

# Low Thermal Sensitivity Hollow Core Fiber for Optically-Switched Data Centers

Kari A. Clark, *Student Member, IEEE*, Yong Chen, *Member, IEEE*,  
Eric R. Numkam Fokua, *Member, IEEE*, Tom Bradley, *Member, IEEE*,  
Francesco Poletti, *Member, IEEE*, David J. Richardson, *Fellow, IEEE*,  
Polina Bayvel, *Fellow, IEEE*, Radan Slavík, *Senior Member, IEEE*  
and Zhixin Liu, *Senior Member, IEEE*

**Abstract**— Optical switches offer several benefits over electronics switches, including better scalability, lower power consumption, and lower latency. However, their implementation faces several challenges, including electronic transceivers to be able to support sub-nanosecond clock and data recovery time. This is required to efficiently handle data center traffic that is dominated by small data packets. Recent research shows that, scalable and sub-nanosecond data recovery can be achieved by synchronizing the clocks of all end-points connected to an optical switch in frequency and phase. In such a system, thermally-induced change of propagation time through standard single mode fiber (SMF-28) necessitates the clock phases to be tracked due to variations in the data centers temperature. Hollow core fiber has been shown to have a thermal coefficient of delay 20 times smaller than SMF-28, offering potential to simplify the clock phase tracking and to increase the distance scale of data center networks. In this paper, we show how the low thermal coefficient of delay in hollow core fiber enables sub-nanosecond optical switching system with only initial frequency and phase synchronization. We obtained error-free real-time transmission of 60 ns packets over 1 km clock and 1 km hollow core data fiber with under 625 ps clock recovery time in both a point-to-point and a 2-to-1 optically switched 25.6 Gb/s real-time system. Based on our results, we estimate that sub-nanosecond clock recovery can be achieved for a 100 m size data center cluster interconnected by an optical switch using hollow core fiber, frequency synchronization and only a single phase calibration at the start-up.

**Index Terms**— Optical fiber communication, Optical fibers, Optical packet switching, Data Centers, Clock recovery

## I. INTRODUCTION

IN today's data centers, data traffic switching is performed predominantly in the electronic domain. Nevertheless, future data center interconnection face challenges of scalability, power consumption, and latency [1]. The current electronic switches and the three-layer Clos switching architecture have difficulty overcoming these challenges due to several

fundamental limits.

Firstly, the physical constraints on the power and density scalability of silicon transistors [2] will ultimately limit the achievable performance of electronic switching ASIC crossbars [3], and is already causing switch performance loss when handling small packet sizes in current electronic switches [3]. Secondly, the aggregate data rate of data entering and leaving each ASIC is limited by the product of the per pin bandwidth and pin density. Increasing the per pin data rate beyond 112 Gb/s will become very challenging due to the high printed circuit board (PCB) dielectric loss at high frequencies [4]. Consequently, increasing the overall switching capacity will require an increase in the number of high-speed serial transceivers with an unacceptable increase in ASIC power consumption. Though techniques such as co-packaged optics offer a temporary solution to this problem for several generations, further increases in data center network bandwidth and scale require alternative solutions.

Optical switching emerges as a viable solution to overcome the density and bandwidth bottleneck caused by electronic switching. It offers many merits including a large number of input/output ports, high bandwidth, and a significantly reduced number of opto/electronic conversions, reducing power consumption, cost, and latency [1], promising the continuous scaling of the data center networks.

Nevertheless, one of the major challenges to implementing optical switching in a data center is the need for fast clock and data recovery (CDR). When two network nodes begin communicating a data packet is transmitted via a momentary optical path through an optical switch. At the receiver side, the clock embedded in the transmitted data must be recovered before the data can be correctly sampled. The time taken for this process to complete is the CDR locking time. In data center networks, the CDR locking time can be much longer than the time required to receive short data packets, with the best

Manuscript received X<sup>th</sup> Month, 2019. This work was jointly funded by EPSRC Grants AirGuide Photonics (EP/P030181/1), TRANSNET (EP/R035342/1), EP/R041792/1, the ERC Lightpipe project (682724), Royal Society PIF\R1\180001.

