Train builder meeting minutes

Minutes of 10th September 2008 meeting, 9:00-17:00, Cosener's House
Author: C.Youngman (Revised 16.9.2008)

1. People present

STFC & LPD: C. Angelsen, J.Coughlan, M.French, R.Halsall, W.Helsby, T.Nicholls and S. Taghavi.
DEPFET: A.Kugel.
HPAD: P.Goettlicher.
UCL: M.Warren and M.Wing.
WP76:C.Youngman.

Absent: S.Esenov (WP 76), I.Sheviakov (FEA) and M.Zimmer (FEA)

2. Agenda

- Review of last meeting and TB design status – JC
- Common timing and control aspects – CY
- 10GE development work progress and plans – MZ
- 10GE PC NIC tests – CY
- HPAD front end status – PG
- LPD front end status – JC
- DEPFET front end status – AK
- Software status; DOOCS etc. – CY
- TB in kind contribution discussion
- Common control interface in kind contribution discussion

The slides and minutes of the meeting are reachable at
http://xfel.desy.de/project_group/work_packages/photon_beam_systems/wp_76_daq__control/train_builder

3. Minutes

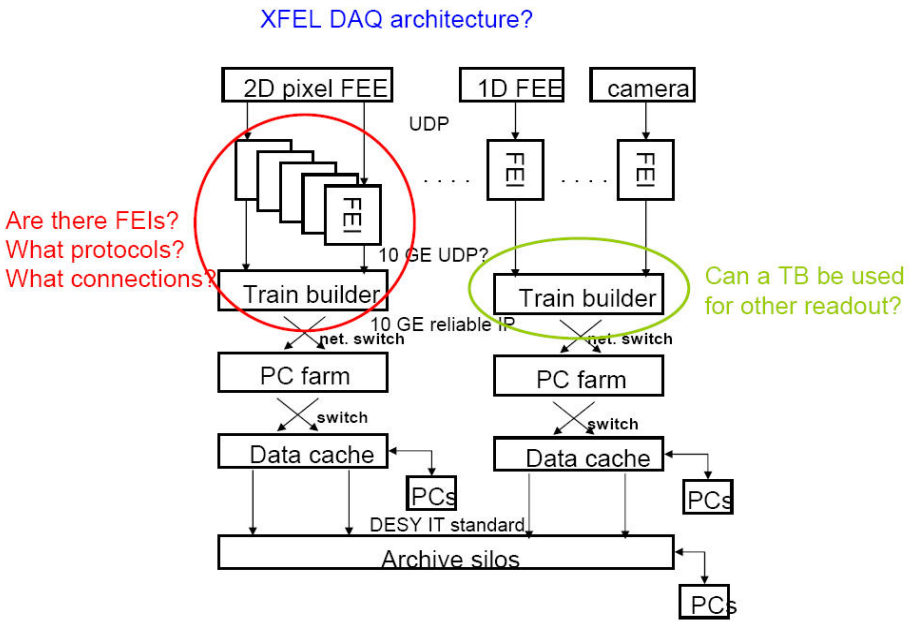
The aim of the meeting was to review the TB and detector front end status and move towards getting the in kind proposal submitted to the XFEL management. The same applied to the common timing and control interface, which is called the clock and control system in the minutes. This distinguishes it from slow control systems like HV and LV, which HPAD want to be provided by WP 76, see section 3.2.

3.1. Review of last meeting and TB design status

The following decisions were made w.r.t. the train builder and detector front end systems at the June meeting:

- 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 10GE contribution from DESY FEA.
- The TB design uses 10GE links over SFP+ fiber transceiver input and outputs modules. The link hardware is fully bi-directional.
- Two boards with each 8 inputs and 8 outputs are needed per Mpixel detector. The ½ Mpixel frames will be combined at the following PC layer. This matches the 16 x 10GE links now supported by all 2D detectors.
- The input data transfer protocol will be UDP without retries. The use of other protocols PGP, Aurora, etc. is not excluded.
- A maximum of 2 bytes per pixel and 512 picture frames per train will be inputted. The detector to TB fibre distance is ≤30m.
- The output protocol is TCP. The TB to PC farm distance is ≤300m (i.e. multi mode fibre)
- Network package loss should be minimal; the data encoding chosen should result in the smallest number of picture frames being affected if a packet is lost.
- The nominal XFEL bunch train repetition rate is 10Hz, but there have been plans for special runs with increased rates up to 30Hz. This can only be handled at the TB by reducing the number of maximum number of frames input.
- The option of providing on board processing (data reduction and compression algorithms) should be kept.
- The ATCA 8U form factor will be the baseline used to implement the TB.
- A common control signal source board should be foreseen.
- A common distribution system might also be possible.

