, Annu. Rev. Nucl. Sci. **43** 333-78 (1993).

[3] S. W. Herb *et al.*, Phys. Rev. Lett. **39**, 252 (1977); W. R. Innes *et al.* Phys. Rev. Lett. **39**, 1240 (1977); **39**, 1640(E) (1977).

[4] N. Brambilla *et al.*, Quarkonium Working Group, CERN-2005-005, (2005).

[5] C. Patrignani *et al.*, Annu. Rev. Nucl. Par. Sci., 47 (2012).

[6] D. C. Peaslee (ed.), Topics in Hardon Spectroscopy, Vol. 2., Nova Sci. Publ. (1995).

[7] Eichten, K. Gottfried, T. Kinoshita, K.D. Lane, Tung-Mow Yan, Phys. Rev. D **17**, 3090 (1978); Erratum-ibid. **D21**, 313 (1980).

[8] D. Besson, T. Skwarnicki, Annu. Rev. Nucl. Par. Sci., 43:333-78 (1993).

[9] C. N. Yang, Phys. Rev. 77 (1950) 242.

[10] C. Klopfenstein *et al.* (CUSB Collaboration), Phys. Rev. Lett. **51**, 160 (1983).

[11] R. Nernst *et al.* (Crystal Ball Collaboration), Phys. Rev. Lett. **54**, 2195 (1985).

[12] M. Narain *et al.* (CUSB-II Collaboration), Phys. Rev. Lett. **66**, 3113 (1991).

[13] R. Morrison *et al.* (CLEO Collaboration), Phys. Rev. Lett. **67**, 1696 (1991).

[14] K.W. Edwards *et al.* (CLEO Collaboration), Phys. Rev. D **59**, 032003 (1999).

[15] M. Artuso *et al.* (CLEO Collaboration), Phys. Rev. Lett. **94**, 032001 (2005).

[16] F. Pauss *et al.* (CUSB Collaboration), Phys. Lett. B **130**, 439 (1983).

[17] W.S. Walk *et al.* (Crystal Ball Collaboration), Phys. Rev. D **34**, 2611 (1986).

[18] U. Heintz *et al.* (CUSB-II Collaboration), Phys. Rev. Lett. **66**, 1563 (1991).

[19] U. Heintz *et al.* (CUSB-II Collaboration), Phys. Rev. D **46**, 1928 (1992).

[20] G. Crawford *et al.* (CLEO Collaboration), Phys. Lett. B **294**, 139 (1992).

[21] M. Kornicer *et al.* (CLEO Collaboration), Phys. Rev. D **83**, 054003 (2011).

[22] P. Haas *et al.* (CLEO Collaboration), Phys. Rev. Lett. **52**, 799 (1984).

[23] D. Cinabro *et al.* (CLEO Collaboration), ICHEP-02-ABS-929 (2002); arxiv:hep-ex/0207062.

[24] H. Albrecht *et al.* (ARGUS Collaboration), Phys. Lett. B **160**, 331 (1985).

[25] G. Bonvicini *et al.* (CLEO Collaboration), Phys. Rev. D **82** 111102(R) (2010).

[26] I. Adachi *et al.* (Belle Collaboration), Phys. Rev. Lett. **108** 032001 (2012).

[27] R. Mizuk *et al.* (Belle Collaboration), Phys. Rev. Lett. **109** 232002 (2012).

[28] B. Aubert *et al.* (*BABAR* Collaboration), Phys. Rev. Lett. **(**101), 071801 (2008). B. Aubert *et al.* (*BABAR* Collaboration), Phys. Rev. Lett. **(**103), 161801 (2009).

[29] J. P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. Phys. Lett. B **(**D84), 091101 (2011).

[30] S. Banerjee *et al.*, Phys. Rev. D **77** 054012 (2008).

[31] Aristotle, The Organon, O. F. Owen (translation), Vol. 1 (1853) 217-218.

[32] Aristotle, The Organon, O. F. Owen (translation), Vol. 2 (1853) 548-553.

[33] M. Kocian, SLAC-PUB-10170

[34] L. S. Brown, R. N. Cahn, Phys. Rev. D **13**, 1195 (1976)

[35] P. Moxhay and J. L. Rosner, Phys. Rev. Phys. Lett. B (D28), 1132 (1983).

[36] S. N. Gupta, S. F. Radford, and W. W. Repko, Phys. Rev. Phys. Lett. B (D30), 2424 (1984).

[37] H. Grotch, D. A. Owen, and K. J. Sebastian, Phys. Rev. Phys. Lett. B (D30), 1924 (1984).

[38] F. Daghighian and D. Silverman, Phys. Rev. Phys. Lett. B (D36), 3401 (1987).

[39] J. P. Fulcher, Phys. Rev. Phys. Lett. B (D42), 2337 (1990).

[40] T. A. Lahde, Nucl. Phys. Phys. Lett. B (A714), 183 (2003).

[41] D. Ebert, R. N. Faustov, and V. O. Galkin, Phys. Rev. Phys. Lett. B (D67), 014027 (2003).

[42] M. Eckmann *BABAR*Note #**572** 2003 "A study of the feature extraction algorithm of the *BABAR*-EMC"

[43] A. Gray *et al.* (HPQCD and UKQCD Collaborations), Phys. Rev. D **72**, 094507 (2005).

[44] R. J. Dowdall *et al.* (HPQCD Collaboration), Phys. Rev. D **85**, 054509 (2012).

[45] G. P. Lepage *et al.*, Phys. Rev. D **46**, 4052 (1992).

[46] R. Lewis, R. M. Woloshyn, Phys. Rev. D **84**, 094501 (2011).

[47] C. T. H. Davies, Lect. at Schladaming Wint. Sch. (1997). arXiv:hep-ph/9710394

[48] B. Aubert *et al.* (*BABAR* Collaboration), Nucl. Inst. Meth. **A479**, 1 (2002); (*BABAR* Collaboration) arXiv:1305.3560.

[49] W. Menges (*BABAR* Collaboration), IEEE Nucl. Sci. Symp. Conf. Rec. 5:1470-1474 (2006).

[50] D.J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. **A 462**, 152 (2001).

[51] T. Sjöstrand, Comput. Phys. Commun. **82**, 74 (1994).

[52] See for example, L.S. Brown and R.N. Cahn, Phys. Rev. Phys. Rev. D **D** 13, 1195 (1976); G. Karl, S. Meshkov, and J.L. Rosner, Phys. Rev. Phys. Rev. D **D** 13, 1203 (1976).

[53] J. Beringer *et al.* (Particle Data Group), Phys. Rev. Phys. Rev. D **86**, 010001 (2012).

[54] S. Agostinelli *et al.* (Geant4 Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. **A** 506, 250 (2003).

[55] *BABAR* Analysis Document 1853.

[56] W. D. Hulsbergen, Nucl. Instrum. Methods Phys. Res., Sect. **A 552**, 566 (2005).

[57] All photon energies discussed are in the center of mass frame of the decaying particle.

[58] An extension of the usual Crystal Ball function from T. Skwarnicki, PhD Thesis, DESY F31-86-02, Appendix E (1986).

[59] J.P. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **84**, 072002 (2011).

[60] W. Kwong *et al.*, Annu. Rev. Nucl. Part. Sci., **37**, 325 (1987).

[61] C. O. Dib *et al.*, Phys. Rev. D **37** 735 (1988).

[62] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. Lett. **108**, 152001 (2012).

[63] W. Kwong and J.L. Rosner, Phys. Rev. D **38**, 279 (1988).

[64] S. Godfrey and J.L. Rosner, Phys. Rev. D **64**, 097501 (2001).