K. A. Clark., P. Bayvel and Z. Liu are with the Department of Electronic and Electrical Engineering, University College London, Gower Street, Bloomsbury, London WC1E 6BT, U.K. (e-mail: kari.clark.14@ucl.ac.uk; p.bayvel@ucl.ac.uk; zhixin.liu@ucl.ac.uk).

Y. Chen, E. R. Numkam Fokua, T. Bradley, F. Poletti, D. J. Richardson and R. Slavík are with the Optoelectronics Research Centre, University of Southampton, University Road, Southampton SO17 1BJ, U.K. (e-mail: yc1m12@soton.ac.uk; eric.numkam-fokoua@soton.ac.uk; t.bradley@soton.ac.uk; djr@orc.ucl.ac.uk; fp@soton.ac.uk; R.Slavik@soton.ac.uk)

research prototypes achieving 8 ns CDR locking time [5]. In data center networks where small data packets dominate the traffic, sub-nanosecond CDR locking time is required to ensure a sufficiently high network throughput [6].

CDR requires recovery of both the frequency and phase of the clock embedded in received data signals. The synchronization of transceiver clock frequencies can be achieved by distributing an optical reference clock to all network nodes [7]. Nevertheless, as shown in [6], clock phase recovery can still take up to 10s of nanoseconds because the CDR module must search through a range of clock phases for each new incoming data signal. This search for the correct sampling clock phase is necessary because of the change of the signal propagation delay due to environmental temperature variation. In a practical data center environment, the operating temperature range could vary from 18 to 27 °C [8] and the rate of temperature change in a typical production cloud data center could be up to 0.03°C/s [6]. For standard single mode fibers (SMF-28), the typical rate of temperature-induced propagation delay change (given by the thermal coefficient of delay, TCD) is about 40 ps/km/°C [9]. Assuming a typical 1-km short-reach data center link, a 0.5 °C temperature change could lead to a data delay change of half a symbol period for 25 Gb/s signals, causing bit errors if the CDR module fails to track the change. To resolve this issue, Clark et al. proposed clock phase caching [6], where the transceivers are configured to memorize and track the relative clock phase offset between different transceiver pairs, enabling sub-nanosecond CDR locking time.

A possible alternative solution is to use optical fiber that is insensitive to temperature changes, such that the clock phase maintains stable regardless of the temperature variation. In such a system, an optically-switched network could be constructed that uses fixed clock phases, calibrated on network startup, and then uses static clock phase values without any update required as data center temperature changes. This approach could eliminate the need for actively clock phase caching, reducing the hardware power consumption and network overhead.

A very promising low-thermal sensitivity optical fiber is hollow core fiber (HCF), which has been demonstrated to have TCD of 2 ps/km/°C, which is about 20 times lower than TCD of SMF-28. A specific design of HCF was even demonstrated to have TCD = 0 [10]. Although this was achieved only at a specific wavelength, it opens a new opportunity for ultra-fast clock recovery that is immune to the impact of temperature variation.

In this paper, we experimentally investigate the performance of HCF in clock-synchronized data center interconnection, with a focus on the tolerance to temperature variation. We first study the temperature tolerance in a point-to-point transmission without any active clock phase tracking. We then measure and compare the clock phase variation of HCF with SMF-28 in a 2×1 optically-switched system demonstration with sub-nanosecond clock recovery achieved without any clock phase update.

## II. THERMAL PROPERTIES OF STANDARD AND HOLLOW CORE FIBERS

Signal propagation time through an optical fiber depends on temperature, which is in standard (glass-core) fibers attributed to two effects [11]. The dominant effect (95% of total change) is due to the change of the refractive index of silica glass with temperature, making light travelling slower through the fiber as the temperature is increased. The other effect is the fiber elongation due to the thermal expansion of the fiber glass with temperature, which forces light to travel a longer distance. Fig. 1 shows the measurement of the delay change using a spool of 1-km of SMF-28 with the propagation delay changing at a rate of 46 ps/°C, corresponding to TCD of 46 ps/km/°C. In Hollow Core Fibers (HCFs), there is no silica glass in the core and the refractive index change of the core material (air at constant volume or vacuum) is very small. Thus, the thermal propagation time change in HCFs is dominated by the thermally-induced elongation. As this effect is relatively small (5% of total contribution in SMF-28), HCFs have thermal sensitivity about 20 times lower than SMF-28. This is shown in Fig. 1 in which the delay change of 1-km long hollow-core photonics bandgap fiber (HC-PBGF) sample used in this study with temperature is shown, changing at a rate of 1.8 ps/°C.