The Train builder fits into the overall DAQ architecture as shown below



, where the 2D detector FEIs (Front End Interfaces) have been drawn but in reality are part of the front end systems of the detectors, they are not part of the TB. The idea of using the TB in the non 2D pixel detectors should not be forgotten, but it is unlikely due to the very much smaller train data volumes, ≤ 1% , originating from these detectors. The DAQ requirements of these detectors should be satisfied using a FEI to buffer data and send it by TCP to the PC farm layer.

The design and prototype development strategy should be a stepwise approach. Building small functionality demonstration and proving units using the AMC format is attractive as these can be exercised within a micro-TCA crate. The AMC can also be used as a mezzanine on ATCA board. The optical 10GE development of DESY-FEA, if packed onto a mezzanine, could be used as is. The work plan could look like:

- Investigate FPGA working with crosspoint switch on a development board. Points of interest: memory interface, board to board connections, serial protocols, working with the switch, …
- AMC demonstration board with FPGA+memory+crosspoint switch and two 10GE SFP+ links
- Full prototype ATCA board with RTMs
- Full production board

The UK in kind committee is very positive about the project and has given the go ahead to develop the project proposal. The XFEL management is very interested in pursuing the project as an in kind proposal and are waiting for the longer proposal, see c) below.

RAL has organized a loan of a Mindspeed crosspoint switch development board and is trying to get a similar system from Vitesse. The group has also been getting up to speed with TCA: quotes obtained for equipment as used by DESY, joined the TCA spec group PICMG, etc.

The following open questions (some are answered in the in kind contribution discussion) remain:

- What are realistic timescales for needing a full TB system?
- Is algorithm processing essential on the TB?
- How best to partition TB work packages? DESY 10 GE optical and UDP/TCP? + RTM?
- How do we share work practically? FPGA Firmware IP (intellectual property). PCB design. Tools.

The final design may evolve with experience on prototypes and emerging technologies, which depend on timescales. The need to stay receptive to improvements and alternatives remains.

3.2. Common timing and control aspects

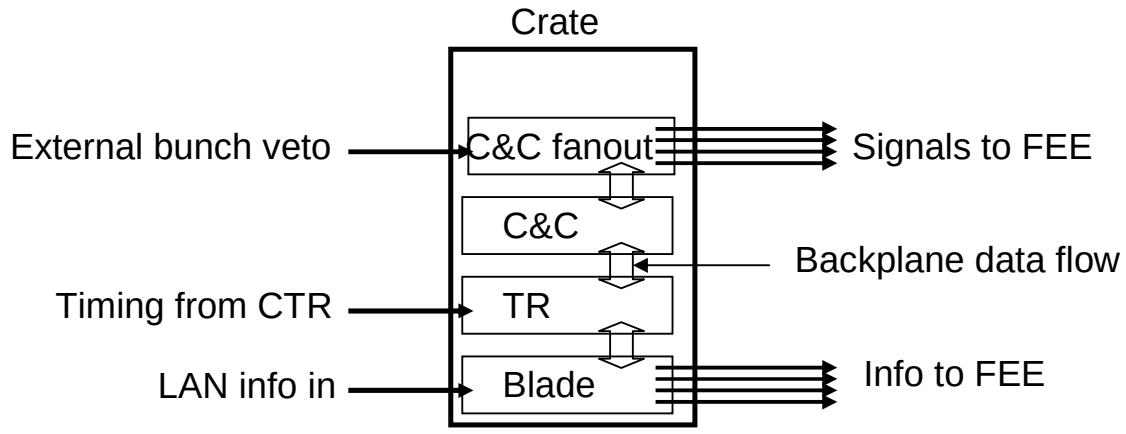
The photon bunch timing structure and the beam distribution structure currently foreseen at XFEL were described. The only change since the last meeting is that the maximum energy has been reduced to 22 GeV. The nominal repetition rate of 10Hz remains as does the possibility of running at higher rates.

