HPS
HEAVY PHOTON SEARCH

A Proposal to Search for Massive Photons at Jefferson Laboratory

December 1, 2010

Authors

A. Grillo, V. Fadeyev
University of California, Santa Cruz, CA 95064

M. Ungaro
University of Connecticut, Department of Physics, Storrs, CT 06269

W. Cooper
Fermi National Accelerator Laboratory, Batavia, IL 60510-5011

A. Micherdzinska
The George Washington University, Department of Physics,Washington, DC 20052

G. Ron
Hebrew University of Jerusalem, Jerusalem, Israel

M. Battaglieri, R. De Vita
INFN, Sezione di Genova, 16146 Genova, Italy

M. Holtrop (Co-Spokesperson), K. Slifer, S. K. Phillips, E. Ebrahim
University of New Hampshire, Department of Physics, Durham, NH 03824

M. Khandaker, C. Salgado
Norfolk State University, Department of Physics, Norfolk, VA 23504

S. Bueltmann, L. Weinstein
Old Dominion University, Department of Physics, Norfolk, VA 23529

A. Fradi, B. Guegan, M. Guidal, S. Niccolai, S. Pisano, E. Rauly, P. Rosier and D. Sokhan
Institut de Physique Nucleaire d'Orsay, 91405 Orsay, France

P. Schuster, N. Toro
Perimeter Institute, Ontario, Canada N2L 2Y5

P. Stoler, A. Kubarovsky
Rensselaer Polytechnic Institute, Department of Physics, Troy, NY 12181

R. Essig, C. Field, M. Graham, G. Haller, R. Herbst, J. Jaros (Co-Spokesperson), C. Kenney, T. Maruyama, K. Moffeit, T. Nelson, H. Neal, A. Odian, M. Oriunno, R. Partridge, D. Walz
SLAC National Accelerator Laboratory, Menlo Park, CA 94025

S. Boyarinov, V. Burkert, A. Deur, H. Egiyan, L. Elouadrhiri, A. Freyberger, F.-X. Girod, V. Kubarovsky, Y. Sharabian, S. Stepanyan (Co-Spokesperson), B. Wojtsekhowski
Thomas Jefferson National Accelerator Facility, Newport News, VA 23606

K. Griffioen
The College of William and Mary, Department of Physics, Williamsburg, VA 23185

N. Dashyan, N. Gevorgyan, R. Paremuzyan, H. Voskanyan
Yerevan Physics Institute, 375036 Yerevan, Armenia

Table of Contents

Abstract

The Heavy Photon Search (HPS) is a new experiment proposed for Jefferson Laboratory to search for new heavy vector boson(s), aka "heavy photons" or "dark photons" or "hidden sector photons", in the mass range of 20 MeV/c^2 to 1000 MeV/c^2. Heavy photons mix with the ordinary photon through kinetic mixing, which induces their weak coupling to electrons, εe, where $\varepsilon \sim 10^{-3}$. Heavy photons in this mass/coupling range are expected on very general theoretical grounds, and motivated by recent astrophysical evidence suggesting they might mediate dark matter annihilations and/or interactions with ordinary matter. Since they couple to electrons, heavy photons are radiated in electron scattering and can subsequently decay into narrow e^+e^- resonances which can be observed above the copious QED trident background. For suitably small couplings, heavy photons travel detectable distances before decaying, providing a second signature. The HPS experiment exploits both these signatures to search for heavy photons over a wide range of couplings, $\varepsilon^2 > 10^{-10}$, and masses, using a new compact, large acceptance forward spectrometer, silicon microstrip vertex tracker, $PbWO_4$ electromagnetic calorimeter, and muon system. The experiment will be positioned behind the CLAS detector in Hall B at Jefferson Lab, and run with 200-500 nA beams at 2.2 and 6.6 GeV on 0.15% - 0.25% X_0 tungsten targets. The experiment achieves great sensitivity by exploiting CEBAF's 100% duty cycle, high luminosities, and 40 MHz continuous readout. With 3-month runs at each incident energy, HPS will explore a large and unexplored domain in the mass/coupling plane with extraordinary sensitivity, enough to exclude or see a heavy photon signal conclusively. As a bonus, HPS should make the first observation of "true muonium", the $\mu^+\mu^-$ "atom" which decays like a heavy photon and has ~ cm decay length.

1 Introduction

Interest in the possibility that a massive photon-like particle could exist has been piqued recently by suggestions from Arkani-Hamed et al. [1] and Pospelov and Ritz [2] and others that such a particle could mediate dark matter annihilations and account for recently observed electron/positron features in the cosmic rays, and perhaps even the putative DAMA dark matter annual modulation. The possible existence of a heavy photon was suggested earlier, however, especially by Holdom [3]. He considered the possibility that Nature has additional U(1)'s, such as those suggested in string theories, which would naturally kinetically mix with *the* photon, generating a coupling between the new heavy photon(s) and electric charge of εe, where $\varepsilon \sim 10^{-3}$. Since it does couple to electric charge, the heavy photon could be produced when suitably energetic electron beams bremsstrahlung off heavy targets, albeit at rates strongly suppressed compared to expected QED processes. For heavy photons of mass ~100 MeV, preferred by those trying to account for the cosmic ray data and in line with more general considerations as well, the decay would be to e^+e^-; at higher masses other channels would open up. The signature for a heavy photon would be twofold: 1) a sharp bump in the e^+e^- invariant mass above the copious QED trident background; and 2) a secondary decay vertex, since the coupling can be weak enough to generate detectable lifetime effects. For heavy photons of mass between 20 MeV and 1 GeV, electron beams of ~few GeV can initiate heavy photon bremsstrahlung.

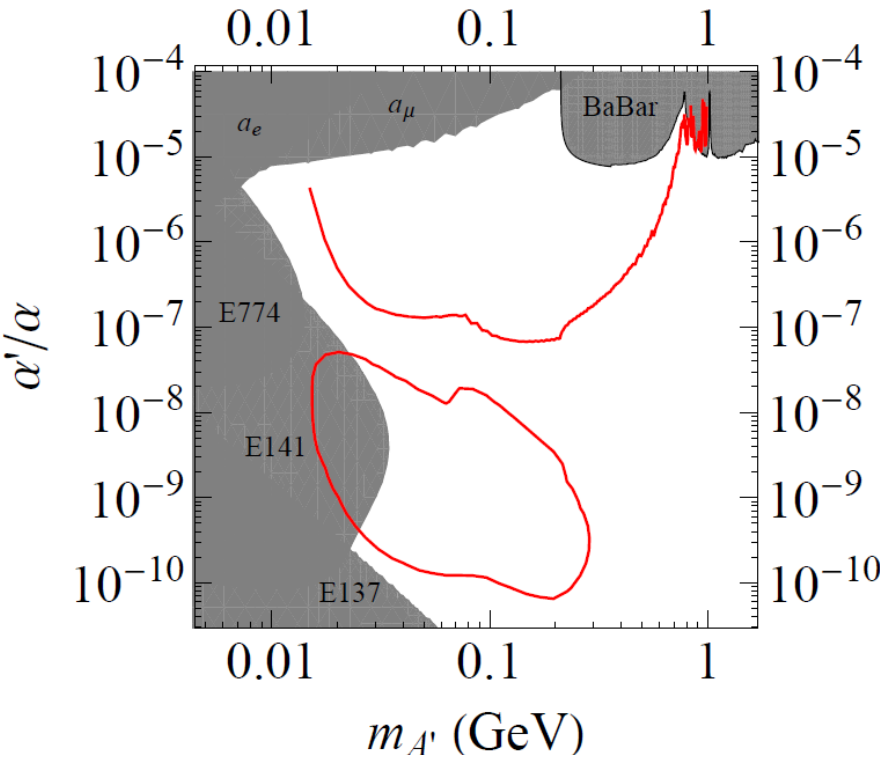

Figure 1.1. Existing heavy photon exclusion regions (gray), and proposed search regions for the HPS experiment in the heavy photon mass-coupling plane, where $\alpha'/\alpha \equiv \varepsilon^2$.

Existing constraints on heavy photon masses and couplings come from axion searches, the anomalous magnetic moments of the muon and electron, and direct searches for heavy photons in the B factory data [4]. These constraints and the reach of the Heavy Photon Search (HPS) experiment are shown in Figure 1.1. Roughly speaking, heavy photons are allowed below a coupling strength of few x 10^{-3} and throughout the mass range of 20-1000 MeV/c^2. As indicated in the figure, the HPS experiment will simultaneously explore two large regions of this parameter space. One HPS search region focuses on a wide range of heavy photon masses and moderate couplings with a traditional bump-hunt search, which other experiments plan to probe as well. The other region is unique, and utilizes both invariant mass and separated decay vertex information to provide unparalleled sensitivity to small couplings over the mass range 20-250 MeV/c^2.

Physicists and engineers from SLAC, the Hall B Group at JLab and their collaborators, Fermilab, and UC Santa Cruz have developed a concept put forward in Reference [4] into this proposal. The experiment utilizes a high acceptance forward spectrometer with precise momentum, vertexing, and calorimetric measurement capability to provide ~1% momentum resolution, ~1 mm vertex resolution, and fast triggering and electron identification. The experiment probes a unique region of the mass-coupling plane where the heavy photon signal would be lost in the trident background without the additional signature of a secondary vertex displaced from the target, and it simultaneously accesses a higher coupling strength region by relying on bump hunting alone. The experiment depends upon the 40 MHz readout capability of the silicon microstrip vertex/tracker which utilizes LHC readout electronics, the matching high rate capability of a highly segmented $PbWO_4$ calorimeter, and high rate triggering and data acquisition systems. Combined with CEBAF's superb duty-cycle, high intensities, and excellent beam properties, this high rate capability lets HPS achieve the high integrated luminosities required to search for heavy photons. As a bonus, HPS should also observe a particle expected in conventional QED but as yet unseen, so-called true muonium. This $\mu^+ \mu^-$ atom is expected to decay into e^+e^- with the heavy photon signature, a sharp mass bump (at ~210 MeV/c^2) and a separated decay vertex, which in this case is often a centimeter or more beyond the target!

In what follows, we discuss the physics motivation for heavy photon searches, the properties of the heavy photon signal and the prominent QED backgrounds, the proposed beamline and apparatus, the performance of the apparatus as determined from detailed full Monte Carlo simulations, and finally the calculated reach of the experiment. Then we motivate a Test Run to check some of the critical assumptions of our design, and sketch the proposed test apparatus. We conclude with the proposed run plan and a summary of the resources needed for the experiment.

1.1 References

1. N. Arkani-Hamed , D.P. Finkbeiner, T. R. Slatyer, and N. Weiner, Phys. Rev. **D79**, 015014 (2009), ArXiv: 0810.0713.

2. M. Pospelov and A. Ritz, Phys. Lett. **B671**, 391 (2009), 0810.1502.

3. B. Holdom, Two U(1)'s and Epsilon Charge Shifts, Phys. Lett. **B166** (1986) 196.

4. J.D. Bjorken, R. Essig, P. Schuster, and N. Toro, "New Fixed-Target Experiments to Search for Dark Gauge Forces", Phys. Rev. **D80**, 2009, 075018, arXiv: 0906.0580.

2 Physics Motivation

We consider new sub-GeV mass vector bosons - "dark photons" A' - that couple very weakly to electrons (similar considerations apply to pseudo-vectors, scalars, and pseudo-scalars with sub-GeV mass that couple to electrons). It is useful to parameterize the coupling g' of the A' to electrons by a dimensionless $\varepsilon \equiv g'/e$, where e is the electron charge. Cross-sections for A' production then scale as $\alpha'/\alpha = \varepsilon^2$, where $\alpha'=g'^2/4\pi$ and $\alpha=e^2/4\pi$ are the fine-structure constants for the dark photon and ordinary electromagnetic interactions, respectively. This experiment will search for A' bosons with mass $m_{A'} \sim 100$ MeV and $\alpha'/\alpha \leq 10^{-5}$, which can be produced by a reaction analogous to photon bremsstrahlung and will decay promptly to e^+e^- or other charged particle pairs. We refer the reader to Figure 2.3.1 for a summary of the reach of this experiment.

2.1 Motivation for New Physics Near the GeV Scale

New light vector particles, matter states, and their associated interactions are ubiquitous in extensions of the Standard Model [1-10]. However, the symmetries of the Standard Model restrict the interaction of ordinary matter with such new states. Indeed, most interactions consistent with Standard Model gauge symmetries and Lorentz invariance have couplings suppressed by a high mass scale. One of the few unsuppressed interactions is the coupling of charged Standard Model particles ψ,

$$\Delta L = g' A'^{\mu} \bar{\psi} \gamma_{\mu} \psi \qquad (2.1)$$

to a new gauge boson A', which is quite poorly constrained for small g' (see Figure 2.3.1). Similar couplings between the A' and other Standard Model fermions are also allowed, with relations between their couplings (anomaly cancellation) required for the A' gauge symmetry to be quantum-mechanically consistent. For example, the A' could couple only to electrons and muons, with opposite charges $g'_e = -g'_\mu$ (a U(1) e,μ boson), or could have couplings proportional to the electromagnetic charges q_i of each fermion, $g_{qi} = \varepsilon e q_i$.

A' couplings to Standard Model matter with the latter structure can be induced by ordinary electromagnetic interactions through the kinetic mixing interaction proposed by Holdom [1],

$$\Delta L = \frac{\epsilon_Y}{2} F'_{\mu\nu} F_Y^{\mu\nu} \qquad (2.2)$$

where $F'_{\mu\nu} = \partial_\mu A'_\nu - \partial_\nu A'_\mu$ is the field strength of the A' gauge boson, and similarly $F_Y^{\mu\nu}$ is the hypercharge field strength. This effect is generic, ensures that the A' interactions respect parity, and (as we discuss below) naturally produces small g' and A' masses near the GeV scale. This mixing is equivalent in low-energy interactions to assigning a charge $\varepsilon e q_i$ to Standard Model particles of electromagnetic charge q_i, where $\varepsilon = \varepsilon_Y/\cos(\theta_W)$ and θ_W is the Weinberg mixing angle.

The A' couplings to neutrinos and parity-violating couplings are negligible compared to Z-mediated effects (see e.g. [12]).

As noted in [1], a new gauge boson A' that does not couple to Standard Model matter at a classical level can still couple through quantum-mechanical corrections. For example, loops of any particle X that couples to both the A′ and Standard Model hypercharge generates mixing of the form (2.1), with

$$\epsilon \sim 10^{-3} - 10^{-2}, \quad \frac{\alpha'}{\alpha} \sim 10^{-6} - 10^{-4} \qquad (2.3)$$

These quantum effects are significant regardless of the mass m_X of the particle in question, which could be well above the TeV scale (or even at the Planck scale) and thus evade detection.

Smaller ε are expected if nature has enhanced symmetry at high energies. For example, it has been conjectured that the strong and electroweak gauge groups of the Standard Model are embedded in a grand unified theory (GUT) with gauge group SU(5) or larger that is broken spontaneously at a high scale $M_{GUT} \approx 10^{16}$ GeV. In this case the mixing (2.1) is suppressed,

$$\epsilon_{GUT} \sim \frac{\alpha_i^2}{16\pi^2} \ln\left(\frac{M_{GUT}}{M_X}\right) \sim 10^{-5} - 10^{-3} \qquad (2.4)$$

where α_i are gauge couplings. An ε of this size leads to effective couplings

$$\frac{\alpha'}{\alpha} \sim 10^{-10} - 10^{-6} \qquad (2.5)$$

As shown in Figure 2.3.1, no experiment to date has probed the range of ε expected in grand unified theories for $m_{A'} \geq 50$ MeV.

An A' mass near but beneath the weak scale is particularly well-motivated, as U(1)' symmetry-breaking and the resulting A' mass may be determined by the same physics that generates the W and Z masses [13]. The best candidate for the origin of the weak scale is low-energy supersymmetry. In this case, the A' can naturally acquire mass suppressed by a loop factor or by $\sqrt{\varepsilon}$ compared to the weak scale, leading to MeV to GeV-scale A' masses [5,12,14-17]. In supersymmetric models, the gauge kinetic mixing (2.1) is accompanied by quartic interactions

$$\frac{\epsilon_Y}{4} g_Y g_D |\phi_D|^2 |h|^2 \qquad (2.6)$$

between the Standard Model Higgs doublet h and any scalar Φ_D charged under U(1)', where g_Y and g_D are the gauge couplings of Standard Model hypercharge and the A' coupling to Φ_D, respectively. Electroweak symmetry breaking gives h a weak-scale vacuum expectation value, so that (2.5) generates a mass term for Φ_D. For positively charged Φ_D, and sufficiently small bare

mass, this mass term is negative and triggers U(1)' breaking by the Higgs mechanism. The resulting induced mass for the A' is

$$m_{A'} \sim \sqrt{\epsilon} \sqrt{\left(\frac{g_D g_Y}{g_2^2}\right)} m_W \sim MeV - GeV \qquad (2.7)$$

where g_2 is the Standard Model $SU(2)_L$ gauge coupling and m_W is the W-boson mass. The resulting mass is precisely in the 50-1000 MeV range targeted by this experiment. The combined resonance and vertexing search will probe a sizable fraction of this parameter space. In particular, the resonance search will probe the small g_D region, e.g. for $g_D = 0.005$ and $\alpha'/\alpha \sim 7 \times 10^{-7}$, we have $m_{A'} \sim 150$ MeV/c^2, which is well within the reach of the experiment. The vertexing search will probe larger values for g_D, e.g. for $g_D = 0.1$ and $\alpha'/\alpha \sim 3 \times 10^{-7}$, we have $m_{A'} \sim 100$ MeV/c^2. Note that the mechanism of U(1)' breaking above does not rely on supersymmetry, as any quartic interaction of the form (2.6), with arbitrary coupling, can transmit electro-weak masses to the A'. Thus, the mass relation (2.7) should not be interpreted too literally.

We stress that the mass of the A' breaks any apparent symmetry between it and the photon: though Standard Model particles have induced ε-suppressed charges under the A', any new matter charged under the A' would not have any direct coupling to the photon, and would thus not have an electric "milli-charge". This is the main reason why such particles could have gone undetected.

An electron beam scattering on a high-Z target such as Tungsten will produce A's through bremsstrahlung reactions with a cross-section

$$\sigma_{A'} \sim 100\ pb \left(\frac{\epsilon}{10^{-4}}\right)^2 \left(\frac{100\ MeV}{m_{A'}}\right)^2 \qquad (2.8)$$

several orders of magnitude larger than in colliding electron and hadron beams [18]. The A' can decay to electrons, and is therefore visible as a narrow resonance in the trident e^+e^- mass spectrum.

Seeing such a new gauge boson would constitute the first discovery of a new gauge force since the observation of Z-mediated neutral currents. Besides the obvious physical interest of a fifth force, the A' like the Z could open up a new ``sector'' of light, weakly coupled particles whose spectrum and properties could be measured in fixed-target experiments and flavor factories. The A' sector would provide a new laboratory for many physical questions, and would be revealing precisely because its interactions with Standard Model particles are so weak. In particular, if nature is approximately supersymmetric near the TeV scale, the mass scale of supersymmetry breaking for the A' sector is naturally suppressed by ε times gauge couplings. In this case, supersymmetry could be studied easily in the A' sector, and possibly even discovered there by relatively low-energy experiments before Standard Model superpartners are seen at colliders.

2.2 Motivation for an A' from Dark Matter

Dark matter interpretations of recent astrophysical and terrestrial anomalies provide an urgent impetus to search for A's in the mass range 50 MeV - 1 GeV, with a coupling $\varepsilon \sim 10^{-4} - 10^{-2}$.

The concordance model of big bang cosmology - the "Lambda Cold Dark Matter" (ΛCDM) model - explains all observations of the cosmic microwave background, large scale structure formation, and supernovae; see e.g. [19-23]. This model suggests that Standard Model particles make up only about 4% of the energy density in the Universe, while ``dark energy'' and ``dark matter'' make up 74% and 22%, respectively, of the Universe's energy density. The concordance model does not require dark matter to have any new interactions beyond gravity with Standard Model particles. However, an intriguing theoretical observation, dubbed the WIMP miracle, suggests that dark matter does have new interactions. In particular, if dark matter consists of ~10 GeV to 10 TeV particles interacting via the electroweak force ("weakly interacting massive particles" or "WIMPs"), they would automatically have the right relic abundance consistent with the ΛCDM model.

In addition to the WIMP miracle, evidence from cosmic-ray data, from balloon-borne instruments and satellites, and the terrestrial direct dark matter detection experiments DAMA/LIBRA and CoGeNT may indicate that dark matter interacts with ordinary matter not just gravitationally. While the WIMP miracle suggests that dark matter is charged under the Standard Model electroweak force, we will see that these observations are also consistent with the hypothesis that dark matter interacts with ordinary matter through a new *force, mediated by a new 20 MeV - 1 GeV mass gauge boson*.

In addition to explaining any or all of these observations, dark matter charged under this new force can easily have the correct thermal relic abundance observed today by virtue of its interactions via the new force carrier, reproducing the success of the WIMP dark matter hypothesis. The reason is that the thermal relic abundance is set by the annihilation cross-section at freeze-out, which is given by $\sigma_v \sim \pi \alpha_X^2 / m_{DM}^2$, where α_X is the coupling of the dark matter to the A′ and is independent of ε; for α_X of order the weak-interaction strength and dark matter masses of order the weak scale, the WIMP miracle is reproduced.

The satellites PAMELA [24] and Fermi [25], the balloon-borne detector ATIC [26], the ground-based Cherenkov telescope HESS [27,28], as well as other experiments, observe a likely excess in the cosmic-ray flux of electrons and/or positrons above backgrounds expected from normal astrophysical processes. While this excess may be due to nearby pulsars or supernova remnants, a plausible origin is dark matter annihilation or decay.

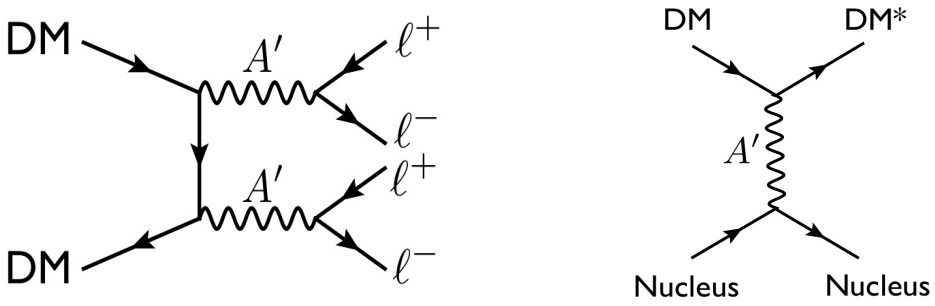

Figure 2.2.1 Left: Dark matter annihilation into the dark photon A', which decays into charged leptons such as electrons and/or muons, can explain the cosmic-ray electron and/or positron excesses seen by PAMELA, Fermi, ATIC, HESS, and other experiments. Right: Dark matter scattering into an excited state off nuclei through A' exchange in direct dark matter detection experiments could explain the annual modulation signal observed by DAMA/LIBRA, and the signal observed by CoGeNT.

However, two features of these observations are incompatible with annihilation of ordinary thermal WIMP dark matter, but compatible with an alternative explanation: that dark matter is charged under a new U(1)' and annihilates into A′-pairs, which decay directly into electrons and positrons, or into muons that decay into electrons and positrons, see Figure 2.2.1 (left) (see e.g. [13,31-37]). These two features are:

- The annihilation cross-section required to explain the electron signal is 50-1000 times larger than the thermal freeze-out cross-section for an ordinary WIMP that is needed to reproduce the observed dark matter relic density. This can be explained if dark matter interacts with a new long range force mediated by an O(GeV) mass gauge boson, which allows the dark matter annihilation cross-section $\langle\sigma v\rangle$ to be enhanced at low dark matter velocities, i.e. $\sim \langle\sigma v\rangle \propto 1/v$. In this case, in the early Universe when the dark matter velocity was high (~ 0.3c), the annihilation cross-section that determines the relic abundance can naturally be the same as that of an ordinary WIMP and reproduce the WIMP miracle. However, in the Milky Way halo now, the dark matter has a much lower velocity ($v \sim 10^{-3}c$), leading to a large increase in the annihilation cross-section that is required to explain the cosmic-ray data. The enhancement at low velocities through a new long-range force is very well known and called the Sommerfeld effect [38]. We refer the reader to [29] for a recent detailed discussion.
- The PAMELA satellite did not see an anti-proton excess [39], which strongly suggests that dark matter annihilation is dominantly producing leptons, and not baryons. If dark matter is interacting via a O(GeV) mass force particle in order to have a large annihilation rate via the Sommerfeld mechanism, then annihilations into the force carrier automatically fail to produce any baryons. Kinematically, the force carriers cannot decay into baryons, and are instead forced to decay into the lighter charged leptons. Thus, annihilation products of dark matter are leptonic in this case.

We note that within supersymmetry, neutralino ("wino") dark matter could be responsible for the electron and positron excesses, but ordinarily produces too many anti-protons. Moreover, if

the neutralino is a thermal relic, its annihilation cross-section is too small to account for the observed signals.

To explain the additional sources of evidence for a new GeV scale force, we briefly summarize the consequence for dark matter mass spectra that follow from dark matter carrying a charge under a new force. If in addition to its U(1)' charge, dark matter is charged under a non-Abelian force (a very natural possibility) and the force carriers acquire a mass, then radiative effects can split all components of the dark matter with size, $\delta \sim \alpha' \Delta m_{W'}$, where α' is the non-Abelian fine structure constant and $\Delta m_{W'}$ is the splitting of gauge boson masses [13]. Typically, these splittings are $\Delta m_{W'} \sim \alpha' m_{W'} \sim$ 1-10 MeV for $m_{W'} \sim 1$ GeV [14]. Thus, $\delta \sim 100$ keV for $\alpha' \sim 10^{-2}$. These splittings are completely analogous to the splittings that arise between the $\pi^{\pm}$ and π^0 from Standard Model SU(2) breaking. If instead a non-Abelian force confines at a scale $\Lambda_D \sim$ GeV, then a heavy-flavor meson can be cosmologically long-lived and thus a dark matter candidate [40]. Hyperfine interactions can naturally induce ~ 100 keV splittings of the dark matter particles in this case.

We emphasize that the GeV-scale non-Abelian force carrier particles mediate quantum corrections that generate the 100 keV and 1-10 MeV splittings of dark matter states [14, 40, 41]. When mass splittings arise, Dirac states split into two Majorana states, and the A' mediated interactions of dark matter with ordinary matter as well as dark matter self-interactions are dominated by inelastic interactions between the two Majorana states [13]. The direct dark matter detection experiment DAMA/LIBRA as well as the INTEGRAL telescope provide intriguing evidence for such interactions. The DAMA/NaI [43] and DAMA/LIBRA [44] experiments have reported an annual modulation signal over nearly eleven years of operation with more than 8σ significance. Modulation is expected because the Earth's velocity with respect to the dark matter halo varies as the Earth moves around the sun, and the phase of the observed modulation is consistent with this origin. A simple hypothesis that explains the spectrum and magnitude of the signal, and reconciles it with the null results of other experiments, is that dark matter-nucleus scattering is dominated by an inelastic process,

$$\chi N \rightarrow \chi^* N \qquad (2.9)$$

in which the dark matter χ scatters off a nucleus N into an excited state χ^* with mass splitting $\delta \approx$ 100 keV [41,42]. The kinematics of these reactions is also remarkably consistent with all the distinctive properties of the nuclear recoil spectrum reported by DAMA/LIBRA. In addition, the INTEGRAL telescope [45] has reported a 511 keV photon signal near the galactic center, indicating a new source of ~ 1-10 MeV electrons and positrons. This excess could be explained by collisions of O(100 GeV~1 TeV) mass dark matter into excited states only slightly more massive (O(MeV)) in the galaxy [46]—the excited dark matter states subsequently decay back to the ground state by emitting an A', which in turn decays to an e^+e^- pair. The 511 keV excess then arises from the subsequent annihilation of the produced positrons.

The above dark matter related anomalies are thus consistent with 100 GeV - 1 TeV mass dark matter interacting with ordinary matter through an A'. By itself, the DAMA/LIBRA anomaly can also be explained with *light* dark matter, a 5-10 GeV dark matter particle scattering off nuclei

through A′ exchange. Such a low dark matter mass cannot explain the cosmic-ray anomalies, but is motivated instead by another direct dark matter detection experiment, CoGeNT, that has recently reported an anomaly. The CoGeNT collaboration reported about 100 events from an unknown source in a low-threshold, high-resolution Ge detector [30]. These events are consistent with a 5-10 GeV dark matter particle scattering off the Germanium nuclei in the CoGeNT detector [60]. The required dark matter mass and scattering cross-section are intriguingly close to the values needed to also explain the DAMA/LIBRA annual modulation signal.

The Standard Model particles that are usually invoked to explain dark matter scattering off nuclei are the Higgs boson and the Z-boson. However, the Higgs boson couples only weakly to nuclei and typically generates scattering cross-sections $<10^{-43}$ cm^2, which is much smaller than the required cross-section ($\sim10^{-40}$ cm^2) to explain DAMA/LIBRA and CoGeNT. The Z-boson can provide a cross-section of the right size, but measurements of the invisible width of the Z at LEP tightly constrain the coupling between dark matter and the Z-boson (since the Z is kinematically allowed to decay into 5-10 GeV dark matter particle if they couple to each other). An A′ mediating the scattering, however, can easily produce the required cross-section and satisfy all constraints.

Remarkably, it is even possible to build models in which both the DAMA/LIBRA and CoGENT signals *and* the cosmic-ray anomalies are naturally explained by dark matter coupling to an A′ [61]. For this we need two dark matter components: the dominant dark matter component consists of a TeV-scale particle annihilating or decaying to an A′, which in turn decays to charged leptons to explain the cosmic-ray anomalies; a subdominant (<1%) component consists of a 5 GeV particle that scatters off the nuclei in the DAMA/LIBRA and CoGeNT detectors via A′ exchange. The required cross-section can be naturally obtained in a supersymmetric model.

In addition to the various dark matter anomalies, the existence of an A' may also help explain various other particle physics anomalies [47] such as the anomalous magnetic moment of the muon ((g-2)μ) [48]. The A′ mass and coupling suggested by this anomaly lies close to the (g-2)μ bound shown in, for example, Fig. 2.3.1. The HPS experiment will probe a large part of this parameter space.

While these experimental hints provide an urgent motivation to look for an A', it is important to emphasize the value of these searches in general. There has never been a systematic search for new GeV-scale force carriers that are weakly coupled to Standard Model particles. Nothing forbids their existence, and their discovery would have profound implications for our understanding of nature. The HPS experiment will probe a range of A' masses and couplings that are expected on general theoretical grounds, and suggested by astrophysical evidence and the muon g-2 anomaly.

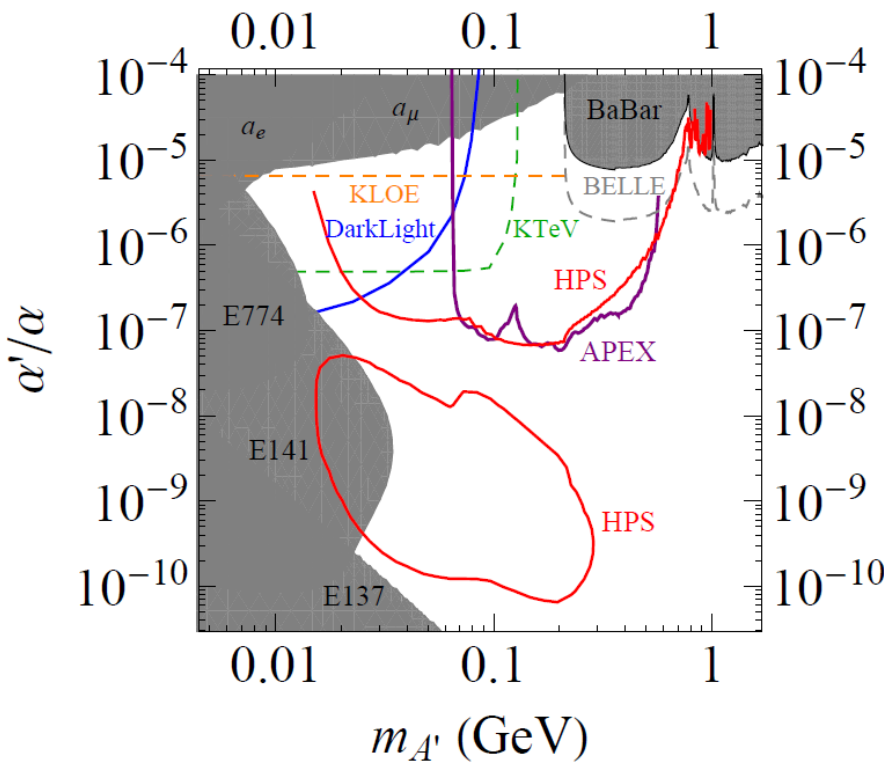

Figure 2.3.1 Anticipated reach in $\alpha'/\alpha = \varepsilon^2$ for the Heavy Photon Search (HPS) experiment at Hall B in JLab (red lines) with *existing* constraints on an A' from electron and muon anomalous magnetic moment measurements, a_e and a_μ (see [54]), the BaBar search for Υ(3S) → γμ$^+$μ$^-$ [49], and three beam dump experiments, E137, E141, and E774 [55, 56, 57] (see [11]). The a_μ and Υ(3S) limits assume equal-strength couplings to electrons and muons. Also shown are *estimates* of *potential* 2σ sensitivities for A' searches in existing data (thin dashed lines), assuming optimal sensitivity as described in the text: KTeV $\pi^0 \to \gamma$ A' $\to \gamma$ e$^+$ e$^-$, (green dashed curve), KLOE $\Phi \to \eta$ A'$\to \eta$ e$^+$ e$^-$ (orange dashed curve) and Belle e$^+$e$^-\to \gamma$ A' $\to \gamma$ μ$^+$ μ$^-$ (gray dashed curve). In addition, we show the projected 2σ sensitivities for the proposed "APEX" experiment in JLab Hall A (purple) [58] and the proposed "DarkLight" experiment using the JLab Free-Electron Laser (FEL) (blue) (we show the 5σ sensitivity taken from [59]). The HPS lines assume an electron beam with a current of 450 nA (200 nA) incident for 9×10^6 seconds on a 0.25% (0.125%) radiation length tungsten target and a beam energy of 6.6 GeV (2.2 GeV). The upper solid red lines correspond to the 2σ sensitivity of a full resonance search. The lower contours denote the sensitivity of a vertex-based resonance search. Here the lines correspond to 2.4 signal events where the vertex requirement was chosen to have 0.5 expected background events in each resolution-limited mass window. The full detector, trigger, and vertexing efficiencies have been included in these estimates. See Section 6 for a more detailed discussion.

.

2.3 Current Limits on Light *U(1)* Gauge Bosons

Constraints on new A's that decay to e$^+$e$^-$ and the potential search reach of HPS are summarized in Figure 2.3.1. Also shown are constraints from electron and muon anomalous magnetic moment measurements, a_e and a_μ [47], the BaBar search for Υ(3S) $\to \gamma$ A' $\to \gamma$ μ$^+$ μ$^-$ [49] and three beam dump experiments, E137, E141, and E774 [11]. The constraints from a_μ and the BaBar

search assume that the A' couples to muons -- this is the case, for example, if it mixes with the photon. If it only couples to electrons, then the constraints on α'/α and $m_{A'}$ in the region to which the proposed experiment is sensitive are weaker than $\alpha'/\alpha \leq 10^{-4}$.

We caution that systematic uncertainties in the A' limit beyond those quoted in [49] may slightly weaken the resulting limit, which should therefore be taken as a rough approximation unless further analysis is done. First, A' production in B-factories is more forward-peaked than the Υ(3S) decay mode considered in [49], so that the signal acceptance is more uncertain. In addition, background distributions in [49] are derived from smooth polynomial fits to data collected on the Υ(4S) resonance, which is assumed to contain no signal. This assumption is not correct for A' production, though the resulting systematic effects are expected to be small.

Several past and current experiments have data that could be used to significantly improve current limits on α'/α, as discussed in [47,51,52]. Here, we estimate the potential sensitivity of searches in three channels ($\pi^0 \rightarrow \gamma$ A' $\rightarrow \gamma e^+ e^-$, $\Phi \rightarrow \eta$ A' $\rightarrow \eta e^+ e^-$, and $e^+e^- \rightarrow \gamma$ A' $\rightarrow \gamma \mu^+ \mu^-$), considering only the statistical uncertainties and irreducible backgrounds. These are likely overestimates, as we are unable to include either systematic uncertainties or significant instrumental backgrounds such as photon conversion in the detector volume.

BaBar, BELLE, and KTeV (E799-II) have produced and detected large numbers of neutral pions, of order 10^{10}, of which roughly 1% decay in the Dalitz mode $\pi^0 \rightarrow \gamma e^+ e^-$. These experiments can search for the decay $\pi^0 \rightarrow \gamma$ A' induced by A'-photon kinetic mixing, which would appear as a narrow resonance over the continuum Dalitz decay background. KTeV has the largest π^0 sample, and its $e^+ e^-$ mass resolution can be approximated from the reported measurement of the $\pi^0 \rightarrow e^+ e^-$ branching fraction [50] to be roughly 2 MeV. This paper also reports the measured mass distribution of Dalitz decays above 70 MeV, from which we estimate potential sensitivity to α'/α as small as 5×10^{-7} for $70 < m_{e^+e^-} < 100$ MeV, as shown by the green shaded region in Figure 2.3.1.

Similarly, KLOE can search for the decay $\Phi \rightarrow \eta A'$, likewise induced by A' kinetic mixing with the photon, in a sample of 10^{10} Φ's. An analysis of this data is ongoing [53]. We have taken the blue dashed curve in Figure 2.3.1 from [51], which assumes that mass resolution σ_m is dominated by KLOE's 0.4% momentum resolution. We have adjusted the contours from [51] to determine a 2σ contour and enlarged the bin width used to determine signal significance from σ_m in [51] to $2.5\sigma_m$. Above the muon threshold, Φ decays are not competitive with B-factory continuum production.

In addition, BaBar and Belle can search for the continuum production mode $e^+e^- \rightarrow \gamma A' \rightarrow \gamma\mu^+\mu^-$ in their full datasets. For example, an analysis of the Belle Υ(4S) data set would increase statistics by a factor of ~ 24 relative to the BaBar Υ(3S) search that we have interpreted as a limit above. We have derived the expected sensitivity (shown as a black dashed line in Figure 2.3.1 simply by scaling the Υ(3S) estimated reach by $\sqrt{24}$. These searches have not been extended below the muon threshold because of large conversion backgrounds.

Finally, we also show the projected 2σ sensitivities for the proposed "APEX" experiment in JLab Hall A (dotted purple) [58] and the proposed "DarkLight" experiment using the JLab Free-

Electron Laser (FEL) (dotted blue) [59]. The vertexing reach of HPS covers a region that is not covered by either of the other two proposed experiments.

2.4 References

1. B. Holdom, Two U(1)'s and Epsilon Charge Shifts, Phys. Lett. B166 (1986) 196.
2. M. J. Strassler, Possible effects of a hidden val ley on supersymmetric phenomenology, hep-ph/0607160.
3. M. J. Strassler and K. M. Zurek, Echoes of a hidden val ley at hadron col liders, Phys. Lett. B651 (2007) 374–379, [hep-ph/0604261].
4. M. J. Strassler, Why Unparticle Models with Mass Gaps are Examples of Hidden Val leys, 0801.0629.
5. K. R. Dienes, C. F. Kolda, and J. March-Russell, Kinetic mixing and the supersymmetric gauge hierarchy, Nucl. Phys. B492 (1997) 104–118, [hep-ph/9610479].
6. S. A. Abel and B. W. Schofield, Brane-antibrane kinetic mixing, mil licharged particles and SUSY breaking, Nucl. Phys. B685 (2004) 150–170, [hep-th/0311051].
7. S. A. Abel, M. D. Goodsell, J. Jaeckel, V. V. Khoze, and A. Ringwald, Kinetic Mixing of the Photon with Hidden U(1)s in String Phenomenology, JHEP 07 (2008) 124, [0803.1449].
8. A. Ringwald, From Axions to Other WISPs, 0810.3106.
9. Y. Nomura and J. Thaler, Dark Matter through the Axion Portal, 0810.5397.
10. J. Mardon, Y. Nomura, and J. Thaler, Cosmic Signals from the Hidden Sector, 0905.3749.
11. J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, New Fixed-Target Experiments to Search for Dark Gauge Forces, Phys. Rev. D80 (2009) 075018, [0906.0580].
12. M. Baumgart, C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, Non-Abelian Dark Sectors and Their Collider Signatures, JHEP 04 (2009) 014, [0901.0283].
13. N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, A Theory of Dark Matter, Phys. Rev. D79 (2009) 015014, [0810.0713].
14. N. Arkani-Hamed and N. Weiner, LHC Signals for a SuperUnified Theory of Dark Matter, JHEP 12 (2008) 104, [0810.0714].
15. C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, Kinetic Mixing as the Origin of Light Dark Scales, 0902.3246.

16. D. E. Morrissey, D. Poland, and K. M. Zurek, Abelian Hidden Sectors at a GeV, 0904.2567.

17. A. Katz and R. Sundrum, Breaking the Dark Force, 0902.3271.

18. R. Essig, P. Schuster, and N. Toro, Probing Dark Forces and Light Hidden Sectors at Low-Energy e+e- Col liders, 0903.3941.

19. WMAP Collaboration, E. Komatsu et al., Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, Astrophys. J. Suppl. 180 (2009) 330–376, [0803.0547].

20. SDSS Collaboration, D. J. Eisenstein et al., Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, Astrophys. J. 633 (2005) 560–574, [astro-ph/0501171].

