Advanced algorithms and WRE upgrade to 10 Gb/s capacity

ASTERICS GA DELIVERABLE: D5.15

Document Identifier:	ASTERICS-D5.15
Date:	March 29, 2019
Work Package:	WP5 - CLEOPATRA
Lead Partner:	University of Granada (UGR)
Document Status:	Final
Dissemination Level:	Public
Document Link:	www.asterics2020.eu/documents/ASTERICS-D5.15.pdf

Abstract

This document describes the implementation of a new White Rabbit systems that are able to work with 10 Gigabit Ethernet networks. Under this context, an upgrade of the White Rabbit technology has been provided and then, a novel fully modular and flexible architecture has been designed. As a result and for the first time in the literature, it has been enabled the utilization of the White Rabbit technology in 10 Gigabit Ethernet infrastructures with a timing synchronization performance slightly better than conventional White Rabbit devices and data services capable of utilizing almost the full 10 Gigabit Ethernet capacity. This document presents the work done for doing this implementation as well as the component used to develop such a solution.

I COPYRIGHT NOTICE

II DELIVERY SLIP

Name		**Partner/WP**	**Date**
From	J. Díaz Alonso	UGR/WP5	March 29, 2019
Author(s)	M. Jiménez López J. Díaz Alonso	UGR/WP5	March 29, 2019
Reviewed by	J. Díaz Alonso	UGR/WP5	March 29, 2019
Approved by	AMST	ASTRON/WP1	May 13, 2019

III DOCUMENT LOG

Issue	**Date**	**Comment**	**Author/Partner**
0.1	March 14, 2019	First draft of the document	M. Jiménez López/UGR
0.2	March 22, 2019	Document improvements	M. Jiménez López/UGR
0.3	March 29, 2019	Minor changes	M. Jiménez López/UGR

IV APPLICATION AREA

This document is a formal deliverable for the GA of the project, applicable to all members of the ASTERICS project, beneficiaries and third parties, as well as its collaborating projects.

V TERMINOLOGY

7S	Seven Solutions
10G	10 Gigabit Ethernet
25G	25 Gigabit Ethernet
ARM	Advanced RISC Machine
ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster
BC	Boundary Clock
BRAM	Block Random Access Memory
CERN	European Organization for Nuclear Research
CPU	Central Processing Unit
CSMA/CD	Carrier-Sensing Multiple-Access with Collision Detection
CTA	Cherenkov Telescope Array
DAC	Digital-to-Analog Converter
DAQ	Data Acquisition
DMA	Direct Access Memory
DMTD	Dual Mixer Time Difference
DDMTD	Digital Dual Mixer Time Difference
ESFRI	European Strategy Forum on Research Infrastructures
FF	Flip Flop
FPGA	Field Programmable Gate Array
GbE	Gigabit Ethernet
GM	Grandmaster
GPS	Global Positioning System
GT	Gigabit Transceiver
HAL	Hardware Abstraction Layer
IP	Intellectual Property
JNCA	Journal of Network and Computer Applications
LAN	Local Area Network
LLC	Logical Link Control
LUT	Look-Up Table
LUTRAM	Look-Up Table as Random Access Memory
MAC	Media Access Control
MAN	Metropolitan Area Network
MTIE	Maximum Time Interval Error
NIC	Network Interface Core

NTP	Network Time Protocol
OC	Ordinary Clock
OHWR	Open Hardware Repository
OS	Operating System
PCI	Peripheral Component Interconnect
PCIe	Peripheral Component Interconnect express
PCS	Physical Coding Sublayer
PHY	Physical
PI	Proportional Integral
PLL	Phase-Locked Loop
PMA	Physical Medium Attachment
PMD	Physical Medium Dependent
PPS	Pulse Per Second
PTP	Precise Time Protocol
PTPv2	Precise Time Protocol version 2
RAM	Random Access Memory
RISC	Reduced Instruction Set Computing
RS	Reconciliation Sublayer
RTS	Real Time Subsystem
SDH	Synchronous Digital Hierarchy
SerDes	Serializers-Deserializers
SFP	Small form-Factor Pluggable transceptor
SFP+	Small form-Factor Pluggable transceptor plus
SKA	Square Kilometer Array
SMA	SubMiniature type A
SoC	System-on-Chip
SOF	Start of Frame
SONET	Synchronous Optical Network
SyncE	Synchronous Ethernet
TDEV	Time Deviation
TLV	Tag-Length-Value
TRL	Technology Readiness Level
TSU	TimeStamp Unit
UGR	University of Granada
WAN	Wide Area Network
WIS	Wide Area Network Interface Sublayer
WP	Work Package
WR	White Rabbit
WRE	White Rabbit Element
WR-PTP	White Rabbit Precise Time Protocol
WRS	White Rabbit Switch
WR-Z16	White Rabbit Zynq 16 ports

A complete project glossary is provided at the following page:
http://www.asterics2020.eu/glossary/

VI PROJECT SUMMARY

ASTERICS (Astronomy ESFRI & Research Infrastructure Cluster) aims to address the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities (SKA, CTA, KM3Net & E-ELT). It brings together for the first time, the astronomy, astrophysics and particle astrophysics communities, in addition to other related research infrastructures. The major objectives of ASTERICS are to support and accelerate the implementation of the ESFRI telescopes, to enhance their performance beyond the current state-of-the-art, and to see them act as an integrated, multi-wavelength and multi-messenger facility. An important focal point is the management, processing and scientific exploitation of the huge datasets the ESFRI facilities will generate. ASTERICS will seek solutions to these problems outside of the traditional channels by directly engaging and collaborating with industry and specialised SMEs. The various ESFRI pathfinders and precursors will present the perfect proving ground for new methodologies and prototype systems. In addition, ASTERICS will enable astronomers from across the member states to have broad access to the reduced data products of the ESFRI telescopes via a seamless interface to the Virtual Observatory framework. This will massively increase the scientific impact of the telescope, and greatly encourage use (and re-use) of the data in new and novel ways, typically not foreseen in the original proposals. By demonstrating cross-facility synchronicity, and by harmonising various policy aspects, ASTERICS will realise a distributed and interoperable approach that ushers in a new multi-messenger era for astronomy. Through an active dissemination programme, including direct engagement with all relevant stakeholders, and via the development of citizen scientist mass participation experiments, ASTERICS has the ambition to be a flagship for the scientific, industrial and societal impact ESFRI projects can deliver.

VII EXECUTIVE SUMMARY

This report provides a complete description regarding the implementation of a new solution capable of deploying White Rabbit (WR) technology using 10 Gigabit Ethernet (10G) networks. This work has been performed within the Work Package (WP) 5.1 of the ASTERICS project. The WP 5.1, Time Synchronization, is centered around the White Rabbit Element (WRE) technology.

There are many scientific applications such as distributed telescopes which require high accuracy timing synchronization and, at the same time, high data bandwidth delivery mechanisms. Consequently, WR is presented as the main cornerstone for the timing synchronization system thanks to its sub-nanosecond accuracy. However, WR was designed to work in 1 Gigabit Ethernet (GbE) infrastructures and it cannot be easily deployed in 10G

networks without an intensive adaptation effort.

In this deliverable, a new WRE solution has been implemented to be integrated in many applications that demand high data bandwidth capabilities. Under this context, a full description of the solution architecture is presented in terms of hardware, gateware and software. In addition to that, some experiments have been performed in order to characterize the system in relation to timing synchronization accuracy, frequency distribution quality, data bandwidth and latency.