The timing system was described. The 2D detectors will be connected to the XFEL machine timing system using Timing Receiver (TR) AMC boards being developed by Kay Rehlich's (DESY-MCS4) group and a Swedish in kind contribution. The TRs are connected to Central Timing Relays (CTR) which are connected to the master timing units. By using round trip delay measurements the TR is expected to generate triggers with jitters in the pico-sec range, well within the nano-sec requirements of the 2D detectors. Petr Vetrov of Manfred Zimmer's FEA group has been involved in the design of other AMC boards and one of these will be the development board used for the TR.

The TR being developed, probably, looks similar to the IP-module used at FLASH, but little documentation exists. In this case the TR is synchronized to the machine timing system and generates trigger pulses (TTL on the equivalent FLASH TR) and interrupts to a host PC (by PCIe on the AMC board?). The interrupt can be used to trigger a read for telegram information (a start bunch train trigger might be accompanied by bunch pattern information) sent with the generating trigger event. Note that this, telegram, information may also be distributed via LAN.

The long term environmental (temperature) stability of candidate chipsets to be used in the TR have been made and a first prototype of the TR can be expected in the Spring of 2009.

The signals, clocks and information required (see slides) by the detectors were listed and a proposal for a common detector clock&control (this name is less easily confused with the overall control system!) system made, see schematic below. The blade PC collects (LAN or telegram) bunch related data, the TR is the timing interface, the C&C generates/processes clock and other signals as needed (5MHz, bunch train from TR, …) , and the C&C fanout distributes all signals to the front ends.



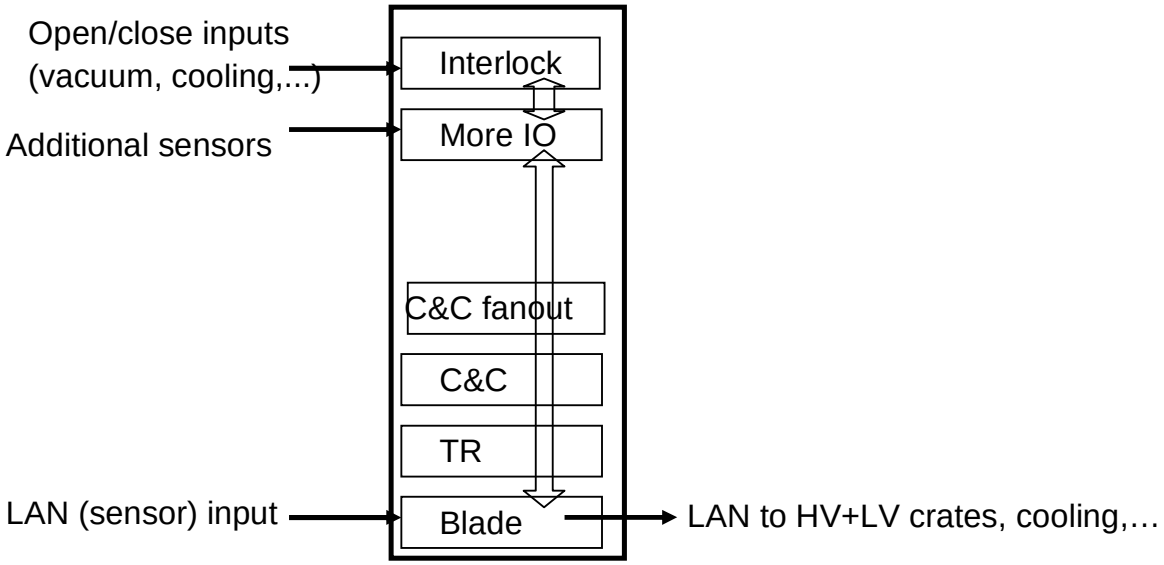
Open questions arising from the discussion were:

- What happens to the C&C system when 1Mpixel detectors are replicated (one C&C and many fanouts?)
- The detector side inputs and outputs need to be fully defined including connectors, etc.
- The TB will also need inputs and outputs to the C&C, they need defining.
- A definition of what the blade PC is and what it does (run control, sets up calibrations, etc.) is required.

