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

D. Calvet and E. Molina Gonzalez

Abstract—**High precision clock distribution is of primary importance for the accurate synchronization of distributed detectors in medium to large-scale physics experiments and in some other scientific instruments as well. Common techniques for distributing a reference clock to a large number of end-points often rely on high-speed serial transceivers embedded in Field Programmable Gate Arrays and use precision phase measurements methods to track, and eventually compensate, distribution delay variations. White Rabbit is a well-known solution that combines these techniques with Ethernet technology. This work explores an alternative approach to clock distribution, specifically in view of the Hyper-Kamiokande experiment. We report the design and test of a cascadable 48-optical port clock distributor in 1U x 19'' standard form-factor. The central element is a commercial System on Module equipped with a Xilinx Zynq UltraScale+ device. Superior port density compared to other designs is reached by using the large available number of ordinary differential I/O pairs instead of the limited count of high speed SerDes. This approach comes at the expense of lower link bandwidth. We detail the concepts and the difficulties of designing fixed-latency high-speed serial communication using ordinary FPGA I/O's and we investigate interoperability with dedicated multi-gigabit FPGA SerDes. We explain how to implement precise clock round-trip latency measurements with the proposed links. We present characterization measurements obtained with this demonstrator and show its main figures of performance. Finally, we outline how these studies will serve to construct the final clock distribution system of Hyper-Kamiokande.**

Index Terms— **Precise clock distribution, jitter, FPGAs.**

I. Introduction

DIGITIZATION electronics for the readout of detectors in modern physics experiments and other scientific applications are very often synchronous sampling circuits that need to be precisely synchronized to a common global reference clock with low jitter and wander. The Hyper-Kamiokande experiment [1] is an example of application that require individual channel clock jitter and inter-channel synchronization controlled at the level of few tens of picoseconds over thousands of electronic channels distributed across thousands of cubic meters. Modern techniques for clock and synchronization data distribution often use the multi-gigabit per second capable SerDes blocks embedded in Field Programmable Gate Arrays (FPGAs). The reference clock is embedded in the serial data. The Clock and Data Recovery

(CDR) section of the receiver SerDes is used to extract a copy of the initial sender clock and deserialize received data. Special configuration of the SerDes block, generally specific to a particular device or family, and delay alignment techniques [2] are used to ensure the fidelity of the reconstructed clock to the original. The Digital Dual Mixer Time Difference (DDMTD) [3] is a commonly used method for the precise measurement of phase differences between synchronous clocks. The White Rabbit technology [4] combines the above techniques with popular Ethernet standards and offers ready-to-use products for building high precision time-compensated clock distribution and data aggregation networks. The goals of this work is to explore alternative techniques and evaluate their benefits and limitations, primarily in view of building the clock distribution system of the Hyper-Kamiokande detector.

II. Motivations for a New Custom Development

The Hyper-Kamiokande collaboration decided from the earliest design stage to have distinct paths for clock distribution and data acquisition. Adding in parallel to the standard Ethernet network planned for data acquisition a White Rabbit network exclusively used for clock distribution does not seem a judicious solution. Currently available White Rabbit switches, 3rd generation, are somewhat outdated (Xilinx Virtex 6) and the 4th generation is still in development. Coming products will support 10 Gbps Ethernet but the number of ports per switch will remain 18. Our target application has modest bandwidth needs (125 Mbps minimum per link), but scalability to ~2000 end-points was the (initial) goal. This scale is easier to reach with switches that have a higher port count. These are among the reasons that led developing new custom hardware.

III. Proposed Clock Distributor Architecture

The Hyper-Kamiokande neutrino detector is composed of a huge underground tank (60 m diameter, 70 m height) filled with ultrapure water and instrumented with around twenty-five thousand PhotoMultiplier Tubes (PMTs). Digitizer end-points are placed under water while the data acquisition system, the power supplies and the clock distribution system are placed above the water tank, in the cavern of the experiment. A schematic view of the clock distribution system of the Hyper-Kamiokande detector is shown in Fig. 1. The proposed topology

for the clock distribution network is a multistage tree of interconnected switches. Assuming that identical switches with one up-link and N down-links are used, a two-stage fanout tree composed of one root switch driving N slave switches scales up to N^2 end-points.
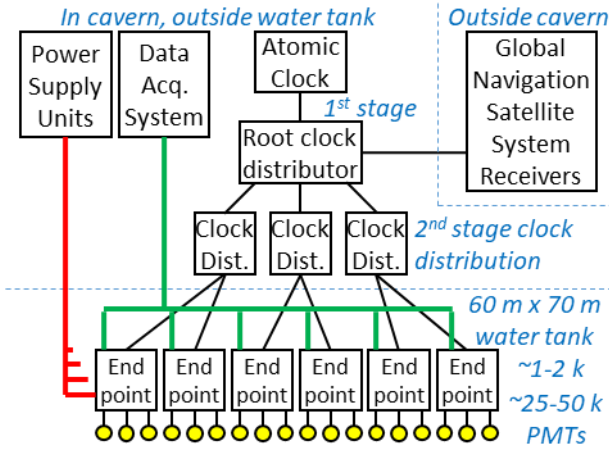

Fig. 1. Schematic view of the clock distribution system of the Hyper-Kamiokande detector.

