Stacked Tracking for CMS at Super-LHC

J. Jones, **A. Rose**, B. Constance, C. Foudas, G. Hall, M. Raymond, K. Zhu

Imperial College London, UK.
john.jones@imperial.ac.uk

Abstract

We report recent work on the design of a pixel detector for CMS at the Super-LHC. This work builds on previous studies of a tracking detector capable of providing track stubs to be used in the Level-1 Trigger (L1T). We now focus on the use of two 'superlayers' of tracking; each comprising a pair of pixel sensors with $50 \times 50 \times 50 \mu m^3$ pitch ($z \times \varphi \times r$) separated by a few millimetres. Preliminary work on track reconstruction in Field Programmable Gate Arrays (FPGAs) is also presented.

I. OVERVIEW

The current design of CMS is based on the nominal beam luminosity $10^{34} cm^{-2} s^{-1}$. It is anticipated that after running for several years, both LHC and the detectors will be upgraded to operate at a luminosity of $10^{35} cm^{-2} s^{-1}$ [1]. This presents a great challenge both in terms of radiation hardness and the increased data rates that will have to be sustained by the detectors and their corresponding DAQ systems.

The background to these issues and the stacked tracking principle were presented in LECC 2005 [2,3], and only a summary is stated here for completeness. Apart from jet vetoing by multiplicity, the simplest useful tracking contribution is a stub from two consecutive barrel layers. The stub can be used in coincidence with the outer detector to indicate whether the hit in the outer detector was caused by a high-p_T particle. The quality of the stub (i.e. whether the hits are matched correctly between the two layers) is dependent on the layer separation; for layer separations of greater than a centimetre (see figure 1), tracks from different events will overlap, producing a large number of track combinatorials during reconstruction.
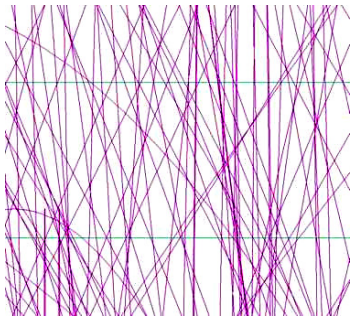

Figure 1: Track overlap in y-z plane (detector coordinates – see Fig. 3). 1cm layer separation is denoted by the two cyan lines. Note the significant overlap of tracks between these two layers, which will hinder tracker reconstruction.

An alternative approach to the combinatorial problem involves bringing two pixel layers together (hereafter refered to as a stack) so that they are separated by a few millimetres. The combinatorials then become manageable; even the limited knowledge of the interaction point is sufficient to make a 1:1 match between many of the hits in the two layers. This enables fast reconstruction using simple binning techniques, which could be implemented in an FPGA off-detector or a radiation-hard ASIC on-detector.

The basic layout of a stacked pixel detector is shown in figure 2. In order to ensure that the signal efficiency is high, the design is hermetic using overlapping segments of detector.
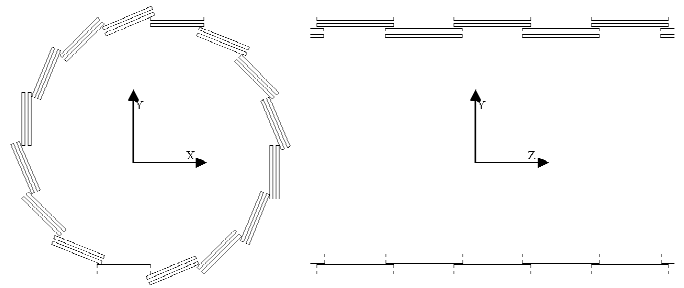

Figure 2: Basic layout of a stacked tracker (not to scale). Left is an x-y view, right is a y-z view.

The single stack approach, while useful for reducing the on-detector data rate, results in several complications. Firstly it increases the material budget in the inner detector; while this can be mitigated with modern materials [4], this is a trade-off that will have to be considered in the design of any new pixel system.

The second issue is fundamental to the stack design. As stated in [2], the ability to cut on transverse momentum by difference analysis of column pixels comes at the price of a lack of ability to actually measure p_T. This follows from the close proximity of the stacks, reducing the lever arm to such and extent that neither transverse momentum nor charge are measurable.

Tied to the ability to measure p_T is the ability to correctly project a track onto a calorimeter trigger tower [5]. While this was previously shown to be possible for particles with transverse momentum greater than approximately 20GeV, the lower momentum particles that are passed through the correlator ASIC are indistinguishable from the higher momentum ones. This results in a potentially serious inefficiency and also results in a high rate of ghost states.

