5G Mobile Fronthaul Over Latency Optimized Optical Networks

Anju Jose[1],Prasanth P.Menon[2]

[1]PG Scholar [2]Assistant Professor
Department Of Electronics and Communication Engineering
Adi Shankara Institute of Engineering and Technology
Kalady,India

Abstract—In 5th generation mobile communications systems (5G), the number of small cells with centralized-radio access network (C-RAN) architecture is expected to increase greatly. Best solutions today are based on expensive point-to-point (PtP) fiber links and PtP wavelength division multiplexing. A time division multiplexed passive optical network (TDM-PON) is less expensive, but has a poor latency of about 1ms. Guidelines are suggested to design an optimized TDM-PON with minimal latency for transport of constant bit rate fronthaul traffic. The total latency is mathematically analyzed using queuing theory, and closed-form formulas are derived which allow study of the trade-offs enabled by the different parameters. Feasibility of acceptable levels of latency of less than 1 ms established in several application scenarios.

Keywords—*5G; mobile fronthaul; passive optical network; applicable range*

*****

I. INTRODUCTION

Telecommunication networks today are highly converged between the fixed and mobile regions. This is particularly evident when fixed networks are used for mobile backhaul and fronthaul. Fig. 1shows how the fronthaul interface is applied in a TDM-PON system to support multiple remote sites for Centralized Radio Access Network (C-RAN).Backhaul refers to the networking between the Serving Gateway (S-GW) and the base stations (eNB). In a (CRAN) architecture, the baseband processing of the eNB is centralized in a Baseband Unit (BBU) pool and linked to the radio units, i.e., Remote Radio Heads (RRHs) via the fronthaul network, which is the focus of this paper.

The Common Public Radio Interface (CPRI) [1] is a commonly used standard for the interface between the BBUs and RRHs. The main problems in designing a CPRI fronthaul network are associated with the high constant bit rate (CBR) and low latency requirements. For example, an LTE 2 × 2 system with 10 MHz bandwidth requires 1.2288 Gb/s per carrier per sector. To support such high bit rates, CPRI defines line bit rate options 1 through 8 ranging from 614.4 Mb/s to 10.1376 Gb/s. In addition, taking into account the total end to-end latency budget, the latency for the fronthaul network should typically be less than 250 us [2]. This is almost 10 times smaller than the ∼ 2 ms latency available for backhaul. Given these tight requirements, fiber-based solutions are well-suited for fronthaul.

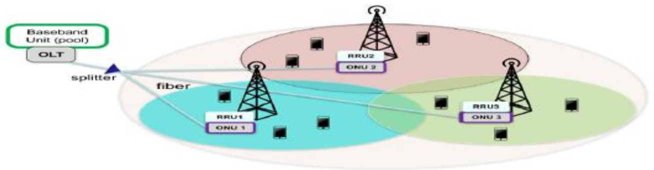

Fig. 1. Mobile PON for C-RAN[2]

The object of this paper is to investigate the feasibility of fronthaul over time division multiplexed passive optical networks (TDM-PONs). In the process, we study how an optimized TDM-PON can be designed for transparent optical transport of CBR traffic with minimal latency. We then mathematically analyze, applying results from queuing theory, all the components contributing to the fronthaul latency including from the TDM of several CPRI links to establish scenarios where the tight latency requirements can be met.

Previous works such as [5]–[8] have proposed and studied the use of PONs for backhaul. Aurzada et al. [9] analyze the latency contributions in a standard TDM-PON with dynamic traffic which our analysis takes some inspiration from. Furthermore, some works have studied the used of PONs specifically for fronthaul which are highlighted in Sec. II. The contributions of paper are as follows: we outline guidelines for a latency-optimized TDM-PON, analyze the total fronthaul latency, and show that a one-way latency less than 250 us is possible to achieve in several application scenarios.

III. TDM-PON BASED FRONTHAUL CHALLENGE

An obvious solution candidate for fronthaul is based on point-to-point (PtP) fibers. However, this is not attractive due to the amount of fiber and transceivers required. When the available fiber is limited, wavelength division multiplexing (WDM) may be used where a dedicated wavelength channel is used for each CPRI link, but this is also expensive. ITUT G.989 (NGPON2) with 40 Gb/s capable PONs based on time and wavelength division multiplexing (TWDM) provisions for fronthaul. Nevertheless, the primary method is via a PtP wavelength overlay or by using the wavelength channels of the TWDM-PON in effect for PtP transport as shown in [10].

