White Rabbit

Erik van der Bij & Maciej Lipiński

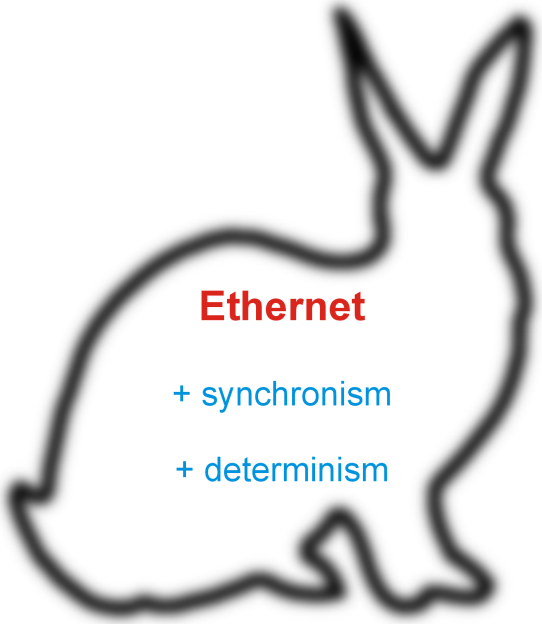
Hardware and Timing Section
CERN

21st November 2012
Network For European Accurate Time & Frequency Transfer

Outline

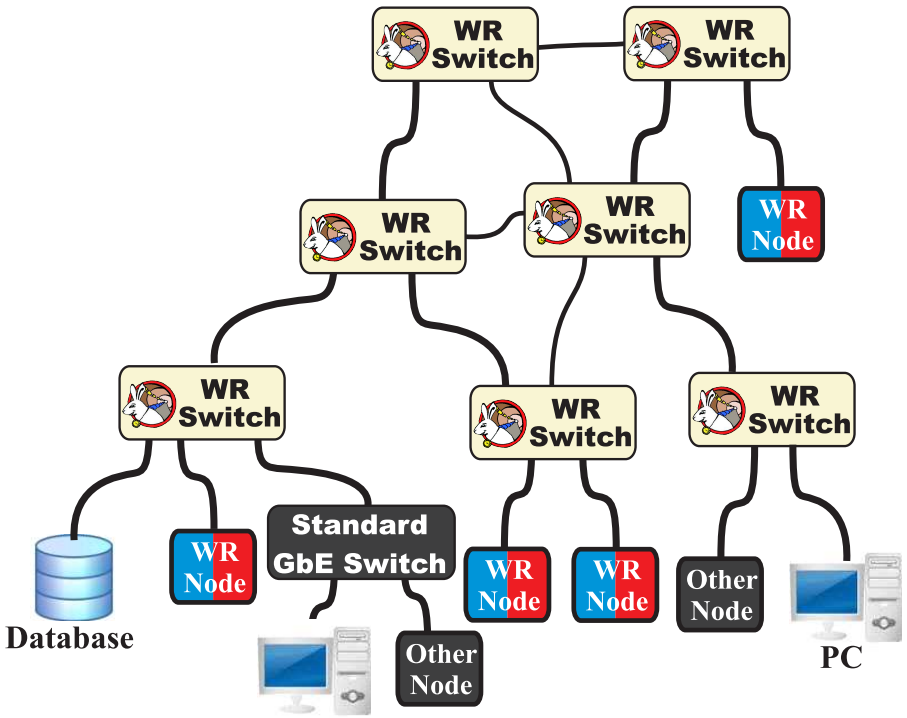
What is White Rabbit?

- Main features of a White Rabbit network:
  - transparent, **high-accuracy** synchronization
  - low-latency, **deterministic** data delivery
  - designed for **high reliability**
- Accelerator's control and timing
- International collaboration
- Based on well-known technologies
- Open Hardware and Open Software



White Rabbit – enhanced Ethernet

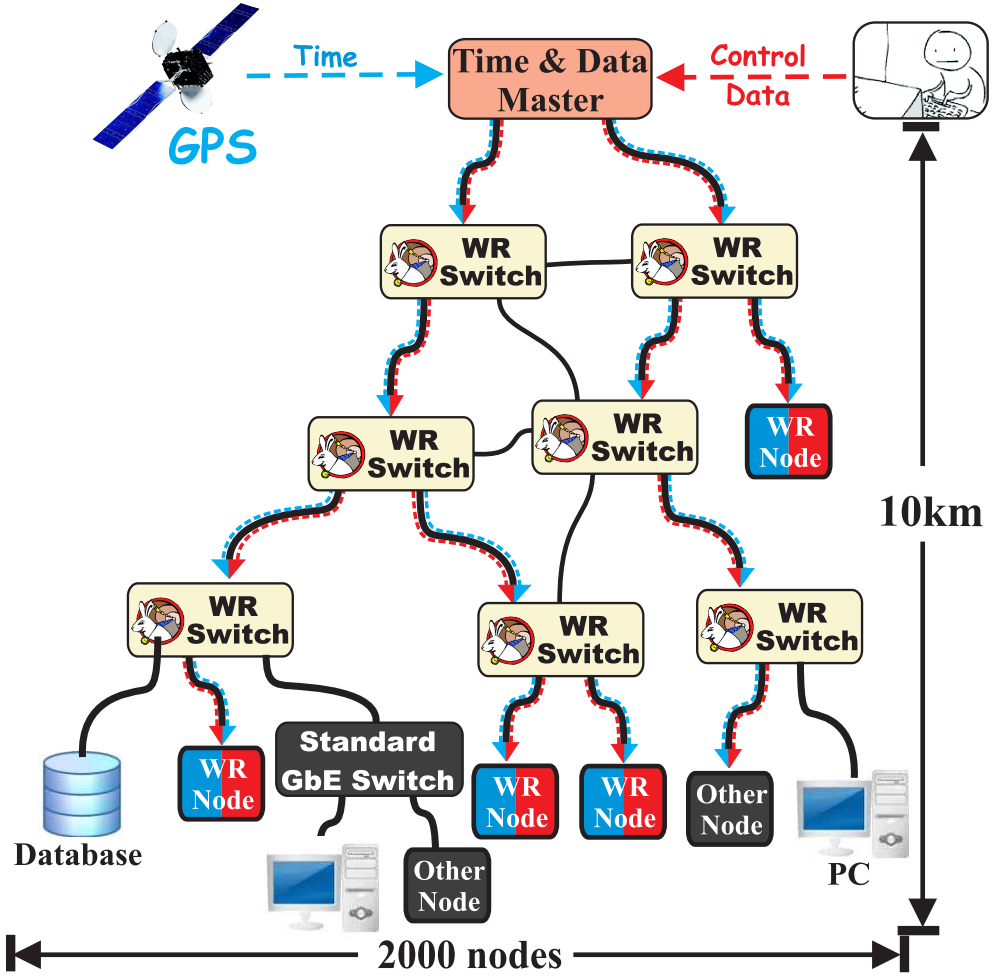
- Few thousands nodes
- Fiber medium
- Up to 10 km fiber links
- Bandwidth: 1 Gbps
- WR Switch: 18 ports
- Non-WR Devices
- Ethernet features (VLAN) & protocols (SNMP)



White Rabbit – enhanced Ethernet

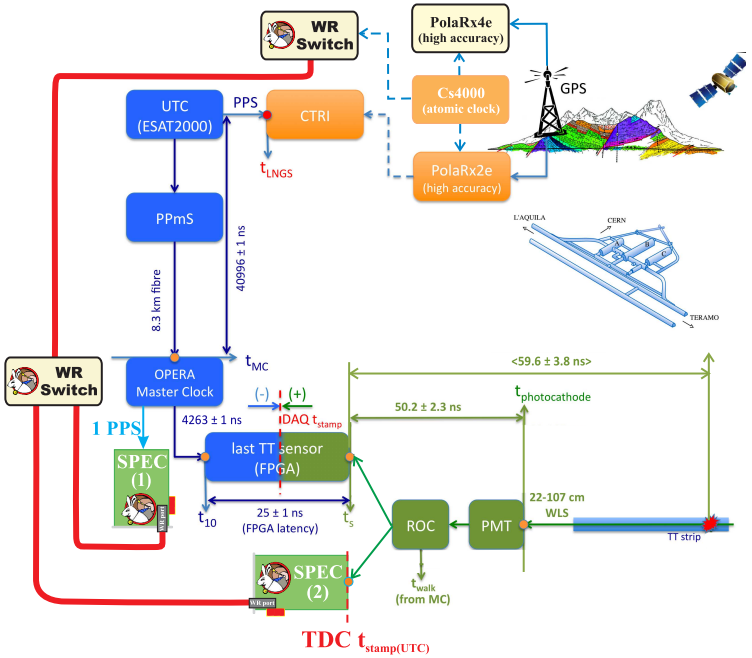
Two separate services (enhancements to Ethernet) provided by WR:

- High accuracy/precision synchronization
- Deterministic, reliable and low-latency Control Data delivery



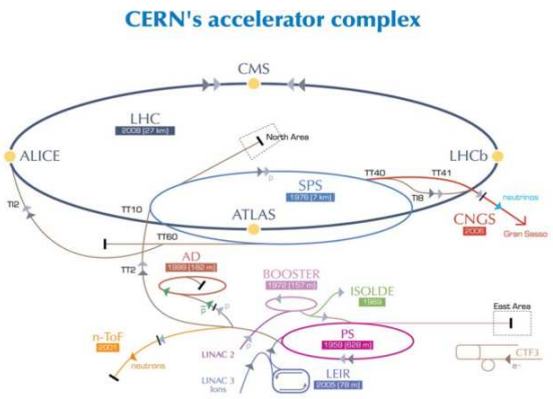
White Rabbit applications

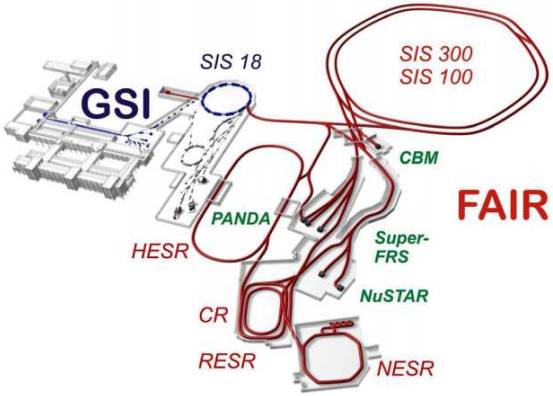
- Existing applications:
  - **CERN Neutrinos to Gran Sasso**



White Rabbit applications

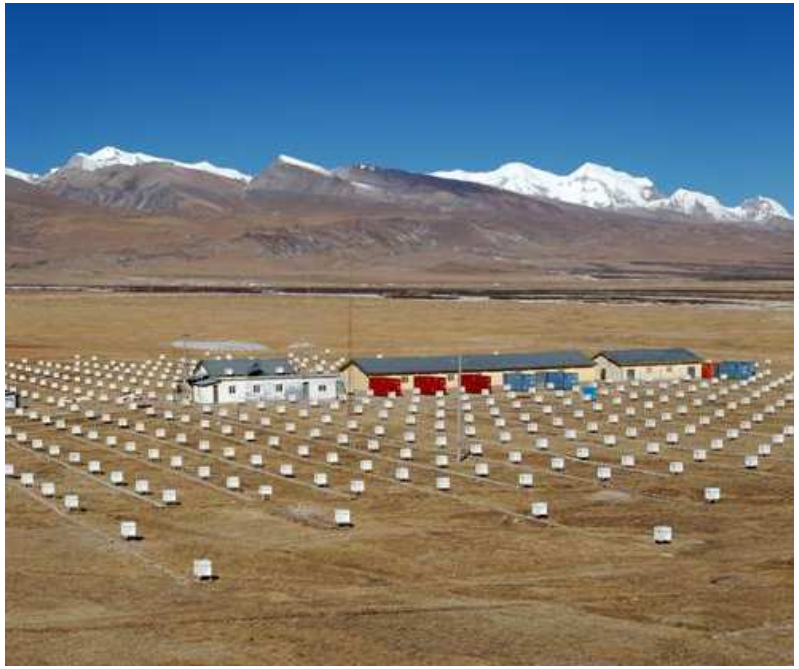
- Existing applications:
  - CERN Neutrinos to Gran Sasso
- Future applications:
  - **CERN and GSI**





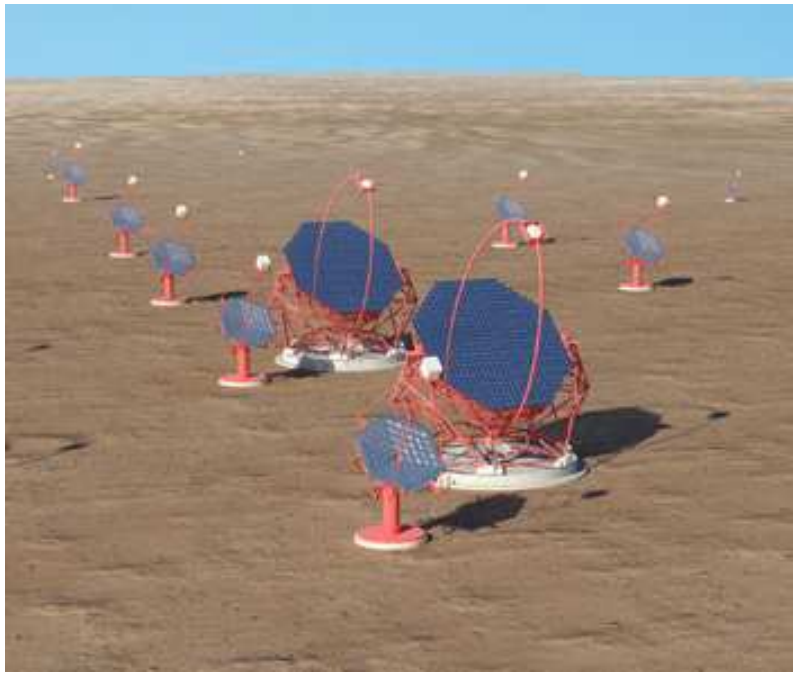
White Rabbit applications

- Existing applications:
  - CERN Neutrinos to Gran Sasso
- Future applications:
  - CERN and GSI
  - **The Large High Altitude Air Shower Observatory (China)**



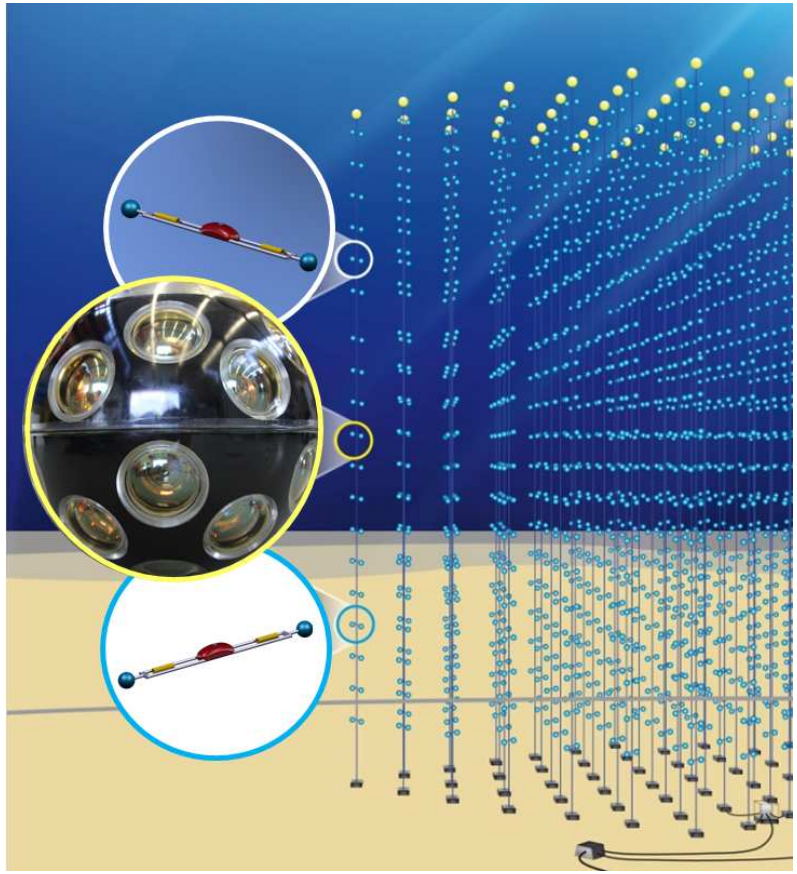
White Rabbit applications

- Existing applications:
  - CERN Neutrinos to Gran Sasso
- Future applications:
  - CERN and GSI
  - The Large High Altitude Air Shower Observatory (China)
- Potential applications:
  - **Cherenkov Telescope Array**



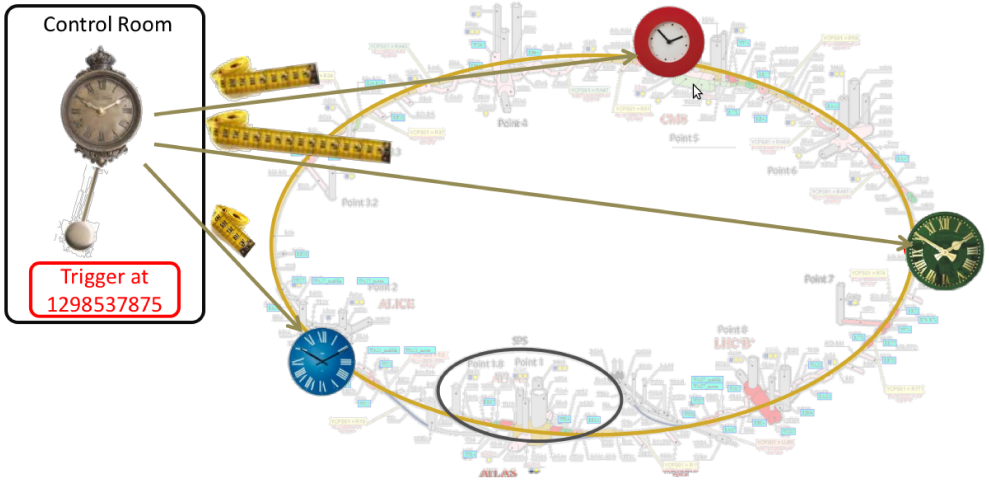
White Rabbit applications

- Existing applications:
  - CERN Neutrinos to Gran Sasso
- Future applications:
  - CERN and GSI
  - The Large High Altitude Air Shower Observatory (China)
- Potential applications:
  - Cherenkov Telescope Array
  - **European deep-sea research infrastructure (KM3NET)**
  - Others



