The MP7 and CTP-6: Multi-hundred Gbps processing boards for Calorimeter Trigger Upgrades at CMS

Dr. Andrew W. Rose[†]
on behalf of
Imperial College, London[†] and University of Wisconsin, Madison

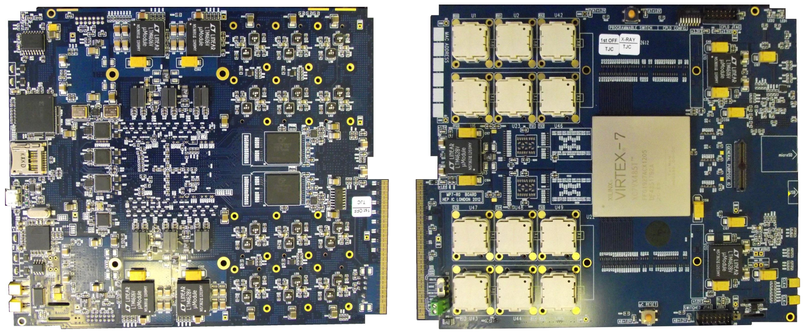


Figure 1: Top (right) and Bottom (left) surfaces of the Imperial MP7. The central component of the MP7 is a single, mid-range Xilinx Virtex-7 FPGA, being fed by six MiniPOD optical receiver modules and feeding six MiniPOD optical transmitter modules, providing a total optical bandwidth of approximately 740GB/s in each direction. The optics are not mounted in these images.

IMPERIAL MP7

In contrast to the typical design methodology employed in particle physics, the Imperial MP7 was not designed to fulfil a specific role, but rather, to be a generic stream-processing engine. This design philosophy manifests itself most clearly in the distinctive data interface, which is (a) all-optical and (b) has symmetric transmit and receive bandwidths.

FPGA

The MP7's processing capability is provided by a Xilinx Virtex-7 FPGA in a FFG1927 package. Xilinx provide four pin-compatible parts in this package with varying amounts of logic and numbers of serial links; the VX415T, VX485T, VX550T and the VX690T. The VX550T and VX690T parts both provide a full complement of 72 optical links in each direction, with the latter part simply offering more logic resources than the former. The MP7 has been designed such that, when using the VX485T part, the 48 optical links available in each direction are all routed to the rear 8 MiniPODs, so that system design is simplified, and a cost saving made, by not having to deal with partially-used MiniPODs.

Optical Interface

Using a single interface for the sending and receiving of all data offers several distinct advantages: Primarily, as stated above, the board ceases to have a specific role and becomes a truly generic stream-processing engine. Application of the board to a specific task is no longer restricted by the bandwidth and compatibility of each type of interface, but only by the total bandwidth, whilst all specialization required for a task is contained within the programming of the board and the interconnections between boards.

The MP7's optical interface is provided by up to six Avago MiniPOD transmitters and six Avago MiniPOD receivers. Each MiniPOD device provides 12 optical links running at up to 10.3Gbps, giving the MP7 a total optical bandwidth of up to 740Gbps in each direction.

PCB

The MP7 has 144 differential-pairs running at 10Gbps and many more running at the order of 1-2Gbps. To ensure signal integrity in the high-speed traces, a ground-plane has been placed between each layer containing high-speed traces, and the presence of these ground planes, necessitated the use a 16-layer stack-up. The number of layers, coupled with the 1.6mm board thickness specified by the µTCA standard, and the sheer number of high-speed traces required the use of NELCO N4000-SI-EP as a board substrate, due to its excellent high-frequency performance and highly-desirable impedance properties.

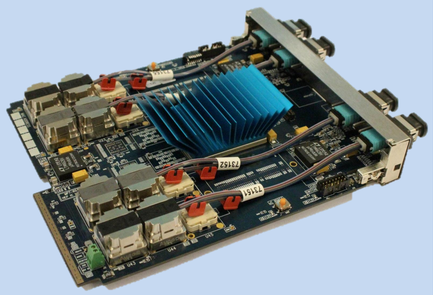


Figure 3: The MP7 card with MiniPOD optics, optical fibres, front panel and heatsink mounted. This configuration of the optics can transmit and receive nearly 0.5Tbps.

