Chapter 8

Luminosity Measurement

8.1 Introduction and overview

8.1.1 Relevant definitions

Instantaneous luminosity is defined in the usual way as

$$R = \mathcal{L}\sigma \tag{8.1}$$

where R is the observed event rate for a process with cross section σ. Integrated luminosity will be denoted by L. We define *interactions* to be inelastic proton-proton collisions (hard-core and diffractive scatterings) leading to significant energy depositions in the main parts of the CMS detector. When running at design luminosity, there will in be (≈ 25) interactions for each bunch crossing (BX).

8.1.2 Goals and requirements

The luminosity measurement is used to monitor the LHC's performance in real time and to provide an overall normalization for physics analyses. The design goal for the real time measurement is to determine the average luminosity with a 1% statistical accuracy in 0.1 s. For offline analyses, the design goal is a systematic accuracy of 5%, although every reasonable effort will be made to produce a more accurate result. Both of these requirements must be met over a very large range of luminosities, extending from roughly $10^{28}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ to $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and possibly beyond.

In addition to providing average luminosity measurements in real time and integrated luminosity values for offline analyses, the luminosity system will produce bunch-by-bunch luminosities useful for accelerator diagnostics and potentially also for accurate modeling of underlying event backgrounds. [1]

Other important and desirable features of the luminosity system include a capability for "always on" operation and a bookkeeping system that is robust and easy to use. Always-on operation means that luminosity information should be available for real-time monitoring of the LHC, whether or not the main CMS DAQ is operational. The bookkeeping system should be such that corrections to the luminosity information can readily be made (for example if more accurate values are obtained through offline analysis). Moreover, users carrying out physics analyses should be able to determine the luminosity corresponding to their samples in a straightforward way.

[1] If the bunch-to-bunch variations in the luminosity are large, the underlying event backgrounds are no longer accurately described by simple Poisson statistics.

8.1.3 General strategy

The normalization of physics analyses will ultimately depend on careful measurements of known cross sections such as the pp total cross section or the production rates for Ws and Zs.

Since the precision pp total cross section measurement can only be done at $\mathcal{L} = 10^{28}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ (see Section 8.4.1 below) and requires a special configuration of the machine optics, it is not suitable for real-time monitoring. Measurement of the production rate for vector bosons can be done at all luminosities, but requires that all elements of the detector be operating and well understood. Moreover, the statistical accuracy of this method will fall well short of what is required to meet the 1% in 0.1 s goal.

The real time monitoring function must therefore be based on the measurement of high cross sections using comparatively simple hardware. These measurements will not only provide real time information during detector operations, but will also be used to "interpolate" between precision measurements to provide statistically precise luminosity information for small samples.

Multiple techniques capable of providing suitable luminosity information in real time have been identified. One technique employs signals from the forward hadron calorimeter (HF) while another, called the Pixel Luminosity Telescope (PLT), uses a set of purpose-built particle tracking telescopes based on single-crystal diamond pixel detectors. At this writing, the PLT has not been formally approved, but is under study. Yet another, which is being pursued by the LHC machine group, employs a Fast Ionization Calorimeters (FICs) positioned at ± 140 m from the IP, at the point of transition between a single and a combined beam pipe. The FICs measure the flux of 0° $\approx$TeV-scale neutrons that are produced in pp collisions.

The HF and PLT techniques, which will be an important focus of this chapter, provide complementary information, which will afford a redundant measure of the luminosity. Since it is based on simple counting, the PLT is expected to provided excellent linearity over a wide range of luminosities. It is, however, potentially subject to saturation at the highest luminosities. The HF presents more of a challenge in terms of linearity, but should continue to provide meaningful information at luminosities well beyond the LHC design luminosity.

In addition to these primary systems, we envisage a number of other luminosity measurements that can be derived from CMS components whose main function is something other than determining the luminosity. For example, the global calorimeter trigger may be capable of providing real-time luminosity information. Similarly, the main pixel detector can be used offline to count primary interaction vertices, which should provide a redundant cross check on other methods.

8.2 LHC beam parameters

The bunch crossing rate at the LHC is 40.08 MHz. An orbit of the machine comprises 3564 bunches, 2808 of which have real collisions when running at design luminosity. The time for a single orbit is 89.92 μs. An important secondary function of the luminosity system is to measure the bunch-by-bunch luminosity and to confirm that collisions are restricted to the proper bunches. Assuming an inelastic and diffractive interaction cross section of $\sigma = 80$ mb[2], and an effective bunch crossing rate of $f_{\mathrm{BX}} = (2808/3564) \times 40$ MHz, the

[2]If one includes elastic scattering, the total cross section is expected to be about 110 mb.

number of interactions per bunch crossing will be

$$\mu = \frac{\sigma \mathcal{L}}{f_{\mathrm{BX}}} \simeq 25 \tag{8.2}$$

at $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Techniques for measuring the luminosity must therefore be able to contend with many overlapping interactions.

8.3 Online techniques

8.3.1 HF

8.3.1.1 Description

The forward hadron calorimeter, or HF, covers the pseudorapidity $3 < |\eta| < 5$ and is described in more detail in Chapter 5. It consists of quartz fibres embedded in a steel matrix. The signals from the HF result from Cherenkov light emitted in the fibres in response to charged particles. The fibres are viewed by phototubes situated on the downstream face. Since most of the visible energy is carried by relativistic particles—i.e., electrons—the calorimeter is most sensitive to the electromagnetic component of the hadronic showers. Thus the measured transverse shower size is determined by the Moliere radius rather than the nuclear interaction length.

Each HF endcap is divided into 36 segments in azimuth and 12 segments in η—i.e., each *physical tower* subtends an angular region of $\Delta\eta \times \Delta\phi \approx 0.175 \times 0.175$. In addition, crude longitudinal segmentation is achieved through the use of long fibres that run from the front face of the HF to the phototube readout at the back end and short fibres that cover only the rear part of the modules. The HF-based luminosity measurement to be described is based solely on the long fibres. For incident electrons, the light yield is 0.25 p.e./GeV.

Signals from the HF PMTs are digitized on a bunch-by-bunch basis using QIE chips, which employ a nonlinear scale to achieve a large dynamic range. The QIE outputs are routed to a set of 18 HCAL Trigger and Readout (HTR) boards, each of which services 48 HF physical towers.

