

Design of a trigger module for the CMS Tracker at SLHC

G. Hall, for the CMS Collaboration

Blackett Laboratory, Imperial College, London SW7 2AZ, UK

g.hall@imperial.ac.uk

Abstract

The CMS experiment is planning a major upgrade of its tracking system to adapt to an expected increase in luminosity of the LHC accelerator to $10^{35} \text{ cm}^{-2} \text{ s}^{-1}$. It will then have to cope with several hundred interactions per bunch crossing and fluxes of thousands of charged particles emerging from collisions. CMS requires tracker data to contribute to the first level trigger, to maintain the present 100kHz rate while increasing the trigger decision latency by only a few μs . A key part of a system to achieve this will be the design of a suitable module to generate trigger primitives.

One possible solution is based on so-called “stacked tracker modules” using closely spaced, coarsely pixellated sensor layers situated at intermediate radius within the tracker volume. A basic readout architecture is proposed and some of the electronic implications are described. Estimates of likely power consumption are given, and data rates and link bandwidth requirements.

I. INTRODUCTION

The upgrade of the LHC accelerator to Super-LHC (SLHC) foresees operating at $10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ luminosity to provide increased statistics, and allow deeper investigations into rare processes including, hopefully, discoveries of new physics. The LHC peak luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ will eventually deliver about $50 \text{ fb}^{-1}/\text{yr}$ [1,2] and CMS was designed for 10 years operation under these conditions. Most of CMS should survive predicted irradiation levels and perform well with few changes at higher luminosities, except for the Tracker which will gradually suffer degradation from radiation damage, probably reducing performance after about 500 fb^{-1} . There are already plans to allow earlier replacement of some layers of the pixel detector, since that will suffer more radiation damage than the outer Tracker. Trigger and data acquisition systems would also take advantage of technology evolution to be improved to cope with SLHC data volumes and rates whenever luminosity increases should occur.

The present Tracker surrounds the interaction point and provides precise, efficient measurement of charged particle trajectories and secondary vertices. It comprises pixels in three barrel layers at radii 4.4-10.2 cm and ten barrel layers of silicon microstrips to a radius of 1.1 m. It includes two endcap disks in the pixel detector and nine in the strip tracker on each side, plus inner barrel disks, extending acceptance to $|\eta|$ of 2.5. With about 200 m^2 of active area the CMS system is the largest silicon tracker ever built [3]. The pixel system is quickly removable, in case of beam-pipe bake-outs. Inner layer replacement was foreseen after several years of high luminosity operation as sensors reach irradiation levels corresponding to $100\text{-}300 \text{ fb}^{-1}$ integrated luminosity. The microstrip and pixel detectors have operated with the rest of

CMS taking cosmic ray data since 2008 and the performance looks very promising.

Most CMS sub-detectors will not change much for SLHC. It is important to maintain compatibility and retain the Level 1 trigger rate limit of 100kHz. Trigger latency can increase from $\sim 3.2\mu\text{s}$ to $6.4\mu\text{s}$, limited by electromagnetic calorimeter pipelines.

The notable exception is the tracking system, whose performance will eventually be degraded by radiation damage caused by immense particle fluxes. Greater radiation tolerance will be required, especially for sensors. In contrast, ASIC electronics should withstand SLHC radiation levels but the $0.25\mu\text{m}$ CMOS technology pioneered by CMS will be superseded by more advanced processes [4].

In the congested SLHC environment of 300-400 events per beam crossing, with thousands of particles emerging from interactions, higher granularity is required [5] CMS also requires to use tracker data in the first level trigger decision.

II. THE UPGRADED TRACKER

The first phase of the machine upgrade might be five or six years after LHC start-up, to reach a peak luminosity of $2\text{-}3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Inner focusing magnets will be replaced, larger aperture collimators installed and the proton linac replaced to reach the ultimate LHC current. Around the same time the inner layer of the pixel system should be replaced. It looks possible to rebuild the pixel detector to achieve improved performance by reducing material [6,7]. In the longer term the pixel system for operation at $10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ is expected to be similar in detector area and material budget to the Phase I device. It is also expected to evolve further, with a new ROC architecture and pixel size optimized for SLHC conditions.

