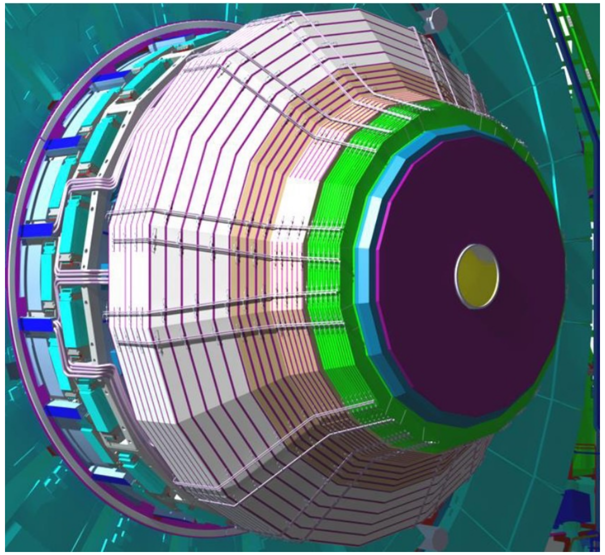
CMS HGCAL Backend: UK contributions and experience

Paul Dauncey, Imperial College London
On behalf of the HGCAL BE group

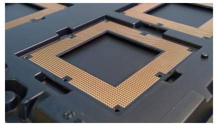
Overview



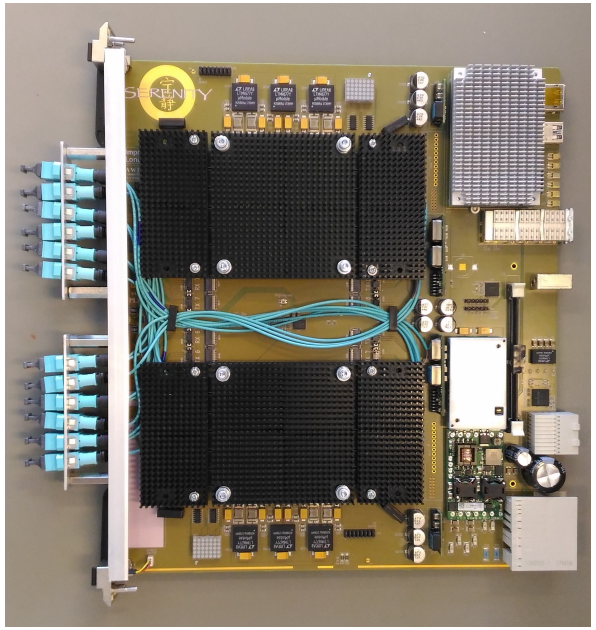
- The CMS HGCAL is the first full-scale high granularity calorimeter which will run at a collider
  - This will provide very useful experience for similar calorimeters in future detectors
- I strongly recommend Dave Barney's talk given in EFCA TF6 meeting
  - Discusses a lot of the issues found during HGCAL construction
  - Folded in with calorimeter experience from construction of original LHC detectors
  - Concentrates on HGCAL detector itself, not off-detector (backend) electronics
  - A link to Dave's talk is on today's Indico page
- Here I will talk about the backend electronics
  - Give some idea of what the UK is doing
  - BE electronics is bound to change substantially for any future collider detector
  - However, some of the unexpected gotchas we have had to deal with might be similar
- Many of the issues arise due to the large number of channels
  - Numbers: DAQ has ~5M cells, trigger has ~1M cells
  - Cost is a very major constraint for anything "per cell"
  - Any future high granularity calorimeter is likely to share some of the same problems

Aimed for uniform BE platform across CMS: Serenity

- Collaboration of institutes from six countries
  - Led by the UK since its inception
  - Serenity boards will used in at least five of the CMS upgrade subsystems, including HGCAL
- Two fundamental ATCA board designs
  - Single very large FPGA (“A” board)
  - Two large FPGAs (“Z” board)
  - Both with high I/O optical bandwidth
- “Z” board has daughterboards
  - Flexibility in choice of FPGAs

