QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer for FTTR Networks

CAI Jinhan, ZAN Mingyuan, SHEN Gangxiang

(School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China)

DOI: 10.12142/ZTECOM.202504007

https://kns.cnki.net/kcms/detail/34.1294.tn.20251126.1148.002.html, published online November 26, 2025

Manuscript received: 2025-09-17

Abstract: As Fiber-to-the-Room (FTTR) networks proliferate, multi-device deployments pose significant energy consumption challenges. This paper proposes a Quality of Service (QoS)-aware energy-saving scheme based on a multi-threshold buffer energy saving (MBES) scheme to reduce consumption while ensuring energy QoS. MBES leverages the centralized control of the main fiber unit (MFU) and the wireless-state awareness of subordinate fiber units (SFUs) for synergistic fiber-wireless energy savings. The scheme assigns independent, dynamic buffer thresholds to service queues on SFUs, enabling low-latency reporting for high-priority traffic while accumulating low-priority data to extend sleep cycles. At the MFU, a coordinated scheduling algorithm accounts for Wi-Fi access delay and creates an adaptive closed-loop control by adjusting SFUs' buffer thresholds based on end-to-end delay feedback. Simulation results show that, while satisfying strict latency requirements, MBES achieves a maximum energy saving of 17.75% compared with the no energy saving (NES) scheme and provides a superior trade-off between latency control and energy efficiency compared with the single-threshold buffer energy saving (SBES) scheme.

Keywords: Fiber-to-the-Room; power consumption; energy consumption; energy saving

Citation (Format 1): CAI J H, ZAN M Y, SHEN G X. QoS-aware energy saving based on multi-threshold dynamic buffer for FTTR networks [J]. ZTE Communications, 2025, 23(4): 48 - 64. DOI: 10.12142/ZTECOM.202504007

Citation (Format 2): J. H. Cai, M. Y. Zan, G. X. Shen, "QoS-aware energy saving based on multi-threshold dynamic buffer for FTTR networks," *ZTE Communications*, vol. 23, no. 4, pp. 48 - 64, Dec. 2025. doi: 10.12142/ZTECOM.202504007.

1 Introduction

The rapid development of emerging services such as ultra-high-definition video, immersive extended reality (XR), smart education, and cloud-edge collaboration has presented unprecedented challenges to indoor access networks in terms of bandwidth, latency, and connection stability[1]. To meet these demands, a next-generation indoor fiber access architecture, Fiber-to-the-Room (FTTR), has emerged[2]. Through the centralized control by a main fiber unit (MFU) and the distributed coverage by multiple subordinate fiber units (SFUs), FTTR integrates the functions of a passive optical network (PON) optical network unit (ONU) with Wi-Fi access points (APs), achieving direct fiber connections and high-performance Wi-Fi coverage across the whole home[3]. This architecture effectively overcomes the signal attenuation, insufficient coverage, and severe interference issues of traditional single-point Wi-Fi in multi-room, multi-device scenarios, providing a transmission guarantee with high bandwidth, low latency, and high reliability for emerging services[4]. However, while FTTR enhances network performance, the significant increase in access devices leads to a sharp rise in system energy consumption. In common low-load or no-traffic scenarios, devices continue to power most of their modules, causing significant energy waste. This issue has become a key bottleneck hindering the green and large-scale deployment of FTTR[5].

Existing research on energy savings in access networks primarily focuses on two separate domains: PONs and Wi-Fi. While research specifically targeting the FTTR architecture remains scarce, findings in these related areas offer important references for its energy-saving design. In the context of PONs, successive standards from Gigabit-capable PON (GPON) and 10-Gigabit-capable PON (XG-PON) to 50-Gigabit-capable PON (50G-PON) have introduced sleep mechanisms for ONUs, such as the doze mode, the cyclic sleep mode, and the watchful sleep mode (WSM), among which WSM, in particular, has become a key reference for standard energy-saving solutions in FTTR[6 - 9]. Significant research has focused on optimizing ONU sleep. For instance, YAN et al.[10] proposed uplink-centric and

This work was partially supported by the National Nature Science Foundation of China under Grant No. W2411058.

downlink-centric scheduling to achieve a trade-off between latency and energy savings. SHI et al.[11] designed a differentiated ONU sleep-scheduling scheme based on service-level agreements to meet multi-level Quality of Service (QoS) demands. ZHANG et al.[12] introduced a modular energy-saving method that reduces power consumption by dynamically deactivating internal ONU transceiver modules. Other works have explored aspects such as adaptive sleep-mode optimization[13], the integration of sleep mechanisms with dynamic bandwidth allocation[14], power-consumption modeling for various PON technologies[15–16], and energy-efficiency analysis for multi-level ONU power states with intelligent sleep-duration adjustments[17–18]. These studies provide valuable insights into the trade-off between energy consumption and QoS in PON systems. In the domain of Wi-Fi networks, energy-saving mechanisms primarily concentrate on stations (STAs). The Wi-Fi 6 standard introduced the Target Wake Time (TWT) mechanism, which allows STAs to negotiate with APs to enter the sleep mode during non-essential periods[19]. LORINCZ et al.[20] and PACK et al.[21] proposed STA sleep-control methods based on load and adaptive intervals, effectively improving the balance between energy consumption and performance. In contrast, research on the energy consumption of APs remains relatively limited. However, existing works have explored aspects such as component-level power consumption[22], power characteristics under different configurations[23], and traffic-based power-prediction modeling[24], providing support for the design and management of energy-saving wireless networks.

In converged fiber-wireless networks, studies have shown that a lack of coordination between the independent energy-saving mechanisms of PONs and Wi-Fi can lead to degraded overall performance[25]. To address this, researchers have proposed various joint energy-saving strategies, including threshold-based ONU sleep with wireless traffic aggregation[26], dynamic scheduling combined with wireless rerouting[27], and coordinated optimization of Wi-Fi access and PON sleep modes[28]. These methods reduce energy consumption while, to a certain extent, guaranteeing latency and throughput. Despite advances in independent and partially joint energy-saving optimizations for PONs and Wi-Fi, these approaches face notable limitations in FTTR scenarios. First, many schemes rely on a single load threshold, which makes it difficult to differentiate services with diverse latency requirements (e.g., voice versus video). Second, they do not fully exploit the MFU's centralized control capabilities or its advantage in monitoring the Wi-Fi-side status.

