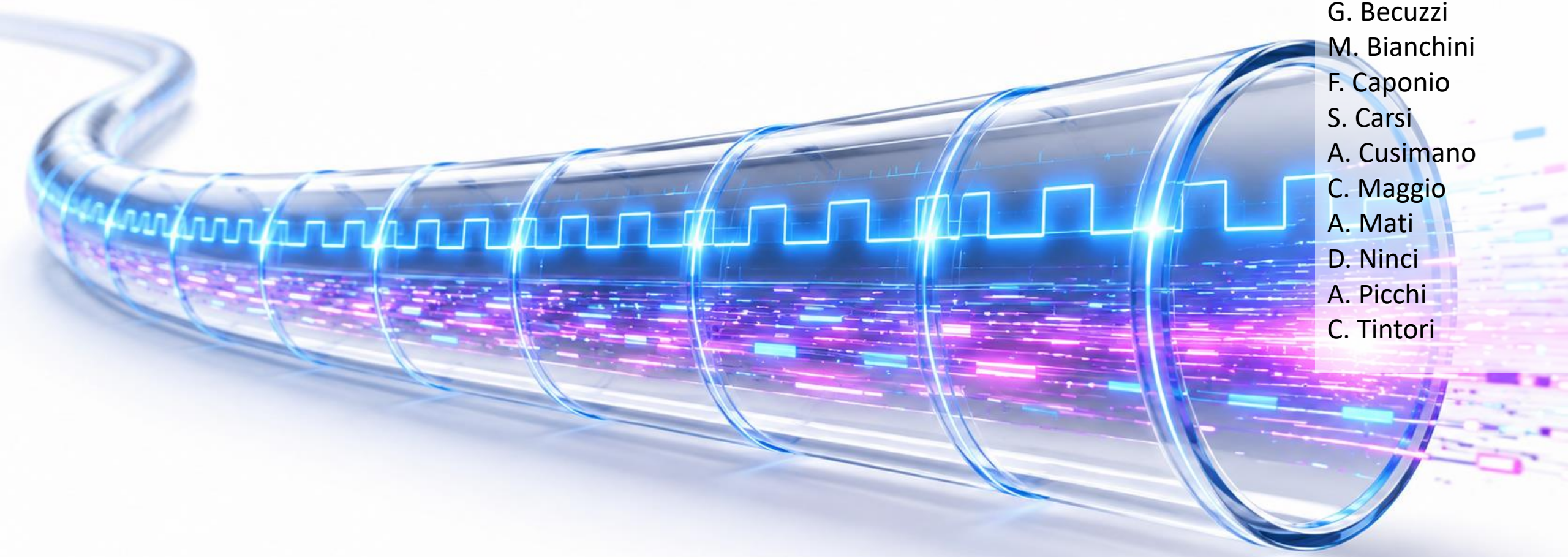
TD-link: A Deterministic Optical Daisy-Chain Link for Synchronous Data Acquisition in Large-Scale Detector Systems

A. Abba
G. Becuzzi
M. Bianchini
F. Caponio
S. Carsi
A. Cusimano
C. Maggio
A. Mati
D. Ninci
A. Picchi
C. Tintori



CAEN
Tools for Discovery

Nuclear Instruments

FERS: detector readout chains for every (photon) need

Detector array - Multichannel

64-channel Frontend unit 1

64-channel Frontend unit 2

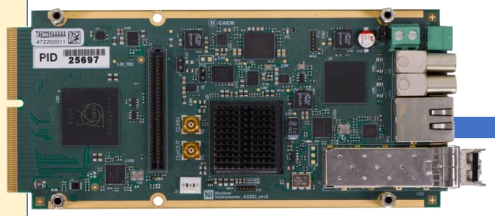
64-channel Frontend unit 4

. . .

TDLink 3.125 Gb/s

optical fibers

1 board/node to maximize readout bandwidth



Node 1 Node 5

Node 2 Node 6

CONCENTRATOR
Sync & Readout
DAQ Interface

Node 3 Node 7

Node 4 Node 8

TDLink 3.125 Gb/s

Daisy chain → Up to 16 boards/node

Up to 128 total boards

Up to 8k channels

Daisy-chain sync

64-channel Frontend unit N-3

64-channel Frontend unit N-2

64-channel Frontend unit N-1

64-channel Frontend unit N

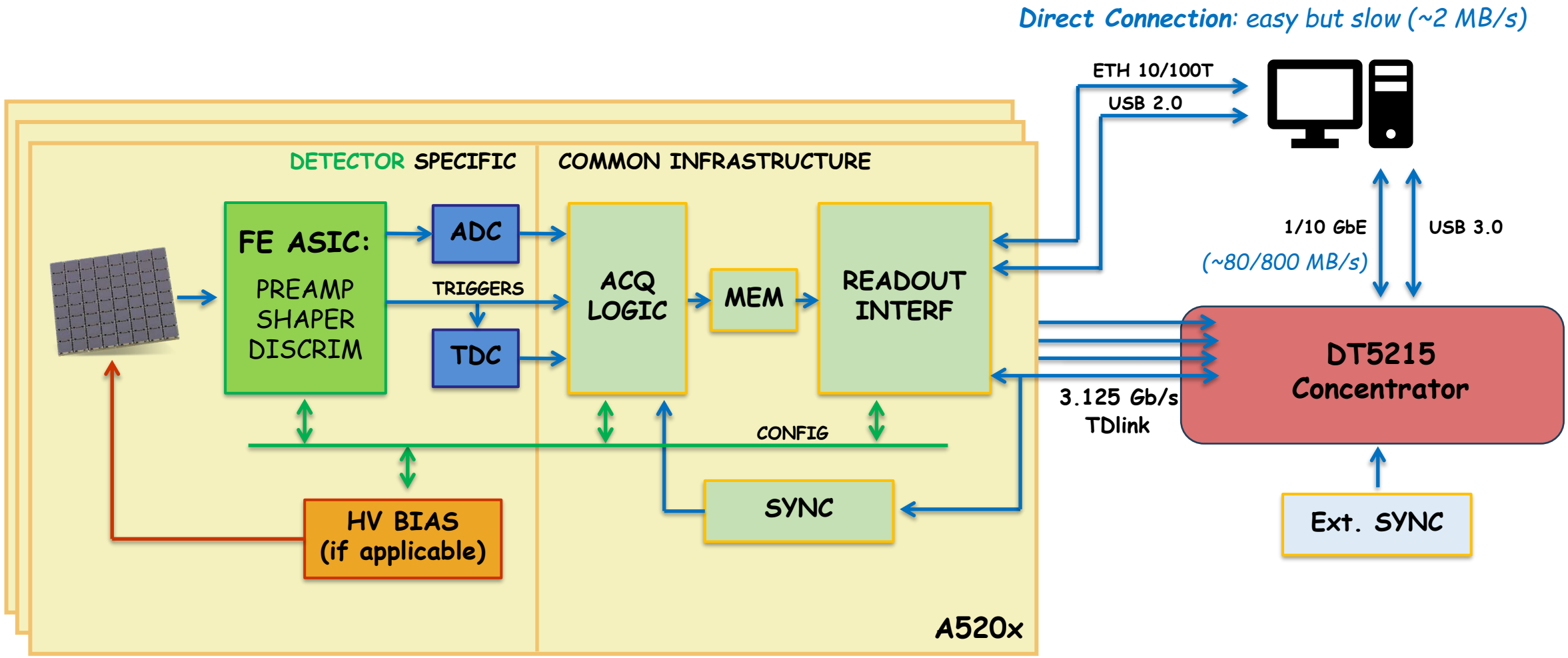
Detector array - Multichannel

10/100MbE – USB 2.0

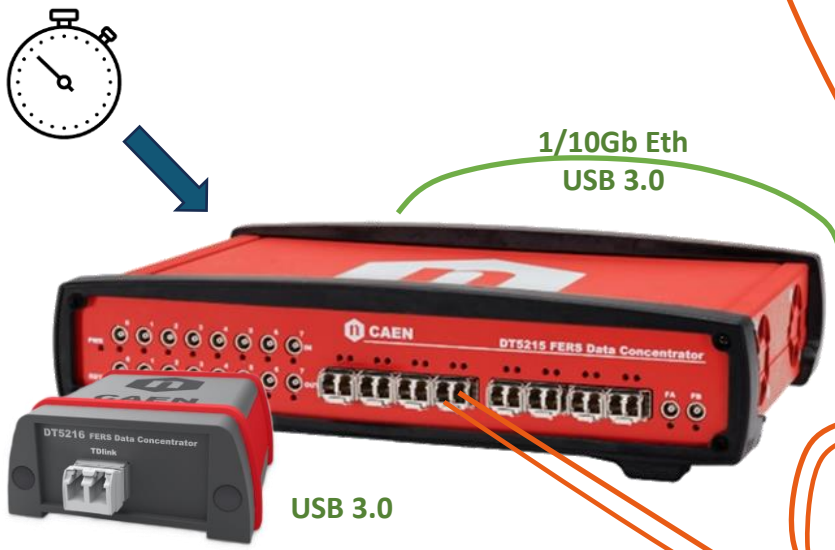
1/10GbE – USB 3.0

DAQ PC
Janus Software

FERS: detector readout chains for every (photon) need

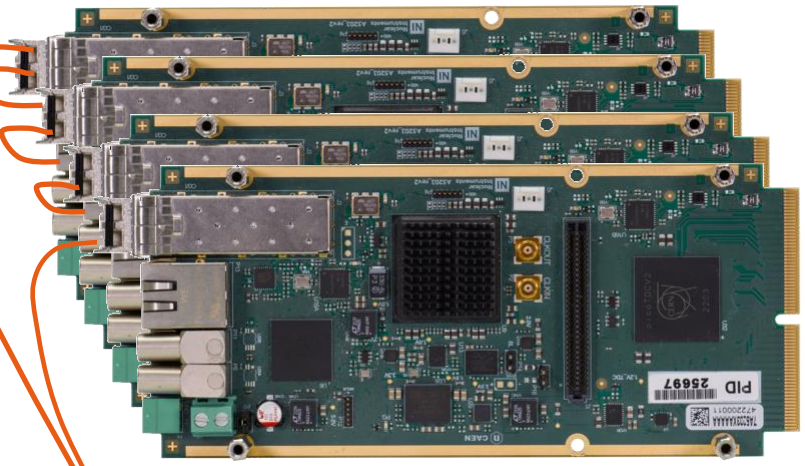


Data Readout and Multi-Unit Synchronization via TDlink



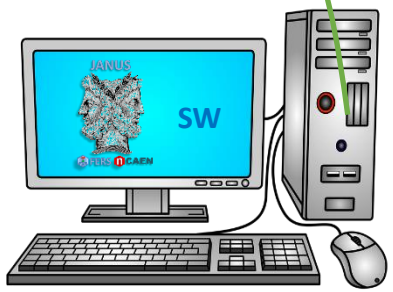
1/10Gb Eth
USB 3.0

USB 3.0

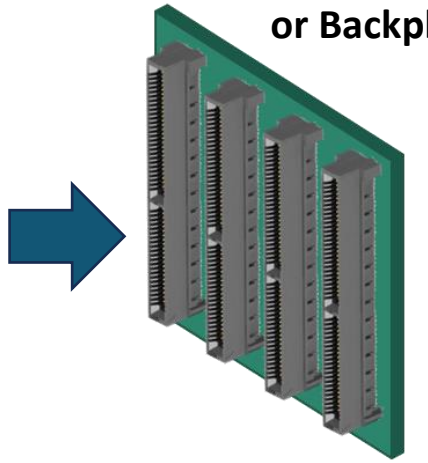


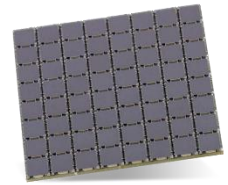
FERS Units

3.125 Gb/s TDlink



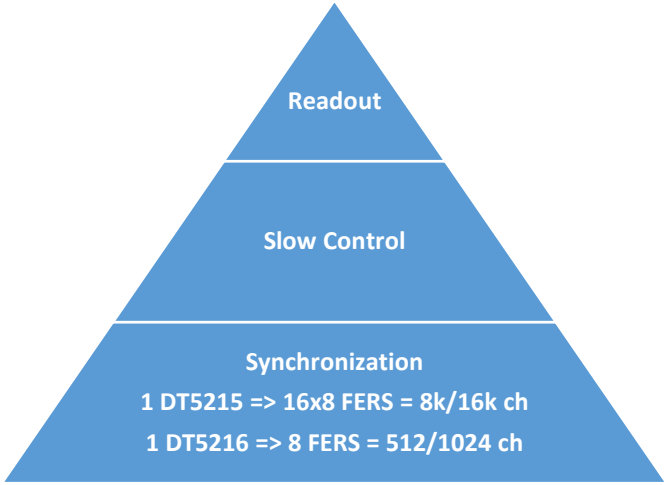
Custom Flange
or Backplane





DT5215/DT5216 Data Concentrators:

- **TDlink** (duplex optical fiber connection) proprietary protocol
- Possibility to sync **multiple DT5215**
- Sync to **GPS** on DT5215 Concentrator Board



What is TD-Link?

