

Development and Test of a 48-optical Ports High Precision Clock Distributor Board

D. Calvet, E. Molina,

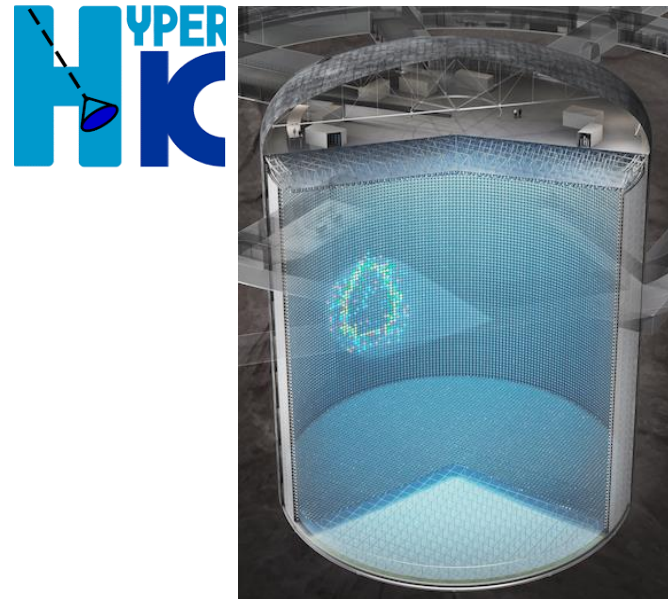
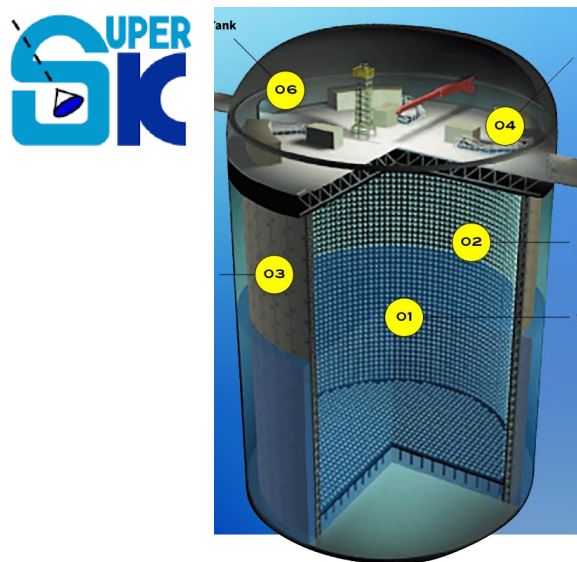
Irfu, CEA Paris-Saclay, France

denis.calvet@cea.fr

emmanuel.molina-gonzalez@cea.fr

High Precision Clock Distribution, Hyper-K case

- Some physics experiment require larger and larger detection volumes and increasing channel counts
- Synchronizing electronics spread across large space with high precision (e.g. few 10's ps) is required
- Example: next generation neutrino observatory Hyper-Kamiokande



- Constructed 1996; in operation
- 50,000 ton water target
- 39 m diameter x 42 m height
- ~12.000 phototubes

- Data taking from 2027
- 260,000 ton water target
- 68 m diameter x 71 m height
- ~20.000 phototubes

October 2023: completion of the excavation of the dome of the cavern (69 m diameter)

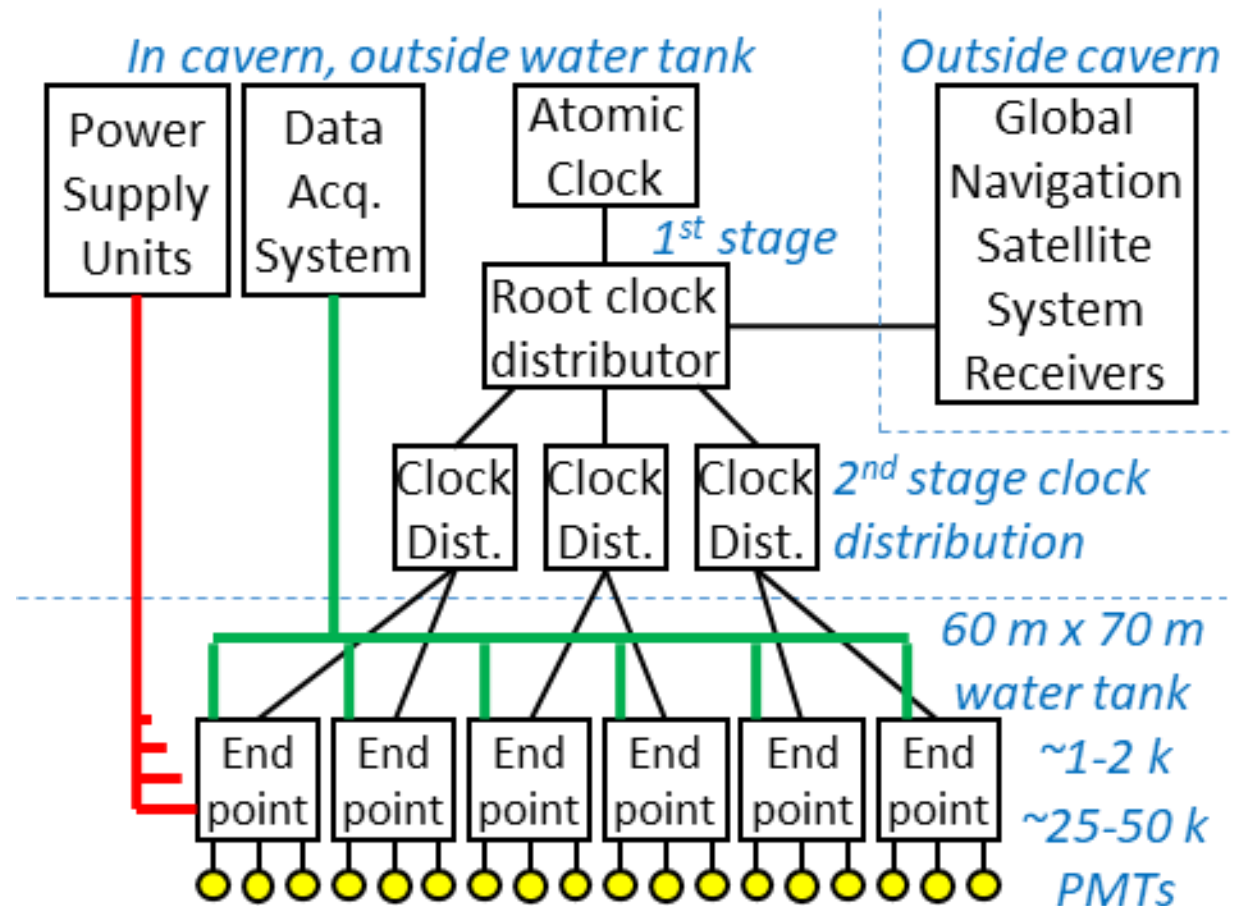
Requirements for Hyper-K, System Architecture

Main requirements

- Distinct path for synchronization and DAQ
- Synchro. path: 125 MHz clock + 125 Mbps data min.
- Stable phase offset after system reconfiguration
- ≤ 100 ps rms clock jitter at all end-points
- Up to 1000-2000 end-points

Proposed architecture

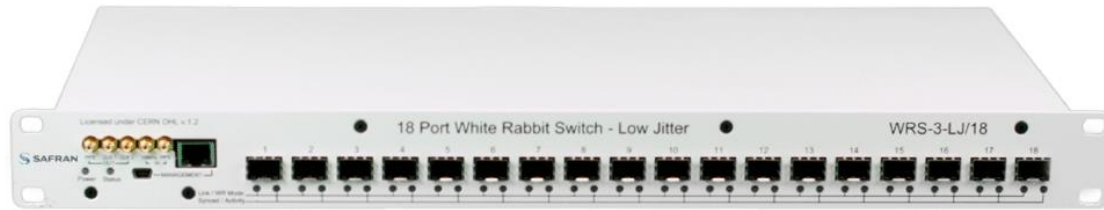
- Two stage clock distribution tree
- 48-port distributors to scale up to 2308 end-points
- Based on modern FPGAs and optical technology



Beyond Common Clock Distribution Solutions

Present landscape

- Many designs rely on FPGA high-speed SerDes for clock and synchronous data distribution
- Clock and Data Recovery circuit (CDR) within SerDes block produces a copy of the transmitter clock
- WhiteRabbit, started at CERN, combines the idea with Ethernet. It's standard, commercially available.

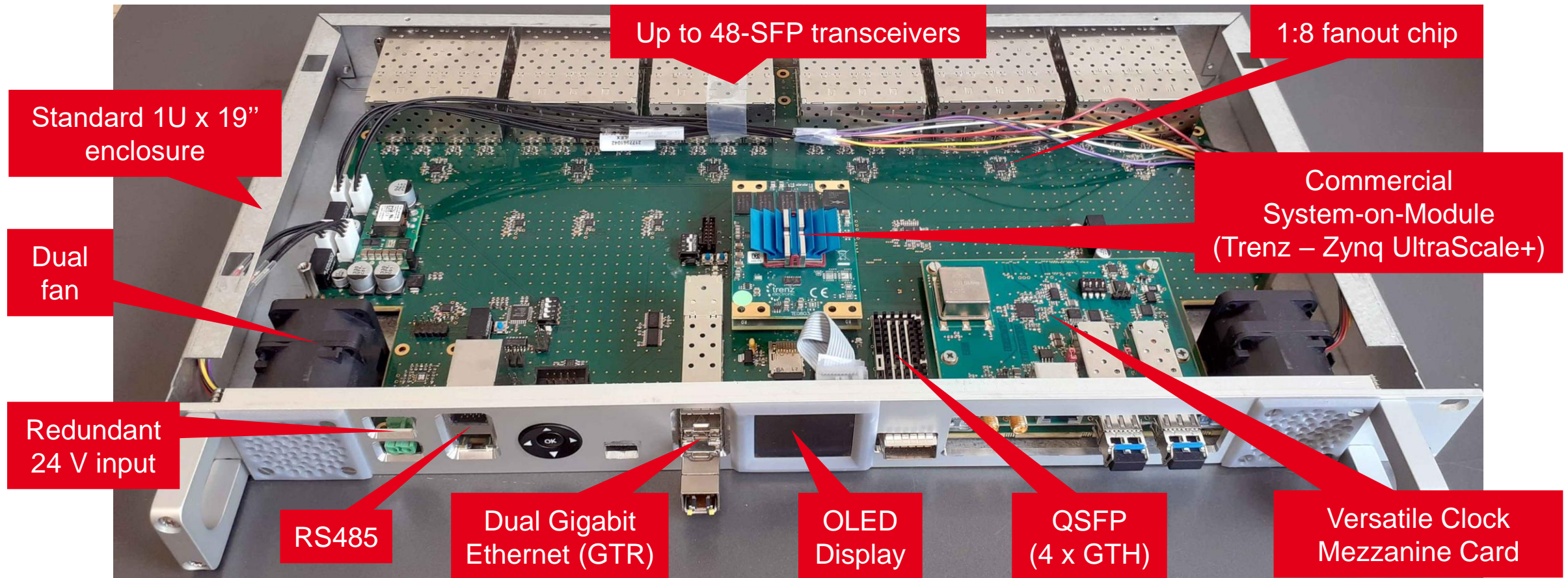


<https://www.white-rabbit.tech/>

Why search for alternatives?