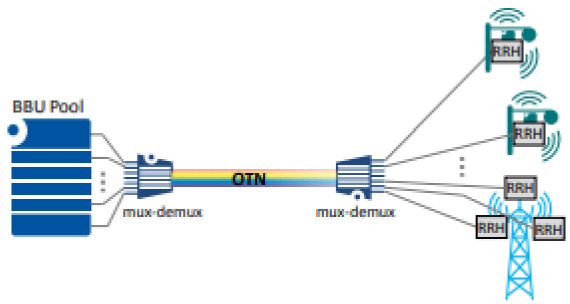

Fig. 2. Fronthaul using ITU-T G.709 Optical Transport Network (OTN)[2]

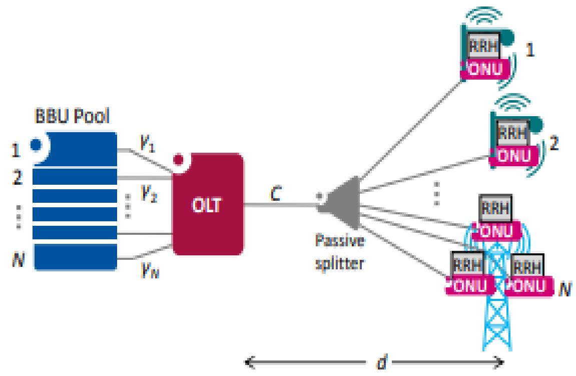

Fig. 3.Fronthaul using TDM-PON[1]

ITU-T has added a framework for transporting CPRI signals in a transparent manner over G.709 Optical Transport Network (OTN) [10], This includes the capability to time division multiplex (TDM) several CPRI links onto a single wavelength channel, e.g., 4 CPRI rate3 (2.4576 Gb/s) links multiplexed onto a 10 Gb/s channel. OTN however requires a symmetric multiplexer-demultiplexer (mux-demux) on each end of the link as shown in Fig 2. Thus, although OTN could be used on a PtP wavelength overlay taking advantage of an existing PON distribution network, a separate PtP fiber would be needed from the mux-demux to each RRH which can be troublesome depending on their location, while instantiating several mux-demux's could lead to high cost.

We focus on a TDM-PON based fronthaul solution because of it's simplicity and commonly available optical distribution network, thus also leading to a lower cost, as also observed in [9]. Fig. 3 shows a TDM-PON based fronthaul network consisting of a BBU pool with multiple CPRI links multiplexed at the optical line terminal (OLT), where the TDM signal travels over the PON distribution network, and is received by several optical network units (ONUs) each connected to an RRH.A standard TDM-PON is not directly suitable for fronthaul due to its large latency in the order of 1ms [6], [8], [9]. The latency is particularly pronounced on the upstream direction as the ONU waits to receive a dynamic bandwidth allocation (DBA) grant sent by the OLT before beginning upstream transmission. Previous works have proposed low latency DBA schemes [8]. Nevertheless, a complete analysis of all the latency contributions when several CPRI links are multiplexed on a TDM-PON is not available in the literature to the best of our knowledge. The need for low latency optical transport over PON is also highlighted in [5] in the context of backhaul and for coordination between base stations. In addition to the latency, standard TDM-PONs also require a quiet window where all transmissions are prevented for ranging new ONUs which can disrupt the CBR traffic flow of CPRI [3].

IV. FRONTHAUL LATENCY ANALYSIS

We use N to denote the number of fronthaul links between the BBU pool and RRHs. Thus, the links, ONUs, and RRHs are indexed by i = 1, . . . , N as shown in Fig. 3. The bit rate on a specific fronthaul link i is given by γi for i = 1, . . . , N, while the capacity of the PON is denoted by C. The capacity and bit rates are symmetric for both downstream and upstream. For example, the value for each γi can be chosen to be any of the line bit rate options 1 through 8, ranging from 614.4 Mb/s to 10, 137.6 Mb/s, specified by CPRI. di denotes the distance over the optical distribution network from the BBU to RRH i. Propagation time (seconds) over a distance d (km) is calculated at 5 us/km and thus given by P = P(d) = d/(2 × 10^5).