Points on how to proceed with the common C&C system:

- It was proposed that UCL should lead an in kind contribution to design and implement the common C&C system.
- need an agreement with all detectors about connections: cable, protocols, etc.
- need a spec. of the final system
- how to interface to other timing systems at other light sources
- interact with Rehlich's group to get precise technical information/documentation timing, is additional functionality required.

The next part of the talk concerned the auxiliary systems needed by the detectors, like HV, LV, etc., and called here slow control components. Heinz Graafsma has asked WP 76 to provide these systems for HPAD and had suggested that they be common systems – it was realized during the talk that this had not been suggested to LPD and DEPFET ! The slides show the current planning for slow control systems for HPAD, which will be discussed at their collaboration meeting 7-8.10.2008. The important point for this meeting is that the proposed hardware system fits into the above crate as shown below and will be driven by the blade PC. Note that the cooling and vacuum components are detector specific and have to be provided by the detector (and will need to comply to interface APIs), whereas the other components (HV, LV, interlock, …) are potentially common.



3.3. 10GE development work progress and plans

Due to an accident at home Manfred was not able to attend the meeting. This was a pity as he'd prepared some nice slides.

FEA's interest in 10GE is to get hands on experience so that a common implementation for the TB and HPAD readout systems can be provided. This effort focuses on using the Xilinx evaluation (e.g. ML510) board's pluggable personality module (XPM) which is large enough and has sufficient RockeIO and lvds connections to host two 10GE channels plus HPAD ADC test systems.

For the 10GE development the Vitesse VSC8486 PHY chipset has been selected to connect the evaluation side XAUI input lines and SFP+ output transceivers.

The status of the proposed XPM is as follows: design and layout is ready, components are ordered but have long delivery times, and UDP software for the Virtex5 is written and tested as is the memory controller software. If all goes according to plan the board should be available for testing in Oct with results by Nov./Dec.

Manfred pointed out that the XPM functionality was similar to that need by the TB's Rear Transition Module (RTM). Additional design effort is required at the RTM (dual channel PHYs, higher density fibre connectors, ATCA zone3 connector selection)

The real question which was left outstanding was:

- Would FEA be willing to do the additional TB RTM design work?

3.4. 10GE PC NIC tests

Two high performance PCs have been purchased (Mainboard D5400XS SKULLTRAIL S771 E5400 EATX, 8 cores dual XEON E5420 2.5GHz FSB1333, Memory 16GB fully buffered 800MHz DDR2, PCIe 1.2) as have two 10GE NICs (a Chelsio S310E-SR+ SFP+ and a Intel Pro/10GBE XF SR XFP). The Chelsio has TOE (TCP offload hardware). It is intended to use the systems to gain experience with 10GE NIC and to use them with the FEA XPM tests.

The systems were put together last week and the two NICs connected back-to-back to each other. A series of quick throughput tests have been made. Some points to be noted:

- OpenSUSE 11 recognized the NICs and loaded default drivers
- It was later realized that the Chelsio driver was not TOE enabled – an update of the driver and a plain vanilla 2.6.23 Kernel was required to get TOE functionality.
- The Intel driver was later updated according to Intel's web site
- Manufacturer system parameter tuning scripts were used – did little?
- Iperf and ttcp test tools produced interesting results and sometimes non believable results.
- At the time of the talk Sergey had invested more time and was getting TCP throughputs of ~800MB/s – but the results should be taken with a pinch of salt as more work is required. UDP write speeds of ~700MB/s were also seen – again more work is required.
- No measurements were made of cpu usage, packet loss, etc.
- Jumbo frames (1500 – 9000 bytes) were played with but appeared not to improve performance. The Intel board supports 32kB Jumbo frames and the Chelsio ~9000.

Additional work is required to get a good understanding of the NIC/PC performance. One additional board of each type will be purchased type so that like-against-like test can be made, currently it is impossible to disentangle the differences in behavior of the two.