To fill this gap, this paper proposes an energy-saving scheme based on a multi-threshold buffer energy saving (MBES) scheme for FTTR networks, which achieves a balance between energy consumption and service performance through differentiated caching at the SFU side and centralized scheduling at the MFU side. The main contributions of this paper are as follows. First, we propose an FTTR uplink energy-saving framework that reduces system power consumption while meeting diverse QoS requirements. Second, we design an SFU-side multi-threshold caching mechanism that differentiates services by priority to balance latency and energy saving. Third, we develop an MFU-side centralized scheduling algorithm that adaptively adjusts buffer thresholds based on Wi-Fi delay feedback, enabling QoS-aware closed-loop energy control.

The remainder of this paper is organized as follows. Section 2 introduces the FTTR network architecture and the uplink QoS assurance mechanism. Section 3 establishes the power consumption models for the SFU and MFU, along with the system energy consumption model. Section 4 elaborates on the proposed MBES strategy, detailing the SFU's multi-threshold wake-up mechanism and the MFU's cooperative scheduling algorithm. Section 5 evaluates and analyzes the performance of the strategy through simulations. Conclusions are drawn in Section 6.

2 FTTR Architecture and Uplink QoS Mechanisms

This section introduces the FTTR system architecture, focusing on its network structure, uplink data transmission process, and QoS assurance mechanisms. This provides the foundation for the design of the energy-saving strategy presented later in this paper.

2.1 FTTR Network Architecture

Fig. 1 illustrates the basic architecture of an FTTR network. The network consists of an MFU, an indoor fiber distribution network (IFDN), and one or more SFUs. Serving as the gateway and control center of the home network, the MFU connects to the external optical line terminal (OLT) through the PON protocol. Meanwhile, it manages SFUs deployed in different rooms via the IFDN using indoor fiber protocols such as G.fin[9]. At the data flow level, the MFU receives downstream data from the OLT and distributes it to the SFUs. For upstream data from SFUs, the MFU either forwards the traffic to the target SFU or transmits it to the OLT according to the destination address. Compared with traditional optical network terminals (ONTs), the MFU can not only receive broadcast information from the OLT but also provide centralized control of the entire FTTR network, including bandwidth allocation, roaming management, and energy-saving control for SFUs. As the terminal devices of the network, SFUs integrate the functionalities of ONUs and wireless APs. Each SFU is equipped with a G.fin module and Wi-Fi/Ethernet modules, enabling the conversion between optical and wireless frames. Specifically, the SFU encapsulates uplink data from user terminals into G.fin frames and transmits them to the MFU, and decapsulating downstream G.fin data from the MFU and delivering it to STAs via wireless links.

2.2 Uplink Transmission and Associated QoS Assurance in FTTR

In the FTTR architecture, all uplink data transmissions are

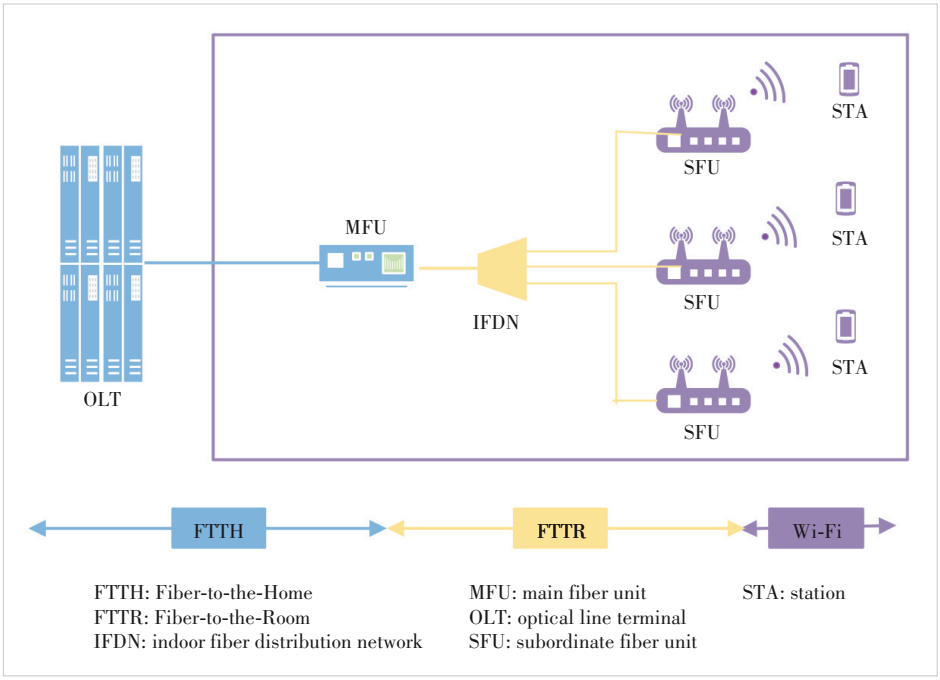

Figure 1. FTTR network architecture

centrally controlled by the MFU. When service data from an STA reaches its associated SFU through a Wi-Fi link, the SFU provides differentiated QoS by classifying packets into transmission container (T-CONT) queues according to service priorities. The four Wi-Fi access categories (ACs)—voice (VO), video (VI), best effort (BE), and background (BK)—are mapped to four T-CONT queues (T-CONT 1 – 4) with different scheduling attributes. Among them, T-CONT 1 is allocated fixed bandwidth to serve the highest-priority services; T-CONT 2 is assigned assured bandwidth, while T-CONT 3 and T-CONT 4 are used to carry low-priority services.