Protocol Overview

- Proprietary protocol for **simultaneous data transport and deterministic timing distribution** over a single ring optical fiber
- Operates at **3.125 Gb/s** line rate, carrying both control and data traffic on the same physical link

Multidrop Daisy-Chain Ring Topology

- Single fiber from **Data Concentrator** → up to **16 FERS front-end boards** → back to Concentrator (closed loop)
- Single concentrator hosts **8 independent links** → readout **hundreds of front-end modules**

Key Advantage

- Dramatically reduces cabling complexity in **large-scale distributed detector systems** where front-end electronics span extended volumes and long distances

TD-Link Synchronization & Timing Architecture

Concentrator (Master) — Clock Generation

- Low-jitter PLL locked to **10 MHz TCXO**, generating **156.25 MHz** system clocks for transceivers and FPGA fabric
- Multiple clock sources: **local oscillator**, **external digitizers**, **GPS** (PPS + NMEA) for absolute global reference
- Inter-quad phase alignment via an **innovative synchronization technique** (detailed later); elastic buffers remain **active**
- External clock sync: **DDMTD** measures phase offset, adjusts 156.25 MHz clock phase via closed-loop correction

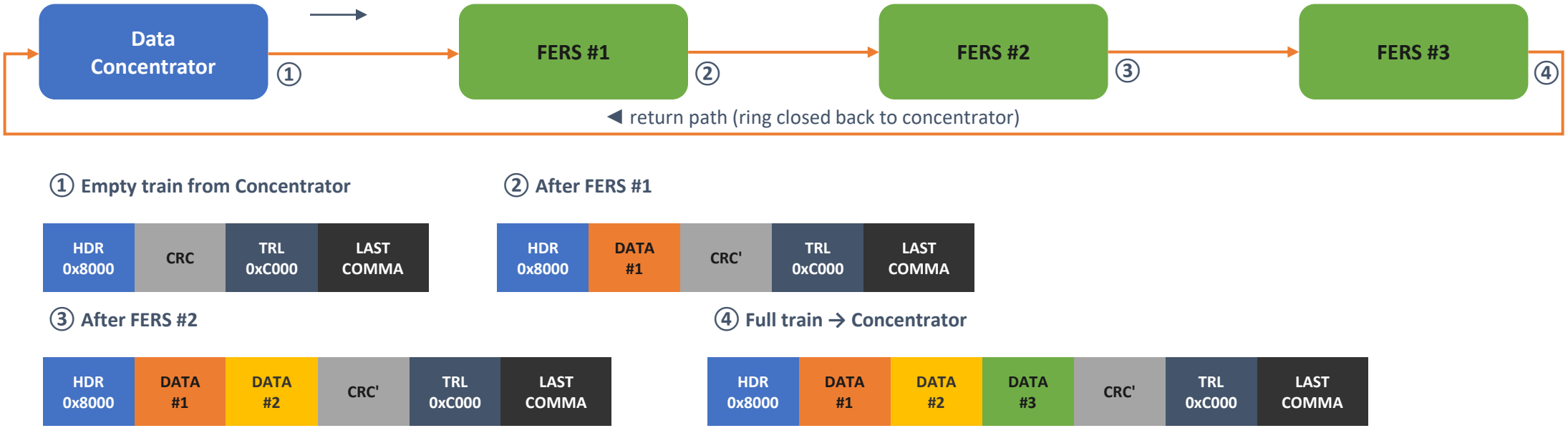
FERS Front-End (Slave) — Clock Recovery

- Elastic buffers **disabled** to enforce deterministic latency
- RX clock recovered from transceiver, cleaned by **external jitter cleaner PLL**
- Cleaned clock, phase-shifted, used as **TX clock** — strict RX-to-TX phase lock along the daisy chain

T0 Global Synchronization

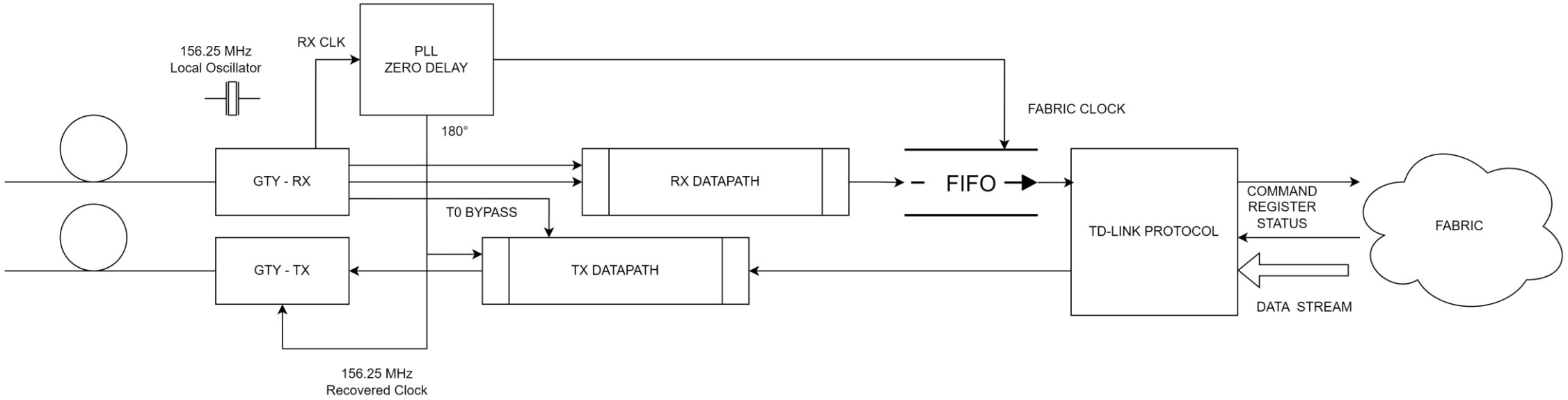
- Round-trip delay measurement → per-board correction → **deterministic global time reset** across all nodes
- Validated: **~25 ps RMS** inter-board synchronization, **stable across power cycles**

TD-Link Data Transport: Token-Based Train Protocol



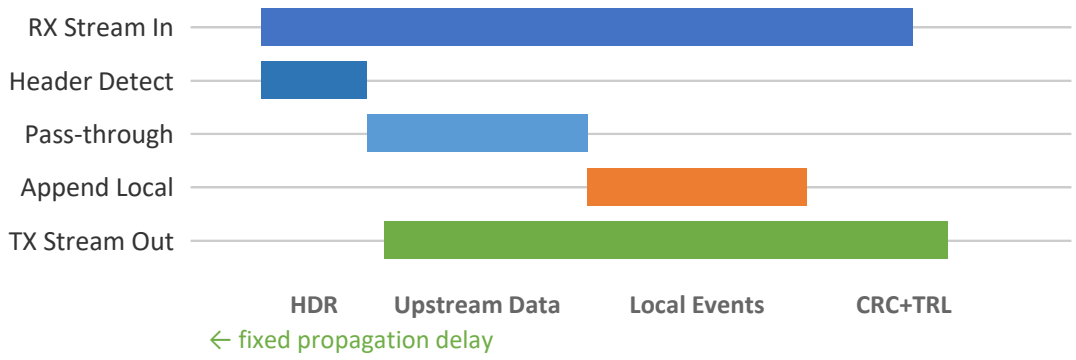
- Train is **never stopped** inside a node — header is forwarded immediately while data is appended in streaming
- Each FERS board appends up to **N events** per train; larger payloads are fragmented across multiple trains
- Max train size: **64 Mword per link** (concentrator DDR4 memory); CRC recalculated at each hop

FERS Front-End Datapath Architecture

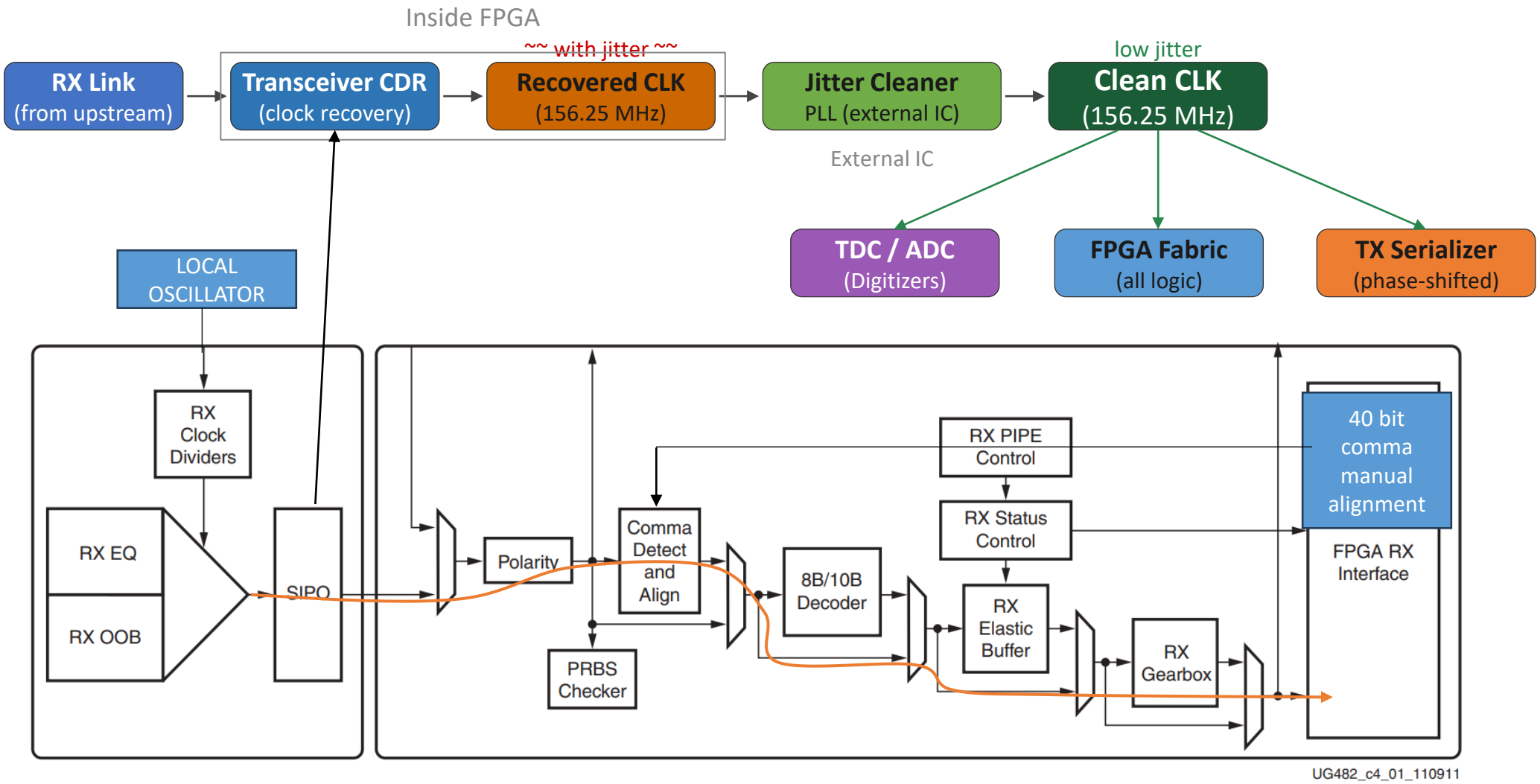


- RX deserializer recovers data **and clock** from the upstream link
- Recovered clock cleaned by **jitter cleaner PLL**, phase-shifted, used as TX clock
- Elastic buffers **disabled** → deterministic, fixed latency through each node
- Protocol FSM decodes header, forwards train, appends **local events** before trailer
- Digitizer data (ADC/TDC) buffered in local FIFO, drained into passing train