3.5.HPAD front end status

Recent changes to the HPAD front end system were reviewed. The mechanical construction (issues: cooling, vacuum, assembly, …) of the front end is being worked on by the collaboration, which effects the control electronics, connection paths, etc.. The idea is currently to have a backplane per quadrant in the forward region of the detector (between ASIC and ADC) and to have common control signals and power distribution of the 250A/quadrant handled there. There is no change to the 16 output links to the TB. More precise information should be available after the 7-8.10.2008 collaboration meeting.

3.6. LPD front end status

The original roof tile geometry of the silicon sensors has now been replaced by a flat tiled geometry. The new mechanical arrangement allows 16 super module readout cards to be used, which is important as this now matches the TB 16 fibre SFP+ channels input requirement per Mpixel. The readout card size remains the same. The increased number of ASICs per card requires a larger FPGA.

The design of the ASIC , its operating modes and interfaces were discussed. Unlike HPAD the analogue part is in the ASIC. The time budgets for data readout rates through from the ASIC through to the backend were shown – the expected link output rate to the TB is 4.7Gbps.

The LPD is expecting the bunch pattern to be delivered with the other fast signals, e.g. bunch start, at a fixed time before each bunch train (order of a few millisecs). This could be restricted to a fixed number of different patterns for a given data taking period, which allows the used pattern to be selected using an index. The need for additional processing on the TB, such as pedestal subtraction, has yet to be decided.

The time schedule for detector development might look like:

- Prototype sensor modules ; 2009 ; COTS dev boards
- First Super module; 2010 ?; 10 Gb TCP link to PC and/or AMC prototypes
- 1 MPix Detector; 2011 ? ; 10 Gb TCP link to PC and/or AMC prototypes
- Full readout 1 MPix in beam test *LCLS*? ; 2012? ATCA prototype
- 1 MPix in XFEL beam ; >2014? ATCA production
- Possible 16MPix in XFEL beam ; >2016? ATCA production

3.7. DEPFET front end status

The current status of: sensor, ASIC, readout and mechanical design were shown.

The readout concept using modified PCIe boards derived from previous ATLAS developments was shown – work is currently underway to incorporate the Vitesse VSC8486 PHYs layer which will allow interfacing the DEPFET readout system to the TB. The layout modifications required are done, FPGA IP-cores for 10GE and XAUI are available, and when the Vitesse chips are delivered test of UDP on 10GE can be started.

The clock and control requirements were restated. A JTAG or I2C protocol could be used to distribute the signals down to the front end(?).

Functionality that needs to be provided by the TB for DEPFET is:

- Arrange UDP frames: Unpack eventually non-word aligned data (e.g. 12 or 14 bit samples)?
- Per-pixel offset/gain compensation: 10 GbE => process 2 pixel @ 250MHz in parallel 16 bit offset; 16 bit gain coefficient=> LUT 64k * 32 * 2 pixel: DRAM or SRAM
- Control signal fanout
- Control message fanout via UDP

The last two items on the list items caused some discussion. There are currently two camps of thought: 1) do not mix data and control (i.e. HPAD and LPD and WP76) and 2) mix them DEPFET. This issue was not resolved at the meeting.

The current DEPFET time schedule is:

- Spring 2009 – 10 GbE prototype tests

- Mid/Fall 2010 – 10GbE/UDP tests (single chip)
- Spring 2011 – XFEL Control Interface
- Late 2012 – 64k pixel module test
- Late 2013 – 1M detector test

3.8. Software status; DOOCS etc.

One slide was shown. The baseline is that DOOCS is not an out-of-the-box system and requires considerable work to get run control system going. A meeting with the DOOCs people is being organized to discuss what their plans are for the next years regarding XFEL preparation, whether we can help, etc. The RAL GDA people expressed their readiness to attend the meeting.

3.9. TB in kind contribution discussion

As a guide to the discussion of the TB in kind contribution JC showed Mindmap slides which breakdown the project work packages, groups, etc. A single summary view is appended to the minutes.