High quality tracking and vertexing performance must certainly be maintained in the congested SLHC environment. From simulations of heavy ion events in the present tracker, with similar track density to SLHC, an extra pixel layer would restore track seeding losses. A new layout can be optimised for track finding and jet reconstruction. Granularity must increase because of leakage currents as well as track recognition.

Multiple scattering, photon conversions, bremsstrahlung and hadronic interactions are undesirable, and depend on limiting material within the Tracker. A major constraint for a new system is that cooling pipes, power cables and optical fibres follow complex, congested routes and installation was time consuming and difficult. It is unlikely they can be replaced.

Another major challenge is the requirement to use Tracker data for the first time in the Level-1 trigger. Single μ , electron and jet Level 1 trigger rates at SLHC will greatly exceed

100kHz and cannot be reduced sufficiently by increasing p_T thresholds or by other improvements in calorimeter and muon system algorithms. Tracking information in the present High Level Trigger (HLT) already provides additional rejection power and motivates future use of Tracker data at L1. However, the constraints are very different, since the HLT has access to all the Tracker data for almost complete track reconstruction and has a relatively long time (~ 40 ms) available. In contrast a L1 track trigger must make decisions in a few μ s and it does not seem feasible to transfer data to HLT processors and fully reconstruct tracks. The data volumes are simply too large.

One proposal has been made to use cluster width information to eliminate low p_T tracks [8]. An alternative which has been simulated in some detail deploys closely spaced, coarsely pixellated sensor layers at intermediate radius and compares hit patterns [9] to eliminate data from low p_T tracks, thus reducing the data volume significantly. The p_T cut is set by the angle of a track in the layer, and the logic might be relatively simple. The development of modules which would allow this is the main subject of this paper.

Presently there is no single design agreed for the Phase II Tracker. Simulations are vital, and alternative layouts are under consideration to investigate performance in detail.

III. THE TRACK-TRIGGER CHALLENGE

The major difficulty implementing tracking triggers at Level 1 is the data volume. It is easy to see that it is not feasible to transfer all data off-detector for decision logic. For example a single layer at a radius of 25 cm with 2.5mm x 100 μ m pixels is expected to have an occupancy $\sim 0.5\%$ at 10^{35} cm $^{-2}$ s $^{-1}$. This would require ~ 20 M channels of coarse pixels, each contributing about 24bits so a data rate of $\sim 96,000$ Gb/s needing an enormous number of optical links, and power. Therefore some method for on-detector data reduction, selective readout, or a combination, is essential. Pixellated trigger layers will be more power hungry than microstrip layers, so the challenge is obvious.

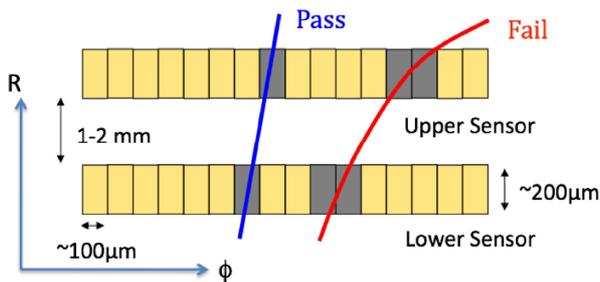


Fig. 1. The principle of selecting high transverse momentum tracks in stacked layers. A stub is a pair of hits passing the selection criteria.

The charged particle transverse momentum spectrum of contains a large fraction of low p_T tracks which are not useful for triggering. It is conceptually simple to estimate the transverse momentum using pairs of closely spaced layers [9], provided sensor element sizes are properly dimensioned, which depend on the radial location of the layer (fig. 1). A double layer identifies “stubs” which are pairs of nearby hits in the two sensors which allow to define a track with transverse momentum above a p_T threshold. The method to find stubs is simply to compare a binary pattern of hit pixels on upper and lower sensors, possibly with a processing step

on the detector which identifies clusters rather than using individual hits since this is expected to lead to extra combinations. These should be the trigger primitives transferred to the L1 trigger system for more sophisticated algorithms to process for the final trigger decision.

