Part II

The BTeV Detector

Chapter 4

The BTeV Baseline Detector

In this chapter, we describe the BTeV detector baseline design. We begin with a review of the important characteristics of hadronic B production and the properties of the Tevatron which influence the design. We then give a description of the C0 Collision Hall and IR. We discuss each major component of the detector in turn, including the key requirements that drive the design, the choice of the baseline technology, the expected performance, and the options which are still under consideration. More details on each major component are presented in later chapters.

4.1 Characteristics of Hadronic b Production

The kinematics of hadronic B and charm production play a major role in the design of BTeV. We review the most important features here.

It is customary to characterize heavy quark production in hadron collisions with the two variables p_t and η. The latter variable was first invented by those who studied high energy cosmic rays and is assigned the value

$$\eta = -ln\left(\tan\left(\theta/2\right)\right), \tag{4.1}$$

where θ is the angle of the particle with respect to the beam direction.

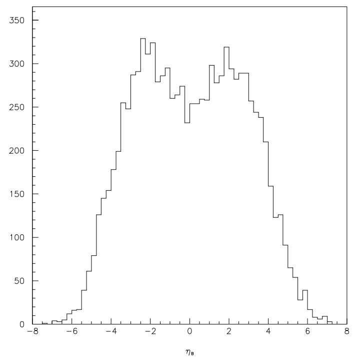


Figure 4.1: The B yield versus η.

According to QCD calculations of b quark production, the b's are produced "uniformly" in η and have a truncated transverse momentum, p_t, spectrum, characterized by a mean value approximately equal to the B mass [1]. The distribution in η is shown in Fig. 4.1.

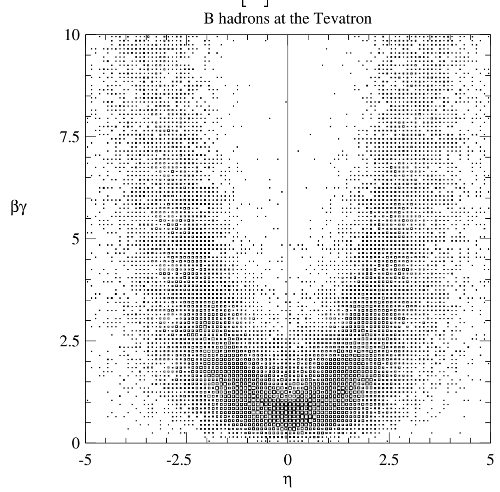


Figure 4.2: $\beta\gamma$ of the B versus η.

There is a strong correlation between the B momentum and η. Shown in Fig. 4.2 is the $\beta\gamma$ of the B hadron versus η from Pythia at $\sqrt{s} = 2$ TeV. It can clearly be seen that near η of zero, $\beta\gamma \approx 1$, while at larger values of $|\eta|$, $\beta\gamma$ can easily reach values of 6. This is important because the observed decay length varies with $\beta\gamma$ and furthermore the absolute momenta of the decay products are larger allowing for a suppression of the multiple scattering error.

Since the detector design is somewhat dependent on the Monte Carlo generated b production distributions, it is important to check that the correlations between the b and the $\overline{b}$ are adequately reproduced. Fig. 4.3 shows the azimuthal opening angle distribution between a muon from a b quark decay and the $\bar{b}$ jet as measured by CDF [2] and compares it with the MNR predictions [3].

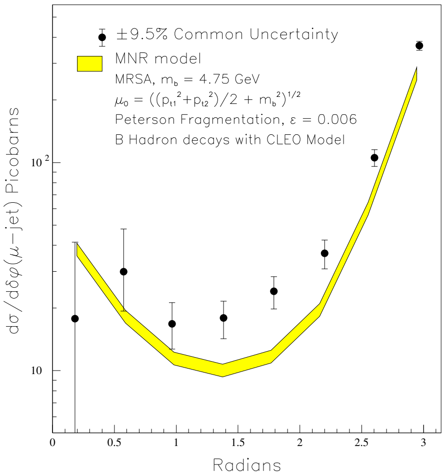


Figure 4.3: The differential $\delta\phi$ cross-sections for $p_T^\mu > 9$ GeV/c, $|\eta^\mu|$ <0.6, $\mathrm{E}_T^{\bar{b}}$ >10 GeV, $\left|\eta^{\bar{b}}\right| < 1.5$ compared with theoretical predictions. The data points have a common systematic uncertainty of $\pm 9.5\%$. The uncertainty in the theory curve arises from the error on the muonic branching ratio and the uncertainty in the fragmentation model.

The model does a good job representing the shape, which shows a strong back-to-back correlation. The normalization is about a factor of two higher in the data than the theory, which is generally true of CDF b cross-section measurements [4]. In hadron colliders all B species, B^o, B^+, B_s^o, b-baryons, and even B_c mesons, are produced at the same time.

The "flat" η distribution hides an important correlation of $b\bar{b}$ production at hadronic colliders. In Fig. 4.4 the production angle of the hadron containing the b quark is plotted versus the production angle of the hadron containing the $\bar{b}$ quark. Here zero degrees represents the direction of the incident proton and 180 degrees, the incident anti-proton. There is a very strong correlation in the proton or the anti-proton directions: when the B is forward the $\overline{B}$ is also forward. (We call both the proton and anti-proton directions forward.) This correlation between B and $\overline{B}$ production is not present in the central region (near 90 degrees). By instrumenting a relatively small region of angular phase space, a large number of $b\bar{b}$ pairs can be detected. Furthermore, the B's populating the forward regions have large values of $\beta\gamma$ and this permits more accurate separation of the primary and secondary vertices and better overall detection efficiency for b hadrons. In addition, the techniques for particle identification in this kinematic range are very well developed. Powerful particle identification is a key to high sensitivity bottom and charm experiments. For these reasons, we have designed a detector with 'forward coverage'.

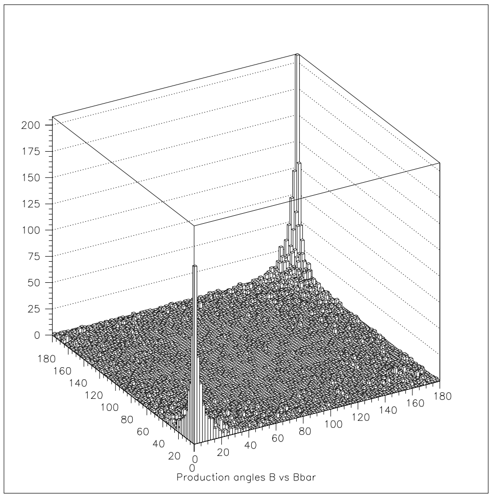


Figure 4.4: The production angle (in degrees) for the hadron containing a b quark plotted versus the production angle for a hadron containing $\bar{b}$ quark, from the Phythia Monte Carlo generator.

Charm production is similar to b production but has a much larger cross section. Current theoretical estimates are that charm is 1-2% of the total $p\bar{p}$ cross-section. The cross section is even more strongly peaked in the forward directions because the average transverse momentum is of the order of only 1.5 GeV/c. The charm cross section has never been measured because texperiments with good acceptance in the central region have very low efficiency for triggering and reconstructing charm. The favorable kinematics in the forward direction gives BTeV a very high efficiency for reconstructing charm.

Table 4.1 gives the relevant Tevatron parameters. We expect to start serious data taking in Fermilab Run II with a luminosity of about $5 \times 10^{31}\mathrm{cm}^{-2}\mathrm{s}^{-1}$; our ultimate luminosity goal, to be obtained in Run III, is $2 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

While BTeV is designed to study $p\bar{p}$ collisions at 2 TeV, it can also run in 'Fixed Target' mode. In that mode, a thin wire is inserted into the halo of the circulating beam and the detector studies the collisions of 800 GeV/c protons (or anti-protons) with the nuclei of the wire. For this reason, Fixed Target mode is also referred to as 'wire mode'.

In this mode, particles diffusing laterally out of the beam eventually reach the wire target where they can interact. Since the diffusion is very slow, each particle will traverse the wire many times until it either interacts or multiple scattering blows its orbit up so badly that it escapes from the machine aperture altogether. We are learning from accelerator physicists what overall interaction rate we can expect.

The cross section for fixed target charm production on a proton at 800 GeV/c is about 30 μb. While this is much lower than in the collider, there are some advantages: first, it

Table 4.1: The Tevatron as a b and c source for BTeV

Luminosity (BTeV design)	$2 \times 10^{32} \mathrm{cm}^{-2}\mathrm{s}^{-1}$
b cross-section	100μb
# of b's per 10^7 sec (Run II)	4×10^{11}
$\frac{\sigma_{b-\bar{b}}}{\sigma_{inel}}$	0.2%
c cross-section	> 500 μb
Bunch spacing	132 ns
Luminous region length	$\sigma_z = 30$ cm
Luminous region length	σ_x $\sigma_y = \approx 50$ μm
Interactions/crossing	$< 2.0 >$

may not require low-β quadrupoles to be installed so it might be available to us earlier than collider mode; second, it is not expected to affect the luminosity at B0 or D0, which is not the case when the beams collide at C0 and this will translate into longer running times; the region along the beam that has to be instrumented with precision vertex detectors is much smaller which means that a small scale vertex detector prototype is all that is needed for initial experiments; and the fact that the center of mass is boosted towards the spectrometer increases the acceptance. However, this mode is not useful to us for B-physics studies because the cross section is of order 10 nb. While we believe that the most sensitive charm experiment will be using the collider, this mode does give us the opportunity to do physics early with a partially instrumented detector.

Bibliography

[1] M. Artuso, "Experimental Facilities for b-Quark Physics," in *B Decays* revised 2nd Edition, Ed. S. Stone, World Scientific, Singapore (1994).

[2] F. Abe *et al*, *Phys. Rev.* D **53**, 1051 (1996).

[3] M. Mangano, P. Nason and G. Ridolfi, *Nucl. Phys.* B **373**, 295 (1992).

[4] F. Abe *et al*, *Phys. Rev. Lett.* **75**, 1451 (1995). Previous UA1 measurements agreed with the theoretical predictions, see C. Albajar *et al*, *Phys. Lett.* B **256**, 121 (1991). Recent D0 measurements agree with both the CDF measurements and the high side of the theoretically allowed range. See S. Abachi *et al*, *Phys. Rev. Lett.* **74**, 3548 (1995).

4.2 C0 Baseline Detector Description

We describe here the baseline detector for BTeV, which can achieve our physics goals, and is designed to have the capability to perform new measurements on new interesting final states that may turn out to be recognized as important in the future.

The challenges for any future hadron collider B experiment are to achieve high efficiency and low background in the hostile and complex environment of a hadron collider. This requires the effective and aggressive use of the most modern technologies. For BTeV this represents a significant step beyond what will be achieved by the upcoming generation of hadron collider B experiments and e^+e^- collider experiments running on the Υ(4S), and their possible extensions and upgrades. Furthermore, we must be able to compete successfully with the LHC-b experiment, when it starts running.

BTeV has begun to take shape based on an understanding of what needs to be done to meet these challenges. These are the key design features of BTeV:

- a dipole located on the IR, which gives BTeV an effective 'two arm' acceptance;
- a precision vertex detector based on planar pixel arrays;
- a vertex trigger at Level 1 which makes BTeV very efficient for many final states, even if they have no leptons in them; and
- excellent particle identification. Especially important is the requirement of very good charged hadron identification. Many of the states that will be of interest in this phase of bottom and charm physics will only be separable from other states if this capability exists. Also, it will allow for the possibility of kaon tagging. Muon and electron identification are important for tagging and for studies of decay modes involving leptons, too.
- A high quality $PbWO_4$ electromagnetic calorimeter capable of detecting final states with single photons, π^o's, η's or η''s.

The parameters of the Tevatron that must be taken into account in creating a design for meeting these challenges are given in Table 4.1. A schematic of the experiment is given in Fig. 4.5. We now discuss the key elements of the baseline design.

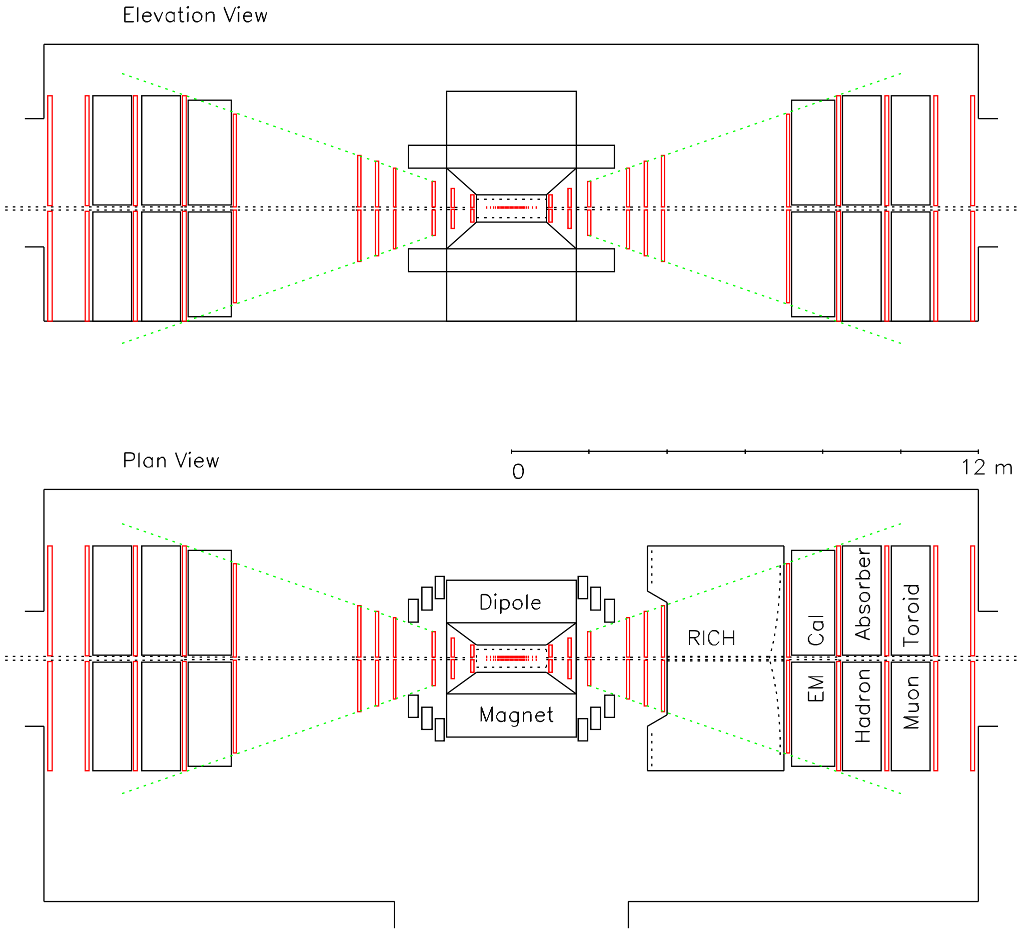


Figure 4.5: Layout of BTeV/C0 Spectrometer

4.2.1 The C0 Experimental Area

The experiment will be carried out in the newly constructed C0 test hall, see Figure 4.6. The test hall is a 216 m^2 (9m wide by 24m long) enclosure centered along the C0 straight section of the Tevatron. (In plan view, the hall is asymmetric transversely, extending 3.5m from the beam on the west side and 5.5m from the beam on the east side.) The Tevatron beam is 2.5m above the floor slab and 4.25m below the roof of the hall. There are flared transition enclosures both upstream and downstream of the test hall. The test hall specifications and dimensions are compatible with the detector described in this document.

To the east of the test hall is the C0 assembly building (see Figure 4.7) a steel framed, industrial type, structure containing a 150m^2 assembly hall at the test hall elevation. Detector components will be constructed in the assembly hall and moved into the test hall through a 6m $\times$ 6m 'shield door'. The C0 assembly building also has space on various levels for the electrical, water and air handling systems needed for the BTeV detector. Figure 4.8 shows a layout of the assembly building and test hall with the magnet and muon toroids of the proposed BTeV detector superimposed.

Detector elements are brought into the area at a ground level loading dock and lowered to the assembly floor using the 30 ton crane which covers the loading dock and assembly hall. There is no crane in the beam enclosure; all detector elements must be designed so they can be constructed or staged in the assembly area and moved through the shielding door into the enclosure. During operation the shielding door is sealed with concrete blocks for radiation protection. The blocks are stored in the south end of the assembly area in an alcove when the door is open.

There are cable ducts from the experiment enclosure to a 150m^2 equipment room at grade level on the north end of the assembly building. Preliminary engineering has been done on converting this equipment room into a two or three level electronics/counting room and office area for experimenters. Electric power, air handling, other utilities, as well as an elevator, lavatories, etc., although not yet installed, have been designed to accommodate BTeV needs in these areas.

Figure 4.6: C0 Test Hall

Figure 4.7: Tevatron berm and C0 Assembly Building

Figure 4.8: Layout of C-0 Test Area

4.2.2 The BTeV/C0 Spectrometer Magnet

We have identified a magnet that satisfies the specifications for the BTeV/C0 spectrometer. The SM3 magnet, currently part of the decommissioned Fermilab MEast spectrometer, is shown on the proposed layout, Fig. 4.5, of the BTeV/C0 spectrometer. This magnet was assembled in 1981 from soft iron blocks recovered from the decommissioned Nevis Cyclotron. The coils for the magnet were built of 5 cm square aluminum conductor by the Sumitomo Corporation under the aegis of the US-Japan Agreement on High Energy Physics. The magnet operated in MEast from 1982 until 1997, at a central field of about 0.8 Tesla, serving experiments E605, E772, E789, and E866.

The SM3 magnet was assembled by welding together, in place, various blocks of steel. It has a total weight of 500 metric tons. The construction and assembly drawings for this magnet have been located. Engineering preparations are now underway for disassembling the magnet. The modified magnet could then be reassembled in the C0 assembly hall and rolled into the C0 collision hall, as shown in Fig. 4.6.

The central field specified for the BTeV/C0 spectrometer is much higher than the past operating excitation of the SM3 magnet. Preliminary studies with magnetostatic modeling programs (POISSON and OPERA) have led to a design for a new pole-piece for SM3. This pole-piece, indicated in Fig. 4.9, yields a central field of 1.6 Tesla, and an integrated dipole field of 5.2 T-m.

The design has an integrated sextupole field component on the symmetry axis of 0.002 T/m. This is small compared to the natural sextupole moment of the Tevatron dipoles and is acceptable for insertion into the Tevatron lattice. In this design the magnet would draw 650 kW of power at 4200 amps, similar to its past operating power level (it was powered by two 500 kW Transrex power supplies). Note that the magnet will be oriented so that charged particles are deflected in the vertical plane. The properties of the magnet, with the pole faces shimmed to the BTeV requirements, are shown in Table 4.2.

Table 4.2: BTeV/C0 Vertex Dipole Properties

Property	Value	Comment
$\int B \times dl$	5.2 T-m	2.6 T-m on each side of center of IR
Central Field	1.6 Tesla	
Steel Length	3.2 m	
Overall length	5.3 m	
Magnet Vert. aperture	± 0.3 rad	
Magnet Horz. aperture	± 0.3 rad	

The magnet is centered on the interaction region in Z thus creating both an 'upstream' and 'downstream' spectrometer. In quark-antiquark production at 2 TeV, the bottom quark

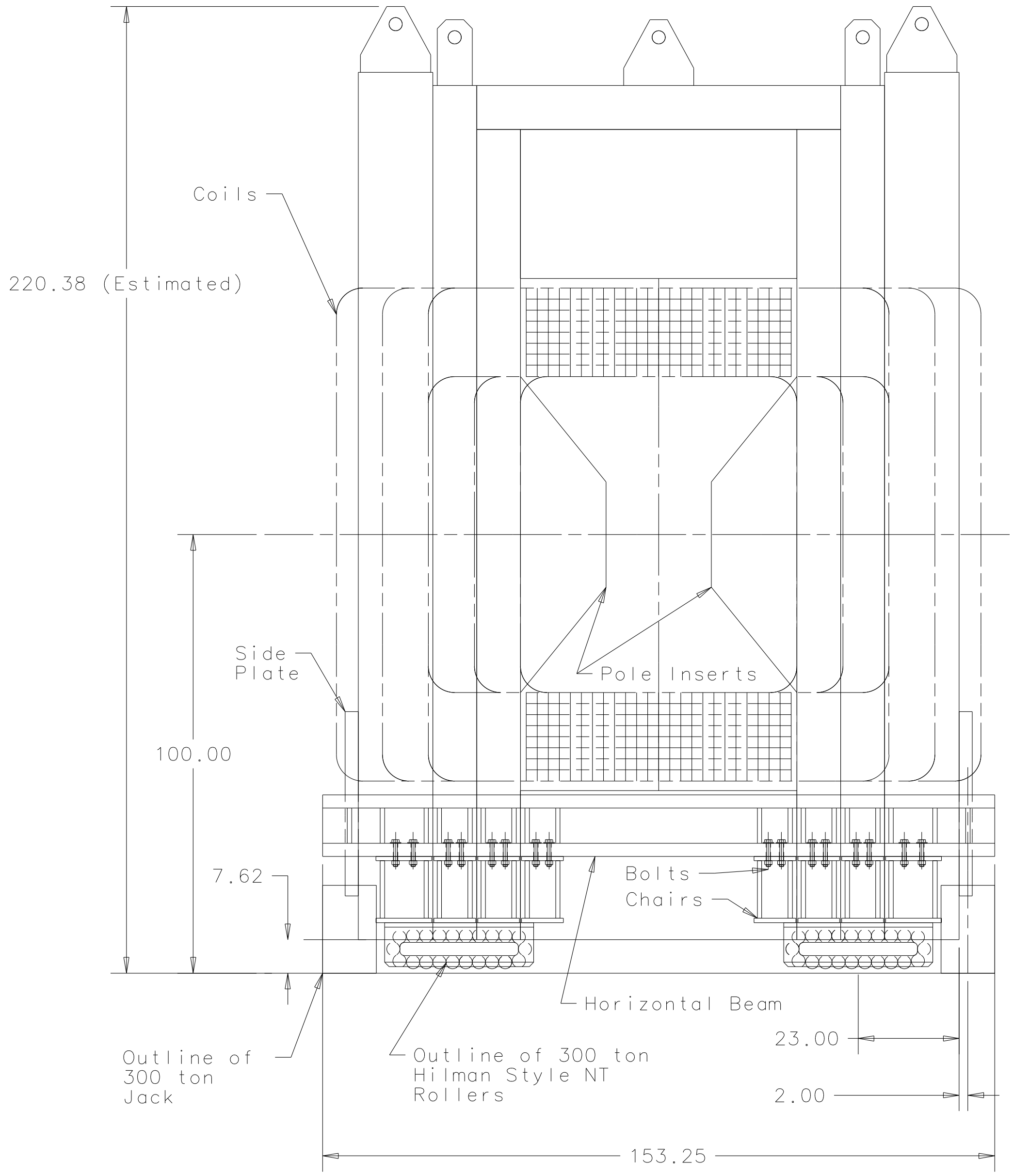


Figure 4.9: Cross section of the modified SM3 dipole with rollers and pole piece inserts. All dimensions are in inches.

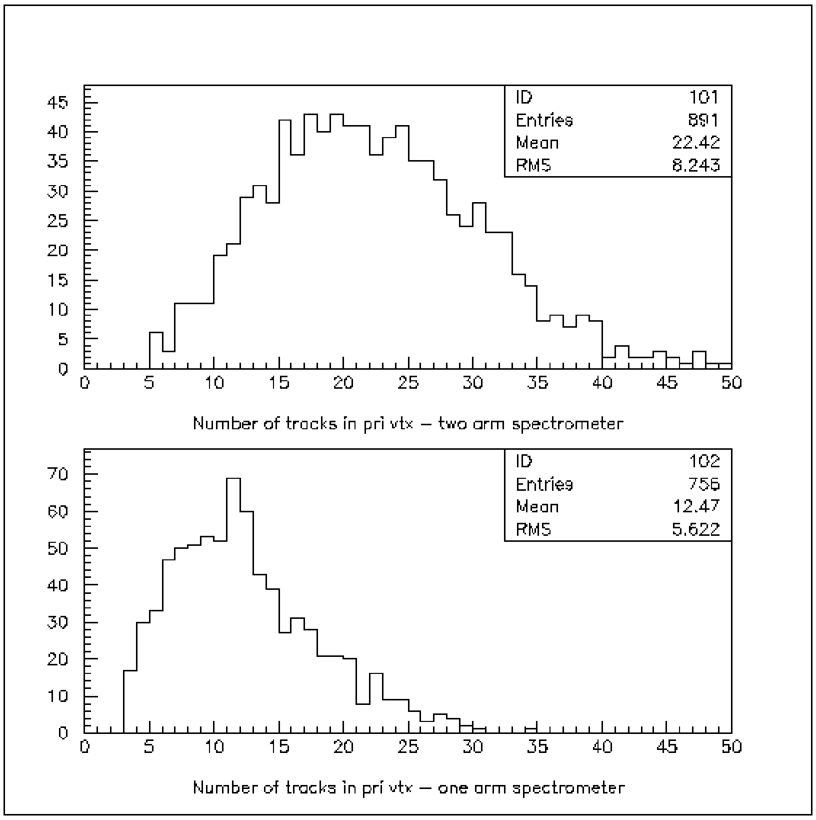


Figure 4.10: Fraction of events with N primary tracks above 1 GeV/c momentum for a) a dipole centered on the IR giving an effective two-arm acceptance and b) a single forward spectrometer arm.

and antiquark are usually either both boosted upstream or both boosted downstream. Thus having two spectrometers doubles the acceptance of the experiment for tagged decays. Having two spectrometers further improves the experiment because the extended coverage increases the number of tracks from the underlying event that can be used to determine the primary vertex. Fig. 4.10 shows the number of high energy primary tracks (low multiple scattering) that contribute to locating the primary vertex for single arm coverage and for two arm coverage. If we require three such tracks to determine the primary vertex, then 2.2 times more events survive with two arms than with one, and the primary vertex resolution is enhanced.

In this central dipole geometry, there is a strong magnetic field at the vertex detector. Because of the excellent spatial resolution of the vertex detector, it is possible to get a crude measurement of the track momentum using the vertex detector alone. This measurement can be used to reject tracks at the trigger level that are at such low momentum that multiple coulomb scattering errors make their assignment to a vertex uncertain.

The second spectrometer arm will also be useful during the R&D and early data-taking phases of the experiment. It is possible, given the lab's budget, that only one arm will be fully instrumented initially. The other side can be used to test new detector concepts or to test prototypes of production components under actual beam conditions until it becomes possible to instrument it fully.

4.2.3 Tracking System

The tracking system is designed to reconstruct tracks and measure their angles and momenta over the full aperture ($\pm$300 mr $\times$ $\pm$ 300 mr) and momentum acceptance (from 1 GeV/c to more than 100 GeV/c) of the experiment. The main requirements that have to be satisfied by the system are:

- Provide adequate resolution on track positions, at least 9μm or better, and angles, at least 0.1 mr, to separate with high efficiency and robustness the secondary vertices of bottom and charm decays from the primary interaction vertex
- Provide sufficient momentum resolution, about 1% at 100 GeV/c, to permit the reconstruction of the invariant mass of multibody bottom and charm decays with good precision
- Deliver its information to a Level 1 trigger processor which implements a trigger based on detached vertices or tracks with large impact parameters
- Handle the rates and tolerate the radiation doses expected in the experiment

To satisfy these requirements, the experiment will employ two tracking subsystems which work together to provide efficient track reconstruction, excellent vertex resolution, and excellent momentum and mass resolution. These are

- The vertex detector which provides the precise vertex measurement and also has momentum measurement capability. The signals from the vertex detector provide the main input to the Level 1 vertex trigger.
- The forward tracker, which provides, along with the vertex detector, precision momentum measurement. It also permits us to reconstruct tracks from K_s^o's and Λ's, which often decay downstream of the vertex detector.

In the following paragraphs, we discuss the design of each of these systems.

4.2.3.1 Vertex Detector

The vertex pixel detector provides the high resolution tracking near the interaction which is required to associate tracks with their proper vertices – primary and secondaries. In designing a vertex tracking system, one must consider the long interaction region (IR) of the Tevatron, which has a σ_z of 30 cm. This forces one to have a rather long vertex detector. In addition, the detector must be placed very close to the interaction region in order to achieve good resolution and acceptance; so radiation damage is an issue. In this experiment, the vertex detector will be used in the Level 1 trigger to select events with secondary vertices. This places special requirements on the detector and its readout.

The baseline BTeV vertex tracking system uses planar arrays of silicon pixel detectors, placed at intervals along the IR and upstream and downstream of it (to catch tracks coming from the end of the IR). Silicon pixel detectors are chosen because of their:

- Superior signal to noise (expected to be better than 100 to 1 for 300μm thick detectors) as compared to other solutions such as silicon strips
- Excellent spatial resolution. This includes the ability to improve the spatial resolution by measuring the charge sharing between adjacent pixels
- Low occupancy due to the large number of channels
- Intrinsically better pattern recognition capabilities compared to strip detectors
- Speed of response
- Radiation hardness

It is especially important for the trigger that the number of spurious noise hits be as low as possible. Also, the vertex tracking system must not produce pattern recognition ambiguities or ghost tracks which would take extra time to sort out at the trigger level. The two dimensional nature of the pixels is an enormous help in this regard.

The baseline vertex detector, shown in Fig. 4.11, has rectangular 50μm by 400μm pixel detectors. It has triplets of planes distributed along the IR separated by 3.2 cm. The individual planes cover ± 5 cm in both transverse dimensions. Acceptance studies show that this is an adequate size for these detectors. They are mounted above and below the beam, and are arranged so that a small square hole of $\pm$6mm $\times$ $\pm$6mm is left for the beams to pass through.

Each triplet consists of one plane that measures the bend view, one that measures the non-bend view, and a third that measures the bend view again. The reason for this particular arrangement of pixel orientations is related to the trigger and is discussed below in the trigger section. A crude momentum measurement can be made using only information from three or four triplets. If the pulse height is read out and made available to the trigger, charge sharing can be used to improve the momentum resolution (the point resolution is expected to be between 5μm and 9μm). The momentum information can be used to reject very soft tracks that would adversely affect the trigger algorithm, and to properly assign multiple coulomb scattering weights to tracks in the vertex fits.

The implementation of this system will be very challenging. Because pixel information must be made available to the first-level trigger, a custom electronic design is necessary. There are also significant mechanical (e.g. alignment and support) and thermal issues that must be addressed. The detector must be shielded electronically from the circulating beam which is a significant rf source. It will be enclosed in a secondary vacuum, isolated from the main accelerator vacuum, by a thin aluminum vacuum container. The silicon should be as close to the beam as possible. A hole of $\pm$6mm on a side will result in the innermost edge of the detector receiving a fluence of close to 2.0×10^{14} per year (10^7 sec) at a luminosity of 2×10^{32} $cm^{-2} sec^{-1}$. The detector needs to be retractable to a distance of 2 cm from the beam while the collider is being filled and until acceleration is complete. Table 4.3 summarizes the properties of the pixel detector.

While pixel detectors have been successfully operated in HEP experiments, no pixel detector of the proposed size which operates at the high rates we will experience in BTeV, and which can provide signals to the first-level trigger has yet been built. A substantial R&D effort is required to meet the BTeV specifications. Although the BTeV pixel detector is unique, pixel detectors are necessary for several other experiments, and we are taking advantage of work being done all over the world. A status report on the R&D to develop this detector is given in Chapter 5.

We recognize that the design presented here is very challenging. We are studying ways of reducing the technical requirements while incurring only minimal loss of physics reach. We are also particularly interested in reducing the amount of material in the pixel detector in order to reduce the number of conversions of photons from B decays. This is discussed in the chapter on the electromagnetic calorimeter below.

4.2.3.2 Forward tracking system

The purpose of the forward tracking system is to:

- Provide high precision momentum measurements for charged tracks. The vertex detector will provide precise position and direction measurements, but it is too short and contains too much material to provide precise momentum measurements.

- Provide the only trajectory measurements for important classes of tracks which do not pass through the vertex detector. The most important such class consists of K_s's and Λ's which decay downstream of the pixel planes. K_s's are daughter particles in some of the most important bottom and charm decays.

- Confirm and locate tracks at the entrance (and exit) of the RICH, the EM calorimeter, and the muon system. In particular, this system allows one to flag charged tracks that interacted or decayed within the spectrometer volume.

The baseline technology for most of the forward tracker is planes of 4mm diameter straw tubes. Seven stations of tubes are distributed along each arm of the spectrometer. Three of these stations are located in the fringe field of the dipole magnet. Three more are placed between the end of the magnet coils and the RICH entrance window (these establish the direction of tracks entering the RICH). The final plane is located between the RICH and the EM calorimeter. A single station has three sets of straws directed along the three views: Y, U, and V (U and V are at $\pm 45^o$ from Y). Each set in a station is made up of three layers of parallel straws.

The straw tube chambers are expected to provide a spatial resolution of $\sim 150\mu$m per plane. Three parallel straw tubes per station will yield a resolution of better than 100μm in each of the three views. This resolution is sufficient to provide a momentum resolution between 0.5% and 1% for momenta between 1 GeV/c and 100 GeV/c.

Recent simulations have demonstrated that at a luminosity of 2×10^{32} $cm^{-2}sec^{-1}$, 4mm diameter straw tubes do not provide adequate segmentation near the beamline in the first three measuring stations. As is discussed in Chapter 6, we have modified the baseline design to include silicon strip detectors to cover the central portion of these stations.

Tables 4.4 and 4.5 summarize the baseline design specifications for the straw tube and silicon strip subsystems of the forward tracker.

4.2.3.3 Diamond Pixel Option

Diamond pixel detectors could have significant advantages over silicon pixel detectors but are in an earlier stage of development. The recombination length of diamond has improved steadily in the last couple of years and is now around 150 μm. The resulting signal is about a factor of three lower than that of silicon. Research on improving the signal collection continues. Properties of diamond that make it superior to silicon if sufficient pulse height can be obtained are:

- lower radiation length leading to lower multiple scattering;
- better radiation hardness; and
- better thermal conductivity which leads to simpler cooling schemes which can result in less material in the detector.

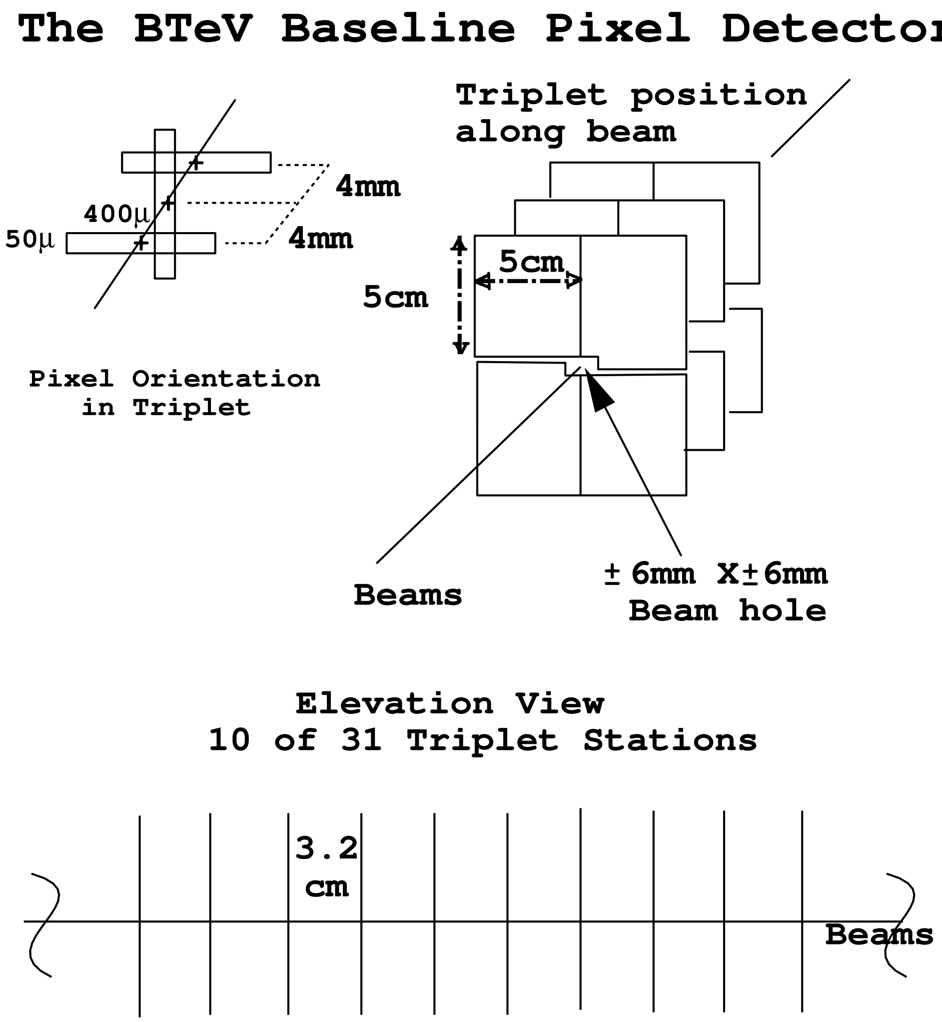


Figure 4.11: Baseline pixel vertex detector

Table 4.3: Properties of Baseline Pixel Vertex Detector

Property	Value
Pixel size	rectangular: 50μm $\times$ 400μm
Plane Dimensions– upper half	10cm $\times$ 5cm
Plane Dimensions– lower half	10cm $\times$ 5cm
Central square hole (adjustable)	nominal setting: 12x12mm
Total Planes	93
Total Stations	31, each consisting of three planes
Pixel orientations	the outer two with narrow pixel dimension vertical the middle one with narrow pixel dimension horizontal
Separation of Stations	3.2 cm
Plane to plane separation (within station)	4 mm
Total Plane Thickness (incl cooling, supports)	1000 μm Si equivalent
Station Thickness (incl. Al shielding skin)	3200 μm Si equivalent
Total Pixels	4.2×10^7
Total Silicon Area	$\approx$1m^2
Readout	analog readout (no. of bits to be decided)
Trigger	signals are used in Level 1 trigger
Rate requirements	beam crossing rate is 132 ns.
Deadtime	$<$3% at $2{\times}10^{32}$cm^{-2}s^{-1}
Noise requirement	desired: $< 10^{-6}$ per channel per crossing required: $< 10^{-5}$ per channel per crossing
Resolution	9μm or better
Alignment accuracy	5μm or better
Radiation tolerance	$\approx 6 \times 10^{14}$ particles/cm^2
Power per pixel	$<$40 μWatt

Table 4.4: Properties of the baseline forward straw tracker

Property	Value
Straw size	4mm diameter
Central hole	5 cm $\times$ 5 cm to 24 cm $\times$ 24 cm
Total Stations	14 (7 per arm)
Z positions (cm)	96, 146, 196, 296, 341, 386, 706
Half size (cm)	30, 45, 60, 90, 105, 118, 210
Views per station	3
Layers per view	3
Total number of straws	78132
Total station thickness ($< L > /X_0$)	0.6%
Readout	ASD + timing chip (6 bits), sparsified
Total channels	118,464
Rate requirements	beam crossing rate is 132 ns.

Table 4.5: Properties of the proposed baseline forward silicon tracker

Property	Value
Wafers	4×6 cm^2 / 2.5×4.75 cm^2
Pitch	100 μm
Thickness	300 μm
Ladder modules	12 per station
Coverage	24 cm $\times$ 24 cm
Central hole	5 cm $\times$ 5 cm
Total stations	6 (3 per arm)
Z positions (cm)	100, 150, 200
Measurement views	4 (X,Y,U,V)
Channels per view	2400
Total channels	57600
Readout	double-sided, sparsified binary

4.2.4 Triggering System

The proposed broad charm and beauty physics program of BTeV/C0 calls for a trigger whose efficiency

- is large for heavy-quark decay events which can be found off-line and
- is relatively independent of decay mode.

Because the program has many goals and will undergo several phases of development, the trigger needs to be flexible. We plan to implement a multi-level trigger with several "branches" that can be independently prescaled and then combined into a global trigger. These will include muon and dimuon triggers, general interaction triggers, and possibly electron and dielectron triggers. However, the demands of a collider B experiment are difficult to meet with simple, conventional approaches to triggering. In particular, the second criterion above is not satisfied by simple triggers commonly used in hadron collider experiments, e.g. lepton triggers. BTeV's main B-physics trigger focuses on the key difference between heavy quark events and typical hadronic events – the presence of detached vertices – at the first trigger level, Level 1. It is very efficient for the B decays that can be successfully reconstructed by the spectrometer. It also avoids the need to focus at Level 1 on specific final states, which can limit the accessible physics. Its implementation requires a very sophisticated trigger architecture which also provides a more than adequate foundation for any of the other triggers we are considering.

In this section, we focus on the Level 1 detached vertex trigger, for which we are carrying out a program of research and development, based on extensive simulation and design studies. Details some of the other triggers – the muon trigger and the electron trigger – are given in the sections on the muon detector and the electromagnetic calorimeter.

4.2.4.1 Level 1 Detached Vertex Trigger

The trigger algorithm has the goal to reconstruct tracks and find vertices in every interaction up to an interaction rate of order 15 MHz (luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at $\sqrt{s} = 2\,\mathrm{TeV}$). This entails an enormous data rate coming from the detector ($\sim 100\,\mathrm{GB/s}$), thus a careful organization of the data-flow is crucial. This trigger must be able to operate both in collider and fixed-target modes. It must be capable of reducing the trigger rate by a factor of at least 100.

The key ingredients for such a trigger are

1. a vertex detector with excellent spatial resolution, fast readout, and low occupancy;
2. a heavily pipelined and parallel processing architecture well suited to tracking and vertex finding;
3. inexpensive processing nodes, optimized for specific tasks within the overall architecture;

4. sufficient memory to buffer the event data while the calculations are carried out; and

5. a switching and control network to orchestrate the data movement through the system.

The pixel detector described above is the first ingredient. Use of pixels for the tracker will produce very high spatial resolution for the trigger (and the final analysis) and, because of their 2-dimensional nature and excellent signal-to-noise, will reduce the level of potential confusion during the crucial track-finding phase of the trigger. Since the pixel detector is located in a magnetic field, some momentum information can be derived from the bending in the pixel detector alone without needing to use the downstream tracking system. If, in addition, the pixels have pulse height readout, they can achieve a spatial resolution of less than 12 μm, which provides rather good momentum information to the Level 1 trigger.

To provide the second ingredient, we employ parallelism both at the subevent level and at the event level. The detector layout described above has "stations" consisting of three planes. Processors organize the hits in each triplet of planes into "station hits". Still at the subevent level, individual processors work on the station hits from a slice of the detector, as will be described below. At the event level, track candidates from the individual slices are brought together in a common processing node, one member of a large farm, for vertex finding.

The third ingredient is provided by using large numbers of inexpensive, commercially available CPU chips. New processors are constantly appearing on the market, often optimized for very specific tasks, and prices are dropping rapidly. Research is in progress to determine the characteristics of the processors needed at each stage.

The fourth ingredient is also commercially available. The time allowed to process the data is limited by the number of events which can be stored while the trigger decision is being made. An event is expected to be significantly less than 200,000 bytes. The memory required to buffer 256 crossings even for 200,000 byte events is only 50 MB and is adequate for a decision time of 33 μs. This system can be off the detector (i.e. not in a strong radiation environment) so commercial (PC) memories are adequate and the costs are small on the scale of the full trigger system. Memory costs do not limit the decision time which could be much longer if necessary.

The final ingredient is switching networks which appear at the input of each section of the Level I trigger and between the Level 1 trigger and next level of triggering. The design of this switching hardware is under study.

The reference trigger scheme was originally developed by the Penn Group - Selove, Sterner, and Isik [4]. While it has undergone major revisions and is still being updated, it has been an excellent starting point for studying this kind of trigger.

The proposed algorithm has four steps:

1. In the first step, hits from each pixel plane are assigned to detector subunits. This can be done very rapidly for all planes in parallel using field-programmable gate arrays (FPGAs). It is desirable to subdivide the area of the detectors in a way which minimizes the number of tracks crossing from one subunit into the next and assures uniform and

low occupancy (< 1 hit/event) in all subunits. We therefore divide each plane into 32 azimuthal sectors ("ϕ-slices"). Simulation shows that at 2 TeV, on average each ϕ-slice contains 0.3 hits per inelastic interaction, approximately independent of the plane's z coordinate.

2. In the second step, "station hits" are formed in each triplet of closely-spaced planes using hits from each ϕ-slice. Since the pixels under consideration have a rectangular shape (e.g. $50\,\mu$m by $400\,\mu$m), the two outer planes in each station are oriented for best resolution in the bend view, and the inner plane for the non-bend view. There is a "hit processor" for each ϕ-slice of each station. The hit processor finds triplets of overlapping pixels, and sends to the next stage a single "minivector" consisting of x, y, $\frac{dx}{dz}$ and $\frac{dy}{dz}$. Given the good position resolution of each measurement, a detector station determines a space point accurate to $5-9\,\mu$m in x and y, and y and x slopes to $\frac{dy}{dz} \approx 1\,$mr (bend-view) and $\frac{dx}{dz} \approx 10\,$mr (non-bend). The use of these minivectors in the track-finding stage substantially reduces combinatorics.

3. In the third step, minivectors in each ϕ-slice from the full set of stations are passed via a sorting switch to a farm of "track processors." The sign of the bend-view slope is used to distinguish forward-going and backward-going minivectors and send them to separate farms. To handle the few percent of tracks crossing segment boundaries, hits near a boundary between two ϕ-slices can be sent to the processors for both slices.

   Each track processor links minivectors into tracks, proceeding along z from station to station. For each pair of minivectors in adjacent stations, the processor averages the y-slopes in the two stations and uses this average slope (which represents the slope of the chord of the magnetically deflected trajectory) to project from the first station into the second. It then checks whether the y-value of the minivector in the downstream station agrees with the projection within an acceptance window. If three or more hits satisfy this requirement, a fast fitting algorithm finds the momentum of the track candidate.

4. In the fourth step, the tracks are passed to a farm of "vertex processors" and used to form vertices. The vertex with $\sum p_z$ closest to zero is designated as the primary vertex. Tracks which have a large impact parameter with respect to this vertex are taken as evidence for heavy quark production in an event. To reduce the effect of multiple scattering on vertex resolution, tracks below an adjustable bend-view transverse-momentum (p_y) threshold are excluded from the vertex finding.

It is desirable to have pulse-height information included with each hit so that interpolation can be used to improve the point resolution beyond the $50\,\mu\text{m}/\sqrt{12}$ otherwise available. Digital (1 bit) readout seems to be adequate for the p_y cut, but the improved momentum resolution with analog information would allow the processor to calculate the mass of the charged particles at a vertex, which may be a useful variable on which to cut if additional rejection is needed. It also can improve the position accuracy for tracks incident at large angles, which cross over rows or columns of pixels.

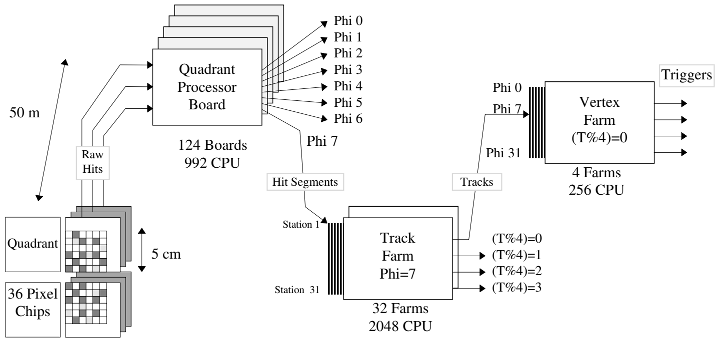


Figure 4.12: Schematic of Proposed Level 1 vertex trigger

The Level 1 rejection factor for events containing only light quarks is required to be approximately 10^{-2} in order to reduce the data rate into Level 2 to sufficiently low levels. The overall acceptance for $B \to \pi^+\pi^-$ events is expected to be 10-20% for the proposed BTeV Level 1 vertex trigger. The expected trigger efficiency for $B \to \pi^+\pi^-$ events that will have both B daughter πs reconstructed should be much higher – about 50%. Results of simulation studies with this trigger will be described below in Chapter 8.

D. Husby and W. Selove have developed a detailed proposal for this trigger system. In this design there are about 1000 hit processors, 2000 track processors, and 128 vertex processors. Since the architecture is scalable, there is flexibility for these numbers to be optimized to meet unanticipated conditions. The design uses a combination of digital signal processors (DSPs), FPGAs and commercial CPUs. Fig. 4.12 shows a schematic diagram of the full system. Details of their current design can be found in Ref. [5]. A hardware design has been worked out for a small-scale prototype; see the trigger research and development section for details.

4.2.4.2 Development of other Level 1 triggers and the global trigger framework

While we have emphasized the “vertex/impact parameter” trigger because it is our most crucial and most ambitious trigger, we will also have other triggers at Level 1. These will include muon and electron triggers and perhaps some triggers derived from the same hardware that is used for the vertex trigger but optimized for different physics. At the output of the first level trigger there will be a ‘global Level 1 trigger’ subsystem that will allow us to combine the various individual trigger elements into final triggers (for example, an event with only weak evidence for a secondary vertex but with an indication of a high p_t single lepton might be accepted). It will also have the ability to apply a programmable prescale to the various triggers before the are ‘or’ed together to form the final trigger. We will probably also include the ability to dynamically change the prescales during the course of the store.

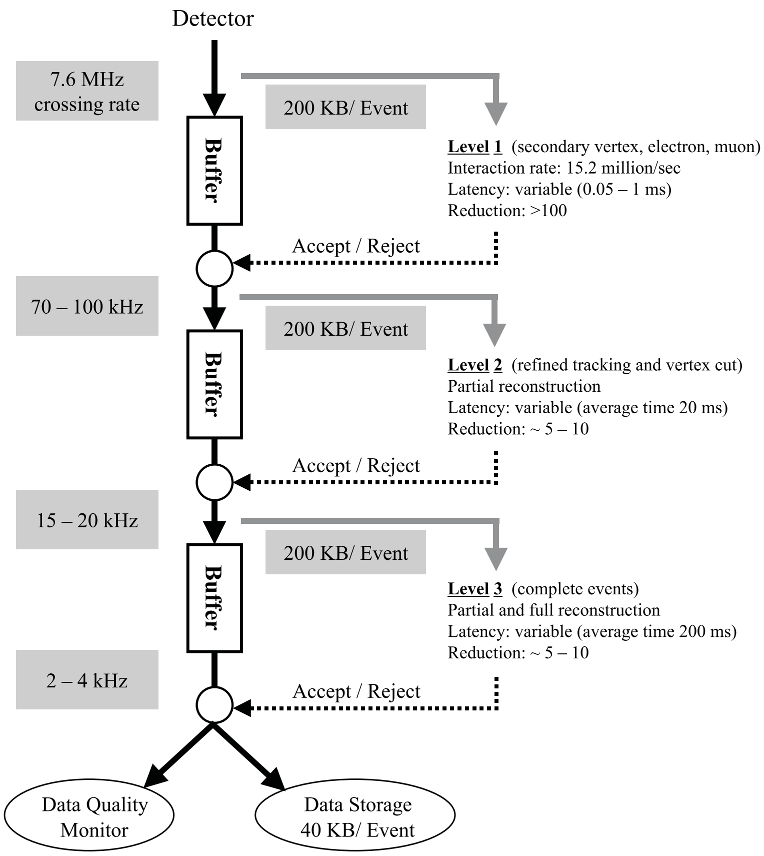


Figure 4.13: The BTeV data flow diagram. The numbers are for the nominal luminosity of $2 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

4.2.4.3 Beyond the Level 1 Trigger

Modern experiments in high energy physics implement hierarchical trigger systems and BTeV is no exception. Many details of the full data acquisition system have to be worked out as the design of the detector components and trigger algorithms become more mature. At the design luminosity of $2 \times 10^{32}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and with a total inelastic cross section of $\sigma_{inel} \approx 60$mb almost every crossing will have an interaction, so events will have to be examined at the crossing rate of 7.5 MHz. Fig. 4.13 gives the outline of the expected data flow at the nominal design luminosity.

At our current level of understanding, the needed level 1 rejection factor seems to be achievable even if the only test made is for a detached vertex in vacuum. The event rate is then reduced to ≈100 kHz in the worst case during the highest luminosity running. It is still possible to move all this data, amounting to a maximum rate of 10 GB/s, to the Level 2 trigger.

Events that pass the first trigger level are forwarded to compute nodes to be further analyzed. First, information from other detectors besides the pixel detector is used to get better values of the momentum of the tracks. The vertex detachment requirement is then imposed with a better knowledge of the error matrices of the tracks and improved vertex assignments are computed. Events remaining after this step are then fully assembled in a switch-based event builder and passed to a processor farm. Even so, it will probably be impractical to write several kHz of events to archival storage.

The final trigger level will combine information from different sub-detectors to further suppress background events. Particle identification will be available at this stage and could be used to obtain a very clean direct charm signal for specific final states. We intend to fully reconstruct the events at this level before the data are sent to a storage system and made available for off-line processing.

Because the event rate surviving this last level may still be close a few kHz, the software will probably have to reduce the amount of data per event to archival storage by writing out an event summary which eliminates much of the raw data. The event summary would be around 50 KB so that the output rate could still be as high as 80 MB/s. This results in a dataset size of 800 TB/yr, somewhat more than is expected from CDF or D0 in Run II. However, initial phases of the experiment will run at much lower rates and will be comparable in dataset size to a current fixed target experiment.

4.2.4.4 Additional Work

Although simulation results are very encouraging, much additional simulation work and a vigorous R&D program are needed to implement such a triggering system. The simulation work involves studying the performance of these proposed Level 1 and Level 2 algorithms under realistic experimental conditions, which include detector inefficiencies, photon conversions, delta ray production, hadronic interactions in the silicon, low energy loopers, multiple interactions, electronic noise, machine backgrounds, and beam-induced backgrounds (e.g. scraping in the support structures and/or beam pipe). Some of these studies have already been done and are presented in Chapter 8.

4.2.5 Particle Identification

Excellent charged hadron particle identification is a critical component of a heavy quark experiment. Even for a spectrometer with the excellent mass resolution of BTeV, there are kinematic regions where signals from one final state will overlap those of another final state. For example, $B_d^o \to \pi^+\pi^-$, $B_d^o \to K^+\pi^-$, and $B_s^o \to K^+K^-$ all overlap to some degree leading to ambiguous classification of the final state. These ambiguities can be eliminated almost entirely by an effective particle identifier. In addition, many physics investigations involving neutral B-mesons require 'tagging' of the flavor of the signal particle by examining the properties of the 'away-side' particle. Our studies show that kaon tagging is a very effective means of doing this.

4.2.5.1 Requirements

In the design of any particle identification system, the dominant consideration is the momentum range over which efficient separation of the various charged hadron types – π, K, and p – must be provided. In BTeV, the physics goal which sets the upper end of the momentum requirement is the desire to cleanly separate $B_d^o \to \pi^+\pi^-$ from $B_d^o \to K^+\pi^-$ and $B_s^o \to K^+K^-$. These two-body decays produce reasonably high momentum pions and kaons. Fig. 4.14 shows the momentum distribution of pions from the decay $B_d^o \to \pi^+\pi^-$ mentioned above for the case where the two particles are within the spectrometer's acceptance. The low momentum requirement is defined by having high efficiency for 'tagging' kaons from generic B decays. Since these kaons come mainly from daughter D-mesons in multibody final state B-decays, they typically have much lower momentum than the particles in two body decays. Fig. 4.15 shows the momentum distribution of 'tagging' kaons for the case where the signal particles are within the geometric acceptance of the spectrometer. About 1/5 of the tagging kaons never exit the end of the spectrometer dipole. Almost all kaons exiting the dipole have momenta above 3 GeV. Based on these plots, we have set the momentum range requirement for the particle identification system to be

$$3\,\mathrm{GeV/c} \;<\; P_{particle\ id} \;<\sim 70\,\mathrm{GeV/c} \tag{4.2}$$

Finally, kaons and pions from directly produced charm decays have momenta which are not very different from the kaons from B-decays. Fig. 4.16 shows the momentum spectra of kaons from accepted $D^o \to K^-\pi^+$, $D^o \to K^-\pi^+\pi^-\pi^+$, and $D_s^+ \to K^+K^-\pi^+$ in both collider and fixed target mode. The range set by the B-physics requirements is a reasonable, if not optimal, choice also for charm physics.

4.2.5.2 RICH radiators

Because of the large momentum range and limited longitudinal space available for a particle identification system in the C0 enclosure, there is really only one choice of detector technology – a gaseous ring-imaging Cherenkov counter. Pions and kaons can be separated in this momentum region with a single gas radiator like C_4F_{10} ($n = 1.00138$ in visible range). Some

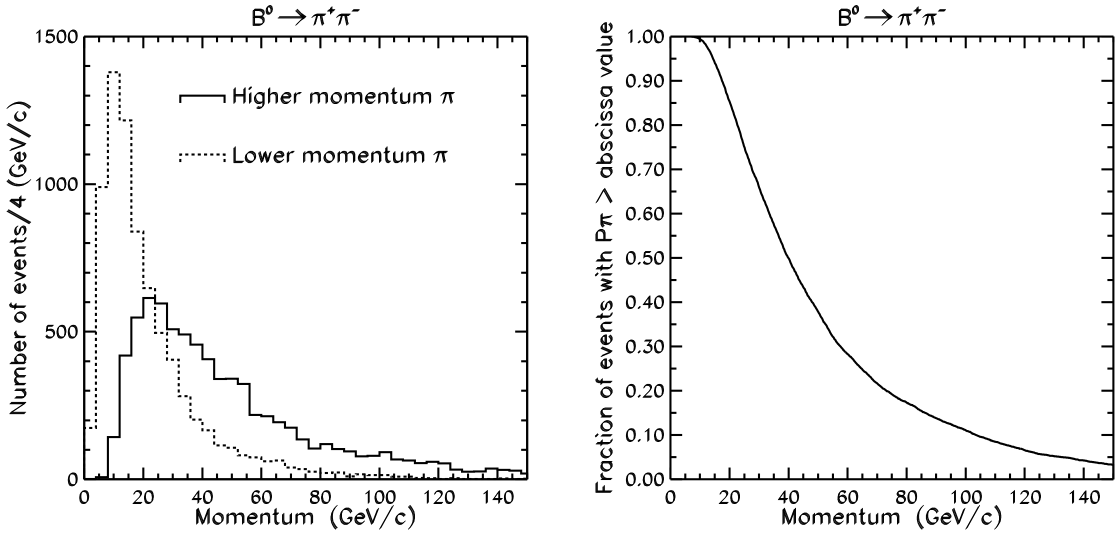
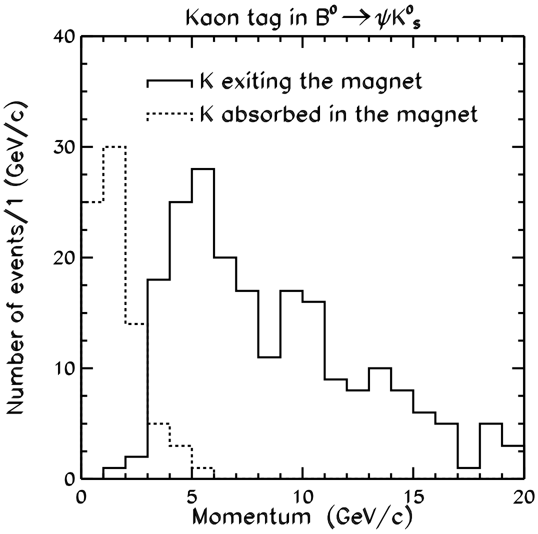


Figure 4.14: The momentum distribution of pions in $B_d \to \pi^+\pi^-$ decays. The left plot shows distributions for the lower (dashed line) and higher (solid line) momentum pion in this decay. The right plot presents the latter distribution in integral form, which gives loss of efficiency as a function of the high momentum cut-off of the particle ID device.

Parameter	C_4F_{10}	aerogel
n	1.00138	1.03
$\Theta_c(\beta=1)$	52.5 mrad	243 mrad
π threshold	2.5 GeV/c	0.6 GeV/c
K threshold	9.0 GeV/c	2.0 GeV/c
p threshold	17.1 GeV/c	3.8 GeV/c
thickness	283cm	5cm

Table 4.6: Several parameters for the proposed gaseous and solid radiators.



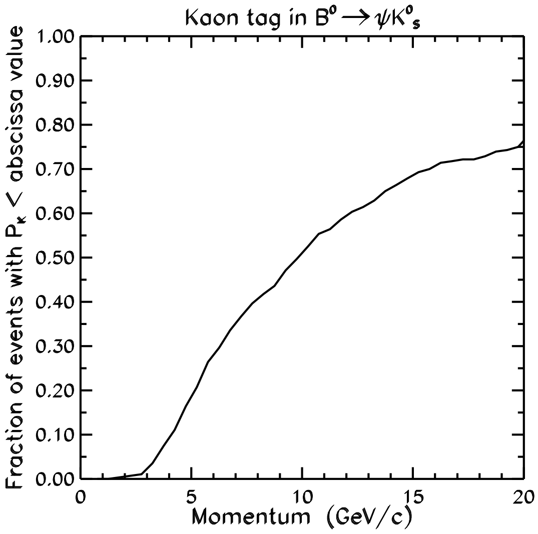


Figure 4.15: The momentum distribution of 'tagging' kaons for the case where the signal particles (ψK_S^0) are within the geometric acceptance of the spectrometer. The left plot shows distributions for kaons absorbed in (dashed line) and exiting from (solid line) the magnet. The right plot presents the latter distribution in integral form, which gives loss of efficiency as a function of the low momentum cut-off of the particle ID device.

parameters of this gas are given in Table 4.6. Among other experiments, C_4F_{10} is used in DELPHI (endcap) [1], HERA-B [2] and HERMES RICH detectors [3]. It was also chosen for one of the LHC-B RICH detectors [4].

Note that below about 9 GeV, no gas can provide K/p separation and that, since kaons are below threshold, the RICH operates in a threshold mode (except that it has much better noise discrimination than a normal threshold counter because it still measures a Cherenkov ring for pions). In order to improve identification of low momentum particles, one interesting possibility is to insert a thin ($\sim$ 5 cm) piece of aerogel at the entrance to the gas RICH as proposed by LHC-B [5] and already implemented by HERMES [3]. For example, aerogel with a refractive index of $n = 1.03$ would lower the π, K, p momentum thresholds to 0.6, 2.0, 3.8 GeV/c respectively as illustrated in Fig. 4.17. Shorter wavelength Cherenkov photons undergo Raleigh scattering inside the aerogel itself. They are absorbed in the radiator or exit at random angles. A thin window between the aerogel and the gas radiator passes photons only in the visible range, diminishing the scattered component.

The same photo-detection system can detect Cherenkov rings produced in both the gaseous and the aerogel radiators if it operates in visible wavelengths.

4.2.5.3 RICH dimensions

The RICH detector can be located behind the tracking chambers just outside the central dipole magnet, about 4 meters away from the interaction point. The length of the RICH

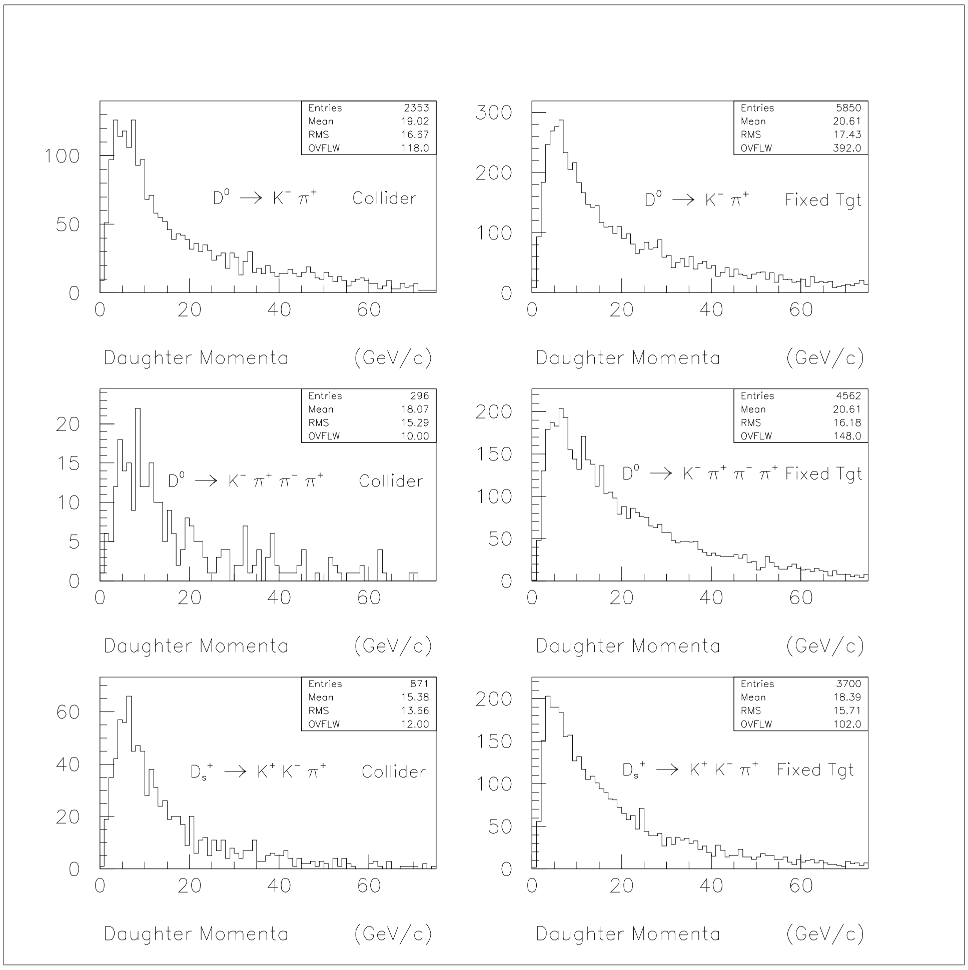


Figure 4.16: The momentum spectra of kaons from accepted $D^o \rightarrow K^-\pi^+$, $D^o \rightarrow K^-\pi^+\pi^-\pi^+$, and $D_s^+ \rightarrow K^+K^-\pi^+$ in both collider and fixed target modes.