8.3.1.2 Technique

8.3.1.2.1 Concept Two methods for extracting a real-time luminosity signal from the HF have been studied. The first is based on "zero counting." The number of interactions, n, in a given bunch crossing is distributed according to Poisson statistics

$$p(n;\mu) = \mu^n \frac{e^{-\mu}}{n!} \tag{8.3}$$

where μ is the mean number of interactions, given by Eq. 8.2 above. At very low luminosities μ is much less than 1 and is approximately equal to the fraction of bunch crossings that contain interactions. Determining this fraction is relatively straightforward using the HF, since nearly all interactions produce summed E_{T} signals well above the noise—see Fig. 8.1.

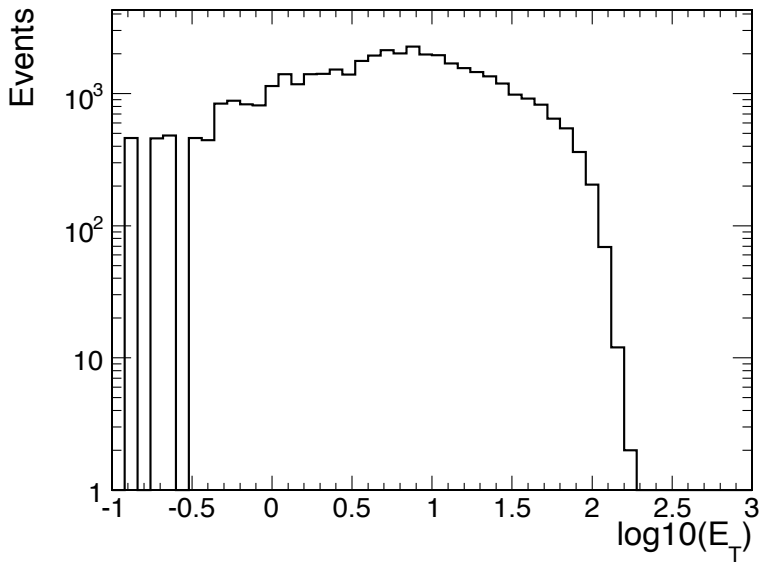

Figure 8.1: E_{T} spectrum of single interactions in the HF. Most events lie in the region $E_{\mathrm{T}} > 1$ GeV.

Even at a luminosity that is only 10% of design, the mean number of interactions per bunch crossing is ≈ 2.5, meaning that $\approx 70\%$ of bunch crossings have 2 or more interactions. Since one cannot reliably distinguish between single and multiple interactions, it is not possible to count interactions in a straightforward way. Fig. 8.1 does, however, show that in a large fraction of cases, one can distinguish between zero and one or more interactions. The mean number of interactions can thus be determined by inverting Eq. 8.3 to obtain

$$\mu = -\ln p(0), \tag{8.4}$$

which is known as "zero counting."

In principle, Eq. 8.4 can be used to infer μ at arbitrarily large luminosities, but for $\mu \gg 1$, the fraction of zeroes becomes very small, a condition here referred to as "zero starvation." We anticipate that well before a zero-counting method succumbs to a loss of statistical power associated with zero starvation, its systematic uncertainties will become unmanagable. We have thus adopted the rule of thumb that the fraction of zeroes should be at least 1%, corresponding to an upper limit on μ of 4.6. Thus, the total E_{T} signal from the HF will not provide a useful luminosity signal at $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. If, however, a *single* $\Delta\phi \times \Delta\eta = 0.175 \times 0.175$ physical tower is used, even at design luminosity the mean number of hits above threshold ($E_{\mathrm{T}} \approx 0.1$ GeV) is roughly $\mu \approx 1$, and zero counting is a viable method. Since the occupancies of all HF physical towers are roughly the same, the statistical power of the technique can be increased by averaging the results from all such towers. In effect, one makes 864 (quasi) independent measurements each bunch crossing.

A second method exploits the linear relationship between the total E_{T} deposited in the HF and the number of interactions and thus the luminosity. Since the HF is very far forward, the maximum E_{T} is kinematically limited to a few hundred GeV, which explains the sharp upper edge in E_{T} visible in Fig. 8.1. This suppresses the large statistical fluctuations that can accompany unbounded power-law distributions.

8.3.1.2.2 Hardware implementation Since the HF and its readout is already part of the CMS baseline design, only a small amount of extra hardware is needed to derive a luminosity signal from it. In particular, a "mezzanine board," called the HF Luminosity Transmitter (HLX) card mounted in each of the 18 HTR boards is used to collect and reduce the HF QIE data for transmission via 100 Mbps ethernet to a router, which further concentrates the data to a single stream that is passed to a dedicated luminosity computer. This is shown in block diagram form in Fig. 8.2.

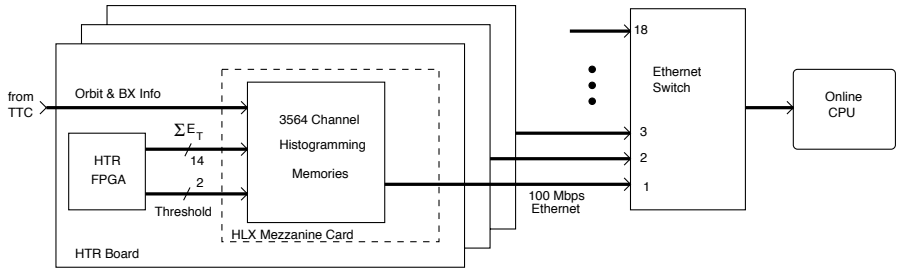

Figure 8.2: Block diagram of the HF luminosity readout.

The HLX boards have the same footprint and form factor as the Synchronization and Link Boards (SLBs) that provide an interface between the ECAL and HCAL readouts and the Regional Calorimeter Trigger system. Each HLX site is connected to the 2 main HTR FPGAs by 36 lines per FPGA. These lines can be driven at up to an 80 MHz cycle rate. At 80 MHz, 1 line must be reserved for framing, effectively leaving 70 bits of data available to describe the state of the 24 HF channels.