Testing of the MP7

The main source of risk in the design of the MP7 was the sheer number and high speed of the optical links and, as such, the main test of the MP7 is the integrity of the optical links. The links have been validated by loop-back test, both as two sets of 24 links running PRBS-31, and all 48 links simultaneously running 8B10B. Running the links at 10Gbps the MP7 transfers 0.48Tb per second in each direction; that is, approximately 30Tb per minute, 1.7Pb per hour, 41.5Pb per day or 0.29Eb per week. To date, the MP7s has transferred in excess of an Exabit of data without an observed error, giving an estimated per-board bit-error rate of approximately 3×10^{-17}.

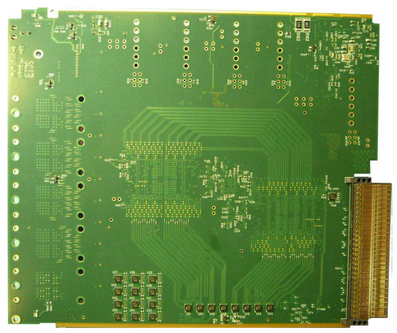
MP7	Card		CTP-6
Virtex 7VX690T	Processor		2× Virtex 6HX250T
690k	Logic Cell Count		2× 252k
72	Input Optical Links	Number	48
10.3 Gbps		Speed	6.5 Gbps
740 Gbps		Bandwidth	312 Gbps
72	Output Optical Links	Number	12
10.3 Gbps		Speed	6.5 Gbps
740 Gbps		Bandwidth	78 Gbps
Ethernet, PCIe×4, TTC, DAQ, SATA/SAS 11× 1.8Gbps LVDS links in 11× 1.8Gbps LVDS links out	Backplane connectivity		Ethernet, PCIe×4, TTC, DAQ, Fat-pipe 12× 6.4 Gbps serial links in 12× 6.4 Gbps serial links out
30× 1.8Gbps link Samtec header	Other electrical connectivity		-
(Optional up to) 2× 144Mb 550MHz QDR II+ SRAM	RAM		2× 2Gb DDR 1333 SDRAM
• 32GB µSD card for multiple firmware revisions and configuration data • USB serial console	Other features		• USB serial console & JTAG through daughter cards • Power supplies on custom designed daughter cards

To maintain or improve the physics performance of the CMS calorimeter trigger under full-energy, high-luminosity conditions, the existing mixed ASIC/FPGA-based will need to be replaced. Two trigger architectures have been proposed, one by each group that built the current calorimeter trigger.

The University of Wisconsin, Madison, favours a conventional trigger architecture, requiring backplane interconnections to handle the boundaries between regions. The CTP-6 is a prototype of the processor card needed for this architecture.

Imperial College, London, favours a time-multiplexed, spatially-pipelined trigger architecture, which requires no backplane interconnection and processes all data from an event in a single FPGA, but which requires a larger optical bandwidth than a conventional trigger. The MP7 is an all-optical, data-stream processor of the final design needed for such a trigger.

A collaboration has been formed between Imperial College and University of Wisconsin to build a system which could be used as either trigger architecture.



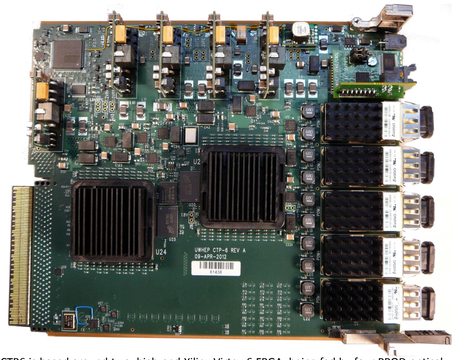


Figure 2: Top (right) and Bottom (left) surfaces of the Wisconsin CTP6. The CTP6 is based around two, high-end Xilinx Virtex-6 FPGA, being fed by four PPOD optical receiver modules and feeding one PPOD optical transmitter module, providing a total optical bandwidth of approximately 310GB/s in and 78GB/s out. The board also has a further 12 serial links going to and coming from the backplane.

