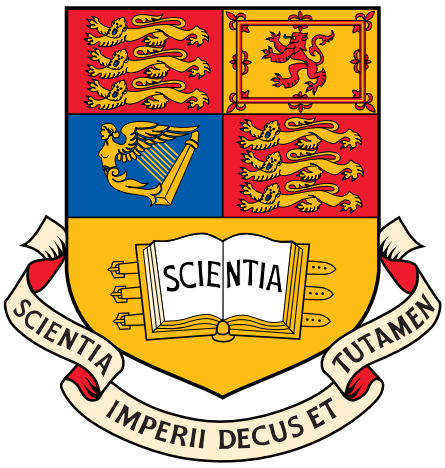
Development of a new Silicon Tracker at CMS for Super-LHC

Mark F. Pesaresi

High Energy Physics Department
Blackett Laboratory
Imperial College
London



A thesis submitted for the degree of
Doctor of Philosophy
of the University of London
and the Diploma of Imperial College.

October 2009

Declaration

Throughout this thesis, the work of other members of the collaboration is referenced.

During the period of this PhD, the author was responsible for:

- Developing of the online software for control and monitoring of the APVe and working in collaboration for the integration and commissioning of the APVe with the tracker readout system. An overview of the APVe and its software is provided in Section 2.3.4. Some of the commissioning tasks that the author was directly involved in are described in Section 2.4. Contributions to these results and conclusions were also made by members of the tracker DAQ community and have been referenced. The author worked in collaboration with the tracker commissioning group in Section 2.4 to identify the nature of the APV high rate noise using a software simulation of the APV25 pipeline process.

- Commencing development on the software required for simulating the performance of various geometries under high luminosity conditions as part of the SLHC tracker simulations workgroup. The author was also responsible for creating the stacked pixel layer in software and integrating this into tracker geometries with a focus on configurability and accurate modelling. In addition, these layouts were tested, maintained and documented by the author over multiple CMSSW versions to ensure that other users could also run simulations under SLHC conditions. The stacked tracking concept described in Section 3.2.1 and previous studies have been referenced. Sections 3.3.1 onwards in Chapter 3 contain the author's own work including the description, modelling and simulation of the concept geometry and stacked pixel layers, occupancy measurements and correlation algorithm.

- Using the software described in Chapter 3 to generate the results presented in Chapter 4. The complete study of single and double stack layer occupancy and triggering performance in Chapter 4 is the author's own work.

Abstract

Tracking is an essential requirement for any high energy particle physics experiment. The Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) employs an all silicon tracker, the largest of its kind, for the precise measurement of track momentum and vertex position. With approximately 10 million detector channels in the strip tracker alone, the analogue non-sparsified readout system has been designed to handle the large data volumes generated at the 100 kHz Level 1 (L1) trigger rate.

Fluctuations in the event rate are controlled using buffers whose occupancies are constantly monitored to prevent overflows, otherwise causing loss of synchronisation and data. The status of the tracker is reported by the APV emulator (APVe), which has now been successfully commissioned within the silicon strip tracker readout system. The APVe plays a crucial role in the synchronisation of the tracker by deterministic calculation of the front end buffer occupancy and by monitoring the status of the Front End Drivers (FEDs), where the tracker data is received and processed. In the event that the buffers are close to overflow, the APVe is required to veto L1 triggers until the system is ready. As such, it is important that APVe is correctly implemented so that the tracker can operate with minimal dead time. The integration of the APVe with the tracker readout and trigger control systems is discussed and the steps taken to ensure its correct operation are presented.

The Super-LHC is a proposed LHC machine upgrade to increase the luminosity by a factor of 10. The increased particle fluxes and radiation environment will necessitate the complete replacement of the current CMS tracker while presenting the design of a new tracker with severe challenges. Power consumption is one of the main challenges for the tracker readout system since a higher granularity detector will be required. Physics performance must not be compromised so the tracker material contribution should be lowered where possible. In addition, it is likely that the Level 1 system will require information from the tracker in order to reduce the trigger rate. A method of reducing the on-detector data rate for input into a L1 trigger using closely separated pixel layers is presented. A detailed simulation of a concept tracker geometry has been developed and the triggering performance has been estimated. The simulations report that the presented tracking trigger layer would be viable for use at SLHC. A layer would be capable of reducing the detector data rate by a factor of $\sim$20 while maintaining efficiencies in excess of 96% for tracks with p_T>2 GeV/c. The information provided by a single stacked layer would not be useful for reducing the L1 trigger rate, but two stacked layers are able to reconstruct tracks with $\delta p_T/p_T < 20\%$ for p_T<20 GeV/c and with sufficient resolution so as to match tracks with L1 calorimeter objects.

Acknowledgements

Working towards my Ph.D has been an extremely challenging but rewarding experience and there are many people I would like to thank for making this possible. First and foremost, my appreciation goes to my supervisor, Geoff, for his continued support, guidance and advice over the last four years and for the countless times he has corrected this thesis. I would also like to thank everyone I have worked with at Imperial and during my time at CERN. Specifically, Mark Raymond for his electronics advice and Matt Noy, Greg Iles and Jonathan Fulcher for their help with the APVe during integration.

I give my thanks to Harry Cheung and Carlo Civinini for all their help and guidance with simulations and CMSSW. I am also very grateful to John Jones for looking after and coaching me in my first year, for all the interesting discussions and for introducing me to his work on stacked tracking. My thanks go to Andy Rose for his software advice and to David Colling for starting my journey in HEP. I should also thank STFC and RAL for providing me with the funding to complete my Ph.D.

Finally but perhaps most importantly, I would like to thank all my friends and family for all their support through the years and keeping me inspired. To Kostas for being far too intelligent and dedicated to physics and to Anastasios for being right about almost everything. For always being there, I owe so much to you both. To Daniela, Barbara and Dan for keeping me motivated when I have needed it the most. And to Mum and Dad, for their unwavering support and interest in my work throughout my entire education. I dedicate this to you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Large Hadron Collider

The Large Hadron Collider (LHC)[1] is a 27 km circumference accelerator based at CERN, designed to collide protons or heavy ions from 2009. The LHC will collide protons at a centre of mass energy of $\sqrt{s}$=14 TeV, 70 times the energy of its predecessor LEP[2] and a significant increase in energy over the Tevatron[3] hadron collider. During nominal operation, bunches of around 10^{11} protons will be collided at 40 MHz resulting in a peak instantaneous machine luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. However it is planned to run the machine initially with a beam energy of 3.5-5 TeV, with fewer protons per bunch and a bunch spacing of 75 ns before increasing the collider energy and luminosity once stable operation is achieved and data are collected[4].

The LHC beam is accelerated in stages before it is collided. Protons are injected into a linear accelerator (LINAC2) before being transferred to the Proton Synchrotron (PS) via a booster accelerator. The PS will provide a 26 GeV beam with the required LHC bunch structure to the Super PS (SPS) which accelerates the protons to 450 GeV before the LHC ring is filled. The two LHC beams are accelerated to 7 TeV in separate beam pipes and circulated in opposite directions using a series of superconducting magnets to focus and guide the proton bunches. The beams cross each other at four interaction points around the ring where the four LHC experiments are situated; CMS[5] and ATLAS[6] are general purpose detectors, optimised

for discovery of new physics at high energy while LHCb[7] and ALICE[8] are dedicated experiments studying CP violation in the B sector and heavy ion physics respectively. The LHC accelerator complex is illustrated in Figure 1.1.

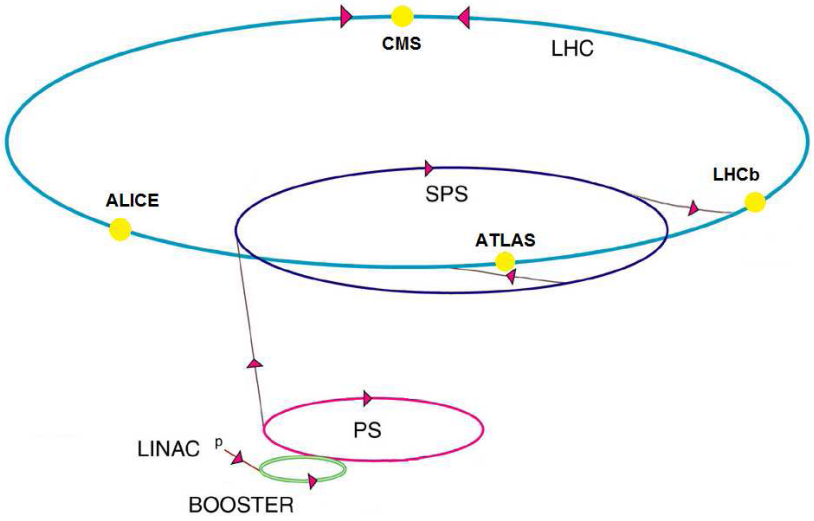


Figure 1.1: The LHC accelerator complex.

The LHC and its experiments represent one of the largest technological challenges ever undertaken in high energy physics. Their primary motivation is to extend our knowledge of physics beyond the Standard Model and in particular to confirm the existence of the long theorised Higgs boson. The high operating luminosity and beam crossing rate put an enormous strain on the detectors. The average number of interactions per second is given by,

$$N = L\sigma \tag{1.1}$$

where L is the machine luminosity in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and σ is the cross section in cm^2. An estimated inelastic proton-proton (including single and double diffractive) cross section at $\sqrt{s}$=14 TeV of 79 mb[9], gives an event rate of $7.9\times10^8\mathrm{s}^{-1}$. For a bunch crossing rate of 40 MHz and a fill ratio of 0.8*, an average of ∼25 inelastic interactions per crossing are expected at peak luminosity.

As a result, pileup of multiple events can become a significant problem, especially if detector components have time resolutions of greater than 25 ns. In addition,

*The bunch fill structure at the LHC will result in 2808 25 ns buckets filled with protons, out of a possible 3564 for a LHC orbit.

due to the short bunch spacing interval, the products from previous bunch crossings can also contribute to the pileup. High granularity detectors with fast response times are therefore required to minimise channel occupancy and hence pileup. As a consequence however, a larger number of channels will increase the on-detector power consumption due to the extra associated readout electronics.

The bunch crossing rate places strict requirements on response and signal times as well as on the speed of associated readout electronics. The high rate also has an important bearing on the design of the detector readout and trigger systems. Since only approximately 100 events/s can be stored for later analysis offline, an online trigger system must be in place to reduce the data rate by selecting the most 'interesting' events. As this cannot take place within a 25 ns period, the system must then be implemented with a pipelined processing stage to buffer data before readout. Hence, maintaining synchronisation between the detector channels becomes an issue.

The detectors must also be capable of surviving the harsh radiation environment the LHC provides. Detector elements and front end readout electronics must be able to withstand the high particle fluences and radiation doses expected in its lifetime with minimal degradation to signal/noise ratios and response times. Readout electronics must also be immune to single event upsets whereby depositions of large amounts of ionisation charge near sensitive circuit nodes occasionally cause the state of bits within a logic cell to flip.

1.2 The Super-LHC (SLHC)

After several years of operation, the LHC machine will be upgraded to increase the luminosity by a factor of $\sim$10. This is motivated by the fact that the quadrupole triplet magnets, which focus the beam at the interaction point, have an expected lifetime of $\sim$700 fb^{-1}[10]. In addition, the time to halve statistical errors increases exponentially if the luminosity remains constant. Increasing the luminosity of the machine extends the discovery reach of the LHC experiments, CMS and ATLAS in particular, as discussed in Section 1.3.

The relation between the luminosity of the machine and the beam parameters is given by,

$$L \propto \frac{n_b N_p^2}{\beta^*} F \tag{1.2}$$

where n_b is the number of bunches, N_p is the number of protons/bunch, β^* is the beta value at the interaction point (IP), and F is the reduction factor due to the crossing angle[11]. Increasing the luminosity requires increasing the number of protons per bunch, currently limited by the injector accelerators and the generated heat loads in the LHC and SPS rings, or decreasing β^* which parametrises the focussing of the interaction region magnets. Increasing the number of bunches (increasing the bunch crossing rate) is not possible due to the electron cloud effect generating unmanageable heat loads within the accelerator[12].

The upgrade will be achieved in two stages. The Phase I upgrade sees a replacement of the LINAC2 with a new linear accelerator (LINAC4) to increase the injection energy by a factor of approximately 3 and the bunch density by a factor of 2. This will make it easier to produce the LHC beam and allow some margin in order to be able to reach the proposed luminosity by increasing the number of protons per bunch. In addition, by upgrading the triplet magnets in the interaction regions around CMS and ATLAS, β^* can be reduced to 25-30 cm allowing a possible increase in luminosity to 2-3$\times 10^{34}$cm^{-2}s^{-1}[13, 12].

The proposed Phase II upgrade may require the replacement of the PS and booster with two new accelerators, the Superconducting Proton Linac (SPL) and the Proton Synchrotron 2 (PS2), which would double the injection energy and protons per bunch into the SPS. The upgrade could also require the replacement of the interaction region magnets to reduce the β^* further depending on the collision scheme agreed for SLHC. Broadly, there are two alternative scenarios to achieve the required luminosity greater than 10^{35}cm^{-2}s^{-1}. The Early Separation (ES) scheme relies on highly squeezed optics ($\beta^*<10$ cm) and placing accelerator elements within the experimental regions ($\sim$3 m from IP) which will impact on detector performance[13]. The Large Piwinski Angle (LPA) scenario however, does not require an upgrade to the interaction region magnets or better beam focussing. Instead, the luminosity

increase would be achieved colliding longer, more uniform, higher intensity proton bunches separated by 50 ns. On the other hand, this means that the peak event pileup is much greater than for the ES scenario. The beam parameters for the two schemes are presented in Table 4.1 in Section 4.1.

The luminosity upgrade will significantly impact the CMS and ATLAS detectors in terms of increased pileup, occupancies, data and trigger rates and radiation. Section 1.5 describes the proposed upgrades to the CMS experiment, if detector performance is to be maintained.

1.3 Physics at the LHC

Our current theoretical understanding of the nature of the universe is provided by the Standard Model (SM) of particle physics, which describes the properties of the fundamental strong and electroweak interactions with elementary particles. Since its formulation, it has been tested experimentally to a high degree of accuracy[14]; first predicting the existence of neutral weak currents and then the masses of the $W^{\pm}$ and Z vector gauge bosons, experimentally observed at the SPS many years later[15].

The Standard Model[16] is described by the $SU(3)\times SU(2)\times U_Y(1)$ gauge symmetry group. In order to explain the non-zero W and Z boson masses in electroweak theory, the SM requires the application of the Higgs mechanism where the original $SU(2)\times U_Y(1)$ gauge invariant symmetry is broken by introduction of a complex scalar doublet field with a non-zero vacuum expectation value[17]. Electroweak symmetry breaking by the Higgs mechanism recovers the gauge invariant electromagnetic field while directly generating the masses for the three gauge bosons. It also predicts a new massive scalar boson, the Higgs, which has yet to be observed experimentally. The Higgs boson interacts directly with the W and Z, controlling the high energy divergences predicted in WW scattering, providing that the Higgs mass is $\lesssim 1\,\mathrm{TeV/c^2}$[18]. The Higgs also interacts with the fermionic fields with a coupling strength proportional to the fermion masses.

The mass of the Higgs cannot be derived directly from the SM, although it can be predicted indirectly from precise electroweak measurements. An approximate upper bound on the Standard Model Higgs can also be inferred from the Higgs self interaction which for self-consistency is required to be finite up to an energy scale Λ where the SM breaks down. If the SM is to be an effective theory up to some unification scale at 10^{16} GeV, the Higgs mass is confined to $M_H \lesssim 160\,\mathrm{GeV/c^2}$. If new physics is expected at the TeV scale, the Higgs mass is bounded to $M_H \lesssim 700\,\mathrm{GeV/c^2}$[18]. Either way, with a seven-fold increase in energy and a hundred-fold increase in integrated luminosity over current hadron colliders, the LHC will be well placed to confirm the existence of the Higgs boson or discover physics beyond the Standard Model instead.

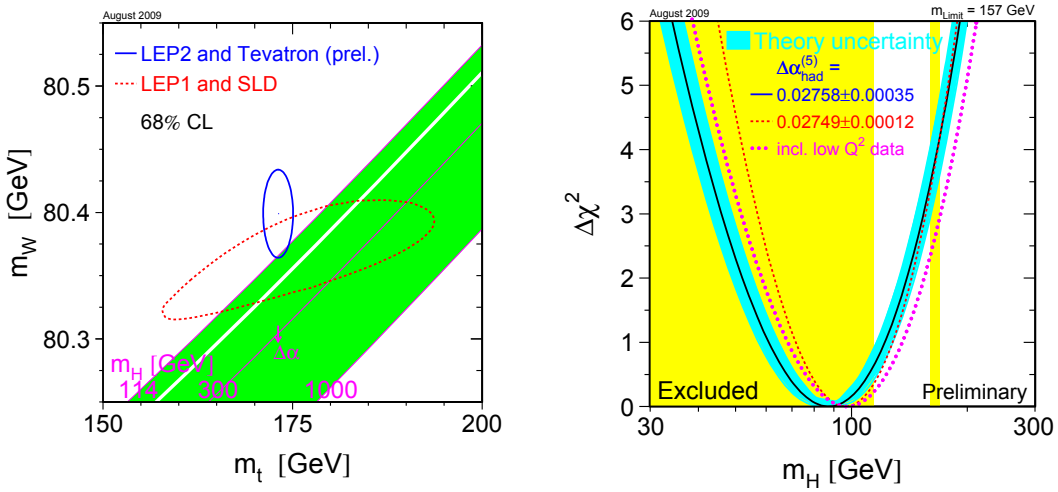


Figure 1.2: Left: Indirect constraints on M_W and m_t based on LEP-I and SLD data (dashed contour) and the direct measurements from the LEP-II and Tevatron experiments (solid contour) at 68% confidence. The shaded region corresponds to the dependence of M_H on these parameters. Right: $\Delta\chi^2$=χ^2-χ^2_{min} distribution of the Higgs mass using all precision electroweak data[14].

Calculation of radiative corrections to the mass of the W boson (M_W) from loops involving the Higgs and top quark, provides an indirect measurement of M_H given the top mass (m_t). Figure 1.2 shows how data from the Tevatron, where the top mass has been measured to 173.1±1.3 GeV/c^2[14], and LEP and SLD experiments have constrained the Higgs mass. Direct searches at LEP using $e^+e^- \rightarrow$ HZ production have excluded a Higgs mass below 114.4 GeV/c^2 with 95% confidence level while

direct searches at the Tevatron have excluded M_H in the region 160-170 $\mathrm{GeV/c^2}$. Using all precision electroweak data, a fit on the Higgs mass places an upper limit on M_H at 186 $\mathrm{GeV/c^2}$ with 95% confidence level.

Despite its experimental successes, the Standard Model cannot be a fundamental description of particle physics. For example, it fails to to predict the spectrum of fermion masses, the Higgs mass and the vacuum expectation value, and is not a true unification of the strong and electroweak interactions since it cannot relate the three gauge coupling constants[19, 20]. Moreover, it does not incorporate the gravitational force which a unified theory would be expected to explain. When single loop corrections to the Higgs mass are applied, the calculations diverge quadratically with the cut-off scale Λ. In order to cancel the divergent contributions to M_H, one must either fine-tune the parameters to an extremely precise degree or be forced to reduce the SM scale to $\mathcal{O}(1\,\mathrm{TeV})$[18].

An extension to the Standard Model would be to introduce Supersymmetry (SUSY), which requires every fermionic SM particle to have a bosonic partner and vice versa. This would solve the Higgs mass divergences by cancellation of the radiative correction terms without the need for fine-tuning[21]. In its most minimal form (MSSM) however, SUSY predicts a whole spectrum of as yet undiscovered particles (sparticles) and at least five different Higgs bosons. SUSY would also introduce a new set of parameters to determine experimentally, meaning that on its own, it would be no more fundamental than the SM[†]. On the other hand, it may provide a candidate for the non-baryonic dark matter present in our universe. The sparticle masses do not necessarily have to equal those of their SM partners, but their masses should be constrained to below $\mathcal{O}(1\,\mathrm{TeV})$ so as to avoid the recurrence of the fine-tuning problem[20].

The LHC will allow the study of Standard Model Higgs and SUSY phenomena, if they indeed do exist and open up a new energy regime for the discovery of beyond SM physics if they do not. It will also be used to perform measurements of Charge-Parity

[†]Although the additional particle content of MSSM is shown to allow joint convergence of the three gauge coupling constants at an energy of $\sim 10^{16}$ GeV.

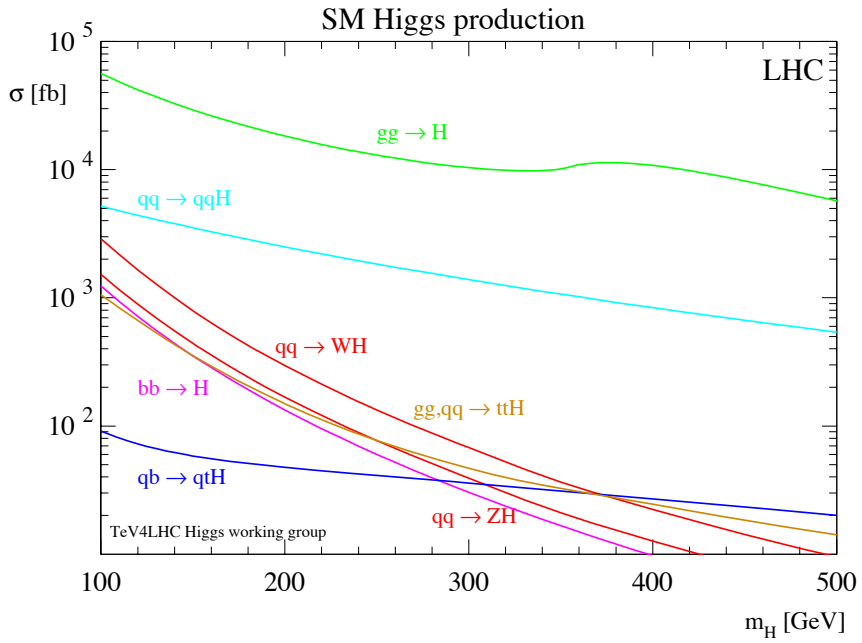


Figure 1.3: Standard Model Higgs production cross sections for pp collisions at the LHC where $\sqrt{s}$=14 TeV[17].

(CP) violation by production of B mesons. Figure 1.3 shows that the dominant Higgs production mechanisms at the LHC will be gluon-gluon fusion, associated Higgs production (HW/HZ) and vector boson fusion (H$t\bar{t}$), although their relative cross sections will be dependent on M_H. The production cross section for the Higgs boson is significant (>10 pb for M_H=200 GeV/c^2) but discovery will be dependent on the Higgs decay mode. Figure 1.4 describes the predicted branching ratios for the decay of the SM Higgs as a function of its mass.

For low masses ($M_H < 2M_W$), the Higgs will predominantly decay to $b\bar{b}$ pairs. Since the QCD production of b quarks is many orders of magnitude higher, this channel is not viable for physics studies. Instead, the more distinctive H $\rightarrow \gamma\gamma$ and H $\rightarrow \tau\tau$ decay modes will be important. In the case of H $\rightarrow \gamma\gamma$ decays, the major background will be from $q\bar{q}$ and $gg \rightarrow \gamma\gamma$[22]. Since a Higgs at this mass will have a width of less than 10 MeV, to extract the narrow signal requires fine grained calorimetry with an excellent mass resolution. Studies have shown that at low luminosities, H $\rightarrow \tau\tau$ decays via vector boson fusion will offer sufficient signal over background by tagging of forward jets[23].

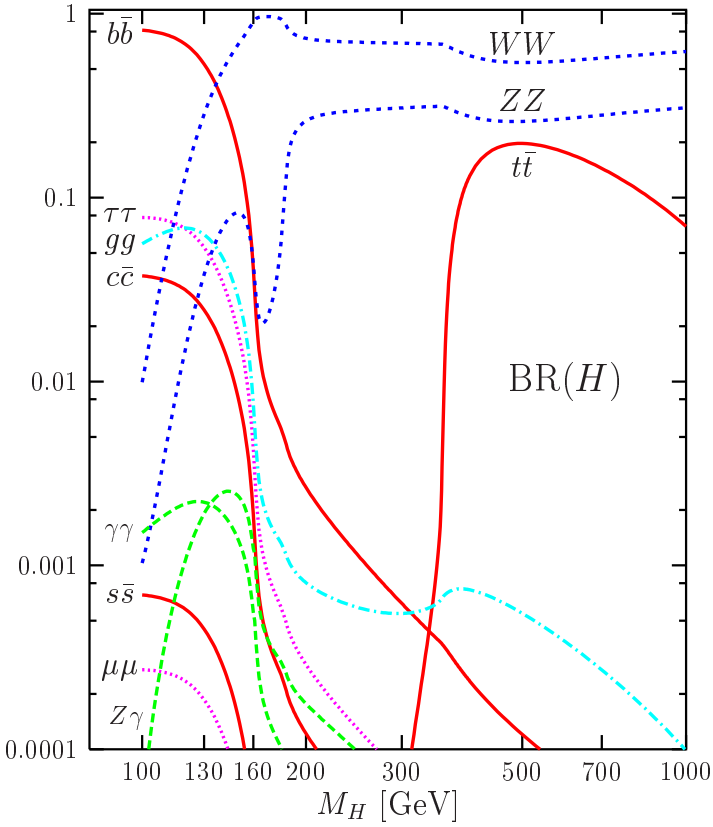


Figure 1.4: The decay branching ratios of the Standard Model Higgs boson as a function of M_H[19].

At higher masses ($M_H < 130\,\mathrm{GeV/c^2}$), H → ZZ(Z*) and H → WW decays will dominate. Isolated leptons from Z decay and an invariant mass at the Higgs mass should offer a clear signal. In the case of the H → WW channel and at very high M_H, detectors must have good missing energy and dijet mass resolutions in order to reconstruct the subsequent WW → $l\nu$+2jet and ZZ → $2l2\nu$ decays. Figure 1.5 indicates the 5σ discovery potential for CMS using various detection channels as a function of M_H. If it exists, evidence for SUSY is expected at TeV scale[20]. Weakly interacting supersymmetric particles from SUSY events cannot be detected directly but can be inferred from large missing energy signals in LHC events. The other detectable products from such events include multiple high p_T jets, leptons and τ jets. Early running of the LHC at low luminosity could reveal SUSY with $\sim 1\,\mathrm{fb}^{-1}$ of data due to the relatively large production cross sections[11] (Figure 1.16).

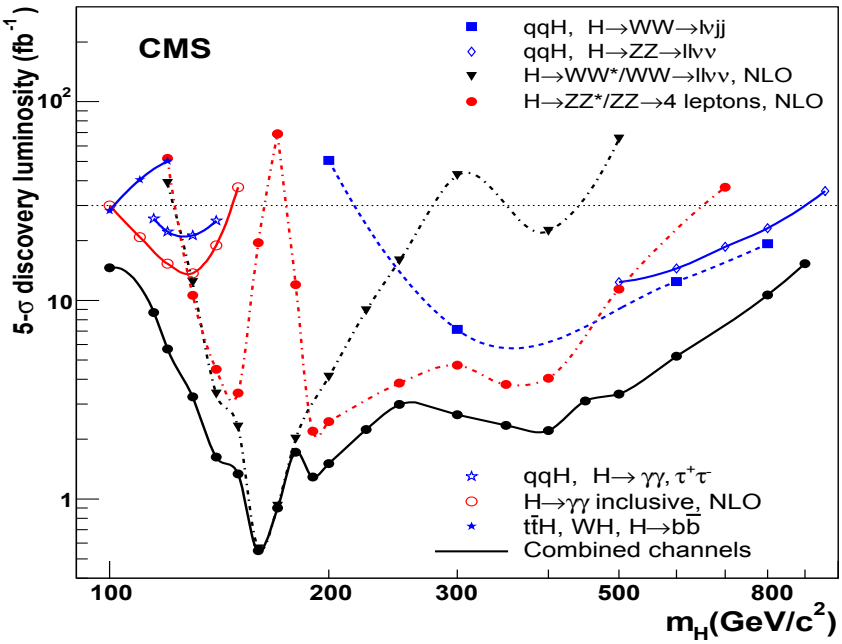


Figure 1.5: The required integrated luminosity needed to achieve a 5σ discovery signal in CMS using various detection channels as a function of M_H[19].

The increased luminosity at SLHC will extend the discovery reach of the LHC for new particles such as those arising from Supersymmetry, or new physics and will allow for detailed measurements of Standard Model processes and any new phenomena discovered during the first phase of LHC running. If the SM Higgs exists, it would have been discovered before the time of the proposed luminosity upgrade. As demonstrated by Figure 1.6, the SLHC will allow improvement in the measurement of the Higgs coupling ratios to ∼10%. It will also increase sensitivity to the measurement of the Higgs self coupling, evident from two Higgs final state events[24]. This will be extremely important in determining the form of the Higgs potential, the final free parameter in the SM. If the Higgs is not found at the LHC, the SLHC upgrade will instead be able to determine the mechanism for preventing divergences in the WW scattering cross section at ∼1 TeV[10].

The SLHC will also extend the mass reach of supersymmetric particles to ∼3 TeV/c^2[10]. If SUSY has already been discovered by the time of the upgrade, the increased luminosity will improve the measurement of the sparticle masses and

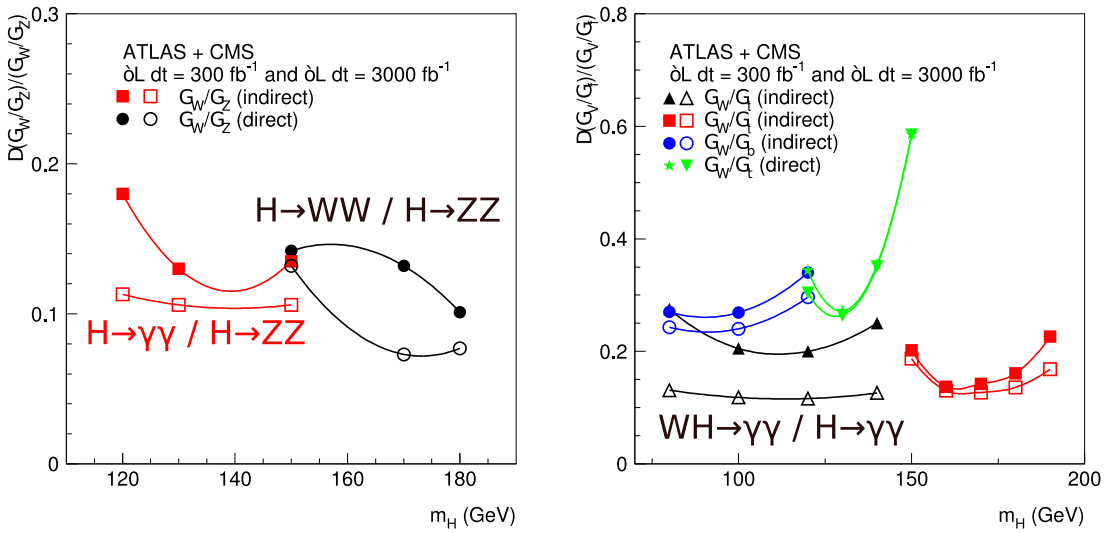


Figure 1.6: Expected uncertainties on the measured ratios of the Higgs widths to final states involving bosons only (left) and bosons and fermions (right), as a function of the Higgs mass for 300 fb^{-1} (closed symbols) and 3000 fb^{-1} (open symbols) integrated luminosity per experiment. The increased statistics will benefit the H $\rightarrow \gamma\gamma$ and H $\rightarrow$WW channels the most, mainly improving the ratio of the HWW to HZZ and HWW to H$t\bar{t}$ couplings, either directly or indirectly (through the W loop in H $\rightarrow \gamma\gamma$). The other couplings see limited improvement due to theoretical uncertainties on the absolute production cross sections at high M_H although this may improve in the future[24].

identify properties such as their spin, decay channels and branching ratios. This will help to narrow the SUSY parameter space and possibly determine the mechanism for Supersymmetry Breaking. In addition, the SLHC will be used to extend the discovery potential for beyond SM physics, such as new forces or heavy vector gauge bosons, at high energies.

1.4 The CMS Experiment

The Compact Muon Solenoid (CMS) is one of the two general purpose detectors in operation at the Large Hadron Collider (LHC) at CERN. A powerful 4 Tesla superconducting solenoidal magnet, the largest ever constructed, houses a central Pixel Detector, a Silicon Strip Tracker and Electromagnetic and Hadronic Calorimeters. The detector also comprises three types of Muon Detectors within the iron return yoke. The barrel region, coaxial to the beam pipe, is complemented by endcaps at each end to ensure detector hermicity (Figure 1.7).

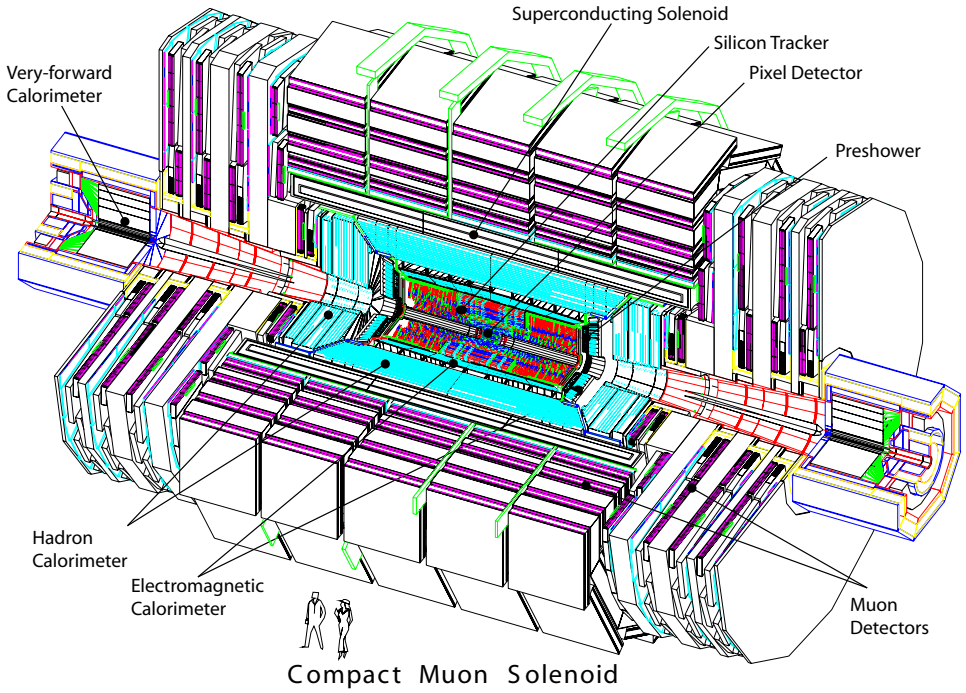


Figure 1.7: Schematic of the CMS Detector[11].

The primary goal of the CMS experiment is to find evidence for the Higgs boson(s); either as described by the Standard Model or by Supersymmetry. As a consequence, CMS has been designed to be sensitive to the various Higgs decay channels over the mass range $100\,\mathrm{GeV/c^2} < M_H < 1\,\mathrm{TeV/c^2}$. The detector has been optimised for the identification of muons, electrons, photons and jets as well as for the accurate measurement of their momenta and energy over a large range within a high luminosity environment. The CMS experiment will also search for Supersymmetric particles, evidence of new massive vector bosons and extra dimensions[11]. In addition, the CMS detector will be used to perform precise measurements of Standard Model physics which at the start of the experiment will be essential in characterising the performance of the detector.

1.4.1 Magnet

Fundamental to the choice of detector layout is the 13 m long, 4 Tesla solenoidal magnet used for providing the central tracking region with ability to measure the momenta of charged particles with excellent resolution. The superconducting coil rests inside a vacuum chamber and is cooled to 4.5 K using liquid helium. Surrounding the solenoid lies the iron yoke for the return and containment of the magnetic field. The return field in the barrel region will be large enough (∼2 Tesla) for muon tracking, so the iron yoke is integrated with 4 muon detection layers, both around the barrel and in the endcaps[‡]. Efficient muon triggering is an important requirement at hadron colliders, hence the large bending power of the magnet provides the $\Delta p/p \sim 10\%$ momentum resolution needed at $p = 1\,\mathrm{TeV/c}$ without placing excessive demands on muon chamber alignment and resolution[11]. Unambiguous determination of the charge of high energy ($p \sim 1\,\mathrm{TeV/c}$) muons is also possible with the high field configuration, further enhancing muon triggering efficiencies. The bore of the magnet measures 5.9 m and houses the detector inner tracking and calorimetry elements. The large field in this region assists the calorimeters by keeping low transverse momentum charged particles within the tracker, improving isolation efficiency and energy resolution[25].

1.4.2 Muon System

Whereas hadronic background is mainly contained within the calorimeters, muons are able to propagate past the magnet with minimal interaction with the detector. Muons therefore provide a strong indication of a signal event over the minimum bias background and are therefore prime candidates for triggering purposes. In addition, dimuon events are important signatures for 'golden' mode Higgs decay channels such as $H \rightarrow ZZ \rightarrow 4l$ and $H \rightarrow WW \rightarrow \ ll\nu\bar{\nu}$. As a consequence, the CMS detector employs a high performance, robust muon system for fast identification of muons with good momentum resolution over a large range of momenta.

[‡]For low luminosity running only three full endcap detection layers will be implemented.

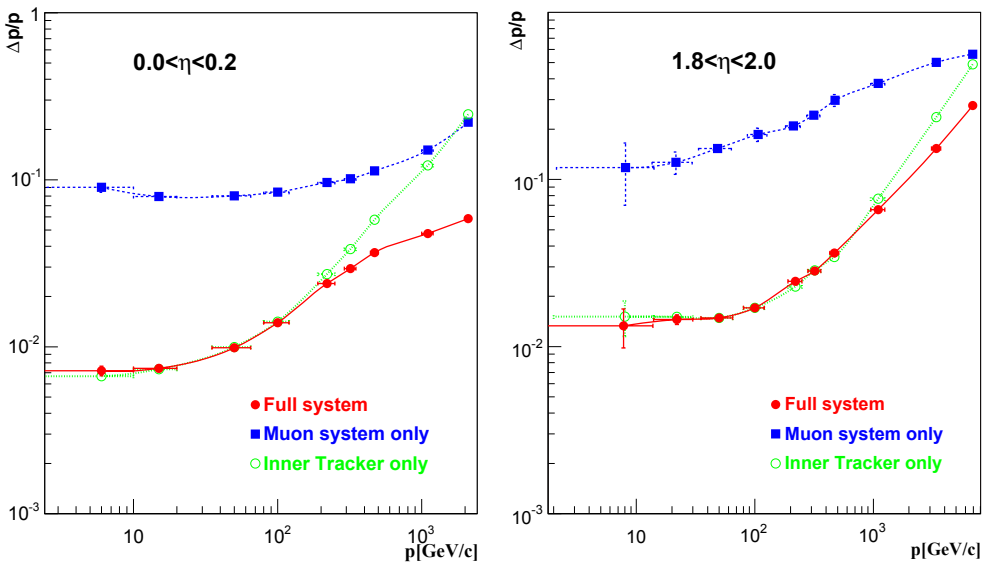


Figure 1.8: Muon momentum resolution as a function of momentum, using either the muon system only, the tracker only or the combined muon-tracker system in the barrel region $0 < |\eta| < 0.2$ (left) and endcap region $1.8 < |\eta| < 2.0$ (right)[11].

Figure 1.8 describes the performance of the system with respect to momentum resolution in the barrel and the endcaps respectively. The momentum is determined by measuring the muon bending angle after the solenoid, using the interaction point as the muon origin. The resolution for high p_T muons is dominated by the position resolution of the muon tracking stations. However, low p_T tracks are significantly affected by multiple Coulomb scattering within the detector, therefore tracking information is used in conjunction with the muon system to improve resolution by up to an order of magnitude. The use of the tracker is also important for increasing the resolution of high momentum muons, as the sagitta of the muon track can be measured both before and after the solenoid with minimal interference from scattering and energy loss, especially within the barrel region.

The muon system as depicted in Figure 1.9 employs three types of detector. Drift Tube (DT) gas chambers are interleaved with the iron return yoke in the barrel section and provide (r,ϕ,z) measurements of passing muons with point resolutions of $\sim$200 μm and a direction in ϕ to within 1 mrad[11]. DT chambers in the barrel region are complemented by Cathode Strip Chambers (CSCs) in the endcaps where

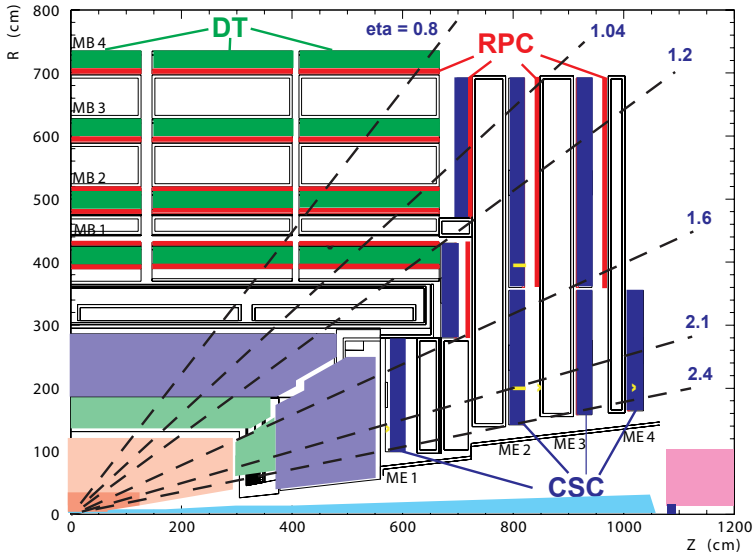


Figure 1.9: One quarter r-z section view of the CMS muon system for low luminosity running. RPC and CSC coverage will be extended after initial operation[11].

background rates are much larger and the magnetic field is highly inhomogeneous. Each CSC is able to provide a set of up to 6 (r,ϕ,z) measurements. Spatial resolution is improved by interpolating hits with gaussian fits so that each chamber attains a point resolution of 100-200 μm[11]. Resistive Plate Chambers (RPCs) are present throughout the muon system to achieve a position measurement with good time resolution for use in the Level 1 (L1) trigger. RPCs are formed from two closely spaced resistive plates with a high electric field applied across the gas gap. Ionisation charge caused by passing muons rapidly undergoes avalanche charge multiplication. This allows RPCs to operate at flux rates of up to 10 kHz/cm^2[11].

1.4.3 Calorimetry

The CMS detector calorimetry is provided by a high performance crystal Electromagnetic Calorimeter (ECAL) and a sampling Hadronic Calorimeter (HCAL). Energy measurements benefit from the placement of the calorimeters within the coil since energy loss due to interactions with the magnet system is eliminated. The calorimeters provide full geometric coverage up to $|\eta| < 5$ in order to maintain hermicity for the accurate calculation of missing transverse energy.