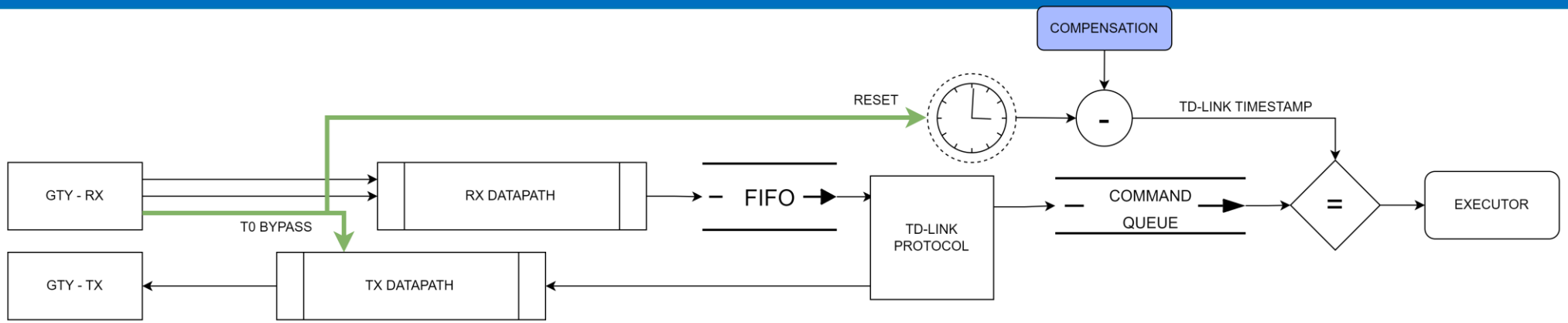
Timing: Data Flow Through Node

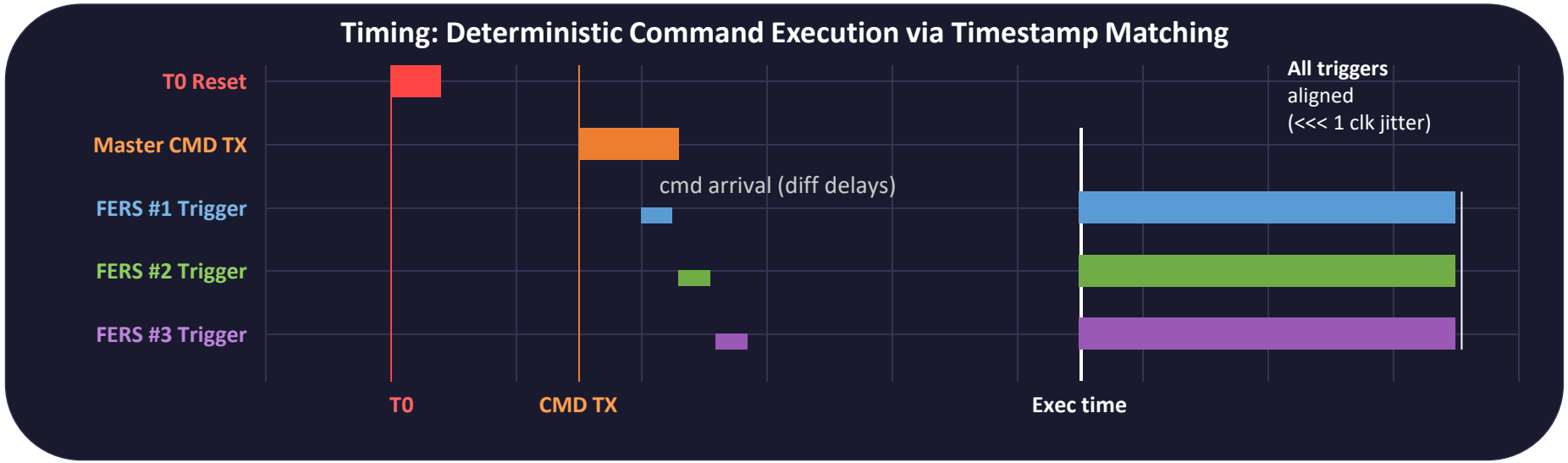


FERS Slave Clock Recovery and Distribution



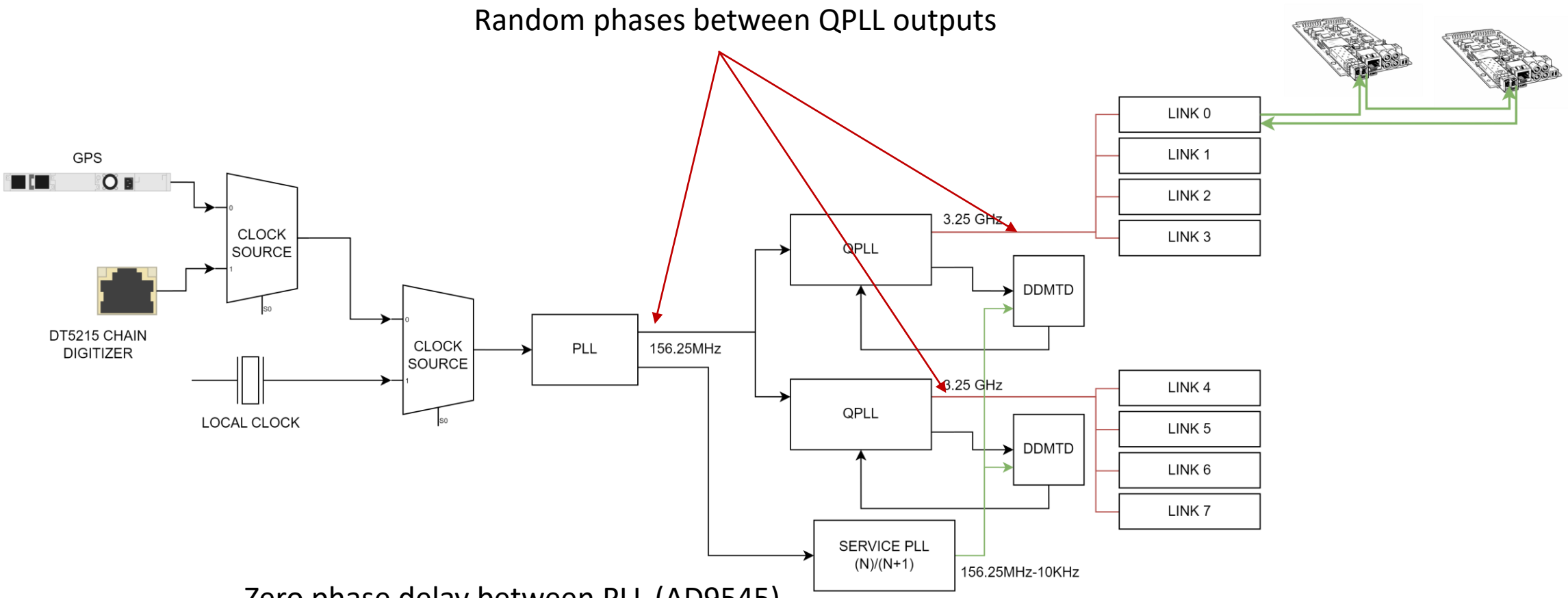
T0 Synchronization & Deterministic Command Execution





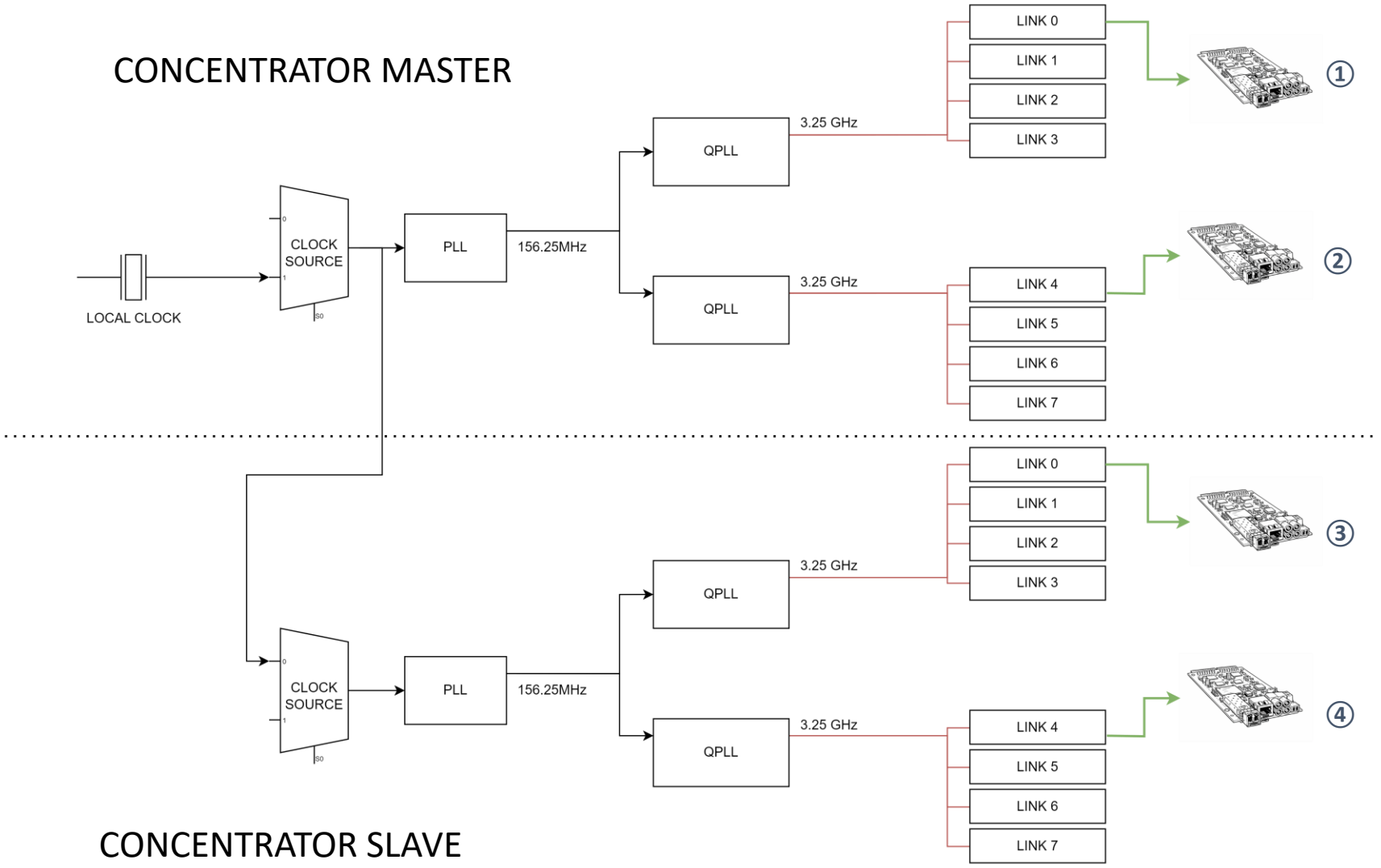
- T0 (K-word 0xFEFE) propagates atomically RX→TX (no FIFO), resets all timestamp counters to zero
- Commands carry a future execution timestamp; each FERS corrects it by CMD_ChainTimeCompensation (derived from RTT)
- Execution fires when local counter matches CMD timestamp → all boards trigger simultaneously regardless of chain position

Multi link and multi-concentrator synchronization

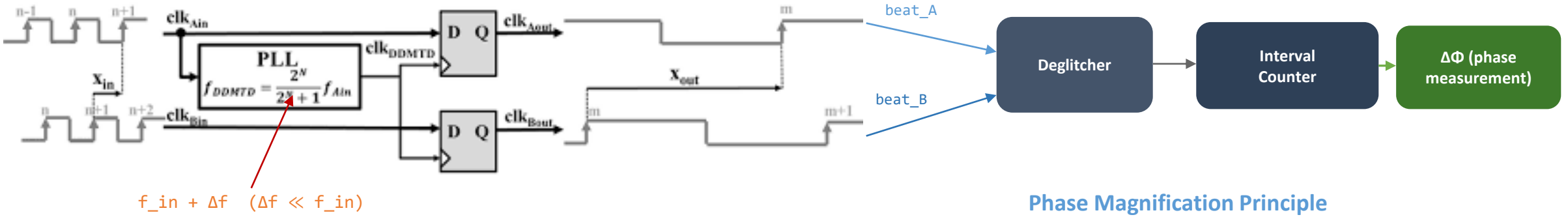


Zero phase delay between PLL (AD9545) Input and output

Multi link and multi-concentrator synchronization



DDMTD Phase Detector: Sub-ps Clock Phase Measurement



Beat frequency: f_beat = f_in − f_DDMTD = Δf
Phase resolution: T_in / (f_in / Δf) = T_in × (Δf / f_in)
e.g. f_in=156.25 MHz, Δf=15.625 kHz → N=10000× magnification → ~0.6 ps resolution

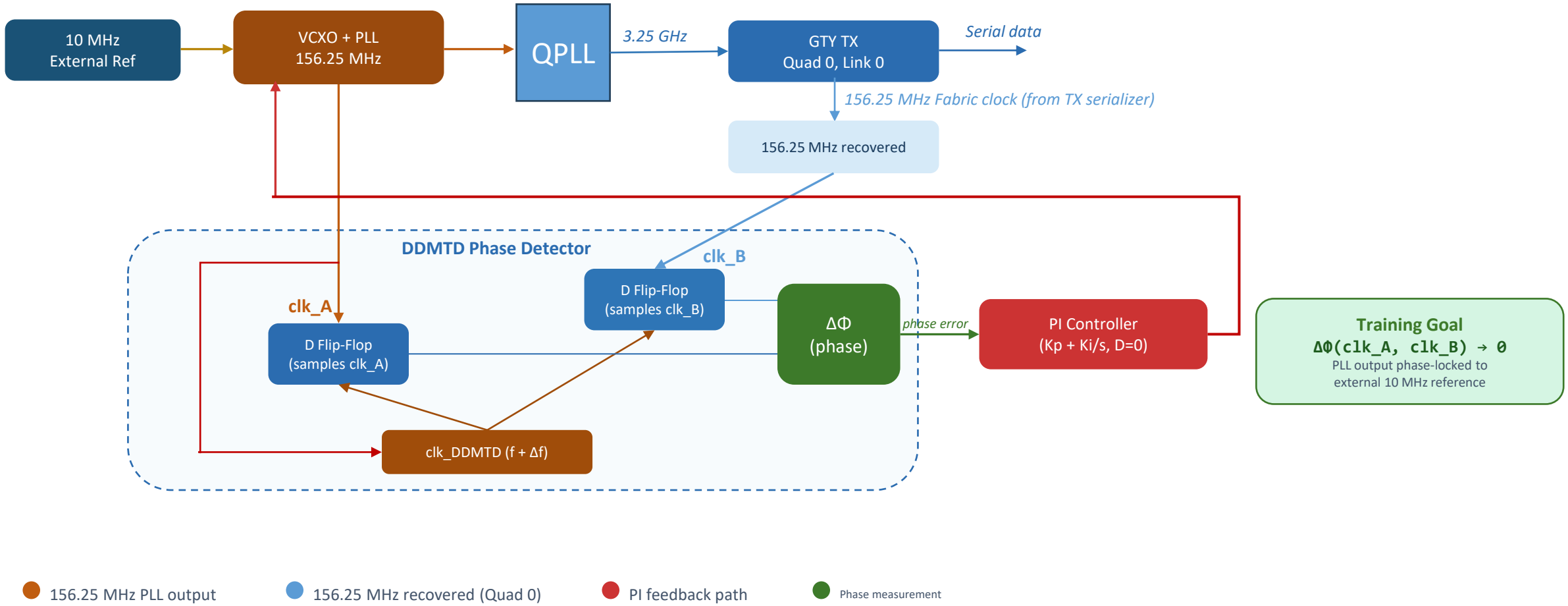
Phase Magnification Principle

↑ Slow beat period = f_in/Δf × T_in (time-magnified phase info)

How It Works

- Two input clocks (clk_A, clk_B) are sampled by D flip-flops clocked with **clk_DDMTD**
- clk_DDMTD has frequency very close to clk_A/B but offset by a small Δf
- Each FF output (beat) is a **slow square wave** at frequency Δf — phase info is **time-magnified**
- Time difference between beat_A and beat_B rising edges = ΔΦ × (f_in / Δf)
- Deglitcher removes metastable glitches at beat transitions
- A simple counter measures the interval → sub-picosecond resolution

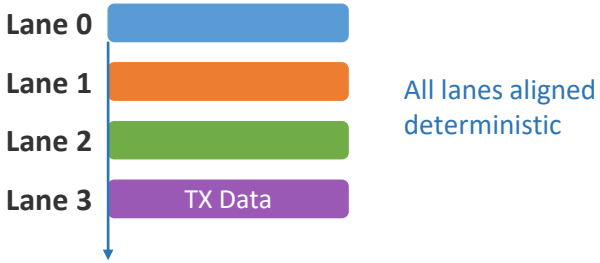
DDMTD Training Loop: PI Controller Phase-Locks PLL to Reference



