Overview of GEM/CSC backend needs

Evaldas Juska
Texas A&M University

CMS GEM/CSC Forward Muon Upgrade Workshop 09.04.2018

1. Overview

Evaldas Juska (TAMU)

- **General**
  - The needs for GE2/1, ME0, and CSC FED are similar
  - But each one also has important differences (covered in later slides)
  - All can reuse the flexible design of GE1/1 firmware
- **So far the design has been based on experience with CTP7**
  - CTP7 resources summary:
    - Optical: 67 RX, 48 TX
    - Virtex7 690T (53Mbit BRAM), 36Mbit external FIFO
    - Zynq processor, DMA to Zynq RAM
  - 1 CTP7 currently handles 12 GE1/1 chamber layers
    - 36 GBTs, DAQ tested up to 800kHz and 3.7Gb/s (limited by AMC13)
    - More details here
  - Expected from ATCA backend:
    - Same or higher optical links count (especially RX)
      - Higher RX count could reduce the total card count needed
    - ME0 needs more logic resources
    - For GE2/1 and CSC not much more logic resources are required
- **Reuse existing ATCA developments from other USCMS projects**

2. GE2/1 overview

Evaldas Juska (TAMU)

- **GE2/1**
  - Main branching point: use LpGBT or GBTX on the frontend
    - Falling back to GBTX would require 3x more bidirectional fiber links
    - This means 1.6x more RX and 3x more TX on the backend in total
  - Otherwise almost identical to GE1/1 (functionally and bandwidth-wise)
    - Board-portable firmware already exists and can be easily reused

- **Main functions of GE2/1 backend**
  - Control of the frontend (VFAT3 chips and OH)
    - Including phase adjustable clock delivery, TTC, and slow control
    - Using either LpGBT or GBTX protocol
    - Fast programming of OH FPGA after hard-reset (more details later)
  - DAQ and error reporting, and TTS
    - Total estimated system output data rate is very low = 4.6Gb/s
  - Trigger data alignment and “concentration”
    - Receive trigger data at 3.2Gb/s and align
    - Retransmit to EMTF on fewer faster links (current plan is for 9.6Gb/s)

- **Optical links will be covered in a later slide**

3. ME0 overview

Evaldas Juska (TAMU)

- **Same functions as GE2/1 with two key differences:**
  - Much higher DAQ data rate (706Gb/s for the full system)
  - Will also perform local 6-layer trigger stub finding

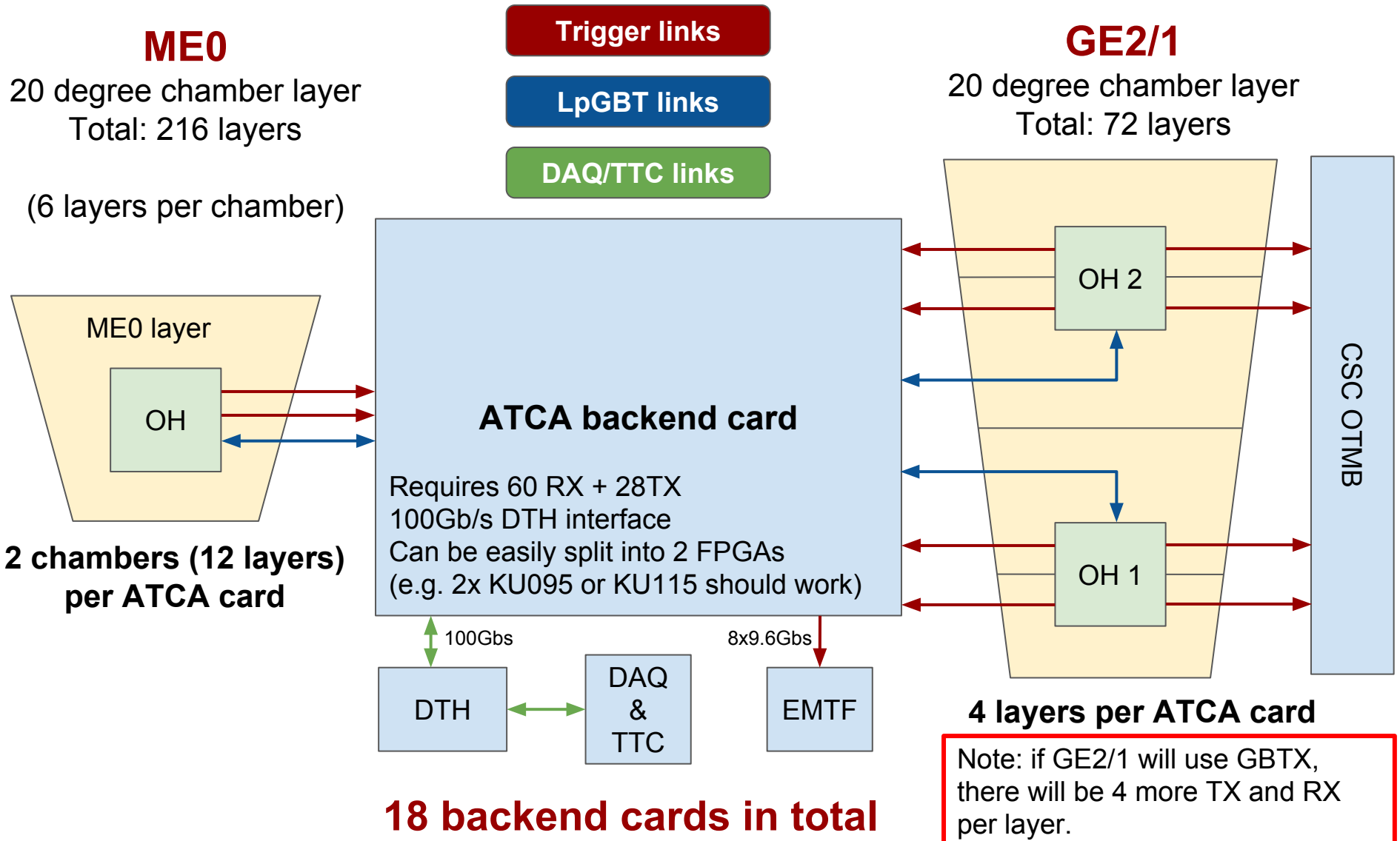
- **Local 6 layer stub finding**
  - May require quite a lot of FPGA resources
  - CSC has a dedicated board (OTMB) for this function (for each chamber)
    - Current firmware for ME1/1 OTMB uses about 80k virtex 6 “logic cells”
  - ME0 has 1536 di-strips vs. 224 half-strips in ME1/1
    - Simply extrapolating linearly results in 549k “logic cells” per ME0 chamber
  - Plan to have 2-3 ME0 chambers per backend card
    - 2 chambers if backend is combined with GE2/1, and 3 if not (see below)
  - Though ME0 stub-finding firmware most likely will be very different from CSCs, and possibly use less resources than this linear extrapolation
    - But still we have to keep this in mind and be conservative