Under SLHC conditions, the hit density means a high rate of combinatorial background if the sensor area which is searched for matching pairs is not carefully defined. This will depend on the radial separation of the two sensor planes (fig. 2).

Quite extensive simulations [10] have been carried out using a layout of the detector which includes a realistic model of individual detectors and the services thought to be required to power and read out the double sensor layer modules producing trigger stubs, which are here referred to as “PT modules”. The objective is to understand better how the PT modules can best contribute to trigger and the overall L1 trigger rate reduction which is achievable. Some results are illustrated in Table 1 and fig. 2 for events containing muons pairs in the presence of high pileup, suggesting sensor separations of less than a few mm could meet the requirements.

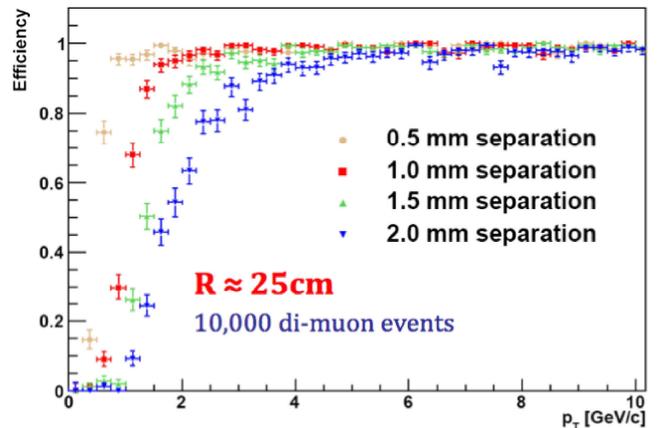


Fig 2. Efficiency for constructing stubs as a function of transverse momentum in high luminosity conditions at SLHC. Here the selection criteria are a row window of 3 pixels, and a column window of 2 pixels for 0.5mm spacing, and 3 pixels for 1mm- 2mm.

Efficiency is the fraction of stubs to tracks above 2 GeV/c, while the fake rate is associated from stubs which are formed when hits from two tracks which would not pass the p_T cut on their own are correlated to generate fake stubs. The reduction factor is the ratio of hits to stubs, which require to be read out. As can be observed, a separation of about 1mm between sensors provides high efficiency and low fake rate with a reduction factor ~ 20 . Efficiency falls as the separation between the layers increases, largely because of geometrical acceptance; the fake rate also increases as it becomes harder to reject accidental combinations.

Table 1. Simulation estimates of stub finding efficiency in 100 μ m x 2.5mm pixels stacked layers at 25cm radius, with 0.5% occupancy.

Radial separation [mm]	Efficiency [%]	Fake rate [%]	Reduction factor
0.5	99.0	0.7	8.0
1.0	99.4	4.1	22
2.0	97.7	17.8	96
3.0	96.0	39.0	210

IV. MODULE REQUIREMENTS

The studies into the definition of the trigger layers have begun to pose questions such as the following

- how are the stubs to be used in the trigger and what rejection factors are achieved?
- how many layers are needed?
- what is their optimal location, allowing sufficient η coverage?
- what is the impact of material, in trigger layers and elsewhere, on trigger performance?
- how important is z-measurement of the primary vertex, and what is the required resolution?
- what is the impact of the trigger layers, which will certainly be more massive than conventional tracking layers, on tracking performance?
- what are the likely cost, power requirements and contribution to the tracker material budget?
- can the layers be read out at the full 40MHz rate or is a L0 trigger, i.e. a signal preceding the L1 trigger, needed to select a region of interest?

Although the answers to some of the questions posed above are needed to guide the design of the PT module, it is also difficult to answer them without concrete details of a module design in mind. For this reason, this must be an iterative process. For the present a couple of concepts are being evaluated, with the hope to identify an optimum design which can be prototyped by a collaboration within CMS.