TX Buffer & Deterministic-Latency Link Design

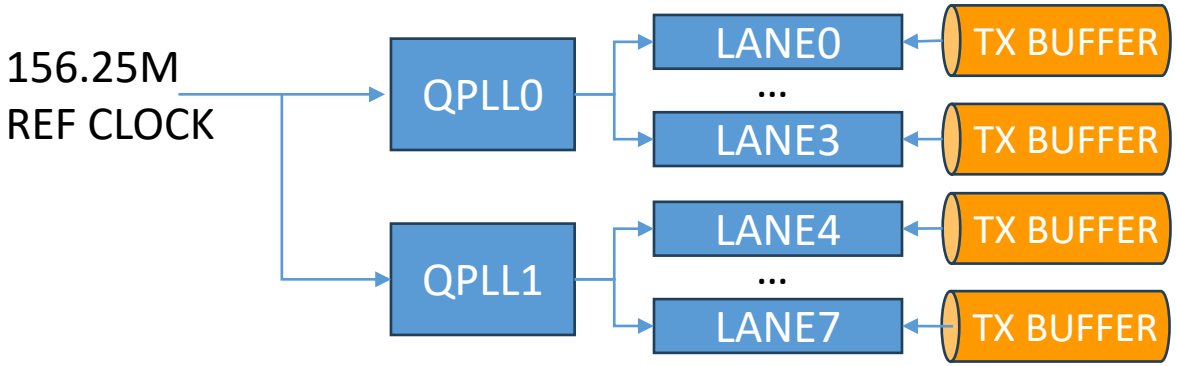
Start-up — Random Phase:

With TX Phase Aligner — Aligned:



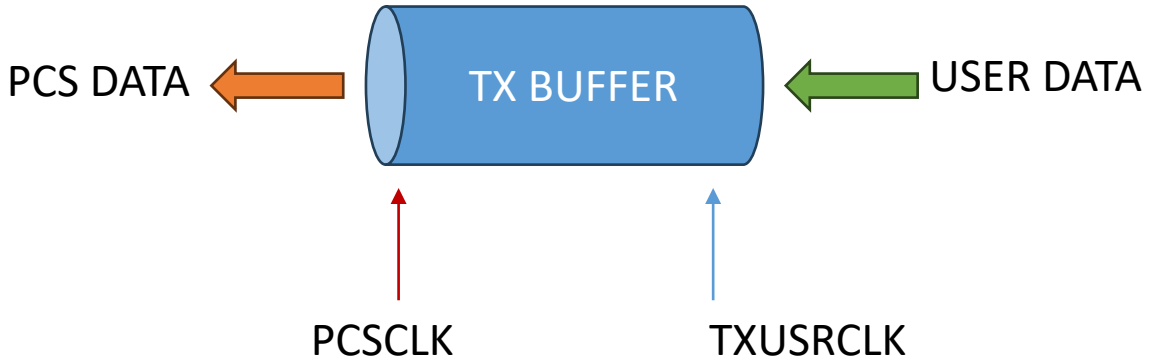
What is the TX Elastic Buffer?

- A **FIFO between two clock domains**: TXUSRCLK (fabric) writes, XCLK (PMA serializer) reads
- Safely handles **CDC (Clock Domain Crossing)** between mesochronous clocks Default mode (enabled) adds **variable latency** (FIFO fill level changes at each reset)
- Bypassing it removes latency variation but **breaks CDC safety** and creates the dual-PI conflict

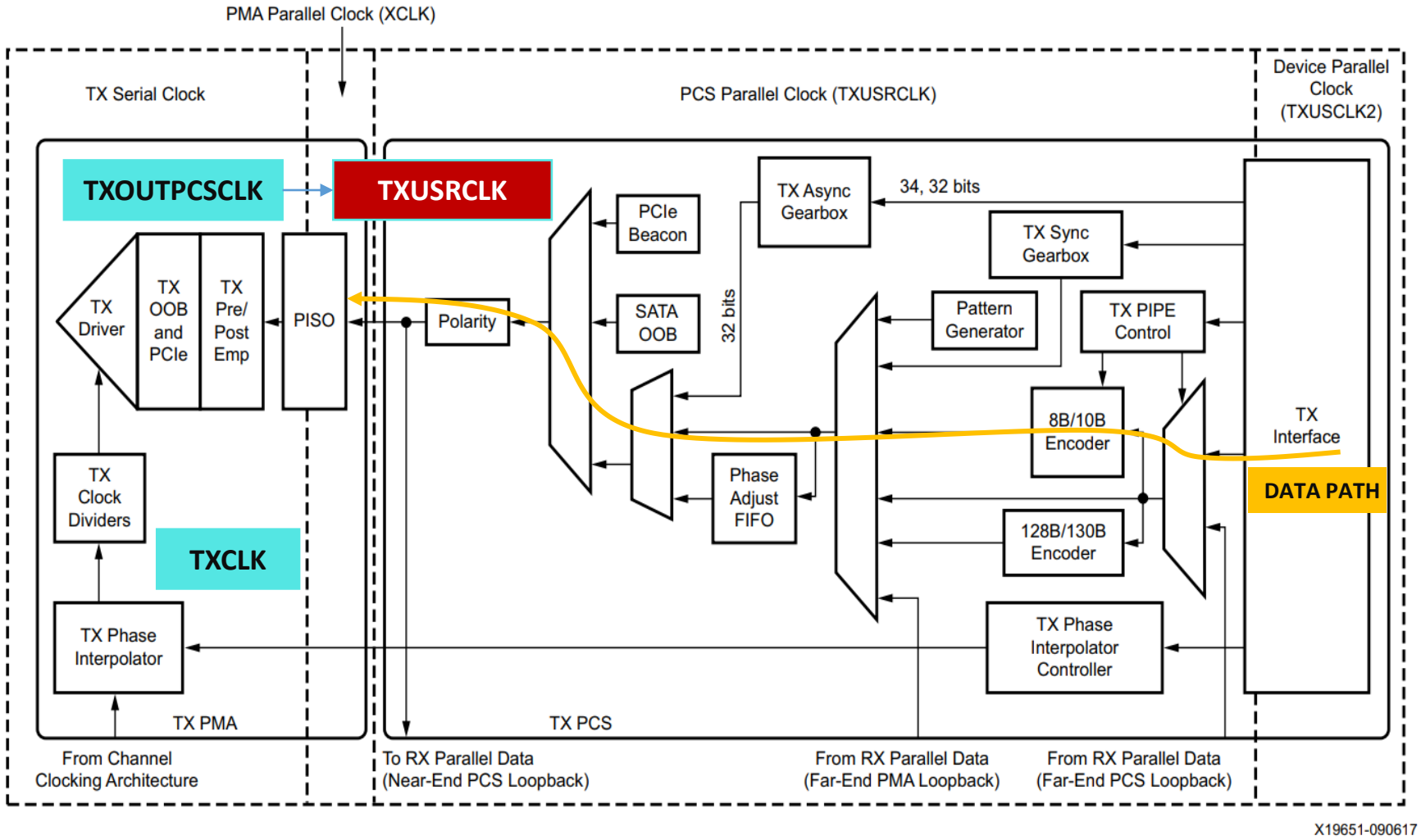


Why Deterministic Latency Matters

- TDlink synchronizes multiple FERS units — each link must add a **known, fixed latency**
- Random FIFO fill → RTT measurement changes per reset → timestamp compensation breaks
- T0 propagation and command scheduling depend on stable, repeatable link latency



TX Phase Interpolator — Buffer Bypass



Buffer Bypass Mode

Elastic buffer (FIFO) is bypassed. The PCS logic runs directly from TXUSRCLK.

The CDC Problem

Two clock domains exist in the TX path:

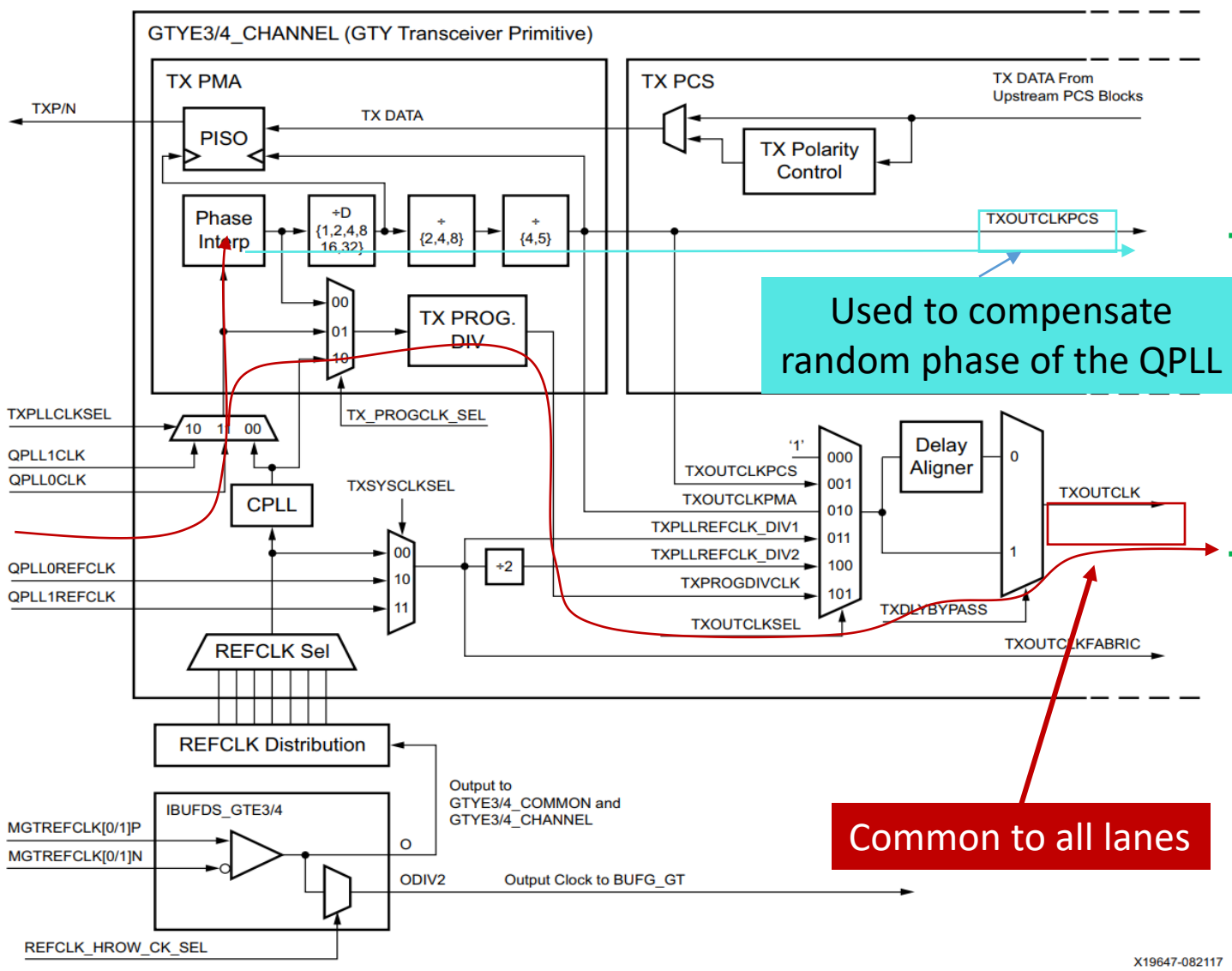
- **TXUSRCLK** (fabric BUFG, user logic)
- **TXOUTPCSCLK** (from PISO divider of TXCLK, PMA-side)

These are **mesochronous**: same frequency but unknown phase offset after each reset.

Phase Interpolator Role

Without the FIFO, there is **no hardware CDC**. The TX Phase Interpolator shifts TXUSRCLK in fine steps until the phase offset between the two domains is within the safe setup/hold window of the PISO flip-flops.

TX Phase Interpolator — Multi Lane synchronization



Used to compensate random phase of the QPLL

Common to all lanes

TXOUTCLKPCS (phase shifted clock) ≠ **TXUSRCLK2** (fabric data clock) — they are mesochronous: same frequency, mutually random phase after each reset.

$\varphi_o \neq 0$ and different for every lane

What is the Phase Interpolator (PI)?

The TX Phase Interpolator is a **mixed-signal circuit inside each GTP transceiver** that can shift the phase of the internal PMA clock (TXOUTCLKPCS) in fine steps.