Switches with at least 45 ports are needed to reach 2000 end-points. Hence, we decided to build a 48-port clock distributor board. This port density cannot be reached with one low to mid-range FPGA using the traditional solution based on high-speed SerDes blocks because too few of these blocks are available. Following our earlier work [5] we took the approach of using ordinary FPGA I/O pins for serial communication. Although such pins are much slower than dedicated SerDes blocks, ordinary FPGA I/O pins in modern FPGAs have now passed the 1.25 Gbps gap, which is the rate required for Gigabit Ethernet. To simplify design, we use a commercially available System-on-Module (SoM) to implement the core functions of the clock distributor. We selected Trenz TE803 [6] family of SoMs because it offers several versions of cost effective Xilinx Zynq UltraScale+ FPGA and provides a large number of I/O pins. The architecture of the proposed clock distributor board is shown in Fig. 2.
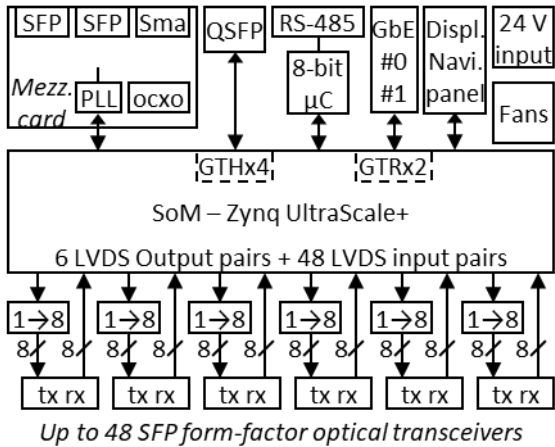

Fig. 2. Architecture of the proposed 48-port clock distributor board.

The clock distributor has an asymmetric structure: in the forward direction, it is composed of six different groups of eight

ports where one FPGA differential output is fanout to eight SFP transceivers via a 1 to 8 external chip (Texas Instruments CDCLVP1208). In the reverse direction, each of the 48 SFP receivers connects to a distinct FPGA input pin pair. The attribute DQS_BIAS [7] is set to TRUE for the LVDS inputs of the FPGA that are connected to the receiver outputs of SFP transceivers (PECL levels, AC coupled internally). This restores the proper DC bias. The board has two standard Gigabit Ethernet ports for system configuration and monitoring (GTR transceivers of the SoM), an optional 40 Gbps QSFP interface (GTH transceivers, only available on some appropriate SoM variants), a RS-485 port, a USB console port, a 1.7'' OLED display, a navigation panel, two redundant external power inputs (24 V), and two counter-rotating fans. A mezzanine card equipped with a state-of-the art PLL (Analog Devices AD9545) provides a replaceable interface to an external clock source: an atomic clock when the board serves as a root node, and serial optical links when the board is used as a second stage clock distributor. All hardware is housed in a 1U x 19'' rack mountable enclosure. A picture of our 48-port clock distributor board is shown in Fig. 3.

Fig. 3. Prototype 48-port clock distributor board.

IV. OPERATION AND PERFORMANCE

A. Transmission of Clock or Serial Data

Each of the six fanout chips is driven by a distinct LVDS output pair of the SoM. Internally, an OSERDESE3 primitive and ODELAYE3 block [7] are used. The OSERDESE3 serializers are set to 8-bit mode and 125 MHz parallel clock rate, leading to 1 Gbps on the optical media. Sub-multiples of that rate can be obtained by sending each bit of data multiples times. Each group of eight optical ports can independently be configured to send a pure 125 MHz clock (selected reference frequency for Hyper-Kamiokande), a 125 MHz clock with a duty-cycle modulated by 125 Mbps user data, serially encoded data at 1 Gbps, 500 Mbps or some lower sub-multiples.