- **ME0 and GE2/1 combined backend**
  - Makes a lot of sense to combine these two systems
  - Distribute the DAQ bandwidth and logic resource requirements
  - Total number of backend cards is reduced as well (see following slides)

4. ME0 and GE2/1 links version 1

Evaldas Juska (TAMU)

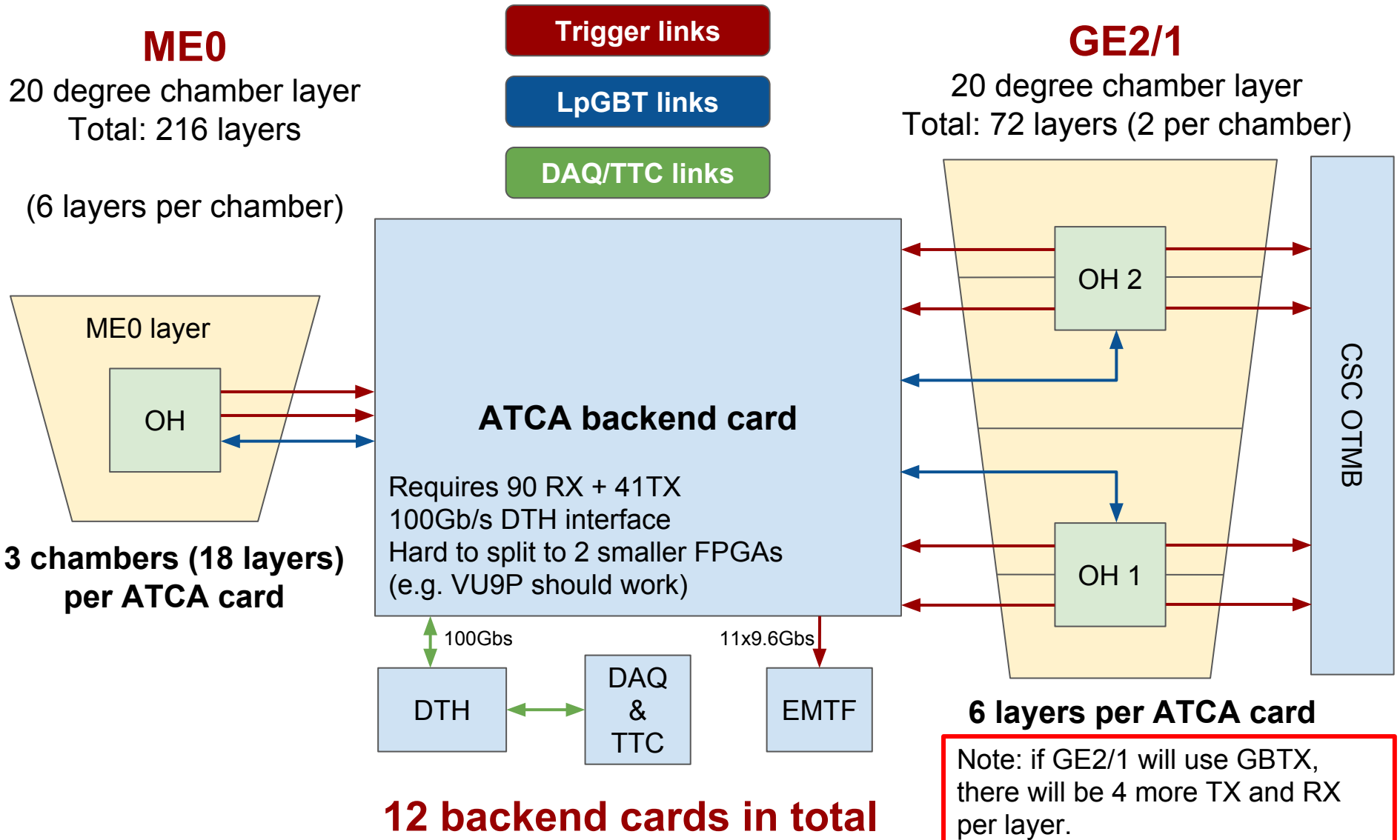
ME0 and GE2/1 as a common system, based on CTP7 link counts