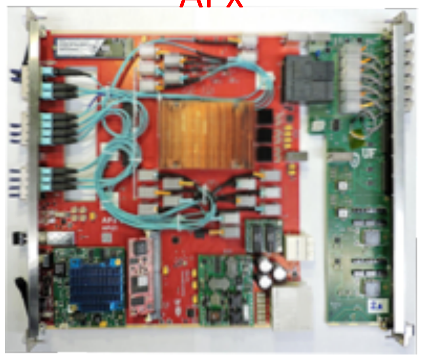


Serenity “A”



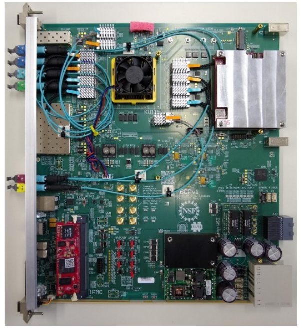
Serenity “Z”

APx



Ocean/X2O

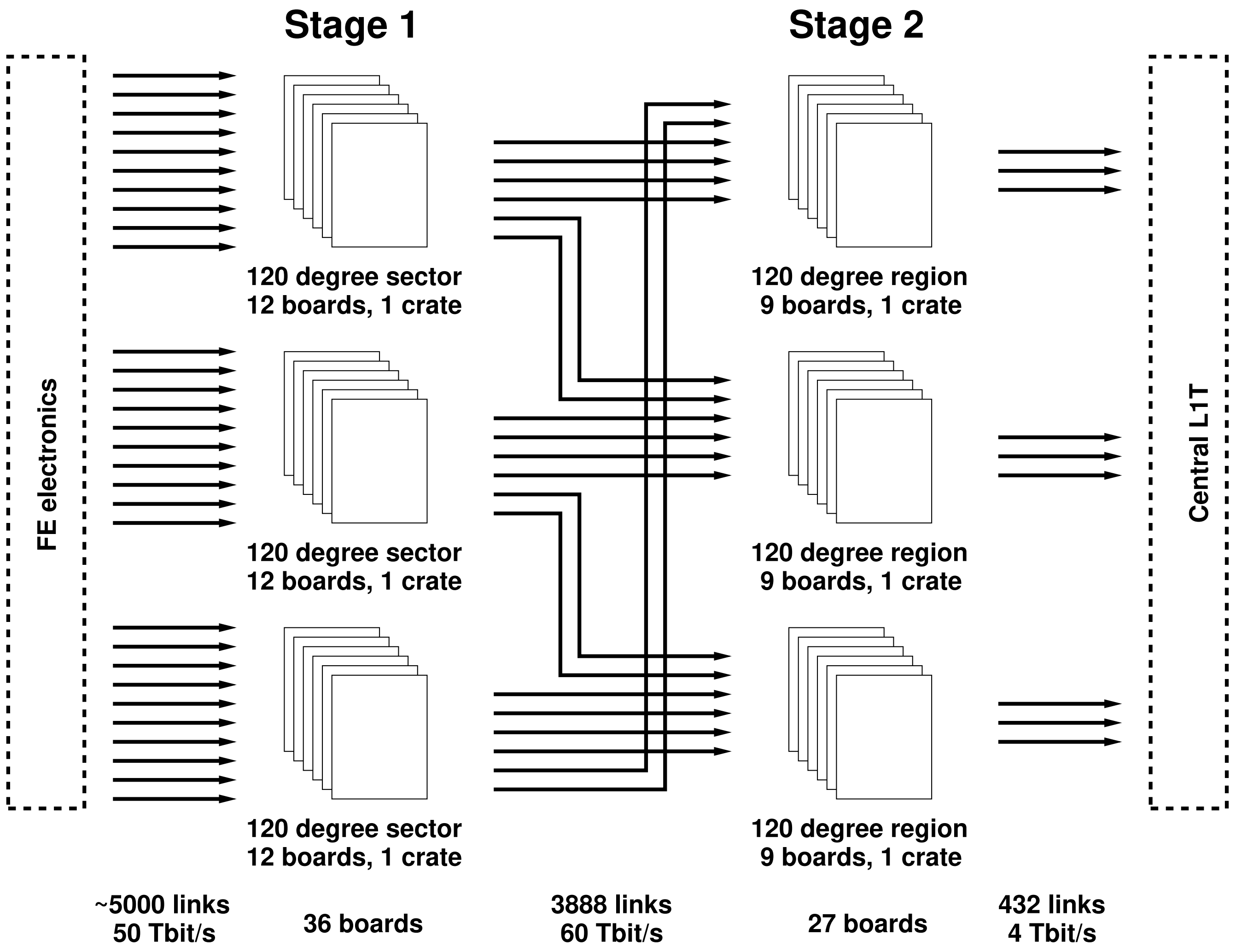
BCP



- Did not manage to include every CMS subsystem
  - Three other boards, targeted for specific applications
  - All conceptually very similar to Serenity “A”
  - Various technical/political/funding/effort reasons for separate designs

UK contribution to BE electronics

- UK co-coordinator of control, DAQ and trigger
  - Maintain commonality for all BE systems in HGCAL
- UK CORE funding contribution to trigger only
  - Collaboration with four countries
  - Large system: ~100 Tbit/s input, ~250 Ultrascale+ FPGAs
  - DAQ system is approximately the same order in terms of scale and rates
- Currently testing last round of prototypes