A. Schematic Module design

The first module type is illustrated in Fig. 3. The pixel size is $100\mu\text{m}$ by 2.5 mm arranged in columns of 256 rows with 32 columns per module, so an approximate active sensor size of $25.6\text{ mm} \times 80\text{ mm}$ contains 8192 pixels. The sensor is expected to be $200\mu\text{m}$ or less in thickness. Hits are read out to the upper and lower sides of the module where the connections between the two sensors are made, which allows a module to be constructed without material under the sensitive sensor area in the interests of minimizing multiple scattering in the measurement paths and reducing heat dissipation in the immediate proximity of the sensors. The data are read out from a column of 128 pixels and transferred to the end of the readout chip on each clock cycle. Typically less than one pixel per column will be hit in each beam crossing and this may be exploited to avoid a high speed serialiser which is expected to be too power hungry.

The readout ASIC (ROC) for each column is assumed to be a 128 channel front-end element, with amplifier and other circuits in each pixel, plus an “assembler” at the periphery where data to be used by the trigger are temporarily stored and comparisons of patterns between the two layers are made. To minimize the interconnections on the module and take advantage of the higher density of metal lines possible at the chip level, the assembler is part of the ROC ASIC, not a separate chip. Probably several columns will be amalgamated into a single chip, perhaps with up to 8 adjacent channels (20mm wide). Note that two ROCs are required to read out the full module width.

It is assumed that the high speed links required for the system will be based on the CERN GBT (Gigabit

Bidirectional Transmission) [11] and Versatile Optical Link projects [12] which are developing a radiation hard bi-directional optical links for use in the LHC experiment upgrades. At the edge of the module there is another ASIC, referred to as a “concentrator” which would be the interface to the GBT, or actually a component of the GBT chip set which would be an economical solution. It would provide inputs to (clock, trigger, control data) and outputs (data for the track-trigger) from the module.

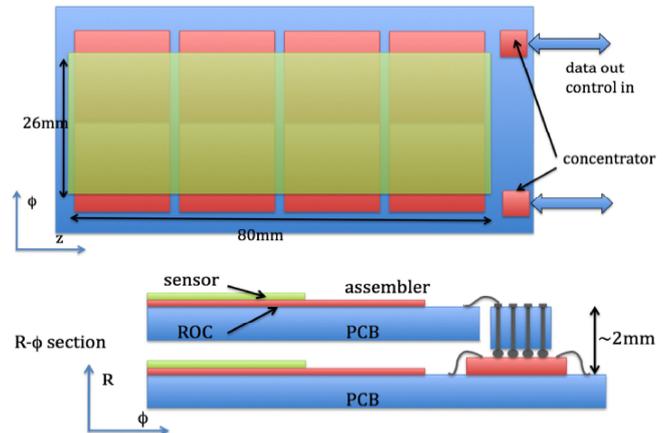


Fig. 3 The PT module seen in plan view (upper) and in section (lower), indicating connections required by the two sensor layers. The sensor is bump bonded to the ROCs. In this example 8 ROCs with a total of 8192 channels are required.

A significant contribution to the total power required for PT layers comes from the links, which are assumed to require $2\text{W}/\text{channel}$ for 4.8 Gbps including error correction, of which 3.2Gbps is available for data. Roughly 3000 GBT links are required to read out a layer of about 40M pixels at radius of 25cm , assuming a data reduction factor of 20 and an occupancy of 0.5% and 24 bits transmitted for each selected pixel (sending data from only one of the two sensors in the double layer). These figures assume 50% use of the GBT bandwidth, so the link power requirement is 6kW for the layer, or $150\mu\text{W}/\text{channel}$. It seems that detector layout will constrain the location of the GBT transceivers to the close vicinity of the module. It is plausible that the link speed might double for the same estimated power consumption, so there is hope to improve on this contribution to the power budget. There are significant uncertainties in these estimates as factors such as ability to select clusters and optimal layout of links, as well as the use of bandwidth must be better understood.