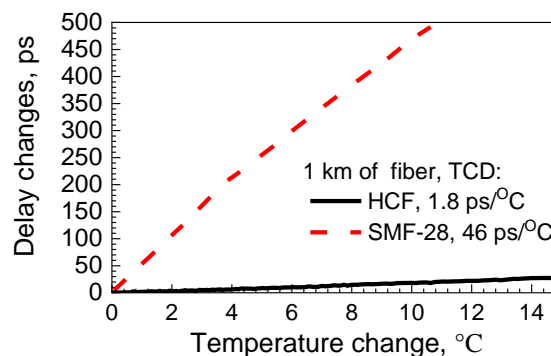


Fig. 1 Change of signal delay propagating through 1-km long SMF-28 (red, dashed) and HCF (solid, black).

It is worth mentioning that the latest generation HCFs have loss below 1 dB/km (recently-published record is 0.65 dB/km at 1550 nm [12]). They are manufactured in km lengths [13]. Further, they can have low chromatic dispersion (e.g., <3 ps/nm/km) and very low non-linearity (>2 orders of magnitude lower than SMF-28). They also have shorter propagation delay (lower latency), as signals propagate about 30% faster in HCFs as compared to SMF-28 [14]. All these features and latest developments makes them particularly interesting for applications in Data Centers.

## III. EXPERIMENTAL SETUP

### A. Point-to-point clock synchronized transmission via HCF

We first study the effect of the low thermal sensitivity of HCF on data transmission in a point-to-point transmission experiment as shown in Fig. 2. An optical clock signal was generated by modulating an 800 MHz reference clock (Si5324, 320 fs rms jitter (50 kHz-80 MHz) [15]) onto a 1550 nm optical carrier, via a LiNbO<sub>3</sub> Mach-Zehnder modulator (MZM). The

optical clock signal was split and sent directly to the receiver (Rx FPGA, Xilinx VCU108) and to the transmitter (Tx FPGA, Xilinx VCU108) via a fully connectorized (FC/APC SMF-28 pigtailed) 1-km HCF. The counter-propagation of the data and the clock signals emulates the worst-case scenario in data center networks, where the clock source is adjacent to the receiver and far from the transmitter, causing the additive change of the clock phase drift in the clock and data paths due to temperature change. The optical clock signals were detected and amplified for frequency synchronization of the two FPGAs through digital phase lock loop (PLL) modules integrated within each FPGA. At the transmitter side, an externally modulated laser (EML) consisting of a distributed feedback (DFB) laser outputting 13 dBm continuous wave (CW) at 1555 nm followed by a 35-GHz electro-absorption modulator was driven with a 25.6-Gb/s non-return-to-zero (NRZ) data of pseudorandom binary sequence (PRBS) length of  $2^{31}$ . The 3-dBm output signal was subsequently launched into a second HCF of 1 km length. The end-to-end loss of the data and clock paths were 6.5 dB and 7.0 dB, respectively, consisting of about 3 dB connector loss and 4 dB fiber loss. At the receiver side, the signal was detected by a 20-GHz bandwidth photodiode followed by a trans-impedance amplifier before the real-time FPGA receiver. The received optical data and optical clock powers were intentionally attenuated to -10.5 dBm to match minimum tolerable power for optical receivers in intra-data center transmission standards [16]. Bit errors were counted in real-time and used to calculate the bit error ratio (BER).

To investigate the impact of clock phase offset due to temperature change, both HCF spools were placed inside a thermally controlled chamber, in which the air temperature was monitored by six temperature sensors surrounding the fiber spools. After an initial stabilization, which took about 5 minutes, the temperature fluctuated by less than 0.05 °C. The chamber was initially stabilized at 33.5 °C and the receiver data sampling clock phase was set to be half a symbol period away from the optimum value. The optimum clock phase value was determined by measuring the Q factor of the received signal. The initial temperature of 33.5 °C was chosen to ensure a significant offset from ambient temperature, maximizing the thermal chamber temperature stability at each point. We then increased the chamber temperature and recorded the BER after the temperature became stabilized (within  $\pm 0.05$  °C). Each BER point was measured from  $7 \times 10^{12}$  bits.