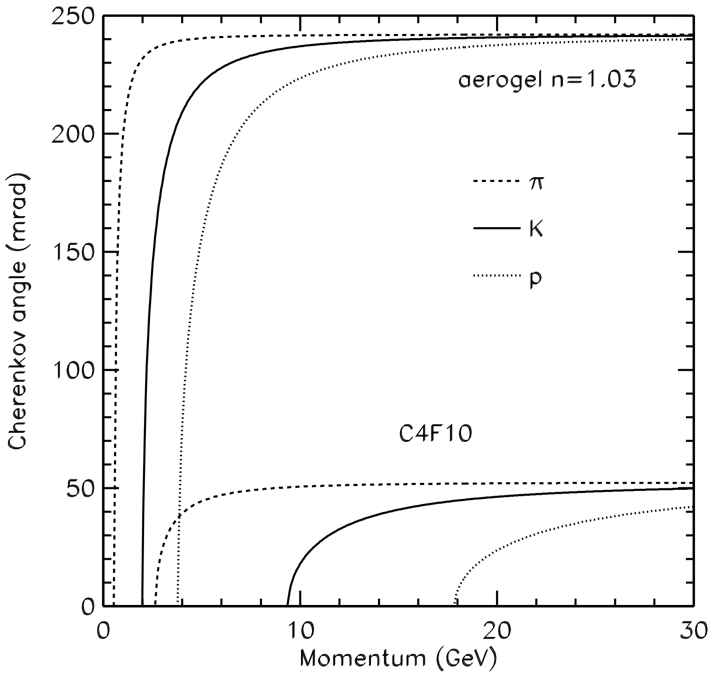


Figure 4.17: Cherenkov angles for various particle species as a function of particle momentum for C_4F_{10} and aerogel ($n = 1.03$) radiators.

detector must be less than 3 meters to allow sufficient space for the EM calorimeter and the muon system. The aerogel radiator with thickness of about 5cm will be mounted at the entrance to the RICH vessel. The gaseous C_4F_{10} radiator will fill the rest of the RICH volume. The transverse dimensions of the RICH entrance window are set by the aperture of the dipole magnet to be about $1.35m \times 1.35m$.

Spherical mirrors at the end of the gas volume reflect Cherenkov photons back and focus them into rings at the photo-detection surface. The photo-detectors should be located away from the flux of particles exiting the magnet, therefore the mirrors are tilted. In order to minimize geometric aberrations from an off-axis mirror configuration, the mirrors will be split along the mid-line of the detector, reflecting photons to photo-detectors located on each side of the vessel in the non-bend view ($x - z$ plane). The longer the RICH detector the smaller the tilt angle. Since the geometric aberrations due to the mirror tilt are significant for the gas radiator, we plan on the longest RICH detector we can accommodate within the space limitations. This also maximizes the photon yield from the gaseous radiator, again improving the resolution of the device. Thus, the gas radiator will be approximately 3 meters long and the mirror radius will be 7 meters. The mirror tilt angle will be about 240 mrads. Note that the mirror tilt angle cannot be further reduced by use of additional flat mirrors at the other end of the detector, like in the designs of HERA-B and downstream LHC-B RICH detectors, since the front of the gas volume must stay transparent to the aerogel photons. The transverse size of the mirror will be about $4m \times 4m$ A possible configuration for the mirrors is to make them from an array of individual hexagons as shown in Fig. 4.18. Each mirror half would consist of 18 full hexagons (76.2 cm tip-to-tip) and 6 half hexagons. A reflectance efficiency over 85% at all relevant wavelengths has been achieved in HERA-B detector with 7mm thick pyrex coated with 200nm Al and 30nm MgF_2.

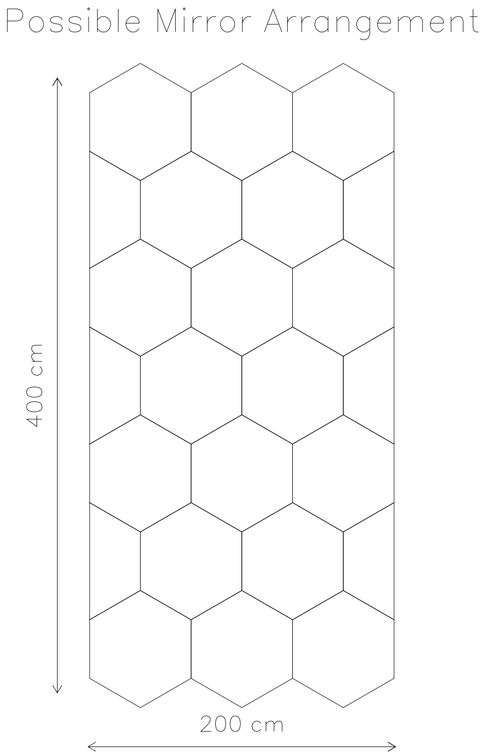


Figure 4.18: Possible arrangement for each half of the mirror using hexagonal mirrors of size 76.2 cm tip-to-tip.

To find the size and optimal position and orientation of the photo-detection surface, we have used a ray tracing Monte Carlo. Even though the true focal plane of a spherical mirror is not planar, non-planar surfaces do not improve resolution significantly and are difficult to realize in practice. Thus, we have assumed that the Cherenkov rings are focused on a plane. We then minimized emission point error for the gaseous radiator with respect to the location of the plane and its orientation using a sample of $B_d \to \pi^+\pi^-$ events. The untilted orientation of the photo-detectors is very close to the optimum. The corresponding contribution to the Cherenkov angle resolution for the gaseous radiator is 0.4 mrad per photon, compared to 0.2 mrad for untilted mirror. As expected for spherical mirrors the photo-detectors must be located about half way between the interaction point and the mirror, which puts them slightly behind the RICH entrance window.

Geometry of the proposed RICH detector is shown in Fig. 4.19.

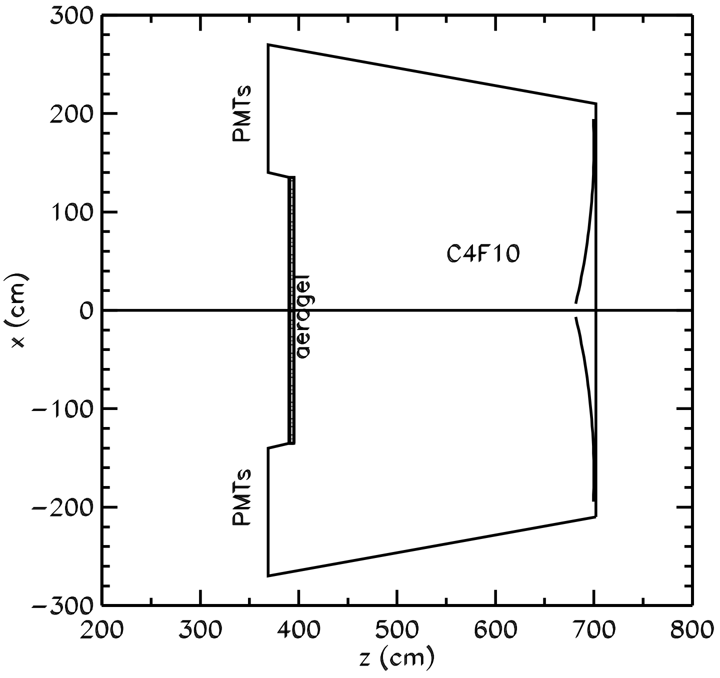


Figure 4.19: BTeV RICH geometry outline.

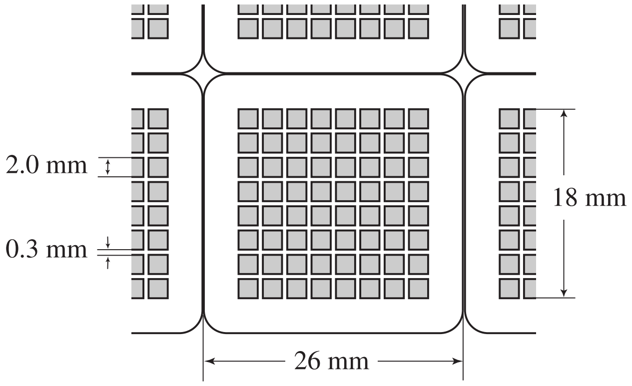


Figure 4.20: Closely packed array of Hamamatsu R5900-M64 tubes as seen from the front. The anode cells are shaded. The picture taken from Ref.[6].

4.2.5.4 Photo-detectors

The photo-detectors must be sensitive to visible wavelengths in order to detect Cherenkov rings from the aerogel. Because of the open geometry of the forward spectrometer and the absence of magnetic field around the RICH detector, arrays of photo-multipliers can be used. Detection of the Cherenkov spectrum in visible wavelengths allows the use of inexpensive window materials, standard photo-cathodes and a simple gas system. The photo-multiplier solution offers durability of the device. Large intrinsic gain and fast response provide for noise resistance and a straightforward read-out system. Hamamatsu has recently developed R5900 multi-anode photo-tubes (MAPMT) which are about 1 square inch in cross-section and are segmented into four, sixteen or sixty-four separate anodes. Thus, anode size can be selected to match other contributions to Cherenkov angle resolution. With the largest segmentation, the anode cell size is 2×2 mm. A closely packed array of R5900-M64 tubes is illustrated in Fig. 4.20. The multi-anode photo-tubes R5900-M16 were adopted for the HERA-B RICH detector [2]. The active photo-cathode area of R5900 tubes constitutes only about 50% of the total area of the tube (not including additional 0.3 mm wide dead spaces in between the anode cells).

To increase the geometrical efficiency, a light focusing system right in front of the PM array is needed. Some losses of light in the light collection system are unavoidable. HERA-B uses a two lens system to concentrate light on the active area of a MAPMT. The magnification of the HERA-B system is 0.5, thus the effective cell size is doubled relative to the true anode cell size and a single photo-tube is serving an area which is larger than its own dimensions (i.e. the tubes are not closely packed). UVT plexiglass lenses have a combined transmission efficiency of about 70% at larger wavelengths (including reflective losses) [2]. LHC-B has proposed a single lens system [6]. An array of convex-plano lenses is in direct optical contact with MAPMT windows (see Fig. 4.21). The MAPMT tubes are closely packed. To recover the inactive area of MAPMTs the magnification of the system is 0.7. Such a design provides almost full coverage of the photo-detection plane with the minimal effective anode cell size (2.8×2.8mm for R5900-M64 tube). There is just one refractive surface to pass, thus the reflective losses are minimized. LHC-B is also studying a light collection system based on light guides concentrating light into active parts of the anodes [8]. Such a system could even avoid small dead areas in between the anode cells which add up to 22% inefficiency for the R5900-M64 tube. Similar solutions will work for our RICH as well.

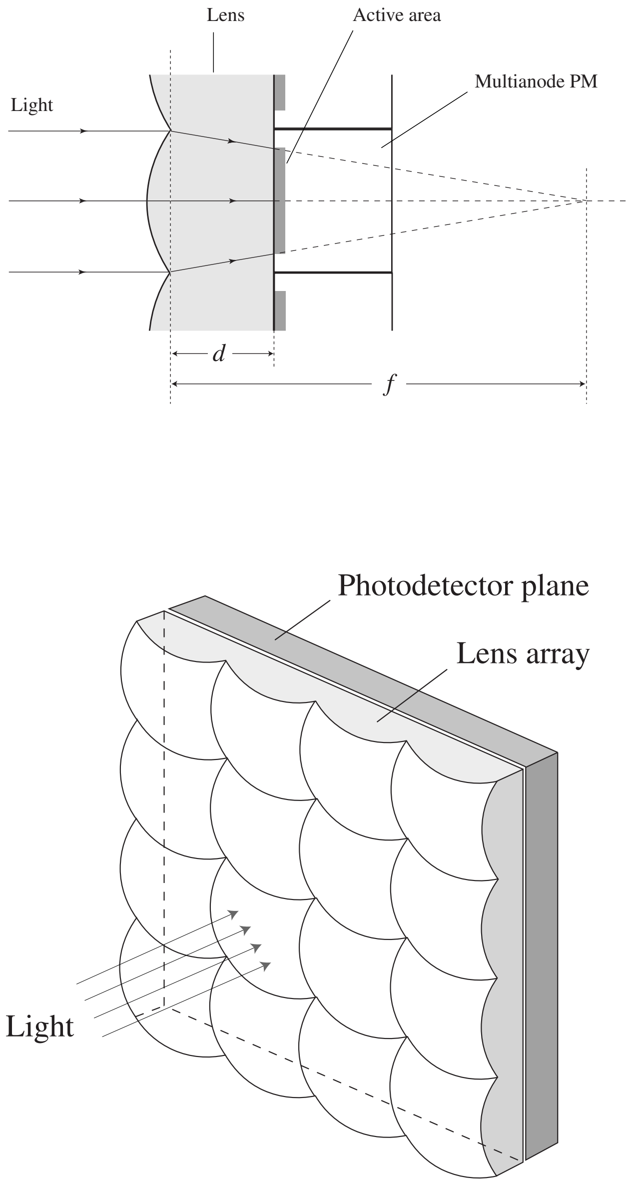


Figure 4.21: Two views of the lens system proposed by R. Forty for use with closely packed arrays of Hamamatsu R5900 MAPMTs. The pictures are taken from Ref.[6].

Hybrid Photo-Diodes offer an alternative photo-detection scheme. In an HPD, a photo-electron emitted by the photo-cathode is accelerated onto a silicon pixel detector by a very high voltage. Commercially available HPDs from DEP in the Netherlands have 61 closely packed hexagonal pixels, each with an area of 3.5 mm^2, enclosed in a single cylindrical tube (see Fig. 4.22). The active area of the tube is small, ≈30% of the total area. These HPDs were successfully used to detect Cherenkov photons in a beam-test setup by the LHC-B group [9]. To recover geometrical inefficiency, a light focusing system with a magnification of < 0.5 could be designed. An aggressive R&D program has been pursued at CERN to develop HPDs with a higher active area for LHC-B RICH detectors. This is mostly achieved by increasing the overall size of the vacuum tube to 80-127 mm in diameter, with a much larger number of pixels inside (1024-2048). To maximize the active area of the front face of the tube, photoelectrons are electrostatically focused onto a smaller diameter silicon pixel array inside, while being accelerated at the same time. Two different focusing schemes are under development. For such a large number of channels per tube it is impractical to have front-end electronic chips situated outside the tube. Thus, the read-out chips are being integrated into the vacuum tubes of HPDs. One of the new HPDs is being developed in collaboration with DEP. Even if suitable HPDs are available in the relevant time scale for BTeV, HPD technology has some drawbacks. Lower intrinsic gain ($5 \cdot 10^3$ vs 10^6 for MAPMT) requires more sophisticated read-out electronics. Also much higher voltages are required (20 kV vs 1 kV for MAPMTs). HPDs are more sensitive to fringe magnetic fields than MAPMTs. On the other hand, if the CERN R&D program succeeds, HPDs with a larger fraction of the active area may be available, perhaps eliminating the need for a light collection system in front of HPD arrays. The overall price for a full size detector is estimated, at present, to be comparable between the MAPMTs and HPDs.

We choose the MAPMT solution for the baseline design, however we intend to continue to evaluate HPD option as well.

4.2.5.5 Expected resolution for gaseous radiator

The emission point error magnified by the mirror tilt is about 0.4 mrad per photon, slightly dependent on the the exact charged track trajectory.

The chromatic error depends on the choice of photo-cathode, window and lens materials. Extending the transparency of the optical elements towards shorter wavelengths increases the chromatic error per photon. At the same time, the number of detected C_4F_{10} photons per track is also increased, making the Cherenkov angle resolution per track almost invariant. It should be noted that the increased wavelength sensitivity has no effect on photon yield and chromatic error for the aerogel radiator. Since materials with good UV transparency (e.g. quartz glass) are expensive, we assume standard borosilicate glass windows on MAPMTs. The dependence of quantum efficiency on photon energy used in the simulations is shown in Fig. 4.23. With this wavelength acceptance, the chromatic error per photon is expected to be 0.39 mrad, well matched to the emission point error.

The third significant contribution to the single photon resolution will come from a finite

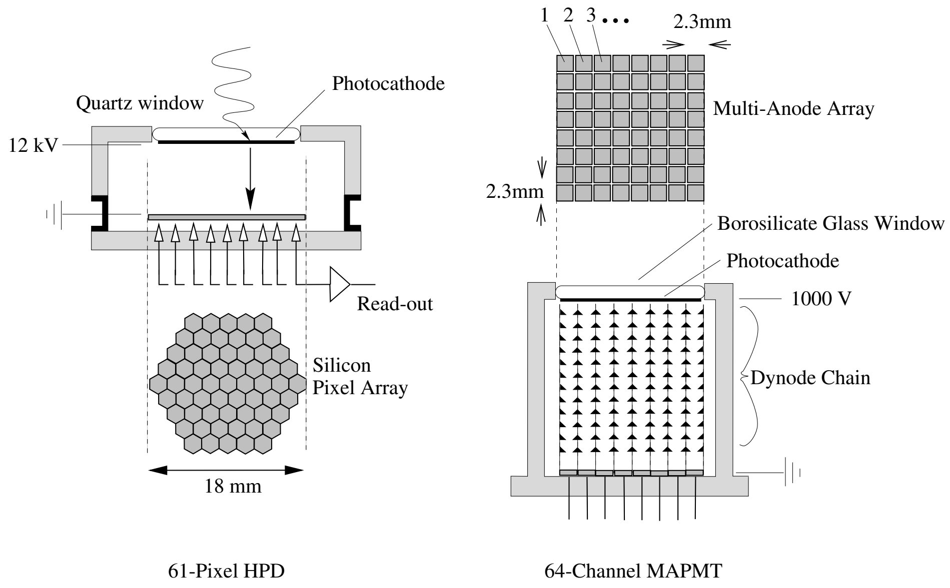


Figure 4.22: 61-Pixel HPD from DEP and R5900-M64 MAPMT from Hamamatsu. The pictures are taken from Ref.[7].

effective cell size of the photo-detector. For $d \times d$ cell size, this contribution is expected to be $\sigma_\theta = d/\sqrt{12} \times 2/R$. Closely packed R5900-M64 photo-tubes with the single lens light collection system described previously would result in a detector granularity error of 0.24 mrad, below the contributions from the emission point and chromatic errors. Covering the entire photo-detection area with the finest segmentation escalates the number of electronic channels and therefore, the cost of the detector. The best possible Cherenkov angle resolution is needed for high momentum tracks radiating photons in the the gaseous radiator. The higher momentum tracks tend to be closer to the beam pipe and illuminate a smaller region of the photo-detection plane than slower tracks. This region also has the highest occupancy of photons since the track multiplicity increases in the beam direction. For these two reasons, the inner region of the photo-detection plane must be populated with the finest segmentation. The outer region of the photo-detection plane is populated by C_4F_{10} photons from slower tracks. All aerogel photons tend to fall into the outer region too, because of the much larger value of the Cherenkov angle. This is fortunate, since low intensity aerogel rings would have been heavily contaminated by high intensity gaseous rings in the inner region. The Cherenkov angle resolution becomes less critical for slower tracks. Since occupancy is lower in the outer regions, it will be possible to populate these areas with coarser segmentation without a significant impact on the performance of the device. We intend to use R5900-M16 tubes here. In this region, the photo-detector segmentation error will be around 0.51 mrad and will dominate the C_4F_{10} Cherenkov angle resolution. To determine the boundaries of

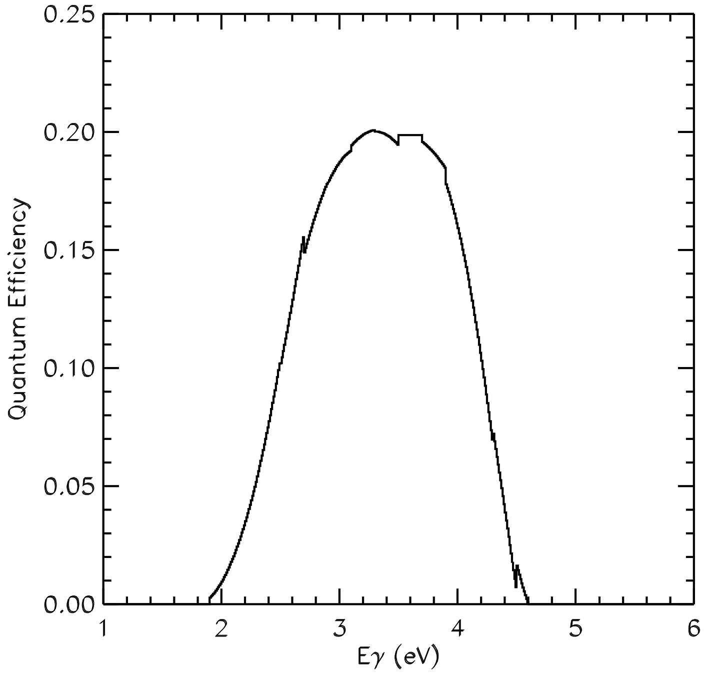


Figure 4.23: Quantum Efficiency of MAPMTs with borosilicate glass windows (as used in HERA-B RICH) used in our simulations.

the regions with fine and coarse segmentation of MAPMTs, we studied the occupancy of the photo-detection plane by various sources of Cherenkov photons. The distribution of photons produced in C_4F_{10} by the faster track in the reconstructed $B^0 \to \pi^+\pi^-$ events is shown in Fig. 4.24. This should be compared to the similar distribution of C_4F_{10} photons produced by all tracks from generic $b\bar{b}$ events shown in Fig. 4.25. The distribution of aerogel photons from generic $b\bar{b}$ events is also shown in Fig. 4.25. The selected regions of the fine and coarse MAPMT segmentation are indicated on these plots. In total, we intend to cover 3.1 m^2 of area in each of the two identical RICH detectors. The inner (1.2 m^2) part of this area will be finely segmented. Table 4.7 summarizes the number of tubes and electronic channels. The coarser segmentation in the outer photo-detection region has almost no effect on Cherenkov angle resolution for the faster track in the $B^0 \to \pi^+\pi^-$ events (2% effect). The resolution for the aerogel radiator is also not effected, since the chromatic error is large (see the next section). To match this error we could use R5900-M4 tubes in the outer region and further reduce number of electronics channels. However, such a scheme would deteriorate the C_4F_{10} high momentum resolution by 10%. Perhaps, another intermediate area with R5900-M16 tubes could fix the problem. Further optimization studies are needed.

Overall Cherenkov angle resolution per photon for the C_4F_{10} radiator is expected to be

PMT Tube type	Area	Number of tubes	Number of channels
R5900-M64	1.22 m^2	1800	115,200
R5900-M16	1.89 m^2	2800	44,800
Together	3.11 m^2	4600	160,000
Two RICH detectors	6.22 m^2	9200	320,000

Table 4.7: Number of Tubes and channels per detector arm and total in two arms.

0.61 mrad. Per track resolution depends on the observed photo-electron yield.

In addition to quantum efficiency losses, photo-electron yield is reduced by imperfect mirror reflectivity, absorption in light collection and PMT window materials and reflective losses on refractive boundaries. The MAPMT tubes also have a small inactive area in between anode cells. This effect has been recently studied for R5900-M64 tubes by the Oxford group [10]. Their measurement (23% light loss) comes close to the naive expectation from the geometrical dimensions of the anode cells and the dead area in between. We simulate this effect explicitly in a geometrical way. Loss of light yield due to finite cell size (photon overlaps) is about 5.5% for saturated rings, and is also modeled explicitly. For mirror reflectivity we assume 87%. In the HERA-B light collection system, there are five refractive boundaries resulting in 15% light loss due to reflections on these surfaces [2]. There is only one refractive boundary in the light focusing scheme described above, thus we assume only 4% loss. Finally, we assume 20% loss for absorption in the light collection system compared to 15% estimated for the HERA-B lens system. After all these corrections the average photo-electron yield from C_4F_{10} for $\beta = 1$ particles is expected to be close to 50. LHC-B expects to achieve a similar photo-electron yield with a 3 times shorter C_4F_{10} radiator. This is mostly (factor of 2.3) due the quantum efficiency assumed by LHC-B, which extends towards much shorter wavelengths. The rest can be accounted for by a more optimistic assumption about light collection efficiency. Even though the extended wavelength coverage increases photo-electron statistics, chromatic error is also increased producing almost no gain in Cherenkov angle resolution per track. To maintain wide wavelength coverage expensive window and light collection materials would have to be used. Therefore, we prefer to work in a narrower range of wavelengths. Our estimated photo-electron yield is larger than projected for the HERA-B RICH (32-35 photo-electrons)[2]. This is partially due to a different light collection scheme assumed here, and the lack of a second mirror in our design.

Our simulations predict the average Cherenkov angle resolution per track from the C_4F_{10} radiator to be 0.0935 mrad. The resulting K/π separation at higher momenta is shown in Fig. 4.26. The projected performance satisfies our physics goals well. Various contributions to Cherenkov angle resolution are summarized in Table 4.8.

Table 4.8: Expected performance of BTeV RICH system.

	C_4F_{10}	Aerogel $n = 1.03$
emission point error	0.4 mrad	0.1 mrad
chromatic error	0.4 mrad	1.9 mrad
segmentation	0.2 mrad	0.5 mrad
total error per photon	0.6 mrad	2.0 mrad
number of photons	50	9
total error per track	0.094 mrad	0.7 mrad

4.2.5.6 Expected resolution for aerogel radiator

The chromatic error for aerogel photons is expected to be about 1.9 mrad. This is likely to be an overestimate since we did not simulate attenuation of shorter wavelengths in the radiator itself. The effect of the lucite filter between the aerogel and gas radiators was simulated as a sharp cut-off at 300 nm.

HERMES reports [3] additional 0.7 mrad error, mostly due to aerogel density variation. The emission point error is expected to be very small (0.09 mrad) because the radiator in only 5cm thick. The detector granularity error is below 0.51 mrad. Putting all these factors together we predict the Cherenkov resolution per photon to be around 2 mrad.

We assume 9 detected photo-electrons per track according to the recent reports from HERMES. This photon yield cannot be increased by using a thicker radiator because of light scattering and absorption in aerogel itself. Further improvements in aerogel transparency would increase the number of photons per Cherenkov ring.

The expected Cherenkov resolution per track is about 0.7 mrad. The resulting K/p separation at lower momenta is shown in Fig. 4.27. The required K/p separation up to the proton Cherenkov radiation threshold in C_4F_{10} should be easily achieved. Various contributions to Cherenkov angle resolution are again summarized in Table 4.8.

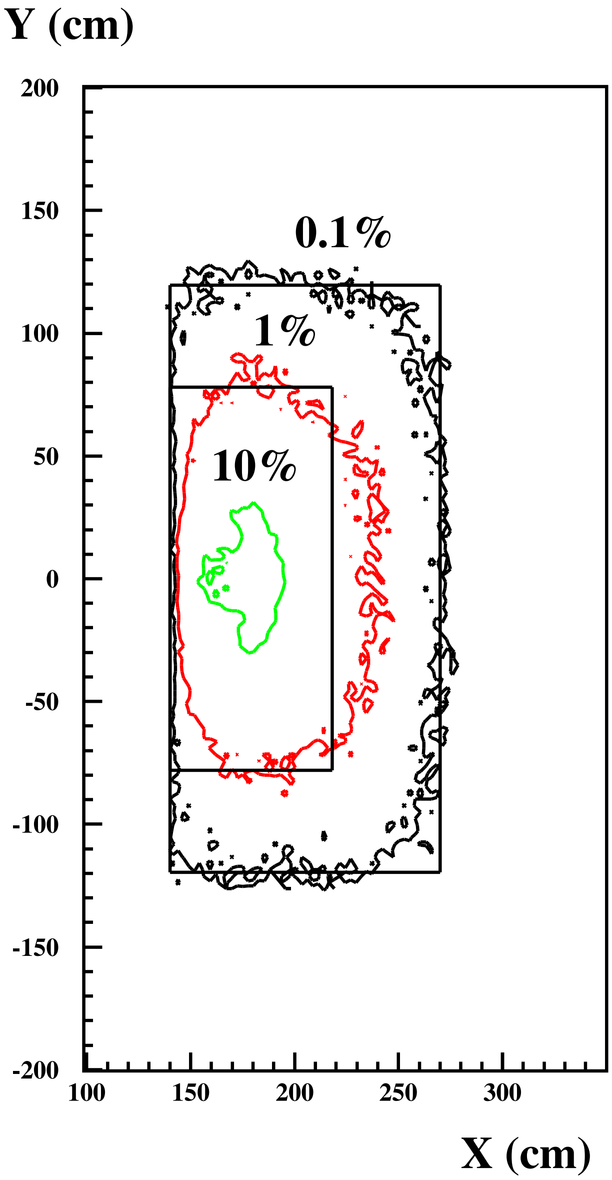


Figure 4.24: Contours of the number of Cherenkov photons per one MAPMT (2.6×2.6 cm^2) per event produced in C_4F_{10} by the faster track in reconstructed $B^0 \to \pi^+\pi^-$ decays. Rectangular boxes illustrate areas to be covered with MAPMTs. The inner box corresponds to the region of finer segmentation.

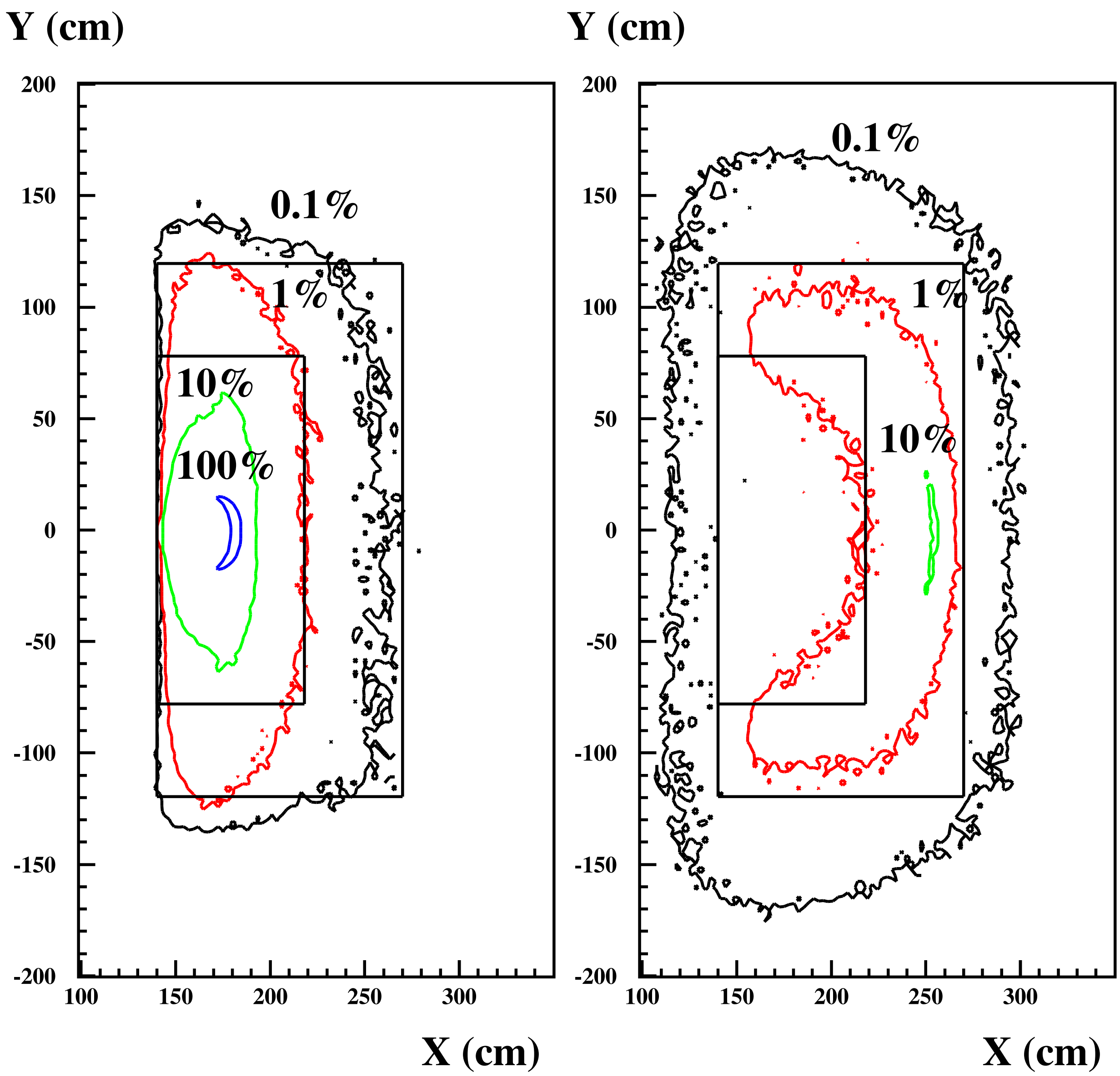


Figure 4.25: Contours of number of Cherenkov photons per one MAPMT (2.6×2.6 cm^2) per event produced by all tracks in generic $b\bar{b}$ event originating from C_4F_{10} (left) and aerogel (right) radiators. Rectangular boxes illustrate areas to be covered with MAPMTs. The inner box corresponds to the region of finer segmentation.

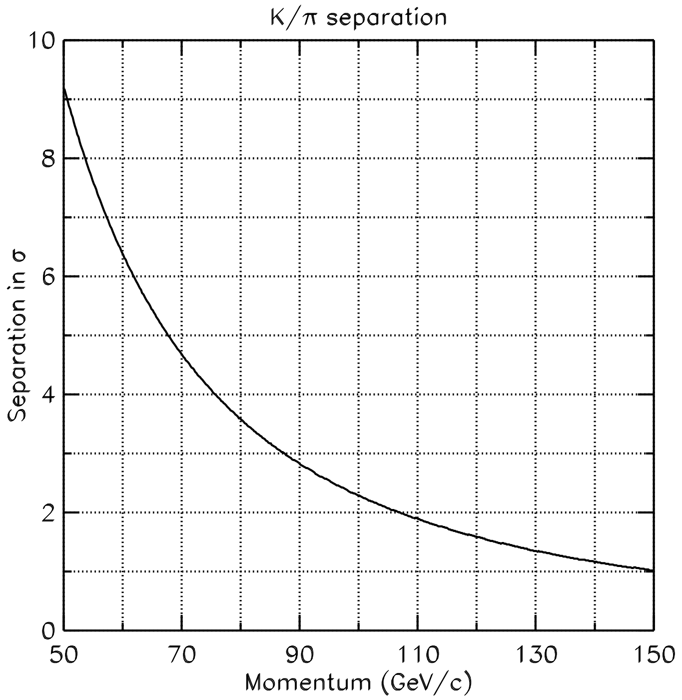


Figure 4.26: Expected K/π separation from C_4F_{10} radiator.

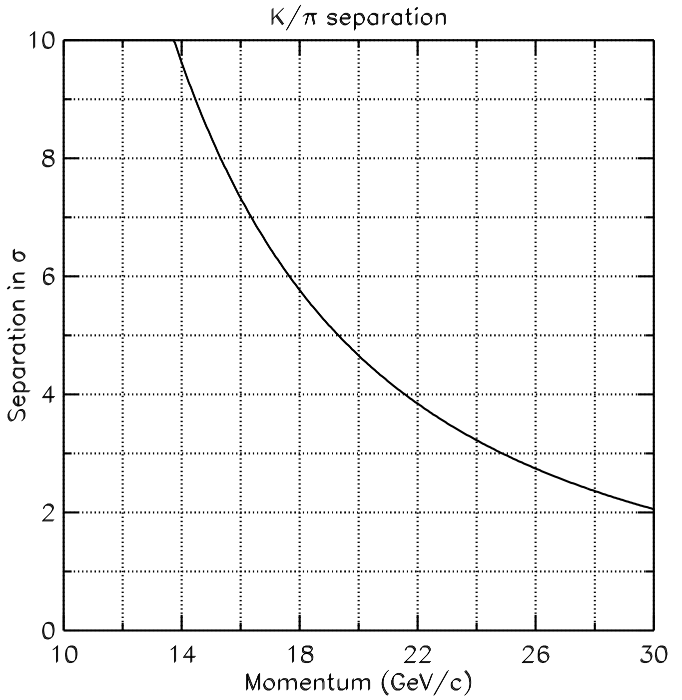


Figure 4.27: Expected K/p separation from aerogel radiator.

Bibliography

[1] W. Adam et al., Nucl. Inst. & Meth. **A343**, 60 (1994).

[2] J.L. Rosen, Proceedings of The Fifth International Workshop on $B-$Physics at Hadron Machines, Los Angeles, USA, Oct. 13-17, 1997, published in Nucl. Inst. and Meth. A408 (1998) 191; S. Kopar, Proceedings of the The 3rd International Workshop on Ring Imaging Cherenkov Detectors Weizmann Institute of Science Ein-Gedi, Dead-Sea, Israel, Nov. 15-20, 1998.

[3] P. Carter, Proceedings of the The 3rd International Workshop on Ring Imaging Cherenkov Detectors Weizmann Institute of Science Ein-Gedi, Dead-Sea, Israel, Nov. 15-20, 1998. "Proposal for Dual Radiator RICH for HERMES", E. Cisbani et al., Hermes Internal Note 97-005.

[4] LHC-B, Technical Proposal, CERN LHCC 98-4, LHCC/P4, 1998.

[5] R. Forty, CERN-PPE/96-176, Sept. 1996 published in Proc. of the 4^{th} Int. Workshop on $B-$physics at Hadron Machines, Rome, Italy, June 1996, F. Ferroni, P. Schlein (Eds.), North-Holland, 1996.

[6] R. Forty, "Use of lenses to increase the RICH photo-detector coverage", LHCb Internal Note LHCB/98-038.

[7] A. Duane et al., LHCb Internal Note LHCB/98-039.

[8] V. Grecco, A. Petrolini, "Study of a possible light collector system for multi-anode photo-multiplier tubes", LHCb Internal Note LHCB/99-005.

[9] S. Easo, Proceedings of the The 3rd International Workshop on Ring Imaging Cherenkov Detectors Weizmann Institute of Science Ein-Gedi, Dead-Sea, Israel, Nov. 15-20, 1998.

[10] N. Smale et al., "Evaluation of the multi-anode photo-multipliers for the LHCb RICH detectors", LHCb Internal Note LHCB/98-066.

4.2.6 Electromagnetic Calorimetry

4.2.6.1 Introduction

A thorough investigation of B decays requires the ability to study decay modes containing single photons, π^o's, and η's. Some of the important decay modes mentioned in chapter 1 are:

- $B^o \to (\rho\pi)^o \to \pi^+\pi^-\pi^o$;
- $B^o \to K^{*o}\gamma$ and $\rho^o\gamma$;
- $B_s^o \to \psi\pi^o,\ \psi\eta,$ and $\psi\eta^{'}$;
- $B^o \to K^{*o}\pi^o$; and
- $B^{\pm} \to K^{\pm}\pi^o$ and $\pi^{\pm}\pi^o$.

We have begun an intensive study of the problems associated with reconstructing electromagnetic showers in the environment of a hadron collider and using them to obtain signals for B mesons containing photons in the final state. While this study is still in progress, we are sufficiently encouraged by our initial results to develop a baseline design for an electromagnetic calorimeter which is capable of reconstructing photons. The electromagnetic calorimeter will also be used to identify electrons which can be used to tag the flavor of the associated B for CP studies or to isolate high mass electron-positron pairs that form exclusive final states. Examples are $B \to D^{(*)}e\nu$ (for electron tagging) and $B \to J/\psi K_s$, $J/\psi \to e^+e^-$, $K_s \to \pi^+\pi^-$.

Below in chapter 11, we discuss the challenges of isolating the small signals of the most interesting decay modes from the large backgrounds that are encountered in hadronic interactions. Here, we present the key requirements on energy and position resolution and on radiation hardness that the calorimeter must satisfy. We then present the BTeV baseline electromagnetic calorimeter design. In chapter 11, we show simulations of key signals and in chapter 14 we present a preliminary cost estimate for the proposed baseline device.

4.2.6.2 Key Requirements and Issues

To determine the appropriate size, segmentation, and technology for the calorimeter requires an understanding of the kinematic properties of the photons in signals, of potential sources of backgrounds, and of the radiation environment.

The area available to the calorimeter is located about 7.0 meters from the nominal interaction point and extends in Z for 1 meter. The distance from the beam to the floor is 2.5m. The transverse dimension which matches the acceptance of the spectrometer is a 4 x 4 meter box centered on the Tevatron beam line.

We have begun very detailed Monte Carlo studies of the following two benchmark processes to understand the signals we hope to measure:

- $B^o \to K^{*o}\gamma$ and
- $B^o \to \rho^+\pi^-, \rho^-\pi^+,$ and $\rho^o\pi^o$.

The results of the full simulation are presented in chapter 11. Here, we present only the 'generator level' study for $K^*\gamma$ and for $B^o \to \rho^+\pi^-$. In the first study, we simply require all tracks of the final state of interest to be within the angular acceptance of the BTeV spectrometer and to satisfy a minimum energy requirement. We chose these two modes because the photon in the $K^*\gamma$ decay has rather high momentum and P_t, while those from the π^o decay of the ρ^+ are much softer. The results of these studies are shown in Fig. 4.28 and Fig. 4.29. Based on these studies, the energy range of interest is extends from 1 GeV to 60 GeV and the angular range of interest extends out to 200-300 milliradians.

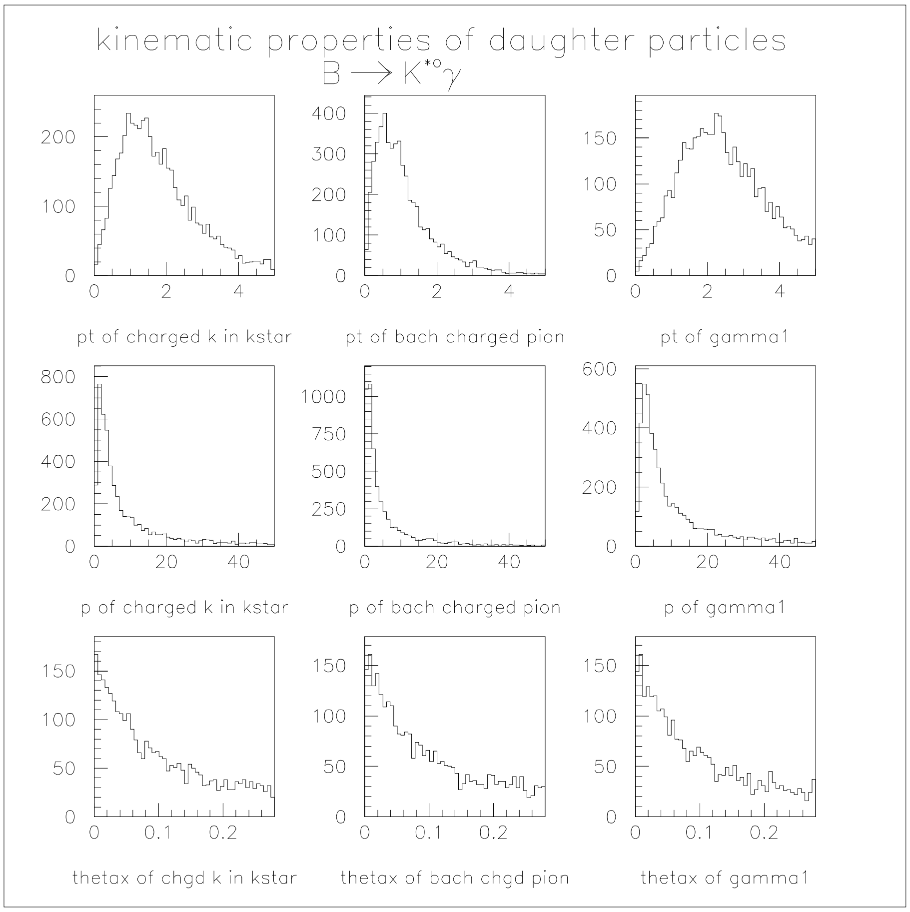


Figure 4.28: Kinematic properties of $B \to K^*\gamma$ where all generated tracks satisfy rudimentary acceptance requirements

To understand the requirements on energy and angular resolution, we have done a more sophisticated simulation. In this simulation, we carried out a full Monte Carlo generation and offline tracking of the charged tracks. For the photons, we smeared the momentum vectors

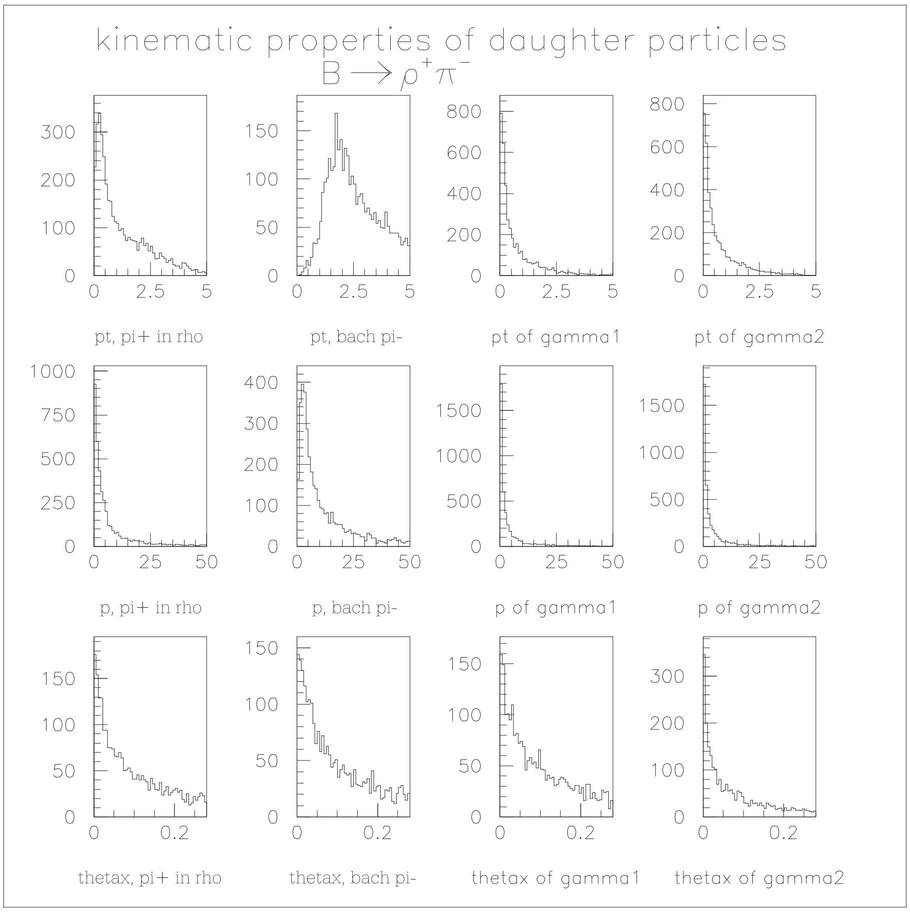


Figure 4.29: Kinematic properties of $B^o \rightarrow \rho^+\pi^-$where all generated tracks satisfy rudimentary acceptance requirements

of the generated photons in energy and angle and used the smeared momentum vectors in mass calculations. The energy is smeared according to

$$\frac{\sigma_E}{E} = \sqrt{\frac{a^2}{E} + b^2} = \frac{a}{\sqrt{E}} \oplus b, \tag{4.3}$$

with E in units of GeV. We refer to the parameters a and b as the stochastic term and the constant term, respectively. The angles were smeared by projecting the generated γ to the surface of the calorimeter, smearing the position there using a Gaussian of width $\Delta X/\sqrt{12}$ and $\Delta Y/\sqrt{12}$, where ΔX and ΔY are the transverse dimensions of the calorimeter elements. The angles were recalculated (using the primary vertex for the upstream point). Finally, a smeared γ momentum vector and its correctly propagated covariance matrix were calculated. These are used together with the charged tracks found in the offline analysis to reconstruct the B mass. The parameters a, b, ΔX, and ΔY can be varied and their effect on the mass resolution and other quantities can be obtained. Some results from these studies are given below in Table 4.9. The results agree well with the results from the full MCFAST simulation and shower reconstruction shown later.

4.2.6.3 Backgrounds

There are two major issues concerning backgrounds which influence the requirements on the calorimeter. The first has to do with problems in the reconstruction of objects – γ's, π^o's, electrons, η's, etc. – which can result in poor efficiency, the reconstruction of 'false' objects, and the reconstruction of objects with seriously degraded resolution. The second has to do with combinatoric backgrounds – that is, not knowing which of the many neutral objects appearing in a typical event to associate with the decaying B. We have learned that the backgrounds due to problems in object reconstruction place the most demanding requirements on the detector, except for the energy resolution which determines the overall mass resolution of the final state. Good mass resolution is a final defense against combinatoric backgrounds. In this section, we discuss the requirements that help reduce the object reconstruction backgrounds. Methods for defeating combinatoric backgrounds are discussed in chapter 11 below.

The dominant processes which mimic a single photon are:

- High momentum π^o's whose two photons are emitted within a very small angle. Because of the finite size of the calorimeter elements and of the electromagnetic shower, these can appear to merge or 'coalesce' into a single shower. Since there is an energy dependent minimum opening angle between the photons in π^o decay, this problem affects mainly the highest energy photons of interest to BTeV. A high degree of segmentation can reduce this background and an analysis which uses transverse shower shape variables can be effective at detecting coalesced photons up to quite high energies. Below we show that coalesced photons can be identified up to over 90 GeV with the proposed baseline calorimeter.

- π^o decays where the less energetic of the two photons is below the sensitivity of the calorimeter or falls outside the calorimeter's geometric acceptance. A π^o can decay so one photon takes most of the energy while the other photon has very little. If the ability exists to detect the low energy photon, it can be paired with the higher energy single photon candidates to see if they form a π^o and if they do, those high energy photons can be rejected as single photon candidates. However, since the BTeV calorimeter is far from hermetic, this approach is not very efficient. Instead, one has to understand which states containing π^o decays are potential sources of fake single photon signals and demonstrate that they have not polluted the single photon signal. For example, the decay $B^o \to K^{*o}\pi^o$ can mimic $B^o \to K^{*o}\gamma$ if the π^o decays very asymmetrically.

The dominant source of background to π^o identification is combinatoric background. All pairs of photon candidates (some of which will be false) must be considered in reconstructing π^o's. Random pairings make a rather smooth background in invariant mass. Since the mean number of photons recorded in the calorimeter in a typical B event is 30, the total number of combinations is very large. Many of the photons are unpaired – that is they come from a π^o, η^o or other parent but one of the daughter photons is outside the calorimeter's acceptance or below its energy sensitivity. Coalesced π^o's also mimic individual photons and participate in incorrect pairings. Shower overlaps both with other photons and hadronic showers in the calorimeter can distort the energy and angles of the photons and push the π^o masses to an incorrect value. Converted photons produce e^+e^- pairs which can bremsstrahlung, producing other photons that create background. The signal to background ratio of the π^o signal depends on the energy resolution since the narrower the π^o mass peak, the smaller is the integrated background under it.

Electrons will be identified by matching their energy with the momentum measured for the track by the spectrometer ($\frac{E}{p}$) and by looking at their transverse profiles. The dominant background which results in the false identification of objects as electrons is expected to be photon conversions due to interactions in the beam pipe, in the silicon near the primary vertex, or in the material of the RICH or forward tracker elements. Other backgrounds will be misidentified charged pions which charge exchange early in the calorimeter, and other rare processes which can mimic the behavior of an electron.

The ability to reject backgrounds will depend on the choice of the calorimeter technology, the segmentation, and to some extent the performance of the other detector elements. These processes are being evaluated with Monte Carlo tools (such as GEANT and MCFAST) and perhaps eventually with test beam studies.

We have identified combinatoric background as a serious problem. A final defense against the combinatoric background is mass resolution. The calorimeter is the only means of measuring the energy and angles of neutral objects in the event. The energy resolution of the calorimeter is a more significant source of mass smearing than its angular resolution. For the range of calorimeter technologies satisfying other BTeV requirements, the energy measurement of the photons will dominate the total mass resolution of the state. Table 4.9 shows the mass resolution for $B^o \to K^{*o}\gamma$, using an energy resolution parameterized as above with stochastic terms of 3.0%, 2.5%, 2.0%, and 1.5% and constant terms of 1.0%, 0.75%, and 0.55%, and 0.25%. Both the constant and stochastic terms are important contributors to the resolution because of the large energy range of the photons.

Table 4.9: Mass Resolution for Different Calorimeter Energy Resolutions (statistical accuracy is 2-3 MeV/c^2)

State	a %	b %	Mass Resolution MeV/c^2
$B^o \to K^{*o}\gamma$	3.0	1.0	54
	2.5	1.0	48
	2.0	1.0	49
	1.5	1.0	49
	3.0	0.75	46
	2.5	0.75	50
	2.0	0.75	42
	1.5	0.75	39
	3.0	0.55	45
	2.5	0.55	41
	2.0	0.55	44
	1.5	0.55	40
	3.0	0.25	41
	2.5	0.25	45
	2.0	0.25	40
	1.5	0.25	36

4.2.6.4 Radiation Hardness Requirements

A full calculation of the radiation environment takes into account radiation deposited in the calorimeter from the direct products of collisions, from the interactions of those products with the material of the spectrometer, from beam halo, and from backscatter and residual radiation from other detectors. We are in the process of carrying out such a calculation, but it is not yet complete. Here, to get an idea of the radiation environment the detector must tolerate, we show, Fig. 4.30 the energy dose in the calorimeter from photons, electrons, interacting hadrons, and non-interacting charged particles per (minimum bias) interaction. We set the radiation requirement based on tripling the deposition from these events to account for sources not included in this calculation. The calorimeter begins at ± 9.88 cm in both X and Y. The annual radiation dose is 60 KGrays at a radius of about 14 cm (20 mr). The dose falls off like a power of the radius to the 2.3 power from 20 cm on out. The detector needs to be radiation hard towards the middle and radiation tolerant on the outside. At 20 mr, the detectors nearest the beam will receive doses that are a quarter of those expected at the LHC (at greater radius but much higher luminosity and energy). We plan to use a solid state chip (the QIE) to digitize the pulse in or near the photomultiplier tube base. The radiation levels at the downstream end of the detector are therefore also of concern.

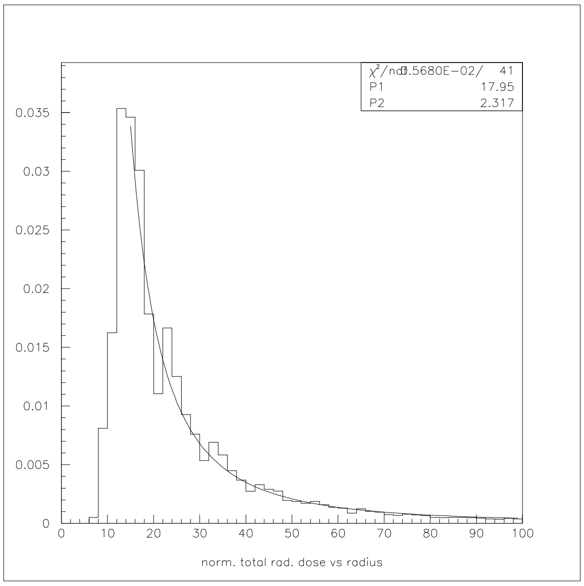


Figure 4.30: Total energy deposition in GeV per cm^2 per interaction (minimum bias) vs radius (cm) due to photons, electrons, interacting hadrons, and non-interacting (minimum ionizing) particles. The curve is the result of a fit to a power law, $A \times R^n$. The best fit gives $n = -2.3$. A rough conversion gives 20 KGrays/year for each 0.01 GeV per cm^2 per interaction.

4.2.6.5 Time resolution

The spacing of beam crossings at the Tevatron will be 132 ns. As long as the occupancy is low and there is some way of tagging which beam crossing each hit goes with, it is not necessary to have a signal which is so fast that it can be collected between crossings. However, it is certainly a convenience if this can be accomplished and may also reduce machine backgrounds as well.

4.2.6.6 Summary of Requirements

The above considerations lead to a range of requirements on various characteristics which are shown in Table 4.10. These requirements are not uniquely determined by our studies. They also reflect our current understanding of available technologies and their limitations. So, they have certain compromises already built in to them.

4.2.6.7 Baseline Technology

Several technologies were considered (from the point of view of performance, cost, and availability), including:

- A sampling calorimeter with lead and either scintillator or liquid argon

Table 4.10: Requirements of Baseline Electromagnetic Calorimeter

Property	Value
Maximum length	1 meter
Angular Resolution	1 mr at 700 cm
Energy resolution:	
Stochastic Term	1-3%
Constant Term	0.55-0.75%
Minimum angle	10-20 mr at 700 cm
Maximum angle	200-300 mr
Length	25 radiation lengths
Pulse pair resolution	must be able to associate pulse height to individual beam crossings with low pile up

- A total ionizing calorimeter based on either CsI or liquid krypton
- A calorimeter based on $PbWO_4$ crystals.

Of these, lead and scintillator was rejected because of poor resolution. Fermilab is reluctant to deal with the numerous safety issues related to operating liquid argon or liquid krypton in the C0 enclosure. These devices also have long charge collection times which can, however, be overcome by careful pulse shaping. Liquid krypton, in particular, looked like an attractive possibility. Liquid detectors are radiation hard. However, the constant term in the energy resolution may be considerably larger due to the relatively large amount of material before the active sampling material. CsI (undoped) was attractive from many points of view. However, it is not radiation hard enough to be used close to the beam. Furthermore, it is not quite fast enough to completely produce all its charge in under one crossing but it this problem is easily overcome with electronics that samples two or three crossings. Lead glass has been rejected because it is not likely to survive the yearly radiation dose.

After study, we have selected for our baseline a calorimeter made of $PbWO_4$ crystals. The BTeV $PbWO_4$ choice for an Electromagnetic Calorimeter (EMC) is largely based on the work of CMS. Our choice is based on several considerations, listed roughly in order of priority:

- It satisfies our requirements on energy and spatial resolution. Blocks of size roughly 2.5 cm can be made. CMS expects to achieve a stochastic term of 2.7% and a constant term of 0.55% using Avalanche PhotoDiodes (APDs) for readout. We plan to use photomultiplier tubes (PMTs) and expect to do better.
- This material is very resistant to radiation damage, especially when doped with either Nb or La. CMS requires the crystals to survive a dose of 50 kGy. This should be sufficient for BTeV.

- They are fast – they deliver 99% of their light output within 100 ns.

The properties of $PbWO_4$ which are important for the calorimeter are given in Table 4.11.

We describe some properties of the CMS calorimeter for purposes of comparison with the device we are proposing. CMS has chosen $PbWO_4$ for both barrel and endcaps. The crystals are tapered to point at the interaction region. The sizes differ, the barrel crystals are 230 mm long, while the endcap crystals, which follow a 5 radiation length preradiator, are 220 mm long. The endcaps are 24.7 mm square in the front, while the barrel counters are smaller at 21.5 mm square. BTeV will use untapered blocks of very similar dimensions. Many details about the $PbWO_4$ calorimeter can be found in the CMS writeup [1].

CMS will use Avalanche Photodiodes (APD) to read out the barrel and Vacuum Phototriodes to read out the endcap. These devices are necessary because of the high magnetic field in which the crystals are positioned. In BTeV we are far from any magnetic fields, so we can use photomultiplier tubes. This provides less noise, better resolution as will be described below, in a well understood and tested technology. In fact, BTeV could use the same tubes that KTeV used for their inner blocks, the Hamamatsu R5364, provided that a radiation hard glass such as C96-1 is used.

As part of its acceptance criteria, CMS requires that all of their crystals emit more than 10 photoelectrons per MeV into a XP2262B photomultiplier tube covering the back face of the crystal. Using a smaller phototube, we expect to collect in excess of 7 photoelectrons per MeV. CMS actually collects ≈4 photoelectrons per MeV into two 25 mm^2 APD's on each crystal.

Table 4.11: Properties of $PbWO_4$

Property	Value
Density (gm/cm^3)	8.28
Radiation Length (cm)	0.89
Interaction Length (cm)	22.4
Light Decay Time (ns):	5(39%)
	15(60%)
	100(1%)
Refractive Index	2.30
Maximum of emission (nm)	440
Temperature Coefficient (%/^{o}C)	-2
Light output/NaI(Tl) (%)	1.3
Light output (pe/MeV into a 2" PMT)	10

Despite all the work that has been done by CMS, there are significant open issues and risks in this approach. The main problem is growing crystals with sufficient quality and reproducibilty (yield), to lead to acceptable costs. There are only four producers working with CMS which currently have the ability to grow these crystals, and they will be kept busy for some time producing the crystals for CMS. These companies are in the Czech Republic, Russia and two in China. We will attempt to look for new sources, however.

It may be possible to use another less radiation hard crystal at the periphery of the detector and we are evaluating such alternatives. However, in this document, we describe a system made totally from $PbWO_4$ crystals.

4.2.6.8 GEANT Studies

The EMC geometry was defined as a rectangular array of crystals 24.7x24.7 mm^2. The length of the crystals was varied. A full simulation of the material properties was done with photons normally incident on the array, but smeared over the front face of the crystals.

i. Crystal length

We first explored the desired crystal lengths. In Figs. 4.31 and 4.32, we show the total energy deposited in the crystal for 2.5 GeV/c and 80 GeV/c incident photons, representing a low energy and an energy near the upper range of interest.

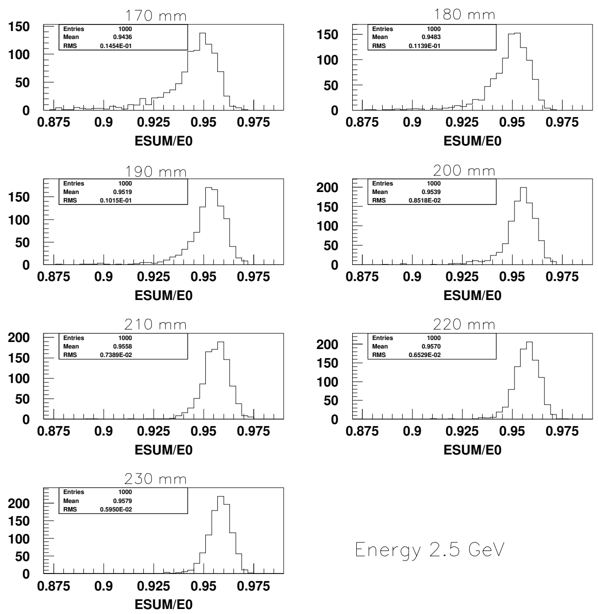


Figure 4.31: Energy deposited for 2.5 GeV/c photons normally incident on $PbWO_4$ crystals of the labeled lengths.

It appears that 220 or 230 mm lengths are required. In order to make this study more quantitative, we clustered the energy in a 5x5 array of crystals centered on the crystal with the largest energy deposit. Then we fit the distributions with a function that reasonably describes the central part of the distribution as a Gaussian with an asymmetric low energy tail. This function has been called the "Crystal Ball lineshape," and was named after a former experiment based on NaI crystals [2]. The function form is

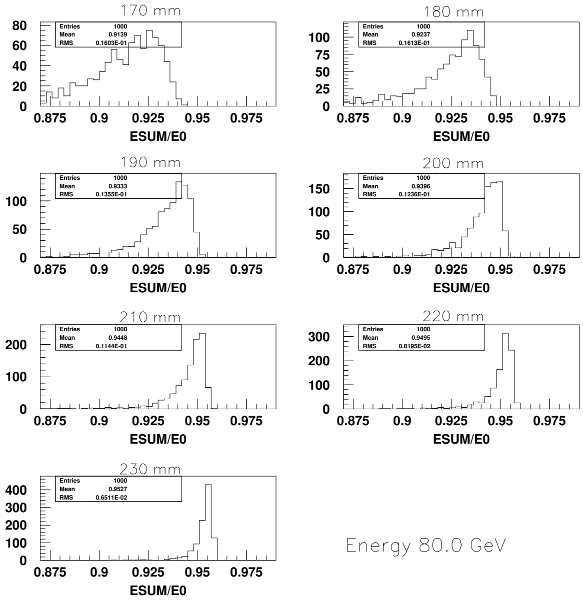


Figure 4.32: Energy deposited for 80 GeV/c photons normally incident on $PbWO_4$ crystals of the labeled lengths.

$$f(E|E^*, \sigma_{E^*}, \alpha, n) = \left(\begin{array}{ll} A \cdot \exp\left[-\frac{1}{2} \left(\frac{E^* - E}{\sigma_{E^*}} \right)^2 \right] & \text{for } E > E^* - \alpha \cdot \sigma_{E^*} \\ A \cdot \frac{\left(\frac{n}{\alpha}\right)^n e^{-\frac{1}{2}\alpha^2}}{\left(\frac{E^* - E}{\sigma_{E^*}} + \frac{n}{\alpha} - \alpha \right)^n} & \text{for } E < E^* - \alpha \cdot \sigma_{E^*} \\ \text{here } A^{-1} \equiv \sigma_{E^*} \cdot \left[\frac{n}{\alpha} \cdot \frac{1}{n-1} e^{-\frac{1}{2}\alpha^2} + \sqrt{\frac{\pi}{2}} \left(1 + \text{erf}\left(\frac{\alpha}{\sqrt{2}} \right) \right) \right] & \end{array} \right. \tag{4.4}$$

Here E is the measured energy, E^* is the "true" (or most likely) energy and σ_{E^*} is the energy resolution. To use this formula, the parameter n is fixed to value of about 5.

