First measurements with the CMS DAQ and Timing Hub prototype-1

September 5, 2019

Jeroen Hegeman on behalf of the CMS DAQ group

TWEPP 2019
Topical Workshop on Electronics for Particle Physics
CMS Phase-2 DAQ overview
Phase-2 CMS trigger-DAQ architecture choice: Stick to the proven two-stage trigger approach, but at higher rate and throughput.

- **Pile-up = 200:**
  - L1A rate: 750 kHz
  - Event size: 7.4 MB
  - Throughput: 44 Tbps
  - HLT output rate: 7.5 kHz

---

**Phase-2 CMS baseline DAQ**

**Trigger Processors**

**Global Trigger**

**Subdetector front-ends**

**Subdetector back-ends**

**Data-to-surface (D2S) network**

**Event-building**

**High-Level Trigger** (SW)

**Storage**

**Data to Surface routers**

**~ 500x200 Gbs switch**

**~ 5 PB local storage**

**60 GBs access**

**120 ATCA crates**

---

**TTC/TTS**

**TCDS/EVM**

**Trigger Processors**

**Global Trigger**

**Storage**

**HLT PC farms/clouds**

**~ 9.2 MHS06**

---

**Event networks**

**Detector Front-Ends (FE)**

**~ 5,000 IO servers**

---

**5x100 Gbs data links**

**40x100 Gbs switches**

---

**Data flow**

---

Jeroen Hegeman

TWEPP 2019
Phase-2 CMS DAQ and Timing Hub (DTH)

DTH: ATCA hub card housed in subdetector back-end crates
Bridge between the synchronous world of L1 trigger and detectors and the asynchronous DAQ world

Diagram:
- Slot numbers: 13, 11, 9, 7, 5, 3, 1, 2, 4, 6, 8, 10, 12, 14
- Point-to-point DAQ (front-panel, optical)
- 40 MHz & 320 MHz clocks (backplane)
- TCDS2 bidirectional (backplane)
- 1GbE for control (backplane)
- 2x10GbE (control)
- 4x100GbE (DAQ D2S)
Phase-2 CMS DAQ and Timing Hub (DTH)

**DTH: ATCA hub card housed in subdetector back-end crates**

Bridge between the synchronous world of L1 trigger and detectors and the asynchronous DAQ world

- Standard ATCA hub providing crate-level network connectivity
- Distributes clock signals, synchronization commands, L1 trigger, and flow control commands from central systems to back-ends
- Bi-directional 10 Gbit/s backplane serial links
- Receives event fragments from back-ends, concentrates and buffers these, and transmits them to the data-to-surface network
- Custom point-to-point links at 16 Gbit/s and 25 Gbit/s via the front panel
- Provides functionality for stand-alone ‘mini’ DAQ and TCDS
- Built-in support for DAQ and TCDS link testing and monitoring
DTH development program
The multi-prong DTH development program (1)

Four ‘units’ to prototype:

1. Baseline DAQ unit providing 400 Gbit/s throughput
2. Baseline TCDS unit handling clock recovery, cleaning, and distribution
3. Ethernet switch with $2 \times 10\text{GbE}$ uplink, providing $1\text{GbE}$ to all node slots
4. ATCA baseboard with power, IPMC, etc.

Current plan for the final system foresees a pair of boards:

1. A DAQ and TCDS Hub $(1 + 2 + 3 + 4)$
2. A DAQ board $(2 \times 1 + 4)$
The multi-prong DTH development program (2)

- **DTH prototype 1 (P1), version 1 (V1)**
  Proof-of-principle platform for DAQ (1), TCDS (2), and baseboard (4) instrumented for clock distribution studies
  
  *Q2 2019: Hardware functional, proof-of-principle well under way*

- **DTH prototype 1 (P1), version 2 (V2)**
  Evolution of P1 V1 for system integration, HMC replaced by DDR4
  
  *Q3 2019: First boards in production, firmware and software utilities being developed in context of the above*