ECAL

The ECAL (Figure 1.10) consists of over 75,000 lead tungstate ($PbWO_4$) crystals present both in the barrel and in the endcaps. The design of the ECAL has been optimised for achieving the excellent energy resolution required for a Higgs search using the $H \rightarrow \gamma\gamma$ decay mode. The background for this channel is particularly large and so demands a mass resolution of $\sim$1% at $M_H \simeq 130\,\text{GeV}$, necessitating a highly granular and well calibrated ECAL. The crystals have a short radiation length (0.89 cm) and small Molierè radius (2.2 cm). With each crystal measuring $22\times22\times230\,\text{mm}^3$ in the barrel region (0.0175×0.0175 in $\Delta\eta\Delta\phi$, $25.8\,X_0$), most of the energy from electromagnetic showers is collected within a few crystals[11].

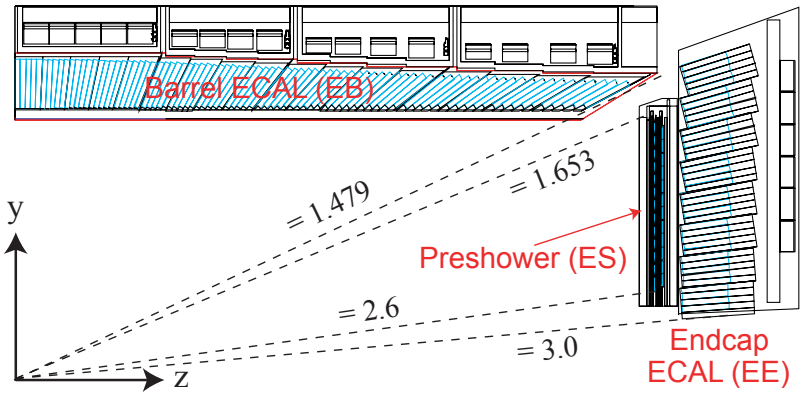


Figure 1.10: Transverse section of the CMS ECAL[11]. Coverage is provided up to pseudorapidities of 1.48 in the barrel and 3.0 in the endcaps. The endcap Preshower, consisting of two layers of alternating lead radiators and silicon strip samplers, is also indicated. The preshower is responsible for rejecting the large $\pi^0 \rightarrow \gamma\gamma$ background that would otherwise compromise the Higgs signal. Only the endcaps are instrumented with a preshower since in that region the angular separation between the two photons from π^0 decays is less than the granularity of the ECAL and thus can fake single isolated photons. With two lead absorber/silicon sensor planes and a total thickness of $2\,X_0$, the preshower helps by resolving the individual showers from incident photons.

Passing photons, pions and electrons deposit energy in the crystals resulting in the production of scintillation light with a typical response time of less than 25 ns. This light is collected by two types of photodetector glued to the crystal ends. Silicon avalanche photodiodes (APDs) convert photons with high gains of $\sim$50 and are used in the barrel region. Vacuum phototriodes (VPTs) provide a more radiation tolerant option in the endcaps even though they operate with lower gains. The signals are

then digitised on-detector and buffered until receipt of a Level 1 trigger accept. The ECAL also provides information to the trigger by summing the energy in 5×5 crystals and calculating the transverse energy in the 'tower' before forwarding this to the Level 1 regional calorimeter trigger, off-detector.

The crystals themselves are extremely radiation hard, with good uniformity over doses of up to 10 Mrad. The energy resolution of the ECAL has been measured to be $\sigma(E)/E<0.4\%$ for photons and electrons of 100 GeV depending on the intercalibration of crystal cells[26].

HCAL

The HCAL is a brass/steel-scintillator sampling calorimeter for the measurement of energy from strongly interacting particles. Its primary requirement is to provide a good measurement of the energy of hadronic jets, containment of showers for the measurement of missing energy and to protect the muon system from contamination. In this respect, it has been designed to maximise the amount of absorber material before the magnet coil whilst providing good jet and missing transverse energy resolution. Plastic scintillator tiles are sandwiched between absorber layers with wavelength shifting fibres to collect and channel the generated light to hybrid photodiodes (HPDs)[11]. The barrel HCAL is complemented by an outer hadronic calorimeter situated immediately outside the magnet so that showers can be sampled by almost 11 hadronic interaction lengths before the first muon chambers. HCAL endcaps provide geometrical coverage up to $|\eta| < 3$, while two forward steel/quartz fibre hadronic calorimeters fulfil the requirement for hermicity by sampling showers up to $|\eta| < 5$.

1.4.4 Tracker

The CMS Tracker (Figure 1.11) comprises a Pixel Detector for track vertexing located near the interaction point and a Silicon Microstrip Tracker (SST) for the calculation of particle momenta within the magnetic field. The tracker design and

performance is motivated by its requirements to provide high spatial resolution tracking of charged particles for accurate momentum measurements of signal decay products. The CMS tracker must also minimise its material budget as multiple scattering and bremsstrahlung processes reduce the performance of the Tracker, Muon and ECAL systems. Radiation levels are also extremely high in this region with particle fluxes of up to $10^7\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, requiring a tracker that can survive the harsh LHC environment for the duration of the experiment.

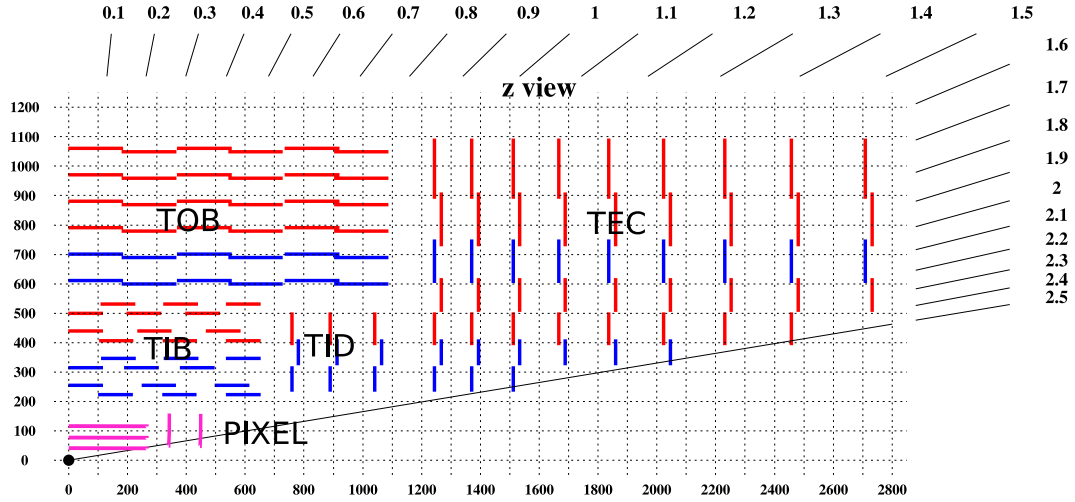


Figure 1.11: Schematic of the CMS Tracker. A quarter section is displayed with the interaction point at the bottom-left corner. The SST is made up of individual modules - indicated here by the straight red and blue lines. Each module consists of either 1 or 2 silicon sensors as well as associated readout electronics. The blue modules are double sided so that a stereo r-ϕ and r-z measurement can be performed. The red modules are single sided. Scale is in mm, with units of pseudorapidity indicating the angular coverage[11].

Silicon Strip Tracker

Figure 1.11 demonstrates the layout of the SST[27, 28]. With a total length of 5.4 m and a diameter of 2.4 m, the tracker is divided into four subsections: the Tracker Outer Barrel (TOB), Tracker Inner Barrel (TIB), Tracker Inner Disks (TID) and Tracker Endcaps (TEC). Subsections are comprised of layers of modules with each module containing a set of silicon strip sensors, a mechanical support structure and readout electronics. A summary of each tracker section with a parameter description of the sensors employed is given in Table 1.1. The sensors themselves are single

sided silicon wafers, with p^+-type strips, with various pitches, on a n-type bulk back layer operated under reverse bias. The voltage is enough to overdeplete the bulk layer so that charged particles passing through the sensors maximise the number of electron-hole pairs collected[29]. The p^+-type silicon strips are capacitively coupled to aluminium readout strips above a thin layer of SiO_2. Signals from 128 strips are recorded by an APV25[30] chip (see Section 2.2.1) which buffers and processes data from each channel before L1 readout. Each module employs between 2-12 APV25s, whose data are multiplexed and sent off-detector via optical links using on-board laser drivers[31]. The analogue readout system used by the silicon strip tracker is described further in Section 2.2.

Section	Layers/Rings	No. Detectors	Sensor Thickness (μm)	Pitch/Pitch Range (μm)
TIB	2 (2 stereo layers)	1536	320	80
	2	1188	320	120
TOB	4 (2 stereo layers)	3528	500	183
	2	1680	500	122
TID	3 (2 stereo rings)	816	320	81-158
TEC	4 (3 stereo rings)	2512	320	81-159
	3 (1 stereo ring)	3888	500	126-172

Table 1.1: Summary of sensor parameters in each of the Silicon Strip Tracker subsections.

The TIB comprises 4 layers, two of which use 'stereo' or double sided modules for the provision of both r-ϕ and r-z coordinates. These modules are made from two back to back sensors aligned with a relative angle of 100 mrad so that a point resolution of between 23-34 μm in r-ϕ and 230 μm in z is possible. The strip pitch within the TIB is 80 μm for the inner two layers and 120 μm for the outer layers while sensor thickness is 320 μm throughout. The strip length is $\sim$12 cm. The fine sensor pitch employed within the TIB allows more strips per sensor and hence reduces the occupancy per strip. In order to keep a similar occupancy in the outer tracker layers where the radiation levels are lower, strip length and pitch are increased. Modules within the TOB use sensors with strip lengths of $\sim$18 cm and pitches of 183 μm for the inner 4 layers and 122 μm for the outer two layers. Due to the increased noise from the resultant higher interstrip capacitance, a sensor thicknesses of 500 μm is

employed to maintain good signal/noise ratio in the TOB. The Outer Barrel also uses two layers of double sided modules allowing point resolutions of 35-52 μm in r-ϕ and 530 μm in z. To avoid resolution deterioration caused by tracks crossing sensors at shallow angles, Tracker End Caps consisting of 9 disks each are used between $|z| > 120\,\text{cm}$ and $|z| < 280\,\text{cm}$. At smaller radii, an Inner Detector comprised of 3 small disks is used and lies in the space between the TIB and the TEC. Sensors in the TID and the TEC are trapezoidal and have strips that point towards the beam axis. Pitches in the tracker end caps vary between 81-205 μm while sensor thicknesses increase from 320 μm before the fourth TEC layer to 500 μm at higher z. Both the TID and the TEC utilise double sided modules. The final system comprises almost 10 million silicon strips covering an area of over $200\,\text{m}^2$ making it the largest silicon tracker ever constructed.

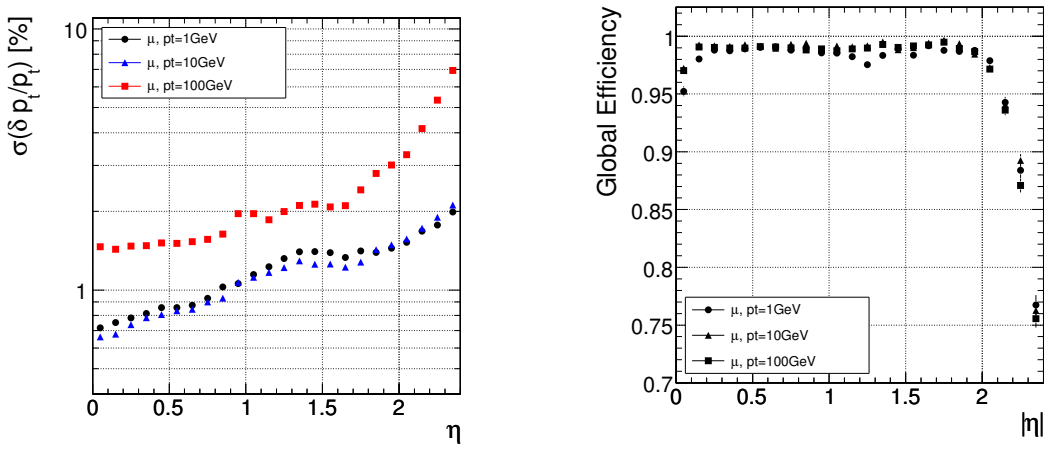


Figure 1.12: Left: Transverse momentum resolution of single muons (at p_T=1,10 and 100 GeV/c) in the SST as a function of pseudorapidity[11], and Right: Track reconstruction efficiency for muons (at p_T=1,10 and 100 GeV/c) as a function of pseudorapidity[11].

Due to its proximity to the interaction point, the CMS Tracker will suffer the effects of an extremely hostile radiation environment (Table 1.2). The silicon sensors and readout electronics within the tracker must be radiation hard enough to survive the 10 year operational lifetime of the experiment with negligible degradation of performance. The sensors use $\langle 100 \rangle$ oriented silicon crystals to minimise the effects of surface damage caused by ionising radiation. However, bulk damage due to the

significant hadron fluence through the sensors results in a reduced signal/noise ratio and an increase in leakage current. This will in turn cause a significant increase in power dissipation throughout the tracker and may lead to a runaway leakage current effect. To reduce the effect of radiation-induced power dissipation and to extract the ∼60 kW generated by the front end readout electronics and cables, the tracker is housed inside a temperature controlled outer support tube with the sensors operating at around -10° C. Bulk damage will also cause the depletion voltage of the silicon sensors to change due to radiation-induced doping concentration changes. At some point, the bulk type will 'invert' to p-type and a rapidly increasing voltage will be required to fully deplete the sensor. With this in consideration, the sensors have been designed to withstand a high biasing voltage (>450 V) without breaking down.

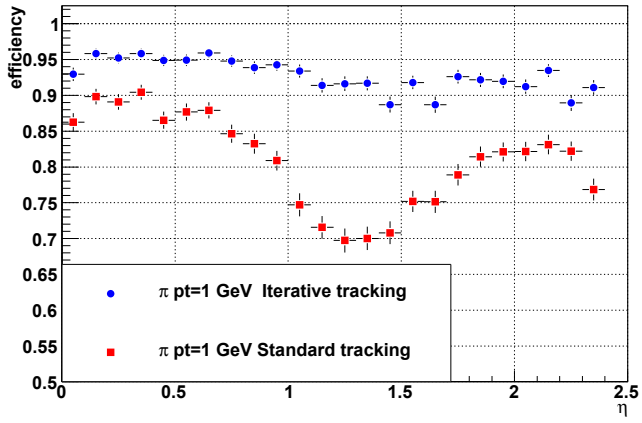


Figure 1.13: Track reconstruction efficiency for p_T=1 GeV/c pions as a function of pseudorapidity[32]. The two efficiencies are from two different tracking algorithms. The Iterative Tracking algorithm allows recovery of pions which interact in the high material regions by reconstructing tracks using looser criteria and better seeding without the expected increase in fake rate or reconstruction time.

A highly granular and precise tracker is required to minimise pileup and occupancy for efficient track reconstruction. With ∼10^7 readout channels, the SST provides the necessary granularity and keeps the strip occupancy between 1% and 3% throughout the tracker. The tracker must be capable of reconstructing charged particle tracks with good efficiencies over a wide range of momenta and pseudorapidity. A single hit finding efficiency of close to 100% is possible, even after irradiation, as the tracker has been designed to keep the signal/noise ratio to above 10[11]. Muon tracks should be

reconstructed with a 98% efficiency and a p_T resolution of <3% (p_T<100 GeV/c) for $|\eta| < 2$ (see Figure 1.12). At high p_T, muon momentum resolution can be improved by including information from the Muon system. Figure 1.13 demonstrates that the track reconstruction efficiency of pions within the tracker is slightly worse than that of muons, especially at $|\eta| \sim 1.5$, depending on the track reconstruction algorithm used. This is due to the larger pion interaction cross section and increased amount of tracker material between $|\eta| > 1$ and $|\eta| < 2$ (Figure 1.14). Increased material within the tracker worsens the effect of electron bremsstrahlung, photon conversions, multiple scattering and nuclear interactions before the calorimeters, reducing the performance of the detector. To maintain sensitivity to the $H \rightarrow \gamma\gamma$ decay channel, a requirement of the CMS tracker was that no more that 50% of Higgs photons should be allowed to convert in the tracker. Where the material thickness peaks at $1.8\,X_0$, $1 - e^{-\frac{7}{9}\times 1.8} \simeq 75\%$ of photons will convert on average.

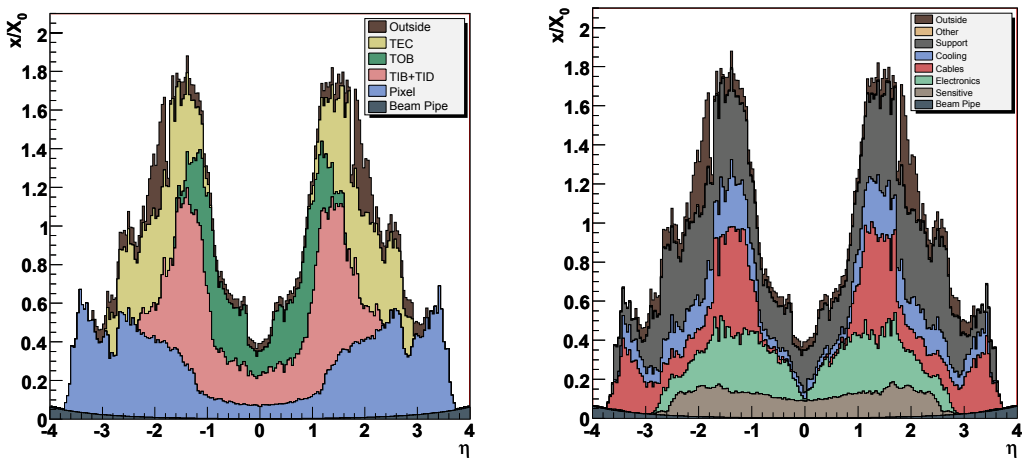


Figure 1.14: Tracker material budget in radiation lengths as a function of pseudorapidity by subdetector (left) and and by material type (right)[26]. The majority of the material is taken up by cabling, cooling equipment, electronics and support structures.

Pixel Detector

The pixel detector (Figure 1.15) is situated within the Silicon Strip Tracker and instruments the region closest to the interaction point. The pixel detector comprises three barrel layers and two endcap layers using pixellated silicon to cover an active

area of $1\,\mathrm{m}^2$. Each pixel measures $100{\times}150\,\mu\mathrm{m}^2$ and uses n^+-type implants on a $285\,\mu$m layer of n-type silicon substrate in contrast to the p^+-on-n sensors employed by the SST[29]. As a result, the pixel detector has a higher sensitivity to the Lorentz drift on charge carriers due to the presence of the magnetic field than in the Strip Tracker. This is a consequence of the fact that electrons provide most of the induced signal current instead of holes, hence their higher mobility induces charge sharing across neighbouring pixels. The pixel detector uses this to its advantage since the hit position within active pixel clusters can be interpolated and hit resolution improved. In the endcaps, where the magnetic field is perpendicular to the pixel plane, the endcap blades are rotated by 20°. In this fashion, the pixel detector will be able to provide single point hits with spatial resolutions of $\sim$10 μm in r-ϕ and $\sim$15-20 μm in z in the barrel layers[11].

The pixel sensor layer is bump-bonded to $\sim$16,000 analogue read out chips capable of amplifying and buffering signals before L1 readout. The final detector consists of about 66 million readout channels. Also essential to achieving such resolutions is the choice of analogue readout for the pixel detector. The pixel data are zero-suppressed and formatted before the analogue signal transmitted for digitisation off-detector[33]. This means that the amount of power dissipated within the tracker is reduced, while immunity to noise is improved and hit positions can be interpolated from the pulse height information from each pixel.

Radius (cm)	Fluence of Fast Hadrons ($10^{14}\,\mathrm{cm}^{-2}$)	Dose (Mrad)	Charged Particle Flux ($\mathrm{cm}^{-2}\mathrm{s}^{-1}$)
4	32	84	10^8
11	4.6	19	
22	1.6	7	6 x 10^6
75	0.3	0.7	
115	0.2	0.18	3 x 10^5

Table 1.2: Radiation & fluence Levels within the CMS detector (barrel) at different radial lengths from the interaction point after 10 years of operation ($500\,\mathrm{fb}^{-1}$ integrated luminosity)[11].

The pixel detector will provide extremely high resolution 3-dimensional measurements of charged particle tracks near the interaction point. This allows a precise calculation of track impact parameters and secondary vertex positions for efficient

tagging of τ's and b-jets. The high granularity and hit resolution will also prove extremely important for track seeding during reconstruction. However, due to its proximity to the primary vertex where particle fluxes will be at their highest, radiation damage mainly to the silicon sensors and read out chips will be substantial (Table 1.2). Although the pixel detector has been designed with radiation hard criteria, to maintain acceptable signal/noise ratios, response times and positional resolutions will require the replacement of the system at least once during the experiment lifetime. This is expected to occur after 2016 as part of the Phase I upgrade as described in Section 1.5.

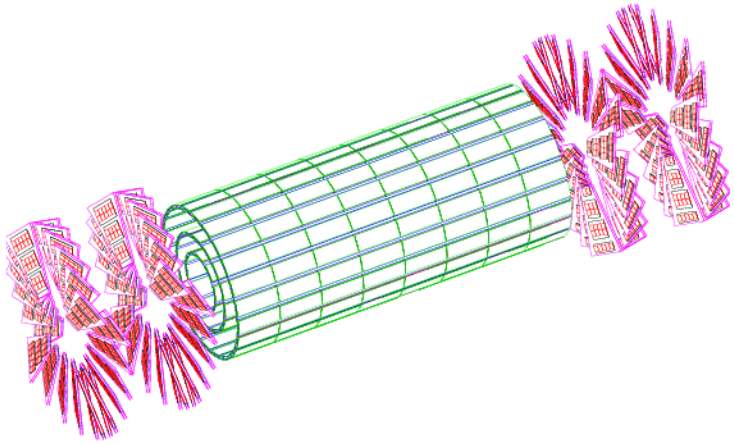


Figure 1.15: Layout of the CMS Pixel Detector[11].

1.4.5 Trigger System & DAQ

At design luminosity the LHC will be able to provide a tremendous amount of data at an interaction rate of 1 GHz requiring a fast, highly performant online data acquisition system (DAQ). Since data from approximately 100 collisions may be stored to mass media every second, a trigger system with a rejection factor of 10^6 is required. CMS achieves this using two physical trigger levels to select events deemed physically significant out of the billion QCD-process dominated background events that occur every second. Fortunately such 'interesting' events are rare (<1 Hz) and

so allows an efficient trigger system that retains as many signal events as possible whilst rejecting a large background.

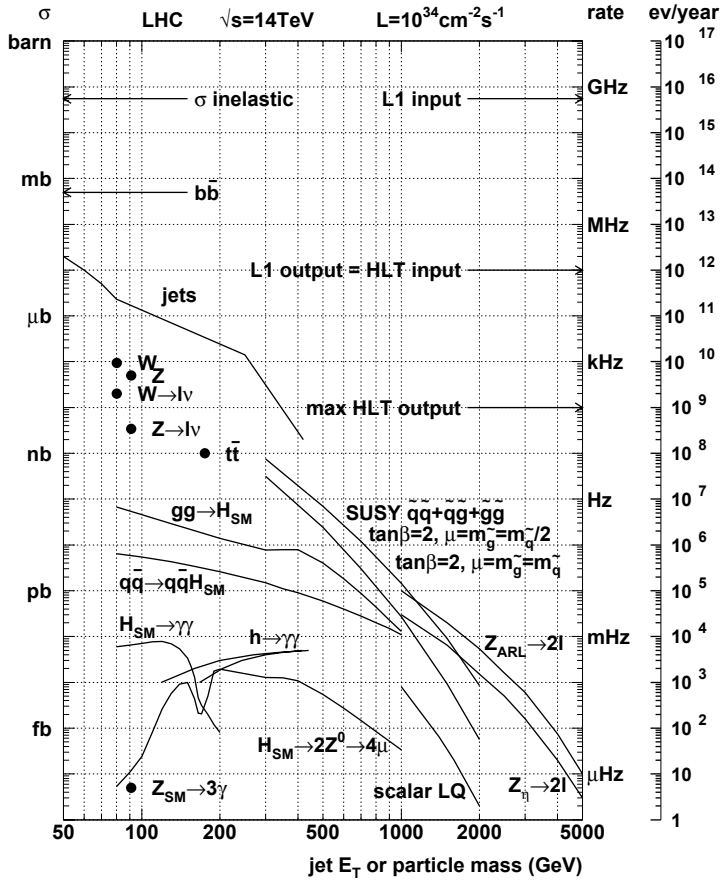


Figure 1.16: Inclusive proton-proton cross-sections as a function of particle mass (or jet E_T) for some physics processes at LHC at $10^{34}\text{cm}^{-2}\text{s}^{-1}$. Rates are provided on the right hand scale[34].

Level 1 Trigger (L1T)

The Level 1 Trigger (Figure 1.17) provides a quick online decision using basic reconstructed objects as to whether an event has provided physically interesting decay products. The primary purpose of the L1 trigger is to reject the significant number of events where only soft QCD processes have followed a proton-proton collision. These are typically characterised by low energy deposits, mainly in the forward calorimeters. In this way, the L1 trigger can be as inclusive as possible while quickly

being able to reduce the rate enough before more detailed processing can be carried out later on.

A Level 1 Accept (L1A) is generated if trigger primitives from the ECAL, HCAL or Muon systems are received and pass the L1 pre-determined cuts. For example, the trigger selects on regional electromagnetic objects (e/γ) or barrel or forward hadronic jets passing transverse energy (E_T) cuts, isolated hadrons (τ-jets) and muon tracks above a certain p_T. The L1 Global Trigger correlates the 128 possible trigger algorithms between the regional and subdetector trigger systems to achieve a rejection factor of 400[34].

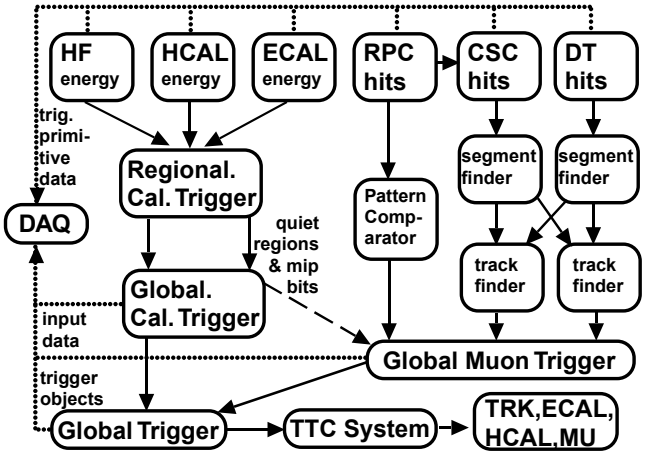


Figure 1.17: Control flow diagram of the L1 Trigger System[34].

Since a L1 decision is required in less than 2 μs, this places a phenomenal challenge on the readout and trigger electronics[11]. Low resolution and easily accessible data are read out through customised hardware processors including Field Programmable Gate Arrays (FPGAs) and high speed low latency multi-Gigabit optical links to achieve the high rejection rates required within a limited time period. The total L1 trigger latency period on detector amounts to $\sim$3.9 μs[§], allowing for the transit time for signals to reach the trigger logic in the services cavern 50-100 m away and a decision to return back to the detector. During this time, the high resolution

[§]The maximum L1 latency is set by the pipeline depth of the CMS readout electronics. The silicon strip tracker APV25 and preshower PACE chips limit the L1 latency to a maximum of 4 μs

data on detector must remain buffered in pipelined memories until the latency time has passed or an L1A is returned upon which the data are read out. The expected maximum L1 event rate will be 100 kHz at high luminosity.

High Level Trigger (HLT) & DAQ

After L1 readout, data from each subdetector are processed by front end electronics and buffered for access by the DAQ (Figure 1.18). Event fragments are buffered in up to 512 Reader Units (RUs) before a network switch, or Event Builder, capable of operating up to transfer rates of 800 Gb/s collates the event fragments. The fragments, totalling around 1.5 MB per event, are then transferred to the Filter Units (FUs) for event reconstruction through the HLT[35]. A single FU may contain multiple processors for running High Level Trigger algorithms in order to reduce the event rate from 100 kHz to the 100 Hz required for storage. All nodes on the Filter Unit farm use the same CMS software (CMSSW) to reconstruct the event within the detector framework and check if the event passes selection. Unlike the L1T, time is less of a constraint in the HLT which has an allocated average processing time of 40 ms per event (at low luminosity). As such, more complex algorithms including those that require tracking information, may be performed on the data. Even so, only partial reconstruction is used at HLT since, for example, full reconstruction of tracks using the entire tracker is extremely CPU intensive[11].

Computing Services collect the data output from the Filter System including events selected for analysis and calibration, for storage to mass media. Data for online monitoring are also obtained from the Filter System and may be processed by the computing services. The Event Manager is responsible for controlling the flow of events through the DAQ and interfaces with the Reader Units, Builder Units and the Global Trigger Processor. The Control & Monitor System (RCMS) interacts with the full DAQ and Detector Control System (DCS) to control and monitor the experiment during data taking and to provide easy configuration and running of the DAQ systems with an Internet accessible user interface[36].

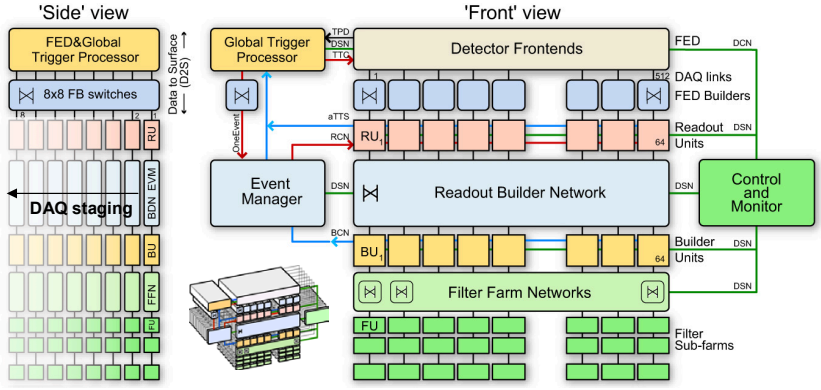


Figure 1.18: Overview of the CMS Trigger and Data Acquisition System[37]. Indicated are the Reader Units (RUs), Builder Units (BUs) and Filter Units (FUs) as well as the Detector Control Network (DCN) and DAQ Service Network (DSN) used by the Run Control/Monitor System (RCMS) for managing the DAQ. On the left, the 'side' view illustrates the principle of DAQ staging where 'slices' of the 'front view' DAQ will be gradually added until the full system is realised.

1.4.6 XDAQ

XDAQ is a generic software package aimed at improving systemwide integration specifically throughout the CMS DAQ and readout systems[38, 39]. Open protocols and standard libraries are used to build a common and simple framework for data acquisition systems operating in a distributed processing network. XDAQ provides a platform independent runtime environment for control of the central DAQ, local subdetector DAQ and readout systems as well as for the trigger and control systems. In addition, the XDAQ suite will aid the configuration, monitoring and calibration of the detector during the commissioning, debugging and operational phases of CMS.

XDAQ is managed by an executive which provides a XDAQ application written in C++ with all the tools needed for control and communication within a distributed processing environment. A copy of the executive may run on multiple processing nodes over a network and communicate using Simple Object Access Protocol (SOAP) over HTTP, ensuring platform independence between applications. XDAQ also provides a number of generic functions and components which may be useful in DAQ applications. Many data structures are available and can be passed between applications using SOAP. Communication through Intelligent Input/Output (I^2O)

binary messaging is also possible, allowing simple peer to peer messages to be sent over the network. Tools are provided to implement Finite State Machines (FSMs), applications whose behaviour is defined by a set of constant states and the transition functions between them, within the DAQ system so that a central controller can manage all DAQ processes using predefined transitions (e.g. "Configure", "Enable", "Halt" etc). The RCMS will issue these commands over SOAP for the quick startup and control of the detector and DAQ. XDAQ also offers a web application interface (HyperDAQ) for easy monitoring or configuration of DAQ systems from a web browser. The XDAQ tools enable the application to dynamically change the web page interface in response to user inputs or state changes.

1.5 CMS Upgrades for SLHC

The proposed luminosity upgrades will have a significant impact on the CMS detector due to the higher expected particle fluxes, channel occupancies, trigger and data rates and of course radiation damage. After approximately 7-8 years of operation, CMS would have collected an estimated $200\,\mathrm{fb}^{-1}$ of data. By this time, the inner layers of the pixel detector would have been subject to a fluence of $\sim 10^{15}\mathrm{n}_{eq}/\mathrm{cm}^2$ which is close to the radiation limit for the current pixel sensors before tracking performance is severely reduced[40].

It is therefore envisaged that during the Phase I shutdown, the entire pixel system will be replaced. The current proposals are to construct an identical system, supplemented with a fourth pixel barrel layer at $r \approx 16\,\mathrm{cm}$ and third pixel endcap pair at $|z| \approx 46\,\mathrm{cm}$[41]. The extra layers will provide better tracking performance, even with the higher occupancies expected after the luminosity upgrade, by reduction of combinatorial fakes using pixel triplets and an improved track vertex resolution[42]. An emphasis will also be placed on reducing the material contribution to the tracker by using a low mass CO_2 cooling system, ultra-light mechanics and by shifting inactive material such as readout electronics out of the tracking volume. The proposed layout is illustrated in Figure 1.19. In order to cope with the higher occupancies when the cabling requirements of the system cannot increase, the pixel ReadOut

Chip (ROC) will be modified to digitise the full pulse height information while the data bandwidth will be increased by a factor of >3 using digital serial links.

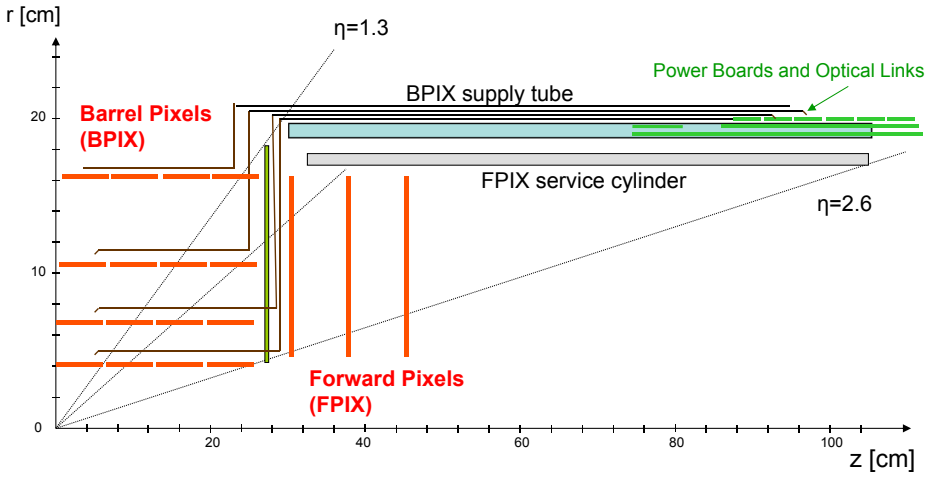


Figure 1.19: Proposed Phase I pixel upgrade. The new pixel system will comprise 4 barrel layers and 3 endcap disk pairs, lighter mechanics and cooling and more inactive material will be placed outside of the tracker coverage.

During the Phase I upgrade, it is also proposed to upgrade the HCAL Barrel and Endcap electronics. The HPDs are set to be replaced by Silicon Photo Multipliers (SiPM), which are similar to APDs in operation, offering better signal to noise ratios in low light conditions and reliability under high magnetic field configurations. In the present HCAL barrel, the scintillation photons from each fibre in a tower (17 layers in the barrel) are optically combined before the electrical signal is generated. For the upgrade, each fibre will instead be read out with a SiPM, before the analogue electrical signals are combined on-detector per tower for a measurement of the energy deposition[43]. The advantage over the current system is that each tower can be segmented to provide a depth readout of the energy (Figure 1.20). This will improve the HCAL energy resolution by removal of low E_T leakage from the ECAL while the extra channels provide redundancy. The front end electronics will also be upgraded to achieve better pulse time resolutions which, with the added depth segmentation, will help to reduce background from detector noise, pileup and out-of-time events such as cosmic muons.

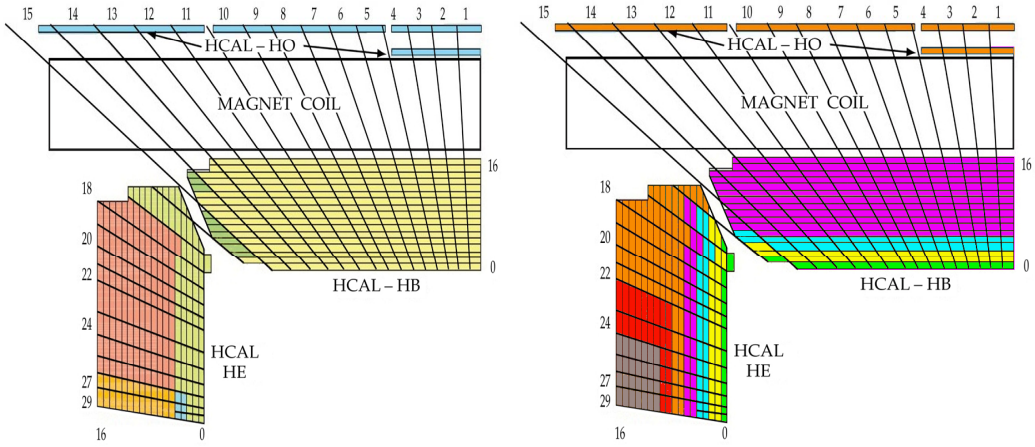


Figure 1.20: Readout segmentation in the current HCAL (left) and proposed Phase I HCAL (right)[44]. Energy deposition per tower will be measured at 4-7 different depths using Silicon Photo Multipliers. The number of segments is limited by the readout bandwidth of the upgraded HCAL which is set by the fixed volume for cabling.

A full fourth CSC station for extra coverage at $|\eta|$ ~1.7 will be installed for the Phase I upgrade. Additionally, a fourth RPC endcap layer will be implemented at the same time so that 4 station RPC coverage can be extended to $|\eta|$ <2.1[11].

The Phase II upgrade is likely to involve upgrades to the Forward HCAL and replacement of some of the radiation damaged scintillators in the HCAL endcaps[43]. The ECAL crystals and readout electronics have been designed to withstand the fluences expected after years of running at SLHC, although there will be degradation of signal with time. ECAL performance will be measured during operation, especially in the endcaps where the fluences are highest[10]. The muon system will probably require replacement of some of the front end electronics and trigger crates to handle the higher rates and radiation background. The RPC detectors at high η are likely to be upgraded for SLHC, since the hit rates are expected to be close to the RPC operational limit. If the upgrade requires placement of magnets near the IP (ES scenario), this may force changes to the detector in the endcaps and impact on the performance of the Forward HCAL.

In order to be able to operate at luminosities of up to $10^{35}\text{cm}^{-2}\text{s}^{-1}$, the full tracker including the pixel system will have to be replaced during the Phase II upgrade.

This is mainly due to the degradation of tracking performance after years of radiation damage. A new tracker however, will have to cope with the congested SLHC environment where charged track multiplicities could be $\mathcal{O}(10^4)$[¶] within the coverage of the tracker. This requires increased granularity, since channel occupancies in the TIB at SLHC will be at least 15-30%, and increased bandwidth for readout of data (or a reduction in transmitted information). Simulations with heavy ions[42], where the particle fluxes are expected to be similar to SLHC, show that efficient tracking will still be possible with higher occupancies indicating that the number of channels may not have to be increased by a factor of 10-20.

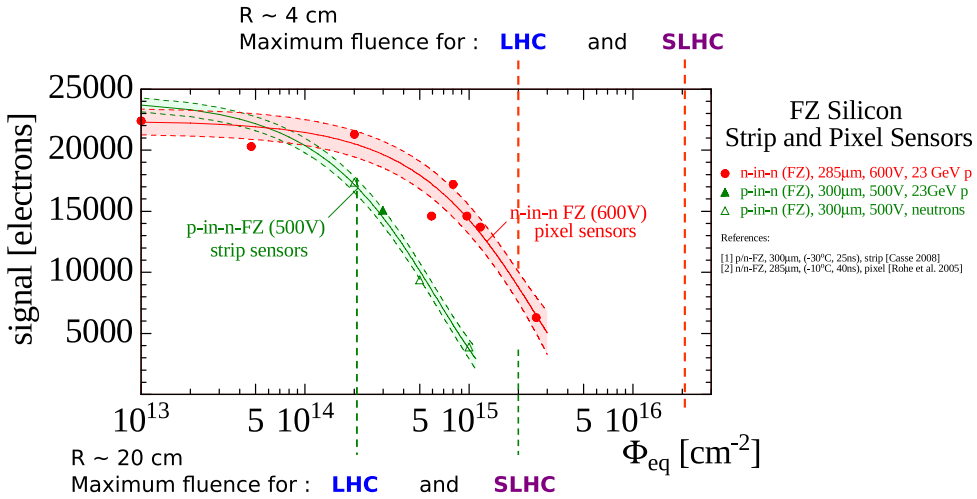


Figure 1.21: Signal performance of the current strip (p^+-on-n) sensors and pixel (n^+-on-n) sensors with fluence[45]. The current pixel sensor technology could be used at radii greater than 20 cm and provide sufficient signal for ~10 years at $10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The exact performance will be determined by sensor parameters such as thickness and operation voltage, and the channel noise which is dependent on the sensor pitch, the leakage current and the performance of the readout electronics.

An upgraded tracker will also need to be radiation tolerant to survive the higher fluences at SLHC. Figure 1.21 shows that the sensor technology used in the current pixels would be sufficiently radiation hard to last in the intermediate and outer tracking regions (r>20 cm) up to a fluence of $\sim 10^{15}\mathrm{n}_{eq}/\mathrm{cm}^2$ (~10 years at $10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}$). Alternative technologies with similar or better performance are also under investigation. In the inner region, it is still unclear if there are any materials which could

[¶]Compared to $\mathcal{O}(600)$ at $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

withstand the extreme fluences expected, implying that the inner pixel layers will have to be situated further out in radius and replaced at regular intervals.

The choice of sensor technology and dimension will impact on the sensor leakage current, especially after radiation, contributing to the detector noise and power. The sensor operating voltage will also affect the power requirements of the tracker. Power consumption is a major concern for an upgraded tracker since power delivery is constrained by the existing volume for cables. Because supply voltages to front end electronics will be lower, total current is likely to increase and therefore heat losses in the cables will be much larger. More channels and higher bandwidth links will also require additional power, as would digitisation on-detector. Cooling would be needed to counter the increased power dissipation within the tracker, adding material to the detector and therefore impacting on tracking and triggering performance. Power requirements will be reduced by ASIC development moving to smaller feature sizes. A binary un-sparsified architecture in 0.13 μm technology is proposed for strip sensor readout at SLHC offering a power per channel reduction of approximately 5 over the APV25[46]. Powering via DC-DC converters will also minimise power losses over cables by reducing the input current, although the technology for use in the CMS tracker at SLHC is still to be proven[47].