After packets enter the T-CONT queues, their uplink transmission is fully scheduled by the MFU through a dynamic bandwidth allocation (DBA) mechanism, in which the status-report-based DBA (SR-DBA) process is commonly adopted. This scheduling operates in a periodic cycle, as illustrated in Fig. 2. Specifically, each cycle begins with the MFU broadcasting a bandwidth map (BWmap), which specifies the exact timeslots and lengths for uplink transmissions of the T-CONT queues at each SFU. Upon receiving authorization, an SFU transmits data and appends dynamic bandwidth report upstream (DBRu) information to the data frame to indicate the real-time queue status. Based on the collected DBRu information from all SFUs, the MFU executes its DBA algorithm to generate a new BWmap allocation for the next cycle, thereby completing one scheduling loop and initiating the next cycle.

Although the SR-DBA mechanism can effectively guarantee bandwidth utilization and QoS, its inherent energy efficiency issues have also become evident. In common home network scenarios characterized by significant spatiotemporal traffic imbalance, many SFUs may carry only a small amount of bursty traffic or even remain idle for long periods. However, under the current SR-DBA process, an SFU must frequently wake up its power-hungry optical transmitter to send DBRu reports, even when its queues are empty, so as to maintain its presence in the scheduling system. This communication leads to unnecessary power consumption and considerable energy waste, which constitutes the key problem addressed in this paper.

3 FTTR Device Power and System Energy Consumption Models

This section establishes the power consumption models of the SFU and MFU, as well as the overall system energy model, providing a quantitative foundation for the energy-saving strategy discussed in Section 4.

3.1 SFU's Power Consumption Model

The data transmission and reception of an SFU are influenced by the operating states of both its optical and wireless modules. Treating the SFU as a single entity for energy control makes it difficult to balance energy efficiency and communication latency. Therefore, this paper adopts a modular approach to model SFU power consumption. As shown in Fig. 3, the overall power consumption of an SFU is decomposed into three core components: the base module, the optical module, and the Wi-Fi module. The base module encompasses critical subsystems, including the processor, buffer, and power supply, all of which

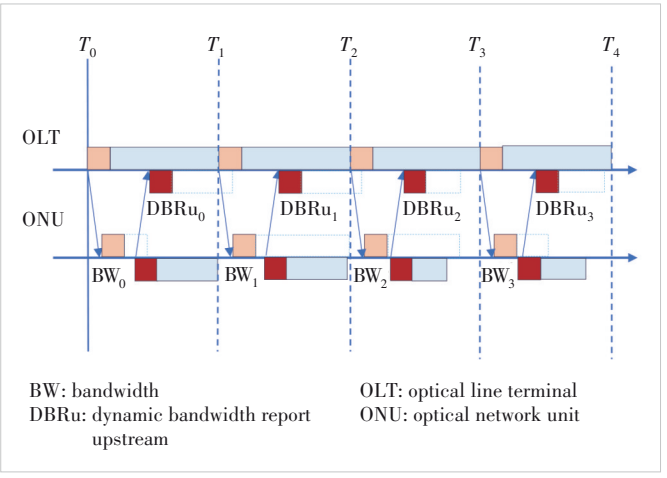

Figure 2. Status-report-based dynamic bandwidth allocation process

are indispensable for sustaining the device's core functionality. The optical module consists of the optical transmitter, receiver, and control circuits, which handles data transmission over the fiber link. The Wi-Fi module comprises the Wi-Fi control unit and radio frequency (RF) units, with the control unit responsible for multiple-input multiple-output (MIMO) operations and transmission power adjustment. Based on the modular division, the total power consumption of an SFU can be expressed as:

$$P_{\mathrm{SFU}} = P_{\mathrm{base}} + P_{\mathrm{PON}} + P_{\mathrm{WiFi}} \quad (1),$$

where P_{base} is the power consumption of the base module, P_{PON} is the power consumption of the optical module dominated by the transmitter and receiver, and P_{WiFi} is that of the Wi-Fi module, primarily determined by the number of RF units as well as their transmission power and maximum data rate.

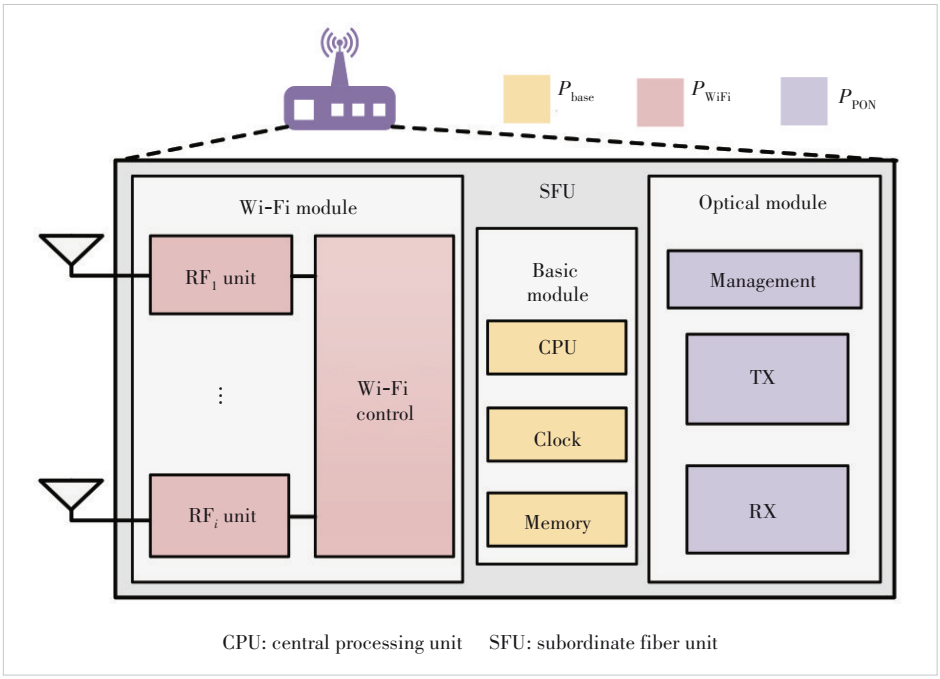

Figure 3. Functional modules and power consumption breakdown of the SFU

The base module of the SFU is configured to remain continuously active in energy-saving design. Keeping this module active allows its internal clock synchronization circuits to respond promptly to wake-up requests, thereby minimizing synchronization latency with the MFU, which is essential for maintaining QoS. In addition, the buffer units inside the base module must operate continuously to process uplink data, preventing packet loss and retransmission that would occur if the module were turned off. More importantly, the power consumption of the base module is much lower than that of the high-power optical and Wi-Fi modules. Thus, considering it as a primary energy-saving target is suboptimal, given the marginal potential gains.

