

CMS Microstrip Tracker Readout at the SLHC

M. Raymond ^a, G. Hall ^a,

^aThe Blackett Laboratory, Imperial College London

m.raymond@imperial.ac.uk

Abstract

The increased luminosity at the SLHC and associated increases in occupancy and radiation levels present severe challenges for the CMS tracker, which will require complete replacement. Inner pixellated regions will expand to higher radii and the outer tracker region will most likely be instrumented with short strip silicon sensors. It is also necessary for the tracker to provide information to the level 1 trigger if the overall CMS trigger rate is to remain at 100 kHz.

Power consumption is one of the main challenges for the tracker readout system, because of the higher granularity necessary. The current status of architectures for a short strip outer tracker readout chip is presented, with projections for performance and power consumption.

I. INTRODUCTION

A major challenge for tracker readout systems at the SLHC is power consumption (and provision). Higher luminosity and hence granularity means more sensor channels and front end chips, and the CMS material budget at the LHC is already dominated by electronics power consumption related material (cabling and cooling).

Another major challenge for the CMS tracker at SLHC is the need to provide trigger information, since without this it is not possible to maintain the average level 1 accept (L1) rate at 100 kHz (the LHC value) [1] unless changes are made in the CMS trigger strategy.

Higher granularity and triggering requirements mean that a complete replacement of the CMS tracker is required. It is hoped that power consumption can be controlled by making use of advances in electronics technology, but savings depend on any additional front end functionality required for the SLHC. Advances in high-speed digital optical link technology may help to reduce the resources required to implement off-detector links (using commercial developments).

II. CMS STRIP READOUT AT THE LHC

The current CMS silicon strip tracker readout system at the LHC is illustrated in figure 1. Analogue readout was chosen, with no sparsification (zero suppression) on-detector, utilising 0.25 μm CMOS technology throughout. The APV25 front end chip [2] instruments the AC coupled silicon sensors and the output data from two APV25 chips (figure 2) are combined onto one optical fibre, by the APVMUX chip, at 40 Ms/s. Opto-electrical conversion is performed in the off-detector CMS FED readout boards [3], where digitization, pedestal and common-mode subtraction, followed by zero suppression, are performed.

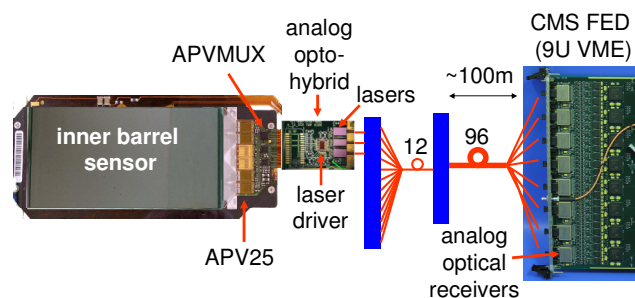


Figure 1: CMS silicon strip readout system at the LHC

Because of no zero suppression on-detector all front end chips operate synchronously. Taking advantage of this, the state of the front end chips is emulated externally, at the L1 trigger control system level, by the APVE VME module [4]. One function of the APVE is to predict the address of the pipeline location in the APV25 which will be triggered, which is then sent to the FED readout boards. This address is compared with the value subsequently transmitted by all APV25 chips in the system, in the output frame digital header information (figure 2), giving a strong check on the correct functioning of all front end chips.

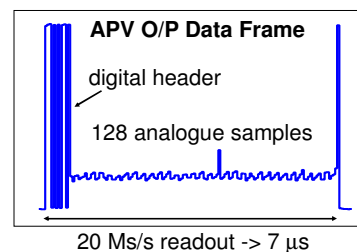


Figure 2: APV25 output data frame format

There are other advantages of a synchronous, non-sparsified, system. Since an L1 trigger sent down will result in the return of a data frame from every front end chip, and the L1 latency is fixed, there is no need to timestamp data on-detector. The data volume per trigger is also occupancy independent which greatly simplifies the functionality required to combine data from more than one front end chip onto off-detector links. Transmitting the raw data off-detector allows the functionality of pedestal and common-mode noise subtraction, and zero-suppression to be performed where the associated power consumption is less critical. The raw data are also available to help set up, diagnose any suspected faults with, and monitor the performance of the front end system.

The analogue, non-sparsified approach provides a relatively simple and robust readout system for the CMS strip tracker at the LHC.