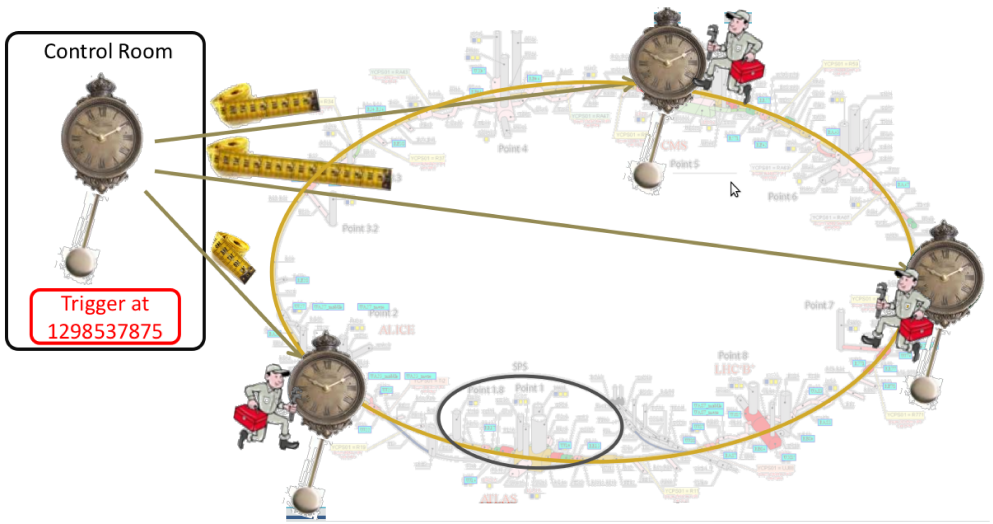
Origin of White Rabbit Project

- Renovation of CERN's General Machine Timing (GMT)
- GMT is great but...:
  - **Manual calibration** (with CS clock)
  - **RS-422**, 500kbps
  - **Unidirectional** communication
  - Separate network required
  - **Custom design, complicated maintenance**
- White Rabbit is meant to solve these problems



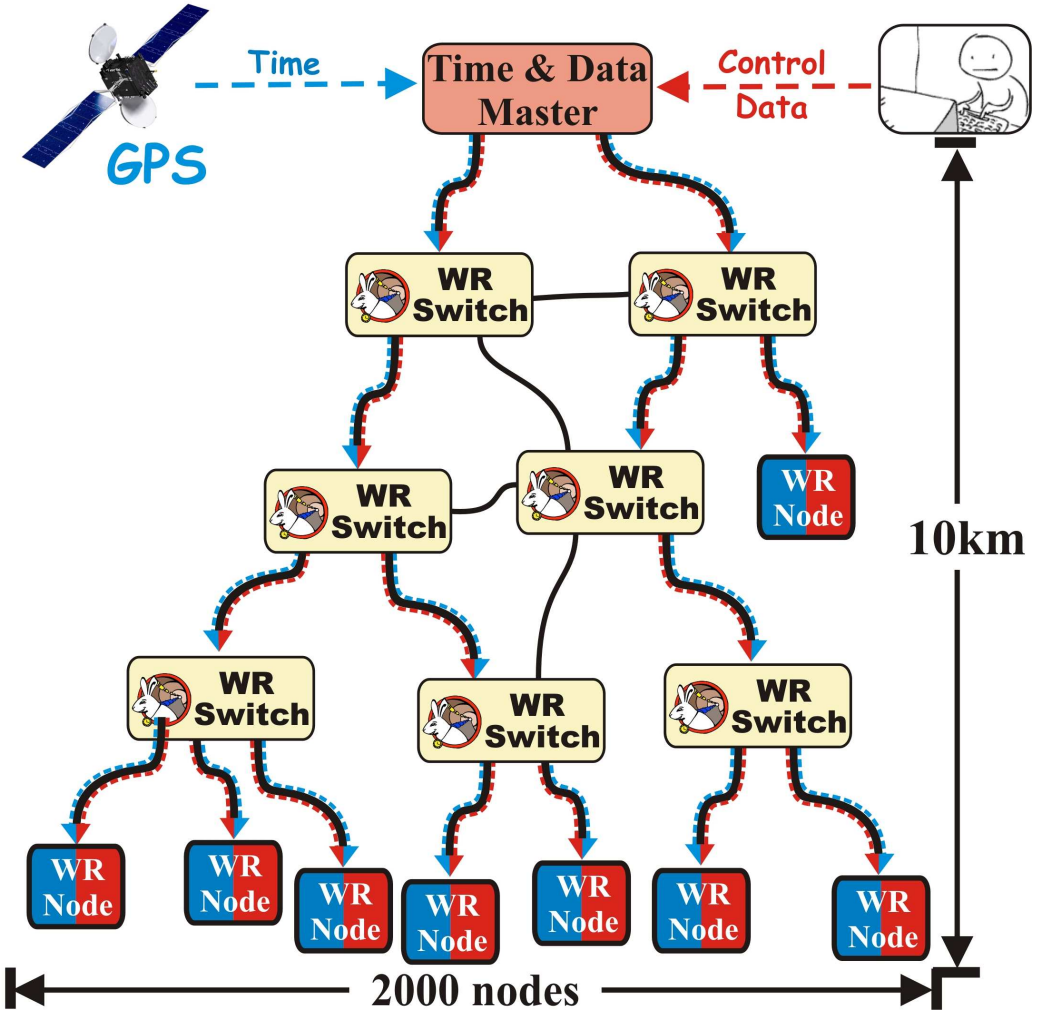
Origin of White Rabbit Project

- Renovation of CERN's General Machine Timing (GMT)
- GMT is great but...:
  - **Manual calibration** (with CS clock)
  - **RS-422**, 500kbps
  - **Unidirectional** communication
  - Separate network required
  - **Custom design, complicated maintenance**
- White Rabbit is meant to solve these problems



Time distribution in White Rabbit

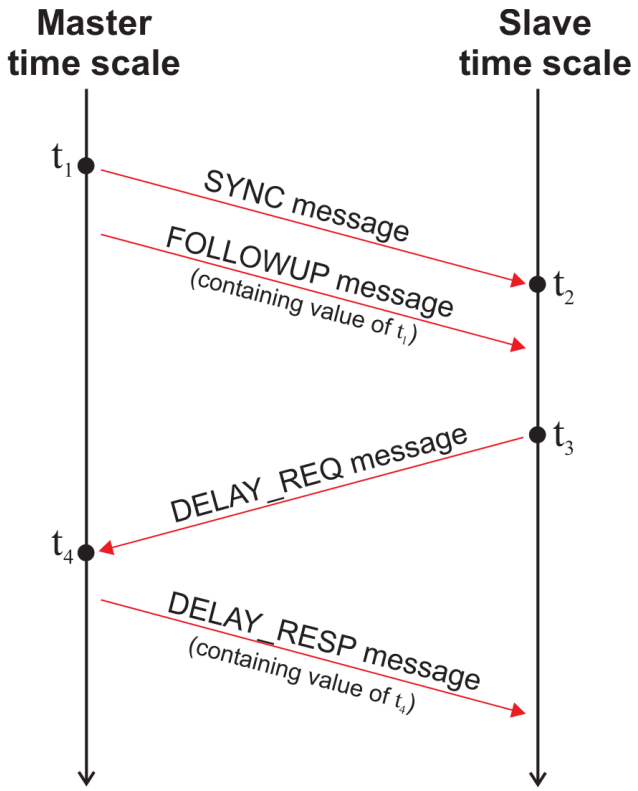
- **High accuracy/precision synchronization**
- Deterministic, reliable and low-latency Control Data delivery



Time Distribution in White Rabbit Network

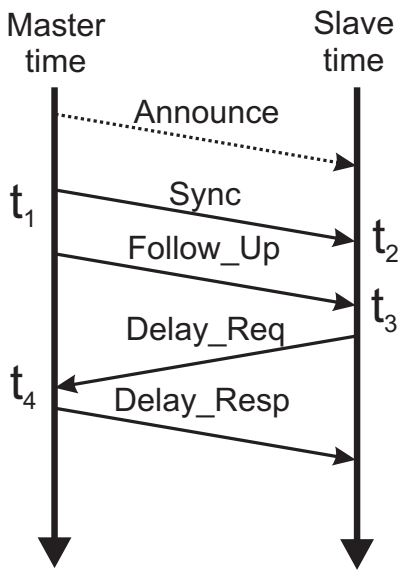
- Synchronization with **sub-ns** accuracy and **ps** precision
- Combination of
  - Precision Time Protocol (**PTP**) synchronization
  - Synchronous Ethernet (**SyncE**) syntonization
  - Digital Dual-Mixer Time Difference (**DDMTD**) phase detection

Precision Time Protocol (IEEE1588)



- Packet-based synchronization protocol
- Synchronizes local clock with the master clock
- Link delay evaluated by measuring and exchanging packets tx/rx timestamps

WR Precision Time Protocol Extension



$$\mu = \frac{(t_4 - t_1) - (t_3 - t_2)}{2}$$

Relative Delay Coefficient: α

Fixed Delays: Δ_{rxs}, Δ_{txs}, Δ_{rxm}, Δ_{txm}

μ → **WR Link Delay Model**

$$\mathbf{delay_{ms} = \mu + asymmetry}$$

$$\mathbf{offset_{ms} = t_2 - (t_1 + delay_{ms})}$$

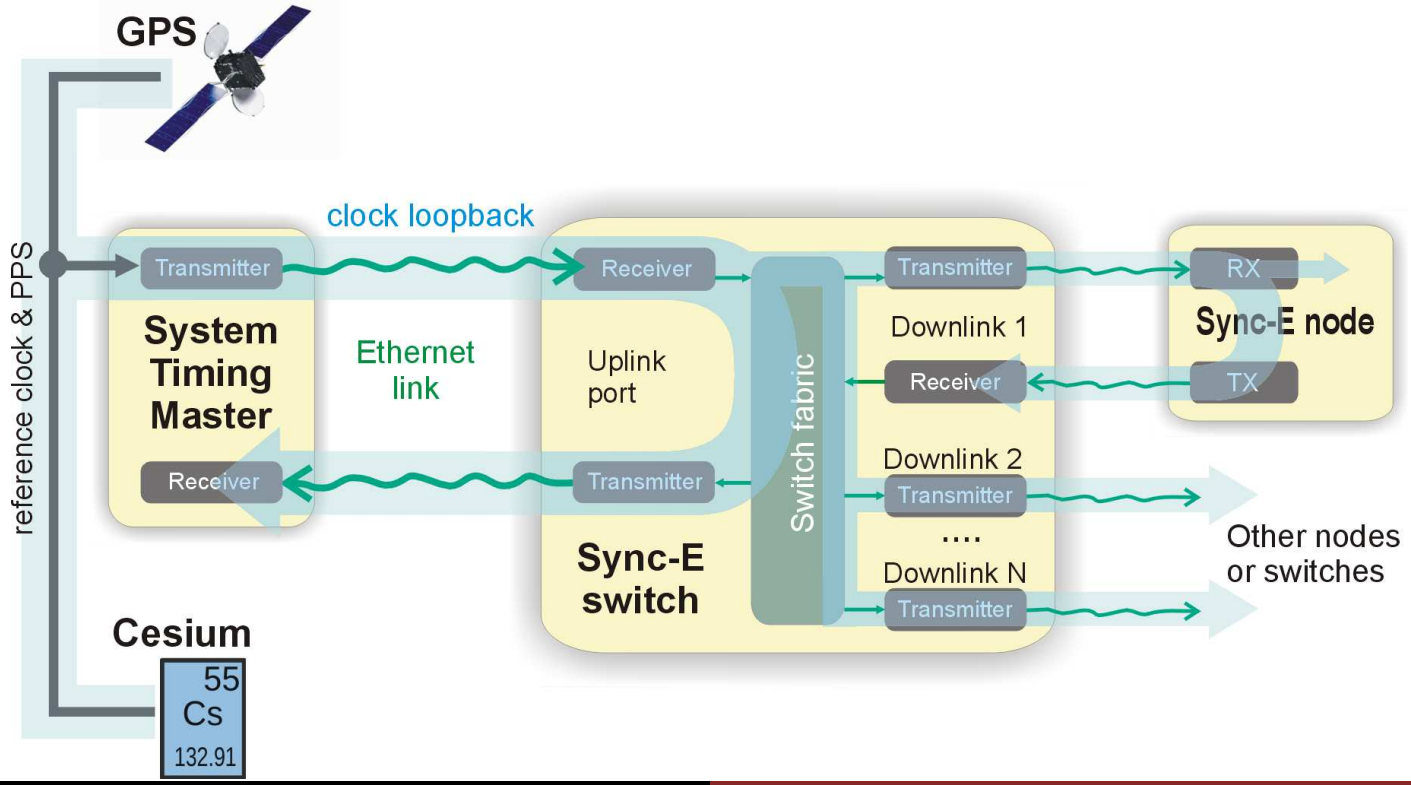
Solution for Ethernet over a Single-mode Optical Fiber

$$asymmetry = \Delta_{tx_m} + \Delta_{rx_s} - \frac{\Delta - \alpha\mu + \alpha\Delta}{2 + \alpha}$$

Synchronous Ethernet (SyncE)

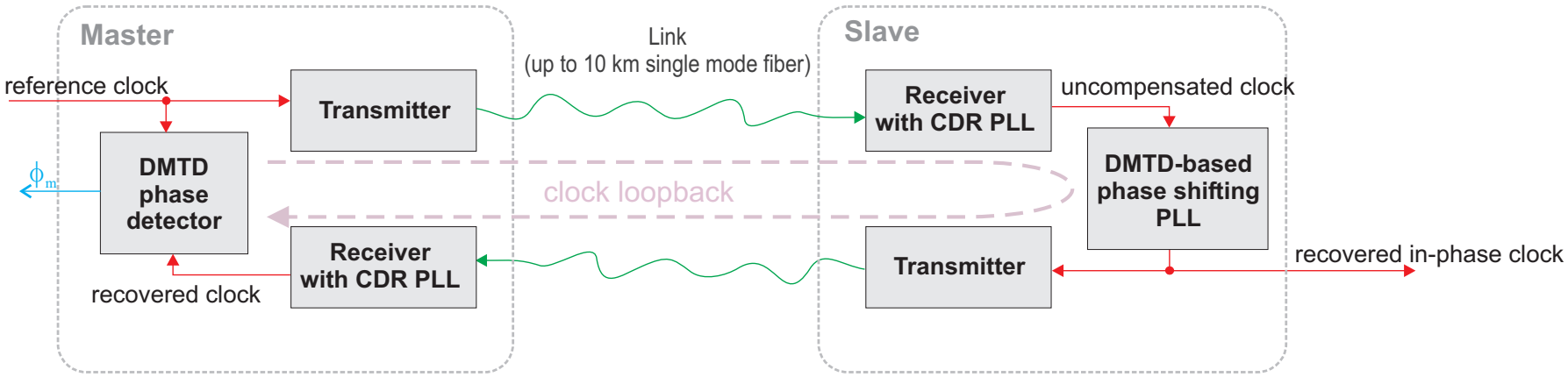
Common clock for the entire network

- All network devices use the same physical layer clock
- Clock is encoded in the Ethernet carrier and recovered by the receiver chip (PHY).