5. ME0 and GE2/1 links version 2

Evaldas Juska (TAMU)

ME0 and GE2/1 as a common system, with more links than CTP7



6. Link summary, Bandwidth estimations

Evaldas Juska (TAMU)

- **Optical links summary**
  - Total ME0 chambers = 36
    - 18 RX + 6 TX per chamber needed (plus 1 TX for output to EMTF)
  - Total GE2/1 chamber layers = 72
    - 6 RX + 2 TX per layer needed (plus 1.3 TX for output to EMTF)
    - Falling back to GBTX would add 4 RX and 4 TX per chamber layer

- **Data rates per chamber (PU=200, L1A rate = 750kHz)**
  - All data rates use 3x neutron background hit rate
  - Includes headers/trailers added by backend
  - Note: numbers quoted are for data rate, not serial link line rate

Detector	GE2/1 layer	ME0 chamber
No optimization	92Mb/s	34.3Gb/s
Using adaptive zero-suppression at the frontend	**64.4Mb/s**	**19.6Gb/s**
Using selective readout requiring loose multi-layer trigger stub	N/A	~3.2Gb/s

7. CSC FED Overview

Evaldas Juska (TAMU)

- **Main function is DAQ, error reporting, and trigger throttling**
  - Total DAQ data rate is very high (598Gb/s)
  - More chambers per card, data packets are more complex than in GEMs
  - Total logic resource needs are not very high, but BRAM is very useful
- **FED optical link requirements**
  - Each FED ATCA card will connect to these chamber types:
  - ME1/1 (72 in total) : 4x 6.4Gb/s RX | 1x 4.8Gb/s TX
  - ME2/1 (36 in total) : 3x 6.4Gb/s RX | 1x 4.8Gb/s TX
  - ME3/1, ME4/1 (72 in total) : 2x 6.4Gb/s RX | 1x 4.8Gb/s TX
  - Outer chambers (360 in total) : 1x 1.6Gb/s RX | no TX required
  - **Total: 900 RX, and 180 TX (plus interfaces to DTH)**
- **TX needed for backpressure and frontend FPGA programming**
  - ODMB backpressure allows extending buffers to ODMB (very useful)
  - Fast frontend FPGA programming via GBT (xDCFEB and new ALCT)
- **Depending on the ODMB FPGA, 10Gb/s may be possible**
  - In this case 1 RX can be dropped for each of ME1/1, ME2/1, and possibly also ME3/1 and ME4/1, **reducing total RX count by up to 180**

8. CSC FED Segmentation and Bandwidth

Evaldas Juska (TAMU)

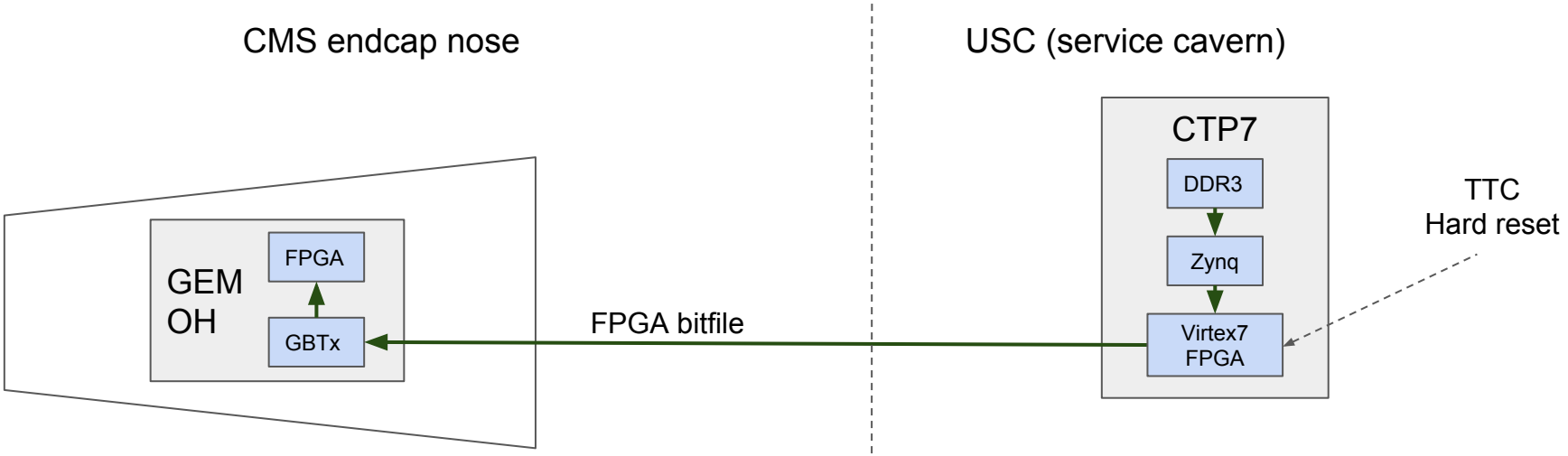
- **Segmentation**
  - Current plan is to have 12 backend cards (75 RX each)
  - Logic in each card could be divided into 2 FPGAs (with some asymmetry)
  - Fewer cards with more fiber links:
  - Link-wise, 9 cards would also divide all chamber types evenly
    - But one card would mix chambers from plus and minus endcaps
    - Requires 100 RX on each card (excluding DTH links)
    - Also total output bandwidth to DTH could be the limiting factor
  - Numbers between 9 and 12 could in principle work, but care would need to be taken to balance the bandwidth
    - Not equal number of each chamber type per card (best to avoid this)