A. The Double Stack Method

The latter issues can be either resolved or at least improved upon by the use of more than one stacked detector. The principal benefit of a stack is the massive reduction of data-rate before leaving the detector. By using two sets of stacked sensors or 'superlayers', one can still benefit from the rate reduction in each individual super-layer by using a geometrical p_T cut, but can also reconstruct stubs in a similar way to a more traditional pixel detector .
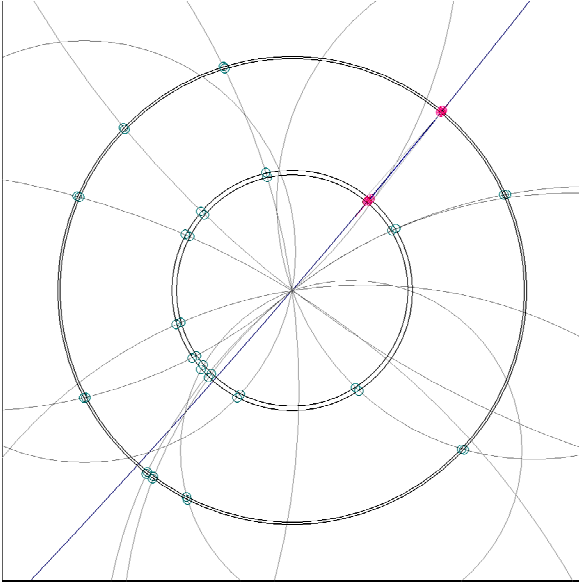

Figure 3: A double stacked detector. One of the layers is placed at a radial distance of 10cm from the beam pipe, and the second super-layer is placed 20cm from the beam pipe.

An example of this detector configuration is shown in Figure 3. For reasons that will later become apparent, the pixel pitch has been relaxed to $50\times50\times50\mu m^3$. While a finer pitch is preferable it is no longer strictly necessary, and larger pixels are easier to design. The inner super-layer is placed at r=10cm with a layer separation of 4mm, while the outermost layer is located at r=20cm with a layer separation of 2mm. The difference in layer separation compensates for the different radii, making the p_T cuts similar for each superlayer (approximately 3GeV).

The fundamental benefit of this design over those previously proposed is that it requires no on-detector communication between the superlayers. Interlayer communication is a crippling limitation of any design due to the limited space available for services and the additional power consumption of interconnections between widely-spaced layers.

B. Double Stack Reconstruction

The reconstruction method for a double stack configuration is similar to that for a single stack (Figure 4). One significant difference is that the performance becomes more significantly affected by the z pitch of the pixel than by the φ pitch. The reason for this is that the z-coordinate of the tracker hits is unaffected by the magnetic field from the solenoid and so the track follows a straight-line path in the r-z

coordinate system. On the other hand the B field reduces the resolution of the track in the r-φ plane and therefore the search window for matching stubs between the superlayers becomes too large. While it is still used, the z-reconstruction becomes dominant and as a result, the reconstruction purity is determined by the intrinsic resolution of the sensor.
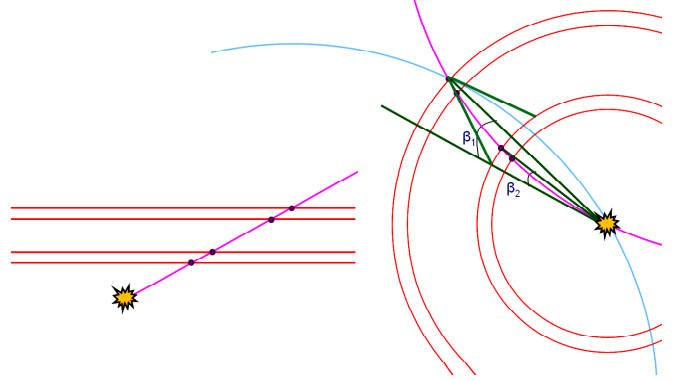

Figure 4: Double stack reconstruction. The track reconstruction in the r-z view is more accurate due to the parallel B field in that coordinate system. The back-searching method is still used in the r-φ plane as an aid, but is insufficient to pair hits on its own.

An additional benefit of this reconstruction is that it gives the precise location of the primary vertex for the event. It should be noted however that there will be an associated inefficiency due to the incorrect management of any secondary vertices, the detection of which is considered too difficult to implement in this design.

