Higher Speed Passive Optical Networks for Low Latency Services

Abstract: Latency sensitive services have attracted much attention lately and imposed stringent requirements on the access network design. Passive optical networks (PONs) provide a potential long-term solution for the underlying transport network supporting these services. This paper discusses latency limitations in PON and recent progress in PON standardization to improve latency. Experimental results of a low latency PON system are presented as a proof of concept.

Keywords: passive optical networks; time-division multiple access; wavelength-division multiple access; low latency

ZHANG Weiliang, YUAN Liquan
(Wireline Product Planning Department, ZTE Corporation, Shanghai 201203, China)

DOI: 10.12142/ZTECOM.202102008

http://kns.cnki.net/kcms/detail/34.1294.TN.20210510.1514.002.html, published online May 10, 2021

Manuscript received: 2021-03-16

Citation (IEEE Format): W. L. Zhang and L. Q. Yuan, "Higher speed passive optical networks for low latency services," *ZTE Communications*, vol. 19, no. 2, pp. 61–66, Jun. 2021. doi: 10.12142/ZTECOM.202102008.

1 Introduction

With the continued growth of new applications and services over fixed and wireless communication networks, requirements for low latency, high bandwidth, timing and synchronization have taken central stage in new network architecture design. In 2015, the ITU-R laid out its vision on the framework and objectives of the development of International Mobile Telecommunications (IMT) for 2020 and beyond[1]. In the IMT-2020 Recommendation, diverse services for ultra-reliable low-latency communications (URLLC), as well as for enhanced mobile broadband (eMBB) and massive machine type communications (mMTC), are envisioned.

These envisioned services translate to stringent requirements for the underlying transport network layer. Many fixed network technologies are being considered for the 5G transport infrastructure, e. g., point-to-point fibers, active wavelength division multiplexing (WDM), and passive optical networks (PONs).

Among these options, the PON stands out as a highly suitable choice. Due to its efficient fiber infrastructure and bandwidth efficiency, the time-division-multiplexed (TDM) PON has been successfully deployed worldwide to over 626 million subscribers as of December 2019. Many of the PON and 4G devices share the same access office. In addition, PON and 5G transport networks share similar network topology. It is therefore attractive to make use of the abundant fiber resources in the PON infrastructure.

Many industry standards development organizations (SDOs) are working on new standardization projects to address the increasing demands of low-latency services[2]. The Full Service Access Network (FSAN) group[3] and the ITU-T Study Group (SG) 15 Question 2 (Q2), which focuses on optical access networks standardization, have conducted several projects to study PONs for 5G mobile x-haul transport, which will be discussed later in the paper.

This paper is structured as follows. We will start with examples of low latency services in Section 2. An overview of passive optical networks and their latency properties are discussed in Section 3, which is followed by recent progress in PON standards to support low-latency services in Section 4. Finally in Section 5, we describe in more detail a latency reduction method for TDM PON and present experimental results as a proof of concept.

2 Low Latency Services

Low latency services are characterized as services of which the latency requirement between service end points is much lower than the latency needed in traditional services, e.g., internet browsing, files/data download, and IPTV. In these traditional services, there is often no clear boundary of time limitation. Some of the most prominent examples of low latency services include 5G x-haul transport for supporting URLLC, virtual reality (VR)/augmented reality (AR) video services, industry applications for factory networks, and robotic control[4].

A 5G x-haul transport network is the underlying transport layer providing fronthaul connectivity between the 5G remote unit (RU) and the distributed unit (DU), midhaul connectivity between DU and the centralized unit (CU), and backhaul connectivity between CU and the 5G core. More details of its latency requirements will be described in Section 4.

VR/AR video services provide immersive viewing experience for end users. They require the data traffic between the server and clients be transported in very short duration to meet the latency requirements of the motion-to-photon (MTP) and motion-to-audio. The maximum end-to-end latency is 20 ms, of which a much lower value is allotted for the access network segment, e.g., the PON.