The proposed TB development schedule was discussed and looks like:

- 2009 evaluation board prototyping (FPGA, crosspoint switch, memory, protocols)
- ?/? – small prototype with mezzanine boards
- Q2/2012 - full prototype ATCA board with RTMs
- ?/2013 - full production board

It needs to be clarified if FEA can lead or at least participate in the RTM (routing…) design.

The TB in kind contribution will deliver two 1/2Mpixel TB boards for each detector (6), one spare board for each detector (3), one board to each detector readout development site (RAL, DESY, HD) (3), etc. A total of 15 – 20 1/2Mpixel boards are likely to be required initially. Boards required later will have to be purchased from the manufacturer. Regarding maintenance and repair support work, the RAL group will provide support within the normal limits associated with maintaining such equipment. It is clear that experience of in the filed operation is required to really understand the maintenance issue.

Who provides and pays for the ATCA crate needed by the TB was thought to be outside the scope of the TB in kind contribution and WP 76 will pay this cost as part of the backend system.

The question of how FTE are costed was raised. An XFEL FTE has a lower cost that a UK based engineer. It was decided to calculate a real cost and expose this as a shortfall which the funding agencies will have to fund.

The in kind long proposal for the TB should be finalized by the end of the year.

3.10. Common C&C in kind contribution discussion

A possible control and timing in kind contribution lead by the UCL group was discussed. It was agreed that such a contribution should be aimed for and all groups were happy with ICL taking the lead in this. A short paragraph proposing the contribution was formulated, the UK in kind committee will be asked for their OK and then this will be sent to the XFEL management.

Here is a brief reminder of how an in kind contribution is turned from a proposal into an accepted project:

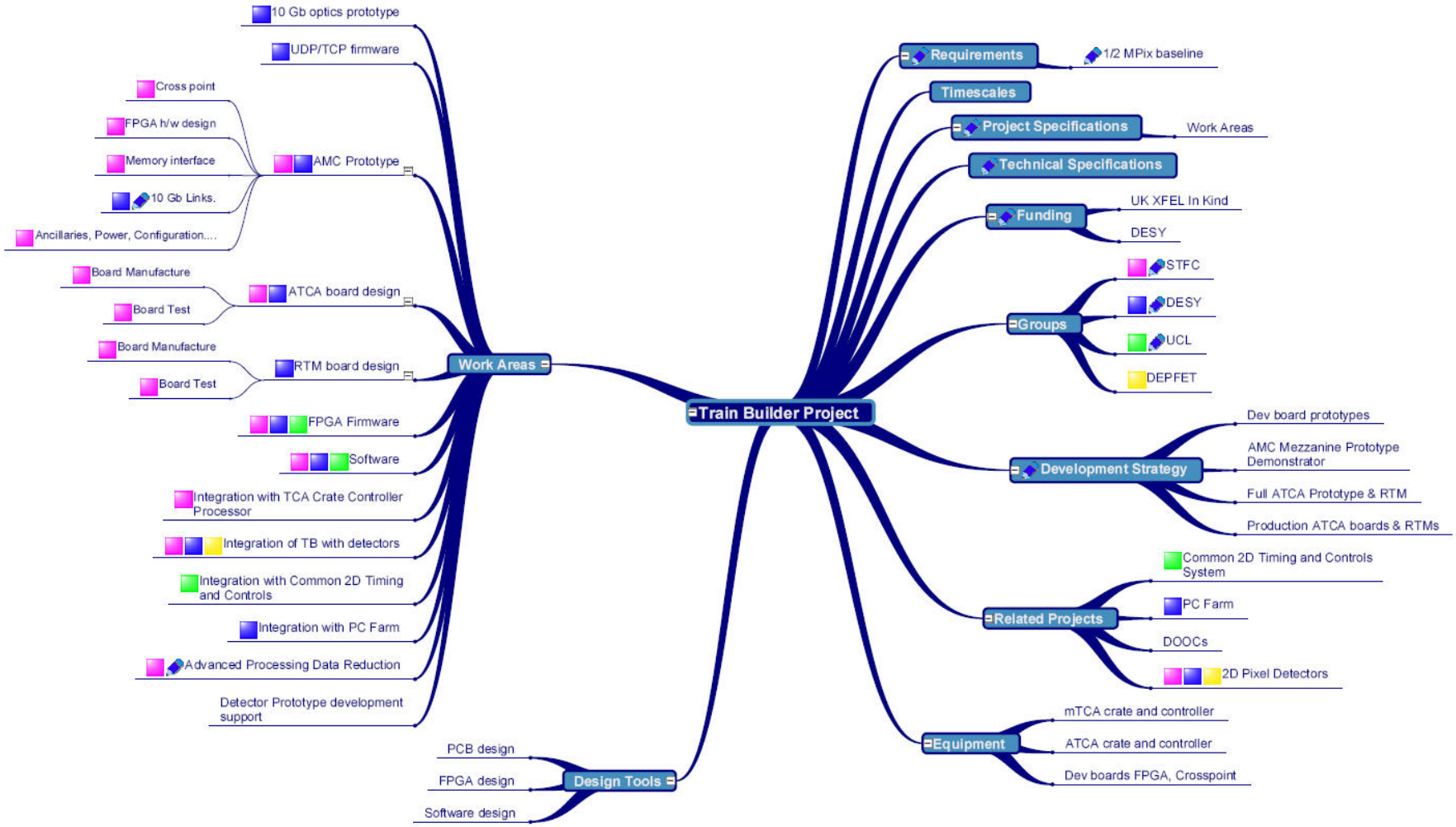
a) A short, single paragraph, description of the proposed in kind contribution is sent to Andreas Schwarz (XFEL director).
b) The short proposal is reviewed and a decision is made on whether the contribution is useful to XFEL.
c) If the decision is positive a longer, 2-3 page, proposal of what will be done including: a detailed description of what the project will provide (initial design ideas), development time schedule (prototypes to final product), FTEs required, resources needed, and deliverables that the contribution will produce. Note that there are detailed rules, including what expenditure can and cannot be included in an in kind contribution (like no travel money)
d) The longer proposal is sent to Andreas, is reviewed by the In Kind review Committee, and is passed to the steering committee for their OK
e) A contract is drawn up and signed.