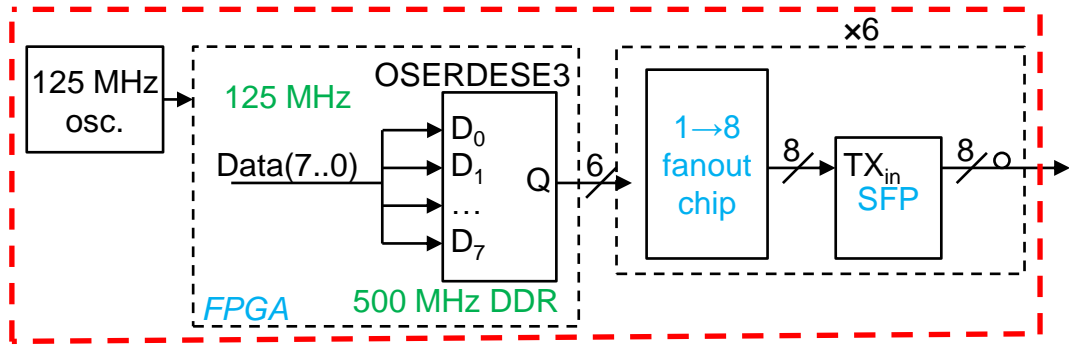
- Based on Super-K experience, Hyper-K, will have distinct clock and DAQ paths. Using a White Rabbit network solely for synchronization + a regular Ethernet network for DAQ does not seem judicious
- Available WR switch (v.3) is an aging product (Virtex 6); v.4 in development following its own planning
→ *R&D on synchronization for Hyper K moved towards custom solutions developed internally*

48-Port Clock Distributor Board

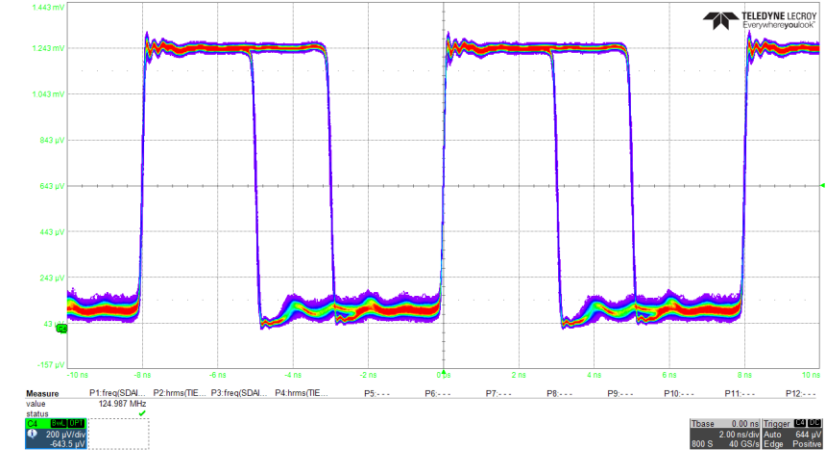
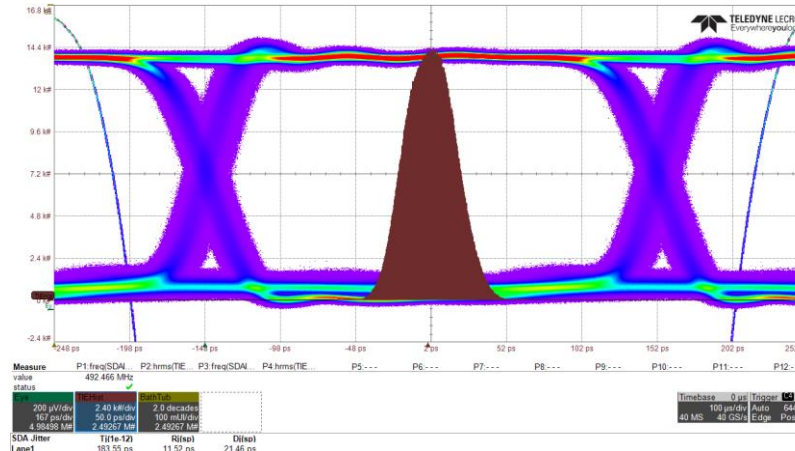
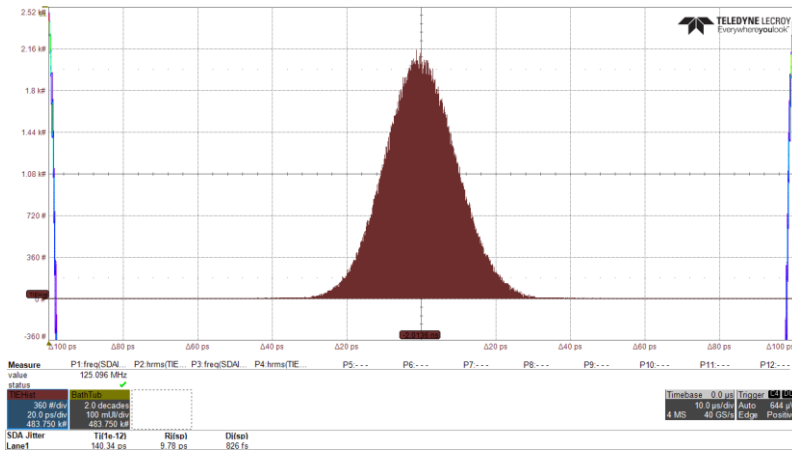


- Small available number of high speed SerDes (4 GTH) assigned to high throughput interface while more numerous ordinary I/O's drive up to 48 SFPs: 48 TX in 6 groups via 1:8 fanout chips + 48 individual RX
- Proprietary serial protocol implemented on UltraScale+ ISERDESE3/OSERDESE3 primitives

Transmit Operation – Jitter Performance



- Depending on serializer inputs (8-bit x 125 MHz) various operating modes: clock up to 500 MHz, serial data up to 1 Gbps and sub-multiples rates
- 6 TX groups of 8 SFPs can be set to a different mode independently

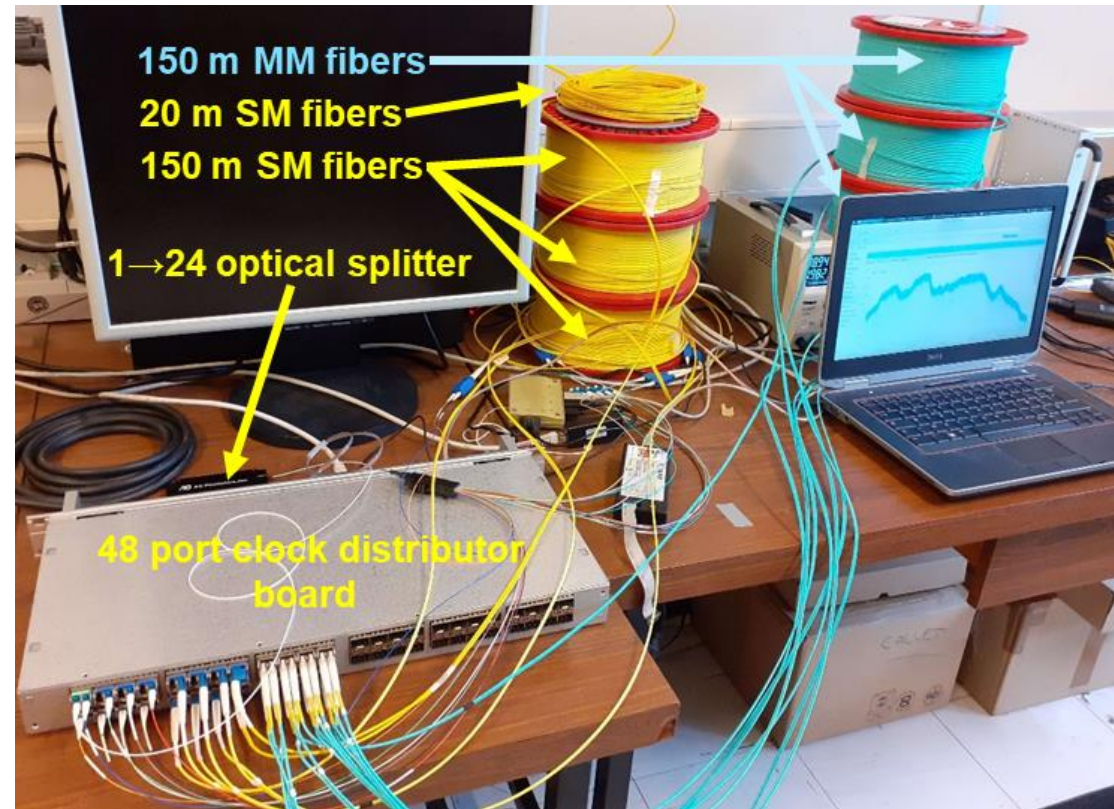
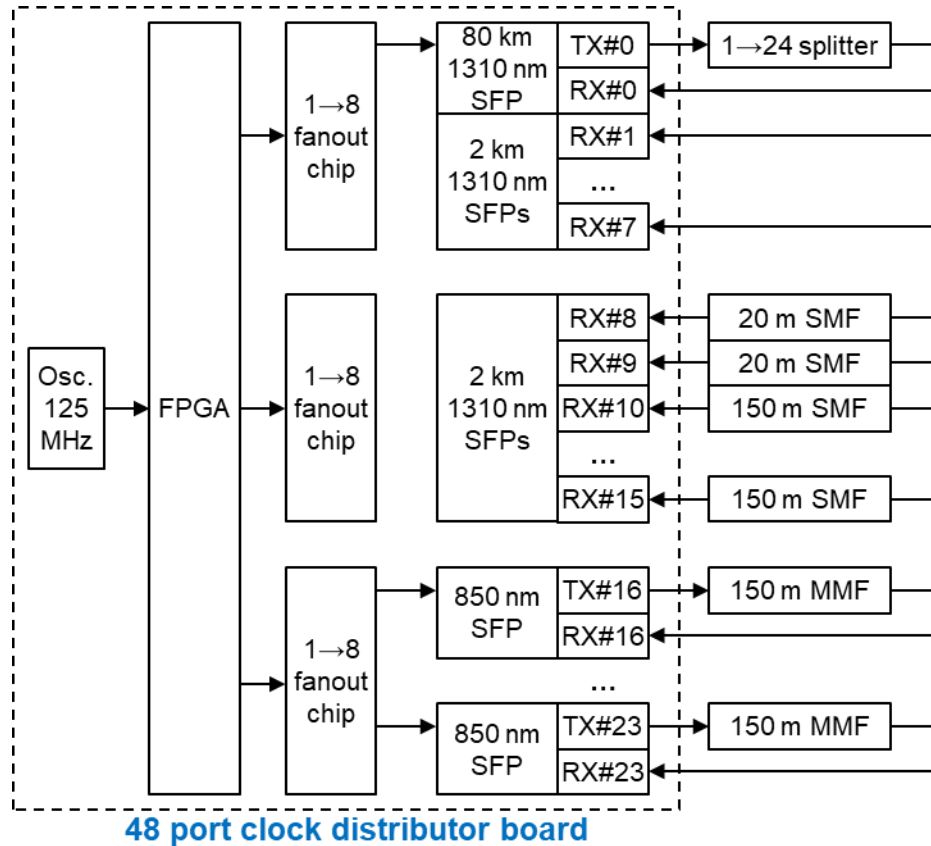