Industry applications are much more complex than 5G x-haul and AR/VR video services. The requirements of timing and latency for the control messages are much more critical than normal services in some cases. For example, for industrial robotic control, there is a need to synchronize the robots with each other, which could limit the latency to less than 10 ms in specific scenarios discussed by the ETSI F5G group. Currently, the ETSI F5G group is studying network design on how to use PON technology for industry applications.

3 Overview of Passive Optical Networks

In this section, we provide an overview of three types of PONs and discuss their latency properties: TDM PON, WDM PON, and time and wavelength division multiplexed (TWDM) PON. TDM and TWDM PONs are mainly used for residential services such as Fiber to the Home, while WDM PONs are for business services due to the higher cost. Therefore, these systems have quite different latency requirements.

3.1 TDM PON

In a TDM PON, as shown in Fig. 1, signals from the optical line terminal (OLT) are broadcasted downstream to all the optical network units (ONUs) in a TDM fashion. In the upstream direction, each ONU transmits its signal in a time slot assigned by the OLT through the dynamic bandwidth allocation (DBA) process.

As specified in the ITU-T G.989.3 Recommendation[5], the DBA engine consists of a bandwidth assignment component and a bandwidth map (BWmap) generation component. The bandwidth assignment component computes the assigned bandwidths for every DBA cycle. The assigned bandwidths are then supplied to the BWmap generator to generate a BWmap once every physical layer (PHY) frame (125 μs).

In a typical TDM PON, the DBA process could result in an upstream latency in the order of several milliseconds because data transmission may take several DBA cycles to complete. When using a conventional TDM PON for wireless x-haul transport, the upstream data from DU and/or RU must wait in the ONU until the completion of DBA.

Another latency causing process for a TDM PON is the need of a quiet window during the activation and registration of new ONUs onto an operational TDM PON. No ONU can transmit upstream data during this period, which can last over a few 100 μs. Such a latency value is incompatible with low latency mobile x-haul applications.

3.2 WDM PON

WDM PON is a logical point-to-point system, as shown in Fig. 2. Signals from the OLT, each transmitted over a different wavelength channel, are combined in a wavelength multiplexer before transmitting to the end user. In the optical distribution network (ODN), a wavelength splitter routes the individual wavelengths to different ONUs. In the case of WDM PON for mobile x-haul services, each of the ONU can be connected to an RU supporting one of the three sectors of an antenna at the cell site. The latency is purely limited by the transmission distance and processing delay. Therefore, no special latency improving mechanism is needed.

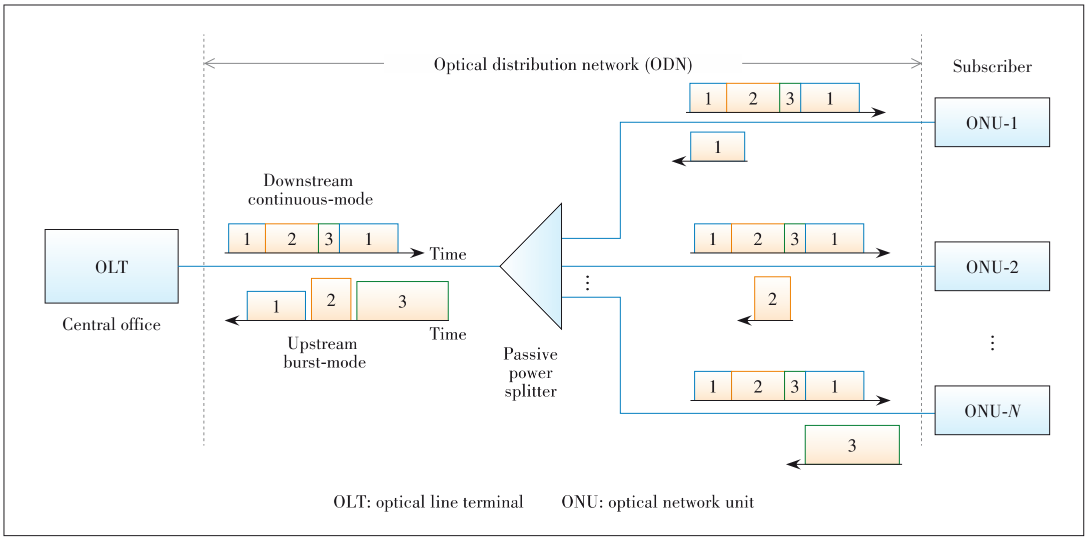
3.3 TWDM PON

TWDM PON is a combination of TDM and WDM PONs[6]. Its latency limitation and methods for improvement thus follow the description in Section 3.1.

