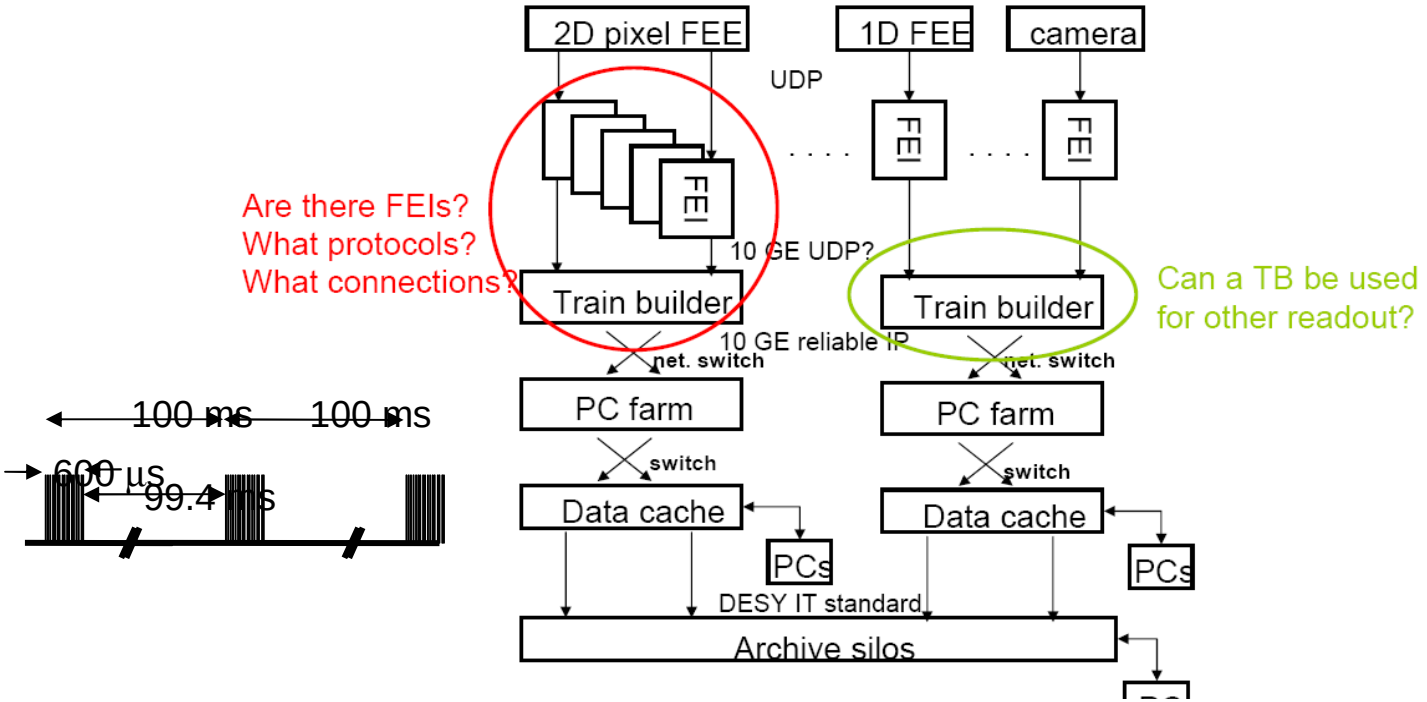
XFEL
Train Builder

XFEL DAQ architecture?



Train Builder (TB) Project

Proposal to provide a common interface for data readout between the three large 2D Pixel detectors and the PC Farm.

The system assembles all the data from 2D readout modules for a complete pulse train (10Hz) ready for processing on the Farm.

Brief History

LPD Expression of Intent pictured ATCA based train builder with 10Gbps optical links. Assumed significant data reduction (LHC).

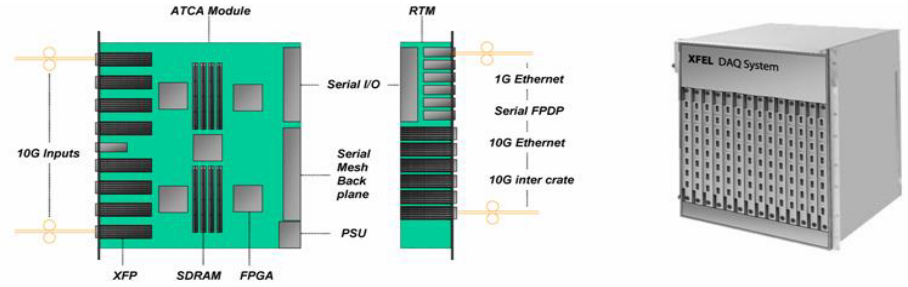
XFEL Advisory Cttee excluded TB from LPD project.

Introduced "Train Builder" concept at XFEL DAQ Workshop at DESY in March 2008.

Produced a Strawman paper design of Train Builder board.

Reviewed TB design at DESY in June. Agreed to proceed with design..

Original Requirements for Train Builder Strawman



1 Mpix Train builder board for 2D detectors
- Max 2 bytes per pixel
- Max 512 FE pipeline
- Max 10 Hz train rate (100 msec)

-> ~ 10 GBytes/sec Data Rate sustained

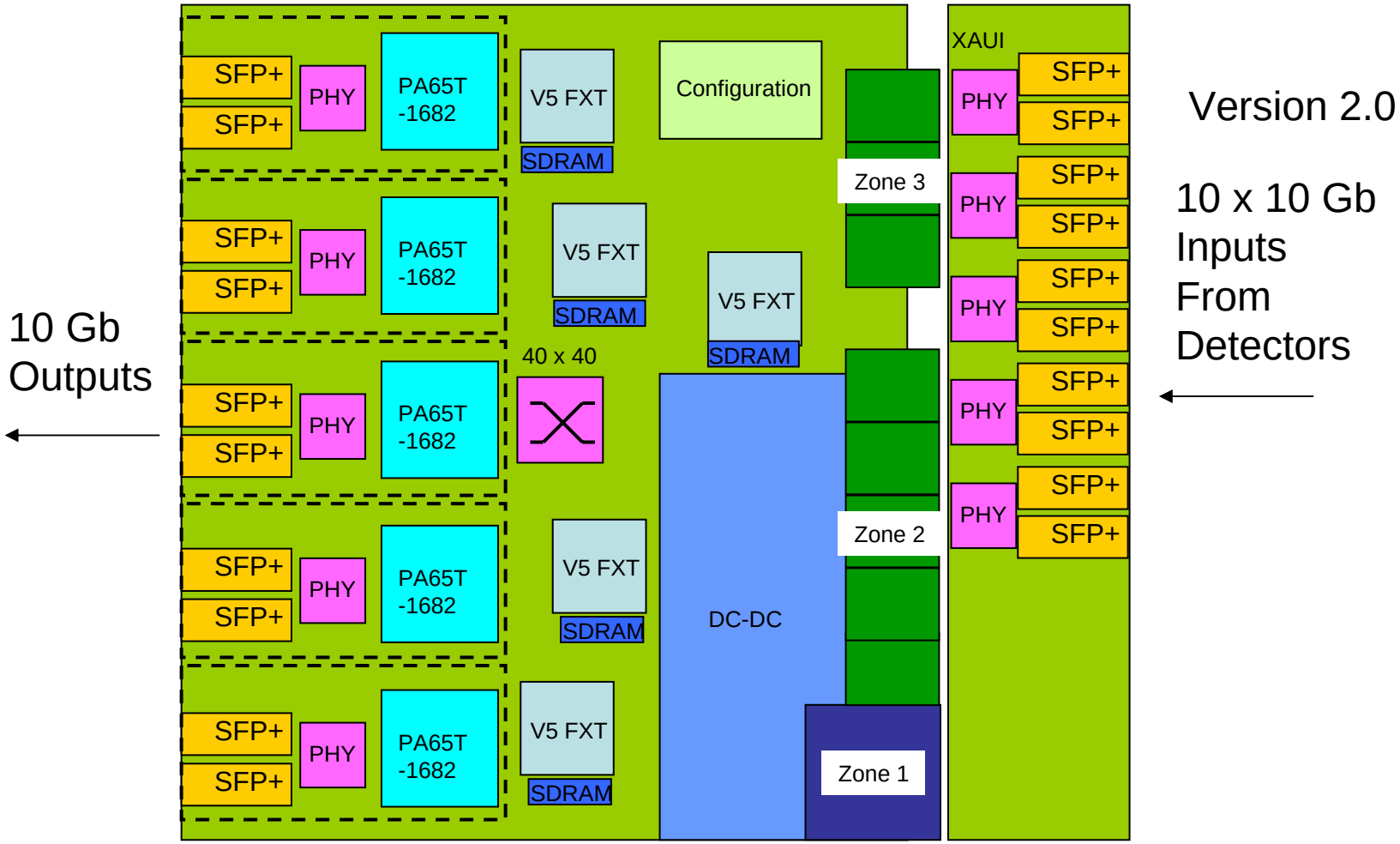
- Readout path only (not Timings and Controls distribution)
- Fixed length frames , Data arriving in Frame order