A typical waveform measured on the optical media when transmitting a 125 MHz duty-cycle modulated clock and an eye diagram obtained when sending pseudo-random serial data at 1 Gbps are shown in Fig. 4a and Fig. 4b respectively.
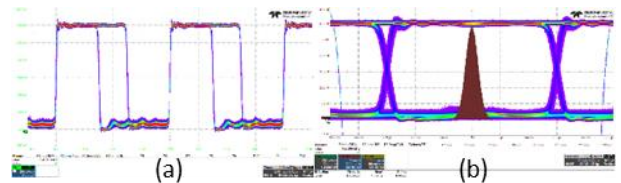

Fig. 4. Transmission of a 125 MHz duty-cycle modulated clock (a). Eye diagram when transmitting pseudo-random data at 1 Gbps (b).

The measured random and deterministic jitters are 10-15 ps rms, which is within our requirements. We checked that the skew between the input reference clock and the output clock on the optical media does not vary by more than few ps when the clock distribution board is power-cycled or reprogrammed.

B. Interoperability with High Speed SerDes Receivers

Contrary to the Xilinx 7 family, the serializer/deserializer block of I/O pins in High Performance (HP) banks of the UltraScale+ family does not support ratio of 1:5 or 1:10. Only 1:2, 1:4 or 1:8 ratios are available. This brings some difficulties for implementing 8B-10B serial encoding and decoding because the parallel domain has to run at $1:8^{th}$ or $1:4^{th}$ of the *encoded* line rate (e.g. 156.25 MHz for Gigabit Ethernet instead of 125 MHz). For simplicity, we chose a proprietary protocol and Manchester encoding for serial data communication. A net data rate of 250 Mbps is sufficient in our case, leading to 500 MBd on the optical media. This speed is the lower limit of the rate acceptable by high-speed SerDes like Xilinx GTP or GTH. We implemented a test receiver in a Xilinx Artix 7, using a GTP SerDes. The parallel side of the GTP receiver is 16-bit wide, leading to a 31.25 MHz recovered clock in the parallel domain at 500 MBd serial input rate. The GTP can output received data with 16 different possible alignment offsets depending on the initial phase of the divider that derives the clock for the parallel domain from the serial recovered clock. In order to provide deterministic latency operation, we embed *a framing pattern* composed of 4 zero's followed by 4 one's at the transmitter side by XOR'ing the data to be sent with this constant repetitive pattern. At link startup, the receiver compares received data with the framing pattern and resets the GTP until the recovered clock reaches the phase offset that leads to the reception of properly aligned framing patterns.

C. Measuring Clock Round-trip Delay Variations

When used as a root node, the clock distributor board can receive copies of the distributed clock looped back by slave distributors. By measuring precisely these round-trip delays, the variations of the phase of the distributed clock can be estimated.

We implemented precise phase measurement logic using the DDMTD technique. The PLL of the SoM is used to generate a clock synchronous to the 125 MHz reference with a period offset of $1/1000^{th}$ (i.e. 8.008 ns or 124.875 MHz). The phase measurement logic we implemented can measure clock round-trip variations with a resolution of 8 ps at 125 kHz refresh rate. All the measurements of one port selected among 48 can be accumulated in a histogram in real time, and simultaneously, the average of a selectable number of consecutive measurements (1, 128, 1024 or 8192) can be recorded at fixed intervals (10 ms, 1 s, 10 s or 60 s) for each port in parallel.

In order to validate the correctness of results, we configured the board for 125 MHz pure clock distribution and looped back with short optical fibers the distributed signals at multiple inputs. By changing the delay setting of the ODELAYE3 block on the transmit path, we could inject some calibrated delay variations. We checked that the values obtained by our measurement logic match the results obtained with an oscilloscope. An example is shown on Fig. 5.

We operated the clock distributor board with various types of loop-back: 1:24 passive optical splitter, 150 m long monomode and multimode optical fibers. The measured phase variations obtained over one week of operation are shown in Fig. 6.
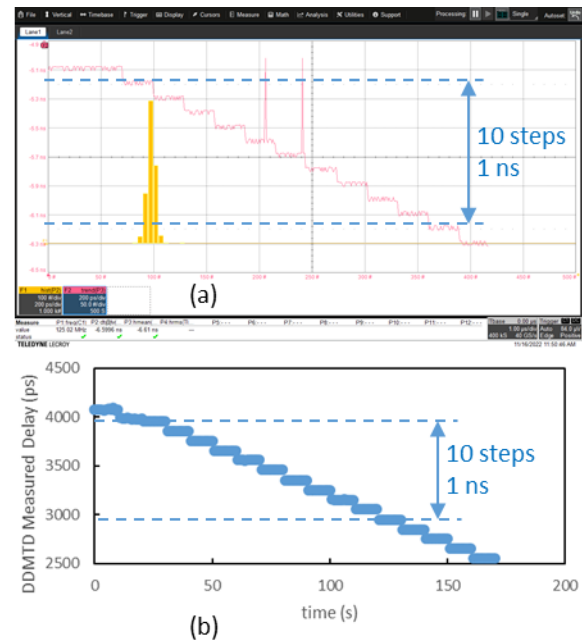