The increase in luminosity will also mean an increase in the Level 1 trigger rate if thresholds remain the same. By the time of the Phase I upgrade, it is hoped that the calorimeter trigger will have been entirely replaced, possibly with a trigger system based on the μTCA[48] standard, fast FPGAs and multi-Gigabit links and switches. The motivation for this is the requirement for a flexible trigger which could take data from different subdetector sources and apply more granular and complex algorithms in order to reduce the Level 1 rate[49]. Since trigger decisions must be achieved within the L1 latency, the hardware must be capable of transferring data between multiple regions and sources at speeds in the range of 4-10 Gb/s[10]. The L1 trigger will also be upgraded for Phase II, and will probably have to process basic information from the tracker in order to keep the rate below 100 kHz[‖]. This will be

[‖]The L1 trigger rate will be maintained at 100 kHz since any increase would require the replacement of all the CMS front end electronics in order to support the higher readout rates. It is assumed that the larger L1 event sizes (due to the increase in occupancy) can be absorbed with an expanded DAQ.

discussed further in Chapter 3. Since the tracker will be replaced during the Phase II upgrade, the L1 latency can be increased to 6.4 μs which is the limit imposed by the size of the pipeline depth of the ECAL readout electronics.

Chapter 2

The CMS Tracker Readout System

The CMS silicon strip tracker comprises almost 10 million channels which are simultaneously read out at the Level 1 trigger rate over approximately 45,000 analogue optical fibres. The tracker readout system has been designed to be able to cope with the high data rates expected at the LHC for both proton and heavy ion collisions. Integration and commissioning tests with the DAQ have proved invaluable in verifying that the entire system can maintain synchronisation with the central CMS control systems.

2.1 CMS Timing & Control Systems

The Trigger Control System (TCS) provides the LHC clock, Level 1 Trigger Accepts (L1As) and maintains synchronisation between the detector readout systems and DAQ. The TCS must primarily control the rate of L1 Accepts generated by the Global Trigger before distribution to the subsystems over the Trigger Timing & Control (TTC) network by monitoring the status of the front end systems. Front end status information is transmitted back through the Trigger Throttling System (TTS) upon which the TCS may temporarily throttle the trigger rate. The percentage of L1As lost due to throttling, or deadtime, is calculated by the TCS. Front end buffers are required to be large enough so that the DAQ inefficiency due to deadtime is less than 1%[50].

The TCS organises CMS subdetector components into TTC/TTS partitions, of which there are 30 (Figure 2.1). Each partition corresponds to a major element in the detector; for example there are four tracker partitions comprising the Inner Barrel, the Outer Barrel and two Endcaps. Control of the TCS is carried out by the Central or Global Trigger Controller (GTC)[51]. The GTC must also receive L1As from the Global Trigger processor and obtain the LHC clock and orbit signals from the LHC control room via the TTC Machine Interface (TTCmi) card.

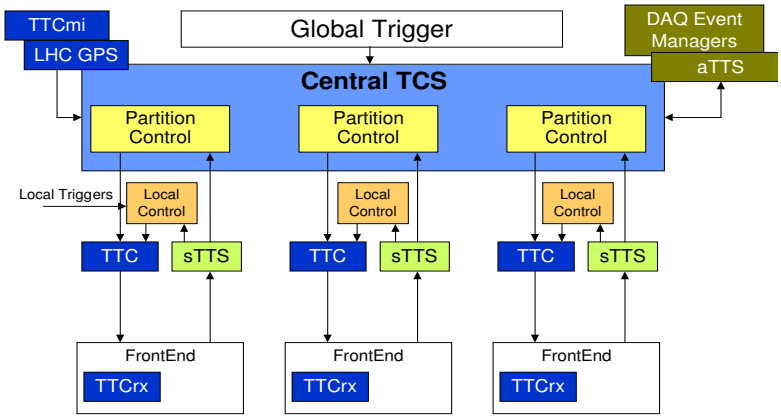


Figure 2.1: An overview of the Trigger Control System including the Trigger, Timing & Control (TTC) network and Trigger Throttling System (sTTS)[50]. Individual partitions can be controlled by a Local Trigger Controller (LTC) which provides the same functionality as the Global Trigger Controller except it can obtain triggers from the relevant local subdetector. In this way, partitions can be operated separately if the central TCS is down for maintenance.

2.1.1 Trigger, Timing & Control System (TTC)

The TTC system provides the CMS detector and readout electronics with the clock and control signals required for synchronous operation. A TTC distribution tree exists for each detector subpartition consisting of a TTC CMS Interface (TTCci) module, a TTC Encoder/Transmitter (TTCex) and possibly a TTC Optical Coupler/Splitter (TTCoc)[50, 51]. Information from the Local or Central TCS and the LHC clock and orbit signals from the TTCmi crate are passed on to the TTCci for distribution over the TTC network using two channels. Channel A is dedicated for the transmission of fast signals; L1 trigger decisions and the LHC clock. Channel

B broadcasts commands from the TCS including Event, Bunch and Orbit Counter Resets as well as L1 Reset (Resync) and Hard Resync signals.

2.1.2 Trigger Throttling System (TTS)

Since data rates from the detector are variable, a number of buffers are used throughout the front end readout systems to de-randomise the flow of data. This however introduces the risk of buffer overflows leading to loss of synchronisation and data through the whole DAQ chain[52]. The CMS detector requires that all readout buffers are monitored through the Trigger Throttling System (TTS) to ensure that buffer overflows are avoided. The TTS provides the Trigger Control System (TCS) with the front end status from which the Level 1 Accept (L1A) rate can be controlled. Buffer overflows can occur in the front end system due to the variation in both the data throughput and the L1 trigger rate.

TTS status signal	Description
Ready	Ready to receive triggers. Applied continuously if connected and working
Warning	Buffers close to overflow
Busy	Temporarily busy/buffers full and cannot receive triggers
Out of Sync	Event fragments do not correspond to the same front-end pipeline position or have different Event IDs or/and synchronous trigger data are out of sync. Requires a resync
Error	An error external to TTS. System requires a reset
Disconnected	Setup not done, cables removed, or power-off

Table 2.1: TTS status signals & descriptions. Modified from [50].

The TTS feedback signals are listed in Table 2.1. In the event that readout buffers are close to overflowing, local monitoring will generate a Warn signal over the TTS line. The TCS should respond by lowering the L1A rate, however further L1As may fill the buffers past a certain threshold at which point a Busy signal should be sent back. This forces the TCS to inhibit L1As so that the buffer occupancy can fall below an acceptable level and a Ready signal can be asserted[50].

2.2 The Silicon Strip Tracker Readout System

The CMS Tracker utilises an on-detector analogue readout design. One benefit of this approach is the improved position resolution obtained by interpolating track hits between adjacent channels. Another is the ability to measure the performance of the sensor and readout electronics in detail before and during operation. The choice of readout design has already proved invaluable in the testing and diagnosis of the tracker during commissioning[53, 54]. An analogue readout system is also advantageous for CMS as on-detector digitisation will require power and therefore extra cooling, increasing the material budget of the tracker. Channel noise can also be improved upon since pedestals may be subtracted just after digitisation off detector, reducing noise contributions from external sources. However, the readout system must be able to cope with the extremely high data rates from the front end which requires extra cabling.

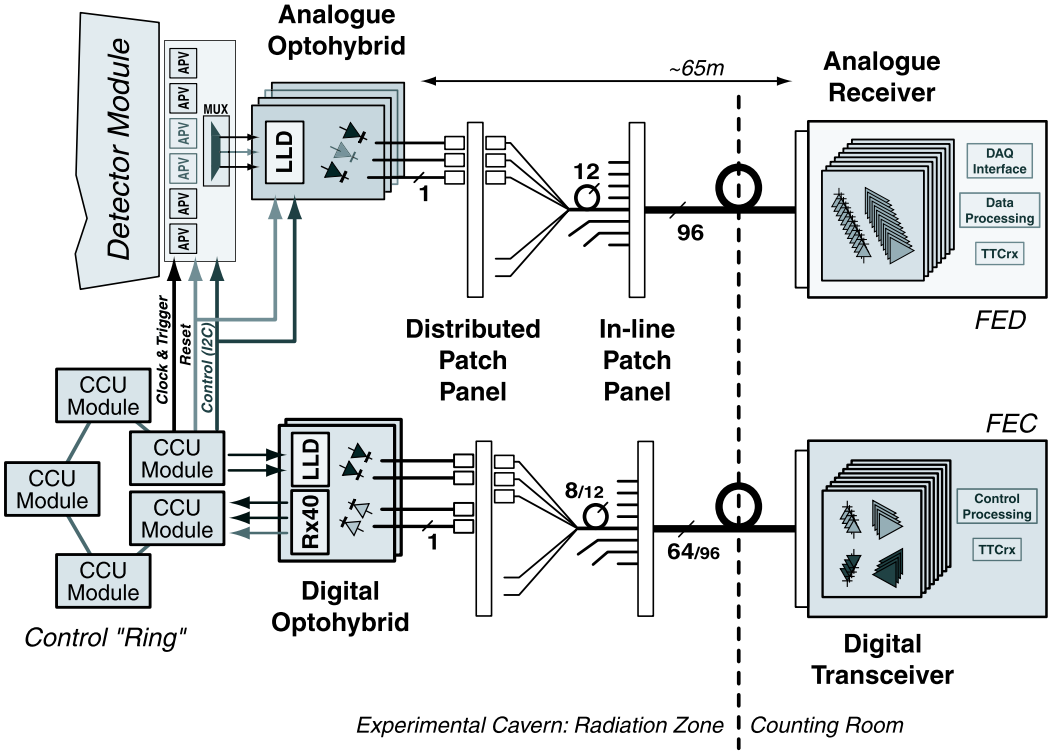


Figure 2.2: The CMS Tracker Front End control and readout architecture[55].

The SST readout system (Figure 2.2) is based on the APV25 chip which is capable of recording and buffering analogue data from the silicon sensors before L1 triggering. Each Front End Module (FEM) comprises a number of APV25 ASICs and APVMUX chips which time domain multiplex data from 2 APV25s onto a single differential line. Data are then converted to optical signals using on-board Linear Laser Drivers (LLDs) for transmission to the Front End Drivers (FEDs). The analogue optical signals are driven at 40 MHz over ∼60-100 m of optical fibre providing the high data rate required whilst minimising the power for data transmission and reducing the contribution to the material budget. Each Front End Module also houses a Detector Control Unit (DCU) for module monitoring and a Phase Locked Loop (PLL) chip for maintaining synchronisation of clock and control signals to the FEM[31].

2.2.1 The APV25

Signals from the tracker silicon strips need to be amplified, processed and buffered before L1 readout. This is achieved with the APV25 ASIC which utilises standard commercial 0.25 μm CMOS technology to obtain both the low noise and low power the CMS detector requires[56]. Radiation hardness is achieved through the small feature size and the thin layer of transistor gate oxide used. The APV25 also employs enclosed gate transistors to counter the increase in transistor leakage current during irradiation. Each APV chip may be exposed to more than the 10 Mrad of radiation expected within the lifetime of the SST without degradation in performance. The expected gain of each APV25 is around 100 mV/Minimum Ionising Particle, where a MIP is approximately 24,000 electrons in 300 μm of silicon. Power consumption has been measured to be 2.7 mW/channel[46] and the ENC (Equivalent Noise Charge) performance as 430+61/pF electrons (Figure 2.3) when operating in deconvolution mode[30].

An APV25 is capable of reading 128 detector channels and sampling them at 25 ns intervals. Each channel consists of a low-noise charge-sensitive preamplifier followed by a pulse shaping CR-RC filter with a peaking time of 50 ns[30]. The signal voltage is then sampled onto a 192-cell analogue pipeline memory at 40 MHz before awaiting

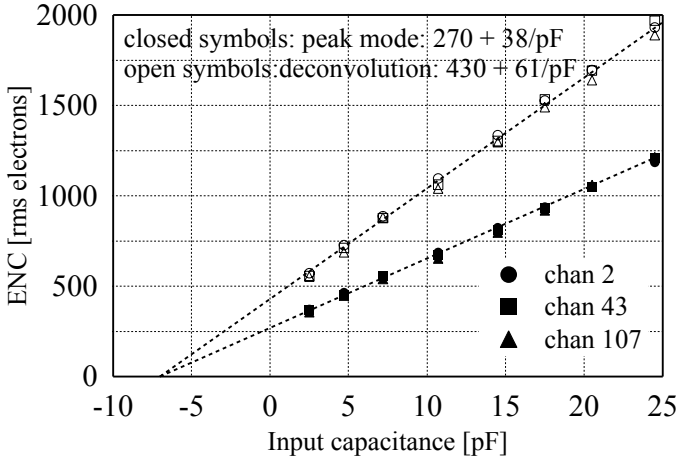


Figure 2.3: Measured noise dependence on input capacitance for the APV25 in peak mode and deconvolution mode[30].

a L1 trigger decision. The pipeline allows up to a $4\,\mu$s L1 latency corresponding to 160 memory locations. This permits a maximum of 32 cells for use as a FIFO to flag data triggered for readout. The APV can also be operated in different modes depending on the luminosity experienced by the detector. Peak mode is used at lower luminosities where the APV selects a single sample from each channel at the peak of the 50 ns CR-RC pulse. In deconvolution mode the APV takes a weighted sum of 3 consecutive pipeline samples using an analogue pulse shape processor (APSP) before readout. This is in order to reduce the effect of signal pileup due to the long pulse shape when the occupancies are high. However, if three pipeline samples are stored per event, the APV will only be able to buffer data from up to 10 L1 triggers. The APV also places a constraint on the spacing between triggers. The first trigger rule implemented by the Global Trigger Controller is that L1As must be separated by at least 2 bunch crossings. This is so that if the APV is operating in deconvolution mode, each of the three APV pipeline samples are reserved to a single trigger.

APV control registers for configuration settings and biases are accessible via I^2C[58] transfers. The APV is operated using a single control line for L1As("100"), Resync ("101") or Calibration Requests ("110") which generate a calibrated signal in the pipeline for readout. This encoding scheme takes advantage of the first trigger rule

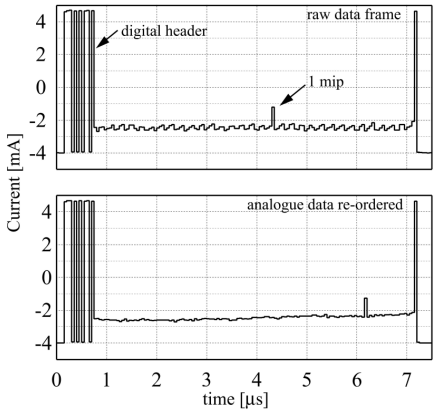


Figure 2.4: An example APV data frame[57] (top) and the same frame if the data are re-ordered by channel number (bottom). The frame is started with a 12 bit header followed by 128 channels of unordered analogue data while a tick mark indicates the start of the next cycle. Pedestals are clearly visible on each APV channel while the re-ordered data show there is a slight baseline droop across the chip.

so that there is no ambiguity between the L1A and Resync signals, since it takes an additional two clock cycles for the APV to register the trigger.

Eleven clock cycles after a Level 1 Resync signal is received by the front end, the APV pipeline logic is initiated. A write pointer circulates the 192 cell pipeline, controlling the sampling of the shaper signal to the pipeline at 25 ns intervals. The trigger pointer follows behind the write pointer a programmable latency period later, equivalent to the L1 trigger latency. If a L1A is received by the APV, the trigger pointer marks the current pipeline location (and the following two locations if operating in Deconvolution mode) in a FIFO. The pointers skip locations marked for readout when circulating the pipeline so that the cell data cannot be overwritten until it has been read out. The readout of triggered data is controlled by a separate cycle with a period of 1.75 μs and phase relative to the L1 Resync signal.

After APSP processing, the analogue data from each channel are multiplexed into a single output frame of length 7 μs, or 4 readout cycles. The frame consists of a 12 bit digital header followed by the 128 channels of analogue APV data. The APV also transmits 'tick marks' every 70 bunch crossings which are used to help the FEDs synchronise the front end readout system. The header contains 3 start bits,

an 8 bit pipeline address indicating the location of the signal data in the pipeline and an error bit signifying the status of the chip[25, 59]. The example APV data frame in Figure 2.4 illustrates this. A deviation from the baseline level in each of the detector channels is also noticeable. These levels or pedestals remain fixed over time and must be subtracted by the Front End Drivers along with the common mode level originating from external sources.

2.2.2 The Front End Driver (FED)

The CMS silicon tracker Front End Drivers are required to digitise and process the raw data from the detector before it is acquired by the Data Acquisition system (DAQ). Each FED (Figure 2.5) is responsible for taking the raw data from 96 optical inputs, amounting to 192 APVs, and digitising it with 10 bit analogue to digital converters at 40 MHz[60]. The FEDs must then buffer and process the digitised data. Delay Field Programmable Gate Arrays (FPGAs) provide independently skewable clocks to each of the ADCs in order to tune the timings of individual channels to within 781 ps[25]. In total, 24 Xilinx Virtex-II FPGAs[61] are used for this purpose, each FPGA using in-chip Digital Clock Managers (DCMs) to provide 4 configurable clocks. Data processing is then handled by 8 Front End (FE) FPGAs, each one merging the data from 24 APVs. The FE FPGA checks synchronisation within the APV channels using the APV tick marks before processing begins.

If an APV header is detected from a synchronised channel, the data are processed under one of three modes. Virgin raw mode outputs the full APV data load with no processing. In processed raw mode, the data are reordered and pedestals are subtracted before output. Zero suppressed mode reorders the APV data and performs both pedestal and common mode noise subtraction. Cluster finding, where algorithms sort through the data to pick out clusters of hit strips, is also performed in this mode, cutting the initial data rate of 3.4 GB/s by more than an order of magnitude. The pipeline address and error bit from the APV25 header are also recorded by the FE FPGA. If some channels are not synchronised or a header was not detected, data processing will still continue but the status of the event will be

forwarded with an error. The same is also true if the pipeline addresses for each event do not match or if the APV presents an error; the Trigger Throttling System (see Section 2.1.2) must instead decide how to proceed if synchronisation is lost.

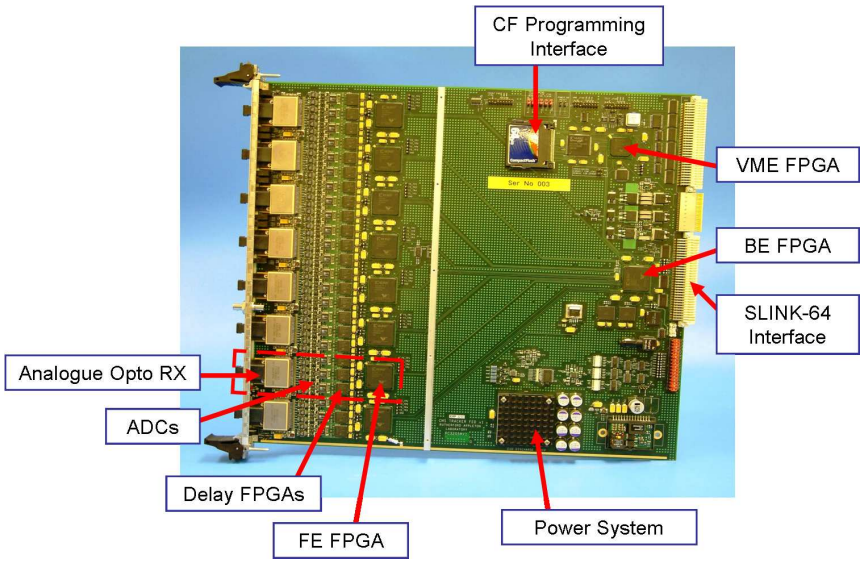


Figure 2.5: Layout of the CMS Silicon Strip Tracker FED. Modified from [62].

After buffering in the FE FPGA, the Back End (BE) FPGA merges the data fragments from each of the FE FPGAs over 80 MB/s links and buffers them using 2 MB of Quad Data Rate (QDR) SRAM memory. The FED data load is prepended with a header containing an event number, orbit number and bunch crossing number of the associated event. An S-LINK transition card interfaces directly with the FED and provides the pathway to the Front-End Readout Links (FRLs) and the rest of the DAQ. The transition card can transmit the event data fragment from the QDR buffer at a rate of 80 MHz providing a theoretical FED data transfer rate of 640 MB/s[63].

2.2.3 The Front-End Readout Link (FRL) & DAQ

Event fragments from pairs of FEDs are passed on to the FRLs[37] via a short (<15 m) S-LINK64 interface[64]. In total, there are a maximum of 512 FRLs for

use with the CMS DAQ, ∼270 of them for the SST alone. Two tracker FEDs output around 1-2 kB of data per event resulting in an average data rate through the FRL of ∼170 MB/s compared with the average Readout Link bandwidth of around 200 MB/s. Consequently, the FRLs may exert back pressure on the FEDs if the data rate increases due to random fluctuations in the L1 trigger rate and the event data size. Back pressure and buffer overflows are however alleviated by merging FEDs with high and low average data rates due to the tracker region they read out.

The FRLs are designed to transmit the data fragments it receives through an on-board network interface card and along 200 m of duplex optical fibre running at 2.5 Gb/s per link. The FRL data will be sent from the services cavern underground to the surface counting room where a set of 64 8x8 network switches, known as FED Builders, will merge the event fragments together to form a super-fragment[35, 37]. Each FED Builder can connect up to 8 Reader Units (RUs) which buffer the super-fragments before event building.

2.3 The Silicon Strip Tracker Control System

The CMS Tracker control system (Figure 2.6) is based around the Front End Controller (FEC) for the distribution of TTC commands. Timing synchronisation, system monitoring and configuration is carried out by the FEC using a token ring control network to communicate with the detector front end modules. Readout buffers and front end status are monitored over the TTS, allowing the TCS to maintain synchronisation and control over the entire Tracker system.

2.3.1 Front End Module Control

The FEM devices, including the APV25, are all accessible and configurable via the Inter Integrated Circuit (I^2C)[58] bus standard. Each device must be configured prior to data taking, this being achieved through the off-detector Front End Controller (FEC) and an on-detector token control ring network (Figure 2.2). The FEC

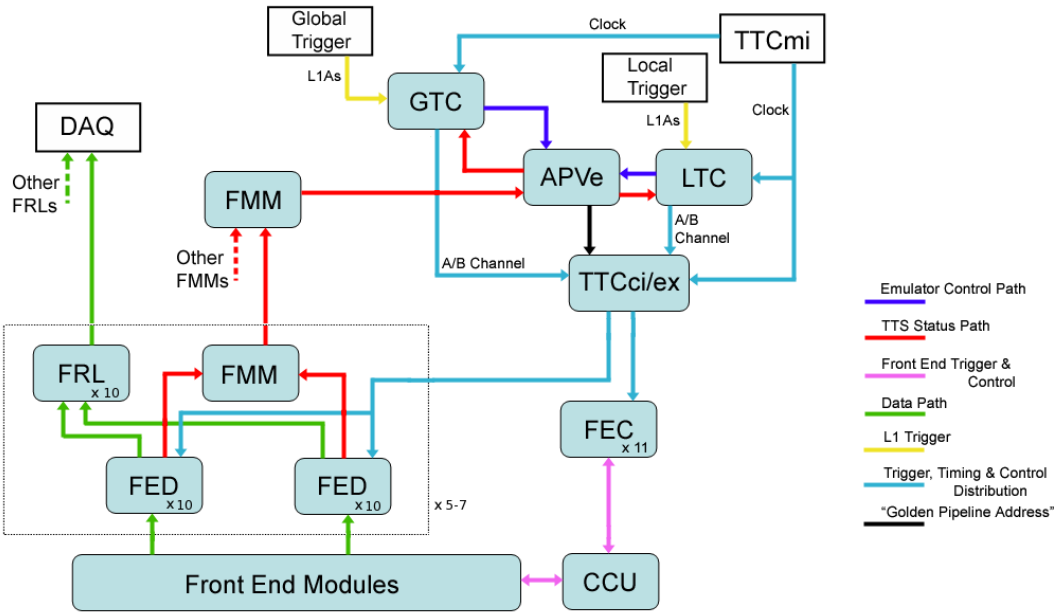


Figure 2.6: The Silicon Strip Tracker Readout and Control System and its interaction with the TCS (one partition). The marked region consisting of the FEDs, FMMs and FRLs is repeated 5-7 times depending on the partition size. There will be approximately 96-134 FEDs, 11 FECs, 5-7 FMMs and 50-70 FRLs per Tracker partition in the final system. All off-detector components are accessible via VMEbus and can therefore be controlled by RCMS and DCS through XDAQ.

also receives clock and control signals from the Trigger, Timing & Control (TTC) interface which it encodes and redistributes to the ∼300 control rings around the detector. Each FEC is able to control up to 8 control rings via digital optical links operating at 40 Mb/s[65].

Clock and trigger signals are distributed over a single channel around each ring to the Phase Locked Loop (PLL) chips located on each FEM. The PLL allows precise timing synchronisation between the modules throughout the detector and the LHC clock by adding fixed delays and phase corrections to the encoded signal. The delay due to time of flight of particles passing through the detector must also be taken into account by the PLL before it regenerates the clock and L1 trigger signals for use in the FEM electronics such as the APV25.

2.3.2 APV Readout Throttling

In deconvolution mode, the APV has only 10 buffers of three samples available for data awaiting readout following a trigger. It takes the APV 7 μs to read out a L1 trigger, so a random increase in the rate of Level 1 Accepts will quickly fill up the APV buffers. However, if more than 10 L1As occur within the 7 μs timeframe, the APV buffer becomes insufficient to hold all the L1 triggers and data would be lost. A buffer overflow will require all the APVs on detector to be reset and resynchronised with the rest of the front end resulting in a significant deadtime in data taking. As a result, the APV pipeline must be monitored during operation and triggers must be halted if the buffers fill up. The APV buffers however, cannot be monitored directly with a fast enough response since the signal latency between the detector and the TCS, located in the counting room, would exceed the length of the APV FIFO.

It should be noted however, that the occupancy of all of the APV buffers is entirely determined by the L1A rate and the APV data readout rate alone. Since all the APVs behave identically, a state machine can be used to emulate the behaviour of the APV and calculate the exact status of its buffers. In the CMS Tracker this is carried out by the APV Emulator (APVe - see Section 2.3.4). The APVe is able to determine the APV25 buffer status within 2 bunch crossings and send fast status signals to the TCS nearby. In this way, the TCS can temporarily stop issuing L1As before the APV buffers overflow. The APVe also supplies the "Golden" pipeline address which indicates the location of the L1 trigger data within the APV pipeline. This allows a verification with the APV header pipeline address that synchronisation has been maintained throughout the front end system.

2.3.3 Front End Driver Readout Throttling

Each FED contains 96 front end buffers, a back end buffer and a buffer for the Trigger Timing & Control header containing event identification data. Buffers may fill up due to fluctuations in the input data rate, corresponding to the mean strip occupancy in the Front End and the APV output rate. Even at a sustained maximum

rate of 140 kHz (back-to-back APV data frames) with occupancies greater than the maximum 3% expected in the Tracker, the FED would be able to cope without any of its buffers overflowing when operating under zero suppression mode. The limiting factor in the transfer rate is the S-LINK interface to the FRLs. The S-LINK will exert back pressure on the FED if sustained data rates of >200 MB/s occur, although typical data rates at full luminosity will average 170 MB/s with a 100 kHz Poisson trigger and 0.5-3% strip occupancy. If the FED is close to overflowing its buffers, it must assert the Busy status. If L1As are still received due to the latency in the TTS loop, the FED is required to drop the event data and only forward the event header with an error code. In this way, it is protected from losing synchronisation with the rest of the front end system, especially if there is still data buffered earlier on in the readout chain.

The FED must also perform a check that synchronisation has been maintained within the APV buffers. It obtains the "Golden" pipeline address from the APVe via the TTCci B-channel which it must then compare against the majority pipeline address from the APV frame headers. The FED buffer and synchronisation status are transmitted to the TCS via a Fast Merging Module (FMM). All 440 Tracker FEDs can be serviced by 25 physical FMMs to provide a merged state for each of the four tracker TTS partitions[66].

2.3.4 The APV Emulator (APVe)

The APVe hardware design is implemented by the IDAQ[67] (Figure 2.7), which can also be used in other applications as a generic data acquisition board. At the heart of the IDAQ is a Xilinx Virtex II Pro FPGA[61], providing the processing power and flexibility for different IDAQ applications including the APVe. The FPGA is booted off an on-board Compact Flash Card, facilitating a quick and easy method of swapping the functionality of the IDAQ. The IDAQ is also equipped with 128 MB of Double Data Rate (DDR) SDRAM memory which can be operated at up to 200 MHz (400 MHz DDR) on a 32 bit wide bus. Although the DDR memory is most beneficial in DAQ applications, it also provides the APVe with a significant amount of external

storage for history recording at fast read/write speeds. The IDAQ also implements a number of interfaces. Up to 300 single ended I/O connections are available either through 5 standard RJ45 jacks on the front panel or through mounting an additional card on the IDAQ. A USB 2.0 interface and a 10/100 Mbit/s Ethernet link are also provided. The IDAQ is implemented as a standard 6U VME board. The APVe will interface with controller PCs via the VME bus standard[68], common to all the electronics boards used in CMS. VME will be used to configure, control and monitor the APVe through custom built software when it is powered up.

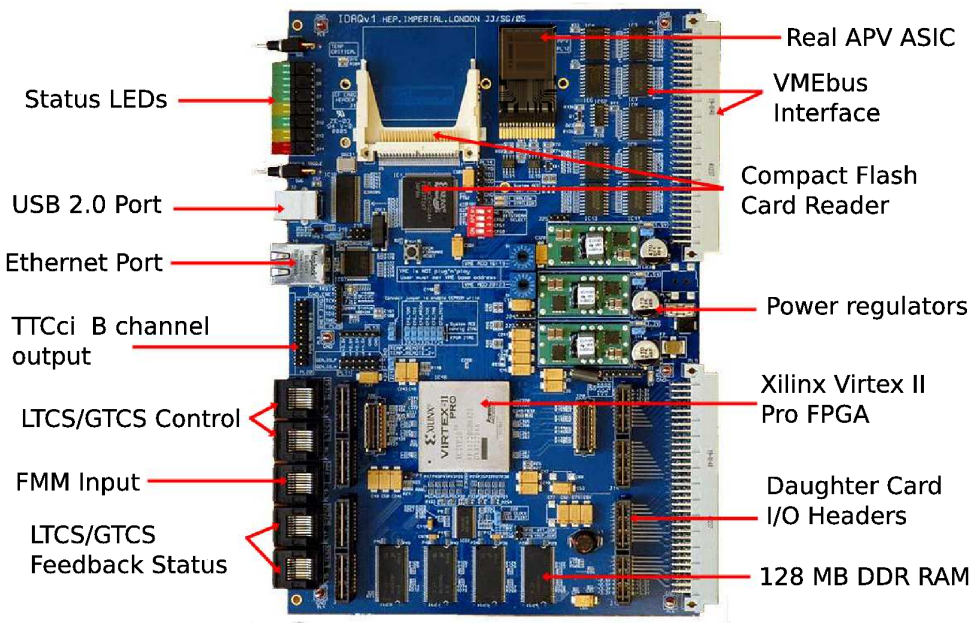


Figure 2.7: Layout of the APVE IDAQ card.

In order to emulate the front end APVs precisely, the APVe must receive clock and control signals from the TCS including L1As. The APVe can switch between a global or local TCS source (GTCS/LTCS), allowing tracker partitions to operate independently if the parts of the detector are undergoing maintenance[69]. A pair of standard Ethernet patch cables provide the TCS signals while another pair carry the APVe TTS status back to the respective control systems. The GTCS or LTCS can also be emulated by the APVe FPGA and interfaced with the APVe I/O using a loopback test card mounted on the IDAQ.

A real APV25 chip is mounted on the APVe and connected to the FPGA for the buffer occupancy calculation. The on-board APV is operated as it would be on the detector and is configurable via I^2C. As an alternative to this, a full VHDL model of the APV25 pipeline logic is implemented in the FPGA, simulating the APV operation. In both methods, the FPGA will keep count of the number of available buffers within the front end APVs, increasing an internal counter whenever a L1A arrives from the TCS input and decreasing the counter if an APV frame (either from the real or simulated APV) is detected. A Busy status signal is sent to the TCS if the APVe determines that the number of available buffers has fallen below a set threshold. The TCS response will be to inhibit further L1As until more buffers are available.

In order to reduce the tracker deadtime caused by blocked L1As, the number of filled buffers before a Busy status is asserted must be as high as possible. This limit is entirely determined by the size of the control loop from the TCS issuing a L1A to the APVe, the APVe determining the buffer status as Busy, the return of this signal to the TCS and the TCS responding with a L1A inhibit. During this time the TCS may issue further L1As and hence reserve buffer space is needed to stop the buffers overflowing within this period. Therefore, for maximum efficiency, this control loop must be as small as possible. The simulated APV offers increased efficiency over the real APV since it has complete knowledge of the internal buffer state and therefore a reduced latency in calculating the number of available buffers.

The control loop at CMS has been measured to be 8 bunch crossings. A control loop of this size reduces the maximum number of buffers to 8 in deconvolution mode*, corresponding to a deadtime of $<0.13\%$ (see Figure 2.8) if the simulated APV is used[69]. This fulfils the detector requirements for a total deadtime of less than $\sim 1\%$; in comparison the first trigger rule yields a deadtime of $\sim 0.5\%$. The thresholds are variable through the software and can be tuned if the control loop size is altered during operation.

*The first trigger rule states that no more than 1 L1 trigger can be sent within 3 bunch crossings, so a control loop of 3 bunch crossings would make maximum use of the APV buffers while a control loop of 9 reduces the number of available buffers by 2.

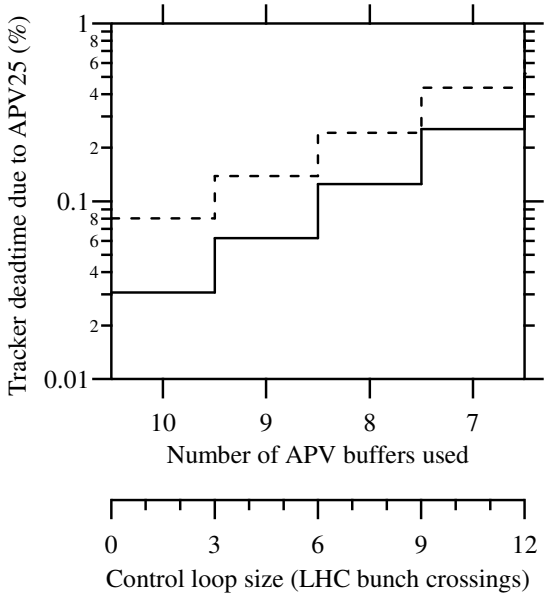


Figure 2.8: Tracker deadtime as a function of the control loop size and the maximum number of APV buffers available when operating in deconvolution mode. The deadtime calculated when the real APV is used is given by the dashed line. The solid line indicates the deadtime when the simulated APV is used[69]. The simulated APV offers increased efficiency since the buffer counter can be decreased as soon as the internal pipeline FIFO is cleared while the FPGA must wait for the detection of the three APV frame start bits (Figure 2.4) when using the real APV.

The APVe is required to monitor the merged FED status signal obtained from the Fast Merging Module (FMM). This status is combined with that of the APV buffers to form a tracker status signal which is received by the TCS. Table 2.1 describes the possible APVe status, where the most significant signal from the FMM and APV is output. The APVe latches on to any status signal presented on the FMM input, even if the event is transient, and will hold OOS or Error signals from the FMM until the signal is cleared and the TCS has issued a ReSync. The APVe may also become Out-Of-Sync if a BC0[†] is not received every orbit (3564 bunch crossings), the APVe configuration has been changed or if a buffer overflow has occurred due to incorrect setting of threshold values. APVe errors are rare but may be caused by FMM errors, unknown input signal codes from the TCS or FMM or incompatible configuration settings.

The APVe also records the address of the APV pipeline cell used to buffer the L1

[†]BC0 = Bunch Crossing Zero; defines the bunch crossing at the beginning of an LHC orbit. By design, no proton interactions take place on a BC0.

event. This "Golden" pipeline address is sent to the FEDs which are required to check the pipeline address with those from the front end APVs to ensure synchronisation has been maintained throughout the tracker. A diagram of the CMS Tracker Control System and its interaction with the APVe is shown in Figure 2.6.

Firmware

The APVe is operated over two clock domains. The local clock uses an on-board 40 MHz oscillator for the control of the APVe via VME. The TCS clock derives from the clock originating from the TCS. Operations relating to buffer occupancy calculations and interfaces to the TTS are run under the TCS domain. Digital Clock Managers (DCMs) within the FPGA are used to lock on to the external clock signals from either a global or local TCS source (GTCS/LTCS) and provides a method of protecting the APVe functionality if the TCS clock is lost[70].

The APVe firmware also implements a simulation of the TCS and FMM to output signals through a daughter card mounted on the board. These signals include TCS control commands (BC0, ReSync), L1As generated either repetitively or pseudo-randomly with Poisson occurrences and TTS signal statuses (Table 2.1) emulating the FMM. By routing the signals back to the APVe inputs, the APVe may undergo testing independent of external electronics. The firmware also stores historical information on the triggered pipeline address, APV buffer and FMM statuses and keeps a record of TCS events using counters and binners.

Software

The APVe can be accessed from a PC via VMEbus using the Hardware Access Libraries (HAL)[71]. An address table provides the HAL with the register space for the control of the APVe. At the lowest level, the ApveObject class uses the HAL libraries and address table to set parameters for configuration of the APVe and access individual registers. Higher level control is implemented by the ApveApplication class which provides user friendly access and control of the APVe by wrapping the

functions found in the ApveObject class. The ApveApplication object requires an XML file to configure the APVe parameters on startup. The ApveDescription class parses this file before handing the data to the ApveApplication. The APVe classes are written entirely in C++.

The ApveSupervisor is a software layer written within the XDAQ framework which implements the ApveApplication. It provides all the basic command sequences to operate the APVe in a full DAQ system and also allows monitoring and history reports to the end user. The main configuration options available through the HyperDAQ web interface are the selection of the external clock source as either GTCS or LTCS, selection of either the real or simulated APV and the mode of operation (Peak, Deconvolution) and the setting of values for busy and warning thresholds. The interface also allows the configuration and operation of the simulated TCS and FMM.

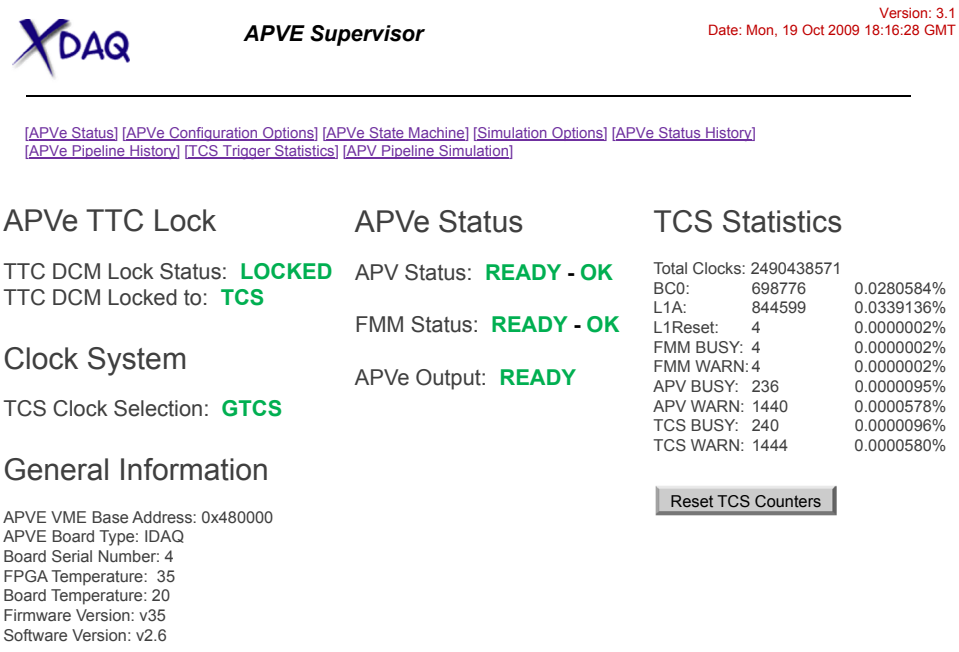


Figure 2.9: The 'Status' page of the ApveSupervisor HyperDAQ interface displaying board status, TCS statistics and the APVe TTS output.

The ApveSupervisor can be used to monitor the APVe reporting on information

such as the board and FPGA temperature, the DCM status and the TTS status of the current APV buffer, FMM and APVe. By accessing the external memory on the IDAQ, the Supervisor can display the log of pipeline, status and TCS events for debugging and for the verification of throttling. A histogram of the trigger distribution is also implemented in XDAQ in order to examine the TCS trigger generation profile.

Communication with the APVe may also be achieved using SOAP messaging so as to allow control under a central XDAQ application running on a separate node. The APVe will be managed by the TrackerSupervisor (via the ApveSupervisor), using system wide State Machine commands such as "Configure" and "Enable" as well as APVe specific commands, to set the TCS source for example. Almost all the configuration options available on the web application can be called over SOAP commands. The TrackerSupervisor, and hence the APVe, is controlled using SOAP by the XDAQ implemented Run Control/Monitoring System (RCMS), responsible for managing the entire DAQ, Timing & Trigger systems, DCS and configuration databases in CMS.

2.4 Integration and Commissioning of the APVe

2.4.1 Integration with the TCS

The TCS is integral to the correct operation of the APVe hence it is important that the interface between the two systems is fully validated. The emulation of the TCS using a mounted daughter card proved extremely beneficial in identifying firmware bugs within the APVe when the trigger controllers were unavailable for testing. This ensured that the integration time with the TCS, especially the global TCS which interfaces with all subdetectors, was kept to a minimum.

The test setup for both the LTC and GTC was identical, requiring a pair of Ethernet patch cables from the APVe to each TCS front panel. The FMM input was disabled on the APVe so that the TTS status reported to the TCS was that of the APV

buffer only. Each trigger controller is capable of sending pseudo random Poisson generated triggers at a configurable mean frequency. The rate was set high enough so that trigger throttling between the APVe and TCS was evident.

No major issues were detected when the LTC was tested, with the APVe successfully able to lock onto the LTC clock source and detect cyclic BC0s after a L1 Resync. With random triggers enabled, the APVe throttled the rate and operated without error until the end of the test. The APVe was also successfully able to lock onto the GTC but consistently lost synchronisation when random triggers were enabled. Using the APVe diagnostic output, the error was traced to a missing BC0 at the APVe input.