- **Data rate**

	ALCT			(O)DMB			Full ring			
Chamber	Rate [Gb/s]	Link [Gb/s]	Util. [%]	Rate [Gb/s]	Link [Gb/s]	Util. [%]	CSCs	Rate [Gb/s]	Fibers to FED	Fibers from FED
ME1/1	0.87	1×4.8	19	4.3	4×6.4	21	72	306.7	288	72
ME2/1	1.16	2×3.2	23	2.8	3×6.4	18	36	102.0	108	36
ME3/1	0.57	2×3.2	11	1.6	2×6.4	16	36	57.5	72	36
ME4/1	0.56	2×3.2	11	1.6	2×6.4	15	36	56.5	72	36
ME1/2	0.08	0.64	12	0.3	1×1.6	20	72	18.4	72	72
ME2/2	0.05	0.64	7	0.2	1×1.6	13	72	11.6	72	72
ME3/2	0.06	0.64	9	0.2	1×1.6	15	72	14.2	72	72
ME4/2	0.12	0.64	18	0.4	1×1.6	31	72	28.7	72	72
ME1/3	0.01	0.64	1	0.1	1×1.6	2	72	1.9	72	72
Total CSC							540	597.5	900	540

9. Remote FPGA programming without PROM

- **Most EEPROMs have limited lifetime (Total Ionizing Dose)**
  - Tested PROMs started becoming unwritable after 10kRad exposure
  - Best option: program the FPGA directly from radhard GBTx chip
    - No EEPROM needed on the front-end boards (OH, xDCFEB, new ALCT)
    - After each hard-reset, backend FPGA reads the bitfile(s) from external RAM (CTP7 uses Zynq DDR3) and streams it to all frontend FPGA parallel programming ports via GBTX
    - Effectively the backend emulates a PROM
  - This feature is included in OHv3, xDCFEB, ALCT LX100
    - Was tested successfully on OHv3, and xDCFEB, ALCT is WIP
    - Programming time is same as from the PROM (70ms, 8bits @ 80MHz)



10. Remote FPGA programming requirements

- **Requirements on the backend**
  - RAM to store the bitfile(s)
    - Depending on the system this could is between 6 and ~10MBytes
    - Impractical to store it in BRAM, so this RAM has to be external to FPGA
  - This bitfile needs to be streamed after hard-reset at 640-1600Mb/s
    - This stream cannot contain any gaps
    - So read bandwidth of the RAM has to be sufficient to sustain this data rate

- **CSC**
  - In CSC two different boards will have to be programmed in parallel
    - xDCFEB and new ALCT
  - Two bitfiles will be interleaved in RAM
  - Requires more bandwidth
    - Baseline option will require 960Mb/s

Detector	GE2/1	ME0	CSC xDCFEB	CSC ALCT
Data rate	640Mb/s	640Mb/s	Two options: **640Mb/s** 1280Mb/s	320Mb/s