4. Conclusion

The next meeting will be on the 4th Dec at DESY. It will be interesting to see how close to completion the in kind proposal for the TB will be and perhaps results from the XPM evaluation board tests, etc. Hopefully we will also see the understanding of the C&C system growing.

The following non ordered list of issues came up during the meeting:

- GDA representatives (Bill and Geoff) should attend the Oct/Nov DOOCS software meeting , they have experience of providing open out-of-the-box control software.
- There is an outstanding offer for K.Rehlich to visit Diamond and see GDA in action.
- A test of RAL's video system with DESY is needed.
- Matt Warren and Martin Postranecky should visit DESY Oct. or Nov. to meet and discuss with the timing and other involved groups the common timing and control system.
- Is FEA willing to be responsible for the RTM development?
- Resolve the issue of sending control data back through the readout link.
- Finalize by the end of the year the TB in kind long proposal.


Train Builder Project
Work Areas
10 Gb optics prototype
UDP/TCP firmware
AMC Prototype
Cross point
FPGA h/w design
Memory interface
10 Gb Links.
Ancillaries, Power, Configuration....
ATCA board design
Board Manufacture
Board Test
RTM board design
Board Manufacture
Board Test
FPGA Firmware
Software
Integration with TCA Crate Controller Processor
Integration of TB with detectors
Integration with Common 2D Timing and Controls
Integration with PC Farm
Advanced Processing Data Reduction
Detector Prototype development support
Design Tools
PCB design
FPGA design
Software design
Requirements
1/2 MPix baseline
Timescales
Project Specifications
Work Areas
Technical Specifications
Funding
UK XFEL In Kind
DESY
Groups
STFC
DESY
UCL
DEPFET
Development Strategy
Dev board prototypes
AMC Mezzanine Prototype Demonstrator
Full ATCA Prototype & RTM
Production ATCA boards & RTMs
Related Projects
Common 2D Timing and Controls System
PC Farm
DOOCs
2D Pixel Detectors
Equipment
mTCA crate and controller
ATCA crate and controller
Dev boards FPGA, Crosspoint