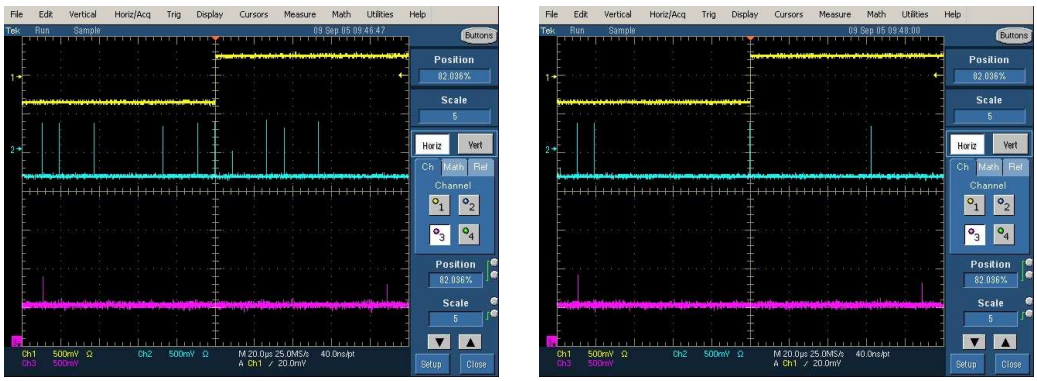


Figure 2.10: Examples of the APVe Ready to Out-Of-Sync output status (Yellow) transition[62]. The error occurs when no cyclical BC0 (Magenta) is detected at the APVe input. It can be seen that a L1A signal (Cyan) is received by the APVe instead.

The oscilloscope plots in Figure 2.10 show that at the moment the APVe asserts the Out-Of-Sync TTS status, a L1A is detected at the APVe input instead of a BC0. It was discovered that the GTC was occasionally generating coincidental L1 triggers and BC0s but prioritised the L1A over the BC0 causing the APVe to lose synchronisation. Although it was initially assumed that, by definition, a BC0 would never coincide with a trigger during normal beam operation, the GTC may generate test and calibration triggers that could replicate this event. The solution required a firmware modification to the global trigger controller so that both BC0s and L1As could be sent at the same time. In addition, in order to permit this, the encoding of

the signals between the APVe and GTC had to be altered as described by Table 2.2. The APVe firmware was modified to reflect the new encoding scheme and implement it only when the GTC input was selected in the software. The new firmware now accepts a simultaneous BC0 and L1A, maintaining synchronicity in the APVe while simulating the front end buffer status correctly.

Signal	Old Encoding (BC0,L1A,L1R)	New Encoding (BC0,L1A,L1R)
Inactive	000	000
L1A	010	010
L1Resync	001	001
BC0	100	100
ECR	011	011
OCR	110	101
BC0+L1A	-	110

Table 2.2: Old and new encoding schemes for control signal lines between the GTC and APVe. The OCR (Orbit Counter Reset) signal has been modified to allow simultaneous L1As and BC0s. The ECR control signal is the Event Counter Reset.

During a later test between the GTC and APVe using the new encoding scheme, random triggers up to rates of 140 kHz were successfully throttled and the system was able to operate without error. Figure 2.11 demonstrates the results of the trigger histogramming function reported by the APVe during this run.

The firmware module operates by counting the number of bunch crossings between triggers using a 32 bit binary counter and binning by the most significant bit. This results in the logarithmic histogram scale. Each of the 32 bins is implemented by an 8 bit counter which can store a maximum of 255 entries. If one of the counters overflows, the histogrammer stops until all the data are read out and reset. The software is able to perform multiple collections of this type so that more statistics can be obtained. Figure 2.11 shows the results for 100 collections during a run with the GTC generating pseudo random triggers at 100 kHz. The GTC can be seen to be implementing the first trigger rule successfully since consecutive triggers are never seen.

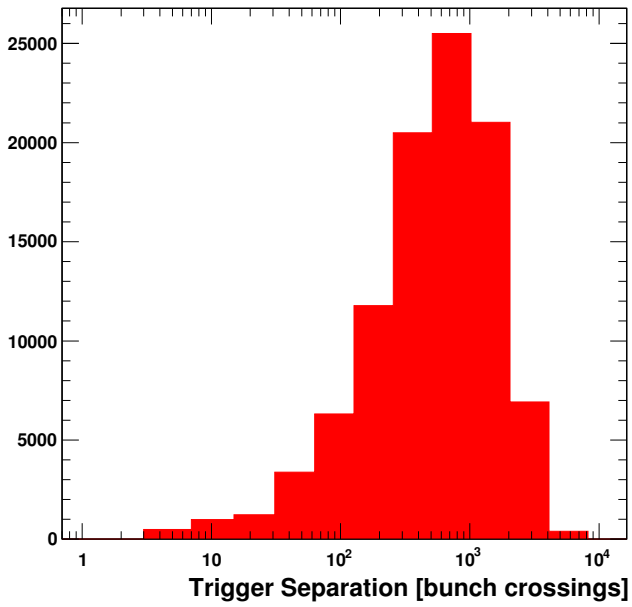


Figure 2.11: Throttled trigger distribution as measured by the APVe during a integration run with the GTC configured to generate pseudo-random Poisson triggers at a mean rate of 100 kHz.

2.4.2 Integration with the TTCci

During operation, the APVe must forward the pipeline address corresponding to a L1 trigger to the FEDs for matching with the majority pipeline address reported by the front end APV25s. This is achieved by transmitting over the TTCci B-channel. The pipeline address is sent from the APVe using an 8 bit data bus plus strobe in LVDS to the TTCci where it is stored in an input FIFO until the B-channel is available. There are requirements on the guaranteed asynchronous bandwidth on the B-channel[69] so that the pipeline address arrives at the FED in time for the synchronisation check.

The integration setup included the readout and control of real APV25 ASICs using FEDs and a FEC, themselves controlled by the LTC via the TTCci (and TTCex and TTCoc). The FMM input was disabled on the APVe and the LTC was configured to generate controlled test triggers.

Initial tests demonstrated a mismatch between the address from the front end APV25s, the TTCci address received by the FED and the address reported in the

ApveSupervisor pipeline history log. The reason for this was incorrect configuration of the TTCci which did not forward the address from the APVe correctly. With the correct settings, it was verified that the pipeline address reported by the APVe history did indeed match that received by the FED over the B-channel but did not correspond to the majority address from the front end.

As described, commands sent to the front end via the TTCci are separated by the A-channel which sends the clock and instantaneous L1A signals to the detector and the B-channel which sends delayed synchronous or asynchronous signals such as the Resync command. The APVe on the other hand receives all TCS commands instantaneously meaning that the Resync to first trigger separation differs for the APV emulation in the APVe and for the APV25s on the front end. For identical setups, this separation is fixed and hence a delay can be implemented in the APVe so that the correct pipeline address is transmitted to the FED.

Initially, it was wrongly assumed that an address offset could be applied to the output of the APVe to match the pipeline locations. However it was quickly determined that after a few triggers, the address calculated by the APVe diverged from the one reported by the APV. The subtlety here lay in the fact that the APV readout cycle is determined by the Resync time and since an APV frame takes $7\,\mu$s to read out, it can take a significant number of bunch crossings before a triggered pipeline address is read and cleared. As a result, the trigger pointer will skip these locations and a fixed address offset would no longer be applicable to subsequent triggers.

For the APV to be emulated correctly, the APVe firmware was modified so that all TCS commands including the Resync could be delayed by a programmable amount. This delay is configurable using the XDAQ interface up to a maximum of 255 bunch crossings. In this way, the Resync to first trigger separation can be made identical for both the front end and the APVe. With the test setup, the APVe had to be configured to delay only the TCS Resync signal by 95 bunch crossings so that the FEDs could successfully match the “Golden” pipeline addresses over a number of triggers. This delay was reconfigured to 102 bunch crossings in the final system due to differing cable latencies.

2.4.3 Commissioning of the Readout System

An important phase of the integration process for the Silicon Strip Tracker Readout System was the commissioning of a vertical slice of the DAQ. This involved testing 32 FEDs, corresponding to ∼6% of the tracker, reading the data out through 16 FRL modules via the S-LINK64 interface. The final stage of the DAQ where data are transferred to the Filter Farm was not tested. Instead, a DAQ PC running the Linux operating system was used to monitor the data output of the FRLs via compactPCI. Throttling of the DAQ data throughput was implemented using two FMM modules as described previously. The TTS feedback to the Trigger Control System from the FMMs was controlled using the APVe, which was also enabled to throttle the trigger rate according to the APV buffer status. The LTC, via the TTCci, was configured to send Poisson generated triggers with an average rate of 100 kHz.

At the time of the test, the tracker was not at a stage where data could be read from the front end using the FEDs. Instead, the FEDs themselves were configured to generate fake data to test the DAQ chain. The data, a replica of a multiplexed APV data frame including digital header and tail, are introduced immediately after the digitisation stage, using the block RAM in each of the Delay FPGAs as a buffer. Using software, any event occupancy may be simulated by programming the appropriate number of hit channels into the data payload. A pseudo-random number generator can be used to add a configurable channel noise to the data on an event by event basis and a common offset can also be applied to the pedestals. Consequently, when the fake event data arrives at the Front End FPGAs, it is processed as real event data would be (see Section 2.2.2). During the commissioning test, the FEDs were configured in Zero-Suppressed mode where only clustered data are forwarded for readout. The pipeline address check in the FED was also disabled in this test.

The event occupancy was varied between 1 and 10% and the throttled trigger rate of the whole system was measured using the LTC. Figure 2.12 demonstrates how

the average measured data loss[‡], calculated as the ratio of blocked triggers to total triggers in run, varies with the occupancy. Two modes of FED operation are presented. In full debug header mode (Zero-Suppressed) most of the information from the APV frame headers is retained for later inspection. In compressed mode (Zero-Suppressed Lite), the minimum amount of header information is transmitted and hence the average data packet size is reduced by ∼0.5 kB.

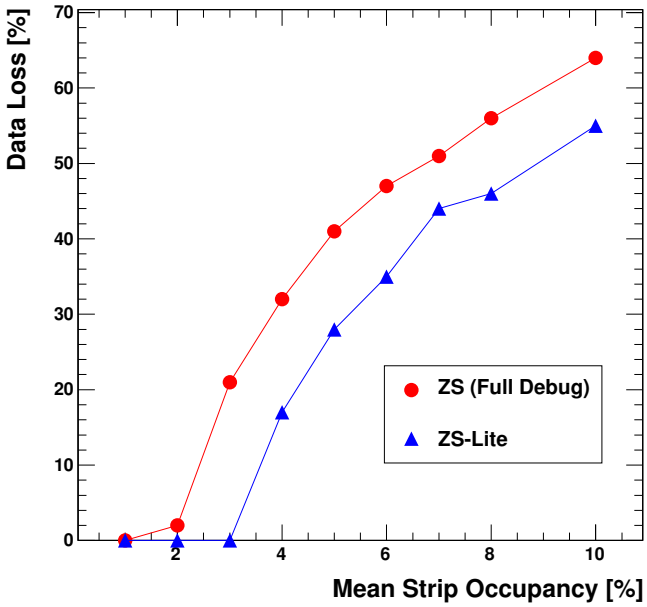


Figure 2.12: Fraction of triggers blocked by the tracker readout system due to rate throttling, as a function of the simulated event occupancy[72].

At low occupancies (<1%), the FEDs can operate in either mode without exerting backpressure. The only deadtime in the system was due to the APVe throttling the rate which was measured to be less than 0.2%. For an average occupancy of up to 3%, which would be the maximum expected in the innermost regions of the silicon tracker, the system was shown to operate without any data loss if the FED data format was set to ZS-Lite. In full debug mode, the trigger rate was observed to drop to below 80 kHz at 3% occupancy due to the backpressure exerted on the system by the S-LINK interface to the FRLs. The results show that the system is capable

[‡]The data loss here is due to deadtime and not to errors or corruption in the data stream, which were not observed.

of running over many events without error and that it will be able to maintain the 100 kHz average trigger rate, even at the highest occupancies (<3%), and absorb any stochastic fluctuations in the instantaneous data rate.

2.4.4 Commissioning of the Silicon Strip Tracker

During the integration and commissioning of the assembled tracker at the CMS Tracker Integration Facility at CERN, tests using the DAQ system described above were performed to qualify the full readout chain from the front end to the Filter Units. Using the data collected from the tests, studies were initiated in order to characterise the tracker performance at high trigger rates. These studies indicated the presence of an effect that had not been seen during the commissioning of the tracker with cosmic muons where the trigger rate was much lower.

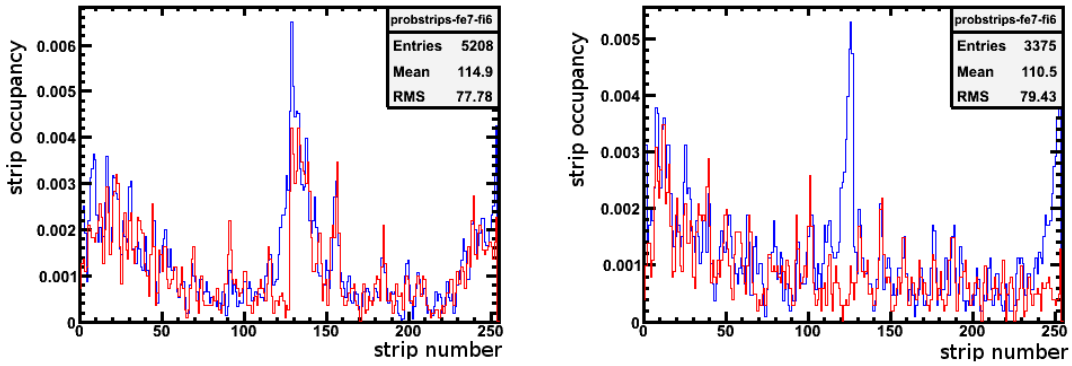


Figure 2.13: Average strip occupancy for a pair of APVs from a single fibre for low (Red) and high (Blue) trigger rates. There is a clear channel dependence. The figure on the right demonstrates that copper shielding between modules and the power bus[73] does not reduce the noise at high trigger rates[74].

When running the system with the LTC generating Poisson distributed triggers at rates above ∼30 kHz, it was observed that channels at the edge of all APV chips throughout the whole tracker experienced an elevated occupancy which increased with trigger rate. Figure 2.13 shows how the average strip occupancy varies by channel number for a single fibre from a TOB rod. It was determined that the effect could not be reduced using an insulation scheme previously used to eliminate

a similar noise problem experienced on TOB modules[73]. In addition, the high rate noise was only manifested when the triggers were Poisson distributed and not when triggers were uniformly distributed.

The noise was found to originate from the operation of the APV25 chip itself. If there are triggered data in the pipeline and the APV is not processing any other data for readout, the marked cell will be sampled onto the next stage in the APV25 at the beginning of a new readout cycle. The APSP stage differs slightly for APV operation in Peak or Deconvolution mode but the circuit can be described as high gain charge amplifier and a network of switched capacitors. In Deconvolution mode for example, the charge on each of the triggered pipeline cells is transferred to the APSP at the beginning of a new readout cycle. Three readout cycles are therefore required to process the data from a single trigger. On each cycle, a series of switches is closed and opened so that the samples are correctly weighted to perform the deconvolution operation. On a fourth cycle, the APSP sums the weighted charges and outputs the data to the Analogue Multiplexer where readout of the APV frame begins immediately.

Figure 2.14 shows that for specific bunch crossing separations between two triggers, the average pedestal output from the edge channels on an APV can fluctuate greatly. By reconstructing the internal operation of the APV, it was determined that specific processes occuring within the APV25 chip at the time corresponded to the trigger spacings where an increased noise was observed. In particular, it was noted that the large fluctuations at a trigger spacing of 157-158 bunch crossings coincided with the closing of switches within the APSP during the retrieval of the first pipeline sample. The other major noise contributions occur at times where the APSP retrieves the second and third samples, during the APSP sample and hold stage before the multiplexer and to a lesser extent during the digital header and tick mark outputs. It is thought that the source of the noise is from current spikes observed during these readout operations, coupling to the channel inputs on the APV via the power bonds before being sampled onto the pipeline. During testing of the APV25, this effect was not observed but it has been shown that by attaching the sensor to the APV, the channel susceptibility to the coupling is increased. At high trigger rates,

it is more likely that one of these noisy events is marked for readout at L1 therefore increasing the average occupancy and affecting detector performance.

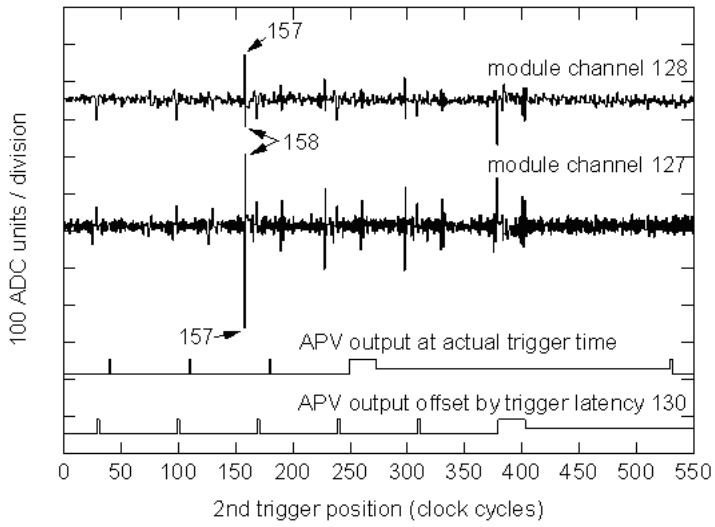


Figure 2.14: Pedestal fluctuations on the most affected APV channels showing a dependence on the spacing between two triggers (where the first trigger occurs at -3)[74]. After taking the L1 latency into account, it can be seen that the noise coincides with the readout operation of the first trigger, including the APSP sample and hold stage and the APV digital header at ∼380 clock cycle spacing. The large fluctuations at ∼158 clock cycle spacing correspond to the APSP readout of the first triggered pipeline sample.

The effect is estimated to generate an additional per event occupancy of ∼1%, which is not an issue for the tracker readout system and DAQ[74]. Instead, the large fraction of uniformly dispersed fake hits causes track reconstruction times to increase to unreasonable levels, especially within the HLT. In order to remove such events from the data stream, the APVe is being commissioned to identify the bunch crossings where the APV noise is expected to be the highest. Using the VHDL simulation of the APV, the APVe will veto triggers when it determines the beginning of an APSP cycle. This however, requires knowledge of the APVe to TCS cable latencies so that the veto can pre-empt the L1 trigger.

Chapter 3

The CMS L1 Trigger at SLHC

After several years of running, the LHC machine will be upgraded to increase its luminosity by an order of magnitude to $10^{35}\text{cm}^{-2}\text{s}^{-1}$. The current proposals are to achieve this in two phases. The Phase I project will see a new pixel detector system installed at CMS in time for a luminosity upgrade to 2-3$\times10^{34}\text{cm}^{-2}\text{s}^{-1}$. After years of radiation damage, and in order to cope with the increased occupancies and data rates after the Phase II upgrade, the entire CMS tracker will then have to be replaced.

3.1 Issues for the L1 Trigger at SLHC

The proposed luminosity upgrade implies that track and particle densities within CMS will be at least a factor of ten greater than those expected at the LHC. As discussed, a replacement tracker will require a higher bandwidth readout system to cope with the increased data generated from pileup. In addition, the L1 trigger system must also be upgraded in order to keep the trigger rate to no more than 100 kHz[10] if replacing the on-detector readout electronics is to be avoided.

The large number of minimum bias events per bunch crossing will affect the ability of the e/γ trigger to perform effective isolation against QCD background within the calorimeter, reducing rejection power for any given efficiency. However, even without

taking this into account, optimistic estimations of the current L1 e/γ algorithm at $10^{35}\text{cm}^{-2}\text{s}^{-1}$ shows that the E_T cut must be at least doubled to around 50 GeV if reasonable trigger rates for single isolated e/γ objects were to be achieved[75]*. Raising the L1 trigger thresholds in this way will reduce the potential of the SLHC upgrade by degrading sensitivity to any low mass discoveries made at the LHC.

For the L1 muon trigger, on the other hand, increasing the p_T threshold provides a negligible further reduction in rate as demonstrated by Figure 3.1. This is partly due to an irreducible combinatorial background, while at high p_T, the transverse momentum resolution is also worsened. Only including information from the tracker which increases the lever arm for the tracking measurement and significantly improves the p_T resolution of the track, allows better control of the muon trigger rate using the p_T threshold. At present, this is performed within the High Level Trigger.

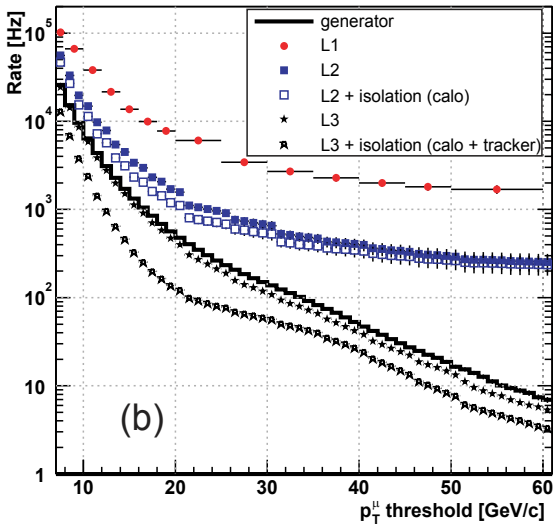


Figure 3.1: The HLT single-muon trigger rate as a function of applied p_T threshold at $10^{34}\text{cm}^{-2}\text{s}^{-1}$[35]. At L3, the addition of tracking information to the trigger allows a significant rate reduction over the L1 and L2 rates due to the improved momentum resolution, especially at high transverse momenta.

One solution could be to increase the use of combined object triggers at Level 1

*Preliminary simulations suggest that the proposed Phase I Level 1 Calorimeter Trigger upgrades may however allow a reduction in the single isolated e/γ trigger rate of a factor of ∼2.

to reduce the output rate, however this risks biasing the trigger strategy at SLHC and ignoring important decay channels. A more systematic approach would be to include all available information from the CMS detector in the L1 decision. Adding tracking objects to the trigger as is currently performed in the HLT would offer a method of reducing the L1 rate without sacrificing physics performance.

3.2 Use of Tracking Information in a L1 Trigger

Tracking information is introduced to the various trigger algorithms in different stages at HLT[35]. In the electron algorithm, hits in the pixel detector are first identified as being compatible with calorimeter deposits. It has been estimated that by correlating calorimeter electron candidates with pixel hits as demonstrated by Figure 3.2, backgrounds can be suppressed by a factor $\sim$10[76] at $10^{32}\text{cm}^{-2}\text{s}^{-1}$. Using hit information in this way provides rejection of the significant number of π^0's generated in proton collisions which pass the existing L1 trigger as well as distinguishing between electron and photon candidates. It also helps to reject calorimeter signals which originate from the products of secondary interactions within the tracker.

The electron HLT algorithm also uses hit information from the strip tracker in order to suppress backgrounds further and reduce the trigger rate[77]. The pixel hits matched to calorimeter clusters are used to seed the track reconstruction of the electron candidate. Following this, a cut is placed on the relation between the measured transverse energy of the ECAL and the measured p_T of the reconstructed track to lower the background rate by an additional factor of $\sim$3. Full tracking information is again used when regional track reconstruction around the HLT electron candidate is performed. This determines whether the electron is isolated within the tracker or not.

The τ-jet HLT algorithm requires hit information from both the pixel and strip detectors in the identification of τ-jet objects. Depending on the luminosity conditions, tracker data are utilised in the trigger decision in slightly different ways. Tracklets can be reconstructed using hits from the pixel detector only. Alternatively, the algorithm can reconstruct full tracks using pixel and silicon strip information albeit

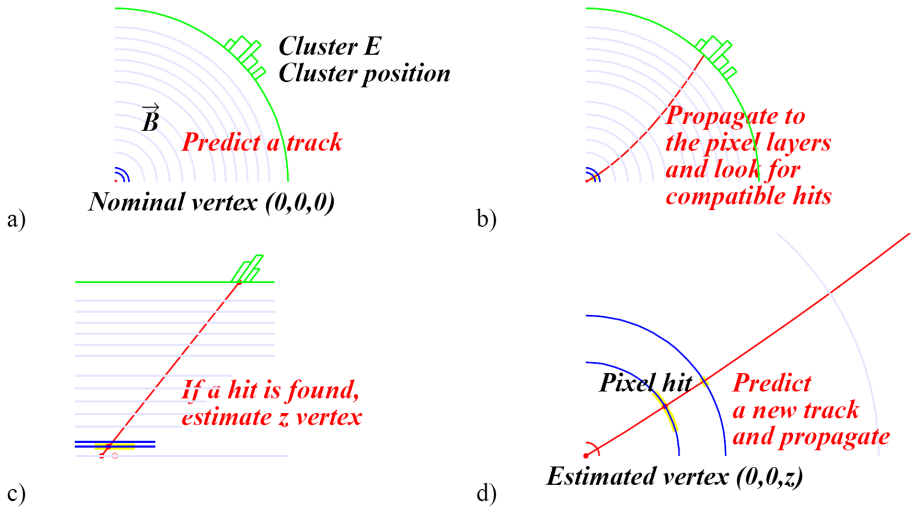


Figure 3.2: ECAL cluster matching to pixel hits in the electron HLT algorithm[78]. The energy weighted position of the electron super-cluster is used to search for compatible hits in either the first or second pixel layers based on the magnetic field and super-cluster energy. If a hit is found, a second compatible hit is searched for using tighter ϕ and z windows based on an estimated z vertex position.

within a region $(\Delta\eta,\Delta\phi)=(0.5,0.5)$ of the L1 τ-jet calorimeter candidate, of which there are up to four allowed per event. Since there is a significant QCD jet background, algorithm performance must be kept under consideration especially when the processing time for full track reconstruction can be substantial. Once the highest p_T track candidate within a cone ($\Delta R<0.2$) of the L1 τ-jet axis is identified, tracker isolation, where no tracks with a transverse momentum greater than 1 GeV/c are allowed within a ring around the leading track as illustrated in Figure 3.3, is performed. If the candidate passes the isolation criteria and silicon strip information has not been used up to this point, regional reconstruction is applied to improve the track p_T resolution and cut on the transverse momentum of the leading track. The isolation algorithm alone offers a factor of $\sim$20 reduction in the background rate while the ability to determine the p_T of the leading track cuts the HLT τ-jet trigger rate significantly[79].

It is clear that including tracking information in its various forms into a L1 trigger decision can offer a method of reducing the rate without sacrificing physics performance. Either by,

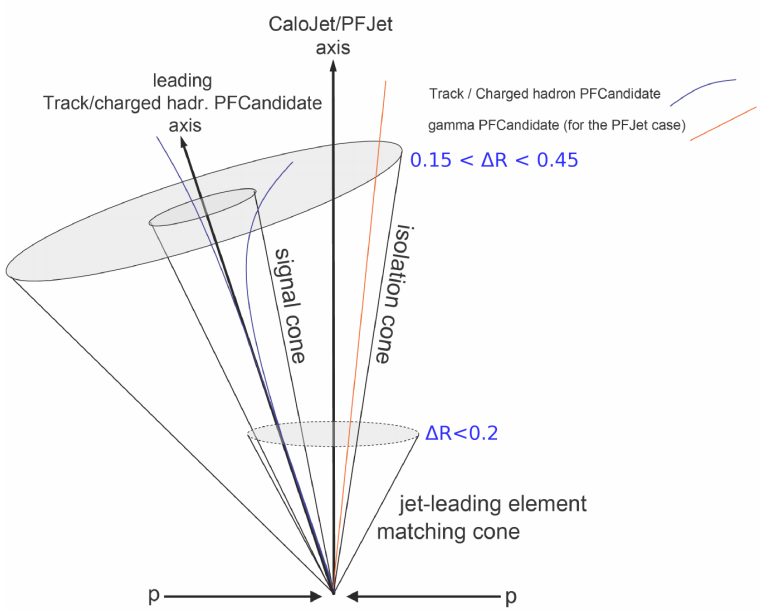


Figure 3.3: The τ-jet HLT isolation cone algorithm[79]. The leading track within a cone of $\Delta R<0.2$ about the L1 τ-jet axis is identified and an isolation requirement is performed within the annulus $0.15<\Delta R<0.45$ (variable) of the track. The leading track p_T cut is luminosity and algorithm dependent but is $\sim$5 GeV during pixel isolation and $\gtrsim$15 GeV after full track reconstruction[80].

- Pixel hit or tracklet matching between the tracker and calorimeters or muon stations. Resolution can be poor and matching could fail at high luminosity if an insufficient number of points are available for reconstruction.
- Pixel tracklet matching including track isolation.
- Regional silicon strip track reconstruction using pixel seeds matched to calorimeter or muon objects. This offers the best track position and momentum resolution for reducing the trigger rate using p_T cuts for example.
- Regional silicon strip track reconstruction including track isolation.

Providing tracking information to the L1 trigger on the other hand poses a significant challenge; for both the upgraded tracker and readout electronics as well as for the L1 trigger logic. Building pixel tracks as performed in the HLT τ-jet trigger requires at least $\sim$1 ms which far exceeds the allocated L1 latency of 6.4 μs[10] and while hit information from the silicon strip tracker can be utilised in the HLT algorithms described above, supplying these data to the L1 trigger within the current

architecture is impossible since signals are digitised, zero suppressed and clustered off detector in the tracker FEDs, as described in Chapter 2. In order to observe the L1 latency requirement, any tracking information provided to the trigger must be digitised and zero suppressed on-detector while any track object reconstruction must be simple enough to be implementable in a L1 trigger even if an architecture using advanced FPGA processing technology and fast data links can be built.

However, the greatest constraint on the amount and type of information the tracker can provide to the trigger will be the occupancy. It is clear that a significant benefit in reducing background rates is obtained by using hits from the pixel detector. On the other hand occupancies are expected to be the highest in this region. At a luminosity of $10^{35}\text{cm}^{-2}\text{s}^{-1}$ and a bunch crossing rate of 40 MHz, simulations indicate that a pixel layer positioned at a radius 10 cm from the beam will experience an occupancy of $\sim$10 hits per cm^2 every bunch crossing. This corresponds to a data rate of $\sim$10-20 $\text{Gbcm}^{-2}\text{s}^{-1}$ assuming a 16bit pixel coding scheme including additional inefficiencies in the data transfer over optical links. The current digital optical link technology envisaged for CMS at SLHC has a maximum bandwidth of 10 Gb/s[81] meaning that the link density and power requirements to read out data for a L1 trigger would be enormous.

As a result, it will only be possible to pass selected information from the tracker to the trigger at Level 1. One solution would be to read out regions in the tracker using information provided by calorimeter or muon objects. However, the resolution of the outer detectors in pointing to the tracker and the length of the luminous region means that the hit densities will be too great[†]. Alternatively, a novel method of reducing the data rate on detector by selection of interesting tracking information must be developed.

[†]With a luminous region of $\pm$16 cm in z and a maximal ECAL pointing resolution at L1 of $\Delta\phi$=0.087, an area of $\sim$25.4 cm^2, or $\sim$250 hits, for a pixel layer at radius 10 cm will be required for readout. With such occupancies, over multiple layers, track reconstruction within the L1 latency will be difficult.

3.2.1 Stacked Tracking Trigger Layers

Collisions at the LHC are predicted to produce a large number of low momentum particles that make up a significant fraction of hit data generated by the tracker. Figure 3.4 shows the averaged p_T spectrum of all charged particles leaving hits in a sensitive layer placed at 25 cm from the beam axis in a typical SLHC event. The right hand figure indicates that ∼85% of these tracks have a transverse momentum of <1 GeV/c. Charged particles with momentum p_T<0.7 GeV/c are considered

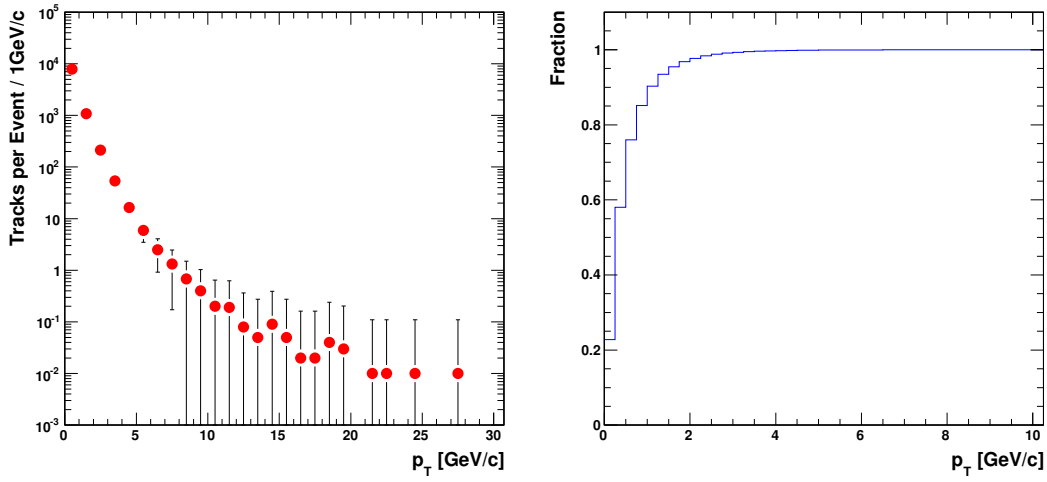


Figure 3.4: The p_T spectrum (normalised per event) for all minimum bias particles that leave hits in a sensitive layer placed at a radius of 25 cm and at an average pileup of 400 p-p interactions per event. The right hand plot indicates, as a fraction of all tracks, the number of tracks with transverse momentum less than the given p_T. Events were simulated within a magnetic field of 4 T and a coverage of $|\eta| < 2.5$.

uninteresting for the purposes of triggering since they fail to reach the outer subdetectors due to the bending power of the 4 T magnetic field. These particles tend to generate multiple hits over many events as they loop within the tracker so eliminating the hits produced by these tracks would reduce the data volume significantly. The simplest method of calculating the p_T of an intersecting charged particle would be to correlate hits between two or more consecutive tracking layers as is performed in current HLT algorithms. As discussed, this method of determining p_T would not be viable using tracking layers with $\mathcal{O}$(cm) separation since the bandwidth requirements to correlate hit information from two layers would be far too large and

impractical. In addition, the excessive layer hit densities expected at SLHC means that the combinatorics produced by correlation over two layers would be considerable. Proposals by J.Jones et al.[62, 82, 83, 84] however, have demonstrated that correlating hits from two closely separated pixel layers (of order $\mathcal{O}$(mm)) is sufficient to distinguish the correct hit combinations and determine if the transverse momentum of the track is above a minimum p_T. The benefit of matching hits over minimally separated layers is that the high bandwidth requirements of transferring information for correlation takes places over a small distance and hence power consumption and material contribution to the tracker can be reduced significantly.

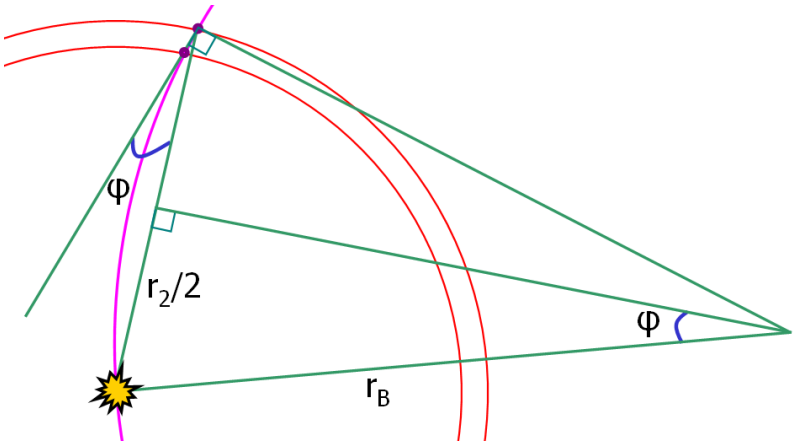


Figure 3.5: Determination of track transverse momentum by reconstruction of the crossing angle ϕ using a pair of closely separated layers at radii r_1 and r_2.[62]

In the CMS tracker, an approximate relationship between the distance traversed by a track and its p_T can be written as:

$$s \simeq \frac{0.6 l r_2}{p_T} \tag{3.1}$$

where s is the r-ϕ distance travelled by the track, l is the radial separation between layers and r_2 is the radius of the outer tracking layer in metres while p_T is measured in GeV/c (Figure 3.5). For two layers separated by 1 mm at a radius of 25 cm from the beam axis, a 1 GeV/c track travels ~150 μm in r-ϕ. The r-ϕ pitch of the current CMS pixel barrel layers is 100 μm. A stacked tracking layer using such pixel sensors would therefore have sufficiently fine pitch to be able to discriminate on

the transverse momentum of passing particles. This can be achieved by correctly matching the hits from each layer and calculating the track crossing angle relative to the surface of the sensor as Figure 3.6 demonstrates.

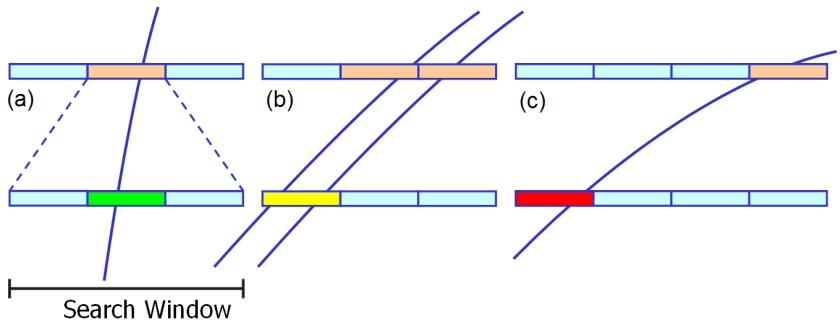


Figure 3.6: Reconstruction of the track crossing angle using a stacked pixel layer[62]. A high p_T track will intersect the layer virtually perpendicular to the sensor. In this case (a), the hits fall within the search window and the correlation reveals a high p_T track. Tracks with lower transverse momenta intersect the layer at shallower angles and hence may (b) either pass or fail the correlation depending on the impact point on the sensor or (c) or always fail the correlation if the angle of incidence is too small.

Hits in the stacked pixel layer can be correlated using a search window; if a pair of hits fall within the window, this indicates the presence of a high p_T track. Low p_T tracks may occasionally pass the correlation cut depending on their impact point on the pixel sensor. In addition, effects such as charge sharing between pixels and the thickness of the sensor will also influence the number of low transverse momentum particles passing the cut.

Monte-Carlo studies[62], using an idealised geometry and not including material effects, have demonstrated the ability of a stacked pixel layer in discriminating against tracks with low p_T. Figure 3.7 shows how the probability of a track passing the p_T cut varies with its transverse momentum for different sensor separations. The width of the transition region in this figure, between where no tracks and all tracks are accepted is determined by the pixel pitch, sensor separation and the size of the search window.

The pixel dimensions assumed in this simulation can be considered optimistic given the limit of current hybrid pixel technology used in CMS. A stacked layer consisting

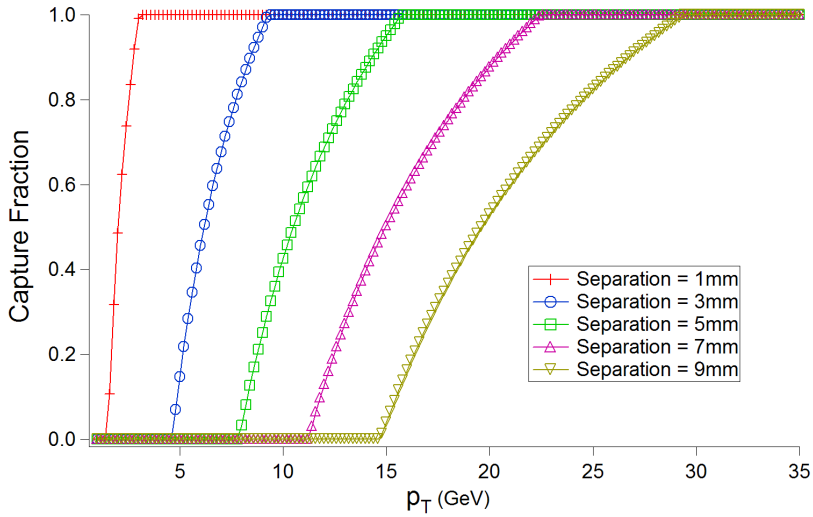


Figure 3.7: p_T discrimination performance of an ideal stacked pixel layer using a toy Monte-Carlo simulation[62]. Results are for a layer at 10 cm with pixel dimensions of $20\times50\,\mu m^2$ ($r\phi\times z$) and a search window of ±1 pixel.

of small sensor elements will also contribute significantly to the power consumption of the tracker unless novel low-power technologies such as Monolithic Active Pixel Sensors (MAPS) can be developed for CMS in time for an upgrade. Consequently, it is necessary to continue these studies and investigate the performance of stacked tracking layers in a high luminosity environment as a function of pixel pitch, search window size and radius.

In addition, more realistic simulations are required to be able to demonstrate the viability of stacked tracking layers for triggering on high p_T tracks at SLHC. Previous studies have neglected material effects such as secondary interactions and multiple scattering, sensor effects such as the finite thickness of the sensor and Lorentz drift and have not considered how a realistic sensor configuration affects triggering performance. It is also important to investigate the robustness of the tracking trigger with respect to pileup and different event topologies. Finally, realistic simulations would be able to help estimate the power consumption, material contribution and cost of a concept stacked tracking layer since these are considered to be the main concerns with building and implementing such a design.

3.3 Simulation of the Stacked Tracking Trigger Layer

3.3.1 The Stacked Layer Geometry

In order to determine the performance and viability of stacked tracking trigger layers in a future tracker at CMS, detailed simulations on a concept layer geometry have been carried out. The basic layer geometry is formed of a number of modules, each covering an area of a few cm^2. A module comprises a pair of pixel sensors stacked one above the other (as detailed in Section 3.2.1) with a separation of $\mathcal{O}$(mm). A ladder is a set of modules arranged end to end along z to form a single structure that can be replicated over r-ϕ to make a layer. Figure 3.8 defines the strawman module layout and demonstrates the inclusion of cooling, cabling, structural support and electronics in the simulation. In the absence of firm proposals for a module layout, the strawman layer has been constructed to be as generic but as realistic as possible, specifically in terms of material. A realistic estimate of the material added to the tracker by a stacked layer is important so that effects such as multiple scattering of hadrons, electron bremsstrahlung and photon conversions are correctly simulated.

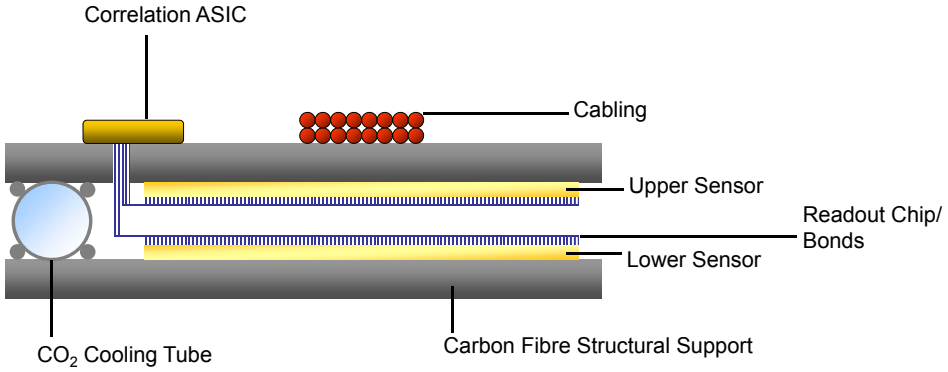


Figure 3.8: Strawman module layout in r-ϕ as used in simulations including cooling, cabling and structural support.