- TX mode: 125 MHz clock fanout
- SFP port #8: $R_j < 10$ ps $D_j < 1$ ps

- TX mode: 1 Gbps PRBS data
- SFP port #8: $R_j < 15$ ps $D_j = 21.4$ ps

- TX mode: 125 MHz clock duty cycle modulated by 125 Mbps PRBS data
- SFP port #8: $R_j < 10$ ps $D_j = 1$ ps

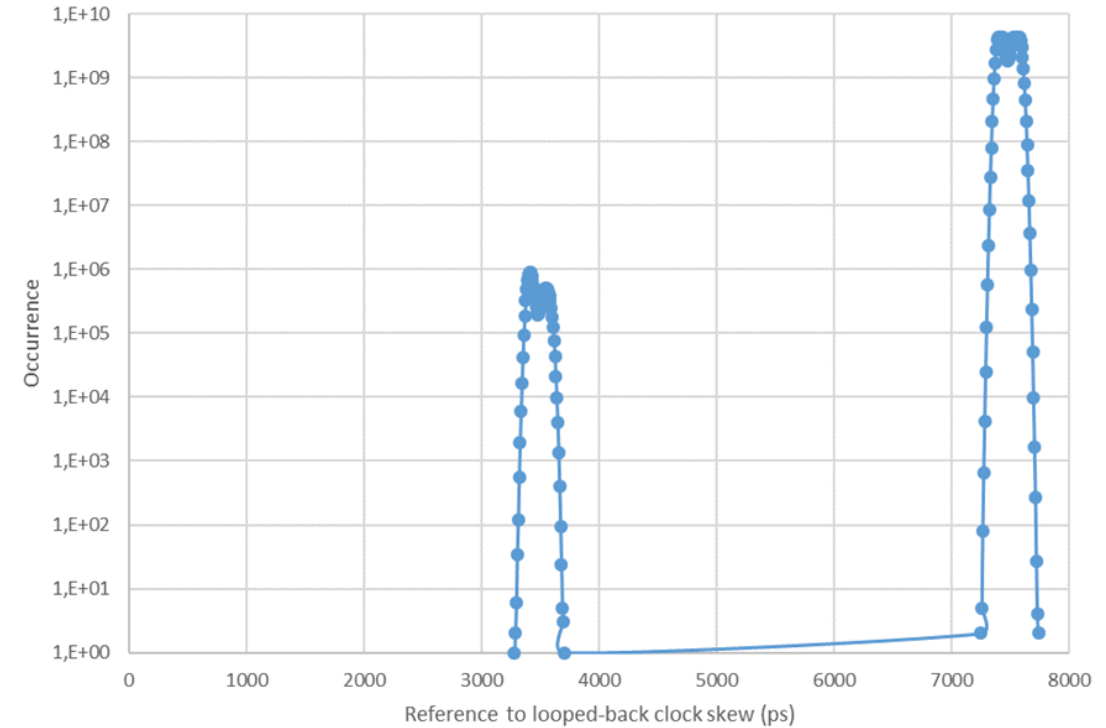
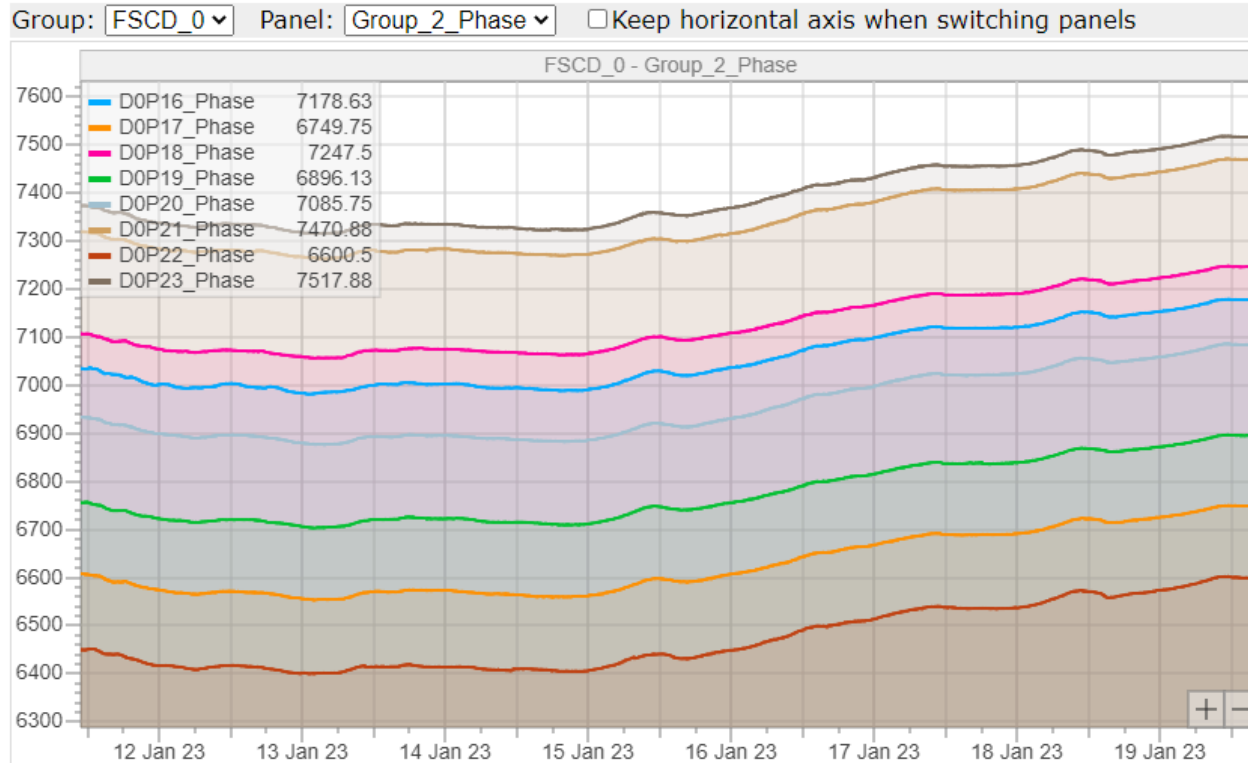
Test Setup for Longer Term Measurements



- Logic based on "Digital Dual Mixer Time Difference (DDMTD)" (used in WhiteRabbit) is implemented in FPGA to track phase shifts between local clock and echo clocks returned by TX to RX loop-back fibers

P. Moreira, P. Alvarez, J. Serrano, I. Darwezeh and T. Wlostowski, "Digital dual mixer time difference for sub-nanosecond time synchronization in Ethernet", in *Proc. IEEE International Frequency Control Symposium*, Newport Beach, CA, USA, 2010, pp. 449-453

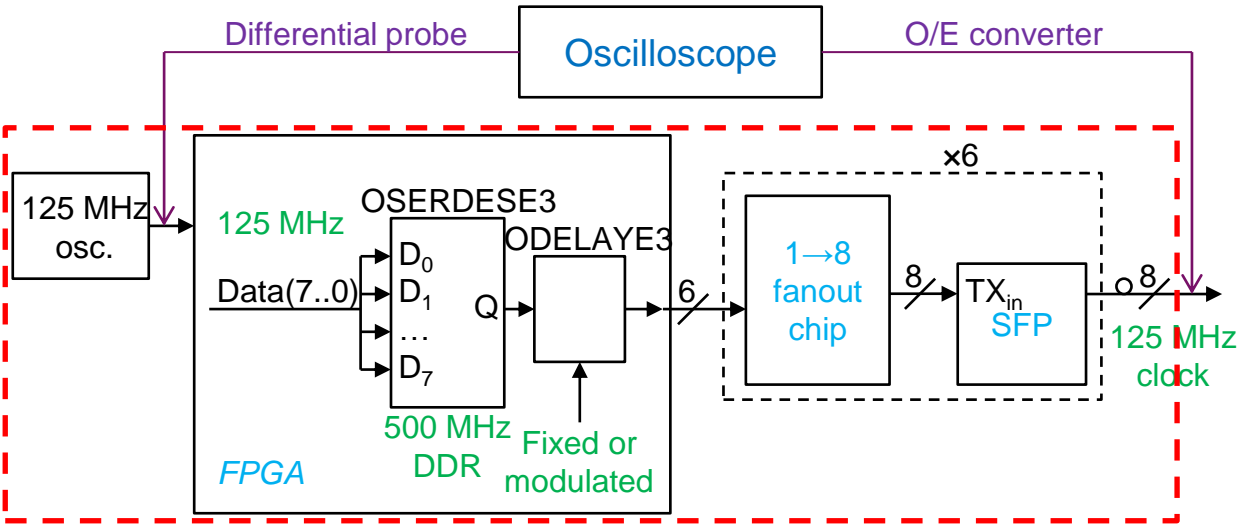
Distributed Clock Phase Variations Over 1 Week