The power consumption of the optical module is mainly determined by its transmitter and receiver, which can be expressed as:

$$P_{\mathrm{PON}} = v_{\mathrm{T}} \cdot P_{\mathrm{Tx}} + v_{\mathrm{R}} \cdot P_{\mathrm{Rx}} \quad (2),$$

where P_{Tx} and P_{Rx} are the power consumptions of the transmitter and receiver, respectively, and v_{T} and v_{R} are binary variables (0 or 1) indicating the on/off states of the transmitter and receiver. In FTTR systems, optical-side energy saving is mainly achieved by turning on/off the transceiver. When no uplink traffic is present for a long period, the SFU turns off the transmitter and periodically activates the receiver to detect MFU downlink data or wake-up instructions, thereby reducing power consumption.

The power consumption of the Wi-Fi module consists of the RF module and dynamic power consumptions, shown as:

$$P_{\mathrm{WiFi}} = P_{\mathrm{RF}} + P_{\mathrm{D}} \quad (3),$$

where P_{RF} is the RF module power consumption and P_D is the dynamic power consumption. Furthermore, P_{RF} is determined by the base power consumption and transmission power, given by:

$$P_{\mathrm{RF}} = P_{\mathrm{RF\_Base}} + P_{\mathrm{RF\_Tx}} \quad (4),$$

where $P_{\mathrm{RF\_Base}}$ is the base power consumption of the RF module and $P_{\mathrm{RF\_Tx}}$ is the transmission power of the RF unit. $P_{\mathrm{RF\_Tx}}$ is an adjustable parameter that determines the coverage of Wi-Fi signals. In practice, the received signal strength indicator (RSSI) at the STA side is commonly used as the evaluation metric. When the RSSI exceeds a predefined threshold, the SFU is deemed capable of maintaining stable communication with the STA at the current transmission power.

Following the RF link power consumption model used in Ref. [5], the transmission power of the i-th RF unit, $P_{\mathrm{RF\_Txi}}$ (dBm), first undergoes antenna path loss σ (dB), then combines with antenna gain G_A (dBi), and subsequently passes through free-space path loss L_p (dB) and obstacle loss L_b (dB) before finally reaching the STA receiver. The received RSSI can be expressed as:

$$\mathrm{RSSI} = P_{\mathrm{RF\_Txi}} - \sigma + G_A - L_p - L_b \quad (5).$$

The free-space path loss L_p is calculated as:

$$L_p = 32.4 + 20\log_{10} D + 20\log_{10} M \quad (6),$$

where D is the distance between the SFU and STA (km), and M

is the frequency (MHz). The obstacle loss L_b is given by:

$$L_b = \sum_{k=1}^{K} \left(j_k \cdot W_k \right) \tag{7},$$

where W_k (dB) denotes the attenuation caused by the k-th type of obstacles, and j_k is the number of such obstacles. Based on the above model, by setting the RSSI in Eq. (5) as the target threshold, the minimum transmit power required to ensure communication quality can be derived as:

$$P_{\text{RF\_Txi}} = \text{RSSI} + \sigma - G_A + \left(32.4 + 20\log_{10} D + 20\log_{10} M\right) + \sum_{k=1}^{K} \left(j_k \cdot W_k \right) \tag{8}.$$

Considering that both the SFU and STA may incorporate multiple RF units, the overall RF transmit power $P_{\text{RF\_Tx}}$ is obtained by summing the transmit power of each unit:

$$P_{\text{RF\_Tx}} = \sum_{i=1}^{N_{SS}} P_{\text{RF\_Txi}} \tag{9},$$

where N_{SS} is the number of RF units. Similarly, the total base power consumption of the RF module, $P_{\text{RF\_Base}}$, is calculated as the sum of the base power of all RF units. In practice, the transmit power is also constrained by the effective isotropic radiated power (EIRP). The limit power P_L is expressed as:

$$P_L \geqslant P_{\text{RF\_Tx}} - \sigma + G_A \tag{10}.$$

Considering the EIRP constraint, the final transmit power $P_{\text{RF\_Txi}}$ is given by

$$P_{\text{RF\_Tx}} = \begin{cases} \sum_{i=1}^{N_{SS}} P_{\text{RFi}}, \sum_{i=1}^{N_{SS}} P_{\text{RF\_Txi}} \leqslant \left(P_L + \sigma - G_A\right) \\ P_L + \sigma - G_A, \sum_{i=1}^{N_{SS}} P_{\text{RF\_Txi}} > \left(P_L + \sigma - G_A\right) \end{cases} \tag{11}.$$

Thus, by substituting the EIRP-constrained transmit power of Eq. (11) into Eq. (5), the actual received signal strength at the m-th STA, RSSI_m, can be obtained.