III. SOME POSSIBLE SLHC CHIP ARCHITECTURES

Figure 3 shows a functional representation of the 128 channel APV25 chip used for strip readout at the LHC. The APV25 [2] uses a relatively slow front end amplifier which produces a 50 ns CR-RC shape pulse, sampled into the pipeline at the 40 MHz bunch crossing frequency. The pipeline is implemented by gate capacitance, which gives the highest capacitance per chip area possible in the process.

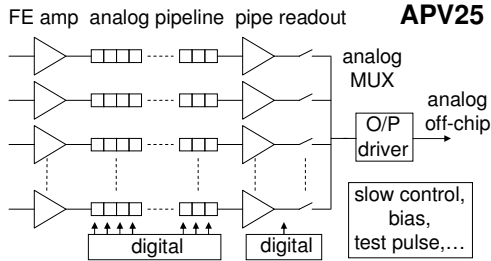


Figure 3: Functional block diagram of the APV25 readout chip

In response to an L1 trigger, the analogue samples stored in the pipeline can be read out in either peak or deconvolution modes [2], where deconvolution mode provides single bunch crossing resolution for signals.

The analogue approach of the APV25 allows pulse height information to be retained and transmitted, but is clearly incompatible with digital off-detector transmission envisaged for the SLHC. If pulse height information is to be retained it becomes necessary to consider where to introduce the digitization step in the analogue chain.

Digitization early in the signal processing chain, before the pipeline, allows analogue functionality to be confined to the front edge of the chip, and a digital pipeline is possible, which should lead to a minimal area requirement for this circuit. This requires an ADC on every channel and table 1 gives power estimates for 6 or 8 bit ADCs running at 20 MHz (one of the proposed bunch crossing frequencies for the SLHC) in 130 and 65 nm technologies. The numbers in the table are based on International Technology Roadmap for Semiconductors (ITRS) 2003 predictions [5].

Table 1: Power estimate, in mW, for a 20 MHz CMOS ADC based on the ITRS roadmap 2003 [5]

Technology	130 nm	65 nm
8 bits	6.4	2.5
6 bits	1.6	0.6

The power consumption per channel for the present APV25 based readout is 2.7 mW, and a significant reduction in this value is required for SLHC. From table 1 it is clear that an ADC on every channel is not a viable option, even at 20 MHz, and it is also possible that the bunch crossing frequency at SLHC may remain at the LHC value of 40 MHz.

Retaining pulse height information without an ADC on every channel means that the required digitization must be performed at a point where the channel information has been brought together, and an obvious option is then to digitize after the multiplexing stage. Figure 4 illustrates this, where the ADC power is shared between all channels. For example,

6.4 mW shared between 128 channels results in a power consumption per channel of only 50 μ W.

Digitization after the multiplexing stage requires that the analogue pipeline and multiplexing stage present in the APV25 is retained, and the slow shaping (plus deconvolution) feature could also be retained. Implementing the pipeline using gate capacitance may not be an option as the gate oxide thickness reduces with feature size, and significant leakage results. It may still be possible in 0.13 μ m technology, but not for finer feature processes, which would have implications for overall chip size as other capacitor implementations tend not to be so area efficient.

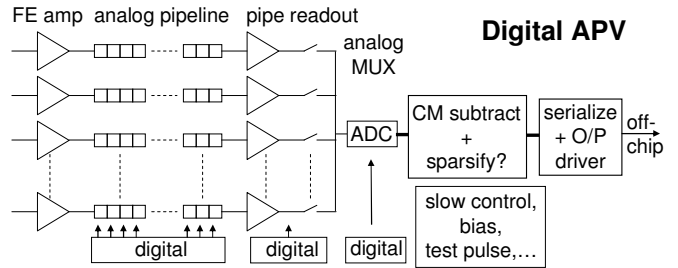


Figure 4: Functional block diagram of an APV-like chip with on-chip ADC.

Figure 4 includes a further block between the ADC and output data serializer stage, where additional functionality could be implemented, such as the pedestal and common-mode noise subtraction currently implemented off-detector in the LHC system. It will be shown in section IV that digital data volumes and associated transmission power at SLHC will dictate that sparsification is necessary if pulse height information is retained.

Comparing the existing APV25 and “digital APV” architectures in figures 3 and 4, it is clear that the digital APV contains all the complexity of the APV25, plus additional complexities of digitization and on-chip sparsification, which will not help to keep power consumption low. On chip sparsification will also add complexity to the overall front end system, losing some of the attractive features of the present system already discussed in section II. The opposite extreme to a digitized analogue, sparsified readout system in terms of complexity is binary, non-sparsified readout.