In order to multiplex several fronthaul links on to the same PON, the CPRI data is segmented into blocks on the PON, and again re-assembled at the PON termination. Once a block arrives on a specific fronthaul link, it is transmitted on a fixed time slot on the PON using a static bandwidth allocation Also, since CPRI rates may be different on each link, but the CPRI frame duration is always the same (66.67 us CPRI hyperframe, 10 ms CPRI frame), the duration of the block, denoted by B, is specified in seconds rather than in number of bits. For example, multiplexing at the hyper frame level, we have B = 66.67 us, while using the typical TDM-PON frame size, we have B = 125 us. In general, B may be chosen to be any value that is convenient for the multiplexing.

The compression ratio and the FEC code rate are denoted by κ and R respectively (with 0 < κ, R < 1). The number of header and guard bits added to the payload of each link is given by H and G respectively. Note that the guard bits factor in only in the upstream direction, i.e., G = 0 for downstream. Therefore, the transmission time from the OLT over the PON for a block from link i is

$$T_i = B.\Upsilon i.\frac{\kappa}{C.R} + \left[\frac{H+G}{C}\right] \quad (1)$$

In general, the overall latency for link i is given by

$$D_i = B + W_i + T_i + P(d_i) + S_i + E_c + E_f + E_m \quad (2)$$

where B is the transmission time from the BBU to the OLT; Wi is the waiting time in the OLT queue; Ti is the transmission time over the PON ; $P(d_i)$ is the propagation time over the PON; Si is defined as any smoothing time that is added e.g., to maintain synchronization across links; and $E_c, E_f, and\ \ E_m$ are the respective electronics delay for compression &decompression, FEC encoding & decoding, and any other miscellaneous electronics (e.g., scrambling & descrambling, reading from & writing to memory, etc.).

A. Case 1: Synchronized phases

In the first case, we consider the scenario where the traffic phase on all the CPRI links are synchronized. This means that the CPRI frames arrive at the OLT on all links at the same instant and need to also arrive at the RRHs in a synchronized manner. This is the typical scenario when all the BBUs in the pool belong to the same operator network. Fig. 4 shows N = 3 links with synchronized CPRI traffic arriving from the BBUs. For simplicity and compactness, we don't show the compression and FEC encoding stages in Fig. 4 and take the traffic as arriving directly at the compressed and encoded bit rate.

We explain the delay computation taking link 2 in Fig. 4 for illustration. After the transmission time to the OLT of B seconds, link 2 experiences a waiting timeW_2 corresponding to the transmission time $T_1$of link 1 over the PON. This is followed by the transmission time$T_2$and propagation time $P(d_2)$ for the block to be received at ONU 2. Each link similarly experiences a delay of W_i+ T_i+ $P(d_i)$. In order to maintain synchronization between the links, the smoothing time Si needs to be added so that all links coincide with the worst delay resulting in

$$D_i^{sync} = B + \max_i(W_i + T_i + P(d_i)) + E_c + E_f + E_m \quad (3)$$

Nevertheless, since the number of links and distances could change between deployments and also over time, a simpler approach could be preferred in practice. Since$W_i + T_i$≤ B is always true and the system is designed with a maximum distance d_{max} in mind, the maximum for $W_i + T_i$and $P(d_i)$ can be separately accounted to yield

$$D_i^{sync} = 2B + P(d_{max}) + E_c + E_f + E_m \quad (4)$$

Some blocks are buffered at the OLT due to contention, but at most N blocks are in the buffer at any given time.

