Readout and DAQ after LS3, possibilities and limits

THEATRE OF DREAMS: LHCB BEYOND THE PHASE 1 UPGRADE,
April 7th 2016, Manchester

Niko Neufeld, CERN/EP (with a lot of help & input from Ken Wyllie)

Before we start

- Don’t believe anybody who claims to be able to predict electronics and computing beyond 6 years from now
- I am going to give you some guesses based on what is known, and on R&D programmes which have started already
- Since the decisions on electronics have to be taken much earlier than on computing, the landscape for future Front- und Back-end electronics is a bit more clear →
  - This is IMHO the most relevant part today and in this one I follow very closely *Ken Wyllie* (although misunderstandings are of course my fault)

Acknowledgments and Apologies

- Thanks to Ken Wyllie, Marco Cattaneo for comments on the slides
- Special thanks to Ken for all the material in the electronics part
- And to the online-team (Beat, Rainer, etc…) for discussions
- And apologies for this rather dry, almost picture-free presentation

Executive summary (Online)

- We (the Online) strongly believe that the Run3 system will work – the same system will be fine for Run 4, 5 and 6
- No reason to (re-)introduce a trigger to save the DAQ
  - Acceleration / Assists with CPUs, GPGPUs, FPGAs or whatever should be done *after* event-building using COTS hardware
- For Run5
  - Continue with full event-building at 40 MHz
  - Output rate (to storage) only limited by what we can afford to store long-term
  - Any increase in number of channels & occupancy can be (relatively) easily absorbed with more and/or faster links

Electronics

Electronics

- Even with another factor 5 to 10 in luminosity will leave most (all) of our FEE considerably below of the (dose) rates anticipated for the GPDs
- So what works for them will certainly work for us

What's going on in the electronics community that could be used in LHCb upgrade^2?

- The main R&D is the massive amount of work targeting LS3 for ATLAS/CMS **phase 2** upgrades (also a lot of work for ATLAS/CMS **phase 1** upgrades, but these are for LS2 (LHCb upgrade #1) so will be already obsolete for LHCb upgrade^2)
- There are many developments that could prove useful for front-end electronics on the detectors, depending on what is envisaged for LHCb upgrade^2. Also remember that much (all) of the infrastructure will be obsolete and have to be replaced (power supplies, optical fibres, monitoring etc…)
  - Continued widespread use of optical links (huge numbers in ATLAS & CMS trackers).
  - Mainstream is again to use common projects driven by CERN:
    - lpGBT (10Gb up-link, 2.56Gb down-link)
    - Versatile Link+ (4/8-way TX array)
    - 1.28 Gb electrical links developed for FE chips to connect to lpGBT.

ASIC technology

- 65 nm will be mainstream (in HEP!): has clear density benefits for smaller pixels with more functionality (eg timing), but radiation tolerance is not obvious and requires careful thought & mitigation.
- 130 nm will continue for some time…… (but will be VERY old at time of LHCb upgrade #2).
- 0.35 um currently used by Calo/RICH will be VERY VERY old at time of LHCb upgrade #2 (obsolete?)
- Attempts being made to share blocks more coherently (IP sharing).

ASIC technology 2: time measurement

- 3 ps resolution TDC targeted on 64-ch chip (looks very nice for TORCH – are the TORCH people actively discussing with this project…?)
  - But so far no proposal for new amplifier/discriminator chip (like NINO)
- In-pixel timing already demonstrated with Timepix3 (1.56ns bins). More precision certainly possible.
  - Being discussed for helping with pile-up in ATLAS/CMS.

FPGAs

- ESA/ESTEC have funded design of a radiation-tolerant FPGA, but only to space levels of radiation → so not an obvious advantage over commercial FPGAs
- Commercial vendors will continue to improve mitigation of SEUs in configuration memory. Unfortunately, it's not clear that they will improve radiation performance of IP blocks like PLLs used for serial links.