Figure 5 shows a binary non-sparsified architecture. For binary readout a fast front-end amplifier and comparator are required. The front-end amplifier speed must be sufficient to enable the hit to be registered in the correct bunch crossing.

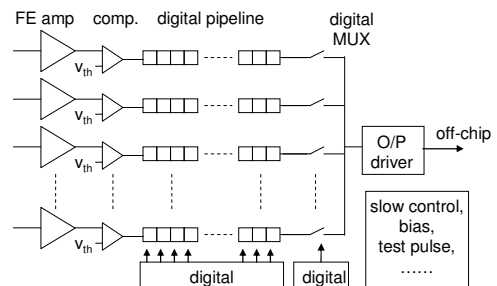


Figure 5: Functional block diagram of a binary, non-sparsified front end chip architecture

Although the architecture in figure 5 does not look greatly different to that in figure 3, the implementation of the functional blocks would be substantially simplified. The pipeline is only one bit per channel and the area it would occupy would be small. The readout would just require the retrieval of a 128 bit digital word from the pipeline and a simple 128:1 digital multiplexer operation, the resulting data stream being transferred directly off-chip.

It seems likely that the digital power associated with an architecture like that in figure 5 would be low. Front end power cannot be as low as for slow shaping, because of the speed requirement and additional comparator functionality.

IV. SLHC CHIP POWER ESTIMATES

A. Front end amplifier simulations

The APV25 was designed for long strips with sensor capacitances, C_{SENSOR} , in the range 15 – 25 pF. The noise performance and rise-time of a charge sensitive preamplifier depends on C_{SENSOR} and g_m , the transconductance of the input FET, according to the formulae:

$$\text{noise} \sim C_{\text{SENSOR}} / \sqrt{g_m}, \quad \text{rise-time} \sim C_{\text{SENSOR}} / g_m$$

For short strips at SLHC, C_{SENSOR} is reduced, so lower values of g_m are possible. g_m depends on drain-source current, and supply voltages are halved when moving from 0.25 μm to 0.13 μm technology so significant savings in input device power are possible.

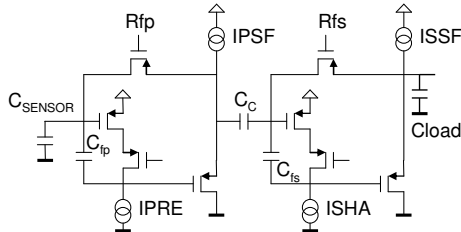


Figure 6: 0.13 μm preamp/shaper circuit for simulations

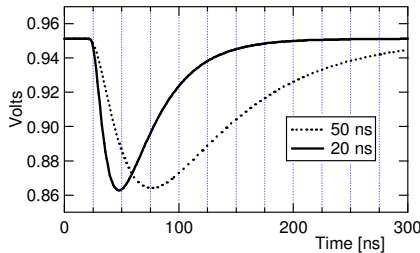


Figure 7: Simulated pulse shapes for the circuit of figure 6, for $C_{\text{SENSOR}} = 5 \text{ pF}$ and with the bias currents in table 2.

Figure 6 shows the transistor level schematic of a simple preamplifier/shaper circuit used to evaluate the performance achievable in 0.13 μm CMOS. A simple two supply rail circuit is chosen with a PMOS input device. The bias currents to the preamplifier and shaper input devices, IPRE and ISHA, together with the shaper feedback resistor Rfs, were varied to achieve the two different CR-RC pulse shapes in figure 7. The 50 ns pulse shape might be suitable for a digital APV type architecture like that in figure 4, whereas the 20 ns pulse shape would be more suitable for a binary front end. A sensor

capacitance of 5 pF was chosen, which would correspond to strip lengths of \sim few cm.

Table 2 shows the values of bias currents chosen to achieve the pulse shapes in figure 7, together with the corresponding power and simulated noise performances. It can be seen that for a 0.13 μm front end for short strips a good noise performance is achievable for substantially less than the $\sim 1 \text{ mW}$ required for the APV25 preamp/shaper.

Table 2: Bias currents, power and simulated noise performance for the circuit and pulse shapes of figures 6 and 7.