DDMTD: Phase tracking

- PTP limitation: timestamping granularity
- Solution: take advantage of SyncE and measure phase shift



WR PTP

- Extension to PTP (IEEE1588) – defined as PTP Profile
- Addresses PTP's limitations
  (granularity, asymmetry, syntonization)
- Compatible with "standard" PTP gear
- Ongoing standardization effort to include WR into PTPv3
- Lab&field-tested for sub-ns synchronization

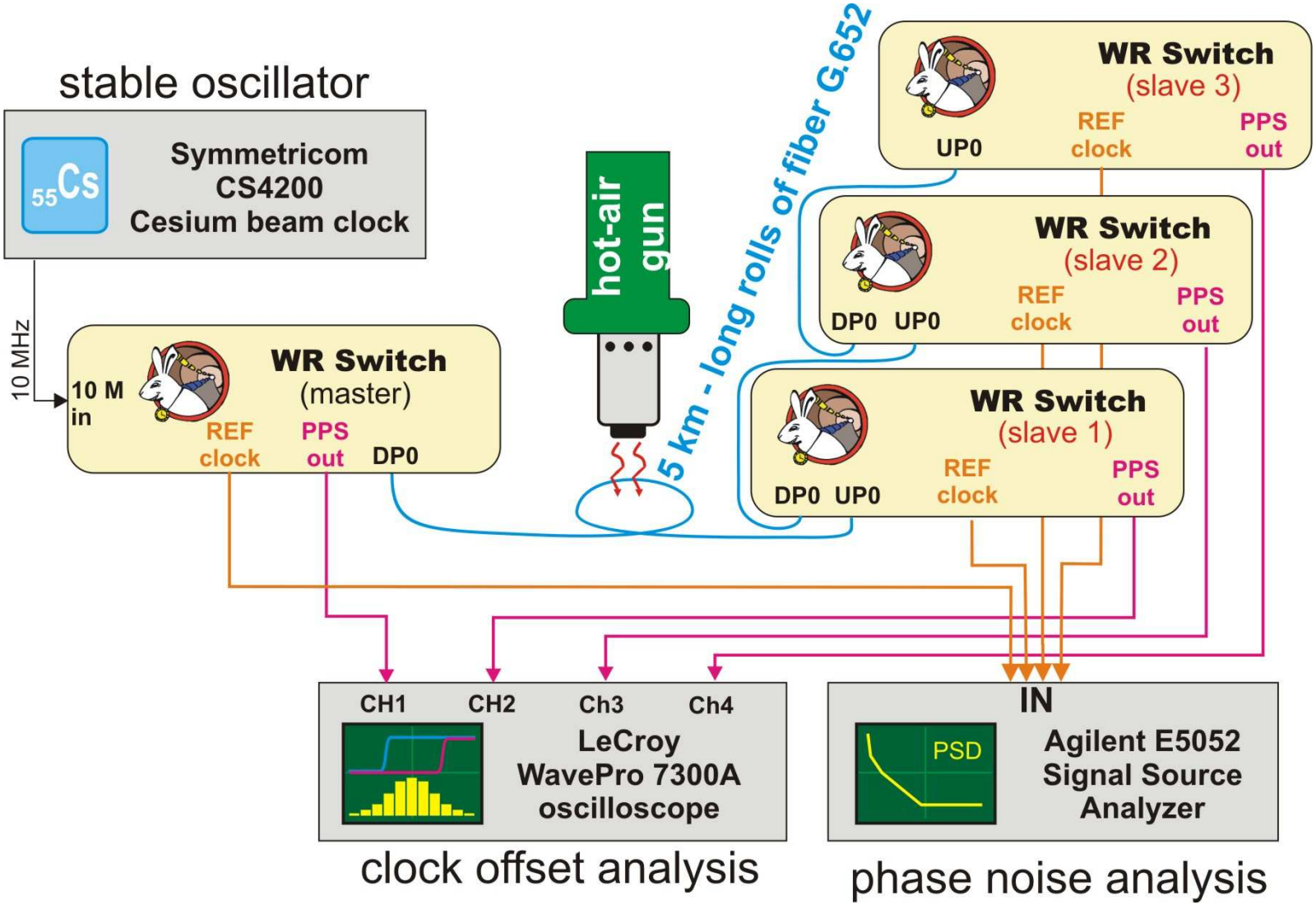
WR PTP

- Extension to PTP (IEEE1588) – defined as PTP Profile
- Addresses PTP's limitations (granularity, asymmetry, syntonization)
- Compatible with "standard" PTP gear
- Ongoing standardization effort to include WR into PTPv3
- Lab&field-tested for sub-ns synchronization

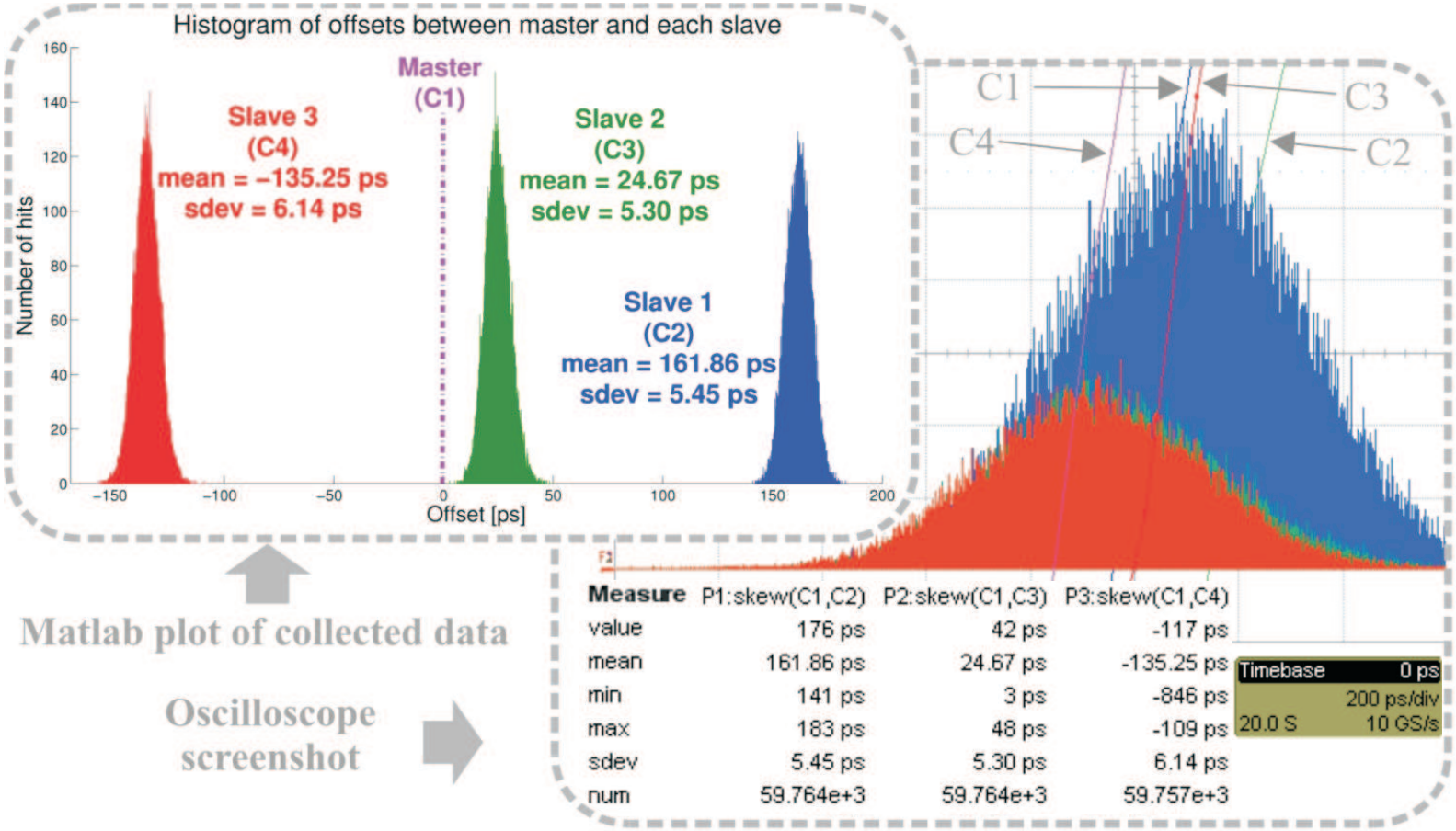
According to ISPCS Plug Fest results ...

... White Rabbit is the most accurate PTP implementation in the world!

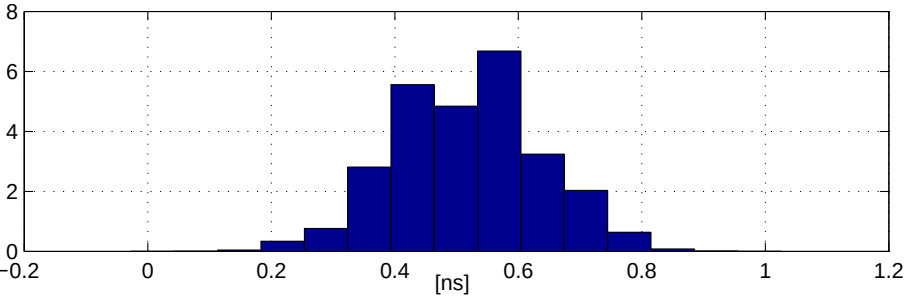
WR time transfer performance: lab tests



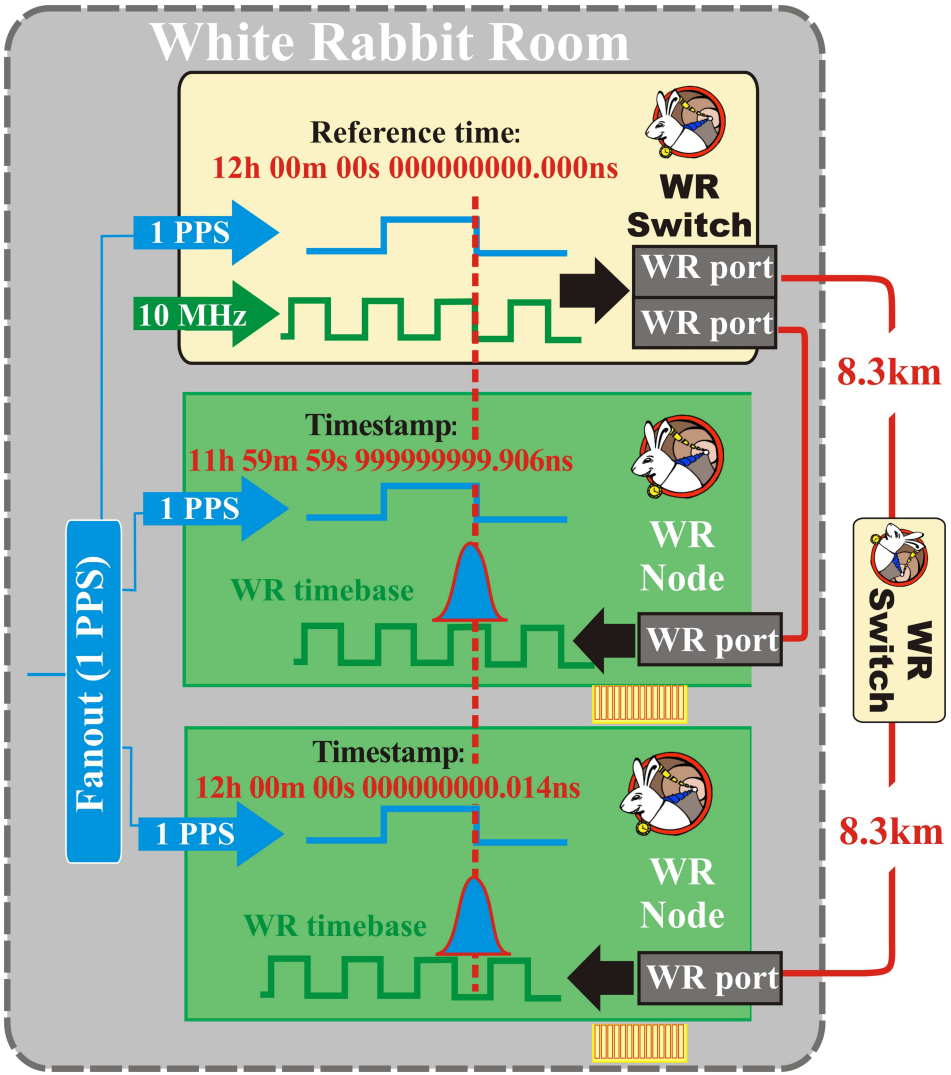
WR time transfer performance: lab tests



WR time transfer performance: deployment for CNGS

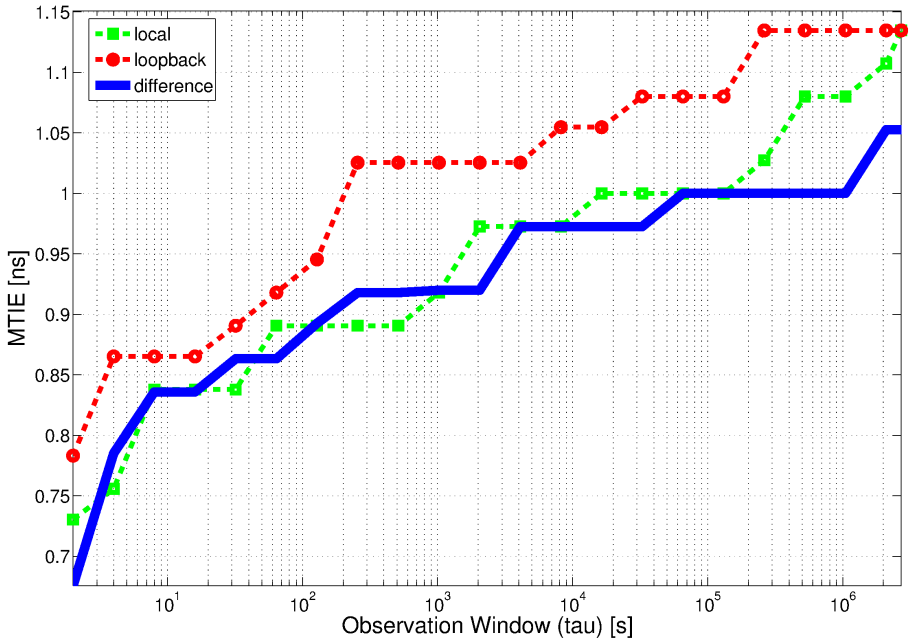


- Duration: 31 d, 7 h, 40 s ($2.7 * 10^6$ samples)
- WR Nodes with TDC used
- Measurement includes inaccuracy of TDC
- Timestamping reference PPS
- Accuracy: 0.517 ns
- Precision: 0.119 ns (std. dev)

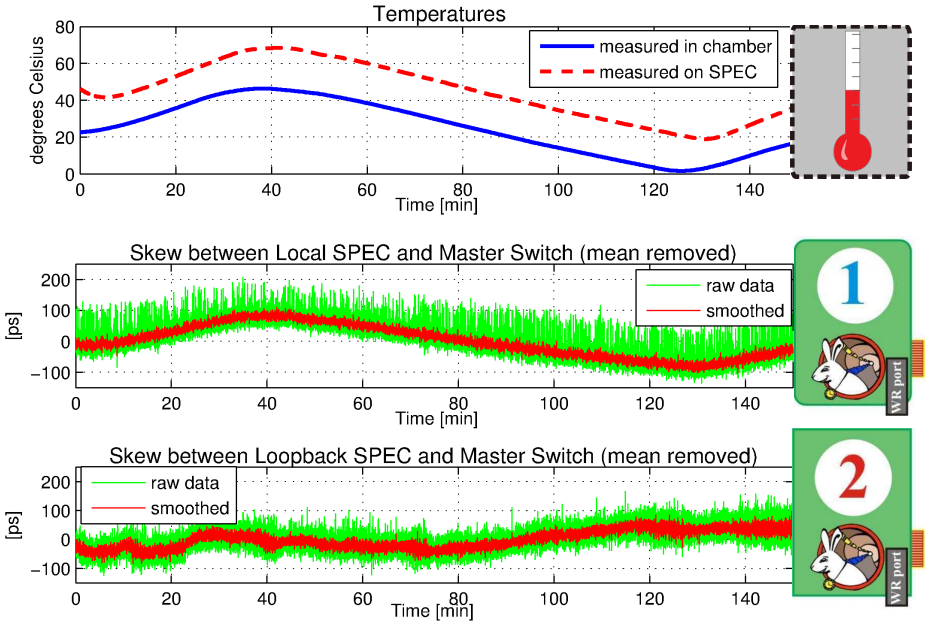


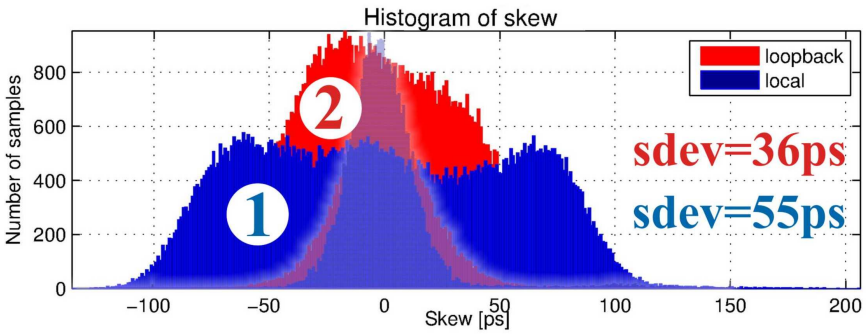
WR time transfer performance: deployment for CNGS

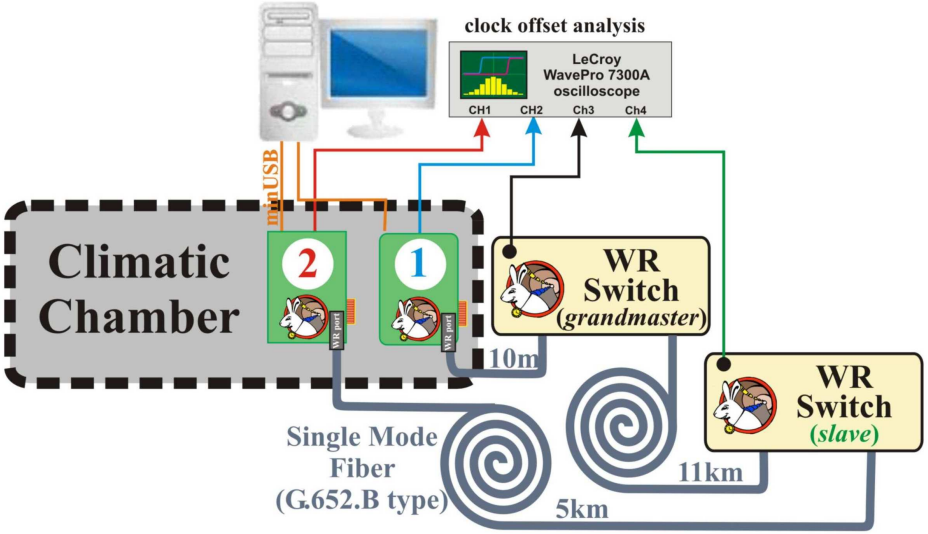
Out of $2.7 * 10^6$ samples
9 values of x_{diff} [0.0003%]
exceeded MTIE=1ns