- Inputs 10 Gbps optical (<30m SR), UDP? (need match from 2D detectors).
- Outputs 10 Gbps Reliable UDP (>30m?) (assume no data reduction).

- Keep option of implementing data reduction algorithms (FPt)

- Kept to an ATCA 8U form factor with Transition card

1 MPix TB Strawman Proposal Architecture based on Cross Point Switch



Inputs to RTM. Processor on Mezzanines

40% scaling

Conclusions from DESY Meeting June 2008

TB related points:

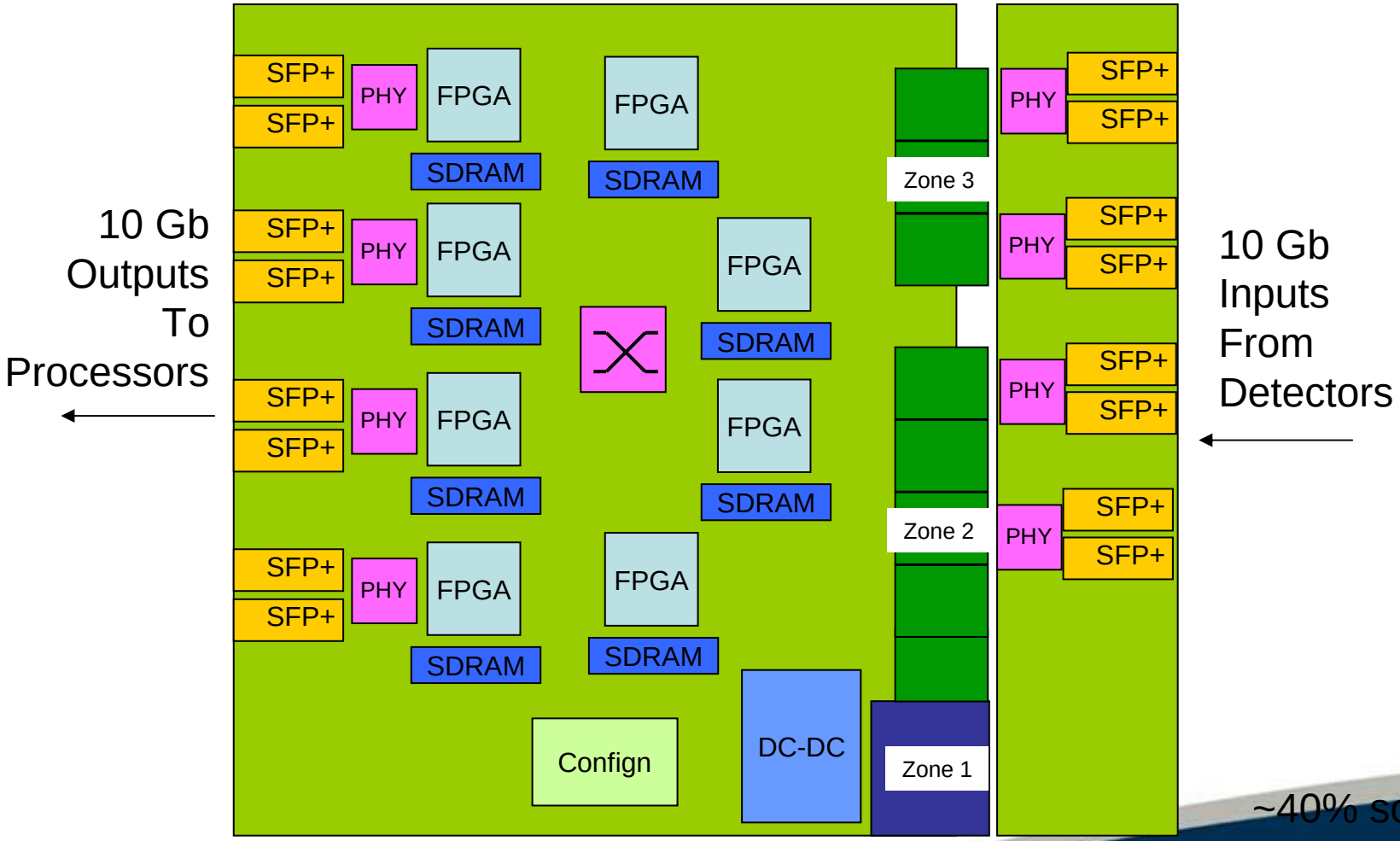
- The decision was made to proceed with the TB design. This will happen as an "in kind" contribution from the UK, lead by RAL, with a contribution from DESY FEA.
- The TB design uses 10GE links over SFP+ fiber tranciever input and outputs modules. The link hardware is fully bi-directional.
- Two boards with each 8 inputs and 8 outputs are needed per per Mpixel detector. The ½ Mpixel frames will be combined at the following PC layer.
- The input data transfer protocol will be UDP without retries. The use of protocols PGP, Aurora, etc. is not excluded.
- The output protocol is TCP.
- Network package loss should affect the minimum number of frames – does this mean that the built block will be transferred frame wise and not boundary independently?

Common control signal interface related points:

- A common control signal source board should be foreseen.
- A common distribution system might also be possible.
- A precise definition of slow machine information is required.

½ **MPix** Train Builder. Matched to 2D 8 x FEI outputs

"Simplest" Train Builder Implementation. All FPGA / No Mezzanines.