Two bits of status information encode the following four possible states for each tower: disabled, enabled-below-threshold, over-threshold-1, over-threshold-2. Inside the HLX, this input will be used to create 4 histograms. Each histogram will have 3564 bins, one for each bunch in the orbit. Each bin will be 2 bytes, so the total data size for a set of histograms is $28\,512$ bytes. An additional 14 bits will be allocated to a running E_{T} sum, which will be used to implement the second method described above.

The baseline design is to add the results from all 48 channels into a single set of histograms (effectively summing the number of zeroes across all channels in a module) the histograms would take at least $2^{16} \div 48 = 1365$, orbits or about 120 ms to overflow (in the worst case, 48 zeroes are accumulated during every bunch crossing). Leaving a safety margin, we will transmit the data to the router at 10 Hz.

8.3.1.3 Expected performance

The zero-counting method has been simulated using a sample of PYTHIA [45] minimum bias interactions. The response of the HF was simulated using OSCAR (Section 2.5), which is based on GEANT4 [30]. The output of OSCAR was analysed using a simple ORCA module that read hit data from the OSCAR output and produced an ntuple that was analyzed using a ROOT script.

Bunch crossings for a given luminosity are simulated by drawing a number of interactions according to Eq. (8.3). The energy depositions (expressed in photoelectron units) for the

next n interactions are then overlaid in an array where each element corresponds to 1 of the 864 physical towers in the HF. These "energy sums" are then smeared to take into account the statistics at the first dynode, scaled by the PMT gain, smeared again by the electronic noise at the QIE input, and quantized according to the QIE ADC calibration. In a final step, the energy values thus obtained are converted to E_{T} values. The $\sin\theta$ values used in this conversion range from 0.016 to 0.10 for $|\eta| = 5$ and $|\eta| = 3$, respectively.

The sum for each physical tower is then compared to a threshold of 6 ADC counts, which corresponds to 2 photoelectrons, or about 8 GeV of incident energy. This translates to an E_{T} ranging from 0.13 GeV for the innermost ring of the HF that is used to $E_{\mathrm{T}} = 0.75$ GeV for the outermost ring.

The zero fraction was computed using

$$p(0) = \frac{\#\mathrm{zeroes}}{\#\text{ of opportunities}} = \frac{\sum \#\text{ of towers with } E_{\mathrm{T}} < E_{\mathrm{thresh}}}{792 \times N_{\mathrm{BX}}} \tag{8.5}$$

where $2 \times 11 \times 36 = 792$ is the number of towers in the forward and backward HFs and the sum extends over all towers and all bunch crossing, except for the innermost and outmost rings. The value of $p(n)$ thus obtained is then converted to a mean tower occupancy (μ) using Eq. (8.4).

In the ideal case, the resulting value of μ should be proportional to the luminosity over a wide range. Fig. 8.3 shows that this is indeed the case. The upper panel shows the observed value of μ plotted as a function of input luminosity. The straight line plotted with the points is not a fit, but rather is chosen to go through the origin and the observed value of μ at $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In the lower panel, the same data are plotted, but in this case the points represent the fractional deviation from the average response.

A slight sub-linearity associated with zero starvation is observed at the highest luminosities. The seemingly simple expedient of raising the E_{T} threshold results in more zeroes, but introduces a superlinearity of different origin. In particular, for thresholds that are too high, the probability of 2 low-energy depositions conspiring to exceed the threshold competes with that for a single large deposition. Since the former mechanism scales with the luminosity squared, a nonlinearity results.

Additional runs of the Monte Carlo show that acceptable linearity is obtained with thresholds between 3 and 8 ADC counts. Although this is a workable range, and it is in principle possible to correct for nonlinearities, any such correction would be Monte Carlo dependent. It is thus desirable to employ alternative algorithms as a cross check. To that end, we have studied the use of a simple linear sum over scalar E_{T}.

Fourteen-bit E_{T} values from the HTR boards with a least count of 0.0625 GeV[3] are summed on a bunch-by-bunch basis. Since the scalar E_{T} deposited in the HF is proportional to the number of interactions, the average E_{T} per bunch is proportional to the luminosity. Fig. 8.4 shows that good linearity is achieved over the full range of luminosities.

8.3.2 Pixel Luminosity Telescope

[3]This least count value, which is 4 times smaller than that used elsewhere in the trigger, is required to avoid nonlinearities in the $\langle E_{\mathrm{T}}\rangle/\mathrm{BX}$ vs. luminosity function.

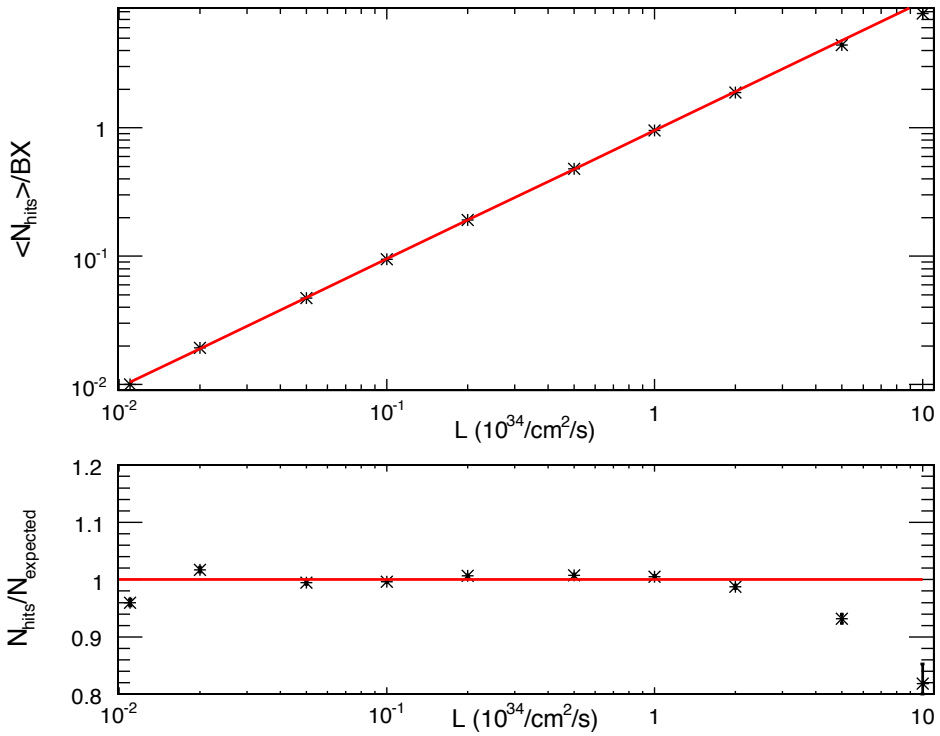

Figure 8.3: Linearity test of HF zero-counting technique. The upper panel shows the mean physical-tower occupancy inferred from counting the number of zeroes vs. the luminosity expressed in units of the LHC design luminosity of $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The lower panel shows the same data plotted relative to what is expected for a linear response. See text for additional details.