GEANT events for different energy photons and lengths were fit with this function for different energies and lengths varying from 210 to 230 mm. An example of such a fit is shown in Fig. 4.33.

The result of this study is summarized on Fig. 4.34, where the σ from each of the the fits is plotted. We see that the energy resolution determined by the crystal length and transverse size is hardly affected by using crystals of 220 mm, but there is a degradation going to 210 mm. This study uses energy deposited in the 5x5 array of blocks without the smearing due to photon statistics, so the resolutions are unreasonably good.

The energy resolution determined without photon statistics or electronic noise for 220

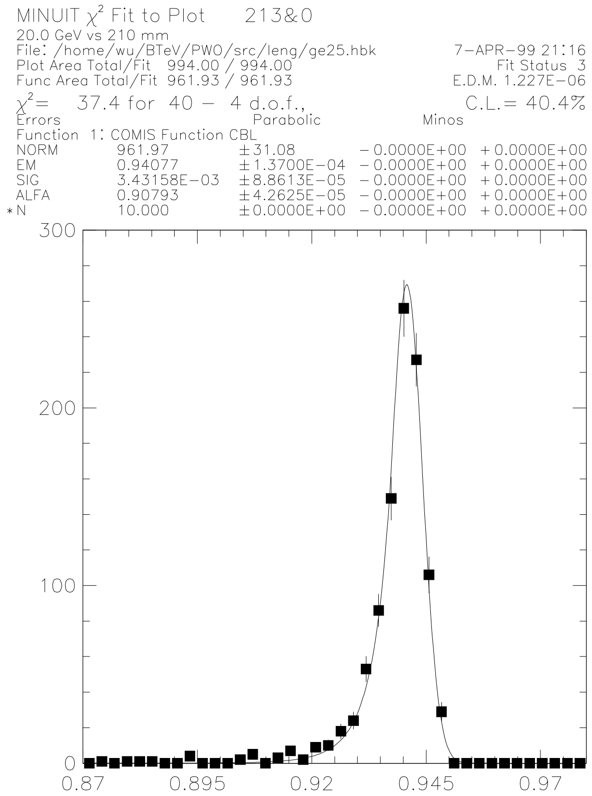


Figure 4.33: Fit of clustered energy deposited for 20 GeV photons normally incident on $PbWO_4$ crystals of 210 mm in length using the Crystal Ball lineshape. The x-axis is normalized to one at 20 GeV.

mm long crystals is shown on Fig. 4.35.

$$\sigma_E = \sqrt{\frac{a^2}{E} + b^2} \equiv \frac{a}{\sqrt{E}} \oplus b, \tag{4.5}$$

with $a = 1.08\%$ and $b = 0.2\%$, and E in units of GeV.

We can determine the component purely due to transverse leakage by fitting the distributions for total energy deposit and subtracting in quadrature the σ's from these fits and those with the 5x5 clustering. The transverse leakage component is shown in Fig. 4.36. We see the effect of transverse leakage can be characterized as

$$\sigma_E^{trans} = \frac{0.7\%}{\sqrt{E}}. \tag{4.6}$$

The transverse leakage component is sufficiently small that it is not necessary to attempt to add in more blocks to the energy sum

ii. Expected Energy Resolution

The final resolution in energy that we can expect from a $PbWO_4$ system will depend on several factors. One of these, is the transverse and longitudinal leakage. Another is

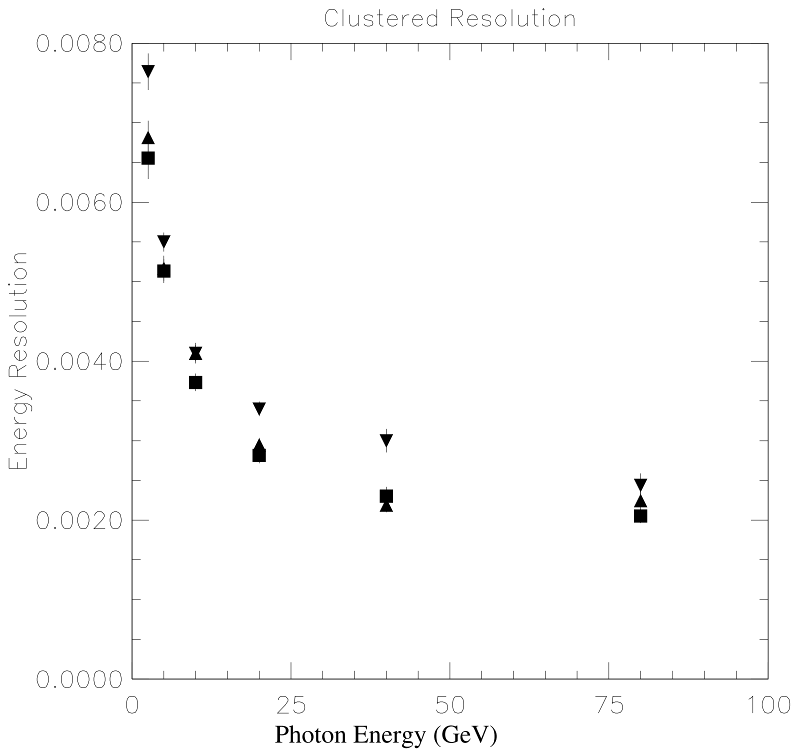


Figure 4.34: Energy resolutions as a function of incident photon energy for different crystal lengths using 5x5 clustering. The squares are for 230 mm, the triangles for 220 mm and the inverted triangles for 210 mm.

photoelectron statistics. These contribute to the so called "stochastic" term which has a $1/\sqrt{E}$ dependence. We will evaluate this term below. Other effects that could contribute include electronics noise, negligibly small for phototubes, variations in light output along the crystal length, errors in the relative calibration between crystals, and possible other yet to be identified sources. These latter effects contribute to the "constant" term.

We can use the energy resolution expected by CMS to guide our expectations. They base their study on 4 photoelectrons/MeV into two APD's. There is an excess noise factor in APD's due to correlated gain in the device. This gives them effectively a factor of two fewer photons. Thus in CMS, photostatistics gives a contribution of $2.3\%/\sqrt{E}$. The term for lateral containment is based on 21.5 mm^2 crystals and a 5x5 clustering algorithm. They find a term $1.5\%/\sqrt{E}$. Adding these together in quadrature gives $2.7\%/\sqrt{E}$. CMS estimates a constant term of 0.55%. They also include extra noise due to leakage currents in the photodiodes, and the effect of summing over ≈25 interactions per crossing each with 7 times more energy than the Tevatron. We do not have these later effects at a significant level.

Additional guidance can be found by viewing the performance of the KTeV CsI calorime-

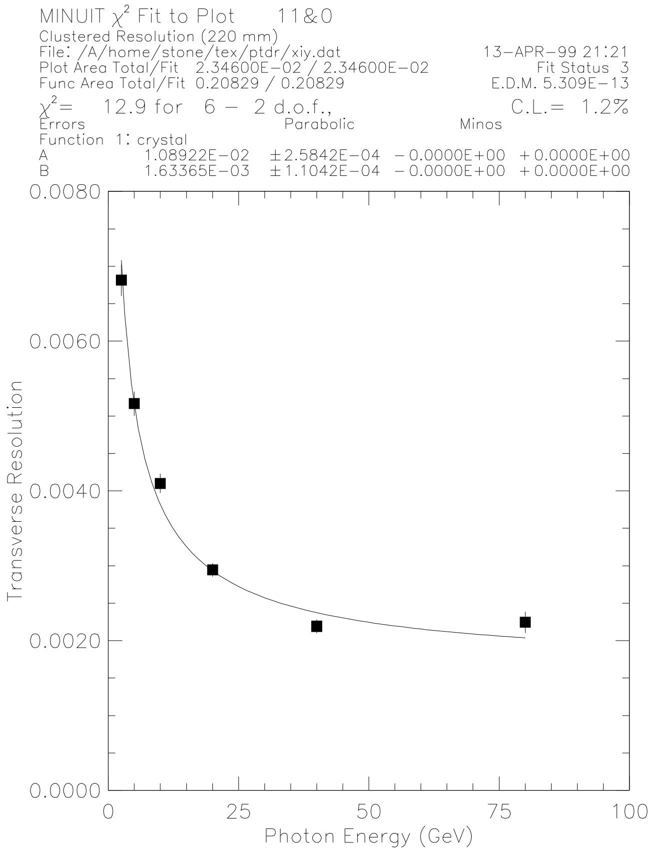


Figure 4.35: Energy resolutions as a function of incident photon energy for 220 mm long crystals using 5x5 clustering.

ter. They have achieved [3]

$$\sigma_E = \frac{2\%}{\sqrt{E}} \oplus 0.45\%. \tag{4.7}$$

The dominant cause of the constant term seems to be a differing crystal response depending on where on the face of the crystal the incident photon strikes. This effect was not observed in the CLEO CsI calorimeter [4].

The photon statistics term for BTeV, taking 7 photoelectrons/MeV is expected to be $1.2\%/\sqrt{E}$. We have already shown that the effect of transverse and longitudinal leakage in 220 mm long crystals is $1.1\%/\sqrt{E}$. Adding these two non-correlated components in quadrature gives $1.6\%/\sqrt{E}$. We have less knowledge of the constant term so we take the CMS estimated value of 0.55%. The expected energy resolution in BTeV then is

$$\sigma_E = \frac{1.6\%}{\sqrt{E}} \oplus 0.55\%. \tag{4.8}$$

To be conservative the physics simulations have been done using 2.5% rather than 1.6%.

iii. $\gamma - \pi^o$ *Overlap*

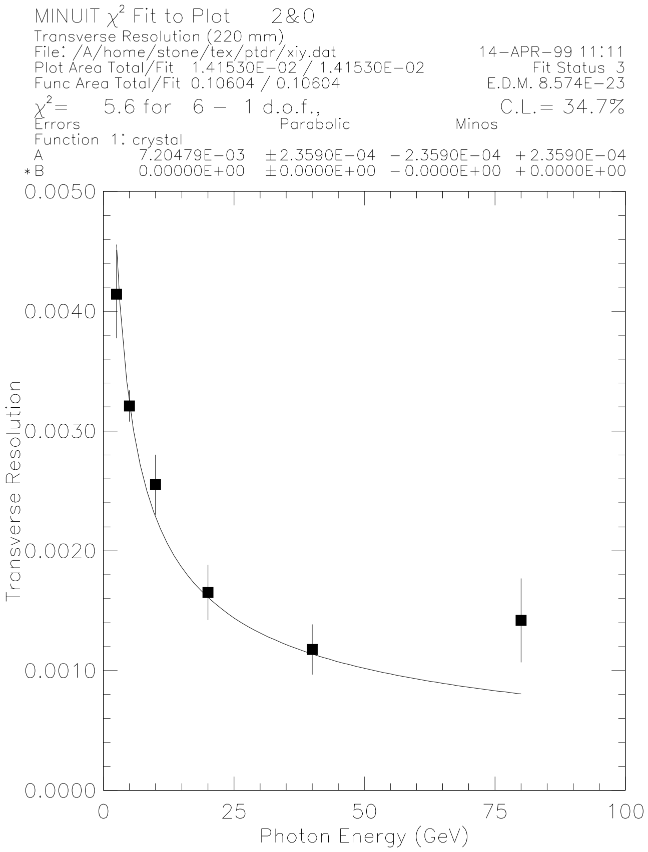


Figure 4.36: The contribution to the energy resolution from transverse leakage using a 5x5 array of crystals of 220 mm length and 24.7 mm^2 in cross-section.

It is clearly quite important for BTeV to be able to distinguish photons from π^o's. Many of the π^o's will be fast and could in principle appear as single photons. To study how well we could separate these two components with the proposed transverse segmentation, we generated both of them at fixed energies letting the $\pi^o \to \gamma\gamma$. We assumed only 4 photoelectrons/MeV, and a constant term of 0.55% thereby forcing a poorer energy resolution than we may achieve.

To study the probability of overlaps quantitatively, we calculate the second-moment of the energy distribution, S_2, defined as

$$S_2 = \frac{\sum_i \left(\overrightarrow{R}_i - \overrightarrow{R}_0\right)^2 \cdot E_i}{\sum_i E_i}, \tag{4.9}$$

where $\overrightarrow{R}_i$ is the position vector to the center of each individual crystal in the cluster, $\overrightarrow{R}_0$ is the energy weighted position vector of the entire cluster, and E_i is the energy deposited in each crystal. The S_2 distribution for both photons and π^o's of fixed energies is compared in Fig. 4.37.

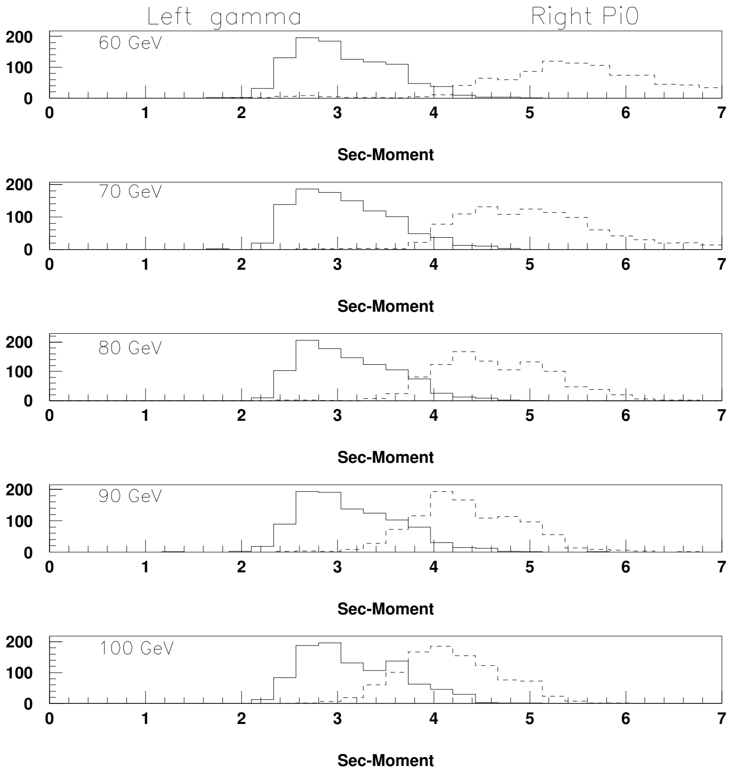


Figure 4.37: Second-moment energy distribution, S_2, for γ's and π^o's, as derived using GEANT and 4 photoelectrons/MeV.

In Table 4.12 we show the mis-identification rates for $\pi^o \to \gamma$ and $\gamma \to \pi^o$, both for 90% efficiency.

As most of the photons and neutral pions of interest in BTeV are below 60 GeV, the performance of the calorimeter in separating them is excellent.

iv. Position Resolution

The electromagnetic calorimeter is 7 meters from the center of the magnet. The resolution on projected angle is given in first order by the error in transverse position divided by 7 meters. The angular resolution enters directly in the mass resolution of fully reconstructed b and charm final states.

We again used GEANT to simulate the energy deposit in 220 mm long crystals taking 7 photoelectrons/MeV. Positions were then determined in both directions transverse to the crystal axes. As we have not seen any difference in the two directions, an advantage of square cross-section crystals, we show results only for one direction, called x. The position of the

Table 4.12: Neutral Particle Fake Rates for 90% Efficiency

Energy (GeV)	π^o to γ fake rate (%)	γ to π^o fake rate (%)
60	2.9	0.8
70	1.3	2.9
80	4.2	5.9
90	13.9	16.0
100	38.5	30.7

reconstructed photon is determined from the energy deposit by using

$$Xcl = \frac{\sum_{i=1}^{5} x_i \cdot E_i}{\sum_{i=1}^{5} E_i}, \tag{4.10}$$

where E_i is the energy reconstructed in each crystal, x_i is the center of each crystal and the sum extends two crystals in the direction of $-x$ from the highest energy crystal to two crystals in the $+x$ direction.

In Fig. 4.38 Xcl is shown versus the known position of the photon, $X0$ for 30 GeV incident photons. The characteristic "S-curve" shape is apparent. To evaluate the position resolution, the overall shape is removed by a fit and the remaining width of the curve gives the position uncertainty. In Fig. 4.39 the central value of the S-curve has been removed and the distribution in $dx = Xcl - X0$ fit with a Gaussian. The width of the distribution then is the r.m.s. position resolution.

The results of performing this study over a range of energies is shown in Fig. 4.40, where we show the r.m.s. position resoluton as a function of energy. The fit gives

$$\sigma_x = 3526\mu\text{m}/\sqrt{E} \oplus 217\mu\text{m}. \tag{4.11}$$

The projected position resolution then is much better than that assumed in the simulations, $\sigma_x = 247 \text{ mm}/\sqrt{12} = 7100\ \mu m$. These position resolutions give excellent angular resolution. At 6 GeV, the projected r.m.s. angular resolution is 0.2 mr.

4.2.6.9 Baseline Design

The baseline BTeV electromagnetic calorimeter consists of two wheels of 2.47 cm X 2.47 cm blocks of $PbWO_4$ crystals – one in each arm of the spectrometer. Each wheel is split down the middle to accommodate the beam pipe and to allow it to be moved away from the IR for servicing if necessary. The block length is chosen to be 25 radiation lengths (22 cm). We consider two options. They differ only in outer radius. The inner radius for both options is currently chosen to be 7 cm, which corresponds to ± 10 mr. For the first option the outer radius is chosen to be 140 cm, which corresponds to 200 mr. The radius is less than we would like, but costs less. The total number of blocks is 10,000 per arm or 20,000 total. In size and

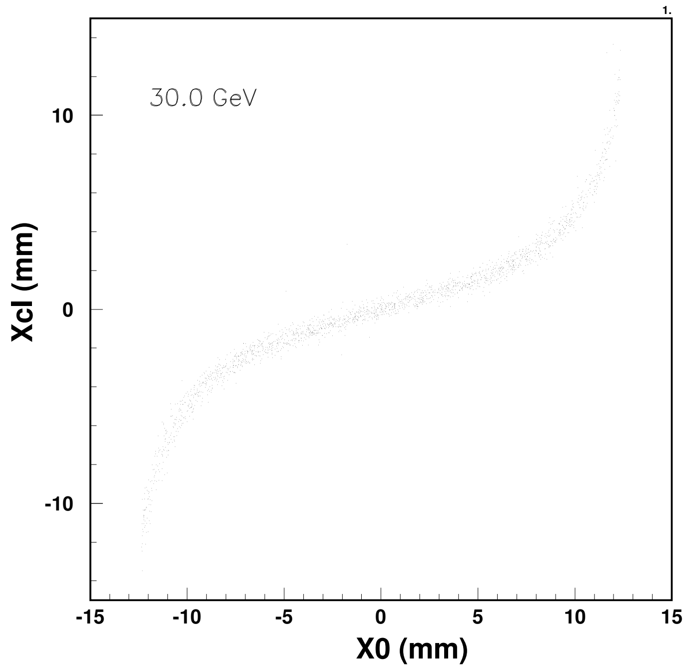


Figure 4.38: Measured photon position Xcl versus known photon position $X0$ for 30 GeV photons.

number of blocks, the system is only a little smaller than the CMS endcap calorimeter. It is this option that we call the baseline and have used in the simulations. The second option extends the calorimeter out to 300 mr. This is not only useful in getting more photon rate, but also extends the electron coverage. It does double the number of blocks, however, to 20,000 per arm.

At present, our blocks are rectangular solids. CMS has tapered its blocks so that the crystals almost point to the interaction point. We may consider doing this if further study shows it to be advantageous. The cross-section is chosen to be the same as the maximum size of the CMS end-cap blocks and the length is the same as the CMS endcap crystals.

There are detectors in each arm of the spectrometer. Each detector is housed in a temperature and humidity controlled hut. Temperature stabilization or some kind of correction is necessary because of the thermal coefficient of the $PbWO_4$. CMS plans to control the temperature to 0.1^oC. We will also employ temperature stabilization but have not yet designed the system. In addition to temperature stabilization, the gains are monitored with a laser pulsing system.

Detectors are read out with PMTs. The light emitted by the crystal peaks at 440 nm. The output of the PMT will go to a modified QIE chip located in or near the photomultiplier base. This chip provides a pulse height for each beam crossing. The expected light output is

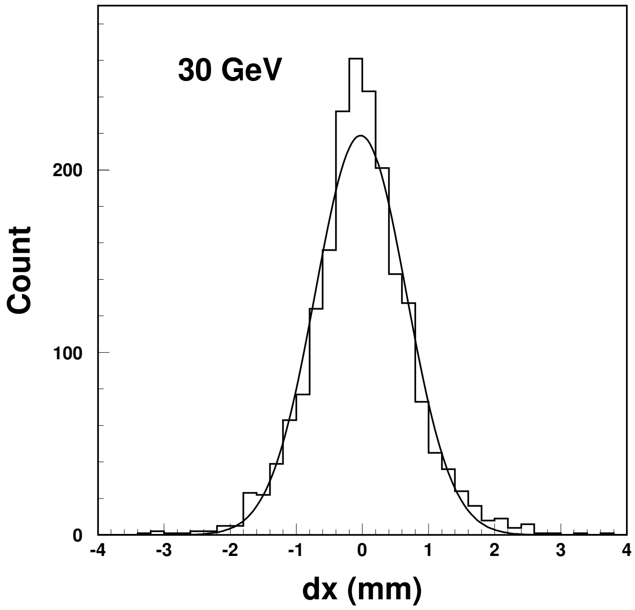


Figure 4.39: Distribution of $dx = Xcl - X0$ for 30 GeV photons. The fit is a Gaussian with mean consistent with zero and $\sigma = 0.7$ mm.

7000 photoelectrons at 1 GeV. The minimum sensitivity of the system is set so that a muon transversing the system, which corresponds to an energy deposition of 300 MeV, will have peak 50 counts above pedestal.

The detector is far enough away from the BTeV dipole so that there should be no need for special magnetic shielding. It is possible that we will magnetize the iron of the muon filter just downstream of the detector. In that case, we will put an iron shield plate between the calorimeter's PMTs and the muon filter.

Based on our extrapolation of information from CMS on the energy resolution, and our own GEANT studies, we expect a stochastic term of about 1.6%. The constant term is expected to be about 0.55%. The spatial resolution in both directions transverse to the Tevatron is expected to be about 3500 $\mu m/\sqrt{E} + 200$ μm, E in units of GeV.

Tests by CMS have indicated that these blocks can sustain integrated doses of 200,000 Grays. The only blocks that receive doses close to this number are the blocks right near the beam.

We have not yet started engineering on the support, but our detector is similar in size and shape to the CMS endcap electromagnetic calorimeter. A solution similar to theirs should work for BTeV. Because our detector is not embedded in a large device, and because at least at present we do not plan to use tapered blocks, our supports will be simpler and less expensive than those for CMS.

The properties of the baseline detector are summarized in Table 4.13. The GEANT studies described above demonstrate that the proposed device will satisfy the requirements of Table 4.9.

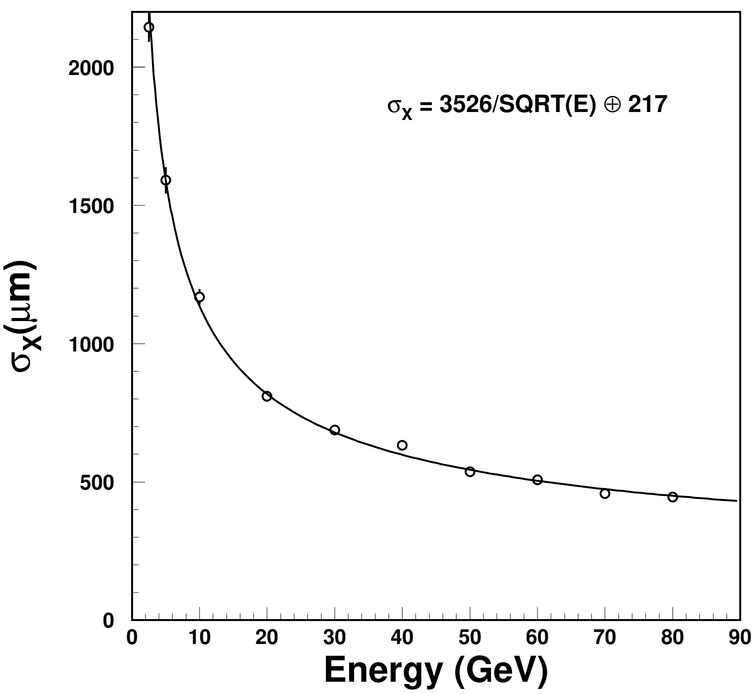


Figure 4.40: The expected r.m.s. position resolution for photons as a function of energy.

4.2.6.10 BTeV Electromagnetic Calorimeter R&D Program

At the time of our original R&D proposal, we did not have a plan for the Electromagnetic Calorimeter. We just listed several options, each of which had problems. Many of the problems we were worried about have been resolved by the development by CMS of $PbWO_4$ crystals as a viable calorimeter technology. We are taking full advantage of the work that CMS has done. We do have to carry out some R&D because we want to use photomultipliers rather than avalanche photodiodes. We will also be using a modified version of the QIE chip to read the data out. We want to obtain a few crystals and to try them with different types of PMTs and with QIE readout. We will have to explore cost issues with phototube manufacturers. Eventually we will want to expose a 5X5 (or better 7X7) stack of crystals to a low energy electron beam to study the resolution, leakage, temperature sensitivity, etc.

It is also critical that we begin to understand the issues related to producing crystals for BTeV. We will need to visit the existing production sites as well as to explore the possiblity of producing crystals with other manufacturers.

Table 4.13: Properties of the BTeV electromagnetic Calorimeter

Property	Value
transverse block size	2.47 cm X 2.47 cm
Block length	22 cm
Radiation Length (cm)	25
Front end electronics	PMT
Digitization/readout	QIE (FNAL)
Inner Dimension	$\pm$ 9.88 cm $\times$ $\pm$ 9.88 cm
Outer Radius (baseline)	140.0 cm
Total blocks per arm (baseline)	10000
Total blocks per arm (option)	20000
Outer Radius (option)	215.0 cm

Bibliography

[1] CMS, “The Electromagnetic Calorimeter Project Technical Design Report,” CERN/LHCC 97-33, CMS TDR 4 (1997).

[2] T. Skwarnicki, “A Study of the Radiative Cascade Transitions Between the Upsilon-Prime and Upsilon Resonances,” DESY F31-86-02 (thesis, unpublished) (1986).

[3] Private communication from Ed Blucher.

[4] E. Blucher, B. Gittelman, B. K. Heltsley, J. Kandaswamy, R. Kowalewski, Y. Kubota, N. Mistry, A. Bean, and S. Stone, “Tests of Cesium Iodide Crystals for an Electromagnetic Calorimeter,” Nuclear Instruments & Methods **A249**, 201 (1986).

4.2.7 Muon Detection

The BTeV muon system has two primary functions:

- J/ψ and Prompt Muon Trigger: Besides providing interesting physics, this trigger performs an important service role by selecting (bottom) events on which the more aggressive and technically challenging triggers (such as the vertex trigger) can be debugged and evaluated.

- Muon Identification: Many of the experiment's physics goals (rare decay searches, CP violation studies which require tagging, studies of beauty mixing, searches for charm mixing, etc.) rely on efficient muon identification with excellent background rejection.

We have selected a magnetic (toroidal) design combined with fine-grained tracking elements (described in detail in a subsequent chapter). This design permits "stand-alone" momentum/mass triggering: *i.e.* a muon trigger based solely on information from the muon detector. In addition, improved background rejection is possible via comparisons of this stand-alone measurement with momentum and tracking information from the rest of the spectrometer. The design of the system, including the fine-grained detector segmentation, has been chosen to reduce and uniformly distribute occupancies and to minimize confusion in pattern recognition (both at the trigger level and in offline reconstruction).

In order to provide a viable trigger, the system must get a rejection at Level 1 of a few 10^{-2}. The goal for muon misidentification is 10^{-3}.

Given the objective of a "stand-alone" trigger and the size limitations set by the experimental hall, one can make fairly general calculations that place specific (and restrictive) constraints on the design of the system. We will first describe these general calculations and use them to motivate the overall design of the muon system. We will then describe trigger efficiency and rejection studies we have performed for $B^0 \to J/\psi K^0_s$, which illustrate the performance of the design for a specific (and important) physics mode.

4.2.7.1 General Design Considerations

In the absence of any special momentum dependent coordinate weighting designed to reduce the effects of multiple Coulomb scattering, the fractional momentum resolution in a magnetic spectrometer can be parameterized by the expression:

$$\frac{\sigma_p}{p} = \sqrt{a^2 + (b\ p)^2} \tag{4.12}$$

In Equation 4.12, the a term depends on the bending power and multiple scattering environment of the detectors, while the b term depends on the bending power and and the detector layout and spatial resolution. Fig.4.41 shows that the potential rejection provided by a trigger for low momentum muons significantly degrades once the low momentum fractional resolution exceeds 25%. An $a = 25\%$ value means that the trigger rejects very soft muons at roughly the 4σ level. At higher momentum, where multiple scattering is less important,

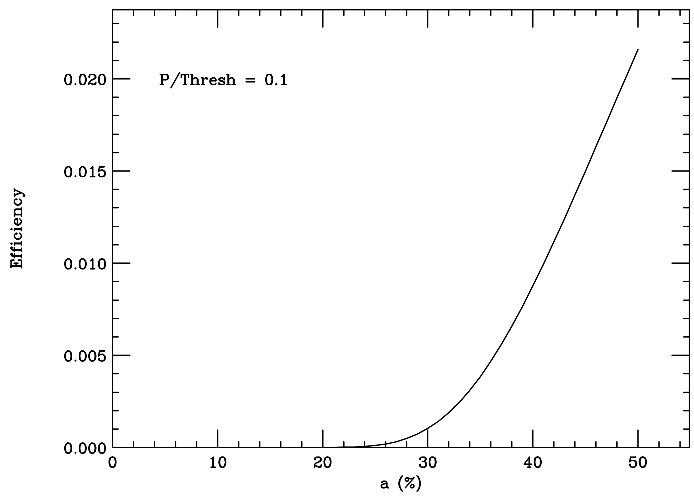


Figure 4.41: The efficiency, calculated in a simple Gaussian model, for a muon with 1/10 the threshold momentum to pass a momentum threshold as a function of a, the MCS term in the resolution formula. We say that a muon at the threshold momentum should fire the trigger 50% of the time. Note the dramatic rise in sensitivity above $a = 25\%$.

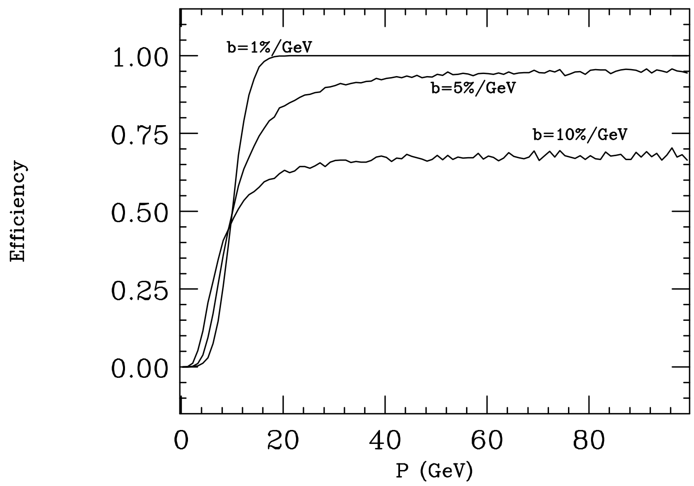


Figure 4.42: Simulated trigger efficiency as a function of muon momentum for a trigger designed to fire with a 50% efficiency at 10 GeV. The multiple scattering dominated term is set to $a = 25\%$ while the measurement dominated term varies from $1\%/GeV < b < 10\%/GeV$.

one becomes sensitive to the b term of Eq. 4.12. The high momentum resolution influences how sharp a momentum threshold one can make in a stand-alone muon trigger. Figure 4.42 illustrates this point by showing the trigger efficiency as a function of momentum for several b values.

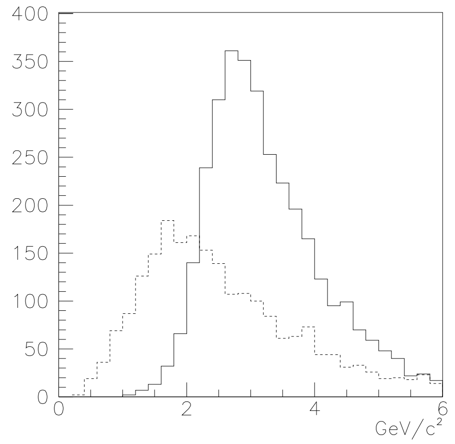


Figure 4.43: Reconstructed $J/\psi \rightarrow \mu^+\mu^-$ mass resolution for $b = 1\%/GeV$ (solid) and $b = 5\%/GeV$ (dashed), where b is defined in Eq. 4.12. Equal numbers of J/ψ candidates were generated for each case.

These considerations suggest minimum performance criteria of $a < 25\%$ and $b \leq 1\%/GeV$. As a further illustration, Fig. 4.43 shows the mass resolution obtainable at the trigger level for $J/\psi \rightarrow \mu^+\mu^-$ events for two values of b.

4.2.7.2 Possible Design Layouts

Several possible measurement and shielding layouts were studied before we reached the baseline design. In assessing possible layouts, we compute the momentum resolution using an error matrix which incorporates Gaussian models for both the detector resolution and multiple Coulomb scattering. The beam is modeled as a Gaussian beam spot with $\sigma_x = \sigma_y = 1$ mm and $\sigma_z = 30$ cm. In cases where there is redundant coordinate information, the error matrix assumes that the fit weighted the space points according to momentum in order to get the best possible momentum measurement.

Our early calculations showed that it is not possible to get a good momentum measurement using just the dispersion provided by the central dipole (and a constraint that muon system tracks point back to the interaction point). The extrapolation distance is too large given the level of multiple scattering up to the muon planes, making it infeasible to point the muon track segment back to the interaction region. A magnetic toroid with magnetic fields circulating in the $\hat{\phi}$ direction is the only (simple) option which allows us to trigger effectively using tracking information from only the muon system.

The first arrangement we considered is illustrated in Figure 4.44. In this design, one determines the momentum of a muon by measuring its slope in $\hat{\theta}$ before and after a toroid. The entering and exiting slopes of the muon are measured using two pairs of measuring stations (μ_1–μ_4) separated by one meter lever arms. At low momentum, where the multiple scattering dominates resolution, a single 1.5 Tesla, 1 meter thick toroid achieves the required

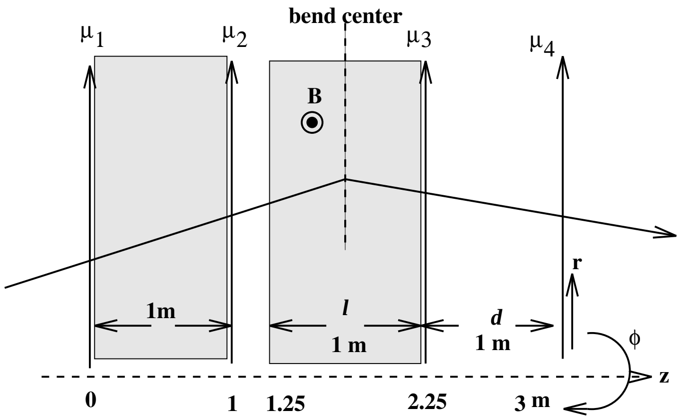


Figure 4.44: Layout of one arm of the muon detector in $2 \oplus 2$ geometry.

fractional momentum resolution of $a = 25\%$. In this arrangement the steel between the measurement stations of the upstream doublet (μ_1–μ_2) primarily serves to shield the muon system from hadrons and need not be magnetized. The requirement of good momentum resolution ($b \approx 1\%/GeV$) dictates a spatial resolution of $\sigma = 2.5$ mm. We will call this the "$2 \oplus 2$" geometry.

Note that the intrinsic rms spatial resolution ($\sigma = 2.5$ mm) required by this (and subsequent) geometries is consistent with what we require from timing considerations alone. Timing information is useful for rejecting background hits such as those associated with accelerator backgrounds (these hits come 60–70 ns before tracks from the interaction point). Assuming a fast gas drift velocity of roughly 9 cm/μs, a drift tube system with a maximum drift time of 30 ns has a maximum drift length of 2.5 mm (leaving 30 ns for charge integration). For a double layer of 1 cm diameter round proportional tubes, offset by half a tube in a "picket fence" design (to avoid dead regions between the tubes), the maximum drift time for the first signal from a track is 30 ns. This arrangement yields an rms resolution of $\sigma = 1.44$ mm.

The major problem with the $2 \oplus 2$ design is that the most upstream of the four detector stations is unshielded against hadronic showers originating in the electromagnetic calorimeter and is thus likely to be too confused for use in a first level trigger. One could, in principle, measure momentum using the downstream doublet μ_3 and μ_4 and a single upstream point μ_2. This geometry gives essentially the same a parameter, but for the same spatial resolution the b parameter is degraded by a factor of 2.8 because the effective upstream lever arm has been reduced to just the distance between the bend center and μ_2.

This led us to the revised design, our "baseline" geometry, illustrated in Fig. 4.45 where the upstream muon trajectory is measured using information from the nominal beam constraint. The effectiveness of this approach is illustrated in Fig. 4.46. The upper black curves give the fractional momentum resolution as a function of muon momentum for the case where only the downstream steel is magnetized and only information from μ_3, μ_4 and the nominal beam center is used. The three curves are for muons emerging at three different azimuthal

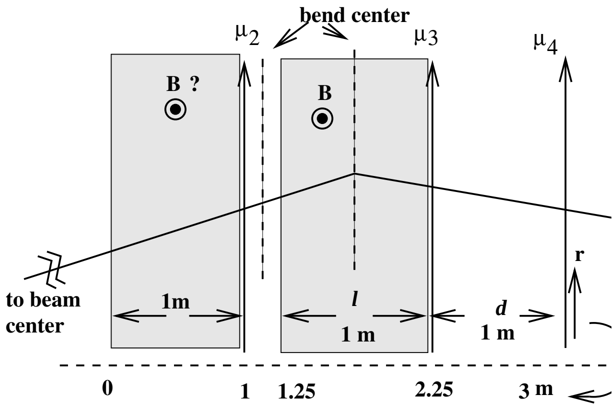


Figure 4.45: Layout showing the baseline geometry. The downstream trajectory is measured by μ_3 and μ_4. The upstream trajectory is measured using the nominal beam center with possible help from μ_2. In order to get sufficient bending power, both steel sections will need to be magnetized.

(ϕ) angles. The pronounced azimuthal variation in fractional momentum resolution arises because of constructive and destructive interference between the dipole and toroid fields. Figure 4.46 shows that at destructive ϕ angles, the a term exceeds our criteria of $a < 25\%$.

The incorporation of information from μ_2 significantly helps the fractional momentum resolution for the case of a single toroid at low momenta as shown in the three blue curves of Fig. 4.46. Essentially the fit used to obtain the "blue" resolution curves ignores the beam constraint at low momentum while exploiting it at high momentum through momentum varying weights. Such a resolution is obtainable off-line, but would be difficult to achieve in a first level trigger.

If one instead magnetizes both the upstream and downstream steel, the bending power of the toroid overwhelms the power of the dipole and a much simpler triggering scenario emerges which is illustrated by the lower seven curves of Fig. 4.46. The case where only a beam constraint and μ_3–μ_4 are used in the trigger is illustrated by the dashed red curves. The green curve shows the resolution one would obtain using a beam constraint and the downstream μ_3-μ_4 doublet if the central dipole were turned off. The green curve shows that the dipole provides very little additional bending power in the ultimate system. When information from μ_2 is included one gets the solid magenta curves. Again, curves are shown for three different azimuths. Each of these curves exhibit good performance relative to our criteria on a and b (over the complete azimuth).

Interestingly enough, Fig. 4.46 tells us that in a two magnet, toroid geometry, μ_2 contributes very little to the momentum resolution. In the absence of noise or confusion, one can design a very simple but effective trigger by correlating information from just the two $\hat{\theta}$ views of the downstream doublet consisting of μ_3 and μ_4 and the nominal beam center. Since this is the minimal amount of information required to obtain a momentum estimate, no momentum dependent weighting of detector information is used. We also have the benefit of deriving this trigger from the most shielded chambers of the muon system, minimizing

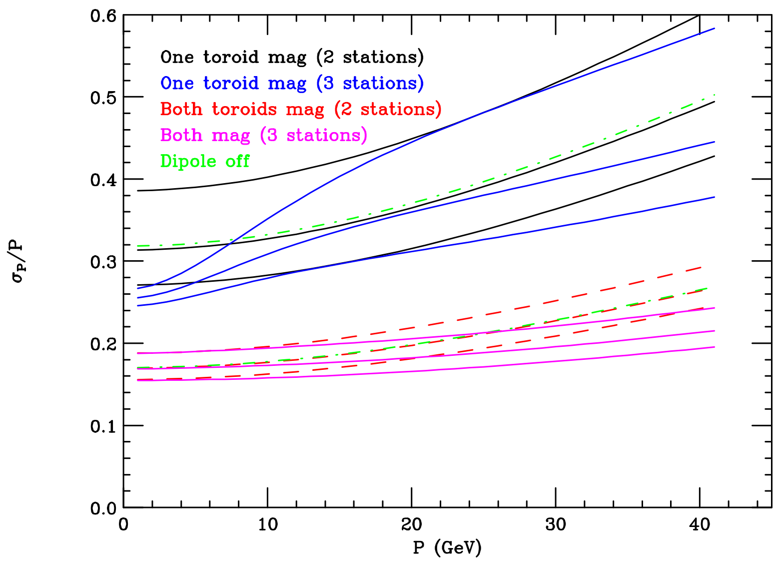


Figure 4.46: Plots of the fractional momentum resolution as a function of momentum for various readout scenario's of the BTeV muon system. All plots assume chambers with 2.5 mm rms spatial resolution. The plots appear in groups of 3 corresponding to $\phi = 270^o$ (best) 0^o and 90^o (worst). The ϕ dependence of the resolution illustrates interference between the central dipole and muon toroid(s). The black and red families of curves illustrate the case where only the two stations after the second toroid are used in the trigger. In this case the trajectory upstream of the toroid comes entirely from the beam constraint. The blue and magenta curves illustrate the case where the station between the two toroids is used to help determine the trajectory prior to the toroids. The green curves assume only the final two stations are used, and that the dipole is turned off .

hadronic punch-through effects.

Although μ_2 is not required to measure muon information in our two toroid scenario, it is likely to be of substantial use in eliminating false triggers due to noise and confusion. Figure 4.47 illustrates just one example by showing how μ_2 can be used at the trigger level to eliminate false triggers from soft muons with an accompanying δ-ray.

4.2.7.3 Baseline Design

Based on the above, our baseline muon system consists of two 1 m thick iron toroids and three measurement stations with spatial resolutions of $\sigma = 1.5$ mm. Each of the measurement stations is separated by approximately 1 meter. The spatial resolution is set by both timing considerations and momentum resolution requirements.

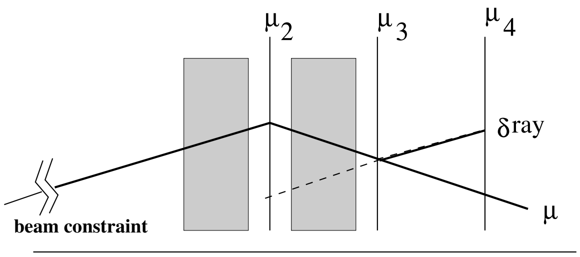


Figure 4.47: This figure illustrates the usefulness of μ_2 in reducing false triggers due to confusion from δ-rays. We illustrate a trajectory for a low momentum μ which is severely bent by the two toroids and therefore extrapolates far from the nominal beam center. As this muon emerges from μ_3 it emits a wide angle δ-ray which leaves an additional hit in μ_4. In the case illustrated, the two hits in stations μ_3 and μ_4 from the δ-ray point back to the nominal beam center and mimic a high momentum muon. However, the δ-ray trajectory extrapolates very far from the muons hit in μ_2 and thus fails to satisfy a trigger which requires a confirming hit in μ_2.

4.2.7.4 Trigger Studies for $B^0 \to J/\psi K_s^0$

We have studied the triggering performance of our baseline muon system using MCFAST. We used minimum bias events to study rejection ratios and $B^0 \to J/\psi\, K_s^0$, where $J/\psi \to \mu\mu$, to investigate trigger efficiency. The minimum bias events were generated with Pythia and include elastic scattering, single and double diffractive, low p_T scattering and semi-hard QCD $2 \to 2$ processes. Events are overlapped to simulate multiple events per crossing. The number of events per crossing is generated from a Poisson distribution with average two. In generating the $B^0 \to J/\psi\, K_s^0$ events, minimum bias events are also added so that the total number of interactions per crossing is also Poisson with an average of two. For the studies presented below, we only consider "real" muons, which can be prompt or can come from decays in flight. In minimum bias events, the muons come almost entirely from pion (70%) and kaon (30%) decays. "Fake" muons from punch through, delta rays, and noise hits in the chamber are not yet simulated.

In the calculations below, efficiency is the ratio of accepted events to generated events. This includes geometrical acceptance and the selection cuts. For our purposes, a signal ($B^0 \to J/\psi\, K_s^0$) event is considered to be in the geometrical acceptance if both tracks from the J/ψ muons are tracked by the silicon and make it through all stations of the muon system. The muon system is assumed to have an outer radius of 2.4 m and the inner radius is varied between 10 cm and 20 cm (10 cm is the smallest possible inner radius given the inner radius of the toroids, while 15–20 cm is the smallest inner radius that can probably be achieved with the baseline design). At that point, both minimum bias and signal events are treated identically. They are accepted if muon(s) in the event pass cuts on momentum, transverse momentum, and/or a minimum invariant mass cut on muon pairs. In all cases where two muons are required, they must have opposite sign and be in different octants in ϕ (each detector plane in the baseline muon system is divided into eight pie shaped octants).

The pair of muons must also be in the same arm of the detector.

The momentum and transverse momentum of each muon in the event is obtained by smearing its generated momentum using the appropriate resolution (see Fig. 4.46) and smearing function. We then vary the cuts on momentum, transverse momentum, and invariant mass of pairs of muons and make two dimensional plots showing rejection of minimum bias events on one axis and J/ψ efficiency on the other. An example plot is shown in Fig. 4.48. We have made similar plots for differing inner radii, one versus two toroids, and so on. Based

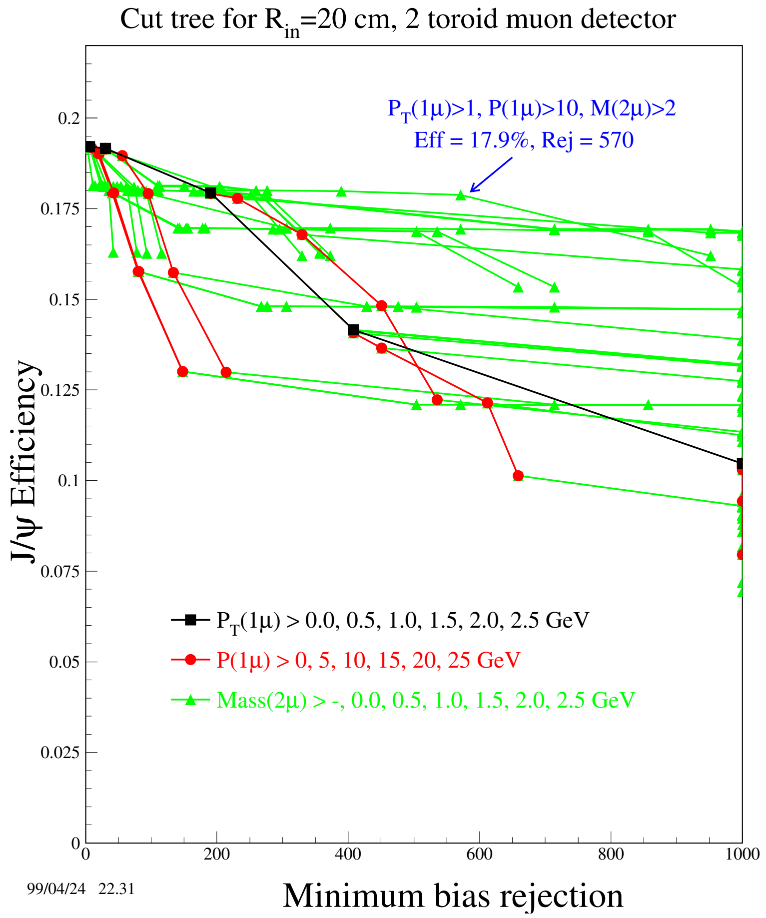


Figure 4.48: Absolute J/ψ efficiency versus minimum bias rejection for cuts on the minimum muon p_T, the minimum muon momentum and the minimum invariant mass. Black lines connect the squares which show the variation as the minimum muon p_T is increased. From each black square a red line connecting the circles shows the change as the minimum muon momentum is increased. From each red circle a green line connecting the triangles shows the variation as the minimum invariant mass requirement is increased. A cut of "-" indicates no cut. Results are for a two-toroid muon detector with inner radius of 20 cm.

on these studies, we find typical rejection ratios of 500-600 with efficiencies of roughly 18%. Note that the geometric acceptance (no momentum or mass cuts) of a 10 cm inner radius system is 22%. Based on these studies, we have found:

1. Most of the rejection power comes from requiring minimum p and p_T for two muons in the event. An invariant mass cut is beneficial, but is not required to obtain large rejections with good efficiency. If it turns out that an invariant mass cut is difficult to implement in the first level trigger, we will not be at a severe disadvantage.

2. An inner radius of 15–20 cm is acceptable. If we insist on a rejection ratio of 500 or more, the efficiency for signal events is 18.8% for an inner radius of 10 cm and 17.9% for an inner radius of 20 cm (20 cm is worst case).

3. Two toroids are needed. If we require a rejection ratio of 500–600, the efficiency for signal events drops significantly (from ~18% to ~14%) in going from two toroids to one. In fact, it appears that the only hope for a single muon trigger is the two toroid case: acceptable rejections are not possible with only one toroid.

4.2.7.5 Summary of Detector Requirements

To summarize, the BTeV muon system should have the following attributes:

1. Robust: It will be very difficult to access the detectors once they are installed.

2. High Rate Capability: Rates of a few kHz/cm^2 are possible in parts of the system (based on calculations of accelerator backgrounds).

3. Good Time Resolution: Timing information on hits with a resolution of 50–60 ns will allow rejection of hits due to accelerator backgrounds.

4. Magnetic: Two toroids of 1.5 Tesla field and roughly 1 m thickness (in each arm).

5. Fine-Grained Spatial Resolution: Tracking resolutions of 2–3 mm are required.

6. Longitudinal Compactness: Given the size constraints, the detectors should be of minimal thickness.

7. Minimal Neutron Susceptibility: We will avoid the use of hydrogenic detector elements.

8. Uniform Occupancy: A geometry designed to minimize pattern recognition confusion and minimize occupancy at smaller radii.

The details of the baseline muon system design are discussed in a subsequent chapter.

4.3 Evolution of the C0 Program

We want to congratulate Fermilab on the completion of a new experimental area at C0 which is capable of supporting a state of the art bottom and charm decay experiment. We have carefully designed BTeV to be a good match to this new enclosure. We propose to develop the experiment in stages. Some of these stages involve R&D into key technologies required for the experiment. Others involve a program of physics and background measurements. Only by doing tests under actual running conditions does one truly confront all the problems of efficiency, backgrounds, and systematic errors in the design.

The rate at which the full experiment can be mounted depends on many things. One is when the lowβ quadrupoles will be available, since this determines the earliest time when we can have collisions at high luminosity. Another is when the R&D for each component is completed. A third is, of course, when manpower for design work and money to build detectors and buy the electronics and computing become available. Finally, all the equipment must be installed, integrated, and made to work.

We assume that the running schedule around which we plan our activities is: each year there will be about 9 months of collider running; that there will be a three month shutdown for maintenance during which installation can take place; and that at some point after a few years of running, the 3 month shutdowns will be arranged back-to-back so that there can be a reasonable period for envisioned upgrades.

Our first goal is to get the 'large components' installed in C0 during the shutdown for Run II installation. At the minimum, this means the SM3 analysis magnet. If we can locate the steel for one or both of the muon detectors, we would like to get that installed too. Once these components are in, all other installation activities are similar to installing components in a fixed target experiment. Individual modules of the detector can be moved into the enclosure and installed in short shutdowns of a few days.

BTeV R&D is proceeding well. Pixel R&D is well along and we would like to complete it in 2002. Two different readout chips, bump-bonded to detectors will be tested in the 1999 fixed target run. Trigger R&D is also proceeding well and will be completed in 2001. The design of the forward tracking system, muon chambers, particle identification system, and electromagnetic calorimetry can be completed in 2001. Prototype modules of the forward tracker and the muon chambers will be tested in the 1999 fixed target run.

Starting at the beginning of Run II we will put counters into C0 to measure backgrounds near the IR and in the region of the muon detector and RICH; to set up the prototype modules of the muon detector and the forward tracker and operate them in the C0 enclosure; to install a wire target and study the interaction rate and backgrounds, and establish the effect on the other collision regions.

In 2000, we also plan to install three pixel planes inside the SM3 magnet and several forward tracking planes, along with trigger scintillators. We will use this arrangement to study the initial full size pixel detector chips in a prototype housing inside the beam pipe. Interactions on a wire target will be used. We are also proposing to establish the ability to test electronics near the beam by placing them in existing Roman Pots.

We plan to have our first $p\bar{p}$ collisions in 2002 using an intermediate $\beta*$ IR constructed from existing quadrupole magnets. This arrangement will not permit us to run simultaneously with both CDF and D0 since separators will need to be turned off to bring the beams into collision at C0. We envisage intermittent running perhaps using the ends of stores. We will install 10% of the final pixel system, a serious full scale system test of both the pixels and the trigger. We will also measure the backgrounds in the collider mode. We will interleave wire mode and collider running and test as much of the vertex tracker and vertex trigger as possible.

In 2003 we plan to start installing the rest of the BTeV spectrometers, and have some collider running after each installation period.

We look forward to full runs in collider mode building up to full luminosity in 2004 and 2005.

With this plan, we can steadily build our understanding of the detector over a period of a few years and can be ready to do physics with a tested analysis program in place when the full detector is complete. It also allows the lab to spread its resources, people and money, over several years. Following this plan, Fermilab will have a serious forward collider B experiment well in progress before the LHC turns on.

4.4 Completing the Simulation of the Baseline Detector

The simulation work on all-charged states is quite far along and has dominated the requirements for the baseline detector. The study of states with photons and π^o's has just begun. More physics channels will be investigated and background calculations for states with photons will be pursued aggressively even though they require us to run a very large number of events. The current baseline detector has been modeled and extensively studied and we are now confronting very detailed problems which involve more global optimizations – those requiring an understanding of the impact of various choices on the whole ensemble of physics topics. For example, the thickness of the pixel detector is a problem not so much for its effect on charged particle tracking but for its impact on photon reconstruction because so many photons undergo conversion to e^+e^- pairs.

We have also made much progress in improving the detail of the simulation and bringing it closer to reality. Significant design work is occuring on nearly all the detectors. As the design of the detectors begins to firm up, we are feeding information into the simulations to make them more realisitic with respect to cell sizes, resolutions, and material budget. Much of the work in the next year will be in modeling the details of the detector elements and understanding more complex issues and detector dependent issues such as pattern recognition and trigger rates to make sure that the proposed design can actually do the measurements we are interested in.

The trigger simulation includes pattern recognition in tracking the charged particles through the silicon detector. More detailed hit generation and digitization have been implemented so that we can more fully understand the efficiency and rejection of the trigger. We have begun to investigate other trigger algorithms as part of the trigger R&D effort. We are pursuing schemes with fewer planes to reduce the amount of material in the vertex detector and to reduce its complexity. Further simulations will be required to fully understand the effects of multiple interactions on the tracking and vertexing efficiencies although much work has already been done in this area. Serious work has been carried out on the Level 2 and Level 3 trigger algorithms and the overall trigger and data acquisition architecture.

Detailed simulations of the expected background rates in the C0 hall have not yet been performed. We will study backgrounds from all sources including those generated in the machine.

There are tools available to study the beam backgrounds [1]. Proton-antiproton collisions are typically simulated with DTUJET93. Beam halo interactions with scrapers and particle tracking through the Tevatron are modeled with STRUCT. Particle loss induced hadronic and electromagnetic showers, and secondary particle transport in accelerator and detector components, including shielding, are simulated using the MARS simulation package. We expect to carry out studies of the beam related backgrounds in cooperation with the Beams Division as the design of the detector and interaction region move ahead. We also are planning to undertake a series of measurements of the environment in the C0 Collision Hall as soon as collider operation resumes.

The simulation tools are now in place to study all the physics and detector issues that need to be addressed to make the final refinement of the BTeV detector. Over the next year, we expect to investigate a wide variety of physics topics and detector issues connected with bottom and charm physics and improve the background studies and to be in a position to make a final "global optimization" that will result in a final design that will give BTeV the greatest possible physics reach.

Bibliography

[1] From the C0 Workshop, Fermilab, Dec 4-6, 1996. See http://fnphyx-www.fnal.gov/conferences/czero/simulations_summary.html

Chapter 5

Pixel R&D

5.1 Introduction

The BTeV pixel detector is central to the capability of the BTeV experiment. At the same time, it stretches the technical frontier beyond what has so far been achieved in other experiments. In these circumstances, the plan has been to gain experience in the various required technologies and to demonstrate the necessary component capabilities that will be needed for the final device. In following this plan, as much as possible we have collaborated with industry and with groups working on LHC experiments developing similar detectors.

The major components for the pixel detector system are the sensor, readout chip, sensor-to-readout-chip connection, mechanical support and cooling, high-density interconnection between the readout chips and the control chips on the detector plane peripheries, and the multiplexing and data transmission to the first-level trigger. Details on the current state of R&D on these areas from various groups and industry can be found in [1]. For some components, we have demonstrated capability beyond that required for the baseline BTeV detector (e.g., 30μm-pitch bump bonding for sensor-to-readout-chip connection, while the baseline calls for 50μm pitch). For others we have demonstrated capability which meets the tough BTeV specifications (e.g., front-end amplifier of the FPIXi designs). For yet other components, we can make straightforward extensions to devices already tested which, when modified, will meet the BTeV needs (e.g., readout architecture). Finally, there are areas where additional work will be needed to get to the final Technical Design Report level (e.g., higher-level integration of components and radiation-hard implementation of the ASIC design). One choice that we will have to make in implementing a radiation-hard front-end is between deep sub-micron CMOS processes and more traditional radiation-hard technologies. We are just beginning to gain experience with yields for our readout chip designs, and have started a program to test yields at the component level too (e.g., for bump bonding).

On the road to gaining local experience and demonstrating necessary components, we have an active program of work in the MT test beam. We are ready to test two versions of Fermilab readout ASICs bump bonded to sensors (with pixel size of 50μm by 400μm). By the time of the TDR, we should have gained experience in assembling and operating multichip modules which will be the basic building block of the final BTeV pixel detector. We also have a physical simulator of the electrical noise from the Tevatron beam for testing electronic components and their shielding inside the beam vacuum-pipe. We are also preparing to place devices into the Tevatron in borrowed Roman Pots as a test of the designs which grow out of the simulator experience.

Finally, we have begun to develop industrial contacts and joint efforts with other experiments on the remaining areas. We have ordered test components for high-density interconnects (HDIs), fiber-optic data transmission, support structures, and cooling channel prototypes. Test stands for these are in various stages of construction, and making rapid progress.

5.2 Sensor development

5.2.1 Introduction

The sensor unit cell size is closely related to the achievable spatial resolution and to the occupancy in the system. The sensor thickness affects the signal-to-noise ratio achievable in the course of the detector lifetime, the overall material budget and the power dissipation of the system, in particular after large irradiation doses. The overall material budget is determined not only by the active elements (sensor and readout electronics), but also by our ability to design a low mass mechanical support and cooling system.

One of the key elements driving our design is the necessity to put these detectors as close to the beam line as possible, in order to optimize the vertex resolution. In our baseline geometry, the inner detector edge is located 6 mm from the beam line. As a consequence, the inner portion of our detector will be exposed to radiation levels comparable to the ones expected for the LHC experiments. One of the effects produced by high-energy particles on semiconductor devices is displacements. These are dislocations of atoms from their normal sites in the lattice, producing defects and trapping sites. Displacements determine the degradation of the bulk material. In detectors, they cause changes of the internal electric field due to the modified doping concentration. For very-high-dosage irradiation, this will eventually lead to inversion of the conduction type of the bulk material (type-inversion), increases in leakage current and depletion voltage, changes in capacitance and resistivity, and charge collection losses. The key to ensure that the sensor could still provide useful signal after high radiation dosage is to be able to operate the detector at high voltage. In order to do this, the breakdown point must be increased and controlled. A common solution to this problem is to use multiple guard rings [2].

In order to increase the useful operating time of the silicon sensors, running with partial depletion has to be considered. This is more suitable for n-type pixel readout, since the depleted region will grow from the n^+ side of the junction after type inversion. Thus, the sensor technology preferred is the so-called "$\mathrm{n^+/n/p^+}$" adopted by both ATLAS and CMS pixel systems [3, 4], where the collection electrodes are $\mathrm{n^+}$ pixels and the $\mathrm{p^+}$ implant biases the back of the detector.

Results from PSI for CMS showed that for a fluence of $6 \times 10^{14} \pi/cm^2$, a depletion depth of 150 μm is obtained for 300 volts. At high bias voltages, charge trapping in the depleted region is below the 10% level. [4].

For an n^+/n pixel detector, the charge-collecting pixels are defined by n^+ implants. Due to the oxide-charge-induced electron accumulation layer between the silicon and silicon oxide, neighboring n^+ implants will be shorted together unless they are isolated from one another. There are two techniques to implement the isolation: p-stop and p-spray. In the p-stop isolation, a high dose p-implant surrounds the n^+ implant. In the p-spray isolation, a medium dose p-implant covers the whole n-side and the collecting n^+ pixels are done by overcompensation with the high dose of n^+ implant [5]. The ideal design for the p-stop (single, common or multiple, width and spacing of the p-implants) needs to be determined. It will have a significant impact on the minimum pixel pitch, the noise, charge collection, and capacitance of the sensor. Recent studies by both ATLAS and CMS also showed that in order to keep the field low for non-connected pixels (failed bump bonding) so as to avoid stochastic discharges, the p-stop has to be designed carefully to create a high resistivity accumulation path between neighboring pixels. The advantages of the p-spray technique is that one or two mask steps and the risk of overlapping implants will be eliminated. A finer pitch will also be possible. A further advantage of the p-sray is that the electric field in the region between the p-spray and the n^+ implant decreases with increasing irradiation, minimizing the risk of micro-discharge.

The use of p^+/n detectors in a high radiation environment also needs to be studied. One interesting option in this case is to go to thinner detectors (say 200 to 250 μm) and lower resistivity silicon(say around 1K ohms-cm). Simulations show that such detectors can be fully depleted at a voltage no higher than 300V up to a fluence of a few times 10^{14} particles/cm^2.

There is a new development on radiation hardened silicon reported by the CERN RD48 (Rose) collaboration. Radiation hardness of silicon detectors can be significantly improved by using oxygen-diffused silicon. Various silicon detector manufacturers, working in collaboration with RD48, have started on a program of optimization of the oxygen diffusion process. We plan to get some pixel detectors made from oxygenated silicon wafers.

5.2.2 Sensor simulation

A detailed understanding of the factors affecting the sensor performance is crucial to its design. We have studied a number of issues through simulation.

In a n^+/n detector the charge carriers collected at the pixel electrode are electrons. We have studied these n^+/n sensors , as well as the conventional choice of p$^+$/n pixel devices, which have holes as the collected charge. In addition, we have considered radiation damage effects. These include the deterioration of the noise performance due to the increased leakage current and the change in detector response induced by the change in the effective donor concentration.

Other factors that affect the ultimate resolution achievable in this system are related more closely to the design approach and the performance of the readout electronics. In particular, the electronic noise, and the threshold that determines the minimal charge deposition that will be recorded as a signal hit, are important. The sensitivity to these parameters has been studied, as well as the tradeoff between analog and digital readout.

In order to understand these effects, we have developed a stand-alone simulation, based on a two-dimensional model of the signal formation in silicon. This program has been interfaced with the Monte Carlo software used to study our physics reach. This integration allows us to have a more realistic model of the detector occupancy, crucial in trigger simulations, and also provides a more realistic implementation of the hit resolution achievable for different track angles of incidence.

While this simulation program has lead us to a much better understanding of the interplay between hit resolution, occupancy and physics reach of our experiment, an additional effort is under way to develop a more accurate model of the static and dynamic properties of the pixel sensors, taking into account the true three-dimensional nature of the electric field distribution and signal charge-flow to the collection electrode. These studies will be crucial to determine our final sensor choice.

5.2.2.1 Spatial resolution studies

We have developed a program to evaluate the spatial resolution achievable with different sensor and readout electronics specifications[6]. Our approach is similar to several other studies performed for silicon microstrip detectors [7], [8]. We have investigated the interdependence of charge diffusion, magnetic field effects, electronic noise, discriminator threshold and number of bits in the back-end electronics.