  - Pre-production (i.e. final design except for bugs) due in second half of 2022
- Discuss some of the biggest issues seen so far
  - But clearly there could be more to come…



Trigger baseline h/w architecture for one of two identical endcaps

Issue: fast-moving market

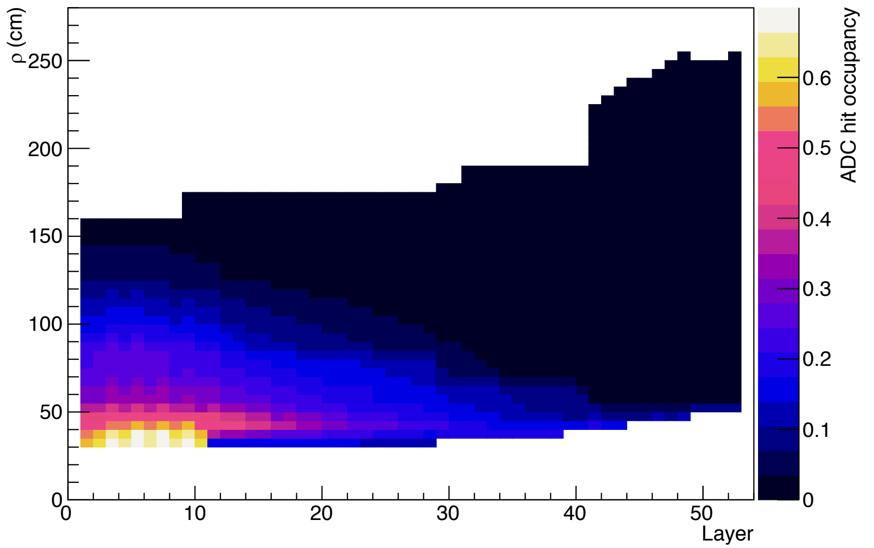
- HGCAL BE electronics is about funnelling very high data rates through big processors
  - The technologies for both the links and the processing nodes are always evolving
  - HEP has ~zero influence on the direction of change so we have to follow the commercial lead
  - Dave Newbold: “Look inside your phone to see the future”
- This has pros and cons
  - There can be large cost savings if you can choose a mass-production COTS component
  - But you must retain flexibility to be able to accommodate this