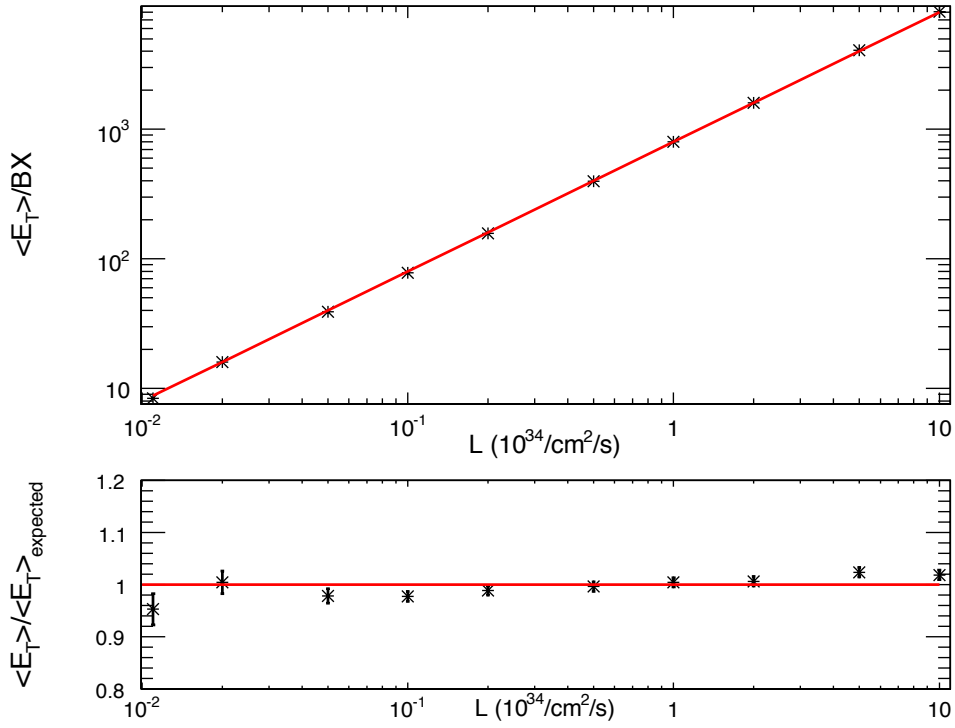

Figure 8.4: Linearity test of the HF E_{T} sum technique. The upper panel shows the average E_{T} per bunch crossing vs. the luminosity expressed in units of the LHC design luminosity of $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The lower panel shows the same data plotted relative to what is expected for a linear response. See text for additional details.

8.3.2.1 Description

The Pixel Luminosity Telescope (PLT) is a proposed, dedicated device for luminosity measurement that is currently under consideration by CMS. The PLT is designed to make fast, stable and precise determinations of the bunch-by-bunch luminosity and of the location of the interaction point centroid. It is comprised of arrays of small-angle telescopes located on either side of the interaction region. Each telescope consists of 3 planes of single crystal diamond sensors bump-bonded to a CMS pixel readout chip. For each bunch crossing, the number of particles traversing the telescope arrays is determined by forming a coincidence of the 3 planes in each telescope from fast out signals provided by the pixel readout chip.

The PLT has an array of 8 telescopes uniformly distributed in ϕ on each side of the interaction region. Each telescope consists of 3 equally-spaced sensor planes. The sensitive area of each plane, 8×8 mm^2, is determined by the active area of the pixel readout chip. In order to maximize the solid angle acceptance, the telescopes are placed as close to the interaction region as possible. They will be located in z between 1.65 m and 1.85 m with 10 cm spacing between sensor planes. Radially, they with be located 10 mm from the beam pipe with the sensor planes at $z = 1.65$ m located between $r = 41$ mm and $r = 49$ mm. The telescopes will be projective at an angle of 1.56° to the interaction point corresponding to a rapidity of $\eta = 4.3$. Figure 8.5 shows the location of a PLT array. The minimum z location is determined by the beam pipe support rods and collar at $z = 1.6$ m. The PLT will be located outside of this so that the telescopes can be installed and serviced without the necessity of removing the beampipe support.

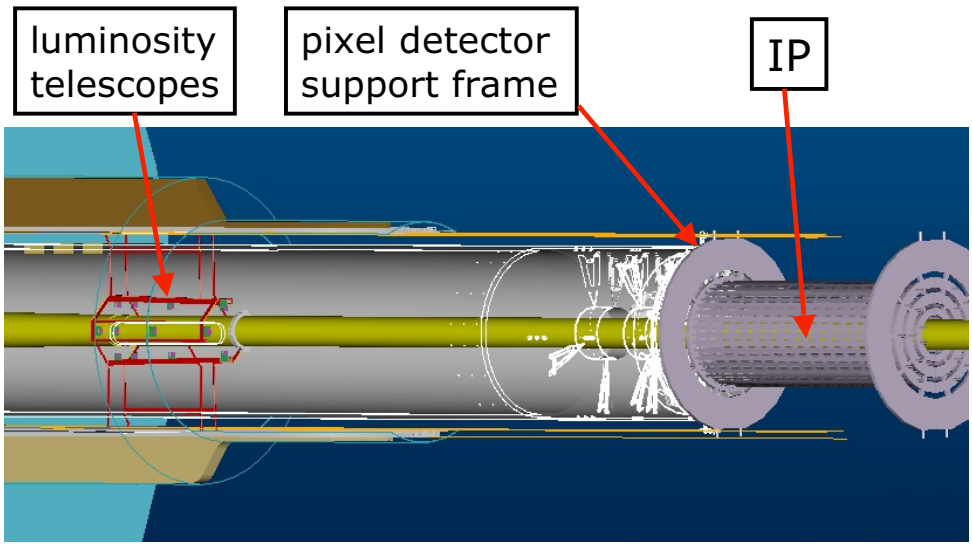

Figure 8.5: Location of a PLT array within CMS.