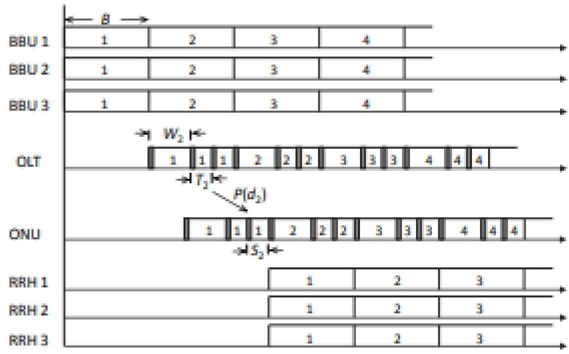

Fig. 4. 3 links with synchronized phases (Case 1)

B. Case 2: Unsynchronized with controlled phases

In the second case, we consider the scenario where the CPRI traffic on the links does not need to be synchronized and the system designer has control over the phase of the traffic on these links. Such control could potentially be available when all the BBUs in the pool belong to the same operator network. Specifically, if the phase of the incoming traffic on each link is offset such that a block arrives at the OLT exactly when a slot is available to transmit over the PON, then the waiting time in the OLT queue is reduced to 0 as all the incoming traffic essentially "passes through". Without loss of generality, we take link 1 as the reference and identify the perfect phase staggering with respect to this link

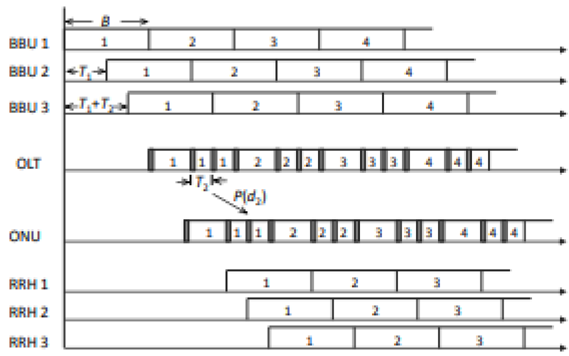

Fig. 5. 3 links with unsynchronized and controlled phases (Case2)

Specifically, this means that if each link i = 2, . . .N is staggered in phase by $\sum_{j=1}^{i-1} T_j$, then a block on link i arrives exactly at the time when it can be transmitted on the PON. So, the waiting timeW_i= 0 for all i. This is illustrated in Fig. 5 using the same example as in Case 1, but with the links offset as described above. Furthermore, since the links are unsynchronized, no smoothing is needed, i.e., S_i = 0. Therefore, the total latency for link i is given by

$$D_i^{unsync-c} = B + T_i + P(d_i) + E_c + E_f + E_m (5)$$

National Conference on Recent Trends in VLSI, Communication and Networks 2018 (RTVCN2K18)
Volume: 4 Issue: 5
ISSN: 2454-4248
37 – 42

C. Case 3: Unsynchronized with random phases

The final case is motivated by a scenario where BBUs could belong to multiple operators. Here, it is expected that the BBUs in a pool may not be synchronized and so may generate traffic independently of each other. In this case, the relative phase difference has a uniformly random distribution over [0, B] due to the periodicity of the frame arrivals. Therefore, in equation (3), $W_i$becomes a random variable, and S_i= 0 since no smoothing is needed as in case 2, so that

$$D_i^{unsync\ -r} = B + W_i + T_i + P(d_i) + E_c + E_f + E_m \quad (6)$$

The uplink transmission of each item of user equipment (UE) isscheduled by a BBU and controlled using a downlink control indicator (DCI). When data arrive at an ONU, the ONU sends a request message to start data transmission. An OLT allocates
bandwidth with request messages from ONUs and sends them grant messages. After that, the ONU finally begins transmission at the time indicated by the grant message. .Transmission division is an effective technique for reducing the waiting time and the sending time. Fig. 4 compares typical examples of transmission division. In the mobile system transmission is managed in transmission time interval (TTI)units. Assuming that the wireless transport rate is lower thanthe PON bandwidth, the data transmission time in MFH islonger than that in a PON. And generally, the OLT allocates the bandwidth after all the data have arrived at the ONU. The transmission division divides data in TTI units to the transportpackets..