Track-trigger and all that

- ATLAS is already using a track-trigger, using back-end cards using associated-memory ASICs (very difficult to emulate in software)
  - This adds track-triggering and -information at Level-2 (first level software trigger in ATLAS)
- CMS tracker phase-2 upgrade will do more in front-end (pixel-strip & strip-strip modules giving 'stubs' for triggering)
  - This is trigger is part of (the improved) Level-1 (corresponds to Level-0 in current LHCb-speak)
  - This is a project driving many interconnect/module developments (flexes, TSVs……) for communicating hit info between tracker layers.
- High-granularity calorimetry:
  - Lots of developments for ILC and now huge project for CMS end-cap based on Si pads.
  - Big challenges for PCB & module construction, and requires low-power ADC & 50ps timing

The (not so exciting but) important rest

- **Power:**
  - Serial powering unbeatable for low-mass stave-based detectors, will be used for CMS pixels -> future UT?
  - New DC-DC convertors in 130 nm
  - New bulk power supplies for long-distance powering
- **Monitoring:**
  - First discussions on new ELMB based on GBT-SCA. No details yet. Do we need this?

Computing, Networking

Moore's law

- Moore's law... holds at the time of this writing: the number of transistors *at equal cost* still doubles every x months
  - Intel has recently increased x from 24 to 30
- After 14 nm today there will be 10 nm, 7 nm, 5 nm, … there will be probably an end to this shrinking (at least for CMOS in Si) by the time of LS4
  - Does not necessarily mean that Moore's law in the sense above ends at the same time!
- In many commercially relevant areas this progress is invested not in more cores but in lower power consumption

Intel on Moore's law

"It's an unwritten law in engineering that every generation thinks their challenges are the most difficult. Although new technical challenges continue to emerge, as they have every generation since the first VLSI chips were created, the outlook for Moore's Law remains the same as it did twenty years ago; the path for the next few generations is visible, and after that, it gets hazy until we move forward."

Mark Stettler (Director of Process Technology Modeling), Intel

What drives computing

- In the consumer area: portable devices and wear-ables --> low power
- In the server area: hyper-scale cloud providers: Google, Amazon, Facebook, Microsoft → scale, power consumption, high network bandwidth

Heterogeneous computing

- There seems to be a slowing down of the number of "cache-coherent" cores → most codes (not only in HEP) cannot really use a coherency domain with more than 32 cores
  - Does not necessarily mean less cores → Cavium ARM v8 server has 96 real cores (on two sockets)
- The ever increasing number of transistors is invested in co-processors, interconnects, assists to accelerate fixed functions (where it makes sense) and integrated FPGAs
  - E.g. motion co-processor in Apple's A series for the i-devices or Intel's integration of the high-speed interconnect in the CPU
- Commercial clouds will offer select heterogeneous compute elements. Current best candidate: FPGAs → we should