21. Supernova Cosmology Project Collaboration, S. Perlmutter et al., Measurements of Omega and Lambda from 42 High-Redshift Supernovae, Astrophys. J. 517 (1999) 565–586, [astro-ph/9812133].

22. Supernova Search Team Collaboration, A. G. Riess et al., Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant, Astron. J. 116 (1998) 1009–1038, [astro-ph/9805201].

23. Supernova Cosmology Project Collaboration, M. Kowalski et al., Improved Cosmological Constraints from New, Old and Combined Supernova Datasets, Astrophys. J. 686 (2008) 749–778, [0804.4142].

24. PAMELA Collaboration, O. Adriani et al., Observation of an anomalous positron abundance in the cosmic radiation, 0810.4995.

25. The Fermi LAT Collaboration, A. A. Abdo et al., Measurement of the Cosmic Ray e+ plus e- spectrum from 20 GeV to 1 TeV with the Fermi Large Area Telescope, 0905.0025.

26. J. Chang et al., An excess of cosmic ray electrons at energies of 300.800 GeV, Nature 456 (2008) 362–365.

27. H.E.S.S. Collaboration, F. Aharonian et al., The energy spectrum of cosmic-ray electrons at TeV energies, Phys. Rev. Lett. 101 (2008) 261104, [0811.3894].

28. H.E.S.S. Collaboration, F. Aharonian, Probing the ATIC peak in the cosmic-ray electron spectrum with H.E.S.S, 0905.0105.

29. D. P. Finkbeiner, L. Goodenough, T. R. Slatyer, M. Vogelsberger, N. Weiner, Consistent Scenarios for Cosmic-Ray Excesses from Sommerfeld-Enhanced Dark Matter Annihilation, [1011.3082].

30. The CoGeNT Collaboration, Results from a Search for Light-Mass Dark Matter with a P-type Point Contact Gemanium Detector, [1002,4703].

31. M. Pospelov and A. Ritz, Astrophysical Signatures of Secluded Dark Matter, Phys. Lett. B671 (2009) 391–397, [0810.1502].

32. J. Hisano, S. Matsumoto, and M. M. No jiri, Explosive dark matter annihilation, Phys. Rev. Lett. 92 (2004) 031303, [hep-ph/0307216].

33. J. March-Russell, S. M. West, D. Cumberbatch, and D. Hooper, Heavy Dark Matter Through the Higgs Portal, JHEP 07 (2008) 058, [0801.3440].

34. M. Cirelli, M. Kadastik, M. Raidal, and A. Strumia, Model-independent implications of the e+, e-, anti-proton cosmic ray spectra on properties of Dark Matter, Nucl. Phys. B813 (2009) 1–21, [0809.2409].

35. I. Cholis, G. Dobler, D. P. Finkbeiner, L. Goodenough, and N. Weiner, The Case for a 700+ GeV WIMP: Cosmic Ray Spectra from ATIC and PAMELA, 0811.3641.

36. I. Cholis, D. P. Finkbeiner, L. Goodenough, and N. Weiner, The PAMELA Positron Excess from Annihilations into a Light Boson, 0810.5344.

37. Y. Cui, D. E. Morrissey, D. Poland, and L. Randall, Candidates for Inelastic Dark Matter, JHEP 05 (2009) 076, [0901.0557].

38. A. Sommerfeld, Ann. Phys. 11 (1931) 257.

39. O. Adriani et al., A new measurement of the antiproton-to-proton flux ratio up to 100 gev in the cosmic radiation, 0810.4994.

40. D. Alves, S. R. Behbahani, P. Schuster, and J. G. Wacker, Composite Inelastic Dark Matter, 0903.3945.

41. D. Tucker-Smith and N. Weiner, Inelastic dark matter, Phys. Rev. D64 (2001) 043502, [hep-ph/0101138].

42. S. Chang, G. D. Kribs, D. Tucker-Smith, and N. Weiner, Inelastic Dark Matter in Light of DAMA/LIBRA, 0807.2250.

43. R. Bernabei et al., Dark matter particles in the galactic halo: Results and implications from DAMA/NaI, Int. J. Mod. Phys. D13 (2004) 2127–2160, [astro-ph/0501412].

44. DAMA Collaboration, R. Bernabei et al., First results from DAMA/LIBRA and the combined results with DAMA/NaI, Eur. Phys. J. C56 (2008) 333–355, [0804.2741].

45. A. W. Strong et al., Gamma-ray continuum emission from the inner Galactic region as observed with INTEGRAL/SPI, Astron. Astrophys. 444 (2005) 495, [astro-ph/0509290].

46. D. P. Finkbeiner and N. Weiner, Exciting Dark Matter and the INTEGRAL/SPI 511 keV signal, Phys. Rev. D76 (2007) 083519, [astro-ph/0702587].

47. M. Pospelov, Secluded U(1) below the weak scale, 0811.1030.

48. Muon G-2 Collaboration, G. W. Bennett et al., Final report of the muon E821 anomalous magnetic moment measurement at BNL, Phys. Rev. D73 (2006) 072003, [hep-ex/0602035].

49. The BABAR Collaboration, B. Aubert, Search for Dimuon Decays of a Light Scalar in Radiative Transitions Y(3S) □ γ A0, 0902.2176.

50. KTeV Collaboration, E. Abouzaid et al., Measurement of the rare decay pi0□ e+ e-, Phys. Rev. D75 (2007) 012004, [hep-ex/0610072].

51. M. Reece and L.-T. Wang, Searching for the light dark gauge boson in GeV-scale experiments, 0904.1743.

52. B. Batell, M. Pospelov, and A. Ritz, Exploring Portals to a Hidden Sector Through Fixed Targets, Phys. Rev. D80 (2009) 095024, [0906.5614].

53. Bossi, F., private communication

54. M. Pospelov, Secluded U(1) below the weak scale, 0811.1030.

55. J.D. Bjorken et.al., "Search for Neutral Metastable Penetrating Particles Produced in the SLAC Beam Dump", Phys. Rev. D38, 1988, 3375.

56. E.M. Riordan et.al., "A Search for Short Lived Axions in an Electron Beam Dump Experiment", Phys. Rev. Lett., 59, 1987, 755.

57. A. Bross et.al., "A Search for Shortlived Particles Produced in an Electron Beam Dump", Phys. Rev. Lett., 67, 1991, 2942-2945.

58. R. Essig, P. Schuster, N. Toro, B. Wojtsekhowski, "An Electron Fixed Target Experiment to Search for a New Vector Boson A' Decaying to e+e-", arXiv:1001.2557.

59. M. Freytsis, G. Ovanesyan, J. Thaler, "Dark Force Detection in Low Energy e-p Collisions", [arXiv: 0909.2862] JHEP 1001 (2010) 111.

60. A. Fitzpatrick, D. Hooper, K. M. Zurek, Implications of CoGeNT and DAMA for Light WIMP Dark Matter, Phys. Rev. D81 (2010) 115005, [1003.0014].

61. R. Essig, J. Kaplan, P. Schuster, N. Toro, On the Origin of Light Dark Matter Species, [1004.0691].

3 Signal and Backgrounds

3.1 Heavy Photon Signal

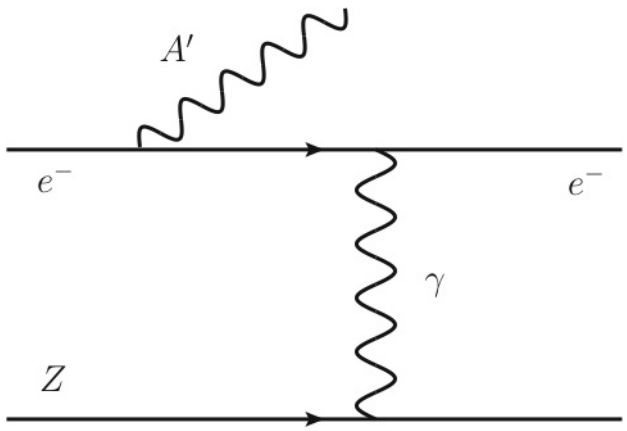

Figure 3.1.1 A' production by bremsstrahlung off an incoming electron as it scatters on a nucleus with atomic number Z.

A' particles are generated in electron collisions on a fixed target by a process analogous to ordinary photon bremsstrahlung, see Figure 3.1.1. This can be reliably estimated in the Weizsäcker-Williams approximation (see [1-4]). When the incoming electron has energy E_0, the differential cross-section to produce an A' of mass $m_{A'}$ with energy $E_{A'} \equiv x\,E_0$ is

$$\frac{d\sigma}{dx\; d\cos\theta_{A'}} \approx \frac{8\,Z^2\alpha^3\epsilon^2{E_0}^2 x}{U^2}\;\tilde{\chi}\times\left[\left(1-x+\frac{x^2}{2}\right)-\frac{x(1-x){m_{A'}}^2\,{E_0}^2 x\,{\theta_{A'}}^2}{U^2}\right] \qquad (3.1)$$

where Z is the atomic number of the target atoms, $\alpha\simeq 1/137$, $\theta_{A'}$ is the angle in the lab frame between the emitted A' and the incoming electron,

$$U(x,\theta_{A'}) = {E_0}^2 x{\theta_{A'}}^2 + {m_{A'}}^2\frac{1-x}{x} + {m_e}^2 x \qquad (3.2)$$

is the virtuality of the intermediate electron in initial-state bremsstrahlung, and $\tilde{\chi}$ is the Weizsäcker-Williams effective photon flux, with an overall factor of Z^2 removed. The form of $\tilde{\chi}$ and its dependence on the A' mass, beam energy, and target nucleus are discussed in [1]. For HPS with E_0 = 6.6 GeV, we find $\tilde{\chi}$ ~ 7 (4, 1) for $m_{A'}$ = 100 (200, 500) MeV/c^2. The above results are valid for

$$m_e \ll m_{A'} \ll E_0, \quad x\,{\theta_{A'}}^2 \ll 1$$

For $m_e \ll m_{A'}$, the angular integration gives

$$\frac{d\sigma}{dx} \approx \frac{8\,Z^2\alpha^3\epsilon^2 x}{{m_{A'}}^2}\left(1+\frac{x^2}{3\,(1-x)}\right)\tilde{\chi}\,. \qquad (3.3)$$

The rate and kinematics of A' radiation differ from massless bremsstrahlung in several important ways:

- **Rate:** The total A' production rate is controlled by $\alpha^3\epsilon^2/m_{A'}^2$. Therefore, it is suppressed relative to photon bremsstrahlung by $\sim \epsilon^2 m_e^2/m_{A'}^2$. Additional suppression from small $\tilde{\chi}$ occurs for large $m_{A'}$ or small E_0.
- **Angle:** A' emission is dominated at angles $\theta_{A'}$ such that $U\left(x,\theta_{A'}\right) \lesssim 2\,U(x,0)$ (beyond this point, wide-angle emission falls as ${\theta_{A'}}^4$. For x near its median value, the cutoff emission angle is

  $$\theta_{A',\max} \sim \max\left(\frac{\sqrt{m_{A'}\,m_e}}{E_0}, \left({m_{A'}}/{E_0}\right)^{\frac{3}{2}}\right),$$

  which is parametrically smaller than the opening angle of the A' decay products, $\sim m_{A'}/E_0$. Although this opening angle is small, the backgrounds mimicking the signal (discussed in Section 3.2) dominate at even smaller angles.
- **Energy:** A' bremsstrahlung is sharply peaked at ≈ 1 , where $U(x,0)$ is minimized. When an A' is produced, it carries nearly the entire beam energy. In fact the median value of $(1-x)$ is $\sim \max\left(\frac{m_e}{m_{A'}}, \frac{m_{A'}}{E_0}\right)$.

The latter two properties are quite important in improving signal significance, and are discussed further in Section 3.2.

Assuming the A' decays into Standard Model particles rather than exotics, its boosted lifetime is

$$l_0 \equiv \gamma\, c\tau \simeq \frac{(3\,E_{A'})}{\left(N_{eff} m_{A'}^2 \alpha\,\epsilon^2\right)} \simeq \frac{0.8\,cm}{N_{eff}}\left(\frac{E_0}{10\,GeV}\right)\left(\frac{10^{-4}}{\epsilon}\right)^2\left(\frac{100\,MeV}{\mathrm{m_{A'}}}\right)^2, \qquad (3.4)$$

where we have neglected phase-space corrections, and N_{eff} counts the number of available decay products. If the A' couples only to electrons, then $N_{eff}=1$. If the A' mixes kinetically with the photon, then $N_{eff}=1$ for $m_{A'} < 2m_\mu$ when only $A' \to e^+e^-$ decays are possible, and $2+R(m_{A'})$ for $m_{A'} \geq 2\,m_\mu$, where [5]

$$R = \left.\frac{\sigma(e^+e^- \to hadrons)}{\sigma(e^+e^- \to \mu^+\mu^-)}\right|_{E=m_{A'}}.$$

For the ranges of ε and $m_{A'}$ probed by this experiment, the mean decay length l_0 can be prompt or as large as tens of centimeters.

The total number of A' produced when N_e electrons scatter in a target of $T \ll 1$ radiation lengths is

$$N \sim N_e \frac{N_0 X_0}{A} T \frac{Z^2 \alpha^2 \epsilon^2}{{m_{A'}}^2} \tilde{\chi} \sim \; N_e C \, T \frac{\epsilon^2 \, {m_e}^2}{{m_{A'}}^2}, \qquad (3.5)$$

where X_0 is the radiation length of the target in g/cm^2, $N_0 \simeq 6\times 10^{23} mole^{-1}$ is Avogadro's number, and A is the target atomic mass in g/mole. The numerical factor $C \approx 5$ is logarithmically dependent on the choice of nucleus (at least in the range of masses where the form-factor is only slowly varying) and on $m_{A'}$, because, roughly, $X_0 \propto \frac{A}{Z^2}$ (see [1,3,4]). For a Coulomb of incident electrons, the total number of A's produced is given by

$$N \sim 10^5 \left(\frac{N_e}{1\ C}\right) \tilde{\chi} \left(\frac{T}{0.1}\right) \left(\frac{\epsilon}{10^{-4}}\right)^2 \left(\frac{100\ MeV}{m_{A'}}\right)^2$$

The efficiency of the detector for detecting these A's is discussed in Section 5.

3.2 Radiative and Bethe-Heitler Trident backgrounds

The stark kinematic differences between QED trident backgrounds and the A' signal can be used to advantage to maximize the signal to background ratio. QED tridents dominate the final event sample, so we consider their properties in some detail here.

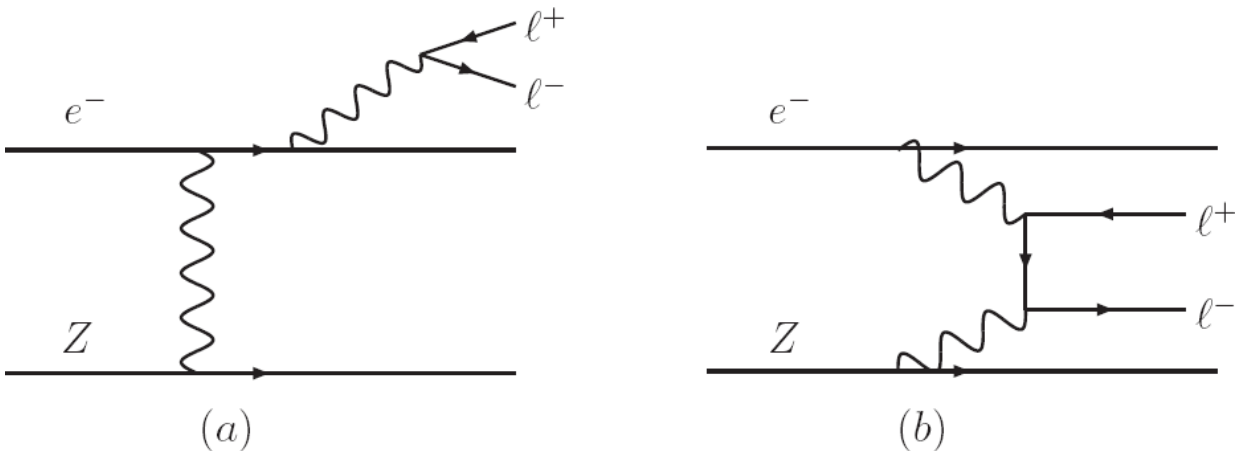

Figure 3.2.1: Sample diagrams of (a) radiative trident (γ*) and (b) Bethe-Heitler trident reactions that comprise the primary QED background to A' → ℓ+ℓ- search channels.

The irreducible background rates are given by the diagrams shown in Figure 3.2.1. These trident events can be usefully separated into "radiative" diagrams (Figure 3.2.1 (a)), and "Bethe-Heitler" diagrams (Figure 3.2.1 (b)), that are separately gauge-invariant (we discuss the simulation of the QED trident background events in Section 6).

The contribution from the radiative diagrams (Figure 3.2.1 (a)) alone is also useful as a guide to the behavior of A' signals at various masses. Indeed, the kinematics of the A' signal events is

identical to the distribution of radiative trident events restricted in an invariant mass window near the A' mass. Moreover, the rate of the A' signal is simply related to the radiative trident cross-section within the spectrometer acceptance and a mass window of width δm by [4]

$$\frac{d\sigma\left(e^{-}Z\rightarrow e^{-}Z(A'\rightarrow \ell^{+}\ \ell^{-})\right)}{d\sigma\left(e^{-}Z\rightarrow e^{-}Z(\gamma*\rightarrow \ell^{+}\ \ell^{-})\right)} = \frac{3\pi\epsilon^{2}}{2\,N_{eff}\alpha}\,\frac{m_{A'}}{\delta m}\ , \qquad (3.6)$$

where N_{eff} counts the number of available decay products. This exact analytic formula was also checked with a MC simulation of both the A' signal and the radiative trident background restricted to a small mass window δm, and we find nearly perfect agreement. Thus, the radiative subsample can be used to analyze the signal, which simplifies the analysis considerably.

It is instructive to compare kinematic features of the radiative and Bethe-Heitler distributions, as the most sensitive experiment maximizes acceptance of radiative events and rejection of Bethe-Heitler tridents. Although the Bethe-Heitler process has a much larger total cross-section than either the signal or the radiative trident background, it can be significantly reduced by exploiting its very different kinematics. In particular, the A' carries most of the beam energy (see discussion in Section 3.1), while the recoiling electron is very soft and scatters to a wide angle. In contrast, the Bethe-Heitler process is not enhanced at high pair energies. Moreover, Bethe-Heitler processes have a forward singularity that strongly favors asymmetric configurations with one energetic, forward electron or positron and the other constituent of the pair much softer.

These properties are discussed further in the Appendix of [4], and illustrated in Figure 3.2.2.

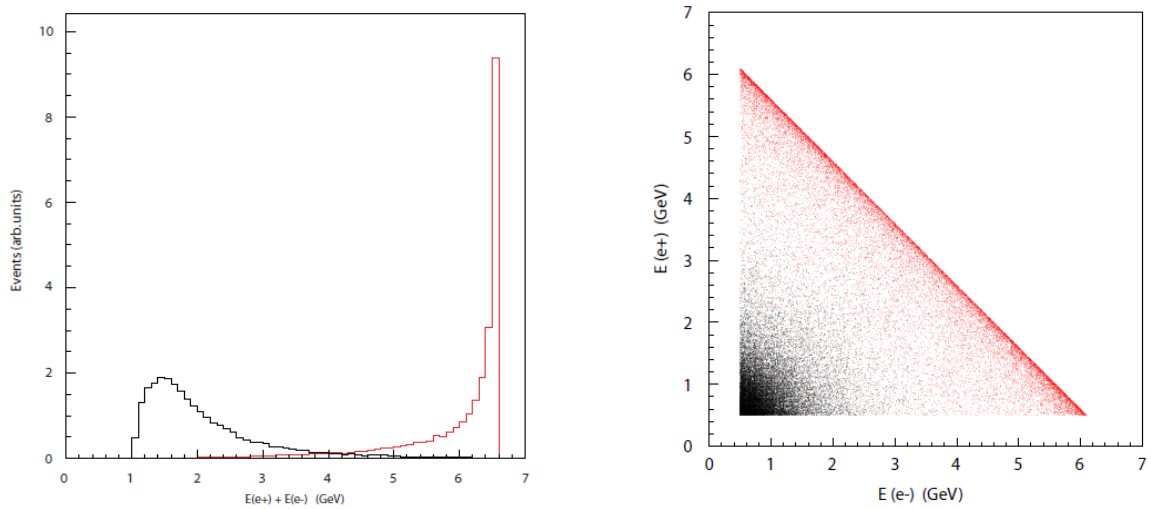

Figure 3.2.2: Left: The distribution of Bethe-Heitler background events (black) and A' signal events (red) as a function of the sum of the electron and positron energy. Note that the signal is peaked at high energies, while the background is peaked at much lower energies. Right: The distribution of the positron versus electron energy for Bethe-Heitler background events (black dots) and A' signal events (red dots). Note that in both plots neither the signal nor background events have been normalized to the correct number. In reality, the number of background events is much larger than the number of signal events. Also, note that

the electron energy here refers to the energy of the electron produced in the reaction, not the recoiling beam electron.

3.3 Beam Backgrounds

The experimental requirements of a high beam current and a large forward detector coverage conspire to make beam backgrounds a significant challenge in the design of the experiment. The multiple Coulomb scattering and bremsstrahlung processes in the target will generate high intensity fluxes of electrons and photons in the very forward direction, while the large Moller interaction cross section with atomic electrons will generate high intensity low energy electrons. The beam backgrounds are simulated using the high energy interaction simulation codes, EGS5 and Geant4.

The beam backgrounds determine a "dead zone" where the high intensity particle flux must be avoided to eliminate secondary particle production and extensive radiation damage in the detectors. All the detector components must be placed outside of the dead zone. To determine its extent, we have considered the detector occupancy and radiation tolerance in the Si tracker, and the trigger rate and radiation tolerance in the electromagnetic calorimeter. Multiple Coulomb scattering in the target increases the primary beam size and angular divergence as it passes through the apparatus, and limits how closely detector elements can be placed to the beam. As the rms scattering angle is $\langle\theta\rangle \sim \sqrt{T/E_0}$, a thinner target and/or higher energy beam can minimize the dead zone. When the target thickness is reduced, however, the beam current (I) must be increased proportionately in order to maintain the A' production rate, $T \cdot I$. See Section 5.2 for details.

An additional process that yields high-energy electron-positron pairs is the "two-step" (or incoherent) trident production, consisting of a high-energy bremsstrahlung photon emission followed by the e^+e^- conversion inside the target. As this is a two-step process, the production rate is proportional to T^2. This process has been simulated using EGS5. The rate becomes comparable to that of the radiative trident background described above for a 1% X_0 tungsten target, and is significantly less for the target thicknesses proposed for this experiment.

3.4 References

1. K.J Kim and Y.-S. Tsai, "Improved Weizsacker-Williams methos and its application to lepton and W-boson pair productions", Phys. Rev. D8, 1973, 3109.

2.] J. Alwall et.al., "MadGraph/MadEvent v4: The New Web Generation", JHEP, 09, 2007, 028, arXiv: 0706.2334.

3. R. Essig, P, Schuster, N. Toro, and B. Wojtsekhowski, "An Electron Fixed Target Experiment to Search for a New Vector Boson A' Decaying to e+e-", arXiv: 1001.2557.

4. J.D. Bjorken, R. Essig, P. Schuster, and N. Toro, "New Fixed-Target Experiments to Search for Dark Gauge Forces", Phys. Rev. D80, 2009, 075018, arXiv: 0906.0580.

4 Experimental Setup

4.1 Overview

The proposed experiment will search for a heavy photon (dark photon) in the mass range from 20 MeV to 1000 MeV in two settings of beam energy 2.2 GeV and 6.6 GeV. Sensitivity to the dark photons relies upon the precision measurement of two quantities in this experiment: the invariant mass of the A′ decay products and the position of the decay vertex. By placing a tracking and vertexing detector immediately downstream of the target inside an analyzing magnet, the complete kinematic information required for A′ reconstruction can be obtained from a single system, whose proximity to the target naturally maximizes the acceptance of a relatively compact detector and provides excellent momentum and vertexing resolution.

Placing the planes of the tracker immediately downstream of the target means that the intense primary beam must pass directly through the middle of the tracking detector. There are two key consequences of this arrangement. Firstly, scattered beam particles and radiative secondaries are bent by the magnetic field to sweep out a "dead zone" where the particle fluxes would be damaging to the sensors as well as creating an environment too dense for pattern recognition. This necessitates a tracking geometry that keeps the sensors out of this region. However, since the energy released in the decay of low mass A′ is small relative to their boost, the opening angle between decay daughters can be quite small. Therefore, to maximize the acceptance for low masses, the size of the dead zone must be minimized. Secondly, interactions of the primary beam with air or even helium at atmospheric pressure gives rise to low-momentum secondaries that generate unacceptable occupancies in the detector. The only way to keep the beam in a vacuum without severely compromising acceptance and vertex resolution is to enclose the entire tracking and vertexing system within a vacuum chamber as well. The background environment and the high spatial precision required for good momentum and vertexing resolution make silicon sensors the obvious choice in this region.

High luminosities are needed to search for heavy photons with small couplings and masses in the 100 MeV range. Utilizing CEBAF's essentially continuous duty cycle, the experiment can simultaneously maximize luminosity and minimize backgrounds by employing detectors with short livetimes and rapid readout. Silicon tracking sensors are ideal from this perspective, since they collect ionization in 10's of nanoseconds and produce pulses as short as 50-100 nanoseconds. Thanks to electronics developed for the LHC, the sensors can be read out continuously at 40 MHz.

The electromagnetic calorimeter just downstream of the tracker uses detectors with comparably short livetimes and high rate capability. It performs two essential functions for the experiment: triggering and electron identification. The device is highly segmented. It is fast, able to readout at rates comparable to those in the tracker, and able to provide good spatial and energy information to the trigger electronics. Like the tracker system, the electromagnetic calorimeter is

split to avoid impinging on the "dead zone". The beam and radiative secondaries pass through the calorimeter in vacuum, to avoid generating unnecessary backgrounds.

The muon detector also exhibits fast response, will be read out rapidly, and will be used to trigger on candidate tracks which pass through ~ 1m iron. It provides an independent trigger, with very different systematics, for A' masses above about 300 MeV/c, and a unique window on A' dimuon decays. The beam will pass through the middle of the detector in vacuum, and eventually be transported to the B-line dump.

The various elements of the experiment are discussed in more detail below, beginning with the beamline, continuing with the tracker/vertexer, electromagnetic calorimeter, the muon system, and concluding with the electronics and DAQ.

4.2 Beamline Elements

4.2.1 Layout

The HPS experiment will utilize a setup located behind the CLAS detector, in the downstream alcove of Hall B, shown in Figure 4.2.1.1. The setup will be based on a three-magnet chicane, the second dipole magnet serving as the analyzing magnet. The dipole field direction (Y) will be perpendicular to the horizontal (XZ) plane. A Hall B "Frascati" H magnet will be used (pole length 50 cm, max-field 1.2 T) as the first dipole of the chicane. The analyzing magnet will be an 18D36 H-magnet from SLAC (pole length 91.44 cm, max-field ~1.2 T) with the gap increased to 14 inches. As the last magnet of the chicane, the C-magnet located in the downstream tunnel (the old pair spectrometer of Hall B) will be used. The distance between centers of the first and second magnet and between the second and the third magnet will be ~500 cm. This distance is defined by the length of the detector package (ECal and muon system) and the space necessary for the shielding wall before the third magnet. The analyzing magnet will be operated at a 1T-m field for the 6.6 GeV run to maximize the momentum resolution. The first and third magnets will be operated at 0.5T-m. For the 2.2 GeV running, the analyzing magnet and the two bending magnets will be set to 0.5T-m and 0.25T-m fields, respectively. The analyzing magnet and the detector system behind it will be mounted on a stand that will allow transverse motion in order to accommodate operations at different beam energies.

The detector package will include multiple layers of silicon detectors, mounted inside the vacuum box in the high field region of the analyzing magnet. They are described in Section 4.3 below. Downstream of the analyzing magnet there will be an electromagnetic calorimeter, for triggering and electron and positron identification (see Section 4.4), and also multiple layers of scintillation hodoscopes sandwiched between iron absorbers for muon identification (See Section 4.5). The target foil will be positioned at the beginning of the high field region of the analyzing dipole. The distance from the target to the first layer of the silicon tracker will be 10 cm. The distance from the target to the face of the electromagnetic calorimeter is ~130 cm. The whole

length of the setup from the target to the end of the muon hodoscope system is approximately 305 cm.

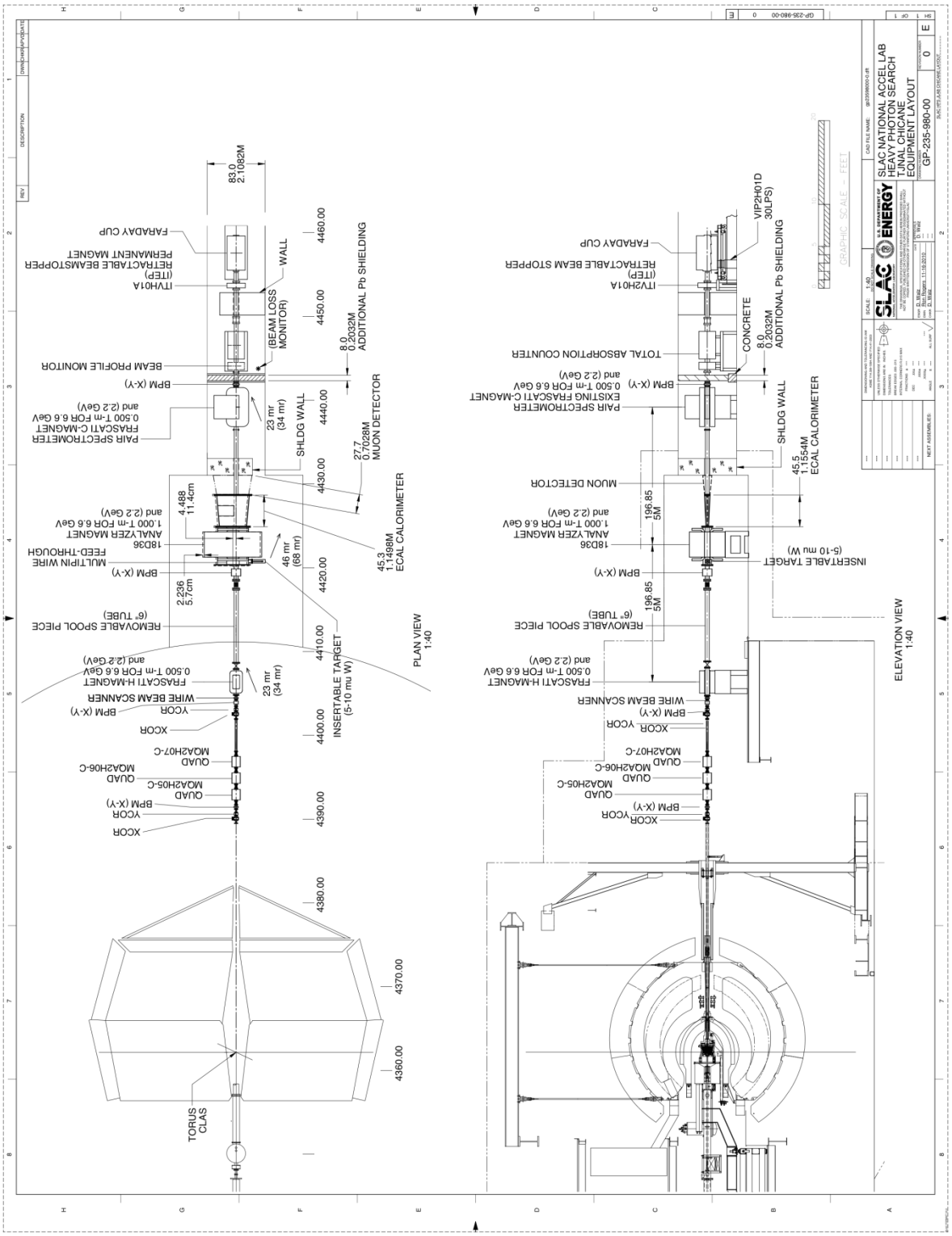

Figure 4.2.1.1. Layout of the HPS experimental setup behind the CLAS detector in Hall B.

4.2.2 **Running Conditions**

Production data will be taken at two beam energies, ~2.2 GeV and ~6.6 GeV, incident on a tungsten (W) target. The proposed luminosity for production running is 0.3×10^{32} cm^{-2} s^{-1} per nucleus at 2.2 GeV and 1.5×10^{32} cm^{-2} s^{-1} per nucleus at 6.6 GeV. These luminosities will be achieved using electron beam currents of 200nA to 450nA, and tungsten target foil thicknesses from 5μm (0.14% RL) to 8μm (0.25% RL). The high intensity electron beam incident on the tungsten target will generate a significant amount of electromagnetic radiation, composed mainly of bremsstrahlung photons, electrons which have radiated, Moller electrons, and beam particles which have multiple Coulomb scattered. In the dipole field, this radiation will create a "sheet of flame" in the bending plane (XZ), at the beam height, Y≈0. Detectors will be positioned above and below the beam plane, leaving a small gap for the bremsstrahlung and Moller background to pass through. The gap between the up and down planes of the first layer of the silicon tracker will be approximately ±15 mrad. Operational experience shows that the CEBAF beam is very clean, and is contained within Y=±1mm with halo at the level of less than 10^{-5}, so it passes through the "dead zone" gap. The beam halo that extends farther, ±2mm, is at the level of 10^{-7}, as seen in Figure 4.2.2.1.

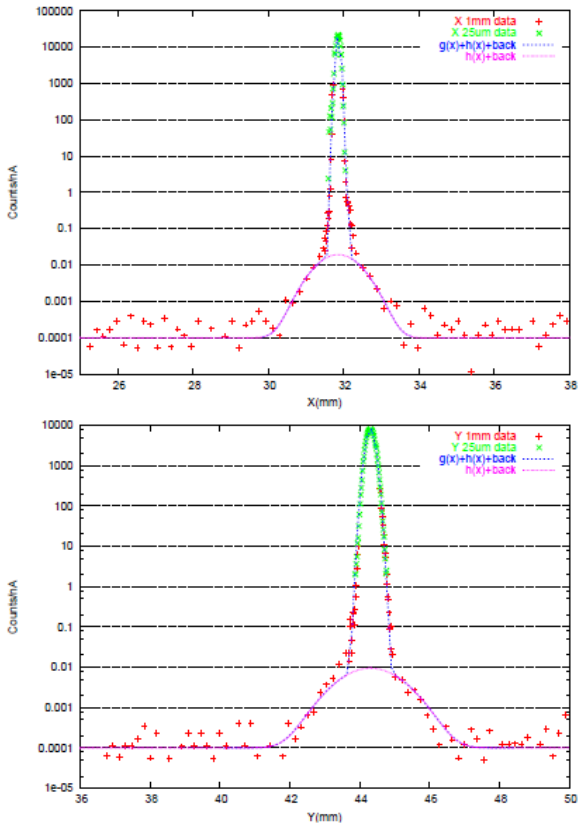

Figure 4.2.2.1: Hall B 6GeV beam profile measured in a large dynamic range profile measurement. Figure taken from A.P. Freyberger Proceeding of the 2003 Particle Accelerator Conference 0-7803-7739-9 ©2003 IEEE.

The gap between the up and down parts of the electromagnetic calorimeter will be about ±15 mm. The beam will be terminated in the Hall B electron beam dump. The transport beam line to the dump will be in vacuum. The electron beam will have clear passage to the dump when the chicane is OFF. The proposed running conditions and the close proximity of detector elements to the beam line impose tight constraints on the beam profile quality. The HPS beam parameter requirements are presented in Table 4.2.2.1.

Parameter	Requirement/Expectation	Unit
E	2200 and 6600	MeV
$\delta p/p$	$< 10^{-4}$	
Current	> 100 and < 1000	nA
Current Instability	< 5	%
σ_x	< 30	μm
σ_y	< 30	μm
Position Stability	< 30	μm
Divergence	< 100	μrad
Beam Halo (> 5σ)	$< 10^{-5}$	

Table 4.2.2.1 Required beam parameters.

4.2.3 **Beamline Proposal**

The proposed HPS target location is approximately 18.6 m downstream of the nominal CLAS target location. This location requires that the optics be altered to account for the new location of the experimental target. Adding additional quadrupole magnets to the Hall-B beamline will reduce the beam size at the HPS target. Small beam size (<30μm) is important for the experiment because it will allow decay angles to be measured with greater precision and will also improve the vertex definition by adding the additional constraint of the A′ trajectory. There is enough physical space to add the quadrupole magnets between the Forward Carriage and the HPS chicane.

The locations of the HPS chicane and other elements used in the optics development are given in Table 4.2.3.1. The optics program ELEGANT is used to determine the optimized element parameters for HPS. The Hall-B ELEGANT deck has been modified to include these new elements for this study. The deck does not include diagnostic elements or correctors at this time.

Expected characteristics of the future 12 GeV machine are used for optimization. It requires that β_x and β_y be less than 400 μm throughout the line and that β_x and β_y are at a minimum at the HPS target (minimum beam size). An emittance of $\varepsilon_x = 3\text{x}10^{-10}$ m-rad and $\varepsilon_y = 2\text{x}10^{-10}$ m-rad are used as input.

Item	Distance in Z (m)	ΔZ(m)	Deck name
CLAS Nominal Target	0	0	ETA2HCLAS
End of Forward Carriage	8.07	0	-
Quad-I	9.1	0.5	MQA2H05-C
Quad-II	9.7	0.5	MQA2H06-C
Quad-III	10.3	0.5	MQA2H07-C
First Dipole	13.2	1	MBX2H90
Target	17.7	0.25	ETA2HTEST
Second Dipole	18.2	1	MBX2H91
Third Dipole	23.2	1.5	MBX2H92
Faraday Cup	28.95	0	IFY2D00

Table 4.2.3.1 Elements used in optics development.

In order to achieve $\varepsilon \sim 2\text{x}10^{-10}$ m-rad the CEBAF accelerator must be well matched, and also the Hall-B line must be properly matched at the start of the line. There exist explicitly for this purpose a wire scanner and matching quadrupole magnets at the start of the Hall-B line. The process of matching the Hall-B line is rarely invoked as empirical tuning suffices to achieve the experimental requirements.

Results from the optics program ELEGANT for the horizontal and vertical beam sizes for a 3 pass beam at 6.6 GeV (12 GeV CEBAF setting) with three additional quadrupoles located on the Forward Carriage are shown in Figure 4.2.3.2. With these additional quadrupoles, a beam size of ~10 μm with 20-30 μrad beam divergence can be achieved at the HPS target, as shown in Figure 4.2.3.3. For lower beam energies, the beam size at the HPS target will be somewhat larger, e.g. 70% larger for a 1 pass beam (2.2 GeV).

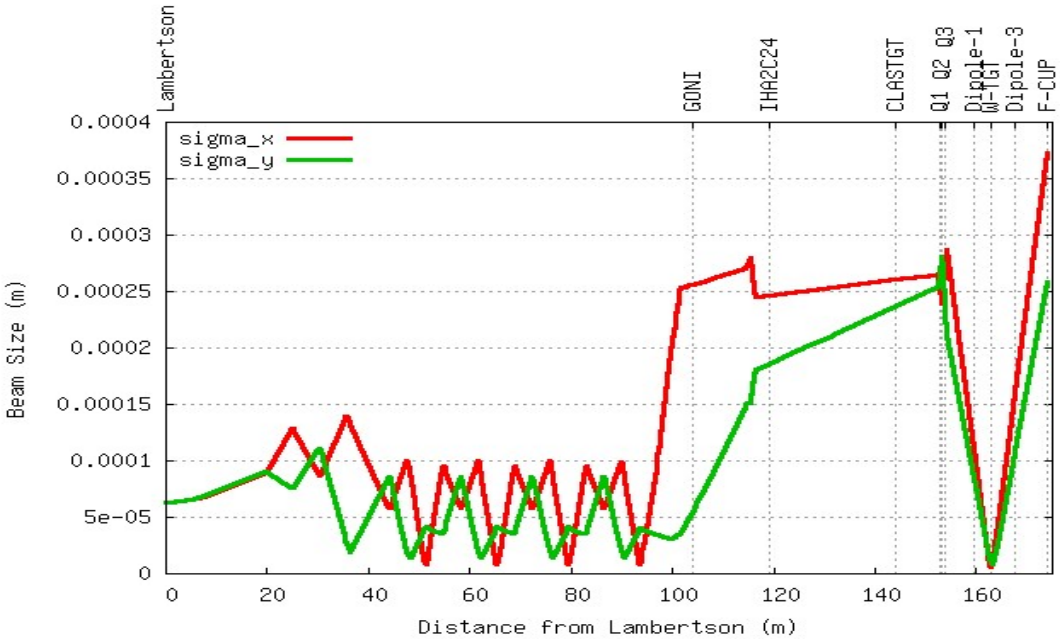

Figure 4.2.3.2. Beam sizes, X and Y, from the CLAS center to the Faraday Cup with additional quadupoles at forward carriage.