Each telescope plane consists of a sensor configured with a pixel pattern electrode and bump bonded to the CMS pixel readout chip. The reason for using pixels is to reduce the capacitance of each channel and, thereby, the inherent serial noise that otherwise would make readout of the sensor in the bunch crossing time of 25 ns prohibitive. With a pixellated sensor the per channel capacitance is reduced by several orders of magnitude and such that the

signal-to-noise ratio will be greater than 20:1 for 25 ns readout. The CMS readout chip [225] provides a fast out signal for each bunch crossing with an output level corresponding to the number of double columns with hit pixels. These signals, 1 for for each telescope plane, will be sent over standard CMS optical fibre analogue links to the electronics room where a three-fold coincidence for each telescope will be determined.

Counting of the fast three-fold coincidence signals constitutes the primary measurement of the bunch-by-bunch luminosity. In addition, since the sensors will be bonded to pixel readout chips, the full pixel information with column and row addresses and pulse height of the hit pixels can also be obtained. The full pixel readout of a group of several chips requires 1 to 2 μs and cannot be obtained for every bunch crossing but can be readout for every CMS L1A trigger, approximately a 30 kHz rate. This less frequent but more detailed information will allow full examination of the telescope hit patterns. Any anomalies seen in the fast coincidence signal data can be investigated and, with tracking information, the location of any "hot" sources of particles, e.g., beam halo impacting the beam pipe, can be determined. Furthermore, this tracking capability will allow the interaction point centroid to be determined. Since the entire readout and control chain from detector to the VME flash ADC module will be essentially identical to that for the CMS pixel detector, this capability can be obtained with modest additional cost or effort.

Single crystal Chemical Vapor Deposited (CVD) diamond [226] is used for the sensor material because of its superiority to silicon in terms of radiation hardness and its lack of need for cooling. Since they are located at a small angle, the charged particle fluence on the telescopes will be large, a few $\times 10^{14}$ cm^{-2} per year at full design luminosity, comparable to that of the inner layer of the barrel pixel detector. Measurements of CVD diamond[227, 228] indicate that the diamond signal amplitude decreases by at most 20% to 30% with no increase in leakage current at fluences of 2×10^{15} cm^{-2} equivalent to several years at full LHC luminosity. Another decisive advantage of diamond over silicon is the lack of a need to cool the diamond sensors. Unlike silicon, that must be cooled to -5° to -10° C in order to maintain operation at high radiation, diamond performs well and maintains negligible leakage current at room temperature and above.

Single crystal CVD diamond has a narrow Landau pulse height distribution that is well separated from zero as seen in figure 8.6 which shows the signal pulse height distribution measured for ^{90}Sr β particles incident on a single crystal diamond sensor of 480 μm thickness. The average pulse height for this diamond corresponds to a collected signal of $18\,100$ electrons, well above the noise level for pixel electronics. Because of the large separation from zero, the single crystal diamond sensors will have an essentially 100% detection efficiency. This is important in order to ensure that the efficiency is well determined and is stable as required for a high precision luminosity measurement. For the PLT, the diamond sensor thickness will be 400 μm yielding an average signal of $15\,100$ electrons for a normally incident minimum ionizing particle.

8.3.2.2 Expected performance

The primary function of the PLT is to provide a prompt and precise determination of the relative bunch-by-bunch luminosity. In addition to providing a real time measurement of the relative luminosity, this measurement when calibrated with an absolute measurement of the luminosity as obtained by TOTEM or by W and Z production will provide a high precision measurement of the absolute luminosity. The PLT also provides a real time determination

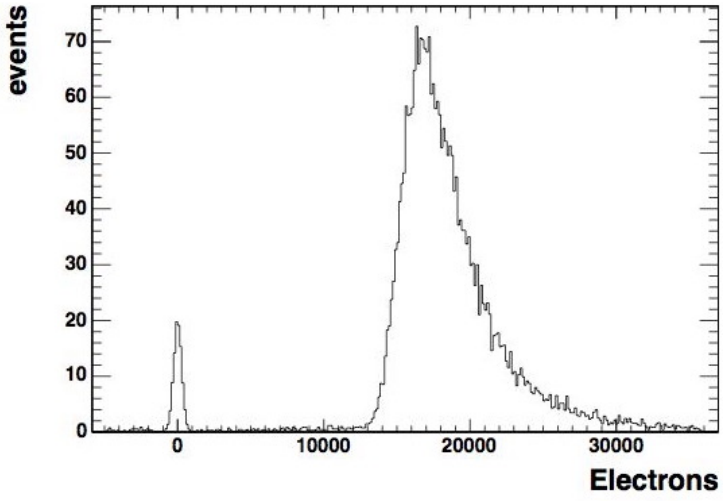

Figure 8.6: Response of a 480 μm thick single crystal CVD diamond to ^{90}Sr β's. The pedestals events are due to particles that fired the trigger scintillator but that missed the diamond sensor.

of the relative location of the interaction point centroid. It will also be able to detect and quantify the particle fluxes due to beam halo and hot spots along the beam pipe. Simulations of the PLT detector have been made to determine: the rates of particle tracks expected in the telescopes, the backgrounds expected from particle interactions in the beam pipe and other CMS components, the uniformity of acceptance around the interaction point region, the resolution achievable on the interaction point centroid from particle track extrapolation and the ability to detect beam pipe hot spots and beam halo.