While the strawman layout is essentially modeled on a pair of the pixel barrel modules used in the current CMS detector, the mechanics and cooling systems are modified so that they are shared between sensors. In addition, the upgrades proposed

for the pixel detector in Phase I have also been included in the module layout[41]. For example, the cooling elements have been modified to simulate the usage of a low mass CO_2 cooling system designed to reduce contribution of material to the tracker.

One of the features of the current CMS pixel system is that it has been designed to allow reconstruction of secondary vertices in a high occupancy environment by use of small pixel cell sizes measuring 100 μm × 150 μm resolving hit positions to within ∼20 μm. This means that in total there are more than 60 million channels in the pixel detector dissipating 3.5 kW of power. While a point resolution of this order might be useful in a trigger layer to identify vertices from the ∼400 interactions every bunch crossing in a worst-case SLHC scenario, a stacked pixel layer at 25 cm covering the required region $|\eta|$ <2.5 would add almost 800 million channels to the tracker. For such a system, simply the pixel front end would require over 20 kW of power. Increasing the pixel pitch in z to $\mathcal{O}$(mm) reduces the channel count while keeping occupancies low (see Section 4.1) and identification of high p_T tracks using stacked sensors is still possible providing the r-ϕ pitch is kept to <100 μm.

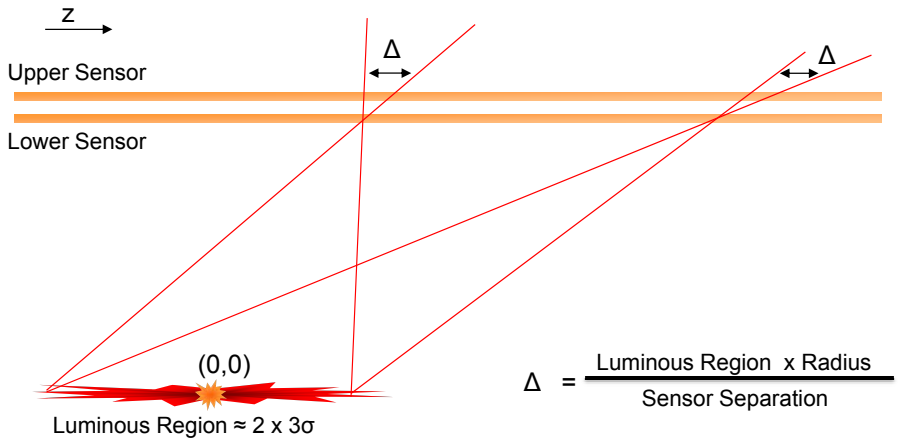


Figure 3.9: Illustration demonstrating the effect of the size of the luminous region on the spread of hit positions (Δ) in a stacked layer.

The lower limit on the z pitch for a stacked layer should be determined by the size of the luminous region at SLHC. It is expected that the longitudinal interaction

region could be as large as ±16 cm[13][‡]. Figure 3.9 demonstrates the effect the z interaction point has on the hit positions within a stacked layer.

The variable Δ denotes the the range of separations along z between hits in stacked sensors, due to the size of the luminous region. While independent of η, it is subject to the radius of the stacked layer and the sensor separation. For example, a layer at 25 cm and a sensor separation of 2 mm gives rise to a range Δ=2.5 mm. In order to maintain the simplicity of the correlation algorithm it is necessary that hit matching takes place over a minimal number of pixels. Increasing the z pitch to match the separation Δ means that only neighbouring pixels will have to be checked for correlation motivating the choice for the pixel length and maximal sensor separation simulated (Table 3.1).

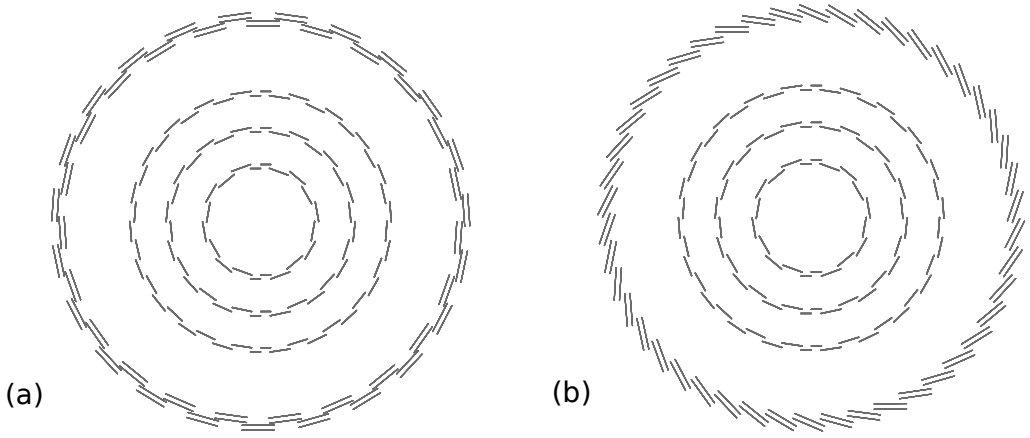


Figure 3.10: An r-ϕ view of the current CMS pixel sensor arrangement and one stacked pixel layer; (a) illustrates the stacked sensor ladders arranged in an in-out layout, (b) demonstrates the tilted arrangement of stacked sensor ladders in r-ϕ.

While the sensor technology for a possible stacked pixel layer has not yet been defined, the current consensus is that the sensors will either employ n^+ implants on n-type bulk as in the current pixel system or n^+ implants on p-type bulk. Studies using p-on-n sensors indicate that for the high fluences expected at SLHC in the region 20 cm>r>40 cm where stacked pixel layers are likely to be placed, charge collection efficiency is severely reduced after type inversion[45]. This effect due

[‡]Assuming a 3σ variation for the 40 MHz scenario at SLHC. The 20 MHz scenario is likely to only give rise to a luminous region of ±11 cm.

to radiation damage is not apparent when the signal charge is collected using n^+ implants.

With this choice of sensor technology, electrons are the predominant charge carrier and hence the pixel detector has a high sensitivity to the Lorentz drift in the presence of the magnetic field. This benefits tracking in the current analogue pixel system, since the hit position within active pixel clusters can be interpolated and position resolutions can be improved. However, not only is it unlikely that a fast, low power hit interpolation scheme could be implemented on the stacked pixel layer at the SLHC bunch crossing rate but due to constraints on power and complexity, the layer will have to employ a binary readout system and this will worsen the resolution of the calculated hit position. It is therefore important to study the effect of how Lorentz drift impacts on the trigger performance of the stacked layer.

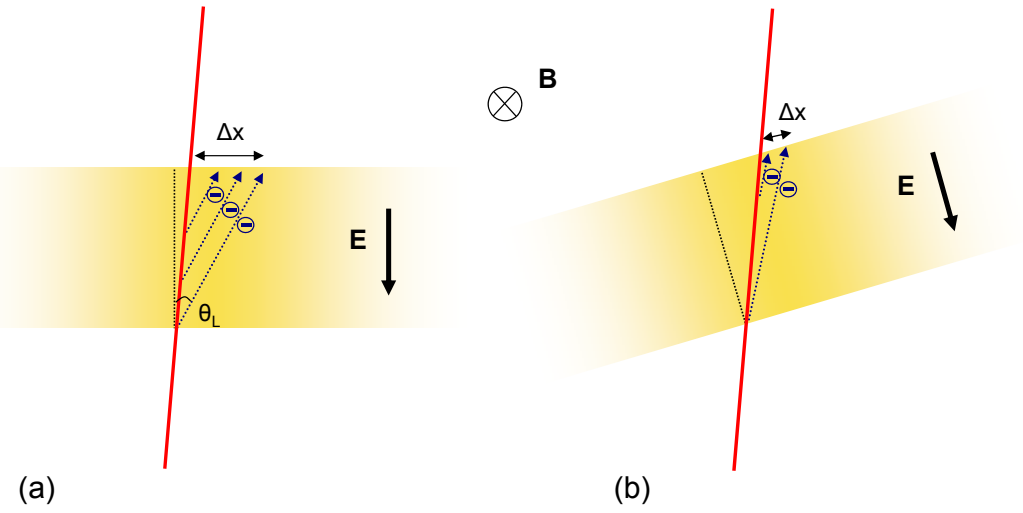


Figure 3.11: Illustration of the effect of Lorentz drift in sensors where electrons are the majority charge carrier; (a) in the untilted sensor case, (b) where the sensor is tilted to coincide with the drift angle. In the tilted case, the resulting signal spread (Δx) is smaller than it is for untilted sensors. The direction of the magnetic field (**B**) is into the page.

Figure 3.10 illustrates two variants of the stacked pixel layer. One geometry arranges sensors in an in-out layout as is currently employed in the pixel detector. In this case the effect of Lorentz drift is observed, leading to large cluster widths. The second geometry arranges sensors which have been tilted with respect to the radial

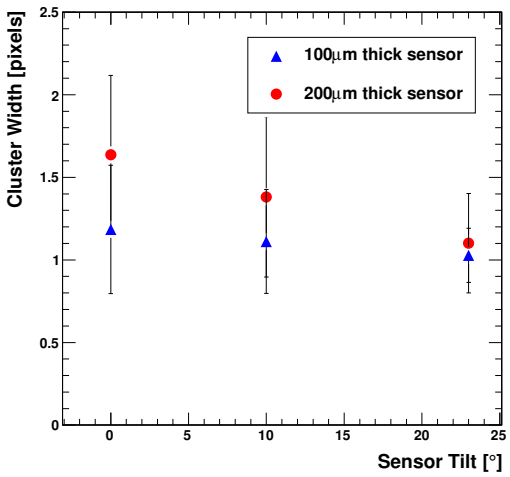


Figure 3.12: Simulated variation of cluster width in r-ϕ as a function of sensor thickness and sensor tilt angle relative to the normal of the radial axis. A tilt of 23° is equivalent to the Lorentz angle for electrons in current CMS pixel sensors operating under a bias voltage of V_{bias}=150 V. Results are for p_T=4 GeV/c muon tracks impinging on a pixel layer at 25 cm with 100 μm pitch in r-ϕ.

axis. Figure 3.11 demonstrates how the effect of Lorentz drift can be negated by tilting the sensors in this fashion.

Figure 3.12 indicates how the cluster width varies with sensor tilt and sensor thickness. The current pixel detector utilises 285 μm thick silicon sensors. However, it would be beneficial if the stacked pixel layers employed thinner sensors at SLHC. Firstly, thinner sensors will evidently generate smaller clusters due to a reduced Lorentz effect. In addition, the leakage current across the sensor will also be reduced which would help to lower the power requirements of the tracker. The disadvantage of a thinner sensor is that a smaller signal is collected at the pixel. However, with a signal charge of 22,000 electrons/MIP against a pixel threshold of ~2800 electrons in the present CMS pixel sensors[85] this may not pose a significant problem[§].

The stacked layer geometry includes overlaps between modules so as to provide hermetic coverage over the full tracker acceptance. In the implementation of a stack it is important that a high p_T track leaves hits in both the upper and lower sensors

[§]The effect of radiation damage to the silicon sensors is not considered here. With increased exposure to the high particle fluxes expected at SLHC, sensor bias voltages will have to be raised in order to maintain signal efficiency. While this will reduce the effect of Lorentz drift, it will also increase leakage current across the sensor

of the same module otherwise the complexity of the system greatly increases. The number of ladders in the layer must be adjusted to provide enough overlap in ϕ to achieve this. Modules are also arranged in z as illustrated in Figure 3.13 so as to allow for overlaps between sensors along the ladder.

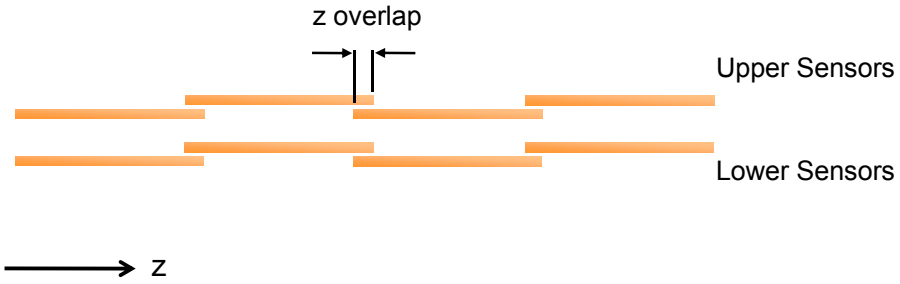


Figure 3.13: An overview of the sensor layout in r-z, demonstrating the overlaps between sensors in z.

However, since the angle of incidence for tracks crossing the stack increases with η, to avoid making hit correlations between adjacent modules an additional offset between the upper and lower sensors in z as a function of z is also required. This offset is demonstrated in Figure 3.14. For a stacked layer at 25 cm and a sensor separation of 1 mm, the offset must increase by ∼1.0 mm every 26 cm along the ladder length from z=0. This means that only 6 module variants are required and the total offset between upper and lower sensors at $|\eta|$=2.5 reaches ∼0.6 cm.

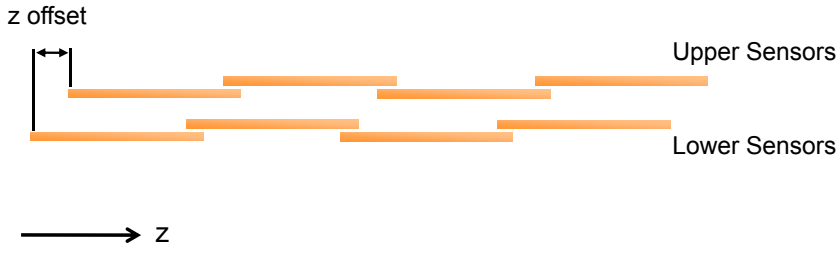


Figure 3.14: An overview of the sensor layout in r-z, demonstrating an applied offset between the upper and lower sensor layers to compensate for the increasing angle of incidence with η.

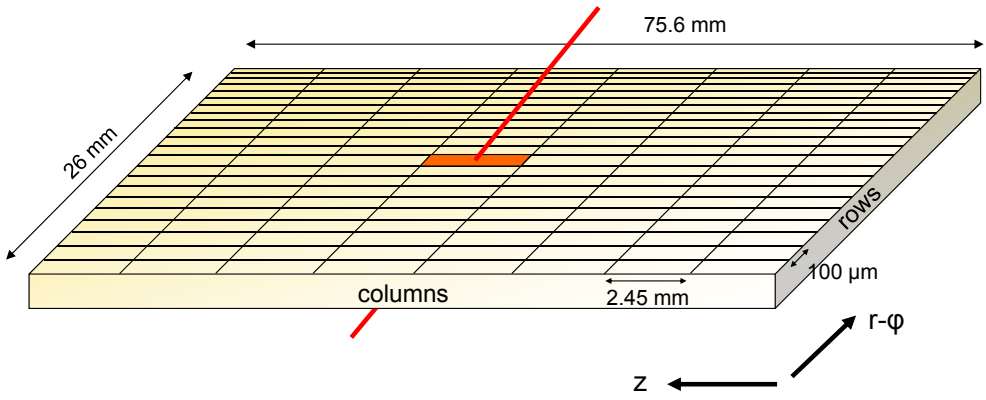


Figure 3.15: A schematic of the sensor layout and dimensions. Pixel rows are defined along the r-ϕ axis and columns along z.

Parameter	Default Value	
Layer Radius	25 cm	35 cm
Ladders/Layer	72	106
Modules/Ladder	44	60
Sensor Dimensions (r-ϕ × z)	28 mm × 75.6 mm	28 mm × 75.6 mm
Sensor Pitch (r-ϕ × z)	100 μm × 2.45 mm	100 μm × 2.45 mm
Rows × Columns (r-ϕ × z)	256 × 30	256 × 30
Sensor Thickness	**100** μm, 200 μm	**100** μm, 200 μm
Sensor Tilt	**0**°, 23°	**0**°, 23°
Sensor Separation	0.5, **1.0**, 1.5, 2.0 mm	0.5, **1.0**, 1.5, 2.0 mm
Layer Length	316 cm	430 cm
Layer Coverage	$\|\eta\| < 2.54$	$\|\eta\| < 2.51$
Module Variants	6	6
Total Modules	3,168	6,360
Total Sensors	6,336	12,720
Total Channels	48,660,480	97,689,600

Table 3.1: List of sensor and layer parameters used in simulations. Selections in bold are used by default unless otherwise indicated.

The dimensions of the sensors (28 mm x 75.6 mm) are similar to those of the current pixel sensors. This allows for 256 x 30 pixels per sensor at a pitch of 100 μm in r-ϕ and 2.45 mm in z (see Figure 3.15 for a definition of pixel row and column). A list of the default layer parameters used in the simulation is provided in Table 3.1.

3.3.2 The Concept Tracker Geometry

In order to estimate the performance of a trigger layer at SLHC, the simulated tracker geometry must be as realistic as possible, especially with regard to material in the inner layers. While a definite tracker design for the luminosity upgrade has not been decided upon at this stage, the main features required in a new tracker with trigger layers can be envisioned in a concept geometry. These include:

- A low mass, low power inner pixel detector for vertex identification and pixel seeding. Providing sensors and readout electronics can be made to withstand the hostile radiation environment at SLHC, a pixel system based on the current detector and occupying a similar region is foreseen. The pixel detector must contribute as little mass as possible since material interactions in the inner layers will be detrimental to any track triggering at larger radii. However, it is also likely that a fourth inner pixel barrel layer and third pixel endcap disk pair will be required. Simulation studies on the performance of the current tracker geometry within heavy ion (Pb-Pb) events at LHC indicate that the particle flux in the pixel detector will be similar to that expected at SLHC[42]. Track reconstruction can be achieved by seeding tracks with three pixel hits instead of two in order to maintain a low fake rate, although efficiency is reduced by 10% due to the geometric acceptance of the pixel system. The addition of an extra layer should recover any efficiency losses due to triplet seeding.

- Stacked pixel trigger layer or layers at intermediate radii. The material contribution by trigger layers to the tracker precludes its installation within the pixel system since multiple scattering will reduce the achievable track momentum resolution. In addition, the angular resolution of a layer with 100 μm pitch

at 25 cm is equivalent to a layer with 40 μm pitch at 10 cm. Consequently, unless the pixel pitch can be reduced to beyond current technological limits, it becomes impossible to distinguish between tracks at low momenta using stacked layers placed closer to the beam pipe. On the other hand, the number of channels (and hence power) scales as r^2 for equivalent pixel sizes and equal coverage. Additionally, due to tracker volume constraints, the maximum radius a single long barrel layer can be placed at before endcaps are required is 50 cm.

- An outer tracker, most probably based on the current silicon strip detector consisting of barrel layers and endcaps. This area of the tracker, where occupancies and radiation levels are lowest, is the most flexible for change and a decision on the final layout, sensor technology and granularity will be based on adopting the lowest power, lowest mass and least complex option.

A concept tracker geometry has been constructed within the CMS simulation software, replacing the current tracker. This geometry layout, defined in Figure 3.16, comprises a four layer barrel, three endcap disk pair inner pixel detector, two stacked pixel trigger layers as previously described in Section 3.3.1 and an outer silicon microstrip tracker.

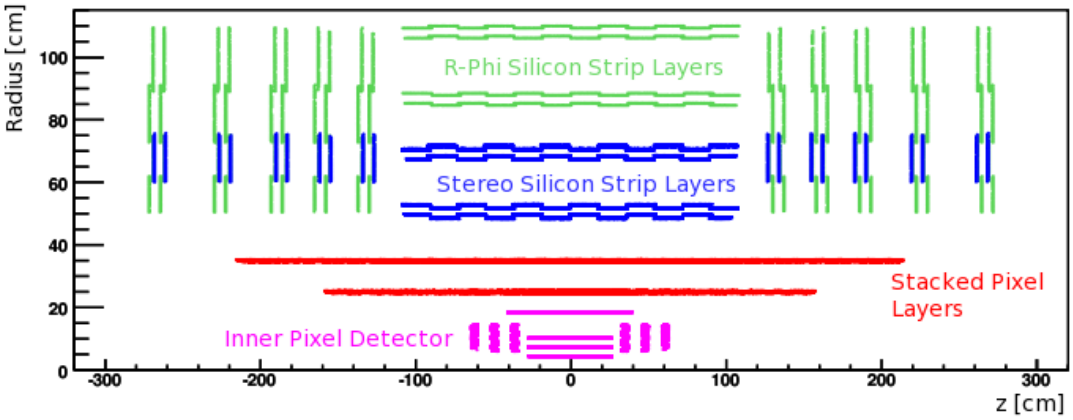


Figure 3.16: The concept tracker geometry, providing coverage over the range $|\eta|$ <2.5. The simulated geometry also includes inactive material including cabling and cooling services, structural support and readout electronics, all of which are not displayed here.

The key point in the simulation of the concept geometry with regards to estimating

the performance of the stacked layer is that a realistic estimate of the material in the inner tracker is made. While an extra pixel layer has been included, the pixel detector has also been updated to include the proposed material reductions for the Phase I upgrade such as a low mass CO_2 cooling system and lighter carbon fibre support structure[41]. It is also possible that the material in the inner tracker could be reduced even further for the Phase II upgrade. However as a worst-case material estimate, the concept geometry can still be used to measure trigger performance.

3.4 Simulation Software

Simulations of physics events within the current CMS detector, including detector response and event reconstruction are carried out using a set of software packages linked together into a common framework. The software suite, known as CMSSW, is written in C++ and is based on the Event Data Model (EDM) whereby event processing takes place using a sequence of independently operating modules. Persistent data from each module are held within a container object called the Event. CMSSW is able to interface with external Monte-Carlo (MC) generators such as PYTHIA[86] and Herwig which are used to simulate single proton-proton collisions at 14 TeV. The MC generator passes the set of stable outgoing particles produced in the interaction to the Event container after which they are propagated though the detector volume. Alternatively, raw detector data can be processed by CMSSW so that higher level objects such as tracks or even physical objects such as electrons can be constructed for use in analysis. This reconstruction stage is equivalent for both simulated and real data.

3.4.1 Detector Simulation

The simulation stage follows directly after the generation of the Monte-Carlo data. Initially, the collision vertex and its associated products are displaced according to a smearing profile that can be specified by the user. Following this, the decay products are propagated through the detector and any applied magnetic field. All

particle interactions with matter, including electromagnetic, hadronic and decay processes, are handled in detail by the GEANT4[87] toolkit which is also interfaced with the CMSSW package in a manner which is transparent to the user. To simulate the detector, a geometry must be provided to the framework so that GEANT can describe particle propagation accurately. In CMSSW this geometry is constructed using Detector Description Language (DDL), a CMS specific XML format, which sets up a family of logical volumes in 3D space that are each tagged with a material description including object density and atomic composition. To define the current CMS detector geometry in detail, the software requires over 400 XML files describing objects down to the sizes and positions of individual readout chips and kapton cables. Active components such as silicon sensors and calorimeter crystals are labelled sensitive volumes so that energy loss, time of flight and hit positions of passing particles are recorded for further processing. The EDM Event identifies these interactions as SimHits.

To model the effect of tracker pileup, pre-simulated data from multiple single inelastic proton collisions are superimposed into the Event. In addition, hit information from previous (-5) and subsequent (+3) bunch crossings is included to emulate the effect of out-of-time pileup during detector operation. Hits from previous bunch crossings can contribute to out-of-time pileup if they arise from low momentum looping tracks within the tracking volume. Only the SimHits with a registered time of flight falling within the 25 ns bunch crossing under consideration are stored in the Event, while all others are discarded. Hits from subsequent bunch crossings are also included due to the time resolution of the tracker readout electronics.

The detector response is modelled using dedicated digitisation modules which convert SimHits into the electronic signals expected to be output from the CMS Front End Drivers (FEDs). These signals are generically known as Digis. In the tracker for example, the energy deposition associated with each SimHit is converted into electrons (or holes) whose drift is simulated through the local electric and magnetic fields to the surface of the sensor. The number of charge carriers to be drifted is fluctuated over the thickness of the silicon sensor and the direction of the track that left the hit. Diffusion of charge carriers along the sensor plane is assumed to be a

Gaussian with a spread proportional to the square of the drift length while the drift direction is given by the Lorentz angle in a 4 T magnetic field (23° and 7° for the pixel and strip sensors respectively).

By dividing the collection plane into discrete units matching the granularity of the physical sensor and integrating the charge arriving at each unit, a signal is induced on the pixel or strip. Finally, the effect of the readout electronics is emulated by manipulating the signal charge and applying miscalibrations, thresholds and noise sources before converting it to a digitised hit described in terms of ADC counts. For example, in the pixel system, a threshold of ∼2000-2500 electrons is applied so as to zero suppress the data, as is performed in the pixel ROC. The total electronic noise (from the Read Out Chip, Token Bit Manager and Analogue Optical Hybrid etc.) is simulated by approximating the contribution as a Gaussian centred at zero with a spread of 500 electrons. Data losses due to finite buffer sizes and fixed readout latencies as a function of luminosity, as well as inefficiencies due to unbonded and noisy pixels are also simulated.

3.4.2 Modelling the Concept Tracker Geometry

The concept geometry with stacked pixel layers as described in Sections 3.3.1 and 3.3.2 has been modelled using DDL for use within CMSSW. While the outer barrel is essentially based on the current tracker TOB geometry and although significant modifications were required to represent the simpler endcaps and extra pixel layers, the stacked pixel layers are a completely new design. The geometry files describing these layers have been coded to be as configurable as possible so that simulations of different layouts are easier to achieve. Parameters such as layer radius and sensor tilt are adjustable, subsequently allowing the details of the geometry to remain transparent to the end user. Such configurability is not available in the CMSSW implementation of the current CMS detector.

The IGUANA visualisation tool, provided by CMSSW, can be used to inspect the geometry layout after each modification. Figure 3.17 illustrates the modelling of a stacked layer in a tilted configuration. Volumes describing passive material, such as

cabling, cooling and readout electronics, are clearly evident. Using the model of the stacked layer, GEANT is able to determine the amount of material it contributes to the tracker.

Figure 3.17: IGUANA representation of the stacked pixel layer geometry in a tilted configuration.

The material budget in terms of radiation lengths for the concept tracker geometry is provided in Figure 3.18. Compared to that of the present CMS tracker there is a large reduction in material, especially in the mid-high pseudorapidity range (Figure 1.14). The majority of the cut is due to a smaller TEC and the removal of the TIB, TID and associated readout electronics and cooling which is currently routed between the Outer Barrel and Endcaps. In addition, the inner pixel detector has been modelled with the thinner structural carbon fibre support structures and smaller lightweight CO_2 filled cooling tubes planned for the next upgrade.

The average thickness of a stacked pixel layer in radiation lengths, measured to be $\sim 0.03\,X_0$, is less than double that of a single upgraded pixel layer since cooling and mechanics are shared between two sensors. Because the stacks extend out as barrel layers only, non-sensitive material can also be moved further out in η so as to avoid inhomogeneities in the material distribution seen in the present tracker. It should be noted however that the geometry only provides an estimate of what an upgraded

tracker using this layout might contribute in terms of material, especially without a definite proposal for the stacked layer design. For example, the layer does not include possible additional material due to the provision of power to modules (DC-DC converters) or high speed optical link transceivers and associated electronics.

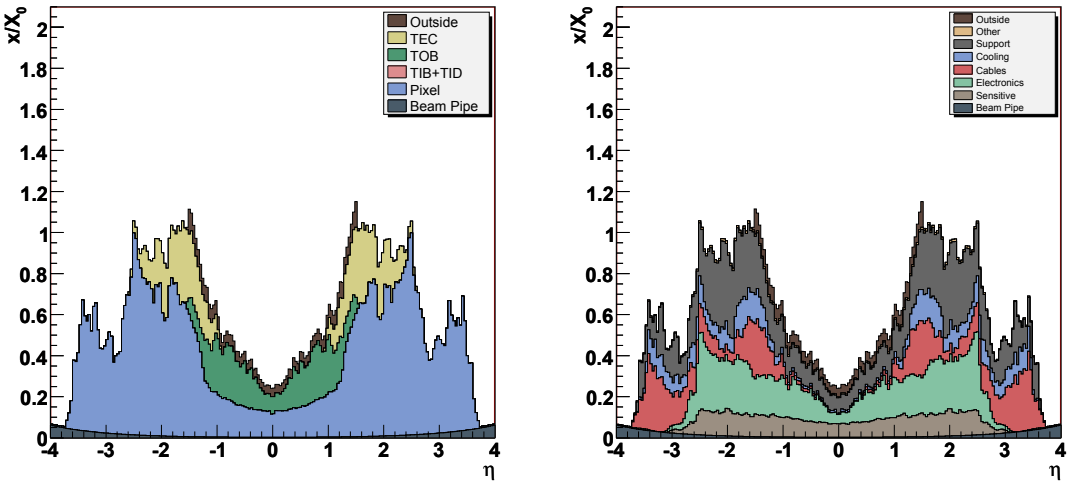


Figure 3.18: Material contribution of the tracker in radiation lengths as a function of pseudorapidity by subdetector (left) and by material type (right).

3.4.3 Fast Simulation

The time to run the simulation process described above increases quadratically with pileup as demonstrated by Table 3.2. Increasing the pileup to above 200 interactions per bunch crossing in the current software leads to memory leaks which causes the simulation to fail. Since simulating events can be extremely time consuming, an alternative method, specific to CMS software, has been developed. The Fast Simulation (FastSim) package is a CMSSW integrated tool that allows accurate simulation and reconstruction of events within CMS but with minimal CPU time and memory usage. This is made possible by parametrising the detector and detector effects, making simplifying assumptions on operating conditions and optimising software code. The Fast simulation and reconstruction chain is designed to be able to run in parallel with the standard GEANT-based (Full) simulation to achieve the same

Pileup (average) (Interactions/BX)	Full Simulation CPU Time (s)	Fast Simulation CPU Time (s)
0	130	0.3
20	210	3.4
100	1220	11.9
200	4350	25.1
400		55.2

Table 3.2: Average CPU timings (seconds) per event for both the Fast and Full (GEANT) Simulation algorithms. Includes Monte-Carlo generation of a p_T=50 GeV/c di-muon pair, pileup, simulation and digitisation with pixel noise within the concept tracker geometry only. Statistics for 100 events using a quad core 2.66 GHz Xeon CPU with 8 GB RAM and the CMSSW_2_2_6 software package.

results. In fact, the FastSim parametrisations are tuned by performing detailed comparisons with the Full Simulation. For its relevance to the simulations performed in this thesis, only the part of the software relating to the parametrisation of the GEANT simulation is discussed here.

The FastSim package is able take Monte-Carlo generated particles, apply vertex smearing and propagate the tracks through the tracker volume and magnetic field as in the Full Simulation. However, the FastSim only propagates a subset of the particles in the event and interacts them with a set of infinitely thin cylinders and disks, coded to emulate the real geometry, with average radiation thicknesses of each provided by the user. For example, low p_T particles below a default cut (300 MeV/c) are not propagated by the FastSim. Most of the important interaction processes are taken into account by the Fast Simulation. Photon conversions, bremsstrahlung, multiple Coulomb scattering and energy loss by ionisation are simulated analytically using probability density functions.

Nuclear interactions of hadrons in material are instead modelled by assuming pion-proton interactions and using experimental measurements of the inelastic cross section as a function of pion energy. Production of δ rays, however, is not taken into account by the FastSim. The result of these simplifications reduces simulation times significantly as Table 3.2 indicates.

Figure 3.19 demonstrates the distribution of interaction hits in the concept geometry for the Fast and Full simulations. To correctly simulate the hit positions, the Full Simulation detector geometry is called and the hit is assigned to the nearest compatible sensors on that layer by propagating the track. By measuring the average bremsstrahlung by electrons in each point of the tracker for both Fast and Full Simulations, the average thickness of each layer in radiation lengths is tuned to reproduce the correct amount of material.

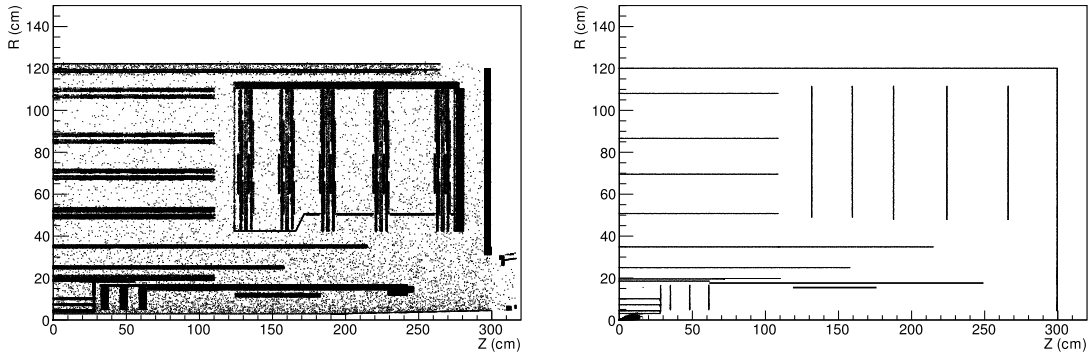


Figure 3.19: Distribution of vertices from electron bremsstrahlung for the Full (left) and Fast (right) simulations demonstrating the comparative interaction geometries.

The FastSim is also able to add pileup to events at luminosities higher than the Full Simulation can handle. However, the Fast Simulation fails to take into account pileup from previous or subsequent bunch crossings and hence out-of-time pileup is not correctly modelled.

3.4.4 Full and Fast Simulation Occupancy Measurements

Figure 3.20 indicates how the average occupancy as defined in equation (4.1) in a stacked pixel layer at 25 cm varies with pileup for both the Fast and Full Simulations. There is an approximate factor of 6 discrepancy between the two packages in their default configuration which remains constant with pileup. This difference can be partly reduced by lowering the minimum p_T cut at which tracks are propagated through the detector in the FastSim. By default this is set at 200 MeV/c corresponding to a charged particle reaching a radius of 33 cm before looping back. Performance studies for a layer at 25 cm require that these tracks are included.

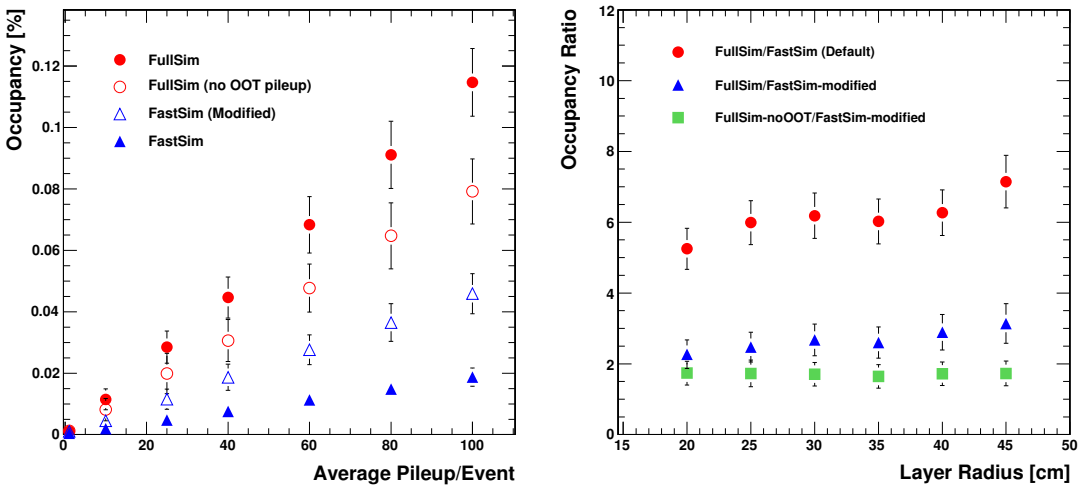


Figure 3.20: Left: Occupancy dependence on the average pileup per event for a stacked pixel layer at 25 cm and 200 μm thickness, Right: Variation of the FullSim to FastSim occupancy ratio as a function of stacked pixel layer radius where sensor parameters and layer coverage remain the same at each radius. The figures denoted by FastSim-default are provided by the default configuration of the Fast Simulation package. FastSim-modified indicates the results for the Fast Simulation where the minimum track p_T cut has been reduced to 50 MeV/c and looping tracks are fully propagated. FullSim-noOOT refers to occupancies generated by the Full Simulation where out-of-time pileup is not included in the event.

Additionally, looping tracks such as these are only propagated by the FastSim up to their maximal radial extent. By forcing the simulation to correctly treat looping low momentum tracks, which dominate the tracker occupancy, the discrepancy factor at intermediate radii can be reduced to ∼2.5 with a 20% uncertainty. This factor appears to be slightly higher for the outermost radii. Running the FastSim with these modifications approximately doubles the CPU time at any pileup. The remaining difference between the simulation packages is shown to be partly due to the failure of the FastSim to take out-of-time pileup into account. By running the Full Simulation without adding out-of-time pileup, the layer occupancies agree to within ∼70% and have no dependence on radius. The remaining difference could be due to the failure of the FastSim to simulate δ rays. A detailed discussion of occupancies is provided in Section 4.1.

3.5 Hit Correlation Algorithm

In order to identify high p_T tracks, hits from the upper sensor of a module must be correlated with those from the lower sensor. The algorithm begins by collecting the digitized hits for each module. A binary zero suppressed readout system is emulated by selecting pixel hits with a charge deposition of greater than 4000 electrons[¶]. Pixels are then either considered as hit or not hit as information on the amount of charge deposited is dropped.

Following the charge threshold cut, hits from each sensor in a module are correlated and high p_T tracks are identified by the generation of a stub if the hits satisfy the correlation windows. Initial studies carried out with untilted sensors in a idealised geometry demonstrated that an effective cut on the p_T of a crossing track can be applied by only selecting hits from each sensor that lie within a pixel or two of each other in r-ϕ[62]. The p_T cut was shown to be determined by the sensor separation and the pixel pitch. However if the sensors are tilted, hits from high p_T tracks are also separated by a fixed number of pixels dependent on the tilt angle and the sensor separation as demonstrated by Figure 3.21.

Since the stacked layer is not a perfect cylinder but is instead comprised of flat modules hermetically arranged to cover the region in ϕ, there is an additional separation between hit pixels dependent on the sensor separation, layer radius and r-ϕ impact point on the sensor. This is true for both tilted and untilted sensors although it can be minimised by keeping the row (r-ϕ) correlation window in pixels to,

$$\text{Row Window} \geq \left\lceil \frac{l}{p}\left(\Delta s^{+} + \Delta s^{-} + \Delta x^{+} + \Delta x^{-}\right) + 0.5 \right\rceil \tag{3.2}$$

where,

$$\Delta s^{\pm} = \frac{\sin(\phi)}{\cos(\phi \pm \theta)} \tag{3.3}$$

[¶]This is a conservative estimate on the charge threshold since the current pixel threshold is set at ∼2800 electrons and could possibly be lowered in the future. On the other hand, the amount of data to be read out or correlated would be reduced since the higher threshold would also reduce the cluster width slightly. Of course this also means that the position resolution of reconstructed hits would be worse, if the data from a stacked layer were to be used for track reconstruction after the Level 1 trigger.

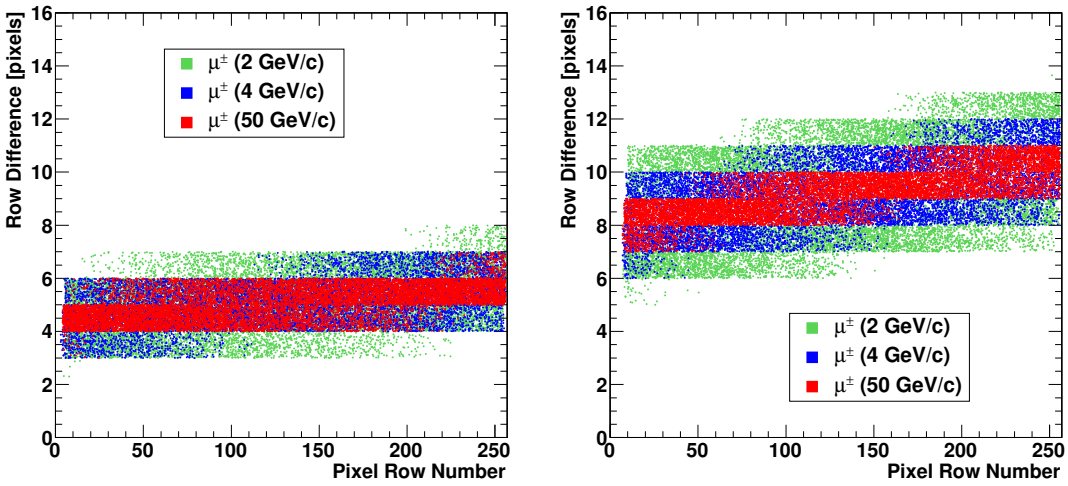


Figure 3.21: Pixel row difference (row difference between hit pixels on the upper and lower sensors) as a function of the hit pixel row on the upper sensor for 50, 4 and 2 GeV/c $\mu^{\pm}$ for a stacked layer at 25 cm in a tilted configuration using a 1 mm separation (left) and 2 mm separation (right). In order to correlate hits efficiently, an offset as a function the active pixel row must be applied. This offset is also dependent on the sensor separation, sensor width, sensor tilt and layer radius.

is the contribution to the separation due to the track angle of incidence (minimum track p_T), and

$$\Delta x^{\pm} = \frac{\sin(\alpha)}{\cos(\alpha \pm \theta)} \tag{3.4}$$

is the maximal contribution to the separation due to using flat modules of width w. The individual signed contributions are from tracks that bend with the tilt angle (+) and against the tilt angle (-). The angles ϕ and α are given by,

$$\phi = \frac{0.6r}{p_{Tcut}}, \quad \alpha = \frac{w}{2r} \tag{3.5}$$

where r is the radius of the stacked layer, p is the pixel pitch in r-ϕ, l is the sensor separation and w is the width of the sensor in metres while p_{Tcut} is the minimum desired transverse momentum a track can possess if it is to always pass the correlation cut, in GeV/c. The expression on the right hand side of equation 3.2 denotes a rounding up to the nearest integer.

The angle θ defines the sensor tilt. In a tilted configuration where θ=23° and α,ϕ are much smaller in comparison, the spread in the r-ϕ separation increases by ≈10%. In addition, a track charge dependence is introduced unless the correlation window is

kept asymmetric in r-ϕ. The overall effect is that tilting the sensors requires that the row correlation window is increased otherwise the p_T at which tracks are selected is raised and a possible track charge bias could be generated in the trigger.

By keeping the sensor width small, the hit separation as a function of the r-ϕ sensor impact point can be reduced. However, the cost of manufacturing and complexity of assembling a large number of narrow modules at intermediate radius within the tracker may be too great. Alternatively, the row correlation window can be increased although if the offset due to the sensor width is too large, this will affect the ability of the stacked layer in rejecting low p_T tracks.