VIII SCIENTIFIC DISSEMINATION

The content of the current document has been submitted as a scientific article in the Journal of Network and Computer Applications (JNCA) of Elsevier publisher. The specific details regarding this contribution are:

M. Jiménez-López, F. Girela-Lopez, J. López-Jiménez, E. Marín-López, R. Rodríguez, J. Díaz, "10 Gigabit White Rabbit: a high bandwidth subnanosecond synchronization solution for data and timing distribution", Journal of Network and Computer Applications.

Table of Contents

1 Introduction

This section exposes an introduction about this document detailing its motivation and scope, objectives and report structure. Tasks here described have been performed for the University of Granada (UGR) team for the Astronomy ESFRI & Research Infrastructure Cluster (ASTERICS) project. Under this context, the results exposed in this report are also available in the PhD dissertation [1] performed by Miguel Jiménez López and in a scientific contribution in JNCA.

1.1 Motivation and Scope

Nowadays, there are many scientific applications that deploy networks where data bandwidth is an important topic to be considered, especially for Data Acquisition (DAQ) systems. In addition to the data delivery service, another key requirement is the high accurate timing synchronization that allows the coordination between different nodes of the network. Under this framework, there are some standard alternatives such as Precise Time Protocol (PTP) and Network Time Protocol (NTP) that can be easily deployed in conventional networks. However, these approaches are not able to provide the timing synchronization accuracy demanded by some applications such as distributed telescopes, for instance Square Kilometer Array (SKA) [2] or Cherenkov Telescope Array (CTA) [3]. Consequently, more sophisticated solutions must be implemented in order to deal with system needs. In this line, White Rabbit (WR) is a very interesting candidate that can be used in optical fiber networks with a sub-nanosecond range accuracy [4]. Nevertheless, WR was originally designed to work only in 1 Gigabit Ethernet (GbE) networks whose data capacity is not enough to fulfil requirements for data delivery services of the DAQ system.

This document presents a new approach for implementing White Rabbit Element (WRE) using high data bandwidth technologies such as 10 Gigabit Ethernet (10G). This solution has been designed to be fully modular and, therefore, it can be adopted in many applications.

1.2 Document objective

This document exposes the specific implementation of WRE system using 10G technology. Design decisions have been explained and properly justified comparing the proposed solution with the current WRE alternatives. This comparison helps to put the emphasis on the advantages of 10G WRE solution in terms of flexibility, modularity and data distribution capacity.

1.3 Structure

This report has been organized in several sections that are described briefly in the following lines:

- **Section 2, White Rabbit overview.** It presents the WR technology and its

implementation details.

- **Section 3, 10 Gigabit Ethernet.** It exposes the main standard versions for the 10G technology.
- **Section 4, 10G solution with WR synchronization.** It describes the WRE solution for 10G networks.
- **Section 5, Experiments & Validation.** It presents a set of experiments that has been designed to measure the performance in terms of timing synchronization accuracy, frequency dissemination, data delivery bandwidth and endpoint latency.
- **Section 6, Conclusions.** It draws the main conclusions in relation to the developed work in this deliverable.

2 White Rabbit overview

WR [5,6] is a synchronization technology that has been created within the framework of an open source and international collaborative project managed by the European Organization for Nuclear Research (CERN) in 2009. The WR designs are hosted in the Open Hardware Repository (OHWR) that contains more than 100 open source projects.

WR technology has been originally designed to be deployed in optical fiber networks that use 1 GbE as physical layer technology. It is able to provide a sub-nanosecond accuracy for timing synchronization and a picosecond range precision for networks connecting thousands of nodes with up to 10 km fiber links. In order to perform this, WR implements three different mechanisms that are briefly discussed in Table. 1.

Mechanism	Description
Frequency synchronization (syntonization)	WR uses a technology similar to Synchronous Ethernet (Sync-E) that is responsible for recovering the reference clock to the network and uses it to adjust the local oscillator by means of Digital-to-Analog Converter (DAC) devices.
Enhanced Precise Time Protocol version 2 (PTPv2)	WR enhances the PTPv2 protocol including its own Type-Length-Value (TLV) definitions, calibration parameters and asymmetry factor. For the rest of the document, this protocol is referred as White Rabbit Precise Time Protocol (WR-PTP).
Fine phase measurement	WR implements Digital Dual Mixer Time Difference (DDMTD) blocks in the gateware design that are in charge of measuring the phase difference between several clocks. Thanks to that information, WR is able to improve the timestamp resolution overcoming the cycle limitation.

Table 1. WR timing synchronization mechanisms.

The first mechanism is known as frequency synchronization (syntonization) whose main goal is to lock the local clock to a reference one coming from the network. For this task, a control servo loop based on a Phase-Locked Loop (PLL) has been implemented and is responsible for adjusting the local oscillator by means of DAC chips. In addition to the syntonization, WR enhances the PTP version 2 including new TLV definitions through creation of new signaling messages. WR-PTP also takes into consideration calibration parameters of different platforms together with the asymmetry factor of the fiber link in order to provide a deterministic behavior

of the Pulse Per Second (PPS) output signal. Under this context, the new revision of PTP for high accuracy profile will be intensively based on the WR implementation. The third mechanism for the WR timing synchronization is the fine phase measurement. In order to perform this, WR includes in the gateware level several IP blocks known as DDMTD ones. They are able to measure the phase difference between two clocks with a precision in the range of femtoseconds. This phase information is used to improve timestamp resolution and, consequently, it enhances the timing synchronization performance.

Regarding network configuration, WR follows a network with tree topology (Fig. 1) where different kinds of elements are placed. The root device is also known as Grandmaster (GM) one which is responsible for obtaining a very accurate timing reference normally coming from Global Positioning System (GPS) or atomic clock. Then, GM element synchronizes its local clock and disseminates its reference to the rest of the network. The intermediate levels are composed of WR switches that act as PTP Boundary Clock (BC) devices locking reference clock from the uplink and distribute it to the rest of network layers using their downlinks. Finally, the end nodes act as PTP Ordinary Clock (OC) devices that recover the reference clock from the network and adjust their local oscillators. In addition to the synchronization process, end nodes normally implement user-defined tasks that depend on the specific application for the WR network.

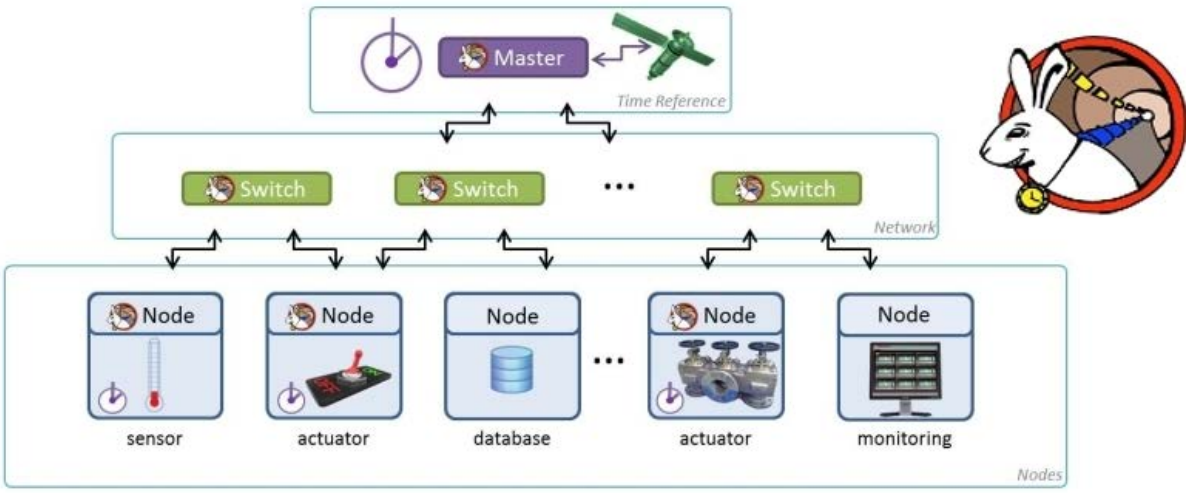

Fig 1. WR network topology. This picture has been taken from [5].