The logic of the readout chip design is illustrated in fig. 4. Identical chips are foreseen in the two layers, with certain elements not operational, because not needed, in one layer to save power. Hit data are transferred to a memory buffer in the assembler area each clock cycle, then data from the lower layer are passed to the upper layer. In the upper layer, hits are also passed from neighbouring columns so a comparison can be made between patterns from the lower layer and three columns in the upper layer to make a decision on valid stubs. Those patterns consistent with high p_T stubs are transferred from the module off-detector.

Very provisional estimates of local power requirements have been made which suggest that, using 130nm CMOS, about $100\mu\text{W}$ per channel might be achievable, leading to a total of $\sim 250\mu\text{W}$ per channel. It should be emphasised that these evaluations are quite uncertain, since chip designs have

not begun and the logic and local data transmission rates are not well understood. However, such estimates are essential in developing the module design.

Hit data should be stored on the pixel for full readout following a Level 1 trigger. Given the likely number of layers in the future Tracker, it is probably desirable to read out all hit data from PT modules despite the low p_T threshold, which will add further to the power required. This functionality would also be valuable for evaluation. It is estimated that a binary pipeline in each pixel would require too much space and an architecture similar to that used in the present pixel detector is under discussion.

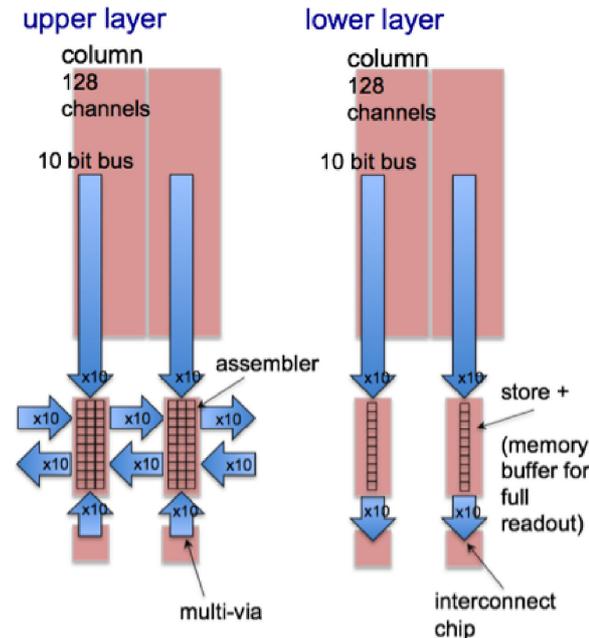


Fig. 4 (left) Schematic of possible layout of ROC chip to read out 128 pixel columns, in this case grouped in units of 4 ROCs per chip. (right) Schematic of the data flow to allow comparison logic to be placed in the assembler area of the ROC, at the periphery of the sensor.

The total power consumption for stacked layers with these pixel dimensions can thus be estimated to be about 10kW for 40M pixels at 25 cm radius, and 19kW for 75M pixels at 35 cm radius. The total number of links required is 2900 and 5600 for the two cases; which does not allow for full readout of the layers, only track-trigger data. These layers will therefore represent the major contribution to power consumption of a likely layout of a new Tracker and great care will be needed not to allow either power, material or numbers of links to increase significantly if the tracking performance is to be maintained.

3.2 Alternative Module design

CMS now has experience of automated assembly but it will be highly desirable to optimise construction to take full advantage of commercial manufacturing. It may be possible to design a module with more advanced technologies, and transfer many of the assembly issues to industry. To do so requires a different approach to the logic and a careful evaluation of commercial methods, where multi-layer technologies continue to advance significantly [13].

The concept is derived from hybrid pixel detectors. The basic module consists of a matrix of read-out chips (TFEA: Tracker Front-End ASIC), each integrated circuit an array of

4 by 160 identical channels, mapped onto a corresponding array of $100 \mu\text{m} \times 2000 \mu\text{m}$ elements on the silicon sensor. (Dimensions are illustrative, chosen to use a 150mm diameter sensor wafer.) The modules proposed are composed of a sandwich as illustrated in fig. 5 and assembled using a combination of standard technologies, such as wire-bonding and bump-bonding.