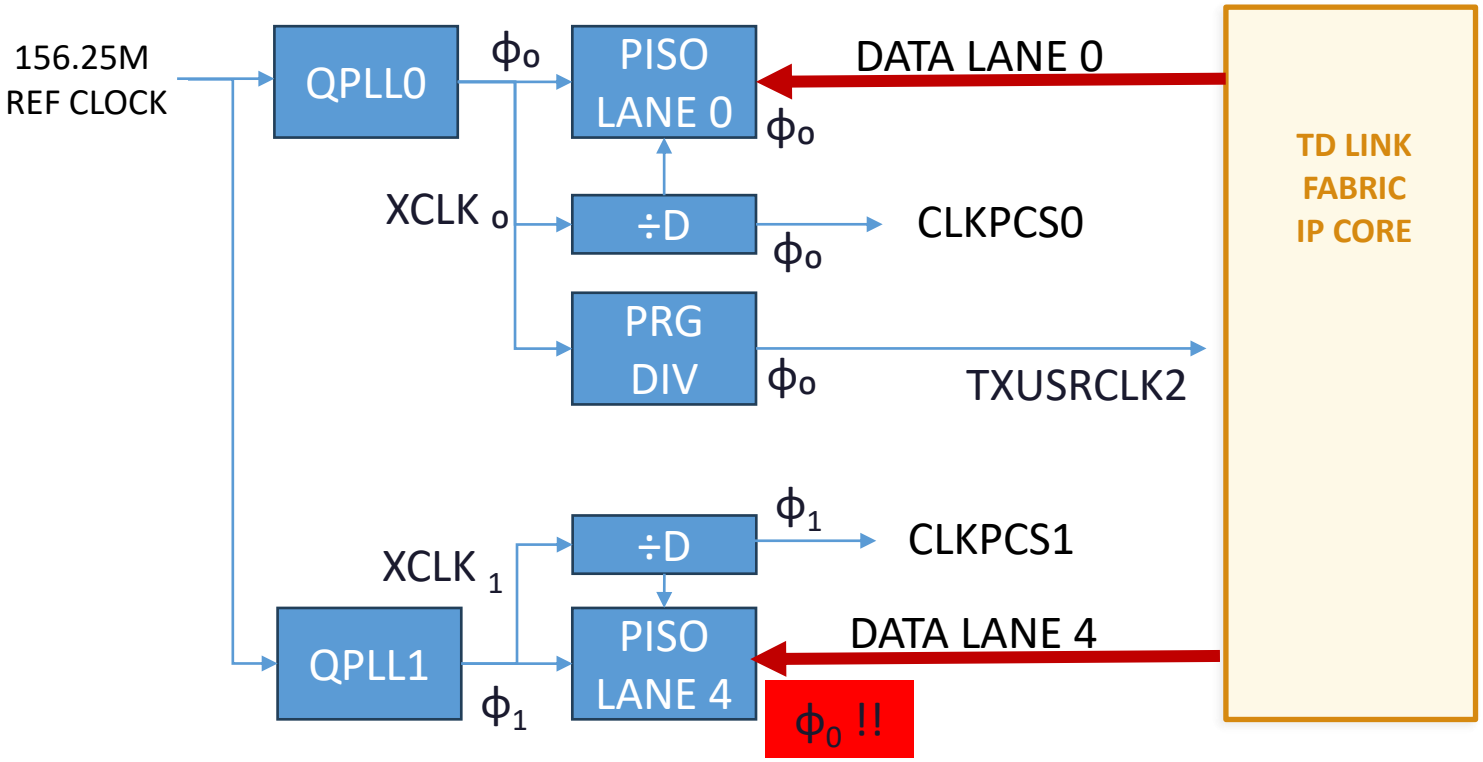
How it works

- The PMA serializer runs on its own clock domain (TXOUTCLKPCS), derived from the PLL
- Resolution: typically **~18.6 ps per step**

Why it matters for TDlink

- Without alignment, the serializer clock phase ϕ_o is **random after each reset** and different per lane
- The multi-lane PI alignment forces all lanes to the **same phase**, eliminating lane-to-lane skew

Buffer Bypass — TX PI Dual-Purpose Conflict



The TX Phase Interpolator has two jobs:

Job 1 — CDC Phase Margin

Shift TXUSRCLK vs XCLK (PISO clock) to guarantee setup/hold for the data→serializer clock domain crossing.

Job 2 — Inter-Lane TX Alignment

Shift each TX serializer clock so the 156.25 MHz recovered at the FERS receivers are phase-aligned across all lanes.

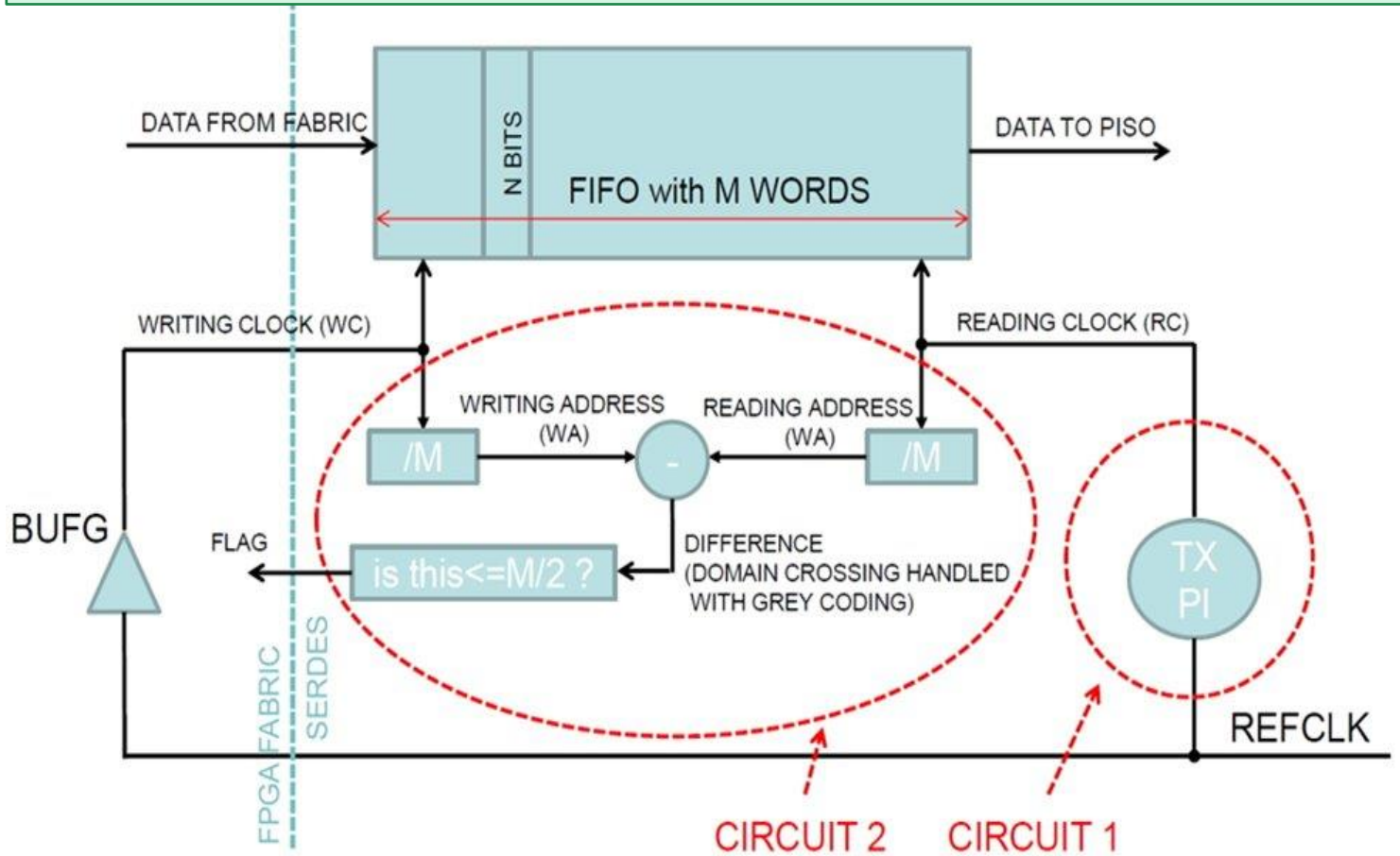
The Conflict:

Each QPLL starts with a **random phase** ($\phi_o \neq \phi_1$ after reset).

If you use the PI to align ϕ_o to ϕ_1 (Job 2), the resulting $\Delta\phi$ on **Lane 4** violates the setup/hold timing between **$XCLK\_PCS_1$** and **$TXUSRCLK_2$** — the PI cannot satisfy both constraints simultaneously.

Deterministic TX Buffer

Solution: keep the elastic buffer enabled. The FIFO handles CDC (Job 1) for each lane independently. The TX PI is then free to perform Job 2 — shift each serializer clock to align recovered 156.25 MHz across all lanes.



Circuit 1 — TX PI

Shifts REFCLK phase before the CPLL/QPLL. Each step moves the XCLK (FIFO read clock) by ≈1.5 ps → 18.6 ps @ 156.25 MHz. Total range ≈ 1 REFCLK period (6.4 ns at 156.25 MHz).

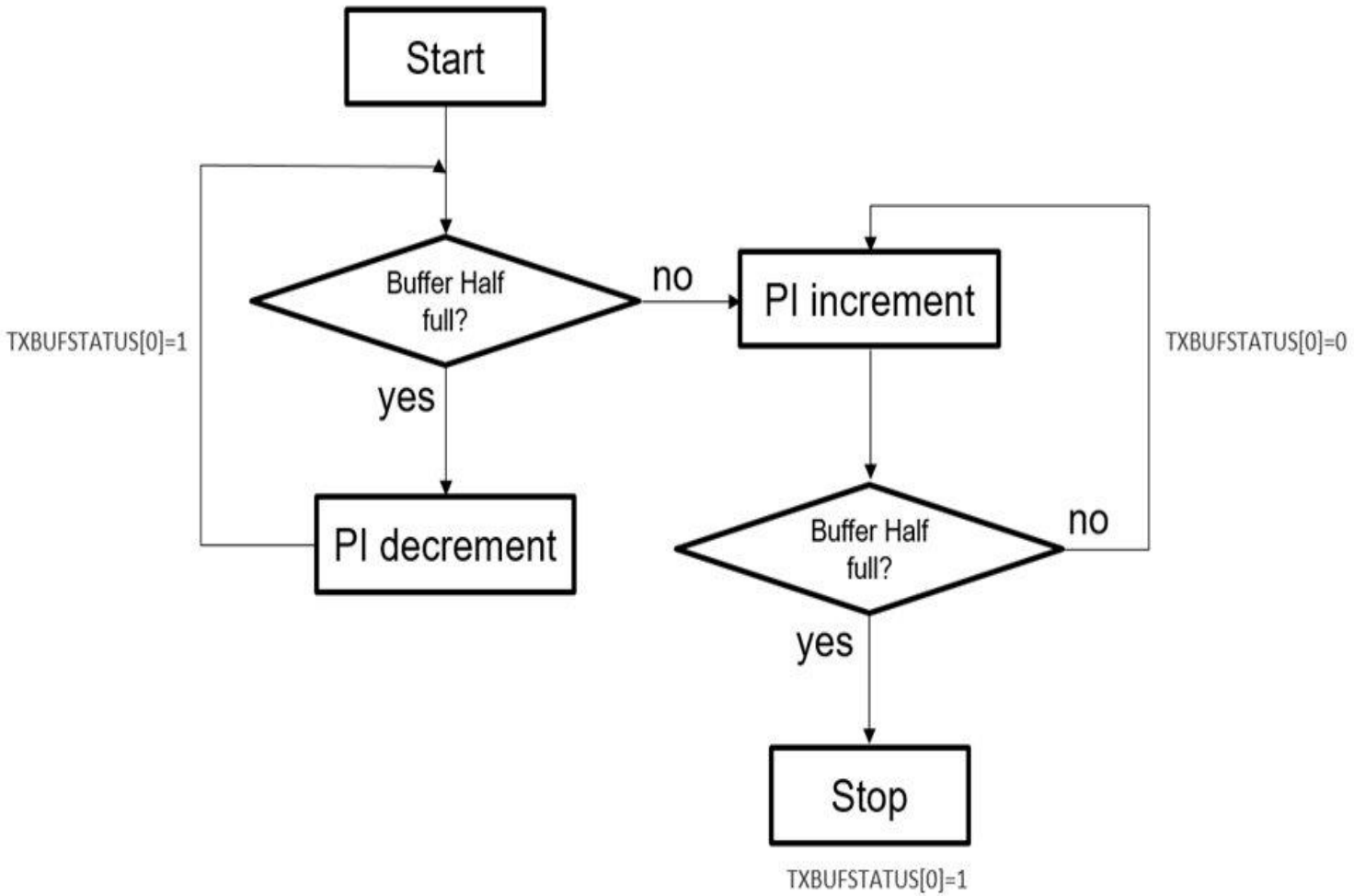
Circuit 2 — Fill level feedback

The FIFO maintains WA and RA pointers (both divided by M for comparison). The difference WA−RA (using Gray coding for CDC) is compared to M/2:

`txbufstatus[0] = 1` → FIFO ≥ half-full
`txbufstatus[0] = 0` → FIFO < half-full

This flag is the **1-bit phase detector** driving the alignment FSM.

Deterministic TX Buffer



Phase 1 — Initial Sweep (find edge)

On start: if buffer is already half-full (**`i_half_full=1`**), decrement PI step-by-step to exit the half-full zone. Then increment PI until the buffer becomes half-full again (0→1 transition).

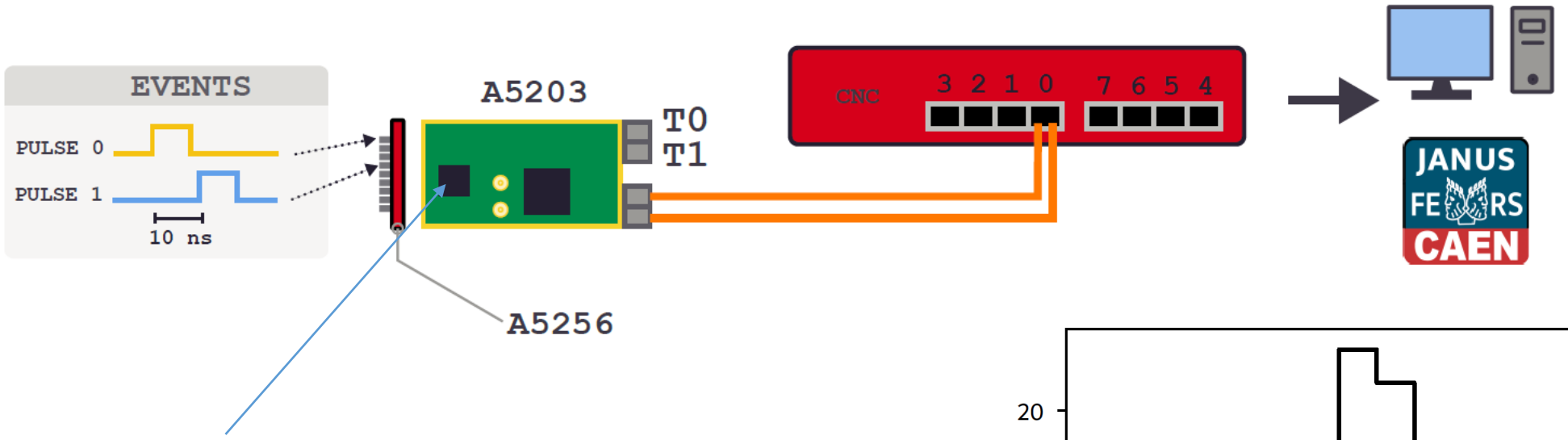
Phase 2 — Lock on Transition (0→1)

The 0→1 transition of **`txbufstatus[0]`** marks the half-fill point. The FSM stops: the PI is locked on this code. The phase between WR and RD pointers is now **deterministic**.

