DE LA RECHERCHE À L'INDUSTRIE

www.cea.fr

STATUS OF R&D ON A CLOCK DISTRIBUTION SYSTEM FOR HYPER KAMIOKANDE

D. Calvet,
CEA Paris-Saclay

denis.calvet@cea.fr

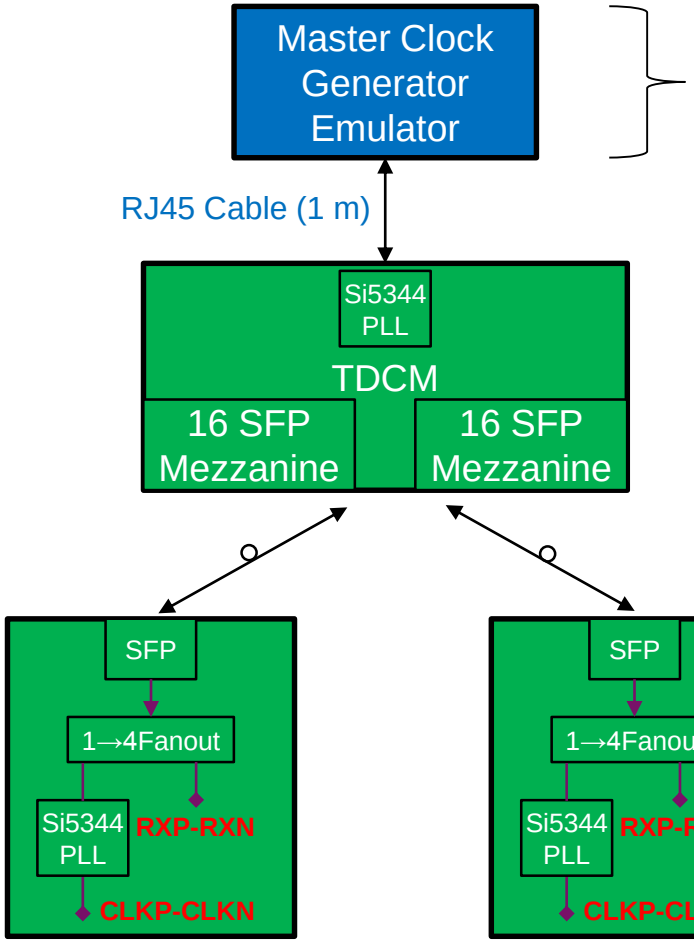
Saclay, 09 July 2021

Irfu

PRESENTATION OUTLINE

- **Characterization of the TDCM**
- Transmit path of TDCM versus new clock distributor
- Proposal for interface and communication with clock distribution system
- Summary & future work

TDCM CHARACTERIZATION SETUP



Enclustra Starter Kit with Xilinx Spartan 6 FPGA Module
50 MHz local oscillator + FPGA internal PLL produces 100 MHz reference clock

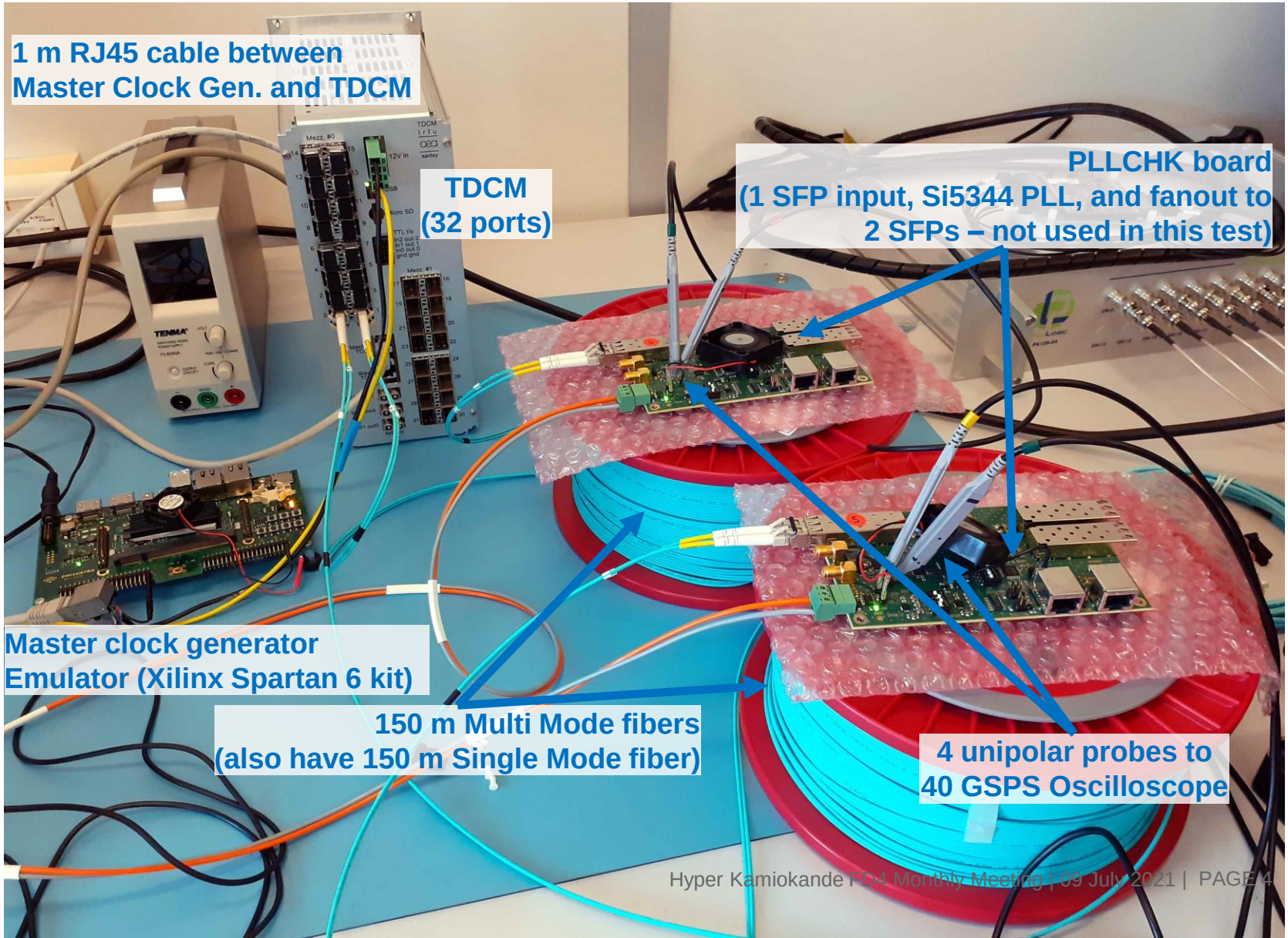
Optical fiber
3m or 150 m (Multi Mode)
150 m (Single Mode)

Differential probing:
- On a copy of SFP RX signal
- At one of PLL outputs

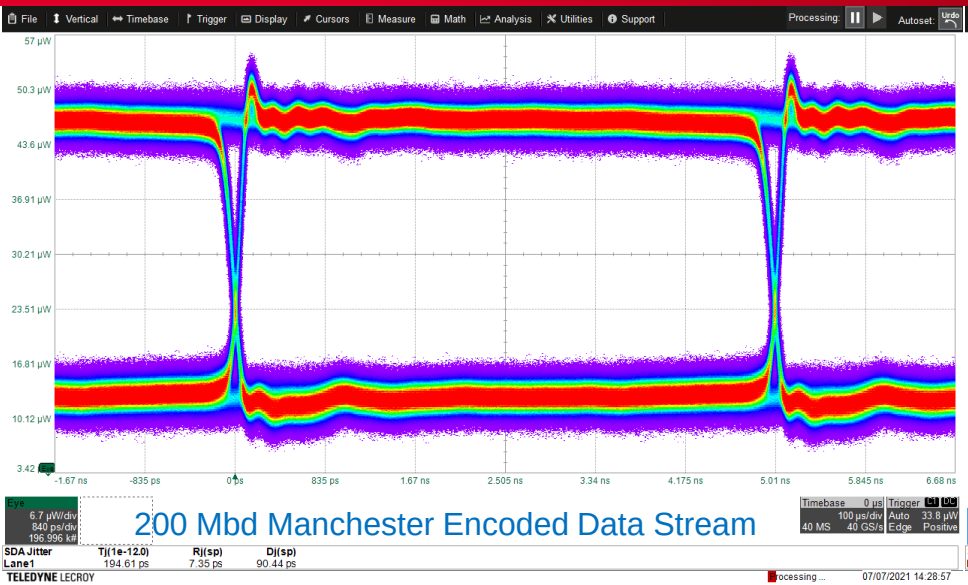
Test equipment

- 80 GSPS LeCroy Oscilloscope / Serial Data Analyzer
- 1 Optical probe; 4 unipolar probes