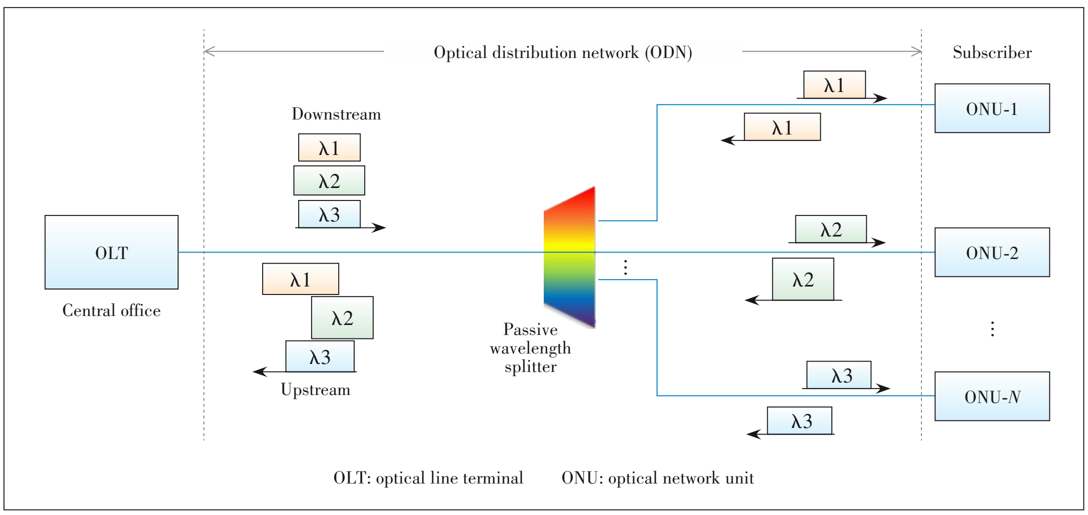
4 PON Standards Supporting Low Latency Services

As mentioned earlier, the ITU-T Q2/SG15 group is leading the efforts on standardizing optical access networks including PON for low latency services. The group began in June 2017 the "5G Wireless Fronthaul Requirements in a PON Context" project to analyze specifications from 5G standards, PON system requirements, and practically realizable PON architectures. Results of the study were agreed in October 2018 and published in the supplementary document G. Sup66[7]. In addition, the Q2 group has completed the standard for single fiber bidirectional point-to-point optical access system covering line rates of 10 Gbit/s, 25 Gbit/s, and 50 Gbit/s in the G.9806 Recommendation to support 5G x-haul and business services[8–9].

In this section, we provide a high-level summary of the findings in Supplement G.Sup66, with a focus on the latency aspect[7, 10]. We will first describe the requirements from the wireless transport network perspective. We then discuss modifica-



▲Figure 1. Schematic of a typical time-division-multiplexed (TDM) passive optical network (PON)



▲Figure 2. Schematic of a typical wavelength-division multiplexing (WDM) passive optical network (PON)

tions to existing PON standards for practical PON implementations to meet the transport latency requirements.

4.1 Wireless Transport Network Latency Requirement

In a centralized radio access network (C-RAN) functional split architecture, the industry has converged on two interfaces: the F1 interface (midhaul/backhaul) for the high layer split Option 2, and the Fx interface (fronthaul) for the low layer functional split Option 6 or 7 (Fig. 3). In both cases, the transport capacity varies with the actual aggregated user traffic on the air interface. This is an essential feature which allows for applying more bandwidth and cost efficient x-haul networks.

The transport at the F1 interface is very similar to backhaul transport. The required end-to-end latency is in the order of tens of milliseconds for eMBB and in the 1 ms range for URLLC services, which leaves a sub-millisecond range for the transport layer.