The ability to achieve a spatial resolution better than the nominal 'digital' resolution (σ=width $/\sqrt{12}$) is closely related to how accurately we can interpolate between the location of the centers of individual pixels, using the information of the reconstructed charge collected at each sensor in a cluster.

We have modeled the charge produced by the minimum ionizing track traversing silicon using a Landau probability distribution. Electrons and holes produced by the energy deposited by the track drift along the electric field lines ($\vec{E}$) in the detector. The equations describing this drift motion are:

$$\vec{J_n} = q\rho_n\mu_n\vec{E} \tag{5.1}$$

$$\vec{J_h} = q\rho_h\mu_h\vec{E} \tag{5.2}$$

where q is the electron charge, μ_n and μ_h are the mobility of electrons and holes respectively, ρ_n is the number of free electrons per unit volume, and ρ_h is number of free holes per unit volume.

In parallel, the charge cloud spreads laterally due to diffusion. The parameters characterizing the drift in the electric field (μ_h, μ_e) are related to the parameter describing the diffusion of the charge cloud (D_h,D_e) by the Einstein equation:

$$D_{h(e)} = \frac{kT}{q}\mu_{h(e)} \tag{5.3}$$

where D_h and D_e are the diffusion coefficients, q is the electron charge and kT is the product of the Boltzmann constant and the absolute temperature of silicon. The average square deviation with respect of the trajectory of the collected charge without diffusion is $< \Delta r^2 >= 2D\Delta t$. In our study we have used $\mu_h = 400\ cm^2/Vs,\ \mu_e = 1450\ cm^2/Vs$.

When the silicon detector is located inside a magnetic field, there is an additional source of charge spreading: the lateral motion induced on electrons and holes by the magnetic field. This effect shifts the centroid of the detection point of the charged cloud and also widens it, because the amount of drift is proportional to the path length of the detected electrons (holes). This so called "$E \times B$ effect" depends on a parameter called the Hall mobility μ_H, proportional to the charge carrier mobility discussed above. In order to take into account the effect of the magnetic field in the bend plane, the following values of the Hall mobility have been taken from experimental data quoted in [9]:

$$\mu_{H,h} = 0.72\mu_h, \tag{5.4}$$
$$\mu_{H,e} = 1.15\mu_e. \tag{5.5}$$

The charge-cloud image appears on more pixels when the incident track crosses the detector at an angle, as the generation point of the electron-hole pair spreads out along the track path length. These various effects are illustrated in Fig. 5.1, which shows the charge broadening due to diffusion, the magnetic field and the combination of the these two effects together with track inclination with respect to the normal to the detector.

An analog readout improves the spatial accuracy by charge weighting. A low intrinsic noise is obviously a key element in this approach, together with a low and uniform threshold. These features are important, not only to maintain the advantages of an analog readout, but also to minimize problems in data flow related to excessive data transfer. Finally, the achievable digitization resolution and the knowledge of the gain of the analog front-end are additional issues that should not be overlooked. We will address the interdependence among these parameters.

Two assumptions are common to all the configurations discussed in this study. We have considered a 300 μm thick Si detector biased in over-depleted mode (200 V for a detector that is fully depleted at 120 V). In this first stage we have assumed full charge collection and we have assumed negligible capacitive cross talk between adjacent pixels. These assumptions will be relaxed in subsequent studies.

5.2.2.2 Factors affecting the spatial resolution

We have carefully examined the different factors affecting the hit resolution in the small pixel direction, both in the bend plane and in the direction orthogonal to it, with reference to a pixel geometry of 50μm $\times$ 300μm, using several track orientations.

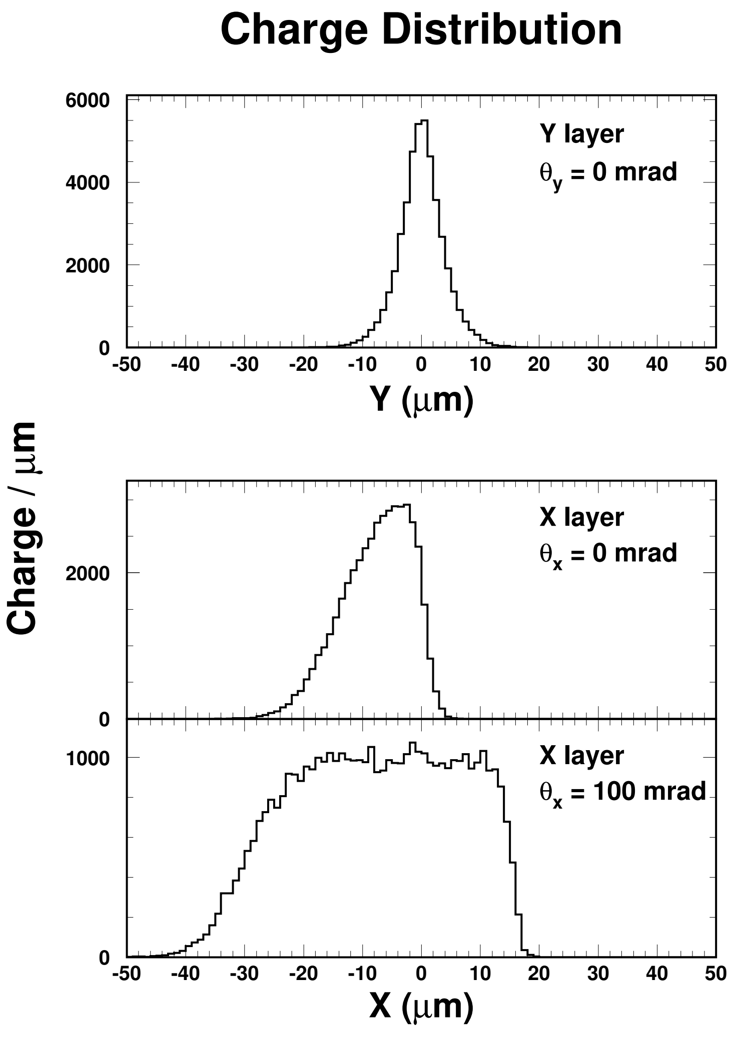


Figure 5.1: Collected charge spreading in a 300 μm silicon detector a) produced by diffusion, b) produced by the interplay of diffusion and the magnetic field for B=1.6T, and c) produced by diffusion and magnetic field effects when the charged track is incident at an angle of 100 mr in the bend plane.

Fig. 5.2 shows the relationship between the position of incidence at the pixel plane and the reconstructed coordinate for tracks at normal incidence on the pixel detector, whereas Fig. 5.3 assumes a track inclined at an angle $\theta_x = 300mr$. We have started our study assuming a readout scheme with the intrinsic electronic noise switched off, and we consider both a binary and an analog readout (with 6-bit ADC resolution). We can see that if the electronics noise is negligible, a digital readout actually can achieve a better resolution than the naive expectation of the width of the cell divided by $\sqrt{12}$. This is due to the fact that when the charge is shared between two or more pixels it is possible to do a digital charge weighting algorithm that provides some interpolation between the position of individual pixels. This improvement is constant, on the average, for different angles.

Let us start our discussion on analog readout with tracks at normal incidence. Fig. 5.2 shows that, in presence of the magnetic field, the charge is actually 'focused' inside the pixel instead of spread out by virtue of the magnetic field. This makes the analog and digital readout equivalent, because there is not a second pixel helping to interpolate the track position using charge sharing. On the other hand, tracks incident on the pixel region where the magnetic field spreads the charge further with respect to diffusion, are helped significantly by analog readout. In order to take full advantage of the analog information, it is necessary to adopt an optimal choice of the charge weighting algorithm that reflects the expected distribution of the charge profile. In particular, the center-of-gravity linear weighting algorithm is suitable only for relatively large angle tracks, where enough charge sharing is produced by the track inclination. For tracks at normal incidence, it is necessary to perform a non-linear weighting [8].

An additional important constraint in the achievable resolution is the discriminator threshold. In this study we have assumed that only the pixels having a signal above threshold are read out, and we have varied the threshold for analog and binary readout. Fig. 5.4.a shows the effect of increasing the threshold for an incident angle $\Theta_x = 300$ mr, for analog and digital readout respectively. Fig. 5.4.b shows the corresponding pixel occupancy. Each curve shows the fraction of events having N pixels hit for a given threshold. For instance, for a threshold of 1000 electrons, less than 5% of the events have 4 or more pixels hit, 45% have 3 pixels hit, and about 50% have two pixels hit. The resolution achievable for binary readout shows a characteristic oscillatory behavior as we change the threshold. The worst resolution is always at the thresholds for which most of the events have clusters with the same size, corresponding to less interpolation power. In the analog readout case, the higher the threshold is, the coarser the charge sharing process is. In several of the results presented below, we have assumed a threshold of 1000 electrons, a somewhat optimistic assumption, supported by promising initial test results on the FPIX0 pixel readout chip [10]. The slope of the threshold dependence of the resolution is not very high below 3000 electrons, as illustrated by Fig. 5.4, so the results are largely unaffected if we use less restrictive conditions.

Each readout circuit features some intrinsic noise, most commonly the series noise associated with the input transistor of the charge-sensitive preamplifier integrating the charge collected from the pixel sensor. In addition, at very high radiation damage, a contribution from the parallel noise produced by the excess leakage current may become non-negligible

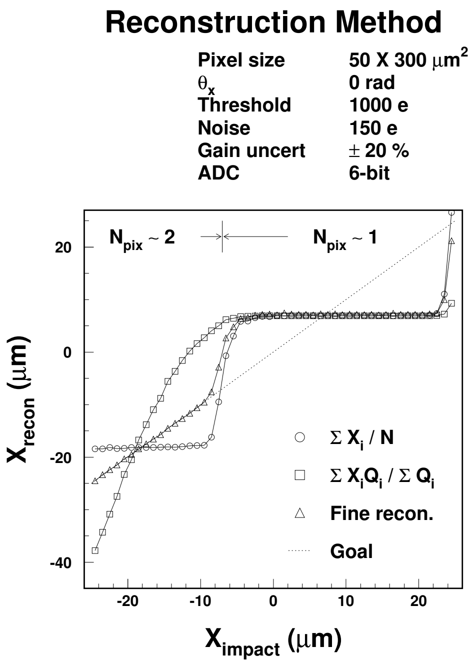


Figure 5.2: The reconstructed x coordinate as a function of the charged track impact point on the pixel cell for a pixel size of 50x300 μm^2 and normal incidence.

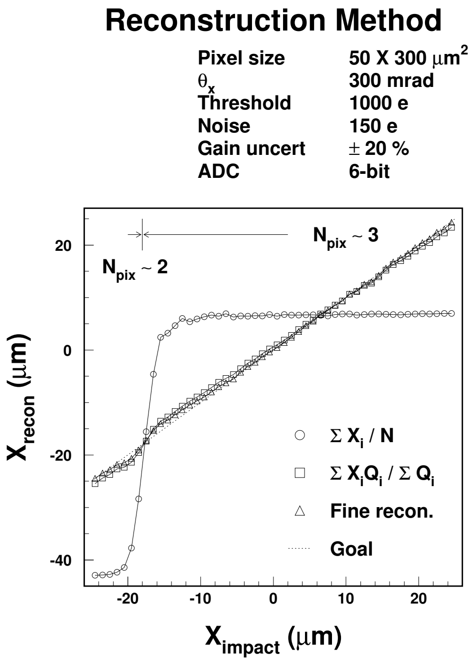


Figure 5.3: The reconstructed x coordinate as a function of the charged track impact point on the pixel cell for a pixel size of 50x300 μm^2 and incidence angle in the bend plane $\theta_X = 300$ mr.

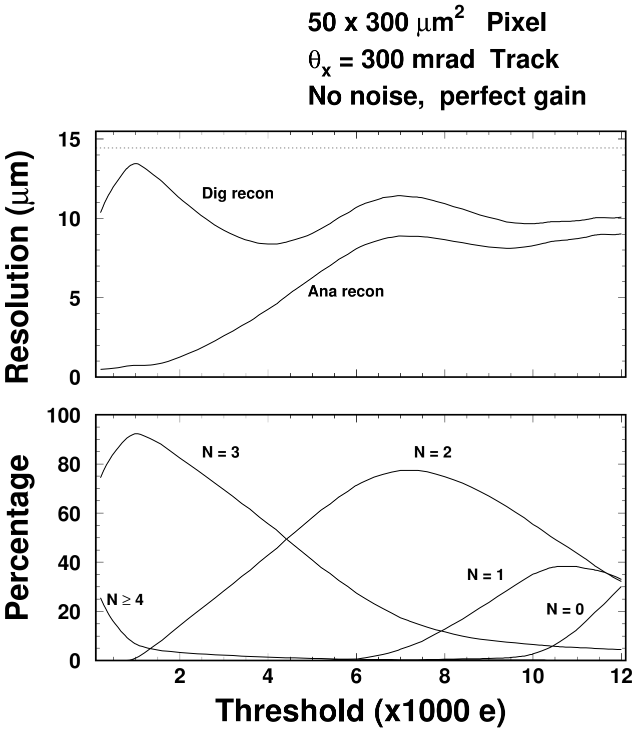


Figure 5.4: a) Spatial resolution in the reconstructed x coordinate as a function of the threshold for a pixel size of 50x300 μm^2 and incidence angle in the bend plane $\theta_X = 300$ mr. b) percentage of events having N pixels hit as a function of the threshold for the same configuration.

or even the dominant effect. A pixel detector has the potential for being a very low noise system, as the capacitive load at the input of the charge sensitive preamplifier is quite negligible. The recent results achieved at Fermilab in the prototype development of the front-end electronics for our pixel system are quite promising, and lead us to assume that an equivalent noise charge of about 150 electrons is an achievable goal. However to take full advantage of the low noise, the minimum threshold spread among channels needs to be comparable, otherwise either our assumption on the threshold needs to be modified or the effective noise is higher than assumed.

Fig. 5.5 shows the number of pixels hit as a function of the track angle for electron and hole collection. Fig. 5.6 shows the resolution as a function of the angle in the bending plane (θ_x), and in the non-bend plane (θ_y) expected with digital readout and with a 4 bit ADC respectively. At angles above ± 50 mr the resolution achieved with analog readout is significantly better.

One of the questions that we have addressed in this study is the interplay between maximum dynamic range and ADC resolution. The known shape of the Landau distribution, with a long tail above the most probable energy loss for a minimum ionizing particle in 300 μm silicon (24,000 electrons), suggests a dynamic range at least of 48,000 electrons. However, the goal of measuring the charge collected in the pixel with the maximum pulse height more accurately obviously conflicts with the accuracy of the charge measurement in the neighboring pixels at any given bit resolution. It is interesting to see the interplay of these two factors. We have studied this issue at two different angles in the non-bend plane: $\theta_x = 100$ mrad and $\theta_x = 200$ mrad. Fig. 5.7 shows the study at $\theta = 100$ mrad and Fig. 5.8 the corresponding results at $\theta = 200$ mrad. In both cases the solid line defines binary readout, quite insensitive to the dynamic range of the front-end electronics. However it can be seen that for the quite limited resolution provided by a 2 bit ADC, there is actually an optimum in the dynamic range at about 25,000 electrons. For higher resolution, this optimum appears to have a threshold, above which there is not much improvement in resolution achieved by increasing the dynamic range. At a larger angle (200 mrad), where the ratio between the maximum pixel charge signal and the signal in its neighbors becomes progressively lower, the results are consistent, but are even more favorable to the lower dynamic range.

Another effect that we have studied is the sensitivity of the hit resolution to the digitization accuracy of the ADC chosen. Fig. 5.9 shows a comparison of the performance of different ADC's for a threshold choice of 2000 electrons. There is a definite difference between the region within 100 mr of normal incidence and the rest of the solid angle. The angle θ_x is the track angle in the non-bend plane and the angle θ_y is the angle in the bend plane. In the former region the resolution improves smoothly from 10 to about 5 μm as the angle increases. The higher number of bits is helpful at larger angles, where a resolution of about 3 μm is achievable with a 4 bit ADC. Finally, a non linear response flattens the resolution achievable in the region of larger angles. Fig. 5.10 shows the effects of adding 450 electron rms threshold spread between different channels and of raising the threshold to 3000 electrons. We can see that it is quite valuable to be able to maintain a low threshold and try to achieve a 4 bit resolution in the ADC.

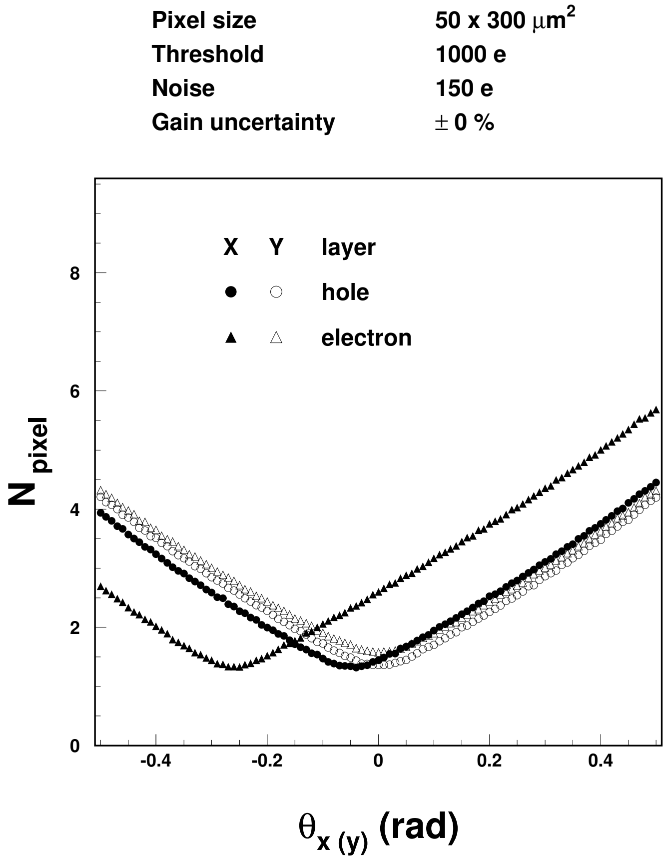


Figure 5.5: Number of pixels hit for a 50 μm pixel small dimension as a function of the track orientation.

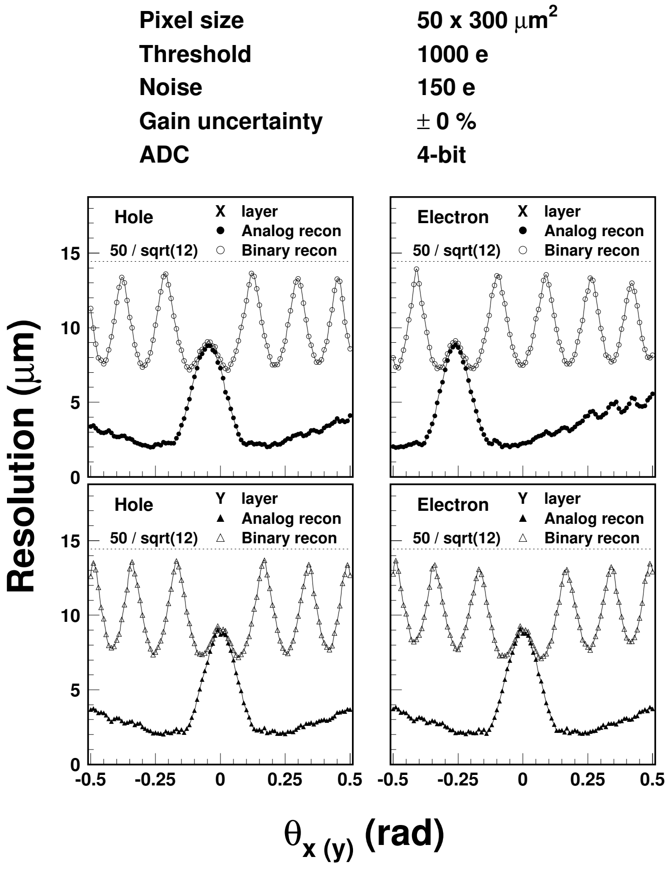


Figure 5.6: Resolution for binary readout and 4-bit ADC readout for 50 μm pixel small dimension.

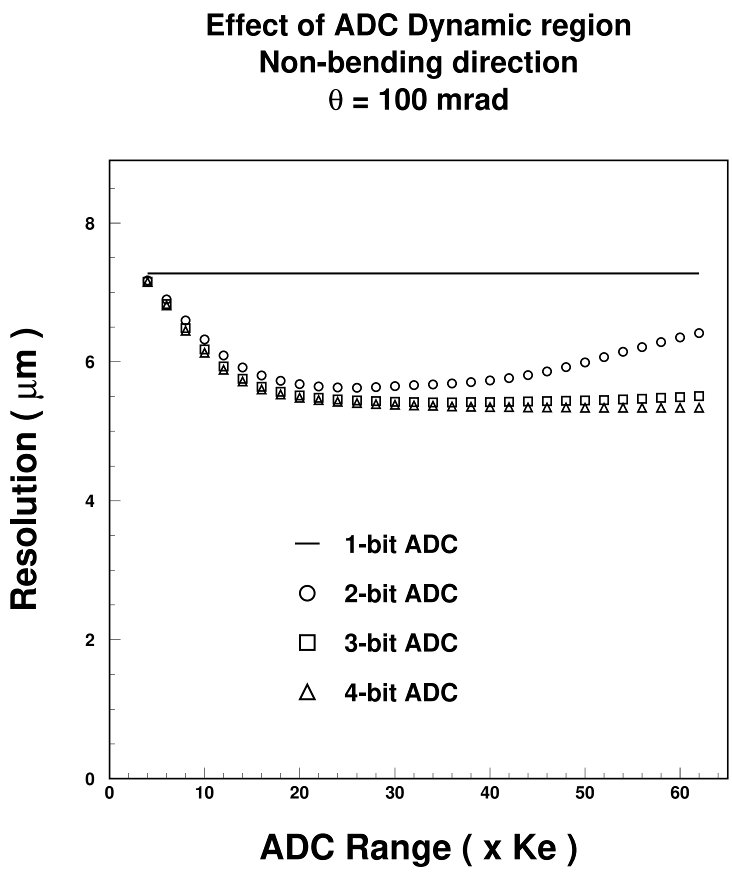


Figure 5.7: Effect of dynamic range of the ADC for different number of bits, for tracks incident at 100 mrad.

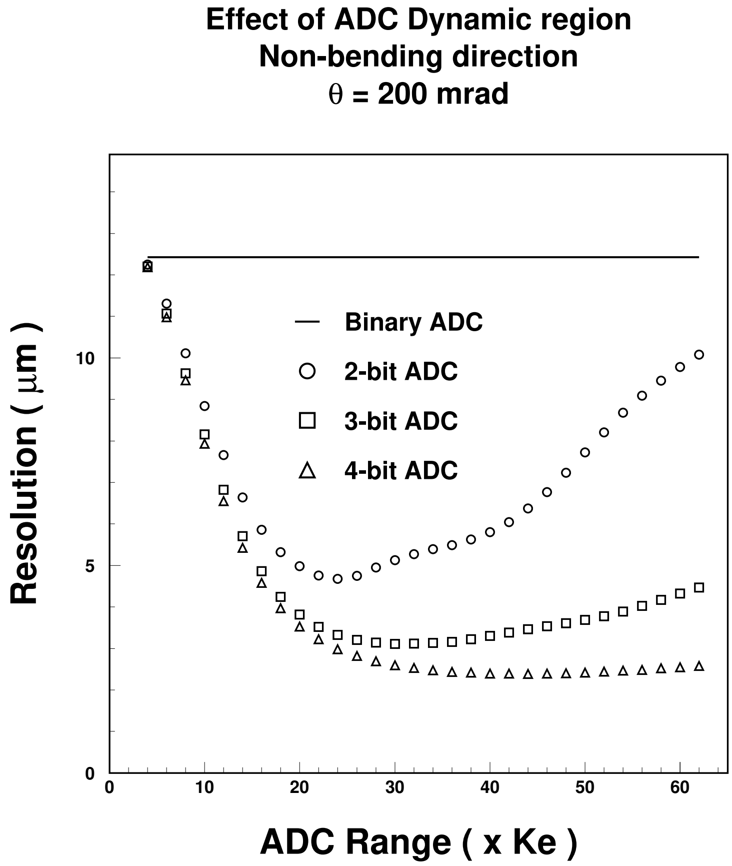


Figure 5.8: Effect of dynamic range of the ADC for different number of bits, for tracks incident at 200 mrad.

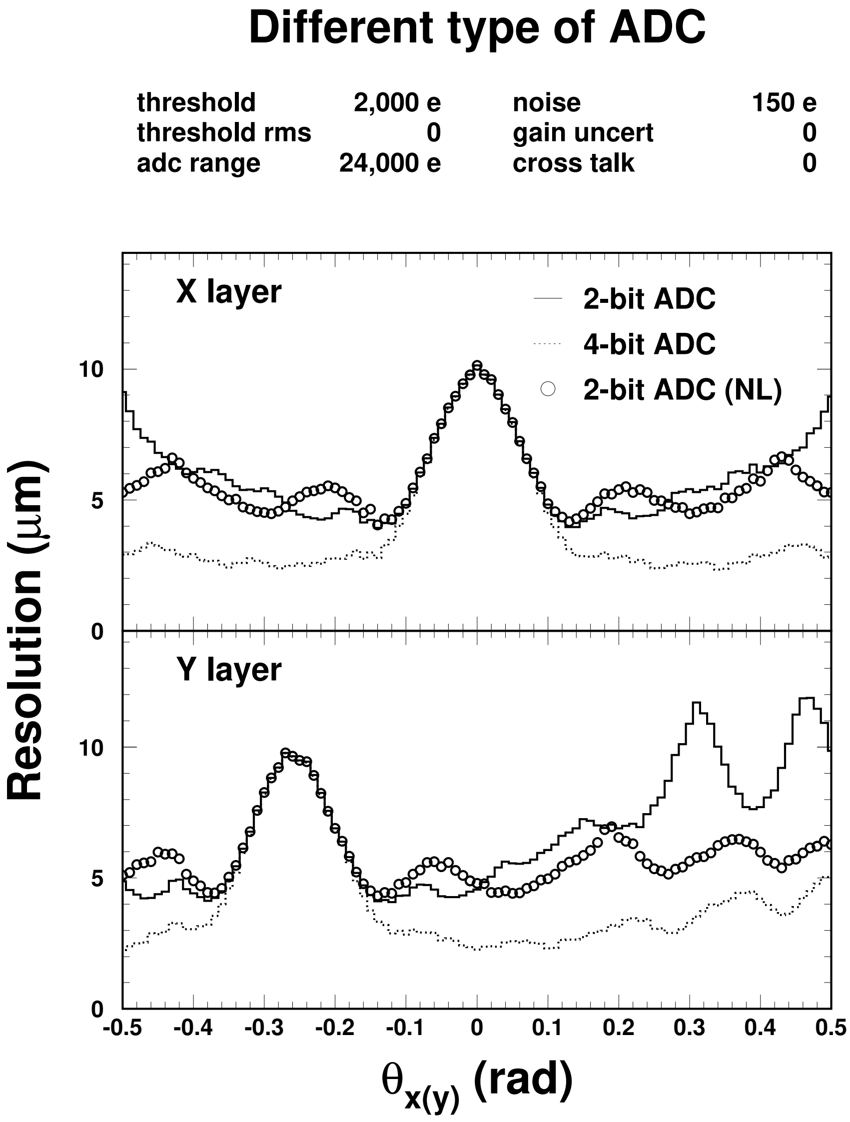


Figure 5.9: Comparison between different ADC resolution, with a nominal threshold of 2000 electrons and no threshold spread.

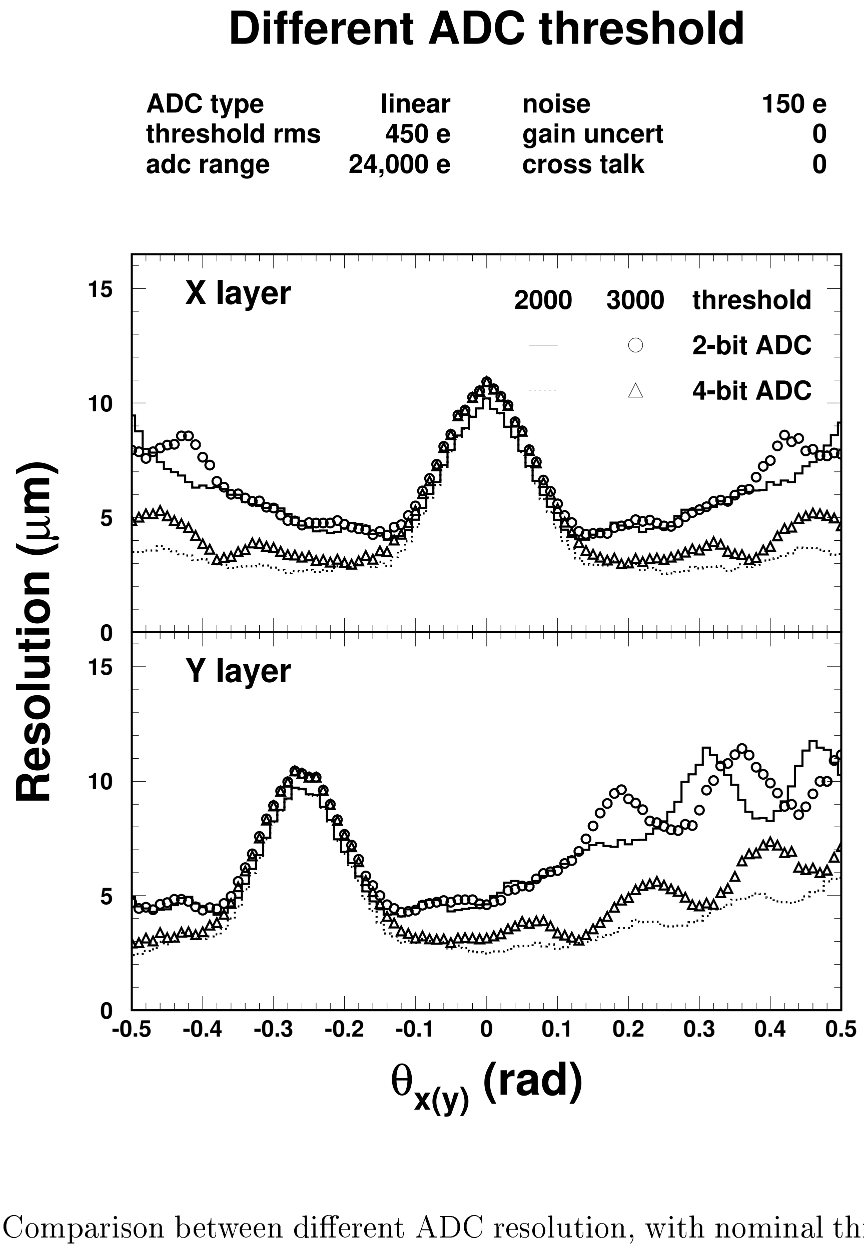


Figure 5.10: Comparison between different ADC resolution, with nominal thresholds of 2000 and 3000 electrons and a threshold spread of 450 electrons.

Table 5.1: Pixel occupancy for electron collection for different cell geometries and event type. One interaction per crossing is assumed.

	(30 μm x 500 μm)		(50 μm x 300 μm)	
	Light Quark Evt	B evt	Light Quark Evt	B Evt
npix/chip	0.87	1.76	0.65	1.30
ntrk/chip	0.26	0.51	0.25	0.52
ncol/chip	0.25	0.50	0.26	0.53
npix/trk	3.5	3.55	2.65	2.59
ncol/trk	1.03	1.03	1.06	1.06

Table 5.2: Pixel occupancy for hole collection for different cell geometries and event type. One interaction per crossing is assumed.

	(30 μm x 500 μm)		(50 μm x 300 μm)	
	Light Quark Evt	B evt	Light Quark Evt	B Evt
npix/chip	0.52	1.11	0.40	0.90
ntrk/chip	0.25	0.51	0.25	0.52
ncol/chip	0.24	0.50	0.25	0.53
npix/trk	2.16	2.24	1.69	1.79
ncol/trk	1.03	1.04	1.05	1.06

5.2.2.3 Study of the correlation between hit resolution and B decay time resolution for different detector configurations

The studies discussed above have allowed us to map the achievable hit resolution for any given geometry as a function of the track incidence angle. They have also provided us with more accurate information on the hit multiplicity associated with a given track angle. We have used this more realistic information to achieve a better understanding of several key features of our detector performance.

We have investigated[11] the occupancy in the inner portion of our detector for different event classes and different assumptions on the interaction rate. This study has been a crucial input for the engineering group responsible for the design of the pixel readout chip and to test the readout architecture and identify possible bottlenecks in the data throughput. The pixel occupancy has been studied for a chip located with its inner edge 6mm from the beam pipe. The chip has been modeled as an 8x8 mm^2 square. Two different cell geometries have been studied, 30x500 $\mu\mathrm{m}^2$ and 50x300 $\mu\mathrm{m}^2$: they correspond to two different choices of the pixel small dimension with the same unit cell area. The results of this study are summarized in Tables 5.1 and 5.2 for electron and hole collection respectively.

In light quark events, it can be seen that the average number of tracks per chip in this "hot" region is only 0.25: thus 75% of the events do not produce any tracks crossing any given chip located near the beam. This study also provides data samples for the readout architecture simulation and the trigger processor generator. Initial studies have shown that, assuming a pixel size of 50 μm x 400 μm and 2 or 3 bit flash ADC's digitizing the pixel cell signal, the proposed column-based data-driven architecture should be capable of achieving the hit readout rate required by the BTeV experiment [12].

The crucial parameter to determine the physics reach of BTeV is the vertex resolution, or equivalently, the resolution in the proper decay time. The time resolution is affected not only by the hit resolution in the pixel sensor and extrapolation distances, but also by the material in the pixel system, including the sensor, the readout electronics and the material used for mechanical support and cooling. The study of the effect of increasing the material budget in the detector system is currently under way [13]. The importance of keeping the material budget to a minimum is illustrated by a comparison of the time resolution achieved for the decay $B_s \to D_s K$ under two assumptions shown in Fig. 5.11. The top plot shows the resolution achieved with a pixel small dimension of 50 μm and pixel planes organized in stations containing 3 detectors whose material budget was described by three layers of thickness 0.6, 0.5 and 0.6 mm of silicon equivalent respectively. The bottom plot corresponds to the same geometry but with an increased equivalent Si thickness (1.1, 1.0, 1.1 mm). The time resolution degrades from 42 to 50 fs. This study illustrates the importance of keeping the material budget to a minimum.

The simulation included in the original EOI was based on an assumed hit resolution that was uniform throughout the sensor. However we have shown that the resolution improves substantially for angled tracks due to the increased charge sharing that allows a better interpolation. This more realistic mapping of the effective hit resolution has been introduced in our simulation and has shown that this improvement is quite important and compensates the adverse effects of having taken a more conservative model of the material included in our detector.

5.2.2.4 Non-Gaussian Resolution Functions

All previous discussions of the physics reach of BTeV, contained the approximation that the measurements produced by the pixel detectors had a Gaussian shape. With the simulation tools now available it is possible to remove this approximation and to demonstrate that a realistic shape for the resolution function does not significantly degrade the physics reach.

The information in the preceding sections was used to implement several different models of the pixel resolution in the MCFast. These models are discussed in detail in section 10.3. The third of these models, the fast clustering model, is the one of interest; it involves the generation of digitized pulse heights and the extraction of a measured coordinate from the pulse height distribution. The shapes of the resolution functions for this model can be seen in Fig. 10.2 parts c) and d). Both of these have significant non-Gaussian components. For reference, two other models will be considered, both of which have Gaussian resolution functions. In the first model the Gaussian has a constant $\sigma = \text{pitch}/\sqrt{12}$, while in the second model σ varies with angle of incidence. The resolution functions for the variable σ model can be see in Fig. 10.1; over the entire parameter space the resolution for the short dimension is better than pitch divided by $\sqrt{12}$. Also, careful inspection will show that resolution function

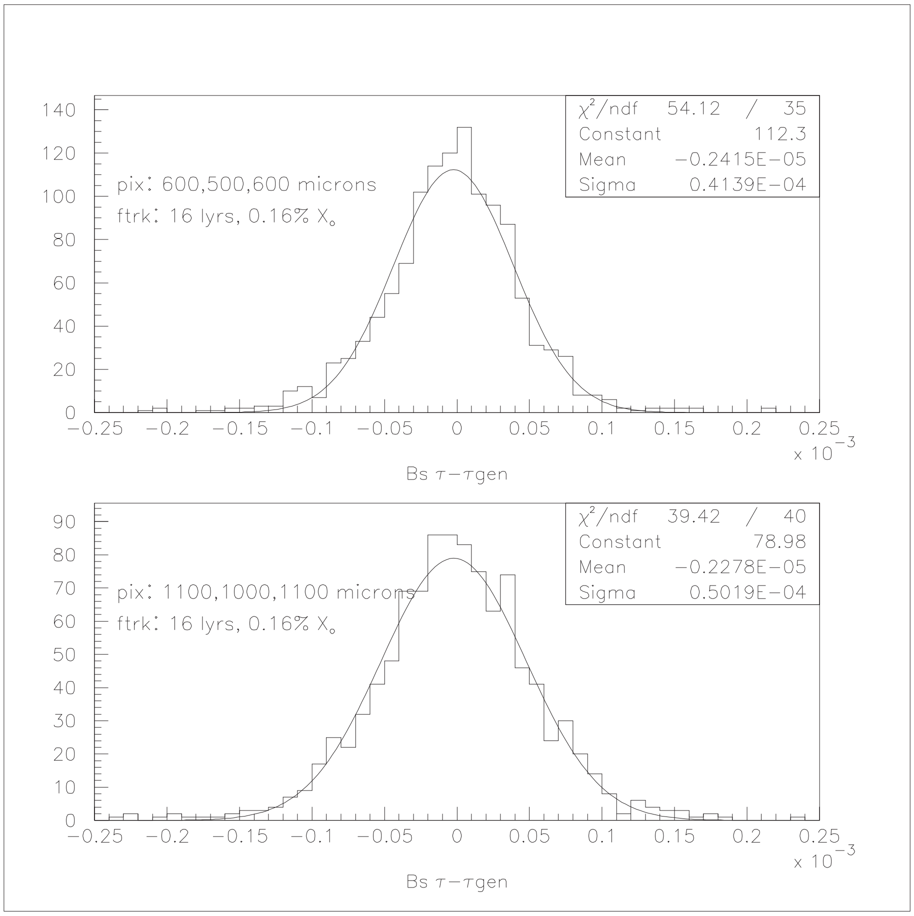


Figure 5.11: Comparison between the time resolution for two different assumptions of the material budget in a triplet station. The decay mode considered is $B_s \to D_s K$.

for the fast cluster model is slightly poorer than the resolution function for the variable σ model; the reasons are discussed in section 10.3.

In order to investigate the physics consequences of these different spatial resolution functions, the decay chain $B_s \to \psi \bar{K}^{*0}$, $\psi \to \mu^+\mu^-$, $\bar{K}^{*0} \to K^-\pi^+$ was considered. After all analysis cuts and background suppression cuts, the resolution on the proper decay time of each B_s candidate was computed. This quantity, in this decay mode, makes one of the most severe demands on the spatial resolution of the pixel detectors. The histograms in Fig. 5.12 show the distribution of this resolution over many reconstructed B_s candidates. In part a), the distribution is shown for the constant σ model of the pixel spatial resolution. In part b) the distribution is shown for the variable σ model and, in part c) the distribution is shown for the fast clustering model. The most prominent feature in these plots is that there is a large improvement in resolution going from constant σ to variable σ. This is expected because the variable σ model, as discussed above, has better resolution than the constant σ model. There is also an increase in yield, which comes from the improvement in detachment and isolation cuts which derives from improved spatial resolution. A third observation is that, when going from the variable σ model to the fast clustering model, only minor changes are seen, consistent with the small differences between the two models. The conclusion of this study is that the expected resolution is not significantly degraded when using a model with explicit pulse height generation.

Part a) of Fig. 5.13 shows the distribution of the quantity $(t_{\text{measured}} - t_{\text{generated}})/\sigma_t$ for the variable σ model of the pixel spatial resolution. Here t is the proper decay time of a B_s candidate and σ_t is its error, computed on a candidate by candidate basis. Notice that the distribution is very close to a unit Gaussian and that there are few outliers. This is expected because the underlying resolution model is Gaussian; the outliers come from the small number of cases in which fit seeds are outside the radius of convergence of the fitter and from the few cases in which the first order error propagation formulae do not work. All entries in these plots are for true signal combinations. Now compare this to part b), which contains the same plot for the fast clustering model, in which the underlying resolution functions are not Gaussian. The core of this plot remains consistent with a unit Gaussian but the number of outliers has clearly increased. The increased number of outliers are attributable to the tails of the pixel spatial resolution functions. The clear conclusion is that the increased number of outliers is small enough to be of no concern.

Parts c) and d) of Fig. 5.13 compare the confidence level of the fit of the $\mu^+\mu^-K^-\pi^+$ system to a common vertex. Again all entries are for true signal combinations. Part c) is for the variable σ model, in which the pixel resolution function is explicitly Gaussian; in this case the confidence level distribution is flat, as expected. Part d) shows the distribution for the fast clustering model which contains two sorts of non-Gaussian contributions, a box-like component in the core and long tails. The combined effect of these non-Gaussian components is the spike in the few bins near a confidence level of zero and a slight downwards slope in the body of the plot. The area in the spike is about 4% of the total area of the plot. Again the clear conclusion is that the introduction of non-Gaussian resolution functions does not raise significant concerns about the physics reach of BTeV.

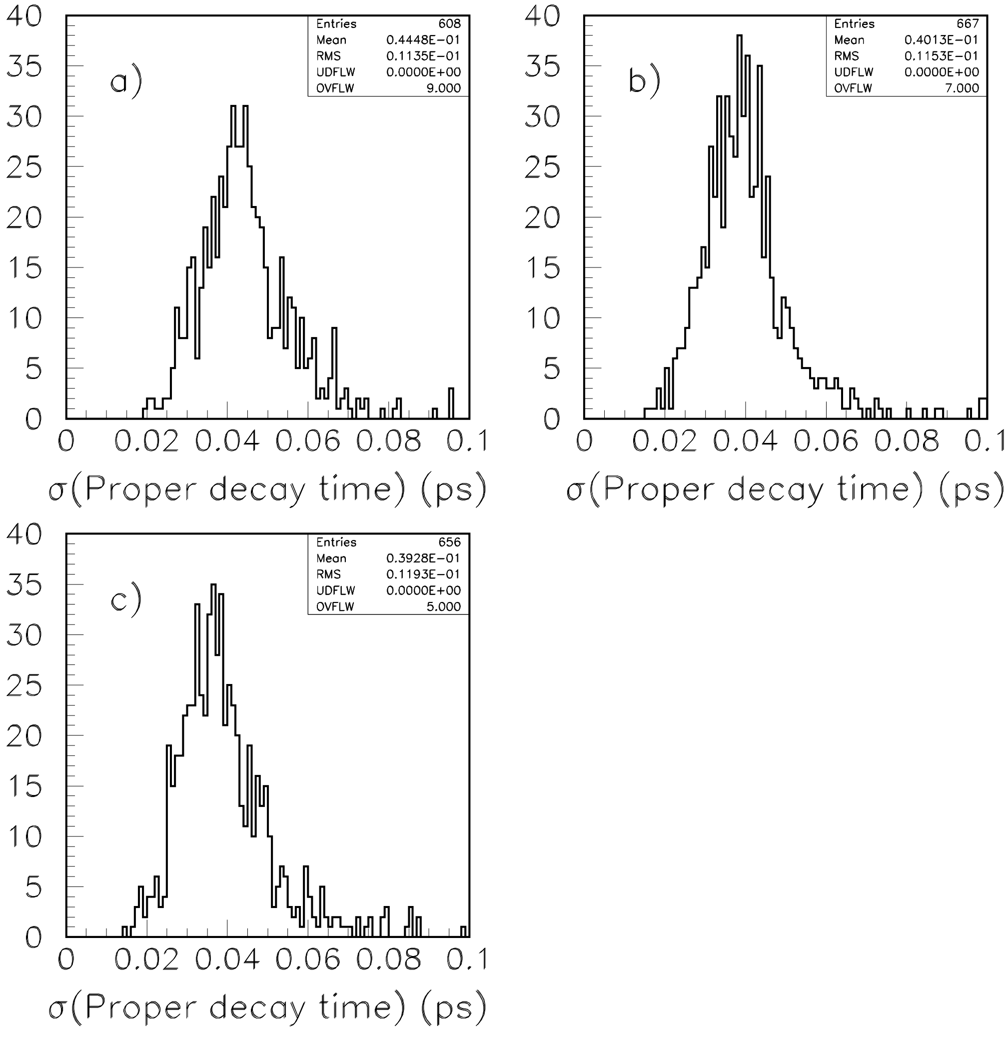


Figure 5.12: The expected resolution on the proper decay time for the decay chain $B_s \to \psi \bar{K}^{*0}$, $\psi \to \mu^+\mu^-$, $\bar{K}^{*0} \to K^-\pi^+$: a) for the constant σ model of the pixel spatial resolution; b) for the variable σ model; and c) for the fast clustering model. All entries are for true signal combinations. These models are described in section 10.3.

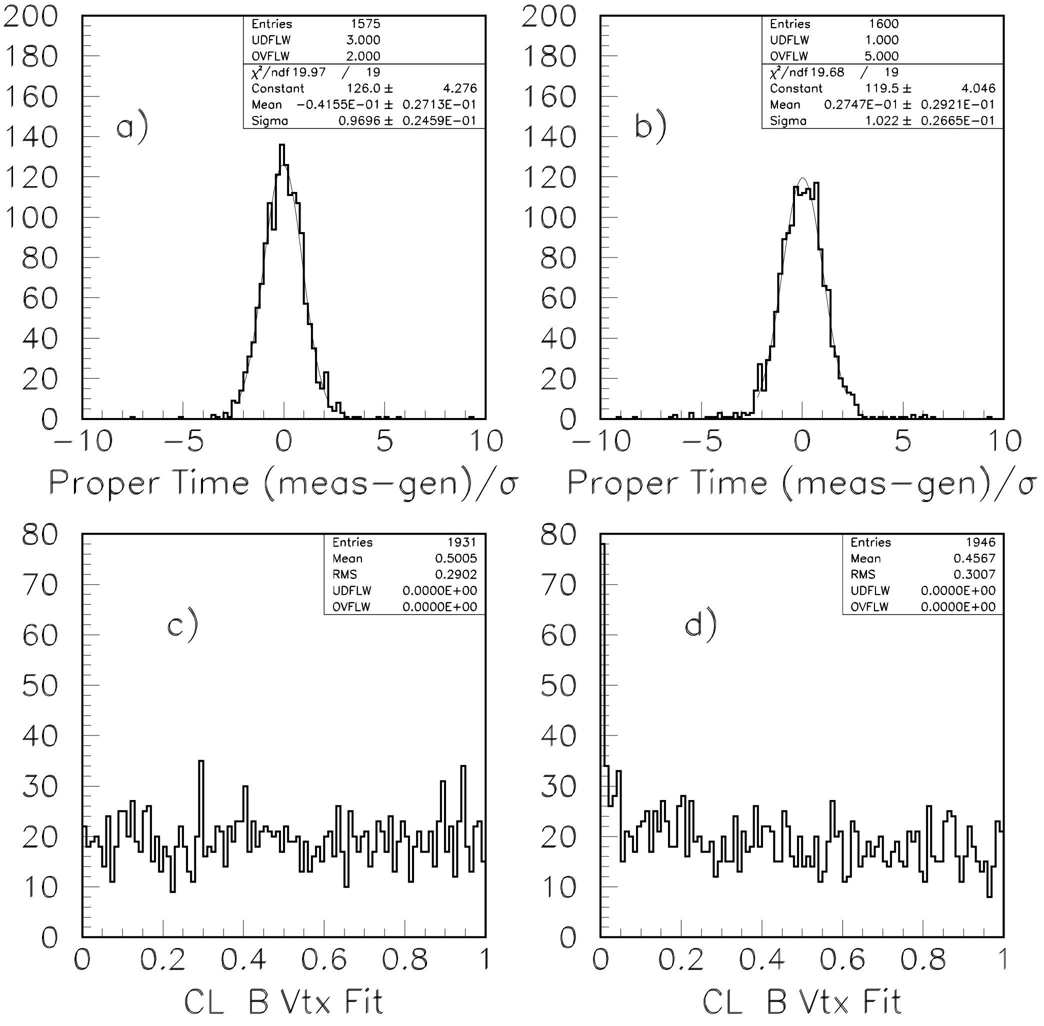


Figure 5.13: Parts a) and c) show quantities computed using the variable σ model of the spatial resolution of the pixels, in which the resolution function is explicitly Gaussian. Part a) shows the normalized residual on the proper decay time of a B_s candidate, while part c) shows the confidence level of the fit four tracks to a common vertex. All entries are for true signal combinations. Part a) is expected to be a unit Gaussian while part c) is expected to be flat. The reasons for the few outliers in a) are discussed in the text. Parts b) and d) show the same quantities for the fast clustering model, which does have non-Gaussian contributions; in these cases the deviations from the ideal behavior present but are small enough not to raise significant concerns.

5.2.2.5 Future sensor simulations

There are several other issues affecting the sensor performance that we would like to investigate. Some of them determine the optimal performance throughout the lifetime of the sensor while radiation damage effects accumulate. Other studies are dynamical in nature and will provide a more realistic model of the charge sharing between adjacent electrodes.

Our baseline sensor has a collection electrode on the ohmic side of the sensor (n^+). This requires a more complex geometry on this side in order to provide electrical insulation between the individual pixels, typically with p-stop or p-spray implants [14], [15]. These structures introduce additional complexity into our modeling of the sensor, by producing a lateral component in the electric field that sometimes has quite dramatic effects in the detector performance. It can introduce local regions of very high field that may determine the break-down voltage. In addition, it can produce modulation of the charge sharing between pixel cells introducing some lateral component to the drift motion that can increase charge sharing when the track crosses the detector at normal incidence [16]. It is intuitive to see that this electric field profile is of true tridimensional nature and thus requires 3-D simulation tools. We have chosen to perform our studies using a very sophisticated simulation program, DaVinci [17]. It has been used successfully by teams at CERN[9] and the University of Hawaii[18] to simulate their pixel sensors. The CERN group has even added an interface to study the charge collection profile in the presence of a magnetic field that we need to implement eventually if we want to model the charge collection in our experiment.

The issues that we are planning to study are:

1- a mapping of the electric field throughout the whole sensor (both inside the active volume and at the periphery for different guard ring configurations). This study is crucial to understand the sensor breakdown voltage and to design sensors that ultimately will withstand high reverse bias. The sensor is likely to be operated at high voltage during some stages of its lifetime. In addition, a detailed mapping of the electric field profile in these regions is an essential input to the the p-isolation design.

2- a study of the profile of the charge pulses in a cluster of neighboring pixels while we change the track impact parameter inside a central pixel. This study is crucial to understand the actual charge sharing between neighboring pixels and the expected pulse height with different sensors and at different irradiation doses. The radiation effects will be modeled with a different space charge density N_{eff} and with a degraded carrier lifetime due to the excess of trapping centers induced by radiation effects. In addition we are planning to introduce a realistic simulation of transport properties in the magnetic field.

These studies will be complemented by the extensive measurement program detailed in other sections of this chapter. In particular, the test beam data and extensive electrical characterization of sensors before and after irradiation will provide crucial tests of our model.

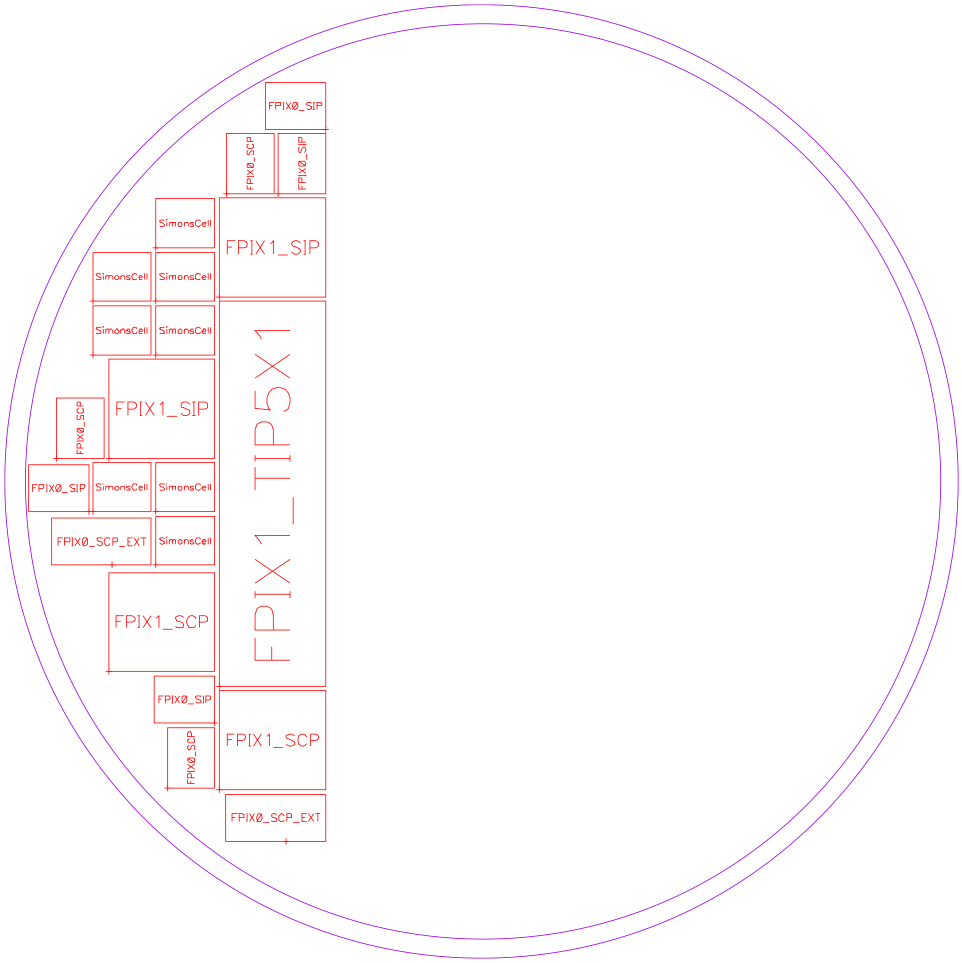


Figure 5.14: Wafer map

5.2.3 Prototype sensor

We have recently sent out for bids to a number of vendors for prototype sensor wafers, jointly with the US-CMS collaboration. The sensors will be n^+ on n with p-stop isolation between n^+ implants on the n-side of the wafer. Fig. 5.14 shows the wafer layout of the BTeV portion.

Our sensor designs are very similar to the p-stop designs of the ATLAS "First Pixel Prototype" [19]. We also borrowed heavily from their experience with the guard-ring design.

The BTeV portion of the wafer contains the following structures:

- FPIX1-SIP: A sensor with individual ("atoll") p-stops - designed to mate to a single FPIX1 readout chip (2 per wafer).

- FPIX1-SCP: A sensor with a single common p-stop - designed to mate to a single FPIX1 readout chip (2 per wafer).

- FPIX0-SIP: A sensor with individual ("atoll") p-stops - designed to mate to a single FPIX0 readout chip (4 per wafer).

- FPIX0-SCP: A sensor with a single (floating) common p-stop - designed to mate to a single FPIX0 readout chip (3 per wafer).
- FPIX0-SCP-EXT: A sensor with a single common p-stop which is extended to one side so that it may be biased - designed to mate to a single FPIX0 readout chip (2 per wafer).
- FPIX1-TIP5X1: A sensor with individual ("atoll") p-stops - designed to mate to 5 FPIX1 readout chips (1 per wafer).
- SimonsCell: test structures(both diodes and pixels) with a variety of guard ring configurations (9 per wafer). Three different guard ring designs will be tested. These are the ATLAS design, the CMS-strip design [20], and a new design from BNL and Johns Hopkins University (JHU) [21].

Additional test structures will also be present on the wafer to monitor flat band voltages, layer thickness, implant resistivity, aluminum sheet resistance, etching uniformity and alignment. Our plan for this submission is to have enough sensors to be mated to the FPIX0 and FPIX1 chips and to gain experience on multichip-module assembly. We will also use the devices to do systematic studies on guard ring designs, p-stop designs, as well as effects such as surface damage, stability and variation from wafer to wafer and vendor to vendor. Irradiation studies will also be carried out so that dependence of the breakdown voltage, depletion voltage, and leakage current on the fluence can be studied.

A number of wafers will be made from oxygen-diffused silicon.

This round of submissions is focussed on p-stop designs. However, recent tests reported by the ATLAS collaboration show that the use of p-spray will significantly improve the performance of detectors after a radiation dosage exceeding 2x$10^{14} p/cm^2$. We plan to have a separate submission to vendors to study this technology. We will collaborate with the ATLAS collaboration on the submission and study of p-spray detectors.

5.2.4 Collaborative effort with BNL/JHU

Because of the commonalities of the problems faced by BTeV and LHC experiments, a close collaboration amongst various institutes is called for. We have collaborated with BNL and JHU on sensor development. BNL possesses facilities to fabricate high quality prototype detectors. The advantage of this path is cost-effectiveness and fast turn-around. As a result of our collaborative effort, a large number of detectors of various configurations have been fabricated and tested under different processing conditions [22, 23]. BNL can now produce good detectors reliably. Various guard-ring designs have been studied and improved upon. So far, we have demonstrated that n^+/n devices can be operated at 500v without breaking down, both before and after irradiation. Single pixel devices with pixel pitch of 60 and 120 μm with common or atoll p-stops have also been studied. One of the pixel cells in these devices is connected to a large probing pad on the periphery. All the other pixels in the devices are ganged together to another probing pad. During measurement, the ganged pixels were grounded together, and I-V measurements were taken on the single pixel cell. Our results showed that breakdown occured more frequently for devices with atoll p-stop than with the common p-stop. This is even more so for devices with small pitch. We will modify the separation between the n^+ implant and the p-stop in a future fabrication.

5.2.5 Diamond option

In addition to the baseline silicon-sensor plan, some members of BTeV are participating in R&D into diamond pixel detectors as an option for BTeV. Test results on diamond from the CERN RD42 collaboration indicate that diamond detectors have a useful range of operation up to fluences exceeding 10^{15} n/cm^2. Diamond has excellent thermal conductivity and mechanical properties. These, coupled with the fact that diamond detectors have no leakage current, make diamond detectors a very attractive option for the innermost region where the radiation dosage is very high. On the other hand, this is a technology still under development, and has never been used in a real high-energy physics experiment. There are concerns about the quality of the diamonds, collection efficiencies, thickness of the diamond, large scale production, and cost.

Our simulation shows that there are significant gains in resolution by moving the detectors close to the beam. In our baseline design, we have chosen to place the pixel detector no closer than 6 mm from the beam because of radiation damage to silicon. The use of diamond pixel detectors would enable us to operate the vertex detectors closer to the beam.

We plan to test diamond pixel detectors using the FPIX0 readout chip in the MT test beam during the fixed target run at Fermilab.

5.2.6 Goals of sensor development

Our goal is to have a final specification for the sensors by May 2000. To achieve this goal, we have to do studies of guard rings, p-stop designs, and p-spray devices before and after irradiation. Electrical characteristics of devices before and after irradiation also will be simulated using computer models.

At the same time, we will keep in close contact with the CERN RD42 collaboration to investigate the possibility of using diamond detectors in the innermost area of our pixel system.

5.3 Pixel readout chip

5.3.1 Pixel readout chip development at Fermilab

One of the defining characteristics of BTeV is the plan to use a vertex trigger as the primary trigger for the experiment. This means that the pixel readout ASIC must be capable of reading out all hit information from every $p\overline{p}$ interaction. MCFAST simulations have shown that an $8mm \times 7.2mm$ pixel sensor located $6mm$ from the beamline (corresponding to the innermost pixel readout chip, given the baseline BTeV design and the current readout chip size) will be hit by one or more tracks in approximately one bunch crossing out of four at a luminosity of $2 \times 10^{32} cm^{-2} sec^{-1}$. The number of pixel cells hit by a single track depends on the pixel size and the track angle with respect to the pixel sensor. Given a pixel size of $50\mu m \times 400\mu$m and a sensor thickness of 300μm, our simulations indicate that an average of five pixels will be hit in the innermost readout chip in those crossings in which one or more tracks hit the chip. Therefore, this pixel chip has to sustain an average readout rate of 1.25 pixels per BCO (beam crossing). To account for statistical fluctuations and allow for effects not included in the simulations, the chip must be capable of even higher data transfer rates, and must be able to buffer hit information which can not be immediately read out.

The BTeV pixel readout chip should be optimized for the 132 ns time between crossings planned for the Tevatron collider. It must be radiation hard so that it can be used close to the beamline. As described above, the baseline BTeV design calls for n^+ on n silicon sensors with appropriate guard ring structures for high voltage operation. These sensors provide adequate signals after significant radiation exposure, but also have rather large radiation-damage-induced leakage current. The BTeV pixel readout chip must be able to tolerate this leakage current at least up to 25-50 nA per pixel.

An R&D program has begun at Fermilab whose goal is the design of a pixel readout ASIC for BTeV[24][25]. The program envisions a series of prototype pixel readout chips, each with specific engineering goals. The current expectation is that the final design will be implemented using the radiation hard Honeywell 0.5μm CMOS Silicon-On-Insulator (SOI) process. For this reason, the first prototype devices have been made using relatively inexpensive radiation soft CMOS processes using only features (number of metal and polysilicon layers, for example) shared by the Honeywell SOI process.

Significant progress has been made in the development of a pixel readout ASIC. Two prototype circuits have been fabricated and are being tested. The first chip, FPIX0, was designed in early 1997, and received at Fermilab in September, 1997. It has been extensively tested, both alone and bonded to a sensor [26]. FPIX0 chips bonded to pixel sensors will be tested in the Meson test beam, as described below. The second prototype pixel readout chip, FPIX1, was designed during 1998, and received in December, 1998. This circuit, which contains $\sim$ 970000 transistors, is the largest ASIC ever designed at Fermilab. It is currently being bench tested. Beam tests are planned for late summer, 1999.

5.3.2 FPIX0

FPIX0 has achieved both of its primary design goals. These were: 1) To establish a front-end design appropriate for use at the Tevatron collider; and 2) To verify that deadtimeless operation is possible. Deadtimeless operation requires that the analog and digital sections of the chip be isolated from one another, and that readout signals be isolated from the sensor so they are not coupled back to the amplifier inputs. In the FPIX designs, the metal 3 layer is used to isolate the sensor from the readout chip.

The FPIX0 amplifier consists of two folded cascode stages, AC coupled to one another.

The first stage is a charge amplifier that uses a current controlled feedback circuit [27]. The second stage provides additional gain and is DC coupled to the discriminator. The FPIX0 amplifier is a continuous time amplifier which does not require front-end reset signals. The feedback current controls the return-to-baseline time and also compensates for sensor leakage current. The circuit was designed to completely compensate for leakage current up to ~ 25 nA. Tests of the FPIX0 demonstrate that the circuit works as designed, and in fact currents of 100 nA or more can be tolerated with only minor signal degradation. This is essential for operation in BTeV close to the Tevatron beam, where radiation damage will result in large sensor leakage current (pixel sensors can not be AC coupled to the readout electronics because the required capacitor would be larger than the pixel).

One of the primary goals of our test beam activity is to verify the simulation results on resolution as a function of the number of ADC bits. FPIX0 chips (which include an analog output) bonded to ATLAS sensors will be used for this purpose.

5.3.2.1 Test results

Fluorescence x-rays from a number of metal foils have been used to provide an absolute calibration of an FPIX0 bonded to an ATLAS (Seiko ST1) pixel sensor. Figure 5.15 shows the pulse height spectrum collected using a single FPIX0 pixel amplifier while the sensor was being illuminated with terbium x-rays. The $K_{\alpha 1}$ peak (44.5 keV) is easily identified at 407 mV. Since it is known that a 44.5 keV x-ray creates 12300 mobile electron-hole pairs in fully depleted silicon, the scale for this channel is 30.2 input electrons per mV peak output signal. We have verified the linearity of the FPIX0 amplifier response to small input signals using a variety of characteristic x-rays.

Measurements of amplifier noise and discriminator threshold dispersion have been made for a number of FPIX0 chips, and also for a smaller test chip (Pre-FPIX1) which contains a variety of FPIX0-like amplifiers. Two conclusions have been drawn from these measurements. First, there is a trade-off between dynamic range on the one hand, and low noise and small discriminator threshold dispersion on the other. Second, the cells which are bonded to sensor pixels are only slightly noisier than identical unbonded cells. If all of this noise difference is attributed to the difference in input capacitance, then a SPICE simulation indicates that the total input capacitance of the (Seiko ST1) sensor pixel and bump bond is approximately 180 fF, which is very close to the expected value.

No pickup of the digital activity associated with readout has been detected. With the metal 3 shield grounded, there is no significant coupling through the detector. The capacitance between the metal 3 shield and a sensor pixel has been inferred to be approximately 1.8 aF/μ^2. This capacitance was inferred by disconnecting the shield from its ground, injecting a pulse on the shield, and measuring the output voltage from a single pixel unit cell. Finally, we have measured the pixel-to-pixel cross talk to be less than 2%.