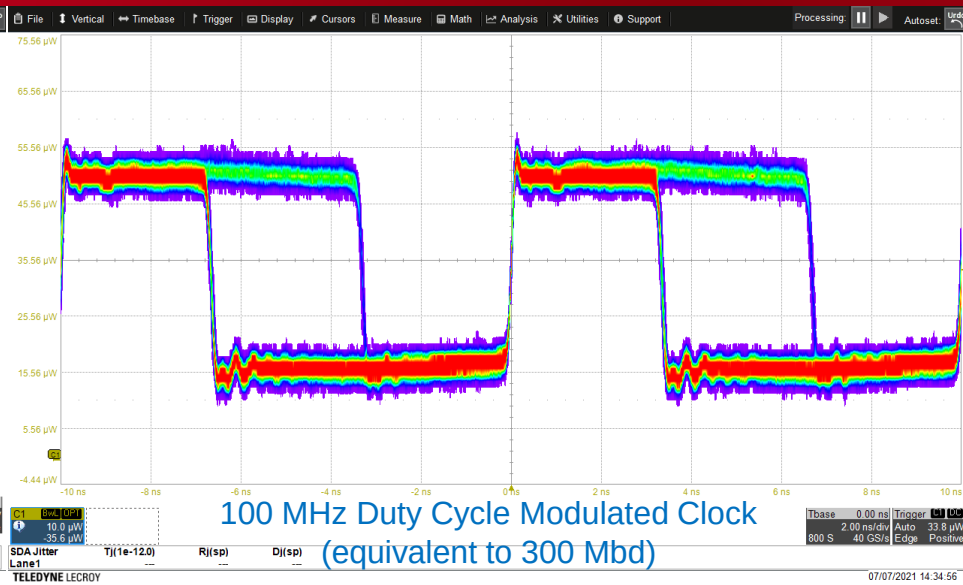
TDCM CHARACTERIZATION SETUP



EYE DIAGRAMS AT VARIOUS SPEEDS



200 Mbd Manchester Encoded Data Stream



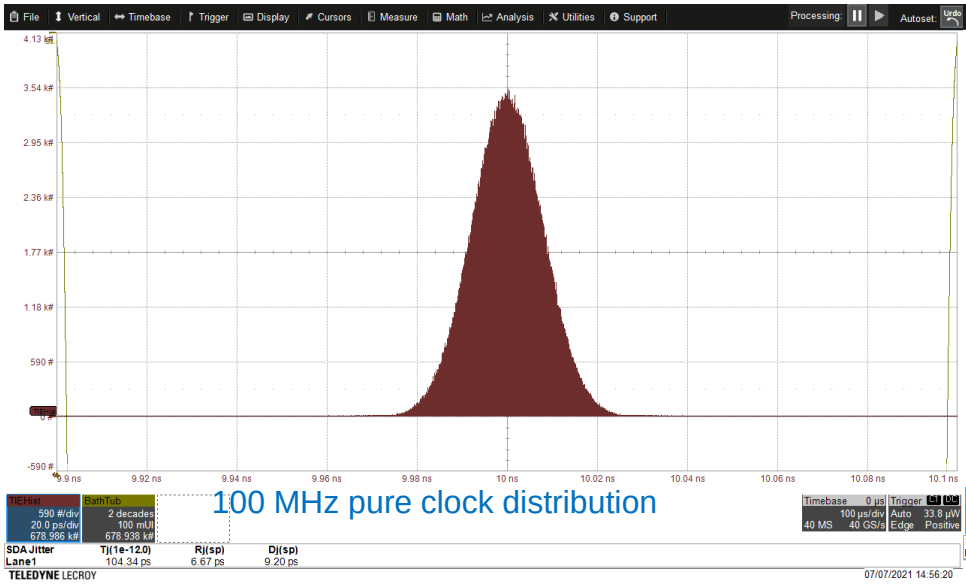
100 MHz Duty Cycle Modulated Clock (equivalent to 300 Mbd)

300 Mbps PRBS31 pattern

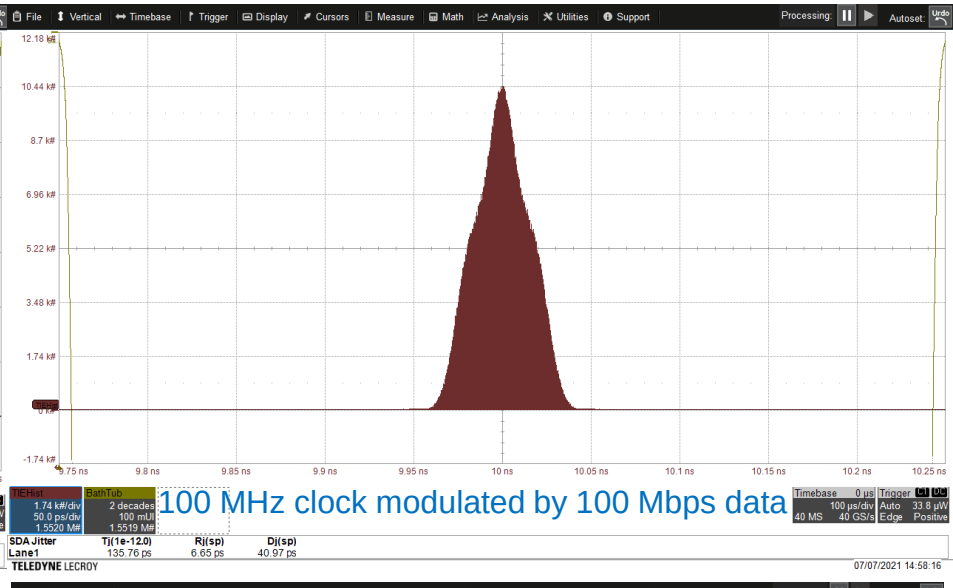
600 Mbps PRBS31 pattern

All measurements done on TDCM output port#0, 150 m Single Mode fiber and O/E converter of oscilloscope (BW limit: 3 GHz)