The particle rates from minimum bias interactions in the PLT were simulated using PYTHIA (version 6.227). With the CMS magnetic field on but without the beam pipe or other CMS detector components, the number of three-fold coincidences per telescope per bunch crossing is 0.043 at a luminosity of 10^{33} cm^{-2}s^{-1}, 10% of design luminosity. These coincidences arise from tracks originating from the pp interaction. With a total of 16 telescopes, the net particle detection rate will be 0.69 per bunch crossing at 10^{33} cm^{-2}s^{-1} luminosity. With this rate, the relative luminosity of each of the 2835 filled bunch crossings within an orbit can be determined to a precision of 1.1% every second (about 10^4 orbits). When the beam pipe and other CMS components are included in the simulation, the three-fold coincidence rate increases by about 10% to 0.048 per bunch crossing. This increase is due almost entirely to photons that interact in the beam pipe at around $z = 1.0$ m. These are photons from the decay of π^0's produced in the pp interactions that are directed at the solid angle of a telescope. When traversing the beam pipe at a 1.5° angle, these photons see about 3.0 cm of beryllium corresponding to 8.5% of a radiation length and some are converted. The increased three-fold coincidence is, thus, due to pp production and will therefore scale with the luminosity.

In order for the luminosity measurement to be precise and stable it is important for the acceptance of the telescope array to be uniform over the interaction region. The luminosity measurement will then be insensitive to any small movement of the interaction point centroid. The acceptance was determined by shifting the interaction point in r, and z and determining the rate of the three-fold coincidences observed at each location. These rates were

then normalized to the rate obtained wsith the interactiion point at $r = 0$, $z = 0$. Fig. 8.7 shows this normalized rate as a function of r and z. The z-projection at $r = 0$ is shown in Fig. 8.8. The acceptance is flat to 1% out to $\pm$ 320 mm. The r-projections at $z = 0$ are shown in Figs. 8.9 a) and b) for the cases in which the interaction point is offset in a radial direction toward 1 of the 8 telescopes and in a radial direction exactly between 2 of the telescopes, respectively. The acceptance is 1% flat out to $r = 12$ mm, in the former, and to $r = 8$ mm, in the latter.

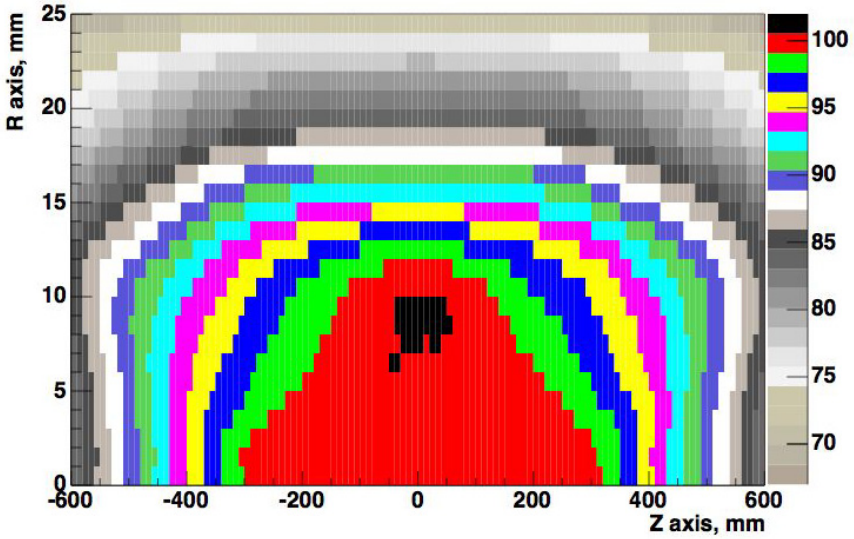

Figure 8.7: Track acceptance as a function of the position of the interaction point normalized to $r = 0$, $z = 0$.

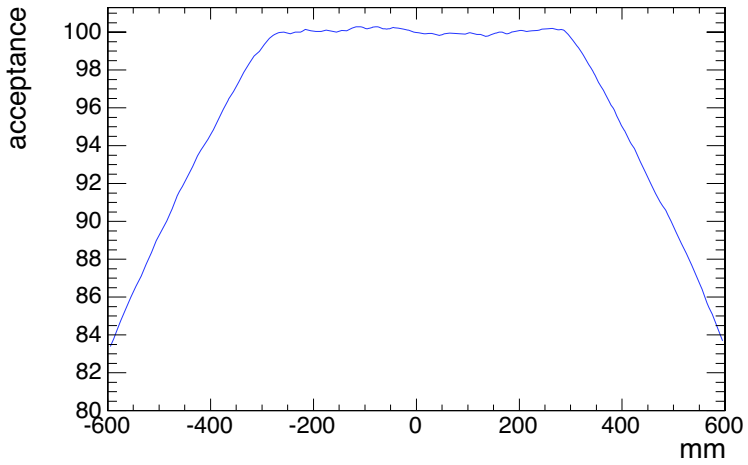

Figure 8.8: Track acceptance as a function of z at $r = 0$.

Another important function of the PLT is to determine the relative location of the interaction point centroid and to monitor any possible drifts on the time scale of a second. For this measurement, the addresses and pulse heights of those pixels above threshold, recorded for every CMS Level 1 trigger, will be used. Extrapolation of tracks to the interaction point will be affected by the bending of the particles in the 4 T magnetic field. For a given telescope, the

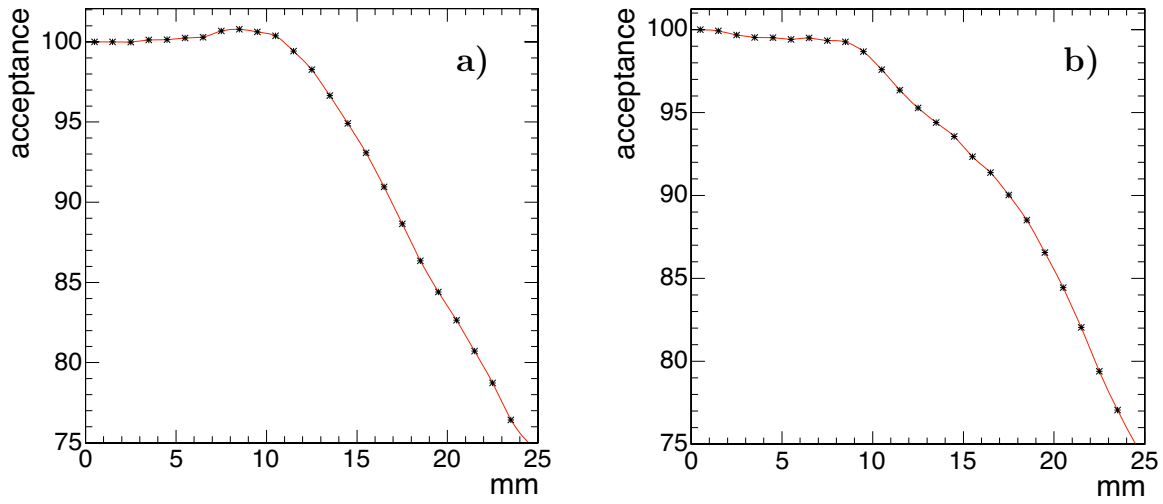

Figure 8.9: Track acceptance as a function of r at $z = 0$ for a) r in direction of 1 of the telescopes and b) r exactly between 2 of the telecopes.