The dynamic power consumption of the Wi-Fi module in the SFU, P_D, mainly arises from protocol processing at the baseband and media access control (MAC) layers, and is positively correlated with the actual data transmission rate. A higher data rate increases the processing load for modulation/demodulation and coding/decoding on the chip, thereby leading to a higher power consumption. For a single STA, its actual transmission rate R_m is constrained by the maximum negotiated link rate $R_m^{\max}$ with the SFU and varies dynamically with the current RSSI_m. This nonlinear relationship between rates and RSSI can be approximated by a sigmoid function[29].

$$R_m = \frac{R_m^{\max}}{1 + e^{-\frac{\left(120 + \text{RSSI}_m\right) - c_1}{c_2}}} \tag{12},$$

where parameters c_1 and c_2 are set according to experimental data in Ref. [29] to model the growth characteristic of throughput as signal strength varies. Accordingly, the total SFU transmission rate R_{total} at a given time is the sum of the rates of its connected STAs.

$$R_{\text{total}} = \sum_{m=1}^{M} R_m \tag{13}.$$

The dynamic power consumption of the SFU is then calculated as

$$P_D = P_{\max} \cdot \alpha_D \cdot \frac{R_{\text{total}}}{R_{\max}} \tag{14},$$

where $P_{\max}$ is the total power consumption of the device at maximum load, and α_D is the proportion of maximum dynamic power relative to total power. Therefore, $P_{\max} \cdot \alpha_D$ represents the upper bound of dynamic power consumption. $R_{\max} = \sum_{m=1}^{M} R_{\max}^m$ is the sum of the maximum negotiated link rates of all STAs, representing the maximum achievable total rate.

3.2 MFU's Power Consumption Model

Following the modeling approach of the SFU, the total power consumption of the MFU in an FTTR system, P_{MFU}, can also be divided into base power consumption P_{M_Base} and dynamic power consumption P_{M_Dynamic}:

$$P_{\text{MFU}} = P_{M\_\text{Base}} + P_{M\_\text{Dynamic}} \tag{15}.$$

As the control center of the network, the MFU is usually required to remain continuously active to manage all SFUs, and its power characteristics differ from those of the SFU. Its base power consumption is a fixed value, while its dynamic power consumption mainly depends on the number of active SFUs under its management, calculated as:

$$\begin{cases} P_{M\_\text{Base}} = P_{M\_\max} \cdot \alpha_{M\_\text{Base}} \\ P_{M\_\text{Dynamic}} = P_{M\_\max} \cdot \alpha_{M\_\text{Dynamic}} \cdot \frac{N_{\text{active}}}{N} \end{cases} \tag{16},$$

where $P_{M\_\max}$ is the maximum power consumption of the MFU, and α_{M_Base} and $\alpha_{M\_\text{Dynamic}}$ are the proportional coefficients of the base power and the maximum dynamic power relative to the maximum total power consumption, respectively. The model assumes that power consumption increases linearly with the number of active SFUs, N_{active}. When all SFUs are active, the dynamic power consumption reaches its upper limit, $P_{M\_\max} \cdot \alpha_{M\_\text{Dynamic}}$.

3.3 FTTR System Energy Consumption Model

The total energy consumption of the FTTR system, E_{FTTR}, consists of the energy consumption of the MFU and the sum of all SFUs, which is expressed as:

$$E_{\mathrm{FTTR}} = E_{\mathrm{MFU}} + \sum_{i=1}^{N} E_{\mathrm{SFU}}^{i} \tag{17}$$

where N denotes the set of SFUs in the system. As the MFU must remain active as the central network controller, its energy consumption can be expressed as the accumulation of its instantaneous power consumption P_{MFU} over the total operational time T_{tot}:

$$E_{\mathrm{MFU}} = \int_0^{T_{\mathrm{tot}}} P_{\mathrm{MFU}}\,\mathrm{d}t = P_{M\_\mathrm{Base}} \cdot T_{\mathrm{tot}} + \int_0^{T_{\mathrm{tot}}} P_{M\_\mathrm{Dynamic}}\,\mathrm{d}t \tag{18}$$

Unlike the MFU, the SFU is the primary target of energy-saving strategies, as its operating state dynamically switches according to the employed mechanism. Therefore, the energy consumption of an SFU must be calculated by summing its consumption in different states. For a single SFU, based on the on or off state of its optical transceivers, three states are defined: active, doze, and sleep. The energy consumption of the i-th SFU is given by:

$$E_i = E_i^a + E_i^d + E_i^s + E_i^o \tag{19}$$

where E_i^a, E_i^d, E_i^s, and E_i^o represent the energy consumed in the active, doze, sleep, and state-transition modes, respectively. When the SFU is active, all modules remain on; in the doze state, the transmitter is turned off; and in the sleep state, both the transmitter and receiver are turned off. The energy consumption in each state can be calculated as:

$$E_i^a = \left(P_{\mathrm{base}} + P_{\mathrm{PON}} + P_{\mathrm{WiFi}}\right) \cdot T_i^a = \left(P_{\mathrm{base}} + P_{\mathrm{Tx}} + P_{\mathrm{Rx}} + P_{\mathrm{WiFi}}\right) \cdot T_i^a \tag{20}$$

$$E_i^d = \left(P_{\mathrm{base}} + P_{\mathrm{Rx}} + P_{\mathrm{WiFi}}\right) \cdot T_i^d \tag{21}$$

$$E_i^s = \left(P_{\mathrm{base}} + P_{\mathrm{WiFi}}\right) \cdot T_i^s \tag{22}$$

where T_i^a, T_i^d, and T_i^s denote the time durations of the SFU in the active, doze, and sleep states, respectively. In addition, when an SFU switches from sleep to the active state for data transmission, the processes of clock recovery, synchronization with the MFU, and laser activation are required. The transition overhead time is calculated as:

$$T_i^o = \max\left\{T_i^{\mathrm{rec}} + T_i^{\mathrm{sync}}, T_i^{\mathrm{laser}}\right\} \tag{23}$$

where T_i^{rec} is the clock recovery time, T_i^{sync} is the synchronization time, and T_i^{laser} is the laser activation time. For downlink reception, the receiver must be woken up earlier, corresponding to $T_{\mathrm{downoverhead}} = T_i^{\mathrm{rec}} + T_i^{\mathrm{sync}}$. For uplink transmission, the transmitter must be activated in advance, corresponding to $T_{\mathrm{upoverhead}} = T_i^{\mathrm{laser}}$. Assuming the instantaneous power consumption during the wake-up process equals the active-state power of the corresponding modules, the transition energy is