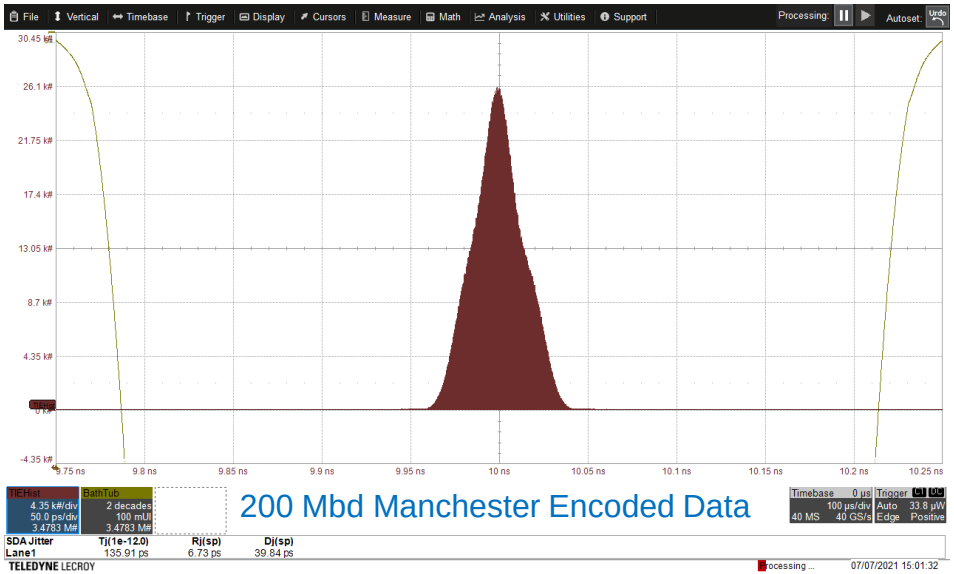
JITTER MEASUREMENTS



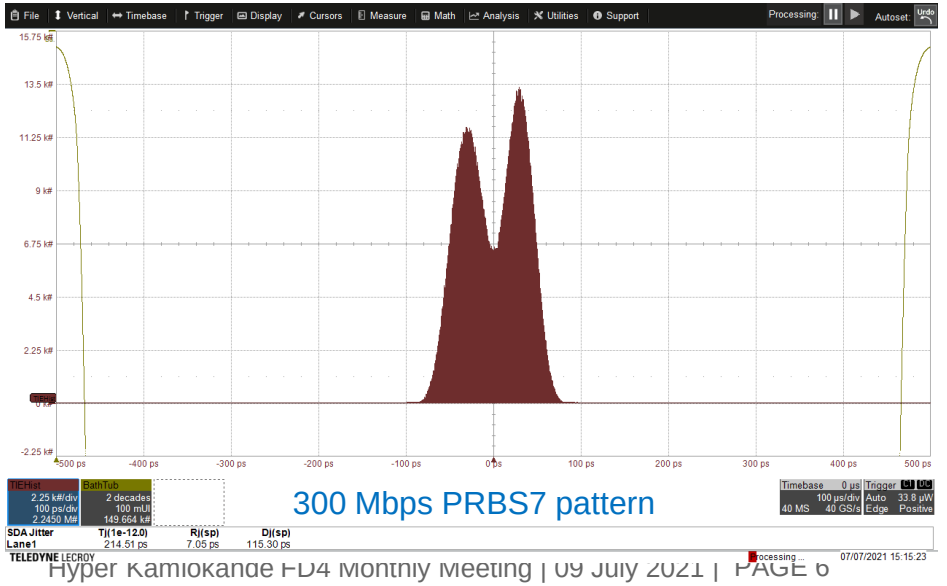
100 MHz pure clock distribution



100 MHz clock modulated by 100 Mbps data



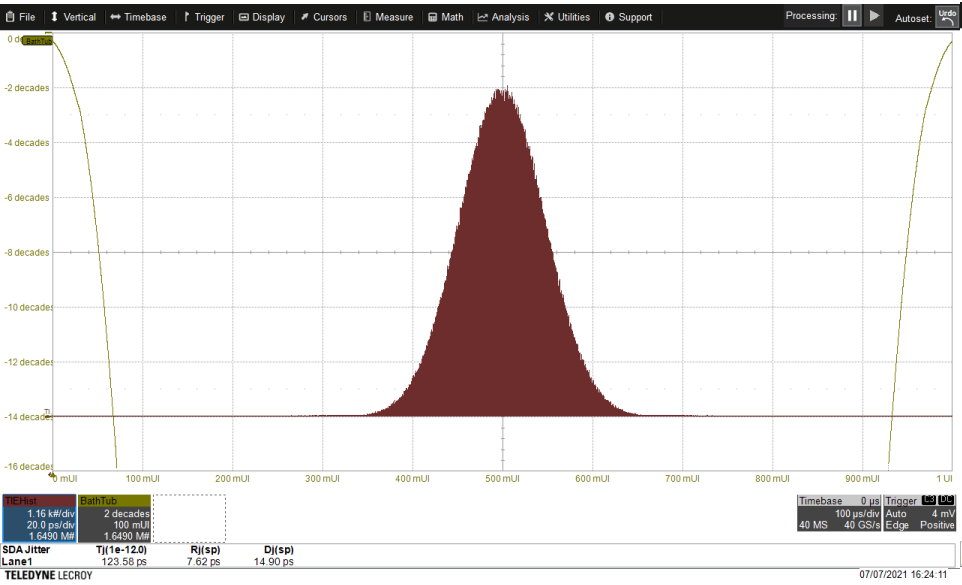
200 Mbd Manchester Encoded Data



300 Mbps PRBS7 pattern

All measurements done on TDCM output port#0, 150 m Single Mode fiber and O/E converter of oscilloscope (BW limit: 3 GHz)

JITTER MEASUREMENTS



Signal declared as “Clock”

Signal declared as “Custom” [Serial Data]

Test conditions

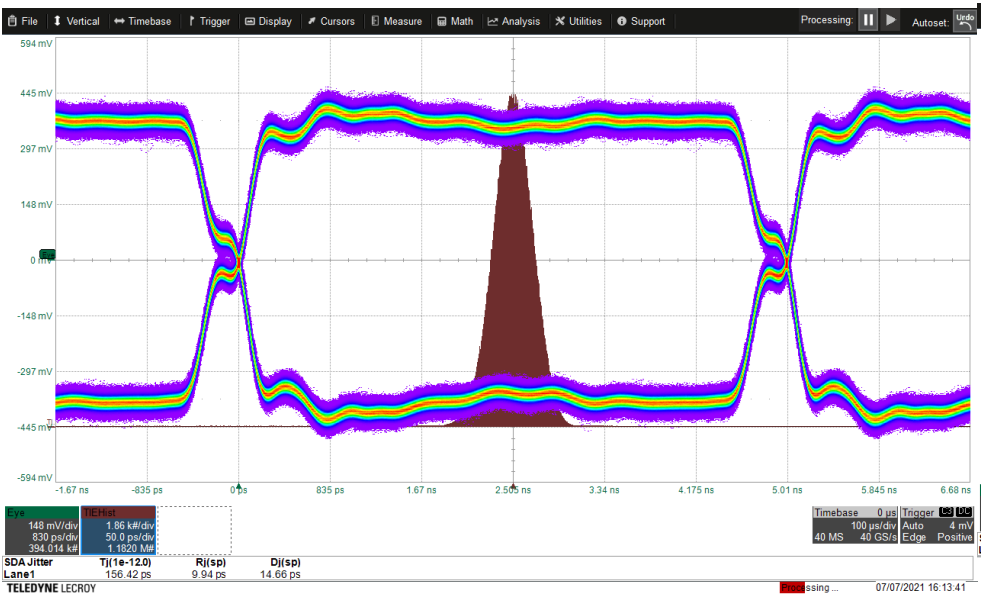
- TDCM configured for 100 MHz pure clock distribution
- Measurements done on TDCM output port#0, 150 m Multi Mode fiber, O/E converter board and 2 unipolar probes connected to RXP-RXN output
- Only oscilloscope setting is changed to declare analyzed signal is “Clock” or “Custom” [Serial Data]

Result and interpretation

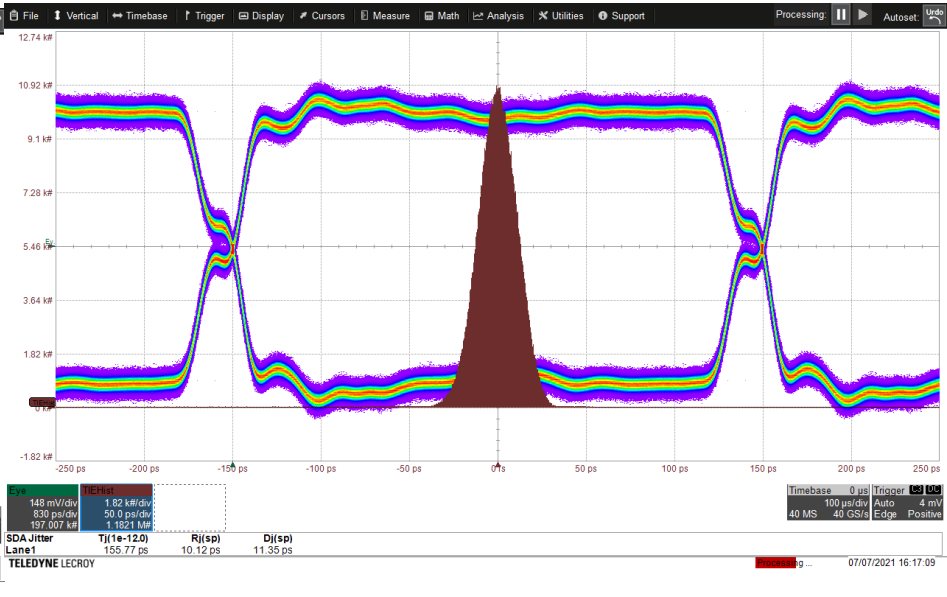
- “Clock” oscilloscope setting: Rj=7.62 ps Dj=14.9 ps. Only the rising edge of the signal is considered
- “Custom [Serial Data]”: Rj=20.34 ps Dj=195.81 ps. Both edges are considered and distorted falling edge causes the double distribution and large deterministic jitter

→ More investigations needed (e.g. with Phase Noise Analyzer) to characterize Dj of TDCM and track its origin