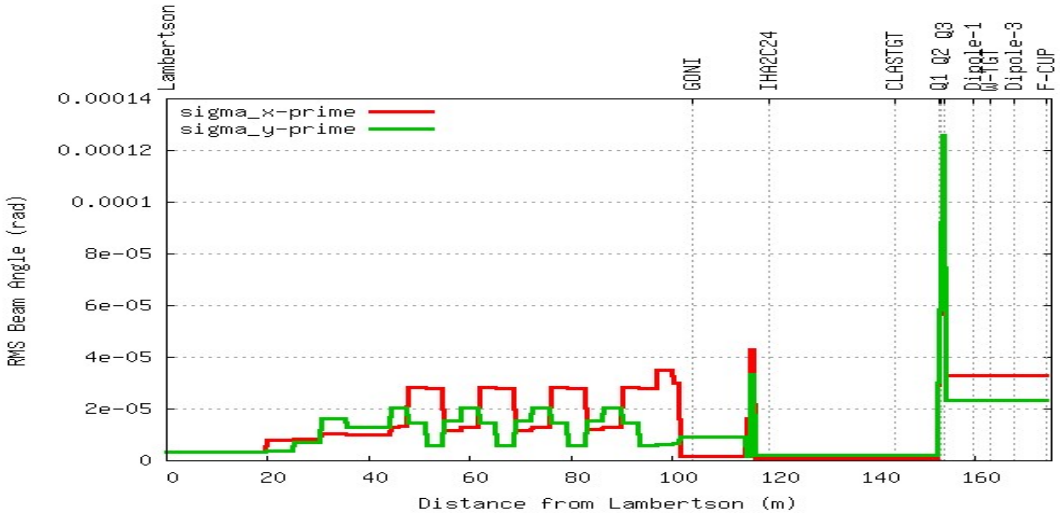

Figure 4.2.3.3. Beam divergence at the HPS setup region.

4.2.3.1 **Diagnostics and Trajectory Control**

Hall-B experiments typically require nA-level beam currents. Measuring beam positions with such small beam currents is difficult; presently Hall-B relies on three sets of cavity beam position monitors (BPM). These monitors are physically large and rather slow. The large (100+ nA) beam currents envisioned for the HPS experiments allow the use of the standard CEBAF antenna BPM cans with transport style electronics. The use of antenna BPMs for the HPS experiment will reduce the space/cost of beam position measurements. The antenna BPM "cans" come in two sizes, M15 and M20. The M15 has a standard 2-3/4" flange and is usually associated with 1" or 1.5" beam tubes. The M20 BPM has a 4-5/8" flange and is usually associated with 2" to 3" beam tube. The BPM measured position is linear within about ±7 mm of the electric center of the BPM can.

Adding M15 and/or M20 to the HPS beamline will allow the beam position to be measured in a straightforward manner. A pair of BPMs upstream of the first dipole will define the incoming trajectory. The third BPM a M20 or M15 will be located at the end of the chicane, just upstream of the third dipole. Trajectory control of the incident beam will require at least four corrector dipoles (two horizontal and two vertical) between the forward carriage and the first dipole. In addition to BPMs, a harp, and a harp and a beam viewer will be installed before the first dipole and before the HPS target, respectively. Harps and the viewer will be used periodically to measure the beam size and check the beam profile.

The three dipoles, which form the chicane, might not be on bi-polar supplies so it will not be possible to degauss them easily. Therefore a corrector should be associated with each of these magnets so that any remnant field can be "nulled" out to achieve a straight-ahead trajectory. The strengths of these correctors are yet to be determined.

4.2.4 **Targets**

Thin tungsten foil will be used. High Z material is chosen to minimize the hadronic production relative to the trident and A' production, since the ratio of QED to hadronic processes goes as (Z^2/A). The target will be located 10 cm in front of the first plane of silicon strip detectors. The primary target will be 0.25% of a radiation length (approximately 9 microns tungsten). Foils of 0.1% and 1% may also be substituted for some of the data taking, adjusting the beam current as appropriate.

Production rate and background studies lead to the use of beam currents up to 500 nanoamps with the 0.25% radiation length target. Simulation studies have also indicated that a beam spot as small as 10 μm rms would lead to enhanced discrimination against backgrounds in vertex reconstruction. These concentrated beam intensities would be accompanied by severe local heating of the target material. Simulation shows that a 10 μm rms beam of 500 nanoamps would cause a temperature rise to melting point within about 7 msec. This situation may be handled by

moving the beam across the target surface or vice versa. Maintaining a stable beam position is considered important for acuity in vertex reconstruction, and so it is proposed to move the target continuously across the beam.

The strength of tungsten drops by an order of magnitude with temperature increases in the range of 1000 C. In addition, the material re-crystallizes above this range, which increases the tendency for cracking where thermal expansion has caused temporary dimpling. For these reasons, we plan to keep the temperature rise below 1000 degrees. Although complete simulations of the heating of a moving target remain to be done, it has been established that moving at 1.2 cm per second would achieve this goal. Study of the long term cooling after the beam passes, and concerns about fatigue lifetime, lead to a requirement that the beam should not return to the same spot on the target for 20 seconds.

The combination of vacuum, magnetic field, potential radiation damage, and the need for reliability, presents some challenges for the target design. As insurance against unwanted effects of the magnet tripping off or related to motion, structural parts that are in the strong fringe field but not firmly tied down will be of insulating material. The device additionally has to be compatible with the silicon strip detector system and its cable and cooling plant, as well as beam line diagnostic equipment. Several options are being considered, of which the three most promising will be mentioned here.

The simplest option under consideration moves the foil by installing it around the periphery of a disc which spins at a few RPM, as shown in Figure 4.2.4.1. The disk is positioned and rotated by a long, stiff, shaft, which is at an angle to the beam axis to allow space for the actuation mechanism. The shaft is constrained radially by a vacuum bearing, lubricated by molybdenum disulfide rather than fluorocarbons (e.g. NSK YS series). At about 40 cm from the front face of the analyzing magnet coils, the field has dropped below 500 gauss, and a double cylinder of iron surrounding the target mechanism reduces the field further, below 50 gauss. This allows the option of using a magnetic coupler to transmit rotational motion through the vacuum pipe wall, or of a bellows sealed wobble plate design (e.g. Inficon FRU040-L). Beyond this is a conventional small gear motor in air (e.g. a motor/feedthrough unit MDC BRM-275-03).

To provide the option of retracting the target from the beam line, the shaft, together with its bearings, coupler and motor, can be driven outwards along its axis, using a bellows in its vacuum housing, and a linear stage that supports the movable section. Controls and read-back for the motor and in-out motion are conventional and will be compatible with the experiment's control system. However, a signal for the proper functioning of the target will be provided, and this must shut off the beam if a fault is detected. This mechanism would allow the beam to scan a 24 cm circumference circle on the target. A better use of the area of the foil would be achieved, with benefit for its longevity, by allowing for a small pitch-angle motion of the rod. This requires a hinge point in the support and a second, transverse, linear actuator. The clearance between the rotating shaft and the bellows has to be assured in this extension of the design, which is under evaluation.

Alternative mechanisms under consideration make use of bellows deflections to permit the full motion of the foil. In one case, a short section of the upstream beam pipe, with bellows at either end, is supported by a gimbal. Inside the vacuum, attached to this ring, is a cantilever stretching along the beam pipe and supporting the target foil at its location in the mouth of the magnet. The ring can be tilted vertically and horizontally by a pair of magnetically shielded actuators in air, and they can implement a rastering motion of the foil. The cantilever can actually support several foils of different thicknesses for systematics studies. However, the fatigue life of the bellows remains to be evaluated, and so the advantages of the technique may not compensate for the operational risk. A version of this device with only one axis of motion was used for a pair of wire scanners in the core of the Mark II detector at SLAC.

The bellows life is being evaluated also for the third alternative, which considers a piston rod extending into the vacuum at right angles to the beam, just outside the coils of the magnet. In this case the flange of the rod is attached to a pair of actuators that can enforce an up-down motion and an in-out motion of the foil. The bellows extension has to cover the full piston stroke in this case. Because of the magnetic field adjacent to the coils, the actuators would be positioned at some distance from the vacuum port. This version would also allow for foils of several thicknesses to be mounted for systematics studies. Its advantage is in keeping beam line space free for beam diagnostic equipment to be installed closer to the target.

One of these options will be selected after their designs and that of their experimental environment are more fully developed. A functioning target will then be prepared for beam tests, and a long operational reliability test of the final version will be made before installation at the experiment.

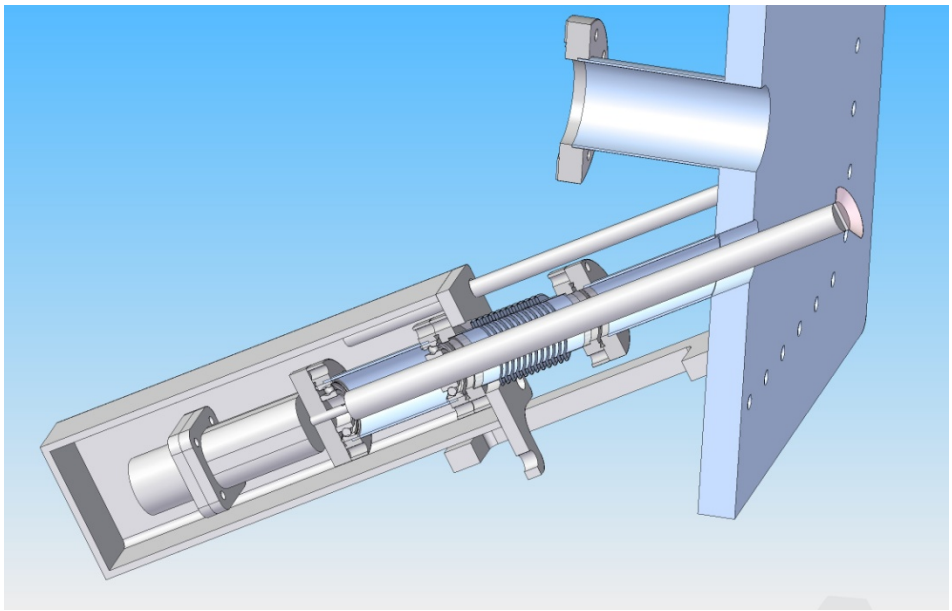

Figure 4.2.4.1 Rotatable and retractable target mechanism

4.3 Tracking and Vertexing System

4.3.1 Design Considerations

At the beam energies necessary to achieve sensitivity to A′ in the most interesting mass range, the decay products will be charged particles with momenta on the order of a few GeV/c. Consequently, multiple scattering dominates the measurement uncertainties for any feasible material budget. In particular, uncertainty in the mass determination and uncertainty in the vertex position are limited by the uncertainties in the directions of the A′ decay products, which results from multiple Coulomb scattering in the first detector layers. As a result, for both the mass and vertexing resolutions, the radiation lengths of material must be minimized throughout the tracker. Doing so also results in the optimal momentum resolution and minimizes tracking pattern recognition errors in an environment of dense hits.

To minimize the impact of beam backgrounds, the tracker/vertexer is split into upper and lower half-planes which avoid the "dead zone" defined by the beam envelope and those electrons which have radiated in the target. To avoid secondary backgrounds, which would arise from the passage of the beam through gas, the beam must be transported in vacuum, and the entire tracker/vertexer placed within a vacuum chamber.

Finally, the position and tails of the beam profile may not be exactly known, and there are significant uncertainties in the background estimates and the radiation tolerance of the sensors for the background environment. So it will be important to be able to adjust the position of the tracking planes in-situ, or even remotely. It is also sensible to provide for access to the tracker with minimal intervention, to accommodate repairs or replacement of tracking planes.

4.3.2 Sensors

Given a high density of hits that results in pattern recognition challenges, it is natural to consider pixellated sensors for this experiment. However, the occupancies per unit time are such that the only technology currently meeting the rate requirements is a hybrid pixel detector. This is not a feasible solution here, primarily because the power and cooling required in the tracking volume would result in a material budget that seriously compromises sensitivity. It is also likely that the cost and effort associated with development of a hybrid pixel system for the project makes it unachievable on the desired timescale. The simple, low-mass choice is silicon microstrips.

Cut Dimensions (L×W)	100 mm × 40.34mm
Active Area (L×W)	98.33 mm × 38.34mm
Readout (Sense) Pitch	60□ m (30□ m)
# Readout (Sense) Strips	639 (1277)
Depletion Voltage	40V < V_{dep} < 300V
Breakdown Voltage	>350V
Total Detector Current at 350V bias	<16 □ A
Bias Resistor Value (both ends of strips)	0.8 ± 0.3 M□
AC Coupling Capacitance	>12 pF/cm
Total Interstrip Capacitance	<1.2 pF/cm
Defective Channels	<1%

Table 4.3.2.1: Specifications of sensors available for this experiment. Many of the sensors exceed the bias voltage and bad channel specifications by a significant margin.

The sensors must to be rectangular and as large as possible to cover a relatively large area with the smallest possible number of sensors. These sensors must also be of a radiation-tolerant design, have a relatively fine readout pitch to minimize occupancies and two-hit resolution and have a very small rate of defects. A sufficient supply of suitable sensors has been identified that meets all of these criteria, remaining from the cancelled Run IIb upgrades of the DØ and CDF detectors at the Tevatron [1]. These are p+ on n, single sided, AC coupled, polysilicon-biased sensors fabricated on <100> silicon. The relevant specifications of these sensors is shown in Table 4.3.2.1 While the specifications only ensure that these sensors may be operated up to 350V bias after irradiation, previous experience indicates that many of them will be operable to 1000V and therefore remain fully depleted to a dose of approximately 1.5×10^{14} 1MeV neq/cm^2. By testing and selecting those that withstand the highest bias voltage for use in the parts of the detector where the radiation field is most intense, the radiation tolerance of the detector can be maximized to allow a smaller dead zone and a longer running time before replacement of planes is necessary.

4.3.3 Readout Electronics

The extreme occupancies in the tracker demand the shortest possible readout integration time with the best possible time resolution. Since the development of a front-end readout chip is well beyond the scope of this experiment, existing front-end chips designed for the LHC tracking detectors are the obvious choices. Of these, the APV25, originally developed for use by the CMS

tracker, is the most attractive option [2]. The relevant specifications of the chip are summarized in Table 4.3.3.1 and the layout of the chip is shown in Figure 4.3.3.1.

# Readout Channels	128
Input Pitch	44 □ m
Shaping Time	50ns nominal (35ns min.)
Output Format	multiplexed analog
Noise Performance (multi-peak mode)	270+36×C(pF) e- ENC
Power Consumption	345 mW
Communication Protocol	I^2C

Table 4.3.3.1: Specifications of APV25 Readout ASIC.

The first desirable attribute is excellent noise performance, largely due to the inherent properties of the 0.25μm fabrication process. High signal-to-noise ratio in the detector results in high single-hit efficiency with a low rate of noise hits and extremely good single-hit spatial resolution. Before irradiation, the signal-to-noise ratio with the previously described sensors will be approximately 34, resulting in full single hit efficiency with a negligible rate of noise hits. Single-hit resolution will be approximately 6.5 μm based upon test beam and simulation with similar sensors [3]. After irradiation, when noise increases and charge becomes lost to trapping centers and a shrinking depletion layer, the extra headroom in signal-to-noise will extend acceptable performance as long as possible and provide a period in which to identify the onset of failure well before replacement is necessary.

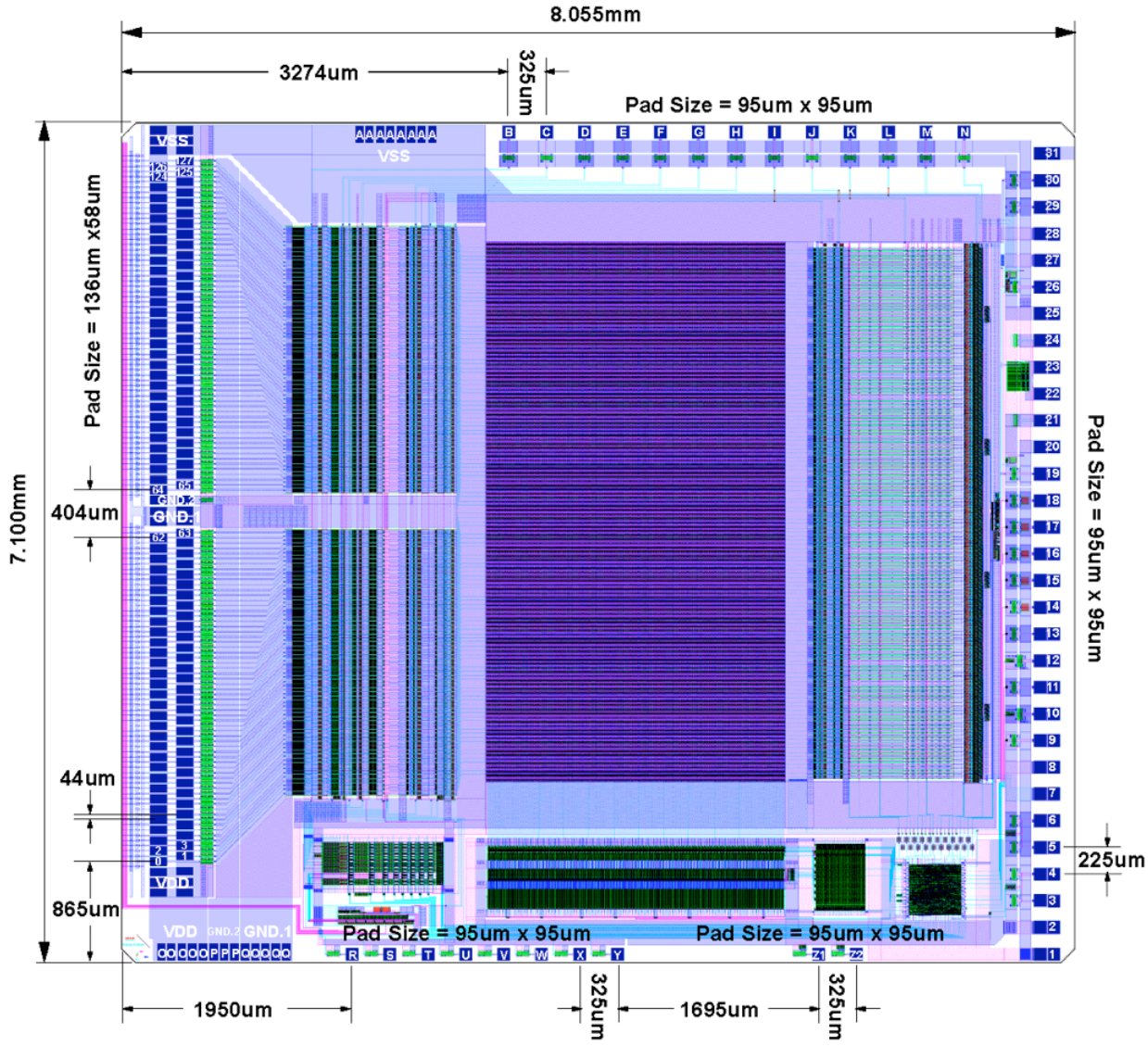

Figure 4.3.3.1: Layout of the APV25 Readout ASIC. The signal input pads are shown at left.

The second desirable attribute naturally arising from the 0.25 micron fabrication process is a high degree of radiation tolerance. Although the readout chips are expected to see much smaller doses than the hottest regions on the sensors, some chips will see doses that would result in performance degradation of chips that are less radiation tolerant.

Finally, and most importantly here, the APV25 has a flexible readout architecture, which can be used to improve the time resolution of the tracker. In addition to the standard readout modes employed at the LHC, the APV25 may be operated in "multi-peak" mode, in which the output of the shaper (with nominal shaping time of 50ns) is sampled at clock-synchronous 25ns intervals, in multiples of three samples, after a trigger is received [4]. Fitting the known form of the shaper output curve to these samples allows the determination of both the peak amplitude and the t0 of the hit. This technique has been developed extensively for the planned use of the APV25 chip for the upgrade of the Belle detector (Belle-II) for operation at Super-KEKB, which has a 2ns

bunch spacing. Tests have shown that at S/N>25, the hit time may be reconstructed with an RMS of 2ns or better for the readout of three samples [5].

Where occupancies are large relative to the decay of the shaper response, one must worry about overlap of hits in time. While this problem has been studied for Belle-II, the effect is small enough that it is sufficient to detect and eliminate such cases and accept a small loss in single-hit efficiency [6]. However, in this experiment, the rate of time overlaps will be larger, and two steps will be taken to mitigate the problem. First, we plan to read out six samples per hit, which will allow the unambiguous deconvolution of pairs of time-overlapping hits. Second, the shaping time of the APV25 can be reduced to 35ns to reduce overlaps once it can be established that this does not compromise the reconstruction of t0 in any way.

The flexibility and availability of the APV25 have made it popular with a number of experiments, which has further advantages. Not only are chips procurable in quantity at reasonable cost, but a reference test stand is widely available to bootstrap the development of DAQ [7]. One of these test stands is being assembled at SLAC for this experiment.

The APV25 chips will be hosted in groups of five on hybrid circuit boards, often simply referred to as "hybrids." In addition to the APV25 chips themselves, the hybrids host a number of passive components that define the power environment of the chips including the bypassing necessary to ensure the low-noise capabilities of the chips are not compromised. In order to reduce the amount of effort in developing them, the design will use existing APV25 hybrid designs as a starting point. In particular, the hybrids for the CMS Tracker Inner Barrel (TIB) were designed with similar constraints in mind and have a very compact format [8]. Mechanically, the hybrids for this experiment will have similar dimensions to the CMS TIB hybrids, with the exception that the TIB hybrids are wider due to the need to support six chips instead of five.

Unlike most trackers, the readout electronics for this experiment will not reside within the tracking volume, so the material budget for the hybrids and the cooling they require is less constrained. It is therefore anticipated that the hybrids will be fabricated in standard FR4 rather than alumina ceramics, which broadens vendor selection, reduces cost and saves turnaround time. In addition to reducing schedule risk, this will allow a prototype run to ensure that the hybrid design is sufficient without incurring significant additional costs.

4.3.4 Detector Layout

There are six measurement stations, or "layers", placed immediately downstream of the target. Each layer is comprised of a pair of closely-space planes and each plane is responsible for measuring a single coordinate, or "view". The details of the six layers are shown in Table 4.3.4.1 and a conceptual rendering is shown in Figure 4.3.4.1. Altogether, this layout comprises 106 sensors and hybrids and 530 APV25 chips for a total of 67840 readout channels. The total power consumption of the system is 186 Watts.

	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
z position, from target (cm)	10	20	30	50	70	90
Stereo Angle	90 deg.	90 deg.	90 deg.	50 mrad	50 mrad	50 mrad
Bend Plane Resolution (m)	≈ 6.5	≈ 6.5	≈ 6.5	≈ 6.5	≈ 6.5	≈ 6.5
Stereo Resolution (m)	≈ 6.5	≈ 6.5	≈ 6.5	≈ 130	≈ 130	≈ 130
# Bend Plane Sensors	4	4	6	10	14	18
# Stereo Sensors	2	2	4	10	14	18
Dead Zone (mm)	±1.5	±3.0	±4.5	±7.5	±10.5	±13.5
Power Consumption (W)	10.5	10.5	17.5	35	49	63

Table 4.3.4.1: Key parameters of the sensor layout for the tracking and vertexing system. Notice that the excellent single-hit resolution in the measured view enables good stereo resolution at the relatively small stereo angles necessary for pattern recognition. The stay clear region for the primary beam creates a "dead zone" in the tracker on either side of y =0.

The requirements of the tracking system are discernibly reflected in the layout. The 90-degree stereo of the first three layers provides a pair of high-resolution 3-d space points for vertexing. The small-angle stereo of the last three layers minimize ghost hits to improve pattern recognition while still providing sufficient pointing resolution into Layer 3 for robust hit association in the dense environment there. Meanwhile, the presence of a bend-plane measurement in every single layer assures the best possible momentum resolution. As discussed in Section 5.3 the pattern recognition of this layout is robust and the vertexing and momentum resolution, though limited by multiple scattering, are sufficient to provide the necessary sensitivity.

Because the area needed to cover a given solid angle varies with depth in the tracker, the sensitive area varies by layer. As can be seen in Figure 4.3.4.2, it would possible to achieve almost full acceptance for low-mass A′ decays outside of the dead zone in the first four layers in a 1T magnetic field, whereas some particles escape detection in the last two layers. Because the extremely pure tracking and vertexing required here necessitates at least five hits, the acceptance of the fifth layer is the key quantity in defining the overall acceptance of the tracker. However, achieving the best purity requires the extra constraint of a hit in Layer 6, as discussed in Section 5.3.

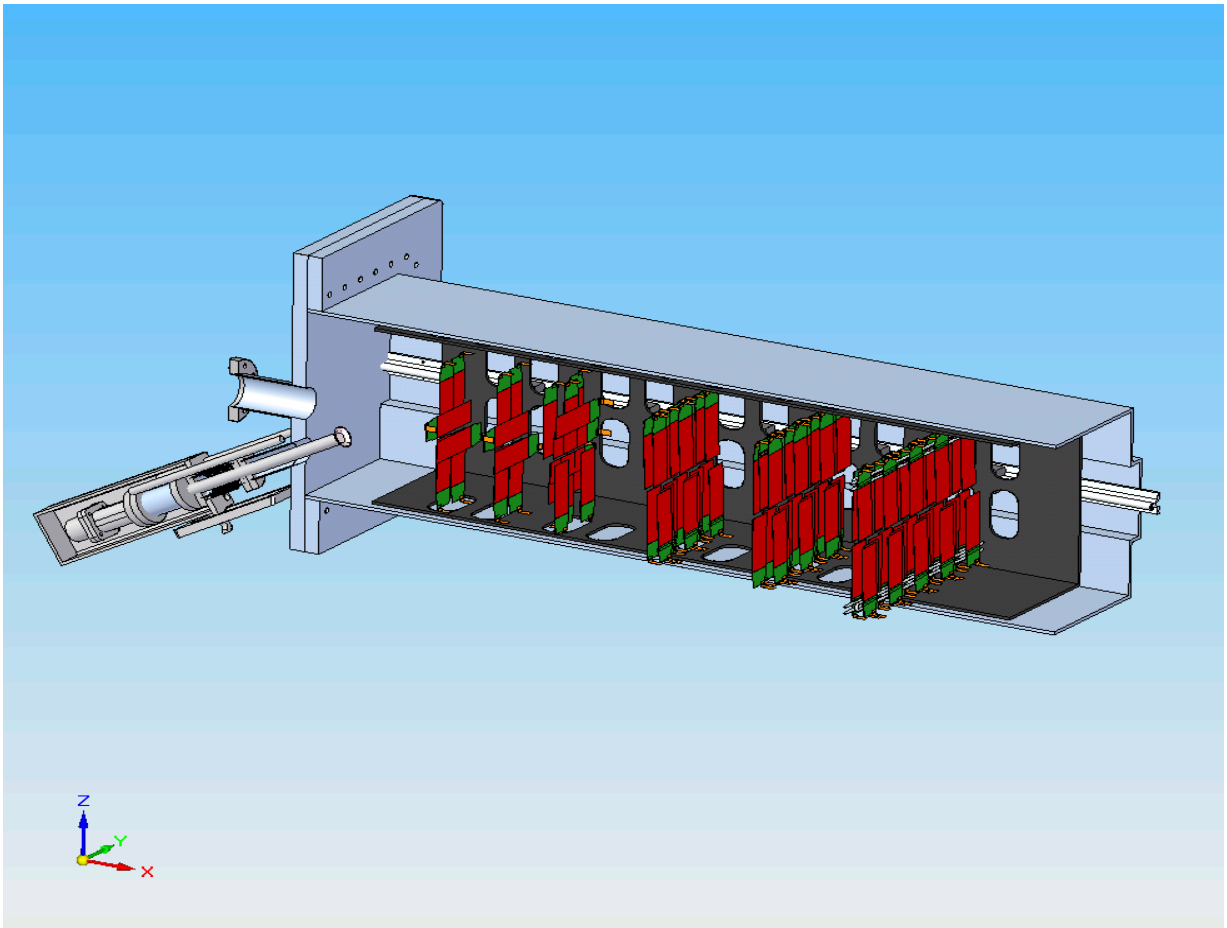

Figure 4.3.4.1: A rendering of the tracker showing the layout of the silicon planes inside a cutaway of the carbon fiber support box and the vacuum chamber. The target assembly is just upstream of the first silicon layer.

The solid angle subtended by the dead zone limits the acceptance for low mass A' decays, which have very small opening angles between decay daughters. For this reason, a great deal of attention has been paid to the optimization of this region of the detector, where careful simulation of the backgrounds has been used to determine the occupancies and radiation doses that limit coverage there.

In particular, three effects have been considered in setting the size of the dead zone. The first is the potential for an accident with the primary beam that would acutely damage or destroy a sensor. The second is the radiation dose from scattered primary beam and radiative secondaries that will render the sensors inoperable over time. The last is the density of hits associated with this radiation field that results in an unacceptable rate of pattern recognition failures. All of these issues are most severe in the first layer, which thereby determines the angular acceptance of the entire detector for prompt decays.

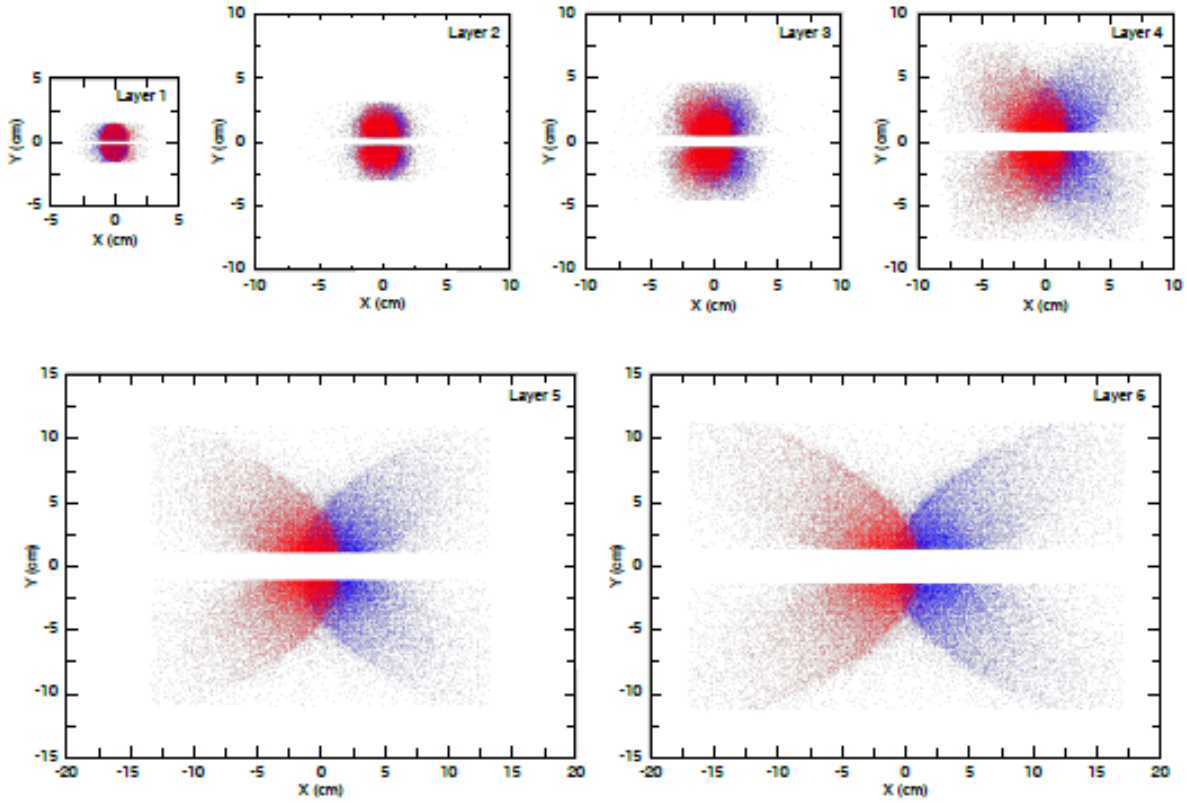

Figure 4.3.4.2: Hits in the silicon tracker in all six layers produced by A′ decays with $m_{A'}$ = 300 MeV/c^2 and E_{beam} = 6.6 GeV. All tracks that hit the first five layers are shown, resulting in a five-layer A′ acceptance of 44% in the 1T magnetic field. Over 90% of tracks within the five-layer acceptance also produce a hit in Layer 6.

The potential for acute damage from the tails of the unscattered primary beam or a steering accident is the least certain but potentially most severe liability of a small dead zone. If a 10 μm beam spot at the target is achievable, these tails should not be a limiting factor in any layer. However, the ability to retract the modules during beam adjustments and change their positions in response to changes in optics will be critically important to ensuring the safety of the detector.

The radiation field at Layer 1 provides the clearest constraint on the size of the dead zone. Under 1000V bias, the sensors will withstand more than 1×10^{14} 1MeV neutrons before the silicon can no longer be fully depleted [9]. Since the bulk radiation damage caused by electrons with energies less than 10 GeV is approximately a factor of 30 less than 1 MeV neutrons, this corresponds to more than 3×10^{15} electrons [10]. If we demand that the silicon remain fully depleted for three months of continuous running at 400 nA with a 0.25% X_0 target, the active silicon may not come closer than 1.5mm to the center of the dead zone, as illustrated in Figure 4.3.4.3. The depletion layer will shrink by a factor of two only after running a factor of two longer, at which time the signal loss will begin to significantly compromise time tagging. It is important to note that the rapid spatial variation of the dose in Layer 1 results in much more intense irradiation of the guard structure. Because little is known about this specific scenario, the ability to replace at least the first layer with minimal intervention is a mandatory element of the design.

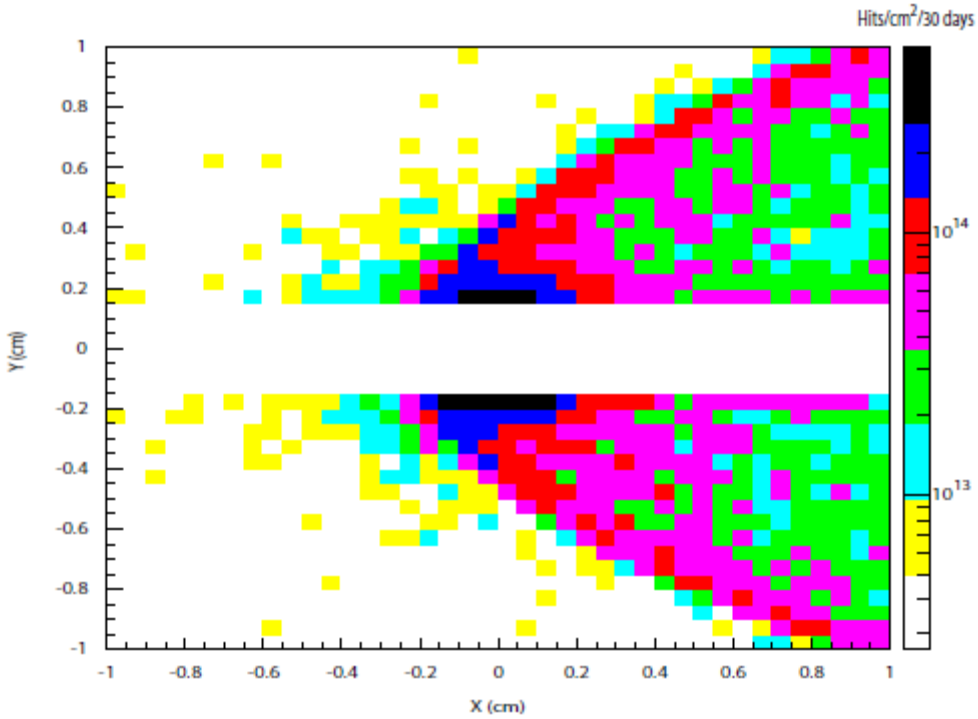

Figure 4.3.4.3: Background hits in the central portion of Layer 1 for one month of 400 nA beam. The active silicon begins at y =±1.5mm.

Limitations on the size of the dead zone due to high occupancies are less clear, and arise from two distinct considerations. First, for robust pattern recognition, it is desirable to limit the hit occupancy that must be considered during pattern recognition to less than 1%. Given hits with 2ns time resolution, hits can be assigned to an 8ns window without significant loss of efficiency. Limiting the occupancy in this time window to 1% places the edge of the dead zone in the first layer at y =±1.5mm, as shown in Figure 4.3.4.4. It should be noted that a more advanced algorithm that fits hit times and positions simultaneously will be developed and should outperform this simple selection. Second, for pristine reconstruction of hit times and amplitudes, it is important to minimize triple coincidences within a time window corresponding to the time evolution of the shaper output; approximately 250 ns for the nominal 50ns shaping time. For the 1% hit occupancy considered above, the rate of triple coincidences in individual strips within the same 250 ns window is 11%. While this is probably acceptable in a small number of strips with the highest occupancy, reducing the shaping time to 35ns will suppress this rate by a factor of two and is therefore highly desirable.

Because these considerations lead to similar conclusions, the nominal dead zone of the tracker has been designated as θ_y < ±15 mrad. It should be noted that occupancies and radiation intensities in the layers with small-angle stereo, where the sensors must be rotated, are somewhat lower and thus will allow complete stereo coverage down to ±15 mrad in all layers.

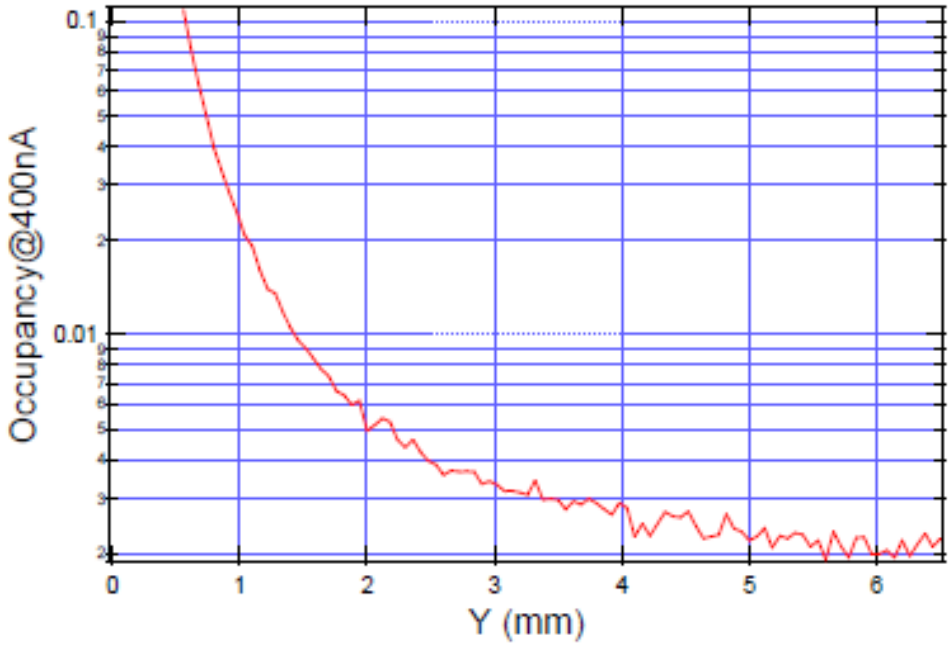

Figure 4.3.4.4: The hit occupancy per strip for 60 μm strips in Layer 1 for 7.5 ns of beam at 400nA intensity.

4.3.5 Sensor Modules and Mechanical Support

Each of the ten planes that measures a single view is split into two modules to accommodate the horizontal dead zone that cuts the entire tracker in half. These modules are the fundamental mechanical units of the detector: the components of a module are permanently bonded together during assembly. Each consists of a number of silicon sensors and their hybrid circuit boards glued to a composite support structure.

Adjacent silicon sensors are placed on alternating sides of the structure to provide overlap that ensures full sensitivity over the surface of the module for incident particles. In the areas between sensors on either side of the structure, holes are cut in the structures making them approximately 50% void in order to minimize the amount of material the modules present to the passage of particles. At the end of each sensor is the hybrid circuit board containing the five APV25 readout ASICs necessary to read out the signals from a single sensor. The pitch of the APV25 and the sensors are similar enough that a pitch adapter can be avoided here: wire bonds can be made directly from the chips to the sensors.

A set of modules of both types for Layer 6 are shown in Figure 4.3.5.1. Although there are ten unique module types in the system, there are only two arrangements of sensors: those that are square with each other and the underlying module support structure and those that are rotated by 50 mrad. As a result, module assembly requires at most two unique sets of tooling. The key mechanical attributes and material budget for the sensor modules are shown in Table 4.3.5.1.

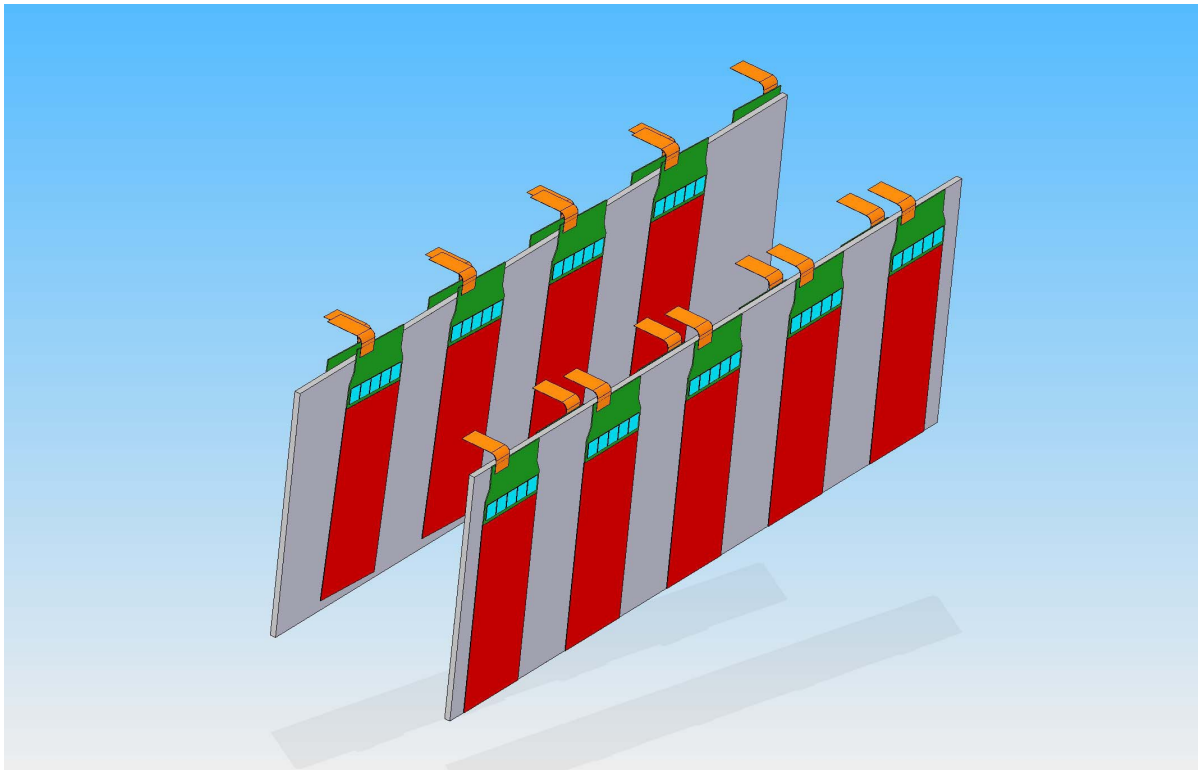

Figure 4.3.5.1: A pair of modules comprising the top half of Layer 6, a 50 mrad stereo layer. Each hybrid has a short pigtail cable attached to a solder array. These pigtails connect to longer cables dressed to the carbon fiber support box.