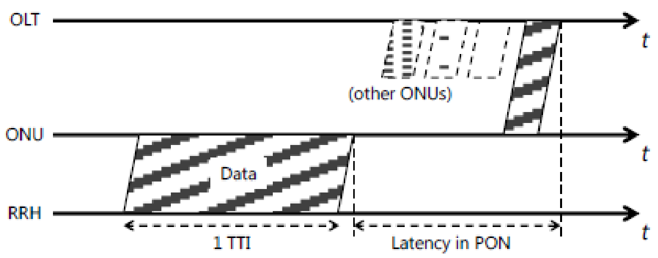

Fig. 6.One transport packet (W/o transmission division)

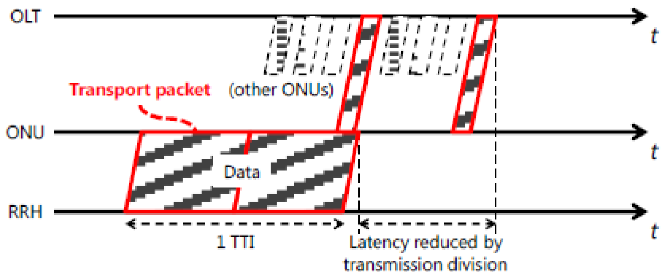

Fig. 7. Two transport packet (With transmission division)

Here, if we define target latency as the maximum time between all the data arriving at the ONU and subsequentlyarriving at the OLT, L_{max} can be expressed as

$$L_{max} = \left[\frac{T_{TTI} . R_W . R_{FEC} R_{MFH} .}{B_{PON} . N_{TP} .}\right] N_{ONU} - T_{OFF} + T_{PROP} + T_{PROC} \quad (7)$$

It is clear that a lower maximum latency can be achievedwith a larger number of transport packets from (7).where $T_{TTI}$is the TTI length, $R_W$is the wireless transport rate, $R_{FEC}$is the forward error correction (FEC) rate,$R_{MFH}$is the rate of the increase caused by MFH overheads, and $B_{PON}$is the PON bandwidth,$N_{TP}$is the no.of transport packets.However, while transmission division can reduce latency, italso reduces the amount of allocable PON bandwidth due to anincrease in burst overheads.The total MFH bandwidth must not exceed the PONbandwidth, so the number of accommodable ONUs (RRHs)N_{ONU}should meet the following equation

$$B_{PON} \geq \sum_{i=1}^{N_{ONU}} B_{MFH}(i) \quad (8)$$

where *i* is the index for ONUs. If the ONUs have the same parameters , can be derived asN_{ONU}

$$N_{ONU} \leq \frac{B_{PON} T_{TTI}}{T_{TTI} R_W R_{FEC} R_{MFH} + T_{BURST} N_{TP} B_{PON}} \quad (9)$$

Transmission division can be realized by allocating multiple grants in a single control message or by shortening the bandwidth allocation frequency. It should be noted thatshortening the bandwidth allocation frequency increases the number of control frames, which occupies more of the PON downlink bandwidth.

ISSN: 2454-4248
37 – 42

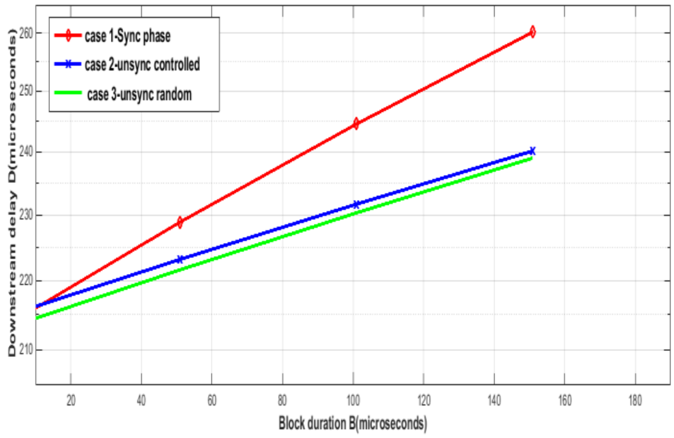

Fig. 8. Downstream delay with 4 CPRI links @1.2G over opt-GPON

V. NUMERICAL RESULTS AND DISCUSSION

It remains to characterize the probability distribution of Wi in the above equation. For this analysis, we assume all γi are equal. Therefore, the waiting times Wi are identically distributed. This identical distribution is represented by the generic random variable W, which can be viewed as the waiting time in the OLT queue for an arbitrary block from the superposition of the arrival streams.