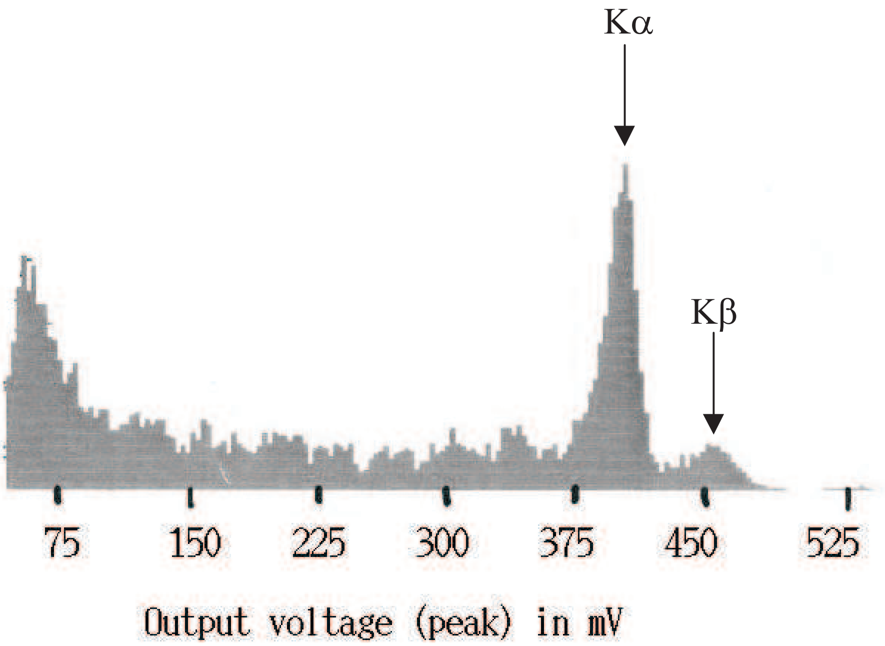


Figure 5.15: Terbium x-ray spectrum recorded by a single FPIX0 channel.

5.3.3 FPIX1

5.3.3.1 Design features and architecture

The FPIX1 readout chip is the first implementation of a new column-based pixel architecture designed to meet the requirements of BTeV. The most stringent requirement is that all pixel hit information from every Tevatron crossing must be digitized and read out so that it may be used to form the primary trigger for the experiment. Simulations indicate that, with a 26.5 MHz readout clock, FPIX1 is capable of reading out an average of more than three pixels per BCO. Relatively straightforward extensions of the FPIX1 architecture should increase the readout bandwidth by a factor of four or more.

In FPIX1, each pixel hit is associated with a six-bit time stamp (the beam crossing number) and the hit information is held in the pixel unit cell until readout. In normal operation for BTeV, the chip itself generates a readout trigger for every beam crossing. The time stamp associated with a hit pixel is held in a register located at the end of the column of pixels containing the hit. All pixels in a given column which are hit in the same 132 ns beam crossing are associated with the same end-of-column timestamp register. Each column in FPIX1 has four timestamp registers.

Figure 5.16 is a photo-micrograph of a single FPIX1 chip. Figure 5.17 is a schematic diagram of the layout. FPIX1 is composed of three building blocks: the pixel unit cell, the end of column logic, and the chip control logic. The active area consists of an array of 160 rows by 18 columns of pixel cells, each 50μm $\times$ 400μm. All "normal" chip I/O is through a double row of pads (shown along the bottom in Figures 5.16 and 5.17). The opposite edge of the chip is used to provide direct access to amplifier and discriminator outputs from one row of pixel cells. These are for diagnostic use during the prototype phase of chip development, and will not exist in the production version of the chip. In addition, one column of pixels is designed to be read out under external control using pads on this "top" edge of the chip.

A block diagram of the FPIX1 pixel unit cell is given in Figure 5.18. The FPIX1 front-end is an optimized version of the FPIX0 front-end described above. The dynamic range of the FPIX1 amplifier is approximately 32000 e^-. FPIX1 was designed with a two bit flash ADC in every pixel cell.

The digital section of the FPIX1 pixel unit cell consists of a command interpreter and a bus controller. Four pairs of lines carry commands from the end-of-column-logic to the pixel unit cells. Each pair of command lines is associated with a particular timestamp register in the end-of-column logic. The commands which control the action of the pixel cells are: "idle," "reset," "output," and "look-for-data." Commands are presented by the end-of-column logic simultaneously to all pixel cell command interpreters in a column. When a pixel discriminator fires, then if the "look-for-data" command is present on one of the four pairs of command lines, the command interpreter in the hit cell associates the hit with the timestamp register corresponding to the "look- for-data" command. Simultaneously, it alerts the end-of-column logic to the presence of a hit. After the association of a hit pixel with a timestamp register is made, the command interpreter in the hit pixel cell waits for an "output" or "reset" command on the line corresponding to the associated timestamp register, and ignores the other command lines.

In normal operation, the FPIX1 chip control logic generates a readout trigger for every beam crossing after waiting for two crossings to allow hit information to settle. A token passes quickly through columns containing no data associated with the requested crossing number, and stops at the first hit column. Another token is generated by the end of column logic in this hit column, and this token drops down the column through unhit cells until it reaches the first hit pixel. At the next rising readout clock edge, the hit pixel cell loads its data onto the column bus and drives it to the end-of-column logic for one clock cycle. In parallel, the column token is transmitted to the next hit pixel. The time required for the token to pass through unhit columns, and unhit pixel unit cells in a hit column, determines the maximum allowable readout clock frequency. A good deal of effort was spent in the design of a high speed token-passing circuit for FPIX1. The time required to skip an unhit pixel cell has been measured to be $\sim$ 190 ps (greater than 5 GHz skip rate). This allows a 26.5 MHz readout clock to be used with a comfortable safety margin.

5.3.3.2 Preliminary test results

The first measurements of FPIX1 performance are very encouraging. The front-end performance is approximately as expected. Both the amplifier noise and the discriminator threshold rms are $\sim 40\%$ smaller than in FPIX0, and acceptable for use in BTeV. The discriminator overdrive required to latch a hit in the proper beam crossing is only $\sim 250e^-$. The two-bit ADCs operate as designed, and all of the readout logic works. An IR drop on the line which

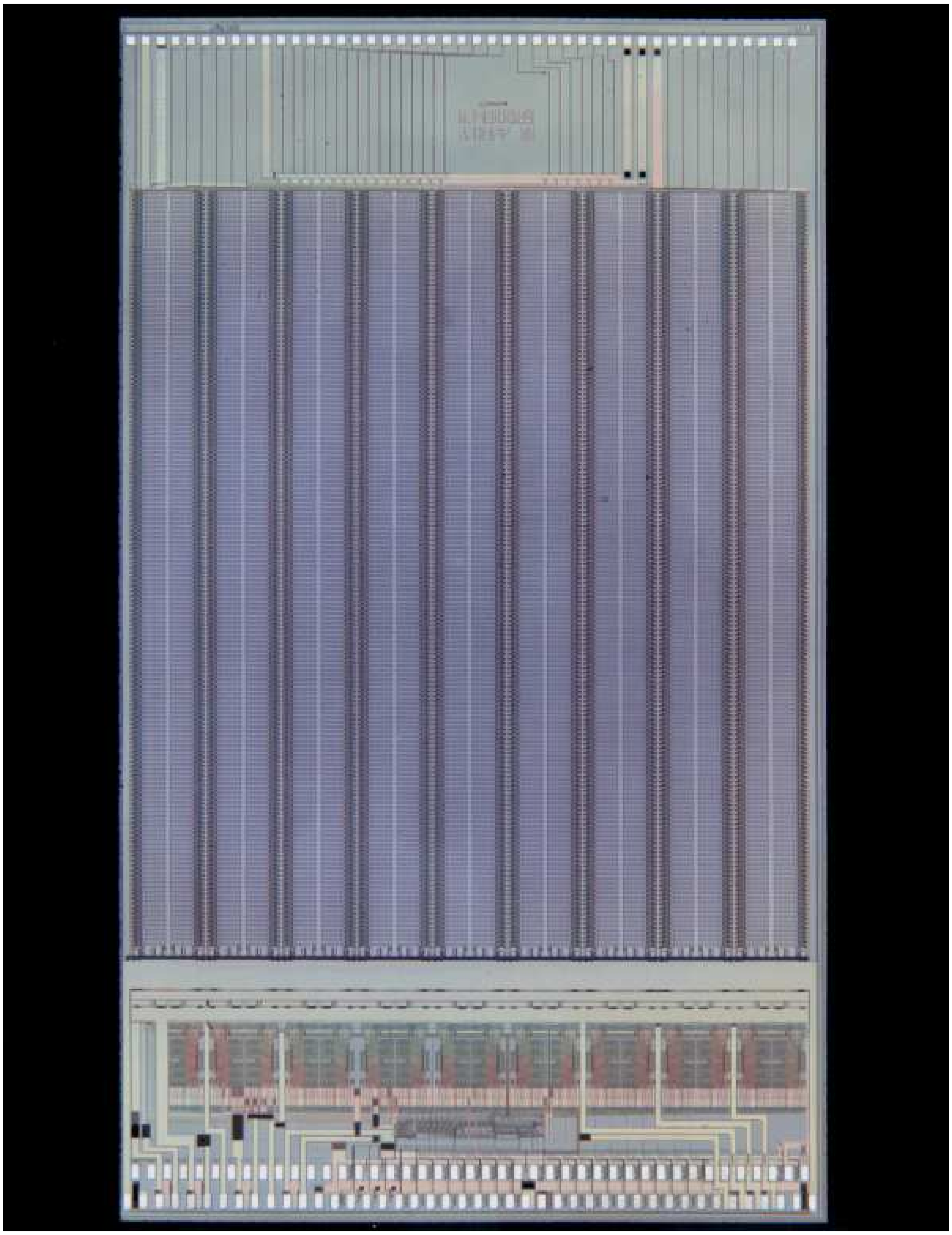


Figure 5.16: Photomicrograph of FPIX1.

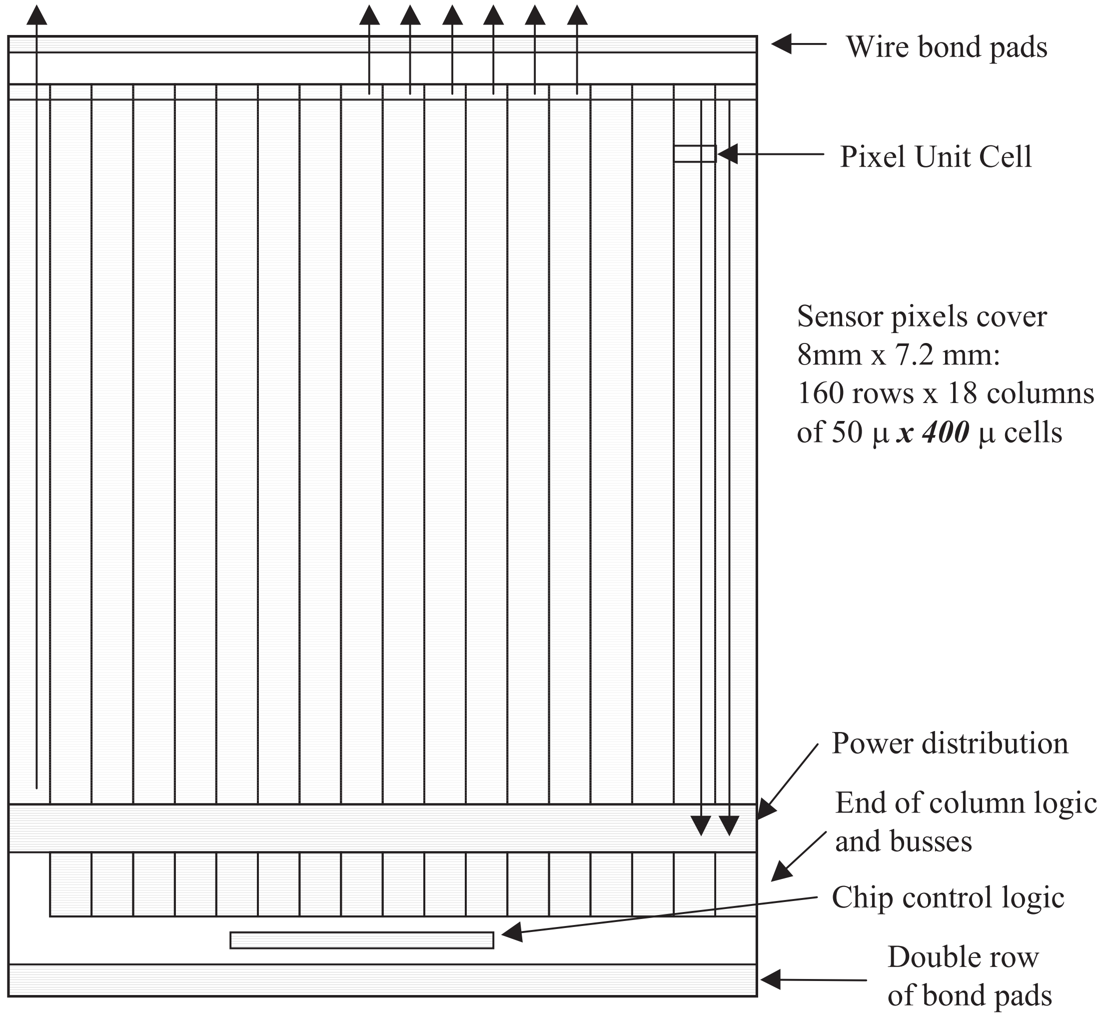


Figure 5.17: FPIX1 Floor Plan.

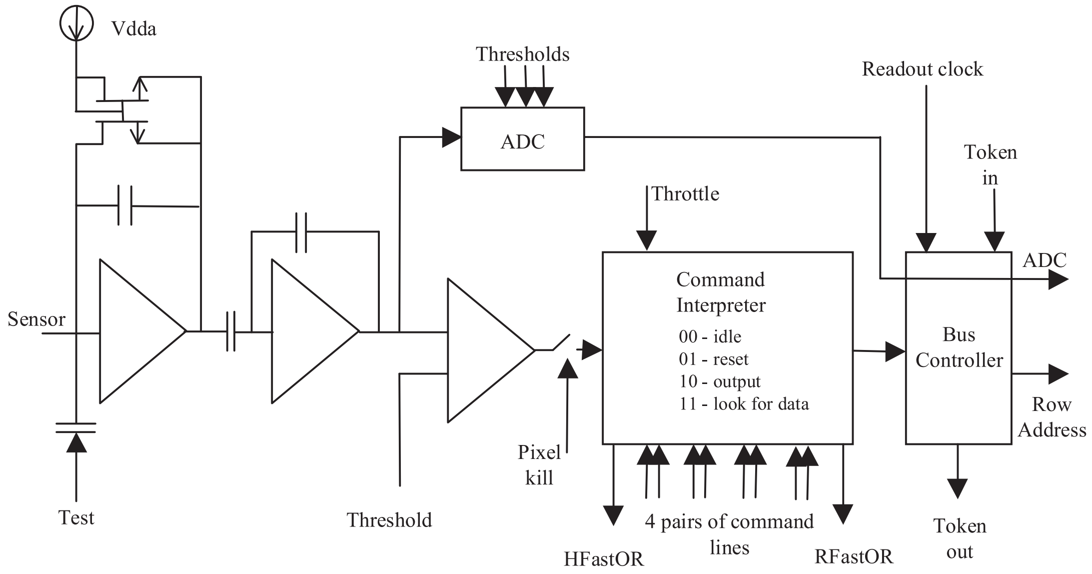


Figure 5.18: The FPIX1 Pixel Unit Cell.

alerts the end-of-column logic to the presence of a hit in the column may mean that only $\sim 125/160$ pixels in each column can be used in the beam test. This error is embarrassing, but completely understood, and trivial to correct in the next FPIX design.

As FPIX1 was designed, simulations were performed using both SPICE and Verilog. SPICE simulations were used in the analog design and also to provide parameters (e.g. time delays) for the Verilog simulations. Layout versus schematic checking was performed both at the cell level and at the full-chip level. The behavior of the entire chip was simulated using Verilog. We are continuing these simulations using pixel hits generated by MCFAST and Verilog models updated with measured FPIX1 parameters. We will rely heavily on Verilog simulations as we decide what features of FPIX1 to modify in the FPIX2 design.

5.3.4 Future plans

FPIX1 was fabricated using the HP 0.5μm CMOS process. This choice was made in order to facilitate the production of a final BTeV pixel readout chip using the radiation hard Honeywell 0.5μm SOI CMOS process. However, recent results from RD-49 at CERN indicate that standard commercial "deep submicron" (0.25μm and below) CMOS processes may be even more radiation hard than military processes such as the Honeywell 0.5μm SOI, provided only that a set of special design rules is followed [29]. An ASIC is currently being designed at Fermilab which contains amplifiers, comparators, and other circuits similar to those required for the front-end of the pixel readout. This chip will be fabricated using a standard commercial 0.25μm CMOS process. Its performance before, during, and after exposure to radiation will be evaluated. Based on this evaluation, we will decide whether to implement the next FPIX chip in HP 0.5μm, or to redirect our efforts towards a deep submicron implementation.

5.4 Bump bonding development

5.4.1 Introduction

The BTeV baseline pixel detector, like all other pixel systems, is based on a design relying on a hybrid approach. With this approach, the readout chip and the sensor array are developed separately, and the detector is constructed by flip-chip mating of the two. This method offers maximum flexibility in the development process, choice of fabrication technologies, and the choice of sensor materials. However, it requires the availability of a highly reliable, reasonably low cost, fine-pitch flip-chip attachment technology.

There are several bonding technologies which may be suitable for this. These include indium bumps, Pb-Sn solder bumps, fluxless solder bumps, and anisotropic conductive films(ACF) or tapes. The technology has to be able to fulfill the following requirements:

- small bump - the typical bump diameter and height for our pixel detector is between $10 - 12\mu$m.
- fine pitch (50μm)
- high yield - a defect rate of better than 10^{-4} is required.

5.4.2 Prototyping experience

Given the fact that we are mostly dealing with dies (for the readout chip) so far, and that our pitch is extremely fine, we chose to go with indium bumps which possess many attractive features. We have successfully deposited indium bumps on 30 micron pitch for an earlier tests of four hybridized arrays. The low melting point of pure indium, 156°C, allows for cold pressure welding, instead of solder reflow, and low forces (2-3 gram/bump) are used during the bonding process. For small prototype runs, there are only a handful of companies that can provide such services to outside customers. Because of the technical challenges (small die, large aspect ratio, fine pitch), we decided to go to two different vendors. These are Advanced Interconnect Technology (AIT) in Hong Kong, and Boeing North America Inc. in Anaheim. Our preliminary results showed that the failure rate in bump deposition from AIT was better than 2×10^{-4} at 90% C.L. There were a couple of dies with poor bump deposition(bridges or shorts, missing bumps) from the BOEING batch. Ignoring those two bad ones, the failure rate from BOEING was comparable to that of AIT. However, since the beginning of last year, BOEING no longer provides bump-bonding service to outside customers. We did, however, manage to get four pixel detectors using the FPIX0 chip bumped and mated at BOEING before they decided to stop their service. So far, we have fully tested one detector and found out that all the channels have been successfully connected. Fig. 5.19 shows the hit-map of a

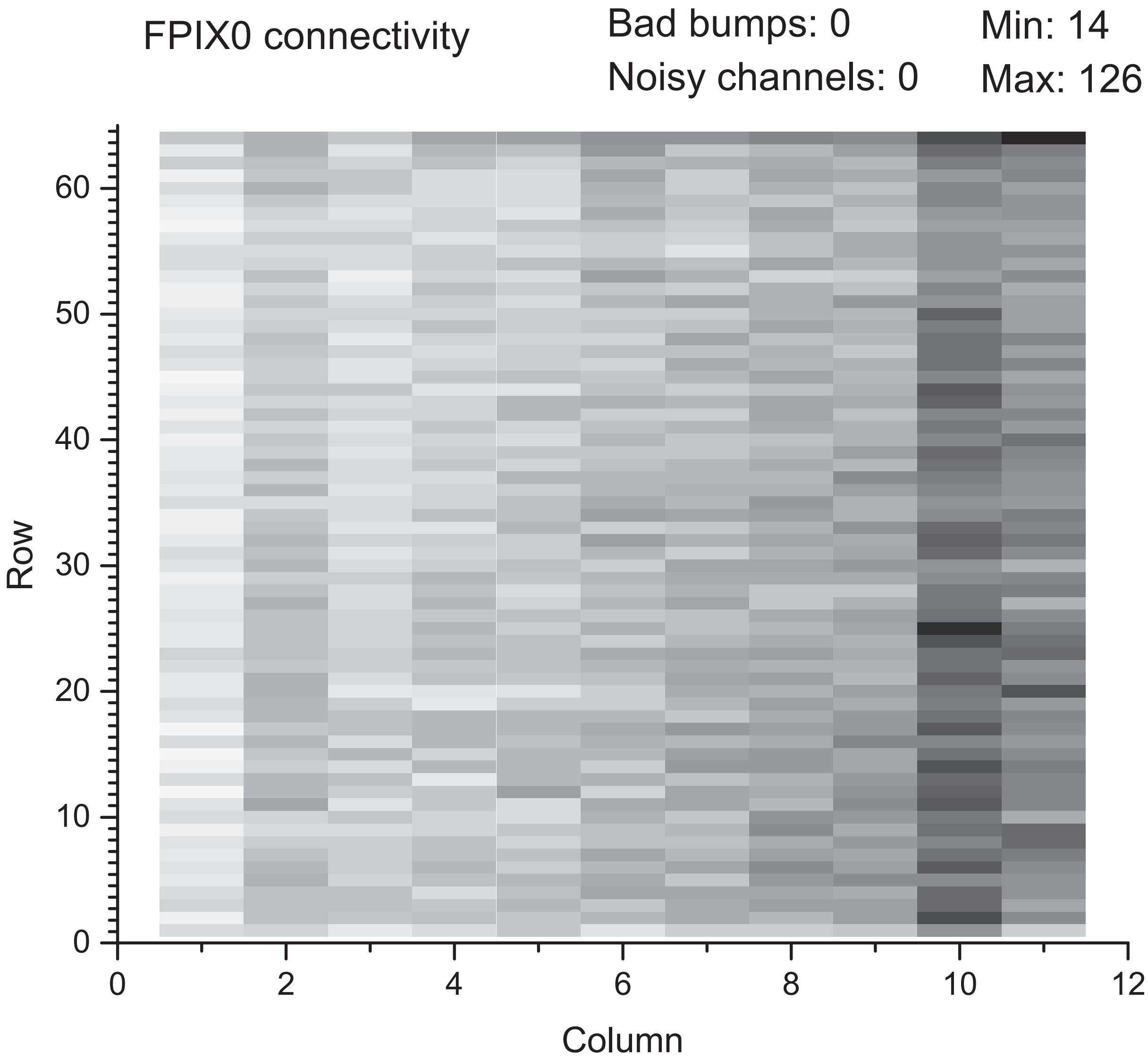


Figure 5.19: Hit map of a bump-bonded FPIX0 pixel detector showing 100% connectivity.

prototype pixel detector using a Sr^{90} source. All the channels were found to work, showing that all the bump bonds were good.

We have also received from AIT four thinned silicon wafers with bumps put on. The wafers were thinned at BNL. Two of the wafers were thinned to 250μm and two to 200μm. They were then sent to AIT where indium bumps were deposited. The pitch in the narrow dimension for the pixels on these wafers was 30μm. Visual inspection under a microscope showed that apart from a few bad or missing bumps on the edge of the wafers, all the bumps were good and fairly uniform. Based on these limited statistics, the failure rate in bump deposition on thin wafers is about the same as for 300μm thick wafers. However, we have not done any tests on flip-chip mating with thinned wafers.

There are lots of companies which can provide flip-chip mating services for large pitch (typically 200μm pitch and 100μm diameter bumps). Our requirements are at the very high end of the technology due to the small bumps, fine pitch, and relative low volume by industrial standards. These rule out high volume companies such as Flip-Chip Technology, Abpac, and Delco.

5.4.3 Large scale study of bump bonding yield

The goals of the large scale bump bonding study are to find a suitable vendor to do reliable and cost-effective bump bonding and to establish a quality control procedure so as to minimize the failure rate. We have established good working relations with industry and institutes working on LHC related projects. We have, and will continue to benefit from, sharing experience and information with LBNL.

Because of our fine pitch and small bumps requirement and that we have been dealing with small readout dies, we have opted to go for indium bumps. However, our recent contacts showed that Pb/Sn solder has been made to work at various places(e.g., IZM in Germany, SII in Japan, and AIT in Hong Kong) for a pitch as fine as 50 μm. Solder bumping is a self-aligning process and is a bit cheaper than indium bumps. However, the process typically involves the use of flux which needs to be removed after flip-chip mating. With small bumps, this may not be easy. Apart from environmental considerations, flux residue will lead to corrosion of the joint. The use of flux-less solder bump is now being pursued by industry. MCNC has a new fluxless, no-clean process called PADS (Plasma Assisted Fluxless Soldering) which has shown considerable success with flip-chip assemblies [34]. This relies on a pretreatment which enables the subsequent solder reflow in inert ambients. The use of this process, however, will pose stringent requirements on the thickness and material of the passivation on the readout and sensor wafers. This needs to be studied. Another concern is that the heavy Pb component in the solder bump may lead to a substantial increase in the electronics single-event upset rate. We plan to study this carefully together with the US-CMS collaboration.

Both pure In and Pb/Sn solder appear to meet our requirements for providing adequate electrical and mechanical bonds at 50 μm pitch. Our task is to start qualifying vendors with facilities to provide such services to outside customers and demonstrate the quality and yield of their bump bonding services by providing bumping and bonding of a large number of dummy pixel detectors. Long term effects on the bump bonding will also be studied. These include mechanical stability, effect of thermal cycles, irradiation, and handling the parts during assembly. The study will be done on different pitches and both technologies (Indium and Pb/Sn) will be tested.

For this purpose, we have designed a dummy assembly module consisting of an upper and a lower array. As much as possible, the two arrays will be close to the size of the real sensor and readout chip that we will have. Each array will have just metallization (Aluminum or gold) which will be covered with passivation with openings for the bump-bond pads and probe pads. The wafers will be made with two pitches, 30μm and 50μm. The upper array mimics the sensor chip and consists of pairs of bump-bond pads connected together. The lower array mimics the readout chip and consists of a large number of bump-bond pads ganged together with traces leading to probing pads on the periphery (Figure 5.20). When the two are mated, one can use an automatic probe station, with the appropriate probe card to test connectivity, and hence determine the presense of bad connections. The probing pads on the periphery are large enough so that they can also be used for wire-bonding. Assembly using these dummy modules can be practiced.

To characterize a vendor, we will send them a number of these dummy wafers. The wafers will be 300μm thick. Depending on the success of these tests, we will send them real sensor wafers and FPIX1 readout wafers so that we can do a real evaluation. We will also do tests on thin wafers. Table 5.3 lists the vendors that we have contacted and have expressed some interests in working with us. Both TRONICS and Rockwell can only handle 4” silicon wafers. This may be a concern for future production runs.

Vendor	*Technology*	*Location*	*Status*
AIT	*Indium* *solder*	*Hong Kong* *Hong Kong/Singapore*	*contract* *pending*
IZM	*Solder*	*Germany*	*SOW sent*
LMCO	*Indium*	*Massachusett*	*SOW sent*
MCNC/Unitef	*solder* *fluxless solder*	*North Carolina* *North Carolina*	*proposal rec'd* *proposal rec'd*
TRONICS	*solder*	*France*	*proposal rec'd*
Rockwell	*Indium*	*California*	*on hold*

Table 5.3: A list of potential vendors that can provide bump bonding services. SOW stands for Statement of Work. Proposal received means that they have already sent in their proposal to our SOW.

Our plan is to select up to four vendors from the list and work together with them. A lot of interaction between us and the companies is needed to understand the possible causes of failures, to rectify them, and to establish a detailed quality control procedure.

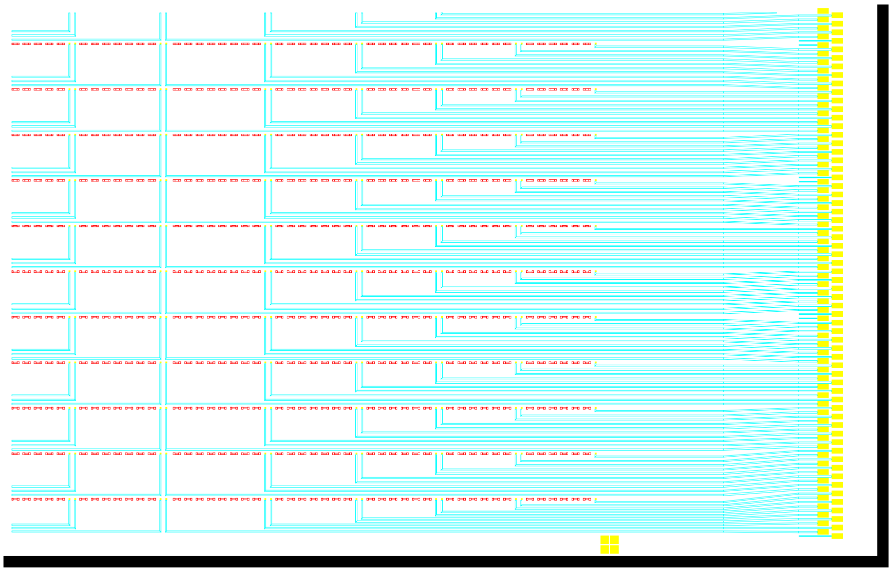


Figure 5.20: Layout of the lower array in the dummy wafer that will be used for large scale bump bonding studies.

5.4.4 Diagnostic tools

To characterize the various vendors and to understand the yield of the process, we need some diagnostic tools. Some of these will also be valuable during the production phase as a means of quality control. Possible tools include:

- optical microscope - Coupled with a precision x-y table, this is a simple and cheap way of detecting defects such as missing bumps, bridges, misplaced bumps, and irregular bump deposition.
- scanning electron microscope - This is useful to measure precisely the bump diameter, height, and placement precision. It also shows the under-bump-metallization.
- wafer profiler system - As a defect inspection tool, the profiler automatically detects bridges, under and oversized bumps, missing and extra bumps, and bump position. Furthermore, the system calculates bump height, diameter, volume, as well as coplanarity.
- acoustic imaging system - Using high frequency ultrasound, high resolution images can be produced to specific levels within a flip-chip mated device.
- X-ray imager - Using a high resolution micro-focus X-ray tube, a highly magnified X-ray image of the flip-chip mated device will show the presence of poorly connected or missing joints between the sensor and the readout chip.

We are working together with local industries and the bump bonding companies to determine which tool will provide the most useful information.

5.5 Data output and control

The input and output links to the pixel planes can be categorized based on the data bandwidth. The highest bandwidth corresponds to the pixel hits generated during the normal operation of the pixel detectors and collected by the pixel detector amplifiers and discriminator chips. However, these chips must be initialized, controlled and monitored by the DAQ system. This latter information will travel to and from the pixel detector planes at a considerably slower speed.

The electronics dedicated to the functions of receiving and transmitting data will be located on the corners of each pixel detector plane. This circuitry must be in charge of adapting electrical protocols, data formats, and data speeds to interface with the pixel detector chips. This electronics, called Data In and Out (DIO) is severely constrained by factors such us physical space, power dissipation, radiation dosage, EMI noise, mass, and outgassing.

The expected occupancy of the pixel detector planes, at the maximum luminosity BTeV intends to run ($2 \times 10^{32} cm^{-2} sec^{-1}$), can generate as much as 16 Gb/s of digitized data. The initialization, control and supervision data will not exceed 160Mb/s for each half-plane of the pixel detector system.

Figure 5.21 shows a block diagram of the proposed scheme. The two links on the upper left corner receive the 53MHz clock and the initialization and control data for the pixel readout chips. The links on the bottom right of the block diagram transmit the pixel hit data to the DAQ and Trigger Processor. These links will also be used to read out the configuration and status information of the pixel readout chip for supervision purposes.

The main building blocks of the DIO electronics are the optical receiver, the data-concentrator and the optical transmitter chips. The optical receiver is an interface chip between the serial optical channels and the pixel readout chips. At the physical level, it must detect the current generated by the pixel detector and generate the CMOS and LVDS signals to interface with the pixel readout chips. Furthermore, it must provide a readout clock with very low jitter (< 150ps). At the logical level, it must decode serial information into initialization and control commands for the pixel readout chips, and decode the BCO clock.

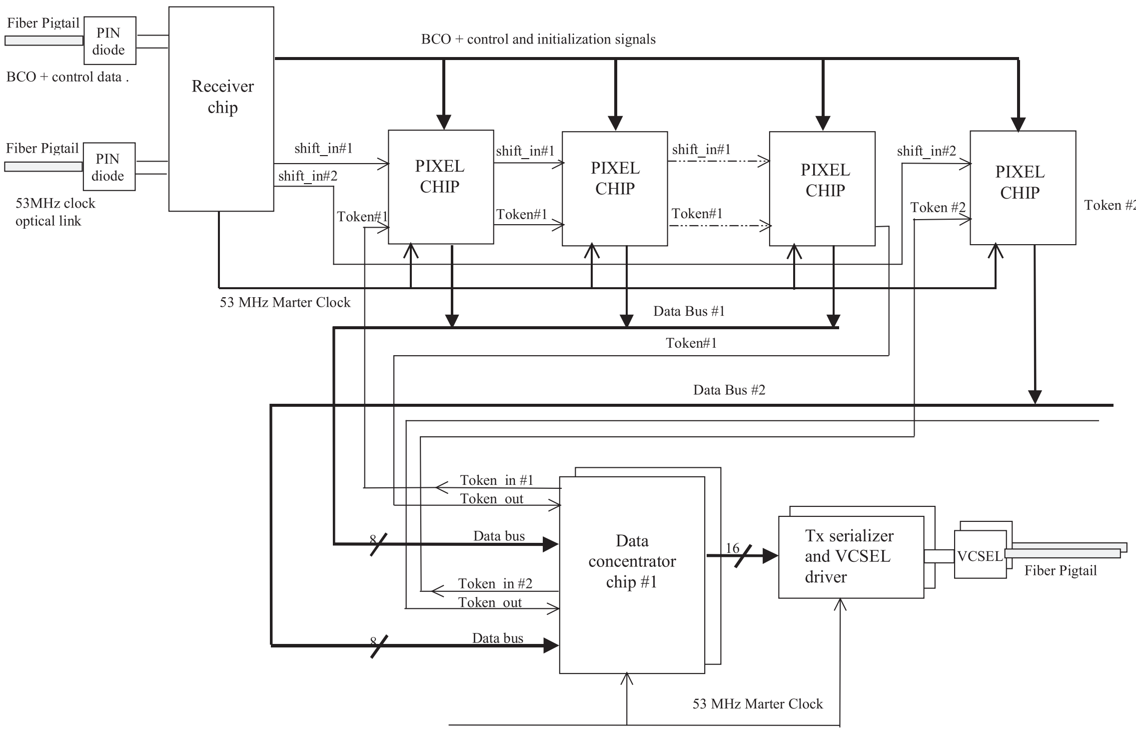


Figure 5.21: Block diagram of the proposed data I/O scheme.

The data-concentrator chip will group the data output of pixel readout chips. Its main function is to control the data flow from a pixel module, manage the token bit passing, equalize the data stream from the pixel chips into a serializer and optical transmitter. The data equalization is needed due to the fact that the hit and data rate is different in every pixel chip according with its proximity to the beam. The pixel chips closest to the beam will generate as much as 400 Mb/s of data and may need a dedicated link to the data-concentrator chip. Other chips, which do not generate as much data, will share a common bus and will be readout following a token passing algorithm. The data-concentrator must keep the token passing algorithm consistency and interface to the serializer and optical transmitter.

The optical receiver and the data-concentrator chips will be designed at Fermilab. The data I/O of the pixel detector planes will be carried through fiber-optic cables. Current studies show that these cables present the best solution to meet the requirements of mass, EMI noise, radiation, power consumption, and physical space. Since other experiments in the HEP community are also designing optical links for pixel detectors and calorimeters, we will benefit by collaborating with other groups working in this area. The following new technologies are currently under study:

- Vertical Cavity Surface Emission Lasers (VCSELs): Allow gigabits of data to be transmitted several hundreds of meters with few mW of optical power. VCSELs have been proven to be very radiation tolerant [30].

- Step Index Pure Silica Fibers: Accept high levels of radiation without increasing attenuation more than a small percentage [31]. Other graded index fibers are also under study. They are not as tolerant as the pure silica core, but they are less costly. Radiation tests will be performed to determine the level of acceptability.

- Custom ceramic and silicon packages: Very small custom designed packages for VCSEL and PIN diodes, and fiber-optic pigtails may allow optimizing for size, along with meeting the requirements of mass budget and radiation tolerance [32].

The fast data output links must deliver up to 16 Gb/s in every half-plane. We are researching the convenience of fast serial links, about 1 Gb/s, versus slower multifiber parallel links. Fast serial links are harder to design, but will reduce the number of fiber channels. No commercial devices exist which can accomplish the task of interfacing to the pixel chips, serializing the data, and transmitting over a fiber optic channel, and meet the pixel detector requirements. Most of the commercial devices are inappropriate for radiation environments. Furthermore, none of the commercial gigabit-speed data serializers meet the power budget. Among the possible alternatives, one of the most attractive is a CHFET-GaAs serializer being designed for the CMS experiment at CERN. We have already started to collaborate with the chip designer to meet BTeV specifications and to support testing at Fermilab. The protocol of this chip will be compatible with the so-called G-Link receiver at a frequency of 1.2 Gb/s, and its power consumption is less than 100 mW [33].

A second possibility is a prototype chip being designed by an R&D section of Honeywell. It codifies four channels of 16 bit words into 3 fiber links. Each link has two fibers for data and a third for clock. The chip is being implemented in a 0.25μm CMOS process. This is not a radiation hard process by design. However, some 0.25μm CMOS processes have proven to be very radiation resistant [29], which may possibly exceed the required levels. Some research groups at the RD47 collaboration at CERN are currently looking into this problem.

The optical components will require a special package. The package must be of low material, non-magnetic, must not outgas and be as small as possible. Furthermore, it must couple the VCSEL or PIN efficiently to the fiber. The fiber must be radiation tolerant as well. Figure 5.22 shows a possible solution where a ceramic substrate is used to host the VCSELs, PINs and fiber pigtails. This kind of substrate has ceramic or silicon alignment blocks, strain reliefs, and wire bond pad to interface to the transmiitter and receiver chips. The fiber pigtails can be ribbonized some distance outside the detector plane and connected to a standard fiber ribbon cable. This connection can be performed either by splicing or through a small-form-factor fiber ribbon connector.

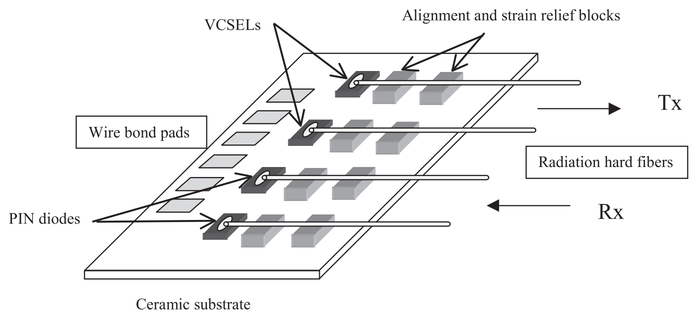


Figure 5.22: Sketch of a possible arrangement of the optical components.

5.6 Pixel Readout and Quadrant Processors

The data read out from the pixel detectors are transmitted through fiber-optic cables to Quadrant Processors. From the Quadrant Processors the data will be delivered to the pixel trigger system and the DAQ system.

There are several factors that impact the number of fiber optic links needed to transmit the data from the pixel detector to the Quadrant Processors: pixel hit and noise rate, forms of data compression of the pixel data, maximum optical-link data rate, and safety margin on the data transmission bandwidth.

The pixel detector will be used for the lowest level trigger, and all the data it generates, whether noise or actual hits, has to be transmitted to the Quadrant Processor. For this reason, it is important to include additional bandwidth.

A simple form of data compression can be achieved directly by the pixel chip, thereby decreasing the data rate. Data alignment by timestamp implemented in FPIX1 already achieves an initial degree of data compression. During readout, the timestamp has to be transmitted just once for all the hits that occurred simultaneously. In the next FPIX generation, we hope to implement what we have named "group" reading. The concept is that instead of reading a column by individually reading each pixel cell, we will read groups of consecutive pixel cells. So, we will not have pixel row address, but actually pixel group address, and the row position of several pixels can be uniquely identified with just one row address. Simulations have shown a data compression by a factor of 2.36 with respect to a pixel by pixel readout if we use a group size composed of four pixels.

To estimate the data that a pixel plane generates, we used the data compression scheme which includes alignment by timestamp and group reading formed by the data of four pixel cells. The readout data is composed of time stamp, group row and column address and two more bytes of pulse heights, each byte with two pulse heights. The pixel size is 50x400 μm^2. Other assumptions used in the simulation were:

1 . There are no noise hits

2 . Charge carriers: electrons

3 . Threshold: 2000 electrons

4 . Magnetic field of 1.6 T

5 . Minimum bias events with two interactions per 132 ns bunch crossing.

With these assumptions, we obtained a data rate for a pixel plane half of 2.5 GBits/s. We will now assume that the data rate for the fiber optic cables is 1.06 GBits/s, and we can now estimate the number of fiber-optic cables required per half plane. Assuming a safety margin of four, we obtain ten fiber-optic data links per half, or five per quadrant.

The Quadrant Processors will have inputs from three groups of five fiber-optic cables coming from the quadrant. These quadrants are the three adjacent quadrants that form a triplet. There will be a total of 124 Quadrant Processors. The main functions of the Quadrant Processors are:

- Sort the data from the pixel detector by time stamp
- Store the data from the pixel detector in its internal buffers until a decision from the Global Trigger
- Find the center of mass of the pixel hits
- Calculate the minivectors from the triplet and send the data to the Track processor
- After a decision from the Global Trigger, if the event was accepted, deliver the data to the DAQ. Otherwise, just discard the event and clear the event buffer.

5.7 Assembly issues

5.7.1 Introduction

The assembly design for the BTeV pixel detector system is dominated by the need to keep the amount of material to a minimum. This requirement is motivated by the adverse effects of multiple coulomb scattering on the achievable spatial and mass resolution. The former parameter is the most critical, since it influences many elements in the final capability of BTeV: separation of decay vertices from interaction vertices, trigger efficiency and enrichment, signal to background levels, proper time resolution, and sensitivity to multiple interactions per crossing. The mass resolution is also important, but mostly influences just the signal to background quality of BTeV data.

Counter to the need to keep materials at a minimum are the need for reproducible, stable position-determining supports, the need to remove significant amounts of heat directly from the active sensor areas, the need to move the detectors back from the interaction region during injection and machine-study periods, and the need to reposition the detectors reliably and accurately for physics data-taking.

In addition to these design criteria, we have to take into account that the pixel readout chips are limited in size, and the final detector assembly must be made as a matrix of communicating, overlapping pieces. The deadspace and/or gaps between materials also need to be minimized to maintain good efficiency for multiparticle charm and, especially, beauty decay modes of interest.

5.7.2 Interconnections

In principle, each half of a pixel plane could be made from a monolithic sensor with a single large readout device bump bonded to it for signal processing. In practice, this concept cannot be used for many reasons. The size of readout chips cannot be arbitrarily large due to the reticle sizes which are repeated to pattern a full wafer. In any event, yields on sensors and readout devices are such that the cost of making pieces so large would be more than prohibitive. Furthermore, the radiation dose varies dramatically from the near-beam region to the edge of each plane (approximately as the distance to the power 1.7 according to experience at CDF). Thus, various regions of sensor need different bias voltages to achieve depletion (as required for sensitivity), and the local breakdown voltages may also vary with radiation dose. Thus, we are necessarily drawn to a matrix of devices and their interconnections.

Current FPIX1 prototypes have sensitive areas of up to 7.4 by 8.0 mm^2. There is additional area on one edge for control and I/O circuitry, and for wire-bond pads. We may be able to increase the readout chip size somewhat, but the basic final element is not going to be much different from what we have now. We do plan to make many of the readout chips double size (back-to-back) and use larger sensors with multiple readout chips in the final detector. This can reduce the complexity of the interconnection scheme to a certain extent.

Nevertheless, each chip will have a full set of pads for interconnection, and the density of control and readout lines remains very high.

In order to gain experience on the technical issues, we are making a five-readout-chip sub-assembly with flexible cable interconnections. This will test industrial capability and allow us to get some design experience in this area.

Fallback options include the more ambitious, and technically challenging solution using traces on the sensor with bump bonds in place of wire bonds for readout-chip-to-control and output-line connection. This technology, referred to as MCM-D technology, is being pursued as an option by the ATLAS collaboration [35].

5.7.2.1 Pixel Plane High Density Circuitry Interconnect

The pixel detector design is severely bounded by several constraints which reflects on the choices of high density cables and pixel plane circuitry interconnect. Some of these constraints are mass, outgassing, distribution of the high voltage for detector bias, radiation exposure, reliability, and cost.

In a pixel detector, the readout chips are interconnected and the output signals are routed to the edges of the pixel planes. There, the output signals are wire bonded to the fiber optics data serializer integrated circuit for data readout.

In this application, the circuit density of the pixel module circuit interconnect is highly associated with the readout speed of the pixel chip. All data generated inside the pixel chip has to be readout for the lowest level trigger decision. In order to accommodate reasonable readout throughput, several readout buses will have to routed from the pixel chips to the data serializers. Based on the space available for routing it is easy to recognize that very high density circuits need to be employed.

We are presently researching the usage of low-mass flex-circuit interconnect. A preliminary analysis and consulting with vendors strongly suggest that this approach should effectively meet the constraints already outlined. For example, Kapton, one of the polyimide dielectrics used for flex-circuit manufacturing supports as much as 118 KV/mm, which is very good for the high voltage distribution of the detector bias. Other advantages include low mass (the radiation length X_o of Kapton is 286 mm), reasonable cost, radiation tolerance, etc.

We have researched a number of flex-circuit manufacturers and decided to work in conjunction with Fujitsu Computer Package Technologies, a division of Fujitsu, on the design, characterization and production of very high density flex-circuits for the pixel-plane interconnect. Fujitsu's nominal design rules include 25μm hole diameter and 20μm line width and spacing in four or more metal layers which represents the state of the art in this technology.

The present proposal for the interconnect is to glue flex-circuits on the top of the pixel-plane substrate, possibly a carbon-fiber plate, which will hold the detector and associated electronics. A similar approach is being pursued by D0 for the silicon detector hybrids and in the implementation of the majority of the pixel plane interconnect for ATLAS [36]. We estimate that the following four layers of flex-circuit would suffice:

- two layers for signal interconnects,
- one more layer for power and other signals and
- one layer for the ground plane.

These layers are quite thin and can be kept within 5μm of copper thickness. However, substantial mass can be saved if we can implement a three layer circuit, two with signals and power and the third, a ground plane made of aluminum. We will investigate the possibility of depositing a thin aluminum layer on the top of the carbon fiber substrate, use this layer as the ground plane, and laminate and interconnect the flex-circuit directly to this substrate. We are also researching vendors that could manufacture aluminum-on-polyimide flex-circuits.

As our first attempt to build a pixel module, we are presently laying out a five-chip pixel module using the FPIX1 readout chip. Figure 5.23 shows the progress on the layout. The wire bonding pads of three FPIX1 chips are already interconnected in this layout. These chips will be bump bonded to a detector and wired bonded to a high density flex-circuit interconnect. Both the chips and the circuit will be glued to a G10 board. The characteristics of this circuit interconnect will be the following:

- Metal layers: 5μm thick Cu, with 20μm line and space design rules. The first metal layer is capped with 2μm of Ni, and 1μm for wire-bondable gold.
- Film: 50μm thick Upilex SGA (which is comparable to Kapton)
- Film via: 25μm via on 108μm capture pad.
- Lamination: 50μm epoxy
- Z-vias: 200μm via on 350μm capture pad. Via filled with $SnSb_5$ solder.
- Size: 27mm x 54mm x 0.2mm

We intend to test this pixel module for several important performance characteristics: noise, threshold dispersion and crosstalk. Specifically for the flex-circuit itself, we intend to measure the electrical and mechanical performance [36, 37] and the effect of radiation, thermal cycle, and outgassing in vaccum.

Future research for the pixel plane interconnect will include technologies like multi-chip-module interconnect type D (MCM-D) and silicon (or possibly diamond) module hybrid to allow greater use of bump bonding and reduce the number of wire bonds. However, important issues such as high voltage distribution need to be carefully considered.

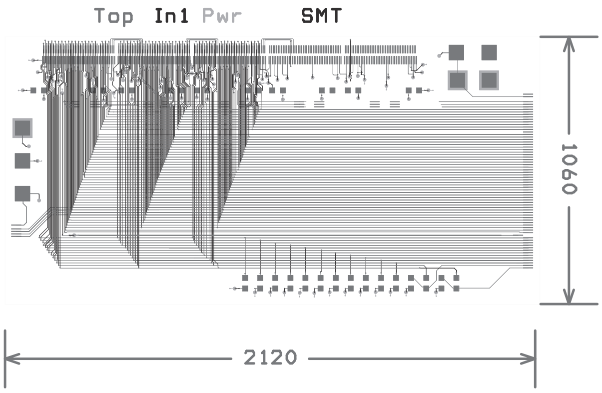


Figure 5.23: Schematic layout of the High Density interconnect prototype. Units are given in mils (0.001 inch).

5.7.3 Mechanical and cooling Issues

The mechanical design is driven by the goal of minimizing the material needed to achieve a satisfactory solution. Carbon-carbon, beryllium and CVD diamond structures appear most suited. We are focusing on plates attached to both sides of cooling pipes. The plates would only be required in the region where they support the pixel devices, with enough periphery to support control chips and pixel-plane-output cables. Rigid support to the outside world can, thus, be limited to a minimum, say the corners of the half-planes.

We will operate with a thermal budget of 40 microwatts per pixel, and doing the engineering as if all the sensors would be placed on the two sides of the above sandwich structure. Furthermore, the sensors will have to be operated at around -10° C and the temperature for each half-plane have to be maintained uniformly to within $\pm 1^{\circ}$ C. We plan to hire Energy Science Laboratories Inc. (ESLI), San Diego, to perform part of the research work for the development of the mechanics and cooling for a pixel detector plane. This contract will be divided in two phases. The first phase is the theoretical study and simulation of different options for the pixel plane mechanics and cooling, while the second phase is the actual implementation of a prototype of the most promising proposal.

Probably the most important item in this interaction will be the communication between ELSI and Fermilab. Since this is an investigative project and it may have unexpected results, we need to keep evaluating and assessing other promising solutions.

This investigation should answer questions about the ability of different substrates and cooling structures to remove heat from integrated circuits attached to their surface. Also, the mechanical stability of the structure is important for the performance. Two basic approaches are to be considered during this design study:

a) a plate that conducts the heat from the integrated circuits to one or more edges that are attached to a cold sink,

b) a plate that has the cooling channels embedded.

Several different schemes can be envisioned for the second option, such as fluid in pipes between plates, spacers with fluid flow "everywhere" inside, and "fuzzy" carbon spacer with fluid everywhere. Figure 5.24 shows the assembly of pixel plane with pipes between plates.

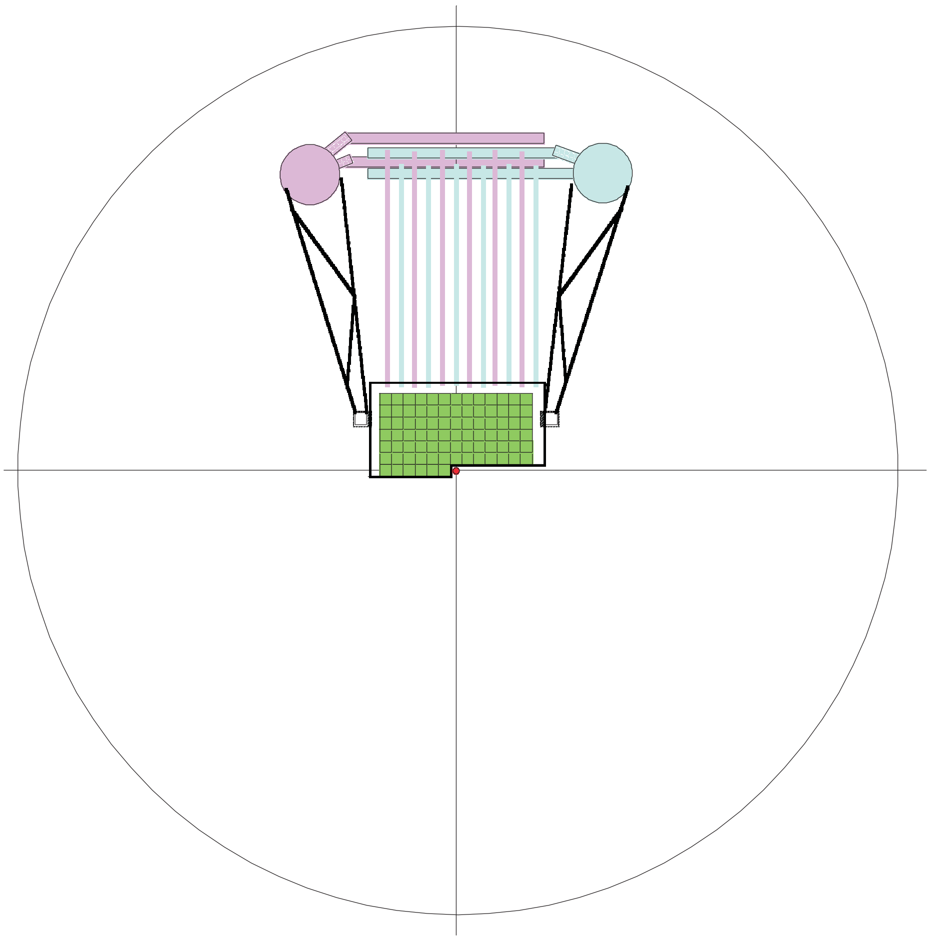


Figure 5.24: Schematic drawing of the pixel plane with cooling channels between the plates

A report from ESLI for the first phase will include the estimated performance of cooling systems and design recommendations, judgements on the feasibility of the economical construction of each option, estimates of the pressure drop of the cooling fluid, heat transfer properties, stiffness, vibration modes, reliability of piping, and temperature profiles. From these, the best solution will be picked.

In the second phase, a prototype of the best solution will be constructed. This model does not need to be a full working unit. However, it will contain enough of the features of the proposed design that it can be tested and compared to simulations.

5.7.4 RF shielding issues

The pixel detector will be installed inside the beam pipe enclosure in the C0 interaction region. The inside of the beam pipe is, in several respects, an aggressive environment for operation of a large, sophisticated detector. We would like to carry out a series of tests both in the laboratory as well as in the Tevatron to study the rf pickup and the shielding that will be required. Our objective with these tests is to investigate the electro- and magnetic (EM) interaction of the Tevatron beam with pixel detectors and associated electronics. When the beam passes close to the pixel detectors and electronics, it generates the equivalent of an electromagnetic pulse which can induce noise and cause other problems in the electronics. Shielding is required to protect the electronics from this pickup. This shielding has traditionally been a metallic skin and adds to the material budget. We are pursuing two kinds of tests, one in a simulator of the Tevatron beam in a laboratory set-up, the other inside the Tevatron itself. We feel that these tests will give us very important insights and will help us to generate guidelines for the design and shielding of the pixel system.

5.7.4.1 Beam simulator

An apparatus has been built to simulate the rf environment of the Tevatron. The basic structure of the test stand is a 5' long cylindrical coaxial pipe made out of stainless steel. It will be used to simulate the electromagnetic field from a bunch of beam particles. The electronics to be tested will be inserted through a slot in the outer conductor. An rf amplifier is used to generate the required current, estimated to be about 8.4Amp. Along with the beam simulator, we have constructed a computer driven test circuit which will feed a controlled bit pattern across a variety of signal paths inside the simulator apparatus. We also anticipate testing various shielding techniques. Our goal is to determine the level of rf pickup in the test circuit and to determine which signal path designs and shielding configurations minimize noise while keeping material use at a minimum. The results of these tests will then be applied to produce a prototype rf shield (if needed) to be tested in the C0 interaction region.

5.7.4.2 Roman Pot Tests for BTeV

We plan to install an existing Roman Pot and castle from E811 into the Tevatron. The pot will make it easy for us to insert various test devices into the region of the beam and make the desired measurements. We will make measurements with pixel detectors and chips placed closed to the beam, and study methods to decrease the EM interference of the beam on the electronics. Our current plans are for a two-year investigation. This will start with simple devices to make EM measurements. Each test will build on the previous tests. The goal is to construct sophisticated electronic test devices with active circuitry similar to whar we could use in the real detector. We hope to learn how to construct a small test pixel detector and take data from it.

The following topics will be investigated:

- layout issues and line terminations;

- rf shielding: its shape, thickness, material, grounding, and support;
- power supplies: their design, location and distribution

Theoretical analysis of the susceptibility of layout traces and transmission lines will be conducted as a first stage of design, to evaluate electromagnetic compatibility rules between the detector and the surrounding media. In such an analysis, it is very difficult to consider all the variables affecting the detector design. Physical prototypes of both general and critical parts of the detector will help to understand the final EM susceptibility of the detector.

We envision two options for the location of the power supplies: outside and inside the enclosure. The first option simplifies the design and procurement of power supplies, but imposes restriction on the distribution cables. The cables can act as EM pick-up elements conducting noise into the detector or generating conductive paths between the different layers of the detector. In these tests we will be able to analyze the performance of this topology, the level of filtering and shielding, and the input-output isolation of the power supplies. Note that the addition of filtering and shielding may require too much mass for the experiment.

The second option imposes more challenges due to the need of locating DC-DC converters near the detector to break conductive paths and decrease the EM pick-up. The pixel detector will be subject to strong magnetic field, which does not allow any magnetic material in the design of such converters. Research will be conducted to evaluate the option of using DC-DC converters based on piezoelectric ceramic transformers [38, 39] and the use of high frequency air core transformers.

5.7.5 Material budget

Table 5.4 summarizes our current understanding of the material budget for one single pixel plane. Each pixel plane will consist of a number of modules. Each module may consist of 4 to 7 readout chips. Figure 5.25 illustrates the way we envisage each plane will be assembled. To account for non-sensitive area in each pixel sensor required for guard rings and scribe edge, the table assumes a 25% overlap of modules on opposite sides of the substrate (this overlap is not shown in the figure). Our baseline design calls for a 5 mm thick Be substrate with integrated cooling channels.

5.7.6 Outlook

Between the time of this PTDR and a full TDR, we anticipate having results from the various prototypes and the engineering studies described above. These will be used to develop engineering calculations and assembly drawings of an actual BTeV pixel detector - optimized for minimum material and stability of positioning. We do anticipate final alignment using charged particles from real physics events for each run (each repositioning of the pixel half-planes), including whatever is needed to make the trigger work effectively.

$Item$	$Thickness(mm)$	$X_0(mm)$	$X/X_0(\%)$
$Sensor$	0.25	93.6	0.30
$Readout\ chip$	0.15	93.6	0.20
$Bumps\ and\ wire\ bond$	0.02	10.0	0.004
HDI	0.224	284	0.08
$Components\ on\ HDI$			0.04
$Glue$			0.02
$Coolant(water/alcohol)$	0.35	40	0.02
$Substrate(Be)$	0.5	353	0.14
$Total$			0.80

Table 5.4: Material budget of a BTeV Pixel plane. This assumes that we can build a pixel plane using only one substrate. Within a plane, we assume that there is a 25% overlap in sensor or readout chip, that the HDI has 100% coverage and that the cooling fluid and cooling channel covers 20% of the substrate. Each HDI has 4 layers with each layer consisting of 50μm kapton and 6μm of Cu.

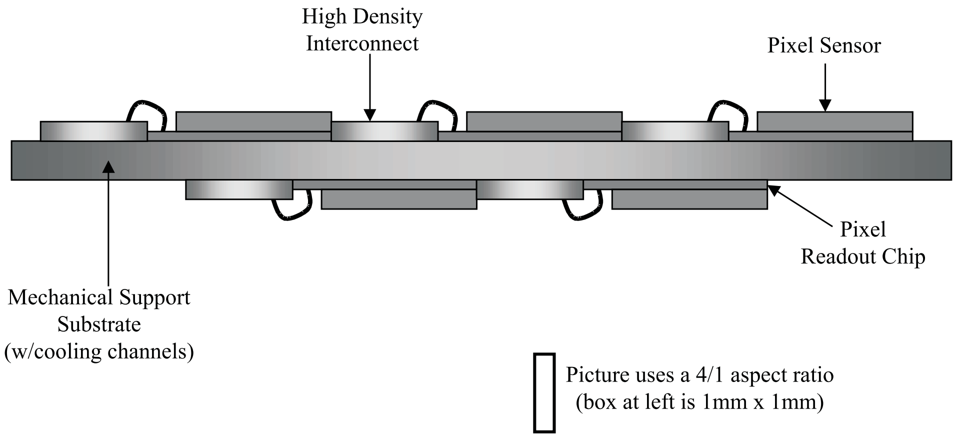


Figure 5.25: Schematic of the assembly of a pixel plane.

5.8 Pixel Beam Test

Beam tests of prototype pixel detectors will take place during the 1999 fixed target run. During this six month period it is planned to test silicon sensors bonded to both FPIX0 and FPIX1 readout chips and a diamond sensor bonded to an FPIX0 chip.

5.8.1 Goals

FPIX0

- To measure the efficiency, charge sharing and resolution as a function angle of incidence, number of ADC bits, and readout threshold. The effect of magnetic field will also be studied.

FPIX1

- To test the fast readout in a high intensity beam.
- To study the resolution as a function angle of incidence, threshold, and different sensor designs.

Diamond Sensor

- To determine whether diamond pixel detectors bonded to FPIX0 chips provide good efficiency, resolution, uniformity, and signal to noise.

5.8.2 Test Stand

The test stand is shown in Fig. 5.26. The central aluminum box at the bottom of the figure has slots for mounting pixel test boards at several different angles. Two triplets of silicon strip detectors form a telescope for tracking the beam particles.

5.8.3 Silicon Telescope

There are six planes of 20 μm pitch silicon strip detectors which are read out with SVX2b chips. A VME based data acquisition system has been adapted from a CDF test stand. It consists of an SGI workstation, a VME CPU, a Silicon Test Acquisition and Readout board (STAR), a Test Fiber Interface board (TFIB), and a Test Port Card (TPC). The STAR has 3 data buffers which can accept data from the SVX2b chips via the TPC. Simulations indicate that we should be able to project track locations at the pixel detector with 4-5 μm precision.

5.8.4 Pixel Readout Board

The pixel readout has been designed to be integrated with the existing SVX2b readout. An RS485 VME interface board is used to send control signals to the pixel readout chip via a pixel readout board. The pixel readout board is physically separated into two boards - an inner board on which is mounted the pixel sensor bonded to the readout chip, an 8-bit flash ADC and power supplies, and an outer board which has all the readout control components. The outer board can accept a trigger either via a NIM or ECL input or through the RS485 cable. After receiving a trigger, the data is read out into one of the STAR buffers. The same outer board can be used for either FPIX0 or an FPIX1 chip. The inner board will be different for each readout chip. For example, the FPIX0 board has a flash ADC whereas the FPIX1 chip itself has a 2-bit ADC in each pixel.

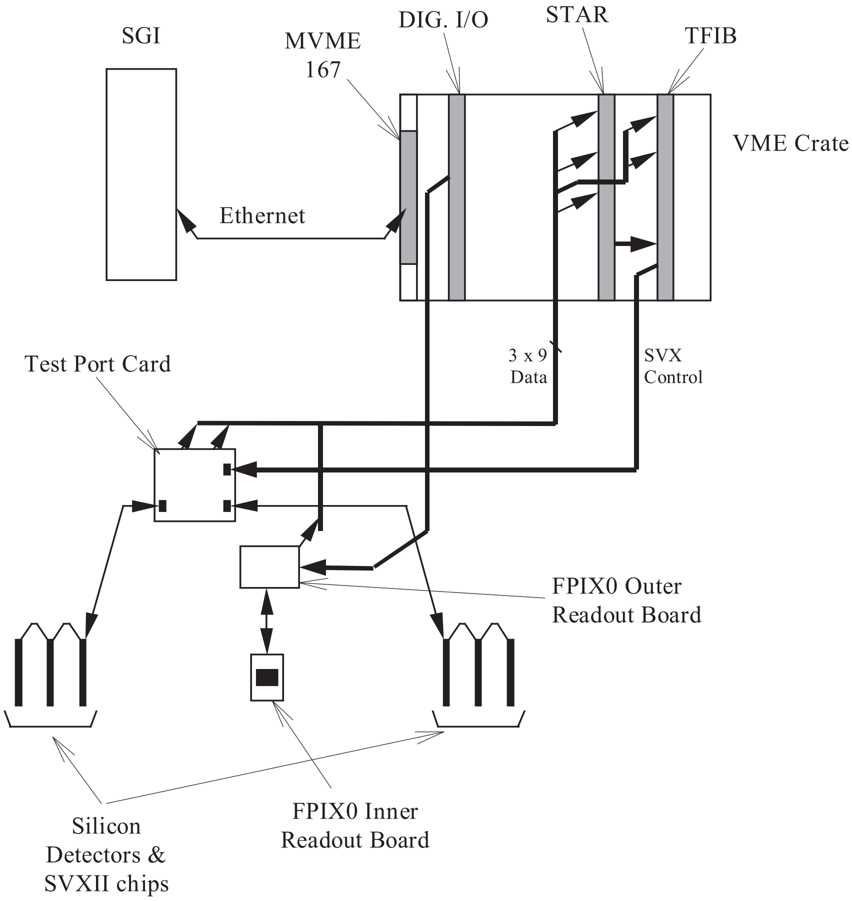


Figure 5.26: Test Stand for Pixel Beam Test.