The estimated performance was calculated using a Monte Carlo simulation derived from the one used in [2].

C. Double Stack p_T Measurement and Track Projection

The projection of a reconstructed track in the r-z plane follows directly from the reconstructed track stub. However, to perform r-φ reconstruction requires the calculation of the transverse momentum. As the inner layer of each stack is very close to the outer layer, the transverse momentum can be calculated to the nearest approximation by using just one hit from each super-layer. The current implementation takes this approach although using both pixel coordinates would provide a small additional benefit, either by the use of linear interpolation or by more complex algorithms that weight the pixels optimally. Only the simplest case is considered here as the algorithm must operate efficiently in hardware.

As there are only two superlayers in this design, the beam spot must be used as an additional constraint. In the simulations described later in this paper, the following 3-point reconstruction equation is then used:

$$p_T = \frac{r_{out}^2 Bc}{8r_{in}\sin(\Delta\varphi)}$$

Where r_{out} is the radius of the outer super-layer, r_{in} is the radius of the inner super-layer, B is the magnetic field strength in Tesla, c is the speed of light, p_T is measured in eV and $\Delta\varphi$ is the angular separation between the hits in the two superlayers. This is an approximation relying on the layers being equidistant and can be optimised.

Once this value has been calculated the track can be projected onto an ECAL trigger tower for matching with detected hits, or forwarded to the muon system for matching with tracklets built using information from those detectors.

In simulation one can calculate the difference between the impact location of the reconstructed track and the 'true' track. Figure 5 shows the momentum resolution. $\Delta\varphi$ is calculated as:

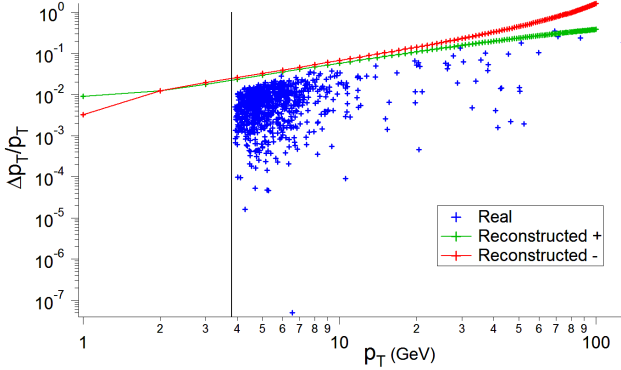$$\Delta\varphi = \frac{p_T^{reco} - p_T^{real}}{p_T^{real}}$$

Figure 5: Resolution of reconstructed transverse momentum plotted as a function of real transverse momentum. The black line represents the reconstruction cut, the 'real' points are from simulation and the red and green lines represent the largest possible $\pm\Delta\varphi$ variation and therefore the worst-case momentum resolution.

For this example the momentum resolution is very good, increasing to approximately 20% at p_T = 100GeV. The approximately exponential degradation of resolution at 100GeV shown by the red curve is the result of the angular separation of the track approaching the intrinsic resolution of the pixel system. The reconstructed position resolution on the ECAL face in $\Delta\varphi$ is shown in figure 6.
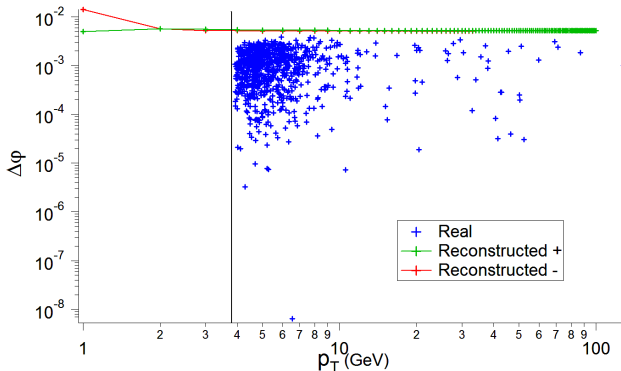

Figure 6: Reconstructed angular resolution at the ECAL-Tracker interface plotted as a function of real transverse momentum. The black line represents the reconstruction cut, the 'real' points are from simulation and the red and green lines represent the largest possible $\pm\Delta\varphi$ variation and therefore the worst-case momentum resolution.

This plot shows approximately flat behaviour over the p_T range of interest. At higher p_T the transverse momentum resolution becomes less important as the track becomes approximately straight, and therefore the position error, as for the transverse momentum case, becomes dominated by the intrinsic resolution of the pixels. In any case it should be noted that the error is far smaller than the size of an ECAL trigger tower, being approximately 0.003 radians, or 4mm when projected (this is in fact smaller than a single ECAL crystal).