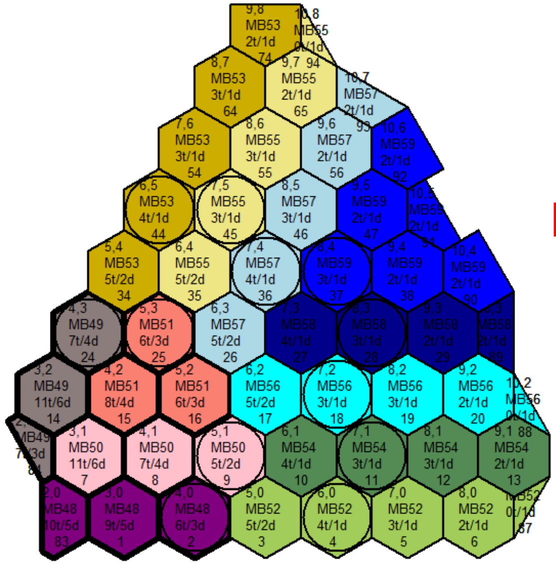
Optocore AOC

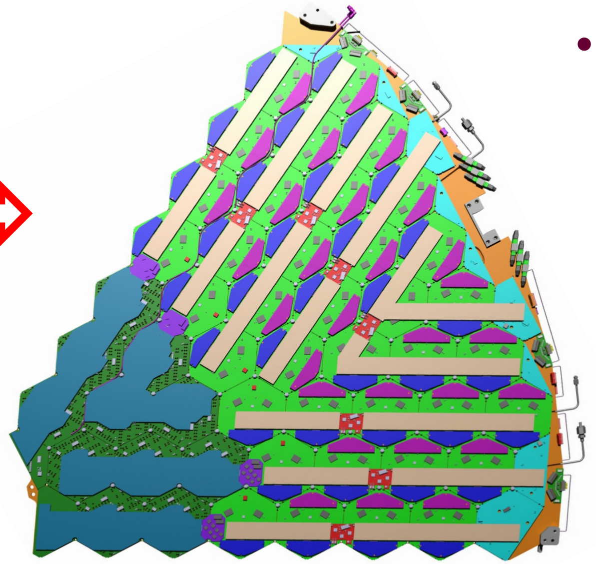
- Example: Serenity boards are designed with Samtec Firefly link connectors
  - Very small footprint so allow a high I/O per FPGA on ATCA boards; essential to achieve cost constraints
  - The Firefly components are unlikely to be available longer-term so must buy all spares up front
- But other low latency point-to-point 25 Gbit/s parts are available
  - Approx half the cost but similar bit error rates
  - Footprint is bigger so cannot simply re-layout the connector areas; would need to ditch ATCA format
  - But CMS central timing/DAQ interface is already designed for ATCA…

Issue: large inhomogeneity in geometry and readout

- The occupancy in the HGCAL varies from ~50% to ~0.5%
  - Building a uniform readout system (DAQ or trigger) for all regions is not trivial
  - One F/O link from each FE silicon hexagonal module would have resulted in 30k links
  - Prohibitive in terms of cost and cooling