Bibliography

[1] D. Anderson and S. Kwan, Proc. of Pixel98 - International Pixel Detector Workshop, May 7-9 1998, Fermilab-CONF-98/196.

[2] B.S. Avset and L. Evensen, Nucl. Instrum. Meth. A377 (1996), 397.

[3] The ATLAS Pixel Detector Technical Design Report, CERN/LHCC-98-13.

[4] The CMS Tracker Project Technical Design Report, CERN/LHCC-98-6.

[5] R.H. Richter, Nucl. Instrum. Meth. A377 (1996), 412.

[6] Marina Artuso and Janchun Wang, BTeV Internal Report BTeV-Int-97-18, February 1998

[7] E. Belau *et al*, Nucl. Instr. and Meth. **124**, (1983) 253.

[8] R. Turchetta, Nucl. Instr. and Meth. **A 335**, (1993) 44.

[9] L.H.H. Scharfetter (RD19), *Active Pixel Detectors for Large Hadron Colliders*, CERN Thesis, 1997.

[10] A. Mekkaoui, Talk given at the 3rd International Meeting on front-end Electronics for High Resolution Detectors, Taos, NM (1997).

[11] Penny Kasper, BTeV Internal Report BTeV-int-98/3, March 1998.

[12] G. Cancelo *et al.*, BTeV Internal Report BTeV-int-98/16, September 1998.

[13] Penny Kasper, BTeV Internal Report BTeV-int-99/4, March 1999.

[14] G. Lutz *et al.*, Talk Presented at the Workshop on Vertex Detectors: State of the Art and Perspectives, Erice, Sicily, (1986).

[15] P. Fisher *et al.*, ATLAS Internal Note INDET-NO-086 (1994).

[16] M. Artuso, Presented at 29th International Conference on High-Energy Physics (ICHEP 98), Vancouver, Canada, 23-29 Jul 1998, Syracuse Un. Preprint HEPSY-98-2, (1998).

[17] Technology Modeling Associates, Inc., Sunnyvale, CA

[18] S. Parker, talk given at Fermilab (1998).

[19] S. Seidel, "Pixel Sensors for ATLAS", in Proc. if Pixel98 Workshop, Fermilab, Batavia, IL, May 1998, ed. D. Anderson and S. Kwan, Fermilab-CONF-98/196, 263.

[20] M. Da Rold et al., IEEE Trans. Nucl. Sci. 44 (1997), 721.

[21] C-Y. Chien, private communication.

[22] Z. Li et al., IEEE Trans. Nucl. Sci. 45 (1998) 348.

[23] Z. Li et al., Nucl. Instrum. Meth. A408 (1998).

[24] D.C. Christian, et al.,"Development of a Pixel Readout Chip for BTeV",FERMILAB-CONF-98-354, 1998. To be published in the proceedings of the 7^{th} International Workshop on Vertex Detectors (VERTEX'98).

[25] S. Zimmermann, et al.,"High Readout Speed Pixel Chip Development at Fermilab", FERMILAB-CONF-98-278, 1998. Published in Rome, 1998, Electronics for LHC Experiments, 528.

[26] A. Mekkaoui, et al.,"Results from an FPIX0 Chip Bump Bonded to an ATLAS Pixel Detector", FERMILAB-CONF-98-278, 1998. Published in Rome, 1998, Electronics for LHC Experiments, 599.

[27] L. Blanquart, et al., Nucl. Instrum. Meth. A395 (1997), 313.

[28] M. Artuso and J. Wang, "Some Further Studies on Factors Affecting the Pixel Resolution", BTeV-int-98/13, BTeV Internal note.

[29] W. Snoeys, "Radiation tolerance beyond 10 Mrad for a pixel readout chip in standard submicron CMOS", Proc. of the 4th Workshop on Electronics for LHC Experiments, Rome, September, CERN/LHCC/98-36, 114.

[30] R. Gomez, M. Johnson, and M. Baert, "Preliminary studies of the use of Vertical Cavity Surface-emitting lasers for data transmission in High Energy Physics", Fermilab-Pub-97/098.

[31] I. Bhandal, "Optical Fiber for telecommuncations", presented at the 2nd Workshop on Optical Readout Technologies for ATLAS, Oxford University, England, Jan. 7-8, 1999.

[32] J. Hall, "Packaging for a high-radiation environment",presented at the 2nd Workshop on Optical Readout Technologies for ATLAS, Oxford University, England, Jan. 7-8, 1999.

[33] P. Denes, “Radiation hard gigabit systems”, presented at the 2nd Workshop on Optical Readout Technologies for ATLAS, Oxford University, England, Jan. 7-8, 1999.

[34] N. Koopman et al., “Fluxless no-clean assembly of solder bumped flip chips”, Presented at 46th ECTC 1996, Orlando, Florida, May 28-31, 1996.

[35] P. Gerlach, “MCM-D Technology for Pixel Detector Modules”, in Proc. of the Pixel98 Workshop, Fermilab, Batavia, IL, May 1998, ed. D. Anderson and S. Kwan, Fermilab-CONF-98/196, 497.

[36] P. Skubic et al., “Flex Circuits for the ATLAS Pixel Detector”, in Proc. of the Pixel98 Workshop, Fermilab, Batavia, IL, May 1998, ed. D. Anderson and S. Kwan, Fermilab-CONF-98/196, 519.

[37] K. Neriyanuri, “ATLAS Pixel Detector High Density Inerconnect design and testing of ASIC analog multiplexer”, Master Thesis (unpublished), 1997.

[38] M. Imori et al., “A High voltage Supply Using a Piezo Ceramic Transformer”, IEEE Nuclear Science Symposium, 1995.

[39] Chih-yi Lin, “Development of a piezoelectric transformer converter”, Virginia Power Electronics Center, Current Magazine, Spring 1993.

Chapter 6

Forward Tracking System

6.1 Introduction

The forward charged particle tracking system will play a central role in nearly every physics analysis in BTeV. The major functions of the system are to measure momenta for tracks found in the pixel system, to measure all parameters for tracks which do not pass through the vertex detector (such as K_s and Λ^0 daughter tracks), and to project tracks into the RICH systems and EM calorimeters. Measurements from the forward tracking system can also be used at Level 2 of the trigger to greatly improve the momentum estimate for muon tracks found in the muon spectrometer at Level 1. Level 2 triggers can use forward tracking information to link these tracks to the pixel system and associate them with vertices.

The requirements of the forward tracking system are:

1. Provide tracking coverage in each arm in the angular range $\theta = \pm 10$ mrad to $\theta = \pm 300$ mrad over a single arm distance of 7m.

2. Achieve a bend-view resolution of $\leq$ 100 μm (per station). This corresponds to a momentum resolution of $\sim$ 1% for 100 GeV tracks reaching the last measurement station just before the EM calorimeter.

3. Provide robust pattern recognition for charged particles, in conjunction with the pixel system.

4. Materials in the sensitive detector region should be minimized in order to reduce the effects of multiple Coulomb scattering, ionization energy losses, and hadronic secondary interactions. A design goal is to introduce less than 1% X_0 per measurement station.

5. Separate information from different beam crossings. This means that the maximum drift time must be less than the time between bunch crossings (132 ns).

6. The detectors should be robust mechanically and electrostatically but should be serviceable during accelerator down times. This implies an ability to remove chambers without removing the beam pipe.

7. Cost effective, proven technologies should be used to the extent possible.

Meeting these requirements in a forward hadron collider environment will be challenging. One reason is that the particle flux varies by three orders of magnitude over the tracking volume. Sensitive detector elements must be kept short in the high occupancy regions. Since the tracking volume is quite large (the most downstream chamber covers an area 4.2m by 4.2m), we require large area, low mass detectors such as drift chambers, straw tube detectors, or honeycomb drift chambers in the forward region. One finds that maintaining short sense wire lengths in the high occupancy region while minimizing the amount of material associated with the necessary readout and detector services (electronics connections, gas manifolds, module/wire supports) becomes especially difficult. We considered straw tube configurations which placed shorter straws in the high occupancy region, growing longer with radial distance from the beam pipe. These solutions however always resulted in unacceptable amounts of material in the sensitive detector volume. Our conclusion was that a divided geometry which employs high granularity detectors with low mass readout in the inner region and large area tracking devices for the outer region was necessary to achieve acceptable occupancy levels while maintaining good momentum resolution.

6.2 Detector Geometry

The baseline BTeV design consists of 14 tracking stations (7 in each arm) employing straw drift tube detectors and silicon microstrip detectors. The straw tracker provides large area outer coverage and is comprised of planar straw tubes arranged in three measurement views: Y (vertical, bend view), and two stereo views U and V oriented at ± 45 degrees to X (these stereo angles are still under study). Each measurement view consists of three physical layers of 4 mm diameter straw tubes in a close-packed configuration. Two such layers are required to resolve the left-right ambiguity in the drift-time coordinate measurement. The third layer provides redundancy and increases the efficiency of the coordinate measurement. The first three measuring stations nearest the interaction region cover only the outer tracking region while the downstream stations have an inner square aperture with dimensions corresponding to the ± 10 mrad acceptance limit. The details of the forward straw tracker are summarized in Table 6.1.

In the first three tracking stations on either side of the dipole, silicon microstrip detectors cover the high occupancy region near the beamline. Design work for the silicon portion of the forward tracker has only recently begun. We have performed studies both with three planes of single-sided detectors, and with two planes of double-sided detectors. We simulated single-sided detectors with 300 micron pitch (resulting in a factor of 10 reduction in hit occupancy with respect to straw tubes) and covering an area of 24 cm by 24 cm arranged as three views: X, Y and a 45 degree stereo view. While single-sided detectors have an advantage in being simpler in design, the main disadvantage is the additional material required to make the same number of position measurements. Since keeping material to a minimum is one of our primary concerns, our baseline proposal is to employ two planes of 300 μm thick

Table 6.1: Properties of the baseline forward straw tracker

Property	Value
Straw size	4mm diameter
Central hole	5 cm $\times$ 5 cm to 24 cm $\times$ 24 cm
Total Stations	14 (7 per arm)
Z positions (cm)	96, 146, 196, 296, 341, 386, 706
Half size (cm)	30, 45, 60, 90, 105, 118, 210
Views per station	3
Layers per view	3
Total number of straws	78132
Total station thickness ($< L > /X_0$)	0.6%
Readout	ASD + timing chip (6 bits), sparsified
Total channels	118,464
Rate requirements	beam crossing rate is 132 ns.

double-sided silicon microstrip detectors with 100 μm pitch. Standard sized silicon wafer modules are mounted on low mass carbon fiber support pads and are arranged in a ladder readout configuration. Four views are provided: X, Y, U, V where the U and V stereo angles are matched to the corresponding straw tracker stereo views. The properties of the proposed baseline forward silicon microstrip tracker are summarized in Table 6.2.

The forward tracking system provides up to 63 measurement points for tracks within the $\pm$ 300 mrad acceptance.

Table 6.2: Properties of the proposed baseline forward silicon tracker

Property	Value
Wafers	4×6 cm^2 / 2.5×4.75 cm^2
Pitch	100 μm
Thickness	300 μm
Ladder modules	12 per station
Coverage	24 cm $\times$ 24 cm
Central hole	5 cm $\times$ 5 cm
Total stations	6 (3 per arm)
Z positions (cm)	100, 150, 200
Measurement views	4 (X,Y,U,V)
Channels per view	2400
Total channels	57600
Readout	double-sided, sparsified binary

6.2.1 Layout Considerations

The important considerations for the forward tracking layout are:

1. Material should be kept to a minimum, to minimize both multiple Coulomb scattering and photon conversions.

2. Coordinate views should be chosen to emphasize position measurement resolution in the bend (vertical) direction.

3. Stereo coordinate views should be chosen to maximize pattern recognition capability without adversely complicating the detector module construction or significantly increasing its cost.

4. Sensitive lengths of straw tubes close to the beam pipe should be kept as short as possible in order to minimize the hit occupancy in a given straw detector.

5. The number of measurement layers per view should be chosen to maintain good efficiency while minimizing the material in the decay volume.

Our baseline layout places three stations in the fringe field region of the dipole magnet, three stations in the field free region just upstream of the RICH, and one station just downstream of the RICH.

6.3 Performance Issues

6.3.1 Straw Hit Occupancies

We anticipate large occupancies in the inner forward tracking region detectors, and have concluded that detectors with higher granularity than straws are necessary to provide adequate performance. To study the occupancies in the forward straw tracker we simulate generic $b\bar{b}$ events using the PYTHIA event generator and the MCFAST detector simulation package. Minimum bias inelastic events are added to the $b\bar{b}$ event to yield a Poisson event multiplicity distribution with a mean of two. This simulates running conditions at a luminosity of $\mathcal{L} = 2 \times 10^{32} cm^{-2}s^{-1}$. Particles from hadronic secondary interactions, photon conversions, and decays in flight are included in the event sample. The occupancies for the baseline geometry are shown in Figs. 6.1-6.3 for the Y layer in four of the seven forward straw tracking stations. Here we define the occupancy as the average number of charged particles traversing a straw cell in an event. The average occupancy is plotted as a function of the straw number offset from the beam center. For comparison, we show the occupancy resulting from the $b\bar{b}$ event itself.

Forward stations which are closest to the interaction region have occupancies approaching 8% for straw tubes nearest the beam pipe. If a 5 cm beam aperture is used, the occupancy rises to $\approx 20\%$, which is an extremely high value. The notch pattern where the occupancy

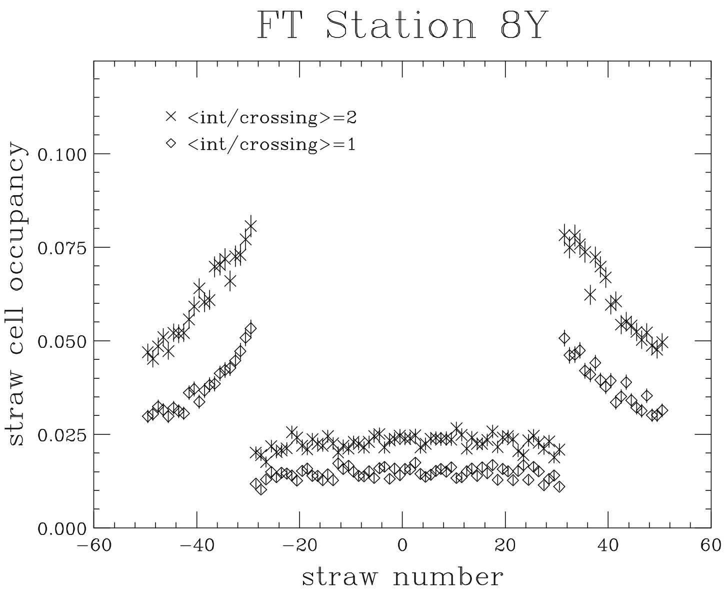


Figure 6.1: Hit occupancy for the forward straw tracking station closest to the interaction region. The occupancy is the average number of particles passing through a given tube in an event.

is reduced to about 2% results from shortening the tubes so that they end 12 cm from the beam. Forward silicon microstrip detectors provide coverage for this $\pm$ 12cm region.

Figs. 6.2-6.3 show the occupancy for stations located immediately upstream and downstream of the RICH. These stations have beam pipe apertures of ± 3.5 cm and ± 7.0 cm respectively, so their straw tubes extend into the high rate region. The occupancy in the station just upstream of the RICH is close to 10%, an uncomfortably large value. The track separation due to the increased distance from the interaction point results in lower occupancies for the station downstream of the RICH.

In Fig. 6.4 we plot the fraction of tracks which are shadowed as a result of the high occupancy in the station upstream of the RICH. Tracks closest to the beam pipe will be shadowed by other tracks in the event about 0.6% of the time. This can be reduced by enlarging the beam gap in the chamber, with a corresponding loss in acceptance.

It should be noted that the occupancies and shadow fractions shown are for a single layer and not for the resolved 3-layer hits. We plan to investigate ways in which the drift time information from all three layers can be exploited to recover track information from these doubly hit cells.

6.3.2 Momentum Resolution of the Forward Straw Tracker

In this section we study the momentum resolution of the proposed baseline forward straw tracking system. For the time being we limit our attention to multiple scattering and the

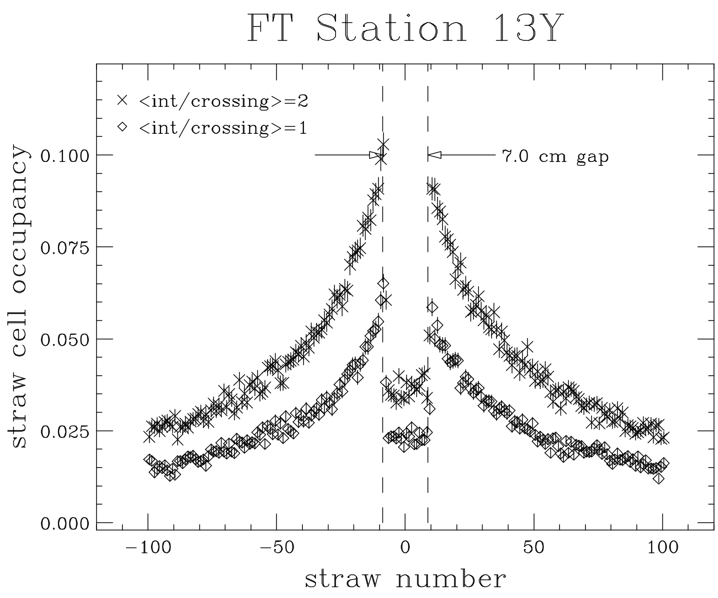


Figure 6.2: Hit occupancy for the forward straw tracking station immediately upstream of the RICH. The occupancy is the average number of particles passing through a given tube in an event.

intrinsic resolution of the straw detectors, since these are the parameters most directly affected by technology choices for the forward tracking system. Additional contributions to the resolution, to be considered later, will come from module misalignments, non-uniformities in the manufacture of detector elements, errors in the dipole field mapping, and non-linearities caused by hit adjacencies. To allow for these errors, we have assumed a conservative intrinsic straw resolution of 150 μm per measurement layer. Drift-time measurement accuracy of 100-120 μm or better has been achieved by the ATLAS TRT group [9] for 4mm diameter straw cells for certain electronic threshold and signal peaking times.

Multiple scattering causes random changes in a particle's direction, producing scattering angles that vary inversely to the particle momentum. This results in a fractional momentum resolution which is roughly constant. The intrinsic position resolution of the detector produces a linear dependence of $\sigma(p)/p$ on particle momentum. These two effects contribute to the behavior seen in the following figures.

In Fig. 6.5 the fractional momentum resolution $\sigma(p)/p$ for the baseline geometry is plotted as a function of p and the polar angle θ of the track at its production vertex. In order to study the performance of the straw tracking system itself, we consider only tracks which pass through the outer straw tracking planes (no forward silicon hit information is used). Furthermore, we select only those tracks which pass through at least two pixel stations, which insures an accurate track vector in the magnetic field region. We also require that tracks pass through at least three forward straw tracking stations. In Fig. 6.5 we show the effects arising from the straw materials and position resolution.

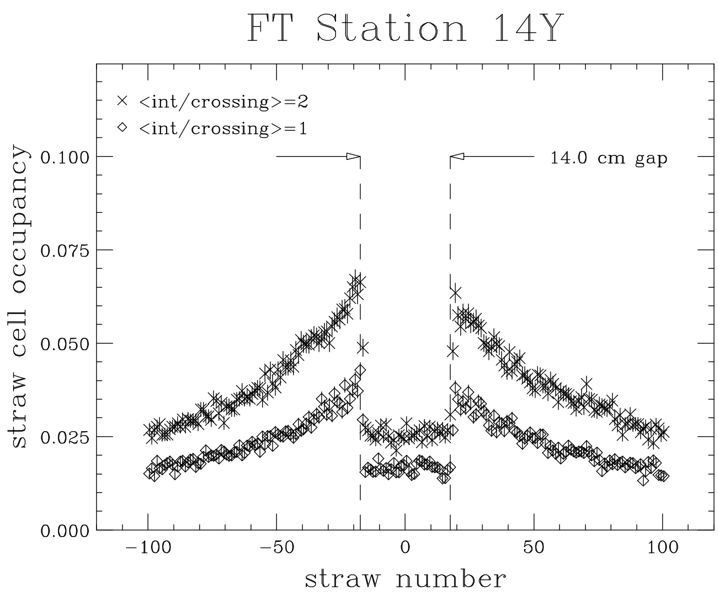


Figure 6.3: Hit occupancy for the forward straw tracking station immediately downstream of the RICH, just before the EM calorimeter. The occupancy is the average number of particles passing through a given tube in an event.

For the baseline system, the fractional momentum resolution is roughly constant (approximately 0.5%) up to approximately 16 GeV/c at which point measurement errors begin to contribute. At approximately 35 GeV/c the contributions from multiple scattering and position resolution are about equal. The curve which plots the resolution for a straw system with zero intrinsic resolution does not approach a constant value at low momentum, since in these plots no constraint on the number of measurement points has been applied.

The bottom plot in Fig. 6.5 indicates that multiple scattering clearly dominates tracks emerging at large angles, since these are mostly low momentum tracks.

6.4 Forward Straw Tracker

6.4.1 Technology Options

There are several technology options for the outer forward tracking system. The large tracking volume ($420 \times 420\ cm^2$ for the most downstream station) suggests use of conventional drift chambers which offer a low mass, high precision tracking solution for designs with small drift distances (eg. $\sim$2 mm). However, supporting the wire tensions over considerable lengths will lead to mechanical difficulties, and most likely will require additional support frame material in the decay region of the detector. There are additional problems: small drift distances over a long length can lead to electrostatic instabilities; space charge effects occur at very high rates; and broken wires are difficult to replace or retrieve and can disable

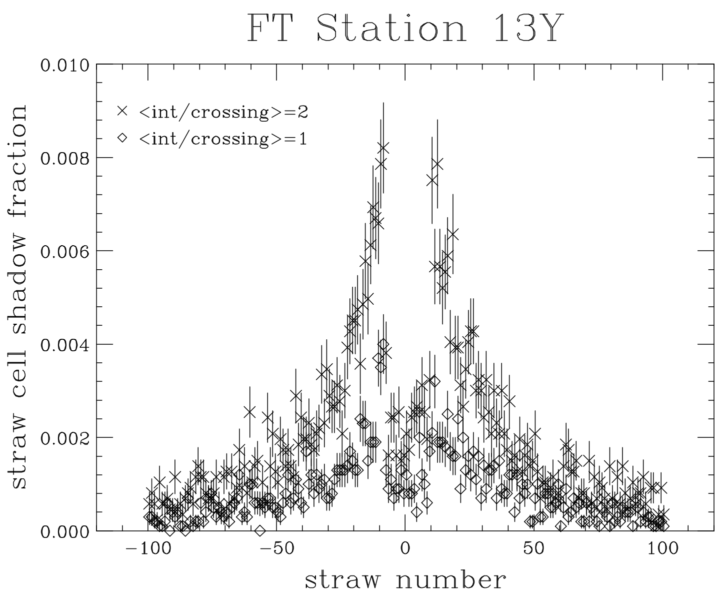


Figure 6.4: Straw cell shadow fraction for the forward straw tracking station immediately upstream of the RICH. A hit is shadowed when another track in the event passes through the same detector.

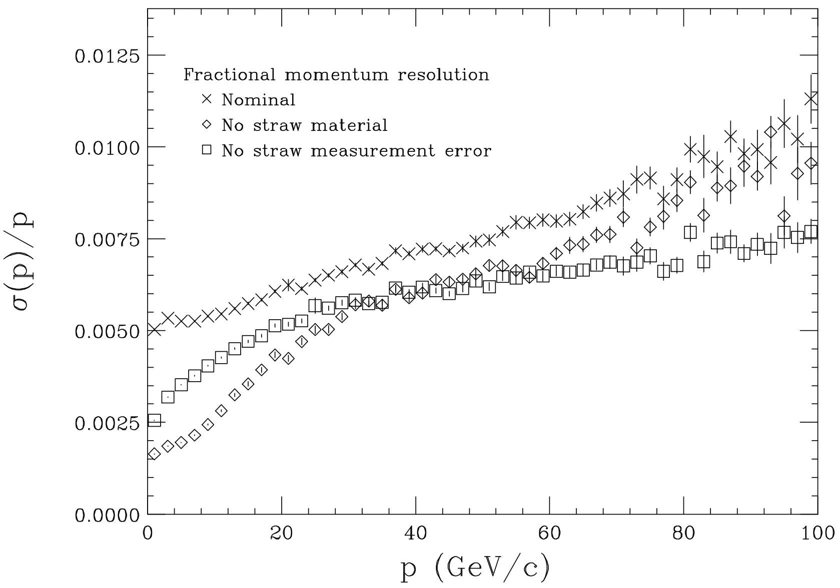
an entire chamber. These and other drawbacks of the conventional drift chamber design for high luminosity hadron collider experiments were considered in detail in Ref. [1].

Honeycomb drift chambers (HDCs) [5] can be used as a low-cost, wide area tracking detector. This technology option is being pursued by HERA-B for their outer tracking system [4]. Layers of an HDC are fabricated by folding three Kapton foils into a hexagonal structure. The result is a planar array of staggered drift cells which are mechanically stable and self-supporting. The construction methods for HDCs are considered to be simpler and more efficient compared to straw chambers, an important consideration for very large tracking systems. However, HDCs have not been thoroughly tested in high rate environments and their long term operation has yet to be proven.

Over the past several years there has been extensive development of straw tube trackers for use in high luminosity hadron collider environments [6, 7, 8, 9]. Straw tube detectors offer the following advantages: electrical and mechanical isolation between cells, ease of wire support, the ability to operate in high rate environments, overall low cost per unit volume, ease of serviceability, and construction of lightweight, low-Z materials. These features make it a suitable baseline technology choice for BTeV.

6.4.2 Straw Detecting Elements

Since we believe straw tracking chambers are the most promising technology choice for the forward tracking system, we have begun a program of research and development to determine



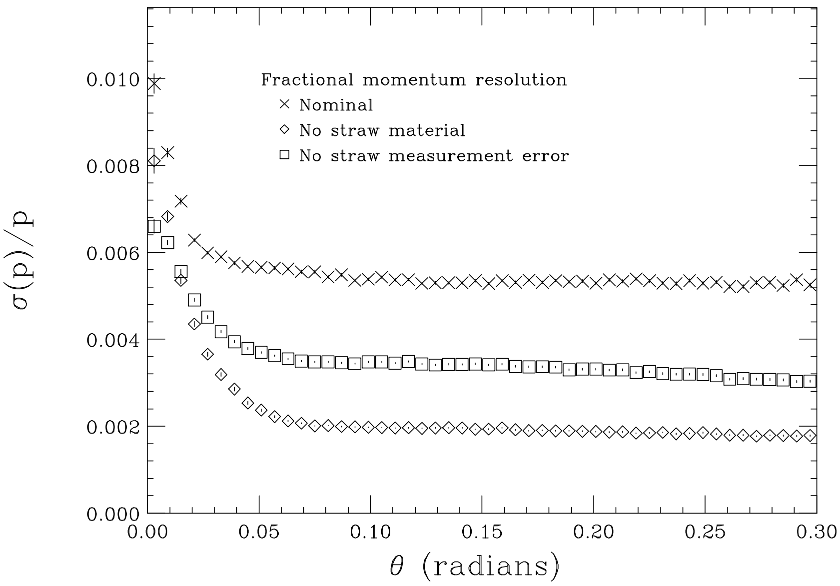


Figure 6.5: Fractional momentum resolution as a function of momentum and angle θ for tracks passing through at least two pixel stations and through the outer straw tracker (averaged over all tracks in $b\overline{b}$ events).

the basic straw drift cell properties required for stable operation in the high rate environment anticipated for BTeV. We plan to develop straw cells capable of achieving high drift time measurement efficiency and drift time accuracy while maintaining low noise counting rates. Where possible we will build on the significant body of research carried out for the SDC straw drift tubes [6] and the ATLAS TRT straw tubes [9].

6.4.2.1 Straw Tube Design

There are several design choices which will need to be specified for BTeV straws. The first is the choice of straw tube diameter and materials. The SDC and ATLAS straw trackers both employ 4mm diameter tubes. Smaller tube diameters reduce the occupancy but there is a limit to practical operation at very small diameters. The material will most likely be a Kapton film with copper or aluminum conduction layers, and possibly protective carbon-polyimide layers. The manufacturing process is rather involved depending on the rigidity requirements and choice of reinforcement scheme. The SDC straws were manufactured by an industrial firm[10] which also applied the conduction layers (copper in this case). The straw manufacture process for the ATLAS TRT is more involved, as carbon fiber threads are applied at CERN for additional mechanical support. This reflects a difference in design philosophy: the ATLAS straw tubes, to some extent, provide some of the tension relief (in addition to the carbon fiber reinforced plastic shell and alignment planes) while for the SDC straw tracker the shell and endplates carried the entire wire tension and the close-packed straws were subject to no external mechanical stresses.

6.4.2.2 Materials

The choice of specific straw materials and accompanying support structure materials has significant implications for possible background counting rates in the straw tubes. The neutron flux in particular needs to be estimated for the chosen BTeV forward system layout.

We are developing a prototype straw tube which places an aluminum conduction layer between two Kapton films, the inner one next to the gas volume being a carbon loaded, low resistivity film (see Fig. 6.6). The idea is that the Kapton forms a protective barrier, similar to the graphite layers deposited on the inner surface of the ATLAS TRT straws. We measured the surface resistivity of the aluminum coated, carbon loaded Kapton film of our prototype to be 6.5 ± 1.0 Ω/square, which is comparable to the specified value for the TRT straw tube. The details of the prototype straw material are listed in Table 6.3. The 1 mil thickness of each film is chosen as a compromise between the 0.5 mil thick films used for SDC straw tubes (which had very little mechanical rigidity) and the thicker and reinforced TRT straws. Our final design will likely have the straws supporting some of their own weight. If this turns out to not to be the case, we may use instead 0.5 mil Kapton films which have been shown to work and would provide a reduction of material in the detector volume.

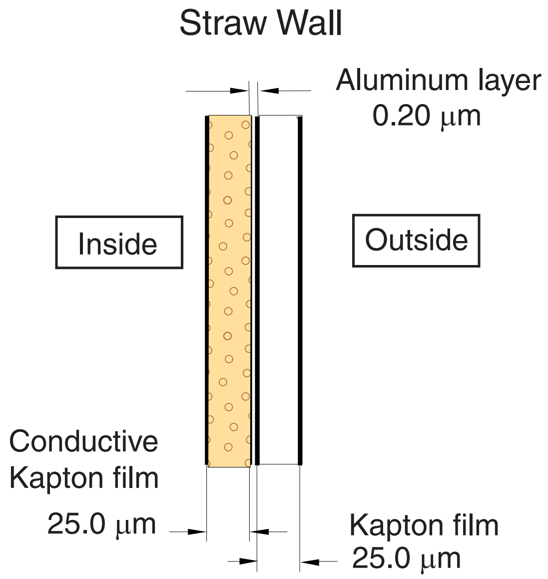


Figure 6.6: Cross section sketch of the straw wall for the prototype BTeV straw tube. The straw is manufactured by winding around a mandrel two 1 cm wide Kapton strips offset by 5mm.

6.4.2.3 Gas Composition

We are considering several possibilities for the optimal BTeV gas mixture. The SDC straw design used 80% CF_4 and 20% isobutane, a mixture chosen for its fast drift velocity ($\approx 100\mu m/ns$) and well known operating characteristics in high radiation rate applications. This very fast drift velocity was targeted for the 16 ns SSC bunch crossing time. The ATLAS TRT gas mixture (70% Xe + 20% CF_4 + 10% CO_2) contains xenon, which is a good x-ray absorber (required for a transition radiation detector), CF_4, to insure high drift velocity, and CO_2 as a quencher. For BTeV straw cells we plan to determine the gas mixture which optimizes drift time resolution, efficiency, and stability of operation. One possibility is to use a TRT-like mixture in which Xe is replaced by Argon: 70% Ar + 20% CF_4 + 10% CO_2. We have observed the analog signal properties of this mixture and found them very similar to the xenon based mixture. Another candidate is Argon-Ethane-CF_4, 50:35:15, used for the Central Outer Tracker (COT) by CDF. This gas has a fast drift time of 88 $\mu m/ns$ when their chamber is operated at 2.5 kV/cm. This configuration gives the COT a 100 ns maximum drift time, suitable for 132 ns bunch crossings. Since our cell size is much smaller than that of the COT, we are not forced to use a fast gas. Mixtures with little or no CF_4 will also be considered.

Closely associated with the choice of gas mixture will be the determination of optimal gas gain at which to operate the detector. Higher values of gas gain are known to introduce space-charge effects which can lead to significant non-linearities in the straw response. The probability that self-limiting streamer discharges occur rises quickly with gas gain. This

Description	BTeV Straw Prototype
Kapton film	Inner: Polyimide type XC $25 \pm 2.5\ \mu m$ thickness Outer: Polyimide type 100 VN $25 \pm 2.5\ \mu m$ thickness
Density	1.42 g/cm^3
Aluminum layer	$(0.2 \pm 0.08)\mu m$ thickness
Resistivity of inner Kapton layer	$6.5 \pm 1.0\ \Omega$/square

Table 6.3: Summary of straw tube material specifications for the BTeV prototype straw tube

in turn leads to an overloading of the front-end electronics and introduces deadtime in the amplifier/shaper/discriminator circuitry.

Perhaps the most important factor for robust operation is mechanical integrity of the straw cell itself. The quality of the anode and cathode surfaces must be high and the position of the anode wire within the tube must be held to strict tolerances. Thus an important part of our program will be to develop reliable construction techniques and quality assurance procedures.

The high voltage range at which the straws may operate without inducing large amounts of dark current will be measured for the selected straw design. At high voltages the streamer discharge rate increases until most of the signal from the straw is from such streamers. Corona discharges occur even without significant wire offsets. This effect is compounded for straw tubes with large wire offsets. Experience from the ATLAS TRT straw tubes shows these offsets must be kept below 300μm to prevent such breakdowns if one expects to be able to vary the voltage over a reasonable range of operating gains. These types of studies will need to be performed for BTeV straws and candidate gas mixtures.

6.4.2.4 Signal Wire, Joints and Supports

The sense wire for the straw cells is 30 micron diameter gold plated tungsten held under 50 grams of tension. In addition, following the technique used for the ATLAS TRT, most of the sense wires will be divided electrically using a small glass capillary bead [9]. This cuts the occupancy rates in half.

The sense wires in straws which span more than 80 cm will require additional support. Fig. 6.7 shows the helical design developed for ATLAS TRT which can be adopted for BTeV straws.

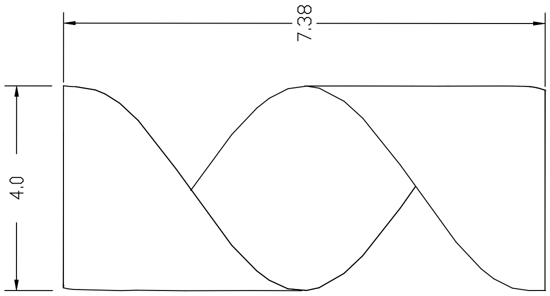


Figure 6.7: Twister wire support used in ATLAS TRT [9]. Dimensions shown are mm.

6.4.3 Construction of Straw Chamber Modules

6.4.3.1 Mechanical Support Considerations

The module concept for straw drift tube arrays was developed for the SDC straw tracker. Rather than constructing large monolithic arrays of several thousand drift tubes, one assembles groups of layers in packages consisting of a few hundred tubes. This method lends itself well to mass production techniques, quality assurance, and ease of repairability. Fig. 6.8 shows a sketch for one of six modules comprising the station nearest the interaction region. The module positions 3 layers of close packed straw tubes via an outer aluminum frame, which provides the gas manifold and wire tension support. Two such modules fit together to comprise a measurement view. The straw tube services are provided at the two ends by the frame, which forms a gas manifold, holds signal wires, interfaces with the front-end boards, and is the attachment and alignment fixture for the module. Fig. 6.9 shows how these modules may be combined to form a tracking station.

Additional services may be required to remove the heat generated in the straws as positive ions flow to the cathode. Straws near the beam pipe will have high counting rates and thus may have significant heat to dissipate. Temperature variations will lead to non-uniformity of the straw tube response and so must be kept to a minimum (yet to be specified). Calculations of the power dissipation rates as a function of straw position will need to be made once the overall design is specified.

6.4.4 Front-end Electronics

The straw tubes require a relatively high gain amplifier, to allow high efficiency operation at moderate gas gain. Current is induced in a proportional chamber anode wire by the motion of positive ions away from the wire. The fall time of this signal is quite slow (relative to the Tevatron crossing time of 132 ns) and must be shaped using a network designed with the appropriate ion velocity in mind. The University of Pennsylvania ASD series of integrated circuits [11] was developed initially for the SDC straw chambers, and more recently for the ATLAS TRT. These ASICs include high gain preamplifiers, pole-zero networks for pulse

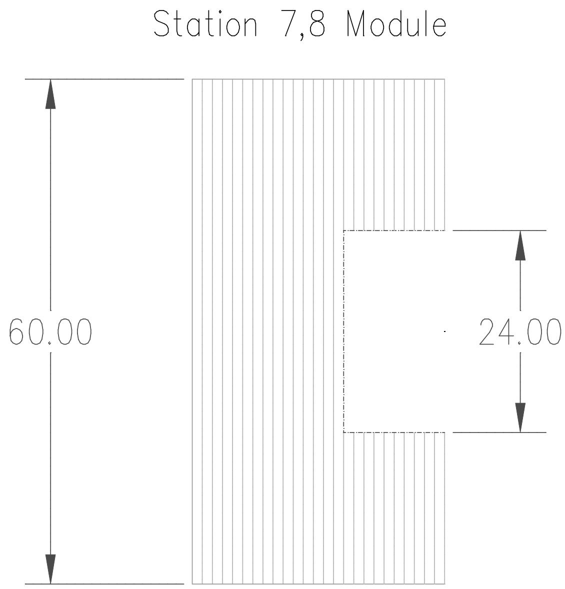


Figure 6.8: View of a straw module for the forward tracking station nearest the interaction region. Dimensions shown in cm.

shaping and ion-tail cancellation, and leading edge discriminators. These ASICs are radiation hard, and are almost exactly suited to our use. We expect to use one of the ASD ASICs for the BTeV straw chamber system.

6.4.5 TDC Requirements and DAQ Readout

The baseline design requires a position resolution less than 100 μm per measurement station (~ 150 μm per plane). The error introduced by the granularity of the TDC bins should be small compared to the intrinsic straw tube resolution. This intrinsic error is a function of drift distance as well as other factors. An RMS error of 150 μm is achieved by having a much larger error for tracks passing very close to the wire, and a significantly smaller error (~ 100 μm) for tracks passing at an optimal distance from the wire. The TDC binning error should thus be much less than 100 μm. If we operate the straw tube with a gas mixture and high voltage setting which results in a 100 μm/ns drift velocity, then the TDC bin width of 1.5 ns (corresponding to a binning error of 44 μm) would suffice. This is not a difficult specification. We have begun discussions with the Fermilab PPD Electrical Support group about the design of a purely digital TDC to meet our needs.

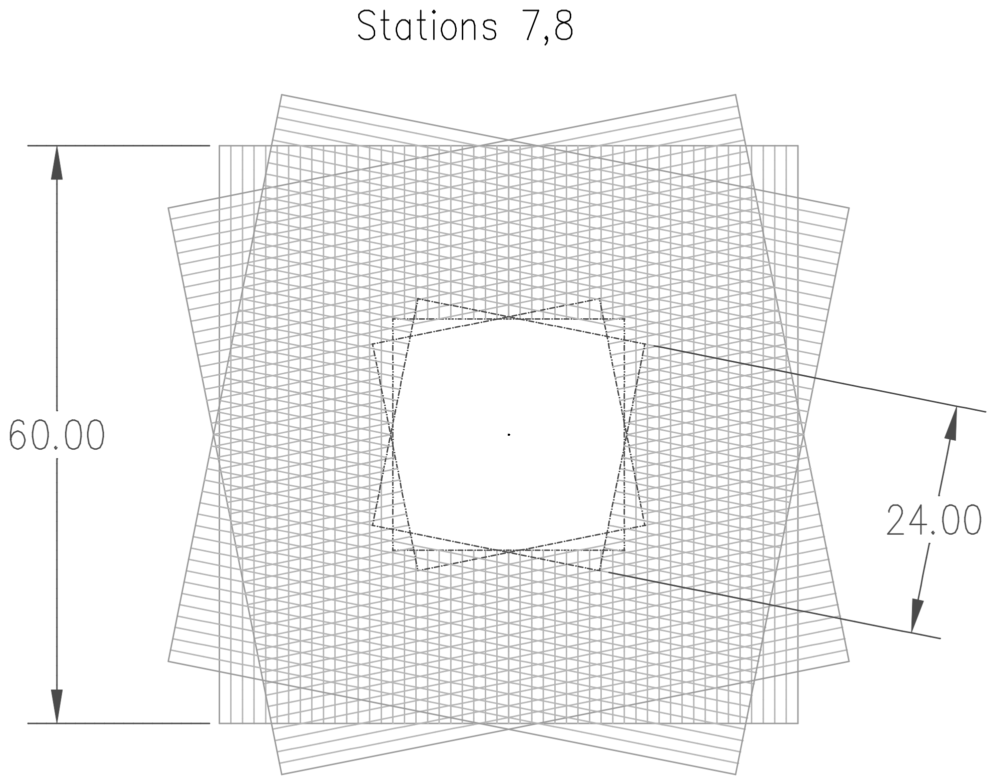


Figure 6.9: Front view of the forward straw tracking station nearest the interaction region. Dimensions shown in cm. The station consists of three measurement views built from six independently built and serviced modules. The station shown, which is a proposed change to the baseline design, consists of an X view and two ± 11.3 degree stereo views. Note that the stereo views mostly measure the Y, bend view coordinate.

6.5 Forward Silicon Tracker

6.5.1 Technology Options

Microstrip Gas Chambers[2] (MSGC's) are being pursued by the LHC-B[3] and HERA-B[4] Collaborations for their inner region tracking detectors. These devices offer good granularity, low occupancy, resolutions better than 80 μm, and high rate capability. However the boro-silicate glass substrates, on which the MSGC's are based, introduce considerably more radiation length per measurement plane than either conventional drift chambers or straw tube cells. Cost is a major factor for large area coverage as well. These considerations make MSGC's an unattractive option for the outer regions of the forward tracking system where large areas need to be instrumented while keeping the radiation lengths to a minimum to maintain good resolution for wide angle, low momentum tracks. MSGC's can be considered for the the inner region in close proximity to the beam pipe where the occupancy rate is high and the tracks tend to be of higher momenta. However, MSGC's are not yet a proven technology.

Silicon microstrip detectors are a well-proven technology. They offer excellent resolution and low occupancy. Additionally they can be operated in high radiation dose environments. The thickness can be kept to a minimum using double-sided silicon sensors. A ladder arrangement of silicon detector modules can provide relatively large area coverage.

6.5.2 Detector Description and Layout

The baseline proposal consists of two planes of 300 μm thick double-sided silicon microstrip detectors with 100 μm pitch. Standard sized silicon wafer modules are mounted on low mass carbon fiber support pads and are arranged in a daisy-chained ladder readout configuration as shown in Fig. 6.10. Four views are provided: X, Y, U, V where the U and V stereo angles are matched to the corresponding straw tracker stereo views. Each layer consists of 16 silicon strip modules measuring 6 cm by 4 cm and four measuring 2.5 cm by 4.75 cm. The configuration shown here is similar to that proposed by LHC-B [13].

Each view consists of 2400 channels and the whole system has 57,600 channels. With 100 μm pitch we need only readout the hits since the hit alone provides $\sigma = 100\ \mu m/\sqrt{12} = 29\ \mu m$ resolution. A new binary readout ASIC, designed for high speed readout, will be required.

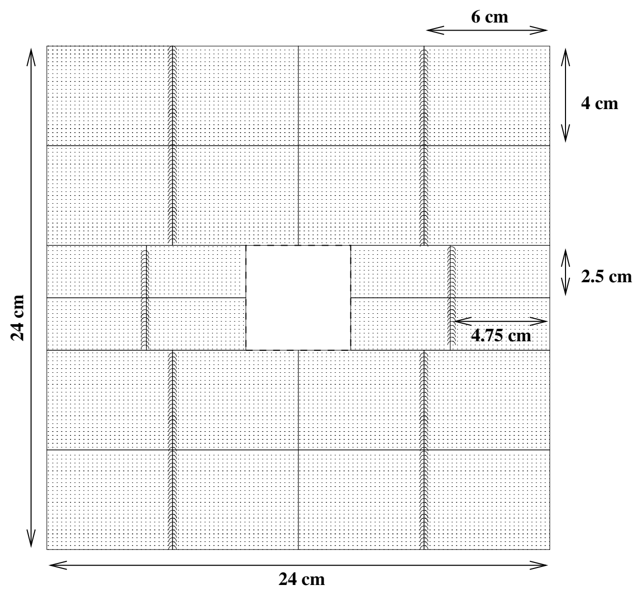


Figure 6.10: Proposed schematic of a forward silicon tracking layer.

Bibliography

[1] D.G. Cassel, G.G. Hanson *et al.*, "Report of the Central Tracking Group", Proceedings of the 1986 Summer Study of the Physics of the Superconducting Supercollider, eds. R. Donaldson and J. Marx, Snowmass, CO (1986) 377.

[2] A. Oed, Nucl. Instrum. Meth. A 263 (1988) 351.

[3] LHC-B Letter of Intent, CERN/LHCC 95-5, August 1995.

[4] HERA-B Design Report, DESY-PRC 95/01, January 1995.

[5] H. van der Graaf, *et al.*, Nucl. Instrum. Meth. A 307 (1991) 220.

[6] Y. Arai, *et al.*, "A modular straw drift tube tracking system for the Solenoidal Detector Collaboration experiment, Part I: Design"; Nucl. Instrum. Meth. A 381 (1996) 355.

[7] Y. Arai, *et al.*, "A modular straw drift tube tracking system for the Solenoidal Detector Collaboration experiment, Part II: Performance"; Nucl. Instrum. Meth. A 381 (1996) 372.

[8] RD6 Collaboration, "The ATLAS TRT straw proportional tubes: performance at very high counting rate", Nucl. Instrum. Meth. A 367 (1995) 143.

[9] ATLAS Inner Detector Technical Design Report, CERN/LHCC/97-16,17.

[10] Lamina Dielectrics Ltd., Billingshurst, W. Sussex, UK.

[11] F.M. Newcomer, R. Van Berg, J. Van der Spiegel and H.H. Williams, Nucl. Instrum. Meth. A 283 (1989) 806.

[12] Y. Arai and M. Ikeno, IEEE J. of Solid-State Circuits SSC-27 (1992) 359.

[13] LHC-B Technical Proposal.

Chapter 7

Design of the Muon Detector

As described earlier, the BTeV muon system should have the following attributes:

1. Robust: It will be very difficult to access the detectors once they are installed.
2. High Rate Capability: Rates of a few kHz/cm^2 are possible in parts of the system (based on calculations of accelerator backgrounds).
3. Good Time Resolution: Timing information on hits with a resolution of 50–60 ns will allow rejection of hits due to accelerator backgrounds.
4. Magnetic: Two toroids of 1.5 Tesla field and roughly 1 m thickness (in each arm).
5. Fine-Grained Spatial Resolution: Tracking resolutions of 2–3 mm are required.
6. Longitudinal Compactness: Given the size constraints, the detectors should be of minimal thickness.
7. Minimal Neutron Susceptibility: We will avoid the use of hydrogenic detector elements.
8. Uniform Occupancy: A geometry designed to minimize pattern recognition confusion and minimize occupancy at smaller radii.

7.1 Detector Design and Fabrication

In this section we describe the detector system designed to satisfy these criteria.

7.1.1 Basic Building Block: Proportional Tube "Plank"

The basic building block in the construction of a detector station is a "plank" of 3/8" stainless steel proportional tubes (see Fig. 7.1). Thirty-two tubes are arranged in a double layer with an offset of half a tube ("picket fence" geometry) and are glued together at two or three points by aluminum ribs ("Fred Ribs" in the figure). This should provide an extremely

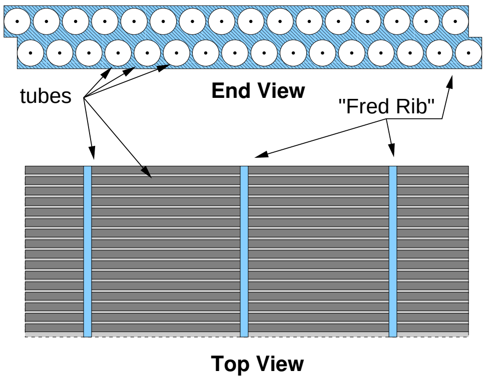


Figure 7.1: End and top views of one "plank" of proportional tubes. The plank is the basic building block of the muon system.

sturdy, self-supporting building block. Proportional tubes have been selected as the detector technology because they are robust and have the necessary rate capability. We intend to use a fast gas (such as 88% Ar, 10% CF_4, 2% CO_2, which has a drift velocity of roughly 9 cm/μs [1]) so the maximum collection time (drift plus charge integration) for a signal should be less than 60 ns. We intend to use thin walled (0.01") stainless steel tubes. This is a proven technology — it has been used successfully by the CDF collaboration where similar tubes [2] had a long lifetime with low failure rate.

The 0.5 cm wire spacing of this design has an effective spatial resolution of at most 1.5 mm, with no dead regions between tubes.

7.1.2 Design of a Detector Station

To minimize occupancy at small radii and to minimize pattern recognition confusion, each detector station will consist of eight overlapping pie shaped "octants," as shown in Fig. 7.2. Each octant is a single unit, which will be hung from a hanging frame. There are three views (r, u, and v) in each octant as shown in Fig. 7.3. The views stack on top of each other, and are built from the planks described above. We are considering adding a thin layer of aluminum between each view to act as a delta ray "shield." While the shield itself will generate delta rays, the idea is to regenerate them near the tubes so that they will tend to fire the same tubes as their parent muons. There will be 15 planks (butted up against each other) in each view of an octant (more than shown in Fig. 7.3).

Currently the angle of the u and v views with respect to the r view is $\pm 22.5^o$. The three views provide redundancy and allow studies to be made in which two views are beat against the third (such as in efficiency studies). The angle of 22.5^o is not yet set in stone, further Monte Carlo studies need to be performed. This geometry gives excellent resolution and redundancy in the bend direction ($\hat{\theta}$) of the toroids, and the resolution in the $\hat{\phi}$ direction is reasonable ($\sigma_\phi = 2.8$ mm, which is approximately twice the single plank resolution).

This octant geometry has several advantages. In all three views the planks are short near

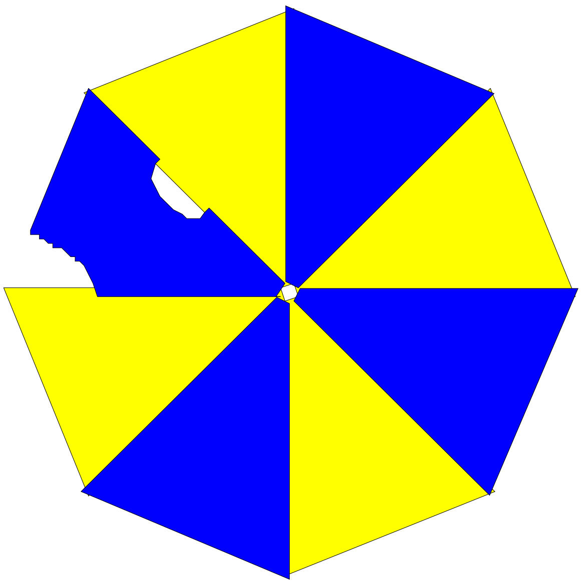


Figure 7.2: Beams-eye view of one muon detector station, which consists of eight overlapping octants arranged in two layers. One octant is cut away in places to show the overlap between adjacent octants.

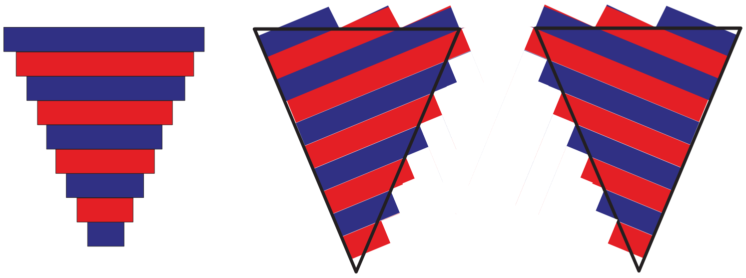


Figure 7.3: Arrangement of planks to form each of the three views in an octant. There will be 15 planks per octant (more than shown).

the beam pipe, and get longer at larger radii. This should result in approximately uniform occupancy as a function of r. This geometry should also help minimize pattern recognition confusion.

With 15 planks per view per octant, and 32 channels in a plank, the channel count for each octant will be 1440 and for each detector station will be 11,520, for a total of 34,560 channels in each arm and 69,120 channels in both arms.

7.1.3 Proportional Tube Details

Each of the 32 tubes in a plank will share a common gas manifold at each end. The wires will be held at positive high voltage and read out through a blocking capacitor by front-end electronics which will attach at one end. High voltage will be fed to the plank at the same end as the electronics. We intend to use a "fast" gas mixture of 88% Ar, 10% CF_4, 2% CO_2 or something similar. For more details on the front-end electronics, expected signals, etc., see the section on front-end electronics below.

The longest tubes in the system will be approximately 6 feet in length, and we are considering several different wire diameters (20–50 microns). We will decide on a wire diameter based on the results of beam tests to be performed this summer (1999, see description below). In any case, our calculations indicate that no wire supports are necessary. Again, we hope to get some feedback on that this summer.

The tubes will be strung individually and the wires held in place via a crimp pin. The arrangement is very similar to that used in the CDF system [2]. The crimp pins will extend out of the gas manifold and will attach to a circuit board which contains the blocking capacitors and an edge connector for the front-end cards.

7.1.4 Miscellaneous Mechanical Details

Because of the overlapping octant geometry of each detector station, there is plenty of room for cabling and front-end electronics everywhere except near the beam pipe. It is contention for space at small radii that limits the inner radius of the system to between 15 and 20 cm.

Weight is a potential concern because we are using stainless tubes. Using thin walled tubes reduces this concern significantly. We estimate the weight of the stainless tubes in one octant (all three views) to be 200 lbs, the same as if we were to use 1/32" thick aluminum tubes. However, the hanging frame used to support the octants and the delta ray shields will probably double the weight of each octant, so that each detector station will weigh 1.5–2 tons.

We plan on building at least 5 spare octants (10% of those in the total system), so that if one develops problems (more than a handful of broken wires, etc.) we can swap a new octant in and work on the problem octant "offline." We do not plan on restringing bad tubes – we will simply make a new plank.

The octants will be fabricated at Illinois (50%) and Vanderbilt (50%). We estimate the cost of the stainless tubing to be \$500,000. Wire will cost \$250,000. Machined parts such as the gas manifolds, gas connectors, wire crimp pins and feedthroughs will cost \$500,000. Including the hanging frame and contingency, we estimate the total cost of materials for the detector planes to be \$1.5 million, not including electronics.

7.1.5 Summary of Mechanical Design

A summary of the baseline BTeV Muon System is given in Table 7.1.

Table 7.1: Parameters of the baseline BTeV Muon System.

Transverse size	$\pm$ 200 mrads
Total Length	3.3 m
Rate Capability Req'd.	few kHz/cm^2
Steel Length	2 m
Toroidal Field	1.5 T
No. of Longitudinal segments (steel)	2
Plane size:	20 m^2
Tube cell size	1 cm (diameter)
Wire spacing:	0.5 cm
Spatial resolution	1.5 mm
No. of channels/octant (3 views)	1440
Octants per station	8
Total channels	35,000 per arm
Momentum resolution	$\sigma_p/p = 19\% \oplus 0.6\% \times p$
Total thickness/station	20 cm

7.2 Front End Electronics

In order to capture all the charge coming from a typical proportional tube for a single incident ionizing particle, one must integrate the charge due to the positive ion drift over a long period of time relative to the expected beam crossing for the BTeV experiment. The time needed to collect a portion of the total charge is a function of the gas used in the proportional tube, the geometry of the tube, and the high voltage bias of the center wire, relative to the outer conductor:

$$\frac{Q}{Q_o} = \frac{\ln(1 + t/t_o)}{2\ln(b/a)} \; [3], [4]$$

where Q is the charge collected in an amount of time, t, Q_o is the total charge from the ionization (depends on the gain of the tube), b and a are the outer and inner radius, respectively, of the conductors in the proportional tube, and:

$$t_o \sim \frac{a}{2v_d} \; [4]$$

where v_d is the drift velocity of the gas for the chosen value of the gain in the tube [1]. Alternatively, for a very fast signal, one could try to just capture the fast signal coming from the electron amplification, but this is only 1–2% of Q_o.

In BTeV we intend to use the muon proportional tubes in the Level 1 Trigger. The natural cutoff for integrating a signal then is on the order of a beam crossing, or 100 ns. However, we would like to reduce backgrounds associated with the collider beams, which pass through about 60–70 ns before interaction products arrive. For a fast gas like $ArCO_2CH_4$ [1], and a wire-to-wire spacing of 0.5 cm, the maximum drift time before the first signals from a track arrive is about 30 ns. This means that if we would like to be able to separate the various "crossings", our electronics should be able to resolve 2 hits occurring 30 ns apart (minimum).

7.2.1 Amplifier choice

Fortunately, electronics to amplify and digitize the analog signal from a chamber hit in a high rate environment were developed for the SSC [3], and continue to be refined for use in detectors for the LHC [5] and at Fermilab [6]. We will take full advantage of these developments.

We intend to mimic the setup used for the CDF COT [6]. There will be a PC board to deliver High Voltage to each Proportional Tube, and a PC board with electronics to amplify and digitize the tube signal. Both boards will be located directly at the end of a group of 16–32 proportional tubes.

We plan to utilize the ASD family of integrated circuits developed at the University of Pennsylvania to amplify and digitize the signals coming from the proportional tubes. These chips (and applications) are described in detail elsewhere [3],[6], but we briefly mention some more important features here.

The ASD8B, which is being used for evaluation and beam tests, amplifies the first $\sim$6 ns of the the chamber signal and outputs an LVDS (equivalent) differential logic signal. This chip, when mounted on a PC board, has a low effective threshold of about $\sim$1 fC. This is the threshold as quoted in the ASD8B reference and verified in tests that we performed on an evaluation board supplied by the University of Pennsylvania in December of 1998. The chip also features a double pulse resolution of $\sim$20 ns. This chip seems to be an ideal choice for the BTeV muon system, and it is important to determine if we have sufficient signal to utilize the ASD8B (or other ASD chip!).

7.2.1.1 Practical Considerations

To calculate the signal we expect in the proportional tube, we will use the data presented in reference [1]. For the same gain as quoted in the paper($\sim 1.8 \times 10^5$), we can use the quoted results to determine our operating parameters for different tube geometries and gas mixtures. In Table 7.2.1.1, we summarize the results for different gases and inner conductor radii.

In the table, we have made some assumptions. We've assumed that the most important factor in calculating the gain is sensitivity to a *single* ionization cluster of a *single* electron–ion pair; we haven't taken into account the effect of adding a termination resistor (see below); we haven't taken into account the effect of a magnetic field; and we've neglected aging effects in the tube. We've also taken for granted that the electronics will be located at the end of the proportional tubes: lowering the electronic noise and increasing the amount of signal we get in the small integration time available. Even so, we can learn a great deal from looking at the table.

Table 7.2: Operating parameters for different gases and inner wire radii (a). The values of t^M_{drift} assume a wire space of 0.5 cm.

Gas	$E_{field}\|_a$	v_d	V_{bias}	$Q_{6\mathrm{ns}}$	t^M_{drift}	a
$\frac{ArC_2H_6}{50/50}$	$150(\frac{\mathrm{kV}}{\mathrm{cm}})$	$4.8(\frac{\mathrm{cm}}{\mu\mathrm{s}})$	3545(V)	7.8(fC)	50(ns)	50(μm)
$\frac{ArCO_2}{80/20}$	164	2.4	3768	5.9	100	50
$\frac{ArCO_2CF_4}{88/2/10}$	150	8.7	3454	9.5	30	50
$\frac{ArC_2H_6}{50/50}$	150	4.8	1656	8.8	50	20
$\frac{ArCO_2}{80/20}$	164	2.4	1806	7.1	100	20
$\frac{ArCO_2CF_4}{88/2/10}$	150	8.7	1656	10.4	30	20

7.2.1.2 Amplifier Observations

A smaller wire size is desirable from an electronics standpoint: the operating voltage (V_{bias}) is lower and the integrated charge collected ($Q_{6\mathrm{ns}}$) is a little higher. The lower operating voltage gives us more flexibility in the choice of a decoupling capacitor, allowing the option of surface mount components in the high voltage system. The higher integrated charge means we can safely raise the threshold of the electronics without losing signal, allowing greater flexibility in our noise budget. Additionally, it is expected that the gain of a tube has less sensitivity to the operating voltage with a smaller wire [1].

Also evident is the advantage of having a fast gas. Not only can one integrate more charge in a smaller period of time, one has less sensitivity to beam "crossings". We can achieve our goal, stated previously, of the first signal to arrive after a maximum time of about $t^M_{\mathrm{drift}} = 30$ ns.

Another way to look at the table is to reason that all the combinations calculated will produce a working detector, and that using the ASD chip allows us a lot of flexibility in our choice of wire diameter, gas, and operating voltage.

At most, a termination resistor will raise our noise by 1 fC and lower our gain by a factor of 2. This is to be offset by the expectation that about 10 primary ion-pair clusters will be produced, on average, for a MIP passing through 1/3 cm of gas volume.

One final consideration in evaluating the ASD8B for BTeV is that the process used to make the chip is obsolete and the chip will not be available in production quantities [7]. We will have to use an improved version (ASDQ) at a slightly higher cost. It integrates for a slightly longer period, has the same double pulse resolution, and has a higher noise level of $2-3$ fC.

7.2.1.3 Front–end Readout

As mentioned previously, the ASD chips produce differential logic signals (LVDS) suitable for driving long (10 m) cable lengths. The LVDS signal consists of 1 bit/channel signifying whether or not a channel is on or off. If we choose to do any logic on the ASD board, we must first convert the LVDS, perform the required function, and finally redrive the signal. This option, though decreasing cable cost, will require more power to be delivered to the front-end, more power to be dissipated in a small area near the detector, and will increase the amount of noise on the board. This option was not used on CDF and is not recommended by the designer of the ASD8B chip [7].

Finally, reading out the raw hits provides for maximum flexibility in designing, prototyping or even upgrading and good utility in re-using these electronics or interchanging electronics with other BTeV detectors. Cards designed with a similar philosophy have been in continual use for 25 years in high energy physics experiments. For the reasons given, we intend to read out the raw hits from the front-end electronics.

7.2.1.4 Feasibility

By utilizing existing electronics packages and packaging for the front-end of the BTeV muon system we will save considerable time and developmental costs. For instance, the major modification needed to utilize the 24 channel CDF COT front-end card for BTeV, is to remove one of the chips from the CDF COT circuit board to accommodate the smallest muon drift tube channel count footprint of 16 channels.

Additionally, such a scheme is easily utilized by other detectors, such as the straw tracker, and we can benefit from standardization. To use this scheme and front-end board in other BTeV gas tracking systems, one has to consider what information is needed for tracking, and build a suitable post front-end circuit board to suit those needs. The front-end can be used right away.

7.2.2 Post Front-end Considerations for the Muon System

As stated, we intend to do no post processing on the front-end amplifier board. We still need to read the signals out, and we would like to keep costs to a minimum.

In order for the front-end to operate correctly, we will need to provide a mechanism to modify the threshold for a channel. Rather than include a digital to analog converter on the front-end card, we would rather provide a single analog threshold for each chip (8 channels) and provide a digital mask for channels that fire continuously.

To reduce cable cost and logistics for long cable runs, we will convert the parallel bitmap coming from the front-end to a serial signal that can be transmitted on a single fiber optic cable. A single optical cable will read out 96 channels and do so in the 132 ns between beam crossings. (The hard limit is 132 bits for a 1 Gbit/s optical link.)

Converting the data from parallel to serial destroys the fine timing information of the front-end signal but allows us to apply a gate to the signals coming from the front-end. By applying this gate we provide both the means to compress the bitmap information in time for the conversion and the ability to filter out-of-time hits from the information sent to the trigger and the DAQ.

All these actions are to be done on a circuit board that intercepts 96 channels of input

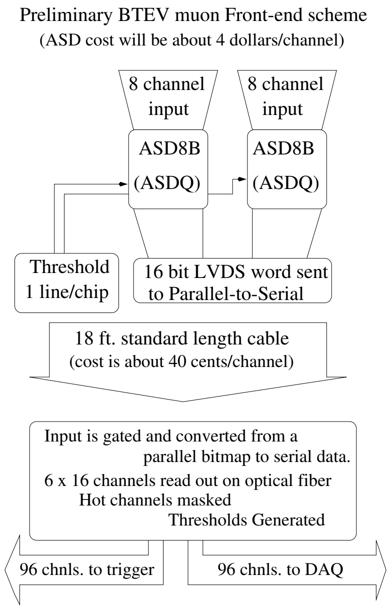


Figure 7.4: Schematic diagram of the front-end and post front-end circuit boards for the BTeV Muon Proportional Tube system. Some highly preliminary cost estimates are included in the figure.

from 6 frond-end cards, and converts the signals for later use. Schematically, the front-end and post front-end card are shown in Fig. 7.4.