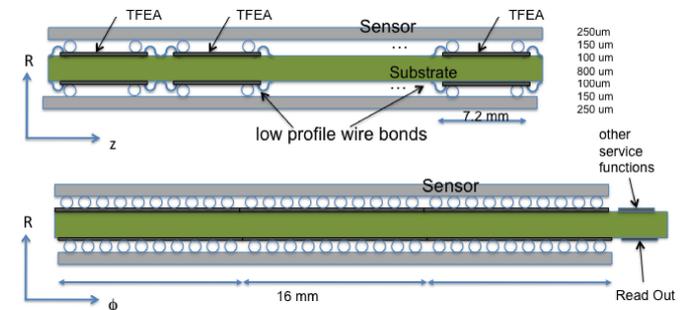


Fig. 5 Cross section of the module in the views along z and in the phi plane. Dimensions are illustrative.

The ASIC should be large enough to cover the sensing area with a minimum of dead space as well as reducing module power consumption, so therefore will avoid moving data at high speed across chips whenever possible.

The integrated circuits are connected using wire-bonding or bump-bonding on a double-sided substrate (fig. 5). The example illustrated requires through-vias in an intermediate layer, of the type used commercially for low cost memory assembly. The read-out chips are then sandwiched between two silicon sensor layers connected to the chips using coarse pitch bump-bonding which should be readily available, e.g. $\sim 200\mu\text{m}$ minimum pitch with relatively large bumps. The choice of an inexpensive and well known interconnection technology minimises costs, risk and investment and simplifies the manufacturing process. In addition, the concept is intended to allow straightforward testing on the ASICs, to enhance module production yield and simplifying manufacturability.

The architecture differs from the previous concept, by aiming to perform all necessary functions *locally* on each front-end chip. No transfer of data to a correlator or assembler area on the chip is necessary as all front-end and triggering functions are performed in or close to each pixel in the TFEA chip.

The Front-End ASIC is composed of a number of identical functional units, all present in each channel but not necessarily activated depending on the position of the ASIC in the module. These units are:

- A front-end amplifier, shaper and discriminator providing a binary yes/no answer at each bunch crossing.
- An Event Store buffer, to store the decision of the previous stage until arrival of the L1 trigger.
- A Data Link unit used to send information retrieved from the Event Store buffer upon arrival of an L1 trigger.
- A Local Trigger link, used to send promptly at each bunch crossing information from the FE to the TFEA chip(s) in the layer below.
- A Trigger Logic block to correlate information from the FE blocks in order to find stiff tracks. Clearly this logic block must have connectivity to adjacent pixels.

- A Trigger Link block to send promptly the result of the previous Trigger Logic block, if positive, to an external trigger logic block.
- A Clock and Control block to receive and regenerate the clock necessary to operate the TFEA. It also contains a slow control interface necessary to address local configuration registers and ancillary logic.

The connectivity of these blocks inside each TFEA and across the two layers of a module is illustrated in fig. 6. Both layers obviously contribute to the formation of a trigger primitive, but only the lower one saves and transmits the data upon arrival of the L1 trigger, which is depicted. It is expected that the hit information from only one of the two layers would be transmitted. As the correlation of the hit position is performed inside the chip on the lower layer of the assembly, this architecture should minimise data movement outside the TFEA chips and therefore reduces power consumption.

This conceptual design poses several important questions, including cost, such as practicality of large scale, low profile wire bonding and the assembly of large area sensors on a multi-layer substrate, with double-sided ASIC assembly, as well as issues concerning the logic and data transmission schemes. Wire and bump bonding meeting these specifications is routinely done in high volume flash memory assemblies at very low cost and with high yield and through via substrates are also in widespread use.