CTP-6

The CTP-6 is a prototype Calorimeter Trigger Processor Card by the University of Wisconsin. It is an AMC card with two Virtex-6 (XC6VHX250T) FPGAs with GTX links, dual SDRAM for dedicated DAQ and TCP/IP buffering, a Module Management Controller (MMC) of the Wisconsin design, and custom-designed power modules providing 1.0V, 1.5V and 2.5V.

Serial Links

On the front panel, the CTP-6 uses 4 Avago AFBR-820B Receiver Modules, and one AFBR-810B Transmitter module to provide 48 optical inputs and 12 outputs running at up to 6.4Gbps. There are 24 inter-FPGA links, also running at 6.4Gbps. The trigger-architecture favoured by Wisconsin also requires back-plane communication. The CTP-6 has 12 bidirectional links to the backplane which can be run at up to 4.8 Gbps.

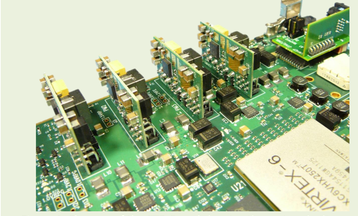


Figure 4: Close up of the CTP-6 card showing the custom-built power modules. The daughter card (top right) and module mounted on the daughter card provide a USB serial-console to the MMC and a JTAG interface.

Custom backplane

A backplane has been manufactured by Vadatech (VT894) to a design specified by the University of Wisconsin. This backplane has the same standard connections as the VadaTech VT892 backplane, but adds 88 differential pairs according for the data sharing requirements of the Wisconsin trigger architecture. This is done using the AMC connector ports which are unconnected on the VT892. The additional connection use the same link technology as the standard backplane connections. Figure 5 shows the connection topology. Slots 2-5 and 8-11have bi-directional links in a 2x4 array for handling boundaries within a processing region. Slots 6 and 7 have links to the upper or lower ranks of the processing array and, in the Wisconsin trigger architecture are used for collating data for inter-crate communication. For completeness, the remaining links are routed to slots 1 and 12, but these links are unnecessary for the Wisconsin trigger architecture.

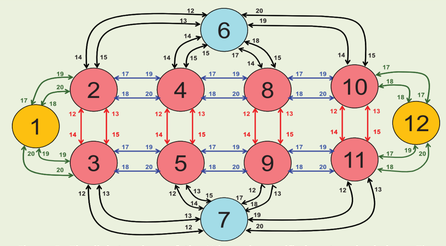


Figure 5: The link topology of the Wisconsin-specified custom backplane. The numbered circles represent backplane slots and the arrows indicate the custom connections between those slots, with the numbers next to the arrows indicating through which AMC-connector fabric pair the connection is made.

Testing the CTP-6

Testing of the CTP-6 is underway at the University of Wisconsin. The parallel FPGA flash memories allow both FPGAs to load in approximately 6 seconds. An SI5388 programmable clock synthesizer allows multiple options for the reference clocks of the GTX links. Gigabit Ethernet is successfully running on a Microblaze embedded processor on the back-end FPGA. The MMC is working with phased power-supply enables, analogue voltage-sensors, and sensors for "power good" and "FPGA load done". The inter-FPGA serial links are operating at 6.4 Gbps. The Wisconsin trigger design also requires data transfer through the standard µTCA fat-pipe. This has been tested, routing though a cross-point tongue (based on a Vitesse VSC3172 72x72-way switch) mounted on a VadaTech MCH, and is working at 4.8 Gbps.

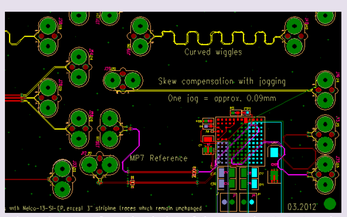


Figure 6: Extract from the design of the second MP7 test card. Of particular interest are the traces running under an LTM4606 low-EMI switch-mode regulator. The trace here are far more extreme than those used on the MP7 and were meant to stress the high-speed traces far beyond any real-world use-case. Other test structures can also be seen – a comparison between softly curved traces and those using 45° bends, traces with jogging for skew-compensation and a structure to test the effects of very closely packed serial links.