7.3 Trigger Electronics

7.4 Test Beam Plans and Goals

During the 1999 test beam running we will investigate the stability, reliability and operating characteristics of our proportional tubes with various tube lengths (1,3,6 feet) and with various wire diameters (20, 30, 50 micron). We will analyze the response of the system at different incident beam angles and will record rate capability, noise, efficiency and resolution. We also will test the performance of prototypes of the front-end electronics and a prototype trigger processor.

We will construct a tracking system consisting of ten planks of proportional tubes. The planks are identical in design to the ones described above: 32 tubes arranged in two, offset layers. The tubes will be 0.01” thick walled stainless, with an outer diameter of 3/8”. We will have enough front-end boards to read out half (160) of these channels.

Our Run Plan is:

1. One week to install planks, trigger scintillators and DAQ. The planks will have been

tested in a cosmic ray test stand at Vanderbilt prior to installation. The DAQ software will have been tested with partial electronics prior to installation.

2. One week of running with Beam #1 (see below) to bring the system up and adjust timing in electronics, etc.

3. One to two weeks of running with Beam #1 to take main data sample. We require several 1-3 hour runs with access time between, to rotate a station or to move other portion of a plank into the beam.

4. Two days of data taking with Beam #2 and Beam #3.

5. After taking a few days to analyze the data and reach conclusions, we may go back and repeat some or all of the above. A third iteration could be necessary. Up to 4 weeks of data taking has been requested.

Beam Requirements:

1. A 1 kHz pion beam with energy of 100 GeV or higher and with a beam spot diameter of 4–8 cm.

2. A beam consisting primarily of muons. Provided that the rate is at least 100 Hz with a beam spot area of 15×15 cm^2, either a beam derived from hadrons hitting a dump or a halo muon beam will be adequate. Muon purity around 99% is desirable. It would be ideal if this could be achieved by moving the stations downstream of the Mtest dump during this part of the test.

3. A 10–100 kHz pion beam for rate studies. This might require a somewhat higher allocation of protons to Mtest area than normal for a small period of time. Depending on the situation, we may require a special high intensity run of 10^{12} protons per pulse for 1 day.

Bibliography

[1] J.M. Butler *et al.*, "Study Of Fast Gases, Resolutions And Contaminants In The D0 Muon System," Nucl. Instrum. Meth. **A290**, 122 (1990).

[2] S. Bhadra, S. Errede, L. Fishback, H. Keutelian and P. Schlabach, "The Design And Construction Of The CDF Central Drift Tube Array," Nucl. Instr. Meth. **A268**, 92 (1988).

[3] "A Fast, Low Power, Amplifier-Shaper-Discriminator for High Rate Straw Tracking Systems," F.M. Newcomer *et al.*, Trans. Nuclear Science, Vol 40, N4, P630-636, August, 1993.

[4] F. Sauli in *EXPERIMENTAL TECHNIQUES IN HIGH-ENERGY NUCLEAR AND PARTICLE PHYSICS (2nd ed.)*, edited by T. Ferbel, World Scientific, Singapore 1991, p. 79.

[5] ATLAS Collaboration,"ATLAS inner detector: Technical design report. Vol. 2," CERN-LHCC-97-17.

[6] R. Blair *et al.* [CDF-II Collaboration], "The CDF-II detector: Technical design report," FERMILAB-PUB-96-390-E.

[7] M. Newcomer, *Private Communication.*

Chapter 8

The BTeV Trigger Performance

8.1 Introduction

The proposed broad charm and beauty physics program of BTeV/C0 calls for a trigger that is very efficient for heavy-quark decay events which can be found off-line and is relatively independent of decay mode.

The baseline trigger is a multi-level trigger with several "branches" that can be independently prescaled and then combined into a global trigger. It is described in detail in chapter 4. The main B-physics trigger focuses on the identification of detached vertices at Level 1. The trigger must be capable of reducing the event rate by at least a factor of 100 at Level 1, and another factor of 25-50 at Levels 2 and 3.

Much work has been done on the Level 1 trigger since the EOI. We present in this chapter the results of studies of the baseline algorithm under more realistic experimental conditions including such effects as electronic noise, additional hits due to secondary hadronic interactions and photon conversions, and multiple interactions per crossing. We have also increased the thickness of the pixel detectors in our simulations to account for the material in support and cooling structures.

We present a status report on Level 1 trigger R&D including an overview of our hardware prototype and timing results for Level 1 trigger software running on digital signal processors. We also present alternative Level 1 designs in which the pixel detector is modified to reduce the material budget.

Preliminary studies of the Level 2 trigger indicate that the event rate can be reduced by a further factor of 10 through the improved measurement of track parameters and rejection of spurious tracks.

8.2 Advanced Level 1 Trigger Studies

The goal of the trigger is to reconstruct tracks and find vertices in every interaction, up to a luminosity of $2 \times 10^{32}\mathrm{cm}^{-2}\mathrm{sec}^{-1}$. It must be capable of reducing the trigger rate by a factor of about one hundred. We employ parallelism both at the subevent level and at the event level. The detector layout has "stations" consisting of three planes of pixel detectors. Processors organize the hits in each triplet of planes into minivectors or "station hits". Still at the subevent level, individual processors work on the station hits from a "ϕ-slice"[1] of the detector to form tracks. At the event level, tracks from individual slices are brought together in a vertex finder node. The corresponding algorithms for these various stages are described in detail elsewhere [1],[2]. Here we report on:

- Trigger diagnostics and 'software instrumentation';
- Performance studies of this baseline algorithm, in presence of noise backgrounds, both random and correlated;
- Performance studies on the baseline trigger in the presence of multiple interactions per crossing.

These studies are performed using a full hit-level simulation and a full pattern recognition algorithm. Although the final criteria for triggering will depend on the luminosity, the algorithm that we have implemented runs well on crossings which have as many as 3 interactions in them.

8.2.1 Trigger Diagnostics and 'Software Instrumentation'

The Level 1 trigger simulation uses MCFast output "hits" and the MCFast geometry as input data. This is a genuine reconstruction package including pattern recognition (i.e., Monte-Carlo generator information is never used to assemble tracks or vertices). However, back-pointers to this Monte-Carlo information are extensively used to monitor the performance of this pattern recognition. In addition, the Kalman filter from MCFast is also used to compare the crude errors on the track parameters to the nominal tracking resolution one could achieve with the detector. Finally, a preliminary version of the $B^o \rightarrow \pi^+\pi^-$ analysis presented in chapter 11, section 4, has been run in conjunction with the trigger simulation, allowing us to study and refine the trigger efficiency and rejection. We have measured the trigger efficiency on a variety of other states presented in chapter 11.

8.2.2 Uncorrelated Noise Study

Here we investigate the effects on the trigger algorithm of adding random noise to the pixel detector. This study was performed by throwing random hits from a Poisson distribution of average value <N>, uniformly distributed over the area of each plane. Electronic noise or neutrons could cause this kind of distribution.

Fig. 8.1 shows an event with no random noise. Fig. 8.2 shows the same event, but with an average of 25 noise hits distributed randomly over each of the 93 pixel planes.

[1] See Chapter 4 for definition of this particular event parallelism

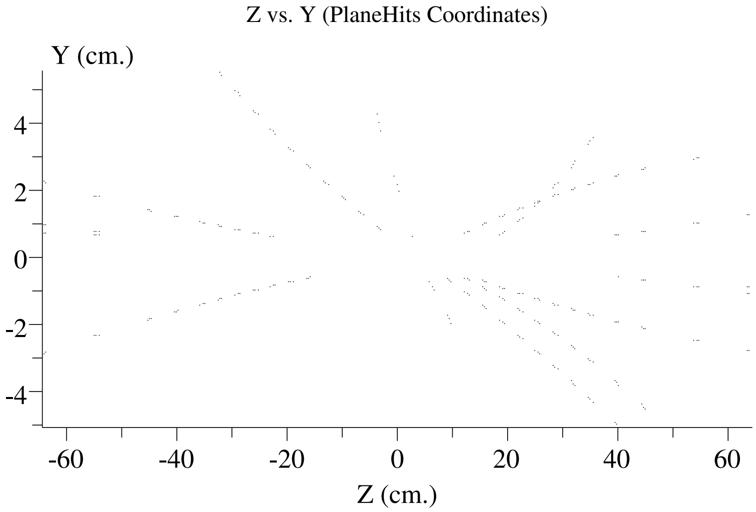


Figure 8.1: Pixel hits in Z-Y projection with no random noise from an MCFAST simulation

Fig. 8.3 shows the efficiency for minimum bias events, which we want the trigger to reject, and signal events (in this case events with a decay $B^o \rightarrow \pi^+\pi^-$), which we want the trigger to accept, plotted against the average number of random noise hits per plane. Three tracks with 3 σ detachment were required in this example.

We believe that it should be possible, and the experience with existing detectors bears this out, to achieve random noise levels below 10^{-5} if hot channel suppression is included in the pixel readout system. The result of our study is that the efficiency and rejection of the baseline trigger are preserved up to $<N> \approx 100$, which corresponds to a contribution of random background to the occupancy of 10^{-4}. We expect the detector to have a much lower noise level than this. However, it was observed that processing time did get somewhat longer when the number of hits rose above a mean of 25. A real estimate of the processing time in the DSP used to reconstruct the minivectors has been made and is described in section 8.6.

8.2.3 Correlated Noise Study

In this study, we used GEANT to simulate the events and turned on all its physics processes. This led to additional hits from hadronic interactions, photon conversions, decays in flight, and curlers (a.k.a loopers). Fig. 8.4 shows that average number of hits per plane rises by about 20% from $\approx$ 8 hits/plane to $\approx$ 10 hits/plane. A degradation of 1.5 is observed in the overall rejection capability of the trigger.

There are other sources of correlated noise. For example, beam-gas interactions or beam halo particles might leave long tracks in the detector. Although the time between crossings is long, there may still be background hits from previous crossings whose source and distribution we are not currently modeling. None of these is expected to be a serious problem but more work does need to be done to prove this definitively.

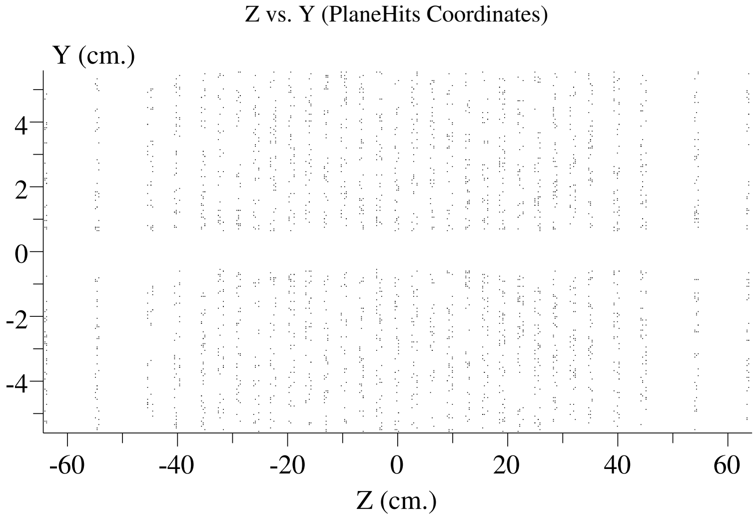


Figure 8.2: Pixel hits in Z-Y projection with an average of 25 random noise hits from an MCFAST simulation

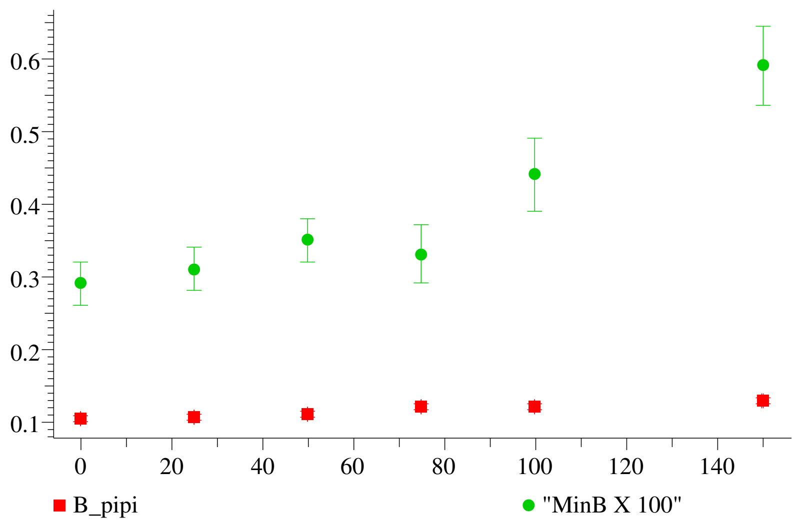


Figure 8.3: Efficiency vs average number of random noise hits for minimum bias and B events. Three detached tracks were required. The acceptance for B events includes all events, even those not falling into the geometrical acceptance.

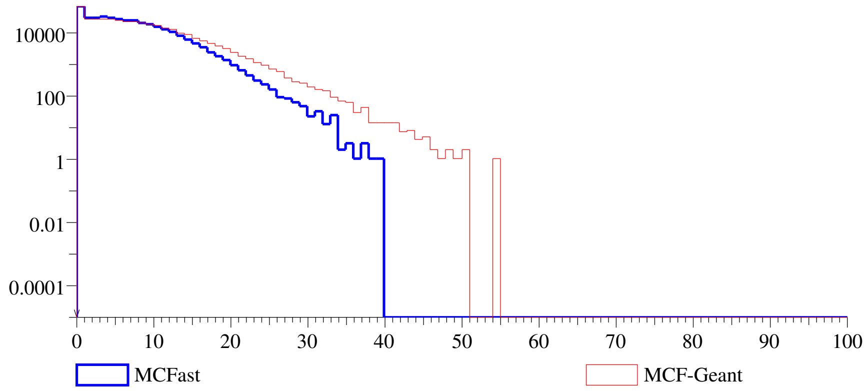


Figure 8.4: Hit multiplicity for minimum bias events in GEANT simulation and MCFAST simulation

8.2.4 Multiple Interactions per Crossing Study

While BTeV will begin running at a luminosity of $5 \times 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, in order to achieve the desired statistical precision, we will eventually want to run at a luminosity of at least 2×10^{32}. Because the Tevatron will have only 108 bunches, this means that the experiment, and in particular the trigger, must be able to handle multiple interactions per crossing. The lower luminosity creates an average of 0.5 inelastic interactions per crossing, while the higher luminosity generates an average of 2.0.

As shown below, even in the presence of multiple interactions, the trigger continues to give good efficiency for states with B decays while providing good rejection against minimum bias events because the occupancy of the pixel vertex detector is very low and the luminous region of the Tevatron is long.

In order to achieve good results when there are multiple primary vertices, we had to make only a small refinement to the trigger algorithm: we put an **upper limit** on the impact parameter (with respect to the primary vertex) which is used to define secondary vertex track candidates. This prevented tracks not used in estimating a given primary vertex (but associated to this interaction) from automatically appearing as 'detached' when compared to another primary vertex. Here we present only a brief summary of the results of an extensive study [3].

Performance on Single Interaction Crossings

The average number of tracks per interaction (primary and secondary) which pass through at least three stations of the pixel detector is 30 for minimum bias events and 38 for $B^o \to \pi^+\pi^-$ events (which will be called B events from now on). After a set of cleanup cuts on the tracks, the average multiplicity of tracks that the trigger uses in the primary vertex search is ≈ 12 for minimum bias events and ≈ 20 for B events. Two additional cuts are placed on tracks which are candidates for use in the primary vertex: they must have a transverse

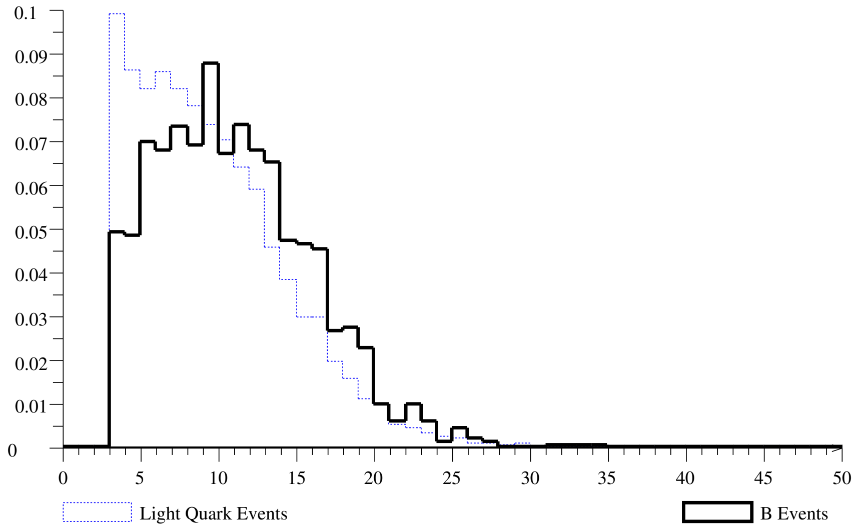


Figure 8.5: The track multiplicity distribution in the primary vertex, for single interaction crossings. The average is 9.6 for minimum bias events, and 11.1 for $B \to \pi\pi$.

momentum of less than 1.0 GeV/c; and they must intersect the nominal beam line within the luminous region. With these requirements, the resultant multiplicity of primary vertex tracks found by the trigger algorithm in a crossing with exactly one interaction is given in Fig. 8.5 for minimum bias and B events. The Z resolution (i.e. along the beam axis) of the primary vertex is given in Fig. 8.6. Note that this distribution has typically more tails for the B sample because of pattern recognition failures in events with multiple vertices: if many tracks from a B decay are mistakenly taken as primary tracks, the B decay vertex could be wrongly taken as the primary. Nevertheless, since some tracks are still likely to be detached, these events could pass the trigger.

To form a primary vertex requires at least 3 tracks. As a consequence of this cut and the transverse momentum cut, the probability to find a second, false, primary vertex is very small, only 0.05%, in minimum bias events. The 'leftover' track multiplicity after primary vertex formation is very different for minimum bias and B events. If one applies an additional detachment cut on the normalized impact parameter $\frac{\sigma_b}{b}$ of 3 and a transverse momentum cut $P_t > 0.5$ GeV/c one has reconstituted the analysis we presented in 1997 [1]. To deal with multiple interactions, we added the requirement that the 'detached track candidate' not be **too far from the primary vertex** – the impact parameter b must be less than 0.2 cm from the primary. This cut also allows us to reject tracks from K_s decays. In addition, further cleanup cuts on these detached tracks are now applied. For instance 4 station hits (or minivectors) on a track are required. Then, one obtains the distribution in Fig. 8.7.

Table 8.1 shows the trigger acceptance for minimum bias events and B events for single interaction crossings. Our goal for Level 1 was to accept fewer than $\approx$1% of minimum bias events, which is achieved in all conditions. Other triggers based on global properties of the

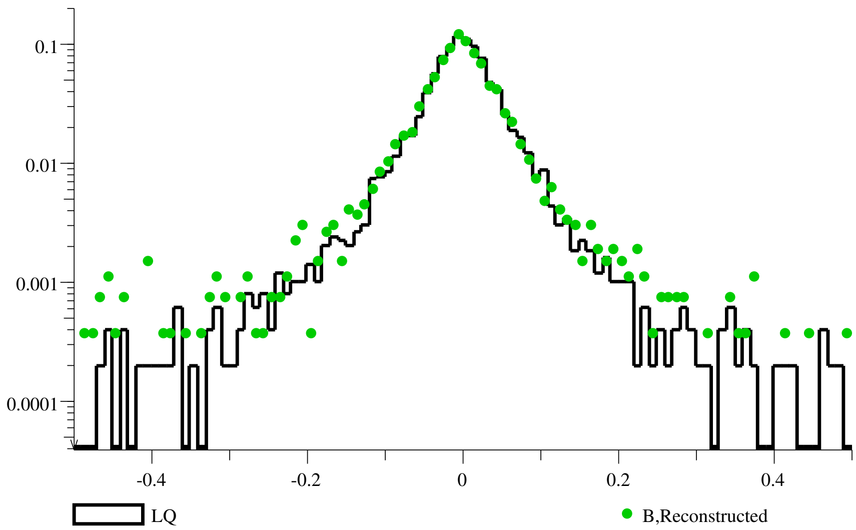


Figure 8.6: The primary vertex Z resolution, for single primary vertex events. The width (σ) of the core of these distributions is approximately 270 μm , while the r.m.s is about 760 microns, reflecting the presence of tails in the distributions.

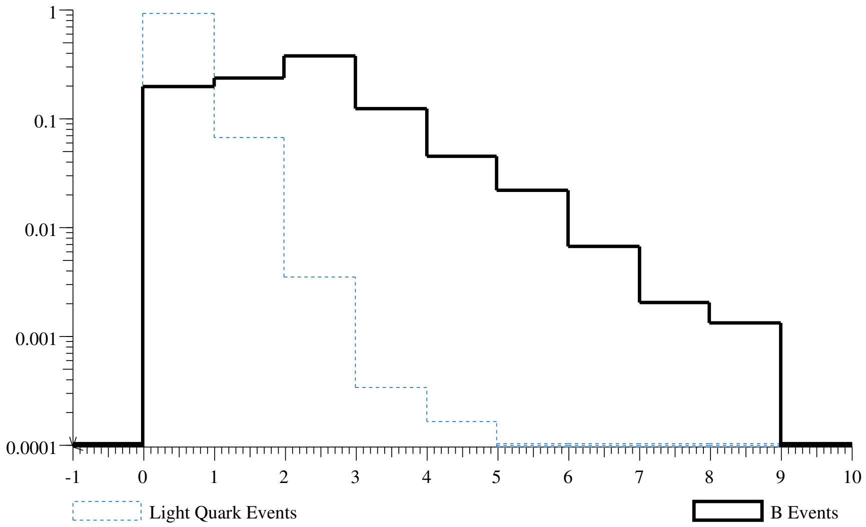


Figure 8.7: The detached, $P_t > 0.5 GeV/c$ track multiplicity for both samples. The average track multiplicity for minimum bias (B) is 0.6 (1.5).

Table 8.1: The trigger acceptance (%) for single interaction crossings, for the minimum bias sample and the $B \to \pi\pi$ sample, versus selection criteria (number of tracks detached from the primary, detachment cut, P_t on each detached track, and impact parameter upper limit in cm). The errors are statistical only.

Criteria	Minimum Bias	$B \to \pi\pi$
2 detached tracks ($4\sigma; P_t > 0.5$ GeV/c; $b_t < 0.2$)	0.9 ± 0.1	54.9 ± 0.9
3 det. tr. ($3\sigma; P_t > 0.5$ GeV/c; $b_t < 0.2$)	0.05 ± 0.02	18.5 ± 1.2
3 det. tr. ($3\sigma; b_t < 0.2$; $Sd_{Pt} > 1.0$)	0.17 ± 0.04	27.0 ± 1.5

vertex and its ensemble of detached tracks have been studied. For instance, the total scalar transverse momentum Sd_{Pt} summed over all these detached tracks could be used, as in the third line of Table 8.1.

Performance on Multiple Interaction Crossings

Fig. 8.8 shows the number of primary vertices reconstructed by the trigger algorithm for an average of 0.5 and 2.5 interactions per crossing, corresponding roughly to luminosities of 0.5×10^{32} and $2.5 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. These plots suggest that the number of reconstructed vertices is well correlated to the number of generated vertices even in these busy events, which can be seen as follows: The distribution can be qualitatively understood as a Poisson distributed number of interactions with a fixed vertex finding efficiency of about 62%. The efficiency is low because the trigger is deliberately made insensitive to very soft collisions such as quasi-elastic and low multiplicity interactions which are unlikely to contain B's.

Fig. 8.9 shows that the primary vertex Z resolution is independent of the number of primary vertices per crossing. No degradation of this resolution is observed as the luminosity increases. Each interaction can be studied independently because the average distance between primary vertices is much larger than the resolution.

Fig. 8.10 shows the 'multiplicity distribution' of detached tracks for different numbers of reconstructed vertices when the average number of interactions per crossing is ≈ 0.5 (corresponding to a luminosity of $\approx 5 \times 10^{31}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$), and a luminosity 5 times this amount. The main point is that there is not a large tail of events in which the number of detached tracks begins suddenly to become troublesome as the number of reconstructed vertices increases. We observe that the efficiency for detecting B events does not degrade as the number of actual or reconstructed vertices increases. Note also that the detached tracks must be associated to the same primary vertex in order to satisfy the trigger.

The total yield of B events per unit time does not degrade as additional detached tracks are found in other primary vertices, provided of course one does not explicitly reject these multiple primary vertex crossings. If one does, the yield follows what one would expect from the simple Poisson law. The probability of triggering on a crossing with at least one inelastic interaction is shown vs luminosity (in units of $10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$) in Fig. 8.11. (Here, we have used a trigger requirement of at least 3 tracks each missing the primary by $\geq 3\sigma$.) This

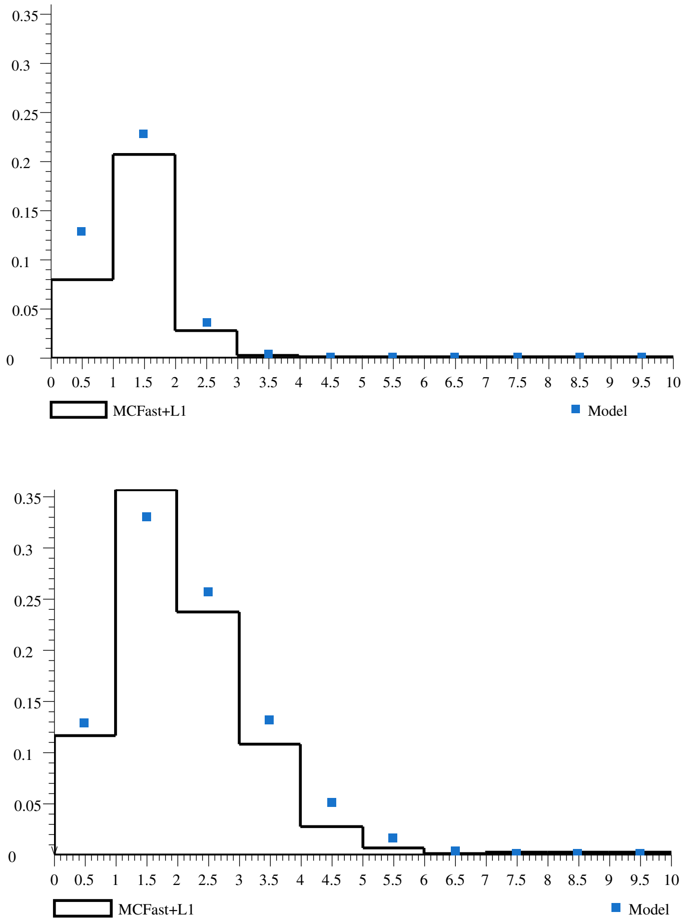


Figure 8.8: The Level 1 primary vertex multiplicity distribution running at luminosities of 5×10^{31} (top plot) and 2.5×10^{32} (bottom plot). These histograms are normalized to the total number of 132 ns crossings in the run. For instance running at 2.5×10^{32}, the probability to reconstruct one and only one primary vertex is about 35 %. Also shown is a simple model assuming a fixed vertex geometrical acceptance and reconstruction efficiency of about 62 %.

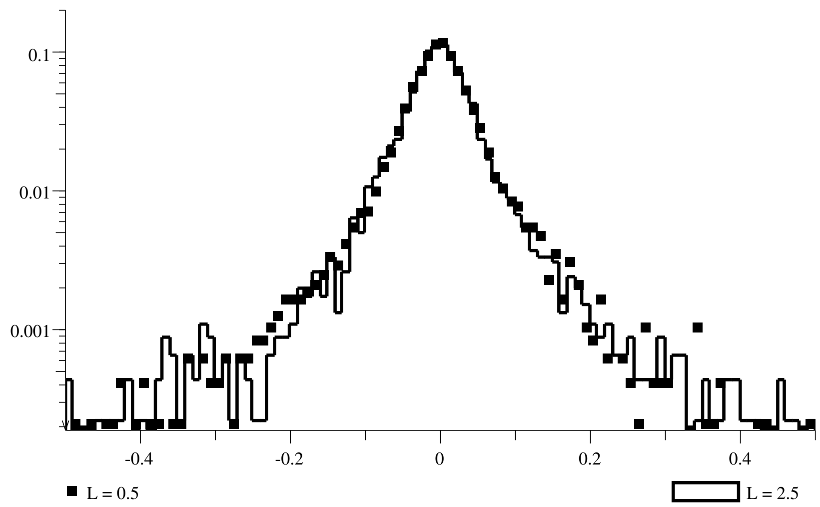


Figure 8.9: The Z position primary vertex resolution plot, running at luminosities of 5×10^{31} and 2.5×10^{32}.

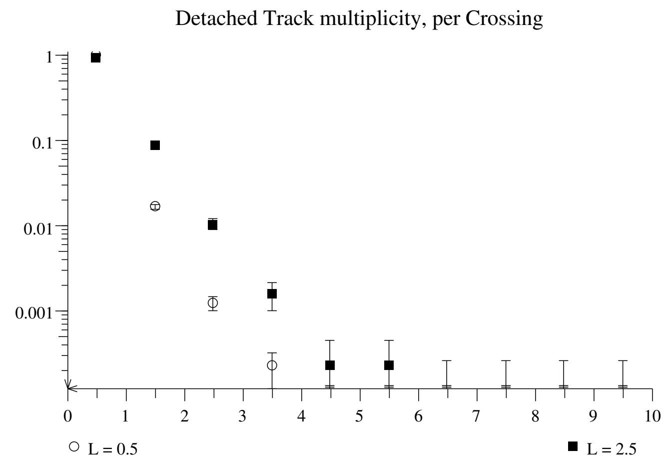


Figure 8.10: The detached (3σ , $P_t > 0.5$) track multiplicity distributions for light quark events running at a luminosities of 5×10^{31} and 2.5×10^{32}. These are normalized to the total number of crossings in the run: the probability to observe 1 detached track at $L = 5 \times 10^{31}$ is $\approx 1.6\%$, and at $L = 2.5 \times 10^{32}$, it is $\approx 8.5\%$

probability is about twice as large if the two detached track at 4σ trigger is used. Ideally, one would expect the yield per unit time to increase linearly and it does increase roughly linearly as long as one does not restrict the number of interactions too severely. The probability for triggering on crossings with inelastic interactions does increase but stays below 1% up to a luminosity of 2.5×10^{32} if we allow at most three reconstructed vertices. For the two tracks at 4σ, in the worst case, the probability is slightly above 1%. There are many additional requirements (such as the presence of a true secondary vertex rather than tracks missing the primary) that can be imposed at Level 1 to reduce the trigger rate even more.

Our results on the performance of the baseline trigger in the presence of multiple interactions per crossing are quite encouraging. In essence,

- The trigger rate is not dominated by unresolved multiple vertices up to at least 2×10^{32} .
- A Level 1 rejection factor of about one in a hundred is possible with a Level 2 bandwith that can be achieved (see chapter 9).
- One gains sensitivity by running above an average of 1 interaction per crossing provided we accept crossings with up to three interaction vertices.

Thus, we believe that it is possible to achieve a rejection against minimum bias events better than 1/100 at Level 1 up to a luminosity of 2×10^{32} with good efficiency for B events we feel confident to reconstruct offline. One can work above 2×10^{32} provided one can handle the data rate of falsely accepted minimum bias events into the Level 2 trigger or one can reduce the rate into Level 2 by adding more requirements on Level 1. However, there is more work to be done in making the simulation more realistic and in refining the algorithm. Also, we need to estimate the required CPU power at each stage of the reconstruction, to balance the data flow in the pipeline. We have started looking into the queuing problems at the front end of the triggering system. We look forward to reporting further on this in the future.

8.3 Efficiency and Rejection of the BTeV Level 1 Vertex Trigger for Selected Physics States of Interest

The results are documented in the above section and in part 3. Here we summarize our results. The rejection factor on light quark events for the two detached track trigger is about 1/100 for single interaction crossing and is proportional to the luminosity if multiple interactions per crossing are accepted. We summarize the efficiency for this trigger in table 8.2. In all cases, the opposite b-hadron decays following the standard QQ decay table (defined in part3) .

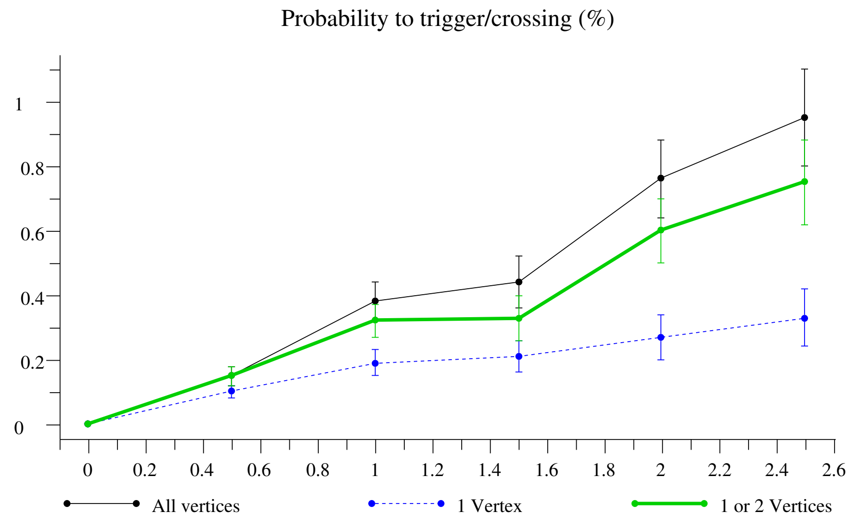


Figure 8.11: The absolute probability for triggering per 132 ns beam crossing. The trigger is based on three detached tracks pointing to the same longitudinal hemisphere, with $Sd_{Pt} > 1.0$. For both plots, the abscissa axis is the luminosity in units of 10^{32} cm^{-2} s^{-1}. The solid line is for events with any number of primary vertices; the dotted line, for events with one and only one reconstructed vertex; and the thick gray line for events with 1 or 2 reconstructed vertices.

Table 8.2: The two detached track at 4σ trigger efficiency for various B decays.

Final State	Efficiency (%)
$B \to \pi^+\pi^-$	55
$B_s \to D_s K$	70
$B^- \to D^0 K^-$	60
$B^- \to K_s \pi^-$	40
$B^0 \to K^+\pi^-$	54
$B^0 \to J/\psi K_s$	50
$B_s \to J/\psi K^*$	69
$B^0 \to K^*\gamma$	40

8.4 Other Level 1 triggers

While we have emphasized the "vertex/impact parameter" trigger because it is our most crucial and most ambitious trigger, we will also have other triggers at Level 1. These will include muon, minimum bias, and (possibly) electron triggers and perhaps some triggers derived from the same hardware that is used for the vertex trigger but optimized for different physics.

8.5 Status Report on Level 1 trigger R&D and trigger prototype

The Level 1 vertex trigger is a key feature of the BTeV detector design. A proposed scheme for the vertex trigger has been worked out by the Penn group in collaboration with D. Husby at Fermilab. The hardware design for this system is described in detail in references [4] and [5]. Although the design can be implemented with DSPs (digital signal processors) and FPGAs (field-programmable gate arrays) that are available today, an intensive R&D program is needed to develop the complete design, test the viability of the system, and perhaps modify some of the hardware components as new hardware becomes available. The R&D program has three main components:

- Simulations;
- Hardware prototyping;
- Software development.

Simulations of the trigger are an important aspect of the trigger R&D program. We generate pixel hits using the MCFast simulation program, and process the hits using a software implementation of our Level 1 trigger algorithm. This procedure allows us to test the trigger algorithm, and to refine the algorithm as we include more realistic running conditions for the experiment. We also use trigger simulations to learn more about the performance of the Level 1 trigger under "normal" and "extreme" running conditions, and to develop the software that is needed for our trigger prototype (see Fig. 8.12).

Although much of the testing of the trigger algorithm can be done with simulations, there are tests that are best done with a hardware trigger prototype. The trigger prototype is a single-board module that implements most of the data paths and processing steps used in the trigger system. It has FIFOs to load data generated by the MCFast simulation program, and it includes four hit processors, four farm (track or vertex) processors, and the merging and switching circuitry that connects them.

The prototype is used to test data flow through the trigger system. There are three primary data paths that are tested:

1. Single station-quadrant to track finder.

   In this test, the prototype is used to test the path from one quadrant of a pixel station to the farm processors that reconstruct tracks. Pixel hits from the three planes of a station-quadrant (see Fig. 8.13) are divided into eight ϕ slices, and fed to an array of eight hit processors. The prototype is used to test paths through four of the eight processors. These hit processors execute the segment-finding algorithm that finds minivectors in the vertex trigger. The minivectors are sent through the sorting network to the farm processors. Each farm processor runs the standard track-finding code to produce track records that are written to a FIFO, which is used to read out and verify test results. The results are used to investigate hit-processor throughput, study the effects of extreme event sizes, and test the ability of the switching network to sort hits.

2. 31 stations to track finder.

   This tests the flow of data for all 31 pixel stations that make up a single ϕ slice. The four hit processors execute code that mimics the data generated by 31 hit processors. We use the results from this test to study the performance of the switching network under a full load.

3. 32 track finders to vertex finder.

   This tests the flow of data from 32 track processors to the vertex processing farm. The four hit processors execute code that mimics data generated by the track processors for all ϕ slices. We use the results to study the switching network, and the vertex finding algorithm.

For each of the three tests, a performance monitor keeps track of statistics on the data paths that are used during each cycle. These statistics are sent to a host computer for graphical display and analysis. The host computer is also used to configure, monitor, and control each of the processors and FPGAs in the prototype.

The design of the trigger prototype entails both hardware and software development. The design of the hardware has been completed. The software that is being developed for the prototype consists of code that is used to implement test scenarios, and the three programs that implement the Level 1 vertex trigger algorithm. The first of these three programs runs on hit processors (each hit processor is a Texas Instruments DSP) and finds track minivectors. The second and third programs run on farm processors (each farm processor is an Analog Devices DSP) and perform track and vertex reconstruction, respectively. The Texas Instruments and Analog Devices DSPs differ in their hardware architecture. Each DSP requires software for data flow and buffer management, and the software to implement the Level 1 trigger algorithm.

The software development for the DSPs encompasses three different aspects:

- Data flow and buffer management;

- Trigger algorithm;
- Optimization.

The code that handles the data flow and buffer management is specific to the hardware architecture of each DSP, and depends on the design details of the trigger prototype. The code that implements the vertex trigger algorithm is based on the trigger code used in our MCFast simulations, but significant modifications are required to produce code for the DSPs. Compared to the trigger code in MCFast, the software for the DSPs is optimized to take advantage of the DSP architecture. This optimization requires assembly language programming to achieve the performance (see timing results in the following section) that we require for the Level 1 vertex trigger.

8.6 Results on Baseline Level 1 Algorithm Timing on Digital Signal Processors

With 100 gigabytes of pixel data coming into the Level 1 trigger every second, a beam crossing every 132 ns, and an average of 2 interactions per crossing, the time required for each pipeline step in the trigger is critical. Timing issues are studied by performing a variety of trigger simulations. These include Monte Carlo studies of the pattern recognition and reconstruction algorithms, and simulations of the trigger hardware. Furthermore, the timing of both the trigger hardware and software algorithms will be extensively tested with the trigger prototype described in the previous section. In this section, we describe timing studies for algorithms on DSPs (digital signal processors).

The BTeV trigger utilizes two types of DSPs: a fixed-point DSP, and a floating-point DSP. The fixed-point DSP is made by Texas Instruments and belongs to the TMS320C6X family of DSPs. The floating-point DSP is an Analog Devices ADSP-2106x SHARC. The trigger code for both the Texas Instruments DSP and the Analog Devices DSP is developed with the aid of simulation software provided by each company. We also use this software to obtain the timing results presented in this section.

Simulations of the Level 1 design show that the first stage of the trigger, the segment finder, requires the most computational power compared to other parts of the trigger. A large number of operations are performed by the segment finder, due to the number of combinations of pixel hits that must be sampled to find the hits that define a mini-vector. To address this problem, we introduce considerable *parallelism* in the trigger architecture. This is a key feature of the Level 1 trigger design. The parallelism begins with the organization of the data readout from the pixel detectors. Pixel hits are read out in parallel for each quadrant of a pixel plane. The hits from the three planes that belong to a pixel station are brought together in a *quadrant processor*, which represents one of 124 identical circuit boards that make up the first stage of the trigger. Additional parallelism is introduced by subdividing the pixel hits in a quadrant into 8 ϕ slices, for a total of 32 ϕ slices per pixel station. The hits in each ϕ slice are processed in parallel by one of the 992 TMS320C6X DSPs

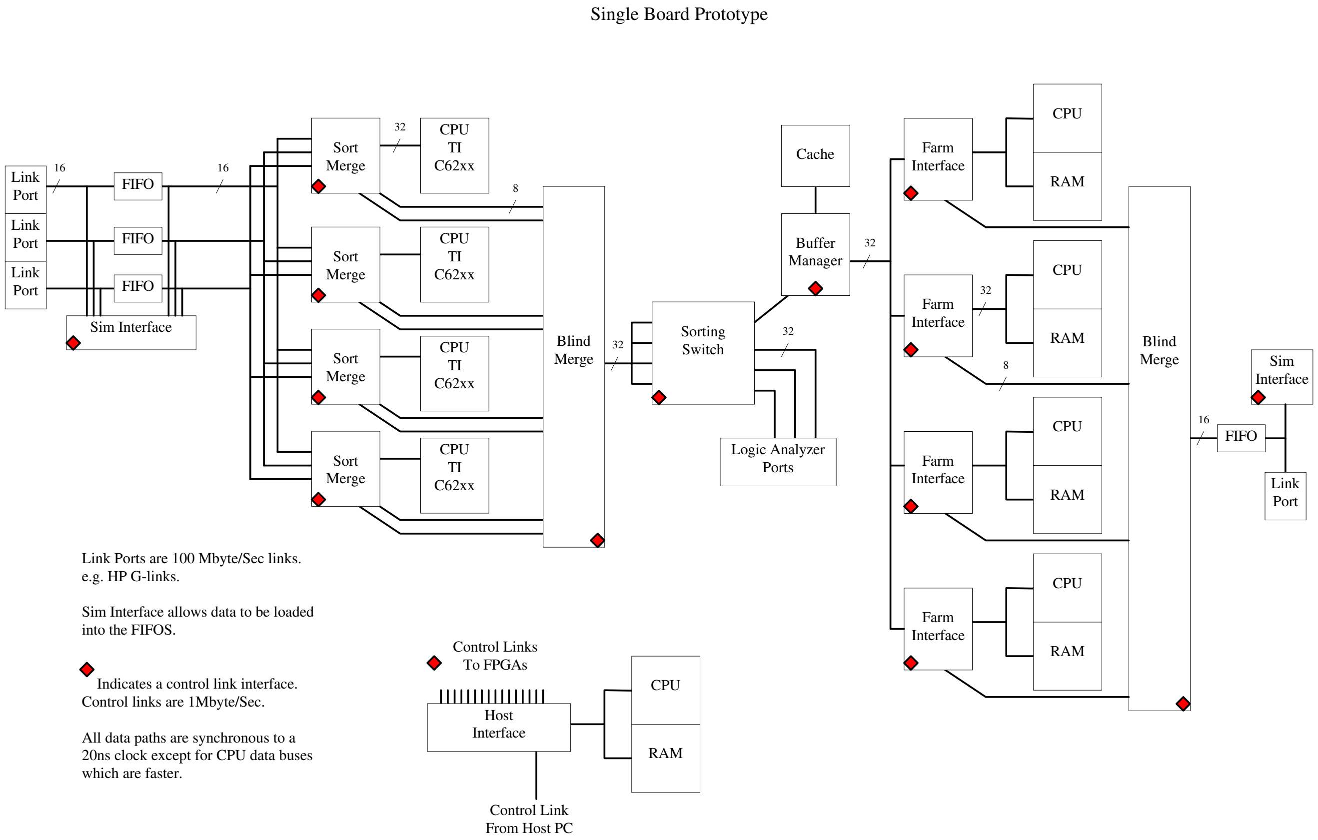


Figure 8.12: Schematic of Single Board Trigger Prototype

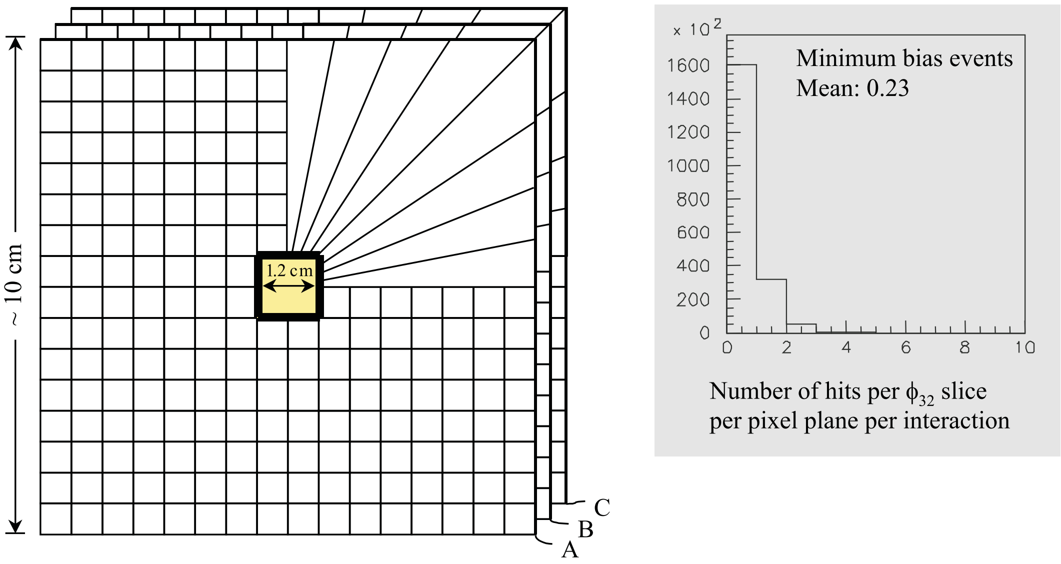


Figure 8.13: Trigger view of one quadrant in a pixel station with planes A, B, and C. The histogram on the right shows the number of hits for each plane per ϕ slice for minimum bias events.

(a total of 992 DSPs for 31 pixel stations, each with 32 ϕ-slice DSPs). Fig. 8.13 shows the "trigger view" of one quadrant, and shows that the subdivision of hits into ϕ slices reduces the mean number of hits per pixel plane to a value of 0.23 hits per ϕ slice for minimum bias events.

Even with the small number of hits per ϕ slice, the timing for the segment finder is critical. Table 8.3 shows timing results for the segment finder algorithm developed for the TMS320C6X DSP. In the table we choose four cases to present our results: (i) For a given ϕ slice and with three planes (A, B, and C) in a pixel station (as shown in Fig. 8.13), we have a single hit in each plane for one track passing through a pixel station. This is the (1,1,1) case in Table 8.3. (ii) For two tracks passing through the same ϕ slice we have two hits per plane, or (2,2,2). (iii,iv) With three and four tracks in the same ϕ slice we get the (3,3,3) and (4,4,4) cases, respectively.

The timing results for these four cases are divided into three categories that represent different levels of optimization for the segment finder algorithm. The first category involves *C* code that has *not* been optimized (the code is compiled without any optimization), and we use this to establish a reference measurement for subsequent measurements involving different levels of optimization. For the second category we allow the *C* compiler for the TMS320C6X DSP to optimize the algorithm, and for the third case the programming is done directly in assembly language. Although programming in assembly language is more cumbersome, we can optimize the code so that all hardware components in the DSP are performing useful operations most of the time. Our timing results in Table 8.3 show that

Number of hits per ϕ slice in planes (A,B,C)	Time for non-optimized segment finder(ns)	Time for compiler optimized segment finder(ns)	Time for assembler optimized segment finder(ns)
(1,1,1)	1915	1630	135
(2,2,2)	5705	3670	315
(3,3,3)	14015	7560	1340
(4,4,4)	28915	13840	2885
weighted average	530	395	37

Table 8.3: Timing results in nanoseconds for the segment finder algorithm running on the TMS320C6X DSP.

the *C* compiler for the TMS320C6X DSP is not very effective at optimizing the segment finder algorithm; the compiler achieves only a minor performance gain for the (1,1,1) case, and at best a factor of two improvement for cases with more hits in a ϕ slice. Table 8.3 also shows that we are able to achieve almost an order of magnitude improvement in the timing by optimizing the algorithm in assembly language. This does not include additional improvements that can be achieved for the (3,3,3) and (4,4,4) cases, which have not yet been fully optimized.

To obtain an estimate of the average time required to find mini-vectors in minimum bias events, we compute a weighted average by using the results in Table 8.3 weighted by the distribution of hits in Fig. 8.13. We use a value of zero for the (0,0,0) case, since the DSP does not perform any operations when there are no hits in a ϕ slice. The average time for each level of optimization is shown in the last row of Table 8.3. Although these results are somewhat oversimplified, they do indicate that an assembly language implementation of the segment finder algorithm is necessary so that the first stage of the trigger stays within a budget of 132 ns per beam crossing. The average time of 37 ns for the segment finder is encouraging, since it suggests that we are close to having a viable implementation for the segment finder in our Level 1 design.

The track and vertex processor timing are also under study. We hope to present results soon.

8.7 Options to the Baseline Level 1 Vertex Trigger

8.7.1 An Alternative Using Two Plane Stations

The choice of three-plane pixel stations in the baseline detector is dictated by the baseline trigger design. That trigger uses the three planes to make minivectors that can then be projected to a small portion of the next station, which substantially reduces the combinatorics. Any attempt to use fewer planes per station will have to cope with larger combinatorics when starting the search for tracks.

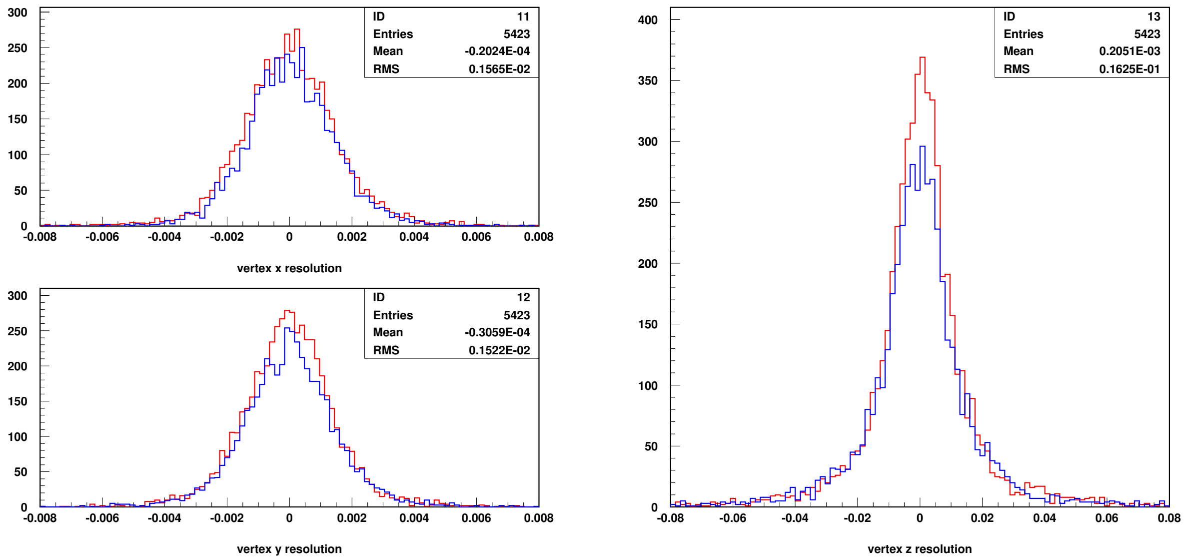


Figure 8.14: Comparison of the vertex resolution for $B^0 \rightarrow \pi^+\pi^-$ for two-plane pixel stations and three-plane pixel stations–left top x, left bottom y, right z.

The removal of one plane per station has advantages of lower cost, a simpler detector, and less material with no loss of physics capability. Simulations using the identical pixel detectors arranged in 31 stations of three planes and 34 stations of two planes showed almost no difference in the effective resolution and efficiency of the two detectors on charged particle states. A full offline analysis was done for several interesting physics channels, $B^0 \rightarrow \pi^+\pi^-$, $B_s^0 \rightarrow \Psi K^*$ and $B_s^0 \rightarrow D_s^+\pi^-$. In the case of $B^0 \rightarrow \pi^+\pi^-$ the secondary vertex resolution is a good figure of merit and is shown in figure 8.14. The vertex resolution is almost identical in the two detector configurations. The 5% difference in efficiency may be recoverable with optimization of station placement.

The time resolution for the B_s decays $B_s^0 \rightarrow \psi K^*$ and $B_s^0 \rightarrow D_s^+\pi^-$ (used to study B_s mixing) is also unaffected by the choice of two-plane or three-plane stations. Simulations show that for the first decay mode the time resolution is 44 fs for both two-plane and three-plane geometries, and for the second decay mode the two-plane is slightly better at 54 fs versus 56 fs for the three-plane geometry. These results can be understood by the fact that multiple scattering is an important contributor to the resolution, and the reduction in the material compensates for the lost measurements.

i. Algorithm Design and Simulation

The basic pattern recognition procedure is still finding straight lines. However they are found across stations instead of within a single station. Because we cannot point into the next station we will have to search over a large section of the next station. It is efficient to restrict the search to quadrants of each station and we do so.

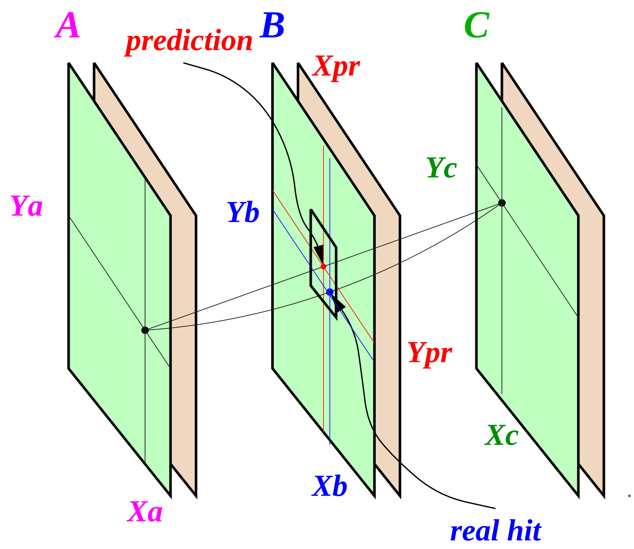


Figure 8.15: Sketch of the basic interpolation algorithm for the planes with fine pitch in x.

We choose three adjacent stations and try to find straight lines by using the hits in the two outer stations called A and C to predict the hit in the middle station B. The prediction is made in both the x and y coordinates since the pixels have 2D information. The algorithm is pictured in Figure 8.15:

$$\begin{aligned} \left|\frac{x_A + x_C}{2} - x_B\right| &< \Delta_x \\ \left|\frac{y_A + y_C}{2} - y_B\right| &< \Delta_y \end{aligned}$$

The size of Δ_x and Δ_y is dictated by the resolution of the pixel detectors, the multiple scattering, and for the y view the bending due to the magnet. In terms of pixel widths, the windows are $\Delta_x = 3$, $\Delta_y = 2$ for x planes and $\Delta_x = 2$, $\Delta_y = 10$ for y planes. The windows are 6 pixels in x planes and 20 pixels in y planes out of 125000 pixels in a quadrant, so the background suppression is excellent. Since the window sizes are larger than the pixels we do not need the full analog information at this stage but only the address of the cluster peak or similar information.

The three stations used to find line segments are overlapped, with each station sending its data to three different line segment finders. At this stage the ensemble of hits is in fact a collection of linked lists. These lists can easily be matched to tracks without combinatorial searches. Track segment fitting is currently less developed, but the momentum and intercept estimators used in the baseline trigger is a fall back option.

ii. Feasibility of Implementing the Design

The basic pattern recognition ability of the pixel detector is more than adequate to find line segments due to real tracks. The main concern is whether the hardware will be fast enough to keep up with the data rate. At a mean of two interactions per crossing we expect to see an average of 8 hits per quadrant and 95% of all quadrants in all events have fewer than 16 hits. If we implement the algorithm using a microprocessor performing loops then the time to complete the search will go like N^3, so we must introduce some parallelism into the problem.

The basic calculation consists of an addition, a multiplication (which can be done as a bit shift), a subtraction, and a test. All of these calculations can be done as integers. Programmable Logic Devices (PLDs, also called FPGAs) contain a large number of simple logic units that can perform hundreds of these calculations in parallel. This is significantly more parallelism than in the baseline trigger design allowing this design to handle the larger combinatorics.

The design can be broken into three sections: an input section that receives and buffers the hits; an interpolation section that does pattern recognition; and an output section that collects line segments to pass on to the next level of processing. The times to complete the various sections scale differently with event complexity. The timing in the input and buffering section scales as the number of hits divided by the number of input paths. If the time for this section is large due to many hits, the hits can be brought in in parallel as long as there are sufficient I/O pins on the chip. The timing of the interpolation section scales with the number of bits used in the interpolation, so it is independent of the size of the event. However, there will be a loss of efficiency should the event exceed the maximum capability of the processor. The timing of the output section scales with the number of line segments found, which makes it stable against random noise and low energy backgrounds.

A design of the line segment finder has been done for a Xilinx PLD. The basic idea is to read all of the hits from the A and C stations into one PLD. These are then used to feed a matrix of interpolation units which will do each hit from the B plane sequentially (see Figure 8.16). We use bit-serial arithmetic to shrink the data bus width and better utilize routing and logic resources. The time do this comparison goes as 1 clock cycle per bit of hit coordinate.

A single chip can check four B plane hits against the interpolations from all A and C plane hits in a quadrant in parallel. Simulation predicts this will take about 120 ns to handle 12 bits of hit information at a clock period of 10 ns. In order to keep up with the data rate 2 chips per station quadrant per view would be required.

The size of a PLD is measured in either gates or configurable logic blocks (CLBs). The design of the interpolation nodes requires 5 CLBs per node so a PLD that can handle a maximum of 16 hits at a time would require $16 \times 16 \times 5 = 1280$. Currently we target a Xilinx PLD that has $8464 = 92 \times 92$ CLBs, so 4 arrays of interpolation nodes can be placed with 40% of the chip left for control logic, input buffers, etc. The new Virtex family chips from Xilinx are more capable and available now.

The I/O capabilities of these PLDs are more than adequate. Seventy-two I/O pins are needed to read in the 12 bits of each coordinate for a hit from each of three planes. The

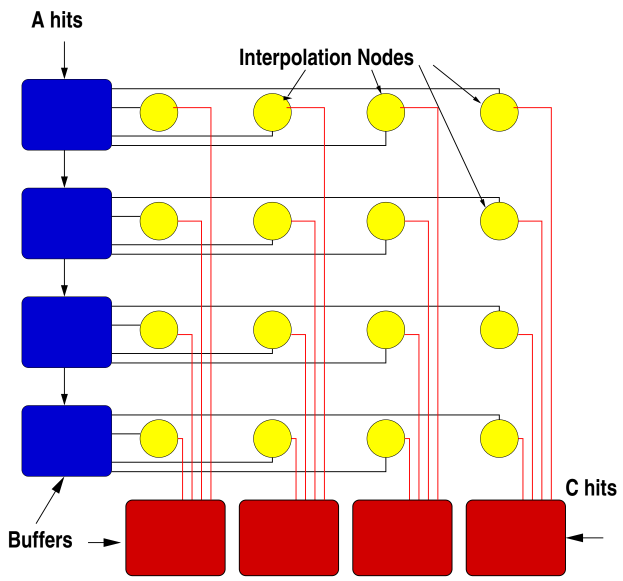


Figure 8.16: Block diagram showing the matrix of interpolation nodes and input buffering for the outer planes.

PLD under consideration has 448 I/O pins. There are more than enough I/O pins to bring in two or three times more hits in parallel. Hits from A and C planes are buffered on the PLD from one event while the previous event is analyzed.

A simple cross check using a Lucent PLD confirms the basic performance. The interpolation portion of the line segment finder was run through the Lucent device simulation and equivalent results were found. The design should work in a variety of devices so while present devices are adequate in both speed and I/O capabilities, we can delay the final choice to take advantage of better and/or cheaper devices as they become available.

8.7.2 The 'Megapixel Trigger'

The Megapixel Trigger is an alternative Level 1 trigger scheme. It benefits from concepts that have been explored in the context of the BTeV baseline vertex trigger and the 2-plane vertex trigger described in the previous section. Like the 2-plane trigger, the Megapixel Trigger is designed for a pixel system with two planes per tracking station. Moreover, the track reconstruction is initiated in a similar manner by finding track segments that span three adjacent pixel stations. Unlike the 2-plane trigger, the Megapixel Trigger limits the track reconstruction to track segments that are found as *inner/outer* tracks, which involve *inner* and *outer* pixel hits (defined below). The inner/outer tracks are used to "seed" the track-finding algorithm. Seed tracks are extended to adjacent pixel planes to find all hits belonging to a trajectory. An important feature of the Megapixel Trigger is that seed tracks are found relatively close to the primary-interaction vertex. This feature can be exploited to extend parallelism in the hardware architecture throughout the Level 1 trigger.

One of the central concepts of the Megapixel Trigger is the distinction between *inner* and *outer* pixel hits, where a single hit corresponds to an (x,y) position measurement made by a clustering algorithm, which combines data from adjacent pixels. Inner hits are defined as hits located in a region close to the Tevatron beams. For hits in a pixel plane, this region is defined as a square with a side of length D centered on the beams and having the same orientation as the square beam hole in the pixel plane. In practice, the magnitude of D is determined by the size of individual pixel sensors, so that the innermost sensors are all within the square region that surrounds the beam hole. The pixel hits from these innermost sensors are the inner hits, and about one third of all hits are inner hits. All of the remaining hits, outside the square region, are outer hits.

The Megapixel Trigger combines hits from three adjacent pixel stations to reconstruct track segments. The track finding is constrained to combinations of three hits, in which the first hit is an inner hit, and the second and third hits are outer hits. By imposing this constraint, the Megapixel Trigger finds only those trajectories that pass through at least three pixel stations, and pass from the inner to the outer pixel region. The 3-station track segments that are found in this manner are referred to as *inner/outer* tracks. The advantages of this approach are that (1) only a limited number of hit combinations need to be considered for pattern recognition, and (2) each trajectory is found *once* by the trigger hardware. By finding tracks only once, the Megapixel Trigger differs significantly from the BTeV baseline trigger, where track segments belonging to a single trajectory are found for *each* pixel station that recorded hits for the trajectory.

The percentage of tracks found as inner/outer tracks is quite high, and is shown in Table 8.4. The table shows the percentage of inner/outer tracks relative to all tracks that pass through at least three pixel stations and have a momentum greater than 1 GeV/c. Results are shown for different types of tracks in minimum bias, B, and charm events for several values of d, where d (which is equal to $D/2$) is the distance from beam center to the border that defines the inner and outer pixel regions. Table 8.4 shows that for most types of tracks the percentage of inner/outer tracks does not vary with d, for the range of d values we have considered. For tracks coming from B decays there is a slight increase in inner/outer tracks for larger values of d, due to the larger p_t for these tracks. From these results we conclude that a large percentage (about 90%) of tracks coming from primary interaction vertices can be found as inner/outer tracks, and used to reconstruct primary vertices in the Megapixel Trigger. Furthermore, we conclude that a large percentage of B-decay tracks (more than 90% for values of d greater than 1.1 cm) can be found, and used to trigger on B events.

After the inner/outer tracks are found, the Megapixel Trigger identifies the *innermost* pixel hits for each trajectory found as an inner/outer track. The goal is to reconstruct the track segments closest to the point of origin. These are the track segments that are least likely to suffer from multiple Coulomb scattering errors, and are required to reconstruct vertices and identify B-decay tracks. For example, Fig. 8.17 shows pixel hits for a simulated $B \to \psi K_s$ event in the $z-y$ projection. In Fig. 8.17a inner/outer tracks are shown as line

Distance of inner/outer border from beam center	Minimum Bias Tracks from primary vtx.	Generic B Tracks from primary vtx.	Generic B Tracks from secondary vtx.	Generic charm Tracks from secondary vtx.
d = 1.0 cm	91%	91%	87%	89%
d = 1.1 cm	91%	91%	91%	90%
d = 1.2 cm	91%	92%	93%	90%
d = 1.32 cm	90%	91%	94%	90%
d = 1.5 cm	89%	89%	95%	89%

Table 8.4: Percentage of tracks found as inner/outer tracks, where tracks are required to have momentum greater than 1 GeV/c and are required to pass through at least three pixel stations.

segments. There are two segments per track: one from the first hit to the second hit on the track and one from the second hit to the third hit. Fig. 8.17b shows the corresponding track segments for the innermost pixel hits associated with each track that was found as an inner/outer track. In many cases, the inner/outer tracks already include the innermost pixel hits, and no additional step is required to find these hits. For the remaining tracks, we are considering two algorithms to find the innermost hits. In the first algorithm, we extend tracks to adjacent pixel stations to locate the innermost hits by using an approach that is based on the linking of minivectors in the BTeV baseline trigger. In the second algorithm, we reconstruct inner/inner tracks in the first stage of the Megapixel Trigger, and then eliminate track candidates that do not include the innermost hits. For both algorithms, the track segments that are found are subjected to a fitting procedure to determine track parameters.