Fig. 5. Delay changes in steps of 100 ps are made on the TX path of the clock distributor board configured for 125 MHz clock distribution. Measurements on one output port made with a high-speed oscilloscope show the correct detection of these phase jumps (a). Measurements obtained with the internal phase measurement logic on a port externally looped-back show that the induced phase jumps are detected consistently (b).
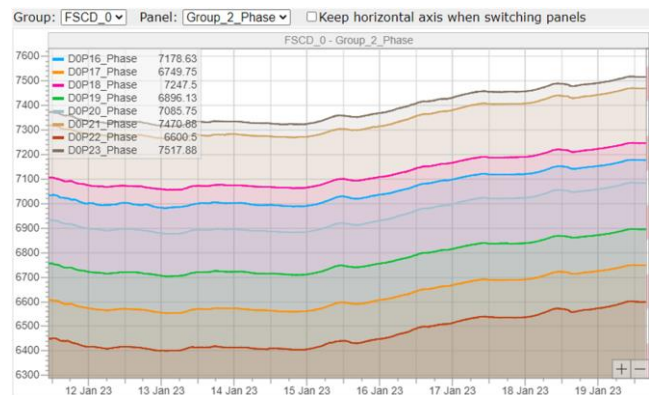

Fig. 6. Phase variations measured over one week of operation of the clock distributor board set for 125 MHz clock distribution. Ports #16 to #23 are in loop back mode over 150 m multimode optical fibers. For each channel, the observed phase deviation is ~200 ps peak-peak. It is highly correlated with temperature variations (6°C peak-peak over this period).

D. Receiving Serial Data and Simultaneously Measuring Round-trip Clock Delay Variations

Receiving serial data and measuring clock round-trip delay variations at the same time on a receiver based on an ordinary FPGA I/O pin pair is not trivial because, contrary to high-speed dedicated SerDes blocks, ordinary input pins do not produce a recovered clock. The DDMTD phase measurement method relies on detecting the edges of the local reference clock and the echoed clock. It cannot operate directly on serially encoded data at input. To work around this limitation, we propose the proprietary serial line encoding shown in Fig. 7.

Sampling serially encoded data as shown at a frequency slightly offset from the repetition rate of the preamble pattern produces a long series of '0' followed by a long series of '1' (corresponding to the sampling of successive preambles) and a fast alternating series of "01" or "10" corresponding to the sampling of user data bits and their complement.
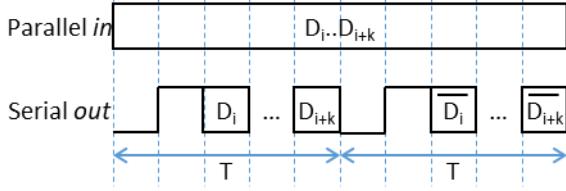

Fig. 7. Serial encoding for the simultaneous transmission of data and phase measurements with the DDMTD technique. At 500 MBd and a primary clock rate of 125 MHz, each pair of consecutive 8 ns period, T, contains two bits of constant preamble "01" followed by two bits of serial data for the first cycle and the complement of these two bits in the second cycle. Coding efficiency is 25%, leading to a net user data bandwidth of 125 Mbps.

To ensure the correct detection of the edges of the carrier clock, a digital filter (upper half of the decoder on Fig. 8) removes the fast alternating part of the above signal before feeding it to the DDMTD phase measurement block. The digital filter determines the presence of rising edges of the carrier clock when the sampled serial input leads to a sufficient number of 0's followed by multiple 1's. Longer or shorter sequences of consecutive 0's or 1's can be rejected by tuning the length of the two parts of this filter. This can be useful when a lower overhead method (e.g. scrambling) is used for encoding serial data instead of Manchester encoding as implemented here.
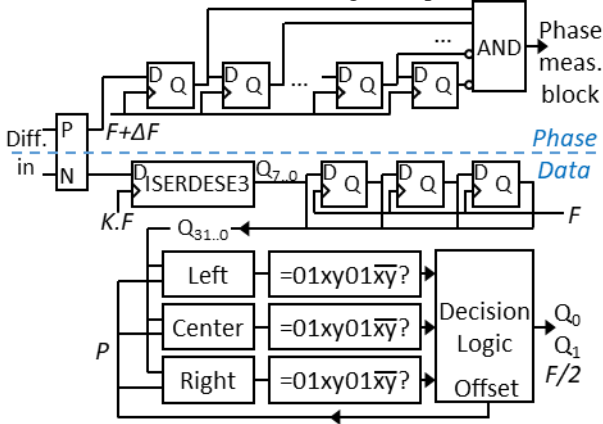

Fig. 8. Block diagram of receiver for simultaneous clock phase measurements and serial data reception. The P side of a differential input buffer drives the phase measurement branch. The N side drives the data reception branch.