11. Summary

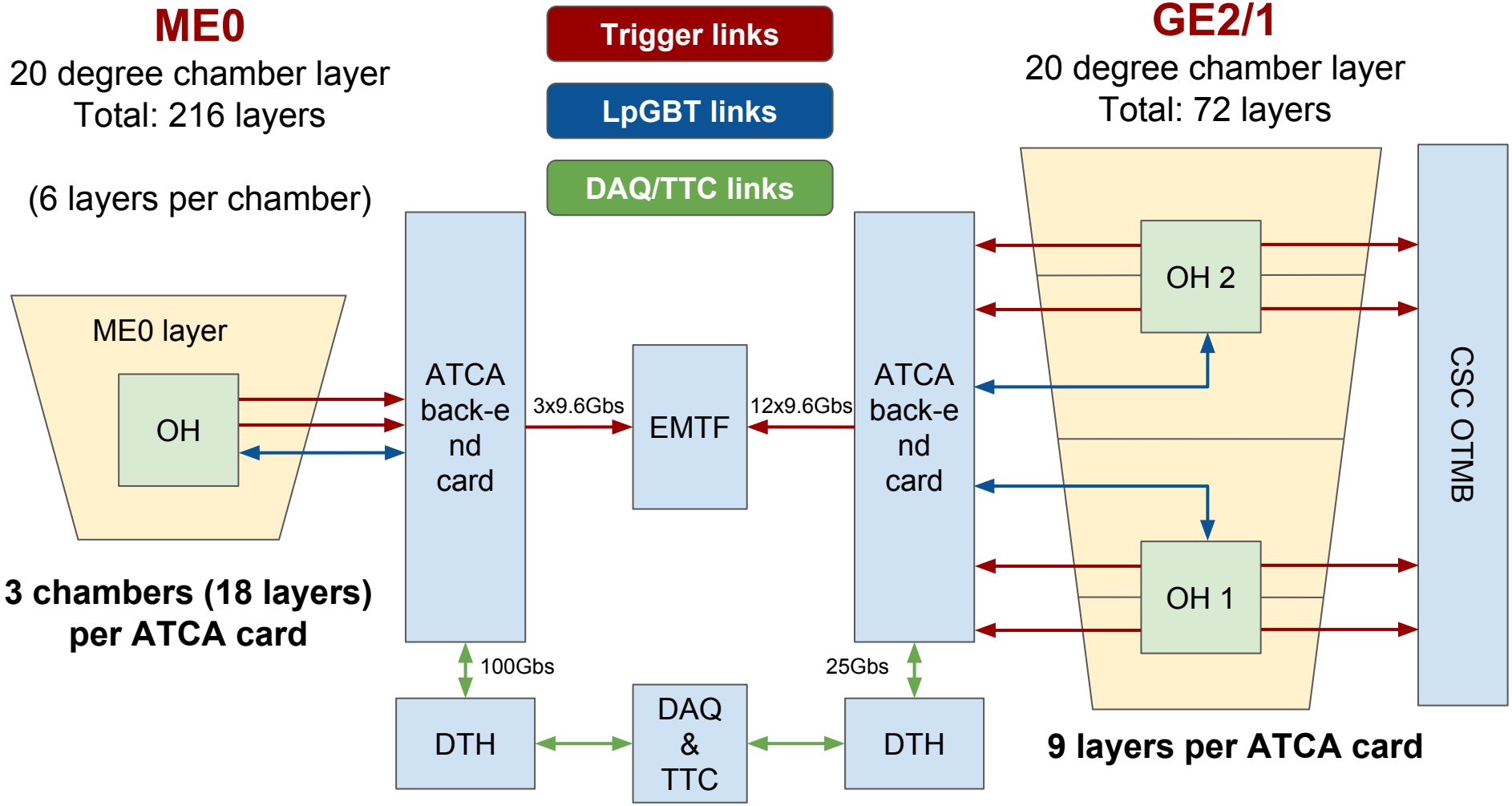
- **ME0**
  - Total 36 chambers
  - 18 RX + 6 TX per chamber needed (plus 1 TX for output to EMTF)
  - DAQ data rate per chamber is 19.6 Gb/s (total output 706 Gb/s)
  - A lot of logic resources needed per chamber
  - Combining backend with GE2/1 would balance the resource needs
- **GE2/1**
  - Total 72 chamber layers
  - 6 RX + 2 TX per layer needed (plus 1.3 TX for output to EMTF)
  - Falling back to GBTX would add 4 RX and 4 TX per chamber layer
  - DAQ data rate per chamber is 64.4 Gb/s (total output 4.6 Gb/s)
  - Logic resource requirements are not very high
- **CSC**
  - Total 540 chambers
  - Total links = 900 RX and 180 TX (possible reduction of up to 180 RX)
  - Total DAQ data rate is 598 Gb/s
  - Logic resource requirements are not very high, but BRAM is useful

BACKUPS BELOW

Backup 1. ME0 and GE2/1 links

Evaldas Juska (TAMU)

ME0 and GE2/1 as separate systems



ME0: 12 backend cards in total

GE2/1: 8 backend cards in total

Backup 2. Bandwidth estimations

Evaldas Juska (TAMU)

- **Nominal HL-LHC luminosity is $5 \times 10^{34} Hz/cm^2$ with 140 pileup interactions per BX**
  - Peak luminosity may be higher, so include rates equivalent up to 200 pileup interactions
  - Many uncertainties in neutron background simulation, so adding a safety factor of 3x
- **All rate estimates are obtained with simulated CMSSW events**
  - PU + neutron background only
- **For every L1A we will readout 3 bx of data**
- **Rates are calculated very conservatively**
  - Using the highest eta of the detector
  - Data rates can be reduced using zero-suppression within a VFAT
    - VFAT3 chip does allow for that
  - Data rates can be further reduced by requiring a loose coincidence between L1A and a multi-layer trigger in the chamber (ME0 only)

Backup 3. GE2/1 bandwidth estimates

Evaldas Juska (TAMU)

- **GE2/1**
  - 0.062 hits per 20 degree layer per BX on average
  - Adding 10% to account for hits on VFAT boundary (conservative)
  - DAQ bandwidth per 20 degree 2 layer super-chamber = 71Mbit/s
    - 2 layers * 750kHz L1A * 0.062 hits * (1.1 * 192bits VFAT data * 3bx + 128bits headers/trailers)
  - With 3x neutron rate we have 0.114 hits per BX per 20 deg layer
    - DAQ bandwidth per 20 degree 2 layer super-chamber = 130Mbit/s
  - There are 8 CTP7s (each covers 4.5 super-chambers)
  - Average data rate from each CTP7 to AMC13 module = **0.83Gb/s**
    - Using 3x neutron hit rate
    - Chamber data + CTP7 headers/trailers (5 * 64bits * 750kHz)
  - **Total system average data output = 6.64Gbs + cDAQ protocol overhead**
    - Safe for uTCA system with one or two crates

Backup 4. ME0 bandwidth requirements

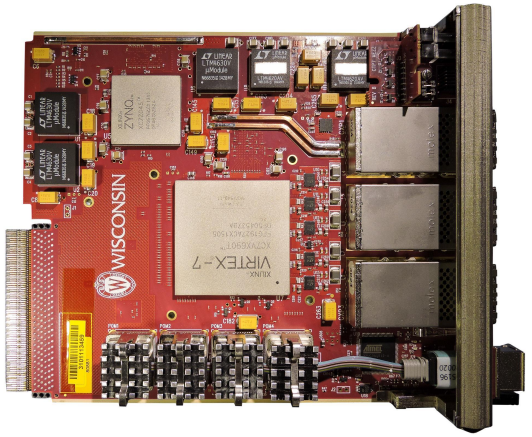
Evaldas Juska (TAMU)