  *Q2 2020: Ready to deliver and support P1 V2 for subsystem integration*

- **Ethernet hub board**
  Test platform for Ethernet switch functionality (3) (hardware and software)

  *Q3 2019: First board assembled, to be tested before software evaluation*
DHT P1 V1 performance evaluation
Presented here are some of the hardware verification tests and measurements related to front-panel and backplane connectivity, and to clock and signal quality.
DTH P1 annotated

- COM-Express on-board controller
- DAQ FPGA
- TCDS FPGA
- 4 × D2S QSFP+
- TCDS SFP+
- 48 V A/B select and 48 V to 12 V
- J20 – J24 (TCDS)
- P10 (power)
DTH P1 V1 QSFP+ connections successfully tested using a 100GbE-equivalent PRBS31 stream into a commercial NIC. Ran for 24 hours without errors.

‘Version 0’ point-to-point sender IP core in alpha testing by back-end developers (including a simple receiver core)

Proof-of-principle TCP/IP firmware exists based on in-FPGA UltraRAM

Firmware development in progress to merge multiple TCP streams in order to efficiently use the data-to-surface network
Simplified DTH P1 clocking scheme

CMS Phase-2 baseline clock distribution path:

- Embedded in the TCDS serial stream from the TCDS master to the DTHs
- Recovered and cleaned clock(s) transmitted on the backplane from the DTH to the back-ends
- Embedded in the (lp)GBT serial stream from the back-ends to the on-detector electronics
Comparing 40 MHz master clock derived from all clock inputs

- Free-running Si5345 disciplined by local TCXO shows significant low-frequency phase noise
- Phase noise on master clock derived from any of the LHC clock input paths is dominated by the external generator
- ‘Grass’ in the 100 Hz to 5 kHz range suspected due to imperfect design around the TCXO. Expected to improve on V2.
High quality LHC clock achieved on the backplane

- Phase noise in all backplane slots comparable
- Minor variations in plateau above $\approx 10$ kHz due to frequency-dependent attenuation on the backplane
- RMS jitter integrated from 1 Hz to 20 MHz: worst case $< 5$ ps
- Two groups, originating from
  - IC53: disciplined by a crystal
  - IC55: disciplined by a TCXO
- RMS jitter integrated from 1 Hz to 20 MHz: worst case < 4 ps
- TCXO seems to introduce unwanted plateau from 5 kHz to 10 MHz
- Signs of high-frequency crosstalk between 40 MHz and 320 MHz clocks in the same slot (Trivially cleaned by back-end PLL)
TCDS – Dual-DTH clock recovery exercise

- Sender DTH driven by external 40 MHz clock generating 10.24 Gbit/s PRBS31 stream on GTH SFP
- Receiver DTH recovering 320 MHz and 40 MHz clocks from received serial stream
TCDS – Recovered 40 MHz and 320 MHz clocks

- Quality of recovered 40 MHz clock on second DTH comparable to original master clock
- Even the raw, recovered 320 MHz clock straight from the MGT has acceptable quality

Note: These results clearly show several undesirable features of the TCXOs. Performance expected to improve with plain crystals.
Backplane TCDS links

Slot 5 to DTH (GTH), with RX retimer

Slot 7 to DTH (GTH) without retimer

DTH ↔ DTH IBERTs at 10.3125 Gbit/s

- Done in both directions
- RX retimers clearly pay off (even with default configuration)
- Effect of Tx re-timer configuration on overall performance to be studied in more detail
Backplane TCDS links

Scope eye diagrams at 10.3125 Gbit/s in node slots

- All slots function well as 10 Gbit/s serial links
- Differences between GTH and GTY are of similar size as slot-to-slot variations.
Summary and short-term plan
• The DTH development program is well on its way:
  • The P1 V1 hardware is fully functional. Basis for characterization measurements and proof-of-principle firmware and software development.
  • The P1 V2 will add DDR4 test and development possibilities and fixes some minor V1 flaws. DAQ and TCDS integration platform for back-end developers.
  • The Ethernet switch board is about to be functionality-tested. Platform for commercial software evaluation and integration.