At the Option 6 (media access control (MAC)-PHY split), Option 7 (Intra PHY split) and Option 8 (PHY-RF split), the acceptable transport latency is in the range of a few 100 µs, similar to LTE, for eMBB and non-realtime mMTC services. Such low latency tolerance plays a critical role in the fronthaul network design. The Common Public Radio Interface (CPRI) Cooperation has defined transport network classes for categorizing the tolerances accordingly, ranging from 25 µs to 500 µs one way[11].

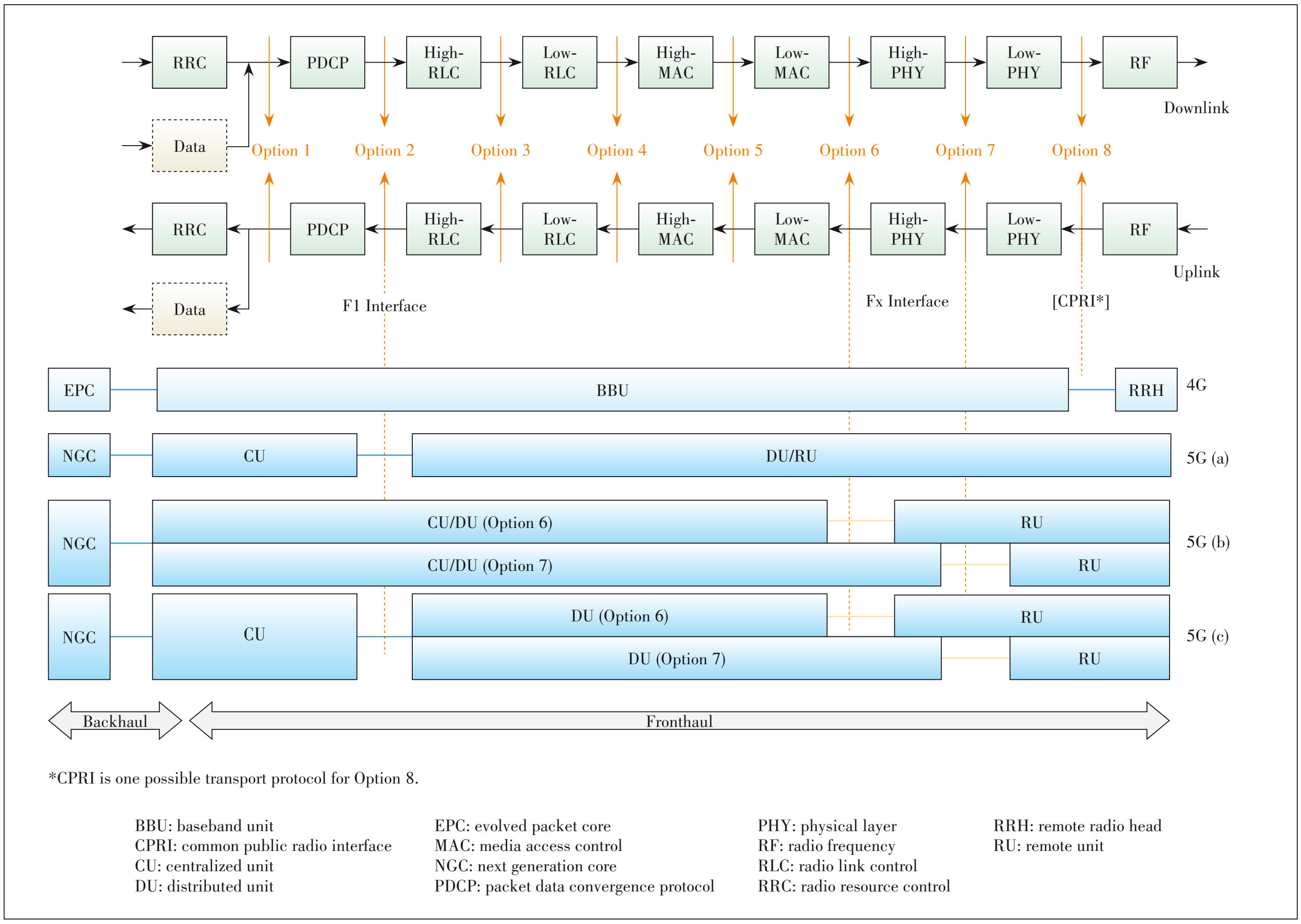
4.2 Modifications to PON Standards

For a practical PON system to meet the transport latency requirements described above, methods to reduce the processing delays must be implemented. Here we consider two typical PON systems: TDM-PON and WDM-PON.

For the TDM PON, as discussed in Section 3.1, there are two latency inducing factors, namely the DBA process and the quiet window during ONU activation.

To mitigate the DBA-induced latency, a straightforward method is to differentiate service classes, where the mobile traffic is assigned the highest priority with fixed bandwidth allocation. However, this leads to low bandwidth efficiency as any unused portion of the bandwidth cannot be reallocated.

Another method is the cooperative (CO) DBA, in which information exchange is introduced between the mobile scheduler (CU/DU) and the PON scheduler (DBA) in the OLT. This method allows the OLT to determine upstream bandwidth allocations in advance and then allocate the bandwidth at the expected arrival time of the upstream mobile traffic based on the actual traffic volume. This method is currently being studied in the ITU-T G.Sup.CODBA Supplementary project[12] in collaboration with the O-RAN group. In addition, traffic descrip-



▲Figure 3. Mapping of CU/DU/RU functions according to the split points: 5G(a) is high layer split (F1); 5G(b) is low layer split (Fx); 5G(c) is cascaded split (Reprint of Fig. 6-5 in G.Sup66[7])

tors in traditional DBA need to be extended to support low latency. These extensions, including jitter tolerance, bandwidth assignment delay tolerance and protection switching delay tolerance, are being added to the existing PON standard[5].