JITTER MEASUREMENTS



TDCM in pure 100 MHz clock distribution mode



TDCM in clock duty cycle modulation mode

Test conditions

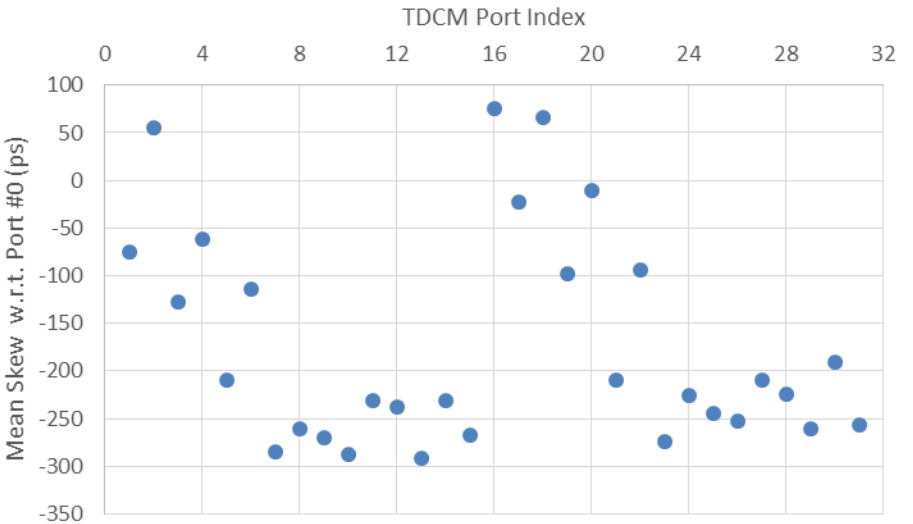
- TDCM output port#0 is used, 150 m Multi Mode fiber, connected to PLL board. Measurement is done at the output of the PLL at the receiving end (CLKP-CLKN outputs) and probed with 2 unipolar oscilloscope probes

Result and interpretation

- After jitter cleaning at receiving end, Rj=9.94 ps Dj=14.66 ps (TDCM in pure 100 MHz clock distribution mode) and Rj=10.12 ps Dj=11.35 ps (TDCM in CDCM-3-1 mode)
- Both edges at PLL output are a distorted, but crossing remain clean (hence low Dj)

→ PLL at receiving end seems effective to filter out Dj coming from TDCM

CLOCK ALIGNMENT BETWEEN DISTINCT OUTPUTS



Test conditions

- TDCM configured for 100 MHz pure clock distribution. 3 m Multi Mode fibers to 2 O/E + PLL boards
- Signals are probed on the two boards after PLL, at CLKP-CLKN outputs. Skew is measured and accumulated by oscilloscope, mean skew is plot taking port #0 as reference and plugging the second board on the 31 other ports successively

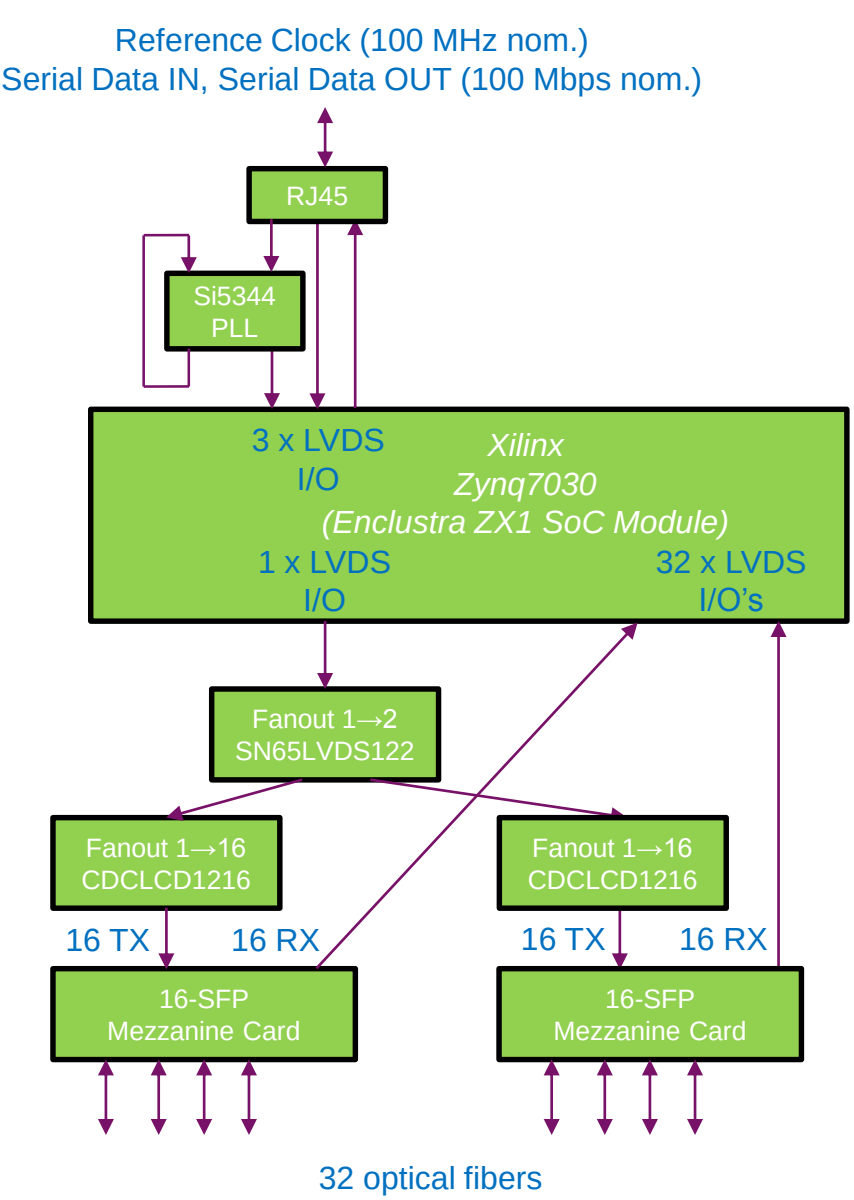
Result and interpretation

- Mean skew between TDCM ports spreads ~400 ps. Between any port couple, skew rms is ~30 ps. Two comparable patterns seen: trace mismatch length on 2 mezzanine cards + mismatch on mother board
- Independent test of multiple power cycling shows repeatability of measurements (within few ps, did not try to make long term tests and temperature variations)

PRESENTATION OUTLINE

- Characterization of the TDCM
- **Transmit path of TDCM versus new clock distributor**
- Proposal for interface and communication with clock distribution system
- Summary & future work