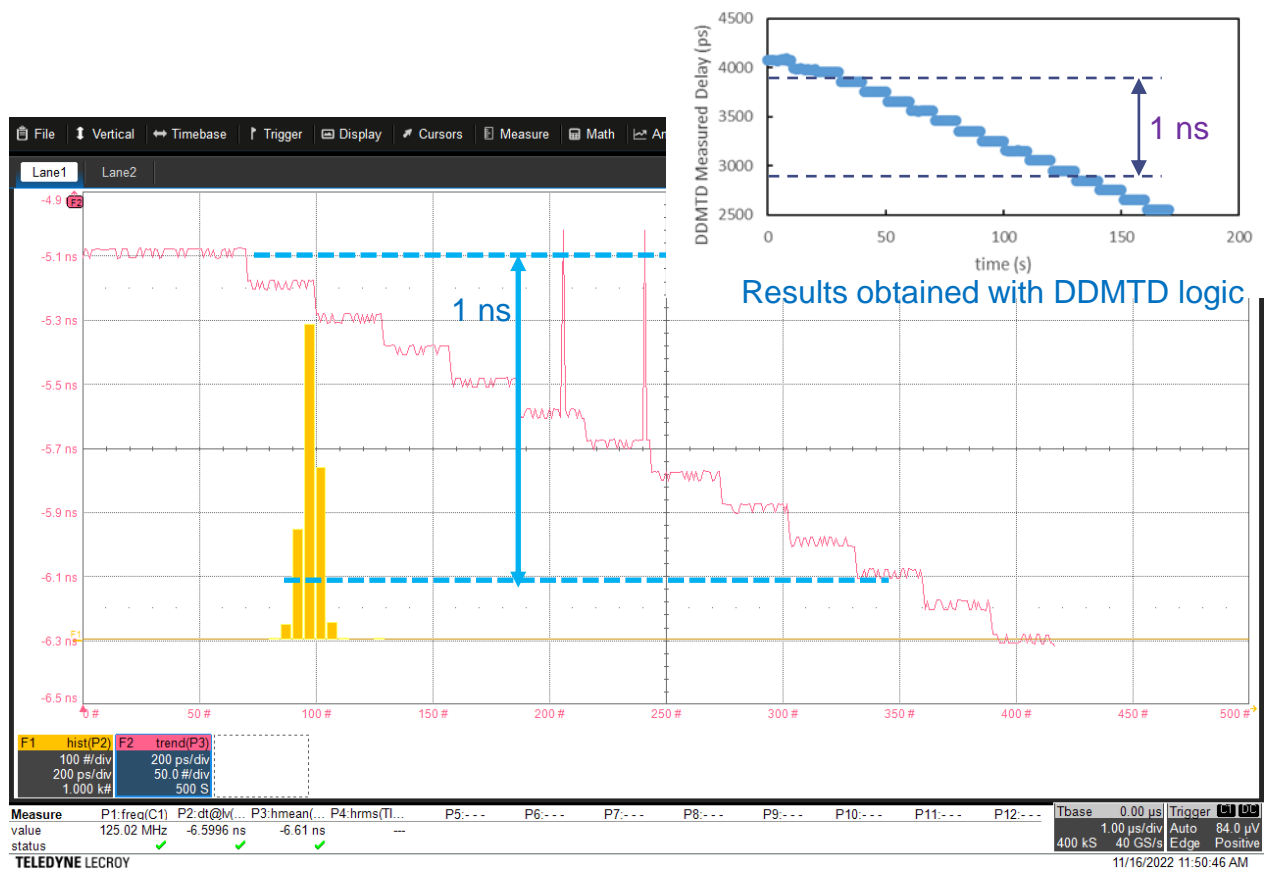
- Group of 8 ports in 125 MHz clock distribution mode
- TX to RX loopback on 8 ports by 150 m MMF fibers
- 200 ps p.p. drift measured; 6 °C p.p. temperature var.

- Histogram of reference clock to echo clock phase differences on one port (125 kHz measurement rate)
- Spurious peak, 4 orders magnitude smaller than main, probably caused by imperfections in DDMTD logic, but no other deviant points

Adding a Programmable Delay on TX Path



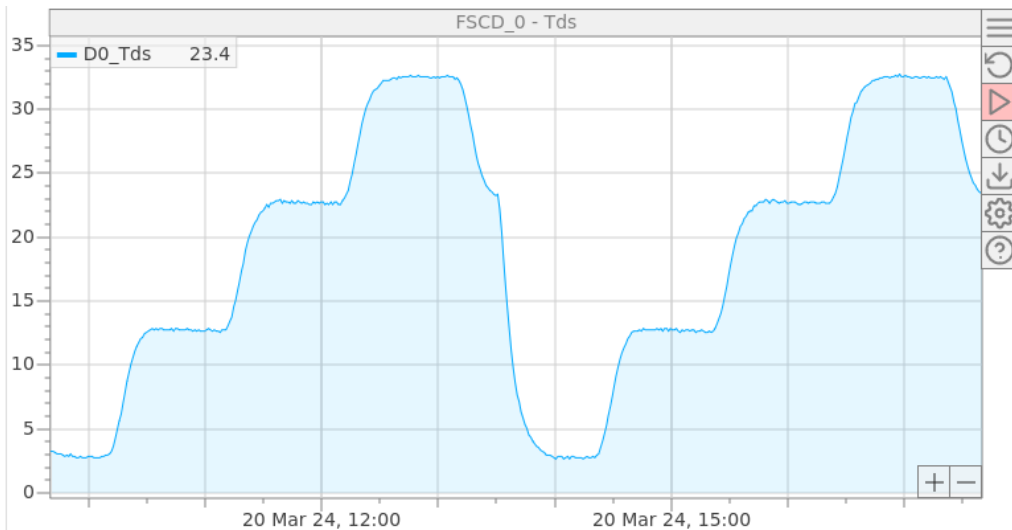
- UltraScale+ OSERDESE3 can be followed by a 512-tap calibrated delay line, ODELAYE3, cascadable up to 3. Delay tap resolution: 2.1-12 ps



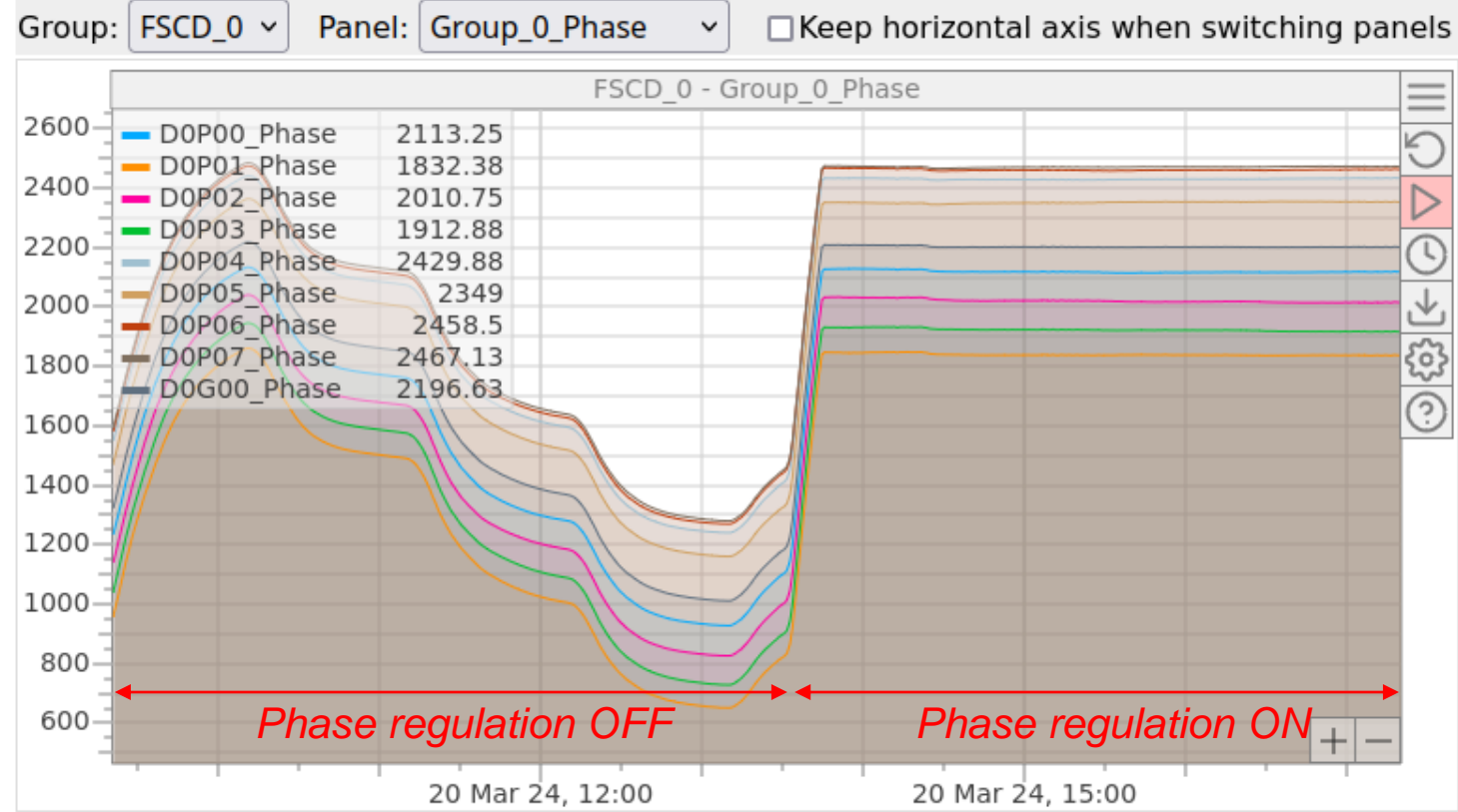
- ODELAYE3 set to perform 12 delay steps of 100 ps
- Expected delay jumps verified with an oscilloscope
- Phase steps also correctly measured by DDMTD logic

Clock Fanout Phase Regulation with ODELAYE3

- Demonstrator set for 125 MHz clock fanout
- 1 TX port drives 150 m SM fiber + 1:8 optical splitter looped-back to 8 RX ports
- All hardware placed in a climatic chamber set at 0°C to 30°C in 4 steps of 10°C



Measured clock distributor board temperature



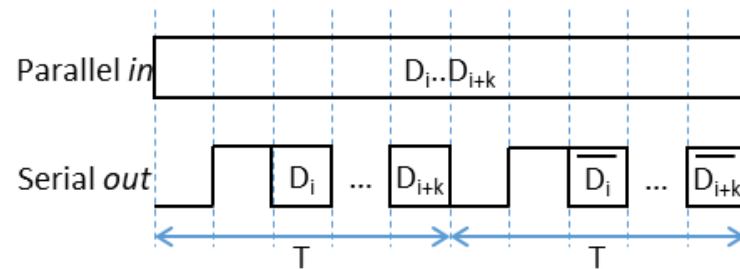
Phase 125 MHz reference to 8 clocks looped backed (DDMTD method)