Each module support structure is comprised of a pair of thin, high-modulus, carbon fiber composite skins sandwiched around a low-mass rohacell core. At the edges of the support structure, running underneath the readout chips, are integrated stainless-steel cooling tubes carrying water-glycol coolant at approximately -5°C to remove the 1.7W heat load generated by each hybrid. The cooling tubes are in contact with the skins on both module faces to cool the silicon to below 0°C, reducing the rate of radiation damage and lowering bias currents to extend the useful life of the sensors. The unidirectional carbon fiber of the skins are oriented to efficiently transport heat from the silicon to the cooling tubes. This arrangement provides sufficient cooling for the silicon without placing cooling tubes inside the tracking volume that would increase multiple coulomb scattering and compromise sensitivity.

	Radiation Length (mm)	Thickness (mm)	Coverage/Unit Acceptance	Scattering Material (% X_0)
Silicon	93.6	0.320	1.2	0.410
Rohacell Foam	13800	3.0	0.5	0.011
Carbon Fiber	242	0.150	0.5	0.031
PGS Passivation	256	0.101	1.25	0.049
Epoxy	290	0.050	0.5	0.009
Total	-	-	-	0.510

Table 4.3.5.1: The material budget for the sensor modules showing the thickness, fraction covered for unit of areal acceptance and average scattering material presented to incident particles. The only full coverage glue gap bonds the carbon fiber to the Rohacell. The silicon will be attached to the passivation with epoxy dots of negligible coverage.

Atop the carbon fiber, there is a self-adhesive layer of pyrolytic graphite sheet that improves thermal conductivity and includes a layer of PEEK passivation rated at 2kV to isolate the back side of the sensors, which are at high voltage, from the conductive graphite [11]. The layers of carbon fiber and graphite will have a low-impedance connection to the analog ground of the readout electronics to ensure the best possible noise performance.

Each pair of modules that make a stereo pair covering half of the acceptance is coupled mechanically around their periphery to ensure that both move together in unison. Along the left and right edges of each pair, they are kinematically mounted to a pair of piezo motors, similar to those used for beam instrumentation. These can be operated remotely to tune the plane positions, optimizing the acceptance for the actual background environment encountered once the experiment is on beam.

The piezo motors are in turn mounted to attachment points in a carbon fiber-honeycomb composite support box that fits inside of the vacuum enclosure, as shown in Figure 4.3.5.2. This support box is the mechanical element that ties all the modules together into a complete tracking system. The cables and cooling lines that attach to the pigtail cables and cooling connections of each sensor module will be semi-permanently installed on the tracker support box to minimize the difficulty of installing and removing sensor modules.

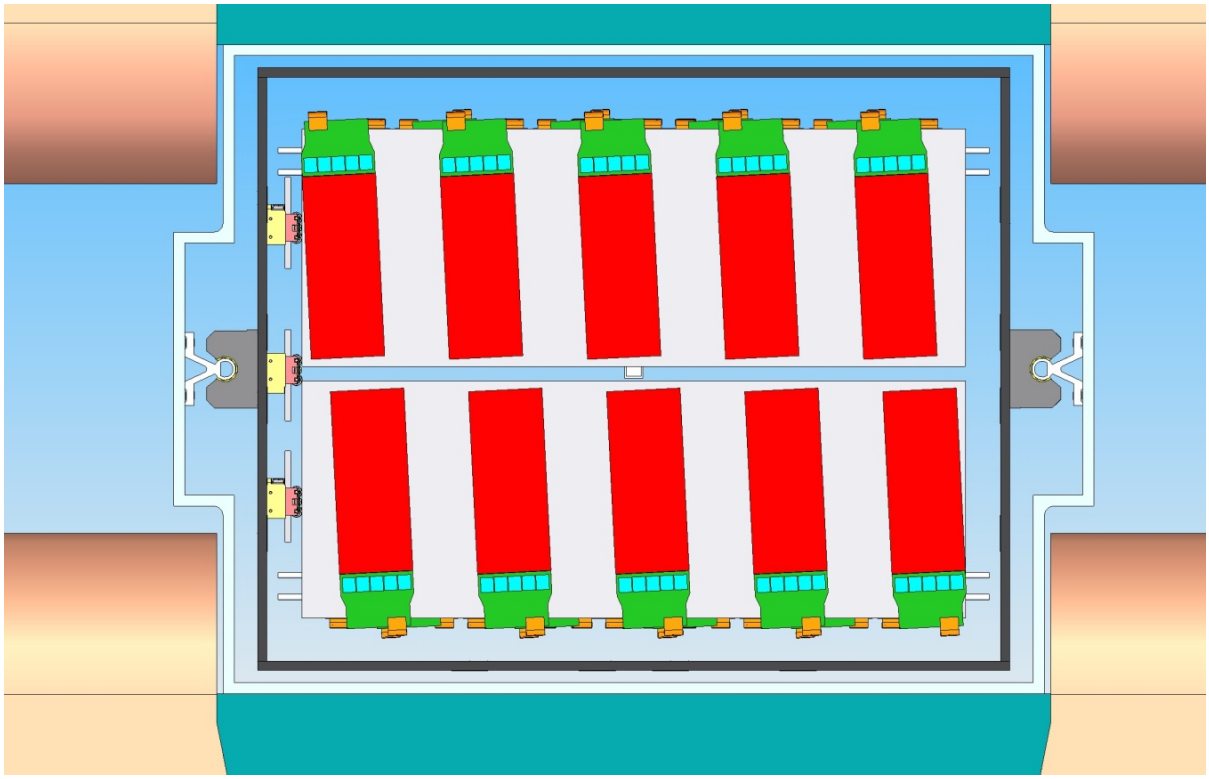

Figure 4.3.5.2: Looking upstream at Layer 6 in the tracker. The piezo motors for moving the half layers are shown at the left, just inside the carbon fiber support box, shown as the dark rectangle. The rail system allows the support box to be extracted from the vacuum chamber to access the silicon detectors and electronics.

4.3.6 **Vacuum Chamber**

The vacuum vessel will be fabricated in non-magnetic stainless-steel, which offers better strength than aluminum at smaller wall thickness. The carbon-fiber support box mounts inside the vacuum chamber on a pair of precision rails installed along the side walls of the chamber, as shown in Figure 4.3.6.1. Carriages attached to the support box glide on the rails to allow insertion and extraction of the tracker from the vacuum chamber.

To make space for this rail system without compromising acceptance, the walls of the chamber incorporate U-channels that extend into the magnet alcoves. These U-channels also stiffen the side walls against deflection when placing the chamber under vacuum, although further study is required to ensure that they are sufficiently braced so that motions induced by placing the chamber under vacuum do not damage the tracker or create alignment instabilities.

In addition to minimizing backgrounds, the vacuum acts to isolate the detector from the environment, which is important since conduction and convection would otherwise dominate the heat load and cooling requirements of the sensor modules.

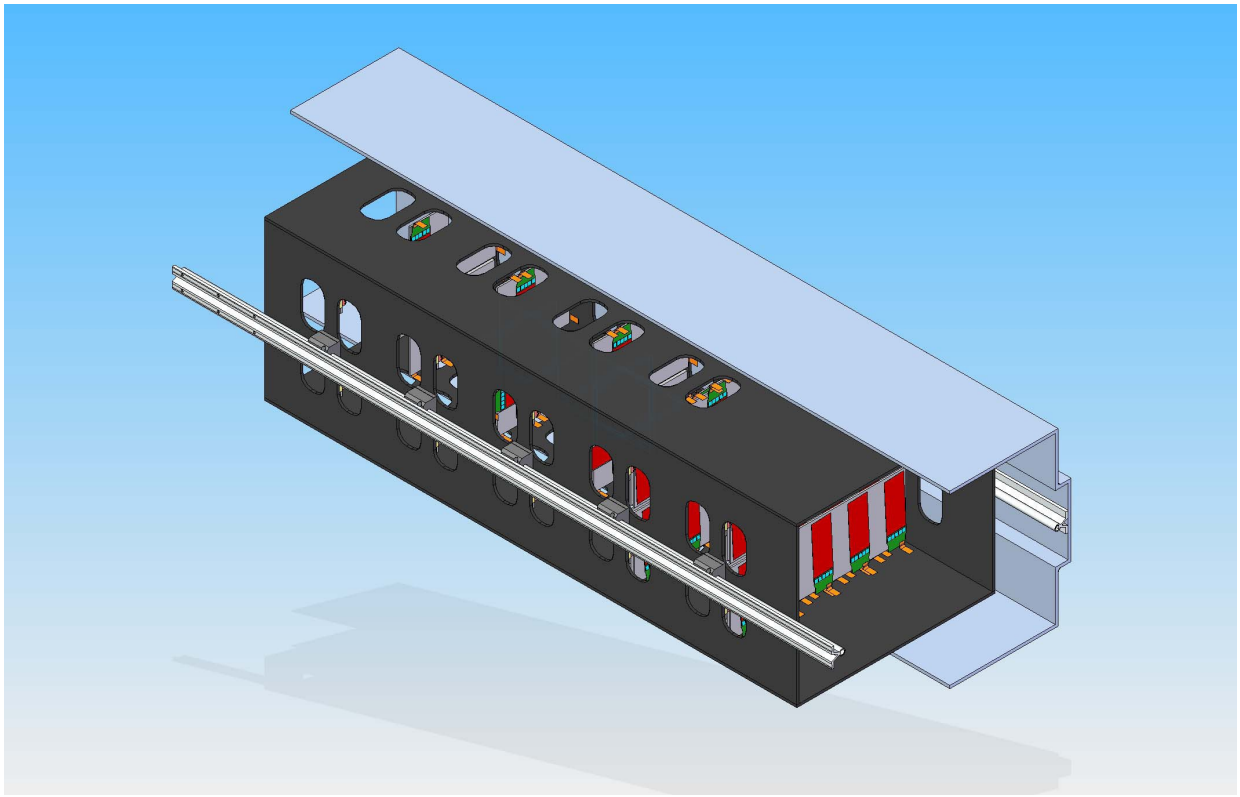

Figure 4.3.6.1: A rendering of the tracker highlighting the rail system that supports the tracker from the walls of the vacuum chamber. This allows the carbon fiber support box to be moved along the beamline with respect to the target.

In order to effectively eliminate these loads, a vacuum of 10^{-4} Torr or better must be maintained, so materials to be used inside the vessel will be chosen to minimize outgassing. In order to achieve acceptable cooling performance for the sensors given the minimal material budget, it will also be necessary to reduce the radiant heat load on the modules from the walls of the vacuum chamber. To minimize aging of the silicon and the possibility of thermal runaway, a shielding blanket of multi-layer insulation (MLI) will be wrapped around the tracker support box. This radiant barrier will reduce the heat load and resulting temperature drops by at least an order of magnitude [12, 13].

The cables and cooling services for the tracker are routed to the upstream end of the vacuum chamber where a patch panel will be located. The design of this panel and the cooling and cable connections are of critical importance due to space constraints and the necessity of maintaining good vacuum. Further, since we will not be able to operate a sub-atmospheric cooling system inside the chamber, extra care will be required for all cooling connections and joints.

The rollers attaching the carbon fiber support box to the vacuum chamber will allow the entire tracker system to be rolled into and out of the magnet with relative ease. This arrangement will enable the tracker to be installed or serviced in a relatively short period of time with only removal of a relatively short spool piece in the upstream beamline. Figure 4.3.6.2 shows the tracker installed within the analyzing magnet.

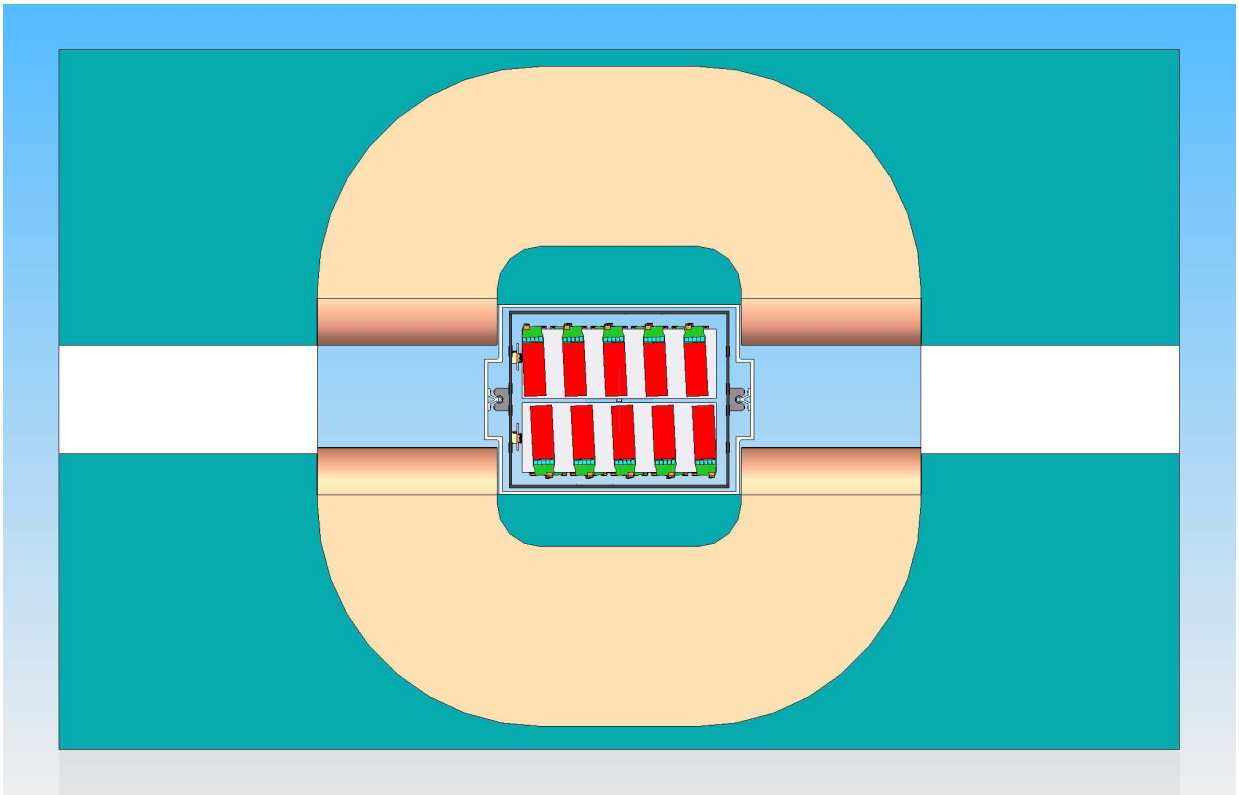

Figure 4.3.6.2: A rendering of the tracker installed within the analyzing magnet. There is significant additional space within the magnet alcoves to accommodate reinforcements to the walls of the vacuum chamber or other necessary services.

4.3.7 References

1. DØ Collaboration, *DØ Run IIb Upgrade Technical Design Report*, http://d0server1.fnal.gov/projects/run2b/Silicon/www/smt2b/Documentation/D0_Run2b_TDR_Silicon_Sept24_02.pdf

2. M.J. French et al., Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker, NIM A 466 (2001) 359–365.

3. P. Azzi et al., *Charge Deposition Model in Silicon*, CDF Internal Note CDF/DOC/TRACKING/CDFR/5080 (1999).

4. M. Friedl et al., Performance of a CMS silicon strip detector module with APV25 readout, NIM A 488 (2002) 175–183

5. M. Friedl et al., Performance of a CMS silicon strip detector module with APV25 readout, NIM A 488 (2002) 175–183

6. Y. Nakahama, Research and Development of a Pipeline Readout System for the Belle Silicon Vertex Detector, University of Tokyo Masters Thesis, (2006)

7. RWTH Aachen CMS Group, *The ARC System*, http://www.physik.rwth-aachen.de/institute/institut-iiib/forschung/cms/detektorentwicklung/arc/

8. U. Goerlach, *Industrial Production of Front-End Hybrids for the CMS Silicon Tracker*, 9th Workshop on Electronics for LHC Experiments, Amsterdam (2003)175-179

9. G. Lindstrom, Radiation damage in silicon detectors,NIM A 512 (2003) 30–43

10. I. Rachevskaiaa et. al, Radiation damage of silicon structures with electrons of 900 MeV, NIM A 485 (2002) 126–132

11. Panasonic Industrial Products Catalog, http://industrial.panasonic.com/www-data/pdf/AYA0000/AYA0000CE2.pdf

12. S. D. Augustynowicz, J. E. Fesmire, and J. P. Wikstrom, *Cryogenic Insulation Systems*, http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19990053342_1999081754.pdf

13. . D. Augustynowicz and J. E. Fesmire, *Cryogenic Insulation Systems for Soft Vacuum*, http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19990062736_1999085329.pdf

4.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECal) will provide the trigger for data acquisition and will be used for electron identification during the data analysis. The ECal will be positioned after the analyzing dipole magnet and will cover the full acceptance region of the silicon tracker. It will consist of two parts, beam-up and beam-down. Each part will cover ~1200 cm^2 area. The gap between upper and lower parts, ~15mrad as seen from the target location, is necessary to avoid the beam electrons, as well as those that have radiated in the target and the bremsstrahlung photons they have produced.

The energy of electrons of interest will be in the range 0.5-6.5 GeV. The ECal modules must have sufficient radiation lengths to absorb the full energy of these scattered electrons and should have fine enough granularity to handle a high rate of electromagnetic background, especially near the beam plane. In a compact detector with tight space constraints, a high radiation environment and the presence of high magnetic field, a lead-tungstate ($PbWO_4$) crystal calorimeter with magnetic field resistant photo detector readout (e.g. avalanche photodiodes (APD)) is the obvious choice. Based on simulations, the optimal cross section of $PbWO_4$ crystals for HPS detector is ~$2cm^2$, that totals to 1200 crystals for ECal[1]. Such calorimeter was built and has been operational in Hall B since 2005 (the CLAS inner calorimeter (IC) with total of 424 $PbWO_4$ crystals with APD readout. We are planning to use IC modules, which including exiting spares will total 460 channels for the ECal. To optimize the performance and cost of the detector we have chosen a hybrid calorimeter design which combines lead tungstate crystals read by

[1] Price for such a module, including the APD and a pre-amplifier, is about $800.

avalanche photodiodes and lead-glass modules with conventional PMT photodetectors. Since the area near the beam plane is under the highest radiation load, the modules near it must be radiation resistant and should have finer granularity. The lead tungstate crystals from the CLAS inner calorimeter fully meet the latter requirements. The outer three rows in each half will be covered by larger lead-glass modules as shown in Figure 4.4.1.

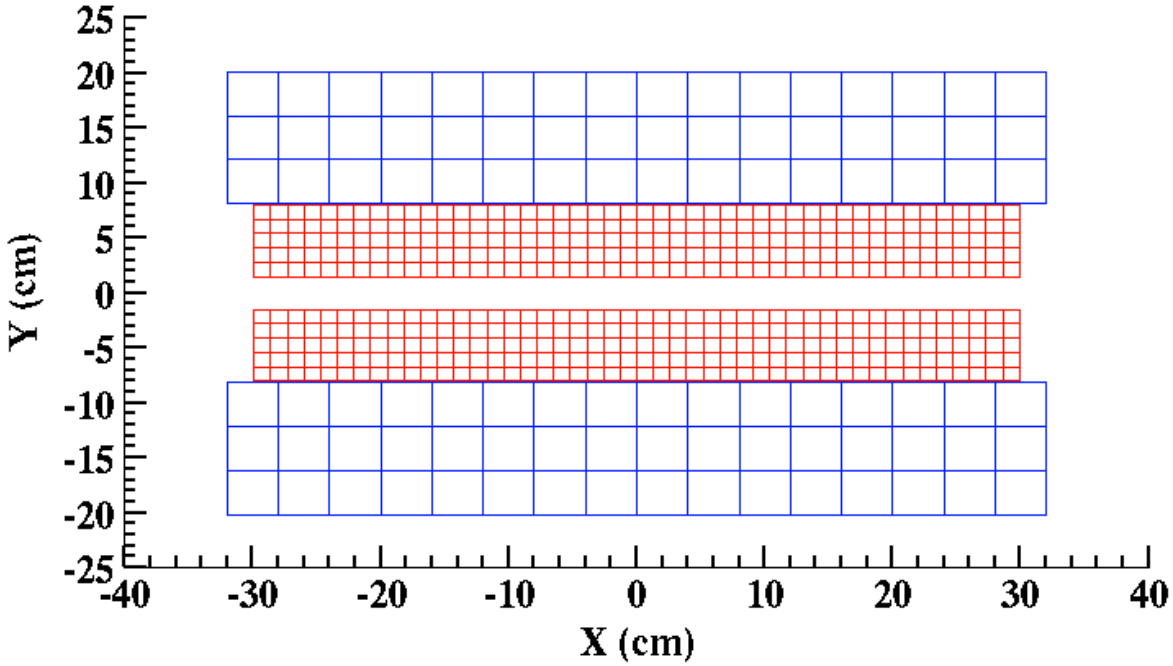

Figure 4.4.1. Beam's eye view of the electromagnetic calorimeter. Red squares represent lead-tungstate modules, blue squares are lead-glass or "shashlyk" type modules.

A rendering of the calorimeter system is shown in Figure 4.4.2. The calorimeter is separated from the beam vacuum by a 1 cm thick aluminum vacuum exit window (made transparent in the figure). The beam goes through a continuous vacuum from the vacuum box in the magnet to a vacuum enclosure made from 1 cm aluminum plates, which is inserted between the top and bottom calorimeters. A different configuration, where the beam exited through a thin vacuum window into a region filled with helium for the calorimeter, was rejected due to the increased background rates. To further reduce the background rates without diminishing the structural integrity, a small area of the aluminum plates is cut out where the most intense part of the electron beam exits to the beam dump.

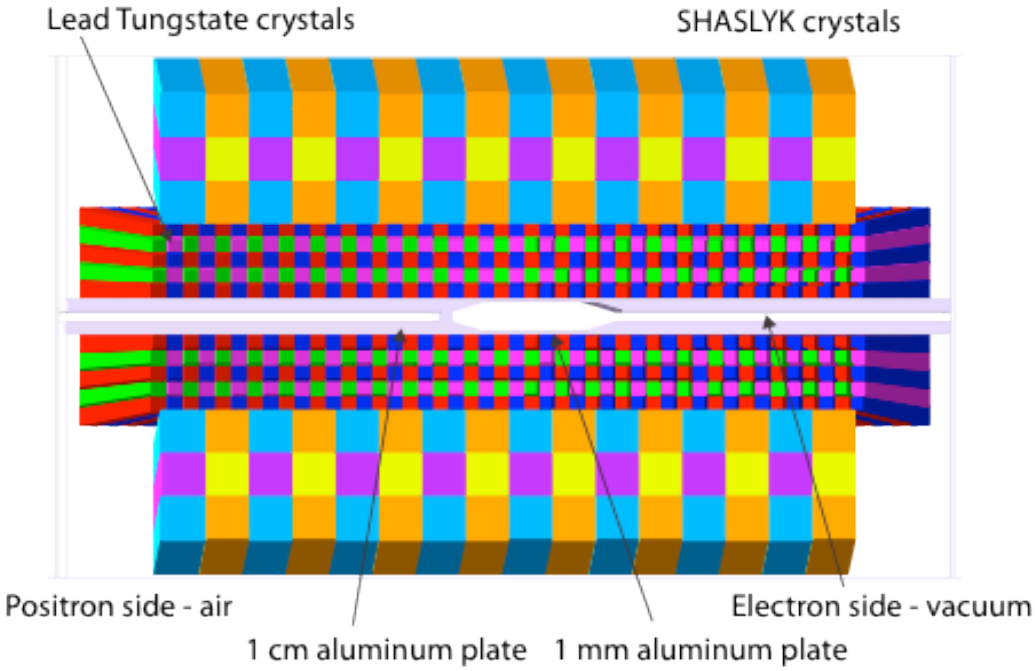

Figure 4.4.2. A rendering of the electromagnetic calorimeter setup looking down the beam line. The front exit window and side plates are rendered transparent to permit a view of the crystals and the vacuum plates.

4.4.1 The Inner Calorimeter

The inner part of the ECal will be built out of 460 lead-tungstate crystals, from the same crystals that were used in the IC. The IC calorimeter was assembled from 424 lead-tungstate tapered crystals with dimensions of 13.3 × 13.3 mm^2 (front face), 16 × 16 mm^2 (rear face) and 160 mm length. There are about 40 more spare crystals that can be used for the ECal. The crystals were fabricated with very tight tolerances: σ = 19 μm for 13.3 mm dimension and σ = 13 μm for the 16 mm dimension. Each crystal is wrapped in VM2000 multilayer polymer mirror film and has a plastic fiber holder glued on its front face for light monitoring.

Avalanche Photodiodes (APD) S8664-55 produced by Hamamatsu Corp. are used as photodetectors. They have a 5 × 5 mm^2 active area and a high quantum efficiency (about 75 %). Single APDs were centered on the back ends of the crystals using MeltMount 1.7 glue. The low-amplitude signals of APDs are amplified using custom-made preamplifier boards. A schematic view of the IC module assembly is presented in Figure 4.4.1.1.

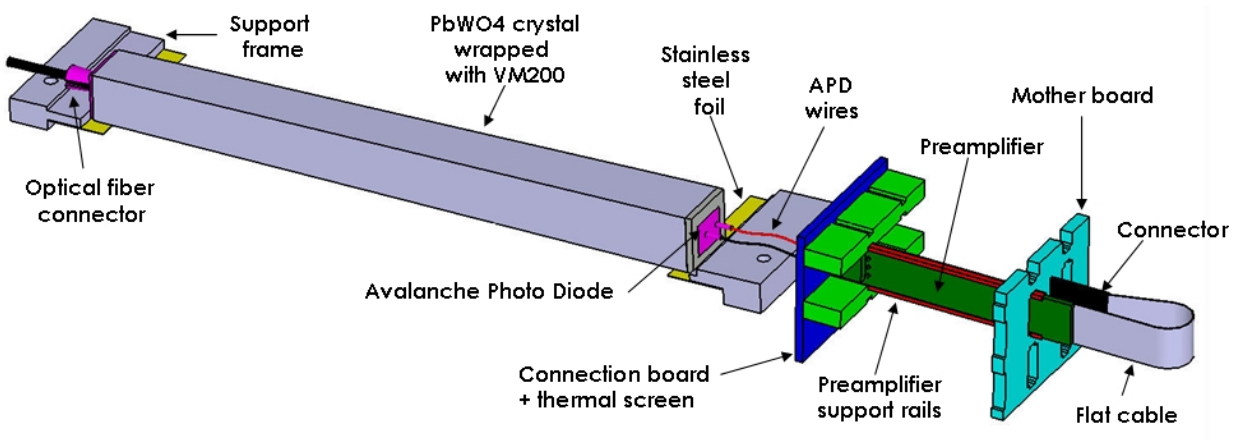

Figure 4.4.1.1. Schematic view of IC module assembly.

In order to maintain stable performance of the $PbWO_4$ calorimeter, the temperature in the ECal enclosure must be stabilized within 0.1 ^{0}C. The expected energy resolution of the system (from operational experience with the IC) is $\sigma_E/E \sim 4.5\%/\sqrt{E}$ (GeV).

4.4.2 The Outer Calorimeter

For the outer layers of the ECal we are planning to use either lead-glass modules (TF-1), available through the Yerevan collaboration. The size of the lead-glass module is 40x40x400 mm^3. A schematic view of the detector assembly with the above modules is presented in Figure 4.4.2.1. Energy resolution of these modules is $\sigma_E/E \sim 5.5\%/\sqrt{E}$ (GeV), which satisfies the experimental requirements. These detectors have been used at JLAB. A total of 96 modules will be needed to construct the outer part of the ECal. ECal Assembly.

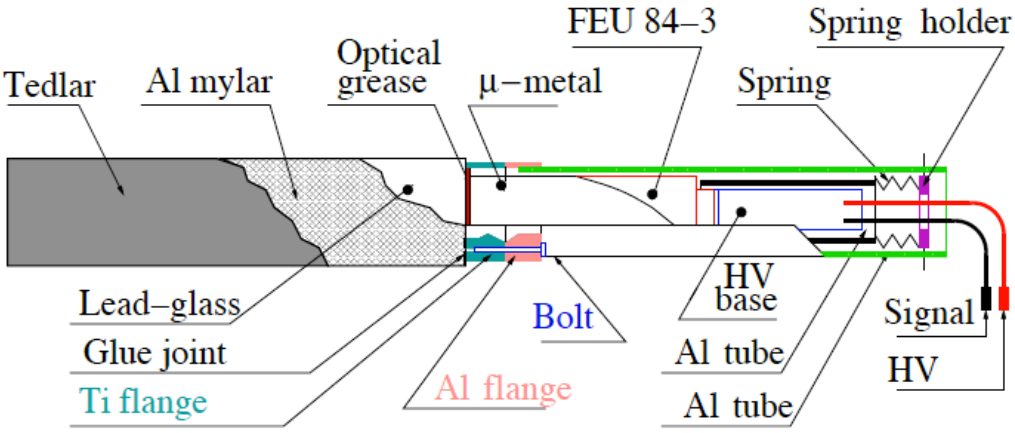

Figure 4.4.2.1. Lead-glass counter assembly used in previous JLAB experiments with the TF-1 module.

4.4.3 ECal Assembly

The ECal will be mounted behind the analyzing dipole magnet, after the vacuum chamber window, as shown in Figure 4.4.3.1. The distance from the front face of the calorimeter modules to the magnet center will be about 75cm. The ECal modules will be assembled inside a calorimeter box, which provides temperature stabilization. The upper and lower parts of the ECal will be separated by a vacuum chamber, which allows unimpeded beam transport to the beam dump and minimizes the material radiative electrons and photons will encounter. The calorimeter box will be sealed and will have thermal insulation to provide the required temperature stabilization (similar to the IC box assembly). As in the IC, $PbWO_4$ modules will be connected to a motherboard that will provide power to and transmit signals from individual APDs and pre-amplifier boards. PMT high voltages and signals for lead-glass modules will be transmitted through side walls of the ECal box assembly.

The CLAS IC box assembly, crystal support frames, pre-amplifier boards and the motherboard were designed and fabricated by the Orsay (France) collaboration in CLAS. That group is interested in participating in this experiment and will carry most of the load in the design and construction of the ECal.

4.4.4 Electronics

Existing low and high voltage systems for the IC will be used for the lead-tungstate part of the ECal. For the PMTs of the lead-glass modules, the CLAS PMT HV system can be used. Signals from each module will be sent to a signal splitter (similar to the IC readout), then one of outputs of the splitter will be fed to a discriminator and then to a TDC channel, while the second one will be sent to the FADC. The trigger from the ECal will be based on FADC information as described in Section 4.6.5.

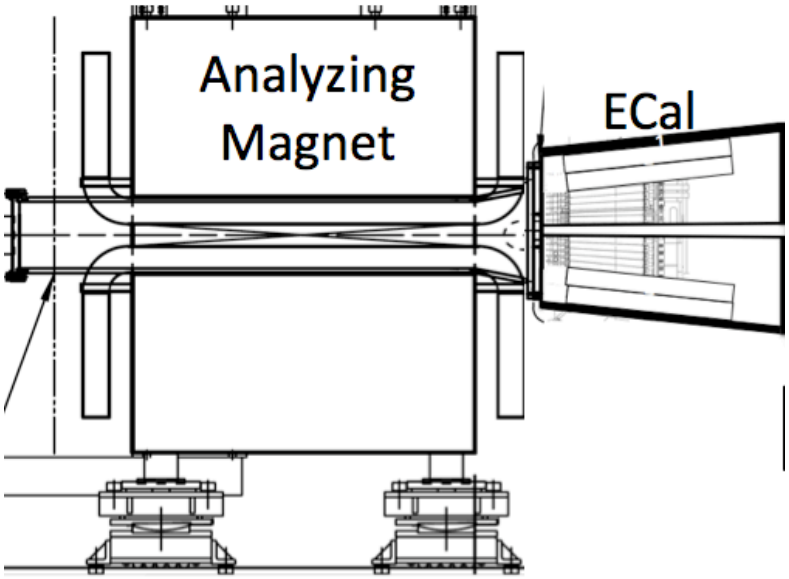

Figure 4.4.3.1. Ecal placement behind analyzing magnet

4.5 Muon System

Searching for the A' in its di-muon decay mode has the advantage of having greatly reduced electromagnetic backgrounds for triggering. The muon detection system will be installed behind the ECal which absorbs most of the electromagnetic background produced in the target. The remainder will be attenuated by the first absorber layer of the muon system. Remaining backgrounds will arise from photoproduction of π+ and π- pairs in the target, which aren't fully absorbed in the ECal or absorber. The muon detector will be about 1 meter long and will consist of layers of scintillator hodoscopes sandwiched between iron absorbers. The number of layers and the thickness of absorbers will determine the π/μ rejection factor.

4.5.1 Simulation of the Muon System

In order to optimize the setup, charged pions and muons were simulated from upstream of the ECal (lead-tungstate part only) in the momentum range from 1 GeV/c to 4 GeV/c using the GEANT-3 model for the CLAS Inner Calorimeter (IC). Five layers of iron absorbers interlaid with layers of 1 cm thick scintillator planes were positioned 1 meter from the front face of the IC, as shown in Figure 4.5.1.1. The figure shows a few of the simulated π+ events. Most pions will shower in the IC and in the iron absorbers and will not reach all the way to the last layers of the scintillation hodocopes, while most muons will leave mip traces in the scintillator and, depending on their momenta, will pass through most or all of the layers of the system.

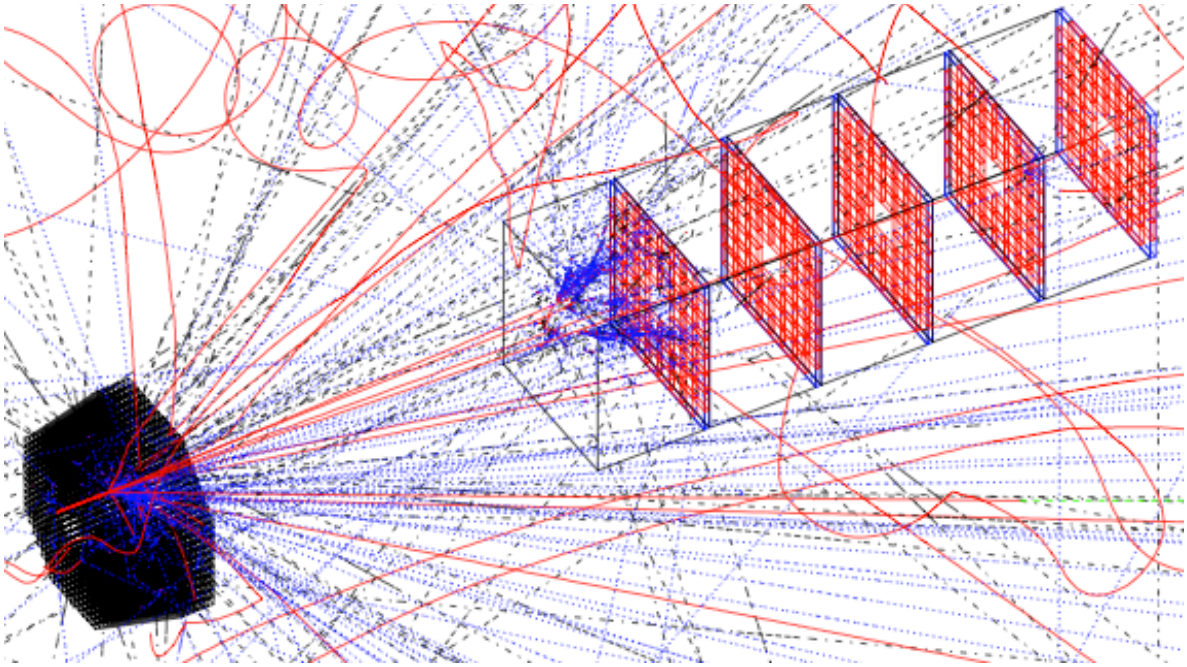

Figure 4.5.1.1. Simulated pion events in GEANT-3 model of IC and in five layers of iron absorbers and scintillators.

The distribution of the deposited energy for muons (top graphs) and pions (bottom graphs) in the 16 cm long lead-tungstate crystals (left graphs) and in 1 cm thick scintillator layers (right graphs) are shown in Figure 4.5.1.2. Detection efficiencies for pions and muons were defined as the ratio of the number of particles that passed through cuts on energy deposition in IC, $\Delta E > 100$ MeV, and in hodoscope layers, $\Delta E > 0.4$ MeV, to the number of simulated events.

In Figure 4.5.1.3, the dependence of pion (open squares) and muon (filled squares) detection efficiencies on the total thickness of the iron absorber is shown. The muon detection efficiency decreases slightly for up to a meter of the total absorber length, while the detection efficiency for pions shows exponential drop up to an absorber thickness of 75 cm.

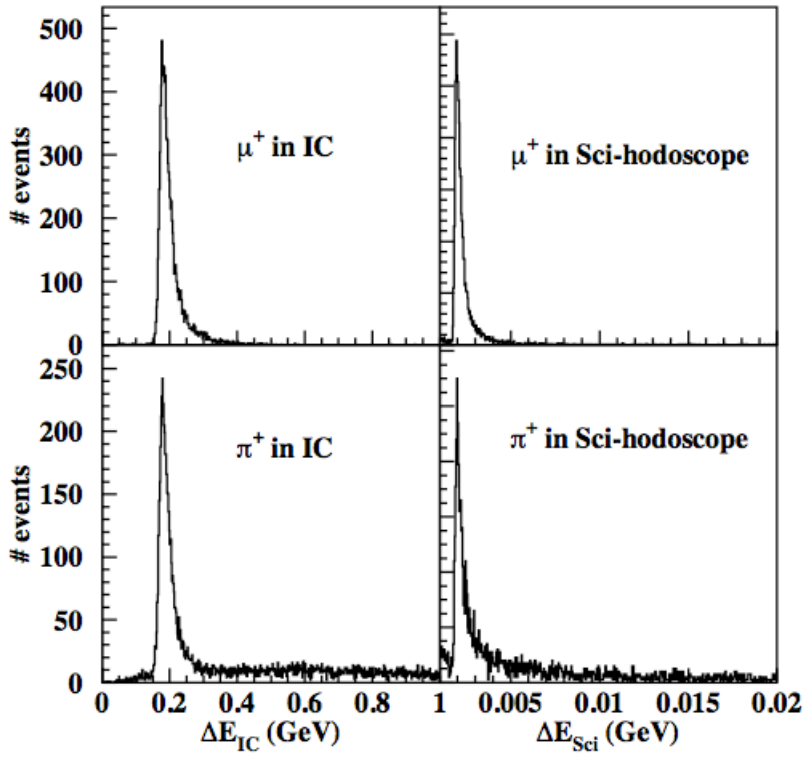

Figure 4..5.1.2. Energy deposition of muons (top graphs) and pions (bottom graphs) in IC (left) and in scintillation hodoscopes (right).

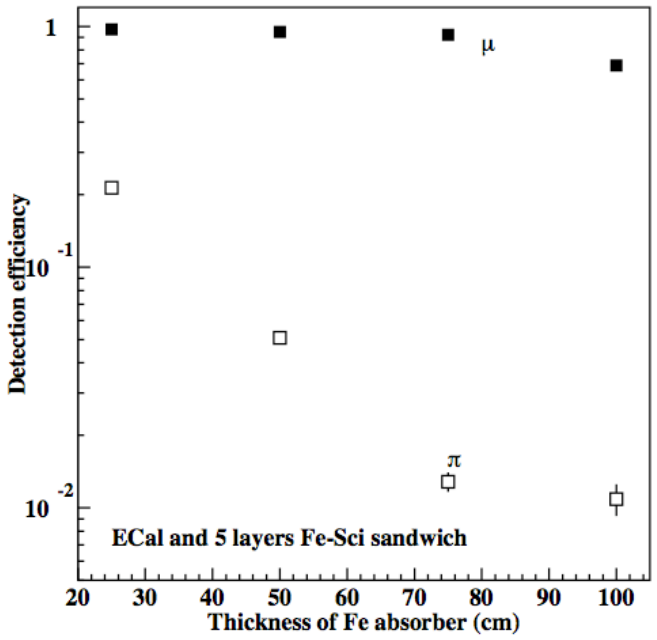

Figure 4.5.1.3. Efficiency of pion and muon detection as a function of total iron absorber thickness in five-layer configuration.