TDCM TX PATH



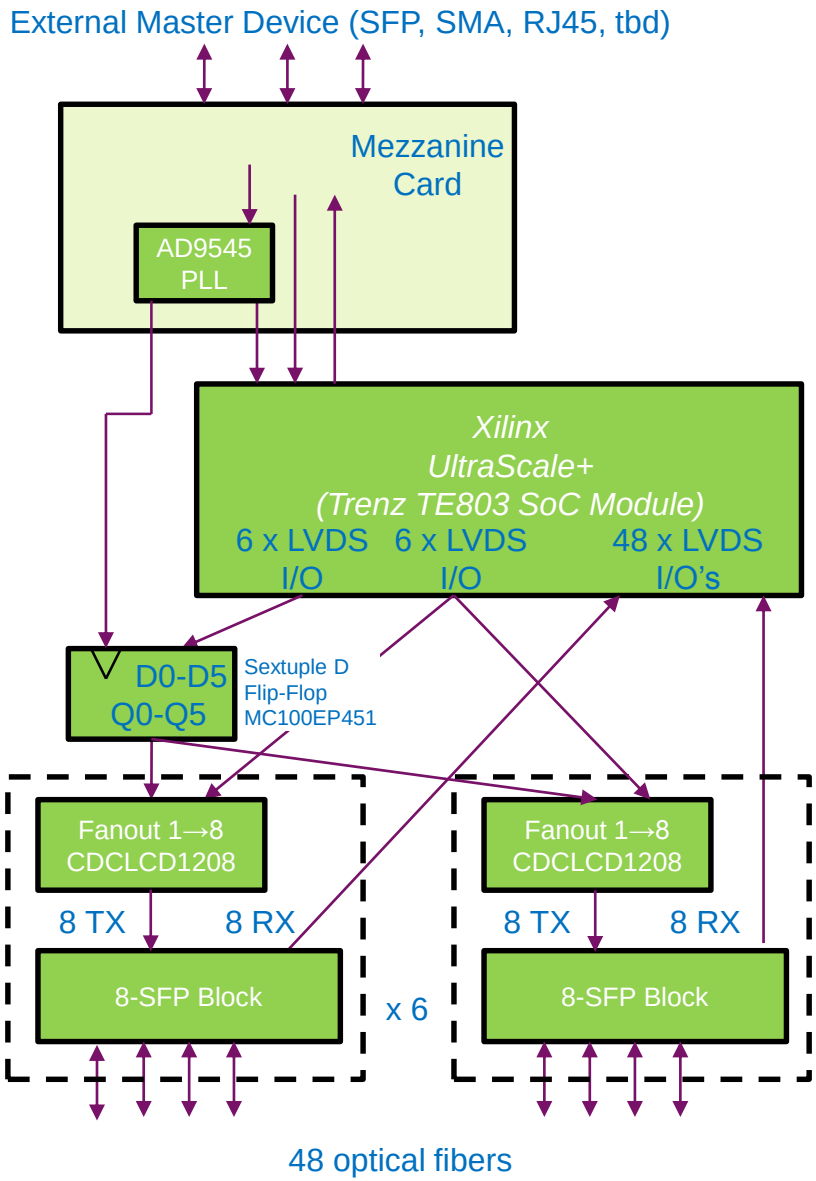
Principles

- Fanout network from 1 to 32 in the back-end to front-end direction
- Set of 32 point-to-point links in the front-end to back-end direction
- All serial communication done via ordinary LVDS FPGA I/O pins + ISERDES2/OSERDESE2 primitives
- Current operation is at 100 MHz – 600 Mbps TX serialization rate and sub-multiples

Limitations

- Same signal is serialized to all 32 outputs
- The FPGA is in the clock distribution path, any skew and jitter caused by the FPGA is directly propagated downstream

HK 48-PORT DEMONSTRATOR TX PATH



Principles

- TX path made by 6 fanout from 1 to 8. Each fanout can take input from one of 2 sources:
  - output pin from FPGA (ODDRE1, OSERDESE3…)
  - output of sextuple D Flip-Flop fed with serial data from FPGA and clocked by purified master clock.
- All serial communication done via ordinary LVDS FPGA I/O pins (1250 Mbps capable, but design aiming at 500 Mbps and sub-multiples)

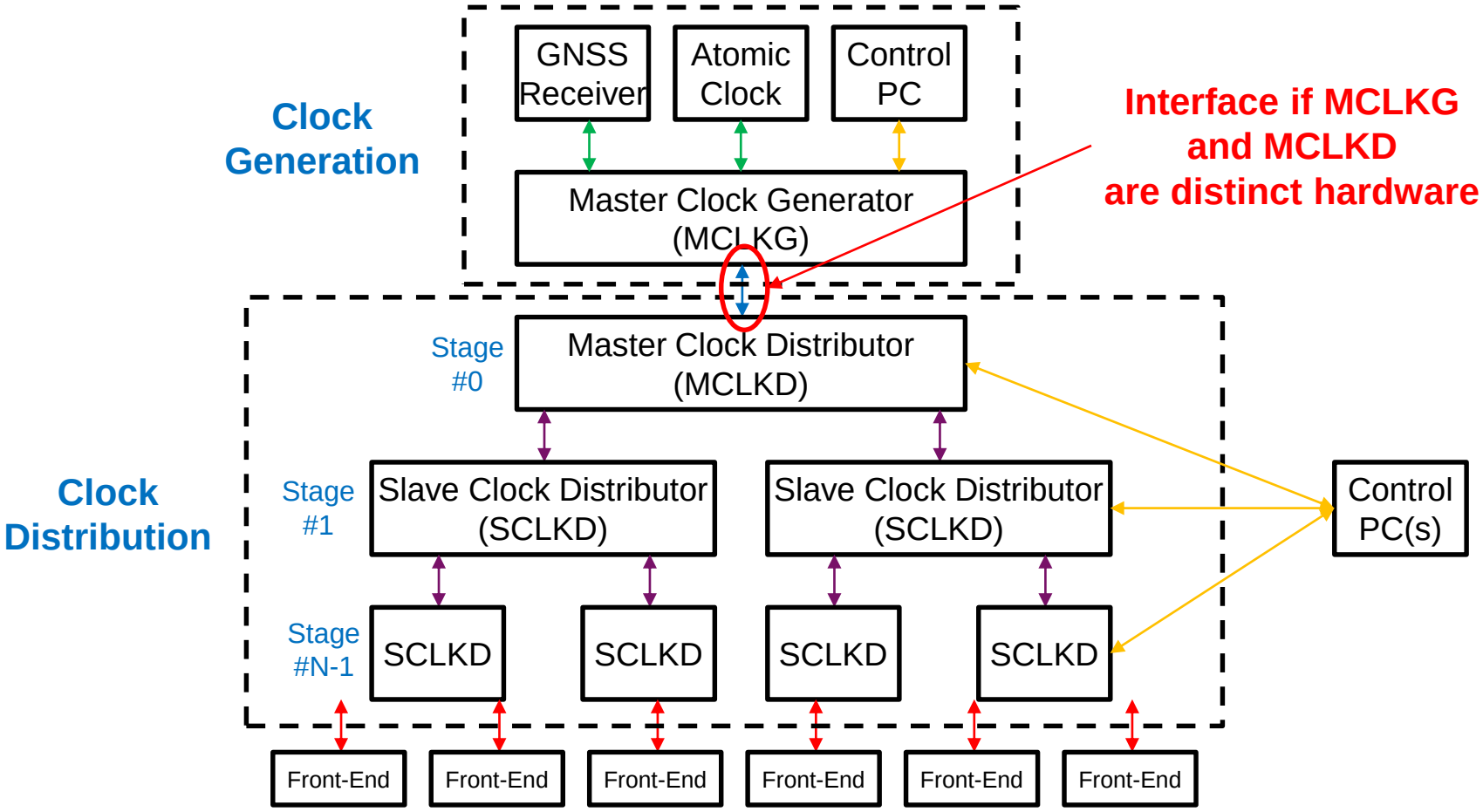
→ Major changes compared to TDCM:

- Using the re-timing D Flip-Flops, the FPGA is no longer in series in the critical path of the distributor
- Granularity of 8 ports for setting links as pure clock distributor, or clock & data distributor

PRESENTATION OUTLINE

- Characterization of the TDCM
- Transmit path of TDCM versus new clock distributor
- **Proposal for interface and communication with clock distribution system**
- Summary & future work