- Designed a corrector on embedded processor tuning ODELAYE3 to stabilize distributed clock phase drift
- For 30°C variations: delay variations ~1.2 ns without regulation and ~6 ps with regulation

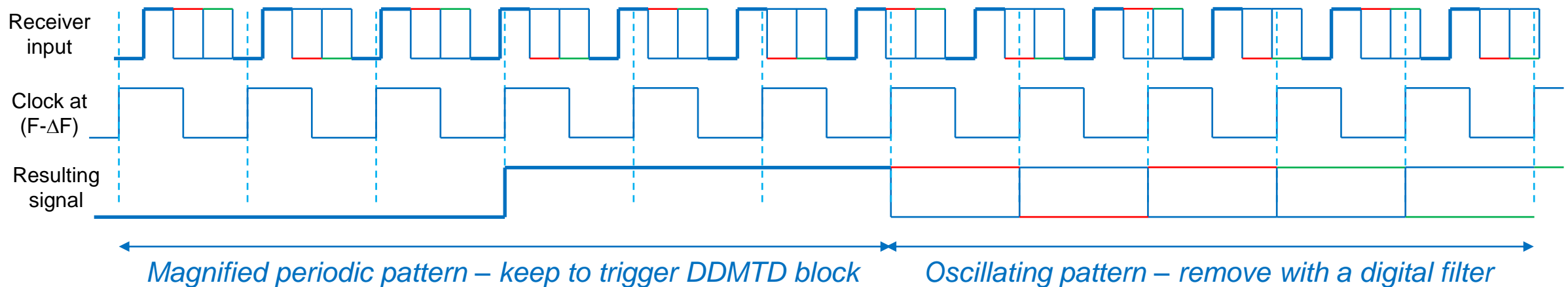
Simultaneous Serial Data Reception and Clock Round-trip Monitoring on FPGA Ordinary Inputs

- Previous tests use loop-back **clock** on RX port to measure phase difference with the reference. But recovered clock not provided by serial data receiver made from ordinary FPGA input pins.

How an ordinary FPGA input pair can simultaneously receive serial data and monitor echo clock phase?



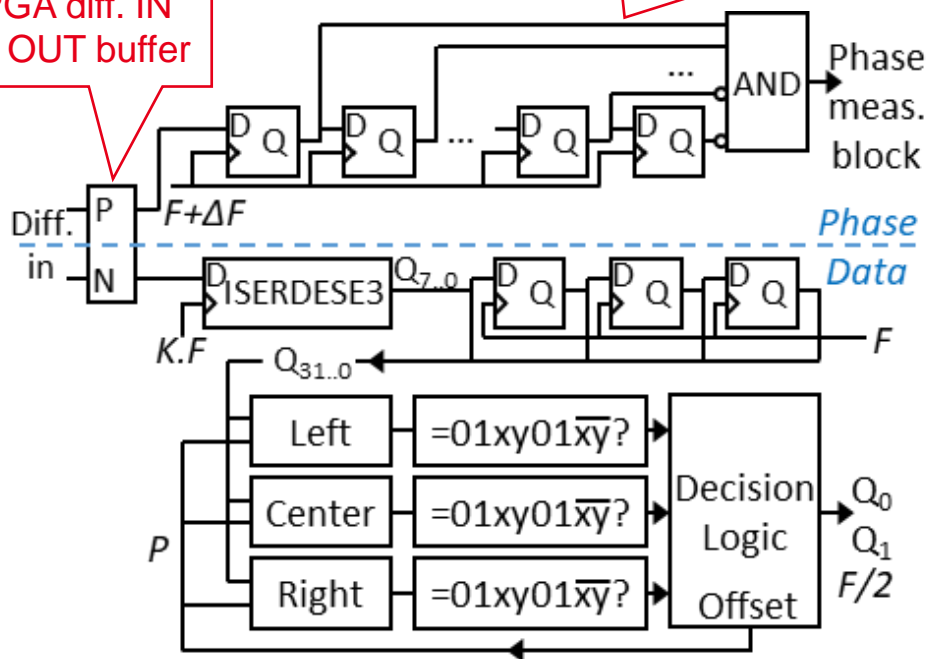
- Proposed solution: use specific encoding to introduce a fixed periodic pattern, e.g. "01" in the serial stream
- Send user data bits followed by their complement at the next cycle



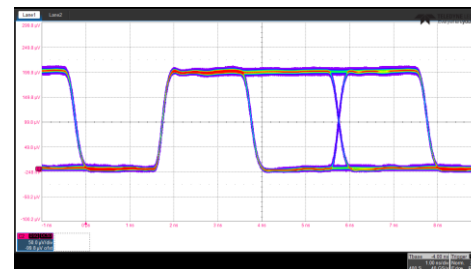
Dual Function Receiver on a Differential Input

Digital filter removes oscillating part of serial stream sampled at $(F-\Delta F)$

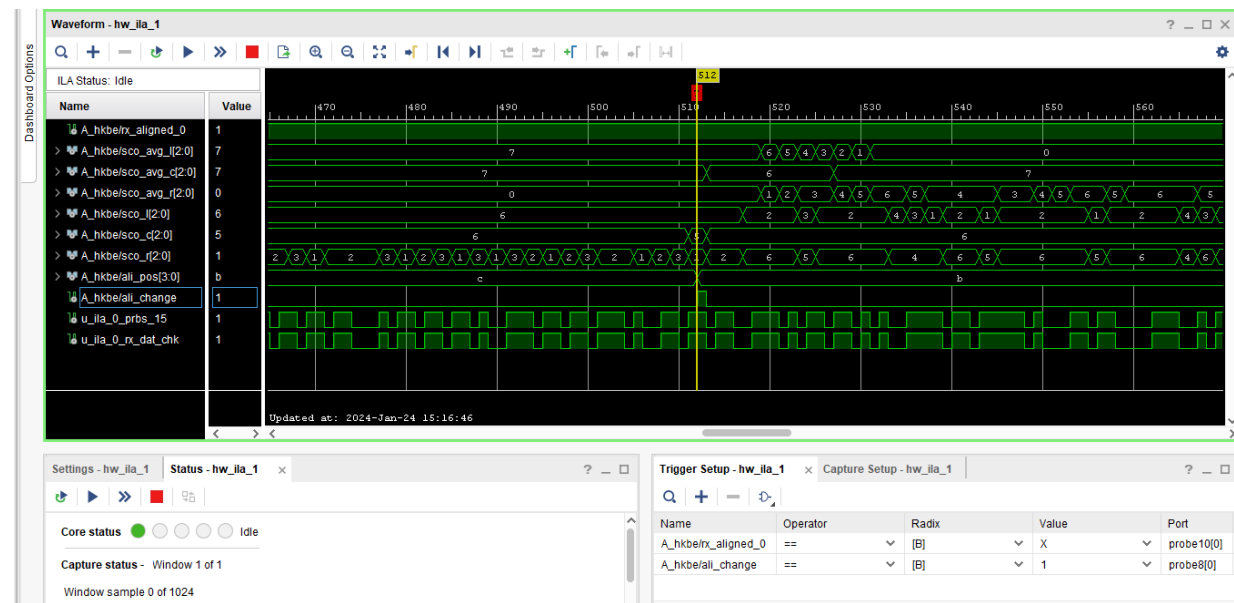
FPGA diff. IN
diff. OUT buffer



Deserializer oversamples received serial input at x2 rate. For 3 adjacent positions, logic calculates how close are received groups of bits from the plausible pattern. Decision logic selects the optimal set of samples and adjusts the sampling offset if necessary.

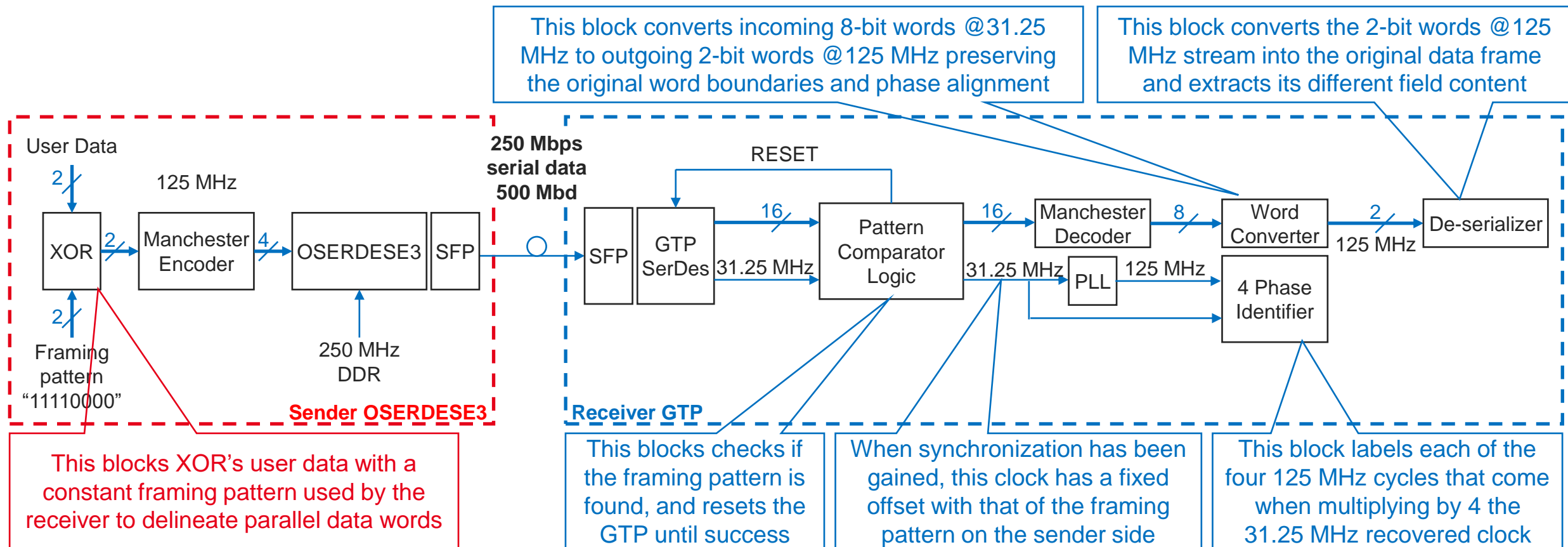


- Measured 500 MBd serial stream embedding 125 MHz periodic pattern and 125 Mbps user data (coding efficiency: 25%)



- PRBS pattern sent is verified at receiver using an Integrated Logic Analyzer (ILA) in FPGA fabric