	50 ns	20 ns
IPRE [μA]	40	90
IPSF [μA]	15	15
ISHA [μA]	10	30
ISSF [μA]	35	15
Power [μW]	120	180
Noise [e]	800	890

B. Overall SLHC chip power estimates

Table 3 shows the power breakdown by functional sub-circuit of the existing APV25 chip. Tables 4 and 5 show estimated power consumptions, in similar format, for digital APV and binary non-sparsified architectures respectively, in 0.13 μm technology. Justifications for the numbers are indicated in the tables, but it should be emphasised that there are considerable uncertainties where estimates are provided for digital functionalities. Nevertheless a target power consumption of close to 0.5 mW per sensor channel seems appropriate for a 0.13 μm chip for short strip readout.

Table 3: 0.25 μm APV25 power breakdown

Sub-circuit	Power [$\mu\text{W}/\text{chan}$]
Preamp/shaper	1050
Inverter	500
APSP (pipe R/O)	200
Mux and O/P stages	550
Digital	400
Total	2700

Table 4: Estimated 0.13 μm digital APV power / channel

Sub-circuit	Power [μW]	Comment
Preamp/shaper	120	Simulated 50 ns CR-RC shaping, $C_{\text{SENSOR}} = 5 \text{ pF}$
Pipe readout	50	APV25 / 4 (guess)
ADC	50	Estimate from [ITRS]
digital	120	(APV25/10) x 3 - see text
Fast serial O/P	230	30 mW / 128 (based on estimate for fast LVDS)
Total	570	

Table 5: Estimated 0.13 μm binary, unsparsified chip power/chan.

Sub-circuit	Power [μW]	Comment
Preamp/ shaper	180	Simulated 20 ns CR-RC shaping, $C_{\text{SENSOR}}=5 \text{ pF}$
comparator	20	Simulated
digital	60	Simpler than digital APV
Fast serial O/P	230	Same as digital APV
Total	490	

C. Estimated link power contribution

Overall system and front end chip architectures are inter-dependent. Data from front end chips must be merged to make efficient use of off-detector link bandwidth, and data volumes depend on whether or not pulse height information is retained, ADC resolution, and whether sparsification is employed.

Table 6: Estimated link power contributions for different front end chip (128 channels) architectures

	Link speed Gb/s	No. of chips / link	Power/ link	Link power/ sensor chan.
APV25 non-sparsified analogue	0.36 (eff.)	2 / fibre	60 mW	230 μ W
Digital APV non-sparsified	2.5	32/ GBT	\sim 2 W	490 μ W
Digital APV sparsified	2.5	256/ GBT	\sim 2 W	60 μ W
Binary non-sparsified	2.5	128/ GBT	\sim 2 W	120 μ W

Table 6 shows estimated link power contributions for the different choices of front end chip architectures discussed in section III. For the APV25 LHC non-sparsified analogue case, the data from 2 APV25 chips are transmitted at 40 Ms/s on one fibre. Analogue samples are digitized off-detector with an effective resolution of 9 bits, giving an effective link data rate of 0.36 Gb/s. The link power contribution / sensor channel is less than 10% of the overall front end channel power budget at the LHC (table 3).

The remaining three rows in table 6 deal with candidate front end chip architectures for SLHC, where it is assumed that the off-detector link will be implemented by the GigaBit Transceiver (GBT) currently under development [6]. The GBT available data bandwidth is taken to be 2.56 Gb/s, with a power consumption of 2 W, organized as 32 x 80 Mb/s lanes.

For the non-sparsified digital APV (128 channels) a 6 bit ADC is assumed, giving 77 Mb/s to transmit for a L1 trigger rate of 100 kHz. This theoretically allows 32 chips per GBT. The power/sensor channel of 490 μ W is too large, being approximately the same as the target power consumption of the front end chip itself (at the SLHC). The ratio of 77 Mb/s data rate to 80 Mb/s available link rate leads to a link bandwidth use efficiency factor (ratio of transmitted data rate to maximum available digital data bandwidth) of 96% which would be unfeasible to implement in practice.

For the digital APV with sparsification case a 6-bit ADC and an occupancy of 4% is assumed, leading to 5 hits per 128 channel chip on average. For each hit 13 bits are required (7 bits address + 6 bits pulse height), and a 20 bit header is added to incorporate timestamp (12 bits) and chip identity. This gives an overall average data packet of 85 bits and so a data rate of 8.5 Mb/s at 100 kHz L1 rate. Combining data from 8 chips on an 80 Mb/s GBT lane gives a combined data rate of 68 Mb/s, with a link bandwidth use efficiency of 85%.