As for the latency due to quiet window opening, one proposal is to use a dedicated wavelength for ONU activation and registration. This dedicated activation wavelength (DAW) may be a newly defined wavelength, a separate PON system operating on a different wavelength on the same ODN, or a subset of wavelength channels in a TWDM PON. Once an ONU is activated in the DAW channel, it is handed over to the low latency operating wavelength channel to begin data transmission.

Another proposal is to use WDM PON, which does not require DBA nor ONU ranging. The latencies depend purely on the processing delays in the end nodes of the PON system and typically range below 10 µs. As such, the ITU-T Q2/SG15 group began a project in February 2020 to standardize WDM PON. The initial target requirements are 20-pair of C-band wavelength channels each at 25 Gbit/s for up to 20 km distance.

5 Quiet Window Elimination Using Dedicated Activation Wavelength

In this section, we describe in more detail the process of using DAW to eliminate the quiet window for TDM PON activation, which is being studied in the ITU-T G.hsp.ComTC project[13]. Experiment results are also shown as a proof of concept.

A DAW could be a newly defined wavelength not used in current PON systems or a legacy PON wavelength. In the latter case, the activation process needs to coordinate the quiet windows and exchange ranging information between the new PON and the legacy PON. Note that the DAW is only used for the activation purpose, not as a service wavelength as in TWDM or WDM PON system. The management of this wavelength channel is much less stringent and different from a service channel.

Here we give the example of the activation process for the scenario using a newly defined wavelength as DAW for a 50G PON, as shown in Fig. 4. In this scenario, three wavelengths are being used:

- λ_{50Gd}: Downstream (DS) wavelength of 50G PON;

- λ_{50Gu}: Upstream (US) wavelength of 50G PON;
- λ_{DA}: DAW in the US.

The activation process for this scenario is as follows.

1) Upon power on, the ONU works at λ_{50Gd} and λ_{DA}, and listens to the serial number (SN) request at λ_{50Gd}.

2) The OLT opens a quiet window at λ_{DA} and broadcasts the SN request at λ_{50Gd}.

3) The ONU responds with its SN at λ_{DA}.