As discussed previously, WR was designed to work in 1 GbE networks. These kinds of networks present incompatibilities in the physical layer implementation in relation to the higher data bandwidth standards such as 10G. This issue represents a significant drawback in terms of interoperability and prevents the utilization of a single network for data and timing synchronization distribution. With the current WR solutions, a separated network must be deployed for timing synchronization purposes while other high data bandwidth network is provided for data delivery increasing significantly system costs and complexity. In order to

solve this issue, a WR technology update has been developed to allow its integration in 10G networks.

3 10 Gigabit Ethernet

The 10G technology was included in the Ethernet standard (IEEE 802.3ae) in 2002 [7-9]. This technology was defined with a speed of 10 Gb/s and can be deployed in different classes of networks such as Local Area Network (LAN), Metropolitan Area Network (MAN) and Wide Area Network (WAN). This section is focused on 10G technology for optical fibers, albeit there are some configurations that can work in copper links. Fig. 2 shows the network stack model of 10G technology with its different layers:

- **Logical Link Control (LLC) and Media Access Control (MAC).** These layers are responsible for implementing MAC capabilities. In this line, 10G standard shares the same Ethernet format, minimum and maximum frame sizes than lower speed Ethernet standards. However, an important difference between 10G and 1 GbE is that the former does not require to implement the Carrier-Sensing Multiple-Access with Collision Detection (CSMA/CD) protocol because it only allows the full-duplex mode.
- **Reconciliation Sublayer (RS).** This converts data from the MAC format into the Physical (PHY) one.
- **PHY.** This layer is composed of four sublayers.
    - **Physical Coding Sublayer (PCS)**. It defines the encoding scheme for data inside the PHY layer.
    - **Wide Area Network Interface Sublayer (WIS)**. It provides Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) encapsulation mechanisms for WAN.
    - **Physical Medium Attachment (PMA) sublayer**. It contains Serializers-Deserializers (SerDes) modules for converting serial data into parallel one and vice versa.
    - **Physical Medium Dependent (PMD) sublayer**. It implements the physical connection and signaling to the media using optical transceivers in optical fiber networks.

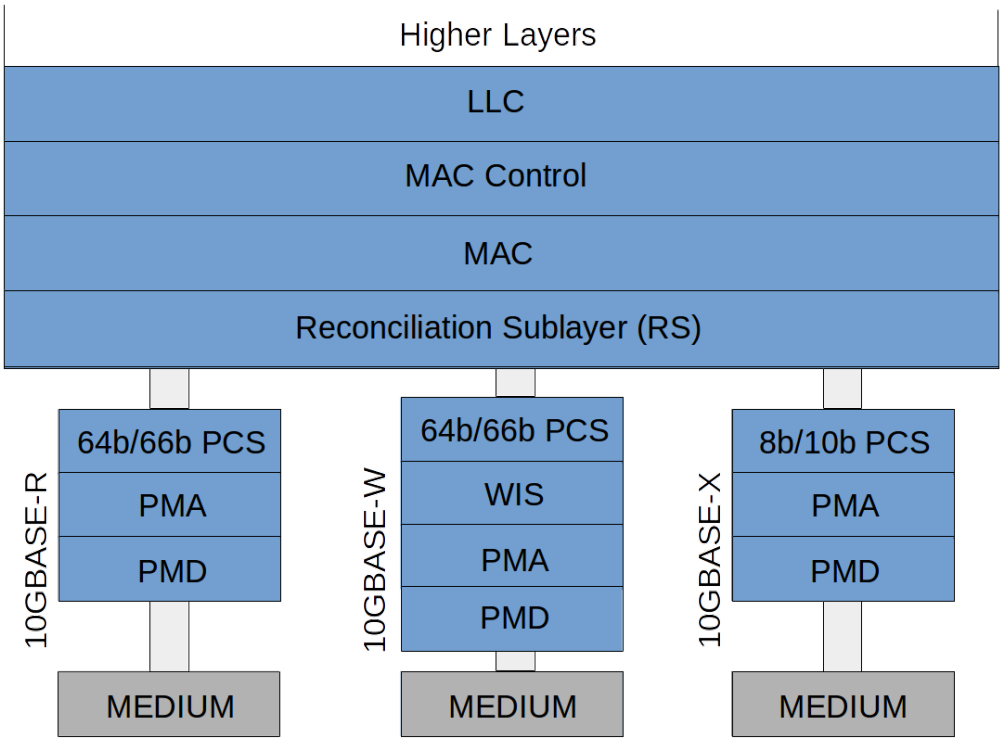

Fig 2. 10G standard layers. This picture has been taken from [1].

As shown in Fig.2, the 10G PHY layer can be implemented using different alternatives. These can be classified into three different types that are summarized in Table. 2.

PHY implementation	PCS encoding	Lines	Frequency line	WAN
10GBASE-R	64b/66b	1	10.3125 GHz	No
10GBASE-W	64b/66b	1	9.58464 GHz	Yes
10GBASE-X	8b/10b	4	3.125 GHz	No

Table 2. 10G PHY implementations.

As a result of the 10G standard study, 10GBASE-R is the most interesting alternative for the WRE solution in 10G networks. The main reasons are:

- It is a serial line standard whose data stream can reach 10 GHz.
- It presents less overhead than the 10GBASE-X version due to the PCS encoding scheme.
- It does not require special hardware to join several data stream into one as in the 10GBASE-X case.

4 10G solution with WR synchronization

This section describes the WRE solution for 10G networks in terms of hardware, gateware and software architectures.

4.1 Hardware

To implement the WRE for 10G system, the White Rabbit Zynq 16 ports (WR-Z16) board (Fig. 3) has been chosen. This platform has been designed by Seven Solutions (7S) and includes a Xilinx Zynq SoC (XC7Z035) that contains a Field Programmable Gate Array (FPGA) device and an Advanced RISC Machine (ARM) processor inside the same chip. Thanks to the Zynq System-on-Chip (SoC), the WR-Z16 becomes into a complete standalone platform where hardware accelerators can be developed for critical tasks and, at the same time, advanced software capabilities can be executed using the hard processor by means of an Operating System (OS) environment. The WR-Z16 board has 16 10G Small form-Factor Pluggable transceptor plus (SFP+) ports, a custom clocking circuitry and some SubMiniature type A (SMA) sockets for PPS, 10 MHz input clock and 10 MHz output clock.

Fig 3. WR-Z16 platform.

4.2 Gateware

The FPGA design configuration is also known as gateware and comprises several Intellectual Property (IP) blocks connected between them to implement a specific set of tasks. This gateware presents some differences with regard to current WRE solutions for 1 GbE networks (Fig. 4).