The hardware architecture of the Megapixel Trigger is organized as a pipeline with parallelism in z (the beam direction), and individual track and vertex processors (digital signal processors, for example) are assigned to a range of z positions in the pixel vertex detector. Processors assigned to a particular range in z receive the data for tracks that originate at a vertex within that range. This can be accomplished with a data-driven architecture that evaluates the apparent origin for each track based on the slope of the track. The purpose is to "funnel" all of the relevant data for an interaction to the same processor, where a trigger decision can be made.

The details of the parallel-pipeline architecture have not been worked out. However, the advantage of this architecture is that the majority of tracks associated with an interaction are funneled to a processor based on the z position of the interaction vertex. This has two consequences. The first is that tracks associated with a primary vertex (call it vertex A), and tracks from B-decay vertices are brought together in a single processor for vertex reconstruction, impact-parameter calculations, and a Level 1 vertex-trigger decision. The second consequence of this architecture is that tracks that occur during the same beam crossing, but are *unrelated* to vertex A, tend to be funneled to other processors. Unrelated tracks

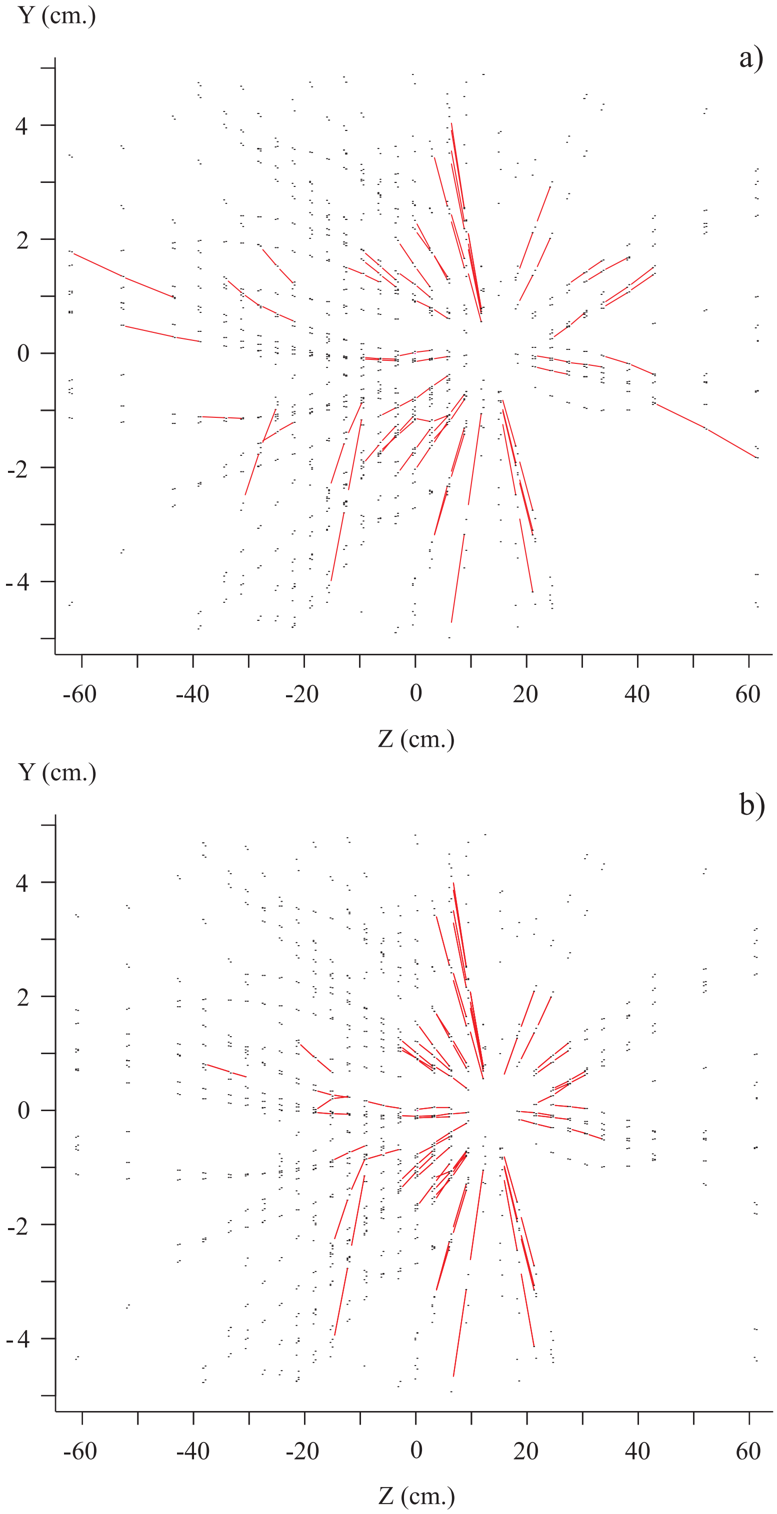


Figure 8.17: Pixel hits (shown as dots) for a $B \to \psi K_s$ event in the $z - y$ projection. The figure shows a) inner/outer tracks (two line segments per track), and b) track segments made up of the three innermost pixel hits for each inner/outer track found in the event.

include those tracks associated with other vertices, and tracks produced by reinteractions in the material of the vertex detector. These tracks are processed independently of tracks associated with vertex A, and do not contribute to the computational load for that vertex.

To conclude, the primary goal of the Megapixel Trigger is to generate Level 1 trigger decisions with high efficiency for B and charm events. A large percentage (more than 90%) of tracks are found *once* as inner/outer tracks. The inner/outer tracks "seed" the track reconstruction, and are used to find the track segments that are made up of the innermost pixel hits. These track segments are close to their point of origin. They are funneled to track and vertex processors by a data-driven architecture. This means that the processing load for a particular vertex is reduced by excluding unrelated tracks from that vertex, and the load for multiple interaction vertices in a single beam crossing is distributed over multiple processors.

8.8 The Global Level 1 Trigger Framework

At the output of the first level trigger there will be a 'global Level 1 trigger' subsystem that will allow us to combine the various individual trigger elements into final triggers (for example, an event with only weak evidence for a secondary vertex but with an indication of a high p_T single lepton might be accepted). It will also have the ability to apply a programmable prescale to the various triggers before they are 'or'ed together to form the final trigger. We will also include the ability to dynamically change the prescales during the course of the store. This device plays a central role in the trigger and data acquisition architecture and is discussed in great detail in chapter 9.

8.9 The Level 2 and 3 Triggers

We have shown that the Level 1 trigger can reduce the trigger rate by a factor of about one hundred. The event rate is then reduced to $\approx$100 kHz in the worst case during the highest luminosity running. While it is technically feasible to move all this data, amounting to a maximum rate of 10 GB/s, to the Level 2 trigger, we discuss below an architecture that allows the Level 2 decision to be made on a subset of the event data, thereby reducing the total bandwidth needed. Only for events passing Level 2 is the complete event assembled. We define further processing done on the complete event to be Level 3. If necessary, the Level 3 software can be quite comparable to the offline analysis in previous-generation experiments.

We anticipate that the data from Level 1 will be dominated by light quark events: the $b-\bar{b}$ production rate is about 20 kHz and not all of these are within the detector or trigger acceptance. The charm cross section is higher, but the triggering efficiency will be poorer by design. Hence, the primary task of the Level 2 trigger will be to reject the remaining light quark events from real collisions at $\sqrt{(s)} = 2.$ TeV. In addition, Level 2 will provide enough kinematic information about the event so that a meaningful decision on the event decay properties can be quickly taken by Level 3.

Level 1 treats each sub-detector completely independently from the others: muon and pixel detectors (Level 1 vertex) for a given crossing are not treated sequentially in the same processor. They are handled asynchronously in different processing units. Events passing the global Level 1 trigger do not come in sequential order. Level 2 decisions are based on partially reconstructed events. By this, we mean that not all information from all sub-detectors is available in the memory of a given level 2 processor. Possibly, the information on a given sub-detector element (i.e., pixel detector or the forward silicon tracker (FSI)) is partial: only a few planes could be used.

While the Level 1 vertex trigger broadly classifies the events in terms of their transverse detachment, looking for a few detached track with moderately high P_t, the task of Level 2 will be to confirm that the event surviving Level 1 cuts do indeed contain a heavy quark decay. Level 3 will classify these heavy Quark candidates in terms of final states, and make a decision according to an exhaustive trigger table.

In this section, we propose a few algorithms for the Level 2 trigger based on the pixel detector, the two most upstream FSI chambers, and the two most upstream straw chambers (STRAW). The basic idea is to firm up the measurement of the transverse momentum of the tracks, and improve the vertex position resolution using this momentum information. The main software tool for this task is the MCFast Kalman filter (or fitter) to (i) test the hypothesis that a track is real and self-consistent and (ii) improve the track parameters. No attempts at doing particle i.d., search for neutral kaons, π^0, etc. are currently proposed at Level 2. Cuts based on the invariant mass of detached tracks could also be successful, but are obviously final-state specific.

8.9.1 Preliminary studies of the Level 2 algorithm

We have simulated a possible Level 2 algorithm. Note that the vertex pattern recognition, currently done as the last stage, could be done at any time after the first stage, as the Level 2 tracks are genuine, complete 3D tracks with a covariance matrix. Level 1 tracks with at least 3 minivectors are used as seeds for the Level 2 pattern recognition. The pion particle identification is always assumed. The Level 2 algorithm works in several sequential phases:

1. Adding pixel hits. The Level 1 pixel pattern recognition has known deficiencies. For instance, a track may cross a ϕ slice boundary and be partially reconstructed. Level 1 algorithms tend to limit the number of hits found on a track to a minimum, in order to gain time in assembling the track. Despite the very low cell occupancy in the pixel detector, there are still a few confused or fake tracks. Adding new hits and performing a Kalman filter allows us to:

    - Reject spurious tracks. About 10 % of the Level 1 detached tracks do not pass the Kalman filter confidence level requirement.
    - Improve the parameters of the track.

- Compute the best possible estimation of the error on these parameters, such that displacement errors at the vertex are correctly estimated. Errors due to multiple scattering are properly taken into account.

2. Adding hits from the straw chamber or the forward silicon tracker. The momentum resolution $\Delta P/P$ using the pixel information only for forward tracks is about 6 %. These tracks are extrapolated to the location of the most upstream tracking station (about ± 1.0 m from the interaction point). Confirming hits are searched for within a window typically 1.0 cm wide. The left-right ambiguities for straw hits are resolved. The second tracking station, located ± 1.5 m from the interaction point, is also used. The momentum resolution, shown in Fig. 8.18, is around 2% for tracks reaching the first or second station. Note that these two detectors do not cover the same rapidity range, tracks reaching the second station tend to have higher momentum and therefore have smaller sagitta and bigger $\Delta P/P$. They are however measured over a longer lever arm. These two effects partially compensate. The transverse momentum resolution also improves, from $\approx 20\%$ at Level 1 down to a few percent.

3. All Level 2 tracks with a good Kalman fit confidence level are considered for the vertex pattern recognition phase. They are required to be "forward", i.e., have straw or silicon strip hits. However, if they point to these detectors, and no hits are found for a few consecutive planes, they will be rejected. The level 2 primary vertex resolution is shown in Fig 8.19.

The detachment criterion at level 2 is based on an exact calculation of the distance of closest approach of the track to the primary vertex and the error on this quantity. The detached track multiplicity, requiring a 5σ detachment, is shown in Fig 8.20, for events passing the Level 1 trigger (based on 2 detached tracks). Note the level 2 detached tracks are not necessarily the same tracks that were called detached at Level 1. A combined Level 1 and Level 2 rejection factor on light quark events of about 1 in 1,000 is achievable, with an acceptance for $B \to \pi^+\pi^-$ of about 40 %. Finally, at this stage, the $B \to \pi^+\pi^-$ mass resolution is about 98 MeV, assuming pion particle identification. The mass resolution is, of course, significantly improved when information is used from all the forward tracking stations.

These Level 1 and Level 2 studies are based on a full pattern recognition code. The Monte-Carlo generator information is used solely to measure the performance of the code, not to select hits on tracks. Future efforts will be concentrated on improving the CPU performance and tuning the algorithm to minimize the computing needed in the Level 2 processors.

8.9.2 Level 3

The final trigger level will combine information from different sub-detectors to further suppress background events. Particle identification will be available at this stage and could be

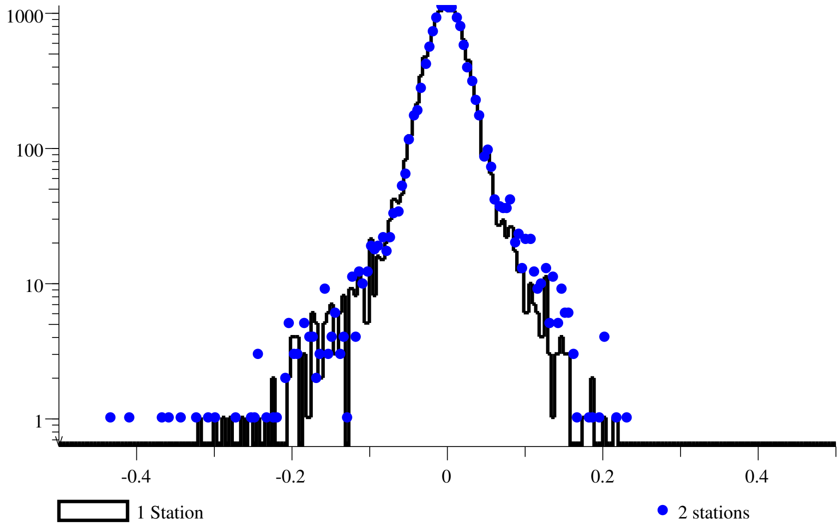


Figure 8.18: The momentum resolution for the light quark sample, at level 2, using the additional pixel hits and the forward silicon strip tracking system. The plot labeled "1 station" refers to track with silicon hits from the most upstream station, the plot "2 stations" uses the two most upstream stations, with at least one hit in either station. The σ of the core distribution is about 2.15 % for "1 station" plot and 2.05 % for the "2 stations" plot.

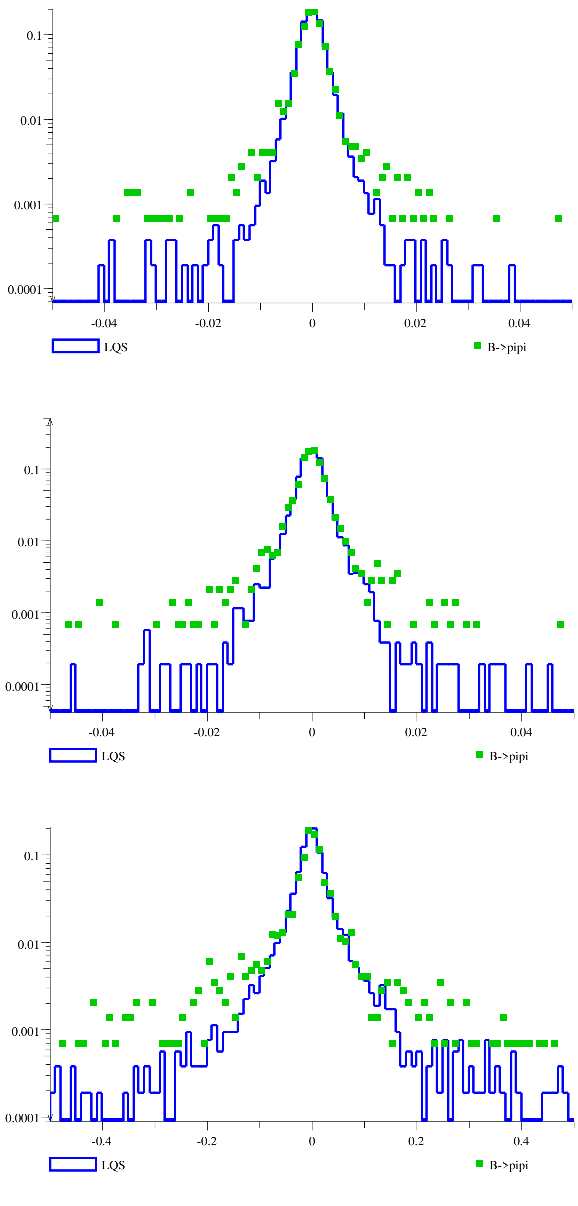


Figure 8.19: The primary vertex resolution along the X (top), Y (middle) and Z (bottom) direction at Level 2, for the light quark sample and for the $B \to \pi\pi$ sample.

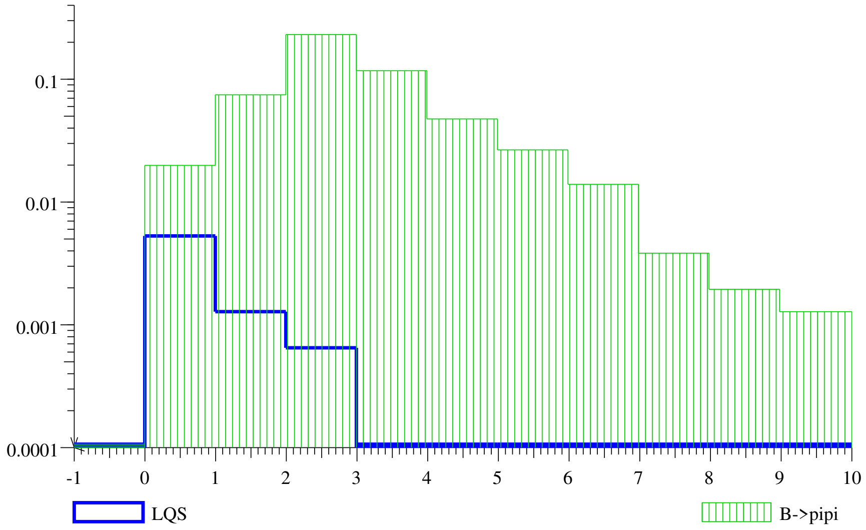


Figure 8.20: The Level 2 detached $(5.\sigma)$ track multiplicity for the light quark event sample and for the $B \rightarrow \pi\pi$ sample selected by the offline reconstruction program. In addition, these Level 2 tracks must have a transverse momentum greater than 0.5 GeV/c. The Level 1 trigger criteria has been applied. The vertical axis is the probability of accepting a beam crossing given the number of detached tracks on the horizontal axis.

used to obtain a very clean direct charm signal for specific final states. In addition, the Level 3 trigger is likely to go through a long list of heavy quark decays to final states of interest. We intend to fully reconstruct the events at this level before the data are sent to a storage system and made available for off-line processing.

Because the event rate surviving this last level may still be a few kHz, the software will probably have to reduce the amount of data per event to archival storage by writing out an event summary which eliminates much of the raw data. The event summary would be around 20 KB so that the output rate could still be as high as 80 MB/s. This results in a dataset size that could be as high as 800 TB/yr. However, initial phases of the experiment will run at much lower rates and will be comparable in dataset size to a current fixed target experiment.

Bibliography

[1] BTeV EOI (May 1997).

[2] These results are based on the work of R. Isik, W. Selove, and K. Sterner,"Monte Carlo Results for a Secondary-vertex Trigger with On-line Tracking," Univ. of Penn. preprint UPR-234E (1996).

[3] P. Lebrun, "On the Vertex Pattern Recognition in the BTeV Trigger," BTeV internal note 97/17 (1997). See also BTeV internal note 99/6 and 99/7. (1999)

[4] D. Husby, P. Chew, W. Selove, and K. Sterner, Nucl. Instr. and Meth. A **383** (1996) 193.

[5] For updated vertex trigger documentation, see http://www-ese.fnal.gov/eseproj/trigger/default.htm and E. Gottschalk, BTeV-pub-99/1, presented at DPF99

Chapter 9

Trigger and Data Acquisition Architecture and Front End Electronics

9.1 Data Acquisition

The BTeV system will digitize, sparsify, and transmit data at the beam crossing rate of 7.6 MHz into off-detector buffer memories. This approach is taken because a large subset of the data is used in the first level trigger and must be digitized in any case and a very sophisticated first level trigger is planned. With this approach applied to all subdetectors, the first level buffers can hold many more beam crossings of data than those found in typical front-end integrated circuits. This means that the decision time of the Level 1 trigger can be extended by as much as two orders of magnitude, if desired, allowing for much more sophisticated trigger processing than would be possible given a fixed Level 1 latency.

Three distinct logical trigger levels are described in chapter 8. From an engineering perspective, there is little difference between Level 2 and Level 3: these levels are both executed in the general purpose processors. The only distinction between these levels is that the start of the Level 3 processing presupposes that ALL data have been transferred to the processor, whereas the Level 2 processing may operate on a subset of the total data. Thus, triggers in the BTeV system are separated into two physical levels, referred as L1 and L2/3. This is typical of most proposed large scale data acquisition systems and acknowledges the greatly increased performance and decreased cost of general purpose processors. In effect, much of the early and intermediate processing in previous trigger architectures has been moved into the first level hardware, Level 1 in our case, while the remaining processing occurs in a greatly expanded array of general purpose processors.

Following the Level 1 trigger, the remainder of the system is sized for a minimum L1 rejection of 75:1, which represents an L1 accept rate of $\approx$ 100 KHz. This rate must be absorbed by the L2/3 processors though a switch. The baseline design assumes that this can be achieved with $\approx$ 4000 general purpose processors of about 1000 MIPS each in the L2/3 array. The use of general purpose processors does not preclude the use of dedicated L2 trigger logic in the form of attached coprocessors.

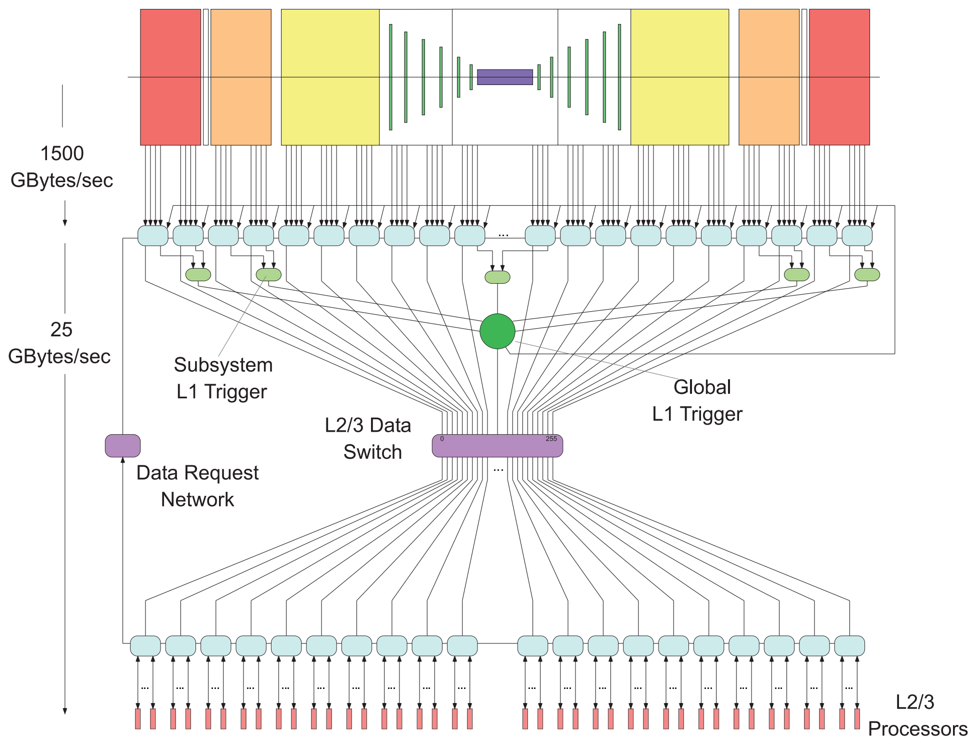


Figure 9.1: Block Diagram of BTeV Trigger and Data Acquisition Hardware

The specifications shown in Table 9.1are used as the baseline for the BTeV data acquisition. These numbers represent both arms of the detector.

event size	200 KBytes
number of detector data links	2048
number of L1 data buffers	2048
number of L2/3 data links	256
number of L2/3 processors	4096

Table 9.1: Estimates of Hardware for BTeV Trigger and Data Acquisition System

The total system buffer memory, assuming 2048 L1/switch input buffers and 256 switch output buffers, is almost 600 Gigabytes. Buffers in the L1/2/3 processors will push total system memory to approximately 1 Terabyte. A block diagram of the trigger and data acquisition system is shown in Fig. 9.1.

9.1.1 Detector Multiplexing

To efficiently balance data rates on the detector data links, a relatively large number of detector channels will be multiplexed into each high speed link. In the case of the RICH subsystem, for example, it is expected that approximately 3000 single bit signals will be multiplexed and encoded to drive one link. The input to this module will depend on the detector subsystem, but the use of programmable logic will allow for some variation in signal widths and sparsification algorithms with common hardware. The detector multiplexer drives the detector data link and receives the coded clock signal for fanout to the front-end components.

9.1.2 Detector Data Links

Each detector data link must transfer approximately 800 MBytes/sec. This rate can be accommodated by a single 10 Gbps fiber, but more likely will be implemented with eight 1 Gbps fibers or thirty-two LVDS copper connections per link. The choice of technology and link speed depends on the physical placement of the first level buffers. The total bandwidth of the detector data links is 1.6 Terabytes/sec.

9.1.3 L1 and Switch Input Buffers

Each Level 1 Buffer must accept data from a subsection of the detector at the 800 MBytes/sec average rate and the data must be held in this buffer pending a Level 1 trigger decision. The BTeV L1 trigger is based in large part on track reconstruction in the vertex detector and decision times may vary over a wide range depending on event complexity. For this reason, it is desirable that the L1 buffer controller support receipt of both asynchronous and out-of-order L1 accepts and rejects. Following an L1 accept, the data are transferred to a switch input buffer. The function of the switch input buffer is to hold data until an L2/3 processor assignment is made, the data have been requested by the processor, and the switch rotation allows transmission of the data. Data requests to the switch input buffer also arrive asynchronously and in arbitrary order.

As a practical matter, the L1 buffer and switch input buffer will likely use the same physical memory, with the data "transferred" between buffers by reassignment of pointers. The size of an individual data buffer is expected to be approximately 256 MBytes and may be implemented using direct Rambus protocol compatible with next generation PC memory architectures.

Data from some detector data links will also be used in the L1 trigger. The L1 buffer module provides a direct datapath from the link receivers to an attached module which may contain local L1 trigger logic. The output of this module connects to the subsystem L1 trigger and finally to the Global L1 trigger.

9.1.4 Ring Buffer Interconnects

Our design approach is sufficiently flexible to accommodate different options for data transmission to the L2/3 processors. If all data is transmitted to the L2/3 processors following an L1 accept, the most natural arrangement of L1 buffers is to place an equal number on each switch input link. If the data is transmitted in steps (staged readout), placement of the L1 buffers should be optimized for the frequency of readout. Using a ring architecture achieves the same multiplexing functionality as a fixed backplane bus, while also allowing reconfiguration of the L1 buffers to match changes in trigger rate, switch size, and access frequency of the buffers.

The ring and switch interconnects utilize 1 Gbps serial links based on the same physical layer standard as Gigabit Ethernet, Fiber Channel, and NGIO. All data are transferred in fixed length packets, similar in size to an ATM cell.

9.1.5 Staged Readout

To reduce the size of the data switch, the Level 2 trigger may operate in stages. In each stage, it would then request only enough information from the switch input buffers to make a decision to continue to the next stage of processing. Events may be then rejected at any stage in the Level 2 analysis, with the remaining data cleared from the input buffers without being transmitted through the data switch.

The reduction in data switch bandwidth requirement is offset by the increase in bandwidth required for the data request messages flowing in the reverse direction. In addition, the input buffer depth must be increased significantly to provide greater latency for event data used in later stages of the L2 algorithm. Buffer management logic also becomes more complicated. The cost reduction comes from the potential decrease in the number of required output buffers. An analysis of the minimum number of required output buffers and the expected fraction of data needed in the average L2 calculation will be performed to determine if the staged readout approach is cost-effective.

9.1.6 Data Switch

A simple, unidirectional, packet synchronous, TDM switch has been shown to be the most efficient for event building applications. The switch configuration follows a standard "barrel shift" rotation and can be expanded to any size in NlogN. The BTeV data switch is expected to require a maximum of 256 channels, which can be implemented using two stages of 16 X 16 switch modules.

The advantage of this switch is that it requires no internal control. Source and destination addresses contained in the data packets are used only to specify buffers on a common ring and are not needed by the switch for data routing. The switch is unidirectional, but packets are recirculated by connecting the last output buffer to the first input buffer in each ring. This provides a mechanism for input to input message transfers (throttle requests to the global trigger), output to output message transfers (L3 accepted events to logging processors) and output to input messages (event requests to the global trigger/event manager).

Because the switch is easily expanded, full implementation can be delayed until the accept rates of the L1 and L2/3 triggers are determined, and the switch sized accordingly. The cost of the switch is not a large factor in the overall system cost.

9.1.7 Data Request Router

If a staged readout is not required, the L2 accept/reject messages can be broadcast to all switch input buffers. For a staged readout, a broadcast router is needed to selectively distribute the accept/reject messages to the switch input buffers. The bandwidth required for these messages can be accommodated by Fast Ethernet (100 Mbps) if multiple requests are buffered in each packet.

9.1.8 Switch Output Buffers

The switch output buffers are similar in function to the switch input buffers. This is the point where all the requested data from a single event is received from the switch and forwarded to a processor. Connection to processors is by way of Fast Ethernet links, with 16 links per switch output buffer. Up to 4096 processors can be connected to 256 switch output buffers, and additional buffers can be added to each ring if the processor array is expanded.

9.1.9 Data Logging

Events accepted by the Level 3 trigger are sent to data logging processors. These processors are attached to switch output buffers in the same way as any other L2/3 processor. Accepted events are placed back on the ring and recirculate through the data switch to the appropriate output port. Given the high trigger rejection, This utilizes approximately 1% of the data switch bandwidth while avoiding the need for a separate data logging network.

9.1.10 Detector Control Links

All detector subsystems transmit data at the crossing rate. This simplifies the delivery of clock and control signals to the detector front-end components because L1 and L2 accept/reject signals are not required at this level. It is assumed that a 53 MHz coded clock will be sufficient to provide crossing and synchronization information. Clock and data recovery should be provided by the detector multiplexing module and then delivered independently to the front-end.

9.1.11 Diagnostics

All system buffers have an Ethernet link for the purpose of sending or receiving data request messages. This provides a convenient path for system diagnostics at little additional cost. We expect to make widespread use of embedded processing and embedded servers in these modules to allow remote test and status monitoring.

9.1.12 General Operation

At startup, buffers can be initialized through the standard Ethernet link. The data switch is initialized by filling the rings with empty data packets, after which it operates in a packet synchronous mode. The front-end systems are synchronized through the clock distribution system.

Event fragments are timestamped with the crossing number which is used for all Level 1 trigger accept/reject messages. Events accepted at L1 are tracked through the remainder of the system by the L1 event number. An L1 accept causes the buffer control logic to move a memory pointer from the crossing number list to the L1 event number list. At the same time, the L1 event number is added to the list of events available for L2/3 processing.

An event request from an L2/3 processor is sent to the Global L1 trigger (or event manager logic attached to the Global trigger). The Global trigger returns a packet containing an L1 event number, along with any trigger information considered useful to the processor. At this point the processor is responsible for disposition of the event. It may request some or all of the data from switch input buffers for use in the L2 decision. Following an L2 accept or reject the processor makes a final data request to all remaining switch input buffers. Buffers then delete or transmit/delete the event data. Each switch input buffer receives exactly one data request message per event.

9.1.13 Component Placement

The detector data links and L1 buffers represent a large part of the overall data acquisition system cost. Placement of the buffers influences the technology and cost of the links. The buffers will occupy the equivalent of 128 6U subracks. The L1 subsystem triggers should be located in the same area as the L1 buffers. Ideally, these components would be placed outside of the collision hall to facilitate maintenance and to allow reconfiguration of rings based on changes in trigger rate and buffer access frequency. This requires the use of optical links. Placement of all other data acquisition components (switch, output buffers and processors) is not critical, since there are a relatively small number of interconnects.

9.2 Global Level 1 Trigger System

The picture of the Level 1 trigger that we have presented focuses on the Level 1 Vertex Trigger which is based on the pixel detector. The actual Level 1 trigger will be more complex. This complexity is managed by the Global Level 1 Trigger, GLV1.

First, there will be several different vertex related triggers. We have described a final selection based a minimum number of tracks, N, each of whose normalized impact parameter with respect to the primary vertex exceeds a value M. In fact, there will be a variety of triggers, some accepting events with a few tracks with large detachments and others accepting events with more tracks that have smaller detachments. We also want to record a sample of events that would otherwise have failed in various ways to understand how the efficiency of the trigger "turns on". This last group of triggers may be prescaled.

Second, there will be triggers that involve other detectors. We plan to have a "standalone dimuon trigger". In addition to improving the overall efficiency for triggering on states containing J/ψ's, this trigger will permit an independent crosscheck on the efficiency of the vertex triggers.

Third, there will probably be triggers which combine information from more than one detector at GLV1. For example, a single high p_t lepton trigger would be interesting but may go at too high a rate to be acceptable. However, the global Level 1 trigger should permit one to accept high p_t single muon triggers if the event also satisfies a relaxed vertex requirement – perhaps one track with a large impact parameter. (At Level 1, one would not know that the high p_t muon corresponded to the single high impact parameter track).

Fourth, there will be special triggers. One example will be a variety of minimum or low bias triggers which will be heavily prescaled. Another example will be special alignment and calibration triggers. We will certainly collect special triggers or use minimum bias triggers to do the quasi real time alignment of the pixel detector.

Fifth, we will probably record a prescaled sample of various 'failures' and 'timed out' events to make sure they are not introducing some kind of physics bias.

There will be one or more primary physics triggers which are not prescaled. Global Level 1 must have the ability to prescale other less important physics triggers or calibration-type triggers. We also need the ability at Level 1 to adjust the prescales dynamically. If the luminosity is not 'leveled', then we want the ability to reduce the prescales on some of the triggers as the luminosity falls. We also want the ability to increase the number of alignment triggers taken at the beginning of a store and to reduce it once enough events have been collected to establish initial alignment constants for the store.

Given these requirements, we have the following picture of the operation of the Global Level 1 trigger. The GLV1 receives 'trigger packets' from each trigger processor. The packet contains a header which has 'trigger primitives' whose format is known to the GLV1. These packets arrive asynchronously. The GLV1 buffers these packets until it receives all the packets it is supposed to receive for a crossing or until a timeout occurs for that crossing. In normal operation, as soon as all packets are received the GLV1 inspects the headers and forms all the various triggers from truth tables that have been downloaded to it. It then applies the appropriate prescale to each trigger and takes the OR of the result. If any of the triggers is true, it issues a Level 1 accept. This results in the event data being transferred to the so-called "switch buffer" for transfer to the Level 2 trigger. If no trigger is satisfied, the GLV1 issues a Level 1 reject. This eventually results in the Level 1 event buffers being freed to be used for other events. In general, the trigger processors which are talking to the GLV1 will have timeouts which are less than the GLV1 timeout. If the maximum allowed time for the Level 1 trigger primitive arrival should expire before a trigger packet from a processor arrives, this generally indicates a problem. A prescaled sample of these events can be recorded for further analysis and diagnostics. Then the rest of the GLV1 decision proceeds as normal.

The fact that the GLV1 receives data packets from the trigger processors allows the processors to send additional data with the trigger primitive information. This could include diagnostic information. It could even include 'event information.' It is quite likely that we will want to write some of the results of the Level 1 trigger calculations out as part of the event going to the Level 2 trigger. The data may be used for diagnostic purposes or may even be used to 'prompt' the Level 2 calculations by directing it to process particular tracks, for example. There are several ways within this architecture to write such data to Level 2. Sending it as part of the trigger processor packet to GLV1 is only one method.

As we have developed the trigger and data acquisition architecture, we have begun to merge functions that are really 'event supervisor' functions into the GLV1. At present, the plan is for the GLV1 to maintain a list of events which have satisfied the Level 1 trigger. When a L2/3 processor becomes available to process a new event, it sends a request for an event to the GLV1. The GLV1 then sends a data packet to the processor which contains the event number to analyze. It also includes the status of all the triggers for that event so that the processor can execute different code depending on how the event triggered. It may also send more extensive event data from the trigger processors. Once the GLV1 has sent an event number and associated data to a Level 2/3 processor, it is done with that event.

The GLV1 is the natural location for accumulating statistics on the performance of the Level 1 trigger. It will periodically send this data up to online analysis computers.

The GLV1 will also play a role in data throttling – that is trying to reduce the rate of data sent to the upper trigger levels if they begin to fall behind and the switch buffer memory approaches its capacity. The 'switch buffer memory' which holds the event data while waiting for a L2/3 processor to become free will be able to issue a 'full warning' to GLV1 when it becomes 90% full. The GLV1 can attempt to undertake some action to alleviate the problem. For example, it could begin to adjust prescales to reduce the total trigger rate. It could also stop sending events. When the 'switch buffer memory' is completely full, it will issue a 'full' indication. The GLV1 will stop issuing Level 1 accepts until this condition is cleared. How exactly the throttling will propagate back to the Level 1 buffers is not yet clear. What is clear is that this will result in deadtime and GLV1 will have to keep track of how much deadtime there is for purposes of normalization and for diagnostic purposes.

We are also discussing 'partitioning.' Partitioning refers to the ability to divide the Trigger/DA system into more than one semi-independent piece, each of which approximates a completely independent trigger/DA system. This capability is especially useful in the debugging stage of an experiment. In this stage, some subsystems may not be installed, some may not yet be operational, some may be down for maintenance, and some may be hanging or crashing. It is important to be able to work on one or a collection of subsystems independently of the state of the others. The BTeV architecture permits a reasonable level of partitioning. The ability of Level 2 processors to request subevents rather than whole events lends itself to partitioning. On the other hand, the requirement that an event be given to one and only one processor creates some difficulties for partitioning.

In our current plan, the GLV1 will maintain separate trigger lists for different 'event' types. It will also maintain separate lists of event numbers for each triggering event type.

It will have an algorithm for assigning events satisfying triggers in more than one list. The algorithm will proably be based on the desired event rates and priorities for each event type. Level 2 processors will request an event of a specific type. The GLV1 will return an event number from the list for that type. The Level 2 processor can then request the subevents that it wants to analyze. If the 'switch buffer' is configured so that it waits a certain minimum interval before it will send an event so that it is sure that it has the whole event, and if it can send a message saying that it doesn't have a requested subevent, then the fact that some (presumably) unneeded subevents are missing does not cause a problem. With this scheme, subgroups can have their own triggers and select the subset of the data they want to read quite independently of each other. However, since each event will go to only one processor, some intelligence in GLV1 will have to decide how events satisfying the requirements of multiple trigger lists should be apportioned.

9.3 Level 2/3 Trigger

The Level 2/3 Trigger is a processor farm. The baseline design of the farm calls for 2000-4000 general purpose processors, such as INTEL/Pentium PCs, running Linux. Each processor runs the same analysis code, and performs the necessary analysis for a single event before it requests data for a new event. Requests for new data are sent to the Global Level 1 Trigger, GLV1, which coordinates the initial data requests for each processor in the Level 2/3 Trigger. The Level 2/3 analysis software for BTeV performs data analysis that is usually part of the "offline analysis" in other experiments. Consequently, the Level 2/3 Trigger can select events using cuts that are usually reserved for offline analyses, and can transform event data into a condensed format (usually referred to as a DST format) before sending the data to event-logging processors.

An important aspect of the Level 2/3 Trigger architecture is that each processor is capable of making trigger decisions for both Level 2 and Level 3. The distinction between Level 2 and Level 3 depends on the memory buffers in which the data for an event reside, and will be described below. A second aspect of the trigger architecture is that each event is assigned to a single processor in the PC farm, so that a single processor selects or rejects an event at Level 2 and, if it passes Level 2, at Level 3.

The Level 2 analysis for an event begins when an idle processor sends a request to the Global Level 1 Trigger (GLV1) for a new event. The GLV1, which maintains a list of events that have satisfied the Level 1 Trigger, responds by sending a data packet to the processor. The data packet contains the event number, and Level 1 trigger information. The processor begins the Level 2 analysis of the event based on the Level 1 trigger information, which summarizes all of the trigger conditions that were satisfied at Level 1. In addition to the trigger information, the processor usually requires data from various detectors to complete the analysis and arrive at a trigger decision. The data reside in switch buffers. The processor requests data from these switch buffers directly. For example, the processor may send a request for data recorded by the forward tracking chambers that are closest to the dipole magnet. The requested data are sent from the switch buffer to the processor's memory and the memory it occupied in the switch buffer is freed. At this point, part of the event is in the processor memory and the remainder is in the switch buffer. In this example, these data are used to improve the momentum determination of tracks found by the vertex detector, so that the processor can calculate more precise impact parameters to select tracks coming from B decays. After the data are received and processed, the processor may request additional data (such as from the other forward trackers or the RICH detectors), or complete the Level 2 analysis by sending a trigger decision to the switch buffer.

A processor that has completed the Level 2 analysis for an event sends one of two trigger decisions to the switch buffer. If the event is rejected, the data for that event are removed from the switch buffers, and the processor is ready to process a new event. If the event is accepted, then the switch buffer sends *all* of the data for that event to the processor for subsequent Level 3 analysis, and removes the data from the switch buffers. This marks the beginning of the Level 3 trigger. At this stage, the data for an event reside entirely in the memory of one processor (the same processor that performed the Level 2 analysis for the event), and nowhere else.

A processor that performs the Level 3 analysis for an event can perform a complete analysis using all of the data that are recorded for that event. The data analysis is comparable to "offline" analyses in other experiments, and includes track reconstruction, vertex reconstruction, muon identification, and possibly charged hadron identification. Events that are rejected by a processor at Level 3 are simply terminated by that processor. Events that are accepted are sent to data-logging processors.

9.4 Data Logging and Online Event Analysis

Each event is handled by only one Level 2/3 processor. (i.e., an event is not scattered across processors nor is it copied to more than one processor). All accepted events need to be moved out of the L2/3 processors and onto tape for long term data storage, and a fraction of these needs to be made available for online monitoring (by "consumer processes").

Events coming out of the Level 3 processors will not be raw data, but already processed data (except possibly for a highly prescaled sample used for monitoring the data reduction algorithm itself). This should shrink the event size from 200 kbytes to 50 kbytes. Assuming an event size of 50 kbytes, and a 4 kHz event rate to the loggers, the data rate to tape is on the order of 200 Mbytes/sec. The consumers themselves should require an additional 5-10% of throughput.

Since the data logging rate out of each Level 3 processor is small, it is less cost effective to attach logging media to these individual nodes than to provide a small number of separate logger nodes. The necessary bandwidth to the logger nodes is also small (1%) compared to the raw data coming up the Level 2/3 farms, so the same switch could be used to pass the event to the logger nodes as well. The switch buffers are small ($<$ 100 bytes) relative to the DST event, so the L2/3 processors can send single events to the logger nodes without incurring any additional penalty. A single event is routed to the next free logging node.

The number of logging nodes themselves is clearly a function of the data rate. It is assumed that data will first be buffered to disk before being written to tape. This serves a dual purpose; bufferring enough data before a transfer to keep the tape drives streaming, and protecting against tape media errors and tape drive failures. With current market technology, a single processor node can handle 10 Mbytes/sec to tape including the initial disk write. We will therefore require 20 data logging nodes.

Consumer events should not interfere with data logging rates and additionally may contain rejected events. The consumer processes will therefore run on separate nodes from the logger processes.

9.5 Front End Electronics

In this section, we discuss various aspects of our front end electronics. We start with basic infrastructure such as racks and then describe our thinking about the front end electronics for the forward tracker, the ring imaging Cherenkov counter, the electromagnetic calorimeter and the muon detector. Electronics for the pixel detector is completely covered in chapter 5. We conclude with a brief discussion of software support for electronics development.

9.5.1 Infrastructure and Other Support Systems

A goal of the BTeV experiment is to minimize hardware and software development. This goal can be accomplished by:

- Choosing cost-effective, commercially-available hardware and software whenever possible;
- Using hardware and software developed elsewhere whenever possible; and
- When hardware and software needs to be developed for BTeV, develop them in such a way that they can be used in the maximum number of BTeV front-end, trigger, and data acquisition systems.

Commonality in hardware and/or software when possible in BTeV's front-end, trigger, and data acquisition electronics will save not only substantial amounts of money but valuable personnel resources. A BTeV workshop was held recently to understand

- what already-developed electronics and software can be used in BTeV and
- where hardware and/or software commonality is possible in BTeV's front-end, trigger, and data acquisition systems.

The following subsections describe our choices for some infrastructure items and other support systems.

9.5.1.1 Electronics Packaging - Subracks and Related Infrastructure

Using a mechanical packaging and bus protocol standard that is commonly used internationally offers several advantages over the lifetime of an experiment. Cost and personnel time savings are the two most important advantages. Standards are well-documented and stable. Various electronics and mechanical components supporting the standard are commercially-available from industry. For example, subracks, processor modules, (module size) adapters, extenders, subrack power supplies and interface ICs are commercially available items not requiring development. This leaves us free to concentrate only on the development of specific front-end, trigger, data acquisition, controls and monitoring electronics. By choosing a commonly-used packaging and bus protocol standard, implementers, as well as people who maintain the electronics over the lifetime of the experiment, have a mature development and testing hardware and software environment. If the standard has been used in previous experiments at Fermilab, we can realize considerable cost and personnel time savings because local expertise is already available and infrastructure is in place.

We have chosen the VME64 Extensions (VME64x) subrack mechanical packaging and bus protocol standard as the baseline implementation for the BTeV front-end, trigger and data acquisition electronics. This standard is an American National Standards Institute (ANSI) standard, ANSI/VITA 1.1-1997. The size of VME64x module chosen is 9U x 400mm. This standard, its physics community VME64xP extension, and the 9U x 400mm module size are being used for all of CDF's and a majority of D0's Run II electronics. Over 250 VME64xP subracks have been purchased by CDF and D0. Other standards used in conjunction with this standard are:

1. IEEE1101.10 - the subrack module mechanical standard for VME64x;
2. IEEE1101.11 - the subrack rear I/O board (transition module) mechanical standard for VME64x;
3. ANSI/VITA 1.3-1997 - the VME64x 9U x 400mm module format standard;
4. ANSI/VITA 25 - VISION, a software VME API (subroutine calls) standard; and
5. ANSI/VITA 23 - VME64 Extensions for Physics (VME64xP), a VME64x-compatible standard which extends VME64x by, for example, specifying 7U and 10U subrack implementations and specifying mandatory control and status registers.

Advantages of these standards include:

1. Substantial increase in subrack user-defined I/O pins - a must in physics applications; low-noise, high-density 2mm metric connectors;
2. Subracks with built-in bussing of six standard voltages and four user-defined voltages - no kludged wiring for bringing power supplies for analog or ECL electronics;

3. Inter-module and subrack EMC and EMI protection - protects sensitive analog electronics from electrical interference from other modules in a subrack or outside noise sources; protects electronics near the subracks from noise generated by subrack modules;

4. Module electrostatic discharge (ESD) protection - protects modules from electrostatic damage during subrack insertion or extraction;

5. Injector, extractor and locking handles - provide a simple method for inserting modules into and extracting modules from subracks; needed because of the increased forces due to the substantial increase in I/O and other module connector pins;

6. Board and subrack keying - protects modules and transition modules from being inserted into incorrect slots potentially damaging the modules; and

7. Optional subrack 48 volt power bussing allowing the use of on-board DC to DC converters to power very sensitive analog electronics.

9.5.1.2 Electronics Packaging - An Alternative

An alternative method of packaging BTeV electronics has been proposed. This method draws on products developed in the world-wide, very price-competitive, PC packaging industry. No communications busses would be used. Slow controls for component initialization, downloading and monitoring would be done by interconnecting Ethernet between PC packages. Electronics would thus not be modularized but, for example, each set of front-end electronics would be housed in its own PC package. Power to that package would be via a 110 VAC line cord and power supplies within the package. This option potentially offers cost savings over bussed systems but would require the development of a completely new infrastructure for hardware testing and software development.

9.5.1.3 Data and Clock/Timing Links and Link Interfaces

Possible and probable areas of hardware commonality in data and control/timing links and other components include:

- Data links from front-end electronics to Level 1 buffers;
- Data links from the Level 1 buffers to the event building switch;
- Data links from the event building switch to event buffers preceding online processors;
- High-speed front-end data links from on-detector electronics, possibly rad-hard, to local electronics;
- Low-noise data links from on-detector electronics, possibly rad-hard, to local electronics;

- Data links to Level 1 trigger subsystems and from those subsystems to the Global Level 1 trigger;
- Control/timing links to front-end electronics and trigger subsystems; and
- Level 1 buffer electronics.

Two ongoing developments and one existing Hewlett Packard (HP) data link series of products should satisfy most if not all of the data and control/timing link applications given above. HP's G-links are transmitters and receivers which operate up to 1.5 Gigabits per second over fiber or copper. CERN is developing a G-link compatible radiation-hard driver for use in its LHC CMS (HCal and ECal) and ATLAS experiments. Fermilab is assisting that development, and we plan to use it for pixel data readout links, and possibly other subsystem data readout links. CDF SVX Run II electronics required the development of a radiation-hard, low-noise transceiver. This transceiver will go into production soon. It is compatible with the IEEE-1596 LVDS (Low-Voltage Differential Signalling) standard developed for the Scalable Coherent Interface standard, IEEE Standard 1596-1992. This standard is actually a differential current standard intended to provide low-noise transmission of digital signals at very high speeds for a few tens of meters maximum.

We will adopt a minimum set of "BTeV data link" standards and, hopefully, a single "BTeV control/timing link" standard. The types of "BTeV links" will include:

- Serial optical G-link 1.3 Gigabit per second data link transmitting up to 20 bits every 1/53MHz (radiation-hard and non radiation-hard); and
- Parallel copper LVDS-like to LVDS 53MHz data link (radiation-hard and non radiation-hard).

The parallel copper LVDS-like to LVDS data link would be used to move data a short distance from the detector to an area near the detector where electronics is more accessible. It provides for relatively low-power, low-noise data transmission off the detector. An alternative for this application is an array of Vertical Cavity Surface Emitting Lasers (VCSELs) and an optical ribbon cable link. VCSELs are inherently radiation hard. If this approach is cost effective and does not add more mass than desired, only minimal electronics need be on or near the detector. All the other electronics (e.g., data multiplexers, Level 1 buffers, trigger electronics, etc.) can be in the counting room and thus very accessible.

Our front-end and trigger subsystem implementers will easily be able to interface their electronics to BTeV standard readout and data acquisition electronics and control/timing signals. This will be accomplished by providing them with data and control/timing link daughterboards. Along with the daughterboards, the implementers will be provided with mechanical and electrical specifications for the daughterboards and component and artwork specifications for circuitry on their boards used to interface to the daughterboards. The following daughterboards are under consideration:

- Serial optical G-link or G-link compatible data link transmitter daughter-board;

- Parallel copper LVDS-like data link transmitter daughter-board; and
- Serial optical G-link data receiver control/timing link daughter-board

9.5.1.4 Slow Controls and Monitoring

In an experiment the size of BTeV, several hundred devices need to be controlled (e.g., high-voltage systems, laser pulsers, interlock systems, etc.). Several thousand 'points' (e.g., power supply voltages, temperatures, gas mixes, interlocks) need to be monitored at regular intervals. Alarm or caution situations need to be established via hard-wiring or in software for equipment and/or personnel safety. Past experiments at Fermilab have chosen to use commercial or in-house-designed hardware and in-house- developed software for slow controls and monitoring systems despite the fact that total process automation solutions (hardware, software and user interfaces) are provided by a multitude of companies worldwide. The beamline cryogenics controls and monitoring system was the first system at Fermilab to use an industry total process automation system. Their hardware controllers and input devices were purchased from Moore Products and compatible software and graphical user interfaces from Intellution. The success of this system was remarkable. Mechanical engineers bought it and implemented it with no Fermilab professional software or electronics engineering help. Significantly fewer people were required to operate the system. As a result CDF will use an identical system for its Run II slow controls and monitoring system. BTeV's baseline slow controls and monitoring system is based on this system.

9.5.1.5 Clock and Timing Distribution

The BTeV experiment is different from all other current and past Fermilab experiments in that a) tracking (pixels) will be used in the lowest level trigger and b) all front-end subsystems will be read out into Level 1 buffers at the 132 nanosecond bunch crossing rate. Most front-end subsystems will require only a bunch crossing clock and a 53 MHz accelerator clock. Some will operate asynchronously not needing a bunch crossing clock. For Run II, CDF has used a modular subsystem for distributing clock and timing information to its front-end and trigger subsystems. Differential ECL signals over high-quality cables are used for long-distance runs from the clock/timing sources to the collision hall. Low-current differential signals (LVDS drivers and receivers) are used to transport clock and timing signals to on-detector front-end electronics and counting room electronics. We will try to use as much of the work done by CDF as practical for its clock and timing distribution subsystem.

An alternative approach to sending clock and timing information to front-end and trigger subsystems is used in the Run II CDF SVX system. A conditioned accelerator clock (with jitter essentially removed) is used to clock timing data (e.g., bunch crossing) into HP G-links via optical cables. Optical splitters can be used to distribute the clock and timing information. At the output of the G-link receiver, the clock is rederived from the output data strobes of the G-link receiver. The bunch crossing timing signal is received as a G-link receiver data bit.

9.5.1.6 Electronics Cooling and Rack Protection

Several systems for cooling and protecting electronics racks exist at Fermilab. Assuming 9U x 400mm VME subracks, our cooling system will be designed so that the exhaust air from a VME subrack that is dissipating 2500 watts will not exceed 95 degrees Fahrenheit. The fans will minimally provide 400 linear feet per minute of air flow. The air/water heat exchangers will reduce the temperature of 95 degrees Fahrenheit input air to 85 degrees Fahrenheit at the above air flow. There is no reason we can't use either CDF's or D0's Run II electronics cooling and rack protection system. This system will be connected to the slow controls and monitoring system described above.

9.5.2 Forward Tracker

9.5.2.1 Straw Tracking

The baseline design has seven stations of straw tubes in each arm. The three stations closest to the collision point on each side of the IR have inner regions instrumented with silicon strip detectors. Each straw tracking chamber at a station consists of six independent modules comprising the three views, for a total of 84 modules. As described earlier, we expect to use one of the U. Penn. ASD chips together with a new TDC to process straw hits. The TDC's may be located on the same printed circuit boards which hold the ASD chips, or on separate boards separated by short copper ribbon cables. Zero suppressed data from a group of straws will be read out to the Level 1 buffers on gigabit optical fiber links.

The straw tracker data output will consist of wire number and TDC data with a resolution of approximately one ns. Preliminary occupancy studies indicate an average of 70 hits per station per event or about 1000 system hits per event. The peak should be 5000 hits per event. The data for each hit include station ID, layer and view ID, Wire ID, and the TDC data for the leading edge; this should fit into three bytes (24 bits). For this estimate, a 4-byte data word will be used. Therefore, the system data will be 4 Kbytes per event, typical, and 20 Kbytes maximum. With an average of two events per crossing, there are 8 Kbytes per 132 ns or 61 Gbytes/sec. If the maximum event size is used to specify the system bandwidth, each of the 84 modules will require 30 Gbit/sec fiber links, for a total requirement of 2520 Fiber Optic links for the straw system.

9.5.2.2 Silicon Strip Tracking

The first three forward tracking stations in each arm will contain silicon strip detectors covering the inner 24 cm by 24 cm area. The silicon strip detectors will be 100 micron pitch sensors in two double sided planes comprising four views, X, Y, U, and V. There will be 57,600 channels of data from six stations of four views. The readout electronics for this system has radiation tolerance requirements. The circuitry design is simple except for the fast readout time necessary.

The silicon detector data will consist of only hit location information, with no analog information. Strips with hits over threshold are read out to the Level 1 buffers. With occupancy of 1%, there would be an average of 600 system hits per event. The peak occupancy is unknown at this time but will be estimated for now at 4%. The data for each hit would be an address containing station ID, view ID, and strip ID. This should fit into two bytes (16 bits). The system data would be 1.2K bytes per event, typical, 4.8K bytes maximum. With an average of two events per crossing, there are 2.4 K bytes per 132 ns, or an average of 18 Gbytes/sec. The maximum estimated rate is 72 Gbytes/sec. Each silicon station will produce 3 Gbytes/sec average, 12 Gbytes/sec maximum. Since there are four views per station, each view will require 24 Gbit/sec fiber optic links. For mechanical reasons, each view will be split at least in half and possibly in quarters. Providing each quarter with eight fiber optic links will allow for peak readout rates. This results in a silicon system total of 576 fiber optic links. Using the option above to move the data off the detector with copper LVDS links requires about four more subracks away from the detector in the collision hall.

9.5.3 Ring Imaging Cherenkov Counter

Simulations of the RICH detector indicate that ~ 3.1 m^2 should be covered with photodetectors per arm. The baseline design assumes use of multi-anode PMTs. With close packing of the PMTs, we expect to deploy 4,600 multi-anode PMTs in each arm. The resolution of the RICH detector is based partly on the segmentation density of the PMTs. Hamamatsu currently offers multi-anode PMTs with several choices of segmentation while the physical size of the tube remains the same (26×26 mm^2):

- a 4 channel tube (2×2);
- a 16 channel tube (4×4); and
- a 64 channel tube (8×8).

The mix of these tubes is physics and budget driven, and is under study. The photo-detection area populated densely with C_4F_{10} photons (~ 1.22 m^2 per arm) will need to be covered with the finest segmentation (1,800 8×8 tubes per arm). The rest of the photo-detection area, populated mostly by aerogel photons, will be covered with the coarser segmentation, at present assumed to be 4×4 tubes (2,800 per arm). With this segmentation the channel count is $160,000$ per arm, $320,000$ total.

Simulations also indicate that for $\beta = 1$ tracks the number of detected photons per ring will be about 50 for gas rings and about 9 for aerogel rings. For a generic $b\bar{b}$ event, we expect each arm to produce ~1000 hits from the gas radiator and ~220 hits from the aerogel radiator. These are underestimates, since the simulation does not include secondary interactions. The probability of 2 photons striking the same the same anode cell is rather low, so it is valid to say that the number of channels fired is approximately equal to the number of photons produced by each radiator.

The PMTs need high voltage of the order of 1kV. Because of the large number of tubes, it is likely that several tubes will share the same high voltage line.

Read-out of multi-anode PMTs will be binary (i.e. we will detect only which anode cells have been fired). The scheme for setting discriminator thresholds needs to be studied. It will depend on gain uniformity of the individual channels, the signal to noise ratio, and on the design of the HV distribution system.

Initiation of a new design project to design 3 styles of front-end boards (a 4, 16, and 64 channel version) is attractive. However, the BTeV Muon readout system is already deploying a PMT readout system based on the ASDx series of ICs developed by Mitch Newcomer's team at the University of Pennsylvania. Each ASDx IC incorporates 8 channels of amplifier, shaper and discriminator in a single rad-hardened ASIC. The RICH system will use a read out system similar to the Muon readout system. The common system components can only simplify design, testing, sparing, installation and support. The Muon system deploys a 4 and 16 channel front-end card, a 64 channel card is needed (incorporating eight 8-channel ASDx ICs). See the Muon system description for a details.

For planning purposes, the 320,000 readout channels routed through 8 channels/IC ASDx ICs implies 40,000 ASDx ICs in total (two arms). They would be installed on 3 types of front-end boards corresponding to the tube segmentation. At one front-end board per PMT, there will be $\sim$9,200 front-end boards, mounted directly behind the PMT bases.

The ASDx ICs require no clocking, and their output signals are low-voltage differential signals (LVDS) which are synchronized as they are latched into Doubler/Serializer boards.

The baseline muon read-out design does not sparsify the data either at the front-end electronics or at the Level 1 buffers. For common electronics to be used between the Muon system and the RICH system, provision for data sparsification will be added to the system.

The figure in the Muon section illustrates the proposed block diagram of the Muon front-end electronics subsystem. The Doubler/Serializer boards will be located outside the detector, but as close as possible to minimize the length of the cable connections from the front-end PMTs. Approximately 3,330 Doubler/Serializer boards are required. Since the Muon system uses 720 such cards, the 4,000 card total is dominated by RICH detector needs.

The output of the Doubler/Serializer boards will be sent to the data acquisition system's Level 1 buffers. The electrical protocol of the serial link to the Level 1 buffers can be differential LVDS or a higher-speed serial optical link. The data can be multiplexed into a single data stream to a high speed serial optical link ($>$ 1 Gbit/s) with headers containing time stamp and channel address, sent to the switched data stream.

9.5.4 Electromagnetic Calorimeter

Our $PbWO_4$ crystal calorimeter is very similar to the KTeV CsI calorimeter. The momentum range of photons that need to be detected is from about 1 GeV/c to 70 GeV/c in both KTeV and BTeV. For both experiments, energy resolution is of prime importance. The light output of the $PbWO_4$ crystals is expected to be only 2.5 times lower than that for CsI crystals. Furthermore, the decay times of both scintillators are comparable, less than 100 nanoseconds for $PbWO_4$ crystals and less than 76 nanoseconds for CsI crystals. The similarity of the two systems and the excellent performance reported by KTeV makes a good argument why we should use the technology developed at KTeV in the BTeV experiment.

The KTeV readout uses special Hamamatsu phototubes, with low gain ($\sim$4000) and good linearity. The tubes (YA2117s) act like a current source which is linear to better than 0.02% for a tube output up to 30 milliamps. Even this small non-linearity can be corrected. The phototube output goes to a Fermilab custom ASIC called the QIE5 (The entire system is called QIE) This IC is used in conjunction with an 8-bit FADC to digitize the output pulse at a fast rate with variable resolution over a 16-bit dynamic range. It accomplishes this by using eight nonoverlapping binary scaled ranges. The device is pipelined so it is deadtimeless.

The output of the phototubes will be connected to a PMT digitizer chassis or subrack (containing the QIE chip or similar device) via time-matched ribbon cable coax. The decision was made to mount the QIE electronics in a separate chassis rather than being mounted on the phototubes due to the radiation levels in the area of the PMTs and the limited space available behind the crystal wall (20,000 - PbWO4 crystal scintillators). Assuming the $PbWO_4$ crystals are 2.47 cm$\times$2.47 cm$\times$ 22.5 cm, a 20480-channel crystal wall approximately 12 ft by 12 ft will be formed in the area near the beam line and requires floor space nearby for either 40 6U subracks or 80 chassis of readout electronics mounted in 10-15 racks. Data from a number of phototubes will be multiplexed and zero suppressed by concentrator modules located in the same racks. These modules will send data to Level 1 bufferers using BTeV standard optical fiber links.

9.5.5 Muon Detector

In order to read out 69120 channels we will use 8640 ASD chips. Each front-end board will host two of ASD chips. Hence the total number of front-end boards is 4320. The output signals are differential LVDS. The ASD chips are amplifiers and discriminators. They provide an asynchronous LVDS output. No clock is used in this chip. The data must be synchronized when it is latched into the "Doubler and Serializer" board. The Doubler and Serializer boards will be located off the detector, as close as possible to minimize the cable connections from the front-end. The function of this module is to multiplex the signals coming from 12 ASD chips. 720 Doubler and Serializer modules are required.

The output of the Doubler-Serializer boards will be sent to the Muon Trigger and DAQ system. The Muon Trigger will receive only one third of the total data, since the trigger is based on these data. The electrical protocol of the serial link to the Trigger and DAQ can be LVDS differential or RS485 over twisted pair cable.

For compatibility reasons the muon trigger and DAQ memory buffers to store the information can be designed on the same kind of boards selected for other BTeV subsystems. VME64 is a potential candidate. The number of boards will depend on the selected standard.

9.5.6 Software for Module and Subsystem Development and the Online Data Acquisition and Readout System

BTeV front-end and trigger subsystem developers can use existing software for developing modules and subsystems. An infrastructure for testing VME-based boards has been developed for CDF, and is the foundation for testing CDF's Run II SVX, Muon and Calorimetry system hardware. This framework, called CDFVME, is Java-based code that communicates to the test subrack(s) via CORBA, a remote procedure-like software tool. The CORBA implementation is ROBIN, and requires a VME processor in the subrack running the VxWorks operating system. VxWorks is the real-time operating system that will be supported at Fermilab throughout Run II for both CDF and D0.

The CDFVME software consists of a series of Java classes to interface to generic VME boards (e.g., read and write register commands), plus templates that users can customize for their specific board. CDFVME also contains a framework in which to run a series of tests in batch mode, and cycle the readout code. Error logging is also provided.

CDFVME uses FISION, the Fermilab supported implementation of the ANSI standard software (ANSI/VITA 25) called VISION to communicate across the VME backplane. FISION is currently supported on Motorola MV16x and PowerPC microprocessor VME CPU modules. CDFVME was jointly developed by Yale, the Online and Database Support Department of the Computing Division and CDF.

The choice of downstream L2/3 processors in BTeV will be driven by price/performance issues. Current market trends indicate that the processors will be some type of PC processor running Linux. Linux PCs will be used for both CDF and D0 Level 3 online processor farms.

Licensing costs and various support issues are generating interest for software implementers to evaluate real time operating system alternatives to VxWorks. Real-time Linux is a possible contender for post Collider Run II data acquisition systems.

Development code for data acquisition should move in the direction of modern programming standards. Run II software has already been designed for object oriented languages such as C++ and Java. CORBA is the standard protocol to communicate between objects on remote nodes.