effect of this curvature will be smallest in the plane determined by z and the radial direction of the telescope and will be an order of magnitude larger in the plane determined by z and the radial direction orthogonal to the telescope. For example, for the telescope at $\phi = 0$ the sagitta in the x-z plane is about 1 μm while the sagitta in the y-z plane is about 20 μm for a 10 GeV/c particle. Since the sensor planes will be essentially perpendicular to the particle tracks, there will be little charge sharing among pixels and the hit spatial resolution will be the digital resolution of 30 to 40 μm set by the pixel size of 100×150 μm ^{2}. Since even for low momentum, 10 GeV/c, tracks the sagitta is too small to be determined with the spatial resolution of the telescope, the tracks will be linearly extrapolated to the interaction point. Due to the variation of curvature with momentum and the longitudinal spread of the breams, the extrapolation point of the tracks will be smeared. Fig. 8.10 shows the extrapolation resolution in the radial direction. It has a fit standard deviation of $\sigma = 2.4$ mm. The longitudinal spread of the beam, $\sigma = 7.5$ cm, is the dominant contribution to the width of this distribution. For a 30 kHz trigger rate, more than 10^3 tracks will be recorded within 1 second at a luminosity of 10^{33} cm^{-2} s^{-1}. Four telescopes, 2 on each side of interaction region, will be used in the measurement of the x (y) interaction point location. This will allow determination of the transverse location of the interaction centroid to 35 μm in both x and y.

Because of the shallow track angle, approximately 1.5°, the RMS of the extrapolated distribution in z is 8.8 cm, 40 times greater than for the transverse distribution and comparable to the spread due to the longitudinal bunch length RMS of 7.5 cm. Using all 16 telescopes, the z interaction point centroid can be located to a precision of 720 μm within 1 second again at a luminosity of 10^{33} cm^{-2} s^{-1}.

For beam diagnostic purposes, the PLT has been designed to be sensitive to beam halo. With a 20 cm total length for the telescope, there is a 32% overlap of the first and last telescope planes for horizontal tracks making the telescopes sensitive to beam halo tracks. Their fraction can be determined by measuring the fraction of horizontal tracks seen in the full pixel data. These beam halo tracks will contribute to the three-fold fast coincidence measurement but once their fraction is determined in the pixel readout data they can be subtracted. The

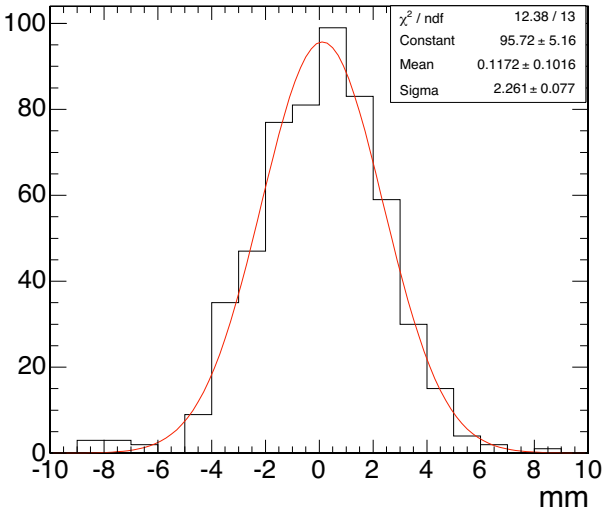

Figure 8.10: Transverse distribution of tracks linearly extrapolated to $z = 0$.

pixel readout chip has the capability of allowing individual pixels to be masked removing them from both the pixel readout and from the fast out signal. If it is discovered that the fraction of beam halo tracks is large and compromises the luminosity measurement, the acceptance for horizontal tracks can be reduced by masking out appropriate pixels.

In summary, on a time scale of 1 second at a luminosity of 10^{33} cm^{-2} s^{-1}, 10% of design, the PLT will perform the following.

- Determine the bunch-by-bunch luminosity to a precision of 1.1%.
- Determine the transverse interaction point centroid to a precision of 35 μm.
- Determine the longitudinal interaction point centroid to a precision of 720 μm.
- Locate possible beam pipe hot spots within ± 1.5 m of the interaction point.
- Monitor the amount of beam halo.

8.4 Offline and normalization techniques

8.4.1 TOTEM

The TOTEM experiment [198, 211], further described in Chapter 7, will measure the total pp cross section and study elastic and diffractive processes at the LHC. Elements of TOTEM will be situated in the far forward regions of CMS. Of particular relevance to the CMS luminosity determination will be a precision ($\approx 1\%$) measurement of the total pp cross section using the so-called luminosity independent method. This approach uses the optical theorem, which holds that the total cross section can be related to the elastic cross section extrapolated to $t = 0$ via the relation

$$\sigma_{\text{tot}} = \frac{16\pi}{(1+\rho^2)} \frac{(\mathrm{d}N_{\text{el}}/dt)_{t=0}}{N_{\text{el}} + N_{\text{inel}}}, \tag{8.6}$$

where ρ is the ratio of the real to the imaginary part of the forward scattering amplitude. The extrapolation of the elastic cross section to $t = 0$ will be made using roman pots placed far (150 200 m) from the IP.

The error on $\sigma_{\rm tot}$ is expected to be of order 1%, coming mainly from the uncertainty in the inelastic cross section measurement. Since CMS and TOTEM will operate simultaneously, it will be possible to arrive at an absolute normalization for the real time techniques described above. However, TOTEM will operate at low luminosity [4] with different machine optics, which will introduce an additional uncertainty in the calibration of the real-time methods at design luminosity.