4) Once the OLT receives the SN response, it opens the quiet window at λ_{DA} and sends a ranging request at λ_{50Gd} directly to the ONU.

5) The ONU responds with the ranging response at λ_{DA}.

6) After receiving the ranging response, the OLT calculates the ranging results at $\lambda_{50Gd}/\lambda_{DA}$ based on the timing difference between the request and the response. The OLT further calculates the ranging results at $\lambda_{50Gd}/\lambda_{50Gu}$ using the ranging results at $\lambda_{50Gd}/\lambda_{DA}$ based on the dispersion difference between λ_{50Gu} and λ_{DA}. The OLT then sends the ranging results at $\lambda_{50Gd}/\lambda_{50Gu}$ to the ONU.

7) The ONU applies the ranging result at $\lambda_{50Gd}/\lambda_{50Gu}$ and starts working on λ_{50Gu}. The ONU tunes from λ_{DA} to λ_{50Gu} in case of tunable ONUs, or switches from λ_{DA} to λ_{50Gu} in case of dual-wavelengths ONUs.

8) The OLT assigns US bandwidth with burst profile of long preamble to the newly activated ONU at λ_{50Gd}.

9) The ONU sends Acknowledge PLOAMu message to the OLT.

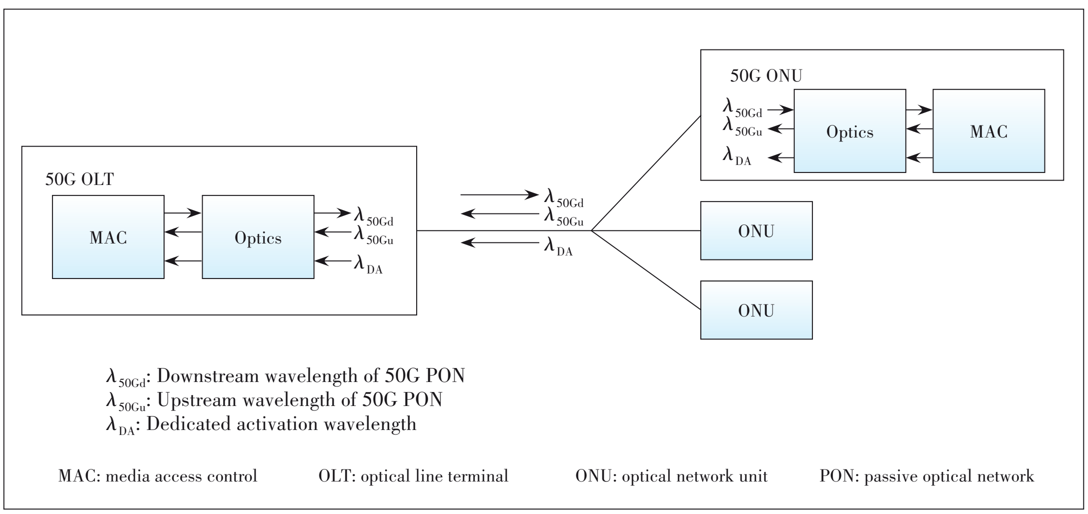
10) The OLT assigns a directed US bandwidth with burst profile of short preamble to the ONU at λ_{50Gd}.

11) The ONU enters the operational state.