Current 1 GbE solutions have been implemented using open source components from OHWR, providing the source codes for each component in the architecture. However, this implementation has not been designed thinking in modularity and flexibility. Therefore, it is not an easy task to exchange some components such as network path ones for other third-party IP ones more suitable for specific applications. For this reason, instead of following an

incremental design based on open hardware modules, a new solution focused on interoperability has been developed. In contrast to this approach, the WRE solution for 10G technology provides a fully modular and flexible architecture whose main cores can be easily updated or replaced. It easies the integration of other third-party IP cores according to the specific application needs. A more detailed description of the FPGA design for 10G infrastructures is presented below:

- **Central Processing Unit (CPU):** It is the brain of the design that is able to execute an OS together with all the software tools. It is referred to as the ARM core.
- **Random Access Memory (RAM):** It is the main memory of the platform that stores data and code information required by processor activity.
- **Direct Memory Access (DMA) module:** It is in charge of transferring packets between the network and the CPU. This avoids the CPU intervention for each packet improving the system data bandwidth releasing CPU resources for important tasks.
- **10G Endpoint:** It contains the MAC and PCS cores for the 10G standard implementation. In the WRE solution for the 10G networks, Xilinx components have been used for MAC [10] and PCS [11]. However, the architecture is flexible enough to use other alternatives without requiring important modifications.
- **Gigabit Transceiver (GT):** It is the physical primitive that contains the SerDes and physical components for the low-level 10G PHY layer.
- **Timing IP:** It contains the timing synchronization related logic, that is, the TimeStamp Unit (TSU) cores and the Timing controller.
    - *TSU modules:* They are able to detect the Start of Frame (SOF) events from the SerDes data interface. It is a requirement in order to guarantee a deterministic packet latency due to some network blocks (MAC and PCS) can expose variable delay. Therefore, if this effect is not considered, time synchronization performance can be degraded. It includes third-party modules used to guarantee the deterministic behaviour of the SOF timestamping operation.
    - *Timing controller:* It is in charge of implementing the needed logic for WR protocol and is composed of similar blocks to the Real-Time Subsystem (RTS) used in the FPGA gateware of the White Rabbit Switch (WRS).
- **Crossbar core:** It allows the processor to access the memory-mapped registers of all the components of the design.

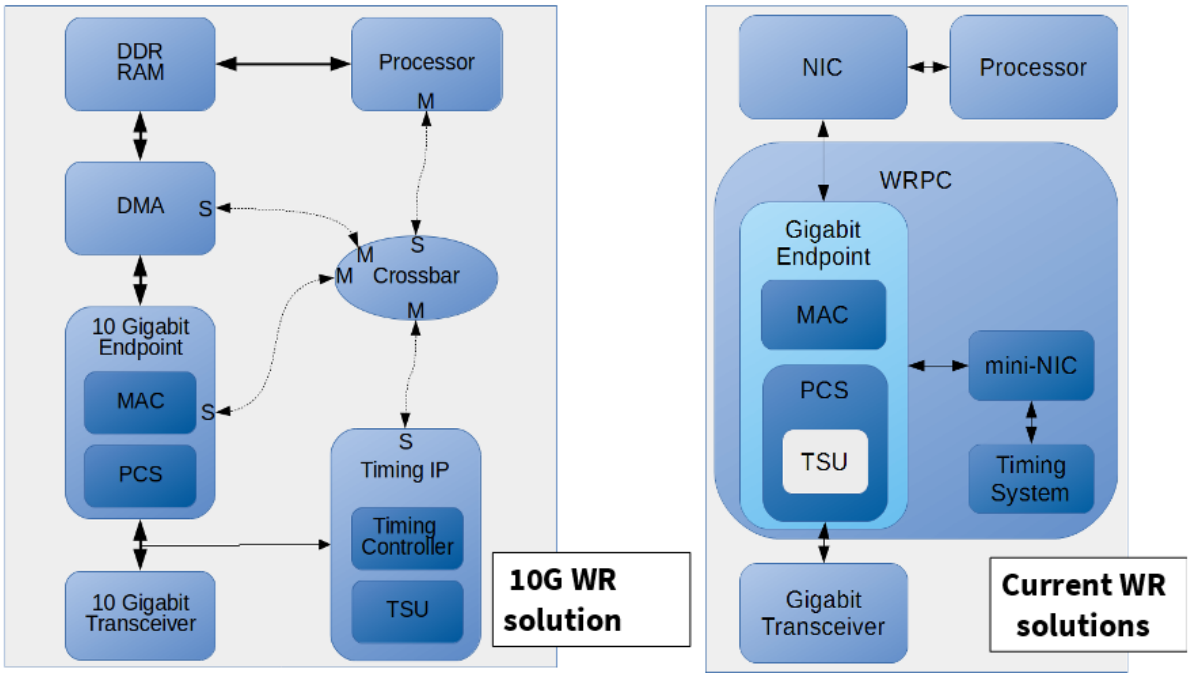

Fig 4. WRE solutions for 10G and 1 GbE networks. This picture has been taken from [1].

4.3 Software

In addition to the hardware and gateware descriptions for the WRE solution, it is necessary to define a software framework. This framework uses Linux OS for embedded systems and it provides different software components. The software components are drawn in Fig. 5 and can be classified into different categories:

- **Userspace daemons:** They comprise several user applications which manage the platform through system calls.
  - *Hardware Abstraction Layer (HAL) daemon:* It is provided with the WR-Z16 board by the company vendor. It handles hardware components of the platform and IP blocks inside FPGA. It acts as a proxy for any application that wants to access these resources ensuring safe mechanisms at the time to access to the critical components of the platform.
  - *Timing daemon:* It implements an enhanced version of the WR-PTP stack giving to the entire system a sub-nanosecond synchronization. In contrast to the current WRE solutions for 1 GbE networks, it is executed by the ARM processor and not by soft-processor.
- **Kernel modules:** They are specific source codes that implement hardware drivers for different components inside the Linux kernel. Under this context, custom drivers have been developed to control the Timing IP core and the network endpoint source code has been adapted from the Xilinx repository to work properly with high accurate timestamps.
- **Phase-control servo loop firmware:** It is an embedded software that runs in a soft-

processor inside the Timing IP core. It implements a Proportional Integral (PI) control servo loop whose main responsibility is the frequency adjustment by means of DAC chips.