The resolution in the z-direction can also be calculated, and is naturally better than for the transverse projection as it only depends on the pixel size and super-layer separation. Figure 7 shows the result for the geometry described previously. The worse case here is in the central region of the detector where the separation between the hits is smallest, and vice versa for the forward region. Even in the central region the resolution is approximately 0.001 in pseudorapidity or approximately 1.3mm, again far better than required.
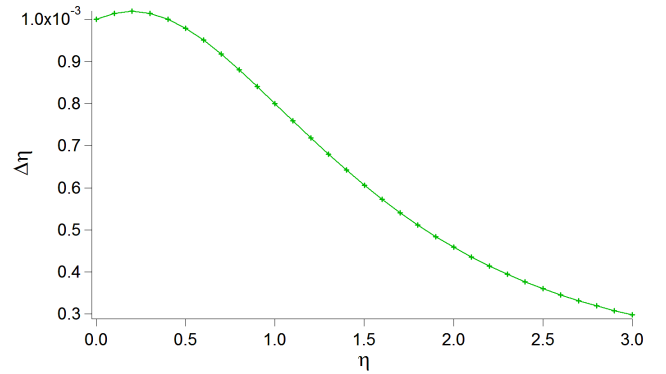

Figure 7: Reconstructed pseudorapidity resolution plotted as a function of hit location.

It is noted that the above methods do not take into account multiple scattering. Although such effects can be minimized by good detector design and low material budget construction, this will also need to be taken into consideration in the development of optimal algorithms. Full Monte Carlo studies for the purpose of testing such algorithms would be of limited value while the design of any future tracker (stacked or 'traditional') is still undecided.

II. DESIGN CONSIDERATIONS

The use of two stacks introduces additional complications in the detector design. While the basic architecture of an individual super-layer remains relatively unchanged from that described in [2], the off-detector processing necessarily becomes more complex as one moves from a stub generator to a simplified form of track reconstruction. The requirement for more precise calibration of the correlation and compensation for mechanical alignment issues necessitates a more comprehensive on-detector processing solution.

A. Detector Structure

As discussed previously, a significant problem is the sensing element for the new detector. At a radius of 10cm from the interaction point the required radiation tolerance is 10^{16}p/cm^2, 300Mrad. This makes full depletion impossible using thick sensors. Furthermore a charge collection speed of

<5ns will be required if SLHC operates in an 80MHz bunch crossing mode.

In this paper a different approach to radiation tolerance is taken. As it is not clear whether a technology exists that can withstand this dose for several years, an alternative approach would be to use a 'throwaway' detector. In this case one should pick an industry-standard technology with proven radiation tolerance approaching but not necessarily meeting the level required to operate in the SLHC environment for several years. Monolithic Active Pixel Sensors (MAPS) [6] are one example of this kind of technology. Some cost studies [7] have shown that in bulk production the technology is significantly cheaper than the custom hybrid pixel technology currently used in CMS [8].

The use of a standard technology such as MAPS also affords the use of fabrication techniques such as stitching. In the design considered here this is taken advantage of to manufacture wafer-size sensors of dimensions 2x10cm^2. This allows one to reduce the material budget by minimising the amount of wastage caused by overlapping detector segments.

Another proposed technology is a lower material budget support structure such as carbon fibre–Kevlar laminates, as currently used in BaBar [4]. By using thin sensors and newer support structures it should be possible to minimise the impact of the new detector on the total material budget. A possible detector layout is shown in Figures 8 and 9.
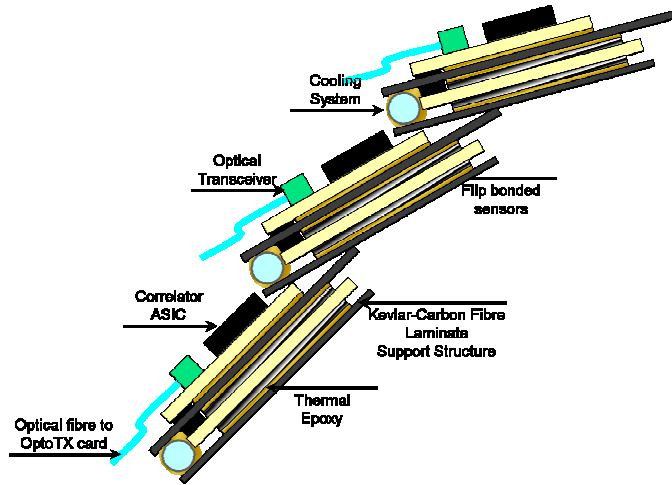