In Figure 4.5.1.4, the rejection factor for charged pions is presented. As can be seen from the figure, for particles from 1 GeV to 4 GeV, the optimal thickness of the iron absorber for pion rejection is ~75 cm (after 16 cm of lead-tungstate). In the efficiency calculation for low energy particles (< 1.7 GeV), not all the layers were considered. Depending on the momentum, particles were not traced behind the third, fourth or fifth absorber.

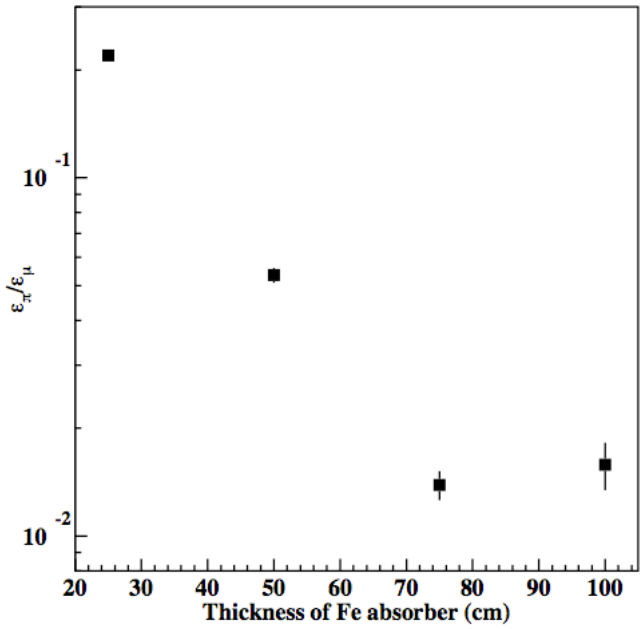

Figure 4.5.1.4 Pion-muon rejection factor as a function of the iron absorber thickness.

4.5.2 **Detector Concept**

Based on the above simulation, the muon detector is proposed as a system of four iron absorbers (total length of 75 cm) and four scintillator planes positioned after each absorber layer. The first absorber will be 30 cm long; the other three will have equal lengths of 15 cm. The muon detector will be mounted behind the ECal, see Figure 4.5.2.1. The pion rejection ($\varepsilon^{\pi}/\varepsilon^{\mu}$) of the proposed system as a function of particle momentum is shown in Figure 4.5.2.2. For the determination of particle detection efficiencies, only signals in the first three layers of the hodoscope were used for momenta < 1.7 GeV/c. In the full momentum range from 1 to 4 GeV/c, $\varepsilon^{\pi}/\varepsilon^{\mu}$ varies from 10^{-2} to $2\text{x}10^{-2}$ and therefore the pion pair suppression factor of the system will be < $4\text{x}10^{-4}$. Similar to the Ecal, the muon detector will consist of two halves, one above and one below the beam. The vertical gap between the first hodoscope layers of the two halves is about 3 cm. The cross section of each half of the first hodoscope is 63x39 cm^2. Each half of the last hodoscope layer is 74x46 cm^2.

The simplest, most economical solution for the hodoscopes is to use layers of extruded scintillator strips with embedded wave-shifting fiber readout. The scintillator strips will be oriented horizontally. The light will be detected from both ends of the strip using multi-anode

photomultipliers. With 5 cm vertical segmentation, a total of 144 readout channels are needed (9x16 channels multi-anode PMTs). Signals from each channel will be sent to a TDC and to a FADC. The FADC information will be used to construct the muon trigger. The TDC information together with information from FADC will be used in offline analysis to measure the hit position along the strip.

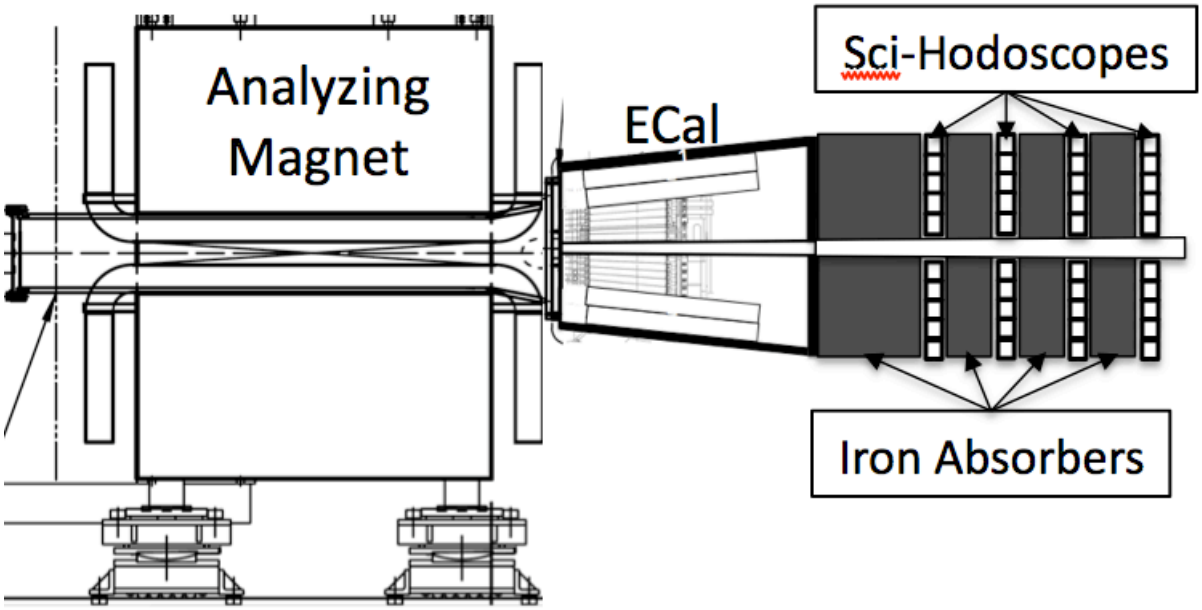

Figure 4.5.2.1. Schematic view of the HPS detector with muon detection system.

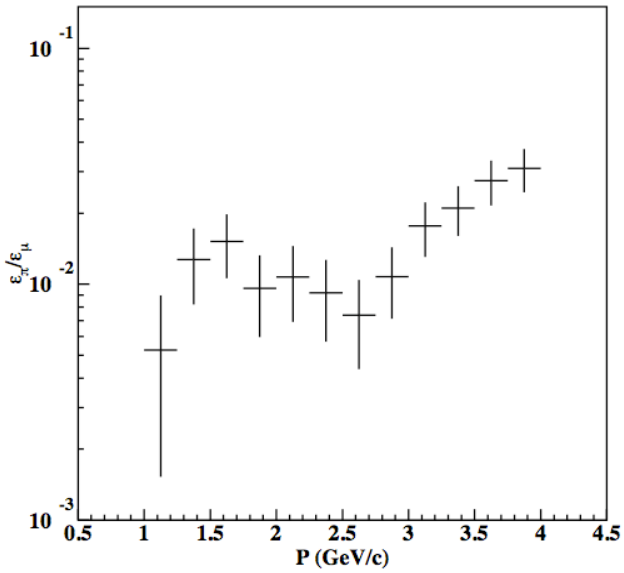

Figure 4.5.2.2 Pion rejection factor in the proposed muon detector system.

4.6 Electronics and DAQ

There are three front-end electronics systems, an Electromagnetic Calorimeter (ECal) system, a Silicon-Vertex Tracker (SVT) system, and a Muon system. A level 1 hardware trigger selects events to be read out. Only the ECal and the Muon system provide inputs to the Level 1 trigger system. The triggered events from the three sub-systems are acquired and processed in the data acquisition and processing system. The events are down-selected in a Level 3 filter processing farm and the remaining events are transferred to the offline storage. The generic scheme for the HPS DAQ system hardware is shown in Fig.4.6.1. This system will be described in detail in the following sections.

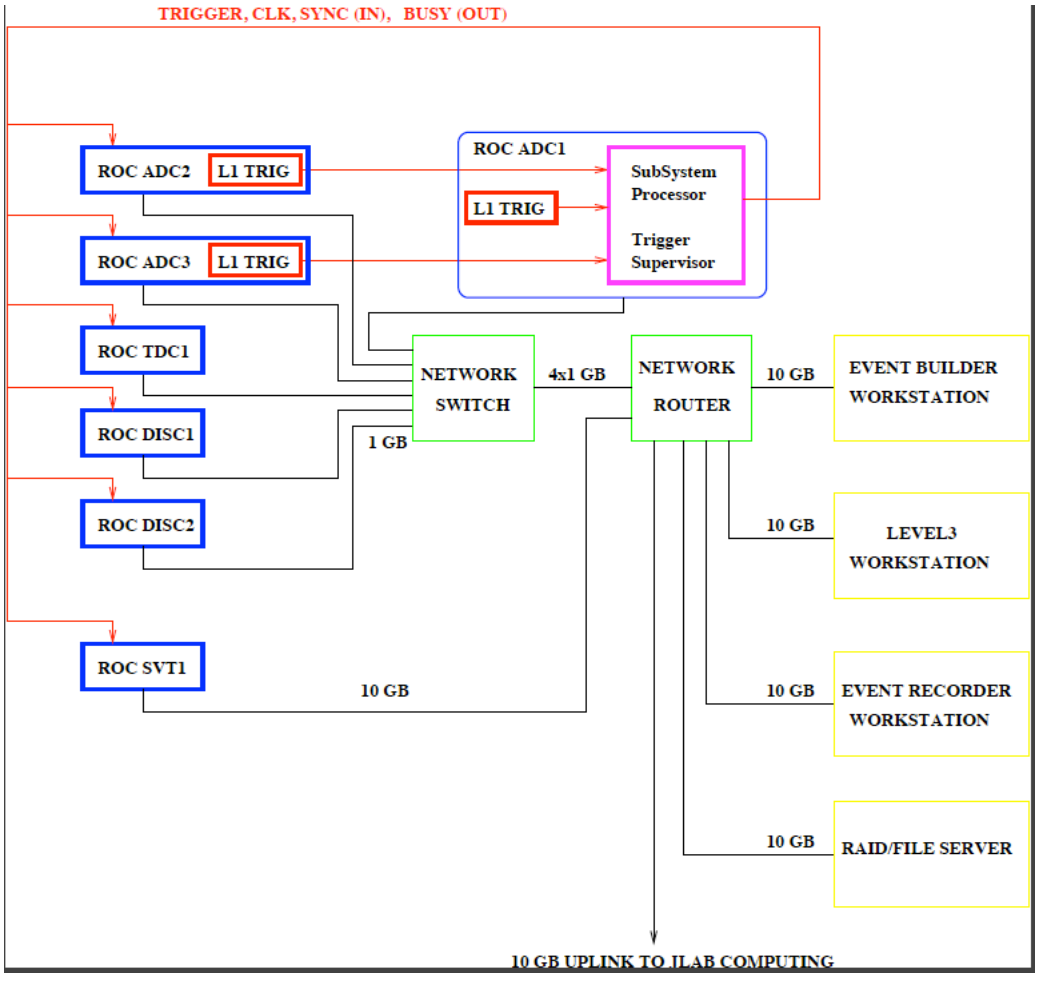

Fig.4.6.1. Readout and processing system block diagram

4.6.1 Electromagnetic Calorimeter Systems

4.6.1.1 Front-end

The Electromagnetic Calorimeter has 556 channels. The system described below can handle up to 560 channels for the ECal plus 144 channels for the Muon System. The charge from APD devices is amplified and shaped in existing JLAB front-end electronics as described in Section 4.4.1. Amplified signals from APDs and anode signals from PMTs will be sent to a splitter. The splitter output will feed FADCs and TDCs.

4.6.1.2 Readout

One of the outputs from every splitter channel is connected to a Flash ADC (FADC) channel packaged in 16-channel VXS modules. Three VXS crates will be used to accommodate 35 16-channel ADC boards plus 9 boards for the Muon System, as shown in Figure 4.6.1.1.

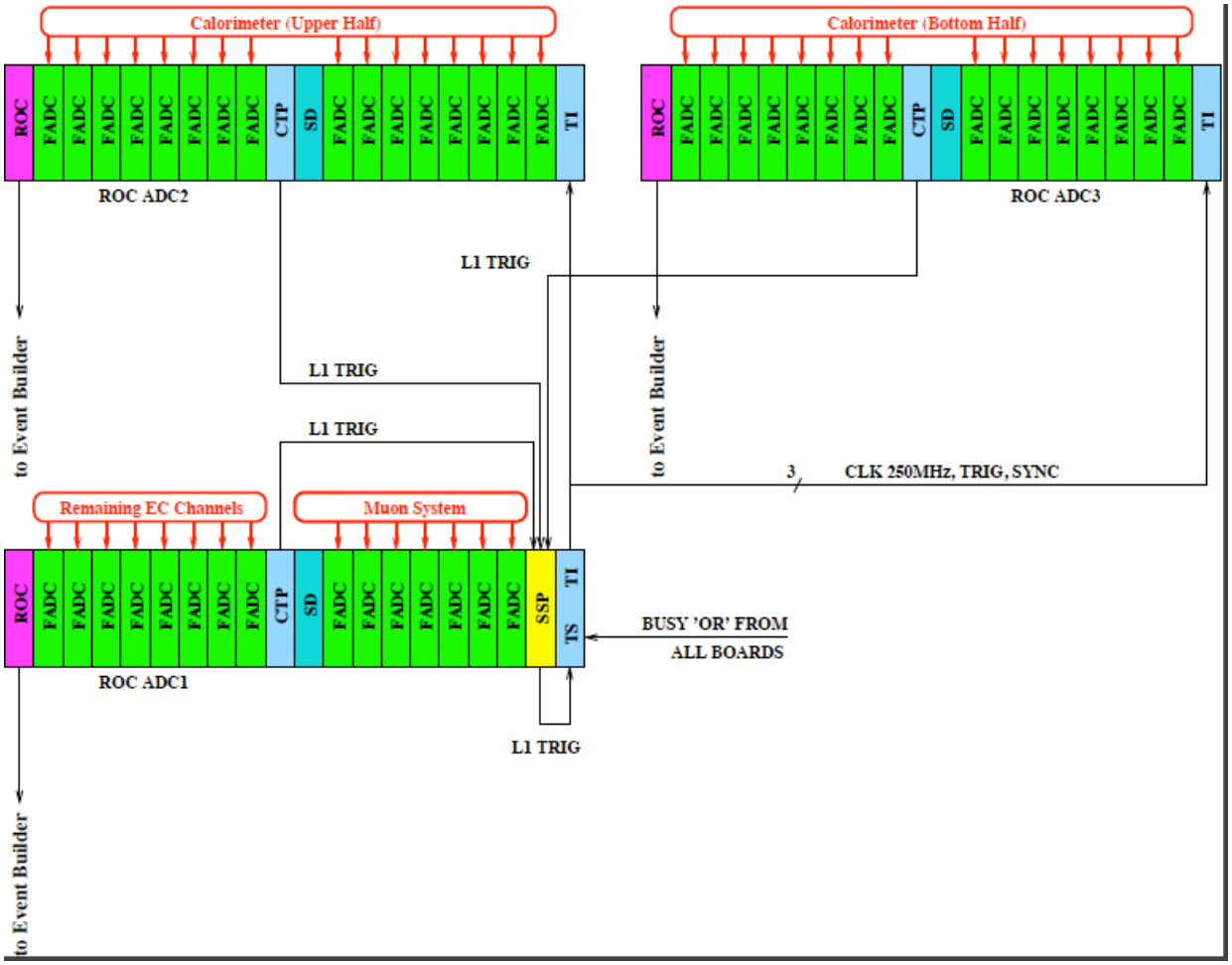

Fig.4.6.1.1 Calorimeter data acquisition and trigger system.

The 16-channel JLAB-made 12-bit 250-MHz Flash ADC boards are still under development and will be in production by the end of 2010 calendar year.

4.6.1.3 **Trigger**

The FADCs are a significant part of the trigger system and will send signal information to the Crate Trigger Processor (CTP) board installed in the same crate. The signal information will include energy for every channel above threshold. A threshold will be applied to the energy sum of several consecutive hits in the same FADC channel as described below. Information will be sent to the CTP board over the back-plane serial bus every 16 ns. The CTP board will perform cluster finding and form the trigger decision. There are three CTP boards in three VXS crates working in parallel searching for clusters in the ECal. Results are reported to the Sub System Processor (SSP) board installed in one of the VXS crates. The SSP board does additional cluster finding on the boundaries between different VXS crates, and forms the final trigger decision.

4.6.2 **Silicon Vertex Tracker System**

The SVT readout system is shown schematically in Figure 4.6.2.1.

4.6.2.1 **Front – End Hybrid**

There are a total of 106 silicon strip sensors, each one connected to a hybrid holding five 128-channel APV25 integrated circuits. The APV25 provide amplification and analog storage for trigger accepted events. There are 5 analog outputs for each 640-channel hybrid. Each hybrid will be connected via a multi-twisted-pair cable to a readout board.

4.6.2.2 **Readout**

The SVT readout system is shown in Figure 4.6.2.1. The outputs of up to 10 hybrids are digitized on ATCA (Advanced Telecommunication Computing Architecture) readout boards. At least 10-bit ADC's are used to digitize the APV25 outputs at a rate of ~40 MHz (system clock of 250 MHz/6 = 41.666 MHz). The board will also contain FPGA's and will be designed at SLAC. Nine of the ATCA readout boards are housed in one 14-slot ATCA chassis. The ATCA crate will be equipped with commercial CPU board and custom TI (Trigger Interface) unit to communicate with JLAB DAQ and Trigger Systems. The TI card will also be designed at SLAC, using essentially the same circuit as the existing JLAB VME card but in an ATCA form factor. The data from the SVT boards will go over Ethernet to a tracker specific CPU board in the same ATCA crate. This board will process the data and send it in one transfer to the rest of the DAQ system.

The destination host and TCP port will be obtained by the CPU board from run control and passed to all SVT readout boards. Also, the CPU board will propagate DAQ commands (download, prestart, go, end etc) to SVT boards.

The crate also contains an existing 10-G switch card, the CIM (Cluster Interconnect Module) which acts as a switch to connect the 10 readout board, the TI, and the CPU to the JLAB network switch shown in Figure 4.6.1.

The trigger and clock and synchronization signals are received by the TI module and distributed via the ATCA backplane to all the modules.

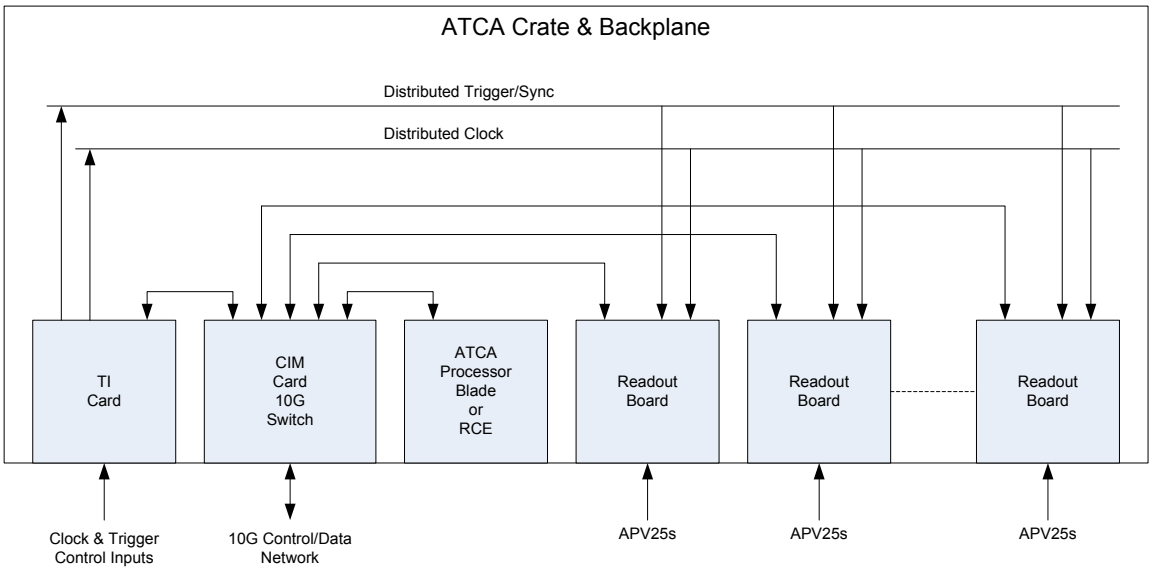

Fig.4.6.2.1. SVT Readout System

4.6.2.3 **Trigger**

The SVT does not provide signals for the Level 1 trigger decision.

4.6.3 **Muon System**

There will be 144 channels which will be read out and also take part in the Level 1 hardware trigger. The data will be input to additional FADC modules, like the ECal System described above.

4.6.4 **System Timing**

All crates receive the 250-MHz master clock. SVT will divide it by 6 for internal use. The ECal and Muon system signals are digitized at 250 MHz (every 4 ns) with a trigger signal generated every 16 ns. The phase of the SVT clock will be latched when the trigger signal is received to correct for the phase of the divide by 6 clock compared to the trigger time.

4.6.5 Level 1 Trigger

4.6.5.1 Trigger System Timing and Layout

Trigger logic will search for a coincidence between different signals every 16 ns; i.e., a single trigger operation takes 4 clocks x 4ns = 16ns.

The maximum trigger decision time (latency) is currently set to 3 μs for Level 1. That value is defined by the SVT readout APV25 chip. A preliminary FPGA algorithm analysis shows that 3 μs for Level 1 should be enough to complete cluster finding, energy reconstruction, and energy selection in the ECal.

It is important to make sure that logically connected regions of the calorimeter will not be wired to the disjoint components of the cluster finder logic, because this would make cluster reconstruction much more difficult. The CTP board contains one FPGA chip collecting information from all 16 FADC modules installed into one VXS crate. That layout should allow searching for clusters in the upper and lower parts of the ECal independently. Unfortunately the total number of channels is slightly bigger than two VXS crates can accommodate. The remaining calorimeter channels will be connected to the FADCs in the third VXS crate, and cluster finding on the boundaries will be performed during the final trigger decision stage in the SSP board, where all the information from three VXS crates will be collected.

4.6.5.2 Level 1 Trigger

The Level 1 trigger system is based on the information which has been input to the ECal Flash ADCs. The first stage components of the trigger logic are incorporated into the Flash ADC board's FPGAs, while the final decision is made in a Crate Trigger Processors (CTPs). Three CTP boards and SSP board will collect data from two parts of the ECal (upper and lower) and from the Muon system, and perform cluster finding independently for those two parts. The trigger decision will be sent to the Trigger Supervisor implemented as a part of one of the TI boards. The Trigger Supervisor generates all necessary signals, and controls the entire DAQ system readout through the Trigger Interface units. The Trigger Interface (TI) units installed in every crate participate in the readout process. The system is free-running and driven by a 16 ns global clock with internal FPGA 4 ns clock.

As mentioned, the VXS crates shown in Figure 4.6.1 contain two trigger system components: processing units on the FADC boards (Figure 4.6.5.1) and Crate Trigger Processor (CTP) boards collecting data from the Flash ADCs over fast serial lines. The CTP boards search for clusters in the ECal, and send the results to the next stage of the trigger system. A signal processing algorithm running in the FADC board processing unit is shown in Figure 4.6.5.2.

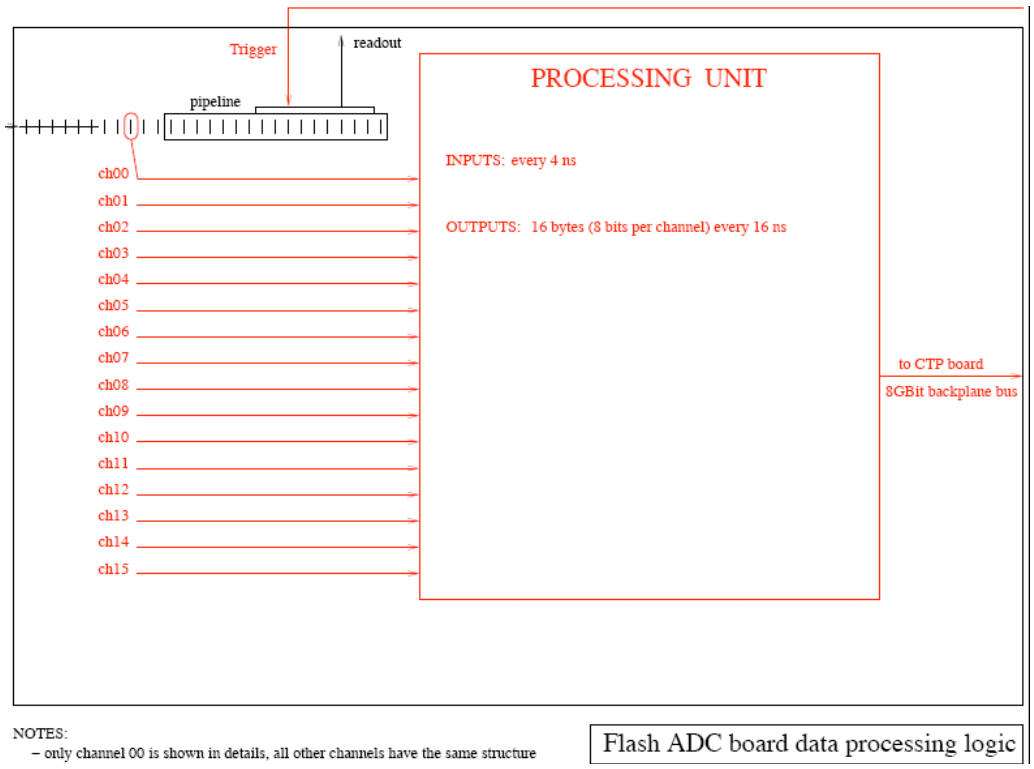

Fig. 4.6.5.1 Processing unit on the FADC board

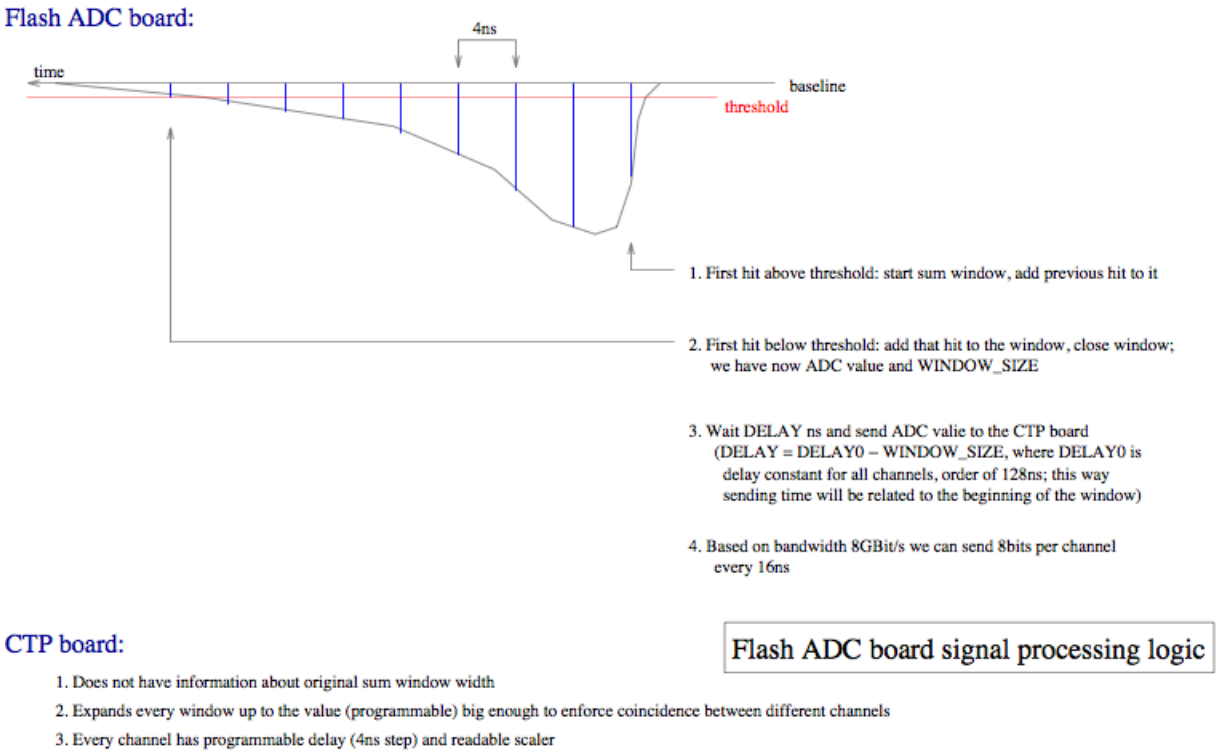

Fig.4.6.5.2.Signal processing algorithm on the FADC board

4.6.5.3 Trigger Rate

The maximum trigger accept rate is 50 KHz.

4.6.5.4 Occupancy and Number of Samples for each Trigger

The assumption is that the system will run at 3 times the noise threshold. The occupancy is thus 0.135%. The above occupancy results in 92 SVT channels over threshold. Background studies show the average track multiplicity in an event is 10. On average, each track produces two "hit" strips in each of the two detectors constituting a layer. So there are 10(tracks) x 2(strips/track) x2 (detectors/layer) x 6 (layers) = 240 samples. Adding the noise pulses, gives 332 samples. Each of these channels over threshold results in 6 digitized values, which record the pulse heights in time bins adjacent to the hit in question, for subsequent pulse shape reconstruction and timing. Thus an average event has 332 x 6 = 1992 samples for each level 1 trigger signal.

4.6.6 Trigger Deadtime

The proposed trigger system is nearly deadtimeless. The trigger decision goes to the Trigger Supervisor every 16ns. The Trigger Supervisor can apply deadtime if necessary, for example on 'busy' or 'full' condition from the front end electronics. There is a limitation on the FADC board on the number of triggers received within a short time interval. For example, if four triggers arrived during about 200ns then the FADC will stop receiving new triggers until at least one of four is processed. We are not expecting to hit that limitation with our expected event rate ~50kHz.

4.6.7 Dataflow

4.6.7.1 SVT Data Rate and Volume

The amount of data for a ~40 MHz clock and a 50 KHz rate (APV25) is estimated as:

- Structure
  - Each chip 12-bit header (3 start, 8-bit address, 1 error) plus 10 bits/channel if 10-bit ADC is used
  - For each ASIC need additional 9 bit chip address
- Number of bits for each hit
  - 9-bit chip address (16-bit)
  - 12-bit header (16-bit)
  - 6 times 10-bit ADC value (72-bit) (assume 10-bit ADC for now since need bare minimum 8 bits)
  - 13 bytes (104 bits) (81 bit minimum)

- Data volume and rate
  - 332 hits x 2 bytes chip address: 664 bytes
  - 332 hits x 2 bytes header: 664 bytes
  - 332 hits x 6 samples x 1.5 bytes: 2988 bytes

The resulting data rate is 4316 bytes x 50 KHz = 215.8 Mbytes/sec.

4.6.7.2 ECal Data Rate and Volume

A single FADC event contains 8 bytes of header for each FADC board, plus 4 bytes per accepted hit (hit means here integral over entire pulse).

With an estimate of 10 % acceptance rate, for each 16-channel ADC board that results in 8 bytes plus 1.6 x 4 bytes = 14.4 bytes average. At a trigger rate of 50 KHz the data rate is then 720 kbytes/sec for each board. A crate holding 16 FADC boards produces 11.5 Mbytes/sec. The limitation will be at the level of 50MB/s data rate from each VXS crate, so there is adequate headroom.

The total data rate from the ECal consisting of 35 FADC boards is 35 x 14.4 bytes x 50 KHz or 25.2 Mbytes/sec. (The total acceptance rate assumed is 10% of 640 channels or 64 channels for each trigger).

4.6.7.3 Muon Data Rate and Volume

The total data rate from the muon system which consists of 9 FADC boards is 9 x 14.4 bytes x 50 kHz, or 6.5 Mbytes/sec. (The total acceptance rate assumes 10% of 144 channels or 14 channels for each trigger.)

4.6.7.4 Event builder

The event builder is a program running on a server assembling the calorimeter, tracker, and muon data into complete events belonging to the same trigger event

4.6.8 Level 3 Sofware Filter

Level 3 will perform full event reconstruction on the Level 3 workstation. The level 3 filter must reduce the data rate received from the Eventbuilder from the expected ~ 215 + 25 + 6 = 246 Mbytes/sec incoming data rate to the less than 100 Mbytes/sec which is the maximum rate to tape. A factor of 4 is the minimum safety margin which should be planned for since the incoming rate is an estimate. In other words the level 3 filter needs to reduce the incoming rate by a factor of 8. An event recorder workstation serves as the event buffer before the events are transmitted to the computer center tape systems.

4.6.9 Monitoring and Calibration

Four practically independent monitoring systems are currently used in JLAB: Nagios-based computer and network monitoring, SmartSocket-based DAQ and Online monitoring, EPICS-based slow controls monitoring, and data monitoring. In addition, a few smaller hardware-specific systems are used. The data monitoring system includes visualization only with almost no alarm capabilities.

4.6.10 Readout Controllers, Computing and Network

Readout Controllers (ROCs) are installed in every VME, VME64X, VXS and ATCA crate. The ROCs are collecting data from the front-end boards, processing it and sending it to the Event Builder over the network. Currently mvme6100 controllers with a prpmc880 or pmc280 co-processor modules are employed. That configuration is fast enough to meet the HPS requirements. By the time of HPS startup, a new generation of multi-processor and multi-core ROCs will be available. ATCA crates will be equipped with commercially available multi-core CPU module.

A Foundry Router is currently used as the backbone of the DAQ system, providing 1GBit and 10GBit connections between components and to the JLAB Computer Center. The Event Builder, Event Recorder, and other critical DAQ components are running on 4-CPU Opteron-based servers, and that configuration is sufficient for HPS as well. ROCs are linked to the Foundry Router through smaller 1GBit switches with 4GBit uplinks. The HPS data storage system (RAID5) is sufficient for up to a 100~MByte/sec data rate. The performance of the link to the JLab Computer Center tape storage will be increased from current 30-40MByte/s to 100MByte/s.

4.7 Offline Computing Model for the HPS Experiment

The following is an outline of the offline computing model envisioned for satisfying the analysis needs of the HPS experiment. The raw data collected over the three month running period must be processed through calibration passes, reconstructed into calibrated objects useful for analysis and separated into analysis streams. Corresponding Monte Carlo will also need to be produced and separated into the same analysis streams.

4.7.1 Data Taking

The primary elements of the HPS detector are an electromagnetic calorimeter (ECal), a silicon microstrip vertex tracker and a muon system. The ECal consists of 5 rows of 46 lead tungstate crystals and 3 rows of 16 lead glass colorimeter blocks on each side of the dead zone. The tracker consists of 6 Si microstrip layers with a total of 67840 readout channels. The muon system has 144 channels. Assuming 30 kHz raw trigger rate, reduction of the trigger rate by a factor of 8 at Level 3, and event size of 4950 bytes, we arrive at data rates of about 18 Mbytes/sec. With 95

days of data collection for each of two runs, the net raw data volume will be 290 Tbytes. The number of events out of the Level 3 trigger is 59 GEvents.

4.7.2 Reconstruction Passes

The raw data must be processed to produce physics data objects that can be analyzed. This reconstruction process will also include filters to select events of physics interest but the output events will be larger due to the size and quantity of the reconstructed objects. Space equivalent to twice that of the raw data volume will be needed. This implies that ~590 Tbytes for data reconstruction output will be needed. About 0.1 CPU seconds are needed to reconstruct an event on a 2.4GHz Xeon core. For such a core, 12 billion CPU seconds would be needed. A typical 2.4GHz Xeon core has a SPECint2000 of 2500. The required CPU hours expressed in SPECint2000 is then 8 G SPEC CINT2000 hrs per pass. A second reconstruction pass may be needed after improvements to the reconstruction used in the first pass are identified and implemented.

4.7.3 Monte Carlo

For modeling signals, estimating backgrounds and confirming the understanding of the detector performance, Monte Carlo simulation will be needed. The number of events needed is estimated to be $1/10^{th}$ of the number of events passing the Level 3 trigger. If the simulated events are comparable in size to reconstructed events then this means that the required space for the simulation output is 43 Tbytes. About 5 CPU seconds are needed to simulate an event. For 5.9 billion events this yields 29 billion CPU seconds per pass or 2 G SPEC CINT2000 hrs. The main simulation will be done off-site.

4.7.4 Analysis Streams

Each analysis will need access to a subset of events relevant to that specific analysis. To minimize the time (CPU and real) required to go through the dataset, streams of analysis specific data may be produced as part of the standard production and/or from users producing their own n-tuples. We estimate that these data will need 20% extra storage space, compared to that needed by the reconstructed data and simulated data.

4.7.5 Analysis CPU

Based on experience with previous experiments, it is reasonable to estimate that the net CPU needed for analysis work (batch and interactive) will be comparable to that needed for production. For CLAS about 30% of physics analysis work was done at home institutions.

4.7.6 Disk Resources

Disk space will be needed for n-tuples, code releases, and scratch areas and approximately 10 percent of the processed data. To accommodate this, 80 Tbytes will be needed.

4.7.7 Data Transfer to/from remote sites

If the option of having a backup of the HPS raw data at JLAB is not feasible then a copy at SLAC will be made. It is very likely that a copy of the reconstructed data will be needed at SLAC and remote sites will be used to assist in the Monte Carlo production. For CLAS about 30% of the simulation production was done remotely.

4.7.8 Summary of Offline Computing Requirements

The HPS offline computing requirements are summarized in the following table:

Silo/Mass Storage (Tape)	**TB**
Amount of Simulated Data Expected	43
Amount of Raw Data Expected	290
Amount of Processed Data Expected	590
Online Storage (Disk) Required	80
Imported Data Expected from Offsite Locations	60
Exported Data Expected to Offsite Locations	590

Computing	SPEC CINT2000 hrs
Simulation Requirements (offsite mostly)	8 G × 2 passes
Production (Replay, Analysis, Cooking) Requirements	2 G × 2 passes

5 Simulated Detector Performance

5.1 Trigger Simulations

The decision to process recorded data in ADCs and TDCs will come from a fast analysis of clusters and hits in the ECal and in the muon detector, respectively, by a Level-1 trigger processing system. Details of the Level-1 trigger electronics and organization were presented in the previous section. Here we discuss the trigger algorithm and the expected Level-1 trigger rate based on GEANT-4 simulations of the detector.

In the experiment, the search for the heavy photon will be conducted in (e^+e^-) or ($\mu^+\mu^-$) final states. The heavy photon is produced predominantly in the direction of the beam and therefore, for the most part, pairs of oppositely charged leptons will be well separated in opposite segments of the ECal. The level-1 trigger algorithm will search for two clusters of energy in opposite segments of the ECal (with respect to the beam direction). Similarly, the Level-1 trigger algorithm will look for minimum ionizing hits in muon hodoscopes in opposite sides from the beam direction. It is expected that the trigger rate from the ECal will be orders of magnitude higher compared to the muon system, and therefore only the ECal trigger selection process is discussed here.

The Bethe-Heitler process will generate most of the electron pairs, however, due to high rates (high accidental coincidences) and overlapping phase space, processes such as elastic scattering, photon bremsstrahlung and Moller scattering will contribute in the trigger (accidental final states (γe^-) and (e^-e^-)). The coincidence time used by the clustering algorithm for grouping hits in individual counters into clusters will play a crucial role in reducing the trigger rate.

5.1.1 GEANT-4 simulations of the ECal

For the study of the calorimeter and the trigger, a full GEANT4 model of the experiment was initiated. This model makes use of a modified version of the Gemc code [1] which is currently being developed to simulate the CLAS12 detector at Jefferson Laboratory. The Gemc code obtains all information for building up the geometry of the virtual detector from a set of MySQL database tables, allowing for a very flexible framework to simulate any detector and making it easy to reconfigure the experiment. Many improvements made to the code for this project are ported back into the original Gemc code so both experiments benefit from this development. The current version of the full experiment simulation is stored in an svn repository [2] and freely available.

All the key elements of the experiment are implemented: the first steering magnet and main analyzing magnet, the target, the silicon trackers, the main components of the vacuum system, the electromagnetic calorimeter, and the muon detector. Figure 5.1.1.1 shows a rendering of the experiment without the muon detector.

Several geometry options have been investigated for the electromagnetic calorimeter in order to optimize its performance. Different arrangements of the calorimeter crystals were tried, with the final version having 5 rows of 46 lead-tungstate crystals and 3 rows of 16 lead-glass calorimeter blocks for each of the top and bottom halves of the calorimeter. The calorimeter is separated from the beam vacuum by a 1 cm thick aluminum vacuum exit window (made transparent in the figure). The beam goes through a continuous vacuum from the vacuum box in the magnet to a vacuum enclosure made from 1 cm aluminum plates, which is inserted between the top and bottom calorimeters. A different configuration, where the beam exited through a thin vacuum window into a region filled with helium for the calorimeter, was rejected due to the increased background rates. To further reduce the background rates without diminishing the structural integrity, a small area of the aluminum plates is cut out where the most intense part of the electron beam exits to the beam dump. A rendering of the simulated calorimeter is shown in Figure 5.1.1.2.