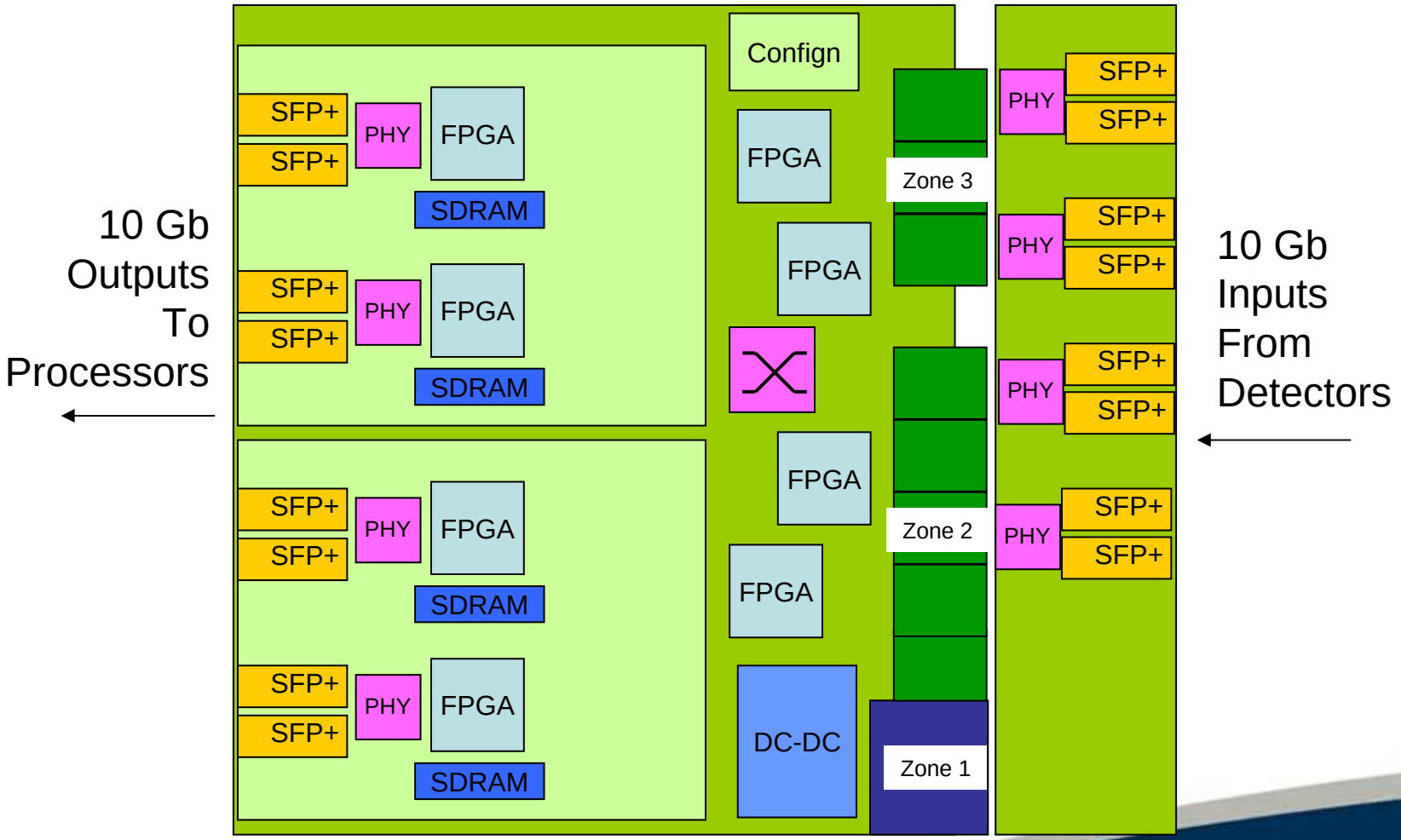
~40% scaling

Science & Technology Facilities Council

FPGA ½ MPix Train Builder
With Double width AMCs mezzanines

Nb Not showing all components on AMCs

SDRAM on component side 2



40% scaling

Final Implementation with AMC Mezzanine cards

Advantages
Upgrade processing units as technology improves
Reduces risk of manufacture of large boards

Mezzanines might allow more channels? Up to 8 single AMCs on some carriers.
Commercial AMC cards market

Prototype AMC in mTCA. Build up system in stages.

Disadvantages

Extra connector for High speed serial.
AMC form factor uses most of board space. Motherboard components under AMCs.
Coordination between FPGAs more complicated.

More cards. More expensive.

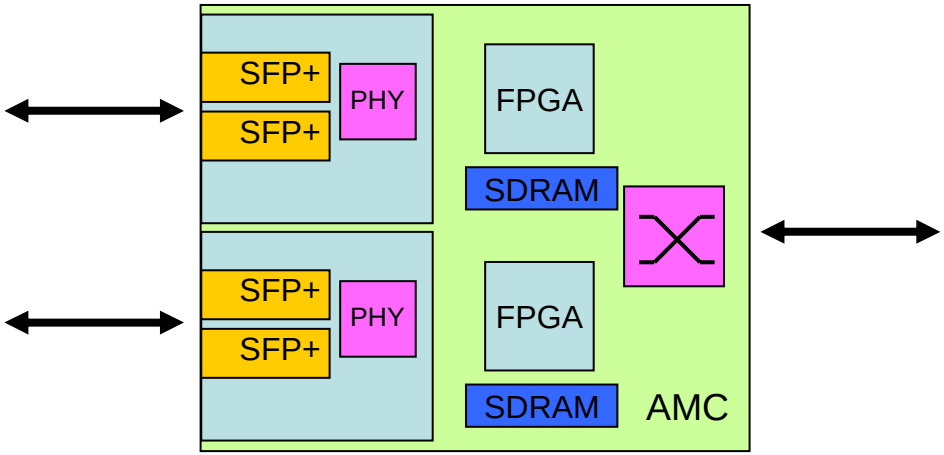
10 Gb Optical Mezzanines? Holding SFP+ and PHYs

Separate Opto cards. Clean partitioning of work. Avoids merging SFP+ PCB designs.
Use same Opto mezzanines on TB and Detector FE readout cards?
Any suitable standards compatible with AMC?

Train Builder Development Plan Proposal

1. Start with FPGA and Crosspoint Switch **development boards**.
   Board to board 10 Gbps.
   Investigate Memory interfaces.
   Serial Protocols. Aurora protocol switching.
   Get experience with driving Crosspoint switches.

2. **Demonstrator AMC card with CrossPoint**. Prototype system using MicroTCA crate.
10 Gb designs from DESY. Consider Optical Mezzanine on AMC?



3. Full **ATCA board Prototypes**. With RTMs

4. Full **ATCA board Production** with RTMs.

UK TB Developments

UK In Kind committee very positive, go ahead to develop a project proposal.

Arranged loan of Mindspeed Crosspoint switch development board. Trying to get similar for Vitesse.

Met with ATCA manufacturer Schroff. Quotes obtained. Same kit as DESY.
Met with ATCA power module suppliers.
Joined PICMG. Access to full ATCA technical specs.

Examining possible mezzanine solutions for optical interface.

Setting up to test switching with Aurora streams on V5 dev board.

Summary

Target is **½ MPix** Train builder board.

TB **well matched to Front End detector links** (16 x 10 Gb links all 1MPix detectors)

TB **Development plan**: Dev boards. 10 Gb. AMC Crosspoint Demonstrator. Full ATCA prototype.

Just a few of the Open Questions

What are realistic timescales for needing full TB system?
Is (FPt) processing essential on TB?

How best to partition TB work packages? DESY 10 G optical and UDP/TCP? + RTM?
How do we share work practically? FPGA Firmware IP. PCB design. Tools.

Final design may evolve with experience on prototypes and emerging technologies.
Depending on timescales. But have to start somewhere.