The recovery of the serial data is done as shown on the lower half of the decoder shown on Fig. 8. An ISERDESE3 deserializer primitive clocked at 1 GHz samples the input signal and delivers 8-bit parallel words at 125 MHz. A history buffer stores the value of the last 32 bits received. Depending on an offset parameter P, three selectors, called Left, Center and Right, pick 8 bits from the history buffer corresponding to relative sampling positions P, P+0.5UI and P+1UI (UI=Unit Interval, i.e. 2 ns at 500 Mbps). The three sets of 8 selected bits are then compared to the expected 8-bit value "01xy01(not x)(not y)" and a score from 0 to 6 is computed for the Left, Center and Right branches depending on how close each group of selected bits matches the expected pattern. During link

synchronization, the offset parameter is changed until reaching the position where the Left and Center selectors stably output the maximum score of 6 (perfect match). During operation, the decision logic constantly monitors the scores obtained at the three adjacent sampling positions to dynamically increase or decrease the offset parameter P and compensate the slow variations of the arrival time of the input signal. One difficulty is ensuring an error-free data capture during the transitions from one alignment position to the next.

V. CLOCK DISTRIBUTION SYSTEM FOR HYPER-KAMIOKANDE

The clock distribution for Hyper-Kamiokande will adopt a two-stage tree structure but collaboration decisions were more conservative than what we explored in this R&D. Second stage distributors and leaf end-points will use the traditional approach based on FPGA dedicated high-speed SerDes blocks for deterministic latency serial communication. Second stage distributors will only have 16-ports, due to the limitations of the selected SoM, but the number of second stage distributors will remain manageable (~60) because the final system will only have ~1000 end points, while our initial hypothesis was more than twice as many. The root distributor will be a re-designed version of the present 48-port demonstrator.

VI. SUMMARY

We reported the design of a 48-optical port clock distributor board and show how ordinary FPGA I/O pins on Xilinx Zynq UltraScale+ family of devices can be used for deterministic latency clock distribution at 125 MHz, serial data transmission at up to 1 GBd and 500 MBd for reception. We implemented precise clock round trip measurement logic based on the DDMTD method and propose specific line encoding and circuitry to extend the use of this method to signals that transport serial data.

The final clock distribution system of Hyper-Kamiokande will inherit some of the concepts of this R&D. The development of this system is a collaborative on-going effort between Irfu (1st stage clock distributor), and other French and Italian partners (clock reference, 2nd stage distributors and end-points).

REFERENCES

[1] F. Francesca Di Lodovico et al., "The Hyper-Kamiokande Experiment", in *Journal of Physics, Conference Ser*ies, vol. 888, 012020, 2017.
[2] E. Mendes, S. Baron, C. Soos, J. Troska, and P. Novellini, "Achieving Picosecond-Level Phase Stability in Timing Distribution Systems With Xilinx Ultrascale Transceivers", *in IEEE Trans. Nucl. Sci.* vol. 67 N°3, pp-473-481, March 2020.
[3] P. Moreira, P. Alvarez, J. Serrano, I. Darwezeh and T. Wlostowski, "Digital dual mixer time difference for sub-nanosecond time synchronization in Ethernet", *in Proc. IEEE International Frequency Control Symposium*, Newport Beach, CA, USA, 2010, pp. 449-453.
[4] J. Serrano, P. Alvarez, M. Cattin, E. G. Cota, P. M. J. H. Lewis, T. Włostowski et al., "The White Rabbit Project", *in Proceedings of ICALEPCS TUC004*, Kobe, Japan, 2009.
[5] D. Calvet, "Back-End Electronics Based on an Asymmetric Network for Low Background and Medium- Scale Physics Experiments", *in IEEE Trans. Nucl. Sci.* vol. 66 N°7, pp-998-1006, July 2019.
[6] Trenz Electronic, "TE803 Technical Resource Manual", 2019.
[7] Xilinx, "UltraScale Architecture SelectIO Resources User Guide", document UG571, August 2019.