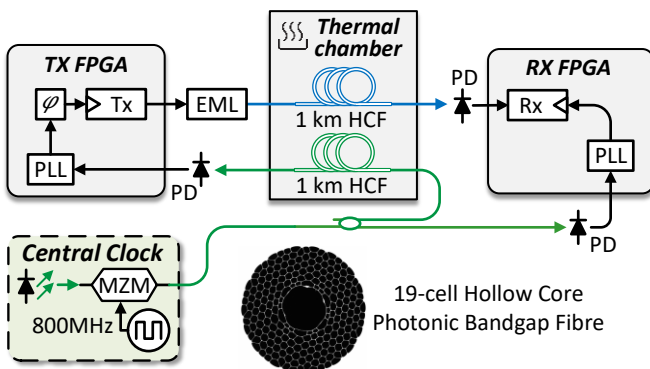


Fig. 2. Point-to-point clock-synchronized transmission using HCF. PLL – Phase locked loop, EML – externally modulated laser, MZM – Mach Zehnder Modulator, HCF – hollow core fiber, PD – photodiode.

### B. Optically-switched prototype system with HCF

Following the point-to-point experiment, we investigated the effect of the performance of the HCF on the BER and CDR locking time in an optically-switched prototype system consisting of a  $2 \times 1$  optical switch, two transmitters (Tx0 and Tx1) and a receiver, Rx, as shown in Fig. 3. In this experiment, the optical clock was sent to Tx0 through 1 km of HCF, and directly to Tx1 and the Rx. The packets were formed by PRBS signals with a length of  $2^9$  to emulate the 64-byte minimum size packets of data center traffic [16]. The optical packets were generated by driving EML0 and EML1 with the 25.6 Gb/s NRZ signals produced by Tx0 and Tx1. Their output optical signals were at 1555 nm and 1553 nm, respectively. The optical packets from both transmitters had 60 ns duration and were interleaved by a  $2 \times 2$  LiNbO<sub>3</sub> MZM with 5 ns gap between the consecutive packets. The  $2 \times 2$  MZM has a rise time of less than 1 ns and was used to emulate fast optical switches. The interleaved packets leaving the switch were amplified by an erbium doped fiber amplifier (EDFA) to compensate for the loss of the  $2 \times 2$  LiNbO<sub>3</sub> MZM (about 6 dB). The use of EDFA is unnecessary in practical systems. Scalable and low loss fast optical switches can be used to provide sufficient power for the optically-switched networks [17]. The interleaved packets then passed through 1 km of HCF before being received by the Rx node. Similarly, to the point-to-point experiment, we attenuated the received optical power to -10.5 dBm to be equal to the minimum required power in the Ethernet standard [16]. Both HCF fiber spools were contained within the thermally controlled chamber that was initially stabilized at 32.5 °C. We measured the BER and CDR locking time at different temperature values. In this experiment, we kept the receiver CDR module running constantly. As a result, only the clock phase offset between the two transmitters contributed to the BER and CDR locking time, which is due to the temperature induced clock and data phase offset through the HCFs.

To calculate the overall BER and CDR locking time, the first and second sequences within each packet are divided into  $64 \times 16$ -bit bins, and the number of bit errors falling in each bin is recorded in real-time at each temperature setpoint until  $10^{13}$  total bits (summed across all bins) have been received [6]. CDR locking time was calculated as the first bin in the packet with a