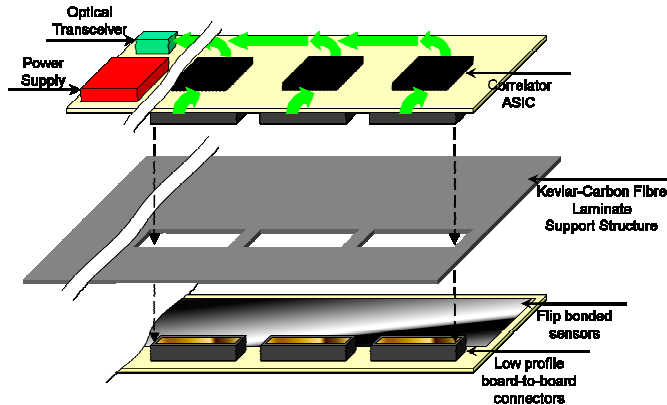Figure 8: Possible detector structure for a stacked pixel array

Figure 9: Possible stack-up for a stacked pixel array

While MAPS are a standard technology that offers good prototyping potential and low cost, one should be prepared to follow industry and change technologies if a superior cost-effective one becomes available. One example of this is 3D ASIC technologies [9].

III. CORRELATION LOGIC IMPLEMENTATION

It was shown in the previous study[2] that the correlation logic could be implemented using a difference analysis technique. While this is a good starting point, it results in two complications. Firstly it does not allow for calibration against the mechanical placement of the detector. This would be useful in order to compensate for fact that the detector is comprised of non-ideal flat segments, as opposed to being a perfect cylinder. Secondly the difference analysis relies on the beam spot location being at or close to r=0.

An additional gain (approximately a factor of four) in detector data rate can also be achieved by filtering in z, by loading calibration into each correlator and filtering in φ and z. It is now assumed that this processing and pixel clustering will be performed on-detector. By encoding only the clusters and the correlated pixel columns in φ rather than the absolute column address, it should be possible to reduce the data rate by a further factor of two. This would require the storage of 1024 calibration constants of 8 bits each for a 256x256 pixel array (8kb).

B. Track Reconstruction Implementation

Once the data has been processed by the correlators on the detector, the data is sent off-detector and drawn into SNAP12 fibre bundles at 40Gb/s/bundle, increasing the data density. By this means the data rate into the first stage of processing can be increased to approximately 200Gb/s/board using five SNAP12 receivers. Figure 10 shows the on-off detector data flow.
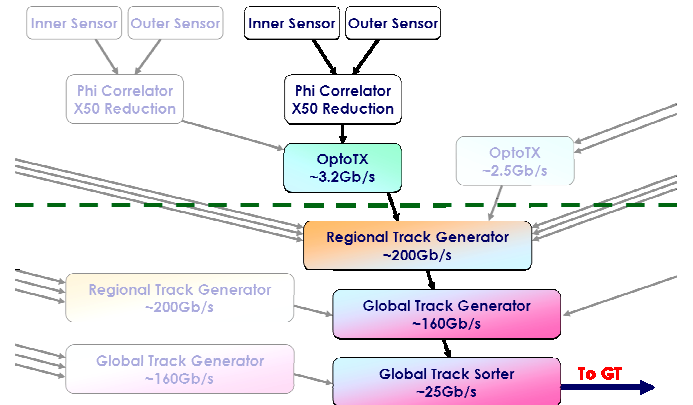

Figure 10: Illustration of different stages of data processing both on and off-detector. The top half of this diagram represents on detector electronics whilst the bottom half is off detector.

1) Regional Track Generator (RTG) – 200Gb/s

The first stage of reconstruction is managed by the RTG. Current firmware development has focused on this part of the system, and began first with an implementation of the correlator. This involves a combination of a column difference analysis and a z-binning method using constants

loaded into the internal FPGA RAM. In the final version of the firmware this is more likely to just use calibrated search windows for both sensor axes. Each RTG handles a single ring of sensors in the r-φ plane.

It is assumed that the data will be channelled directly into FPGAs on the RTG using the Multi-Gigabit Transceivers (MGTs) that are often integrated in modern devices. The Global Calorimeter Trigger Leaf Card [10] offers a possible prototyping platform for this board.