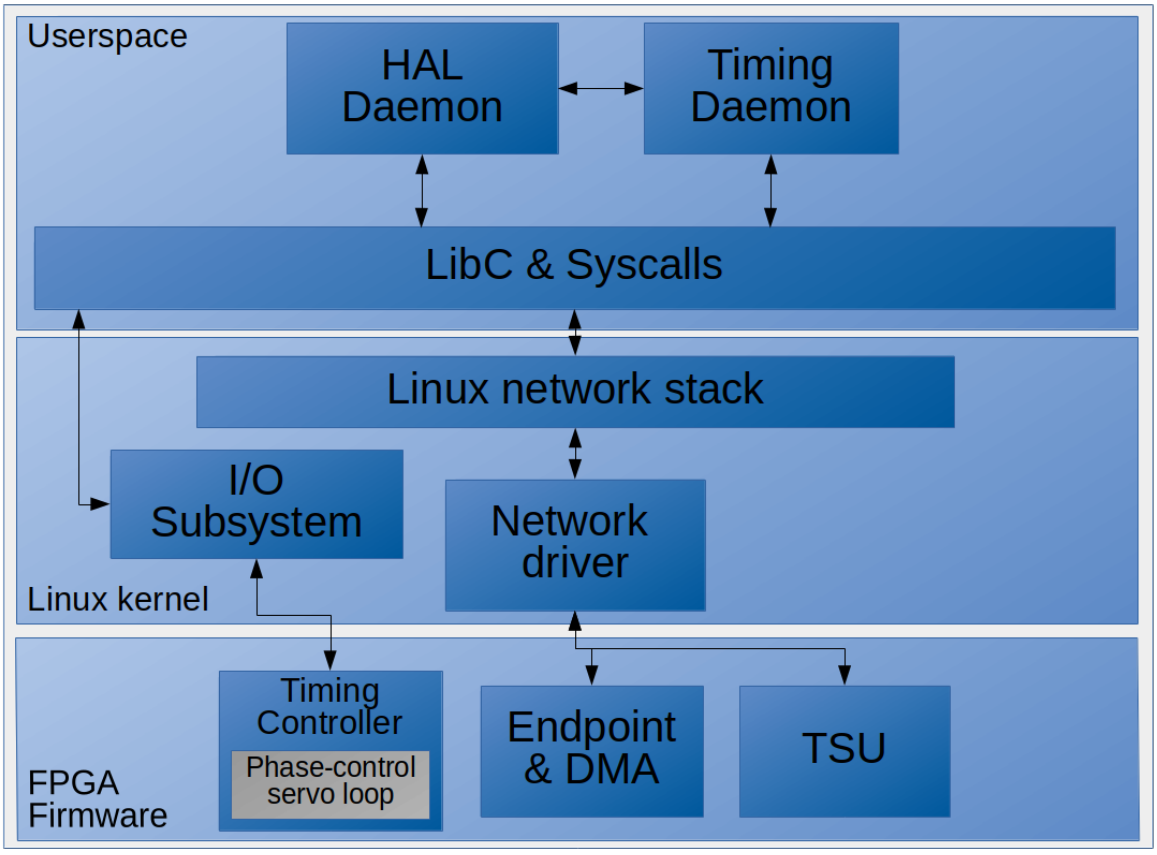

Fig 5. WR-Z16 software. This picture has been taken from [1].

The proposed solution has required different design cycles in order to get a successful implementation. The non-deterministic behaviour of the endpoint has required the investment of a significant amount of engineering resources as well as the final utilization of IP blocks from third parties. The proposed system has been developed from scratch taking into consideration the specifications of the WR protocol and the modifications needed by the 10G technology but not following an incremental design from the OHWR. This design is fully modular and its network elements such as MAC and PCS cores can be easily replaced by existing industrial third-party IP cores. For compatibility reasons, the WR-PTP stack has been adapted to work with 10G networks instead of coding a custom one. Consequently, a solution with high reliability and interoperability is achieved maintaining the compatibility with the original stack from OHWR.

5 Experiments & Validation

This section presents a set of tasks performed to characterize the proposed solution for WRE in 10G infrastructures. Firstly, a resource consumption and CPU impact studies are presented. Then, some experiments have been defined in order to evaluate the system performance taking into consideration the timing accuracy, frequency dissemination, data bandwidth and latency.

5.1 Equipment for experiments

In this section, the required equipment is presented in the following lines:

- 2 x WR-Z16 boards
- 1 x Microsemi 3120A for phase noise analysis
- 1 x Keysight 53230A for long term synchronization stability analysis
- 1 x Morion MV89 for GM mode
- 2 x Avago AFBR-709SMZ SFP+ for 10G links
- 2 x AXCEN 1310/1490 nm Small form-Factor Pluggable transceiver (SFP) for 1 GbE links
- G.652D optical fiber strands of different lengths
- 1 x Endace DAG 10X2-S 10G network interface card
- 1 x Solarflare Communications SFC9120 network interface card
- 1 x Arista DCS-7150S-24-R switch
- 2 x Workstations with Linux OS and Peripheral Component Interconnect express (PCIe) slots

5.2 Resource characterization

A complete study regarding the resource consumption has been performed comparing the 10G WRE solution with the current WR implementations using a WR-Z16 platform. Under this context, the main results are shown in Table. 3 which demonstrates that the 10G solution requires more hardware resources than current 1 GbE WR implementations.

Design	LUT	LUTRAM	FF	BRAM	DSP	MMCM
10G	19398	836	26762	47.5	3	3
1 GbE	6157	166	6513	39.5	3	2
Total	171900	70400	343800	500	900	8

Table 3. FPGA resource utilization for 1 GbE and 10G in the WR-Z16 platform.

Moreover, a deeper analysis has been accomplished extracting resource utilization by block. The blocks that have been defined to be included in this analysis are: MAC, PCS, Network

Interface Core (NIC)/DMA and TSU cores. Results have been represented by means of Fig. 6. The main considerations are briefly described below:

- MAC block presents a more complex design in 10G design than in current 1 GbE ones. Therefore, it demands more resources in terms of Look-Up Table (LUT), Look-Up Table as Random Access Memory (LUTRAM) and Flip Flop (FF) ones.
- PCS core requires more LUT, LUTRAM and FF resources in 10G solution due to the inclusion of new elements in the datapath and the difference in the width of the data bus.
- In current WR solutions, a very basic NIC has been implemented. It has not been designed to provide high data bandwidth services. Consequently, the resource consumption is significantly lower than 10G DMA block. This DMA module is able to provide high data bandwidth transfer capabilities demanding many Block Random Access Memory (BRAM) resources for packet buffering purposes.
- TSU core in 10G solution must implement some SOF recognition logic because it acts as sniffer in order not to interfere in the network elements. Consequently, it requires more LUT, LUTRAM and FF blocks.

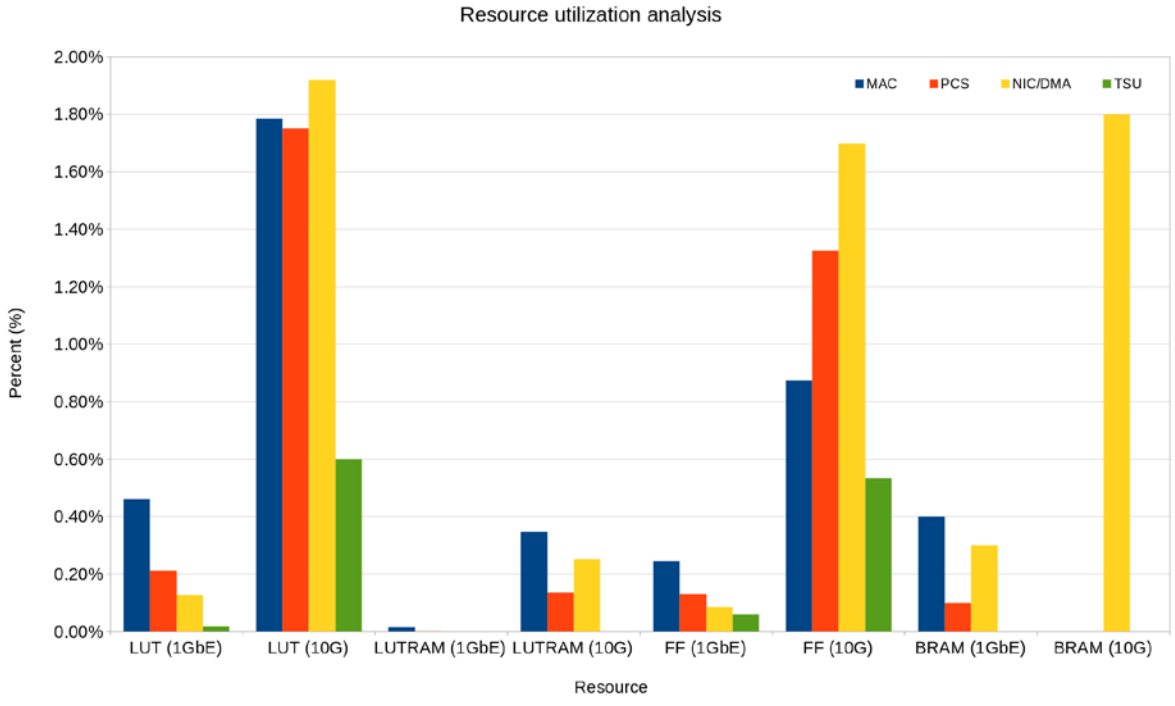

Fig 6. FPGA resource consumption by component.