- **It's not a small amount...**
  - Simulations for PU=140 show 2.5 hits per layer per BX
  - Reading out 3BX of data at L1A rate = 750kHZ
  - Each CTP7-like module with 3 ME0 chambers will output **39Gb/s**
  - Increasing PU to 200 raises the output rate to **53Gb/s**
  - Many uncertainties in neutron background simulation, so also consider scenarios with 3x the nominal neutron hit rate:
    - PU=140, output rate = **80Gb/s**
    - PU=200, output rate = **103Gb/s**

- **Way to control the output data rate**
  - Require a loose multi-layer trigger stub near the readout region
  - Similar to CSC pre-trigger for (D)CFEB readout
  - We will be building stubs in the same board
  - Dramatic reduction in readout rate:
    - PU=200 and 3x neutron background scenario **103Gb/s becomes 9.6Gb/s**

- **Safe for ATCA based backend (100Gb/s readout link)**

Backup 5. CTP7

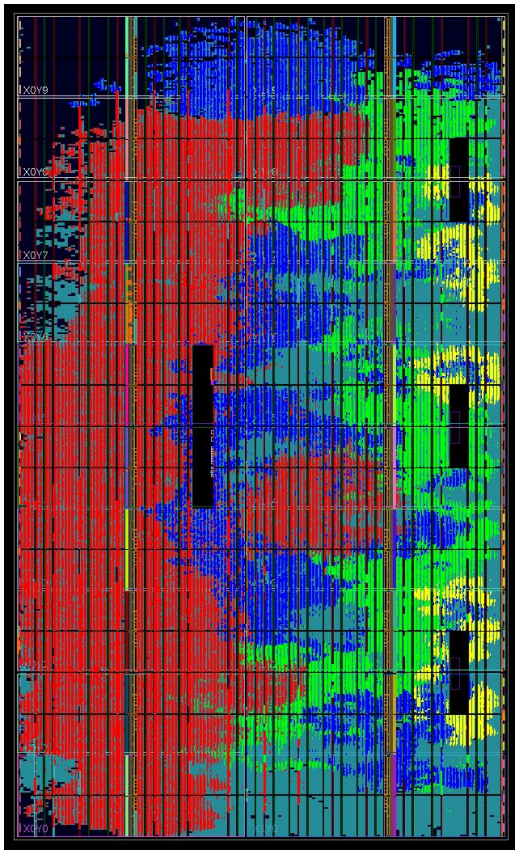


- FPGA
  - Virtex 7 690T, speed grade -2
    - Optical interface: 67 RX and 48 TX (10Gbs)
      - Perfect fit for 60 degree GE1/1 sector
    - 53Mb block RAM for large DAQ buffers
      - Additional 36Mb external FIFO is available
- Zynq 7000 processor as on-board controller (unique feature)
  - Embedded linux, powerful monitoring & control features
    - Fast memory mapped interface to the Virtex7 FPGA (32bit @ 50MHz)
    - Direct Memory Access from FPGA to 512MB of DDR 3 RAM
    - Embedded applications (monitoring, control, calibration)
      - Reg access heavy applications profits a lot (~100k accesses in 1s)
    - Remote programming and debugging (virtual Xilinx cable)
    - Remote procedure call service, and uHAL via TCP/IP interface
- Great support from University of Wisconsin
  - Setup all infrastructure needed to use the boards
  - Provided support for firmware development and solved some issues
- In use at P5: calorimeter trigger, GE1/1 slice test

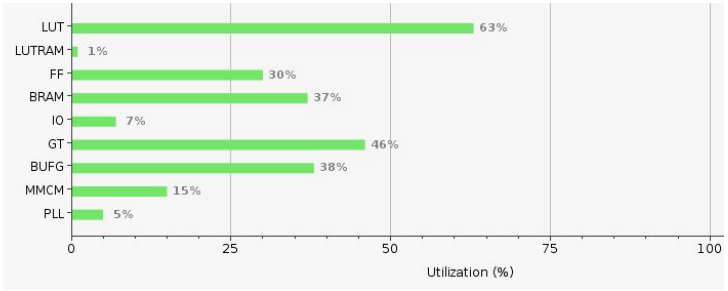
Backup 6. GE1/1 CTP7 Firmware

Evaldas Juska (TAMU)