Implementation studies have shown that a serial correlator algorithm can be implemented that can pipeline-process hit pairs at 120MHz (~4Gbit/s/correlator), occupying approximately 0.8% of a Xilinx Virtex-II Pro 70 FPGA. In later generations of FPGAs this will operate at a higher clock speed and the algorithm itself will be further optimised in the future. It is unclear whether this part of the algorithm will be duplicated in the RTG or only implemented in the correlator on-detector.

The second purpose of the RTG is to pass paired hits in each super-layer forward to the Global Track Generator (GTG) for track building. The method used to achieve this is to project hits from the outer super-layer to the inner super-layer and subdivide the processing into pseudorapidity segments. For inner super-layer pairs the data in that segment is also forwarded to the same GTG. In this way all the possibly matching stubs naturally go the same card.

2) Global Track Generator (GTG) – 160Gb/s

The GTG finishes track building by pairing stubs from the two superlayers, calculating the transverse momenta for each track found and applying a second p_T cut at the detector level. The reduction in rate extrapolated from Monte Carlo studies is approximately a factor of forty, although this depends somewhat on the initial detector rate for each stack. Track candidates from this board are forwarded to the Global Track Sorter.

3) Global Track Sorter (GTS) – 25Gb/s

After the GTG the final rate decreases to a more manageable value. The card is responsible for housekeeping duties in this design and any final processing required. It also sorts the candidate tracks by detector region and measured transverse momentum. These candidates are then forwarded to the Global Trigger to be combined with track candidates from the Global Muon Trigger and hit candidates from the Global Calorimeter Trigger.

IV. POSSIBLE DETECTOR IMPROVEMENTS

The double stack method described above shows several benefits over the single stack method, most notably the proper (albeit crude) calculation of p_T and accurate projection of tracks to the calorimeter.

The design still leaves questions of mechanical calibration. By using a correlation based on calibration coefficients it will be possible to compensate for non-ideal detector geometry and misalignment of the detector. It also offers the possibility of compensating for beam vertex misalignment in the r-φ plane. This needs studying and is not discussed further here.

V. SUMMARY

It has been shown that the use of small layer separations in a pixellated detector system can be used to both reduce tracker combinatorials and reduce the data rate from the detector by means of a simple correlation algorithm. This algorithm could be implemented on-detector using relatively simple electronics; more advanced algorithms could be implemented off-detector in FPGAs.

By the use of more than one stack in several superlayers, the rate reduction can be achieved, and high-resolution track reconstruction and transverse momentum measurement also becomes possible. The design also provides a margin to compensate for real-world inefficiencies such as non-optimal resolution, malfunctioning pixels and system noise.

Our current work involves further development of the off-detector firmware and feasibility studies. Future work will require more refined simulation studies based on technologies such as MAPS. We are also considering a full-scale replacement of the tracker based on a minimum of four superlayers and new materials that can be used to reduce the material budget of the detector.

VI. REFERENCES

[1] F. Gianotti, M.L. Mangano, T. Virdee et al., Physics Potential and Experimental Challeges of the LHC Luminosity Upgrade, CERN-TH/2002-078 hep-ph/0204087.

[2] J. Jones. C. Foudas, G. Hall and A. Rose, A Pixel Detector for Level-1 Triggering at SLHC, 11th LECC Workshop, Heidelberg, Germany.

[3] A. Rose, C. Foudas, J. Jones and G. Hall, A Tracking Trigger for CMS at SLHC, 11th LECC Workshop, Heidelberg, Germany.

[4] C. Hill, US tracker in SLHC, 4th CMS Workshop on Detectors and Electronics for the SLHC, 4th April 2006.

[5] W. Smith, Trigger Considerations, 3rd CMS Workshop on Detectors and Electronics for the SLHC, 16th July 2005.

[6] R. Turchetta, Monolithic Active Pixel Sensors for the CMS Tracker, 2nd CMS Workshop on Detectors and Electronics for SLHC, 12th July 2004.

[7] R. Horisberger, Considerations for Future Large Pixel Layers, 4th CMS Workshop on Detectors and Electronics for the SLHC, 3rd April 2006.

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

[9] V. Suntharalingam et al., 3-D Integrated Circuit Fabrication Technology for High Density Electronics, 12th LECC Workshop, Valencia, 25th September 2006.

[10] M. Stettler, G. Iles, M. Hansen, C. Foudas, J. Jones, A. Rose, The CMS Global Calorimeter Trigger Hardware Design, 12th LECC Workshop, Valencia, 26th September 2006.