WR time transfer performance: temperature tests

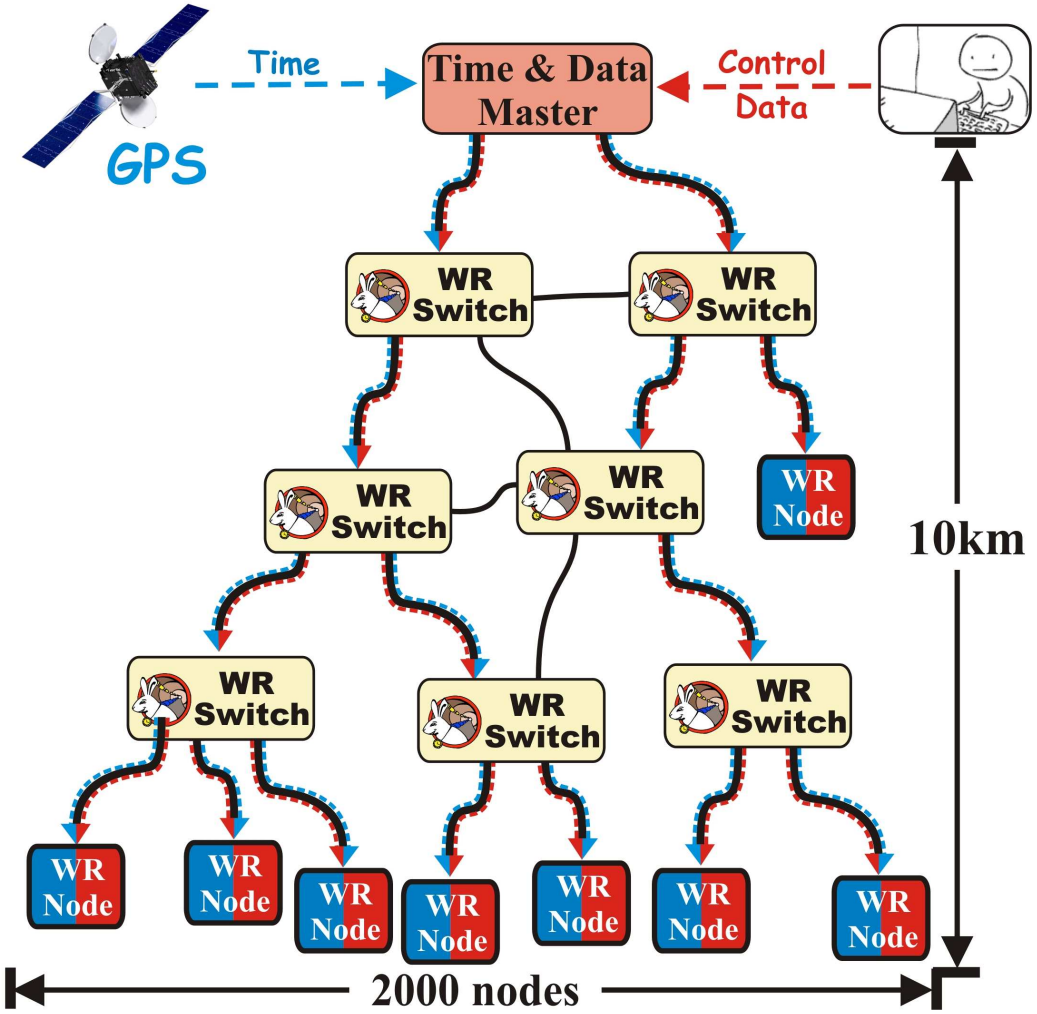






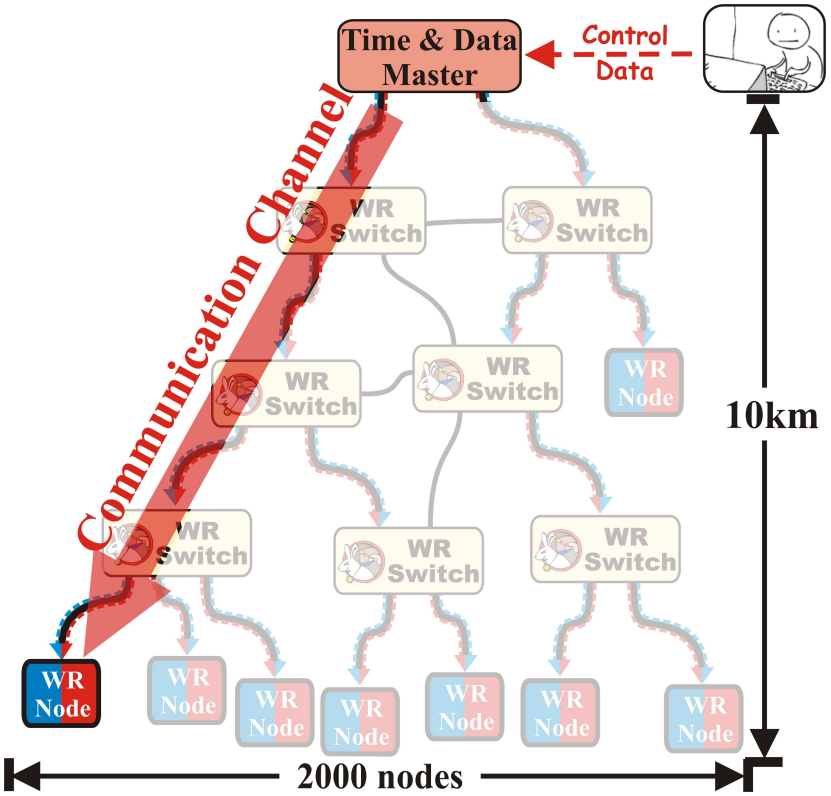
Data distribution in White Rabbit

- High accuracy/precision synchronization
- **Deterministic, reliable and low-latency Control Data delivery**



Control Data

- Two types of data:
  - **Control Data** (High Priority, HP)
  - Standard Data (Best Effort)
- Characteristics of **Control Data**
  - Sent in Control Messages
  - Sent by Data Master(s)
  - Broadcast (one-to-alot)
  - Deterministic and low latency
  - Reliable delivery

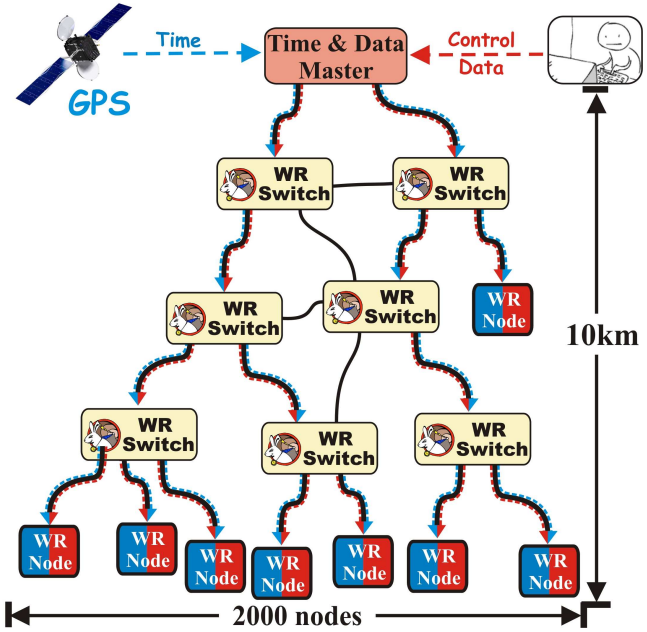


Data Redundancy

- Re-transmission of Control Data not possible
- **Forward Error Correction** – additional transparent layer:
  - One Control Message encoded into N Ethernet frames,
  - Recovery of Control Message from any M (M<N) frames
- FEC can prevent data loss due to:
  - **bit error**
  - **network reconfiguration**

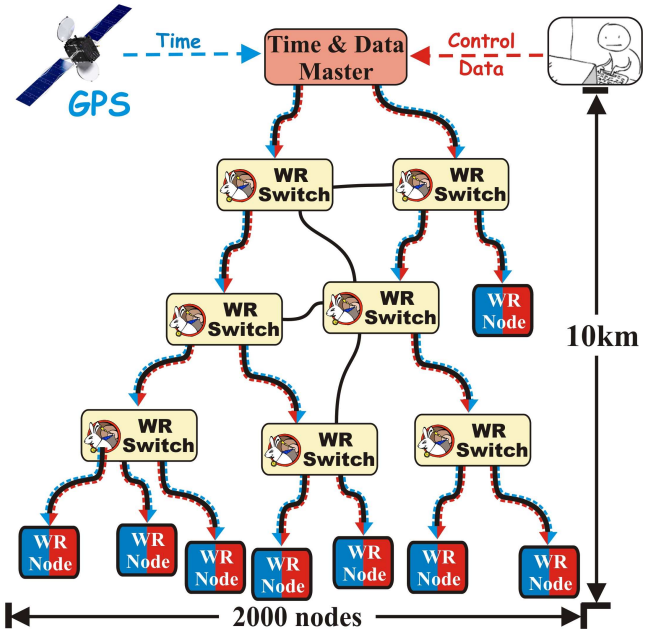
Topology Redundancy

- Standard Ethernet solution: Rapid/Multi Spanning Tree Protocol
- Reconfiguration time: $\approx$ **1s** (best: milliseconds)
- **1s =** $\approx$ **82 000** Ethernet Frames lost
- Solution:
  - take advantage of FEC
  - speed up (R/M)STP$->$**eRSTP** or
  - use multiple paths$->$**eLACP**



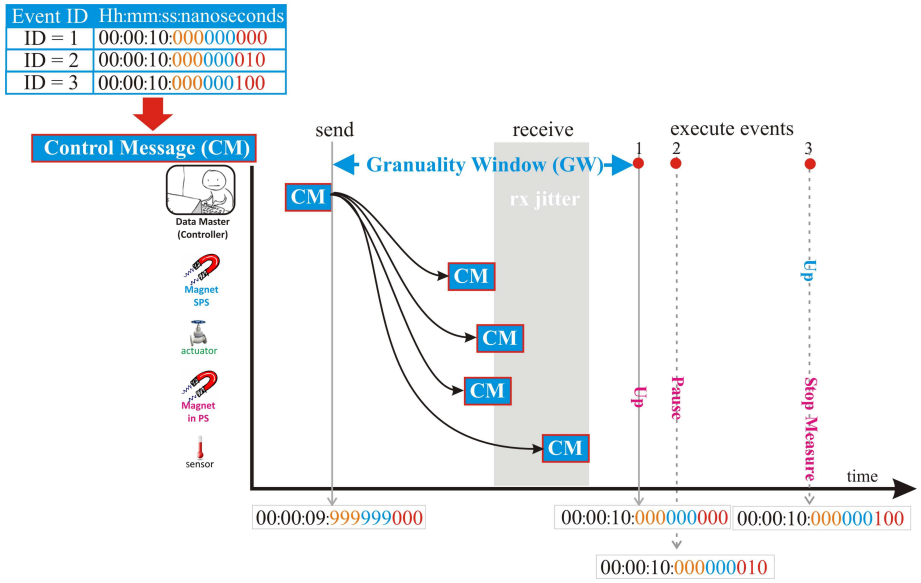
Determinism and low latency

- Control Data:
  7th Class of Service (priority)
- WR Switch:
  - Quality of Service: resource reservation
  - Upper bound latency by design: $< 10us$
  - Cut-through
- Careful diagnostics

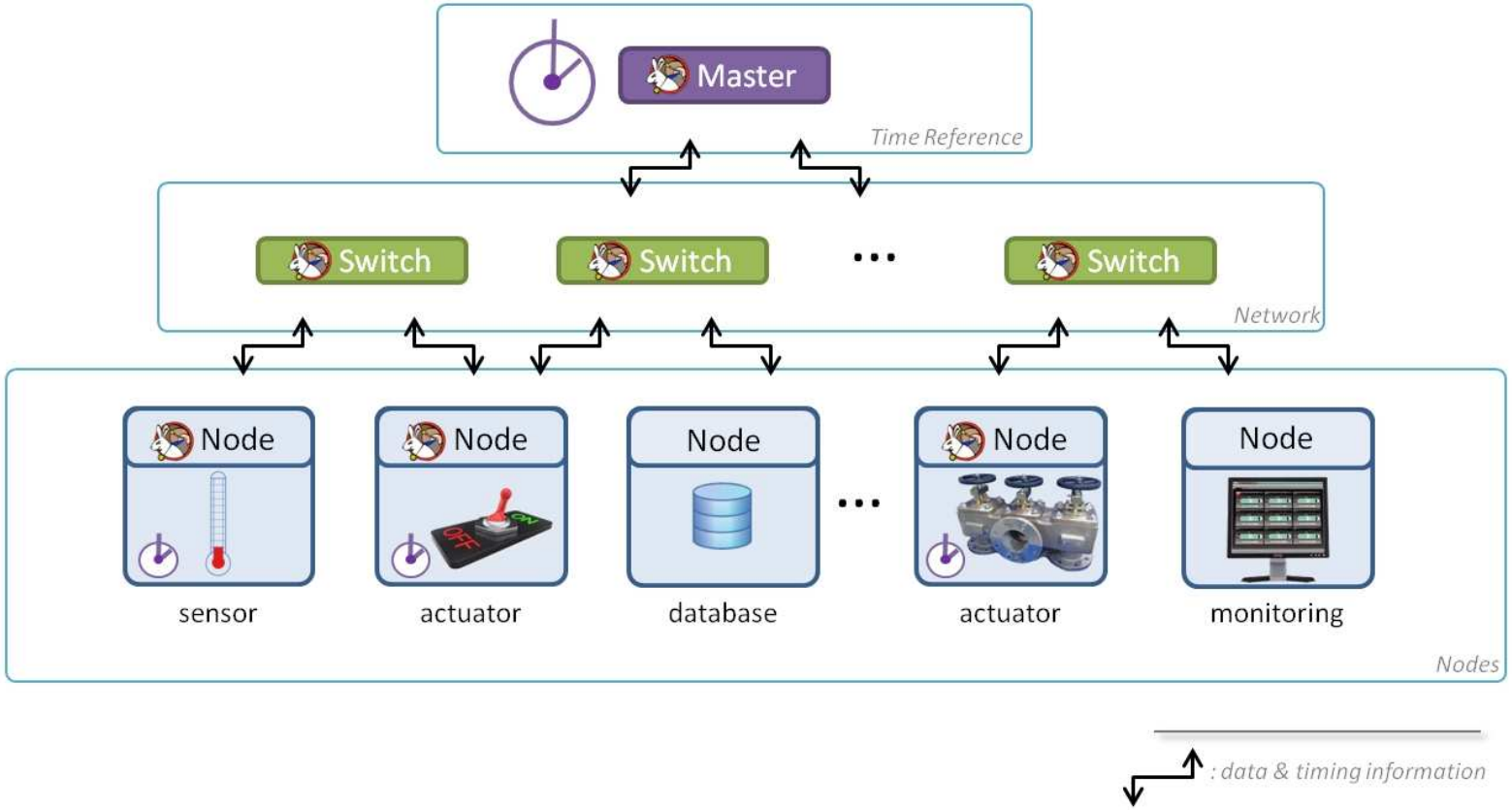


Determinism and low latency

- Control Data:
  7th Class of Service (priority)
- WR Switch:
  - Quality of Service: resource reservation
  - Upper bound latency by design: $< 10us$
  - Cut-through
- Careful diagnostics