CLOCK DISTRIBUTION SYSTEM ARCHITECTURE



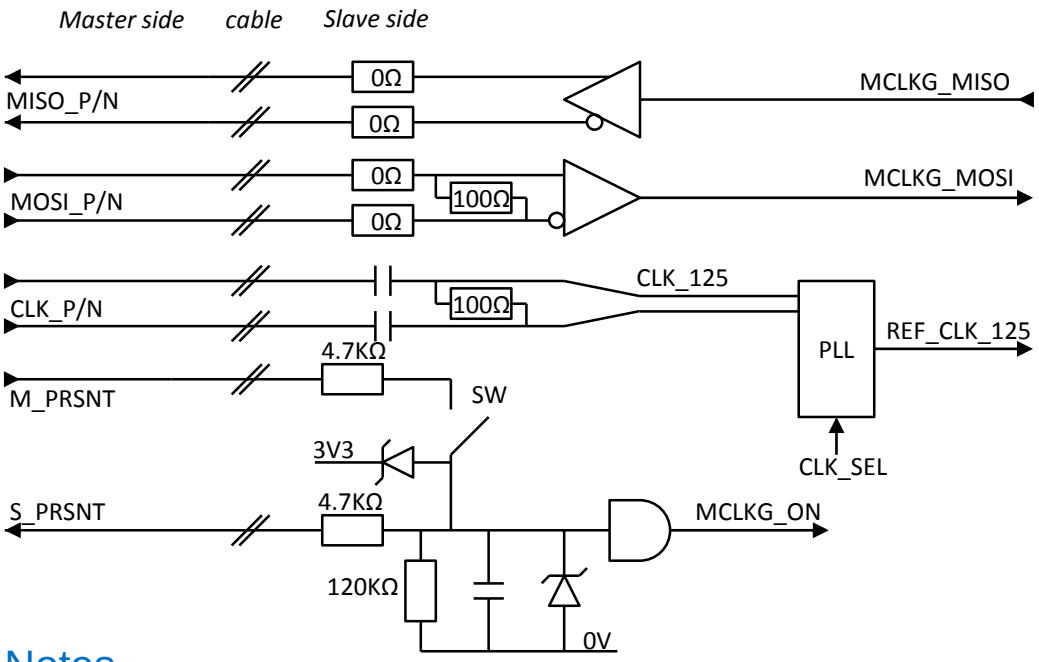
Notes

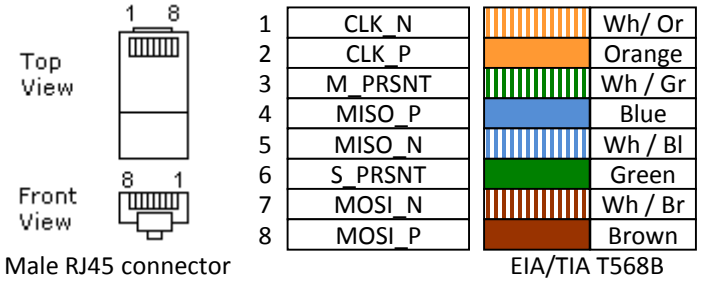
- Master Clock Generator and Master Clock Distributor may be two separate physical entities or could be merged on the same hardware

MCLKG TO MCLKD DEMONSTRATOR INTERFACE PROPOSAL

Concept

- RJ45 cable, LVDS: 1 pair 125 MHz clock + 1 pair 125 Mbps data from Master to Slave. 1 pair 125 Mbps data from Slave to Master. Two optional unipolar signal for partner presence detection



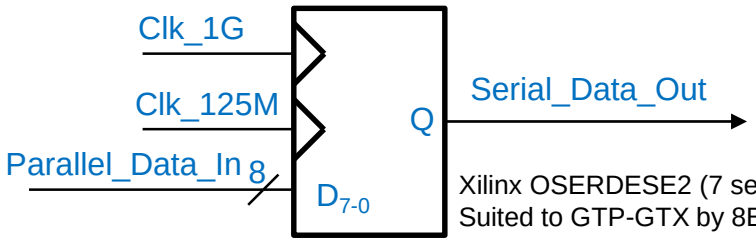


(Note: pin definition similar to that used in T2K nd280m; If adopted for HK R&D, it allows use of TDCM for prototyping phase in the easiest way)

Notes

- M_PRSNT is set high by the master to indicate its presence. S_PRSNT is set high by the slave when the presence of the master is detected and the slave is set to operate with the clock from the master. A switch may be used to instruct the slave to run on its own local clock even when it is connected to the master. Differential outputs should be tri-stated on DC coupled lines when partner is not detected
- 0 ohm resistors may be replaced by capacitors if AC-coupling is desired
- The use of ESD diodes for protection and common mode chokes for reduced EMI is encouraged

DISTRIBUTOR DEMONSTRATOR “UNIVERSAL” SERIAL TRANSMITTER



Xilinx OSERDESE2 (7 series) or OSERDESE3 (UltraScale+)
Suited to GTP-GTX by 8B-10B encoder bypass and 16 or 32 bit parallel interface

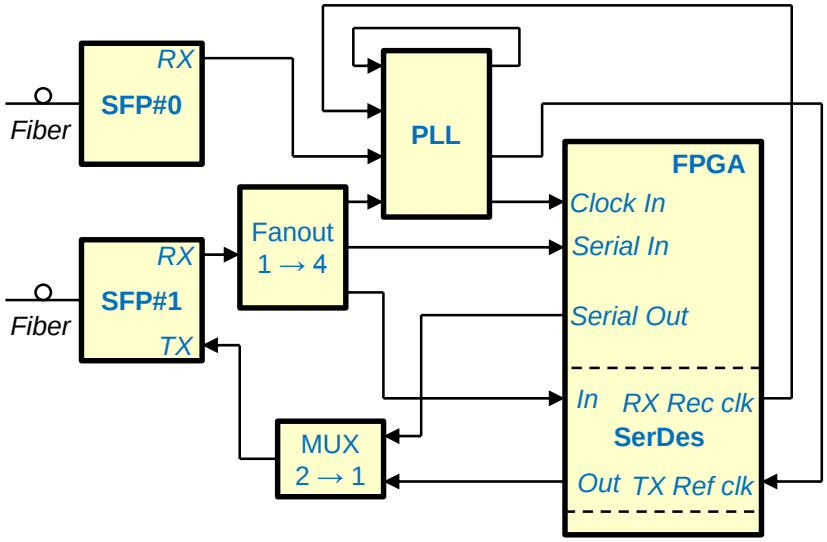
Note: in the table $\underline{d_x}$ means $\overline{d_x}$

	D7-D0 Serializer Input	Output on media
Parallel Input	$d_7\ d_6\ d_5\ d_4\ d_3\ d_2\ d_1\ d_0$	1 Gbps (need encoder for DC-balance, e.g. self-synchronizing scrambler)
Parallel Input	$d_3\ \underline{d_3}\ d_2\ \underline{d_2}\ d_1\ \underline{d_1}\ d_0\ \underline{d_0}$	500 Mbps (data) 1 Gbaud (symbol)
Parallel Input	$d_1\ d_1\ \underline{d_1}\ \underline{d_1}\ d_0\ d_0\ \underline{d_0}\ \underline{d_0}$	250 Mbps (data) 500 Mbaud (symbol)
Parallel Input	$d_0\ d_0\ d_0\ d_0\ \underline{d_0}\ \underline{d_0}\ \underline{d_0}\ \underline{d_0}$	125 Mbps (data) 250 Mbaud (symbol)
Parallel Input	$0\ 0\ 0\ 1\ 1\ 1\ d_0\ d_0$	125 MHz clock modulated + 125 Mbps data
Parallel Input	$0\ 0\ 0\ 1\ 1\ 1\ d_0\ d_0$ $0\ 0\ 0\ 1\ 1\ 1\ \underline{d_0}\ \underline{d_0}$	125 MHz clock modulated + 62.5 Mbps data (DC balanced)
Parallel Input	$0\ 0\ 0\ 0\ 1\ 1\ 1\ 1$	Pure 125 MHz clock

Notes

- Manchester encoding allows simple encoding/decoding; best DC-balance; error detection up to 50% BER and message timing resolution of ½ line rate (instead of 1/10 line rate for 8B/10B encoding)

SUGGESTED FRONT-END "UNIVERSAL" RECEIVER AND TRANSMITTER



- A PLL candidate with 4 inputs and zero-delay capability is Silicon Labs Si5344
- Other types of PLL are possible; if the number of inputs is not enough, it could be extended using a multiplexer
- A buffer may be added in the feedback loop of the PLL to compensate for the delay of the 1 → 4 fanout buffer

Clock Fiber and Data Fiber use case

- Pure clock enters SFP#0. Jitter is filtered by PLL and serves as reference clock for data reception and transmission by I/O primitive or SerDes hard IP of FPGA

Clock embedded in Data use case

- Serial data with embedded clock enters SFP#1. Clock and data are recovered by SerDes hard IP of FPGA. Clock is output to PLL for purification and fed back to FPGA to clock TX side of SerDes

Data embedded in Clock use case

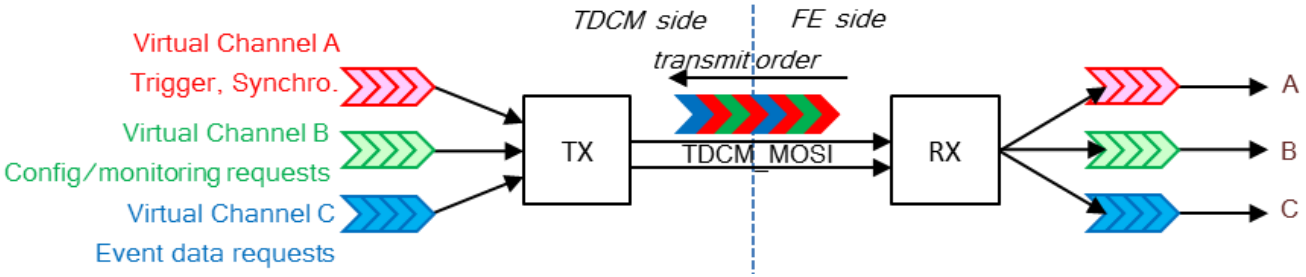
- Clock with embedded data enters SFP#1. Data are extracted from modulated clock using purified clock from PLL. This clock also drives the transmit side of an I/O primitive or high speed SerDes IP