The solution then, would be to apply a row correlation cut as a function of the row pixel number hit. In this way, the window is optimised at any point in the sensor and would allow the design of wider modules for a stacked layer if required. In addition, the ability to assign an offset and window to each pixel row means that the correlation algorithm can be calibrated on detector minimising any inefficiencies due to the misalignment of sensors.

```
// Get offset and window as a function of the upper hit pixel row number
rowOffset = row_offset[upperSensor_row];
rowWindow = row_window[upperSensor_row];

rowDifference = upperSensor_row - rowOffset - lowerSensor_row;

if(hit_globalZposition > 0) {
  columnDifference = upperSensor_column - lowerSensor_column;
}
else {
  columnDifference = lowerSensor_column - upperSensor_column;
}

if((rowDifference >= 0) && (rowDifference < rowWindow)) {
  if((columnDifference >=0) && (columnDifference < columnWindow)) {

    ...// Generate Stub

  }
}
```

Figure 3.22: Pseudo C code describing the basic correlation algorithm. upperSensor_row/column are the pixel row/column numbers for a hit in the upper sensor, lowerSensor_row/column are the pixel row/column numbers for a hit in the lower sensor. The algorithm also takes advantage of the fact that the column window can be applied asymmetrically depending on the detector side in z.

The algorithm also applies a column window cut, based on the difference between

hit pixel columns, to reduce fake correlations if the occupancy is high. By applying the correct z offset as described in Section 3.3.1, the column window cut can be kept to within a small range over the full layer as shown by Figure 3.23. This is important since keeping hit correlations to within a few pixels of each other will reduce the requirement for high speed, high power links to transfer hit data between modules. The algorithm itself is described by the simple piece of pseudo C code provided in Figure 3.22.

The optimal row and column windows as well as row offsets are calculated by simulating with a sample of high p_T tracks. It is also necessary that track vertices are correctly smeared over the luminous region so as to emulate the range of column differences in an entire stacked layer. Plots such as those demonstrated in Figures 3.21 and 3.23 are used to determine the set of calibration constants which must be recalculated for every change in sensor parameter.

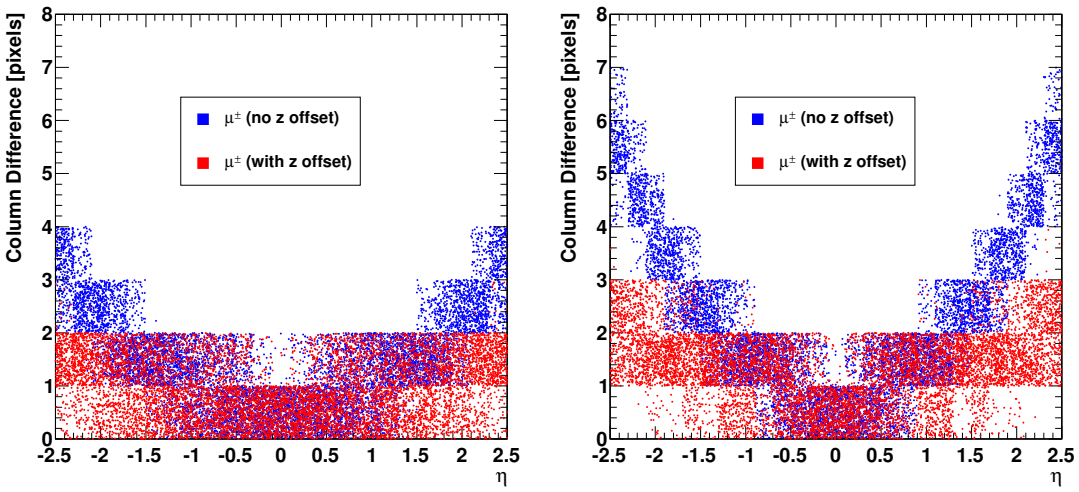


Figure 3.23: Pixel column difference (column difference between hit pixels on the upper and lower sensors) as a function of η for 50 GeV/c $\mu^{\pm}$ for a stacked layer at 25 cm in a tilted configuration using a 1 mm separation (left) and 2 mm separation (right). The blue distribution is for a stacked layer geometry which does not implement the z offset described in 3.3.1. Efficient correlation would require the matching of columns as a function of module position and at high η hit data must be transferred between modules. The red distribution is for a stacked layer geometry which implements the z offset so that the correlation window can be kept fixed to within 2-3 pixels depending on the sensor separation.

3.6 Summary

In order to maintain the L1 trigger rate to 100 kHz without sacrificing physics performance, it is clear that information from the tracker must be included in the trigger decision at SLHC. However, the amount of tracking information generated in every bunch crossing at $10^{35}\text{cm}^{-2}\text{s}^{-1}$ poses a significant problem in terms of readout bandwidth. A method for reducing the on-detector data rate by rejection of low transverse momentum tracks using closely separated pixellated sensors and a simple correlation algorithm has been proposed. Using the CMS software framework, a GEANT description of one of these stacked pixel layers has been implemented for use in simulations.

The trigger layer simulation is designed to be as realistic as possible in terms of material and geometry. Since a strawman module layout is yet to be proposed, properties of the stacked layer such as sensor thickness and separation are easily configurable within the simulation. With current available pixel technologies, a trigger layer would be constrained to a region of radius $r \lesssim 40$ cm in the tracker. In addition, in order to minimise impact on tracking performance, the layer must be placed away from the interaction point. A concept geometry based on an inner pixel detector, stacked pixel layers, and an outer strip tracker has been implemented to satisfy these constraints while providing the simulation with a realistic estimate of the tracker material for tracking and triggering studies.

The simulation itself can use either the Full GEANT description or a parametrised Fast version of the geometry and interactions with matter, the latter of which is able to process high pileup events on much shorter timescales. It has been demonstrated that the Fast simulation underestimates the occupancy compared to the Full simulation by a factor of $\sim$2.5. A detailed study of the performance of stacked pixel layers under SLHC conditions using the Fast simulation is presented in the next chapter.

Chapter 4

Stacked Tracking Trigger Performance at SLHC

Some of the issues and challenges facing the Level 1 trigger for the luminosity upgrade at Super-LHC were highlighted in Chapter 3. In order to maintain the L1 trigger rate at below 100 kHz, it will be necessary to include information from the tracker which is also due to be upgraded. The concept of a stacked pixel trigger layer was introduced as a solution to selectively read out hit information from high transverse momentum tracks thereby reducing the data output from the tracker. Using the simulation tools and geometries described, the following chapter presents the performance of such a layer and its viability as a method for reducing the L1 trigger rate.

4.1 Estimation of Occupancies at SLHC

It is important to have a measurement of the expected occupancies at SLHC and their relative uncertainties in order to estimate the performance of a future tracker design including trigger layers. The predicted occupancy will also dictate the overall system architecture since data transfer rates at the level of a module, ladder and entire tracker will be affected. This in turn must be set against the bandwidth

capability and number of links within and from the detector and their power requirements. A design which is robust against fluctuations is desirable. However overengineering the system to cope with improbable occupancies will cost power and add to its complexity.

Current figures place the expected total inelastic p-p cross section at $\sqrt{s}$=14 TeV at σ_{inel}=79 mb[9]*. QCD events contributing to σ_{inel} are defined here as minimum bias events and include both single and double diffractive interactions since these are also expected to increase tracker occupancies[88].

Until track multiplicity distributions as a function of p_T and η in minimum bias events are measured at the LHC, there is some uncertainty in the expected occupancy at SLHC. Measurements from CDF showed that charged track multiplicities were in disagreement with PYTHIA[86] simulations by ∼30%[89][90]. Low momentum tracks with p_T<0.6 GeV/c can also loop inside the tracker and hence make a greater contribution to the occupancy. These studies show that the variation in these low p_T track multiplicities could be as much as ∼12%.

At present, there are two alternative scenarios for beam operation to achieve the factor of 10 increase in luminosity at SLHC. These are detailed in Section 1.2 and their parameters are listed in Table 4.1.

Either scenario is a equal possibility, therefore it is apparent that the number of collisions per bunch crossing and hence occupancies could vary significantly. It should be noted that the figures quoted so far have been for peak luminosity, and that the instantaneous luminosity decreases over the course of a run. At SLHC, run times are expected to be much shorter so the initial pileup must be increased in order to achieve an average luminosity an order of magnitude greater than at the LHC. One option under consideration for the SLHC upgrade allows the amount of pileup at the beginning of a run to be reduced while the pileup at the end of a run is increased. The possibility for "Luminosity Leveling" would lower the number of

*Estimations are based on results from PYTHIA tuned to fit experimental data from lower energy colliders. While an uncertainty in σ_{inel} exists, the PYTHIA simulations also estimate the total p-p cross section at σ_{tot}=101.5 mb, appearing to agree with experimental data[17].

		LHC	SLHC (ES)	SLHC (LPA)
Peak Luminosity	L $[10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}]$	1.0	15.5	10.6
Bunch Spacing	Δt_{sep} [ns]	25	25	50
Number of Bunches	n_b	2808	2808	1404
Protons/Bunch	N_b $[10^{11}]$	1.15	1.7	4.9
Beta Function	β^* [m]	0.55	0.08	0.25
RMS Luminous Region	σ_{lum} [mm]	45	53	37
Interactions/Bunch Crossing		25	389	531

Table 4.1: Beam parameters[13] for nominal LHC operation and the two upgrade scenarios. The Early-Separation (ES) scheme requires strongly focused bunches at 25 ns spacing. The Large Piwinski Angle (LPA) scenario envisions longer intense flat bunches at 50 ns spacing. Calculation of Interactions/Bunch Crossing uses the p-p inelastic cross section of σ_{inel}=79 mb and factors in the LHC bunch structure which specifies that only 2808/3564 (1404/1782) bunches are filled.

interactions per bunch crossing so that no events within a run need to be discarded while physics studies can operate in a consistent pileup environment.

The following studies run CMSSW_2_2_6 to simulate events in order to estimate the performance of the stacked layer geometry as defined in Section 3.3.1. Using the 25 ns beam operation (ES) scenario as a baseline, SLHC pileup conditions are defined as an average of 400 minimum bias interactions per bunch crossing unless otherwise specified. The simulation includes Poissonian fluctuations in the number of interactions per event under these conditions. All collision vertices are smeared along the z direction to approximate a Gaussian distribution centred at z=0 with σ=53 mm. As discussed in Section 3.4.3, the Full (GEANT) Simulation is not able to handle such a large number of interactions per crossing hence the FastSim is used to model the detector and simulate the PYTHIA generated events. Consequently, out-of-time pileup is not included. The default FastSim configuration has been modified so that the minimum track p_T simulated is lowered to 50 MeV/c and looping tracks are correctly propagated. Digitisation is performed with the standard CMSSW module and a binary readout system in the stacked pixel layers is emulated by selecting hits with a charge deposition of >4000 electrons. Electronic noise in the pixel system is included.

Figure 4.1 demonstrates the expected layer and module occupancies for stacked pixel layers under SLHC pileup conditions using the simulation environment described

above. The average occupancies are defined as

$$\text{Occupancy}[\%] = \frac{N_{\text{Hits}}}{N_{\text{Pixels}}} \times 100 \tag{4.1}$$

where N_{Hits} is the number of hit pixels over the layer/module and N_{Pixels} is the total number of pixels in the layer/module. Both N_{Hits} and N_{Pixels} cover each sensor in a stack. The occupancy for a stacked pixel layer at 25 cm in the untilted configuration and using 200 μm thick sensors is 0.19%±0.01%. Section 3.4.4 showed that the FastSim underestimates the occupancy by a factor of 2.5 with an uncertainty of 20%. Given the combined uncertainty in the expected track multiplicity distributions is 32%, the average expected layer occupancy at 400 interactions per bunch crossing is 0.48%±0.17%. If the SLHC is to operate under the 50 ns (LPA) scenario, the average occupancy could be as high as 0.63%±0.23%.

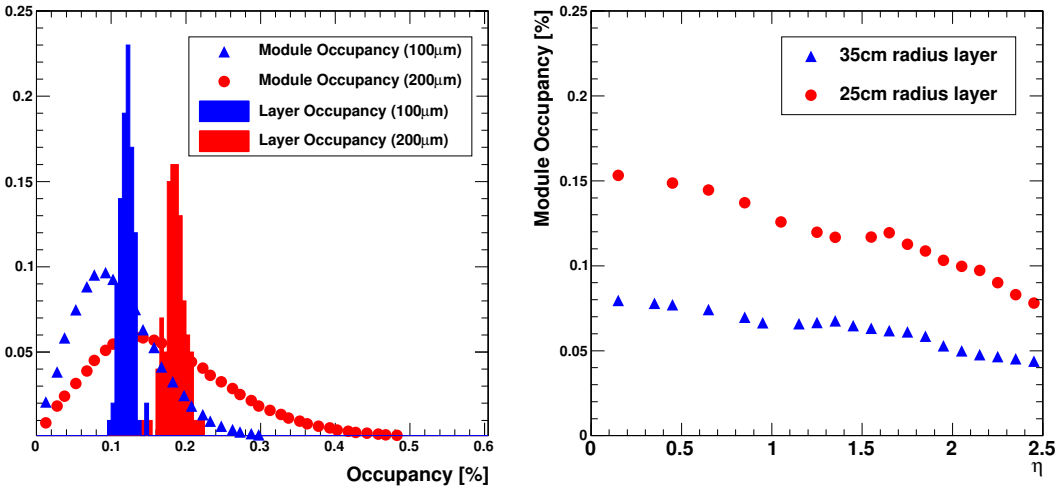


Figure 4.1: Left: Layer and module occupancy distributions for a layer at 25 cm. Right: Average module occupancy as a function of module position in η for layers at 25 cm and 35 cm. Results are for SLHC pileup conditions (minimum bias events with an average pileup of 400 interactions per bunch crossing) and untilted sensors.

Local fluctuations in track density can lead to large variations in occupancy as demonstrated for modules in Figure 4.1. High energy, high multiplicity jets can also contribute significantly increased local occupancies. Simulations need to show that the performance of a stacked pixel layer is insensitive to such effects. In addition, the average module occupancy is shown to vary with η with the highest occupancies to be found in the central region.

Figure 4.2 indicates the influence of sensor tilt and sensor thickness on the average layer occupancy. Tilting the sensors to reduce the effect of Lorentz drift clearly reduces the occupancy by minimizing the cluster width as described in Section 3.3.1. This is also demonstrated by Figure 4.3.

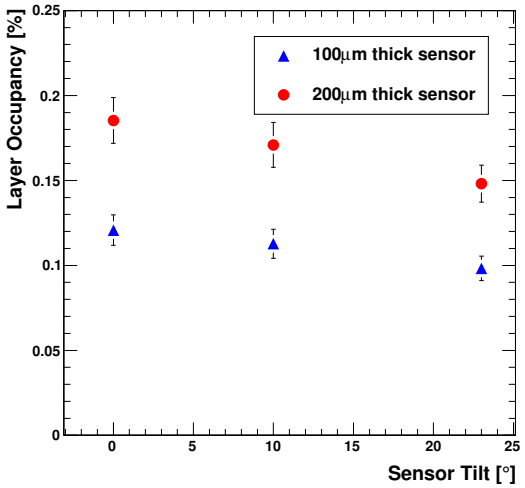


Figure 4.2: Average layer occupancy as a function of sensor tilt and sensor thickness for a layer at 25 cm. Results are for minimum bias events with an average pileup of 400 interactions per bunch crossing.

It is also observed that using thinner sensors lowers the average occupancy by reducing the drift distance and hence cluster width in r-ϕ. However, according to Figures 3.11 and 3.12, reducing the thickness of a tilted sensor should not have a large impact on cluster width. Figure 4.3 provides an explanation as to why thinner sensors which are tilted generate a significantly lower occupancy than a thicker sensor in the same configuration. For tracks with high transverse momentum (p_T>2 GeV/c), thinner tilted sensors have little effect on reducing the cluster width, as is expected. However, there is a large fraction of low momentum tracks in minimum bias events (see Figure 3.4) which traverse the layer at shallow angles. As a result, reducing sensor thickness has a large effect on layer occupancy by minimising the cluster width from low p_T tracks. A stacked layer at 25 cm using 100 μm sensors tilted at 23° reduces the average layer occupancy by ∼50%.

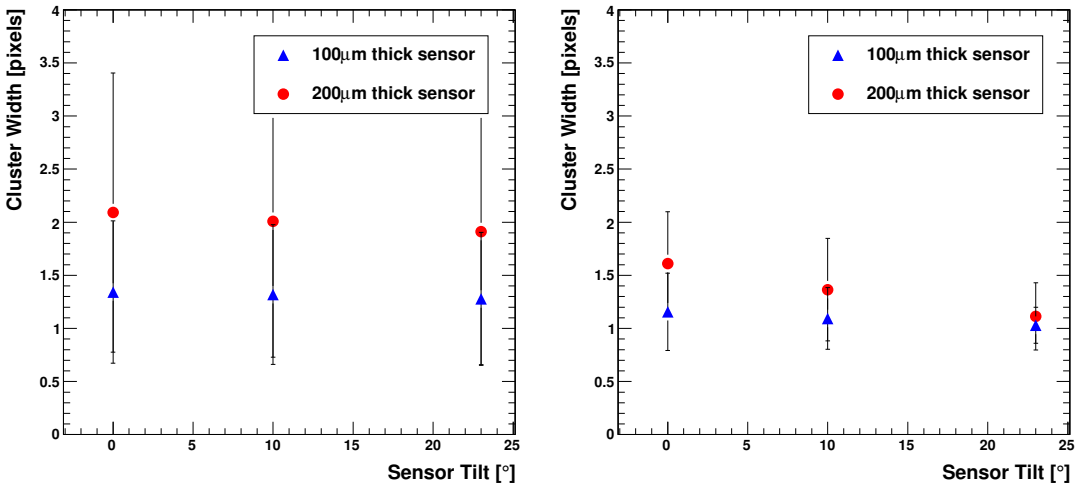


Figure 4.3: Cluster width r-ϕ in as a function of sensor tilt and sensor thickness for a layer at 25 cm; Left: for minimum bias events, Right: for minimum bias events but where tracks have transverse momentum p_T>2 GeV/c.

4.2 Performance of a Stacked Layer at SLHC

4.2.1 Effect of Sensor Separation

For a fixed row correlation window, increasing the sensor separation has the effect of increasing the p_T cut at which stubs are generated. For each sensor separation, a new column correlation window must be applied in order to maintain efficiency. Figure 4.4 demonstrates how a stacked layer at 25 cm is expected to perform at discriminating against the transverse momentum of tracks for various sensor separations and a fixed row correlation cut. The result of the simulation using single muon tracks appears to validate those from previous studies (Figure 3.7).

The efficiency ε described in Figure 4.4 and throughout Section 4.2 is defined as follows,

$$\varepsilon(p_T) = \frac{S_{\text{Stubs}}(p_T)}{N_{\text{Tracks}}(p_T)} \tag{4.2}$$

where N_{Tracks} is the total number of tracks with Monte Carlo transverse momentum (p_T) which generate at least one pixel hit in the stacked pixel layer. S_{Stubs} is the number of tracks with Monte Carlo transverse momentum (p_T) which generate at least one stub in the stacked pixel layer.

The right-hand side of Figure 4.4 indicates the effect of sensor separation on the total average number of stubs per event generated by the trigger layer when simulated under SLHC pileup conditions. A small separation will reduce the effective p_T cut and therefore increase the number of generated stubs while a larger separation has the opposite effect. A large separation should be beneficial for a trigger since only high p_T tracks are likely to be passed. However, if effective isolation of tau jet candidates is required, tracks with a transverse momentum of at least 2 GeV/c must be recorded by the layer[79]. The efficiency for triggering of tracks with p_T>2 GeV/c is provided in Table 4.2.

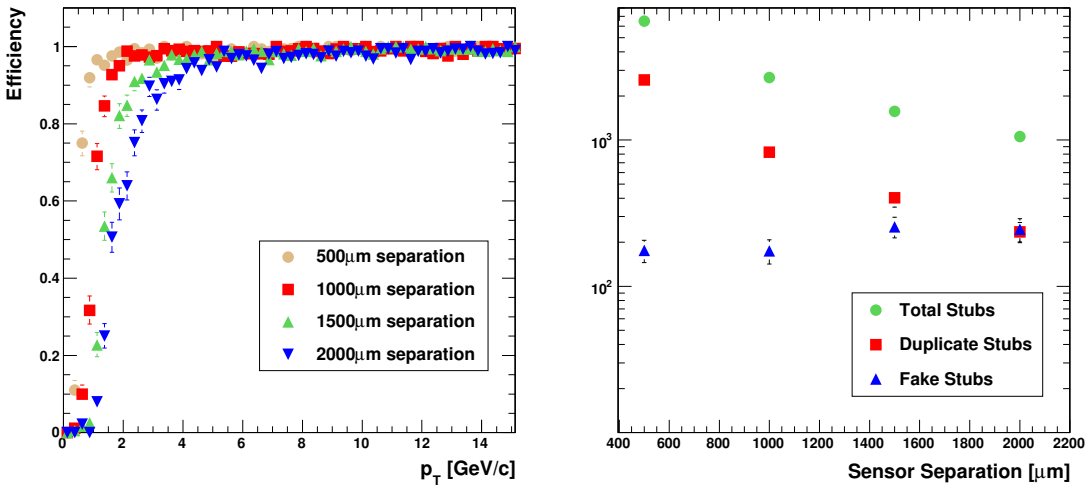


Figure 4.4: Left: p_T discrimination performance of a stacked layer for single $\mu^{\pm}$ tracks at various sensor separations and a fixed 3 pixel row correlation window, Right: Average number of generated stubs per event under SLHC pileup conditions as a function of sensor separation. Results are for a stacked layer at 25 cm in the untilted configuration. Column window is 2 pixels for sensor separations of 1000 μm and less while a 3 pixel window is used for larger separations.

The transition region is the region over which a track of a given p_T may or may not be passed and is dependent on charge sharing between pixels, the pixel pitch, the sensor thickness, the sensor separation and the track impact point on both sensors. Ideally this region width should be as small as possible in order to have a well defined efficiency above the cut. This however requires a small sensor separation which means that more tracks will be passed due to the lower p_T cut.

A chance exists that, in a high occupancy environment, hits from two tracks which would not pass the p_T cut on their own are correlated to generate fake stubs as

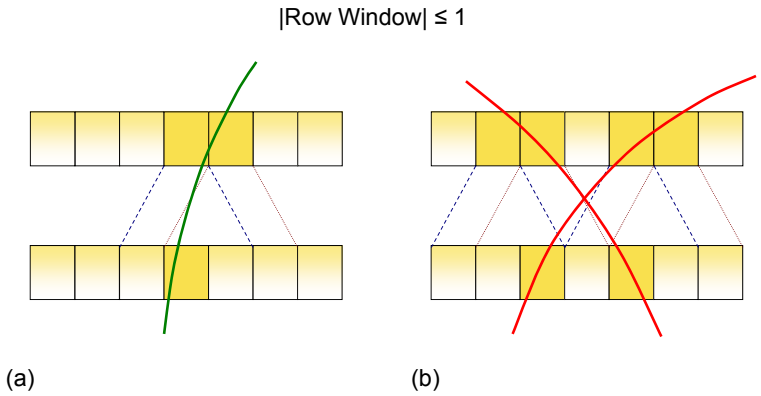


Figure 4.5: Illustration on the origin of duplicate and fake stubs; (a) demonstrates that if the row correlation window is $\geq \pm 1$, clusters of hit pixels can give rise to multiple stubs. An additional or duplicate stub is generated in this example. Tracks which would not normally pass the correlation cut may still produce a stub if hits are incorrectly matched with those from another track. Case (b) demonstrates how four of these fake stubs are generated.

depicted in Figure 4.5. Figure 4.4 shows that the average number of fakes increases slightly with sensor separation, mainly because the column window must be enlarged for greater sensor separations. Also indicated, is the average number of duplicate stubs generated per event as a function of sensor separation. Duplicates, or the number of additional stubs per high p_T candidate, are inherent due to the simplicity of the correlation algorithm when more than one pixel per incident track is hit. The number of duplicates could be reduced to zero using a clustering algorithm either before or after correlation[†]. If clustering could be performed before the module is read out, removing the duplicates would decrease the trigger layer data rate by around 20-30%.

Table 4.2 also provides the average ratio of duplicate and fake stubs to total stubs alongside the reduction factor which is defined as

$$\text{Rate Reduction} = \frac{N_{\text{Hits}}}{N_{\text{Stubs}}} \tag{4.3}$$

[†]Clustering pixel hits by row before correlation takes place has the added benefit of increasing the hit position resolution by calculation of the centroid (charge interpolation would not be possible in a binary readout system). The transition region could then be minimised and hence fewer stubs would be generated per event. Clustering stubs after correlation has taken place would only remove duplicates.

where N_{Hits} is the average number of hit pixels and N_{Stubs} is the average number of stubs generated in a single stacked layer per event, under SLHC pileup conditions. It is an indication of the reduction in the number of hits to be read out if correlation was to be performed on detector. The efficiency ε is defined as

$$\varepsilon(\mathrm{p_T} > 2\ \mathrm{GeV/c}) = \frac{S_{\text{Stubs}}(\mathrm{p_T} > 2\ \mathrm{GeV/c})}{N_{\text{Tracks}}(\mathrm{p_T} > 2\ \mathrm{GeV/c})} \tag{4.4}$$

where S_{Stubs} and N_{Tracks} have the same definitions as in equation (4.2) for tracks with transverse momenta $\mathrm{p_T}$>2 GeV/c.

Sensor Separation (μm)	Row Window (pixels)	$\varepsilon_{\text{Muon}}$ $\mathrm{p_T}$>2 GeV/c (%)	$\mathrm{N_{Stubs}}$	Fake (%)	Duplicate (%)	Rate Reduction
500	3	99.4	6497.1	2.7	39.6	9.2
1000	3	99.2	2670.5	6.6	30.9	22.0
1500	3	98.6	1570.5	16.3	25.6	37.0
2000	3	97.1	1054.1	23.3	22.4	54.4

Table 4.2: Trigger performance of a stacked layer at 25 cm. $\varepsilon_{\text{Muon}}$ is the efficiency for triggering on $\mu^{\pm}$ tracks with $\mathrm{p_T}$>2 GeV/c. The percentage of fake and duplicate stubs and the rate reduction factors are calculated from simulating the stacked layer in minimum bias events at SLHC pileup.

A layer at 25 cm with a 1 mm sensor separation is able to trigger on muon tracks with $\mathrm{p_T}$ greater than 2 GeV/c with an efficiency of 99%, reducing the data volume by a factor of 22.

4.2.2 Effect of Correlation Window Cuts

The row correlation window cut is another method of controlling the transverse momentum at which tracks are discriminated against. This is evident from Figure 3.21. However, the difference is that while varying the sensor separation modifies the $\mathrm{p_T}$ cut continuously, changing the correlation window will modify the $\mathrm{p_T}$ cut in discrete steps. Figure 4.6 demonstrates how increasing the row window with sensor separation means that $\mathrm{p_T}$ discrimination performance is maintained, even at large separations. Table 4.3 defines the row window cuts used to achieve this.

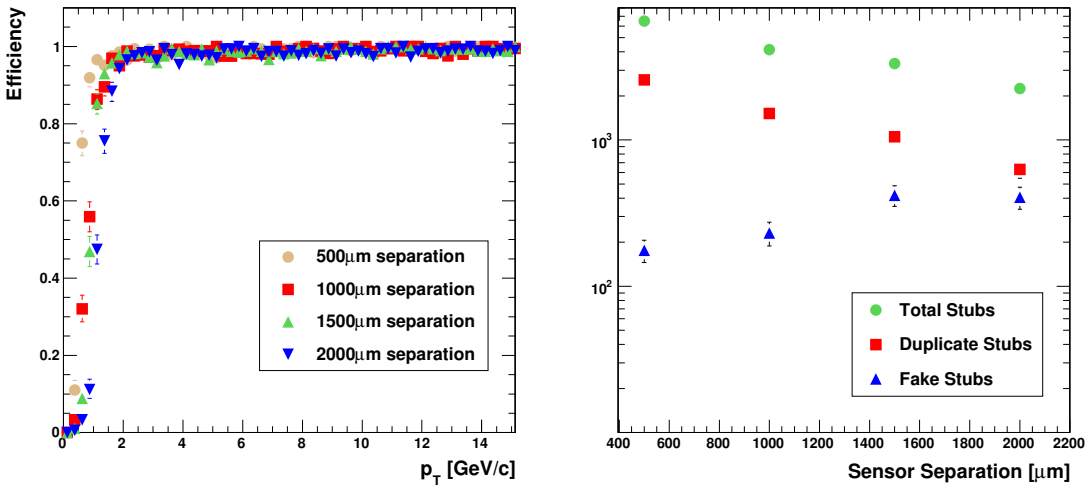


Figure 4.6: Left: p_T discrimination performance of a stacked layer for single $\mu^{\pm}$ tracks and, Right: Average number of generated stubs per event under SLHC pileup conditions as a function of sensor separation where the correlation window is widened with sensor separation. Results are for a stacked layer at 25 cm in the untilted configuration. Compared to Figure 4.4, the p_T discrimination curve does not change with sensor separation, however the total number of stubs including duplicates and fakes does increase with sensor separation.

Hit matching over the minimal number of pixels naturally leads to a lower power and simpler correlation implementation. However, if a low p_T cut is required and the sensor separation is fixed by construction constraints, it may be necessary to increase the correlation window. The ability to change the window cut on the module during operation will also provide a useful and relatively quick method to adjust the transverse momentum of tracks passing the trigger and hence the trigger rate. The right-hand side of Figure 4.6 shows that if the row window is widened with sensor separation, the number of stubs generated per event under SLHC conditions increases (compare with Figure 4.4). This is expected if the effective p_T cut is lowered. It is also observed that there is a clear increase in the number of fake and duplicate stubs due to the larger correlation windows. Table 4.3 summarises these results.

It is interesting to note that a stacked layer with a 2 mm sensor separation and row correlation window of 5 can attain a similar performance as a layer with 1 mm sensor separation and row correlation window of 3, albeit with more fakes. Building a layer with a larger sensor separation but with a correlation window that can be varied

Sensor Separation (μm)	Row Window (pixels)	$\varepsilon_{\text{Muon}}$ p_T>2 GeV/c (%)	N_{Stubs}	Fake (%)	Duplicate (%)	Rate Reduction
500	3	99.4	6497.1	2.7	39.6	9.2
1000	4	99.2	4150.9	5.6	36.6	14.2
1500	5	99.1	3326.0	12.6	31.7	17.5
2000	5	98.7	2248.3	18.1	28.0	25.5

Table 4.3: Trigger performance of a stacked layer at 25 cm where the row correlation cuts are increased with sensor separation so as to maintain a high efficiency. Compared to Table 4.2 the average number of stubs generated per event is a factor of ~2 higher. There is a smaller fraction of fake and duplicate stubs per event although the total number of each is larger, as confirmed by Figure 4.6.

may be more practical in terms of robustness to triggering demands and the physics at SLHC.

Figure 4.7 describes the performance for the stacked layer when selecting pion and electron tracks by p_T. Although the difference appears minimal, compared to muons, the layer is less effective at rejecting low transverse momentum electrons and especially pions. Table 4.4 also indicates that the stack is slightly less efficient at selecting high momentum electrons and pions with p_T>2 GeV/c.

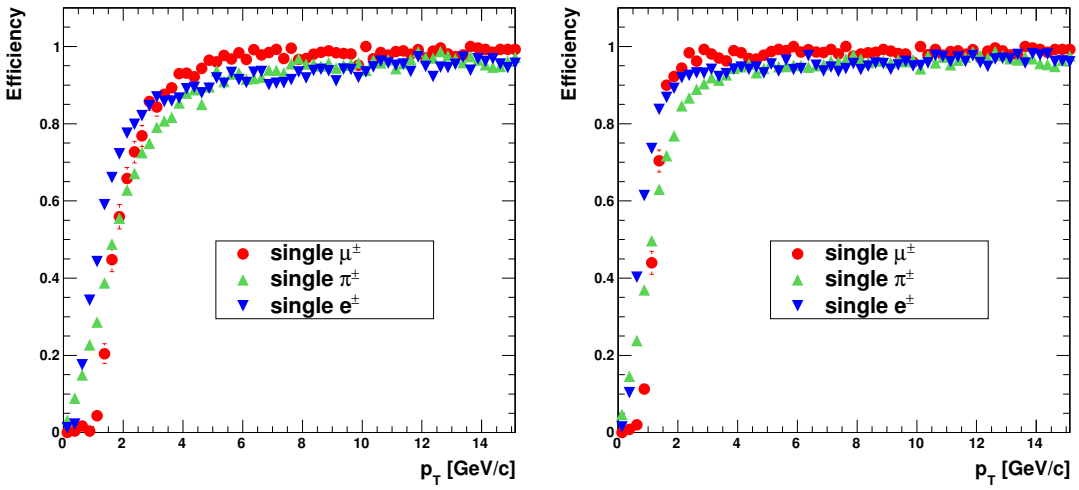


Figure 4.7: p_T discrimination performance for single $\mu^{\pm}$, $\pi^{\pm}$ and $e^{\pm}$ tracks using a stacked layer at 25 cm in the untilted configuration with a 2 mm sensor separation; Left: for a correlation window of 3 pixels, Right: for a correlation window of 5 pixels.

Due to their larger interaction cross-section pions are more likely to interact in the

pixel system, which for the current detector the probability is estimated to be >2-10%[11], before they reach the stacked pixel layer. This means that a low p_T charged hadron resulting from a secondary interaction somewhere within the inner detector could be triggered as a high p_T track, as depicted in Figure 4.8. Since the principle of the stacked tracking layer correlation is based on the assumed track vertex at the beam spot, any low transverse momentum particle originating from a radius r>0 can be incorrectly reconstructed and selected as a high p_T track. Conversely, there is a small chance that a high p_T charged hadron may result from a secondary interaction and fail the correlation if its angular deflection relative to the primary pion is large.

Row Window (pixels)	ε_{Muon} p_T>2 GeV/c (%)	ε_{Pion} p_T>2 GeV/c (%)	$\varepsilon_{Electron}$ p_T>2 GeV/c (%)
3	97.1	94.7	96.1
5	98.7	96.3	97.4

Table 4.4: Trigger efficiencies for single muons, pions and electrons with p_T>2 GeV/c using a stacked layer at 25 cm with 2 mm sensor separation in the untilted configuration.

Similarly, an electron can interact before reaching the stack by radiation of photons so that its transverse momentum is reduced at the layer and hence fail the cut[‡]. In addition, the surplus of low p_T electrons passing the trigger is due to pair production from bremsstrahlung photons in the pixel system. These electrons may be passed by the trigger because of their displaced vertex. This is discussed further in Section 4.4.2.

4.2.3 Effect of Sensor Tilt

The effect of Lorentz drift on a stacked layer and its impact on triggering performance is an important case to investigate. Tilting the sensors as described in Section

[‡]Due to a peculiarity in the simulation tools, the primary electron transverse momentum defined here is that of the original generated Monte Carlo electron, i.e. without bremsstrahlung energy losses. The same is also true for the p_T of secondary electrons (due to photon conversions) although their transverse momenta are defined at the point just after conversion.

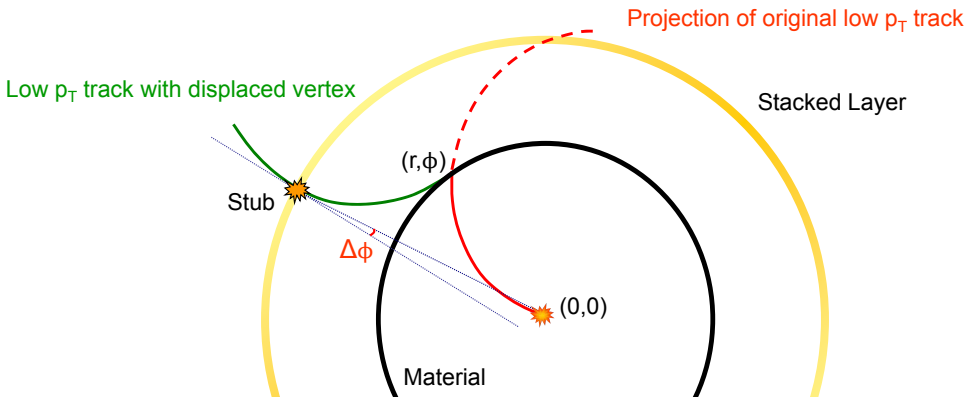


Figure 4.8: Origin of low transverse momentum background in pion events. Interactions with material in the inner detector can give rise to low p_T tracks that pass the correlation due to their displaced vertex. In this case, the reconstructed track crossing angle relative to the normal of the layer ($\Delta\phi$) is small enough to generate a stub even though the p_T of the original track would not have passed the correlation.

3.3.1 eliminates electron drift due to the magnetic field within CMS, therefore reducing cluster sizes (see Figure 3.12). This could be an important benefit when attempting to design low power readout and correlation electronics for the module. In addition it should also help reduce data rates around and from the stacked layer. Figure 4.9 reveals the p_T discrimination performance of a stacked layer at 25 cm in the tilted configuration for both fixed and variable row correlation window cuts as function of sensor separation.

While tilted sensors are still able to trigger on tracks above a certain transverse momentum, the p_T cut is higher and the transition region wider than in the untilted case, especially at larger sensor separations. Equation (3.2) shows that, compared to an untilted layout, a tilted configuration requires a larger correlation window in order to maintain the same p_T cut. Hence, a fixed correlation window will raise the p_T cut as well as introduce a bias between tracks of opposite charge (Section 3.5), increasing the range of momenta where tracks may or may not be triggered on. Nonetheless, the right-hand side of Figure 4.9 demonstrates that the row correlation windows can again be adjusted to reduce the p_T cut and achieve good efficiencies above a p_T of 2 GeV/c. Figure 4.10 provides a direct comparison of p_T discrimination performance for tilted and untilted sensors and identical correlation cuts.

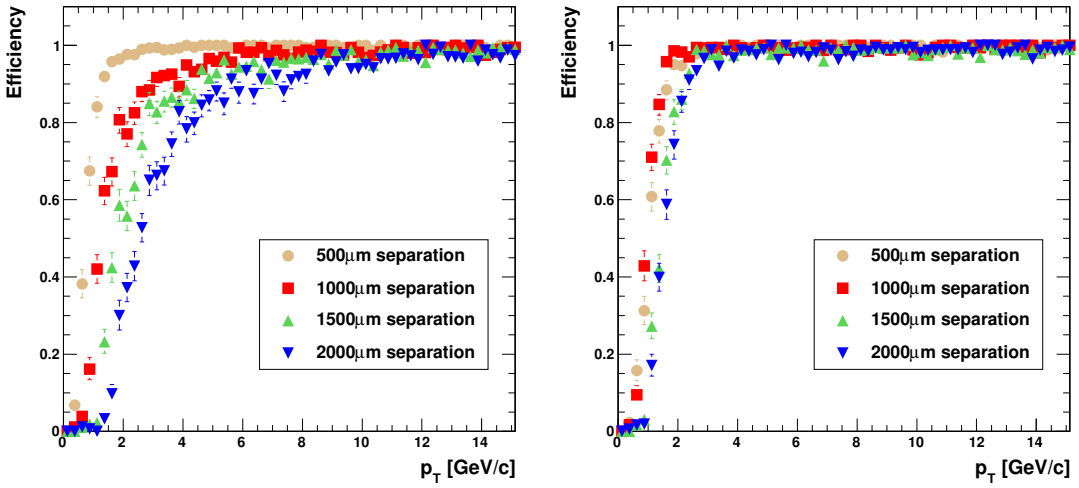


Figure 4.9: p_T discrimination performance for single $\mu^{\pm}$ tracks using a stacked layer at 25 cm in the tilted configuration; Left: for a fixed 3 pixel correlation cut, Right: for variable correlation cuts.

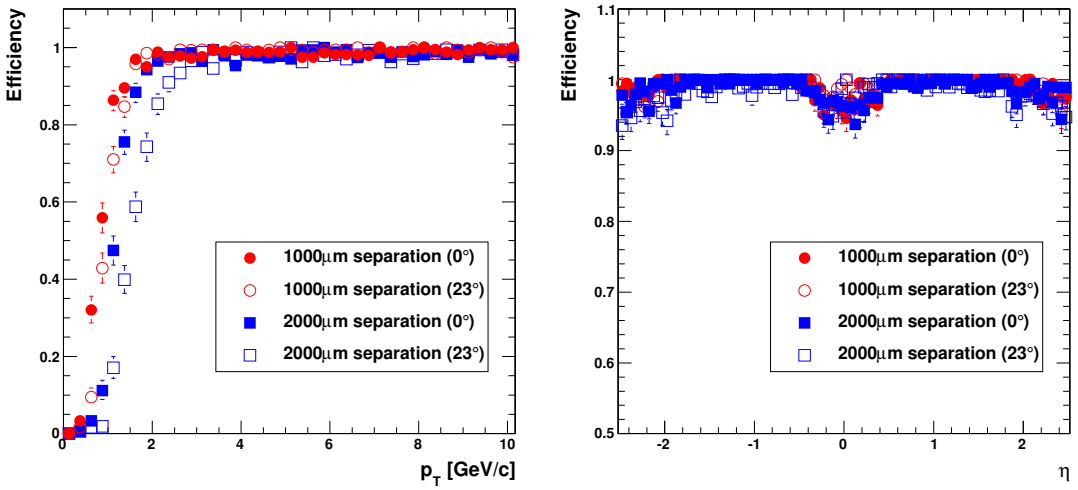


Figure 4.10: Left: pt discrimination performance comparison of untilted and tilted sensors in a stacked layer at 25 cm. Right: Efficiency vs. η comparison of untilted and tilted sensors for tracks with p_T>2 GeV/c using a stacked layer at 25 cm. Results are for single $\mu^{\pm}$ events and identical correlation cuts. The dip at η=0 is due to non overlapping stacked modules in the central region.

The fact that tilted sensors tend to raise the p_T threshold at which tracks are triggered is reflected in the reduced number of stubs generated per event under SLHC conditions in comparison to the untilted geometry. This is illustrated in Figure 4.11 for both fixed and variable row correlation window cuts. It should also be noted that tilting the sensors decreases both the number of duplicates and the number of fakes per event. This is to be expected when cluster widths are reduced. Table 4.5 also indicates the higher rate reduction factors that can be achieved if sensors are tilted. A large sensor separation and small correlation window is the least efficient configuration for triggering of tracks with p_T>2 GeV/c but will however significantly reduce the data rates on detector.