Result

Repeatable inter-lane skew after each reset ≈ few PI steps (**≈10 ps**). PI is fixed post-alignment → no additional jitter.

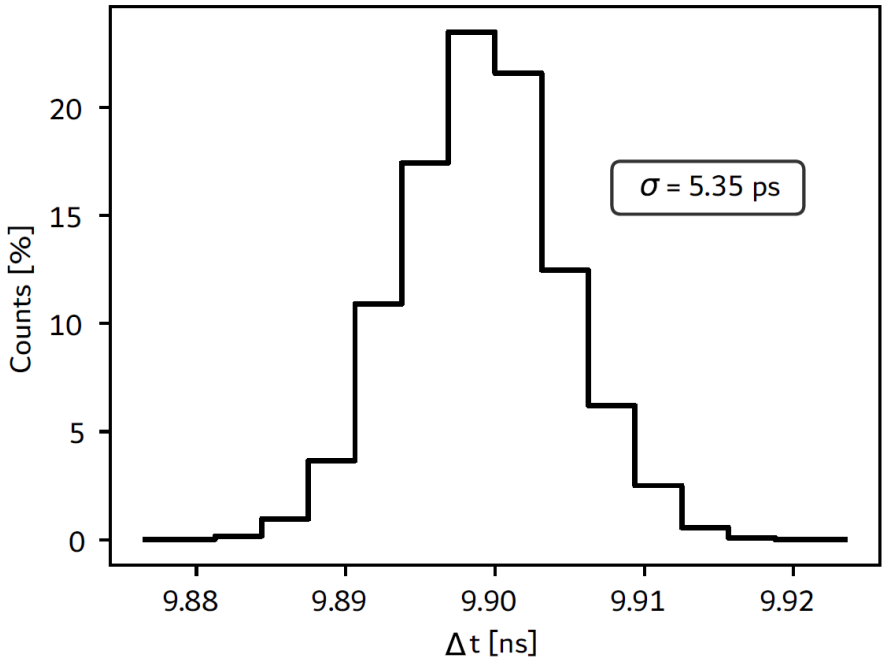
Link timing jitter measurement: TDC Performance



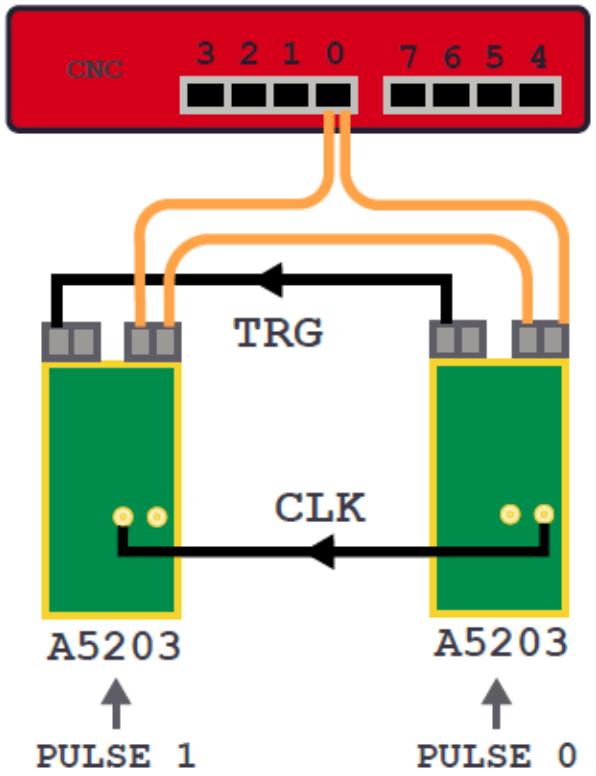
CERN PicoTDC

$$t_i = N_{\text{CLK}}\, t_{\text{timestamp},i} + t_{\text{ToA},i}$$

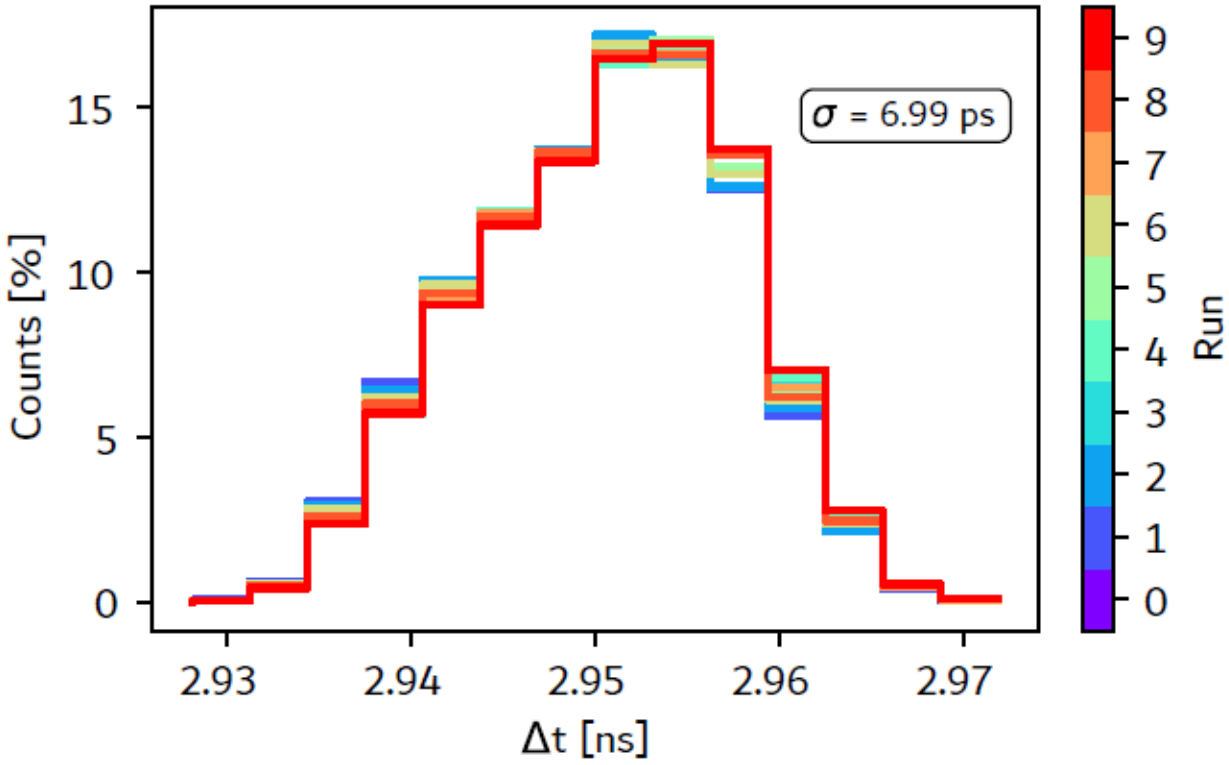
$$\Delta t_i = t_{\text{CH3},i} - t_{\text{CH0}.i}$$



Link timing jitter measurement: Clock distributed via cable



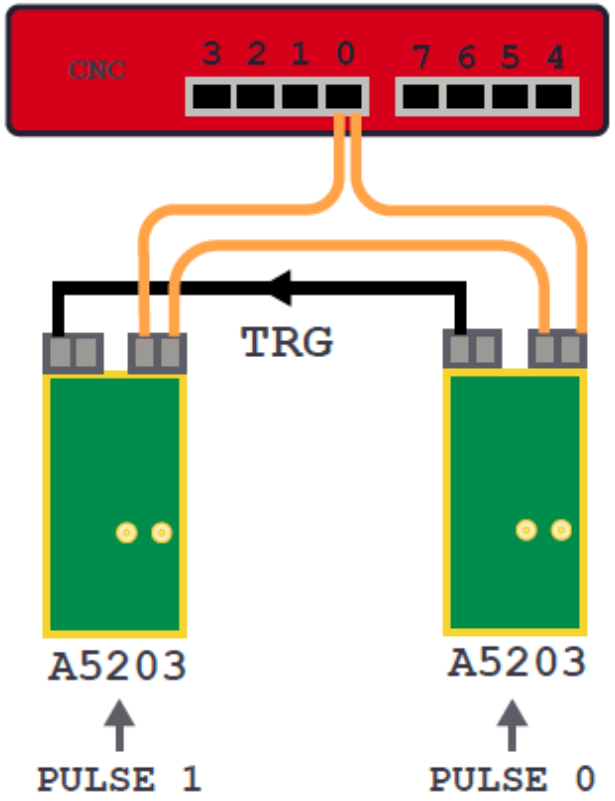
(a) Board connection schematic (w/ HRES CLK)

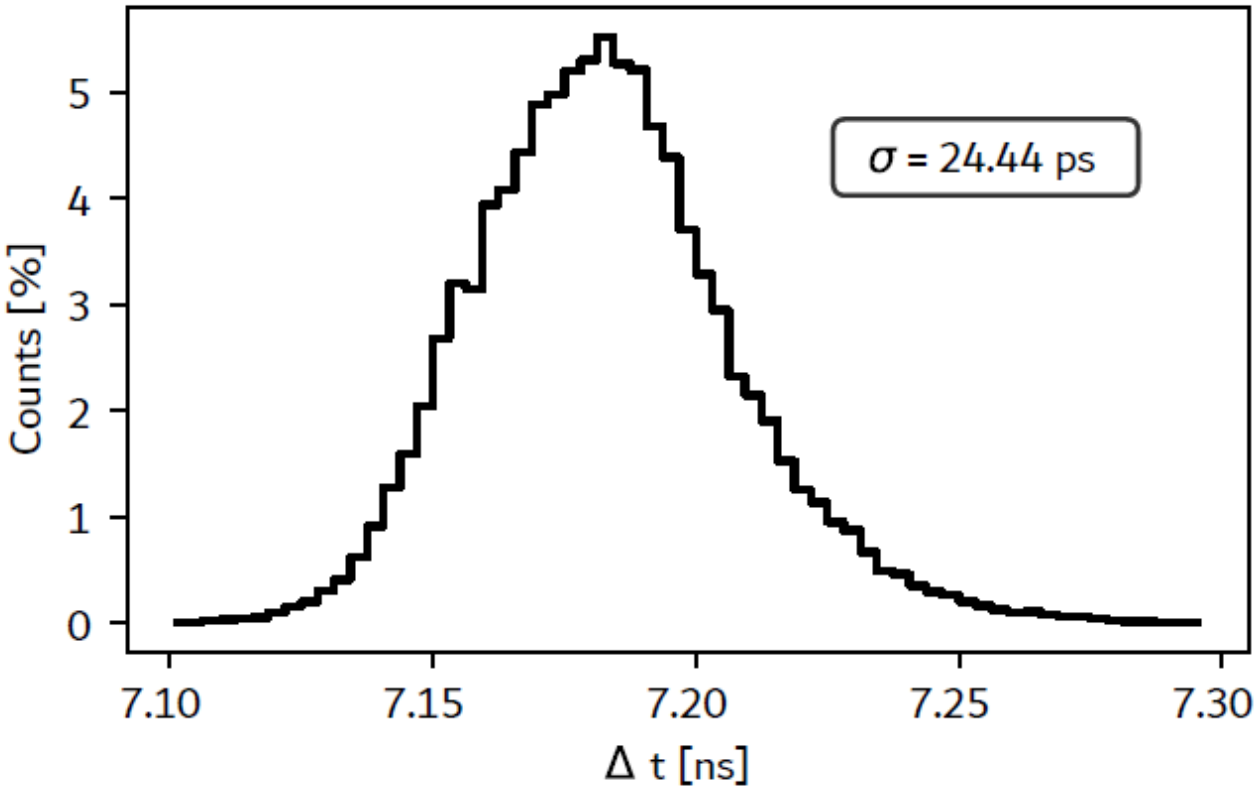


(b) Time difference distribution (w/ HRES CLK)