Stay receptive to Improvements and Alternatives
To this design? To crosspoint architecture? To ATCA form factor?
Keep eye on COTS market. Use commercial systems if it does the job.

Spare Slides

Power

Guesstimates

½ MPix

	Component	Model	Number	Power ea W	Power total W	
ATCA						
	FPGA	LXT100	4	8	32	
	Processing Unit	PA6T	4	12	48	
	Memory	DDR3 2GB SODIMM	8	3	24	
	Opto Transceiver	SFP+	8	1	8	
	Opto PHY	Dual	4	3	12	
	Switch		1	15	15	
	Others?				20	
				Total	**159**	**W**
RTM						
	Opto Transceiver	SFP+	8	1	8	
	Opto PHY	Dual	4	3	12	
				Total	**20**	**W**
	DC-DC inefficiency				1.4	
				Total Power	**250.6**	**W**
Limits	ATCA	200	W			
	RTM	20	W			
	Total	220	W			

ATCA power units @ 250W (300W). Cooling?

Next Generation Commercial Technologies

FPGA with hard core PCIe gen 2, memory ctrls DDR3.

Cross point switches PCIe gen 2 speeds

FPGA with 10 Gbps serial links

Cheaper, lower power, denser optics standards SFP++ , PHYs

Low power processors on AMCs

Detector Inputs

Via Rear Transition Module

Decouples board from connection to detectors

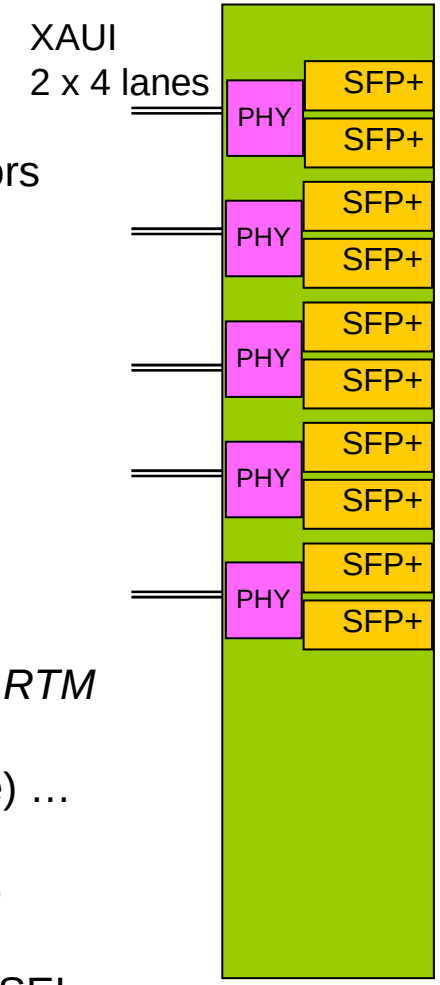
N pairs of SFP+ 10 Gbps Transceivers
+ PHYs (dual)
Each 2 x 4 XAUI lanes to board
Feeding 1 Train Input FPGA

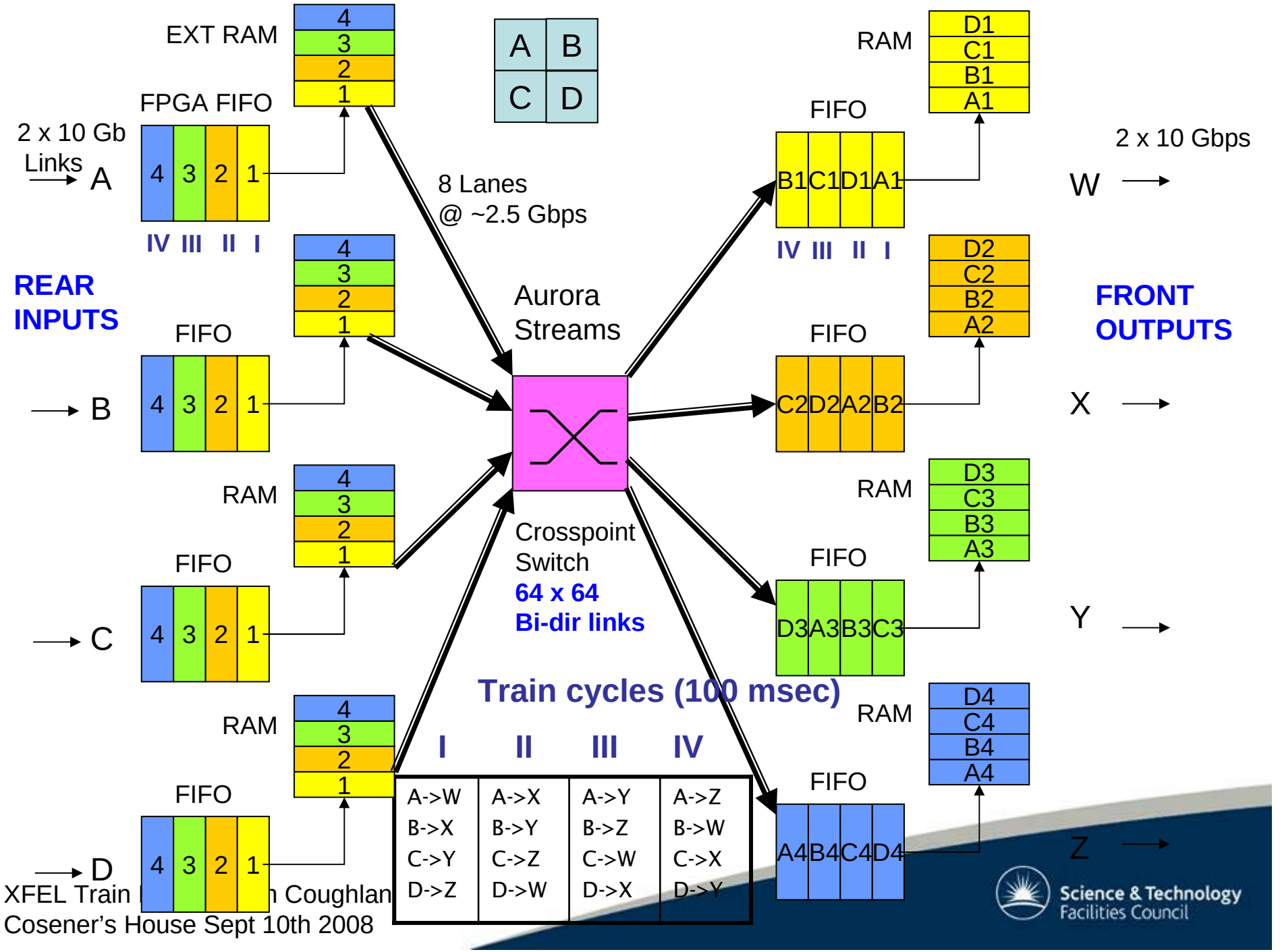
Match 32 modularity.
Multi channel Opto transceivers?
Quad SFP.
Need to merge 4 fibres to one connector on RTM

Protocols Aurora, PGP, 10 GbE (standalone) …

Zone 3 connectors (same as zone 2 fabric?)