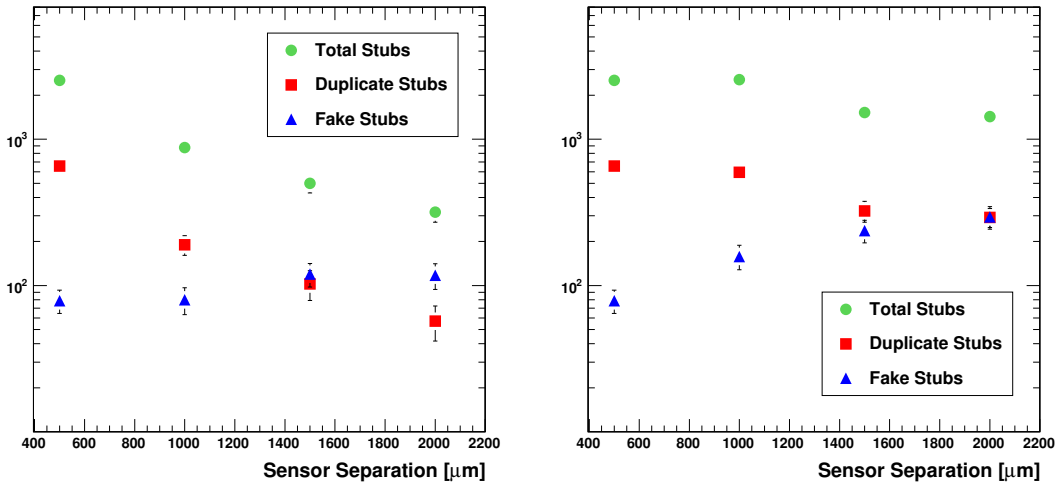


Figure 4.11: Average number of generated stubs per event under SLHC pileup conditions as a function of sensor separation; Left: for fixed correlation cuts, Right: for variable correlation cuts. Results are for a stacked layer at 25 cm in the tilted configuration.

A tilted sensor layout is one under consideration since a reduction in data rate is sure to lower the power requirements of a trigger layer. Eliminating Lorentz drift and minimising cluster width appears to provide a cut at a higher transverse momentum while simultaneously reducing duplicate and fake stubs. For a layer with 1 mm sensor separation and row correlation window of 4, data rates can be a factor of ∼2 lower using a tilted compared to an untilted layout for equivalent pileup conditions.

Alternatively, larger clusters may be advantageous in improving the position resolution of hits in the layer if the full event information could be kept in buffers until a

Sensor Separation (μm)	Row Window (pixels)	$\varepsilon_{\mathrm{Muon}}$ p_T>2 GeV/c (%)	N_{Stubs}	Fake (%)	Duplicate (%)	Rate Reduction
500	3	99.6	2526.6	3.1	26.0	19.0
1000	3	96.6	877.9	9.1	21.7	54.4
1500	3	92.5	499.7	23.9	20.5	95.6
2000	3	89.0	317.7	37.0	18.0	150.2
500	3	99.6	2526.6	3.1	26.0	19.0
1000	4	99.4	2553.4	6.2	23.3	18.7
1500	4	98.7	1524.4	15.6	21.3	31.3
2000	5	98.1	1429.2	20.6	20.5	33.4

Table 4.5: Trigger performance of a stacked layer at 25 cm in the tilted configuration.

Level 1 accept and used during higher level track reconstruction. Since the trigger layer will contribute material to the tracker, if it is unable to provide hit information in this way then it could degrade tracking performance, even though the buffering and reading out of data will add to the power requirements of the layer. As a result, it is possible that an untilted sensor layout will be required to benefit tracking rather than triggering. In addition, the mechanics of constructing a layer with a tilted arrangement of sensors is more complex. One possibility of reducing the data volume from an untilted stacked layer would be to apply a clustering algorithm either before or after correlation.

4.2.4 Effect of Layer Radius

According to equation (3.1), the radius at which the stacked pixel layer is placed also determines the effective p_T cut that can be achieved. While it would be useful to study a number of layers at different radii, due to time constraints, results from only two radii in the most likely radial region are provided here.

Figure 4.12 indicates the apparently minimal change in p_T discrimination performance when moving from a stacked layer at 25 cm to a layer at 35 cm. This can be seen for layers in both the tilted and untilted configuration. Equation (3.1) explains the small shift in the p_T cut. Only a 40% increase is expected at 35 cm with respect to the layer at 25 cm when all other parameters are identical.

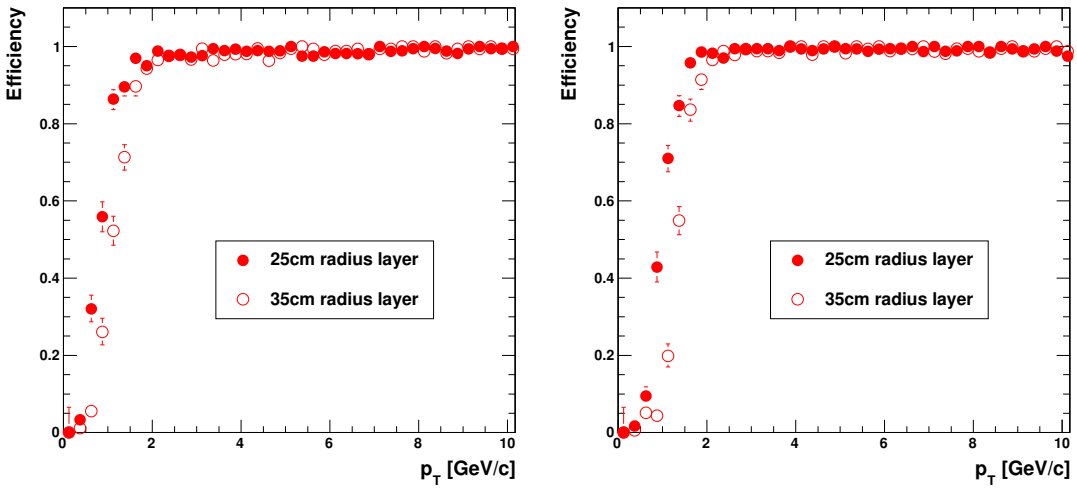


Figure 4.12: p_T discrimination performance for single $\mu^{\pm}$ tracks using stacked layers at 25 cm and 35 cm; Left: in the untilted configuration, Right: when the sensors are tilted.

Although the p_T discrimination performance appears to be similar between the two layers, the effect on the performance of the layer in a typical SLHC event is noticeable. Table 4.6 demonstrates this for layers in both the tilted and untilted configuration. In both cases, the number of stubs generated by the layer at 35 cm is almost halved with respect to the layer at 25 cm. This is specifically due to the slightly higher p_T cut and large proportion of tracks in minimum bias events with low transverse momenta.

Sensor Tilt (°)	Layer Radius (cm)	ε_{Muon} p_T>2 GeV/c (%)	N_{Stubs}	Fake (%)	Duplicate (%)	Rate Reduction
0	25.0	99.2	4150.9	5.6	36.6	14.2
0	35.0	99.2	2578.1	4.4	40.5	22.6
23	25.0	99.4	2553.4	6.2	23.3	18.7
23	35.0	99.2	1514.0	4.8	28.6	30.4

Table 4.6: Trigger performance of a stacked layers at 25 cm and 35 cm in both the tilted and untilted configurations for a 1 mm sensor separation and correlation window cut of 4 pixels.

Simultaneously, the high global trigger efficiencies for muon tracks above 2 GeV/c are maintained with increasing layer radius. The ratio of fake and duplicate stubs to total stubs for layers at 25 cm and 35 cm are also comparable. Consequently, stacked layers at larger radii should be able to operate with lower trigger data rates and with

fewer links therefore reducing the power requirements of sending data off detector. On the other hand, the number of pixels per layer increases with $\sim r^2$ hence the total power consumption may be greater depending on the fraction of power required by the links. Estimates of link power consumption in stacked pixel layers are given in Section 4.3.

4.2.5 Effect of Sensor Thickness

It was demonstrated in Section 4.1, that the occupancy of a layer is dependent on the thickness of the sensor. In addition to reducing the power consumption of a module, thinner sensors also have the benefit of reducing cluster sizes and therefore the detector occupancy and data rate. On the other hand, thinner sensors generate smaller signals so a choice of the sensor thickness will ultimately depend on if a sufficient signal to noise ratio is achievable as well as on the availability of such a thin sensor technology. Figure 4.13 demonstrates that a stacked pixel layer equipped with 300 μm thick sensors has a similar p_T discrimination performance compared to those using thinner sensors.

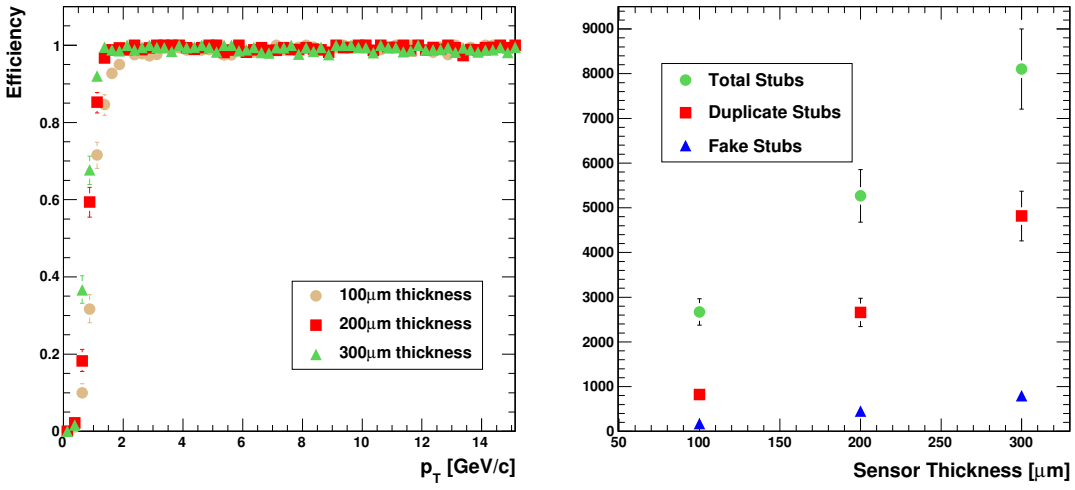


Figure 4.13: Left: p_T discrimination performance of a stacked layer for single $\mu^{\pm}$ tracks and, Right: Average number of generated stubs per event under SLHC pileup conditions as a function of sensor thickness. Results are for a stacked layer at 25 cm in the untilted configuration and a sensor separation of 1 mm.

While the efficiencies are comparable, the right hand plot indicates that the number of stubs increases linearly with sensor thickness. The dominant contribution to this increase is from the duplicate stubs which grow due to the broadening cluster size. As discussed previously, the ability to cluster on detector, either before or after correlation, would reduce the number of stubs generated per event greatly, especially if the module design dictates thicker sensors or an untilted configuration.

4.2.6 Effect of Pileup

It is essential to understand the effects pileup introduces in the performance of the stacked layer. The trigger algorithm needs to be able to operate efficiently at any luminosity while still offering the same reduction in data output and must also be robust against any local or global fluctuations in occupancy. One of the disadvantages of using the Fast simulation is that it underestimates the occupancies in a typical minimum bias event by a factor of $\sim$2.5 compared to the Full simulation. In the most extreme cases, for example in the LPA scenario and using 200 μm thick sensors, it may be possible that the stacked layer will be subject to peak event occupancies of up to 0.63%$\pm$0.23%. Consequently, simulating the extreme occupancies that one might expect in a worst case scenario at SLHC is difficult and requires that the pileup is increased beyond nominal.

Figure 4.14 describes the performance of the algorithm as a function of occupancy. The left hand plot indicates that the trend of number of generated stubs increasing with occupancy is approximately linear and well behaved up to occupancies of 0.9%. The number of duplicate stubs also scales in the same way while there appears to be a small non-linear increase in the fraction of fake stubs at the highest occupancies. This is expected when the hit density becomes large enough for random correlations to be made. The right hand plot from Figure 4.14 demonstrates that for occupancies of up to 0.9%, the performance of the stacked tracker is robust against pileup, with respect to the trigger efficiency (for muons p_T>2 GeV/c) and rate reduction factor. At occupancies approaching 1%, it is observed that the number of stubs generated by fake correlations begins to adversely affect the rate reduction factor.

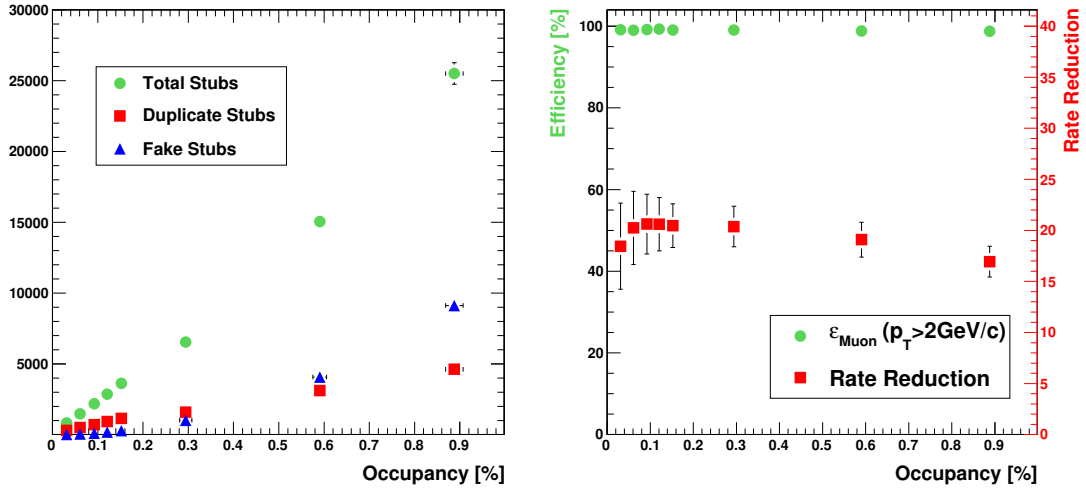


Figure 4.14: Left: Average number of generated stubs per event and, Right: Trigger efficiency for muons with p_T>2 GeV/c (left scale) and rate reduction factor (right scale) under SLHC pileup conditions as a function of average layer occupancy. Results are for a stacked layer at 25 cm in the untilted configuration and a sensor separation of 1 mm.

4.3 Using a Stacked Layer in a Level 1 Trigger

Realistic simulations of stacked pixel layers in high pileup events have demonstrated that the concept of identifying high p_T tracks using correlated hits from closely separated sensors is viable. A pair of pixel sensors with 100 μm pitch and separated by 1 mm can be used to cut on tracks with p_T>2 GeV/c with an efficiency of 99% (for muons) if placed at a radius of 25 cm. Such a layer would be able to reduce the total amount of data by at least a factor of 22.

Assuming an average layer occupancy of 0.5% at 20 MHz[§], a reduction factor of 20 would require the readout of around 12,000 stubs per layer. Using the current SLHC proposals for a customised radiation hard gigabit link architecture (GBT)[91], each optical link is estimated to allow a bandwidth of 3.2 Gb/s to transmit data from the detector to the DAQ or trigger. Assuming a 16bit stub encoding scheme, approximately 1200 links would be needed, requiring over 3.4 kW of power if each

[§]While this is realistic, the occupancy could be factor of 2 greater depending on sensor configuration, beam operation and uncertainties in the inelastic proton-proton track multiplicity distributions.

link consumes ∼2.9 W[92][¶]. While these figures are large, it demonstrates that the rate reduction factors offered by a stacked layer concept are enough to feasibly allow the readout of tracking information for input into the L1 trigger.

Simulations have also demonstrated that there is a certain amount of flexibility in the choice of parameters for the stacked pixel layer. This could prove important in the design of the layer so that constraints due to construction, power and data rates as well as physics can be all satisfied. It has been shown that by choosing a larger sensor separation, rates can be further reduced if the correlation window is maintained. With a larger separation, there is flexibility in choosing the window size so that the effective p_T cut can be modified during operation to suit trigger or data rate levels. On the other hand, choice of sensor orientation to optimise the correlation and reduce the number of duplicate and fake stubs is more likely to be determined by how each design could be implemented mechanically since simulations indicate that there is little difference in performance between the choices.

The study has also proved that the performance of a stacked pixel layer is not adversely affected by pileup and would be robust over a range of occupancies up to at least 0.3%. If the major constraint on the system design is power, this can be used to determine the maximum number of stubs that can be read out per event and hence maximum operating occupancy. If data from initial LHC collisions at $\sqrt{s}$=14 TeV indicate that simulations are currently underestimating occupancies or if the proposed SLHC beam conditions are changed, having the effect of worsening pileup, the stacked layer design must be modified accordingly. As presented, this can be achieved by either increasing the p_T threshold cut by increasing the sensor separation or decreasing the correlation window or by optimising the sensor parameters such as the sensor thickness or layer radius.

While simulations have shown that a stacked pixel layer is effective at identifying tracks with high transverse momentum and reducing the data rate from the tracker, it is important to demonstrate that the information provided by the tracker could be

[¶]This can be compared to the requirements for the current TIB which has a total power dissipation of over 9.4 kW and over 6900 links. It is possible that the link bandwidth could double in the future, reducing the link power per layer even further.

useful for a L1 trigger. The most significant drawback of using a single stacked pixel layer is that while the data output contains hits from high p_T tracks, a measurement of their momenta cannot be provided since the pixel position resolution is too large with respect to the minimal radial separation between sensors. As a result, a future L1 trigger would only have access to the positional information from high p_T stubs for correlation with trigger primitives from other CMS subdetectors.

As simulations indicate, a stacked pixel layer in a typical configuration would generate $\mathcal{O}(1000)$ stubs per event, ruling out the possibility on simply triggering on stub events unless the p_T cut was significantly increased. Instead, a L1 trigger would have to match stubs to calorimeter or muon objects in order to reduce the trigger rate if threshold cuts were to be maintained at their present levels. While it is difficult to set out a definitive trigger architecture at L1, it could be assumed that tracker information can be incorporated into the L1 decision at an early stage; at primitive or regional level for example. This would be the only possibility given the large number of stubs per event and the lack of transverse momentum information.

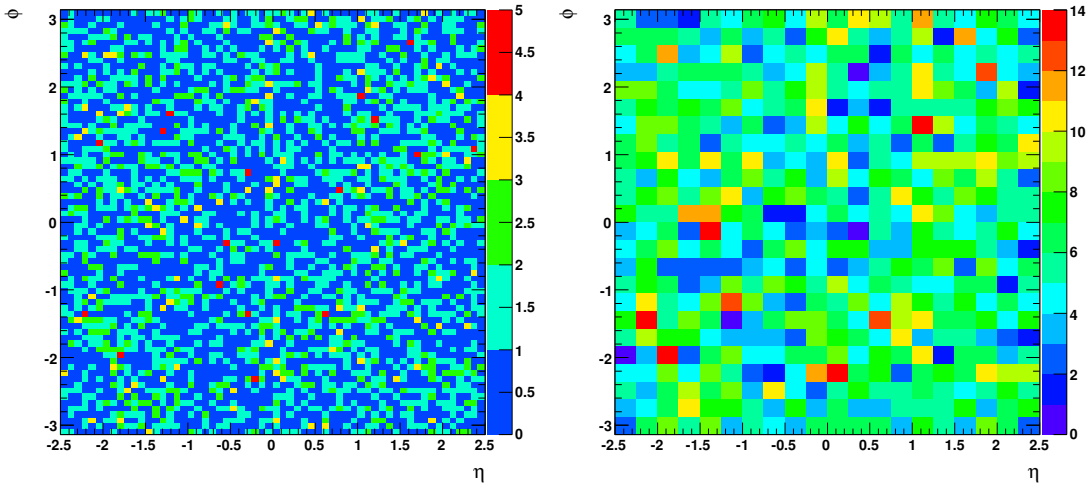


Figure 4.15: Distribution of stubs in; Left: approximate η-ϕ trigger towers ($\Delta\eta\Delta\phi$ = 0.087×0.087) or Right: approximate η-ϕ ECAL trigger regions (3×3 towers).

In the absence of momentum and primary vertex information, stubs will most likely have to be matched to high energy deposits in calorimeter trigger towers using a straight line fit to the origin. Figure 4.15 describes the distribution of trigger stubs in

approximate η-ϕ trigger towers ($\Delta\eta\Delta\phi$=0.087×0.087) or ECAL regions (3×3 towers) for a typical SLHC event. On average, 48% of calorimeter towers are matched to a stub while almost every calorimeter region[‖] matches at least one stub. Since one would naïvely expect a requirement for at least a 5-10% coincidence in order to reduce the trigger rate at high luminosity, this demonstrates that the information from a single stacked layer is unlikely to prove useful in a L1 trigger.

In addition, even if the p_T cut was high enough to pass fewer stubs per event, the lack of momentum information from the layer means that a straight line fit to the calorimeter fails at low p_T, as Figure 4.16 shows. While triggering on these low momentum tracks may not be important, a poor matching performance will inevitably lead to a significant number of fake correlations thereby increasing the trigger rate. Multiple scattering due to tracker material will also decrease the resolution requiring a larger window in ϕ while size of the interaction region in z will mean a window in $\Delta\eta$ of >0.6.

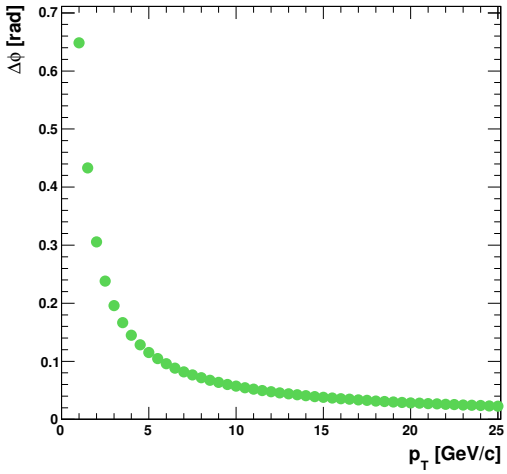


Figure 4.16: ϕ difference ($\Delta\phi$) between the straight line projection of a stub (at radius 25 cm) from the origin to the calorimeter surface and the true impact point of the track at the ECAL as a function of p_T for the barrel region only. For tracks with p_T<6.5 GeV/c, the resolution is larger than that of a trigger tower.

Figure 4.17 plots the p_T spectrum of tracks which generate stubs in SLHC events

[‖]A 3×3 tower region is the current area within the calorimeter used to construct L1 $e\gamma$ candidates of which there will be one candidate per region before selection based on transverse energy. This region size could conceivably be reduced to 2×2 towers for SLHC.

and hints at the origin of the large stub background. It is clear that there is still a sizeable contribution from tracks below the p_T=2 GeV/c threshold. The dominant contribution to this background is from products of secondary interactions within the pixel system as indicated by the p_T discrimination performance for pions in Figure 4.7. Only a small $\mathcal{O}(10\%)$ fraction of low p_T tracks need to be passed in this way in order to generate the distribution seen in Figure 4.17 due to the large number of low transverse momentum charged pions generated in minimum bias events at SLHC.

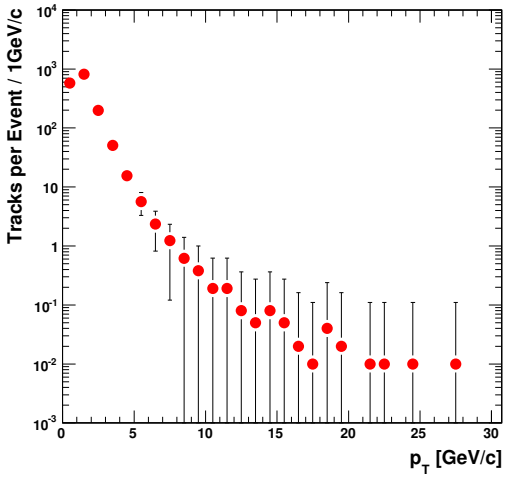


Figure 4.17: The p_T spectrum (normalised per event) for all minimum bias particles that generate stubs in a stacked layer placed at a radius of 25 cm with 1 mm sensor separation under SLHC conditions. Over 80% of generated stubs are from tracks with transverse momentum below 2 GeV/c.

4.4 The Double Stack Geometry

In order to provide the L1 trigger with track transverse momentum and primary vertex information and eliminate the low p_T background, the tracking trigger design could be supplemented by an additional stacked pixel layer. Each layer would be able to provide the necessary data rate reduction required for transmitting tracking information off detector by selection of high p_T track hits with good efficiency. Stub data from each layer could then be correlated as a standard track reconstruction

algorithm would at higher level. The advantage of this design would be that track p_T can be measured but no on-detector communication between layers would be needed, removing the need for high bandwidth links and a complex interconnection scheme between modules which greatly increases the power consumption and material of the system. Additionally, the use of stubs from stacks as opposed to hits from standard tracking layers would significantly reduce the number of combinatorics during reconstruction.

4.4.1 Track Reconstruction

The double stack track reconstruction algorithm will have to be implemented using high speed, low latency customised trigger electronics capable of calculating track parameters to within a few bunch crossings and therefore must not be too complex. The concept is presented in Figure 4.18.

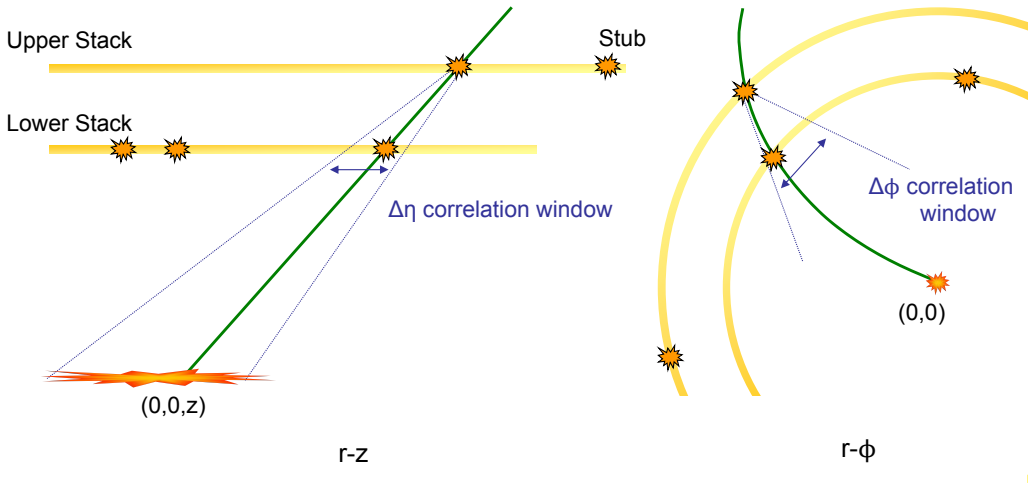


Figure 4.18: The double stack reconstruction method. Stubs from the inner stacked pixel layer which fall within a $\Delta\eta\Delta\phi$ window of a seed stub in the outer layer are correlated.

Correlation is initiated from stubs on the outer stack since studies indicate that at larger radii, fewer stubs and fakes are generated per event. This is partly because low momentum tracks fail to reach layers further out in radius. Stubs from the inner stack are successfully correlated if they fall within a $\Delta\phi\Delta\eta$ window of the seed

stub. Multiple correlations with the same seed stub are allowed. The window size in $\Delta\phi$ must be large enough to accept low p_T tracks and hence combinatorics are dependent on the minimum p_T cut required. In addition, the $\Delta\phi$ window must also be expanded to allow for multiple scattering within the inner layers. On the other hand, the $\Delta\eta$ window size is dominated by the size of the interaction region in z which limits the effectiveness of the correlation and purity of reconstruction.

Using the matched stubs, or track primitives, the transverse momentum can be calculated. Since the double stack provides only two hits, a third point is required to make the p_T measurement. This is taken to be the beam spot at (0,0) which is assumed to be the origin of the track. This approximation could affect the efficiency of the algorithm, especially since it neglects secondary interactions within the pixel system.

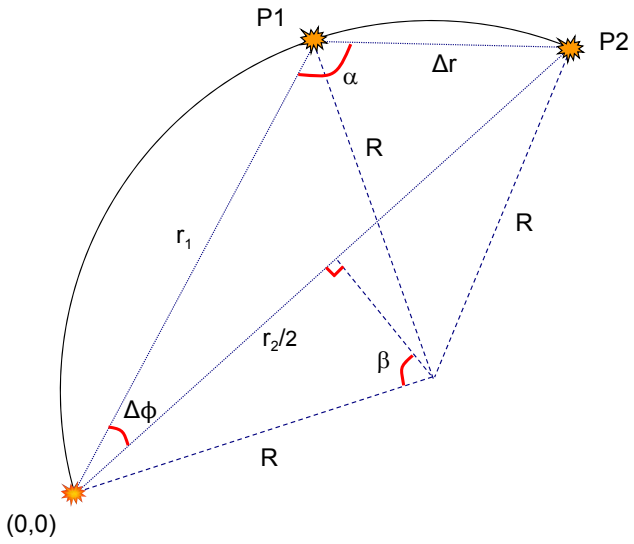


Figure 4.19: The three point reconstruction using two stub positions P1(r_1,ϕ_1), P2(r_2,ϕ_2) and the assumed vertex position (0,0). Using circle theorems, 2π-2β=2α.

Figure 4.19 defines the geometrical relationships between the three points used in the reconstruction of the track transverse momentum. Using the equation for the cosine,

$$\Delta r^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos(\Delta\phi) \tag{4.5}$$

and sine,

$$\frac{\sin(\alpha)}{r_2} = \frac{\sin(\Delta\phi)}{\Delta r} \tag{4.6}$$

substituting into the equation,

$$\sin(\beta) = \frac{r_2}{2R} \tag{4.7}$$

and using the relation sin(α)=sin(β), the radius of curvature R of the track can be written as,

$$R = \frac{\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\Delta\phi)}}{2\sin(\Delta\phi)} \tag{4.8}$$

where r_1 and r_2 are the radii in metres of the inner stack layer and outer stack layer respectively and $\Delta\phi$ is the angular separation in ϕ between the two stubs. This can then be used to calculate the transverse momentum which is related to the radius of curvature by,

$$\mathrm{p_T} = cBR \tag{4.9}$$

where B is the magnetic field strength in Tesla, c is the speed of light (m/s) and $\mathrm{p_T}$ is measured in eV/c.

The $\mathrm{p_T}$ resolution is defined as the spread in,

$$\frac{\delta(\mathrm{p_T})}{\mathrm{p_T}} = \frac{\mathrm{p_T^{reco}} - \mathrm{p_T^{true}}}{\mathrm{p_T^{true}}} \tag{4.10}$$

where $\mathrm{p_T^{reco}}$ is the transverse momentum reconstructed using the double stack and $\mathrm{p_T^{true}}$ is the Monte-Carlo particle transverse momentum. The $\mathrm{p_T}$ resolution is dependent on the rϕ pitch of the pixels within the stacked layer, however at low transverse momenta it is expected that the resolution will be limited by multiple scattering.

The track primitive can also be used to calculate the approximate z vertex. In this case, the resolution is dominated by the z pitch of the pixels and the radial position of the stacked layers. Increasing the separation between stacks or placing the stacks closer to the vertex would minimise the uncertainty in the z vertex due to the pixel z pitch.

The simulation conditions used to determine the performance of the double stack concept are identical to those previously defined for the single stack study (see Section 4.1). The two stacked layers are placed at radii of 25 cm and 35 cm with

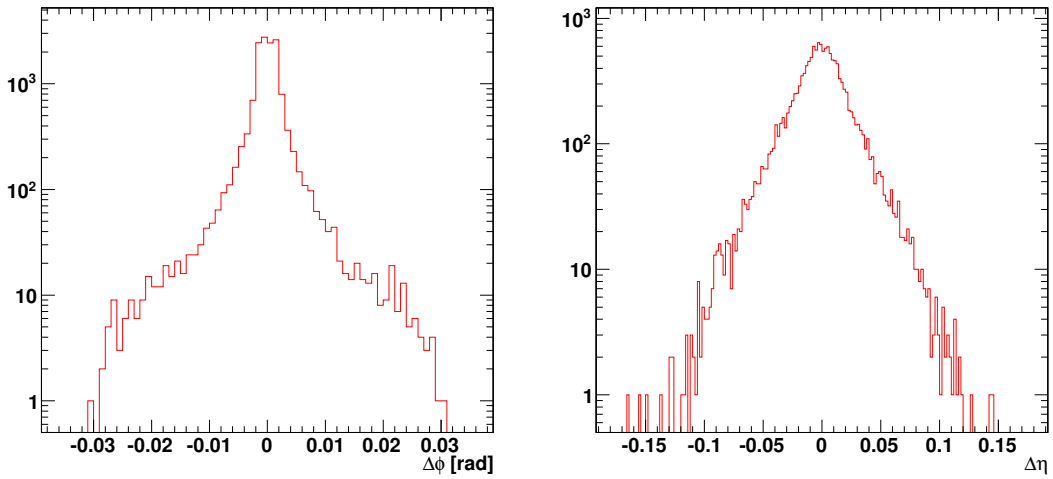


Figure 4.20: Distribution of; Left: $\Delta\phi$ and Right: $\Delta\eta$ difference between stubs on a stacked layer at 25 cm and another at 35 cm using single $\mu^{\pm}$ tracks with p_T>2 GeV/c.

coverage up to $|\eta|$=2.5 as set out in Table 3.1. Both layers use 100 μm thick sensors arranged in an untilted configuration with a sensor separation of 2 mm. Two window cuts for the single stack correlation are simulated; at 3 pixels and 5 pixels. Choosing identical stack parameters for each layer means that there is a small performance difference between the two (see Section 4.2.4). While the efficiency of the trigger is then limited by the the outer layer, the cost and complexity of manufacturing stacks with different layouts means that this is likely to be a necessary design choice.

The simulations also assume that some form of stack layer clustering is implemented, whether it is before or after correlation or possibly even off-detector, hence duplicate stubs are not input into the double stack reconstruction by selection of only one stub per Monte Carlo track for any given sensor. The $\Delta\phi\Delta\eta$ windows have been defined by simulating single muon tracks with p_T>2 GeV/c within the geometry as demonstrated by Figure 4.20. Stubs falling within a range of $|\Delta\phi|$ <0.02 and $|\Delta\eta|$ <0.1 of the seed stub are correlated. An additional cut is placed on the reconstructed z vertex where $|z_{vertex}|$ <0.15 m.

4.4.2 Reconstruction Performance

Figure 4.21 demonstrates the p_T discrimination performance of the double stack configuration for single muon tracks. The efficiency ε described here and throughout Section 4.4.2 is defined as follows,

$$\varepsilon(p_T) = \frac{S_{\text{RecoTracks}}(p_T)}{N_{\text{Tracks}}(p_T)} \tag{4.11}$$

where N_{Tracks} is the total number of tracks with Monte Carlo transverse momentum (p_T) which generate at least one pixel hit in the lower stacked pixel layer. $S_{\text{RecoTracks}}$ is the number of tracks with Monte Carlo transverse momentum (p_T) which are reconstructed successfully using the double stack geometry at least once.

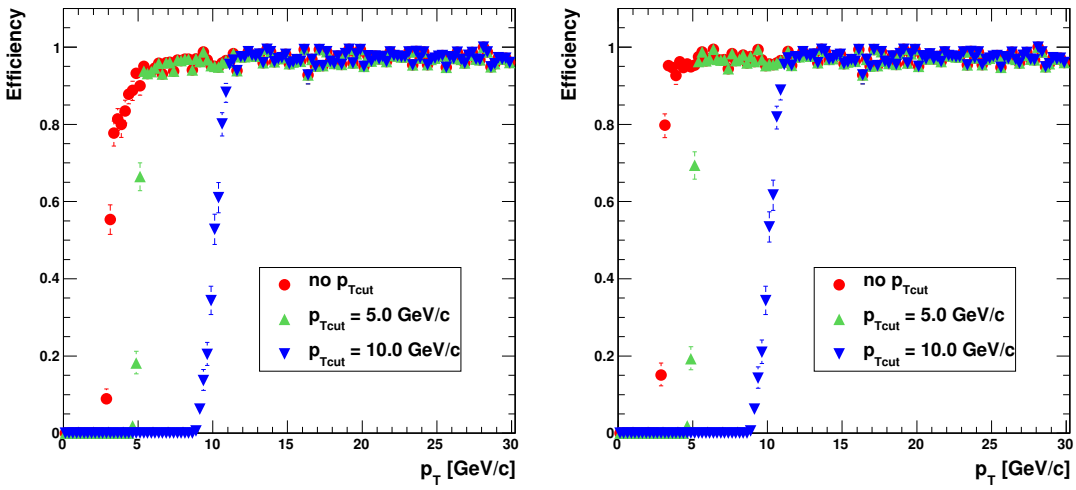


Figure 4.21: p_T discrimination performance using a double stack geometry for single $\mu^{\pm}$ tracks where; Left: the individual stacks use a tight correlation window of 3 pixels, and Right: the individual stacks use a loose correlation window of 5 pixels. The performance is shown when no cut is placed on the reconstructed transverse momentum and then when a cut of p_T=5,10 GeV/c is used.

The efficiency is of course primarily determined by the product of the individual stacks. However, since the correlation now provides a measurement of the track transverse momentum, a p_T cut is able to select tracks with good resolution. This is beneficial as the cut can be varied to adjust the data throughput within the trigger. With no cut on the reconstructed transverse momentum, the inefficiency at low p_T due to the tight single stack correlation window is visible.

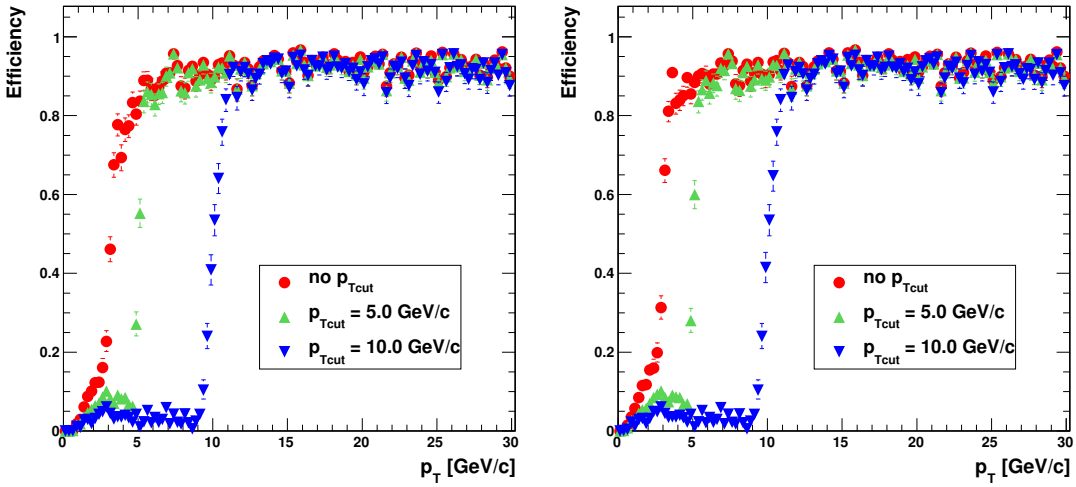


Figure 4.22: p_T discrimination performance using a double stack geometry for single $\pi^{\pm}$ tracks where; Left: the individual stacks use a tight correlation window of 3 pixels, and Right: the individual stacks use a loose correlation window of 5 pixels.

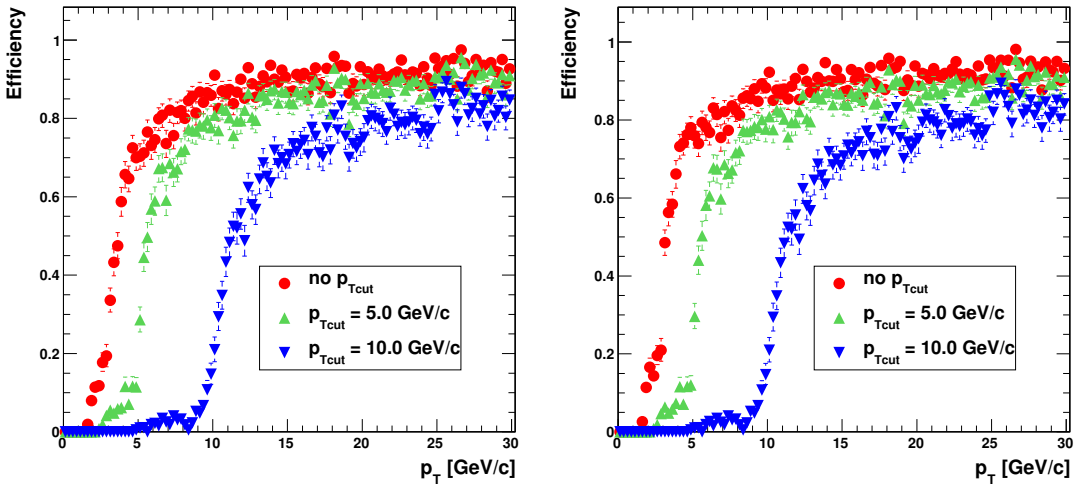


Figure 4.23: p_T discrimination performance using a double stack geometry for single $e^{\pm}$ tracks where; Left: the individual stacks use a tight correlation window of 3 pixels, and Right: the individual stacks use a loose correlation window of 5 pixels.

The choice of correlation window has little effect on the discrimination performance if a p_T cut >5 GeV/c is used.

Figures 4.22 and 4.23 show the same p_T discrimination performance for single pions and electrons respectively. As with the single stack, pions are reconstructed with a lower efficiency compared to muons due to a small fraction of tracks interacting in the inner detector and failing the single stack correlation. Similarly, pions with transverse momentum below the p_T cut can also be passed if they interact hadronically in the pixel system. Increasing the cut has a minimal effect on reducing this small percentage of low p_T tracks. In the case of electrons, the slow turn on in the efficiency remains although the double stack correlation removes the lowest p_T tracks. However, it is evident that the transverse momentum resolution for electrons is poor and that raising the p_T cut means that a significant fraction of high p_T tracks fail the algorithm. This poor resolution is due to bremsstrahlung of electrons so that the simple three point reconstruction algorithm with assumed beam spot underestimates the electron p_T.

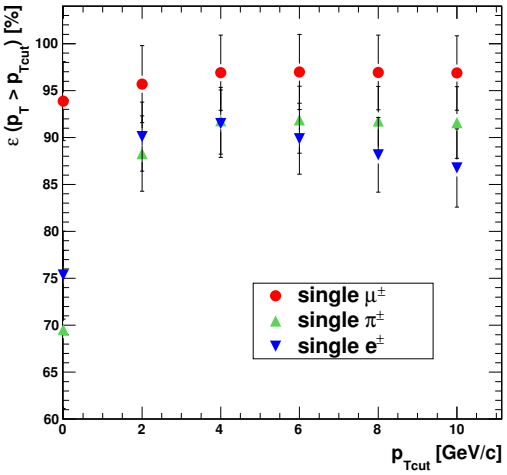


Figure 4.24: Global trigger efficiency for $\mu^{\pm}$, $\pi^{\pm}$ and $e^{\pm}$ tracks with reconstructed transverse momenta greater than p_{Tcut}, as a function of the p_T cut. Efficiencies, as defined as in equation (4.11), are calculated using tracks with Monte Carlo transverse momentum $p_T > p_{Tcut}$ and a Monte Carlo track distribution flat in p_T from 0 to 50 GeV/c. In this case, the individual stacks use a tight correlation window of 3 pixels.