Link timing jitter measurement: Clock distributed via TD-LINK

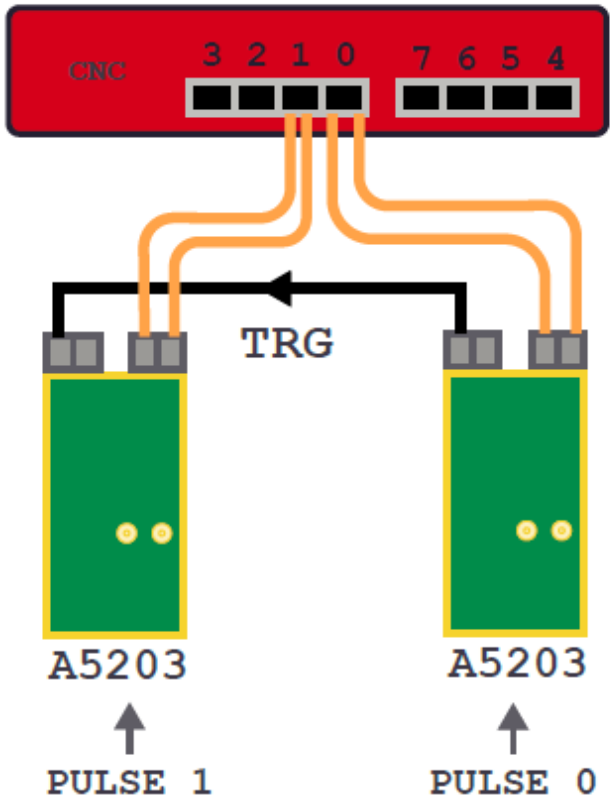
Same Link – Daisy Chain Board

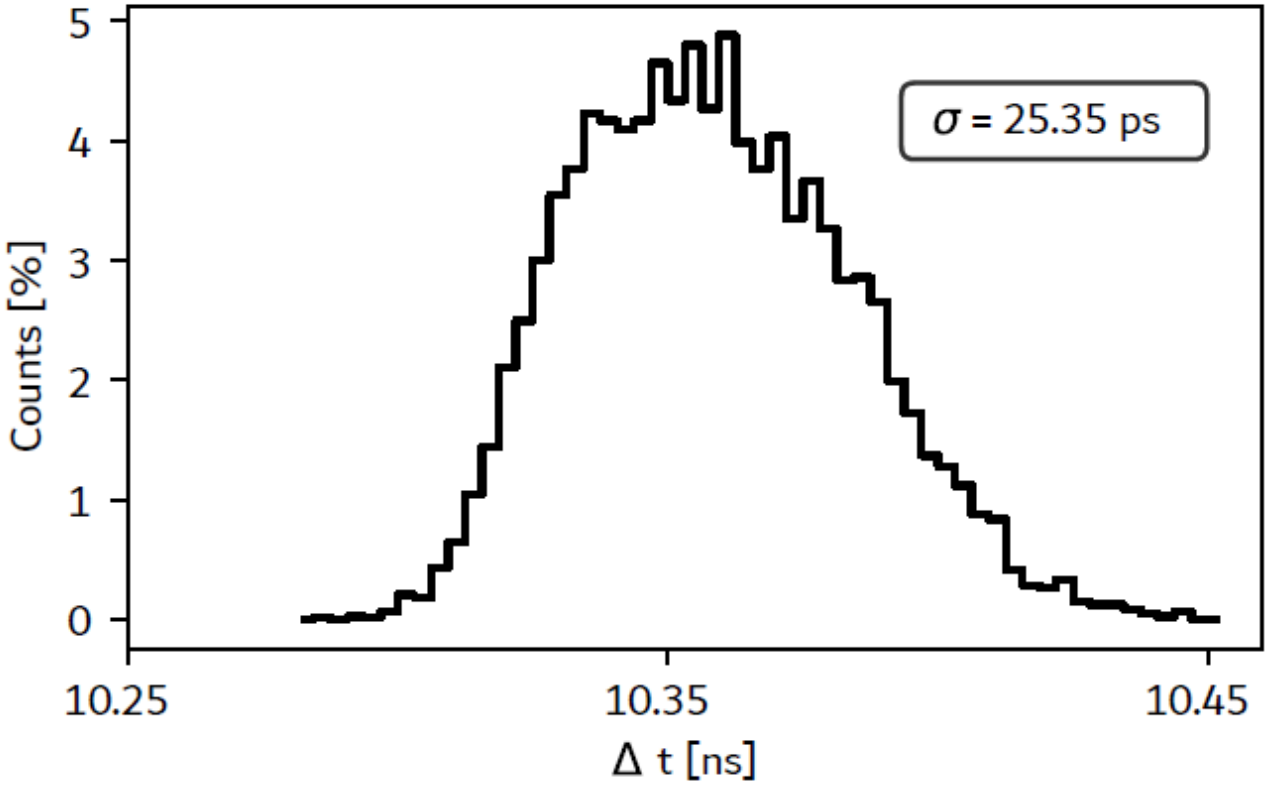




Link timing jitter measurement: Clock distributed via TD-LINK

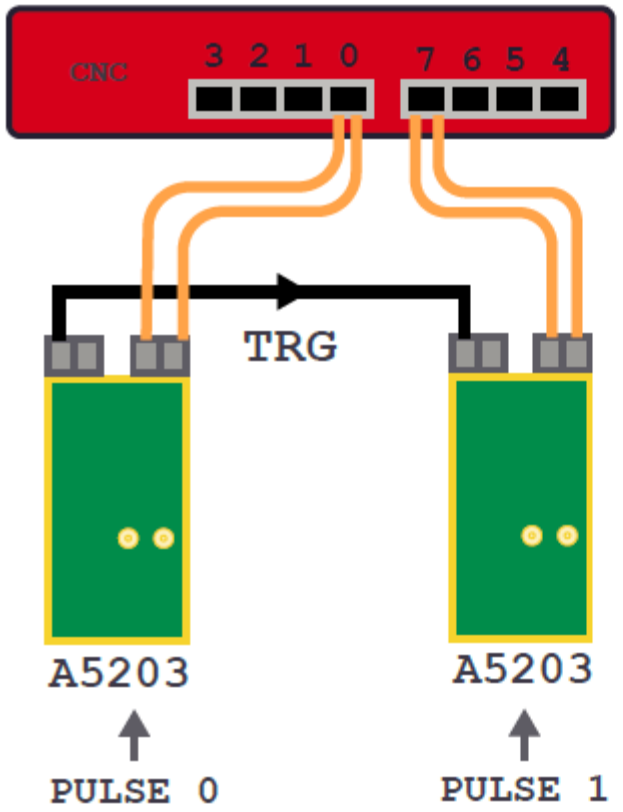
Link 0 and 1 – Same Quad

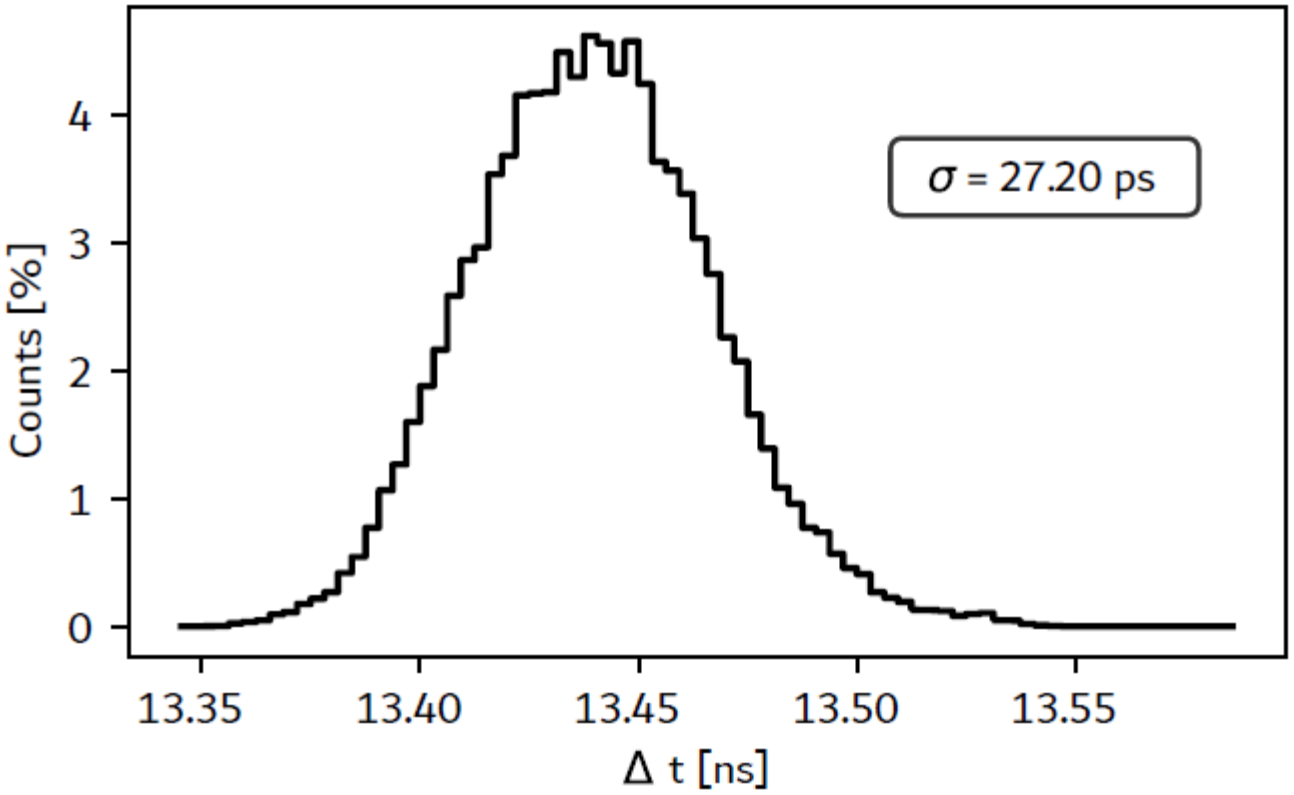




Link timing jitter measurement: Clock distributed via TD-LINK

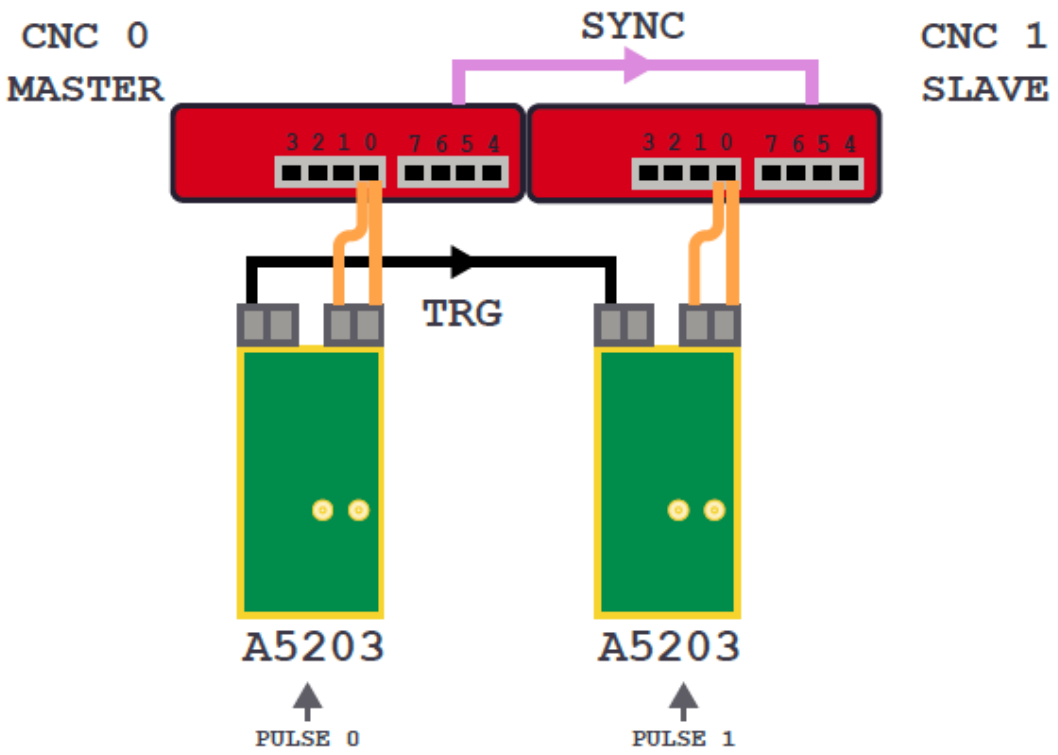
Link 0 and 7 – Quad 0 and 1, same data concetrator