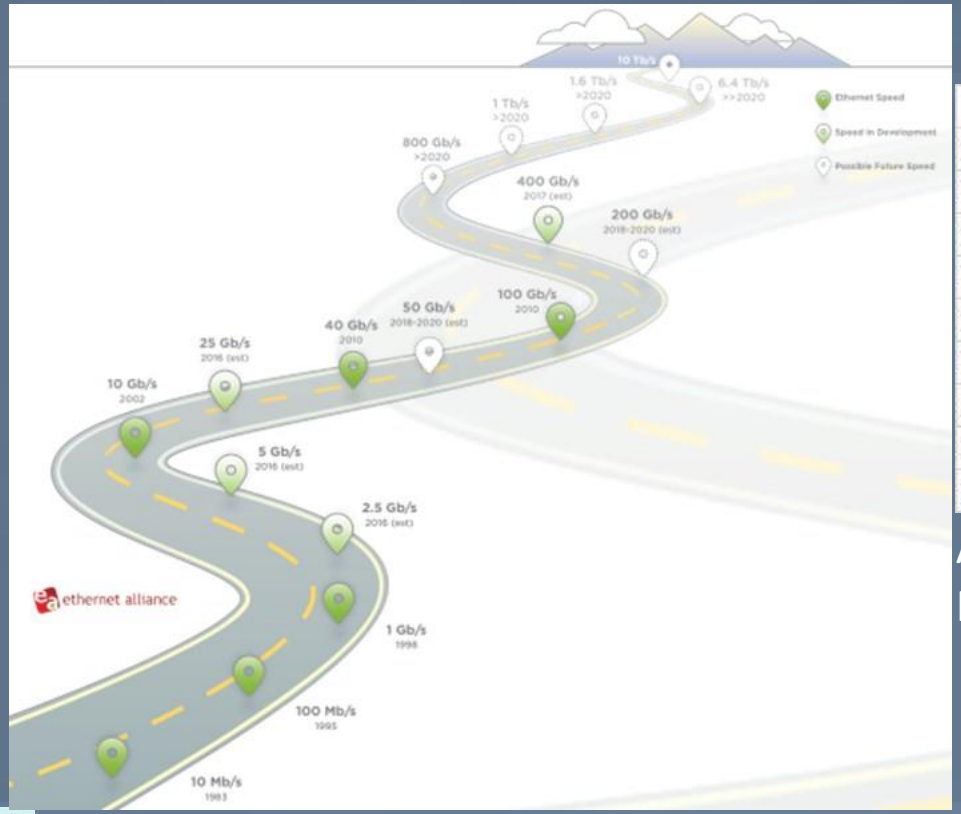
Multi-architecture

- The current server-world is x86 only
- It could be that there are more architectures
  - Impossible to say today if and which, but we should (already today!) be ready to run on something else
- Crucial to test for reproducibility between architectures and with and without accelerators

FPGAs, GPGPUs and all that

- There are many technologies being proposed already today. Time will tell which one is the most successful one
- We should use it in a commercially available form → do *not* build our own boards (e.g. if we want to use FPGAs)
- Should always look at the overall cost-effectiveness → multi-purpose has almost always the lowest Total Cost of Ownership (FPGAs over ASICs)
- Power-cost will be and should be a concern (FPGA over GPGPUs…. ASIC over FPGA)
- When choosing if and which accelerator to use in HLT or offline we should orient ourselves on industry == the commercial cloud providers, because this gives confidence in reproducibility, good software support and portability

The Ethernet road-map



Device Family	Maximum SerDes Bit Rates (Gbps)
Virtex UltraScale+	32.75
Kintex UltraScale+	32.75, 16.3
Virtex UltraScale	30.5, 16.3
Kintex UltraScale	16.3
Virtex-7	28.05, 13.1, 12.5
Kintex-7	12.5
Artix-7	6.6
Zynq Z-7000	12.5, 10.3125, 6.6, 6.25
Spartan-6 LXT	3.125