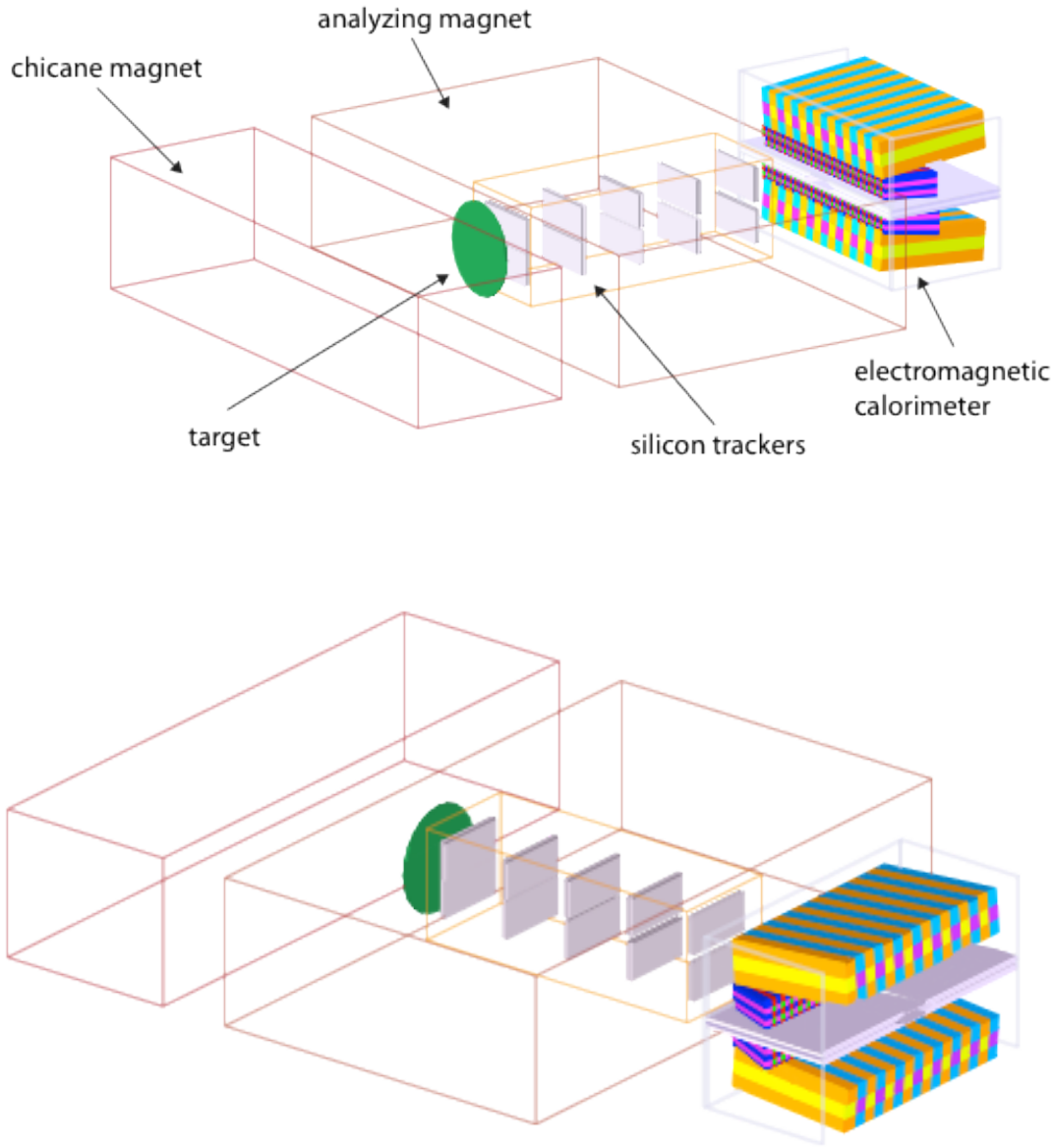

Figure 5.1.1.1. Rendering of the detector simulation used for trigger studies. The red boxes represent the magnets (outlined only), the green circle the target position, the gray rectangles represent a preliminary version of the silicon tracking layers, and the larger blue-gray rectangle represents the vacuum exit window (outlined only). The final object shows the calorimeter, with the crystals colored in alternate colors for clear visibility. The muon detector is not shown on this image.

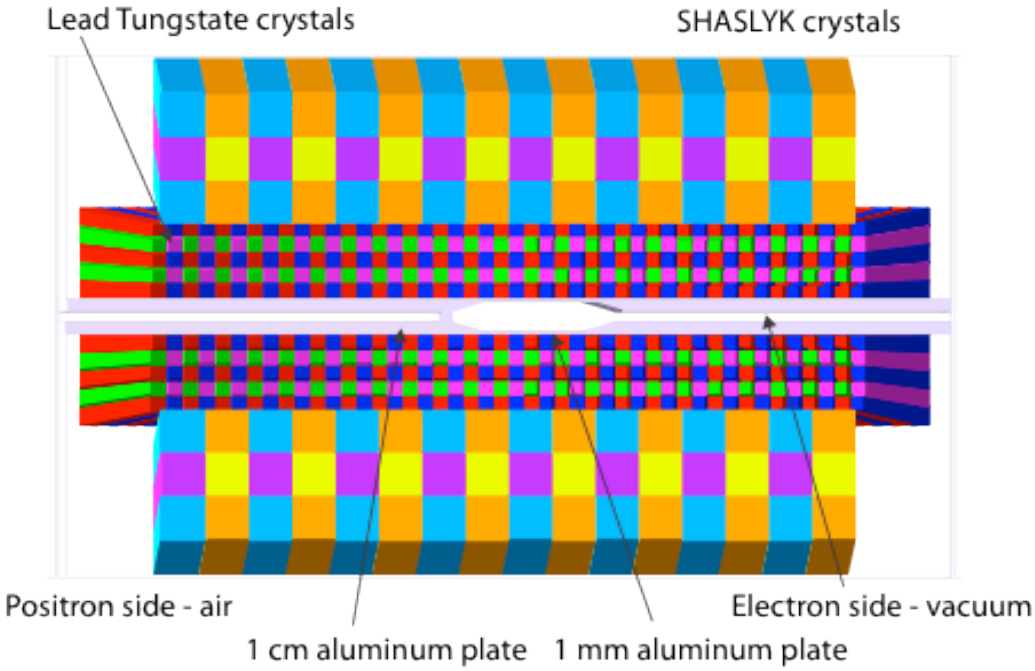

Figure 5.1.1.2. A rendering of the electromagnetic calorimeter setup looking down the beam line. The front exit window and side plates are rendered transparent to permit a view of the crystals and the vacuum plates.

To study the performance of the calorimeter and the trigger logic a large number of background events were simulated. Background events were generated by sending 10,000 6.6 GeV electrons through the 0.25% X_0 tungsten target (equivalent to 4ns of a 400 nA beam) and letting the GEANT4 physics models generate the expected backgrounds (scattered electrons, produced photons and other particles). The physics models used in this simulation include all electromagnetic and hadronic processes implemented in GEANT4. To simulate longer integration times, several events were added together. A significant amount of simulation data was also generated with a 25 ns time window and consistency was checked between the two data sets.

For the trigger studies, input events were used that were generated with MadGraph/MadEvent. These events simulated A' masses of 50, 100, 200, 300, 400, 500 and 600 MeV/c^2. The output files of these simulations could be merged with the background simulation output files to create realistic data samples for different scenarios.

5.1.2 Calorimeter Performance

Two aspects of the calorimeter performance were studied with the simulation. The hit rates on the individual crystals were looked at to make sure these rates will be within a reasonable range for the expected running conditions, and the expected trigger rates for simulated A' particles and for background events were studied to make sure they do not exceed the maximum allowable DAQ trigger rate. This section details the hit rates, the next section the trigger rates.

As one would expect, the highest rates occur on the crystals that are closest to the exiting electron beam. A map of the percentage of events with at least one hit, the hit occupation, for each crystal, using a time integration window of 32 ns and a hit threshold of 10 MeV, is shown in Figure 5.1.2.1. The maximum hit occupation occurs for crystal number 3 in the first row, with close to 75% occupancy.

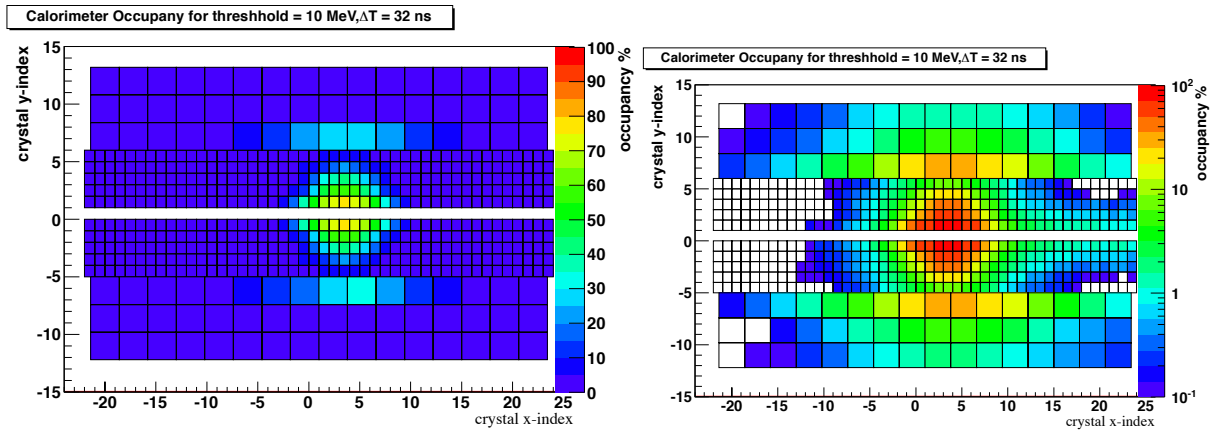

Figure 5.1.2.1 Hit occupancies in the electromagnetic calorimeter for a time window of 32 ns and a threshold of 10 MeV. The figure on the left has a linear z-scale, the figure on the right is identical except for the logarithmic z-scale. Note that the x and y axis are the index of the crystal, not the position. The figure clearly shows that only a few crystals in the first row around the beam exit see very high occupancies.

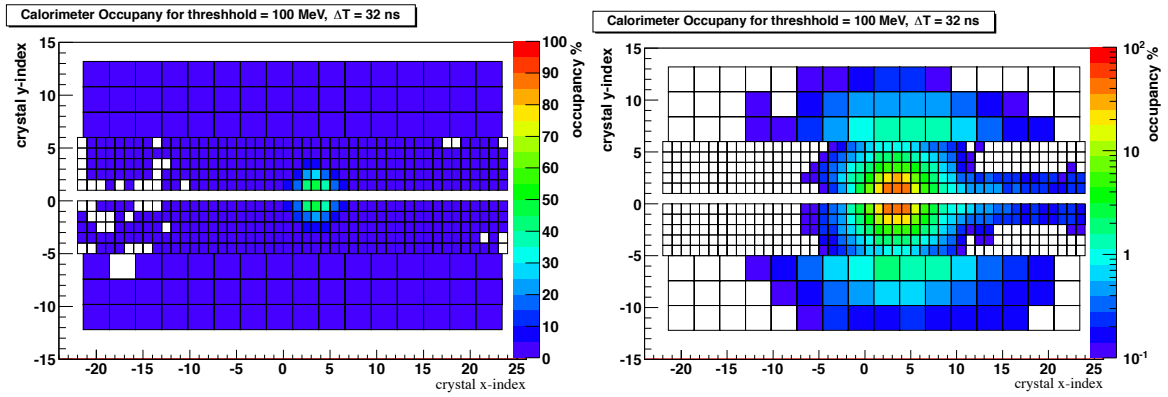

Figure 5.1.2.2 Hit occupancies in the electromagnetic calorimeter for a time window of 32 ns and a threshold of 100 MeV. The figure on the left has a linear z-scale, the figure on the right is identical except for the logarithmic z-scale. Note that the x and y axis are the index of the crystal, not the position. The increased threshold reduces the occupancies on the hottest crystal (#4 row 1) to around 50%.

Increasing the threshold on the crystals improves the occupancy significantly, as can be seen in Figure 5.1.2.2. If the time integration window is also reduced, the crystal occupancies come down to reasonable levels, even for the hottest crystal, see Figure 5.1.2.3.

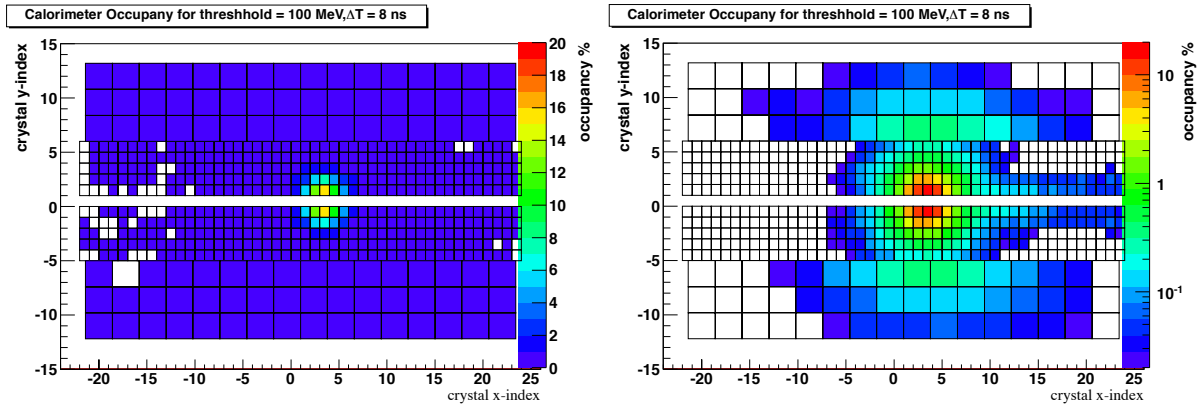

Figure 5.1.2.3 Hit occupancies in the electromagnetic calorimeter for a time window of 8 ns and a threshold of 100 MeV. The figure on the left has a linear z-scale, with a maximum of 20%, the figure on the right is identical except for the logarithmic z-scale. Note that the x and y axis are the index of the crystal, not the position. The reduced time integration window further reduces the occupancies on the crystals to a maximum around 15%. If experimental conditions dictate it, the threshold could be increased further on individual crystals.

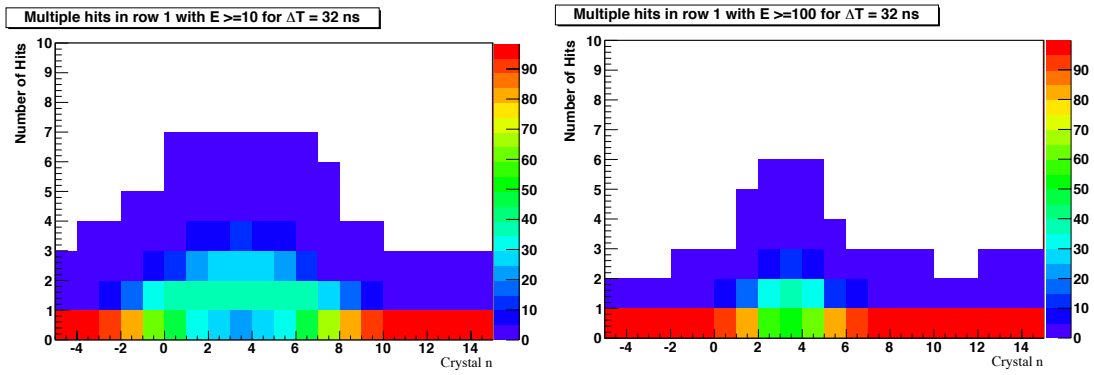

Figure 5.1.2.4 Hit multiplicities for the first row of crystals around the area of the beam exit. The left plot is for a threshold of 10 MeV, the right plot for threshold of 100 MeV. The y-axis shows the number of hits in the 32 ns time window, the z-axis (colors) shows the percentage of events with this hit occupancy.

A more detailed picture of the running conditions for the crystals of the electromagnetic calorimeter can be obtained by looking at the hit multiplicities, the number of hits in a specified time window with energies above some threshold. Figure 5.1.2.4 shows the multiplicity for crystals in row 1 and a time window of 32 ns, with a threshold of 10 MeV (left) and a threshold of 100 MeV (right). The y-axis indicates the number of hits in an event with the specified time window, and the color indicates the percentage of the events where the crystal had that many hits, thus a sum over a column of the histogram adds up to 100%. One can see that for a threshold of 10 MeV and a time window of 32 ns, crystals number 2, 3 and 4 have 2 hits per event for 30% of the time. When the threshold is raised to 100 MeV, the number of events with 2 hits is reduced to fewer than 20%. A plot of the projection on to the y-axis for crystal 4 is shown in Figure 5.1.2.5 (light blue line.) This figure illustrates how reducing the time window reduces the multiplicity on the crystal, exactly as expected. At a threshold of 100 MeV and time

integration windows shorter than 16 ns, the multiplicities on the hottest crystal are reasonable, less than 5% of the events.

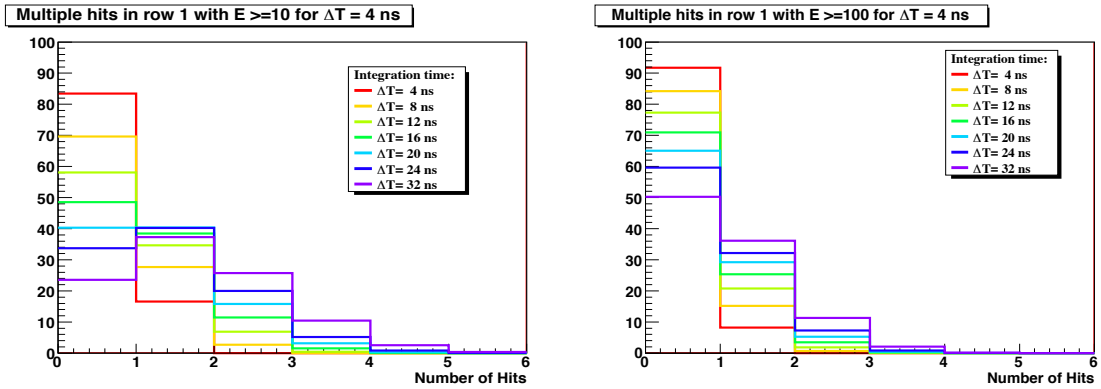

Figure 5.1.2.5. Projection onto the y-axis of the multiplicities of figure 5 for crystal number 3 (violet), and similar projections for shorter time windows.

5.1.3 **Level 1 trigger simulations**

To accurately estimate trigger rates, a Level 1 trigger algorithm was developed and optimized using simulated background data and simulated A' events with masses ranging from 50 MeV/c^2 to 600 MeV/c^2. The events were processed with the full GEANT4 based simulation described in the previous section. For the background events, two simulated 4 ns events at 400 nA were summed to simulate a single 8 ns event. This allowed for the simulation of background events with a trigger coincidence window of 8 ns.

To accomplish the short coincidence timing, the CTP will run in a modified mode. In the standard operational mode, with a predefined integration time and delay, the time of the hit reported to the CTP will have about 16 ns jitter. With a 16 ns time resolution for each channel, a reasonable minimum time window for triggering cannot be less than 32 ns (2 trigger time buckets). To shorten the trigger time window the CTP can use 6 bits for the energy sum (FADC sum), instead of the standard 8 bits, leaving two remaining bits of an 8 bit word to determine the 4 ns clock cycle when summing started. In this arrangement, time jitter at the CTP for each channel will be 4 ns and the coincidence time between channels can be reduced to two 4 ns clock cycles, 8ns. Differences in signal propagation times between channels can be accommodated using internal delays in CTP (with 4 ns steps).

The first step in the Level 1 trigger algorithm is the identification of clusters in the electromagnetic calorimeter. A very simple cluster finding algorithm was used which maximized the number of clusters that would be found. The cluster finding algorithm followed the following logic steps:

1) For each hit in the calorimeter with at least 50 MeV of energy,
2) Search the 3x3 square around the hit for other hits (smaller regions for hits on borders).

3) If a hit with more energy is found, the original hit is not the cluster center,
4) Else add up the energies of the hits over threshold of 30 MeV, in the 3x3 square if these hits were within 8 ns of the center hit.

The resulting cluster samples were then studied to define the most effective trigger, for which the criteria are the largest acceptance of simulated A′ events and the highest rejection of background events. Thus the ratio of accepted A′ events to background events was maximized, with the additional constraint that the background trigger rate would not exceed 30 kHz.

Table 5.1.3.1 shows each of the subsequent cuts and their effect on the number of accepted A′ trigger candidates together with the effect on the background trigger candidates. The simulated A′ mass used was 200 MeV. Numbers are given as a percentage of the total number of simulated events. A more detailed description of each trigger selection cuts follows.

At the lowest level, a trigger required two good clusters in opposite quadrants of the calorimeter. Many events would have multiple clusters in at least one of the two quadrants, in which case all combinations of clusters were tried for trigger candidates. This double counting is not shown in the first row of the table, and eliminated in the last step. This accounts for the increase from the first to the second row in the table.

Trigger Cut.	200 MeV/c^2 A′ Acceptance	Background Acceptance	Background rate
Events with least two opposite clusters	42.35%	2.30%	2.9 MHz
Cluster energy > 500MeV and < 5 GeV	44.25%	0.123%	154 kHz
Energy sum <= E_{beam}*sampling fraction	44.25%	0.066%	82.5 kHz
Energy difference < 4 GeV	44.20%	0.062%	77.5 kHz
Lower energy - distance slope cut	43.46%	0.047%	58.8 kHz
Clusters coplanar to 40°	42.33%	0.0258%	32.3kHz
Not counting double triggers	38.58%	0.0210%	26.3 kHz

Table 5.1.3.1.Trigger selection cuts and their effect on the A' acceptance and background rate, as a percentage of the total number of simulated events. An A' mass of 200 MeV/c^2 was used for this illustration.

As the table shows, a large fraction (2.3%) of the background events have at least two clusters in opposite quadrants of the detector. This would correspond to a background trigger rate of 2.9 MHz. A further refinement on the trigger conditions requires that each of the clusters has an energy of at least 500 MeV, but no more than 5 GeV. This eliminates low energy background hits and hits from electrons with energies close to the beam energy, while having little effect on the A′ acceptance. (Note that in the table the counting of the double triggers give the false impression that the acceptance is slightly higher after this cut.) The fraction of accepted background events now drops to 0.123% (154 kHz). The algorithm now identifies the more energetic and less energetic hit which make up this trigger pair and requires that their sum is less than the beam energy multiplied by the calorimeter sampling fraction (in this case 6 GeV). This cut removes some of the pileup and accidental events. A further cut requires that the two

hits do not differ in energy by more than 4 GeV. These cutes reduce the background acceptance to 0.062% (77.5 kHz).

Next a two dimensional cut is made in distance (of the hit from the beam) plane versus the energy (of the hit). These distributions are shown in Figure 5.1.3.1. The previously mentioned lower energy cut for the clusters is shown as the vertical black line. The histograms show that an additional cut for the least energetic cluster along the red sloped line (E + d 5 (GeV/cm) < 1 GeV), effectively eliminates background events, reducing the background acceptance to 0.047% (59 kHz).

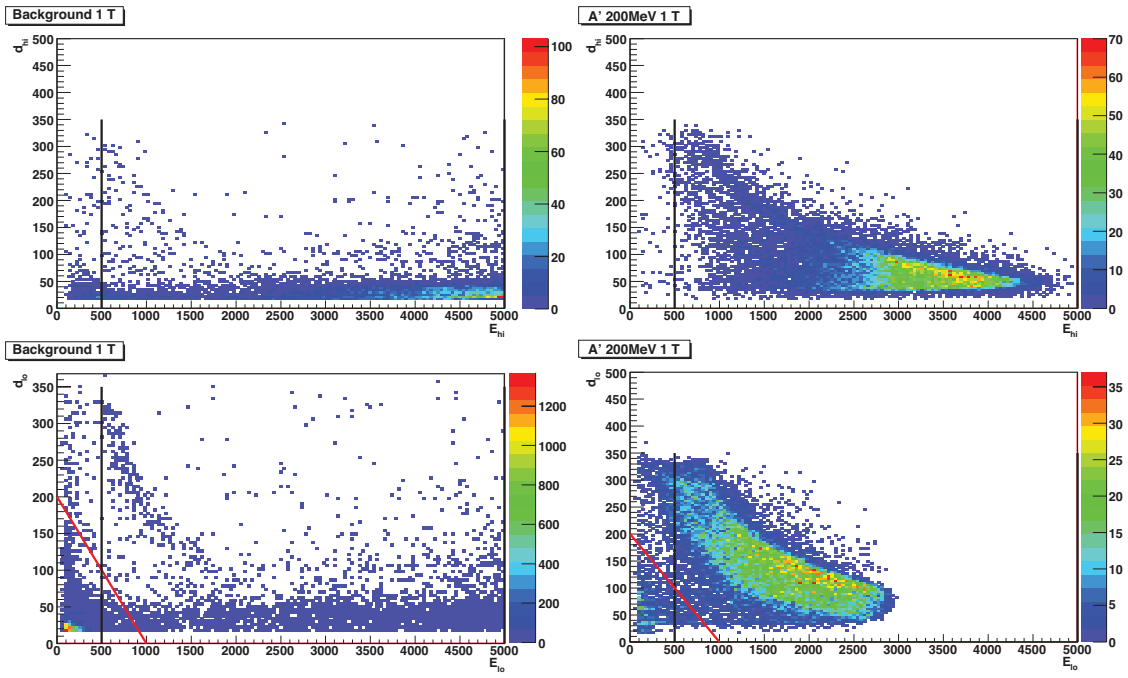

Figure 5.1.3.1. The distance of the cluster center to the beam is plotted versus the energy of the cluster for the most energetic cluster of the trigger pair (top) and least energetic pair (bottom). The left column plots background events, the right hand column plots simulated A' events with a mass of 200 MeV/c^2. As can be seen from the plot, an effective cut to eliminate background events without significantly affecting the A' acceptance is indicated by the red lines.

The requirement that the two clusters are coplanar with the beam within 40 degrees further eliminates background events, leaving 0.026% (32 kHz). Eliminating the remaining double triggers reduces the background acceptance to 0.021 % or 26 ±1 kHz.

These cuts leave us with a trigger rate of close to 30 kHz, which is acceptable to the trigger electronics, which has a maximum rate of 50 kHz. It still leaves some headroom for less than perfect beam conditions, a slight increase in the beam current, or inaccuracies in the GEANT4 physics model. If an additional reduction of the trigger rate is desired, it is possible to remove a few crystals from being considered for the center of a cluster. Removing the crystals with the highest occupancies from the trigger (crystals in row 1, columns 0, 3 and 4) reduces the final

trigger acceptance to 0.011% or 13.8±1 kHz, while reducing the A′ acceptance by only a few percent.

The same trigger algorithm was run on simulated data for a number of different A′ masses. The resulting acceptance estimates are listed in Table 5.1.3.2. The proposed trigger algorithm retains high efficiency for signal events and suppresses the background triggers by a factor of 100.

A′ Mass (MeV/c^2)	50	100	200	300	400	500	600
Ecal	2.9%	15.2%	38.6%	45.2%	43.3%	39.3%	34.8%

Table 5.1.3.2. Trigger acceptance estimates for different A' masses for runs with a beam energy of 6.6 GeV.

The whole process of optimizing the cuts for an efficient trigger algorithm was repeated for simulation runs with a beam energy of 2.2 GeV. These runs where simulated with the magnetic field in the analyzing magnet at half strength (0.5 T) to increase the acceptance for low mass A's. The simulated current for the background runs was 200 nA on a 0.125% X_0 Tungsten target. This reduction in luminosity is needed to maintain reasonable occupancies in the silicon strip detectors. The resulting background trigger rate is 30.8 kHz corresponding to a background acceptance of 0.025%. Table 5.1.3.3 shows the acceptance for simulated A′ masses from 25 to 250 MeV/c^2.

A′ Mass (MeV/c^2)	25	50	75	100	150	200	250
Ecal	4.9%	23.8%	32.1%	34.8%	34.6%	26.2%	18.3%

Table 5.1.3.3. Trigger acceptance estimates for different A' masses for runs with a beam energy of 2.2 GeV.

The trigger rates for trident backgrounds (Sec 3.2 and 3.3) at 6.6 GeV were estimated using radiative and Bethe-Heitler events generated with MadGraph/MadEvents and incoherent (two-step) trident events generated with EGS5. Although these trident processes have a large cross section, a large fraction of events do not satisfy the trigger conditions either because electrons/positrons are produced in a small polar angle or the energies are soft. Table 5.1.3.4 summarizes the trident trigger rates. Since GEANT4 does not include these trident events, this rate needs to be added to the total background rate, resulting in a total rate of 32±1 kHz if no crystals are eliminated from the trigger.

Trident	Estimated trigger rate
Coherent trident	
Bethe-Heitler	7.8 kHz
Radiative	130 Hz
Incoherent trident	180 Hz

Table 5.1.3.4 Trident trigger rates

5.1.4 GEANT-4 simulations of the muon detector

GEANT4 simulations have been used to study the trigger rates in the muon system due to background hits. It is expected that the true di-muon rate will be quite small compared to the ECal trigger rate and should not cause problem for the DAQ. The proposed muon detector model was implemented into the GEANT-4 simulation of the HPS setup. Background events were generated using GEANT-4 physics models by sending 10K, 6.6 GeV electrons through the 0.25% X0 tungsten target (equivalent to 4 ns of a 400 nA beam). Rates of single hits in each horizontal scintillation paddle and coincidence rates in hodoscope layers were studied.

In Fig. 5.1.4.1 rates in the first paddle, closest to the beam plane, are presented without (left) and with (right) 0.2 MeV energy cut. The singles rate in that strip is ~20MHz without any energy cut and gets reduced to ~2MHz after 0.2 MeV cut is introduced (in the experiment this cut will be done at the FADC level). Note that 0.4 MeV cut was used in the studies of p/m rejection. A factor of 10 reduction of rates is observed for the second paddles in the layers.

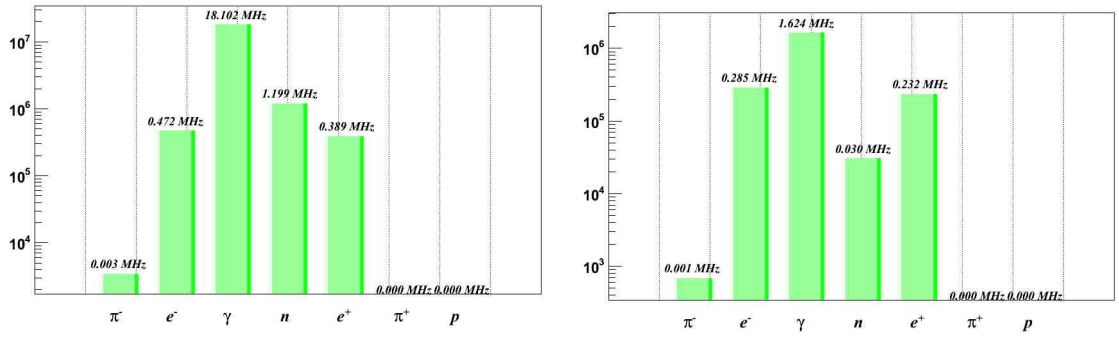

Figure 5.1.4.1 Rates in the first paddle (closest to the beam) of the Layer-1 hodoscope for various particle species. On the Left, there is no energy cut; on the right, the rates are with 0.2 MeV energy cut.

As was described above, selecting coincidence hits with MIP energy deposition in at least three layers of the scintillation hodoscope will identify muons. In Fig. 5.1.4.2 the coincidence rate between the first three hodoscope layers of the top detector in a 4 ns time window and with 0.2 MeV energy cut for a single hit is presented. The rate is ~20kHz. The coincidence rate between these three layers and the corresponding three layers of the bottom module was zero coincidence rate in the same 4 ns time interval. Hence that trigger rate from the muon system will be very low and should be dominated by di-muon pair production with a small mixture of two pion events.

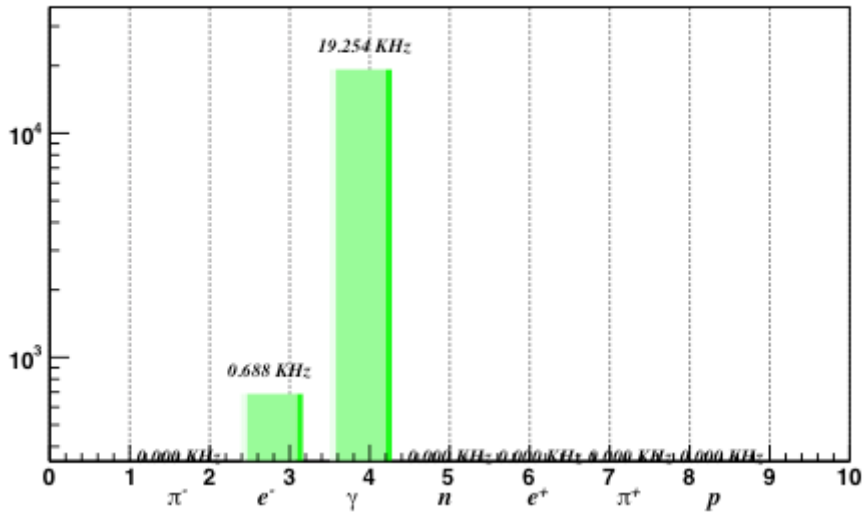

Figure 5.1.4.2 Rates in the first three layers of the top part of the detector.

5.1.5 References

1. See http://clasweb.jlab.org/wiki/index.php/Gemc

2. Anonymous checkout of all the code: "svn checkout svn://gourd.unh.edu/Aprime/trunk".

5.2 Tracker Occupancies and Acceptance

Figure 5.2.1 shows the distribution of charged particle hits in Si tracker layer 1 which is located 10 cm from the target as generated by EGS5. The beam energy is 6.6 GeV, and the target thickness is 0.25% X_0. Multiple Coulomb scattered beam electrons are confined within 0.5 cm of the beam axis (x=y=0), while the low energy Moller electrons are distributed in a parabolic shape. There are very few positrons. From these distributions, the detector occupancy in the horizontal Si strip sensor in the 7.5 ns time window is calculated for a 400 nA beam current and five different target thicknesses, 1.0% X_0, 0.5% X_0, 0.25% X_0, 0.1% X_0, and 0.05% X_0, and is shown in Figure 5.2.2. As described in Section 4.3.4, the dead zone is defined by using a criterion that the maximum occupancy in Layer 1 is 1%. For a 0.25% X_0 target and 430 nA beam, the occupancy is 1% at a distance of 1.5mm from the beam in Layer 1, which corresponds to a dead zone of ± 15 mrad. As long as the product of target thickness (T) and beam current (I) is constant, the same A′ production rate is maintained. Since the multiple scattering and hence the effective beam size is reduced in a thinner target, it is advantageous to use a thinner target and a higher current. Using the constraint that the occupancy is 1% at 15 mrad, we find the beam current I which gives this occupancy for each of several potential target thicknesses T. The quantity $(I \cdot T)^{1/2}$, which is approximately proportional to the sensitivity $S/\sqrt{B}$, is given in Table 5.2.1, showing how the sensitivity improves as the target thickness decreases.

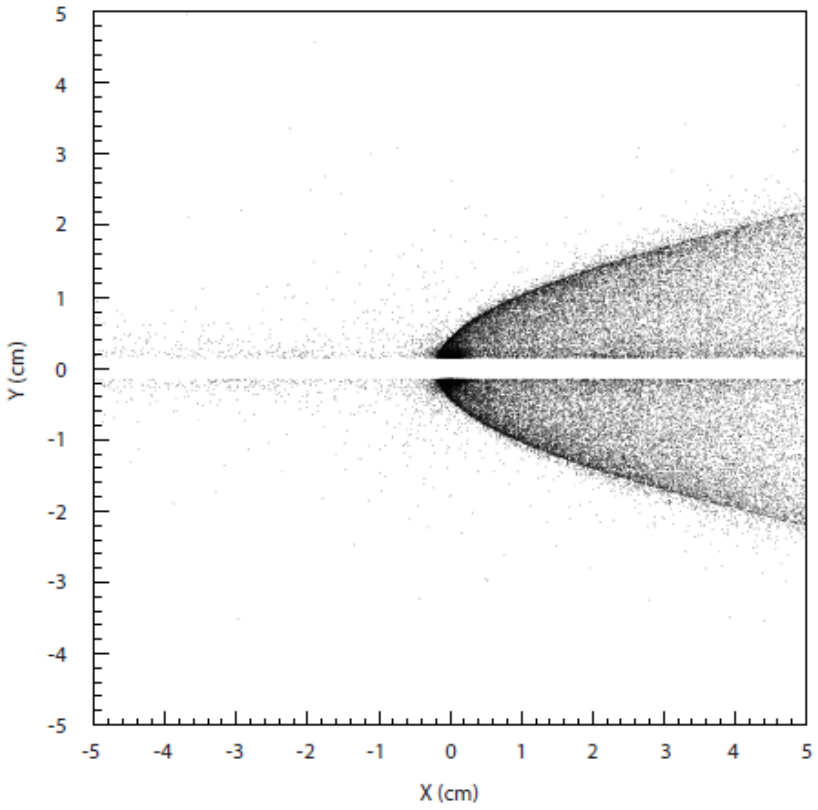

Figure 5.2.1 Charged particle distribution in layer 1

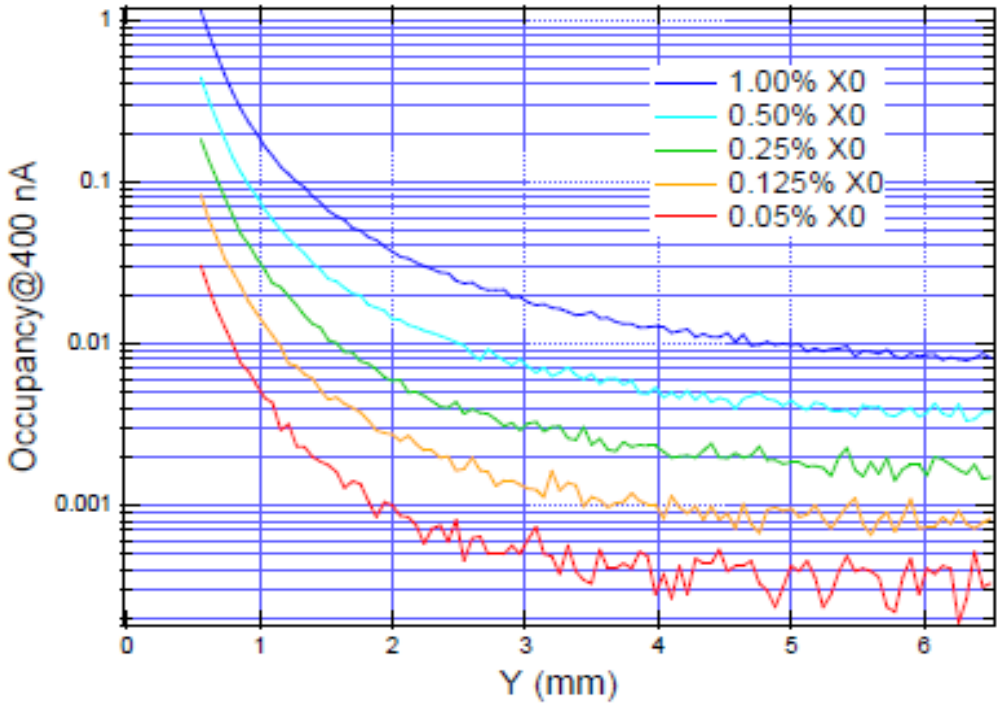

Figure 5.2.2 Silicon sensor layer 1 occupancy at 400 nA vs distance from the beam in mm.

Target thickness (% X_0)	Beam Current (nA)	$\propto S/\sqrt{B}$
1.0	60	7.7
0.5	170	9.1
0.25	430	10.4
0.10	1330	11.6
0.05	2860	11.9

Table 5.2.1. Beam current yielding 1% occupancy in Silicon sensor layer 1 for various target thicknesses at 6.6 GeV, and the relative experimental sensitivities which result.

Once the dead zone is determined, the tracker acceptance can be calculated by requiring that both e+ and e- from A′ decay are detected in the first five silicon layers. The tracker parameters given in Table 4.3.4.1 are used. Figure 5.2.3 shows the tracker acceptance as a function of A′ mass at beam energy of 2.2 GeV and 6.6 GeV for several A′ decay lengths, including when the A′ decays at the target (Zv = 0 cm). The tracker has useful acceptance from 20 MeV to 1000 MeV; lower (higher) beam energies can probe A's of lower (higher) mass. At the lower mass side, the dead zone limits the acceptance, while the transverse tracker size limits the acceptance at the higher mass side. As the A′ decay distance becomes longer, the first-five layer acceptance gets smaller as the A′ decay products remain within the dead zone at Layer 1. To detect these long-lived A′ decays, the last four silicon layers can be used. Figure 5.2.3 shows the last-four layer acceptance when A′ decays at 10 cm and 20 cm from the target. The tracker design based on six silicon layers has a large acceptance over a wide range of decay lengths.

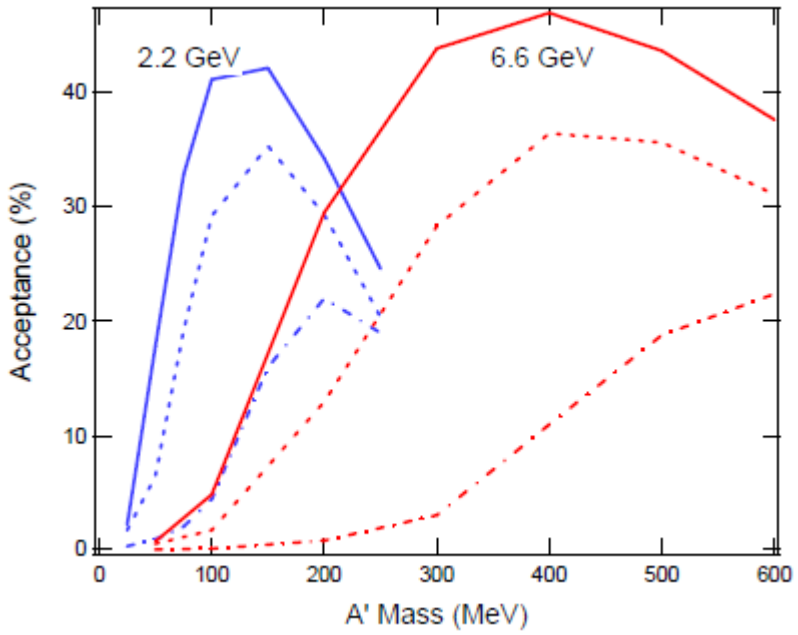

Figure 5.2.3 Tracker acceptance as a function of A' mass at 2.2 GeV (blue) and 6.6 GeV (red). The solid lines are for A' decay vertex at Zv=0 cm, the dash lines at Zv=10 cm, and the dash-dots at Zv=20 cm.