10 G Optical < 30 m. 850 nm MM fibres VCSEL
short reach 10 G





XFEL Train [illegible] Coughlan
Cosener's House Sept 10th 2008

Detector Inputs

Via Rear Transition Module

Decouples board from connection to detectors

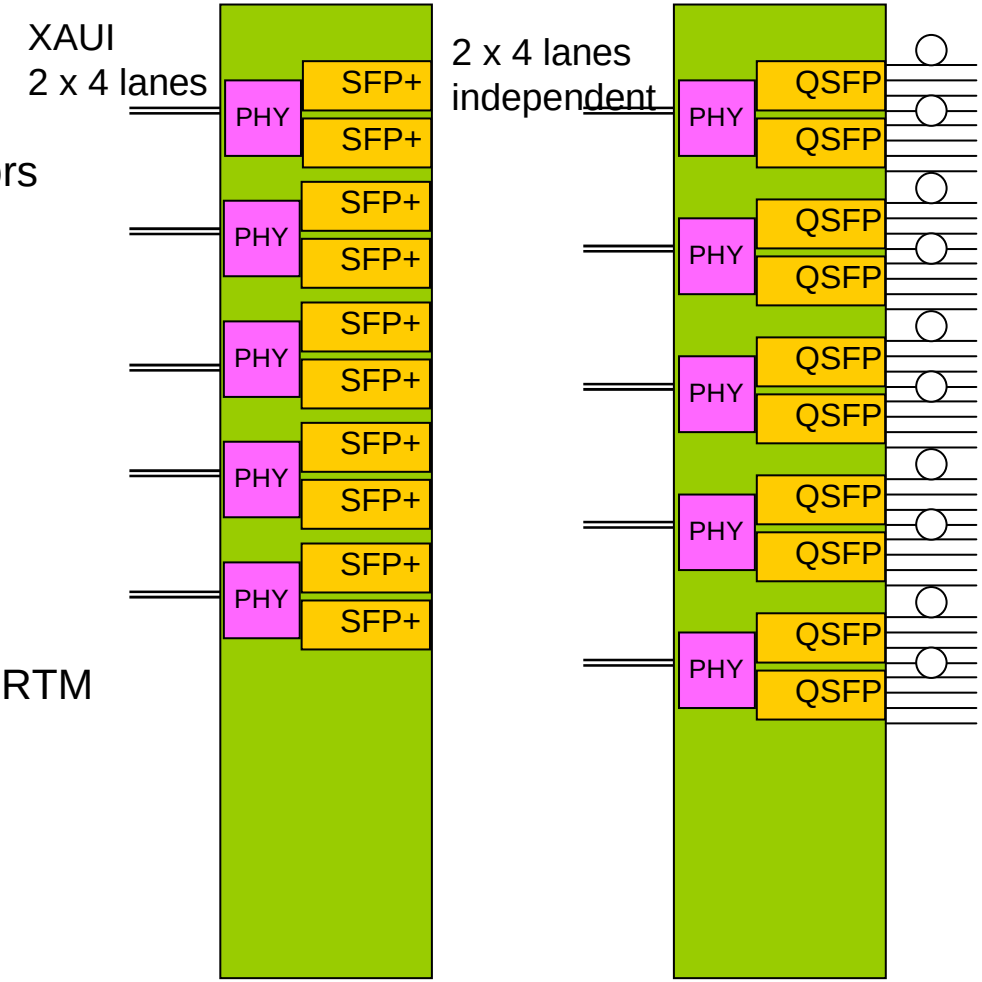
N pairs of SFP+ 10 Gbps Transceivers
+ PHYs (dual)
Each 2 x 4 XAUI lanes to board
Feeding 1 Train Input FPGA

Match 32 modularity.
Multi channel Opto transceivers
Quad SFP?
Need to merge 4 fibres to one connector on RTM

Protocols Aurora, PGP, 10 GbE …

Zone 3 connectors (same as zone 2 fabric?)

Optical 20 m fibres. 850 nm MM
short reach 10 G



Train Input FPGAs

Each processing ~ 2 GB/sec

Receiver. Decode serial protocol

Detector Dependent.
•Merge streams.
•Reorder appropriately for processing units.

Traffic shaping across switch

Serial out through switch eg Aurora, PCIe
I/O = 16 MGTs

Fast Ext memory controller @ 2 GB/s in and out
Deep Buffer train fragments before builder crosspoint switch
~ 2 GBytes ?
DDR2/3

Xilinx V5. (V6 40 nm?)
Altera Stratix 4 (40 nm 8 Gbps transceivers)
Embedded PCIe

Builder Switch

Analogue crosspoint switch.
Vitesse, Mindspeed < 144x144

Serial Protocol agnostic.

All serial links need to be bi-dir.

Simple traffic shaping. Train (sub)fragments.

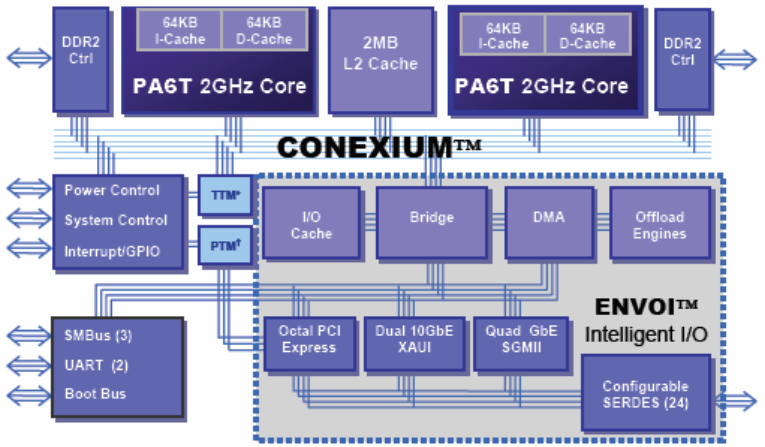
Processor Units

Could be FPGA based. Limited processing. Simple train building.

Conventional multicore Freescale
Field Programmable Computing Arrays. Daresbury Labs.

PA-SEMI device. Very low power.
Dual PPC processor. Dual DDR2 interface. PCIe and XAUI I/O.
PA-SEMI just acquired by Apple!