For the binary non-sparsified case only 1 bit per hit is transmitted and data volume is occupancy independent. An extra 16 bits per trigger are added to allow for the

transmission of header information, including the address of the triggered pipeline location as is done in the digital header for the present system (figure 2). This leads to a data volume per L1 trigger of 144 bits and a data rate of 14.4 Mb/s at 100 kHz L1 rate. 128 chips per GBT are possible with a comfortable link bandwidth use efficiency of 72%

While the sparsified digital APV architecture gives the least link power contribution per sensor channel, the added power and system complexity associated with merging fluctuating trigger-to-trigger data volumes must be taken into account.

V. TRIGGERING

The overall L1 trigger rate at SLHC cannot be maintained at 100 kHz without transverse momentum (P_T) information from the tracker [1], assuming the same trigger strategy planned for LHC luminosity. Ideas presented here are based on the assumption that there will probably be one or more P_T layers, dedicated to providing information for the L1 trigger decision.

Some concepts which have been previously presented include the stacked tracking approach [7], where P_T discrimination is achieved by correlating hits in closely spaced layers, and cluster width discrimination [8], where high P_T tracks in a single sensor layer can be identified by their narrow cluster width. The concepts are clear, but issues associated with practical implementations (construction details, power consumption, cost) need further understanding.

Figure 8 shows a possible implementation of the stacked tracking approach, for an inner layer at 25 cm radius, which could extend over the full pseudo-rapidity range of the existing tracker. A P_T module with dimensions 25.6 mm x 80 mm is constructed from 2 silicon sensor layers, each tiled with 32 readout chips, each chip instrumenting 256 pixels (2 x 128 columns) with a pitch of 100 μ m and 2.5 mm long. The readout chips could be wire-bonded to the sensor. A correlator chip receives signals from both layers.

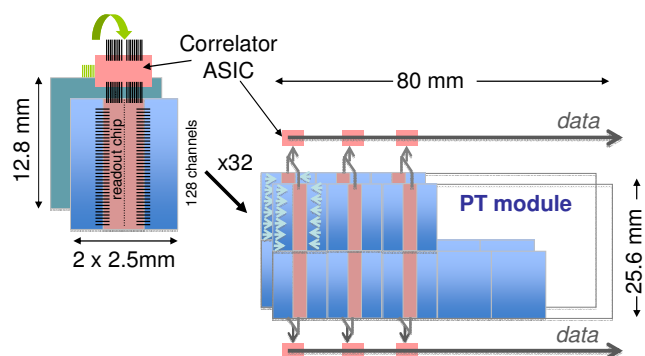


Figure 8: A possible P_T module for a 25 cm radius layer.

Rather than transmitting single hit channel addresses to the correlator chip, more information can be provided by organizing channels into groups, e.g. 32 x 4 per column, so that a hit would consist of the pattern within the group, plus a 5-bit group address. Cluster width discrimination can also be implemented to reduce the number of valid hit patterns. The correlator chip compares the hit pattern and address from both layers (no address decoding is required) and if there is a match then the result is transmitted off-detector.

The P_T module contains 8192 pixels (per layer) with a predicted occupancy of 0.5% at 40 MHz and $10^{35} \text{ cm}^{-2}\text{s}^{-1}$ luminosity. The correlation operation is expected to reduce the hit rate by a factor ~ 20 , giving a “high P_T occupancy” of 0.025% (0.5%/20). Thus the number of positive correlation results per P_T module per bunch crossing should be only two (0.025% x 8192). It is necessary to transmit all the positive correlation results every bunch crossing, and 64 bits are available every 25 ns, for a 2.56 Gb/s off-detector link, so one link can handle data from 2 P_T modules.

Approximately 3000 P_T modules would be required to tile the surface of a 3m cylindrical layer at 25 cm radius (allowing for modules overlapping) so 1500 off-detector links would be required, consuming 3 kW in total for 2 W per link. Extrapolating the readout power for the current pixel system to $50 \mu\text{W}/\text{pixel}$ for the P_T layer pixels gives 8192 pixels/layer x 2 layers/ P_T module x 3000 modules x $50 \mu\text{W} = 2.4 \text{ kW}$. This does not include any extra power for the other digital functionality (correlation operation and short distance digital transmission), so it can be seen that a P_T layer implemented in this way would be a high power layer.