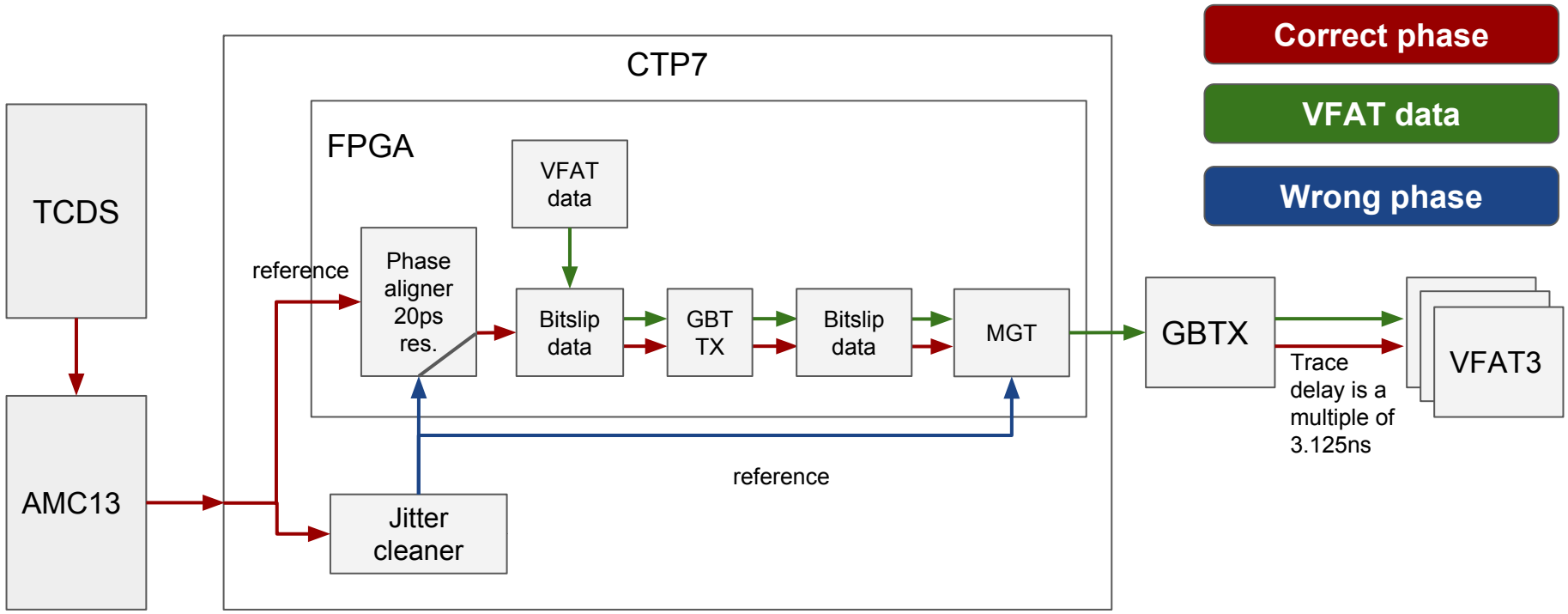
- **Fully functional CTP7 firmware**
  - Supports 12 OHv3s
    - 36 GBT links (fixed clock phase & latency)
    - Fast control
    - Slow control
    - SCA: hard-reset, JTAG, monitoring
    - DAQ (tested with 800kHz L1A, and 3.7Gb/s)
      - Decoding, Buffering, Event Building
      - Optional zero suppression
      - Error checking, status reporting (TTS)
    - Receives trigger data
      - Monitoring, rate counting, local trigger
    - Fast promless programming of OH FPGA (70ms)
  - Work in progress:
    - Trigger data alignment
    - Output to EMTF
  - Easily portable to different boards (e.g GLIB)
  - Chip utilization and timing is very good
  - Similar firmware for v2 electronics passed DAQ stress test at CMS (slice test)