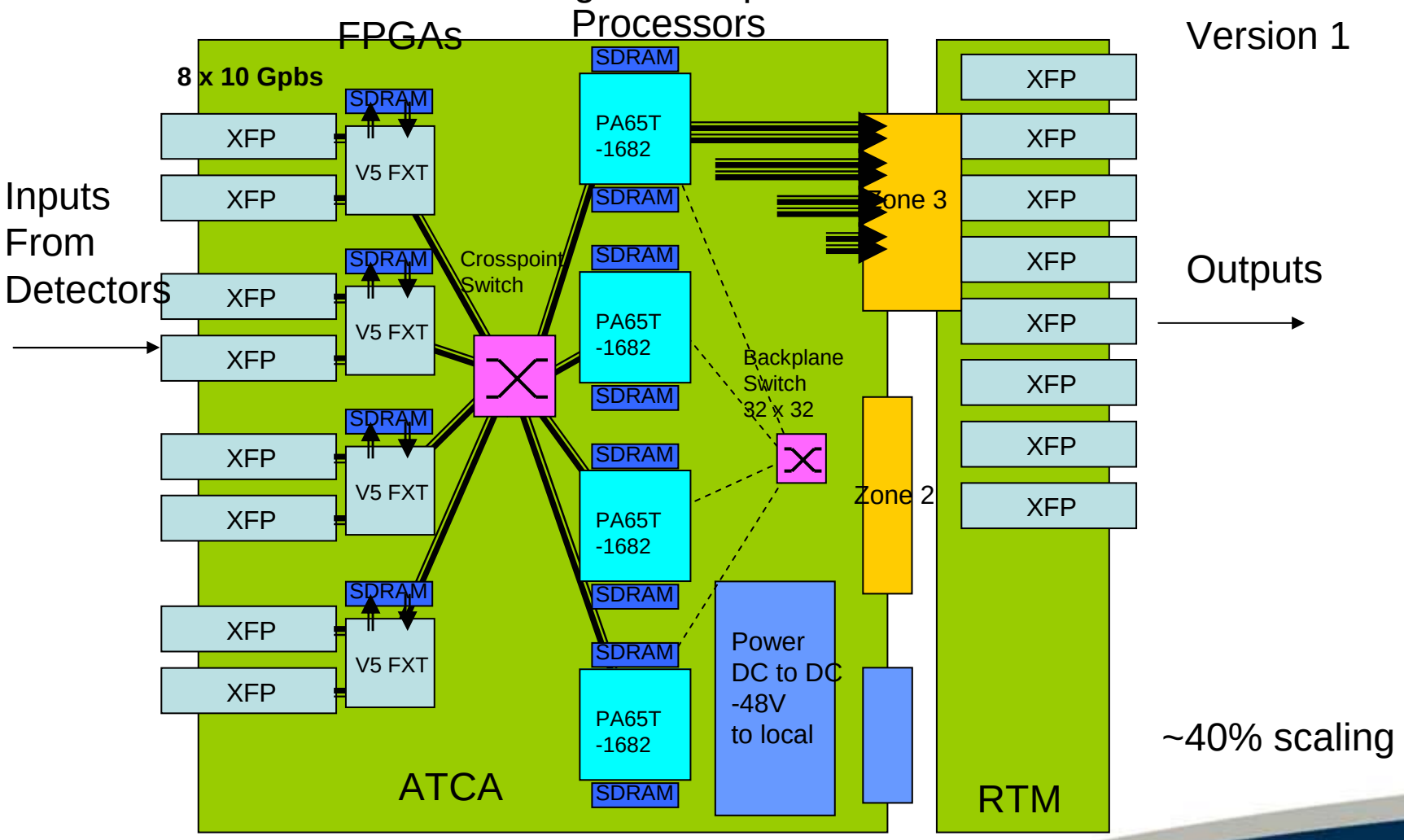
Put on Mezzanine.

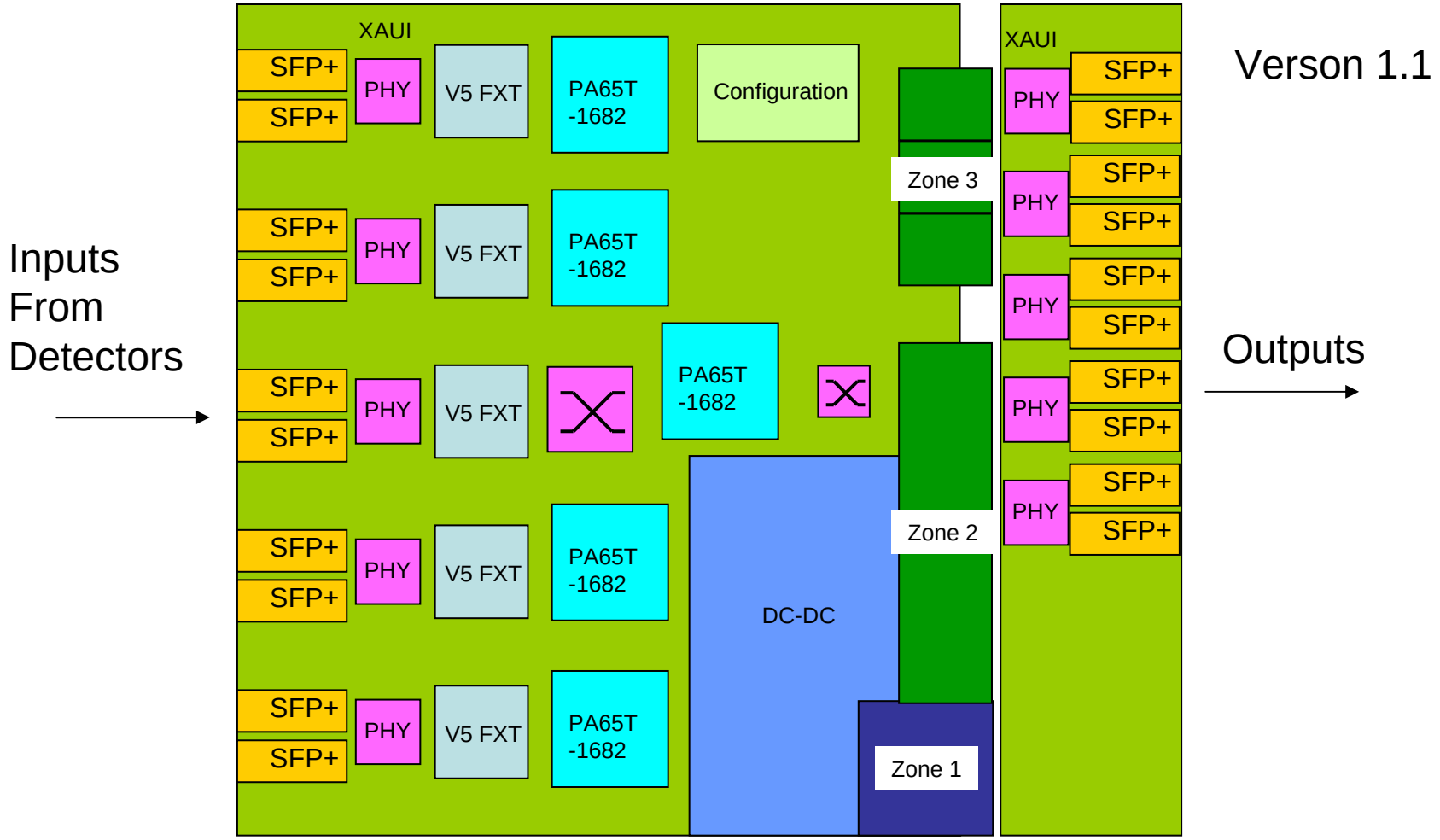


General Purpose GPUs NVIDIA Tesla 10 240 cores, Double prec FPt (put in Farm?)

Architecture

Train builder based on analogue crosspoint switch.