• All P1 V1 characterization measurements performed so far:
  • Reinforce the feasibility of the baseline clock distribution scheme via the encoded TCDS stream
  • Confirm the feasibility of driving several 100 Gbit/s TCP streams out of a single FPGA
Next steps

The short-term DTH plan targets a few remaining open questions:

- **DAQ:**
  - Expand proof-of-principle to full 100 Gbit/s on VCU128 using the in-FPGA HBM
  - Solidify the DTH throughput estimate

- **TCDS:**
  - Verify measurements using a true lpGBT link
  - Prove phase stability of the distributed clock across SerDes resets and FPGA reloads (see also the ‘TCLink’ contribution by E. Brandao)

- **DTH project:**
  - Deliver the P1 V2 to subsystem integrators with adequate firmware and software utilities
Thank you
DTH P1 technical summary

- Dual KU15P: one for DAQ, one for TCDS
- COMExpress on-board controller running CC7
- CERN IPMC
- Main connections:
  - DAQ:
    - Six SLinkRocket FireFlys (6 × 16 Gbit/s or 4 × 25 Gbit/s)
    - Four Ethernet QSFP+s (4 × 100 Gbit/s)
  - TCDS:
    - Two SFP+ inputs connected to a GTH and a GTY
    - 10 Gbit/s backplane TCDS links to all node slots and the second hub slot
    - 40 MHz and 320 MHz clocks distributed to all node slots and the second hub slot
As part of the R&D towards the final DTH the HMC on the P1 V2 has been replaced by DDR4 (without changing FPGA!)

- Based on number of GPIO pins could allocate eight 16-bit chips
- At 900 MHz this gives a bandwidth of 230 Gbit/s
- Sufficient for a single 100 Gbit/s proof-of-principle Ethernet link

Note: Given the recent good news on Virtex UltraScale+ pricing, will now first look at using VU35P HBM in-FPGA memory
For completeness – General test conditions

- All ‘simple’ tests done with a single board in a two-slot mini-crate
- All ‘less simple’ tests done in a CERN-standard ATCA crate, powered by a CERN-standard 48 V supply
- Slots numbered from left to right: 13, 11, 9, 7, 5, 3, 1, 2, 4, 6, 8, 10, 12, 14
- DTH in slot 1 (i.e., the first hub slot)
- For two-DTH tests: second DTH in slot 7
- Most clock studies done with round numbers (e.g., 40 MHz instead of the real LHC bunch clock frequency). Irrelevant for characterization, but convenient for calculations and instruments.
- Backplane link tests done at 10.3125 Gbit/s with default retimer settings
External clock: 40 MHz generated by a Silicon Labs Si5345 evaluation board, disciplined by a 10 MHz reference from an SRS CG635

Master clock: 40 MHz at output of IC50

RMS jitter in [1 Hz, 20 MHz]:
- External clock: 0.8 ps
- Master clock: 2.2 ps
Corresponding RMS jitter numbers on slide 25.
Standard deviation jitter as calculated from phase noise spectra on slide 24.