The analysis shows a clear dependence of the total latency Di on the value chosen for the block duration B. We first discuss this dependence in some examples. Suppose we have 4 BBUs each transmitting at γ_i = 1.2288 Gb/s (CPRI rate2). The total rate after compression, FEC coding, and multiplexing is sent over an optimized PON with capacity of 2.48832 Gb/s (similar to GPON but symmetric), which for convenience will be referred to as opt-GPON. Since the γ_i are equal, the delays D_i are the same for all the links

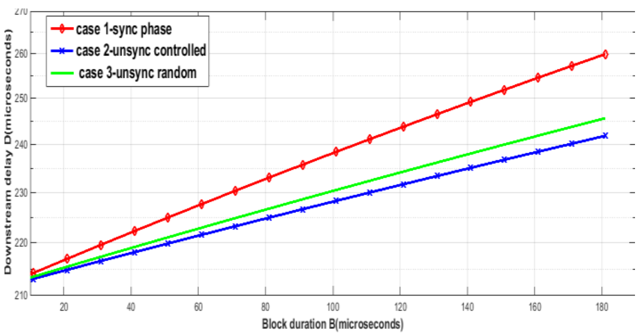

Fig. 9.Downstream delay with 19 CPRI links @1.2G over opt-XGPON

Fig. 8 shows the downstream delay as a function of the block duration B showing a linear dependence in all cases. Case 1 has the highest delay, while Case 2 has the smallest. In Case 3 with random phases, we account for the mean waiting time given by equation (5), which turns out to be quite close to the smallest delay. However, if a delay guarantee is needed, accounting for the mean does not suffice and this is discussed later. It can be observed that taking B = 125 us will violate the 250 us requirement in Case 1 and will be very close for the remaining cases. Taking B = 66.67 us provides a convenient option corresponding to the CPRI hyper frame while satisfying the 250 us requirement in all cases. A larger propagation distance (say 20 km) can be accommodated by choosing B to be even smaller (say 50 us), but at the expense of some increase in the overhead due to the headers. In order to confirm that the above behavior is not particular to the scenario considered, we repeat this with 19 BBUs each transmitting at 1.2288 Gb/s (CPRI rate2) over an optXGPON with capacity of 9.95328 Gb/s. Fig. 9 confirms these observations.

TABLE 1.SIMULATION PARAMETERS

Parameter	Value
TTI lengthT_{TTI}	1[ms]
Wireless transport rate R_W	86.4,500,1000,3000[Mbps]
PON bandwidth B_{PON}	10 [Gbps]
FEC rateR_{FEC}	1.14
Guard time for laser ON T_{ON}	25.6[ns]
Synchronization timeT_{SYNC}	742.4 [ns]
Time for burst delimiterT_{BURST}	6.4[ns]
Guard time for laser OFFT_{OFF}	25.6[ns]
Increasing rate caused by MFH overheads R_{MFH}	1.14
Propagation timeT_{PROP}	10[μs]
Processing timeT_{PROC}	10[μs]

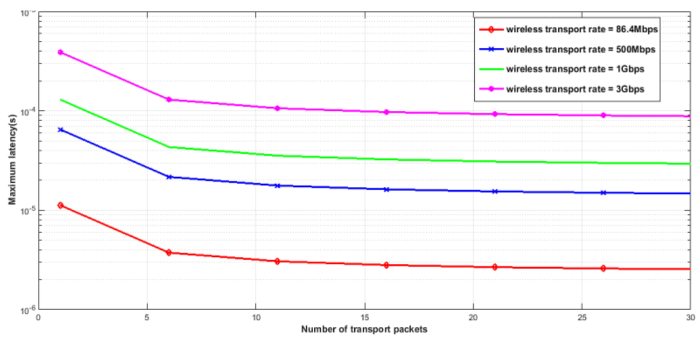

Fig. 10. Maximum uplink latency versus number of transport packets

Fig. 10 shows the MATLAB simulation of achievable maximum latency performance as a function of the number of transport packets. The values of the main parameters are shown in Table I. We assumed that all the ONUs (RRHs) had the same parameters. The wireless transport rates of the RRHs were set at 86.4, 500 Mbps, 1 Gbps and 3Gbps assuming uplink transmissions of LTE, LTE-A, and 5G,respectively. Thisresult shows that a latency of even less than 150 µs is achievable.A lower maximum latency can be achieved as the number oftransport packets increases. The achievable maximum latencyincreases as the wireless transport rate increases because ittakes a longer time to send data.