5.3 **Tracking Performance**

We use a GEANT4 Si tracker simulation based on SLAC's org.lcsim infrastructure for full simulation of the passage of charged and neutral particles through the target and tracker. It creates realistic energy deposits in the silicon microstrip detectors, accounts for dead material, accurately digitizes energy deposits into strip hits, creates clusters, and performs track finding and reconstruction. This simulation is used to get realistic estimates of tracking pattern recognition efficiencies and purities in the presence of all the expected electromagnetic backgrounds, and to evaluate momentum, invariant mass, and vertex resolution.

The standard pattern recognition algorithm is designed to find tracks efficiently using strip hits in the tracker. The first step in track finding is to convert the digitized hits into a common hit format. This format encapsulates all the information needed by the standard pattern recognition algorithm, while insulating the track finding from differences and changes in the digitization algorithms. We consider two types of hits: axial strip hits that have one measured coordinate and one bounded coordinate and stereo hits formed from a pair of strip hits in the same tracker plane. Track finding is controlled by a set of strategies. A strategy consists of the list of detector layers to be used, the role of each layer (seed, confirm, or extend), kinematic constraints (momentum, impact parameters), requirements on the number of hits, and the χ^2 cut.

The track finding algorithm is exhaustive in the sense that all combinations of hits that could potentially lead to a successful track fit are considered. The algorithm proceeds in four steps:

1. The first step is to form a 3-hit track seed candidate by taking all 3-hit combinations possible among the 3 seed layers. A helix fit is performed on the seed candidate, and those seeds that fail the χ^2 cut are eliminated. The helix is defined by five parameters: the curvature, angle and distance from the z-axis at the DOCA (ω, ϕ_0, and d_0 respectively) describing the circular part of the helix, and the slope and point on the a-axis at the DOCA ($\tan\lambda$ and Z_0) describing the trajectory in the s-z plane.

2. The second step tries to "confirm" the seed by adding additional hit(s) from the confirm layer(s). A helix fit is performed on the new seeds and those that fail the χ^2 cut are eliminated. Typically, it is found that good performance is achieved with one confirmation layer.

3. The third step seeks to "extend" a confirmed seed by trying to add additional hits from the extend layers. Each time a new hit is considered, a helix fit is performed and the hit is discarded if it fails the χ^2 cut. If no hits in a given extend layer give a satisfactory helix fit, then the original track seed is kept and the next extend layer is tried.

4. Track seeds that meet the strategy's requirement on the minimum number of hits are merged to form a list of distinct tracks. Two track candidates are allowed to share a single hit, but if a track candidate shares more than one hit with another candidate, an arbitration scheme is used to select the better candidate. Precedence is given to the candidate with the greatest number of hits, while the candidate with smaller χ^2 is selected when the number of hits is equal.

Consistency checks and hit sorting algorithms are used to minimize the number of helix fits performed, substantially improving the performance of the algorithm. Furthermore, a "bad hit " χ^2 cut is used to identify track candidates with an outlier hit and allows preference to be given to track candidates without an outlier hit. A key component of the pattern recognition algorithm is a fast helix fitter. The helix fitter takes as input 3 or more tracker hits. The hits can be any combination of axial strip or stereo hits in the barrel. The fast fitter is used to estimate the helix parameters and helix fit χ^2. First, a circle fit to the x,y coordinates of all hits is performed using the Karimäki algorithm to determine the helix parameters ω, ϕ_0, and d_0. If there are two or more stereo hits, then a line fit in the s-z plane is used to determine the Z_0 and $\tan\lambda$ helix parameters.

For all of the particles we will be measuring, the multiple scattering errors will exceed the intrinsic hit resolution. Multiple scattering errors for both the active and dead materials are estimated and included in the helix fit. Correlations in the multiple scattering errors are ignored, leading to an under-estimate of the helix parameter errors by a factor of ≈1.5. For stereo hits, full account is taken for the separation between the two stereo layers in the calculation of both the hit position and hit covariance matrix.

The performance of the standard pattern recognition algorithm is shown in the sections below. Unless otherwise noted, the tracking strategies require 1 confirmation hit, stereo hits in at least the first five tracker planes (the last plane is added if it passes the extend criteria), $0.5 < P_{\neq} < E_{beam}$ GeV/c, and both the x and y distance of closest approach to the beam axis are less than 0.5 mm.

In order to study the tracking performance of the detector (described below), we use samples of A' events at a variety of energies and decay lengths. On top of each event, we overlay backgrounds produced by the passage of beam electrons equivalent to 450nA in 7.5ns on a 0.25% X_0 target for the 6.6GeV/c^2 samples and 200nA in 7.5ns on a 0.125% target for the 2.2GeV/c^2 samples[2], each with a beamspot of radius 20um.

5.3.1 Tracking Efficiency, Pattern Recognition and Fake Rates

Due to the requirements imposed on the tracks, the efficiency for finding tracks in the geometric acceptance is not 1. The plot below (Figure 5.3.1.1) shows the efficiency to find an electron (or positron) from an A' decay which fall in the detector acceptance, and thus should be found by the tracking algorithm, as a function of track momentum. The average track reconstruction efficiency is 98% and the bulk of the inefficiency comes from the cut on the total χ^2.

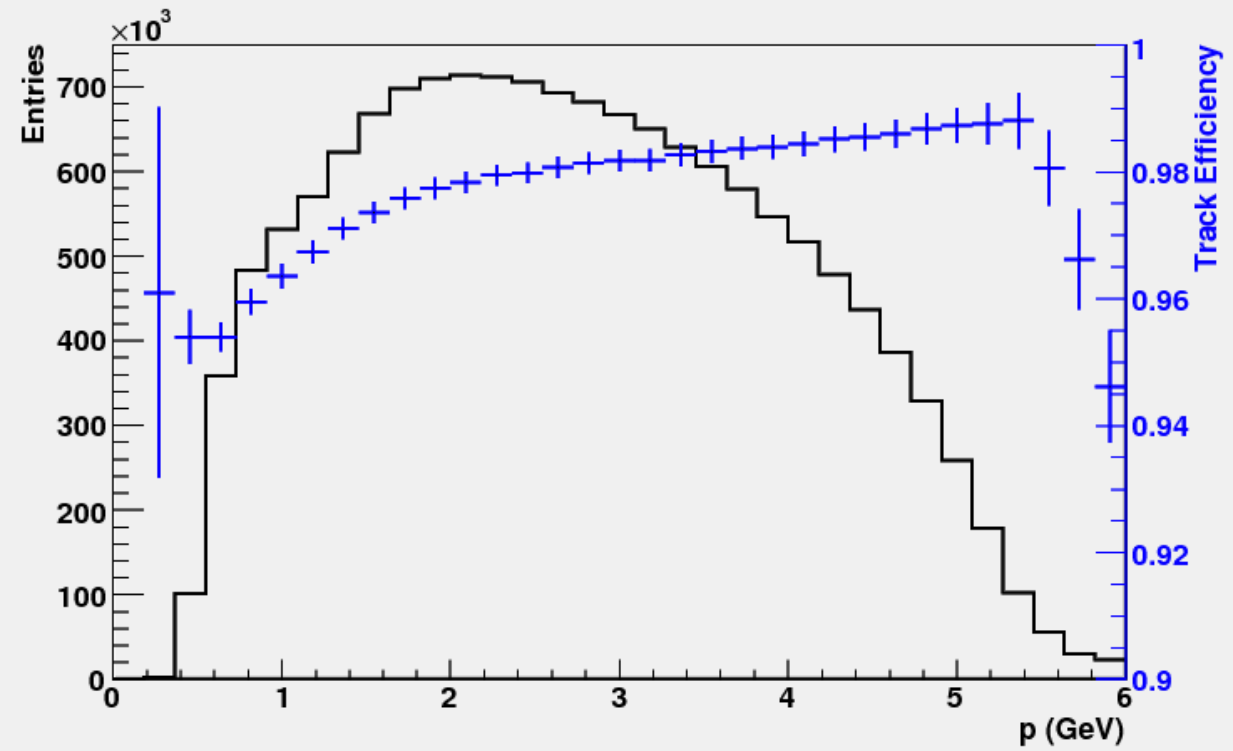

Figure 5.3.1.1: Distribution of A' decay electrons (black) and track finding efficiency (blue) versus momentum for electrons in the detector acceptance.

Of the reconstructed tracks, a small percentage of them include a hit that is not from the correct electron. These "bad" hits may be from one of the high energy beam electrons scattered from the target into the detector or from a lower energy secondary. The left plot of Figure 5.3.1.2. shows the number of bad hits/track for both the electron and positron from the A' decay. The number of tracks with 0 bad hits is >95% and the positrons are slightly cleaner since occupancy of the positron side of the detector is smaller. The right plot of Figure 5.3.1.2 shows the layer number

[2] The magnetic field used for this section was 1T for both energies instead of 0.5T for 2.2GeV/c^2 and 1T for 6.6GeV/c^2. For the experimental reach calculation (Section 6) we use resolutions and efficiencies from samples with the correct magnetic fields.

of the bad hit. They tend to be in the closer layers, particularly in the planes that measure the non-bend coordinate. We'll show how these bad hits affect the track parameters in the next section.

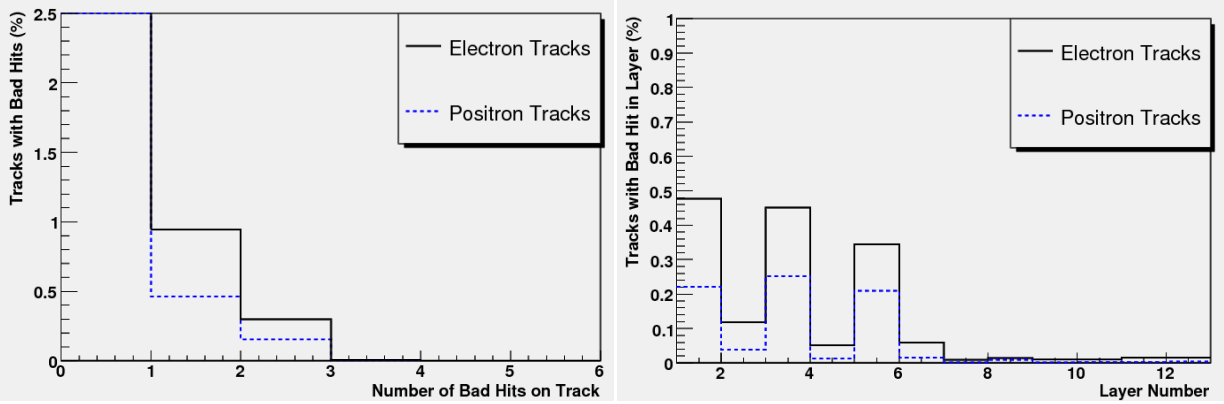

Figure 5.3.1.2: The number of bad hits (left) and the layer number of the bad hit (right) for electron (black) and positron (blue) tracks prior to vertex selection.

5.3.2 Track Momentum and Spatial Resolution

The momentum resolution is shown in Figure 5.3.2.1 as a function of momentum for tracks with 0 bad hits and for tracks with one or more. The momentum resolution for well-reconstructed tracks is $\sigma_p/p = 1.5\%$. Generally, tracks with only bad hits in the non-bend layers have the same momentum resolution as well reconstructed tracks; only bad hits in the bend layers affect the momentum.

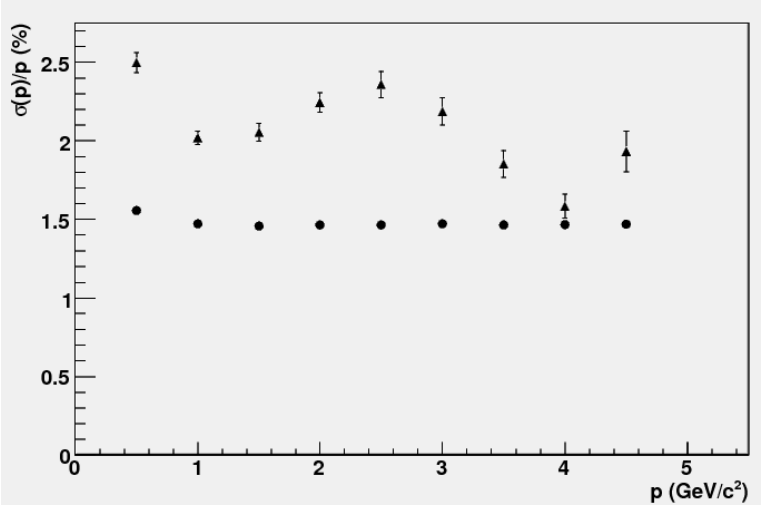

Figure 5.3.2.1: Fractional momentum resolution versus momentum for 0 bad hit tracks (circles) and tracks with 1 or more bad hits (triangles).

One quantity we use to determine track quality is the distance of closest approach (DOCA) to the beam axis. We use this instead of the DOCA to the target beam spot since we are interested in long-lived decays and tracks from those will not point back to the target. We separate the distance into the bend plane (XOCA) and non-bend plane (YOCA) distances. Below, in Figure 5.3.2.2, is the resolution of these quantities as function of momentum for tracks with 0 bad hits. The resolution is, on average, about 50μm but increases significantly at low momentum. The position resolution for tracks with one or more bad hits is somewhat worse, depending on which layer the bad hit is. In particular, when the bad hit is in the first non-bend layer (layer 1), the YOCA is very poorly determined as shown in Figure 5.3.2.2. Tracks with a bad hit in layer 1 are a major contribution to the tail of the vertex position distribution.

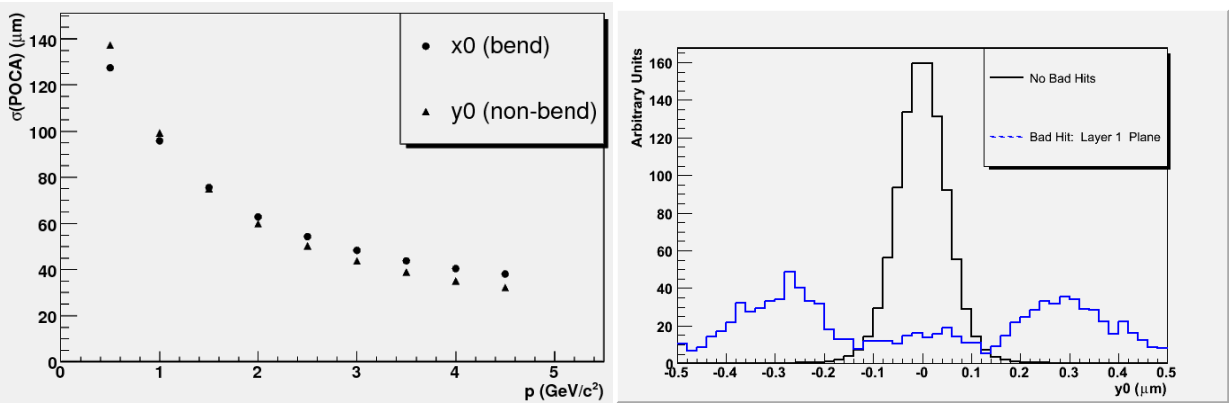

Figure 5.3.2.2: LEFT: The resolution of the position of closest approach to the beam axis versus track momentum. RIGHT: The YOCA resolution for tracks with 0 bad hits (black) and with a bad hit in layer 1 (blue).

For long lived A′ decays, the position of the decay vertex is an important discriminating variable. The dominant background to A′ production is radiative events which originate in the target. Distinguishing A′ decays from the background therefore depends on the vertex resolution and in particular on the tails of the vertex distribution. In order to study the tails, we use large samples of A′ events decaying promptly overlaid on top of the simulated beam background events.

Each pair of oppositely charged tracks is fit to a common vertex using a Kalman filtering method first suggested by Billior [1,2] and used in many experiments. The method uses the measured helix parameters and their correlations to determine the most likely decay position of the A′ and also returns fitted momenta for each particle. We actually fit each pair twice with different hypotheses of their origin. We constrain either the vertex to be consistent with an A′ :

- which originates in the 20um beamspot at the target, and moves off in the direction given by the measured A′ momentum. This fit will be used for the vertexing search.
- which originates and decays at the target within the 20um beamspot. This fit will be used for the bump-hunt only search.

For each electron/positron pair reconstructed in the tracker, we compute the invariant mass based on the fitted momenta of the tracks. The mass resolution depends on the invariant mass of the pair and is shown in Figure 5.3.2.3. The right-hand plot in Figure 5.3.2.3 shows the improvement in the resolution for the second fit, where the decay is assumed to occur in the target.

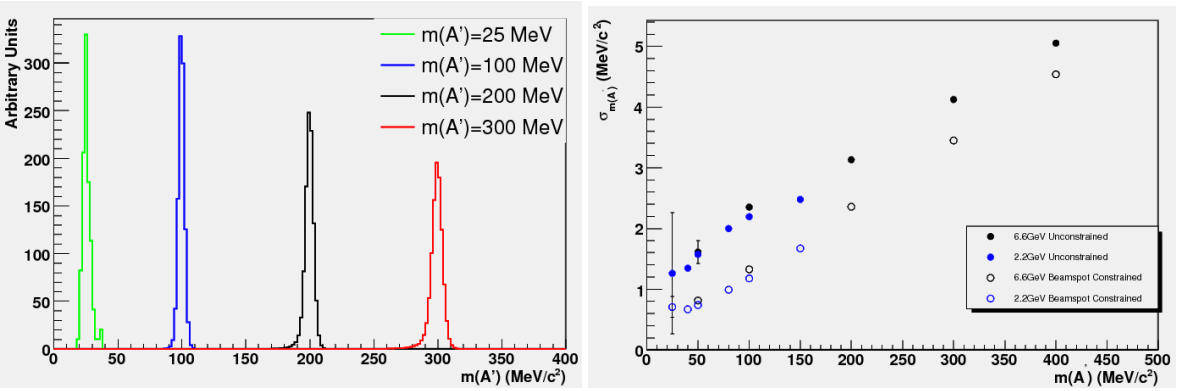

Figure 5.3.2.3 Left: The mass distributions for different generated A' masses. Right: The gaussian width of the mass distributions vs generated A' mass. The open circles are the resolutions when the decay is constrained to the target beamspot and the closed circles are without this constraint.

Even for prompt decays, the z vertex position (V_z) distribution of all reconstructed e+e− pairs (solid black histogram, Figure 5.3.2.4) shows a long tail, still significant beyond 5cm. This tail is primarily comprised of events where one or both of the tracks use one or more bad hits. Fortunately there are a number of quantities we can use to minimize the tails. Namely, for purposes of this proposal, we make the following cuts:

- The χ^2 of each track is less than 20
- The total momentum of the A' candidate is less than the beam energy
- A very loose cut on the reconstructed vertex position $|V_x|$<400um and $|V_y|$<400um
- The clusters in layer 1 of each track must be isolated from the next closest cluster by at least 500 um
- A χ^2 cut on the vertex fit of less than 15

Figure 5.3.2.4 shows the vertex resolution for a sample of 200 MeV A' events generated from a 6.6 GeV beam. The cuts above remove almost all of the tail past ~1.5cm (points with errors in Figure 5.3.2.4) while retaining ~55% of the e^+e^- pairs from the A' candidate. The RMS of the final resolution distribution is 1.2mm while the high side tails of the distribution (i.e. the background for the long lived A' decays) contain $1x10^{-4}$ (>5mm) and $5x10^{-7}$ (>1cm) of the events. The events on the tail are enhanced with vertices where there are one or more bad hits on the track

(represented by the blue histogram in Figure 5.3.2.4), although there is still a contribution from well-reconstructed tracks. The rejection of tracks with bad hits depends strongly on the precision of the virtual A' trajectory, which in turn depends on the size of the beamspot. For vertices reconstructed beyond 1 cm, the rejection factor with a 100 μm beamspot is ~5 times worse than for the 20μm spot. Having a small beamspot is very important.

The vertex resolution depends on the invariant mass of the particles being vertexed. Lower masses have worse Gaussian resolutions as shown in Figure 5.3.2.5. This is expected since the error on the opening angle (θ), due to multiple scattering, scales like: $\sigma(\theta)/\theta \sim (1/E)/(m/E) \sim 1/m$. The fraction of reconstructed A' candidates greater than a given Zv value for four different mass/beam-energy combinations is shown in Figure 5.3.2.6. Interestingly, even the tails of the resolution appear to also scale by 1/m as shown in the right-hand plot of Figure 5.3.2.6. The x-axis of this plot is normalized such that $Z_{norm}=(m(A')/100MeV)\ Z_{cut}$.

In practice, there is much more we can do to clean up the vertex and mass resolution both at the track level (e.g. remove hits that are clearly from scattered beam electrons) and at the vertex level. These will be pursued in the near future.

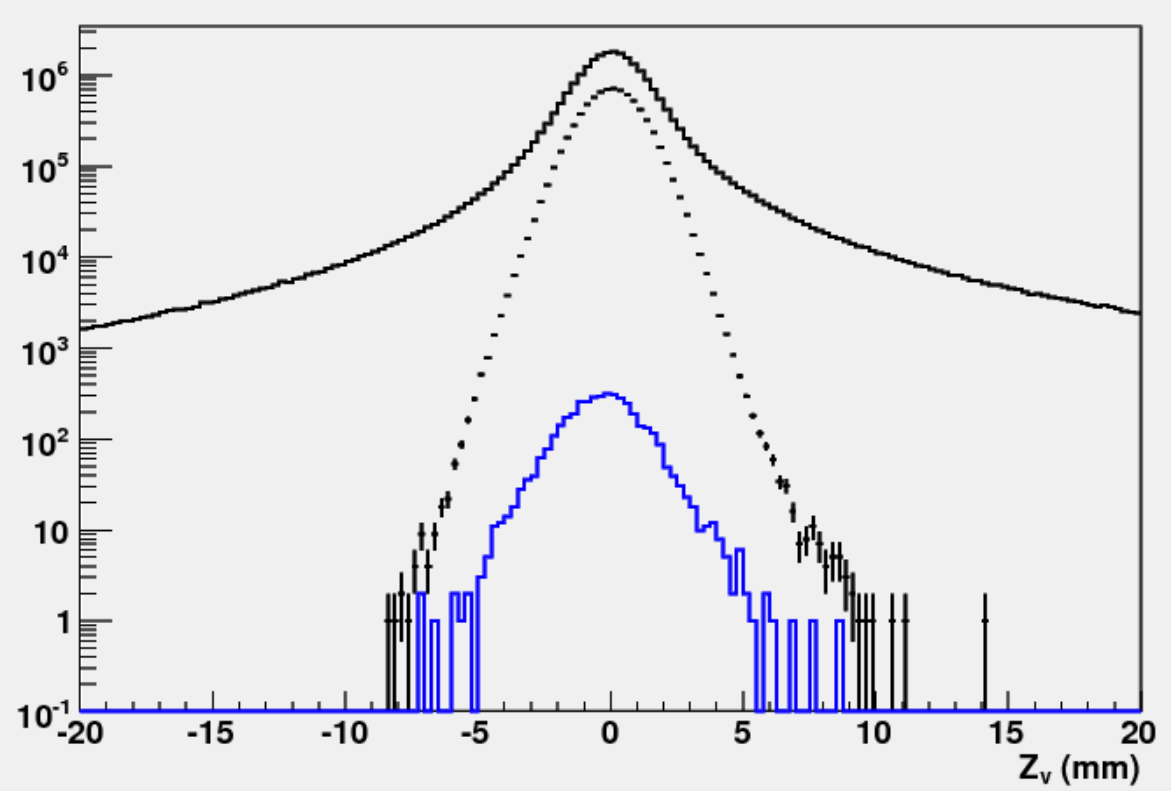

Figure 5.3.2.4: Distribution of the reconstructed vertex position along the beam axis for 6.6GeV 200MeV A' events before (solid black) and after (points with errors) selection. The blue histogram shows the distribution for pairs that have at least one bad hit after selection.

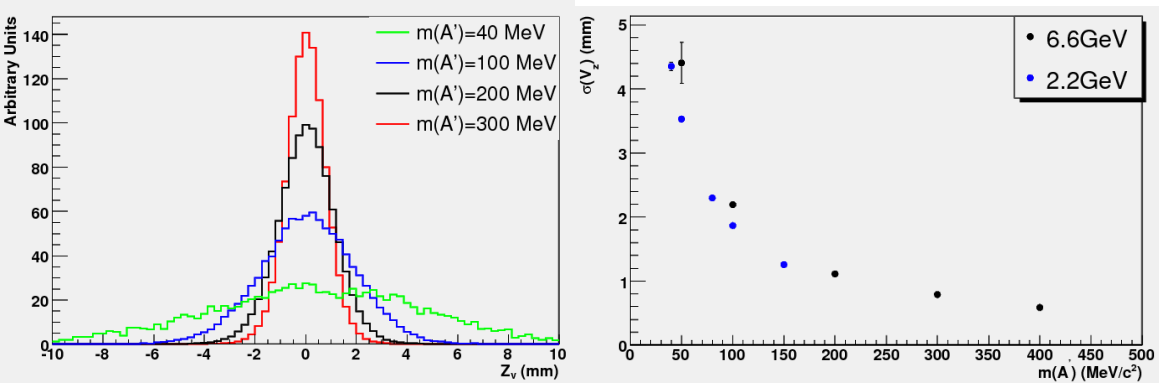

Figure 5.3.2.5: Left: The distribution of reconstructed vertex positions for A' different masses. Right: The (Gaussian) resolution dependence versus A' mass for signal-only events.

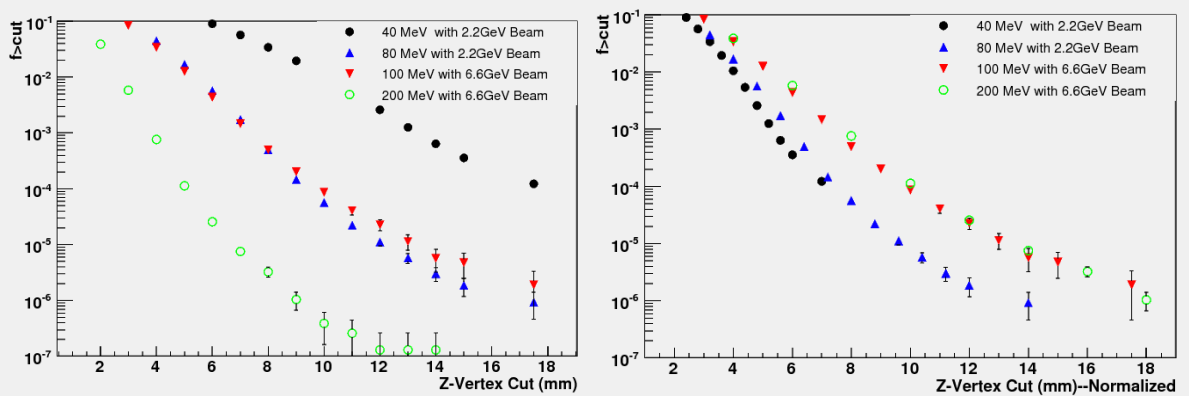

Figure 5.3.2.6: Fraction of A' candidates with reconstructed Z-vertex position greater than the given cut. The x-axis left hand plot shows the actual value of the Z-vertex cut, while the x-axis of the right hand plot is scaled by (m(A')/100MeV) Z_{cut}

5.3.3 References

1. P. Billoir, R. Fruhwirth, and M. Regler, Nucl. Instr. And Meth. A241 (1985) 115.

2. P. Billior and S. Qian, Nucl. Instr. And Meth. A311 (1991) 139.

6 Experimental Reach

The primary search channel for this experiment is, $A' \rightarrow e^+e^-$, with or without a displaced vertex, depending on the magnitude of the coupling α'. As such, the primary irreducible background is QED trident production, with rate given by the diagrams shown in Figure 3.2.1. Trident events can be usefully separated into ``radiative'' diagrams (Figure 3.2.1(a)), and ``Bethe-Heitler'' diagrams (Figure 3.2.1(b)), that are separately gauge-invariant.

The contribution from the radiative diagrams (Figure 3.2.1(a)) alone is a useful guide to the behavior of A' signals at various masses. In particular, the kinematics of A' signal events is identical to that of radiative trident events restricted to an invariant mass window near the A' mass. Moreover, the rate of the A' signal is simply related to the radiative trident cross-section within a small mass window of width $\delta m_{A'}$ by [1],

$$\frac{\mathrm{d}\sigma\left(\mathrm{e^-Z \rightarrow e^-Z\,(A' \rightarrow e^+e^-)}\right)}{\mathrm{d}\sigma\left(\mathrm{e^-Z \rightarrow e^-Z\,(\gamma^* \rightarrow e^+e^-)}\right)} = \left(\frac{3\pi\epsilon^2}{2\,N_{eff}\,\alpha}\right)\left(\frac{m_{A'}}{\delta m_{A'}}\right) \qquad (6.1)$$

where N_{eff} counts the number of available decay states. A fraction ϵ_{bin} of signal events will have reconstructed masses within the mass window, because of the finite mass resolution (for a $2.5\times\sigma$ mass resolution window, $\epsilon_{bin} = 0.8$). Equation (6.1) corrected for ϵ_{bin} allows us to conveniently express the sensitivity to A' signals in terms of the radiative portion of the total QED trident statistics, which we will do shortly.

The Bethe-Heitler process has a much larger total cross-section than either the signal or the radiative trident backgrounds, but exploiting its different kinematics can significantly reduce it. In particular, the A' carries most of the beam energy (see the discussion in Section 3.2) while the recoiling electron is very soft and scatters to a wide angle. In contrast, the Bethe-Heitler process is not enhanced at high pair energy. Moreover, Bethe-Heitler processes have a forward singularity that strongly favors asymmetric configurations with one energetic, forward electron or positron and the other constituent of the pair much softer. These properties are discussed further in the Appendix of [1]. The geometric acceptance and trigger requirements select the region of phase space where signal is dominated, and the Bethe-Heitler background is smallest, as illustrated by Figure 3.2.2 (it should be emphasized, however, that even in this region the Bethe-Heitler background rate exceeds that of radiative tridents by roughly a factor of 5).

To compute the reach of the HPS experiment, we simulate the production of irreducible trident reactions in the detector. We additionally apply a mock-up of the geometric acceptance for the tracking and of the trigger requirements. In addition, high-statistics Monte Carlo samples at particular invariant masses have been used to estimate the background rejection efficiency for a vertex-based search.

We produce generator-level events using MadGraph and MadEvent [2] to compute the full matrix elements for $e^-Z \to e^-(e^+e^-)Z$ in leading order QED, but neglecting the effect of nuclear excitations on the kinematics in inelastic processes. We use the QED nuclear elastic and inelastic electric form-factors in [3]. The MadEvent code was modified to properly account for the masses of the incoming nucleus and electron in event kinematics.

We use a "reduced-interference" approximation that simplifies our analysis and is much less computationally intensive. In this approximation, we treat the recoiling e^- and the e^- from the produced pair as distinguishable. Furthermore, we separate trident processes into the radiative diagrams (Figure 3.2.1(a)) and the Bethe-Heitler diagrams (Figure 3.2.1(b)), and we calculate the cross-section for both of these diagrams separately. Within the acceptance and signal region for the HPS experiment, the Bethe-Heitler reactions dominate the trident rate by 4:1. We have checked that the "reduced-interference" approximation does not correct the trident cross-section by more than 10% in a representative kinematic region [4].

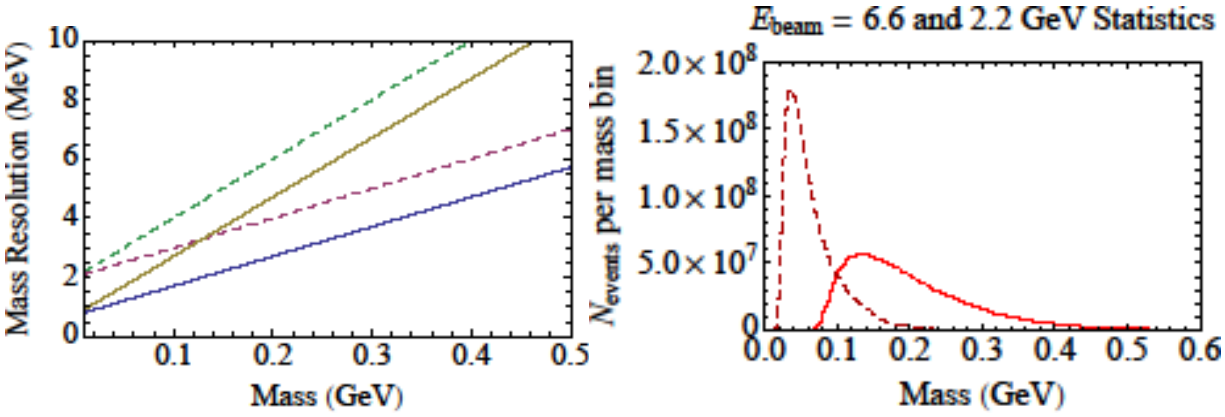

Figure 6.1.1. Left: Mass resolution $\sigma(m_{A'})$ assumed for reach calculations, estimated from momentum resolution achieved in Monte Carlo and expected angular resolution when using a 20μm beam spot constraint. The red (green) dashed line is the resolution used for the vertexing region and the blue (yellow) for the bump-hunt only region for 6.6 GeV (2.2 GeV) beam energy. Right: Distribution of statistics in full resonance search. The curves indicate the number of background QED trident events expected in a resolution-limited mass window of width $\delta m_{A'} = 2.5\,\sigma(m_{A'})$. The red curve corresponds to the distribution of statistics for a 9×10^6s run at 6.6 GeV beam energy, with a current of 450 nA on a 0.25% X_0 target, while the dashed maroon curve corresponds to 2.2 GeV beam energy and a 0.125% X_0 target.

6.1 **Resonance Search**

Equation (6.1) is used to compute the reach for a resonance search in the e^+e^- or $\mu^+\mu^-$ final state. We start by simulating radiative and Bethe-Heitler trident events and require that e^+e^- or $\mu^+\mu^-$ pairs pass the detector acceptance cuts. We additionally require that the total energy exceed 80% or the beam energy and that each track have at least 0.5 GeV of energy. We will refer to these cuts collectively as the "detector/trigger mock-up". We compute the total differential cross section, as a function of invariant mass, for radiative and Bethe-Heitler trident events to pass the detector/trigger mock-up cuts, and from this the final statistics is computed assuming a

run duration of 9×10^6s, beam current of 450 nA (200 nA), and a 0.25% X_0 (0.125% X_0) target at 6.6 GeV (2.2 GeV) beam energy. The assumed mass resolution and the background statistics expected in each resolution-limited mass window are shown in Figure 6.1.1.

To quantify statistical sensitivity, we assume that the continuum background in the resonance search can be modeled by a smoothly varying function and subtracted off. Exclusion power is then determined by the ratio of the signal within an invariant mass window to $\sqrt{N_{bin}}$, where N_{bin} is the total background statistics in the same window.

Using equation (6.1), the sensitivity for a resonance search is determined by

$$\left(\frac{S}{\sqrt{B}}\right)_{bin} = \left(\frac{N_{radiative}}{N_{total}}\right)\sqrt{N_{bin}}\left(\frac{3\pi\epsilon^2}{2\,N_{eff}\,\alpha}\right)\left(\frac{m_{A'}}{\delta m_{A'}}\right)\epsilon_{bin} \qquad (6.2)$$

Here, $\left(\frac{N_{radiative}}{N_{total}}\right)$ is the fraction of radiative reactions among all QED trident events in the search region. This quantity is determined by simulation as described below. N_{bin} is the total number of QED trident events residing in a given invariant mass search bin, and is determined by

$$N_{bin} \equiv \epsilon_{reco}^2 \times \epsilon_{stat}(m_{A'}) \times \sigma_{trigger} \times L.$$

Here L is the integrated luminosity, $\sigma_{trigger}$ is the trigger cross section, $\epsilon_{stat}(m_{A'})$ is the fraction of the total statistics in an invariant mass window centered on $m_{A'}$ of size $\delta m_{A'} = 2.5\,\sigma(m_{A'})$, and $\epsilon_{reco} \cong 0.85$ is the efficiency for reconstructing each track that is within the geometric acceptance of the detector.

6.2 Displaced Vertex and Resonance Search

A search for resonances that decay with cm-scale displaced vertices opens up sensitivity to much smaller couplings than can be observed through a resonance search alone. The vertex reconstruction and quality selection is discussed in Section 5.3. The track and vertex quality cuts retain only about ϵ_{vtx}=70% of both signal and background events. For the purpose of computing reach, the vertex quality requirements reduce the signal efficiency by a factor ϵ_{sigeff} discussed below, which includes this. Likewise we use the high-statistics Monte Carlo studies described in Section 5.3 to model the tails of the vertex distribution for decays at the target. These vertex distributions have been generated for e+e- invariant masses of 100 MeV and 200 MeV with 6.6 GeV beam energy, and 40 MeV and 80 MeV with 2.2 GeV beam energy. The vertex distributions thus obtained scale linearly with invariant mass. Away from these masses, we parameterize the background rejection factor $\epsilon_{rejection}(zcut)$, the fraction of events with a fake vertex beyond a beam line distance of $zcut$, by a smooth interpolation.

In Figure 6.2.1, we show the distribution of fake vertices in the z-direction along the beam line, as well as the distributions for signals corresponding to $\frac{\alpha'}{\alpha} = 10^{-8.5}$ and $10^{-9.5}$, as well as for true muonium (see Section 6.4).

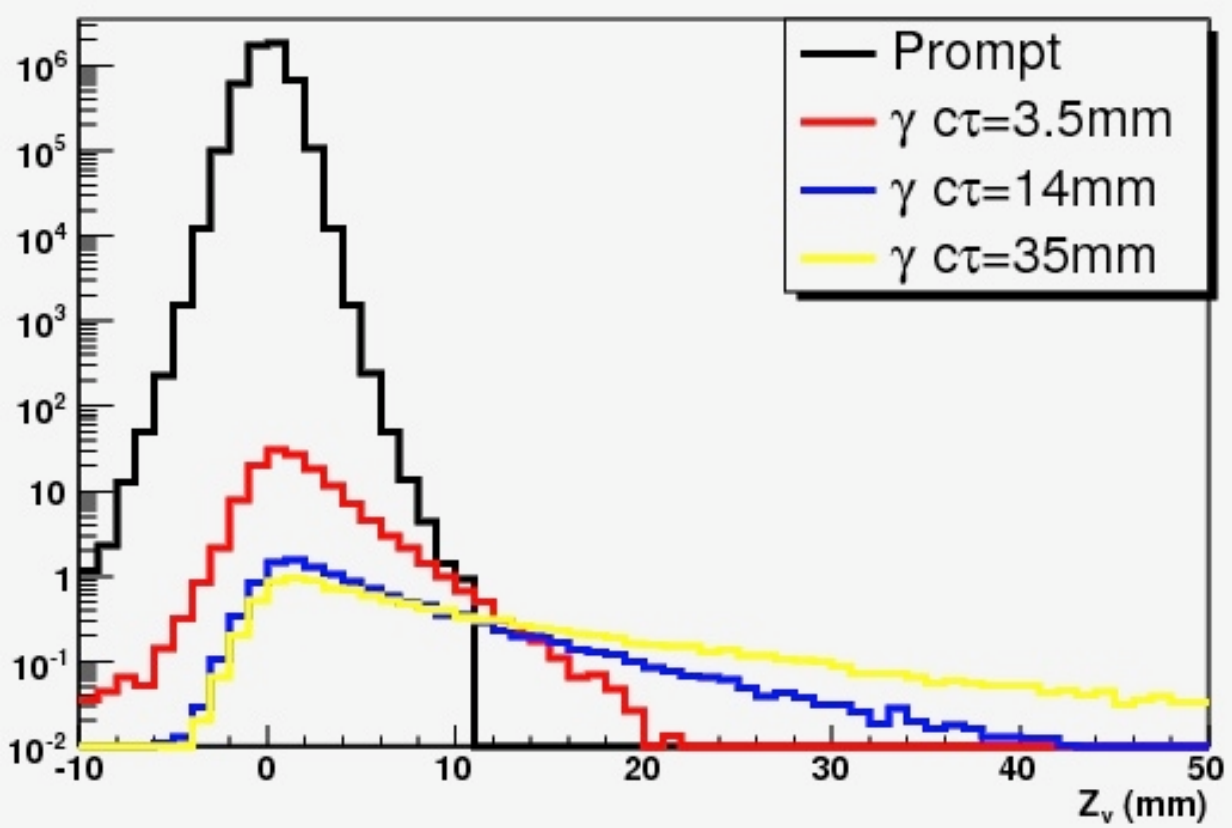

Figure 6.2.1 Signal and background vertex distributions in a resolution-limited invariant mass window, over the 3×10^6s run period at 5.5 GeV (note the different run conditions). The black curve represents the fake vertex distribution from trident events with mass in a $2.5\sigma(m_{A'})$ window about 200 MeV. The red and yellow curves are the vertex distributions for signal events from an A' of the same mass, with $\alpha'/\alpha = 10^{-8.5}$ ($\gamma c\tau = 3.5\,mm$) and $\alpha'/\alpha = 10^{-9.5}$ ($\gamma c\tau = 35\,mm$), respectively. The blue curve corresponds to the rate and vertex distribution expected from $l = 1$ (triplet) state of true muonium, with $\gamma c\tau = 14.2\,mm$ (see Section 6.4).