$$E_i^o = \left(P_{\mathrm{Tx}} + P_{\mathrm{Rx}}\right) \cdot T_i^o \tag{24}$$

In summary, by substituting Eqs. (18) – (24) into Eq. (17), the total energy consumption of the FTTR system can be expressed as:

$$E_{\mathrm{FTTR}} = \left(P_{M\_\mathrm{Base}} \cdot T_{\mathrm{tot}} + \int_0^{T_{\mathrm{tot}}} P_{M\_\mathrm{Dynamic}}\,\mathrm{d}t\right) + \left[\begin{array}{l}\left(P_{\mathrm{base}} + P_{\mathrm{Tx}} + P_{\mathrm{Rx}} + P_{\mathrm{WiFi}}\right) \cdot T_i^a + \left(P_{\mathrm{base}} + P_{\mathrm{Rx}} + P_{\mathrm{WiFi}}\right) \cdot T_i^d + \\ \left(P_{\mathrm{base}} + P_{\mathrm{WiFi}}\right) \cdot T_i^s + \left(P_{\mathrm{Tx}} + P_{\mathrm{Rx}}\right) \cdot T_i^o\end{array}\right] \tag{25}$$

4 QoS-Aware Energy Saving Based on Multi-Threshold Dynamic Buffer

Based on the FTTR system and energy consumption models described in the previous section, this section elaborates on the MBES scheme. It first analyzes the limitations of existing energy-saving mechanisms and then presents the key concept and implementation of MBES, including the SFU-side multi-threshold wake-up mechanism and the MFU-side cooperative scheduling algorithm.

4.1 Limitations of Existing Energy-Saving Mechanisms in FTTR

The current G.fin standard for FTTR networks inherits WSM from the XG-PON protocol for energy management. WSM employs an instant wake-up strategy: whenever traffic arrives at the Wi-Fi interface of an SFU, the device immediately transitions from the low-power mode to the full-power mode. However, this approach exhibits two major limitations in addressing the spatially and temporally unbalanced traffic patterns commonly observed in FTTR networks. First, the strategy is coarse-grained and lacks differentiated QoS support. In scenarios with only small amounts of low-priority traffic, frequent instant wake-ups lead to unnecessary energy consumption. A buffer-based delayed wake-up mechanism, where wake-up is triggered only after traffic accumulates beyond a certain threshold, can mitigate this issue. Nevertheless, traditional single-threshold schemes risk degrading high-priority services (e.g., voice and video), which may experience excessive latency while waiting for low-priority packets to fill the buffer, thereby failing to meet stringent QoS requirements. Second, the strategy lacks cross-domain coordination and adaptive capability. Conventional PON energy-saving decisions rely solely on optical link status, overlooking conditions on the Wi-Fi side. One of the key advan-

tages of FTTR architecture is that the SFU, as a converged node for optical and wireless domains, can collect cross-domain metrics such as Wi-Fi access latency and relay this information to the MFU for centralized analysis[3]. However, the existing WSM fails to leverage this architectural benefit and cannot dynamically adjust energy-saving parameters based on end-to-end service performance, making it difficult to achieve an optimal balance between energy efficiency and QoS assurance.

In summary, existing schemes are primarily single-threshold-based, lack cross-domain coordination, and seldom incorporate end-to-end latency feedback for closed-loop control. These limitations motivate the design of MBES, which differentiates queue priorities, leverages the MFU's centralized control capabilities, and dynamically adjusts threshold settings based on latency feedback.

4.2 SFU's Multi-Threshold Wake-Up Mechanism

On the SFU side, a differentiated wake-up mechanism based on multi-threshold buffering is adopted. This mechanism performs fine-grained buffer management and wake-up control according to QoS. Specifically, an independent buffer threshold is configured for each T-CONT queue, which corresponds to the four service categories: VO, VI, BE, and BK. High-priority services (e.g., VO) are assigned lower thresholds so that even a small amount of data can trigger the optical transmitter to wake up and request bandwidth, thereby meeting their low-latency requirements. In contrast, low-priority services (e.g., BK) are assigned higher thresholds to aggregate more packets and reduce unnecessary wake-ups, thus maximizing the energy-saving cycle.

Fig. 4 shows the process of the SFU multi-threshold wake-up mechanism. When the buffered data of any T-CONT queue reaches its preset threshold, the optical transmitter of the SFU is awakened and reports the current buffer status of all queues to the MFU to request bandwidth. After completing the data transmission, the transmitter is turned off again. For the downlink, since the MFU operates with a fixed broadcast cycle, the receiver of the SFU adopts a periodic listening strategy. It only wakes briefly when it needs to receive downlink data or signaling, while remaining off during other times to save energy.

To ensure the effectiveness of this mechanism, its operating model and constraints are constructed from three aspects: buffer-capacity constraints, transceiver switching strategy, and the relationship between buffer threshold and latency.

1) Buffer constraint

Assume there are N SFUs in the system, and each SFU supports K T-CONT queues with different priorities. For the k-th priority queue of the i-th SFU ($i \in \{1,\cdots,N\}, k \in \{1,\cdots,K\}$), let the current buffered data be denoted as $Q_{i,k}$, and the preset wake-up threshold as $Th_{i,k}^{\text{base}}$.