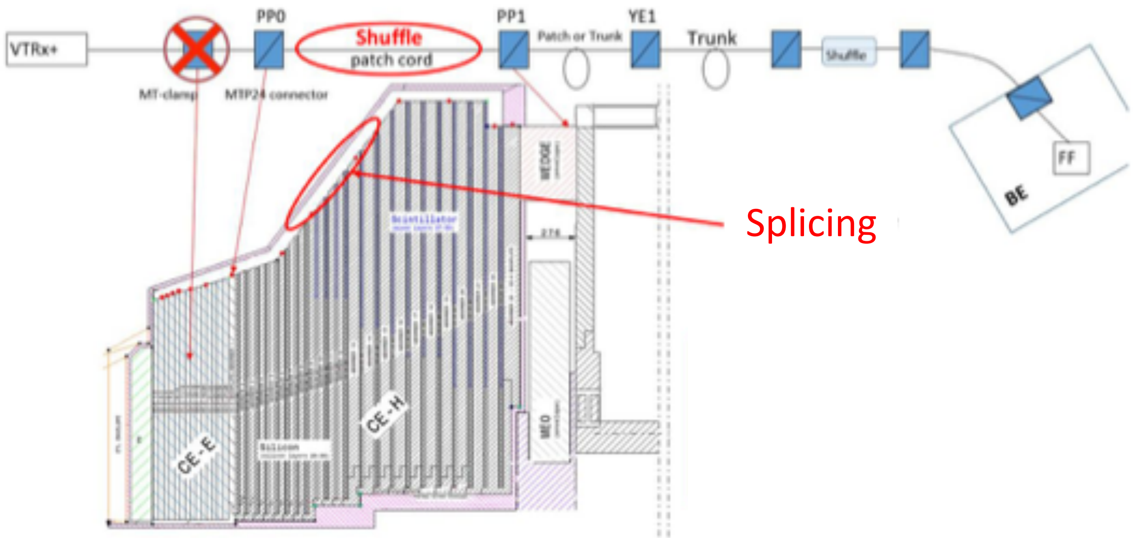


Single layer 60^o sectors

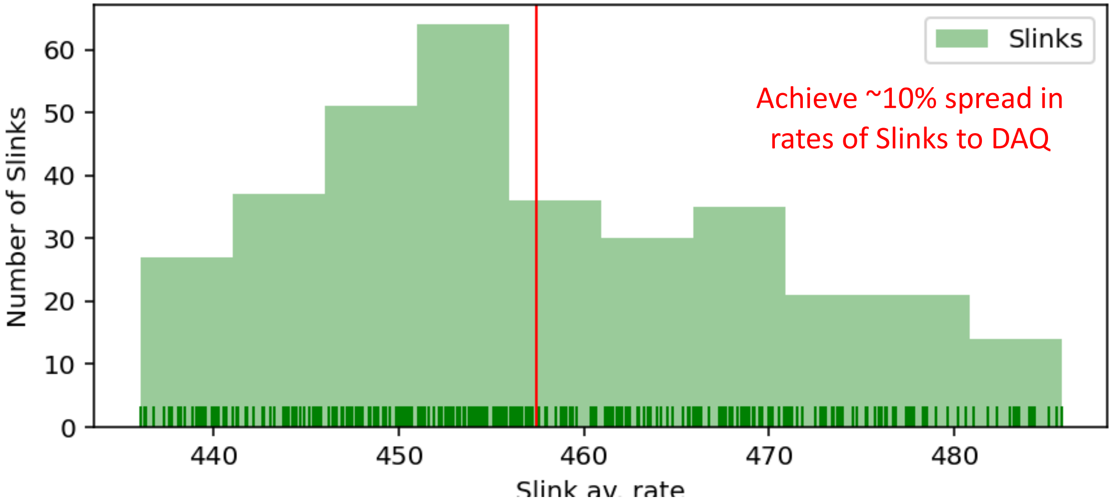
- Need to group modules together for readout
  - As well as keep link TXs out of highest radiation region
  - Colours show readout groups, but every layer is different
  - Some readout groups have rates which need more than one link
  - We even may have to split single data packets across two links
  - But big cost gain: number of links can be reduced to below 10k

Issue: large resulting inhomogeneity in rates

- All BE boards are identical
  - They have the same bandwidth and total buffer size
  - Essential to build-in load balancing from the beginning; homogenise by mixing links
  - Complicated by potential drop-off of FE TX link power with irradiation
  - For example: we will splice fibres on-detector to gather links with different rates into the same F/O connectors