Once analyzed the resource consumption of 10G solution, a CPU load study has been performed in order to ensure that the main CPU activity does not degrade the timing synchronization quality. For old WR architecture, the entire WR protocol (WR-PTP stack and phase servo control loop algorithm) was implemented using a soft-processor. This fact allows

the utilization of the main CPU for user applications exclusively avoiding the interference between them and WR synchronization mechanisms. However, in the 10G system, the WR-PTP stack is implemented in the main CPU processor that is shared for user applications. In this new scenario, the CPU activity could degrade the synchronization performance. To check this assumption, a CPU load experiment has been designed. The first step is to measure the CPU load consumption by WR-PTP daemon in order to obtain its CPU utilization requirements. Results show that WR-PTP requires less than 1% of the total CPU capacity so it does not impose high constraints at this point. Then, the synchronization accuracy (PPS offset between two WR-Z16 boards) has been measured in different CPU load conditions using *stress* and *cpulimit* tools. Table. 4 summarizes the main results of this experiment and demonstrates that synchronization quality is not affected by CPU load conditions as a result of user applications activity.

CPU Load (%)	PPS offset (ps)	
	Mean	Stdev
0	34.421	11.450
25	34.045	13.317
50	35.852	12.906
75	32.991	13.170
100	31.523	13.254

Table 4. CPU load impact study results.

5.3 Timing experiments

This section presents the frequency dissemination experiments together with the long-term synchronization stability analysis.

5.3.1 Frequency dissemination

The frequency dissemination is a very important topic for many applications both in scientific and industrial domains. Consequently, a study regarding the topics that impact the dissemination quality is a crucial factor. In the WR solutions both 10G and 1 GbE implementations, a servo control loop process is implemented to adjust the frequency of the local oscillators. Internally, it is based on a PI control algorithm with proportional (k_p) and integrator (k_i) parameters. These parameters can be optimized in order to obtain the best performance in terms of frequency dissemination. Therefore, in this section, results of the complete analysis of the influence of different parameter values are presented.

The experiment setup is composed of a Morion MV89 as GM reference and two WR-Z16 boards, one acting as GM device and another as Slave one. These boards are connected using optical fiber strands and SFPs and the 10MHz outputs have been connected to the Microsemi 3120A to measure the phase noise. For the sake of completeness and for comparative purposes, the analysis has been performed using 10G and 1 GbE configurations for WR-Z16 platforms. Note that the utilization of a more accurate oscillator for the timing reference could provide better performance results.

This experiment has been performed in two different stages. The first one has consisted of figuring out the convergence region of the PI parameters or, in other words, the specific values in which system is able to lock frequencies. Then, a fine sweeping process has been performed to obtain the best values of proportional and integration parameters (Table. 5).

	1 GbE		**10G**	
	GM	Slave	GM	Slave
k_p	500	500	500	400
k_i	4	4	2	1

Table 5. Optimal values for the PI.

Then, the best values of the PI parameters have been applied to the 10G and 1 GbE systems and the phase noise has been measured. Those results are shown in Fig. 7.

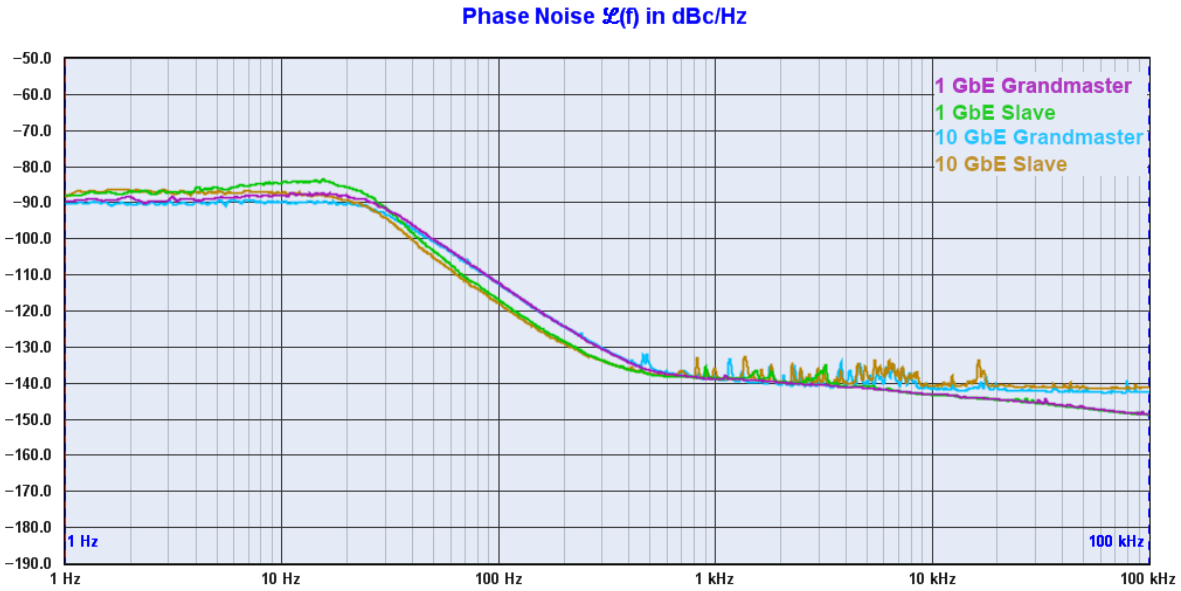

Fig 7. Phase noise results for the best values of PI parameters. This picture has been taken from [1].

Finally, Table. 6 presents a comparison between different platforms: WR-Z16 in 10G configuration, WR-Z16 in 1 GbE configuration, standard WRS and enhanced WRS with low-jitter daughter board [12]. The main conclusion is that the WR-Z16 in 10G configuration exposes the best results in terms of phase noise without considering the enhanced hardware version for the WRS. Please note that the hardware platform for the WR-Z16 board lacks the electronics improvements included on the enhanced WRS. It is a realistic assumption that with equivalent electronics the 10G solution would produce a solution equivalent or slightly better to the WRS enhanced one.

	Bandwidth	1 Hz - 100 Hz	100 Hz - 100 kHz	Total
GM	1 GbE	4.9e-12	4.9e-13	5.4e-12
	10G	4.0e-12	6.6e-13	4.7e-12
	WRS	7.7e-12	2.5e-12	1.0e-11
	Enhanced WRS	6.0e-13	8.0e-13	1.4e-12
Slave	1 GbE	6.4e-12	4.2e-13	6.8e-12
	10G	4.7e-12	6.9e-13	5.4e-12
	WRS	1.1e-11	1.2e-12	1.2e-11
	Enhanced WRS	1.1e-12	8.0e-13	1.9e-12

Table 6. Integrated random jitter comparison between platforms.

5.3.2 Long term synchronization

The long-term synchronization stability analysis is presented in this section. For this experiment, two WR-Z16 boards have been used in 10G configuration, one acting as GM device with the Morion MV89 as timing reference and the other as slave. In this setup, the Keysight 53230A is used to obtain the PPS difference between GM and slave. The first step is to calibrate the 10G link in order to compensate for any fixed delay between master and slave PPS outputs. This procedure has been performed manually but following some recommendations from the WR calibration method [13]. Once calibrated WR-Z16 boards, the synchronization is established and its accuracy is measured for more than ninety hours by means of the PPS difference.

Time Deviation (TDEV) information is depicted in Fig. 8 and represents the phase difference

stability between both devices. In this result, the τ = 1s is below the 2e-11 s level and it decreases until a minimum value around 1e-12 s. These values are very close to those obtained for current WR nodes [14], and even improving some of them.