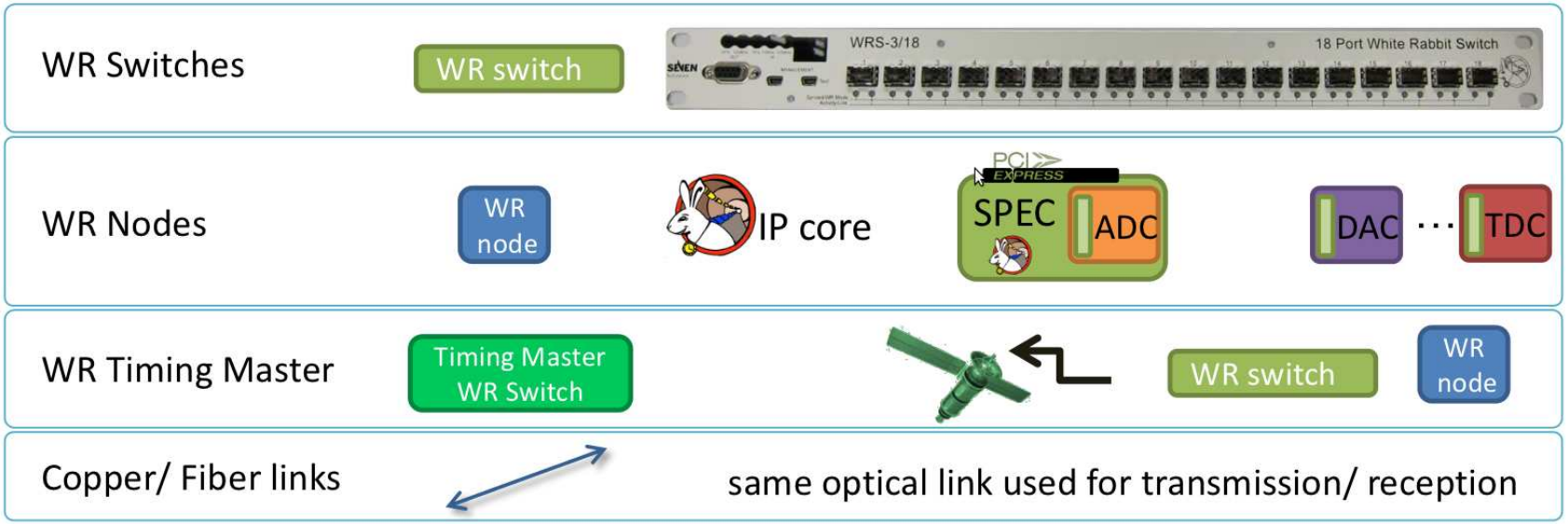


White Rabbit Network Components



White Rabbit Network Components

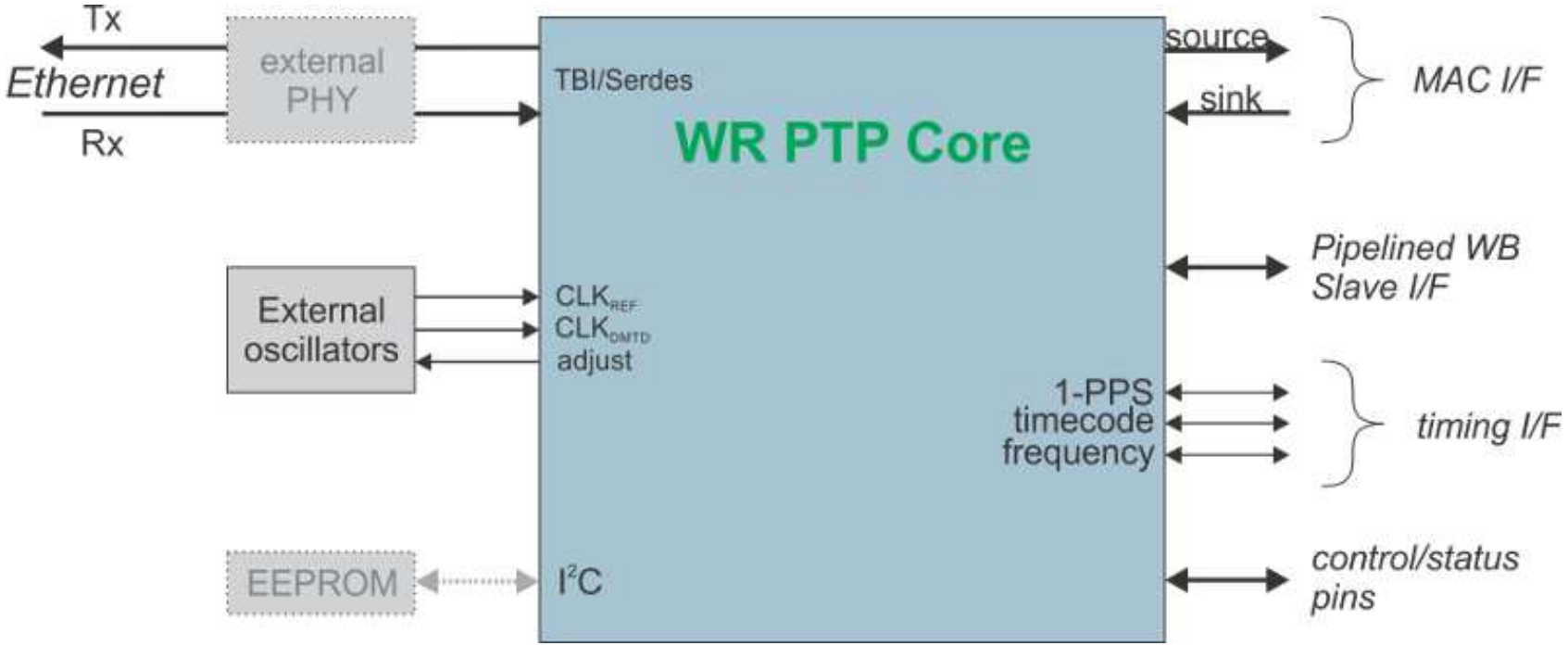
A White Rabbit network is composed of



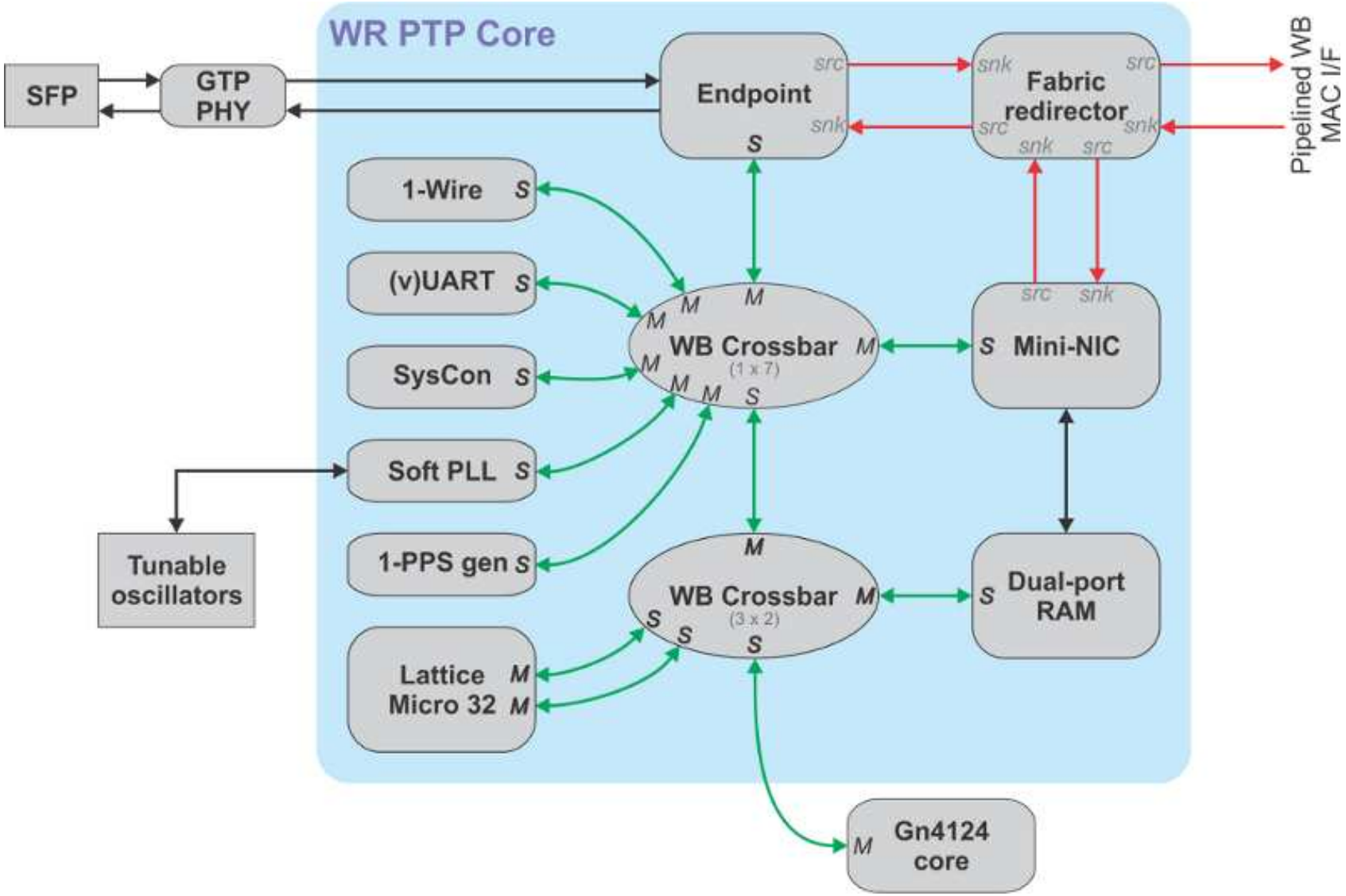
White Rabbit Switch (V3)

- Central element of WR network
- Original design optimized for timing, designed from scratch
- 18 1000BASE-BX10 ports
- Capable of driving 10 km of SM fiber
- Open design (H/W and S/W)

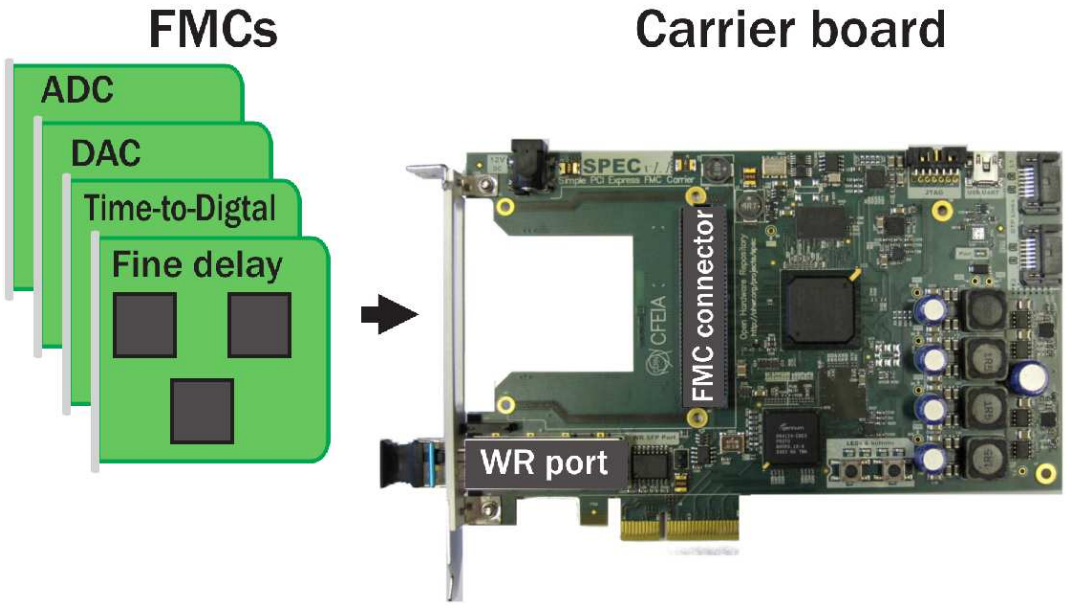
WR Node: WR PTP Core



WR Node: WR PTP Core



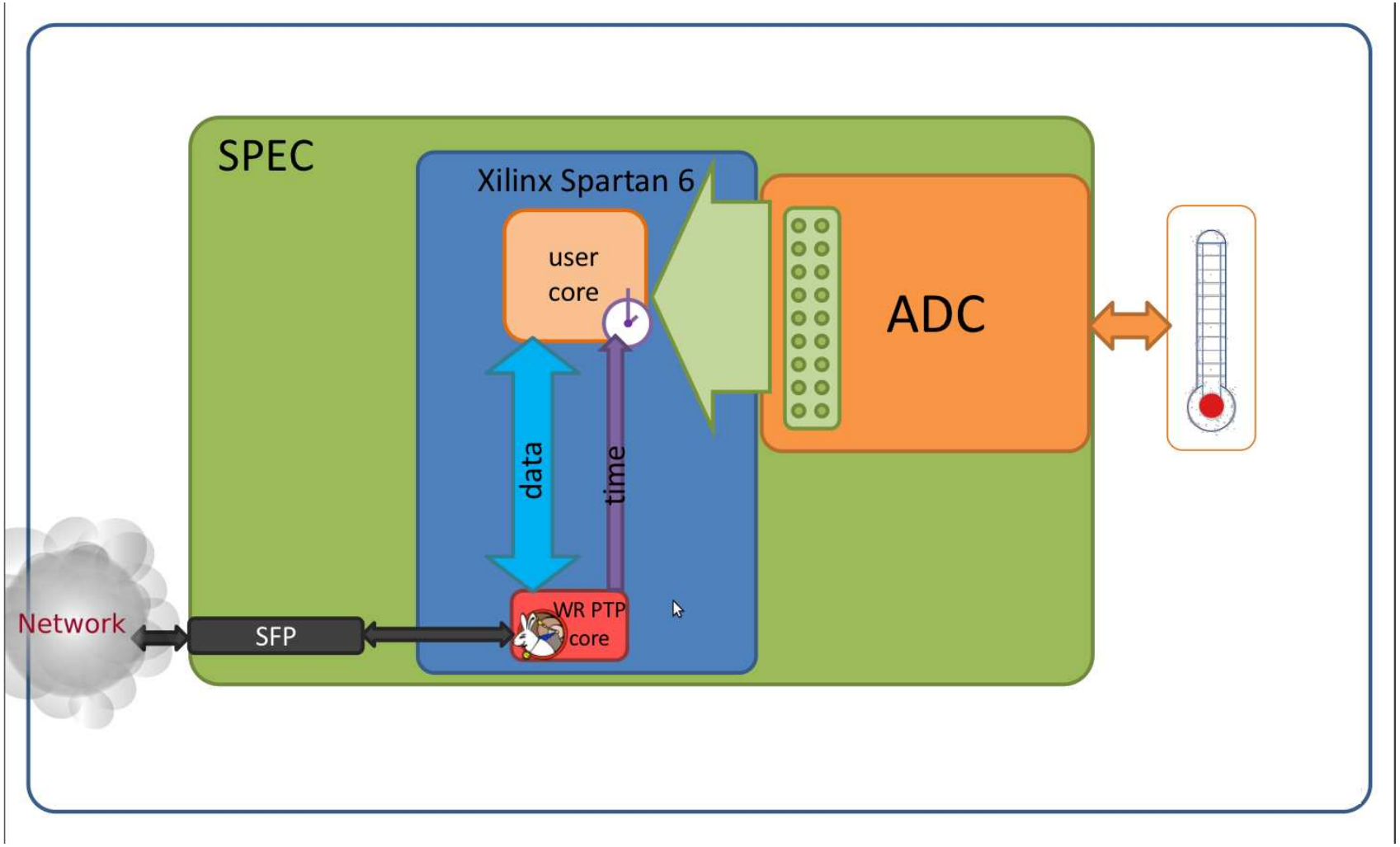
WR Node: SPEC board



FMC-based Hardware Kit:

- FMCs (FPGA Mezzanine Cards) with ADCs, DACs, TDCs, fine delays, digital I/O
- Carrier boards in PCI-Express, VME
- All carriers are equipped with a White Rabbit port

WR Node: SPEC board



WR Node: ongoing efforts

- **Simple VME FMC Carrier (SVEC)**
- Compact Universal Timing Endpoint Based on White Rabbit
- PXI express FMC Carrier Board (SPEXI)

WR Node: ongoing efforts

- Simple VME FMC Carrier (SVEC)
- **Compact Universal Timing Endpoint Based on White Rabbit**
- PXI express FMC Carrier Board (SPEXI)

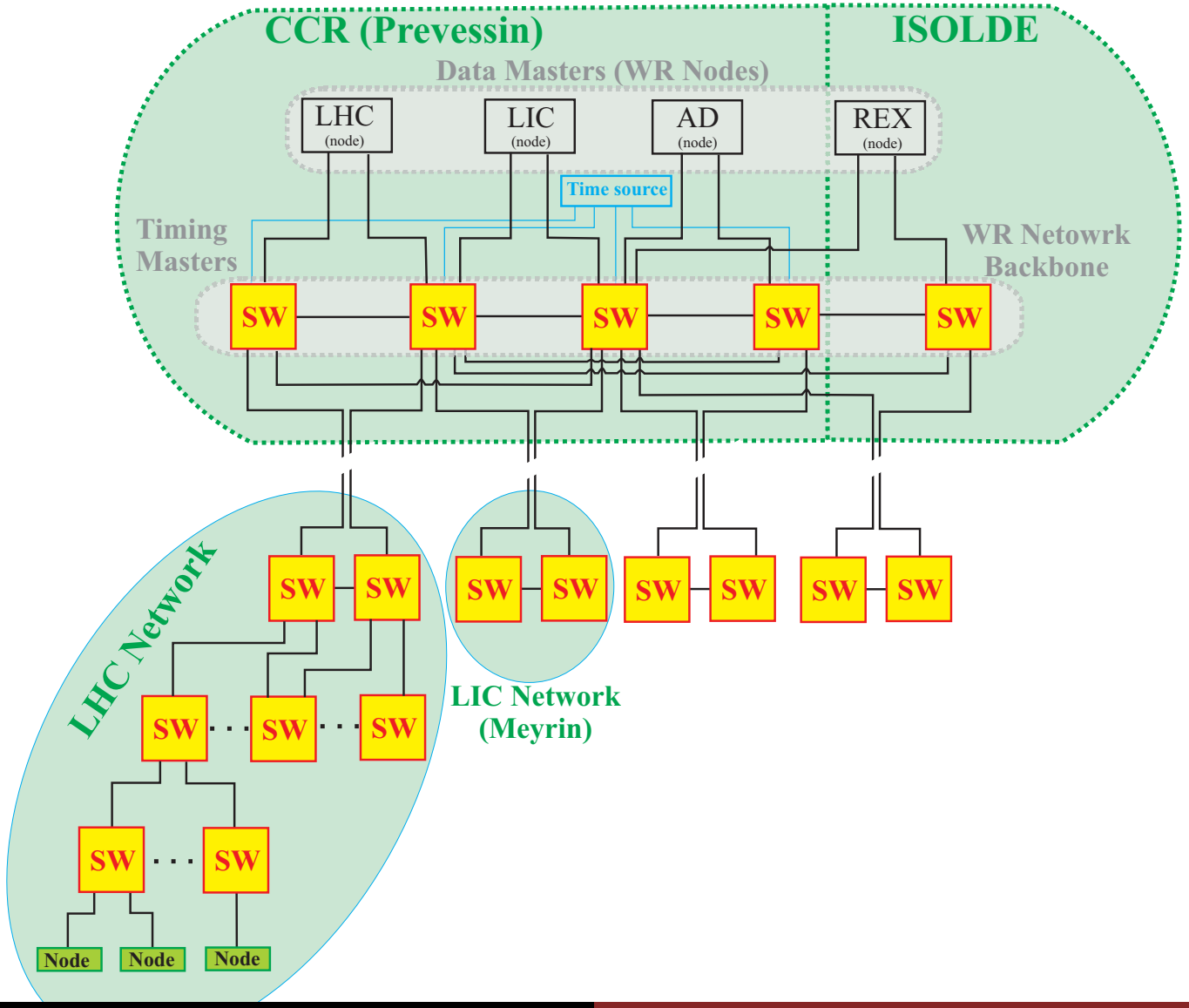
WR Node: ongoing efforts

- Simple VME FMC Carrier (SVEC)
- Compact Universal Timing Endpoint Based on White Rabbit
- **PXI express FMC Carrier Board (SPEXI)**

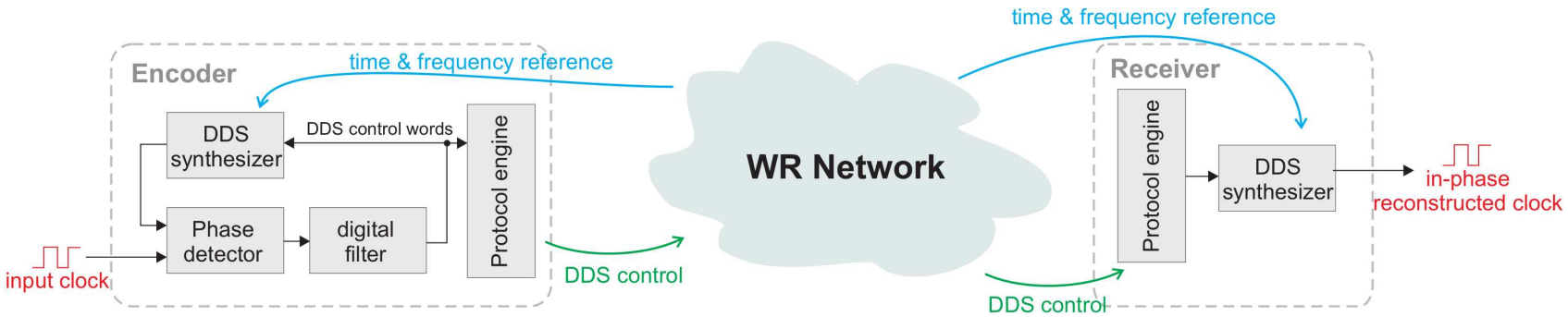
White Rabbit applications

- Control and timing system
- Field bus recommended at CERN
- Time Transfer
- RF distribution
- Distributed oscilloscope
- ...