V. CONCLUSION

We suggest guidelines on how a TDM-PON can be designed for minimal latency transport of CPRI CBR traffic. We mathematically analyze the total fronthaul latency consisting of all the contributions and obtain closed-form formulas allowing to explore the trade-offs between different parameters. We show several scenarios where TDM-PONs can achieve a latency less than 250 us in downstream and less than 150 µs in upstream direction and thus establishing them as a valid candidate for mobile fronthaul. While the analysis here establishes a first level of feasibility, implementation and experiments are further needed to test the concept, while addressing practical challenges such as jitter, drift, and synchronization.

REFERENCES

[1] Anthapadmanabhan, N. Prasanth, Anwar Walid, and Thomas Pfeiffer. "Mobile fronthaul over latency-optimized time division multiplexed passive optical networks." *Communication Workshop (ICCW), 2015 IEEE International Conference on*. IEEE, 2015

[2] Zhou, Siyu, Xiang Liu, Frank Effenberger, and Jonathan Chao. "Low-Latency High-Efficiency Mobile Fronthaul With TDM-PON (Mobile-PON)." *Journal of Optical Communications and Networking* 10, no. 1 (2018): A20-A26.

[3] M. Maier, “FiWi access networks: Future research challenges and moonshot perspectives,” 2014 IEEE Intl. Conf. on Commun. Workshops (ICC), Jun 2014

[4] Zhou, Siyu, Xiang Liu, Frank Effenberger, and Jonathan Chao. "Mobile-PON: a high-efficiency low-latency mobile fronthaul based on functional split and TDM-PON with a unified scheduler." In *Optical Fiber Communication Conference*, pp. Th3A-3. Optical Society of America, 2017..

[5] F. Aurzada, M. Scheutzow, M. Reisslein, N. Ghazisaidi, and M. Maier, “Capacity and delay analysis of next-generation passive optical networks (NG-PONs),” IEEE Trans. on Communications, vol. 59, no. 5, pp. 1378–1388, May 2011.

[6] Y. Luo, J. Gao, and F. Effenberger, “Next generation hybrid wirelessoptical access with TWDM-PON,” 2014 23rd Wireless and Optical Communication Conference (WOCC), pp. 1–5, May 2014.

[7] T. Tashiro et al., “A novel DBA scheme for TDM-PON based mobile fronthaul,” 2014 Optical Fiber Commun. Conf. (OFC), Mar 2014.

[8] Kani, Jun-ichi, Jun Terada, Ken-Ichi Suzuki, and Akihiro Otaka. "Solutions for future mobile fronthaul and access-network convergence." *Journal of Lightwave Technology* 35, no. 3 (2017)

[9] Tashiro, Takayoshi, Shigeru Kuwano, Jun Terada, Tomoaki Kawamura, Nobuyuki Tanaka, Satoshi Shigematsu, and Naoto Yoshimoto. "A novel DBA scheme for TDM-PON based mobile fronthaul." In *Optical Fiber Communications Conference and Exhibition (OFC), 2014*, pp. 1-3. IEEE,2014.

[10] Okumura, Yukihiko, and Jun Terada. "Optical network technologies and architectures for backhaul/fronthaul of future radio access supporting big mobile data." In *Optical Fiber Communication Conference*, pp. Tu3F-1. OpticalSocietyofAmerica,2014.

[11] Liu, Xiang, and Frank Effenberger. "Emerging optical access network technologies for 5G wireless." *Journal of Optical Communications and Networking* 8,no.12(2016):B70-B79.

[12] Bae, S. H., H. K. Shim, U. H. Hong, Hoon Kim, A. Agata, K. Tanaka, M. Suzuki, and Y. C. Chung. "25-Gb/s TDM optical link using EMLs for mobile fronthaul network of LTE-A system." *IEEE Photonics TechnologyLetters* 27,no.17(2015):1825-1828.