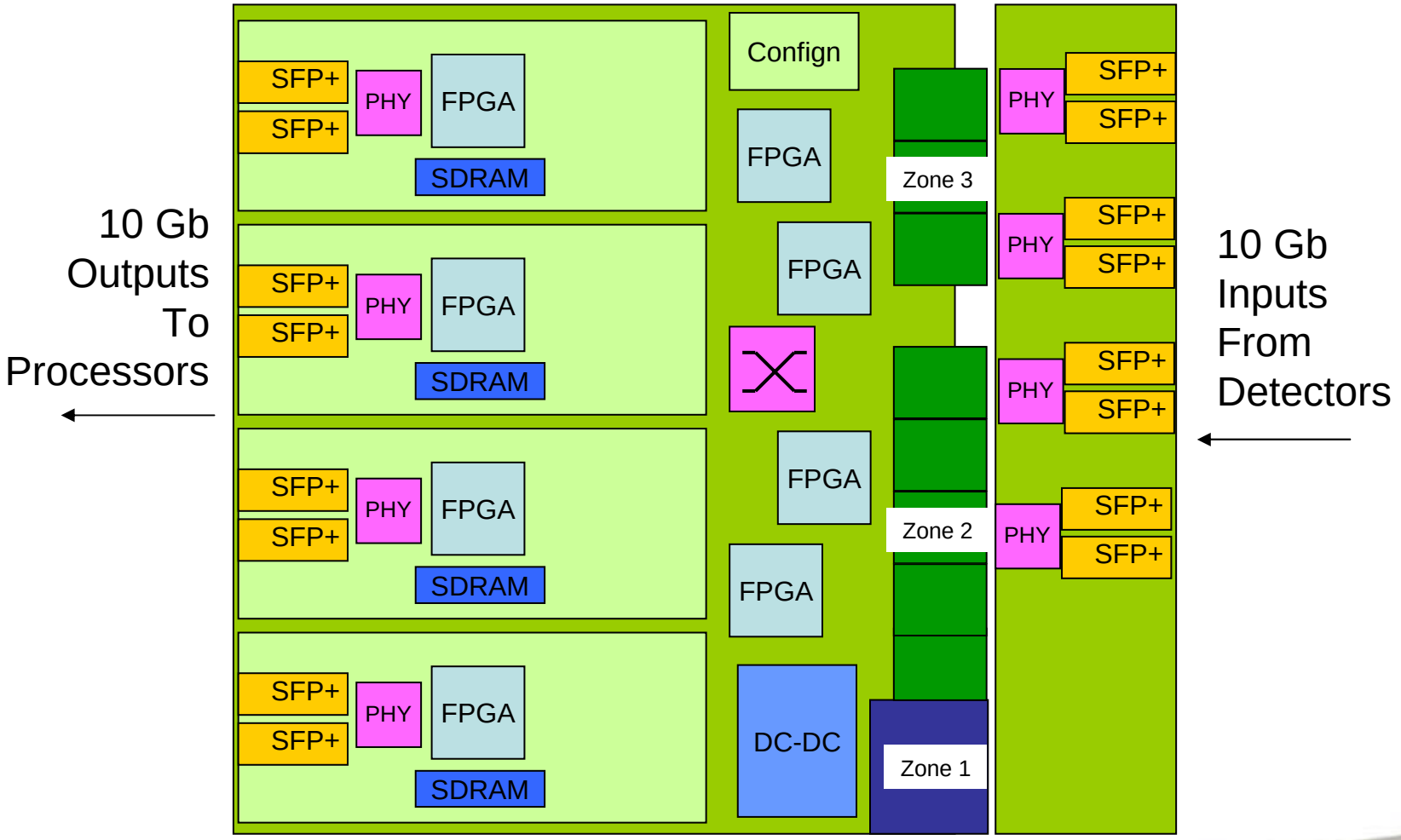
10 x SFP+ 10 GBytes / sec

Implementation on ATCA

With AMC mezzanines (up to 8)

Components under AMC?