WR at CERN


CCR (Prevessin)
ISOLDE
Data Masters (WR Nodes)
LHC (node)
LIC (node)
AD (node)
REX (node)
Time source
Timing Masters
WR Netowrk Backbone
SW
LHC Network
LIC Network (Meyrin)
Node

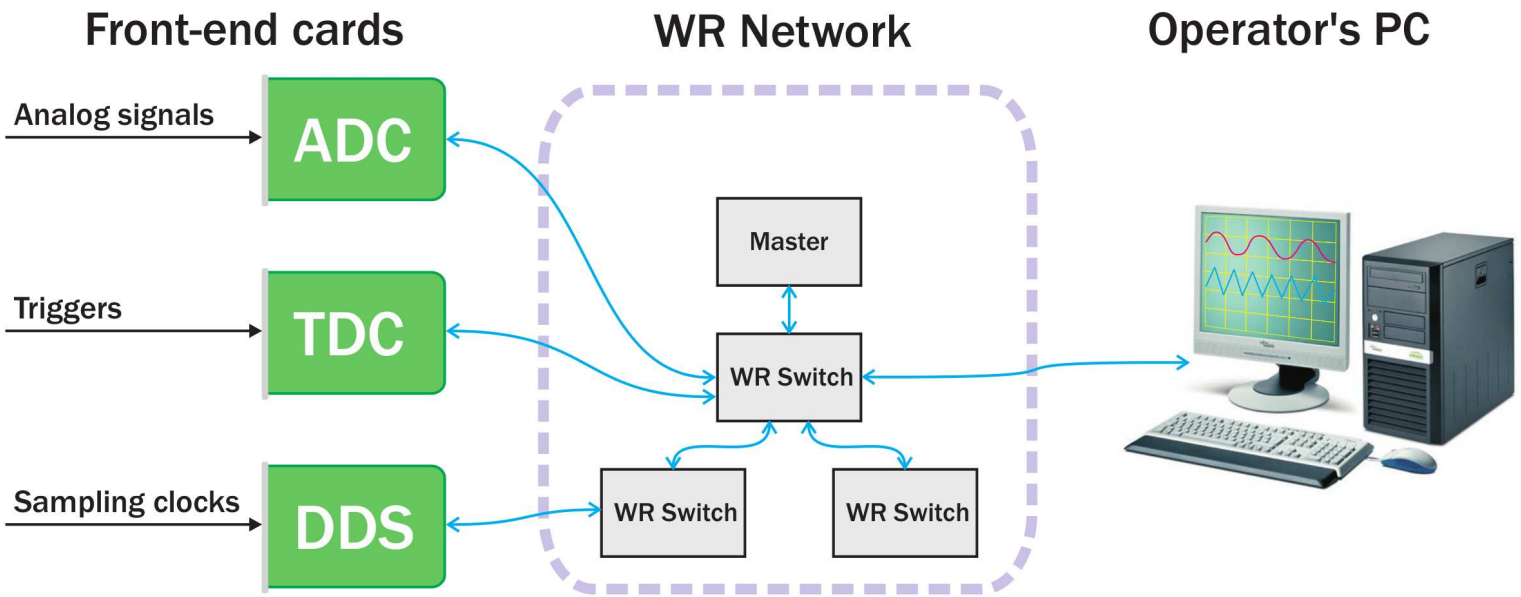
Ethernet Clock distribution a.k.a. Distributed DDS



Distributed Direct Digital Synthesis

- Replaces dozens of cables with a single fiber.
- Works over big distances without degrading signal quality.
- Can provide various clocks (TTC, RF, bunch clock) with a single, standardized link.

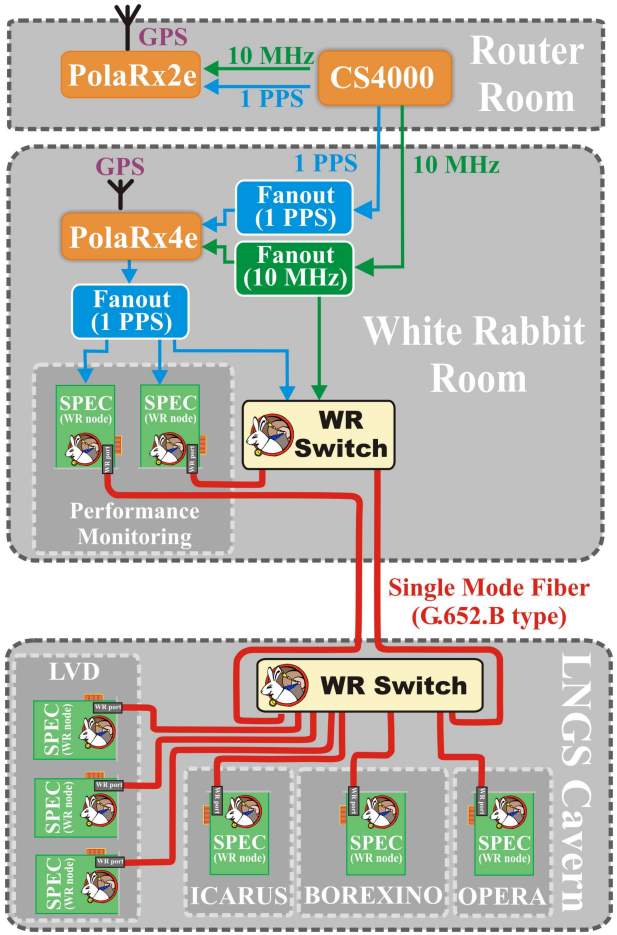
Distributed oscilloscope



- Common clock in the entire network: no skew between ADCs.
- Ability to sample with different clocks via Distributed DDS.
- External triggers can be time tagged with a TDC and used to reconstruct the original time base in the operator's PC.

Time Transfer

- WR Network connected to Cs Clock and GPS
- SPEC + Fine Delay FMC module
  - Time-to-Digital Converter
    - 28 ps resolution
    - 55ps precision (std.dev)
    - 300 ps accuracy (WR-timebase)
  - Pulse Generator:
    - Produces pulses at UTC/TAI
    - Pulse width: 50ns-1s (res: 10ps)
    - Pulse spacing: 100ns-1s (res: 10ps)



Open Hardware

- Open Hardware advantages:
  - Get a design just the way we want it
  - Peer review
  - Design re-use
  - Healthier relationships with companies
- Open Hardware Repository (www.ohwr.org)
- CERN Open Hardware License (OHL)

Open Hardware Repository – www.ohwr.org

CERN Open Hardware License – ohwr.org/cernohl

Provides a solid legal basis

- Open Software licences not usable (GNU, GPL, . . .)
- Developed by Knowledge and Technology Transfer Group
- Defines conditions of using & modifying licenced material

Same principles as Open Software

- Anyone can see the source (design documentation)
- Anyone is free to study, modify and share
- Any modification and distribution under same licence
- Persistence makes everyone profit from improvements

Open Hardware works

Projects

- 46 active projects
  - 38 initiated by CERN groups, 8 by other institutes
- 3.6 developers on average

Types of designs

- About 30 hardware designs (of which 20 FMC projects)
- About 20 re-usable IP blocks
- General tools
  - Production test environment (Python based)
  - ADC performance test

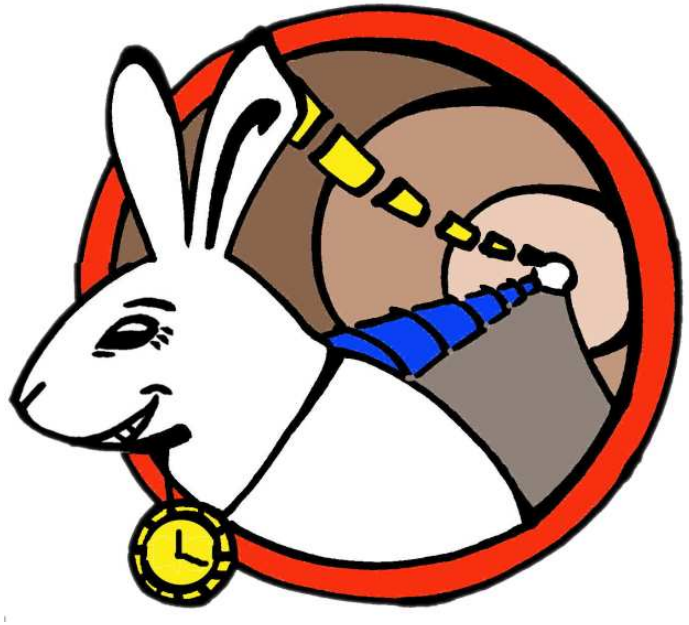
data from October 2011

Summary

- White Rabbit
  - 2000 nodes
  - $<$ 1ns accuracy
  - Tested up to 10km
- WR components available from industry
  - PCIe, VME, Switch
- Fully open design: H/W, G/W, S/W
- Large community
- Proven results
- Project still under development, open for improvements

ohwr.org/projects/white-rabbit

Questions and answers



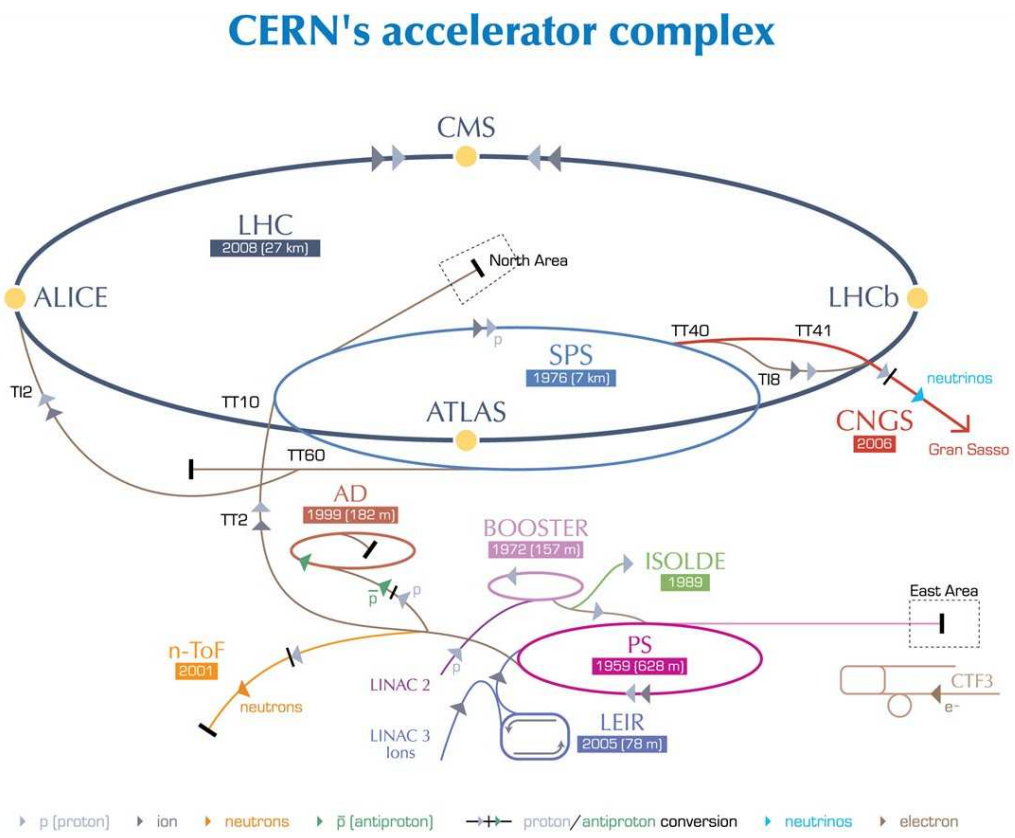
ohwr.org/projects/white-rabbit

Extras

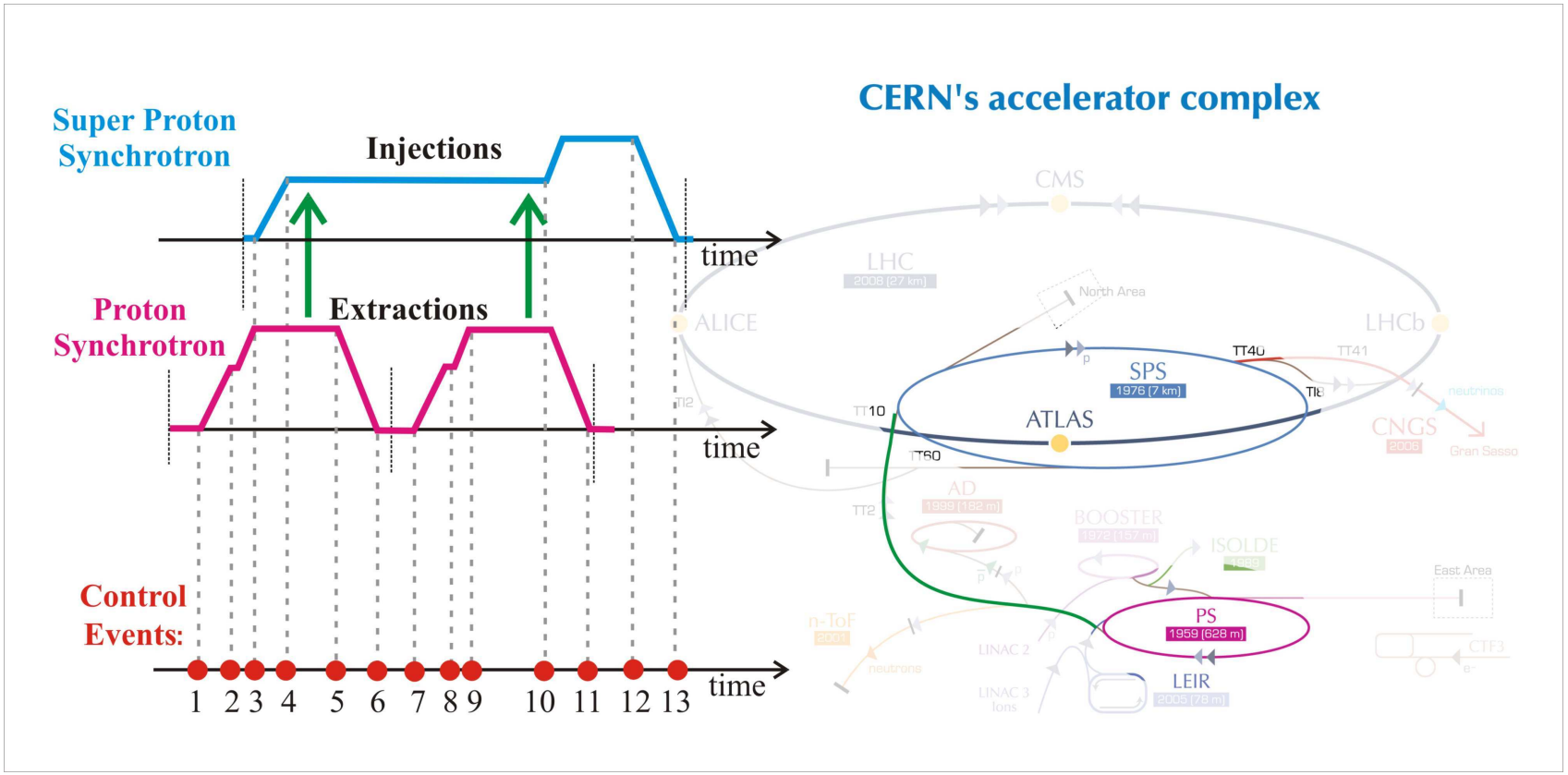
Extras

CERN

- 6 accelerators
- LHC: 27km perimeter
- Thousands of devices to be controlled and synchronized
- A huge real-time distributed system

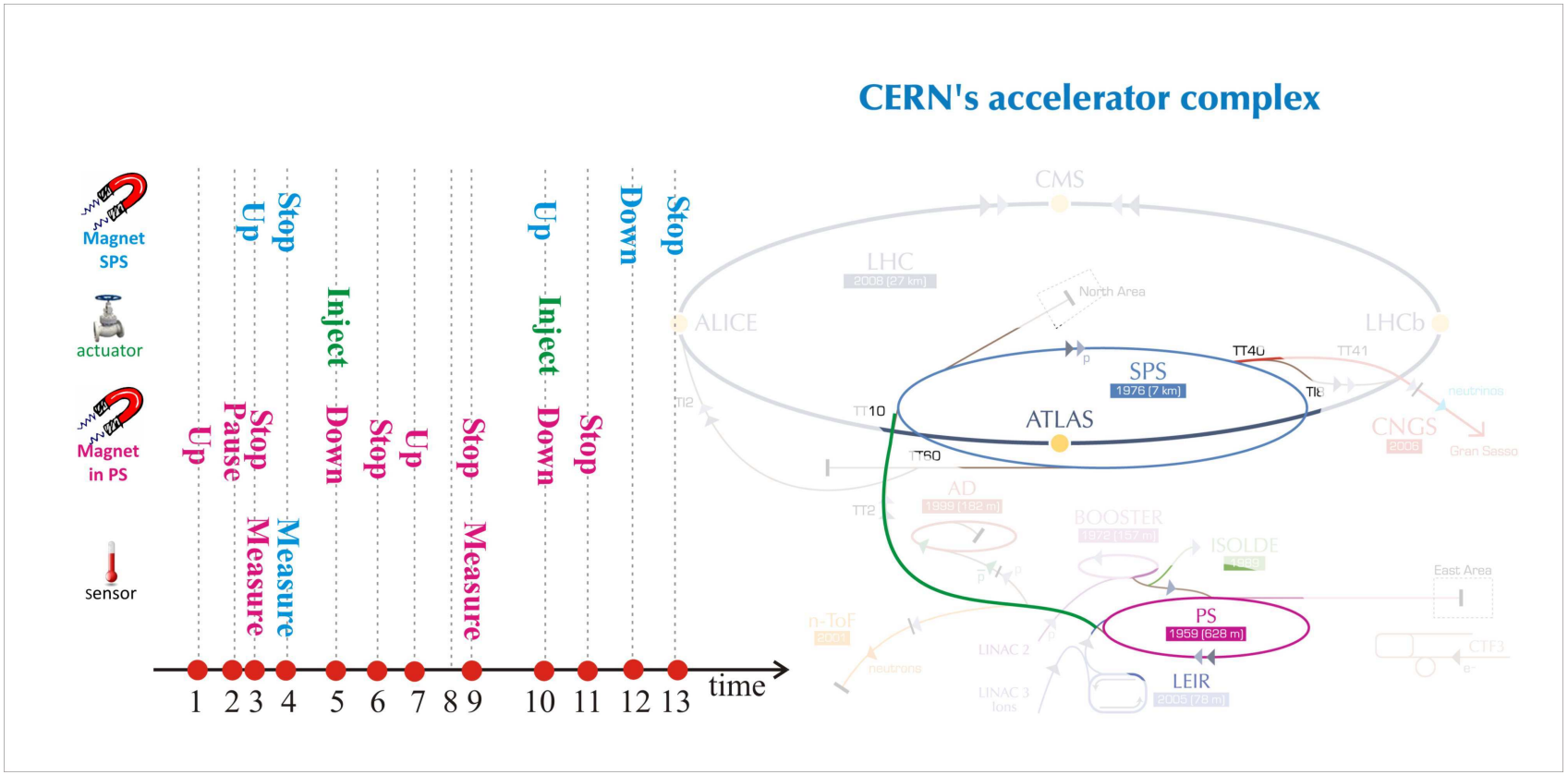


A simplified explanation of CERN control system (1)