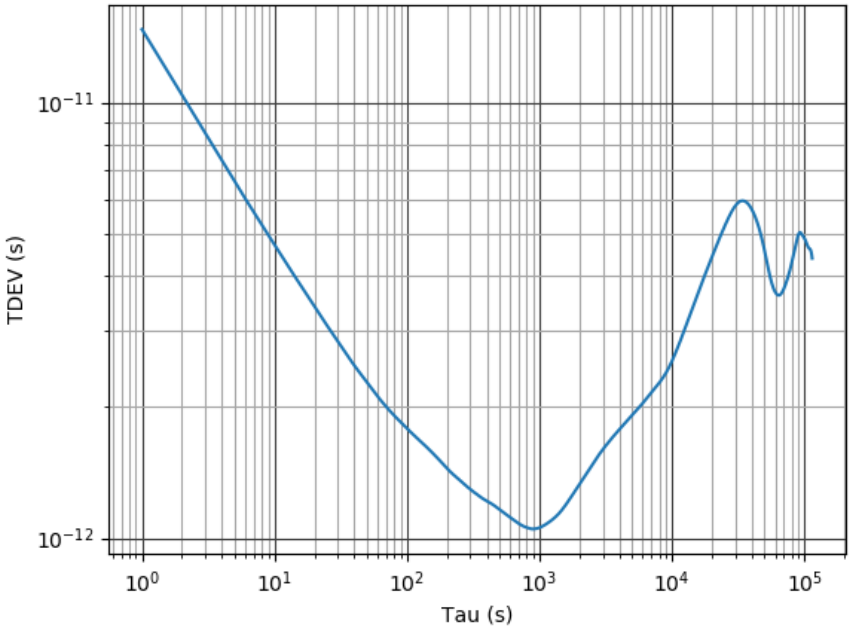

Fig 8. TDEV results. This picture has been taken from [1].

The worst value for the long-term stability analysis can be calculated using the Maximum Time Interval Error (MTIE). Results are shown in Fig. 9 which demonstrates that 10G WRE solution is able to accomplish the WR requirements.

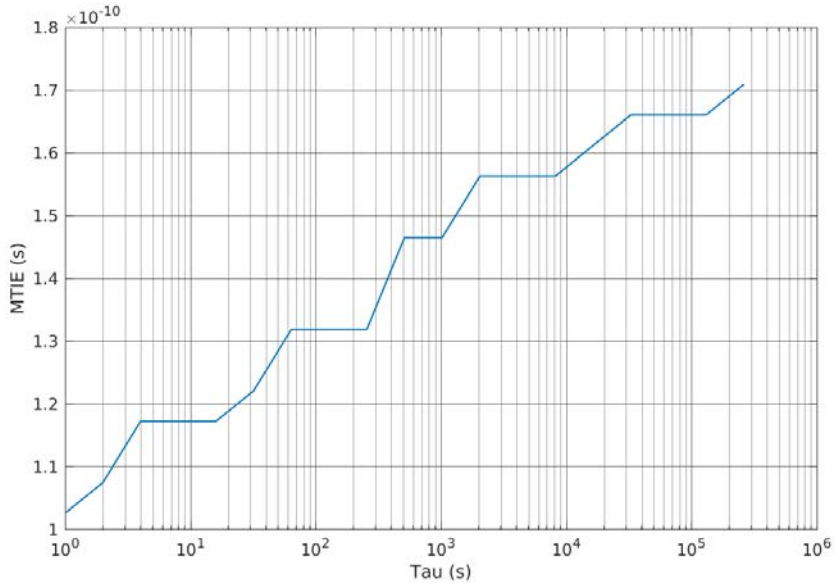

Fig 9. MTIE results. This picture has been taken from [1].

5.4 Data transfer performance & latency experiments

In addition to system characterization and timing experiments, the 10G WRE solution has been evaluated in terms of interoperability, data bandwidth and endpoint latency. The first part of this validation is focused on the interoperability with commercial 10G devices. For this purpose, two 10G network interface cards (Edance DAG 10X2-S and Solarflare communications SFC9120) and a 10G switch device (Arista DCS-7150S-24-R) have been used. Interoperability results show that the 10G WRE system is able to establish a network link and exchange data successfully. Consequently, this solution can be integrated into many applications that deploy commercial 10G devices.

Then, data bandwidth and endpoint latency experiments have been addressed. For data bandwidth measurement, two external workstations equipped with previous 10G network interface cards have been used. These workstations have been connected to the WR-Z16 system using two different 10G optical fiber strands. In this configuration, a workstation can send data packets that are forwarding by the WR-Z16 to reach the second workstation. To obtain results, *nload* application has been used. It can generate a huge amount of data traffic in the form of packets and, at the same time, it provides a data bandwidth measurement.

The 10G WRE system takes advantage of 91.2% of the total capacity of the 10G link as shown in Fig. 10. An important remark to note is that the 10G network interface card is only able to use 9.17 Gbps and, therefore, the WR-Z16 output data bandwidth is also limited by this value. This means that 10G WRE solution could obtain better results if another 10G network interface card with more data bandwidth capabilities is used for experiments.

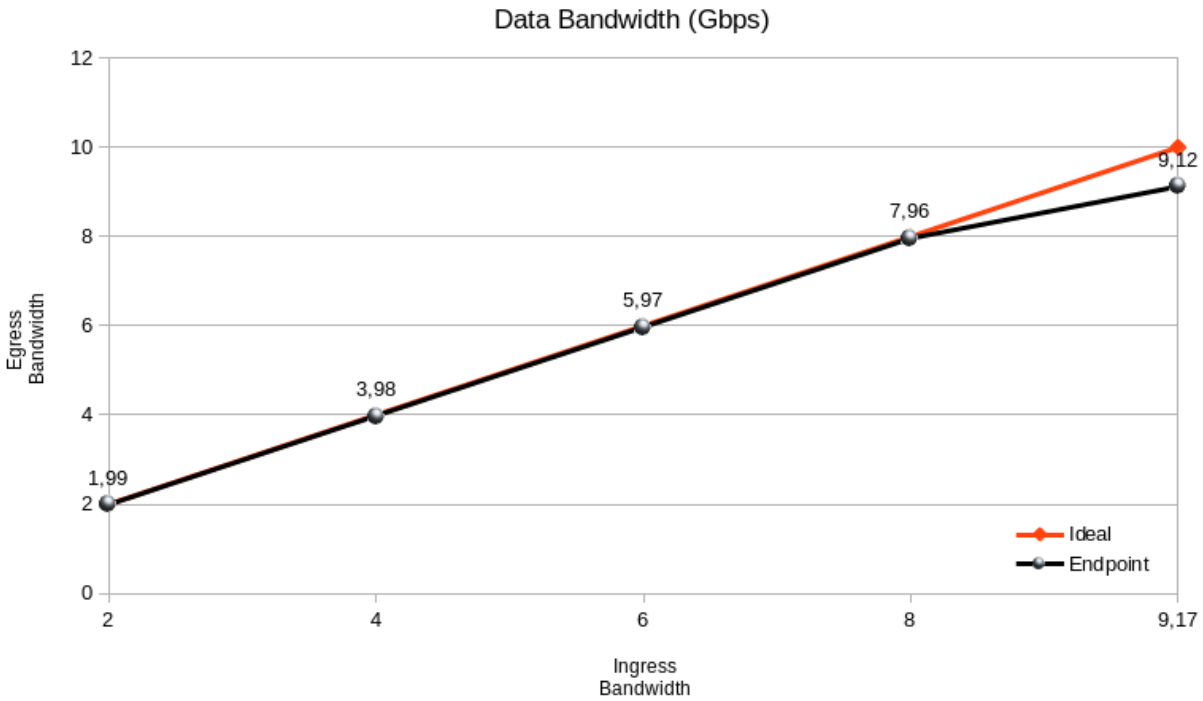

Fig 10. Data bandwidth results. This picture has been taken from [1].