<table>
<thead>
<tr>
<th>Slot #</th>
<th>Jitter (ps) [1 Hz, 100 Hz]</th>
<th>Jitter (ps) [100 Hz, 1 MHz]</th>
<th>Jitter (ps) [1 MHz, 20 MHz]</th>
<th>Jitter (ps) [1 Hz, 20 MHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.9</td>
<td>0.9</td>
<td>3.9</td>
<td>4.4</td>
</tr>
<tr>
<td>4</td>
<td>1.9</td>
<td>0.7</td>
<td>2.6</td>
<td>3.3</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>0.8</td>
<td>3.3</td>
<td>3.9</td>
</tr>
<tr>
<td>6</td>
<td>2.0</td>
<td>1.0</td>
<td>4.3</td>
<td>4.9</td>
</tr>
<tr>
<td>7</td>
<td>2.1</td>
<td>0.9</td>
<td>4.1</td>
<td>4.7</td>
</tr>
<tr>
<td>8</td>
<td>1.8</td>
<td>0.9</td>
<td>3.7</td>
<td>4.2</td>
</tr>
<tr>
<td>9</td>
<td>1.8</td>
<td>0.7</td>
<td>2.3</td>
<td>3.1</td>
</tr>
<tr>
<td>10</td>
<td>1.8</td>
<td>0.6</td>
<td>1.8</td>
<td>2.6</td>
</tr>
<tr>
<td>11</td>
<td>1.8</td>
<td>0.7</td>
<td>1.9</td>
<td>2.7</td>
</tr>
<tr>
<td>12</td>
<td>1.9</td>
<td>0.7</td>
<td>2.5</td>
<td>3.2</td>
</tr>
<tr>
<td>13</td>
<td>1.9</td>
<td>0.7</td>
<td>2.6</td>
<td>3.3</td>
</tr>
<tr>
<td>14</td>
<td>2.0</td>
<td>0.7</td>
<td>2.7</td>
<td>3.4</td>
</tr>
</tbody>
</table>
Corresponding RMS jitter numbers on slide 27.
Standard deviation jitter as calculated from phase noise spectra on slide 26.

<table>
<thead>
<tr>
<th>Slot #</th>
<th>[1 Hz, 100 Hz]</th>
<th>[100 Hz, 1 MHz]</th>
<th>[1 MHz, 20 MHz]</th>
<th>[1 Hz, 20 MHz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.6</td>
<td>2.6</td>
<td>0.2</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>2.7</td>
<td>0.2</td>
<td>3.6</td>
</tr>
<tr>
<td>5</td>
<td>2.4</td>
<td>2.7</td>
<td>0.2</td>
<td>3.6</td>
</tr>
<tr>
<td>6</td>
<td>2.3</td>
<td>2.7</td>
<td>0.2</td>
<td>3.6</td>
</tr>
<tr>
<td>7</td>
<td>2.2</td>
<td>2.7</td>
<td>0.2</td>
<td>3.4</td>
</tr>
<tr>
<td>8</td>
<td>2.3</td>
<td>2.7</td>
<td>0.2</td>
<td>3.5</td>
</tr>
<tr>
<td>9</td>
<td>3.3</td>
<td>0.5</td>
<td>0.1</td>
<td>3.4</td>
</tr>
<tr>
<td>10</td>
<td>2.9</td>
<td>0.5</td>
<td>0.1</td>
<td>2.9</td>
</tr>
<tr>
<td>11</td>
<td>2.7</td>
<td>0.4</td>
<td>0.1</td>
<td>2.7</td>
</tr>
<tr>
<td>12</td>
<td>3.1</td>
<td>0.5</td>
<td>0.1</td>
<td>3.2</td>
</tr>
<tr>
<td>13</td>
<td>3.1</td>
<td>0.5</td>
<td>0.1</td>
<td>3.2</td>
</tr>
<tr>
<td>14</td>
<td>2.8</td>
<td>0.5</td>
<td>0.1</td>
<td>2.8</td>
</tr>
</tbody>
</table>
Crosstalk between the 40 MHz and the 320 MHz backplane clocks in the same slot
Clock quality definitions

Different clock quality metrics:

- **Phase noise spectrum**
  Shows noise power as a function of frequency, relative to the carrier. Requires a dedicated instrument. Spectrum integral gives RMS jitter.

- **Time Interval Error (TIE)**
  Oscilloscope measurement of jitter w.r.t. ideal clock reconstructed from the signal. Not sensitive to low frequency noise. Implicitly integrates up to the scope bandwidth.

- **Edge-to-edge**
  Oscilloscope measurement comparing corresponding edge pairs between two channels. Sensitive to lower noise frequencies than TIE.
Phase noise vs. RMS jitter

\[ T_{jRMS} = \frac{10^{A(f)/10}}{\sqrt{2\pi}f_0} \]

https://www.silabs.com/documents/public/application-notes/AN256.pdf,