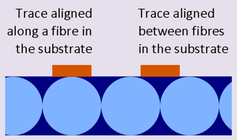


Figure 7: Figure demonstrating a trace aligned along the fibres in the substrate and one aligned between the fibres in the substrate. The different dielectric constant of the fibre and the epoxy result in different impedances for each member of the differential pair and an infra-pair skew. On a 6 inch trace, a skew of 50ps was observed, rendering a 10Gbps unusable. By rotating the trace to an angle of 22.5° relative to the weave, the dielectric constant is averaged over epoxy and fibre and the source of skew removed.

Test boards

The unprecedented (in our experience) high speed and sheer density of the signal traces for the optical links on the Imperial MP7 board led us to believe it prudent to develop a test board for the purpose of examining the effects of various design factors on the integrity of 10Gbps signals. Of particular interest were the effects of stripline and micro-stripline structures, layer transitions using both vias and blind vias, and stubs associated with both full and blind vias.

The most unexpected result from the initial test board was a 50ps skew between members of a differential pair (figure 8a), far greater than that expected from the trace length-matching tolerance, rendering a 10Gbps signal unusable. The cause of the skew was determined to be the different impedance seen be traces aligned along the fibres in the substrate from that seen by trace aligned between the fibres in the substrate (figure 7). The suggested solution for this problem was to have the PCB manufactured at an angle of 22.5° relative to the weave of the substrate, so that no trace, whether at 0°, 45°, 90° or 135°, could run parallel to the fibres and, by so doing, the impedance seen by all traces would average out to approximately the same value. To test this solution, the second test board was manufactured using this technique, and the trace-trace skew was seen to vanish (figure 8b).

During layout of the MP7, it became apparent that to meet both the tight spatial constraints on the board and the strict conditions on the Virtex-7's multi-gigabit transceiver power supplies, some of the switch-mode power supply modules would need to be located directly beneath the high-speed traces. Concerns about the risk to the high-speed traces of EMI from the power modules prompted the inclusion on the second test board of a high-speed trace running under the chosen regulator to test for such effects (figure 6). Although the test structure was a considerably harsher design than that on the MP7, being longer, having more bends, having more passes under the power module and having fewer power-planes between the regulator and the high-speed traces, no more significant degradation of the signal was seen than was expected for a trace of that length.

As well as directly testing various trace structures, the test boards were also used to test the accuracy of the Mentor Graphics HyperLynx PCB analysis software without full EM simulation. It was found that, overall, the software gave very accurate predictions, with one exception: because the software uses a transfer function approximation of vias, rather than a full EM simulation, these approximations would, in some cases, give inaccurate results. By combining the results from the test boards with the predictions made by the simulation software, a good understanding of the 10Gbps traces was gained prior to the manufacture of the MP7, and the performance of the final board is as good as, or better than, expected.

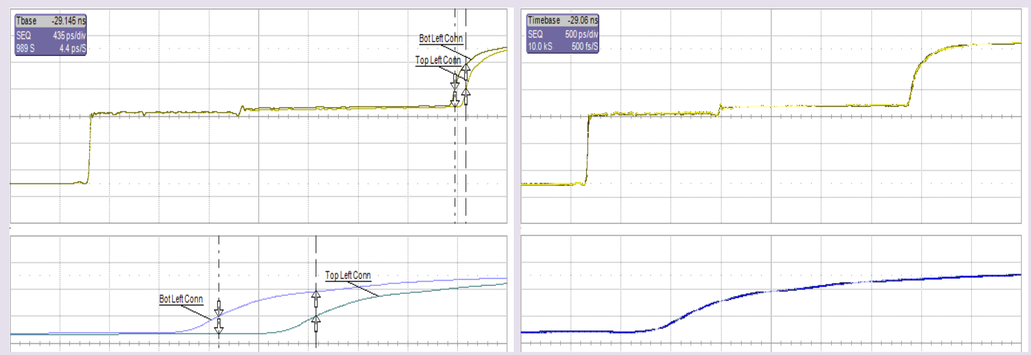


Figure 8 : The trace-trace skew measured on the MP7 test board. The left-hand plot shows the 50ps infra-pair skew measured on the first test board, where the traces were aligned with the weave of the substrate. The right-hand plot shows that, by rotating the traces to an angle of 22.5° relative to the weave, the second test board did not suffer from the same infra-pair skew.