SDRAM on component side 2

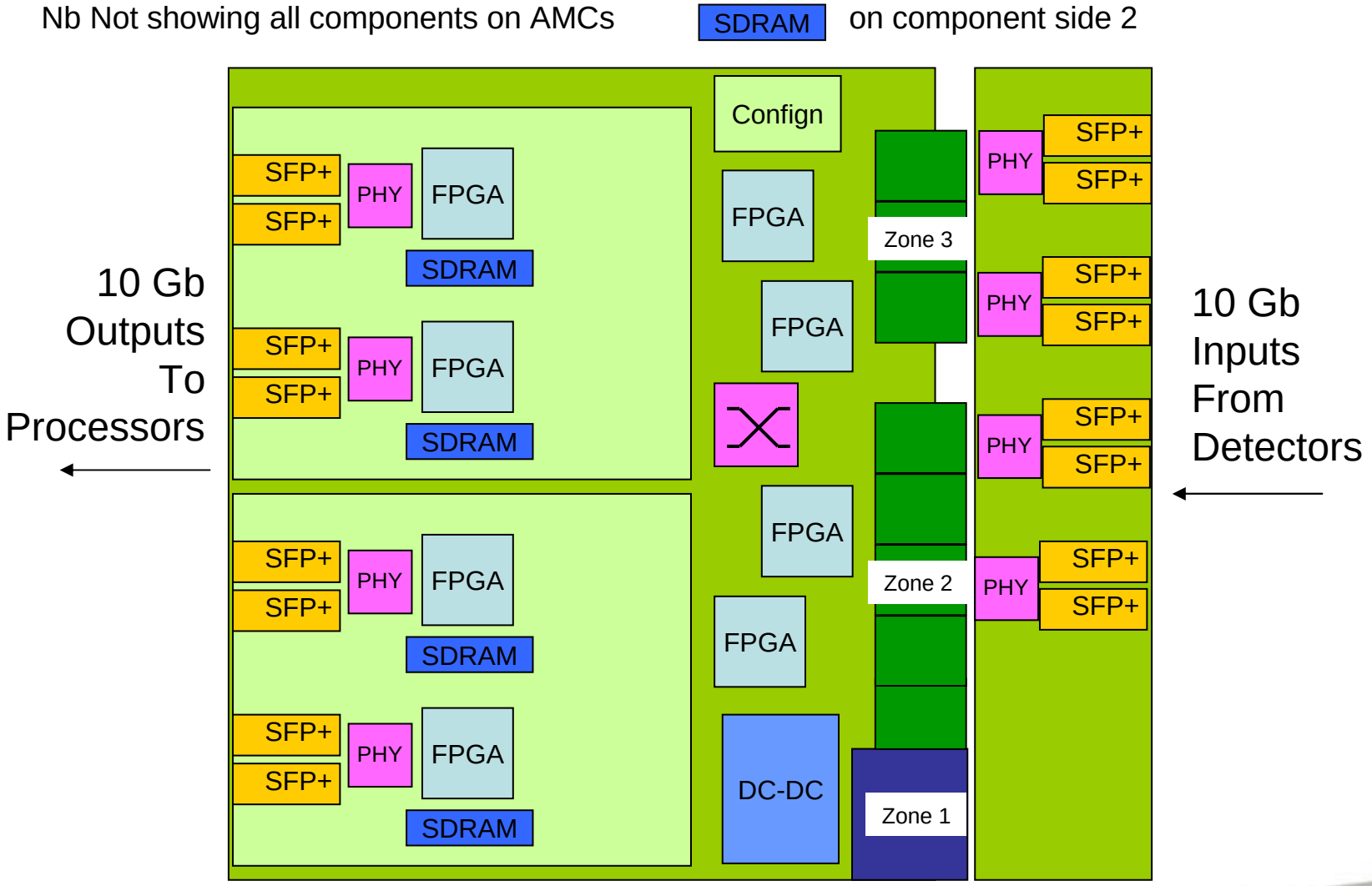


FPGA ½ MPix Train Builder
With Single Width AMCs

40% scaling

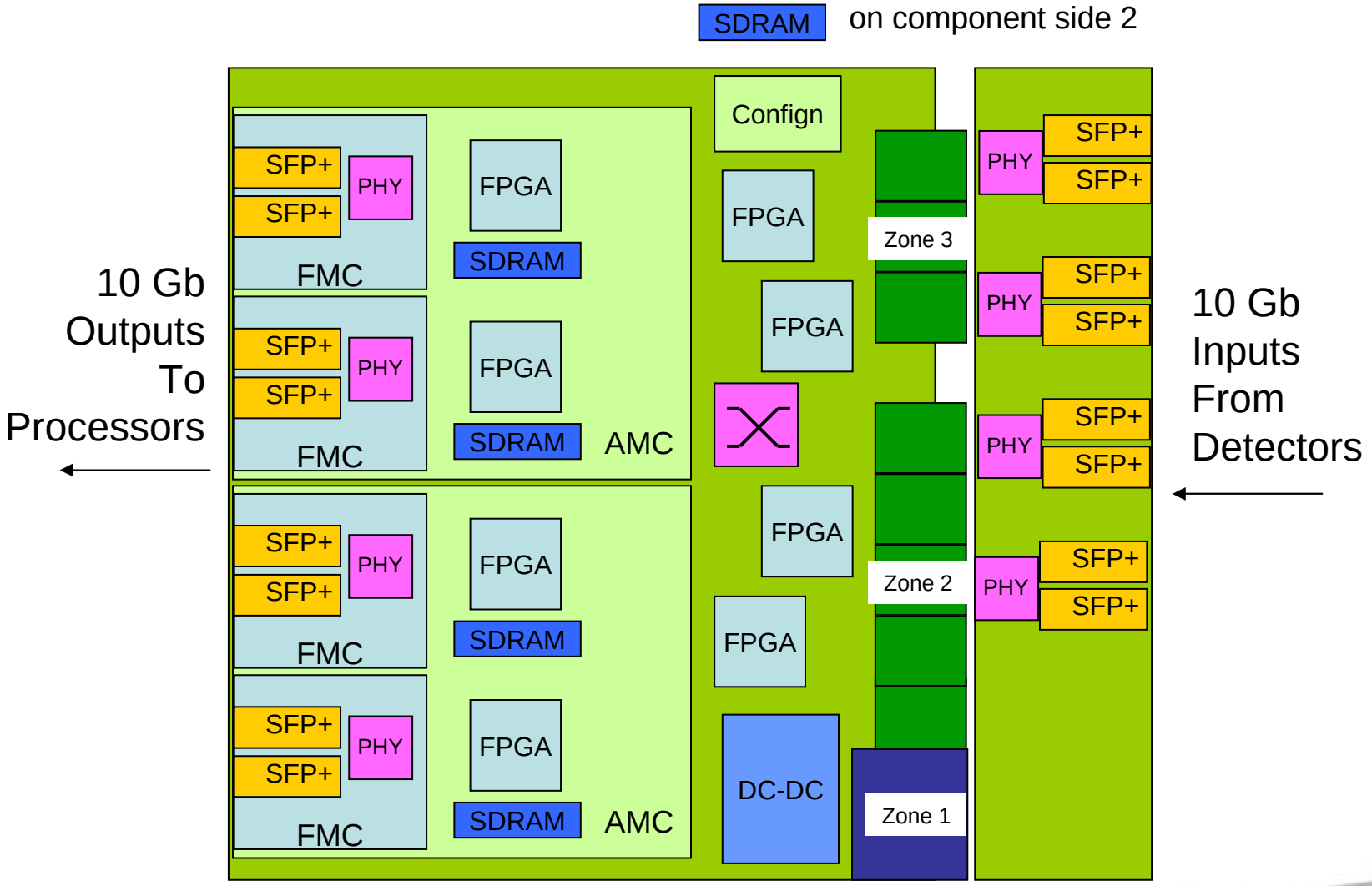
Nb Not showing all components on AMCs

SDRAM on component side 2



FPGA ½ MPix Train Builder
With Double width AMCs

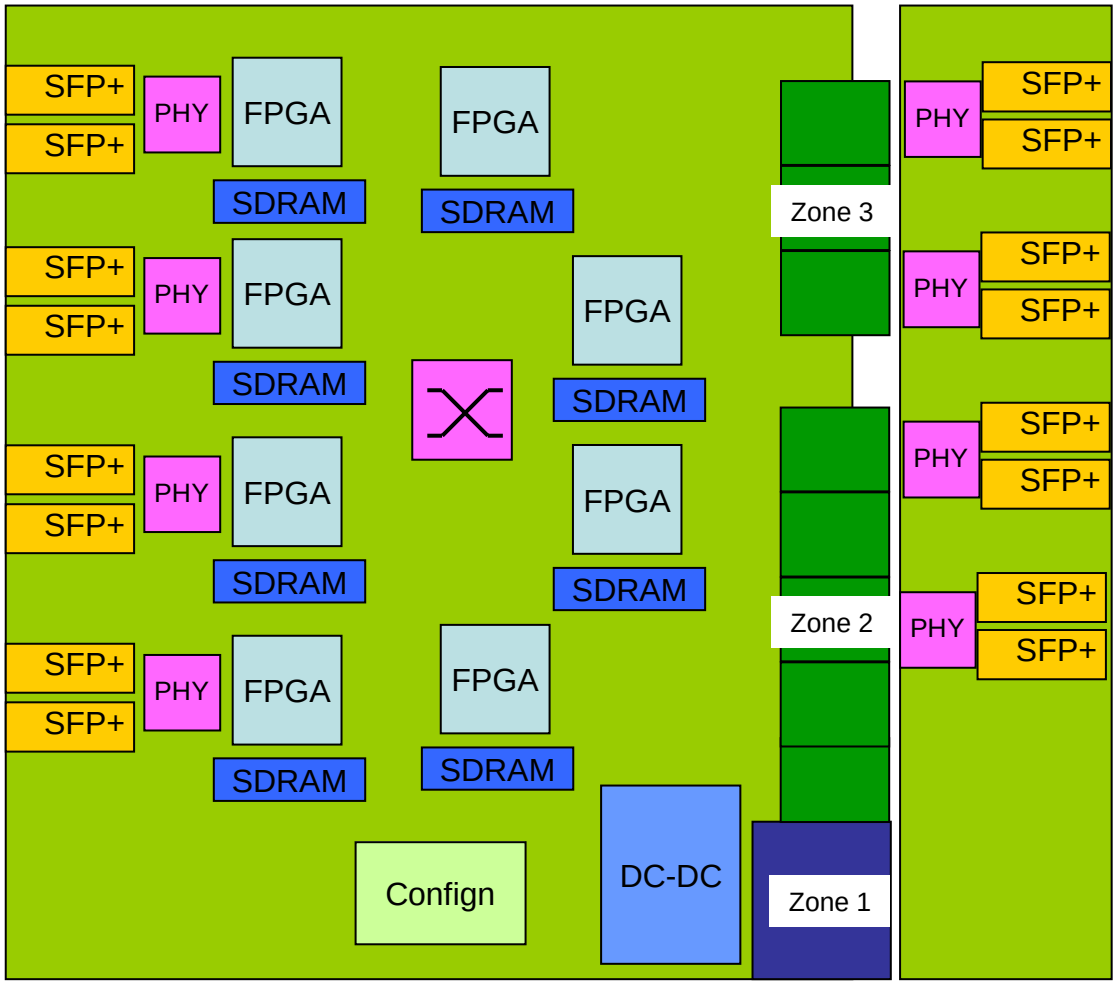
40% scaling



FPGA ½ MPix Train Builder
With Double width AMCs
And FPGA Mezzanines FMCs

40% scaling

FPGA Only Train Builder



10 Gb Outputs To Processors

10 Gb Inputs From Detectors

~40% scaling

Timing and Controls

Assume Train builder is purely data path.
Assume some TCS info in data stream.
Added at FEIs or Train builder (or both)

How would Train builder interface to TCS?
Via backplane from AMC card in same crate?

What data is needed. Train nr, expt id, run type, bunch info….
XFEL event format.
Standard headers.