8.4.2 W and Z rate measurements

Since the pp total cross section measurement described above can only be carried out at low luminosity, using it to normalize the luminosity at $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ will involve a sizeable extrapolation. Moreover, it will not be possible to detect and correct possible long-term drifts in the normalization constants for these systems. For these reasons, it will be useful to have a normalization technique that is based on production data taking.

The production rate for Ws and Zs provides just such a "standard candle." Moreover, the rates are high enough [5] datasets of only several minutes can be normalized with good statistical accuracy, even when realistic trigger and reconstruction efficiencies are taken into account.

The main challenge will be to control the uncertainties associated with theoretical estimates of the cross section and the modeling of the detector acceptance. Frixione and Mangano have argued that uncertainties associated with the interplay between detector acceptance and the parton density functions (PDFs) can be controlled at the few percent level [229]. Moreover, since PDFs are needed to relate parton-level cross sections to the cross sections actually observed in pp collisions, uncertainties in the PDFs represent an irreducible uncertainty in many comparisons between theory and experiment, even if the pp cross sections are perfectly measured. Indeed Dittmar, Pauss, and Zürcher have argued that rapidity distributions in W and Z production can be used to determine the x distributions and consequently the corresponding "parton luminosity," and that this raises the possibility of $\approx \pm 1\%$ accuracy in the determination of certain parton-parton cross sections[230].

8.5 Sources of systematic effects

Both HF luminosity measurements involve the raw rate for energy (or hit) depositions. There is little possibility to carry out event selection, since at design luminosity, each "event" already represents the superposition of several interactions. This leaves the HF techniques vulnerable to non-pp interaction backgrounds, notably "beam-halo" and beam-gas interactions.

Beam gas interactions are thought to be negligible, but this will depend on the vacuum conditions in the ring and will only be verified through special single-beam runs.

[4] For the total cross section measurement, the LHC will operate with 43 bunches at a luminosity of $10^{28}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$, which corresponds to a luminosity per bunch 15 000 times lower than design.

[5] At design luminosity, the raw rate for $Z \to \ell^+\ell^-$ is 30 Hz.

Beam halo in the HF was estimated using a model for beam halo[231]. The resulting background was found to be negligible (at the 10^{-5} level) using the default parameters provided in the model. Beam-halo estimates are, however, notoriously difficult and a quantitative estimate will require single-beam running.

Although the PLT is also subject to beam-gas and beam-halo backgrounds, by using the its fine-grain readout capability an estimate of these backgrounds should be possible even without single-beam operation of the LHC.

8.6 Luminosity monitoring, reporting, and logging

An important design requirement is that the luminosity system must be capable of providing luminosity information whether or not the main CMS DAQ is operational. To satisfy this requirement, an autonomous DAQ system will be used to acquire and analyze real time luminosity information from the HF (via the HLX boards) and the PLT. This system will serve luminosity information to various consumers—e.g., the LHC control room, a publicly availably web page, and the database system used to log luminosity information.

A carefully thought out and reliable luminosity database will be an essential element of the luminosity reporting system. To avoid unneccessary development costs and to facilitate long-term maintenance, we will use the same database planned for maintaining calibration constants in CMS. Preliminary discussions with the CMS database group indicate that the demands of the luminosity system will be modest.

Our current plans call for an approach similar to that employed at the Tevatron, where runs were broken into small segments of order 1 minute duration[6]. These run sections are short enough that the luminosity can be taken to be constant throughout. Moreover, trigger prescaler factors will be allowed to change only on run segment boundaries. To accommodate "on-the-fly" changes, run segments will comprise an integer number of LHC orbits and will start and end during the abort gap. Scaler and luminosity tallies from the real time systems will be written for each run segment.

8.6.1 Luminosity accounting

The cross section for a given process, j, is given by

$$\sigma_j = \frac{N_j^{\text{yield}}}{\epsilon_j L_0}. \tag{8.7}$$

where N_j^{yield} is the number of events selected after background subtraction, ϵ_j is the detection efficiency of the complete selection chain (trigger and offline), and L_0 is the integrated luminosity, assuming no dead time or prescaler losses. Multiplying and dividing by the number of Level 1 Accepts (L1As) for the relevant trigger channel, N_j^{L1A}, we obtain

$$\sigma_j = \frac{1}{\epsilon_j \mathcal{L}_0} \left(\frac{N_j^{\text{yield}}}{N_j^{\text{L1A}}} \right) N_j^{\text{L1A}} \equiv \frac{1}{\epsilon_j \mathcal{L}_0} \rho_j N_j^{\text{L1A}} \tag{8.8}$$

[6]CDF calls these segments "run sections," and DØ calls them "luminosity blocks."

where $\rho_j \equiv N_j^{\rm yield}/N_j^{\rm L1A}$ is the fraction of L1As that end up in the final event sample. Note that ρ_j can be estimated from *any unbiased sample* of L1As. Thus prescaler and dead time losses need not be explicitly taken into account, although it would certainly be advisable to do so as a redundant cross check. Note that since the rejection of events in the high level trigger (HLT) is not unbiased (there is by design a strong bias against background events), rejection factors in the HLT must be tracked and logged. Put another way, ρ_j can be factored into 2 terms—i.e., $\rho_j = \rho_j^{\rm HLT}\rho_j^{\rm offline}$—and it is only possible to determine the second term from the events recorded in mass storage.

Calculating the cross section for any process thus requires knowledge of L_0, which will be determined from: (i) the online monitors, calibrated using the pp total cross section and/or the yield of Ws and Zs; (ii) the raw number of L1As ($N_j^{\rm L1A}$), which will be determined by reading the scalers; [7] (iii) ρ_j determined as outline above; and (iv) ϵ_j the trigger and reconstruction efficiency, which will be determined via the usual Monte Carlo and data based studies.

The analysis above assumes that just 1 trigger provides acceptance for the process under study. That assumption simplifies things for purpose of illustration, but is not strictly necessary. The analysis can be extended to multiple triggers with overlapping acceptance.

[7] In the CMS trigger, the raw number of L1As will be determined from reading the prescale factor, and the number of counts at the prescaler output. There is potentially another small dead time associated with Level 1 trigger rules, which will be accounted for separately.