- **Events** – points in time at which actions are triggered
- Each event is identified by an **ID**

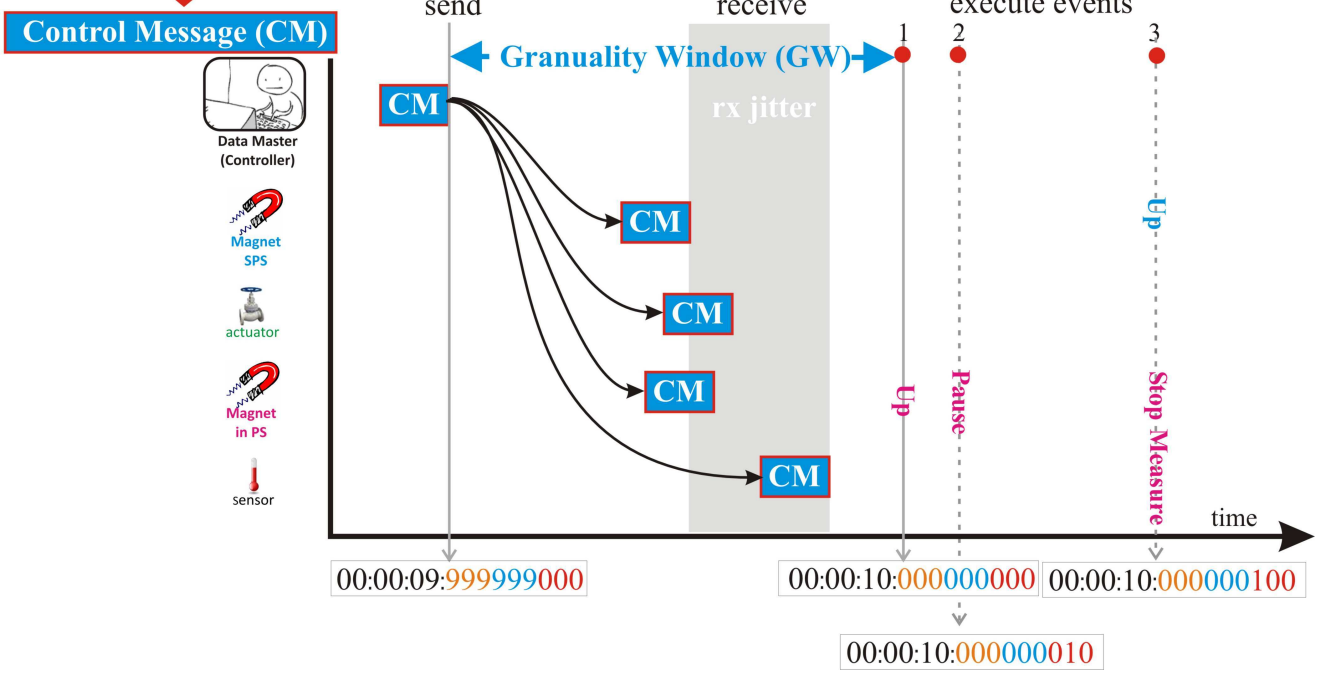
A simplified explanation of CERN control system (2)



- Devices are subscribed to events
- Each device ”knows” what to do on particular event

A simplified explanation of CERN control system (3)

Event ID	Hh:mm:ss:nanoseconds
ID = 1	00:00:10:000000000
ID = 2	00:00:10:000000010
ID = 3	00:00:10:000000100

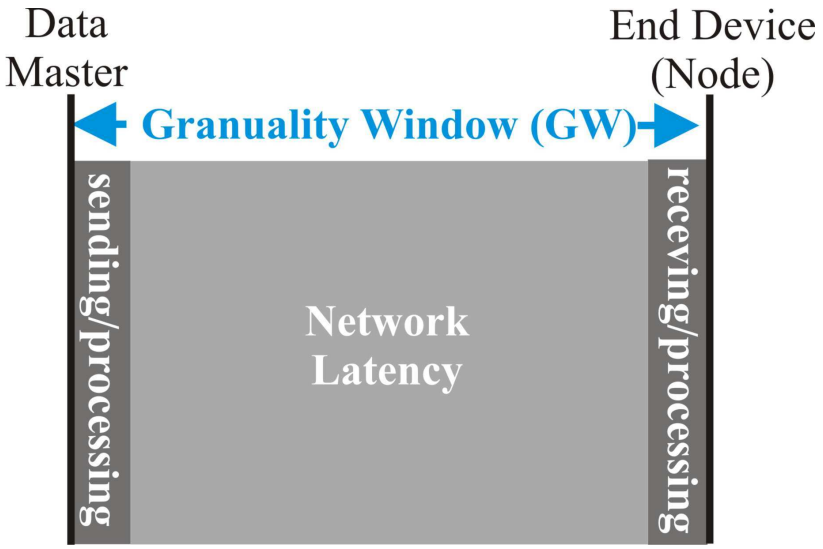


- Each event (ID) has a trigger time associated
- A set of events is sent as a single **Control Message (CM)**
- CM is broadcast to all the end devices (nodes)

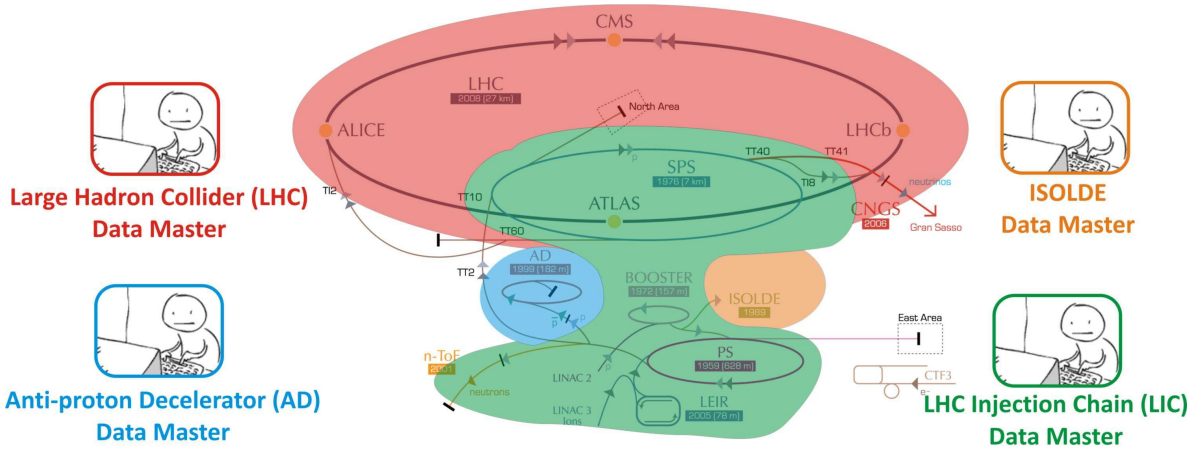
A simplified explanation of CERN control system (4)

Granuality Window:

- Controller-input to node-output (i.e. pulse)
- Maximum bound latency **guaranteed** by the system
- Processing and network latency included



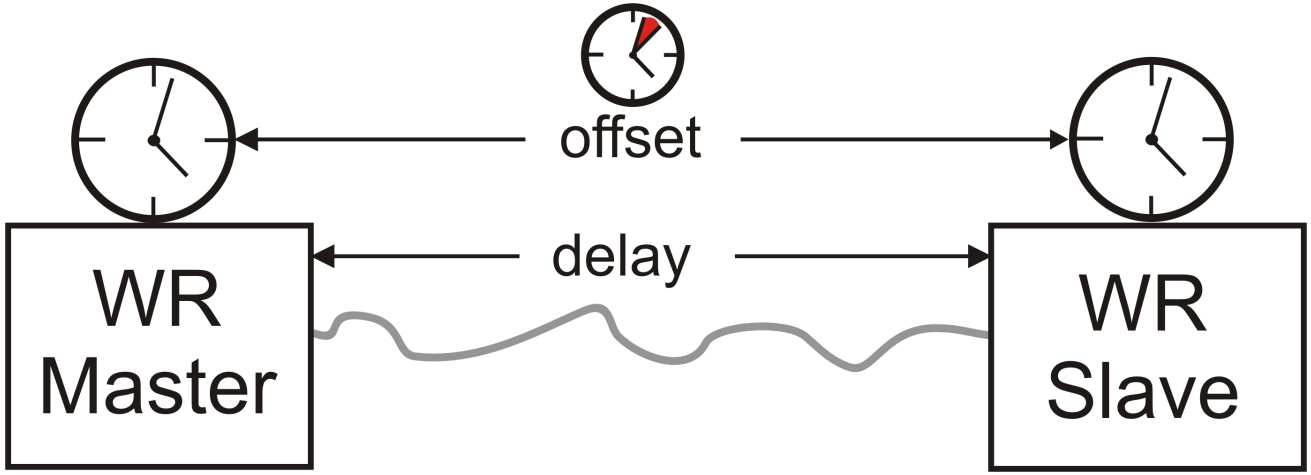
A simplified explanation of CERN control system (5)



- 4 accelerator networks
- Separate **Data Master (DM)** for each network
- LIC Data Master communicates with other DMs and control devices in their networks
- Broadcast of **Control Messages** within network(s)

Time Distribution in White Rabbit

- Synchronization with **sub-ns** accuracy over fiber
- Combination of
  - Precision Time Protocol (**PTP**) synchronization
  - Synchronous Ethernet (**SyncE**) syntonization
  - Digital Dual-Mixer Time Difference (**DDMTD**) phase detection
- WR Link:

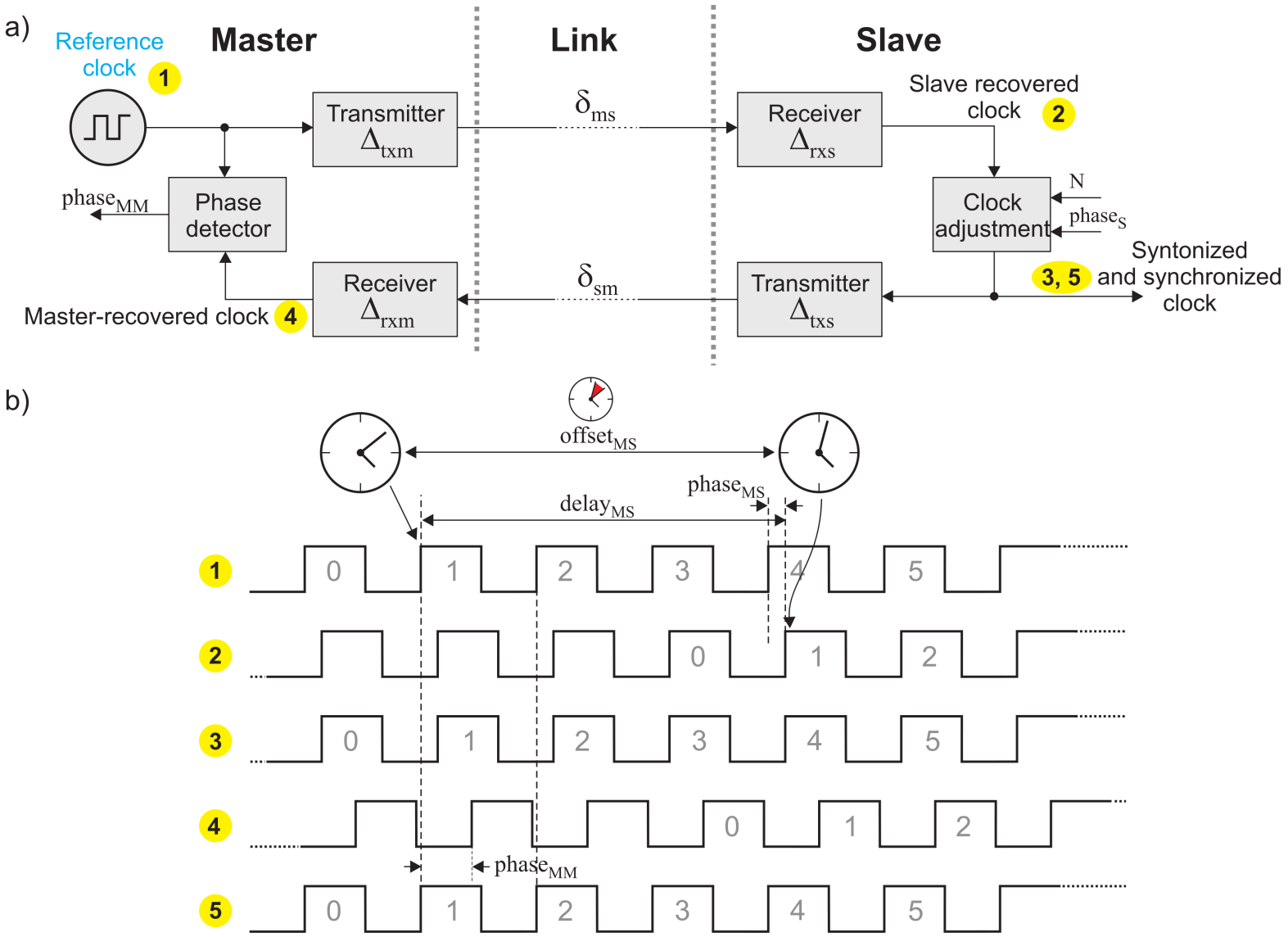


PTP is OK but ...

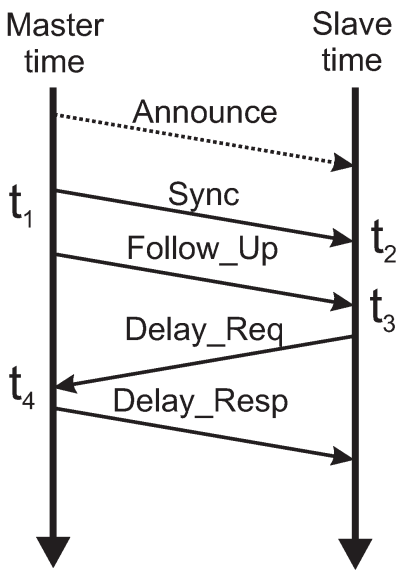
What are the issues...	**and**	**... how we address them**
PTP-base syntonization	$\Rightarrow$	SyncE
limited precision and resolution	$\Rightarrow$	SyncE DDTMD phase detection
unknown link asymmetry	$\Rightarrow$	SyncE DDTMD phase detection WR Link Delay Model