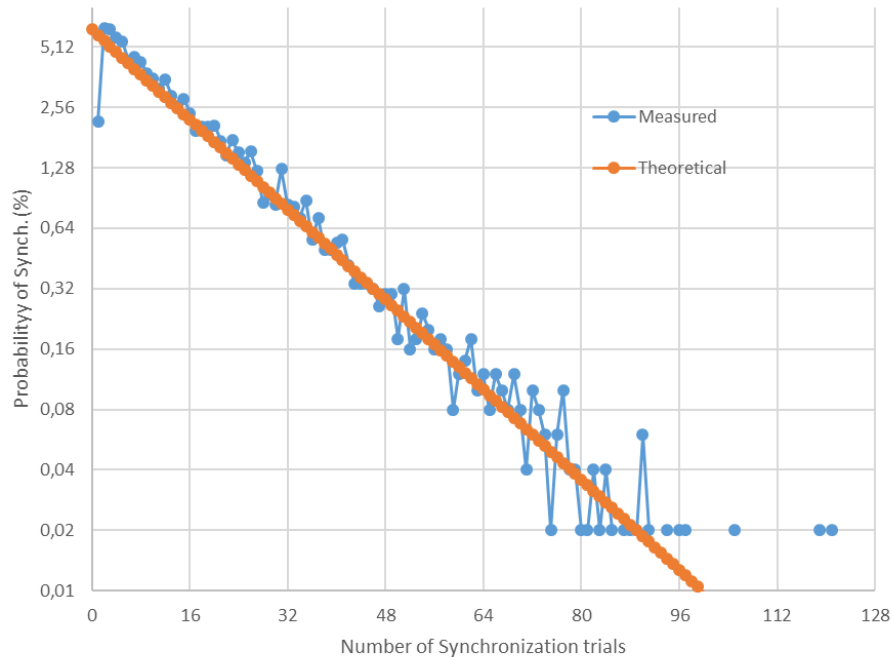
GTP Transceiver Reception from OSERDESE3



- Typically, FPGA ordinary I/O's and high speed SerDes overlap in the 500-1500 Mbps region
- Example shown uses 250 Mbps user data with Manchester encoding leading to 500 Mbd on media
- Instead of comma insertion, user data are XOR'ed with a framing pattern to align received data

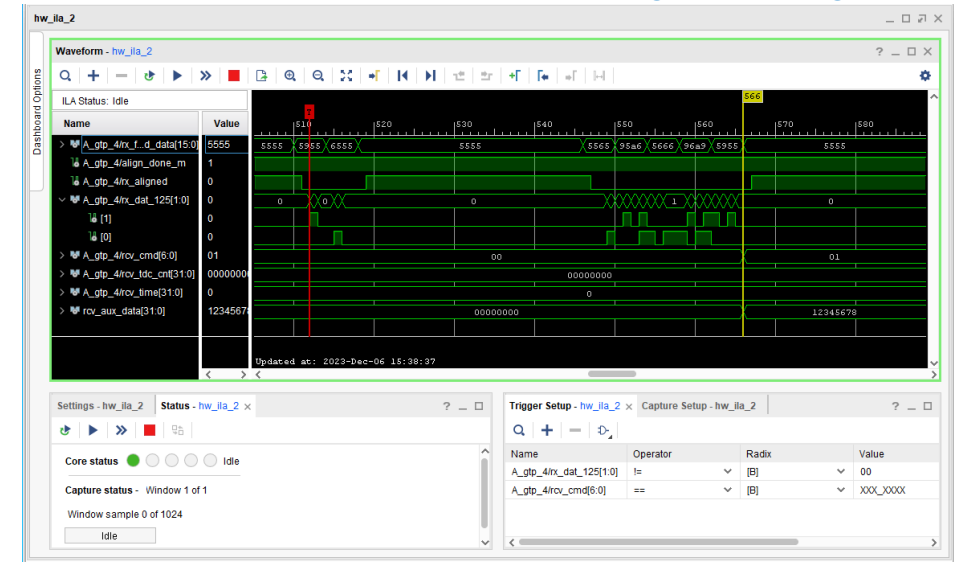
GTP Transceiver Reception from OSERDESE3

Number of trials to reach proper alignment



Artix 7 GTP receiver used for interoperability tests with OSERDESE3 sender

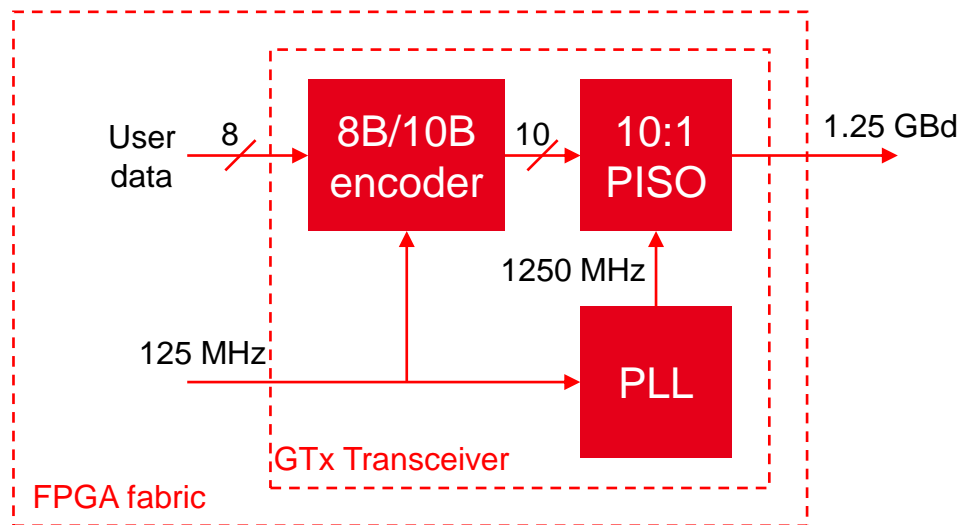
User data reception checked with Integrated Logic Analyzer



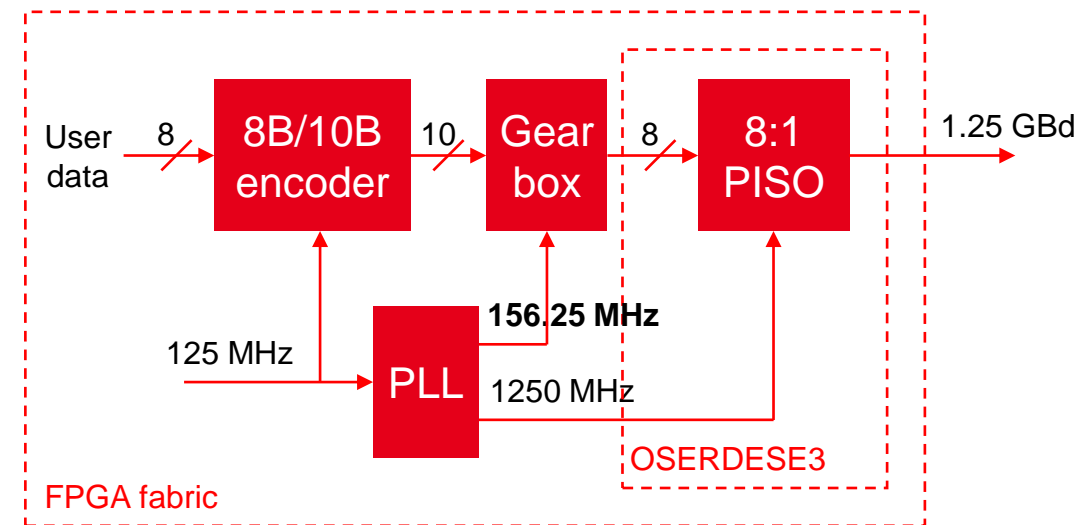
- At 500 MBd, a GTP in 16-bit parallel interface mode produces a 31.25 MHz recovered clock
- Recovered clock can have 16 possible phase offsets, 2 ns apart, compared to the 500 MHz serial clock
- For deterministic latency, select the only recovered clock that makes the framing pattern appear at output
- At link startup, GTP is reset until the recovered clock has the desired phase offset (probability $p = 1/16$)
- Probability of correct synchronization after exactly N trials: $p * (1 - p)^{(N-1)}$

8B/10B Serial Transmission with OSERDESE3

- 8B/10B is a popular method for serial communication. Supported in FPGA high-speed transceivers
- FPGA I/O SerDes in most recent devices, e.g. Xilinx UltraScale+, only support serialization factor 2:1, 4:1 and 8:1 but no longer 5:1 or 10:1 (Xilinx 7 family). User has to build the equivalent function in fabric



FPGA high speed transceiver scheme

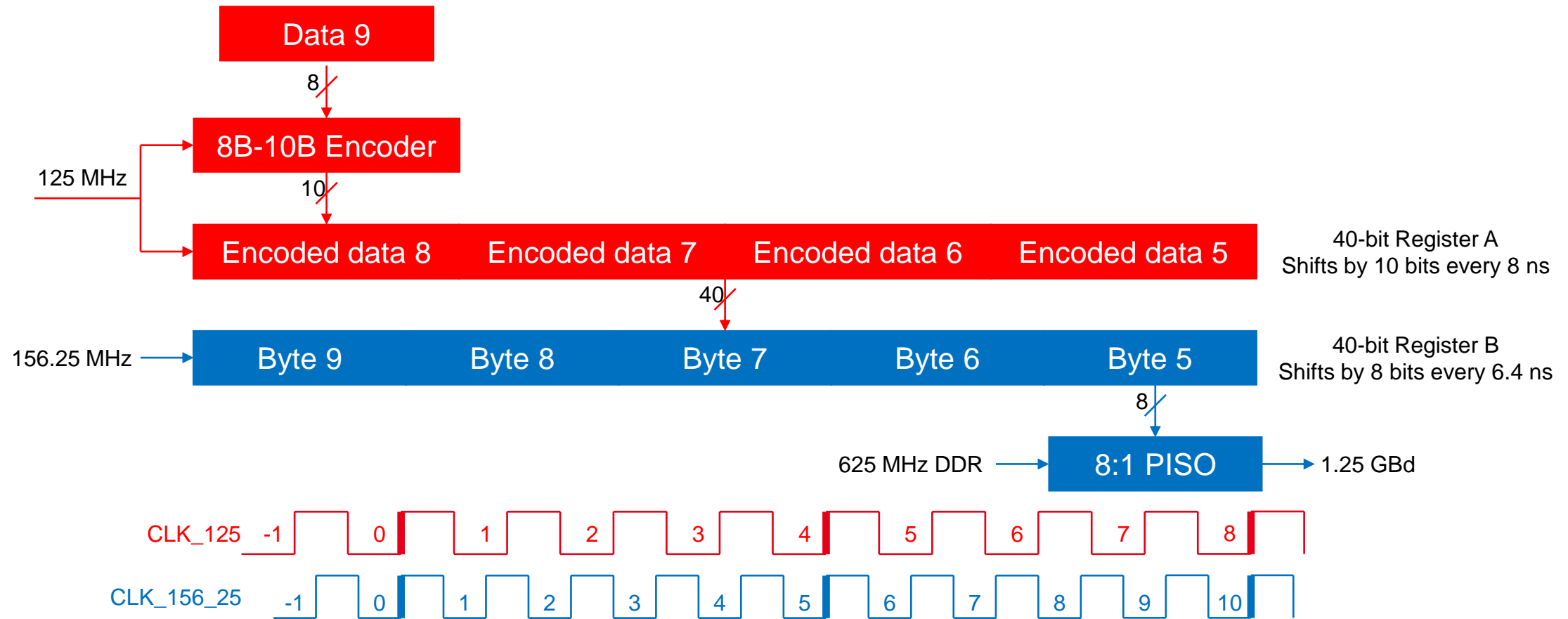


FPGA ordinary I/O scheme

- For precise clock distribution applications, logic and clock domain crossings must have fixed latency