Because the fake vertex distribution falls quite rapidly, the greatest sensitivity is achieved far on the vertex tail, where less than one background event is expected. For the purpose of computing reach, we have determined a mass-dependent choice of $zcut(m_{A'})$ such that the expected background in each resolution-limited mass window $\delta m_{A'}$, with reconstructed vertices beyond this cut, does not exceed 0.5 events in the 9×10^6s run period. This requires rejection $\epsilon_{rejection}(zcut)$ of background events from the target at the level of 10^{-6} to 10^{-7}, achieved for $zcut \sim 5-30$ mm (see Figure 6.2.2).

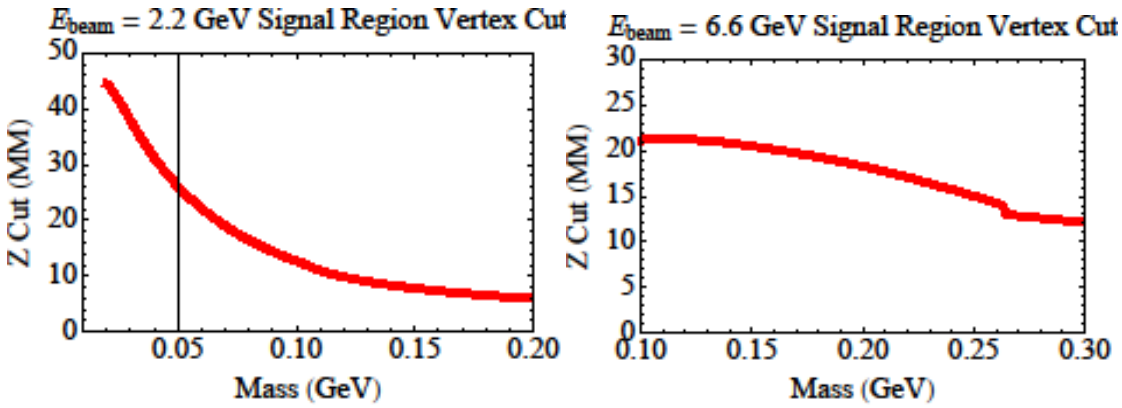

Figure 6.2.2 The minimum vertex displacement zmin (in mm) along the beamline, required for the vertex-based resonance search at 2.2 GeV (left) and 6.6 GeV (right). These are chosen to bring the expected background to 0.5 events in each resolution-limited mass window.

The geometric acceptance falls off at large decay lengths, as shown in Figure 5.2.3. For simplicity we compute reach using the geometric acceptance for z=0, but only considering decays with $z < zmax$=20 cm, so that the fraction of signal events included in the vertex search is

$$\epsilon_{sigeff}(zcut) \cong \epsilon_{vtx} \times \left(e^{-\left(\frac{zcut}{\gamma c\,\tau}\right)} - e^{-\left(\frac{zmax}{\gamma c\,\tau}\right)} \right)$$

Accounting for the reduced acceptance of both signal and background events, the statistical significance expected for a given value of can be computed from that of the pure resonance search as an expected signal can be computed by

$$\left(\frac{S}{\sqrt{B}}\right)_{bin,zcut} = \left(\frac{S}{\sqrt{B}}\right)_{bin} \frac{\epsilon_{sigeff}(zcut)}{\sqrt{\epsilon_{rejection}(zcut)}} > 2, \qquad (6.3)$$

Where $\left(\frac{S}{\sqrt{B}}\right)_{bin}$ is given by (6.2). For the small expected background rate (0.5 events/bin), however, this formula becomes irrelevant, as the exclusion sensitivity of the experiment is limited by the probability of a downward fluctuation in the signal. Thus, for the vertex reach contours in Figures 6.3.1 and 6.3.2, we additionally require an expected signal

$$S_{bin,zcut} = \left(\frac{N_{radiative}}{N_{total}}\right) N_{bin} \left(\frac{3\pi\epsilon^2}{2\,N_{eff}\,\alpha}\right) \left(\frac{m_{A'}}{\delta m_{A'}}\right) \epsilon_{sigeff}(zcut) > 2.4 \text{ events} \qquad (6.4)$$

6.3 Results for Reach Calculations

In this section, we present the expected reach of the HPS experiment, in both a full resonance search and the vertexing resonance search. We also discuss the impact on reach of further rejection of tracks with bad hits, which presently dominate the tails of the background vertex distribution.

Figure 6.3.1 illustrates the expected reach in two 9×10^6s run periods, one with 6.6 GeV beam energy and 450 nA on a 0.25% X_0 target and the other with 2.2 GeV 200 nA on a 0.125% X_0 target. It also shows the regions excluded by previous experiments. The upper solid curves are the lower-limits of 2σ sensitivity for the full resonance search, while the lower solid contours are the outer limit of sensitivity for the vertex-based resonance search, corresponding to 2.4 events in a resolution-limited mass window, where the vertex requirement has been chosen so that 0.5 background events are expected. The dashed curves correspond to the 5σ sensitivities.

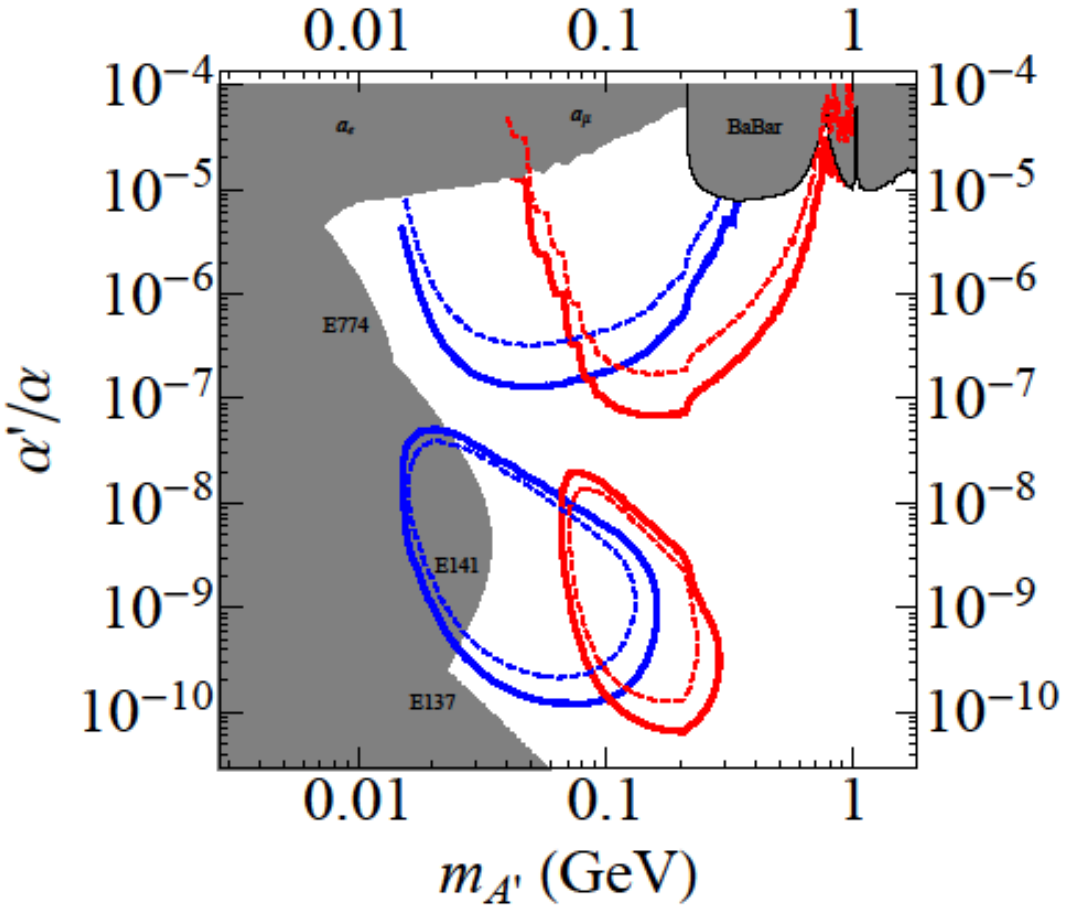

Figure 6.3.1 Anticipated reach in α'/α for the Heavy Photon Search (HPS) experiment, with existing constraints. The solid (dashed) curves in the upper region of the plot correspond to the 2σ (5σ) sensitivity of a full resonance search with 450nA, 6.6 GeV beam and a 0.25% X_0 target (red), along with a preliminary estimate of sensitivity for a 200nA, 2.2 GeV beam and a 0.125% X_0 target (blue). In each case a 9×10^6s run period is assumed. The lower contours in the same colors denote the sensitivity of a vertex-based resonance search. The resultant reach of the combined 2.2 and 6.6 GeV runs is given in Fig. 1.1.

In Figure 6.3.2, we show the sensitivity of the vertex reach as a function of ε for three fixed masses, which represents a one-dimensional vertical slice through the lower contours in Figure 6.3.1. This plot shows that the significance of the reach is highest towards the center of the contour, and drops off for higher and lower values of ε. This illustrates why the contour are closed curves.

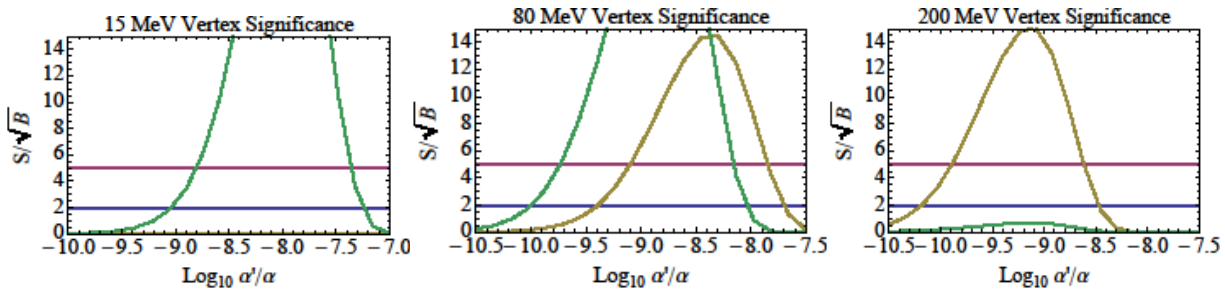

Figure 6.3.2 The sensitivity of the vertex reach as a function of ε for a fixed mass of (left-to-right) $m_{A'}$ = 15, 80 and 200 MeV, which represent a one-dimensional vertical slice through the lower contours in Figure 6.3.1. The green (yellow) curves correspond to the 2.2 GeV (6.6 GeV) contour with the same assumptions as in Figure 6.3.1. The blue (purple) line corresponds to a significance of S/√B of 2 (5), which corresponds to about 2.4 (7) events in a resolution-limited mass window, where the vertex requirement has been chosen so that 0.5 background events are expected.

6.4 **Potential for Discovering True Muonium**

The proposed experiment has the potential to discover ``true muonium", a bound state of a $\mu^+\mu^-$ pair, denoted here by $(\mu^+\mu^-)$. In this section, we review the motivation for such a search and discuss the experimental design considerations.

Positronium and muonium, bound states of (e^+e^-) and (μ^+e^-) pairs, respectively, have been produced and studied [5,6,7], but true muonium, or dimuon, has not yet been detected (see e.g. [8-16]). Together with tauonium $(\tau^+\tau^-)$ and tau-muonium $(\tau^\pm\mu^\mp)$, true muonium is the most compact pure QED system. While $(\tau^+\tau^-)$ and$(\tau^\pm\mu^\mp)$ are difficult to detect since the τ has a weak decay that competes with the QED decay, the μ is very long lived so that the decay of true muonium is purely a QED process. The detection of true muonium would be a significant discovery and would constitute a further important test of QED. In addition, it would demonstrate the capability of the HPS experiment to identify rare separated vertex decays.

The $(\mu^+\mu^-)$ ``atom" can be produced by an electron beam incident on a target such as tungsten [8,9], although we will see below that high Z targets are not important. The $(\mu^+\mu^-)$ atom has an infinite number of states with different binding energies; these states are distinguished by different values for the principal quantum number n. The binding energy of these states are E = -1407 eV/n^2, while their production cross-section scales as 1/n^3, so that higher-n states will be more difficult to produce [10]. In addition to states being distinguished by different principal quantum numbers, they are also produced either in a triplet or singlet state with very different decay properties. The singlet states decays to $\gamma\gamma$ with a rest frame lifetime of 0.602 n^3 ps, or 0.181 n^3 mm, while the triplet states decays to e^+e^- with a rest frame lifetime of 1.81 n^3 ps, or 0.543 mm. The triplet state thus decays exactly like an A′, and can be searched for in the same way. Of course, since its mass is known to be $2m_\mu$ within resolution, a precise search window can be specified. We will not consider the singlet state any further in this proposal, focusing exclusively on the triplet state.

The triplet production cross-section is given by [8]

$$\sigma_{triplet} = 1.20\, Z^2 \frac{\alpha^7}{m_\mu^2}\left(1.79 \ln\left(\frac{\mathrm{E_{beam}}}{\mathrm{m_\mu}}\right) - 6.12\right) \quad (6.5)$$

$$\sim 6.2\, Z^2 10^{-41}\ \mathrm{cm}^2 (\text{for } \mathrm{E_{beam}} = 6.6\mathrm{GeV}) \quad (6.6)$$

We see that this cross-section scales like Z^2 (note that this simplified formula predicts that no true muonium states will be produced at the other nominal beam energy of 2.2 GeV, so that we focus exclusively on the 6.6 GeV beam energy). However, the true muonium bound state breaks up very easily inside the target with a dissociation cross-section that is not only very large but also scales like Z^2 [8]

$$\sigma_{diss} \sim 1.3\, Z^2 10^{-23}\ \mathrm{cm}^2 \quad (6.7)$$

This means that only the bound states produced in the last fraction of the target actually make it out of the target before breaking up, and that the total production rate is effectively independent of Z. The effective thickness through which muonium can penetrate is

$$t_b = \frac{1}{N\, \sigma_{diss}} \quad (6.8)$$

where N is the number density of atoms in the target. The choice of target thickness is not an important parameter when designing an experiment to search for true muonium. Target thicknesses larger than t_b do not help to increase the production rate of true muonium pairs, since those created at the beginning simply will not make it out the back end of the target before breaking apart. Note that for tungsten, t_b = 2.2 μm (0.064% r.l.), while for Carbon graphite, t_b = 190 μm (0.01% r.l.).

Using the triplet production and dissociation cross-sections, the total production rate of true muonium states is [8]

$$Rate_{(\mu^+\mu^-)} = 0.021\ \left(1.79 \ln\left(\frac{\mathrm{E_{beam}}}{\mathrm{m_\mu}}\right) - 6.12\right)\ I\ (mA). \quad (6.9)$$

For the nominal running conditions of E_{beam} = 6.6 GeV, 450 nA beam current, 3 months (~7.8×10^6 s) and a single foil, this experiment would produce about 95 n=1 triplet true muonium states. The typical decay length of these states would be 1.7 cm. The optimal search for true muonium follows the approach of Section 6.3, requiring a vertex cut at about 1.5 cm to reject almost all QED background events, then searching for a resonance at 2 m_μ. Accounting for all efficiencies as discussed in 6.2 and 6.3, we would expect to see only about 10 true muonium events (we caution that the acceptance parameterization here is uncertain at the 50% level).

The production of true muonium events can be increased in several ways. While increasing the target thickness does not help (as discussed above), the production rate scales linearly with both the current and the number of target foils (spaced by 2 or more cm so that true muonium decays

between foils). Using 800 nA and 2 or 3 target foils, instead of 450 nA and a single foil, would produce ~340 – 500 events, of which about 35 – 50 would be detectable in a vertex search.

Seeing the $n=2$ state (which has a very large decay length of about 13.7 cm) will be more difficult, as we would produce only 1/8 as many of these states. However, the large decay length would boost the efficiency for detecting separated vertices, and many decays would be virtually background free, so that even a small number may be sufficient for detection

Finally, since much of the true muonium which is produced will in fact dissociate before leaving the target, the resulting collinear, equal-energy μ+μ- pairs provide another excellent signature. These muons should be identifiable in the HPS muon system.

6.5 References

1. J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, New Fixed-Target Experiments to Search for Dark Gauge Forces, Phys. Rev. D80 (2009) 075018, [0906.0580].
2. J. Alwall *et al.*, MadGraph/MadEvent v4: The New Web Generation, JHEP 0709 (2007) 028
3. K. J. Kim and Y.-S. Tsai, Improved Weizsäcker-Williams methos and its application to lepton and W-boson pair productions, Phys. Rev. D8 (1973) 3109
4. We thank M. Freytsis, G. Ovanesyan, and J. Thaler for independently verifying our cross-section calculation and for computing the full-interference trident cross-section for a representative kinematic region.
5. M. Deutsch, Phys. Rev. 82, 455 (1951).
6. J.I. Friedman and V.L. Telegdi, Phys. Rev. 105, 1681 (1957).
7. V.W. Hughes, D.W. McColm, K. Ziock and R. Prepost, Phys. Rev. Lett. 5, 63 (1960).
8. E. Holvik and H.A. Olsen, Creation of Relativistic Fermionium in Collisions of Electrons with Atoms, Phys. Rev. D35 2124 (1987).
9. N. Arteaga-Romero, C. Carimalo and V.G. Serbo, Phys. Rev. A 62, 032501 (2000) [arXiv:hep-ph/0001278].
10. S.J. Brodsky and R.F. Lebed, Production of the Smallest QED Atom: True Muonium ($\mu^+\mu^-$), Phys. Rev. Lett. 102 , 213401 (2009), arXiv: hep-ph/0904.2225.
11. S. Bilen'kii, N. van Hieu, L. Nemenov, and F. Tkebuchava, Sov. J. Nucl. Phys., 10, 469 (1969).
12. V.W. Hughes and B. Maglic, Bull. Am. Phys. Soc. 16, 65 (1971).
13. J. Malenfant, Phys. Rev. D 36, 863 (1987).

14. S.G. Karshenboim, U.D. Jentschura, V.G. Ivanov and G. Soff, Phys. Lett. B 424, 397 (1998).

15. D.A. Owen and W.W. Repko, Phys. Rev. A 5, 1570 (1972).

16. U.D. Jentschura, G. Soff, V.G. Ivanov and S.G. Karshenboim, arXiv:hep-ph/9706401; Phys. Rev. A 56, 4483 (1997) [arXiv:physics/9706026]; S.G. Karshenboim, V.G. Ivanov, U.D. Jentschura and G. Soff, J. Exp. Theor. Phys. 86, 226 (1998) [Zh. Eksp. Teor. Fiz. 113, 409 (1998)].

7 Test Run for HPS

The projected sensitivities depend critically upon simulations of the background environment and detector performance that have been used to optimize the setup and develop the trigger and reconstruction strategies. While there has generally been good agreement among the background estimates of the various simulation tools, there are some key instances of significant disagreement. For example, the flux of bremsstrahlung photons within the HPS angular acceptance generated by EGS5 is roughly a factor of 5-10 higher than GEANT4 predicts, as shown in Figure 7.1. Conversely, GEANT4 predicts a broader angular distribution of scattered electrons than EGS5, generating twice the occupancy in the tracker near the dead zone, as shown in Figure 7.1. Similarly, the trigger and reconstruction simulations implement idealized detector models that cannot account for every conceivable effect that might be encountered in an actual experiment. For these reasons, we believe it will be important to operate the key detector systems under conditions similar to those required for the experiment to gain valuable operational experience and confirm background estimates. Toward that end, we are proposing a test run including a target, prototype tracking detectors and a portion of the electromagnetic calorimeter in order to study trigger rates, backgrounds and tracking issues at extreme occupancies. Of similar importance, such a test will allow us to gain experience operating in the beam at JLab; firsthand knowledge that will be invaluable in preparing for an extended run of the full experiment.

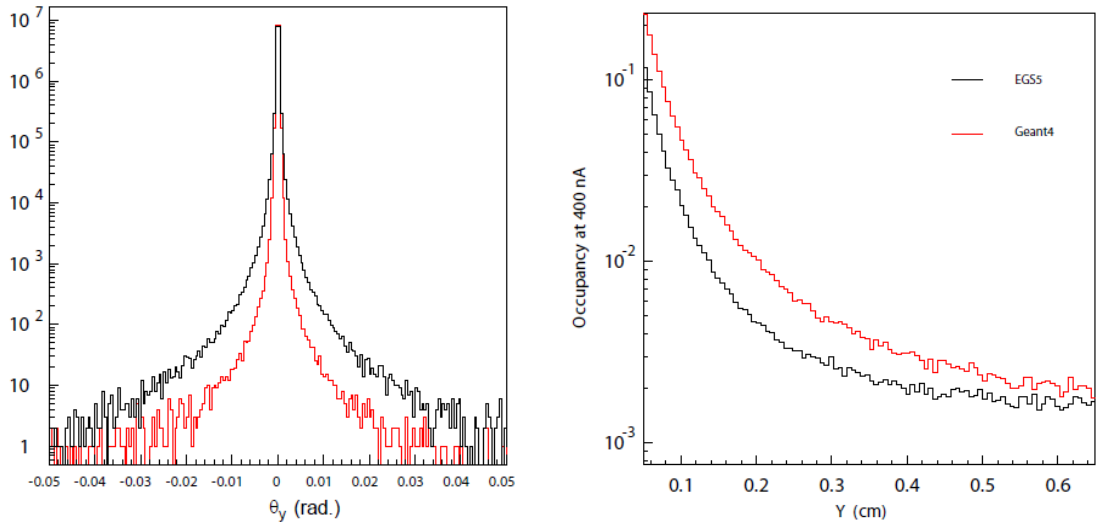

Figure 7.1 LEFT: Angular distribution of bremsstrahlung photons predicted by GEANT4 (red) and EGS5 (black) RIGHT: Occupancy in non-bend plane at Layer 1 predicted by GEANT4 (red) and EGS5 (black).

There are several critical questions to be resolved during a test run. First, we need to establish rates and occupancies in the calorimeter and silicon tracker in a controlled environment that can be used to validate and refine our simulations. Second, we need to establish the viability of our calorimeter-based trigger algorithms in the face of backgrounds that are modeled inconsistently in various Monte Carlo simulations and about which little is known from previous experiments. Third, we need to demonstrate that we can extract precise timing information and track cleanly

in the high occupancies close to the dead zone, which are similarly uncertain. Finally, we need to test the entire DAQ system that combines JLAB (ECal and trigger) and SLAC (Si-tracker) front-end readout systems under realistic conditions to identify any obstacles to successful integration of the two systems. The apparatus we are proposing for the test run is designed to address all of these issues and is outlined below.

Beamline and Target

We intend to use the existing three magnet chicane in Hall-B from the E-07-005 (Two-Photon-Exchange) experiment, located upstream from the CLAS detector as shown in Figure 7.2. As the second magnet in the chicane, the Hall-B pair spectrometer magnet, a standard 18D38, is operated as a 0.5T analyzing magnet. The outer magnets of the chicane are a pair of "*Frascati*" magnets. A target and four planes of silicon tracking will be located inside an existing vacuum chamber, with the ECal positioned directly downstream. This setup utilizes existing magnets, stands, magnet power supplies, power leads, and beam-line monitoring systems.

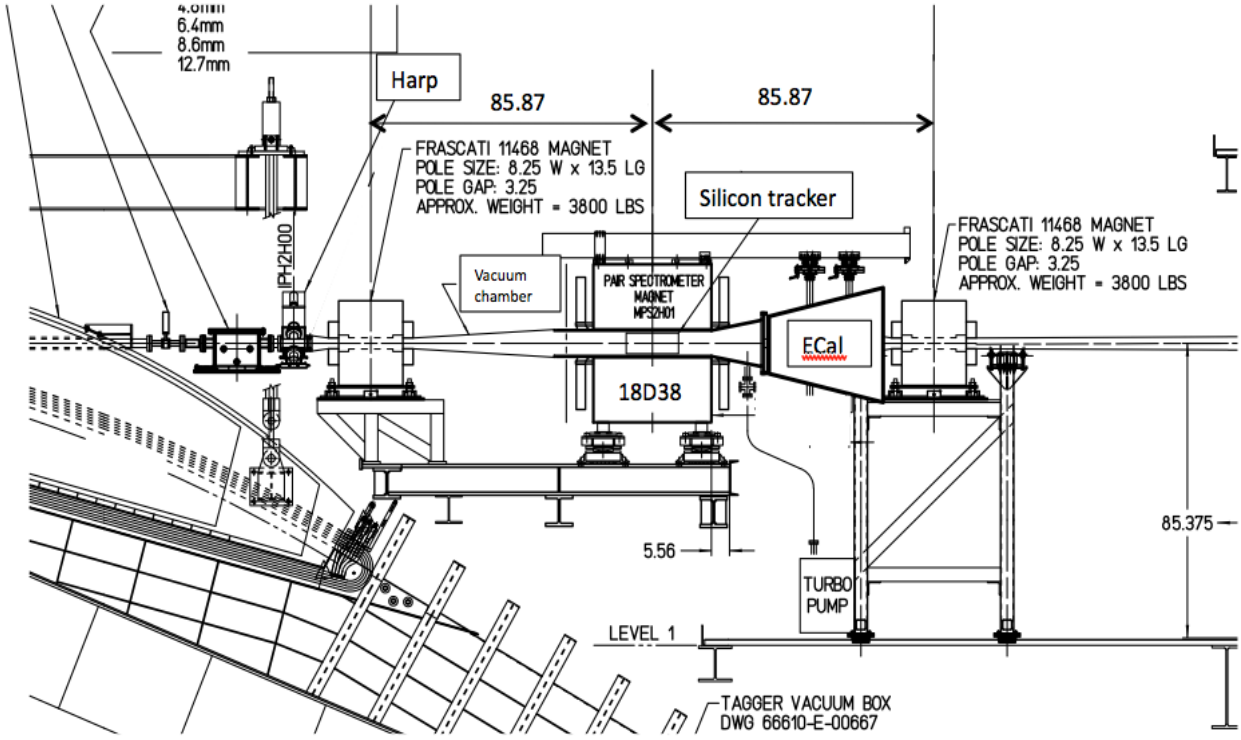

Figure 7.2 Test setup layout in Hall B

To configure the chicane for the test run a new larger diameter beam pipe will be needed upstream of the analyzing magnet, to be connected to the vacuum beamline at the wire harp. Modifications to the upstream and downstream flanges of the scattering chamber in the analyzing magnet will be needed in order to accommodate connections to the calorimeter enclosure and upstream beamline as well as for passing electrical and cooling lines through to the tracking system. The analyzing magnet will be raised by about 3 inches to open up a space

above beam plane for silicon tracker, which will only instrument the region above the beam due to space constraints. A pair of prototype targets consisting of 0.125% and 0.25% tungsten, selectable via motion of the target stand, will be installed in the vacuum chamber with the tracker. By using a larger beamspot and lower beam intensity, we will eliminate the need for the target motion system required for the full experiment.

Silicon tracker

The four layers of the tracker include two layers each of the vertexing and pattern recognition sections to allow tracking of particles that also trigger the electromagnetic calorimeter. The tracking planes instrument only the upper side of the dead zone due to space constraints in the vacuum chamber and only the central portion of the bend-plane, where occupancies are the highest, to minimize costs. The modules will be prototypes utilizing readout electronics and support structures as similar as feasible to those being proposed for the full experiment. However, the modules will not be capable of individual motion nor will they be mounted in a carbon fiber support box to reduce the cost and complexity of the setup. Instead, the modules will be mounted to cooling blocks on an aluminum plate that will be moved vertically to adjust the position of the entire tracker relative to the beam. The instrumented region includes portions of layers 1, 3, 4 and 5 as outlined in Table 7.1. Studies indicate that tracking at full occupancies is possible with this configuration, albeit with a higher rate of pattern recognition failures and a commensurate reduction in the rejection of prompt vertexes, deficiencies that are of no concern for the test run. Although most vertexes will not be reconstructed with coverage on only one side of the dead zone, studies of impact parameter resolution will suffice to establish the pattern recognition and vertexing performance predicted by Monte Carlo with high statistics.

	Layer 1	Layer 2	Layer 3	Layer 4
$\approx$ position, from target (cm)	10	30	50	70
Stereo Angle	90 deg.	90 deg.	50 mrad	50 mrad
Bend Plane Resolution (m)	$\approx$ 6.5	$\approx$ 6.5	$\approx$ 6.5	$\approx$ 6.5
Stereo Resolution (m)	$\approx$ 6.5	$\approx$ 6.5	$\approx$ 130	$\approx$ 130
# Bend Plane Sensors	2	2	3	5
# Stereo Sensors	1	1	3	5
Dead Zone (mm)	<1.5	<4.5	<7.5	<10.5
Power Consumption (W)	5.25	5.25	10.5	17.5

Table 7.1: Key parameters of the sensor layout for the tracking and vertexing system of the test run. These are abbreviated versions of Layers 1, 3, 4 and 5 of the full experiment and provide rudimentary tracking capabilities.

Electromagnetic Calorimeter

The CLAS Inner Calorimeter will be reconfigured to provide the electromagnetic calorimeter for the test run. The 460 lead-tungstate crystals, read out by avalanche photodiodes, will form the inner, highest-occupancy, portion of the full HPS calorimeter. Because the trigger algorithm depends on coverage for both legs of trident decays, it is critical that the layout cover both sides of the dead zone and both the electron and positron sides of the primary beam. The acceptance below the dead zone will be truncated due to the placement of the beam close to the bottom of the magnet bore, as dictated by the tracker, but will still be sufficient to allow for critical trigger studies, as shown by the distribution of radiative decay products shown in Figure 7.3.

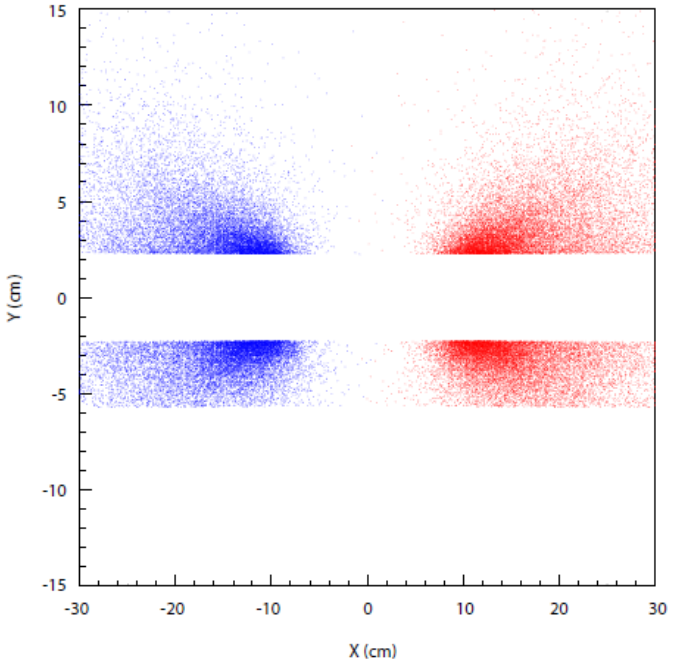

Figure 7.3 The distribution of radiative trident products at the face of ECal for the test run. The distribution below the bend plane is truncated by the wall of the vacuum chamber.

DAQ and trigger electronics

The DAQ and trigger electronics for the test run will be prototypes of the system proposed for the full experiment. Operation of as much as possible of the final DAQ will be critical in identifying any inter-system incompatibilities and exercising the full readout chain as much as for triggering the tracker, which will record tracks in abundance even with random triggers.

8 Run Plan

The search for heavy photons provides an exciting and topical research program for Jefferson Laboratory. The Heavy Photon Search experiment has the potential to make game-changing discoveries. To take advantage of this opportunity, Jefferson Laboratory needs to pursue this research aggressively. We hope to move rapidly to designing, building, commissioning, and running the HPS experiment. Our plan is to follow a staged approach. Stage I is a test run, in which we plan to test critical experimental assumptions as soon as possible, ideally before the 2012 shutdown, by installing a test apparatus in Hall B as described in the previous section. This apparatus will measure critical occupancies in the tracker and trigger rates in the electromagnetic calorimeter, and demonstrate the high rate capability of the silicon vertexer/tracker and electromagnetic calorimeter data acquisition systems. Doing this as early as possible, and preferentially before the 2012 downtime, would allow us to incorporate what we learn and address any problems or surprises which we uncover in our final design. If running before the 2012 downtime is impossible, we would hope to schedule the test run as early as possible in the 12 GeV era.

Stage II is the experimental run proper, which we hope will occur early in the 12 GeV era. The first step will be installing and commissioning the full detector and demonstrating the necessary beam optics and beam control. Then we proceed to data taking runs at 2.2 and 6.6 GeV. In all, we want to take data for 3 months (2160 hours) at each energy, which will extend the search for a heavy photon into the large and unexplored region in the mass/coupling plane favored by the theory. In addition to extending the two sigma exclusion area very significantly, a multi-month run allows good coverage of the parameter space for the 5 sigma discovery region. HPS can see or exclude a heavy photon signal over a very wide range of masses and couplings. The run at 6.6 GeV extends our reach to higher masses, and should be adequate to discover the true muonium signal as well.

Additional follow-on running can be imagined, but it is too early to outline a definite plan. Studies of true muonium will probably benefit from dedicated running at the highest available energies, and may require innovative target designs. Extending the heavy photon search to yet higher masses will require higher energy running and likely need the development of a pion trigger.

The successful execution of this run plan depends on rapid approval and funding for the test run, and timely funding for the experiment proper.

8.1 HPS Test Run

In the 6 GeV era, reconfiguration of the beamline and magnet chicane and installation of the tracker and ECal will take approximately 2 weeks. The test run will require ~7 days for

commissioning and will take data for 7 days at 2.2 GeV beam energy to provide the broadest possible inputs to our simulations for the full experiment.The time required for preparation of a test run in the 12 GeV era will be longer since in this case partial installation of all beam line elements downstream of CLAS will be needed. The test run can then take place as part of the commissioning for the full experiment.

8.2 **HPS Installation, Checkout, and Physics Runs in the 12 GeV Era**

In the 12 GeV era, installation of the beam line elements, analyzing magnet, tracker, electromagnetic calorimeter, and muon system will take 3-4 weeks. For definiteness, we have considered installation in Hall B.

HPS will need 2-3 weeks for commissioning, which will include delivering and monitoring small, stable beams at 2.2 and 6.6 GeV, aligning the detector elements, testing trigger rates and tracker occupancies and exercising the data acquisition system. HPS will then be ready for its major data taking.

For data taking, run time will be divided between two energies, 2.2 GeV and 6.6 GeV incident electrons. A 3 month run at each energy provides excellent discovery potential for HPS over a wide range of possible heavy photon masses and couplings, and an even larger exclusion region if no signal is discovered. Detailed running conditions, energies, currents, and target thicknesses, are given in the Beam Requirements List, which is appended to this proposal. In sum, HPS requires <30 micron spot sizes, good beam stability (<25 micron), minimal beam tail ($< 10^{-5}$), excellent energy stability ($< 10^{-4}$), and currents in the range 100-600 nA. HPS will utilize thin tungsten target foils, 0.125% and 0.25% X_0.

9 Costs, Schedule

Cost estimates for engineering, designing, fabricating, assembling, testing, and installing the Heavy Photon Search experiment in the Hall B at JLab are given below. The costs have been divided into those associated with the Test Run and those associated with the HPS Experiment Proper. The costs reflect considerable savings coming from the donation of about 160 silicon microstrip sensors from Fermilab, the use of a refurbished analyzing magnet from SLAC, and many contributions from JLab, including $PbWO_4$ calorimeter crystals, chicane magnets, and magnet power supplies. Much of the calorimeter readout electronics utilizes designs which are already in place for the Hall B 12 GeV upgrade, eliminating engineering and design expense. Very significant cost savings come from utilizing the FADCs and data acquisition system being developed for the upgraded CLAS detector. Some of the modules needed for CLAS can be shared with HPS during HPS data taking runs. Collaborating institutions will donate engineering and design efforts for the ECal and Muon systems, affording additional savings.

The costs are given in an accompanying WBS summary table, below, which itemizes the major items subsystem by subsystem, and indicates whether JLab (J) or SLAC (S) takes responsibility for construction. Engineering, design, and technician labor rates include lab overheads, and differ between the two laboratories. Contingencies have been set at 30% for labor and 35% for M&S at SLAC, and somewhat lower at JLab, since full engineering designs are not yet available for the SLAC items, and many of the JLab items are similar to items recently constructed. The contingency for commercial items is generally about 10%. No overheads have been applied to M&S at either laboratory. A 10% overhead on the M&S would add roughly $150k to the total cost of the project.

9.1 Test Run Costs

Beamline expenses for the Test Run are held to a minimum by using the 18D36 magnet currently installed in Hall B as the analyzing magnet, thereby using much of the existing beamline, and only adding chicane magnets and beam pipe before the CLAS detector. Minor engineering and design will be required, and a test run beam pipe fabricated. A thin target must be provided as well. Total beamline expenses are about $20k.

Testing occupancies in the silicon tracker requires development and construction of prototype sensor modules, a simple sensor module support system, some modifications of the existing vacuum chamber to allow cable and coolant connections, and power supplies and monitoring. SVT Electronics and DAQ are also needed and are costed below. The tracker/vertexer for the test run will cost about $150k.

The SVT readout requires hybrid and readout board engineering and prototyping, APV25 and chip procurement, fabrication, and test. The SVT DAQ requires designing and prototyping the T1 ACTA. SLAC will provide the ATCA crate, CPU, and Server. These components permit tests

of the entire triggering chain and commissioning of the high rate data acquisition system. The expenses are dominated by engineering development, and total $184k.

JLab will donate the PbWO4 crystals used in the electromagnetic calorimeter, but will need to engineer, design, and fabricate a new enclosure for the crystals, new readout boards, and vacuum chambers for beam passage. Much of the design work will be donated by collaborators. The total expense will be roughly $140k, including fabrication and assembly costs. Reconfiguring the ecal is critical to HPS, to provide tests of the triggering rate and occupancies, and to allow beam transport in vacuum.

Trigger and DAQ electronics for the ECAL are being developed for the CLAS upgrade, so relatively little engineering and technician time will be needed in preparation of the HPS Test Run. Components, including the 250 MHz FADC boards, will be provided at no cost since they can be borrowed from the CLAS upgrade. The system test expenses will also be borne by JLab Hall B. The costs total $32k.

Altogether the Test Run will cost $523k. The test will provide valuable experience with beam control issues, and measure the background and trigger rates for the HPS experiment. The test run will also shakedown the proposed high rate readouts for both the SVT and Ecal, and provide an invaluable opportunity to integrate and debug the DAQ.

9.2 HPS Experiment Costs

The total HPS experiment costs are given below. They are significantly reduced by virtue of the engineering and prototyping expenses already assigned to the Test Run. Some of the items needed for the test run will serve in the experiment proper, e.g., the ecal enclosure and part of the ecal readout electronics expense has already been accounted. Other items are unique to the full experiment. The final beamline and magnet expenses and the muon system expenses are examples. Here are the costs, subsystem by subsystem.

The final beamline will require engineering and design, in addition to the assembly of three girders, which hold the quadrupoles, BPMs, corrector magnets, viewers, and harps. The other major expenses include the bending magnet support stands, beam pipes and windows, analyzing and chicane magnets, power supplies, and power leads. Including the labor for installation and contingencies, the beamline will cost about $590k.

Much of the prototyping for the silicon tracker/vertexer will have been carried out for the Test Run, but some engineering and a lot of technician time is required to finalize design, fabricate, and test the sensor modules. The CF support box, final vacuum chamber, and rotatable target must be engineered, designed, and fabricated. In addition, power supplies, a cooling system, and motors and controls must be provided. The remaining expenses for the SVT total about $630k.

Much of the the EM calorimeter will have been constructed for the Test Run, but additional counters and readout must be designed, constructed, and installed. Cables, and power supplies, and monitoring systems complete the list of needed items. We assume engineering and design is done by an HPS collaborator, but include the cost of supporting a visiting physicist at JLab to do this work. The cost, including technician time for calorimeter assembly and testing, is $90k.

The muon system costs include iron absorber; the beam transport vacuum chamber; scintillator, wavelength shifting fibers, and PMTs; support stands; and cables and connectors. Again we assume engineering and design is done at JLab by a visiting physicist, but include technician labor for system fabrication. The total is about $180k.

This experiment relies on very fast data transmission and high trigger rates. JLab systems will provide the calorimeter and muon system readout and trigger, and define the overall DAQ architecture. A SLAC system will readout the silicon tracker/vertexer at ~40 MHz, select triggered events, and interface to the overall data stream. Beyond what was required for the Test Run, the SLAC SVT system will cost about $340k. The additional JLab ECal and trigger elements will cost about $55k, and JLab will absorb the overall system DAQ and data storage costs with its existing infrastructure.

Altogether, the HPS experiment will cost approximately $1.88M, including Lab overheads on labor and contingencies. That estimate assumes that collaborating institutions will help engineer and design the calorimeter and muon systems, and it reflects considerable savings arising from using existing apparatus and donations of major components.

The costs associated with storing and accessing the data and with data analysis expenses at collaborating institutions are not included.

9.3 **Timetable**

Our goal is to be ready to test critical features of the experiment and its apparatus in a Test Run which would take place Spring 2012, before the down for the 12 GeV Upgrade. This will require approval and funding for the Test Run in the near future. Following the test run, we move to a final design and construction phase for the experiment proper. Having funding in place by early FY13 can ensure that the experiment will be ready for beam time early in the 12 GeV era.