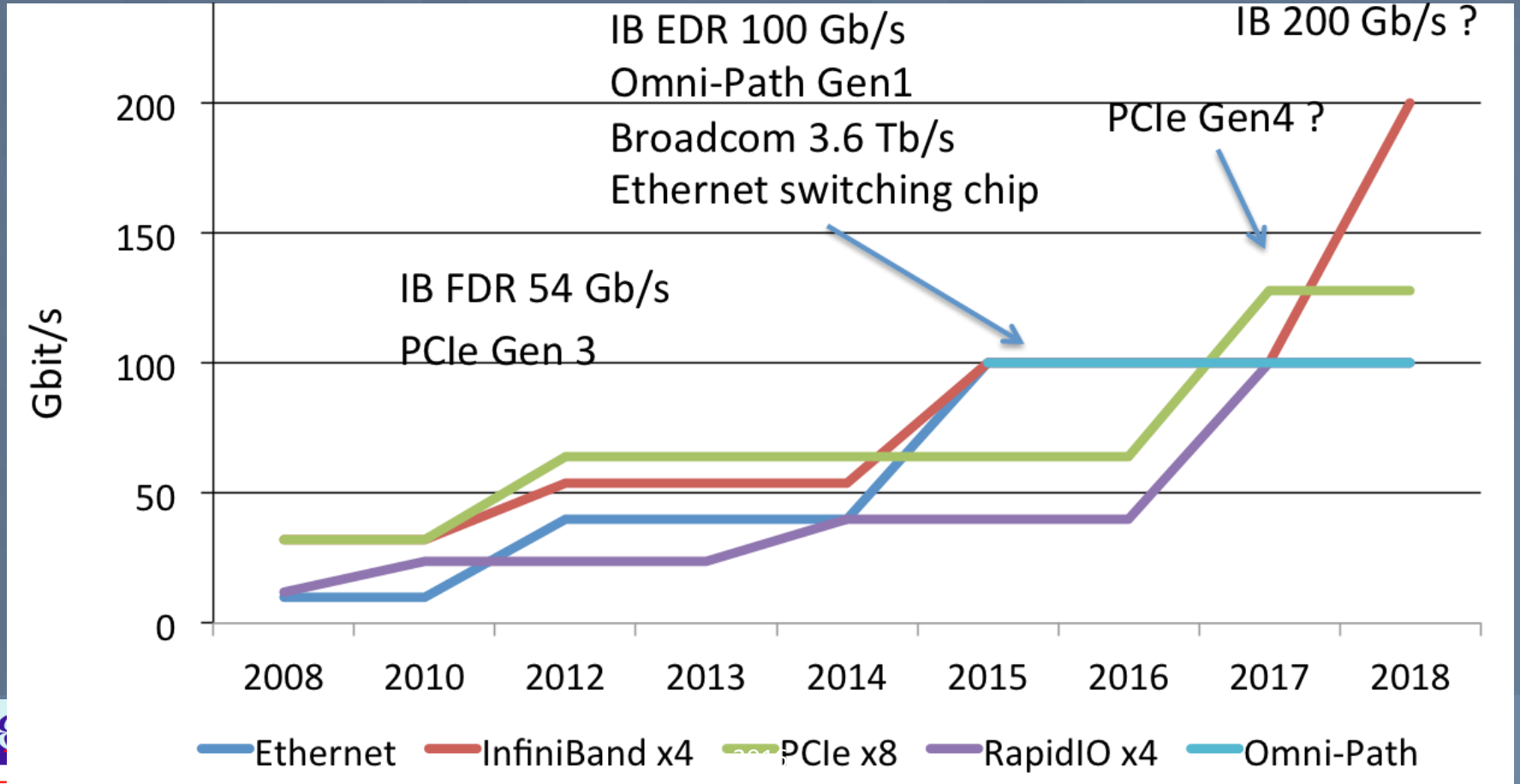
All available today

Network speed is typically 4x the SerDes

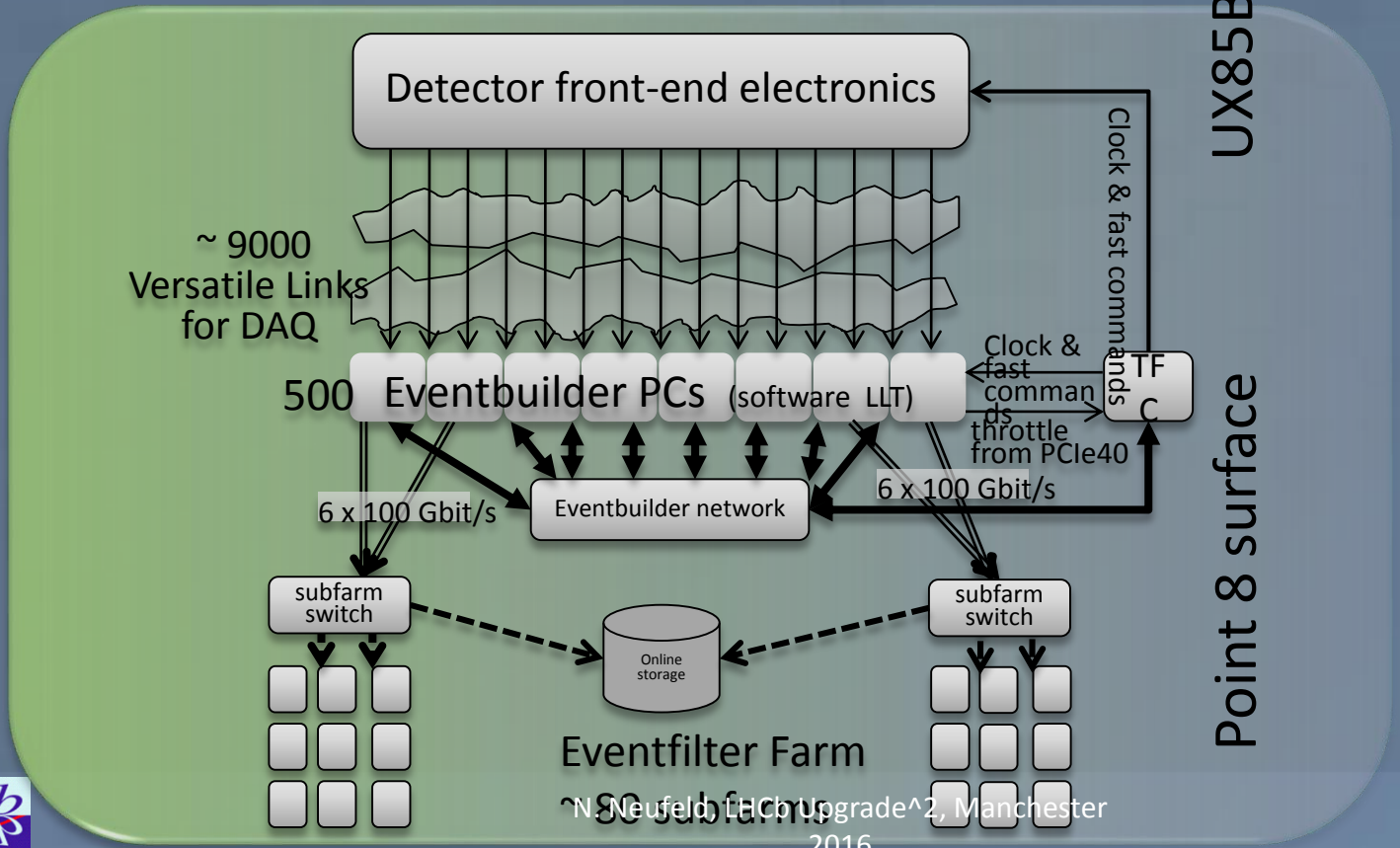
Networking

- Video drives the global internet (Netflix…)
  - It is safe to assume that this will continue
- Bandwidth required for the DAQ should be extremely cheap by LS4, even if we add a factor 5 or so
- With a very modest data-reduction (factor 3, 2 or even 1) it should be possible, and probably commercially reasonable, to send assembled data off-site → could use commercial or scientific (private) clouds instead of running our own HLT farm
  - Trade in CAPEX for OPEX: what do the funding agencies prefer?

Network evolution



The Online system for Run #3



Scaling the DAQ up for Run4

- The DAQ is designed to easily scalable
  - Just add Eventbuilder servers with PCIe40 cards
- By LS3 we will have:
  - Lp-GBT (10 Gbit/s), PCIe40 Gen 2+ > 200 Gbit/s,
  - Cheap 400 Gbit/s links in DAQ network
- Cost of eventbuilder should have come down by at least 50% so we could double the 2020 system at equal cost, part of this cost will be covered by natural refresh (servers after 6 – 7 years, network equipment after 8 – 10 years)

Grid and offline computing

- The current LHC computing grid model does not make much economical sense – it is a political wish
- Commercial clouds (hyper-scale providers) will drive the server market and the smaller we are relative to these the less economical it will be to run our own infrastructure
- My personal guess: Tier2 and Tier3 centres wil have disappearing by end of Run4, Tier1s to be seen , storage is still expensive in cloud and as of today there is no commercial driver for massive storage

Summary

- For electronics a lot will be done for/with us by ATLAS and CMS, need to carefully justify own developments over adaptations where necessary, and in case start not too late!
- For DAQ there is no problem (up to a factor 10 at least), in particular no problem to integrate a 25-50% increase for Run4
- HLT resources could be (partially) off-site → depends on the data-volume
- Computing will be more heterogeneous in the next few years, difficult to know who will be the “winners” at Run5
- Accelerators look attractive today to save power and potentially also cost. Watch out for COTS and TCO