DAQ | GBT RX | VFAT logic | GBT TX



Back 7. GE1/1 Clock distribution & phase alignment

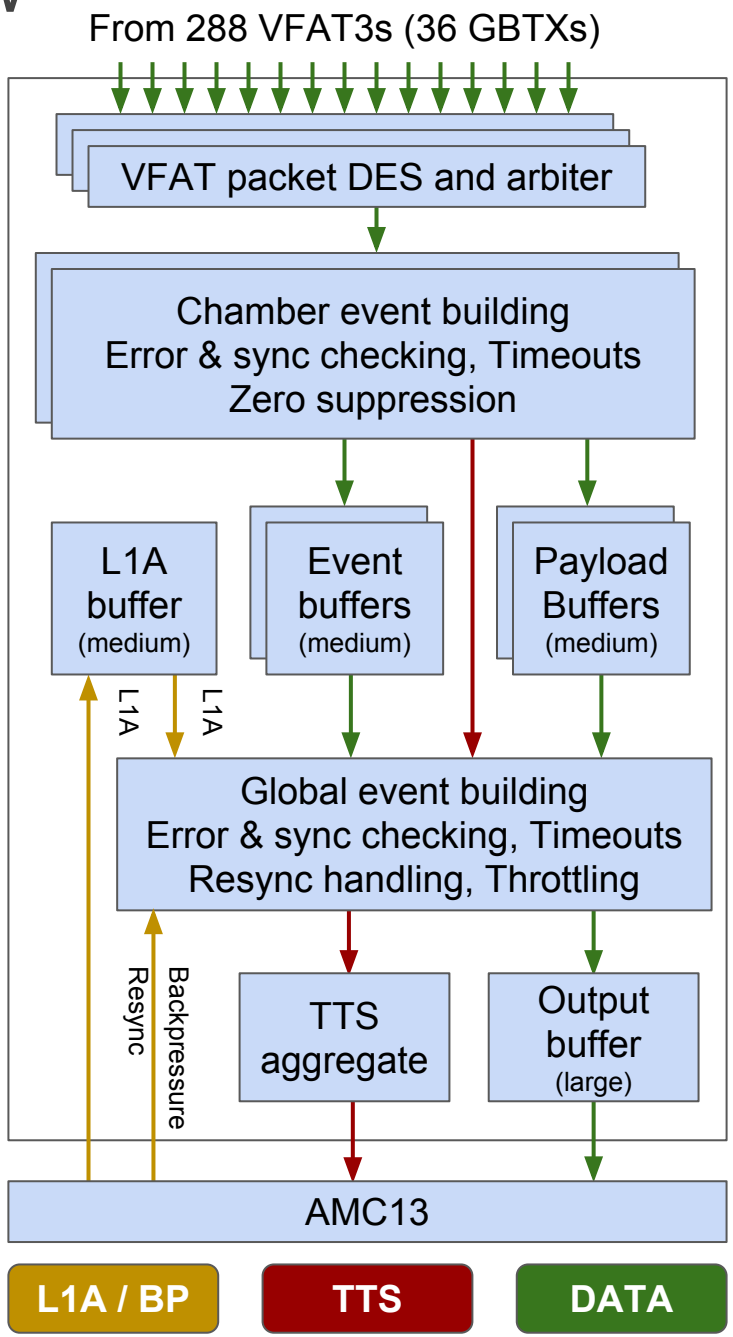


- **Deterministic phase w.r.t. TCDS phase**
  - Phase control per CTP7 (resolution 20ps)
  - Phase control per GBT link (resolution 312ps)
  - Phase control per VFAT (resolution 3.125ns)
  - GBT to VFAT3 trace lengths are a multiple of 3.125ns

Backup 8. GE1/1 DAQ Overview

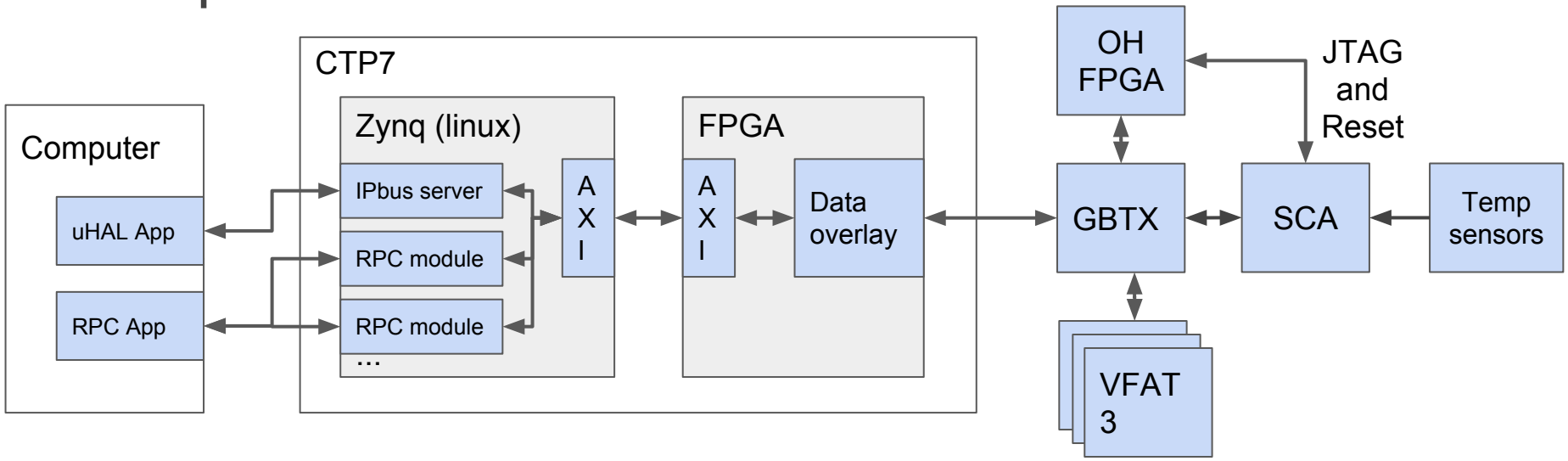
Evaldas Juska (TAMU)

- Supports all features needed for CMS
  - Almost identical to v2
    - Except in v2 VFAT buffers were in OH
    - V2 is running at P5 and passed CMS stress tests (backpressure, resets, etc)
  - Buffers
    - Small buffers for each VFAT
    - Serialized payload buffer for each OH
    - Event buffer for each OH (EVN, flags..)
    - One big L1A buffer
    - One big output buffer for throttling
  - Extensive error and sync checking
    - CRCs, buffer status, format checks
    - EC and BC sync between VFATs and between OHs
    - Aggregation to global TTS state
  - Data timeouts for VFATs and OHs
  - Output throttling, and resync handling
  - High L1A / data rate capability
    - Output limited to 4Gbps by AMC13
    - **Tested with 800kHz / 3.7Gb/s**
      - No problems



Backup 9. GE1/1 Slow Control

Evaldas Juska (TAMU)



- VFAT and OH registers are mapped to CTP7 reg address space
  - Sent over the same GBT link as DAQ and TTC during idle periods
- Zynq has a fast communication channel with FPGA (AXI 32b@50MH)
- All FPGA registers are mapped to Zynq linux virtual memory
  - For a Zynq application it's just like local memory access (simple)
- IPbus server provides access over TCP/IP for uHAL applications
- Remote Procedure Call (RPC)
  - Parts of computer software can be packaged as RPC modules and deployed on the Zynq operating system
    - Access to registers from an RPC module are very fast (~10us per access)
    - Heavy reg access code profits a lot (100k accesses in 1 second)
    - E.g. calibration routines, SCA JTAG driver (can program the FPGA in 10s)