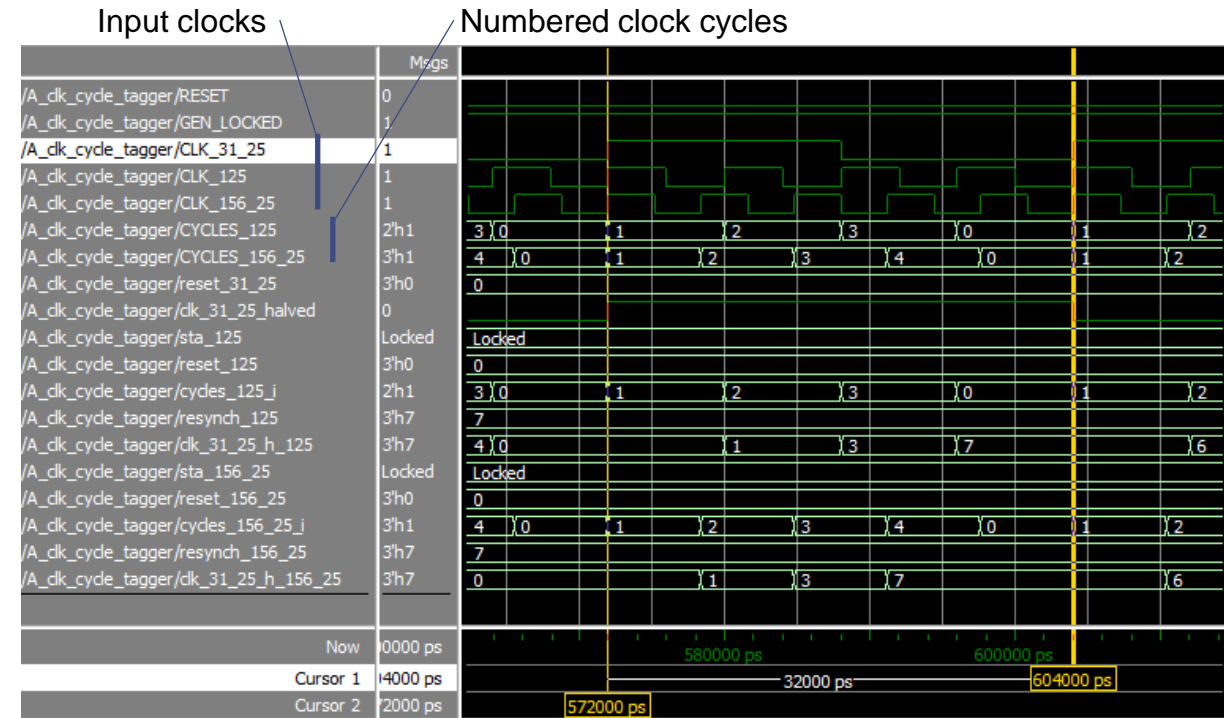
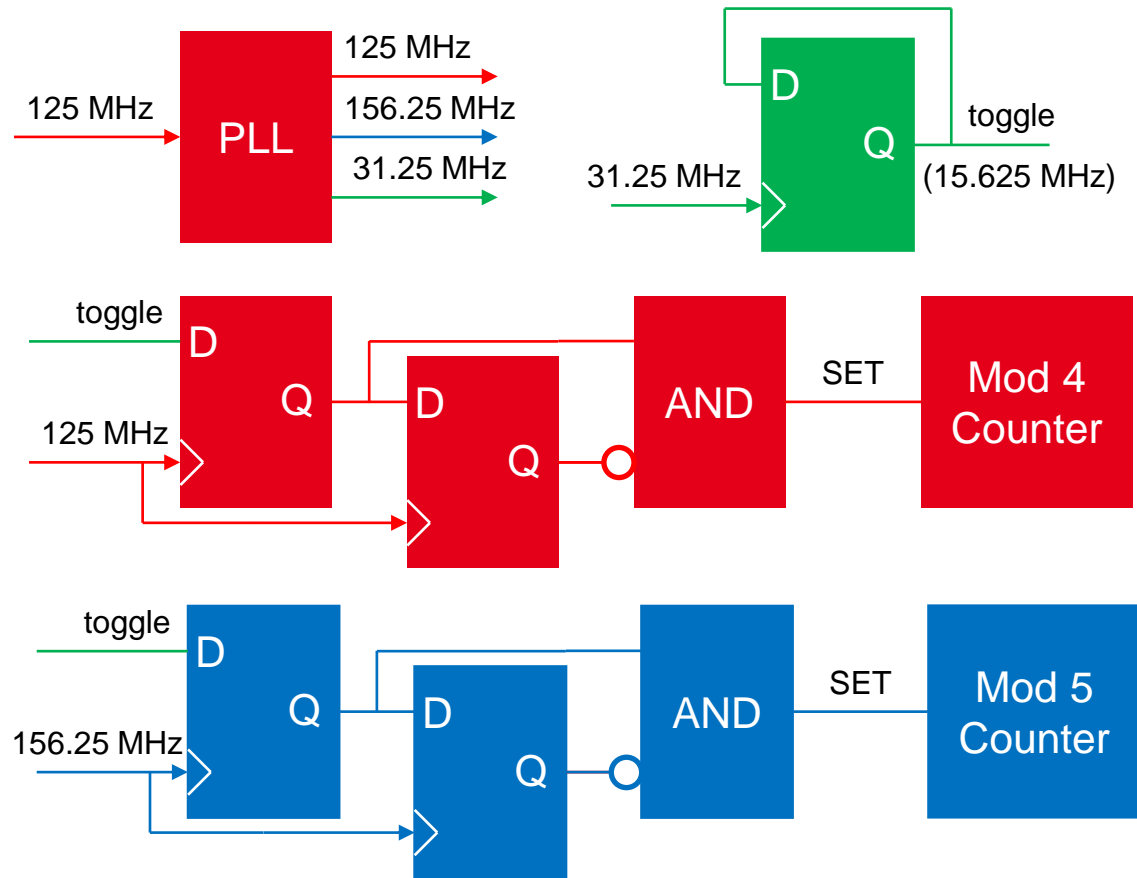
How can this be implemented?

Deterministic Latency 10B-8B Gear Box



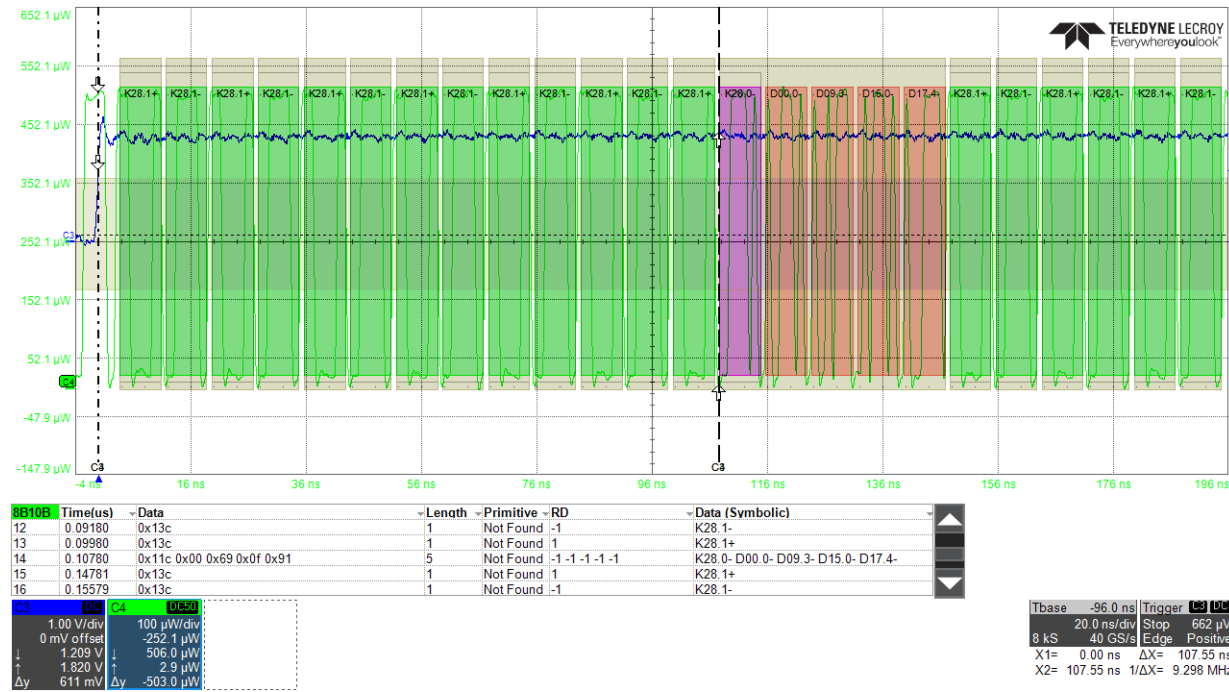
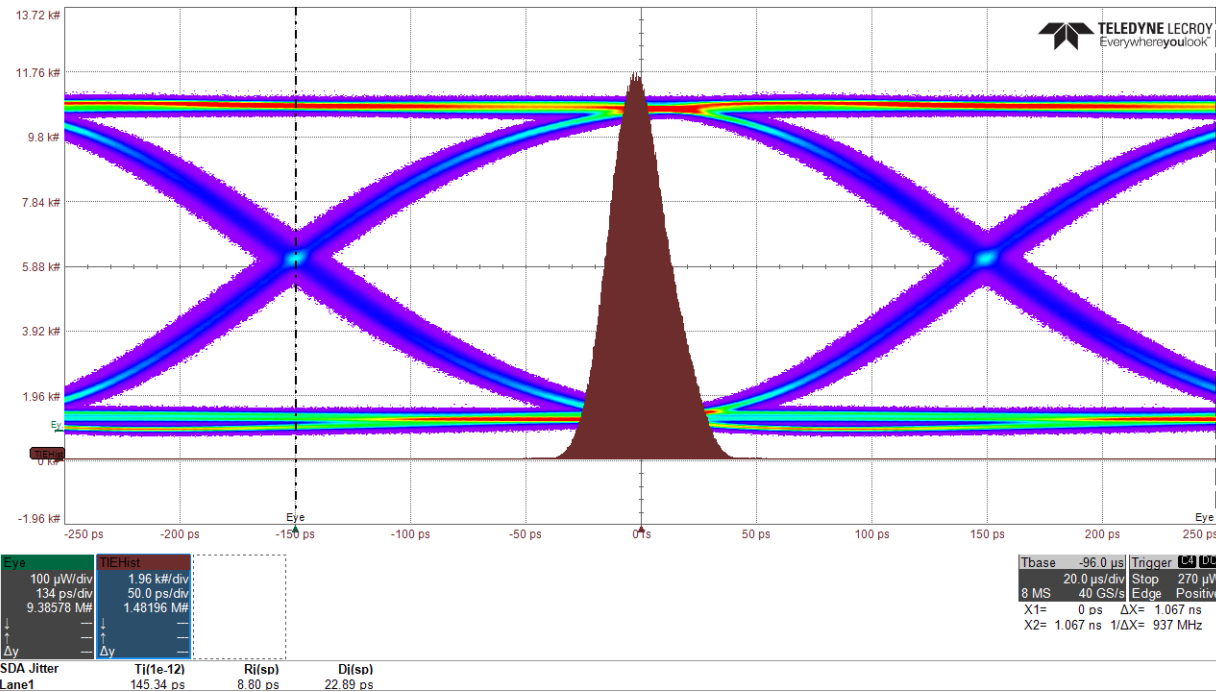
- Clocks are edgewise aligned every 4 periods of clock A (125 MHz) and 5 periods of clock B (156.25 MHz)
- 40 bits parallel data transferred from Register A to B when both clock edges are aligned (31.25 MHz)