The efficiency for triggering on muon, pion and electrons tracks with transverse momentum above the p_T cut is given in Figure 4.24. The plot indicates that the

efficiency for detecting muons and pions is approximately constant (>96%) with the choice of p_T cut for values greater than 4 GeV/c. As demonstrated by Figure 4.23 already, the efficiency for triggering on electrons with transverse momentum above the p_T cut drops as the cut is increased beyond 4 GeV/c. This is due to the poor reconstruction of bremsstrahlung electrons.

The average number of reconstructed tracks in a SLHC event as a function of the p_T cut is presented in Figure 4.25 for the two single stack correlation window options. If the stack correlation window is relaxed to 5 pixels so as to read out stubs from tracks with transverse momentum down to 2 GeV/c for L1 tracker isolation cuts, the average number of reconstructed tracks per event is greater than 100, even if a p_T cut of 10 GeV/c is applied. The reconstructed tracks are predominantly fake tracks; defined here as combinatorial fakes where stubs from the upper and lower stacks are incorrectly correlated.

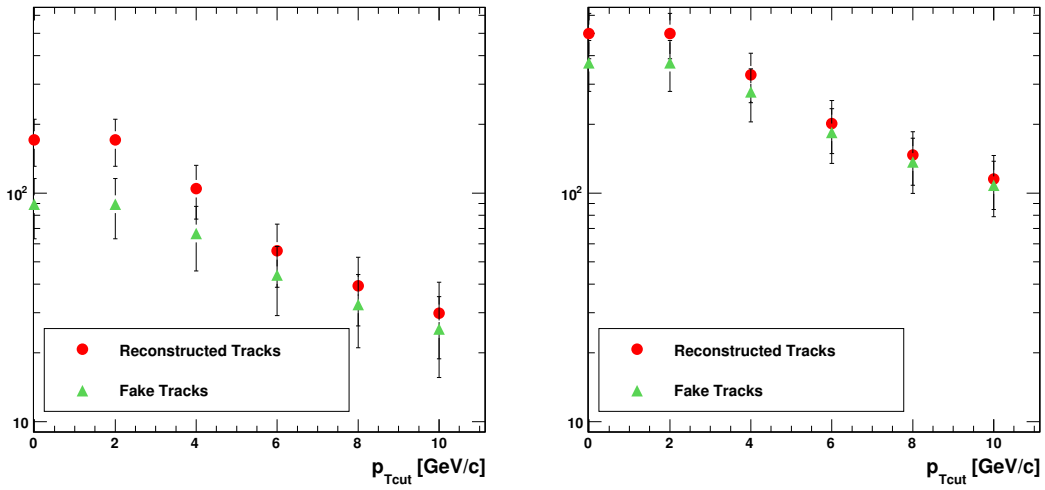


Figure 4.25: Average number of total and fake reconstructed tracks per event under SLHC conditions as a function of the p_T cut where; Left: the individual stacks use a tight correlation window of 3 pixels, and Right: the individual stacks use a loose correlation window of 5 pixels. The number of combinatorial fake tracks per event dominate the rate.

Combinatorial Fakes

The high fraction of fake tracks per event is due to the large number of stubs generated by each layer using the loose stack window as well as the size of the

$\Delta\phi\Delta\eta$ matching window between the two layers. Tightening the matching window would be difficult since this would lead to a loss in reconstruction efficiency (Figure 4.20). Reducing the $\Delta\phi$ window to increase the minimum p_T at which track could be reconstructed is feasible, however it is the much larger irreducible window in $\Delta\eta$ which contributes the most to the combinatorial background**. This comes from the fact that the pixel z pitch is large so that stubs are not able to identify the z vertex and therefore the window must be wide enough to accept tracks originating from anywhere inside the luminous region.

If instead the stack correlation window is tightened to 3 pixels so that each stack produces less than half the number of stubs (see Tables 4.2 and 4.3), the number of combinatorial fakes and hence total reconstructed tracks per event is drastically reduced as demonstrated in Figure 4.25. On the other hand, this comes at the price of losing tracking information from tracks with p_T<4 GeV/c (Figure 4.4) which one might require for tracker isolation cuts at L1††.

The reconstructed transverse momentum distribution of combinatorial fakes for the two single stack correlation windows is provided by Figure 4.26. There is a clear bias towards a low p_T (an artefact of the equation for the three point transverse momentum reconstruction) but the p_T cut would have to be raised significantly in order to remove the majority of fakes, especially when using a loose single stack correlation window.

Misreconstructed Fakes

In addition to combinatorial fake tracks, the double stack can also reconstruct low p_T tracks, which would otherwise not pass the algorithm, with an apparent high transverse momentum. Figure 4.27 demonstrates the relationship between the Monte-Carlo track p_T and the reconstructed p_T for minimum bias events at SLHC and how

**A $\Delta\phi$ window of ±0.02 corresponds to an rϕ window of ±0.5 cm when projecting to a stacked pixel layer at 25 cm but a $\Delta\eta$ window of ±0.1 corresponds to a z window of ±5 cm (at η=0).

††If stub information containing the row difference between the hit pixels could be read out from the tracker, a two stage rate reduction could be implemented if required. A wide single stack correlation window would allow low p_T tracking information to reach the trigger so that isolation could be performed. A tighter cut could then be placed on the row difference recorded by the stub in order to identify higher transverse momentum tracks before the double stack matching and p_T reconstruction.

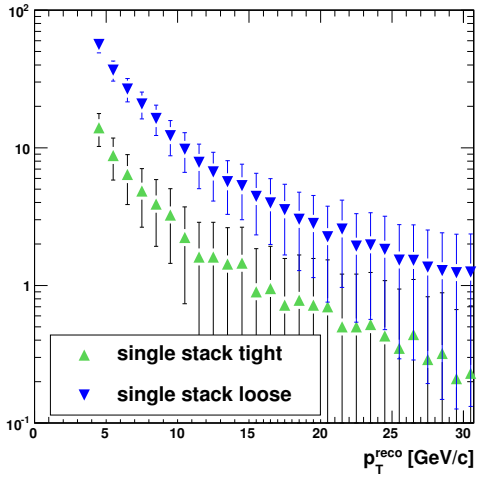


Figure 4.26: Average distribution of combinatorial fake tracks per event under SLHC conditions as a function of their reconstructed transverse momenta for both tight (3 pixel) and loose (5 pixel) individual stack correlation cuts. The p_T cut is set at 4 GeV/c.

a number of low transverse momentum tracks pass the p_T cut. The right hand plot shows that if only tracks originating from the beam spot are selected, the low p_T background is removed. This indicates that these particles are badly reconstructed because they have interacted in the tracker or are products from secondary interactions.

Since the success of the reconstruction algorithm depends on the assumption of the track vertex lying at the beam spot, the low transverse momentum background is irreducible without a third point in the reconstruction. The fraction of fakes could again be reduced by imposing a higher p_T cut. For example, if the tight single stack correlation window is used, a p_T cut of 4 GeV/c passes approximately 8 low momentum tracks per event while increasing the cut to 10 GeV/c reduces the background to ∼2 per event.

Figures 4.28 and 4.29 show the true versus reconstructed track transverse momentum performance for single pions and electrons respectively. By selecting tracks originating from the vertex, the low p_T background is again removed. It is clear that pions are more likely to interact in the detector and be reconstructed incorrectly as earlier results have indicated. The poor reconstruction of bremsstrahlung electrons

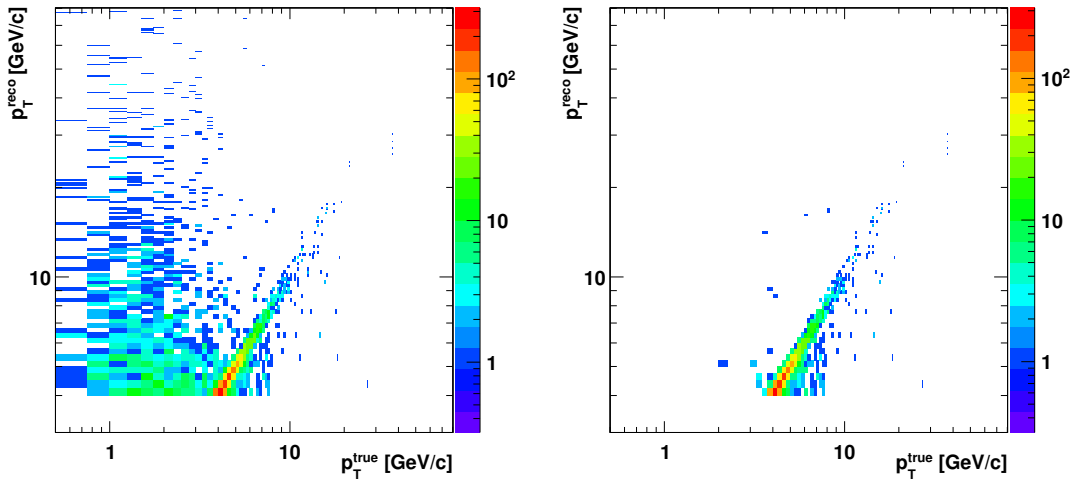


Figure 4.27: Distribution of Monte-Carlo track transverse momentum (p_T^{true}) and measured transverse momentum (p_T^{reco}) for minimum bias particles under SLHC conditions reconstructed by the double stack and passing the 4 GeV/c p_T cut. Left: for all tracks, and Right: for tracks only originating from the beam spot. The individual stacks use a tight correlation window of 3 pixels.

where track momentum can be significantly underestimated is also apparent. The transverse momentum resolution measurement is presented in the next section.

The double band observable for low momentum electrons in Figure 4.29 is shown to be due to tracks not originating from the beam spot. The second band is in fact caused by bremsstrahlung photons from the original electron converting in the inner detector to electron pairs and being reconstructed as high p_T tracks due to their displaced vertex. The fact that these tracks are reconstructed with a transverse momentum ratio of p_T^{reco}/p_T^{true} ~2 can be explained by the photons converting at a particular radius, as described by Figure 4.30. Most of the conversion vertices are found at r~18 cm and r~20 cm where the last inner pixel layer and pixel support tube are placed. This is expected since, as more material is traversed, the number of bremsstrahlung photons generated (and then converting) increases with radius. The ratio between reconstructed and true p_T can be calculated using a similar method to the one used in Section 4.4.1. Assuming that the products of the photon conversion travel in the direction of the original photon, the calculated ratio is given alongside the simulation data in Figure 4.30. This effect highlights the fact that the material in the inner tracking region must be minimised in order to provide good resolution of track transverse momentum and a low fake rate.

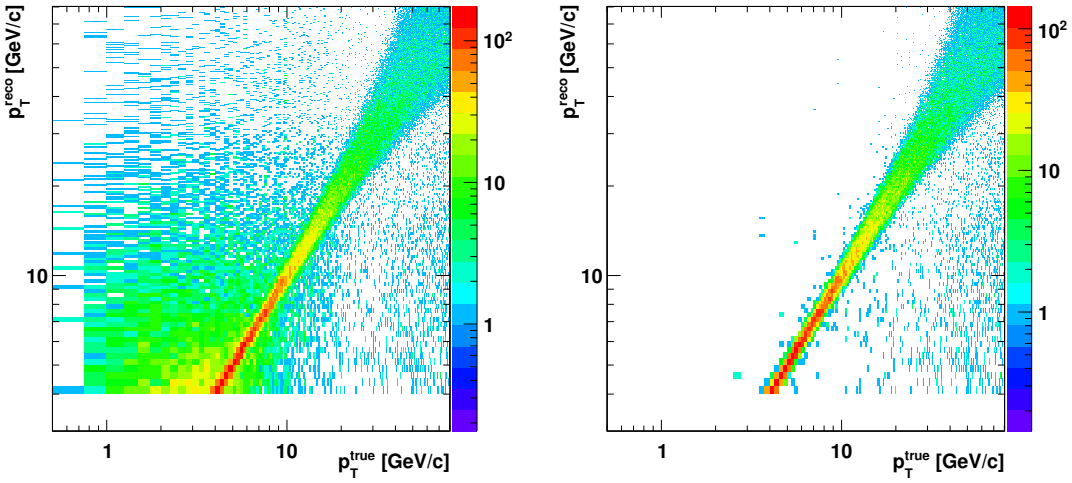


Figure 4.28: Distribution of Monte-Carlo track transverse momentum (p_T^{true}) and measured transverse momentum (p_T^{reco}) for single $\pi^{\pm}$ tracks reconstructed by the double stack and passing the 4 GeV/c p_T cut. Left: for all tracks, and Right: for tracks only originating from the beam spot. The individual stacks use a tight correlation window of 3 pixels.

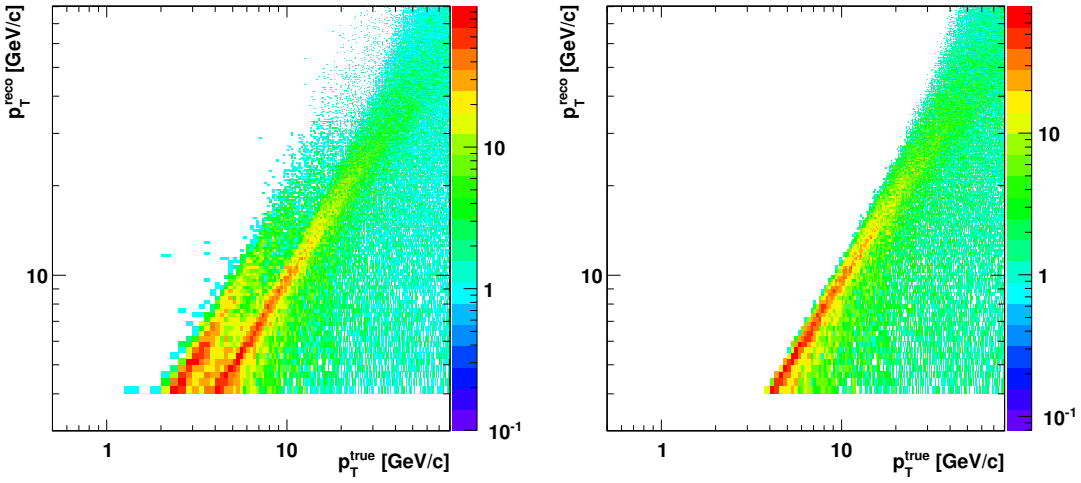


Figure 4.29: Distribution of Monte-Carlo track transverse momentum (p_T^{true}) and measured transverse momentum (p_T^{reco}) for single $e^{\pm}$ tracks reconstructed by the double stack and passing the 4 GeV/c p_T cut. Left: for all tracks, and Right: for tracks only originating from the beam spot. The individual stacks use a tight correlation window of 3 pixels.

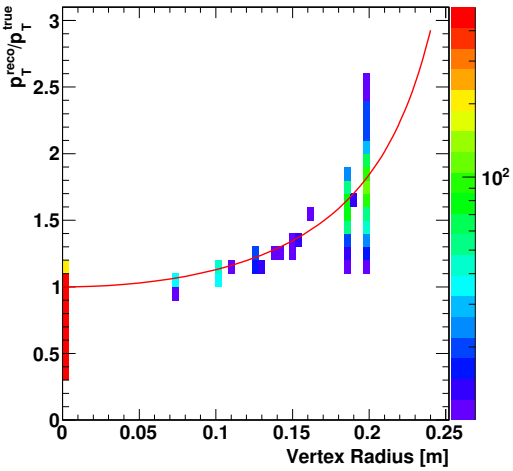


Figure 4.30: Relationship between ratio of reconstructed to true track transverse momentum p_T^{reco}/p_T^{true} to track vertex radius for single electron events. The p_T of the original electrons starting from the beam spot are well reconstructed although due to bremsstrahlung the transverse momentum can be underestimated. Electrons from photon conversions are reconstructed with increased p_T depending on the radius of their conversion vertex. An analytical calculation of the ratio (red curve) demonstrates good agreement.

It is difficult to estimate how many reconstructed tracks per event would be acceptable in order to allow the Level 1 trigger rate to be reduced sufficiently without matching studies between tracker and calorimeter or muon objects. While the p_T cut offers a method to reduce the number of tracks, including fakes, it is clear that it also affects the efficiency of detecting high p_T particles, in particular electrons. For this reason, the cut must not be set too high when matching tracks to calorimetric clusters at L1.

Assuming naïvely that the increase in the number of calorimeter objects at L1 scales with luminosity for identical cuts, one would expect that the addition of tracking information is required to reduce the rate by a factor of ∼10-20. If the positional resolution at the surface of the ECAL is approximately that of a calorimeter trigger tower, the number of tracks should not exceed ∼180-370. However, if the track resolution is significantly worse or if the tracker occupancy has been severely underestimated, two stacked pixel layers may then not be sufficient for reducing the L1 trigger rate. Track position resolution at the ECAL surface is discussed in the following section.

If two stacked pixel layers is not enough to reduce the amount of tracking information to the L1 trigger, a third layer could be implemented. A third stack would have the benefit of increasing the track transverse momentum resolution but more importantly it would reduce the number of combinatorial fakes and tracks from secondary interactions due to knowledge of the track charge, transverse momentum and z vertex. On the other hand, a third stacked pixel layer would have the penalty of increased power dissipation within the tracker, the additional associated material due to cooling and electronics as well as other factors such as cost. If a stacked layer using a tight pixel correlation window were to be placed at ∼15 cm from the beam pipe instead of a fourth pixel layer, some of these concerns may be alleviated so long as the stacks were able to contribute to tracking after L1.

4.4.3 Resolution Performance

It is important to have an estimate of the position resolution of the reconstructed track at the ECAL surface so that tracking information can be combined with calorimetric objects within the Level 1 trigger efficiently. A resolution better than that of a ($\Delta\eta\Delta\phi$=0.087×0.087) calorimeter trigger tower would be sufficient for matching and would minimise the number of fake correlations passing the trigger. Figure 4.31 shows the $\Delta\phi$ and $\Delta\eta$ resolutions at the ECAL surface for reconstructed single electrons, pions and for reference, muons.

The $\Delta\phi$ resolution should be dependent on the track transverse momentum, as equations (4.9) and (4.8) demonstrate. However it is clear that interactions within the tracker limit the resolution for electrons and pions at high p_T in contrast to the result for muons which do not interact. The spread in $\Delta\phi$ is at maximum 0.02 radians for low p_T tracks, corresponding to ∼2.6 cm at the ECAL, much smaller than the trigger tower size in ϕ. The maximal spread in $\Delta\eta$ on the other hand is to within two trigger towers in η, at least in the central region. In the region $|\eta| > 1$, the resolution improves to within a single trigger tower. It is the large pixel pitch in z which is the dominant contributor to the poorer resolution in $\Delta\eta$, however this is negated at high pseudorapidity where the z projection per unit η increases.

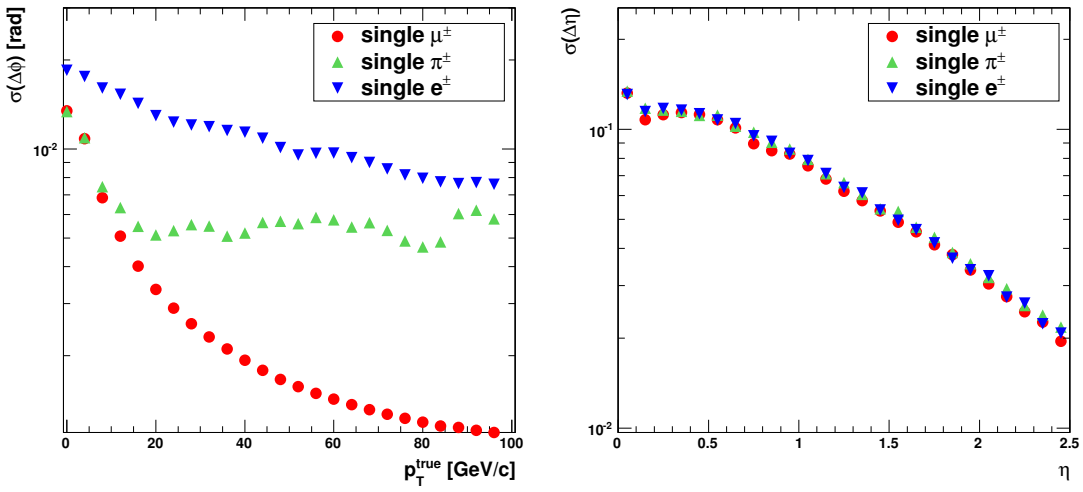


Figure 4.31: Left: $\Delta\phi$, and Right: $\Delta\eta$ (rms) resolutions at the ECAL surface for reconstructed $\mu^{\pm}$, $\pi^{\pm}$ and $e^{\pm}$ tracks using a double stack layer. Results are for real tracks passing the 4 GeV/c p_T cut and a tight correlation window of 3 pixels for the individual stacks.

The transverse momentum resolution for single muons, pions and electrons with momenta up to 80 GeV/c is provided in Figure 4.32. The double stack is able to measure track p_T to within 20% for pions and muons up to 50 GeV/c, although the resolution is slightly worse for electrons. This is because the algorithm underestimates the transverse momentum of the electron if it has radiated energy before reaching the stacked layers.

While the p_T resolution is certainly not acceptable for tracking, it does offer a coarse method for cutting on the transverse momentum so that trigger rates can be reduced if required. It also provides the trigger with an additional cut when matching tracks to electromagnetic clusters by calculation of the E_T/p_T ratio or when matching to muon objects. Increasing the separation between the two stacks would improve the p_T resolution.

Figure 4.32 also shows that the z vertex resolution as a function of η is approximately constant and that the reconstruction is able to determine the vertex point to within 3 mm. Again, the vertex resolution is determined by the pixel pitch in z but it could be improved by bringing the layers closer to the interaction point or by radially spacing the stacks in equal intervals. While a 3 mm z resolution would not be able

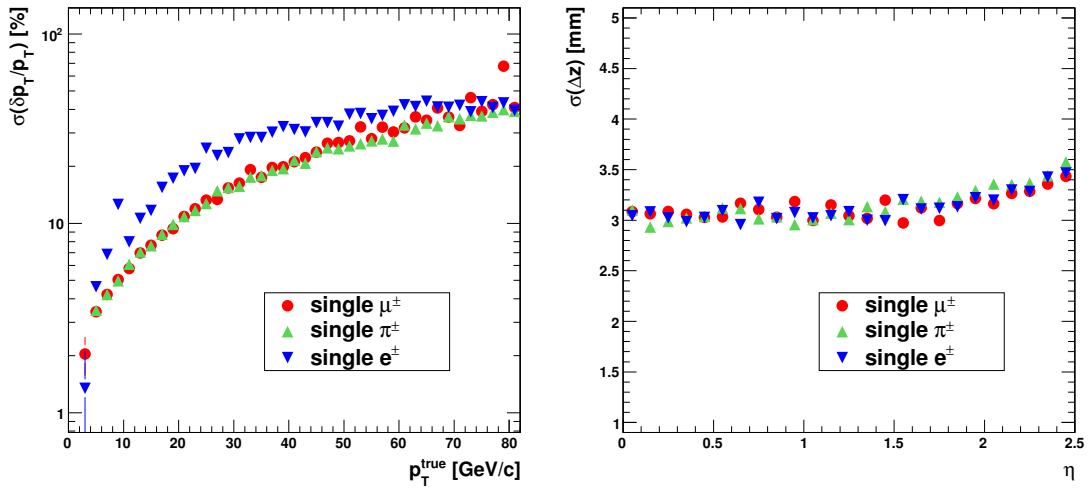


Figure 4.32: Left: Transverse momentum, and Right: z vertex (rms) resolutions for reconstructed $\mu^{\pm}$, $\pi^{\pm}$ and $e^{\pm}$ tracks using a double stack layer. Results are for real tracks passing the 4 GeV/c p_T cut and a tight correlation window of 3 pixels for the individual stacks.

to identify the primary vertex out of the ∼400 interactions per event, it could identify a region of interest within the luminous region for a combined trigger vertex cut.

4.5 Summary

A detailed study has been performed in order to determine the viability of stacked pixel layers for providing reduced tracking information to the L1 trigger. It has been demonstrated that realistic simulations including tracker material before and within the trigger layer verify results from previous studies which did not take these considerations into account.

Simulations have shown that the stack correlation algorithm is able to reject low p_T tracks while maintaining good efficiencies for tracks with p_T>2 GeV. This performance has been measured for a number of different layer configurations. For the given geometry, an untilted stacked layer at 25 cm with 2 mm sensor separation and a 5 pixel correlation window would provide a sufficient reduction in rate while offering some flexibility in reducing the rate further by applying a more stringent correlation cut. An untilted layout may be preferred due to its simpler implementation

and better position resolution if hit data were to be used in later track reconstruction. However, if the readout bandwidth is severely limited or if occupancies at the LHC have been underestimated, a tilted stacked layer would be able to cut the data volume by a factor of two over an identically configured untilted layer.

Because of the large fraction of low transverse momentum hadrons generated in minimum bias events at SLHC, a significant number of low p_T tracks interact within the tracker and pass the correlation cut. This large irreducible background means that a single stacked pixel layer would not provide the L1 algorithms with useful information with which to lower trigger rates, unless the effective p_T cut is significantly raised by increasing the sensor separation and/or radius. This motivates a tracking trigger solution which requires two stacked pixel layers, where the stubs generated by each layer are correlated to reduce the low p_T background and used to reconstruct the track transverse momentum.

Results show that two stacked layers could be used to provide useful information to the L1 trigger in order to reduce rates. Reconstructed track angular resolutions would be small enough to match tracks to trigger towers while the reconstructed track p_T resolution would be able to provide a second transverse momentum cut. Track reconstruction can occur off-detector and algorithms could easily be implemented in FPGAs supplemented with fast multi-Gigabit serial links to match stubs between layers and regions (although processing time will be determined by the number of stub combinations).

The double stack tracking trigger is affected by a large fake background, again due to secondary interactions of low p_T hadrons within the tracker. Further studies matching tracks to calorimeter deposits or muon tracks will determine whether this background will affect the L1 trigger rate or if a third stacked pixel layer is required to eliminate the fake tracks[93]. A third layer will carry the penalty of increased power dissipation within the tracker and contribute added material so it is vital that detailed two versus three stack trigger and tracking studies are performed in the near future.

Chapter 5

Conclusions

The CMS experiment is a general purpose LHC detector which has been designed and optimised to discover the Higgs boson and signatures of new physics beyond the Standard Model. At the time of writing, the CMS detector is fully installed and is ready for first LHC collisions in the coming months. The tracker has been commissioned under full magnetic field (CRAFT09), successfully reconstructing over 12 million tracks and aligning the detector to within 30 μm. More recently, the tracker readout system and DAQ have been stressed to over 100 kHz without error and has further demonstrated that the full system will be able to throttle the trigger rate during nominal operation.

The success of these tests owes much to the earlier integration and commissioning efforts which have proved invaluable in verifying that the entire system is qualified for data taking. Some of these operations have been described in Chapter 2. The APVe, which plays a crucial role in the synchronisation of the tracker by deterministic calculation of the front end buffer occupancy and by monitoring the status of the Front End Drivers, has been successfully commissioned with the tracker readout and trigger control systems. Tests with the Global Trigger Controller have demonstrated that the APVe is capable of throttling L1 trigger decisions at rates up to 140 kHz. During a full readout slice test including the APVe, it was shown that the tracker could be operated up to occupancies of 3% without data loss due to bandwidth throttling. During tracker commissioning at the Tracker Integration Facility,

a previously unknown effect where the channel occupancy was seen to increase with trigger rate was investigated. The early diagnosis of this high rate noise effect has meant that testing of the tracker with the rest of the CMS detector has not been disrupted, allowing successful operation in recent runs such as CRAFT09. Since the noise is due to the internal operation of the APV25 and is therefore deterministic, the APVe will be used to veto triggers where it calculates these effects will occur.

At Super-LHC, it is proposed that the machine luminosity will be increased to $10^{35}\text{cm}^{-2}\text{s}^{-1}$. The increased particle fluxes and radiation environment will necessitate the complete replacement of the current CMS tracker while presenting the design of a new tracker with severe challenges. Power consumption is one of the main challenges for the tracker readout system since a higher granularity detector will be required. Physics performance must not be compromised so the tracker material contribution should be lowered where possible. In addition, it is likely that the Level 1 system will require information from the tracker in order to reduce the trigger rate. It has been shown that in order for this to occur, the tracker will have to significantly reduce the on-detector data rate before readout due to the huge power and cabling requirements.

The stacked tracking concept, where only hits from tracks with transverse momenta above a cut are read out, has been shown to be a viable method to achieve this data rate reduction. A detailed and realistic simulation of a concept tracker geometry has been developed and the simulations report that a stacked tracking layer at 25 cm would be capable of reducing the detector data rate by a factor of $\sim$20 while maintaining efficiencies in excess of 96% for tracks with $p_T > 2\,\text{GeV/c}$. This performance has been measured for a number of different layer configurations showing that the concept is flexible enough to adapt to the requirements of a new tracker at SLHC. The simulations have also shown that the performance of a stacked layer is robust against pileup which is essential since our uncertainty on the expected occupancy at SLHC is large.

It has been shown that the information provided by a single stacked layer would not be useful for reducing the L1 trigger rate, unless the effective p_T cut is raised. This

is due to the large number of low p_T hadrons interacting in the inner tracker and passing the correlation. Apart from the effect on triggering performance, a higher p_T cut may also be difficult to implement since it will require a layer at larger radius (increasing power consumption), or with larger sensor separation (increasing the number of fakes), or pixels with larger dimensions (affecting tracking performance).

Two stacked pixel layers could possibly be used to provide useful information to the L1 trigger in order to reduce rates. Tracks can be reconstructed with transverse momentum resolution $\sigma(\delta(p_T)/p_T) < 20\%$ for p_T<20 GeV/c allowing a variable cut to be placed on the track p_T within the L1 trigger algorithm, in whatever form it takes. As an additional example, it could be used to provide an additional E_T/p_T cut when matching tracks to ECAL calorimeter objects. Simulations have shown that tracks could be reconstructed with sufficient resolution so as to match tracks with L1 calorimeter objects.

The double stack tracking trigger is affected by a large fake background, again due to secondary interactions of low p_T hadrons within the tracker. Further studies matching tracks to calorimeter deposits or muon tracks will determine whether this background will affect the L1 trigger rate or if a third stacked pixel layer is required to eliminate the fake tracks. A third layer will carry the penalty of increased power dissipation within the tracker and contribute added material so it is vital that detailed two versus three stack trigger studies with matching to other L1 trigger objects are performed in the near future. In addition, simulations will be required to parametrise the impact of multiple stacked layers on tracking performance. Nevertheless, it has been important to demonstrate that a future tracker instrumented with stacked pixel layers could provide a L1 trigger with tracking information at SLHC.

References

[1] LHC Project http://lhc.web.cern.ch/lhc/.

[2] LEP Electroweak Working Group http://lepewwg.web.cern.ch/LEPEWWG.

[3] Tevatron Project http://www-bdnew.fnal.gov/tevatron/.

[4] LHC Commissioning with Beam http://lhc-commissioning.web.cern.ch/lhc-commissioning/.

[5] CMS Collaboration http://cms.cern.ch/.

[6] ATLAS Collaboration http://atlas.web.cern.ch/Atlas/index.html.

[7] LHCb Collaboration http://http://lhcb.web.cern.ch/lhcb/.

[8] ALICE Collaboration http://aliceinfo.cern.ch/.

[9] A. Moraes, C. Buttar, and I. Dawson, "Prediction for Minimum Bias and the Underlying Event at LHC Energies," *Eur. Phys. J. C* **50** (2007) no. 2, p435–466.

[10] The CMS Collaboration, "CMS Expression of Interest in the SLHC," CERN-LHCC/2007-014.

[11] The CMS Collaboration, "CMS Physics Technical Design Report Volume 1: Detector Performance and Software," CERN-LHCC/2006-001.

[12] F. Zimmermann, "CERN Upgrade Plans for the LHC and its Injectors," Tech. Rep. CERN-sLHC-PROJECT-Report-0016, CERN, 2009.

[13] W. Scandale and F. Zimmermann, "Scenarios for sLHC and vLHC," *Nuclear Physics B - Proceedings Supplements* **177-178** (2008) p207–211.

[14] ALEPH CDF D0 DELPHI L3 OPAL SLD Collaborations, "Precision Electroweak Measurements and Constraints on the Standard Model," CERN-PH-EP/2009-023.

[15] D. Griffiths, *Introduction to Elementary Particles*. John Wiley and Sons, 1987.

[16] T. Cheng and F. Li, *Gauge Theory of Elementary Particle Physics*. Oxford University Press, 2000.

[17] C. Amsler et. al. (Particle Data Group) *Phys. Rev. Lett.* **B** (2008) no. 667, p1.

[18] S. Dawson, "Introduction to Electroweak Symmetry Breaking," hep-ph/9901280.

[19] A. Djouadi, "The Anatomy of Electro-Weak Symmetry Breaking, Tome 1," hep-ph/0503172.

[20] A. Djouadi, "The Anatomy of Electro-Weak Symmetry Breaking, Tome 2," hep-ph/0503173.

[21] T. Greenshaw *et al.*, "Proceedings of the School for Experimental High Energy Physics Students," Tech. Rep. RAL-TR-2007-004, RAL, 2007.

[22] The CMS Collaboration, "CMS Physics Technical Design Report Volume 2: Physics Performance," CERN-LHCC/2006-021.

[23] A. Nikitenko *et al.*, "Search for the Standard Model Higgs Boson Produced in Vector Boson Fusion and Decaying into $\tau\tau$ Pair in CMS with $1\,\mathrm{fb}^{-1}$," CMS-PAS-HIG-08-008.

[24] F. Gianotti, M. Mangano, T. Virdee, *et al.*, "Physics Potential and Experimental Challenges of the LHC Luminosity Upgrade," hep-ph/0204087.

[25] M. Noy, "Development and Characterisation of the Compact Muon Solenoid Silicon Microstrip Tracker Front End Driver,". PhD thesis, London Univ., 2004.

[26] The CMS Collaboration, "The CMS Experiment at the CERN LHC," *Journal of Instrumentation* **3** (2008) no. 08, S08004.

[27] The CMS Collaboration, "The Tracker Project Technical Design Report," CERN-LHCC/98-006.

[28] The CMS Collaboration, "Addendum to the CMS Tracker TDR," CERN-LHCC/2000-016.

[29] J. D'Hondt, "The CMS Tracker System," in *Nuclear Science Symposium Conference Record, 2005 IEEE*, pp. 1084–1087. 2005.

[30] M. Raymond *et al.*, "The CMS Tracker APV25 0.25μm CMOS Readout Chip," CERN-LHCC/2000-041.

[31] The CMS Collaboration, "The CMS Silicon Strip Tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **573** (2007) no. 1-2, p257–259.

[32] M. Pioppi, "Iterative Tracking," CMS IN-2007/065.

[33] D. Kotlinski *et al.*, "The Control and Readout Systems of the CMS Pixel Barrel Detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **565** (2006) no. 1, p73–78.

[34] The CMS Collaboration, "The TriDAS Project Technical Design Report, Volume 1: The Trigger Systems," CERN-LHCC/2000-038.

[35] The CMS Collaboration, "The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger," CERN-LHCC/2002-026.

[36] V. Brigljevic *et al.*, "Run Control and Monitor System for the CMS Experiment," cs/0306110.

[37] A. Racz *et al.*, "CMS Data to Surface Transportation Architecture," CERN-LHCC/2002-034.

[38] L. Orsini and J. Gutleber, "The XDAQ Framework." `http://xdaqwiki.cern.ch/`.

[39] V. Brigljevic *et al.*, "Using XDAQ in Application Scenarios of the CMS Experiment," `hep-ex/0305076`.

[40] T. Rohe *et al.*, "Radiation Hardness of CMS Pixel Barrel Modules," `arXiv:1001.0666 [physics.ins-det]`.

[41] W. Erdmann *et al.*, "Upgrade Plans for the CMS Pixel Barrel Detector," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **In Press** (2009) , `DOI:10.1016/j.nima.2009.10.029`.

[42] C. Roland, "Track Reconstruction in Heavy Ion Collisions with the CMS Silicon Tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **566** (2006) no. 1, p123–126.

[43] J. Whitmore, "CMS Hadron Calorimeter Front-End Upgrade for SLHC Phase I," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **In Press** (2009) , `DOI:10.1016/j.nima.2009.08.088`.

[44] S. Dugad, "Upgrade Plans for Hadron Calorimeter in the CMS Detector," `CMS-CR-2009-057`.

[45] M. Moll, "Recent Developments to Improve the Radiation Tolerance of Semiconductor Sensors." `http://sirad.pd.infn.it/scuola_legnaro Presentazioni_Web/04_ScuolaLNL2009_Moll_Web.pdf`.

[46] M. Raymond and G. Hall, "CMS Microstrip Tracker Readout at the SLHC," in *Topical Workshop on Electronics for Particle Physics*, pp. 354–358. 2008.

[47] M. Weber, "Power Distribution for SLHC Trackers: Challenges and Solutions," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **592** (2008) no. 1-2, p44–55.

[48] Micro-TCA Standard
http://www.picmg.org/pdf/MicroTCA_Short_Form_Sept_2006.pdf.

[49] M. Bachtis, "SLHC Calorimeter Trigger at Phase I."
http://indico.cern.ch/conferenceDisplay.py?confId=45544.

[50] J. Varela, "CMS L1 Trigger Control System," CMS-NOTE-2002-033.

[51] CMS TTC Homepage
http://cmsdoc.cern.ch/cms/TRIDAS/ttc/modules/ttc.html.

[52] A. Racz *et al.*, "Trigger Throttling System for CMS DAQ,"
CERN-LHCC/2000-041.

[53] CMS Tracker Collaboration, "Tracker Operation and Performance at the Magnet Test and Cosmic Challenge," CMS-NOTE-2007-029.

[54] CMS Tracker Collaboration, "Performance Studies of the CMS Strip Tracker before installation," arXiv:0901.4316 [physics].

[55] J. Troska *et al.*, "Optical Readout and Control Systems for the CMS tracker," *Nuclear Science, IEEE Transactions on* **50** (2003) no. 4, p1067–1072.

[56] L. Jones *et al.*, "The APV25 Deep Submicron Readout Chip for CMS Detectors," CERN-LHCC/99-033.

[57] M. Wingham, "Commissioning of the CMS tracker and preparing for early physics at the LHC,". PhD thesis, London Univ., 2008.

[58] The I^2C Specification http://www.i2c-bus.org/.

[59] R. Bainbridge, "Influence of Highly Ionising Events on the CMS APV25 Readout Chip,". PhD thesis, London Univ., 2004.

[60] G. Iles *et al.*, "Performance of the CMS Silicon Tracker Front-End Driver," in *Tenth Workshop on Electronics for LHC Experiments and Future Experiments*, p. 222. 2004.

[61] Virtex-II Platform FPGA Handbook http://www.xilinx.com.

[62] J. A. Jones, "Development of Trigger and Control Systems for CMS,". PhD thesis, London Univ., 2006.

[63] C. Foudas *et al.*, "The CMS Tracker Readout Front End Driver," physics/0510229.

[64] S-LINK Specification http://www.cern.ch/HSI/s-link/.

[65] F. Drouhin, L. Gross, *et al.*, "The CERN CMS Tracker Control System," in *Nuclear Science Symposium Conference Record, IEEE*, vol. 2, pp. p1196–1200. 2004.

[66] A. Racz *et al.*, "The Final Prototype of the Fast Merging Module (FMM) for Readout Status Processing in CMS DAQ," in *Tenth Workshop on Electronics for LHC Experiments and Future Experiments*, p. 316. 2004.

[67] The IDAQ Card Overview http://www.hep.ph.ic.ac.uk/calice/maps/IDAQV1Manual.doc.

[68] VME Bus Specification http://www.vita.com/specifications.html.

[69] G. Iles, "The APV Emulator to Prevent Front-End Buffer Overflows Within the CMS Silicon Strip Tracker," CERN-LHCC/2002-034.

[70] APVe Firmware Manual http://giles.web.cern.ch/GILES/projects/apve/APVE_Firmware_Technical_Description_v1.0.pdf.

[71] The Hardware Access Libraries http://cmsdoc.cern.ch/∼cschwick/software/documentation/HAL/index.html.

[72] J. Fulcher *et al.*, "Recent Results on the Performance of the CMS Tracker Readout System," in *12th Workshop on Electronics for LHC Experiments*, pp. 187–190. 2006.

[73] J. Lamb, "CMS Silicon Strip Tracker Integration and Slice Test," *Nuclear Physics B - Proceedings Supplements* **177-178** (2008) p300–301.

[74] K. Hahn *et al.*, "Studies of the CMS Tracker at High Trigger Rate," CMS-NOTE-2009-015.

[75] J. Varela, "ECAL Off-detector Upgrade at SLHC." http://agenda.hep.wisc.edu/conferenceDisplay.py?confId=31.

[76] The CMS Collaboration, "CMS High Level Trigger," CERN-LHCC/2007-021.

[77] S. Baffioni *et al.*, "Electron Reconstruction in CMS," CMS-NOTE-2006-040.

[78] K. Lassila-Perini, "Jet Rejection with Matching ECAL Clusters to Pixel Hits," CMS-NOTE-2001-021.

[79] S. Gennai *et al.*, "Tau Reconstruction and Identification with Particle-Flow Techniques using the CMS Detector at LHC," CMS AN-2008/043.

[80] K. A. Petridis-Torres, "Reconstruction and Selection of $Z \to \tau\tau \to e + \tau$-jet Decays at the Compact Muon Solenoid experiment,". PhD thesis, London Univ., 2009.

[81] F. Vasey, "Versatile Link Status." http://indico.cern.ch/conferenceOtherViews.py?view=standard&confId=47853.

[82] J. Jones, C. Foudas, *et al.*, "A Pixel Detector for Level-1 Triggering at SLHC," physics/0510228.

[83] J. Jones, C. Foudas, *et al.*, "A Study for a Tracking Trigger at First Level for CMS at SLHC," physics/0510227.

[84] J. Jones, C. Foudas, *et al.*, "Stacked Tracking for CMS at Super-LHC," in *12th Workshop on Electronics for LHC Experiments and Future Experiments*, pp. 130–134. 2006.

[85] S. Das, "The Status and Performance of the CMS Pixel Detector," CMS-CR-2009-058.

[86] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 Physics Manual," hep-ph/0603175.

[87] "G4–A Simulation Toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) no. 3, p250–303.

[88] I. Tomalin, "Expected Data Rates from the Silicon Strip Tracker," CMS-NOTE-2002-047.

[89] R. Field *et al.*, "Pythia Tune A, Herwig and Jimmy in Run 2 at CDF," hep-ph/0510198.

[90] The CMS Collaboration, "Measurement of the Underlying Event in Jet Topologies using Charged Particle and Momentum Densities," CMS-PAS-QCD-07-003.

[91] P. Moreira and A. Marchioro, "The GBT: A Proposed Architecure for Multi-Gb/s Data Transmission in High Energy Physics," in *Topical Workshop on Electronics for Particle Physics*, pp. 332–336. 2007.

[92] P. Moreira, "GBT Project Status." http://indico.cern.ch/conferenceOtherViews.py?view=standard&confId=47853.

[93] A. W. Rose, "The Level-1 Trigger of the CMS Experiment at the LHC and the Super-LHC,". PhD thesis, London Univ., 2009.