TECHNIQUES FOR DATA TRANSMISSION WITH DETERMINISTIC LATENCY

Hypothesis: at least one type of info from MCLKG has to reach every FE at the same 125 MHz clock cycle

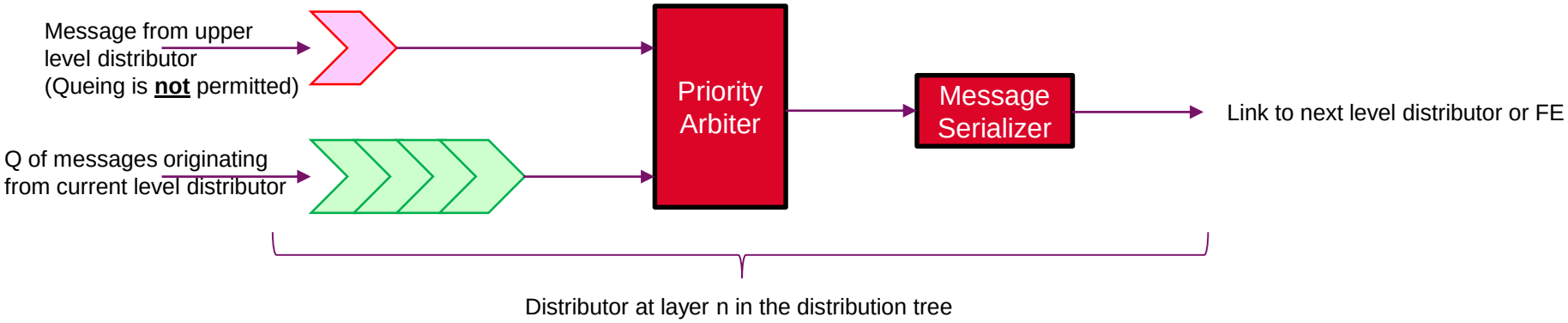
Multiple virtual channels using time division multiplexing

- Messages from each Virtual Channel are sliced in pieces (1 bit, 1 byte or more) sent in round-robin
- Deterministic latency messaging with short latency and mixing fixed/variable size messages



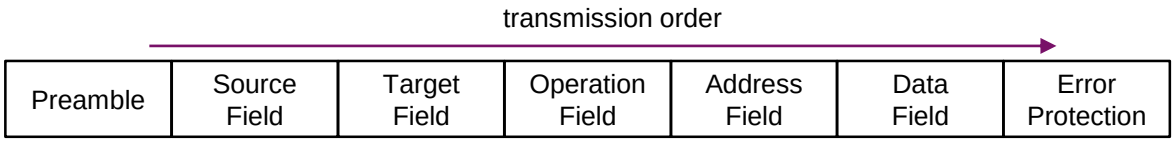
Priority queue

- If no message is under reception from the high priority Q, one message from local Q can be sent
- Store and forward scheme. Can only work with fixed size packets. Allows only one source of deterministic latency messages. Local message Q can only carry asynchronous messages



→ Priority Q scheme has more limitations but probably sufficient for HK and is simpler to implement

PROPOSAL OF MESSAGE FORMAT FROM MASTER CLOCK GENERATOR TO DISTRIBUTORS AND FE



The clock cycle when the message ends gives the absolute time reference with respect to the master clock. Assuming 125 MHz, message dating resolution is 8 ns.

1 bit: parity. More sophisticated error detection and error correction may be added in the final design

32 bit (64 bit?) data to write, or all zeroes in case of a read request

16 bit or more: address of the RD or WR transaction to perform in the target

8 bit typically: 1 Read bit, 1 Write bit, 4 byte-enable bits to address each byte in a 32-bit word individually, 1 bit to specify if the write transaction requires an acknowledge or not. A read request obviously requires a response which is the data read

12 bits is the minimum for addressing 4096 nodes. 1 bit to indicate broadcast, or set target to all 1's for broadcast. 2 bits for broadcast field allows making groups, e.g. all front-ends, all distributors, all front-ends and distributors

2 bits allow addressing 4 layers of distributors. Each FE need to know to whom send a response. Only possible targets for a FE are the top level clock distributor or intermediate distributors. Communication is not supported between distinct FE or between distributors that are at the same level in the fanout tree

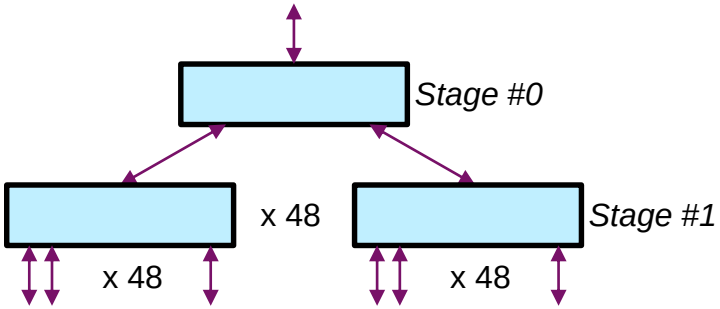
The minimum required is 1 start bit. A longer preamble may be defined if it is beneficial.

Principles

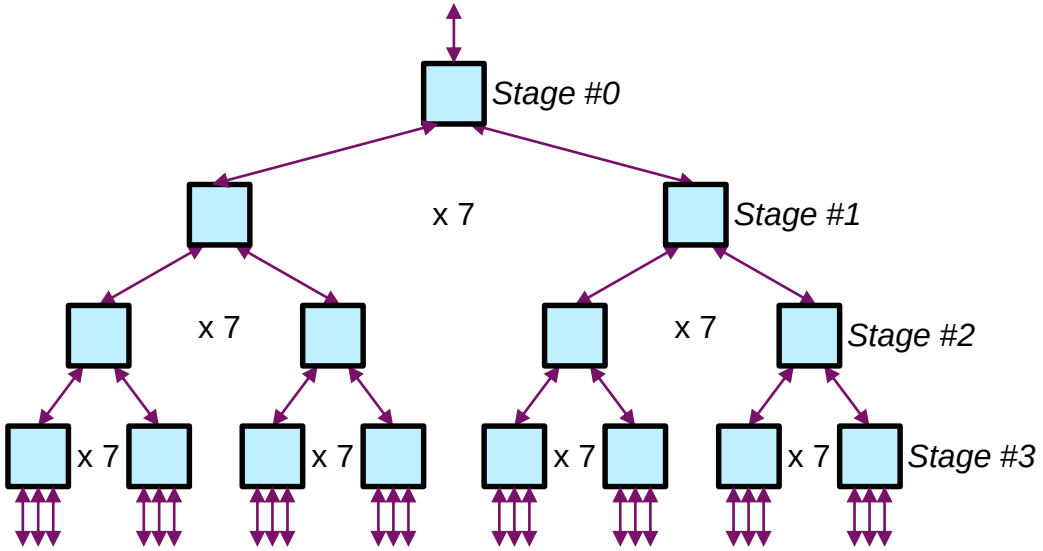
- Only one type of message supported. Fixed size (~75 bits) and fixed format
- Each FE is seen by Master Clock Generator and distributors as a remotely accessible memory space. This virtual memory space is divided into multiple regions: Trigger, Digitizer config, HV, etc.

→ This provides the required level of abstraction for various groups to develop their part in parallel

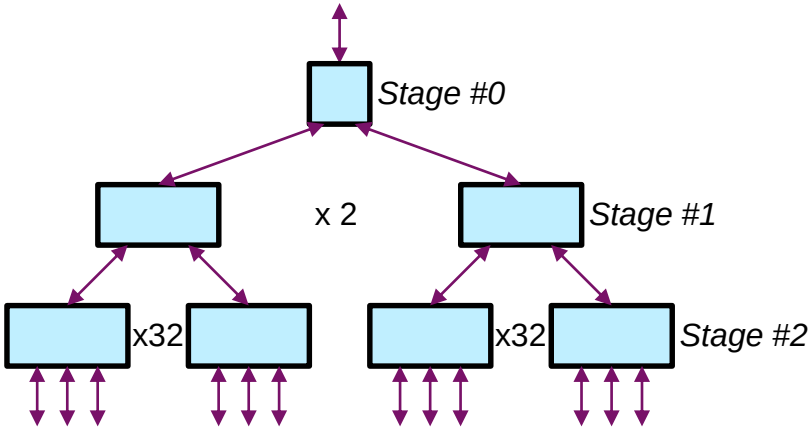
DISTRIBUTOR ARCHITECTURE OPTIONS



2-stage tree of 49-port distributors
(2304 ports max.)