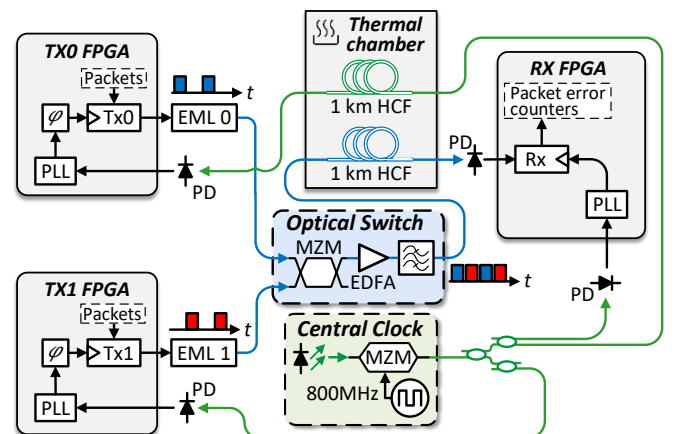


Fig. 3. Synchronous HC-PBGF  $2 \times 1$  optically switch experimental setup. EML – externally modulated laser, EDFA – erbium doped fiber amplifier.

BER of  $<10^{-10}$ , if all following bins also have BER of  $<10^{-10}$ . The overall receiver BER was calculated by summing the errors falling in all 16-bit bins.

#### IV. RESULTS

Fig. 4 shows the impact of temperature change on the BER of the point-to-point clock-synchronized transmission system described in Section IIIA. The blue markers show the measured BERs using the HCFs. The inset shows the eye diagram captured from the receiver FPGA. When temperature increases, the counter propagated clock path and data path led to a total change of the receiver clock phase shift of about 4 ps/°C (equivalent HCF length of 2 km). We observed an error free (BER  $<10^{-12}$ ) over temperature change of 2 °C (35 to 37 °C). To compare with SMF-28, we assumed a thermal coefficient of 40 ps/km/°C and calculated the tolerance to temperature variation based on the measured BER at different clock phase offset values from our previous work [6]. Shown as cross markers in Fig. 4, the error free temperature range using SMF-28 was only about 0.1 °C, which in practice would require active tracking of CDR phase.

Fig. 5a and 5b show the BER and CDR locking time, respectively, of the 2×1 optical switching system experiment described in Section IIIB. The CDR locking time was calculated from the BER as discussed in section 2A. In contrast to [6], in this experiment, the phase of the transmitters was not actively updated, i.e. no phase caching function that we previously employed to track the clock phase that changes with the temperature. This allowed us to directly compare the performance using the two different fiber types. Using the HCF, we observed a window of 4 °C temperature change over which the received data had BER  $<10^{-12}$  and a CDR locking time of  $<625$  ps. The tolerance to temperature variation is consistent with the clock phase offset caused by the 1 km HCF length difference between the two clock to receiver via transmitter paths, as shown in the experimental setup in Fig. 3.

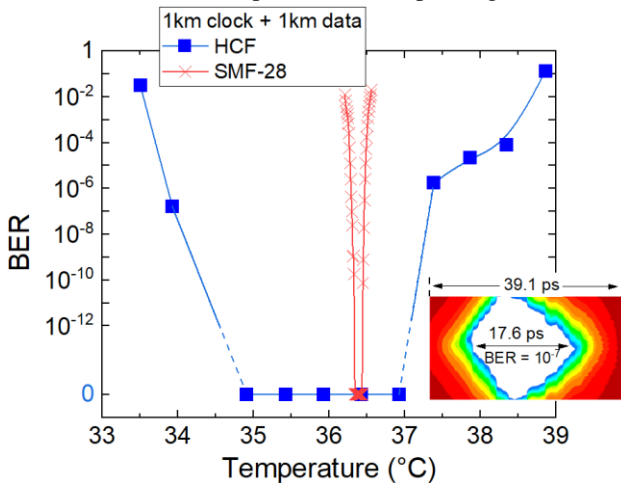


Fig. 4. Impact of temperature variation on the performance of a point-to-point clock-synchronized transmission system. Blue marker: measured BER using the HCF; Red cross marker: estimated when using SMF-28.