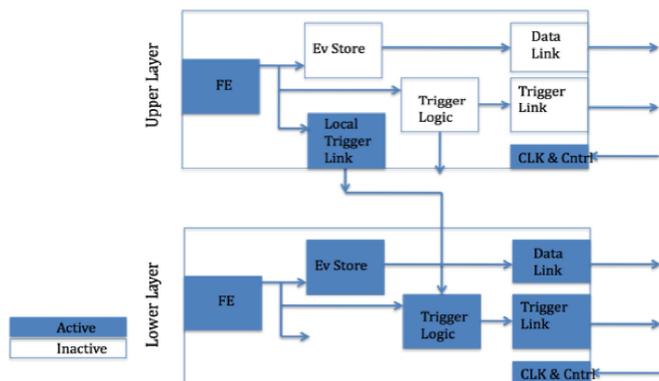


Fig. 6. Block diagram for the logical functions of a pair of TFEA chips

3.3 Developments ahead

The next steps in developing these ideas is to evaluate the approaches to module development in more detail and to compare and contrast the pros and cons. For example it is important to

- understand the impact on the material budget
- understand the implications of different choices for power or logic
- identify and design building block circuits
- understand the requirements for commercial manufacture, including costs and the scale of technological challenges
- evaluate many issues for module construction, especially power and cooling.

There are also many practical details which must be better understood, such as the handling of z-offsets, implementation of comparison logic before hopefully arriving at single concept for prototyping. It will be essential to evaluate a real

module in a beam test, even with limited features, since this type of module has never been used in a previous experimental system.

V. CONCLUSIONS

Modules which will provide trigger primitives for use in CMS look feasible. Once prototyped they will provide a new part in the detector toolbox but will contribute a large fraction of a future Tracker power and material budget. The physics objectives will become clearer in the next few years and may evolve so designs which are flexible will be needed. It is also crucial to improve understanding of power consumption, which is sensitive to occupancy and achievable rejection factors in the selection.

Benefits from ASIC technology feature size reduction are expected in implementing the features required but no dramatic performance gains are yet anticipated. In addition, the questions of how to process off-detector very large volumes of tracker data may not be straightforward, as well as the trigger algorithms required to utilize the information, so there are further challenges ahead. “Conventional” assembly methods may be feasible but it will be important to evaluate commercial manufacturing, exploiting technology progress, which may have a very important role in building these novel features into a future tracking detector. Prototype module development is now very timely.

ACKNOWLEDGMENTS

Many colleagues in CMS developed the outstanding tracking detector which has now been completed and are contributing to preparations to improve it for an even more demanding future. The ideas for the second module concept originate with A. Marchioro, who I would like to thank along with D. Abbaneo, K. Gill, M. Pesaresi, M. Raymond, A. Ryd for very valuable discussions in developing these ideas.

REFERENCES

1. The CMS Collaboration. *CMS Tracker Technical Design Report*, CERN/LHCC 98-6 (1998) and Addendum CERN/LHCC 2000-016 (2000)
2. The CMS estimate of integrated luminosity assumes 10^7 s operation annually at nominal luminosity with a 50% operational efficiency factor.
3. The CMS Collaboration *The CMS experiment at the CERN LHC* 2008 JINST **3** S08004 [doi: 10.1088/1748-0221/3/08/S08004]
4. G. Hall *Recent progress in front end ASICs for high-energy physics* Nucl. Instr. and Meths in Phys. Res., A541 (2005) 248-258
5. F. Gianotti, M. Mangano, T. Virdee, et al., *Physics Potential and Experimental Challenges of the LHC Luminosity Upgrade* Eur. Phys. J. **C39** (2004) 293–333. [doi:10.1140/epjc/s2004-02061-6.]
6. G. Hall, *The upgrade programme of the CMS Tracker at SLHC*. Vertex 2008. To be published in Proc of Science
7. R. Horisberger. Private communication and internal CMS presentations.
8. F. Palla 2007 JINST **2** P02002
9. J. Jones, G. Hall, C. Foudas, A. Rose. *A Pixel Detector for Level-1 Triggering at SLHC*, arXiv:physics/0510228. CERN Report CERN-2005-011(2005) 130-134
10. M. Pesaresi. Private communication and internal CMS presentations. PhD thesis in preparation.
11. <https://espace.cern.ch/GBT-Project/default.aspx>

12. <https://espace.cern.ch/project-versatile-link/default.aspx>
13. A. Marchioro, who developed the ideas in this section. Private communication.