4-stage tree of 8-port distributors
(2401 ports max.)



Dual 2-stage tree of 33-port distributors
(2048 ports max.)

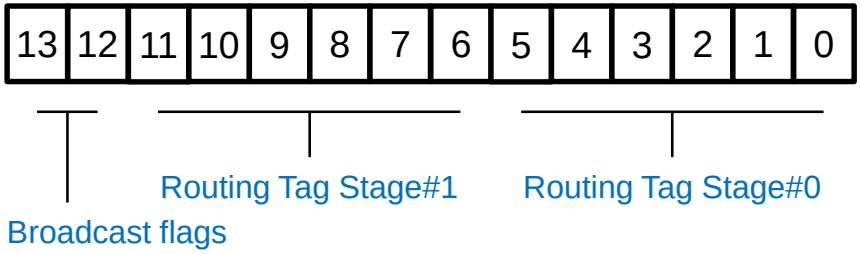
Principle of node addressing

- Message target address is the concatenation of distributor port# of each successive stage
- Advantage: routing in each distributor is trivial (no routing table needed), addressing is independent of distributor architecture and also independent of the physical elements, their location, their role, etc.
- Besides this logical node numbering, a correspondence table is maintained to map to physical units (e.g. detector at position X, Y, Z)

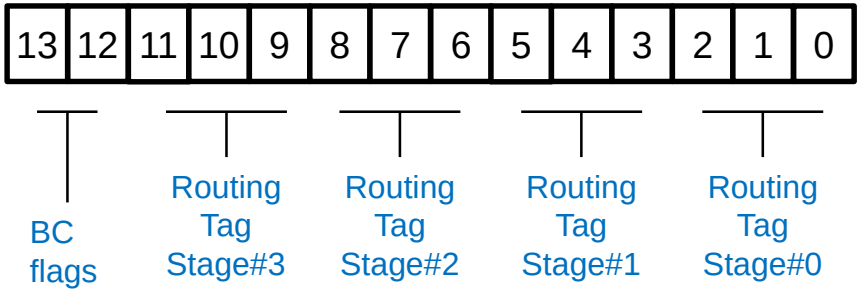
TARGET NODE ADDRESS SCHEME AND MESSAGES ROUTING

Distributor Architecture Variant

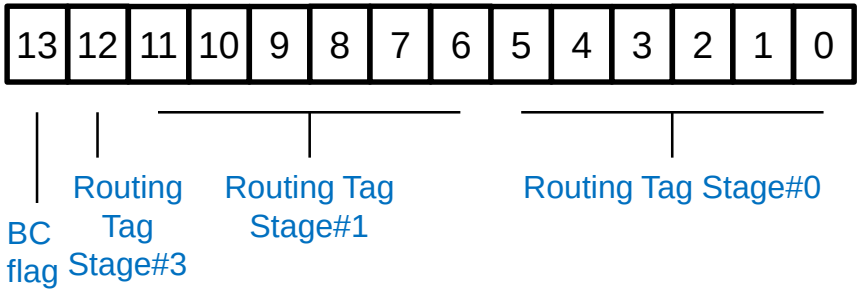
- 2-stage Tree of 49-port distributors



- 4-stage Tree of 8-port distributors



- Dual 2-stage Tree of 33-port distributors



Proposed 14-bit Target Node Field

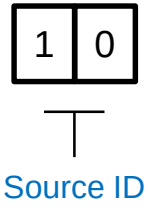
- Routing Tag = 0 to 47 correspond to concentrator downstream port 0 to 47
- Routing tag = 48 means message is for the concentrator itself
- Broadcast flags: 00 – unicast; 01 – all FEs; 10 – all distributors; 11 – all FEs and distributors

- Routing Tag = 0 to 6 correspond to concentrator downstream port 0 to 6
- Routing tag = 7 means message is for the concentrator itself
- Broadcast flags: 00 – unicast; 01 – all FEs; 10 – all distributors; 11 – all FEs and distributors

- Routing Tag = 0 to 31 correspond to concentrator downstream port 0 to 31
- Routing tag = 32 means message is for the concentrator itself
- Bit 12 of target field determines the first distribution branch among the two
- Broadcast flag to all FEs and distributors

SOURCE NODE ADDRESS SCHEME

Proposed 2-bit Source Node Field



- The source that is given for a message is the index of the stage of the distributor in the tree
- Source “00” corresponds to the root distributor
- Source “01” is a leaf distributor in a 2-stage tree architecture
- Source “01” and “10” are intermediate distributors in a 4-stage tree architecture

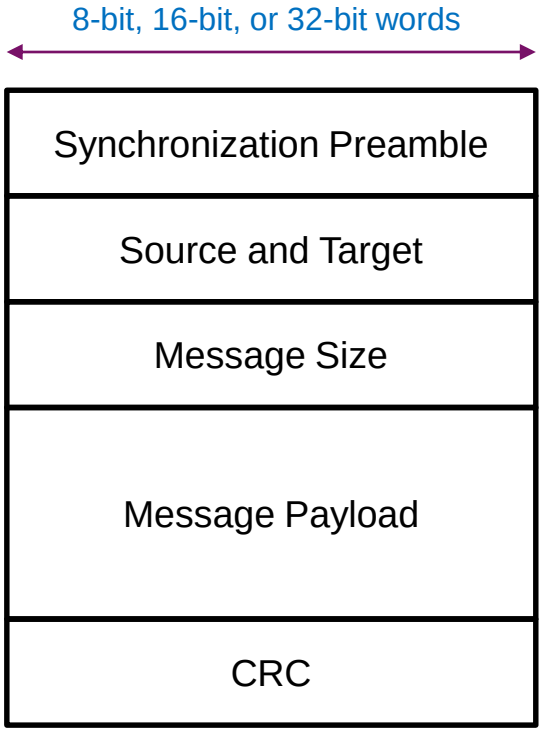
Note

- This scheme does not identify each distributor uniquely
- It is sufficient because a target node can only receive messages from one parent distributor at each stage of the tree, and not from any distributor within that stage

Other notes and source/target fields

- Only the root distributor can broadcast messages
- Distributors other that root propagate messages from upper layer (broadcast and point-to-point) but can only issue messages themselves that are for an individual target node (i.e. not broadcast)
- Root distributor handles deterministic latency messages while distributors at the last stage handle front-end monitoring, slow control, alarms, firmware updates download, etc
- Distributors at intermediate layers (if any) mostly perform message relay from upper layer of distributor(s) to lower layer

MESSAGES FROM FE TO CLOCK DISTRIBUTORS AND MASTER CLOCK GENERATOR



Typically a word like 0xAA55

Same as for outbound messages; source and target fields meaning are exchanged. Broacast bit not needed

Useful to have this in the header

Application dependent. Should be an integer number of words

Error detection, e.g. CRC32. Error correction requires more complex logic

Notes

- Variable size packets may be more appropriate in this direction due to the diversity of information
- Only workable if deterministic latency (to 1 or few clock cycle precision) is not required in this direction
- Otherwise, using multiple virtual channels is a possible solution - at significant additional complexity
- Packet encoding should use a more bandwidth efficient scheme than Manchester, e.g. self synchronizing scrambler/descrambler, but 8B-10B decoding in UltraScale+ FPGA fabric seems difficult to implement because ordinary I/O pin SerDes does not support 1:5 or 1:10 ratio (only 1:2, 1:4 or 1:8)

SUMMARY & FUTURE WORK

Summary

- Some characterization measurements done on TDCM. Good performance in terms of clock latency determinism, repeatability, random jitter, but probably out of HK specs on deterministic jitter (the TDCM was not designed for HK). Still, current tests indicate this Dj can be removed by PLL at the receiving end
- Proposal of interface, data encoding, message format, etc. for demonstrator
- PCB layout of 48 optical port clock distributor demonstrator progressing (not simple!)

Future work

- Investigate deterministic jitter on TDCM. Evaluate performance of cascaded TDCMs. Adapt firmware for operation at 125 MHz instead of 100 MHz
- Pursue development of 48-port HK clock distributor demonstrator board + clock mezzanine
- Initiate/continue discussions to make clock distributor demonstrator interoperable with developments of other groups