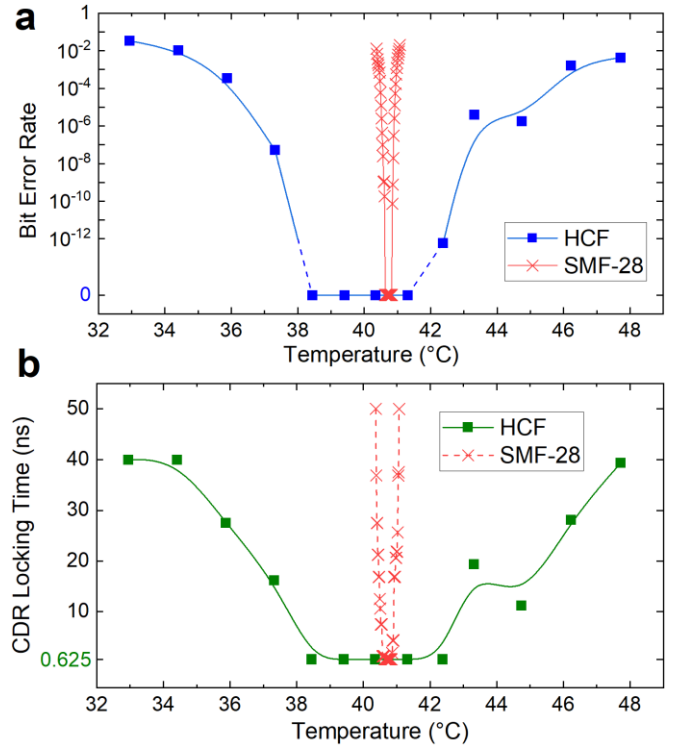


Fig. 5. Impact of temperature variation on the performance of an optically-switched system. (a) BER and (b) CDR locking time. Blue marker: measured BER using HCF; Red cross marker: estimated for SMF-28. Green: CDR locking time.

#### V. DISCUSSION

Our results show that the use of HCF should allow for short-reach data center interconnection (e.g. less than 2 km fiber length) without any active phase tracking that is required in SMF-28 based clock-synchronized networks. As discussed in the Section II, the change of time-of-flight through a fiber is linearly proportional to both the change in fiber temperature and the fiber length [9]. Therefore, based on our experimental results, i.e.  $<625$  ps CDR locking time over a temperature range of 2 °C for 2 km of HCF, an HCF based optically-switched system would be able to achieve 20°C temperature tolerance for fiber length of up to 200 m. This would provide sufficient temperature tolerance [8] for connecting servers within a cluster, with only an initial calibration of the clock phase values.

In addition to the benefits from low TCD, the HCF has the advantages of ultra-wide transmission window (e.g.,  $>100$  nm [12]), promising potential coarse wavelength division multiplexing (CWDM) for low-cost 800 Gb/s and above data interconnection. The additional features like low nonlinearity and low chromatic dispersion across a broad wavelength range potentially promises high power budget for low-cost interconnections without consumed power and latency degradation necessary for the impairment compensation [18].

#### VI. CONCLUSION

We demonstrated that clock-synchronized data center systems can have 20 times greater tolerance to temperature variation using HCF. Error free transmission with  $<625$ ps clock recovery

was achieved without the need to actively track the clock phase of incoming optical packets. This confirms that HCF provides a robust solution to short-reach optically-switched data center interconnection in which stable fiber latency enables high-capacity, high throughput, modulation format and data rate agnostic, and low-latency solution. We believe that the feature of stable latency could support the growing need for many other latency-stable optical fiber communication systems such as passive optical networks and fiber-wireless systems.