Clock Cycle Numbering Logic for Gear Box



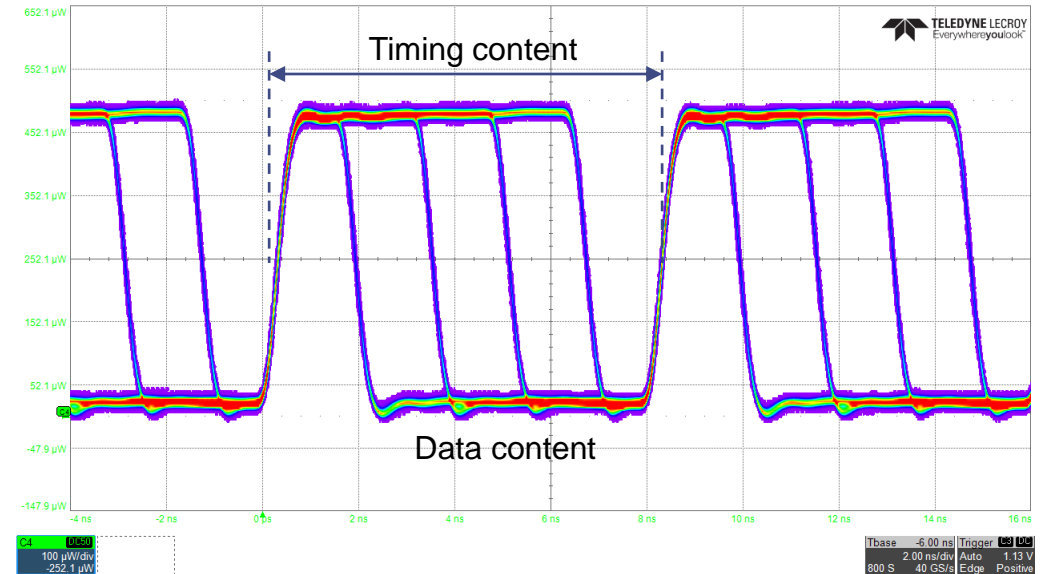
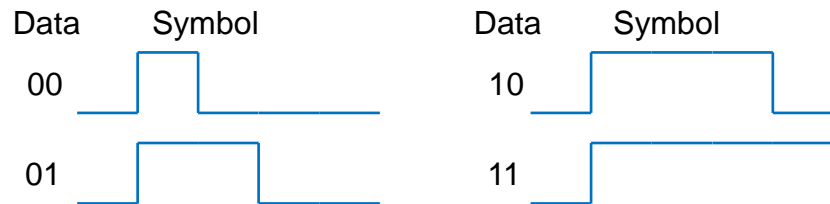
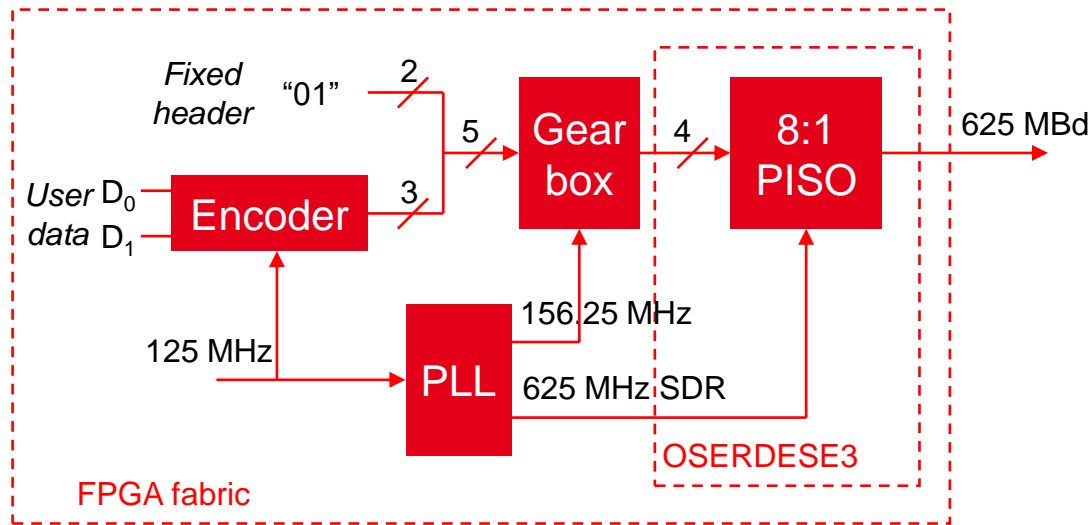
- PLL generates clock at frequency of greatest common divisor of two other clocks. Divide it by 2 in logic
- Detect rising edges of this toggling signal to send SET signals to the counters numbering clock cycles

Fixed Latency 8B/10B Encoder on OSERDESE3



- Serial line rate: 1 Gbps user data; 1.25 Gbd on fiber
- Optical probe with bandwidth limited to 2.5 GHz
- Deterministic latency upon board power cycling verified using a debug pin of the sender FPGA
- Demonstrates capability of UltraScale+ HP I/O for deterministic latency 8B-10B encoding @1.25 Gbd
- Pathway to implement synchronous Gigabit Ethernet on plain FPGA I/O's (only TX part shown here)
- TX stream: idle commas (K28.1); Messages: start comma (K28.0) + 32-bit counter (4 Data symbols)
- Correct data reception and decoding in oscilloscope

Sending 250 Mbps Data on top a 125 MHz Clock



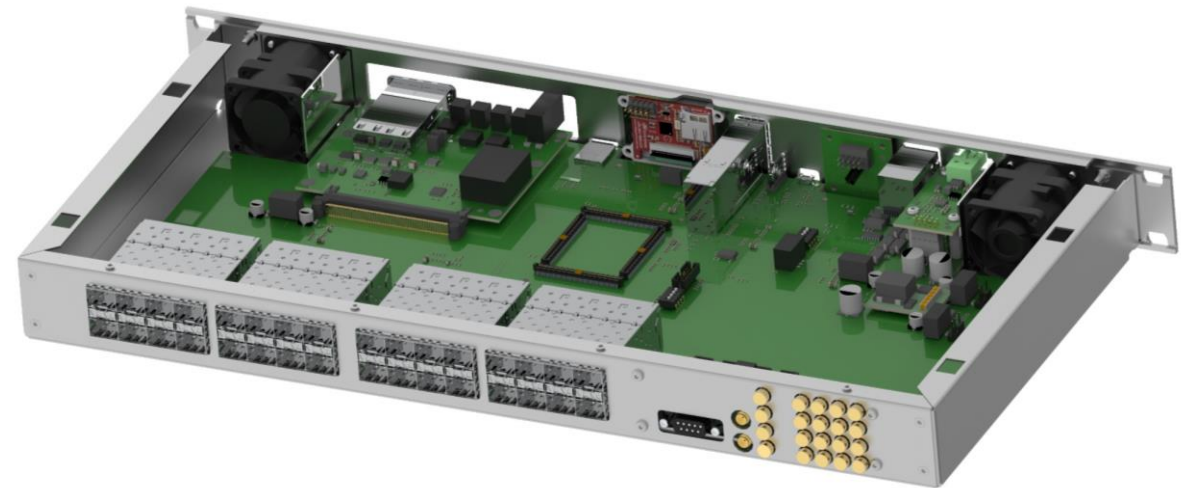
- 650 MBd stream is 125 MHz clock with modulated duty cycle: 20%, 40%, 60%, 80%
- 2 bits of user data per 8 ns period, i.e. 250 Mbps
- A scrambler shall be used to keep DC balance

- Advantage of clock duty cycle modulation for data transport: clock extraction by ordinary PLL (no CDR)

D. Calvet, "Clock-Centric Serial Links for the Synchronization of Distributed Readout Systems", IEEE Trans. Nucl. Sci., Vol. 67, N°8, pp. 1912-1919, Aug. 2020

Summary and Perspectives

- Successful construction of a 48-port clock distributor demonstrator
- Using FPGA I/O primitives for serial communication with deterministic latency are a viable option
- After review, the Hyper-K collaboration opted for a synchronization system composed of custom-made distributors arranged in a 2-stage tree
- The 2nd stage distributors (Lpnhe) and end-points (INFN) will use the traditional and more established approach based on high speed FPGA SerDes (1 Gbps net bandwidth + 8B-10B encoding)
- The 1st stage distributor (Irfu) will re-use some of the concepts and techniques developed in this R&D
 - Same overall architecture
 - Number of optical port reduced to 32
 - Serial communication using FPGA I/O's
 - Additional low speed I/O's at rear panel



Hyper-K First Stage Clock Distributor board (in design)