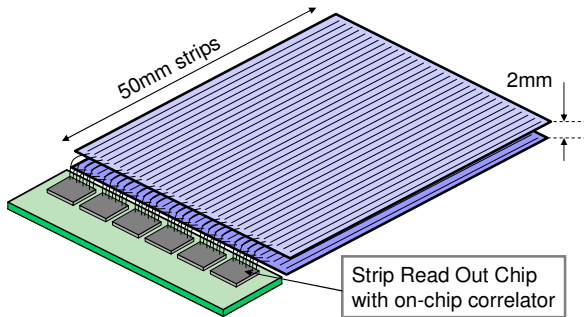


Figure 9: A possible P_T module for an outer layer [9].

Figure 9 shows another possible idea for implementing a stacked tracking P_T layer module [9], this time for an outer tracker layer. For the lower outer layer occupancy, the module could be implemented using relatively long strips with readout chips wire-bonded at the end. The novel idea in this case is to bond neighbouring chip channels alternately to strips in the upper and lower sensor planes, such that correlations between planes can be made on the same readout chip, between neighbouring channels, and no extra correlator chip is required. It is likely that this would save power and material, and costs could be minimised if DC coupled sensors could be used. The correlation operation, which also uses cluster width discrimination so that the correlation between planes is only performed for high P_T tracks, can be implemented with relatively simple logic.

The two triggering layer ideas presented, figures 8 and 9, are included to show recent progress in this area. More work is needed to understand details of performance and practical implementation, and alternative ideas may be preferred or develop in the future.

VI. SUMMARY

The CMS silicon strip readout architecture for SLHC is not yet defined, and major challenges of power consumption and provision of information to the first level trigger decision must be confronted. The pros and cons of different front end

chip architectures are under consideration, involving compromises between power consumption, front end chip and system complexity, system robustness and performance.

A readout chip development programme will begin soon, beginning with front end test structures matched to different sensor options (polarity, strip length, DC coupling), progressing to a full chip prototype in the second year, when decisions will be needed on system issues (e.g. binary/sparsified analogue architecture, powering scheme (serial/parallel), and sensor choices).

It seems likely that a binary, non-sparsified architecture could lead to minimum front end chip power consumption, while also retaining some of the valuable features of the existing LHC system, but the disadvantages of abandoning the pulse height information must also be considered.

A simpler strip readout chip and system architecture would require less resources to develop, and resources will be needed to confront the triggering issues, where ideas are still evolving. It is clear that there will be dedicated chip developments required in this area.

VII. ACKNOWLEDGEMENTS

We would like to thank the UK Science and Technology Facilities Council for supporting this work. Thanks also to R.Horisberger and W.Erdmann for permission to show their outer layer P_T module design (figure 9).

VIII. REFERENCES

- [1] J.Jones et al, A Pixel Detector for Level-1 Triggering at SLHC, 11th LECC Workshop (2005), Heidelberg, Germany, CERN-LHCC-2005-038, p.130.
- [2] M.Raymond et al, The CMS Tracker APV25 $0.25 \mu\text{m}$ CMOS Readout Chip, 6th LEB Workshop (2000), Cracow, Poland, CERN/LHCC/2000-041, p.130.
- [3] J.Coughlan et al, The CMS Tracker Front-End Driver, 9th LECC Workshop (2003), Amsterdam, the Netherlands, CERN-LHCC-2003-055, p.255.
- [4] G.Iles et al, The APVE Emulator to Prevent Front-end Buffer Overflows within the CMS Silicon Strip Tracker, 8th LECC Workshop (2002), CERN-LHCC-2002-34, p.396.
- [5] International Technology Roadmap for Semiconductors 2003 edition, System Drivers section, page 18, table 11: <http://www.itrs.net/Links/2003ITRS/SysDrivers2003.pdf>
- [6] P.Moreira et al, The GBT, a Proposed Architecture for Multi-Gb/s Data Transmission in High Energy Physics, TWEPP-07, CERN-2007-007, p.332.
- [7] J.Jones et al, Stacked Tracking for CMS at Super-LHC, 12th LHC Workshop (2006), CERN-LHCC-2007-006, p.130.
- [8] G. Barbagli et al, Track Momentum Discrimination Using Cluster Width in Silicon Strip Sensors for SLHC, TWEPP-07, CERN-2007-007, p.482.
- [9] R.Horisberger and W.Erdmann, Proposal for a short strip based P_T – trigger or stereo module design for TOB, talk given at CMS SLHC upgrade meeting, 8th July, 2008: <http://indico.cern.ch/getFile.py/access?contribId=3&sessionId=0&resId=0&materialId=0&confId=36580>