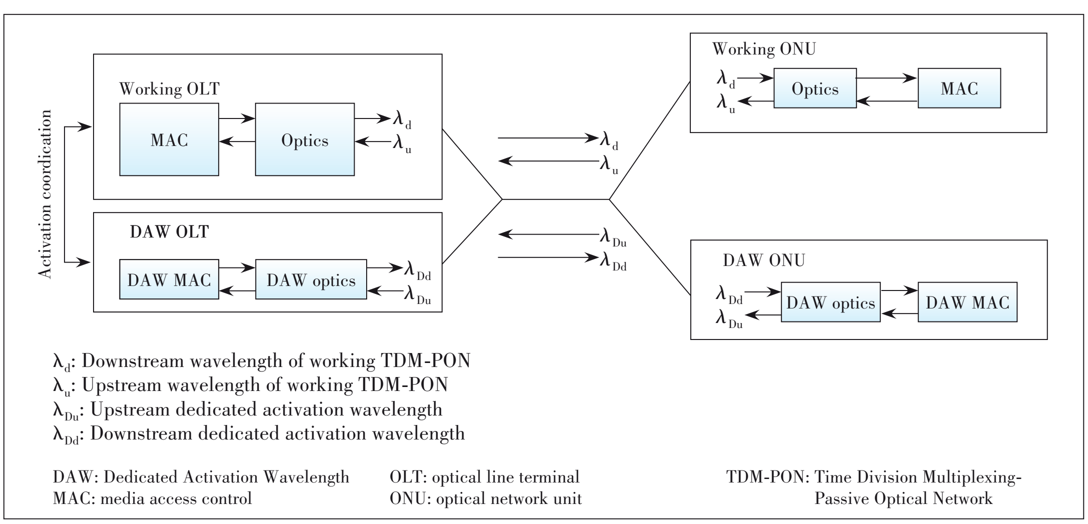
Another parameter to consider is the number of burst allocations per ONU within a PHY frame (125 μs) in a BWmap. The current ITU-T standard specifies a maximum 16 burst allocations per ONU per 125 μs, which corresponds to an ONU buffering delay of 7.8125 μs. When more burst allocations in 125 μs are allowed, lower buffering delay can be achieved, e.g., 4 μs if 31 bursts are allocated[14]. This latency reduction comes at the cost of bandwidth efficiency due to guard time and preamble per burst.

An experiment was set up to test low latency in a TDM-PON using DAW. The experimental configuration is shown in Fig. 5. Two pairs of wavelength channels are used in this system. One pair is working wavelengths and the other is DAW. The latency of quiet window is eliminated from the working wavelengths.

Furthermore, in this experiment, the fixed bandwidth for the ONU is split into multiple (N) mini-slots. The gap between mini-slots as well as the latency due to DBA is reduced when N increases. In this experiment, the traffic flow per direction is 950 Mbit/s, the total US bandwidth is 1 000 Mbit/s, and the fiber distance between the ONU and OLT is less than 100 m. Two cases of packet sizes are measured: a fixed length of 128 bytes and random length between 64 bytes and 1 518 bytes.



▲**Figure 4. Dedicated activation wavelength for 50G PON**



▲**Figure 5. Experimental set-up**

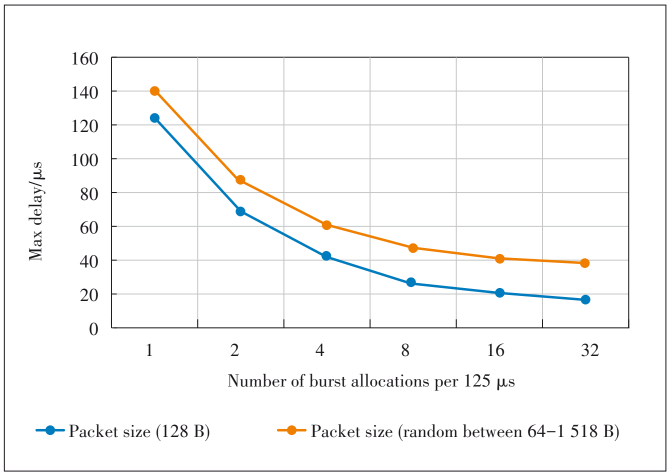
The results of maximum delay versus the number of burst allocations are shown in Fig. 6. The results show that the maximum latency reduces when the number of burst allocations per 125 μs increases in both cases of packet sizes. The reduction of maximum latency becomes more gradual as the number of burst allocations increase. Note that the maximum delay in current TDM-PON system is up to system configurations, e.g., the DBA duration is M×125 μs ($M \geqslant 1$, typically 4) and the quiet window size is up to the differential distance (about 250 μs when the differential distance is 20 km). The typical maximum delay in TDM-PON system is higher than 750 μs. Obviously, it is much higher than that in this experiment and not shown in Fig. 6.