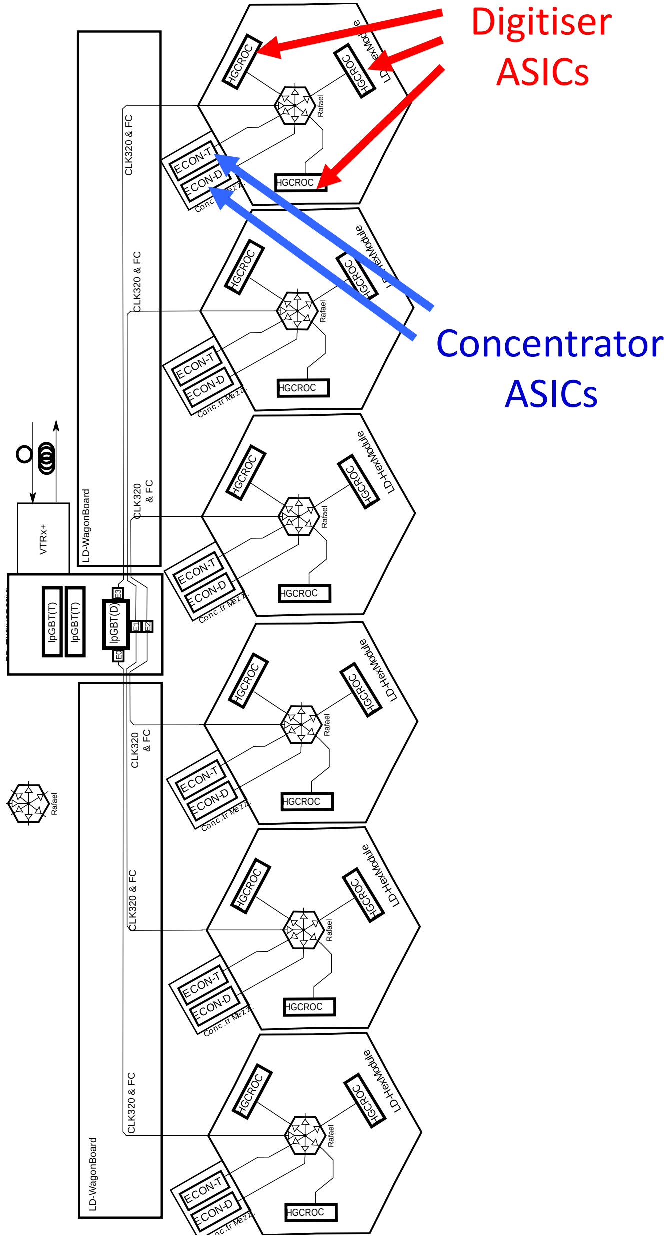
CMS preliminary simulation



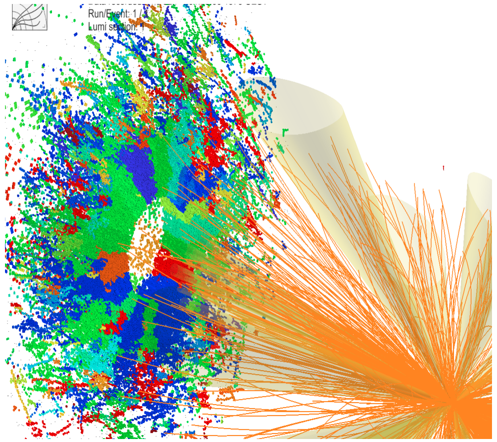
- Even then will require careful BE setup
  - Input links to BE have an average of 2 Gbit/s
  - Central DAQ interface is via 25 Gbit/s “Slinks”
  - Tune connector mapping to BE boards to bundle input links, so as to minimise spread of rates per Slink
  - Subdivide available buffers per input link to match incoming rates on a link-by-link level
  - Allow variable number of input links per Slink