## REFERENCES

- [1] H. Ballani, P. Costa, I. Haller, K. Jozwik, K. Shi, B. Thomsen and H. Williams, "Bridging the last mile for optical switching in data centers," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018.
- [2] I. L. Markov, "Limits on fundamental limits to computation," *Nature*, vol. 512, p. 147, 2014.
- [3] N. Zilberman, G. Bracha and G. Schuzkin, "Stardust: Divide and conquer in the data center network," in *16th {USENIX} Symposium on Networked Systems Design and Implementation (NSDI) 19*, 2019.
- [4] H. J. S. Dorren, E. H. M. Wittebol, R. Kluijver, G. G. Villota, P. Duan and O. Raz, "Challenges for optically enabled high-radix switches for data center networks," *Journal of Lightwave Technology*, vol. 33, pp. 1117-1125, 2015.
- [5] A. Cevrero, I. Ozkaya, T. Morf, T. Toifl, M. Seifried, F. Ellinger, M. Khafaji, J. Pliva, R. Henker, N. Ledentsov and others, "4x40 Gb/s 2 pJ/bit Optical RX with 8ns Power-on and CDR-Lock Time in 14nm CMOS," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018.
- [6] K. Clark, H. Ballani, P. Bayvel, D. Cletheroe, T. Gerard, I. Haller, K. Jozwik, K. Shi, B. Thomsen, P. Watts and others, "Sub-nanosecond clock and data recovery in an optically-switched data centre network," in *2018 European Conference on Optical Communication (ECOC)*, 2018.
- [7] G.8261, *Timing and synchronization aspects in packet networks*, ITU, 2008.
- [8] A. S. H. R. A. E. TC9.9, *Data Center Networking Equipment--Issues and Best Practices*, ASHRAE Atlanta, GA, 2011.
- [9] R. Slavík, G. Marra, E. N. Fokoua, N. Baddela, N. V. Wheeler, M. Petrovich, F. Poletti and D. J. Richardson, "Ultralow thermal sensitivity of phase and propagation delay in hollow core optical fibres," *Scientific reports*, vol. 5, p. 15447, 2015.
- [10] E. N. Fokoua, M. N. Petrovich, T. Bradley, F. Poletti, D. J. Richardson and R. Slavík, "How to make the propagation time through an optical fiber fully insensitive to temperature variations," *Optica*, vol. 4, pp. 659-668, 2017.
- [11] D. N. Payne and A. H. Hartog, "Determination of the wavelength of zero material dispersion in optical fibres by pulse-delay measurements," *Electronics Letters*, vol. 13, pp. 627-629, 10 1977.
- [12] T. Bradley, J. Hayes, L. Hooper, H. Sakr, G. Jasion, M. Alonso, A. Taranta, A. Saljoghei, M. Fake, I. Davidson, Y. Chen, N. Wheeler, E. Fokoua, W. Wang, S. R. Sandoghchi, D. Richardson, F. Poletti and H. C. Mulvad, "Antiresonant hollow core fibre with 0.65 dB/km attenuation across the C and L telecommunication bands," in *2019 European Conference on Optical Communication (ECOC)*, 2019.
- [13] A. Nespoli, S. Straullu, T. Bradley, H. C. Mulvad, J. Hayes, G. Jasion, M. Gouveia, S. R. Sandoghchi, S. Bawn, F. Forghieri, D. Richardson, F. Poletti and P. Poggiolini, "Record PM-16QAM and PM-QPSK Transmission Distance (125 and 340 km) over Hollow-Core-Fiber," in *2019 European Conference on Optical Communication (ECOC)*, 2019.
- [14] F. Poletti, N. V. Wheeler, M. N. Petrovich, N. Baddela, E. N. Fokoua, J. R. Hayes, D. R. Gray, Z. Li, R. Slavík and D. J. Richardson, "Towards high-capacity fibre-optic communications at the speed of light in vacuum," *Nature Photonics*, vol. 7, p. 279, 2013.
- [15] *Si5324 Any-Frequency Precision Clock Multiplier / Jitter Attenuator*.
- [16] 802.3bm, *IEEE Standard for Ethernet - Amendment 3: Physical Layer Specifications and Management Parameters for 40 Gb/s and 100 Gb/s Operation over Fiber Optic Cables*, IEEE, 2015.
- [17] K. Shi, S. Lange, I. Haller, D. Cletheroe, R. Behrendt, B. Thomsen, F. Karinou, K. Jozwik, P. Costa and H. Ballani, "System Demonstration of Nanosecond Wavelength Switching with Burst-mode PAM4 Transceiver," in *2019 European Conference on Optical Communication (ECOC)*, 2019.
- [18] Z. Liu, B. Karanov, L. Galdino, J. R. Hayes, D. Lavery, K. Clark, K. Shi, D. J. Elson, B. C. Thomsen, M. N. Petrovich, D. J. Richardson, F. Poletti, R. Slavík and P. Bayvel, "Nonlinearity-Free Coherent Transmission in Hollow-Core Antiresonant Fiber," *Journal of Lightwave Technology*, vol. 37, pp. 909-916, 2 2019.