6 Conclusions

In summary, we provided an overview of low latency services and their corresponding requirements. As a potential solution to these requirements, PON technologies to support 5G xhaul transport are presented. Recent progress in PON standardization projects by the ITU-T Q2/SG15 is discussed. A proof-concept experiment and its results are described.

Acknowledgment

We would like to acknowledge Dr. Jun Shan WEY for the discussion and extensive review of the manuscript, and colleagues in the fixed media research and product teams of ZTE Corporation who contribute to the low latency PON project.



▲**Figure 6. Maximum latency under different numbers of burst allocations per 125 μs**

References

[1] ITU-R. IMT vision – framework and overall objectives of the future development of IMT for 2020 and beyond: ITU-R M.2083-0 [S]. 2015
[2] WEY J S. The outlook for PON standardization: a tutorial [J]. Journal of lightwave technology, 2020, 38 (1): 31 – 42, 2020
[3] The Full Service Access Networks Group. About FSAN [EB/OL]. [2021-03-01]. http://www.fsan.org
[4] GlobalData and ZTE Corporation. White paper: precision 5G transport—the foundation of future mobile network [EB/OL]. (2021-02-26)[2021-03-01]. https://www.zte.com.cn/global/solutions/201905201708/201905201738/GlobalData_Precision_5G_Transport
[5] ITU-T. 40-gigabit-capable passive optical network (NG PON2): transmission convergence layer specification: ITU-T G.989.3 recommendation amendment 3 [S]. 2020
[6] ITU-T. 40-gigabit-capable passive optical network (NG PON2): G.989.x series of recommendations [R]. 2015
[7] ITU-T. 5G wireless fronthaul requirements in a PON context, edition 3.0: ITU-T G.Sup66 supplement [S]. 2020
[8] ITU-T. Higher speed bidirectional single-fiber point to point optical access systems: ITU-T G.9806 recommendation amendment 1 [S]. 2020
[9] IEEE. IEEE P802.3cp bidirectional 10 Gbit/s, 25 Gbit/s, and 50 Gbit/s optical access phys task force [EB/OL]. [2021-03-01]. http://www.ieee802.org/3/cp
[10] WEY J S, LUO Y, PFEIFFER T. 5G wireless transport in a PON context: an overview [J]. IEEE communications standards magazine, 2020, 4(1): 50 – 56. DOI: 10.1109/MCOMSTD.001.1900043
[11] eCPRI Transport Network Group. Common public radio interface: requirements for the eCPRI transport network, v1.2 [EB/OL]. (2018-06-25) [2021-03-01]. http://www.cpri.info/downloads/Requirements_for_the_eCPRI_Transport_Network_V1_2_2018_06_25.pdf
[12] ITU-T SG 15 (G.Sup.CODBA). ITU-T rec. series G supplement CO DBA: OLT capabilities for supporting CO DBA [Z]. 2021
[13] ITU-T SG 15 (G.hsp.ComTC). Higher speed passive optical networks: common transmission convergence layer specification living list [Z]. 2021
[14] OU H, TAKAHASHI K, NAKAMURA H. A proposal to change a maximum number of burst allocation series for improving upstream latency in G.989.3, contribution D3 [C]//ITU-T e-meeting. Geneva, Switzerland: ITU, 2016

Biographies

ZHANG Weiliang (zhang.weiliang@zte.com.cn) received his Ph.D. degree in communication and information system from Tsinghua University, China in 2001. He is currently a senior expert of fixed networks at ZTE Corporation and engaged in the research, standardization and product planning of fiber access and home networking. He has led or participated in a number of national "863" projects, as well as provincial and ministerial key projects. He has published more than ten papers and held more than 100 authorized patents.

YUAN Liquan received his master's degree in information engineering from Harbin Institute of Technology, China in 1999. As a project leader at ZTE Corporation, he is responsible for pushing forward the research and standardization in fiber access and home networking of the company, cooperating with the standardization bodies including ITU-T SG15/IEE802.11/IEEE802.3/Broadband Forum/CCSA/ETSI. As an editor, he has participated in drafting more than 10 optical access standards and held more than 40 authorized patents.