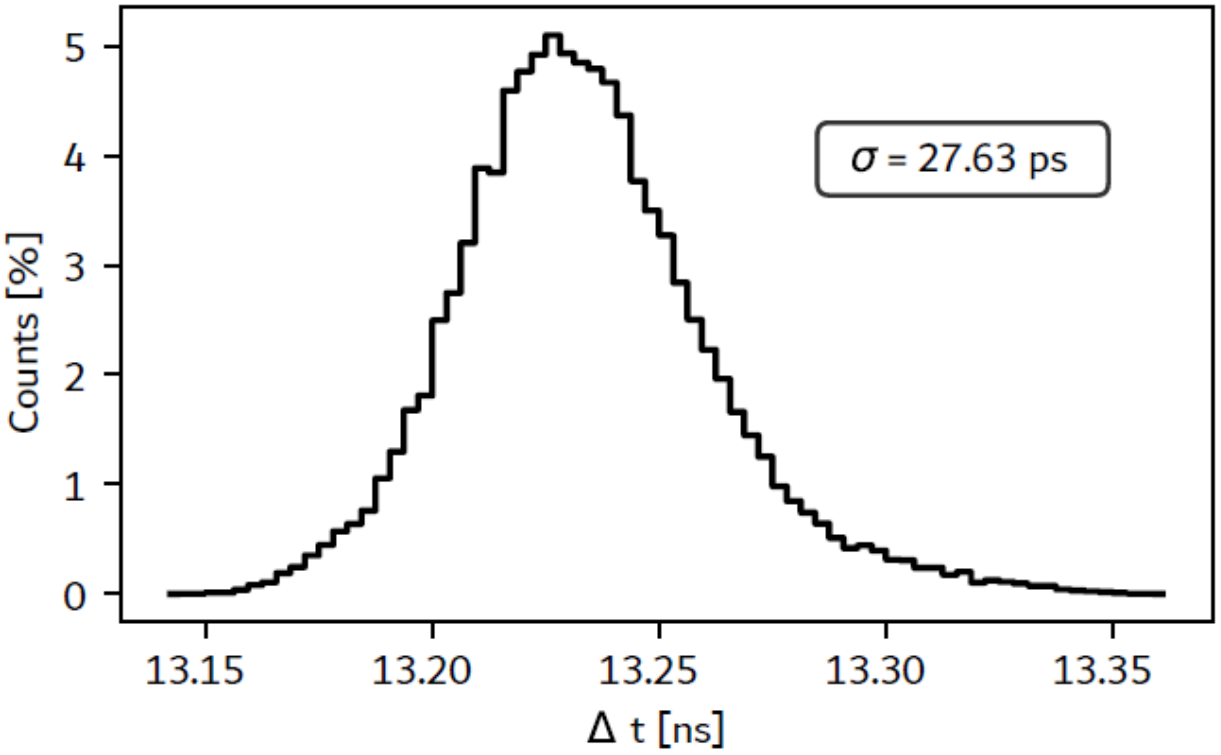




Link timing jitter measurement: Clock distributed via TD-LINK

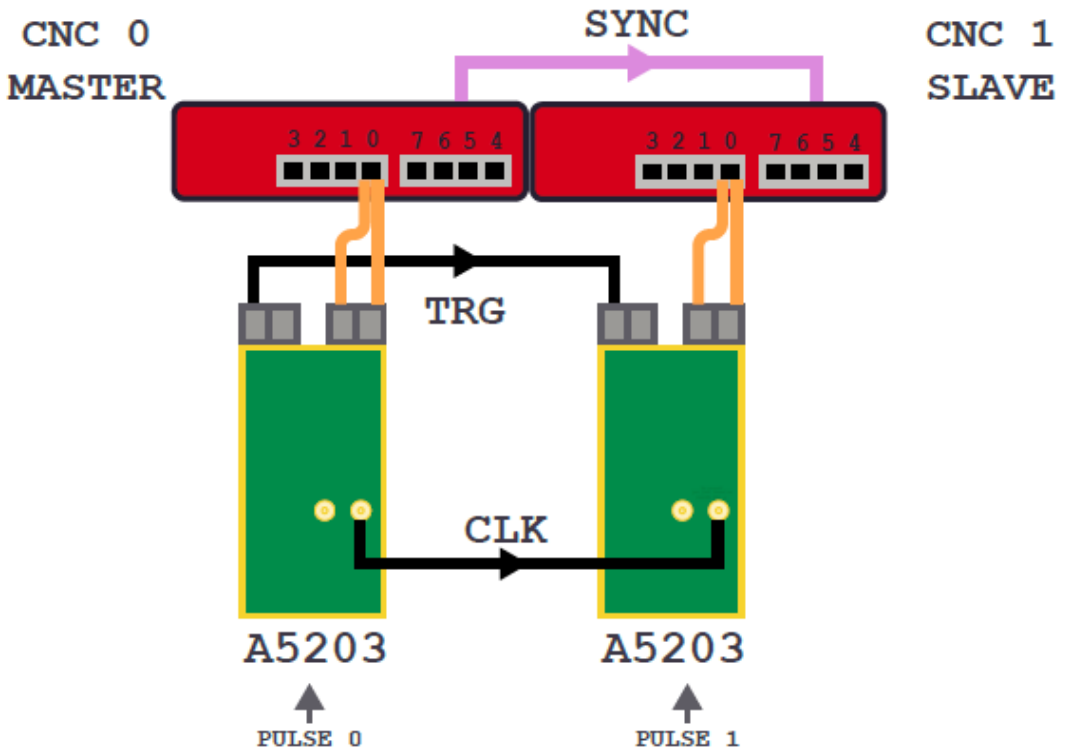
Link 0 and 0 – Two different data concentrators

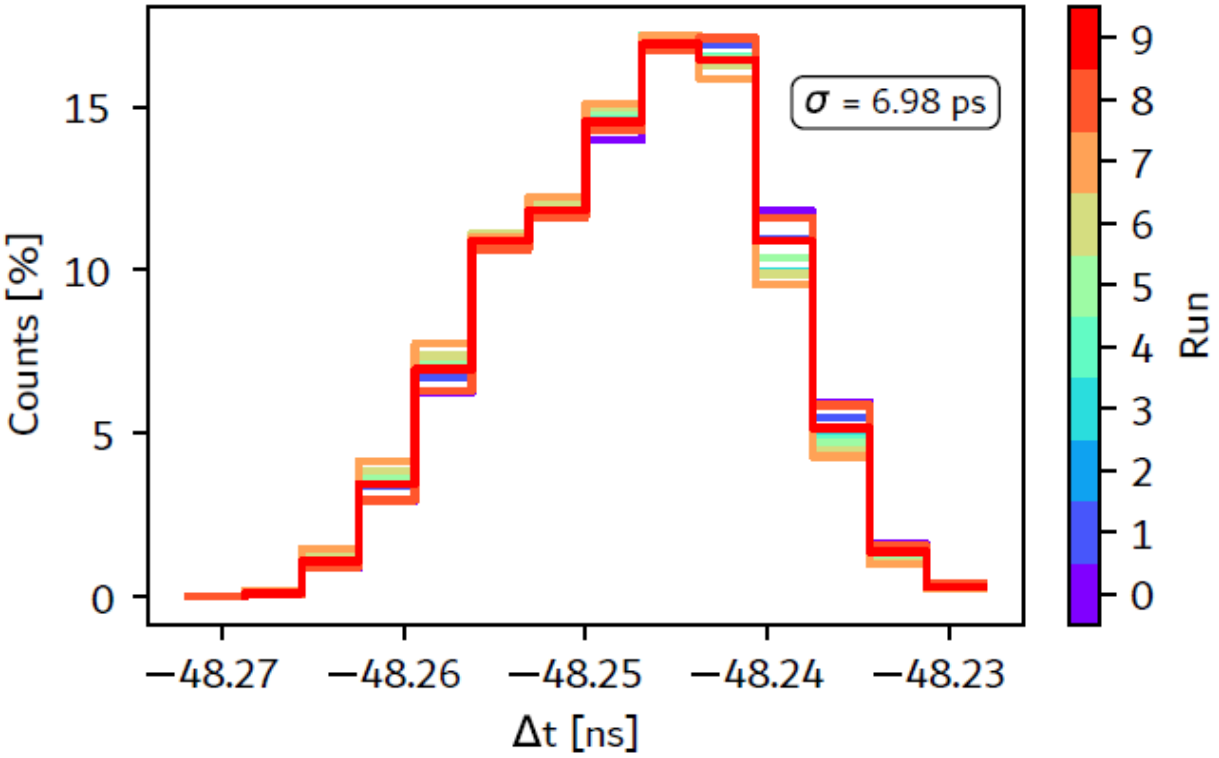




Link timing jitter measurement: Clock distributed via cable

Link 0 and 0 – Two different data concentrators





Link timing jitter measurement: Summary

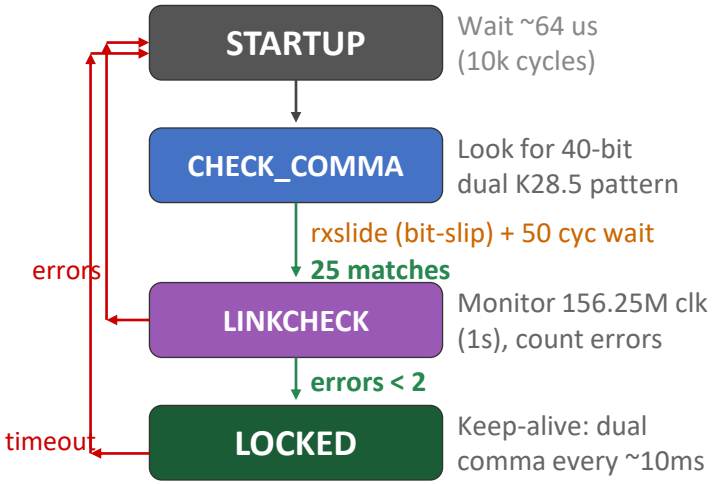
Setup	Resolution [ps] (w/o HRES)	Resolution [ps] (w HRES)
Same board	5.3	-
Same link	≈24	7.0
Same QUAD	≈25	6.7
Same CNC	≈27	7.0
Different CNCs	≈28	7.0

Feature	TD-Link	White Rabbit
Timing precision	~25–30 ps RMS	1ns/100 ps
Clock distribution	Embedded in serial data	SyncE + PTP (IEEE 1588)
Topology	**Daisy-chain ring**	Star / tree (switches)
Link speed	3.125 Gb/s per ring	1 Gb/s (GbE)
Max nodes / link	16 per ring unlimited number of rings	~1000 (network)
Max distance	SFP 300 m per hop	Up to 10 km fiber (SMF)
Data + clock on same link	Yes	Yes
Dedicated switch needed	No	Yes (WR Switch)
Target application	Detector front-end DAQ	General lab / accelerator timing

BACK-UP

Physical Layer: CDR, 40-bit Comma Alignment & Link Lock

CDR State Machine



40-bit Comma Alignment

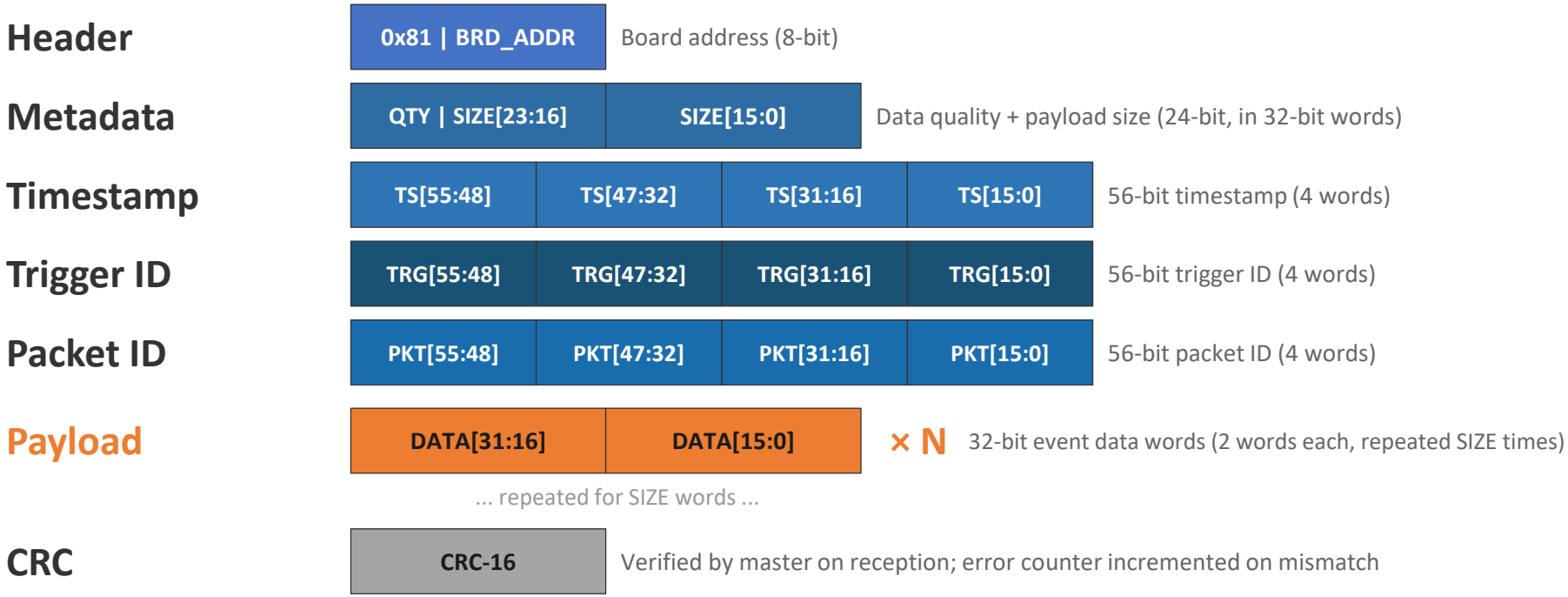
- GTP transceiver outputs **20-bit raw** data per clock (no internal 8b/10b decoder)
- CDR checks **current + previous** 20-bit words (40 bits total) for a dual K28.5 comma pair
- A single K28.5 (20-bit) has a symmetric 10-bit pattern → shifted by 10 bits it can still match
- Two consecutive K28.5 with **forced alternating disparity** (RD+, RD−) break the symmetry
- Master forces **disparity control**: commas are sent with a deterministic RD+/RD− sequence
- Searching **before 8b/10b decoding** on the raw 10b stream: only one alignment out of 20 matches → no 180° phase ambiguity

Expected raw 40-bit pattern:

Word [N-1] (20 bits)	Word [N] (20 bits)
`10100000 11100 11110 0` K28.5 (RD+)	`01100000 11010 11110 0` K28.5 (RD−)

LINKCHECK: after comma lock, monitors 156.25 M clock cycles (1 second). If error count ≤ 1 → LOCKED. Otherwise → full GTP soft reset + STARTUP.
Keep-alive: once LOCKED, master periodically sends dual commas. If no comma received within ~9.6 ms → STARTUP (link loss detection).

TD-Link Data Packet Structure (0x81 Response)



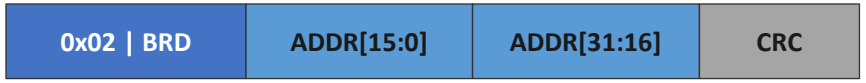
- Each FERS board appends one 0x81 response per train pass; packet is forwarded immediately (streaming, never stopped)
- SOP (Start of Packet) and EOP (End of Packet) signals generated for upstream data handling
- Train terminated by 0xC000 (EOB — End of Block); master releases pending token

TD-Link Control Packet Formats

Register Write

6 words × 16-bit | Writes a 32-bit value to a 32-bit addressed register on target board

Register Read Req



4 words × 16-bit | Requests a register read from target board

Register Read Resp

6 words × 16-bit | Board responds with register value; CRC verified by master

Command

5 words × 16-bit | ARG[47:16] offset by LinkTimeCode for deterministic execution

Enumeration TX

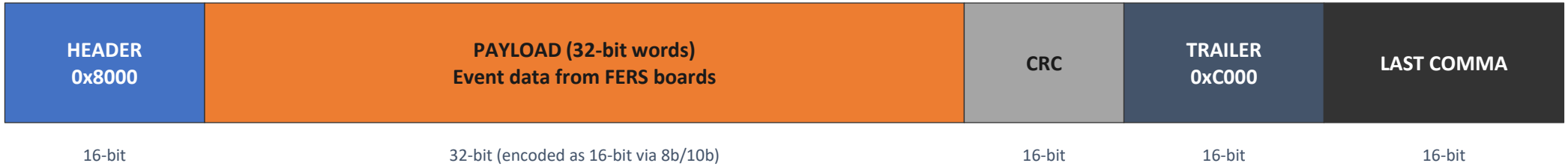
Enumeration RX

Each board increments CNT as packet traverses the chain; round-trip time measured

All fields are 16-bit words on the link (8b/10b encoded)

Address and data are 32-bit values split across two consecutive 16-bit words

TD-Link Packet Structure



Encoding Scheme

- Payload carries **32-bit data words**
- 8b/10b encodes them on **16-bit symbols**
- Header, CRC, Trailer, LAST COMMA are **16-bit control words**

Streaming Behavior

- Each FERS appends up to **N events** per train pass
- CRC **appended at the end of each hop**
- Payloads stored in concentrator **DDR4** (64 Mword/link)