Issue: high number of individual electronics elements

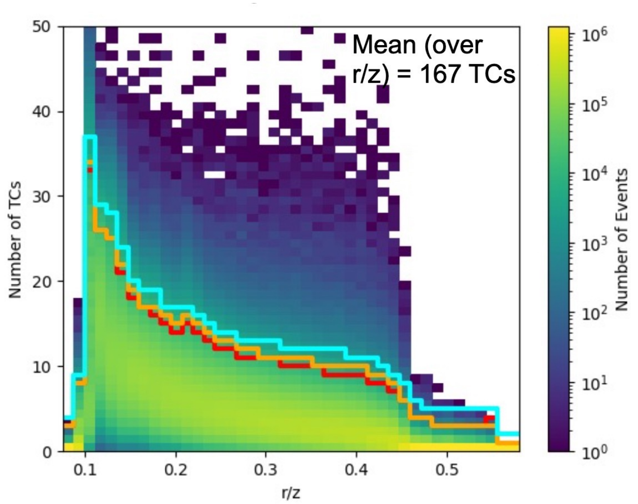
- FE electronics has ~100k digitiser ASICs and ~50k concentrator ASICs
  - Failure/buffer overflow rates not yet known so must assume non-negligible
  - Working on assumption that individual ASICs problems will not stop data-taking
  - Disabling data path, recovery procedure and resynchronisation must be done while rest of HGCAL still running
  - Requires very close integration of DAQ and control; effectively a single system
- A single HGCAL BE FPGA handles up to ~200 FE modules
  - Individual failure and recovery procedures need x200 copies in firmware
  - Includes separate event counters, etc, for alignment checks after recovery
  - Stripped down, highly resource-efficient implementation needed
  - Offload to software where possible; MHz speed not required
- Also need to keep track of dead areas
  - A tracker with reasonable redundancy can handle a missing hit in a layer, giving a small degradation in resolution
  - But a sampling calorimeter missing a layer gives a low (i.e. biased) result
  - Recovery offline is possible with clever reco as long as dead regions are known event-by-event



Issue: complications for triggering



- No layer-to-layer data exchange on-detector
  - Probably true of all future calorimeters; don't want to cut holes in absorber material
  - Some ideas for non-projective grooves, but may complicate assembly
- Services area around layer edges is very limited and difficult to cool
  - All trigger data selection uses local information; all sophistication pushed to BE



- Around ~50k hits (out of ~1M total trigger cells) above threshold in total per interaction
  - Very hard to push these through a trigger processing ML algorithm given current technology; have gone for explicit clustering algorithm (so far)
  - Again very non-uniform distribution so must load balance system
  - Tuning link layout to BE so as to minimise loss of trigger hits
- Single particle clusters can be made of large numbers of hits
  - Typical O(100) but tails up to O(1000); have to be able to handle worst case
  - Even in modern fast FPGAs running at ~400 MHz, accumulating each hit in turn breaks latency budget
  - Must limit to parallelised, decentralised cluster property calculations which restrict what can be calculated

Conclusions

- UK groups heavily involved in the CMS HGCAL backend electronics
  - Providing generic BE Serenity boards for use more widely across CMS
  - HGCAL BE systems will use Serenity throughout
- BE of a future detector will look very different
  - But difficult geometries, rate inhomogeneities, limited connectivity, etc, will still be there
  - None of the issues are unique to HGCAL but the combination of many of them together makes it non-trivial
  - Costs can blow up very quickly if not kept strictly under control

Backup

Recommend Dave Barney's talk given in EFCA TF6 meeting

- Discusses a lot of the issues found during HGCAL construction
  - Folded in with calorimeter experience from construction of original LHC detectors
  - Concentrates on HGCAL detector itself, not backend electronics
  - A link to Dave's talk is on today's Indico page
- Some of the issues he raises
  - 8-inch silicon wafers for sensors gave unexpected issues; e.g. only one irradiation facility was big enough
  - Inhomogeneities in the calorimeter drain effort and money; commonality is much better even if apparently more expensive initially
  - Never say never; assume an upgrade will be needed and design accessibility in from the start
  - The frontend electronics in HG calorimeters must be very tightly integrated with the mechanics; FE ASICs are needed much earlier than in previous projects (and HGCAL has three)
  - Intrinsic stability is more valuable than calibration capability; the latter without the former requires a large and continuous effort
  - Reliability, longevity and long-term availability of connectors are very difficult to guarantee
  - Scheduling for a long burn-in is critical; much better to find problems before installation than after