WR extension to PTP (**WRPTP**) for extra data exchange and logic

Fine Delay Measurement



Link Asymmetry



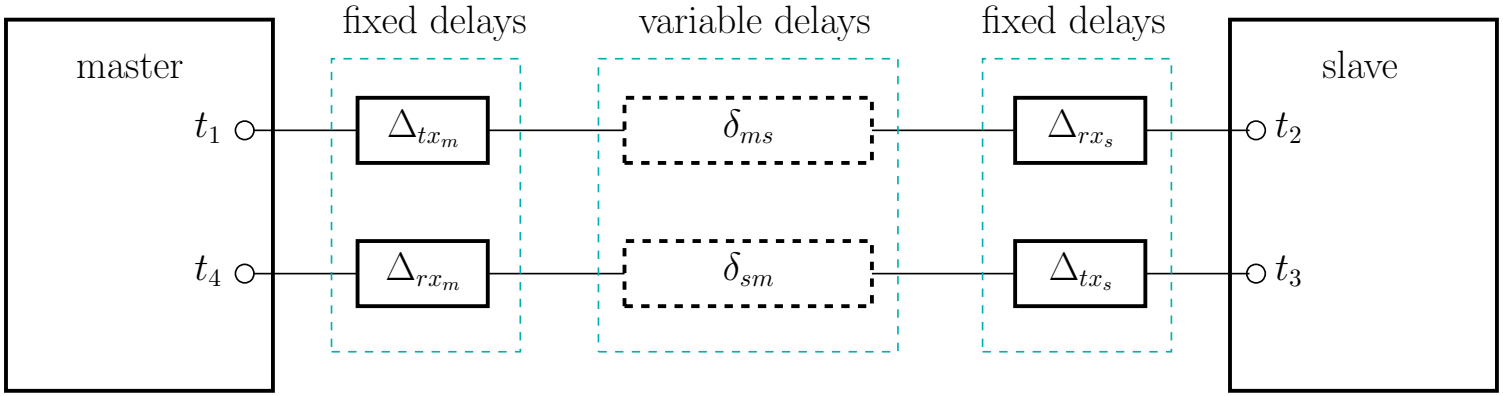
$$\mu = \frac{(t_4 - t_1) - (t_3 - t_2)}{2} \xrightarrow{\mu} ?$$

$$delay_{ms} = \mu + asymmetry$$

$$offset_{ms} = t_2 - (t_1 + delay_{ms})$$

Link Delay Model

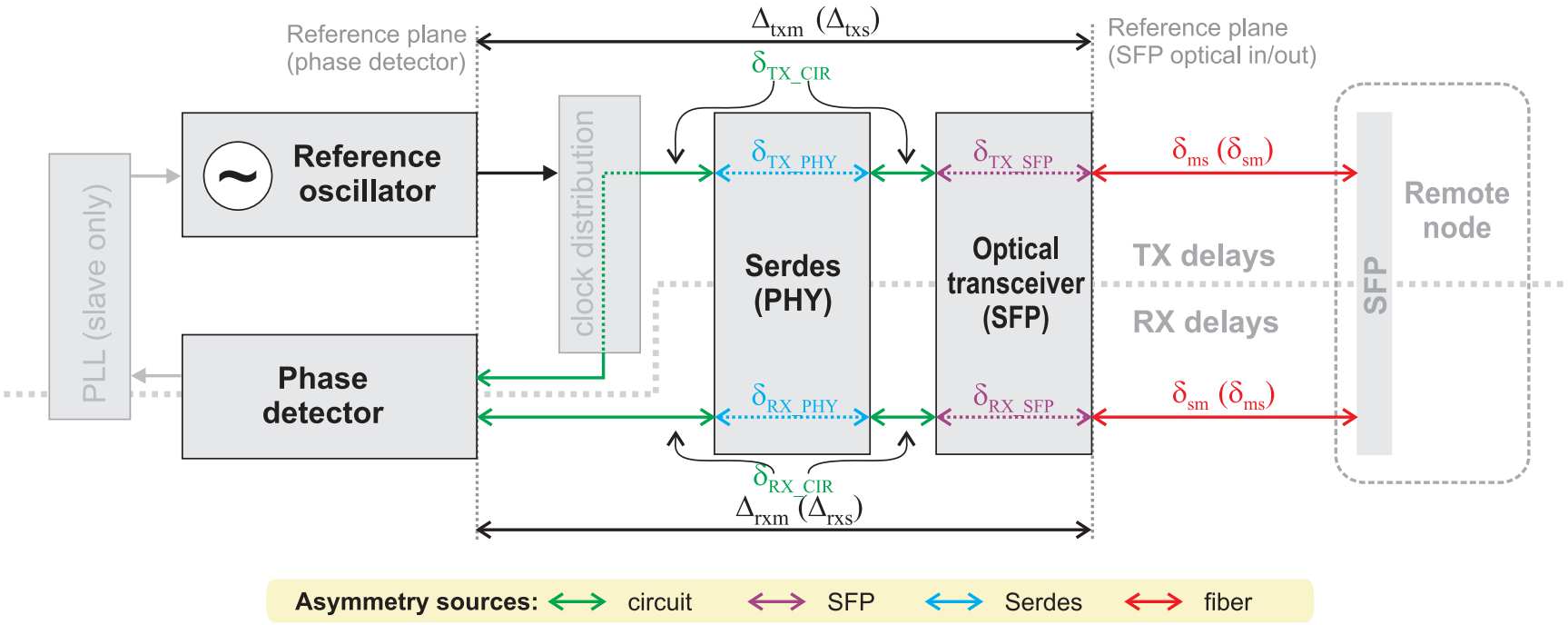
$$delay_{ms} = \Delta_{tx_m} + \delta_{ms} + \Delta_{rx_s}$$
$$delay_{sm} = \Delta_{tx_s} + \delta_{sm} + \Delta_{rx_m}$$



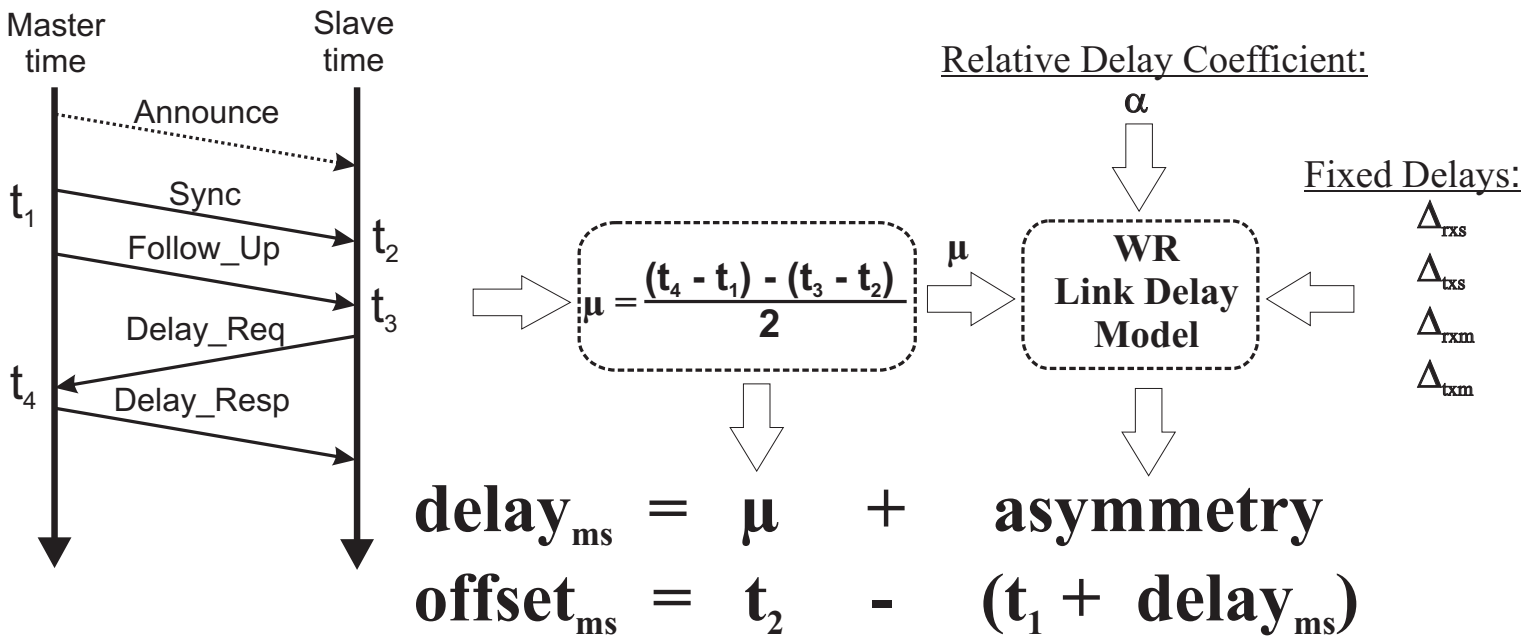
Relative Delay Coefficient (α)
for 1000BASE-BX10 over a Single-mode Optical Fiber

$$\delta_{ms} = (1 + \alpha)\,\delta_{sm}$$

Fixed Delays



Link Delay Model: fiber optic solution



Solution for Ethernet over a Single-mode Optical Fiber

$$asymmetry = \Delta_{tx_m} + \Delta_{rx_s} - \frac{\Delta - \alpha\mu + \alpha\Delta}{2 + \alpha}$$

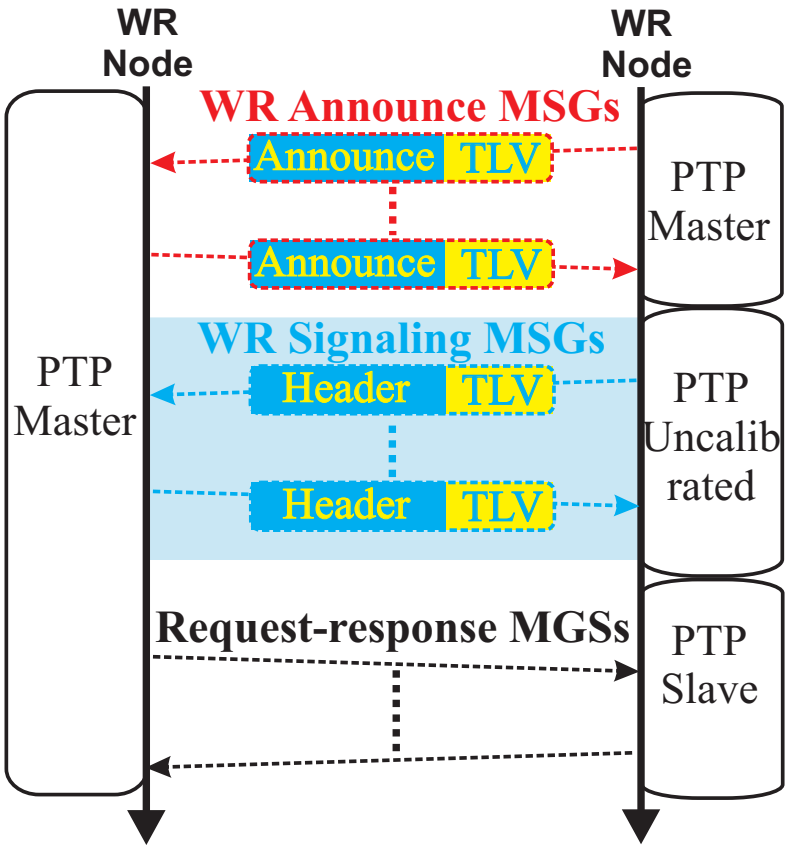
PTP and SyncE in WR

- Compatibility with PTP verified
- Compatibility with SyncE for further study
- Frequency distribution aligned with PTP's logic topology
- PTP's Announce messages used for WR-peers recognition

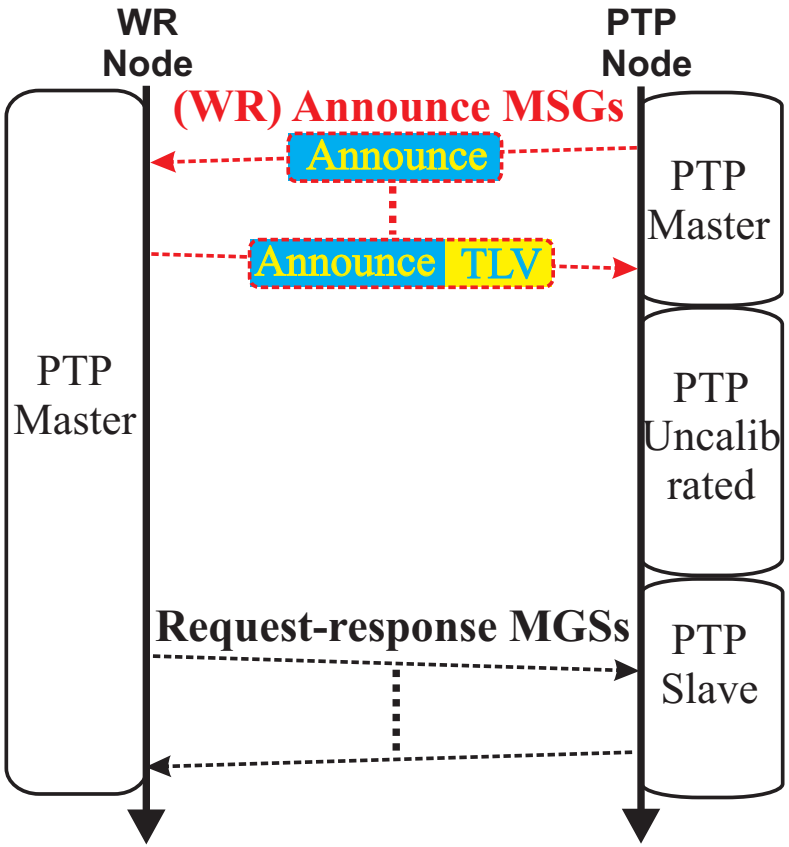
White Rabbit extension to PTP (WRPTP)

- WR-peers recognition
- Calibration (fixed delays measurement)
- Exchange of WR-data
- Support of redundancy
- Mapping over IEEE802.3/Ethernet

WR-peer recognition and WR-data exchange

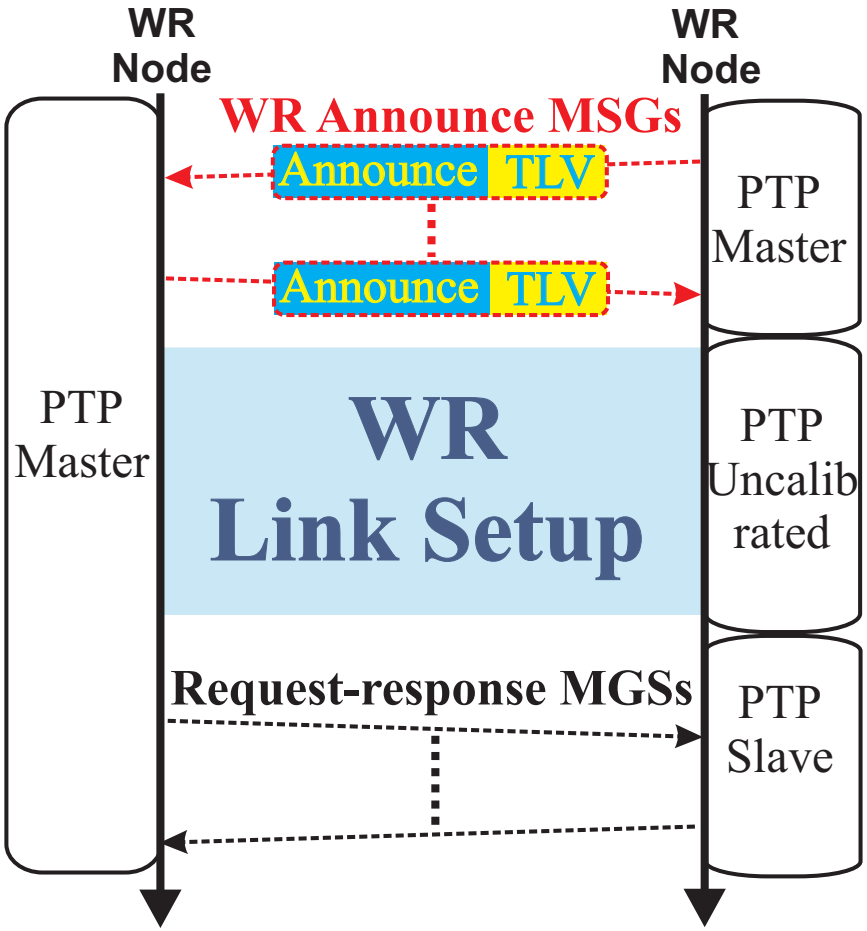


Two WR devices



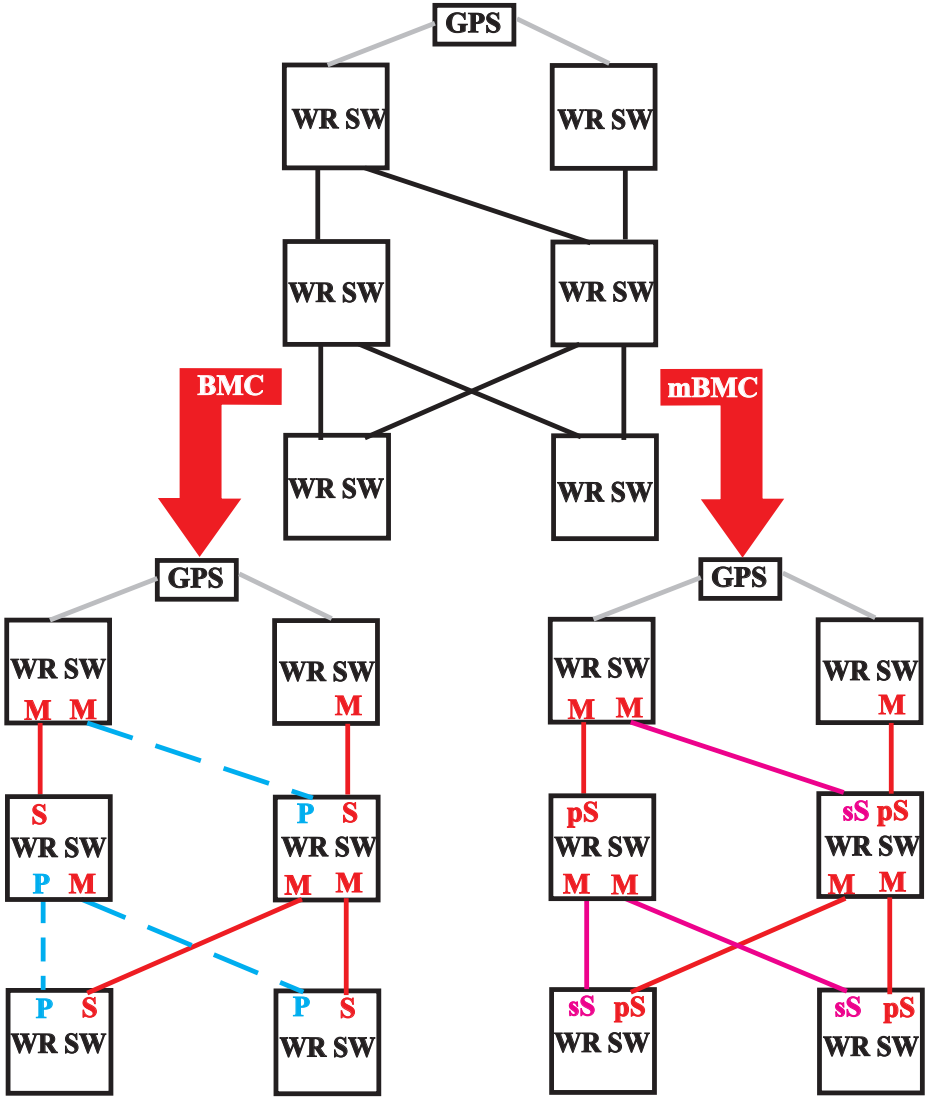
WR and non-WR device

WR Link Setup



- Frequency locking
- Calibration
- Exchange of WR-parameters
- WR Finite State Machine
- WR Signaling Messages

Modified Best Master Clock Algorithm (mBMCA)



Clock Recovery System and mBMCA