For the endpoint latency experiment, external workstations are no longer required. Instead of this, some additional IP cores have been developed in the FPGA in order to generate different kinds of data packets and to receive them. Moreover, time baseline block that provides a common time base for time measurement and two TSU are also mandatory to check the time difference between packet transmissions and receptions. Thanks to these new components, the endpoint latency can be computed subtracting two timestamps (one for the transmission and the other for the reception of a specific packet) and dividing the result by two. This procedure is an approximation that considers that transmission and reception paths are symmetric ones and that the fiber propagation delay is negligible compared to the endpoint latency for short links.

Finally, results for the endpoint latency show that the 10G WRE system presents a value between 198 and 205 ns with a standard deviation of 2.59 ns. This is a good result but It is important to remark that the goal of this system has not been achieving ultra-low latency features and therefore better results can be achieved in this case. Nevertheless, the modularity of the design allows including third party IP block that can be customized for low latency applications in case the future application has this goal.

6 Conclusions

The present document describes one of the contributions of the UGR that has been partially funded for the ASTERICS project. The main task was focused on the development of a new 10G system able to provide performance timing synchronization together with high data bandwidth capabilities. For the high performance timing synchronization, WR technology has been proposed. However, it was originally designed to work only in 1 GbE networks. Under this context, a significant effort has been performed to allow WR to be deployed in 10G platforms.

Then, a 10G WRE system has been implemented using the WR-Z16 platform. The gateware part has been designed following a modular and flexible architecture. The design has required the integration of third-party modules and because network components can be easily replaced, this modularity open the door to customization if the specific application has different requirements. In addition to this, some software components have been integrated and developed to control hardware resources (HAL daemon and drivers) and to implement WR protocol (Timing daemon).

Once described the 10G WRE system, several experiments have been accomplished to characterize the solution in terms of resource utilization, CPU load impact over timing synchronization, frequency distribution, long term stability, data bandwidth and latency. For the frequency distribution, the 10G WRE system is able to disseminate frequency over the network with a phase noise lower than most current WR devices. Moreover, the 10G WRE system provides an accuracy whose worst value is bounded by 200 ps, fulfilling WR requirements and presenting a similar performance than the most evolved WR devices.
In addition to the timing experiments, the data bandwidth and latency have been measured. Under this context, the 10G WRE solution is able to take advantage of 91.2% of the total capacity of the 10G link and presents an endpoint latency comprises between 198 and 205 ns with a standard deviation of 2.59 ns.

Thanks to this work and for the first time in the literature, an operative 10G WR system has been developed. It allows that WR technology can be exploited on the future on many applications from scientific to industrial domains such as data acquisition instruments, astrophysics infrastructures, data centers or telecommunication networks using a single network for data and timing packets. Furthermore, experiments not only shows that 10G WR system is able to provide a timing performance better than the existing 1 GbE version but a significant improvement on the data handling operation. Under this context, UGR team has also performed some developments for scientific infrastructures such as SKA [14] and CTA [15].

Finally, it is important to remark that although this results represents a remarkable scientific contribution, its utilization by astrophysics facilities still requires a deeper validation, current

Technology Readiness Level (TRL) is 4, and a proper 10G WRE ecosystem of devices so the adoption for the community will require further research and developments in order to be fully exploited.

Future work will address the improvement of the timing performance with enhanced timing circuitry as well as the evaluation of the feasibility of 25 Gigabit Ethernet (25G) interfaces.

Bibliography

[1] M. Jiménez-López, Distributed control system based on high accurate timing synchronization, PhD dissertation, 2019.
[2] Square Kilometer Array telescope project website, Available online at https://www.skatelescope.org/, Last visited on March 2019.
[3] Cherenkov Telescope Array project website, Available online at https://www.cta-observatory.org/, Last visited on March 2019.
[4] White Rabbit wiki page, Available online at https://www.ohwr.org/project/white-rabbit/wikis/home, Last visited on March 2019.
[5] T.Włostowski, Precise time and frequency transfer in a White Rabbit network, 2011, Available online at https://www.ohwr.org/project/white-rabbit/uploads/6a357829064b9e27a46fbce4cb4398b4/mgr.pdf, Last visited on March 2019.
[6] G. Daniluk, White Rabbit PTP Core the sub-nanosecond time synchronization over Ethernet, 2012, Available online at https://www.ohwr.org/project/white-rabbit/uploads/479c4673c48eae8fe17816d144921217/GD_mgr.pdf, Last visited on March 2019.
[7] IEEE, IEEE 802.3ae-2002 - IEEE Standard for Information technology - Local and metropolitan area networks - Part 3: CSMA/CD Access Method and Physical Layer Specifications - Media Access Control (MAC) Parameters, Physical Layer, and Management Parameters for 10 Gb/s Operation, 2002, Available online at https://standards.ieee.org/standard/802_3ae-2002.html, Last visited on March 2019.
[8] D. Cunningham, The status of the 10-Gigabit Ethernet standard, in Proceedings 27th European Conference on Optical Communication (Cat.No.01TH8551), vol. 3. IEEE, sep 2001, pp. 364–367. Available online at http://ieeexplore.ieee.org/document/989668/, Last visited on March 2019.
[9] Intel, 10 Gigabit Ethernet Technology Overview, 2003, Available online at http://web.cse.ohio-state.edu/$\sim$panda.2/788/papers/1h_pro10gbe_lr_sa_wp.pdf, Last visited on March 2019.
[10] Xilinx, Xilinx 10G MAC IP core website, 2018, Available online at https://www.xilinx.com/products/intellectualproperty/do-di-10gemac.html, Last visited on March 2019.
[11] Xilinx, Xilinx 10G PCS/PMA (Base-R) IP core website, 2018, Available online at https://www.xilinx.com/products/intellectualproperty/10gbase-r.html#overview, Last visited on March 2019.

[12] M. Rizzi, M. Lipinski, P. Ferrari, S. Rinaldi, and A. Flammini, White Rabbit Clock Synchronization: Ultimate Limits on Close-In Phase Noise and Short-Term Stability Due to FPGA Implementation, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 65, no. 9, pp. 1726–1737, 2018, Available online at https://ieeexplore.ieee.org/document/8400550/, Last visited on March 2019.
[13] G. Daniluk, White Rabbit calibration procedure, Available online at https://www.ohwr.org/documents/213, Last visited on March 2019.
[14] M. Jimenez-Lopez, F. Torres-Gonzalez, J. L. Gutierrez-Rivas, M. Rodriguez-Alvarez, and J. Diaz, A Fully Programmable White-Rabbit Node for the SKA Telescope PPS Distribution System, IEEE Transactions on Instrumentation and Measurement, pp. 1–10, 2019, Available online at https://ieeexplore.ieee.org/document/8412745/, Last visited on March 2019.
[15] M. Jiménez-López, J. Machado-Cano, M. Rodríguez-Álvarez, M. Stephan, G. Giavitto, D. Berge and J. Diaz, Optimized framegrabber for the Cherenkov Telescope Array, Journal of Astronomical Telescopes, Instruments and Systems, 5(1), 014001, 2019, doi: 10.1117/1.JATIS.5.1.014001, Available online at http://dx.doi.org/10.1117/1.JATIS.5.1.014001, Last visited on March 2019.