An SFU reports its buffer demand for all queues to the MFU only when at least one queue buffer reaches or exceeds its threshold; otherwise, the SFU remains in an energy-saving mode without reporting. This relationship can be expressed as:

$$R_{i,k} = \begin{cases} Q_{i,k}, \exists Q_{i,k} \geqslant Th_{i,k}^{\text{base}}, k = 1,\cdots,K \\ 0, \text{otherwise} \end{cases} \tag{26}$$

where $R_{i,k}$ represents the buffer information of the k-th queue of the i-th SFU reported to the MFU. In addition, the configuration of all thresholds must comply with the physical buffer-capacity constraint of each SFU. Let the maximum buffer capacity of the i-th SFU be B_i. The sum of thresholds of all priority queues must not exceed this limit:

$$\sum_{k=1}^{K} Th_{i,k}^{\text{base}} \leqslant B_i \tag{27}$$

2) Transmitter shut-down strategy

The switching strategy of the transmitter follows certain strict timing constraints. We define the start and end time of the j-th bandwidth allocation cycle as $[T_{j\,\text{start}}, T_{j\,\text{end}}]$. Within this cycle, if the i-th SFU is assigned an uplink timeslot starting at $T_{i,j}^{\text{start}}$ with

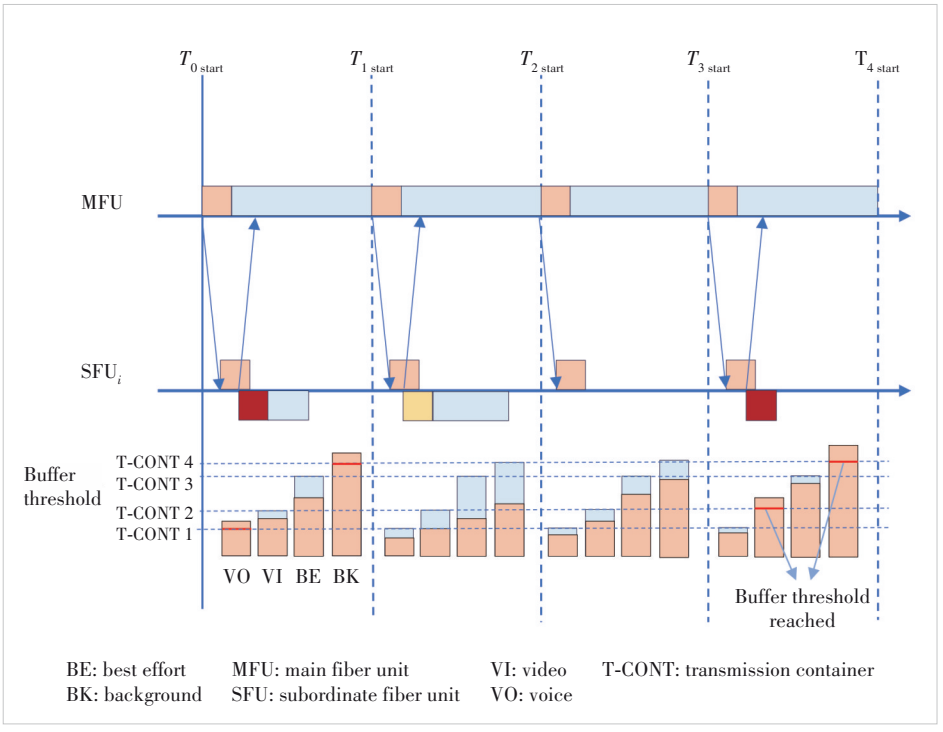

Figure 4. Process of the SFU's multi-threshold wake-up mechanism

length $\mathrm{Len}_{i,j}$, this timeslot must be fully contained within the cycle, satisfying the following boundary conditions:

$$T_{j\,\mathrm{start}} \leqslant T_{i,j}^{\mathrm{start}} \tag{28}$$

$$T_{i,j}^{\mathrm{start}} + \mathrm{Len}_{ij} \leqslant T_{j\,\mathrm{end}} \tag{29}$$

To transmit data on time, the SFU must wake up its transmitter in advance. If the transmitter requires a start-up overhead $T_{\mathrm{upoverhead}}$, the wake-up time is determined as:

$$T_{i,j}^{\mathrm{wake}} = T_{i,j}^{\mathrm{start}} - T_{\mathrm{upoverhead}} \tag{30}$$

After completing one uplink transmission, the SFU can shut down its transmitter to save energy. However, this operation is beneficial only if the idle interval between two consecutive transmissions is sufficiently long. Specifically, the idle time from the end of transmission $j-1$ to the beginning of transmission j must be greater than or equal to the transmitter start-up overhead $T_{\mathrm{upoverhead}}$.

$$T_{i,j-1}^{\mathrm{start}} + \mathrm{Len}_{i,j-1} + T_{\mathrm{upoverhead}} \leqslant T_{i,j}^{\mathrm{wake}} \tag{31}$$

In addition, to ensure fairness in energy saving among multiple SFUs, the MFU should maintain a fixed relative order in the allocation of uplink timeslots.

3) Receiver shut-down strategy

Let $T_{i,j}^{0}$ denote the moment when the i-th SFU parses the downlink frame from the MFU and obtains its allocated timeslot in the j-th bandwidth allocation cycle. It is calculated as:

$$T_{i,j}^{0} = T_{j\,\mathrm{start}} + T_{\mathrm{DBA}} + T_{i}^{\mathrm{tran}} + T_{G}^{\mathrm{proc}} \tag{32}$$

where $T_{j\,\mathrm{start}}$ is the start time of the j-th cycle, T_{DBA} is the computation overhead of DBA, T_{i}^{tran} is the propagation delay from the MFU to the i-th SFU, and T_{G}^{proc} is the processing time for the SFU to parse the downlink frame. If the duration of receiving the downlink data frame is denoted as T_{down}, the completion time of the i-th SFU receiving all downlink traffic in the j-th cycle is:

$$T_{i,j}^{1} = T_{i,j}^{0} + T_{\mathrm{down}} \tag{33}$$

After completing downlink reception, the SFU decides whether to shut down its receiver based on energy-saving benefits. The prerequisite for executing the shutdown operation is that the idle interval from the completion of this reception (T_{ij}^{1}) to the start of the next reception ($T_{i,j+1}^{0}$) must be greater than or equal to the receiver start-up overhead $T_{\mathrm{downoverhead}}$:

$$T_{i,j+1}^{0} - T_{i,j}^{1} \geqslant T_{\mathrm{downoverhead}} \tag{34}$$

4) Relationship between SFU buffer thresholds and latency

In the FTTR energy-saving scenario, the end-to-end latency of the k-th priority packet at the i-th SFU, denoted as $T_{i,k}$, can be expressed as:

$$T_{i,k} = T_{i,k}^{\mathrm{WiFi}} + T_{i,k}^{\mathrm{PON}} + T_{\mathrm{trans}} = T_{i,k}^{\mathrm{WiFi}} + (T_{\mathrm{report}} + T_{\mathrm{upoverhead}} + T_{i}^{\mathrm{sleep}}) + T_{\mathrm{trans}} \tag{35}$$

where $T_{i,k}^{\mathrm{WiFi}}$ is the latency on the Wi-Fi side caused by contention or retransmission, and T_{trans} is the optical transmission delay. Since the distance between the SFU and MFU in FTTR systems is very short, T_{trans} can be ignored. The optical-side latency $T_{i,k}^{\mathrm{PON}}$ is further decomposed into three parts: the reporting overhead T_{report}, the actual sleep waiting time in the queue T_{i}^{sleep}, and the transmitter start-up overhead $T_{\mathrm{upoverhead}}$.

To satisfy QoS requirements, $T_{i,k}$ must not exceed the latency bound D_k, i.e., $D_k \geqslant T_{i,k}$. When a packet arrives at an SFU, its $T_{i,k}^{\mathrm{WiFi}}$ is already fixed. Therefore, the maximum tolerable optical-side latency can be derived and translated into the maximum allowed number of waiting cycles, denoted as $n_{i,k}$. This value represents the maximum number of scheduling cycles that a packet can wait in the SFU buffer without violating its latency bound:

$$n_{i,k} = \left\lfloor \frac{D_k - T_{i,k}^{\mathrm{WiFi}} - T_{\mathrm{report}} - T_{\mathrm{upoverhead}}}{T_{\mathrm{cycle}}} \right\rfloor \tag{36}$$

The SFU does not need to track the countdown of each packet individually. Instead, it maintains a global minimum waiting cycle, denoted as n_i, which is the smallest value among all $n_{i,k}$. This represents the "most urgent" packet in the buffer. When a new packet (the m-th arrival) enters, the SFU calculates its $n_{i,k,m}$ and updates the global counter.

$$n_i = \min\{n_{i,k,1}, n_{i,k,2}, \cdots, n_{i,k,m}\} \tag{37}$$

The counter n_i acts as a mandatory countdown timer and decreases by one at the end of each scheduling cycle. When n_i reaches zero, the SFU must wake up and request bandwidth to meet latency targets, regardless of buffer thresholds. Thus, the unified bandwidth-reporting condition that combines latency-triggered and buffer-threshold-triggered events is

$$R_{i,k} = \begin{cases} Q_{i,k}, \exists Q_{i,k} \geqslant Th_{i,k}, k = 1, \cdots, K \\ Q_{i,k}, n_i = 0 \\ 0, \text{otherwise} \end{cases} \tag{38}$$

Under the QoS constraint, buffer thresholds are further adjusted dynamically to optimize energy saving, with $Th_{i,k}^{\mathrm{base}}$ serving only as the initial value. The threshold directly affects the balance between energy saving and latency: a high threshold yields better energy saving but larger latency, while a low threshold does the opposite. For the i-th SFU in the j-th bandwidth allocation cycle, if the bandwidth allocated to priority k is $G_{i,j,k}$, the newly arrived data is $Q_{i,j,k}$, and the bandwidth demand

reported in the previous cycle is $R_{i,j-1,k}$. The total amount of data at the beginning of the current cycle is:

$$Q_{i,j,k}^{\text{total}} = R_{i,j-1,k} - G_{i,j,k} + Q_{i,j,k} \tag{39}$$

When $Q_{i,j,k}^{\text{total}} \geqslant Th_{i,k}^{\text{base}}$, a bandwidth report is triggered. It is evident that the time to reach the threshold depends on the packet arrival rate λ_k. To adapt the thresholds to traffic variations, the arrival rate is estimated by averaging over the previous r cycles:

$$\lambda_k = \frac{1}{r}\sum_{j=n-r+1}^{n}\frac{Q_{i,j,k}}{t_{i,j}} \tag{40}$$

where $t_{i,j}$ is the effective observation duration in cycle j. Accordingly, the adaptive threshold update mechanism driven by the arrival rate is defined as:

$$Th_{i,k} = \alpha_k \cdot \lambda_k \cdot (n_{i,k} \cdot T_{\text{cycle}}) \tag{41}$$

where α_k is the weight factor of the priority-k service, and $(n_{i,k} \cdot T_{\text{cycle}})$ is its maximum waiting time. With this design, the thresholds decrease adaptively for high-priority services or traffic bursts to suppress queueing latency, while the thresholds increase for low-priority or low-traffic services to maximize energy savings.

4.3 MFU's Cooperative Scheduling and Dynamic Threshold Adjustment

The MFU allocates bandwidth to multi-priority services based on buffer reports and Wi-Fi-side latency information from SFUs. During this process, it dynamically adjusts buffer thresholds to meet QoS requirements. Since wake-up is performed locally by SFUs, the MFU does not require additional state decisions, which reduces signaling overhead and allows it to focus on global scheduling. Specifically, the MFU first collects the bandwidth requests $R_{i,k}$ from all active SFUs. It then calculates the end-to-end latency $D_{i,k} = T_{i,k}^{\text{WiFi}} + T_i^{\text{sleep}}$, which combines the Wi-Fi-side latency $T_{i,k}^{\text{WiFi}}$ and the sleep-induced latency T_i^{sleep}. The MFU sorts the set $\{D_{i,k}\}$ in descending order to obtain the index mapping $\sigma(i)$, and reorders the bandwidth request set $\{R_{i,k}\}$ accordingly into $\{R_{\sigma(i),k}\}$. This ensures that SFUs with longer latency are prioritized in bandwidth allocation. For SFUs whose buffer thresholds are not reached, the MFU retains their scheduling order but does not assign bandwidth, thereby maintaining overall fairness. Bandwidth is then allocated from T-CONT 1 to T-CONT 4 in descending order of priority. Let the uplink frame length be L_f (μs), the total available uplink capacity be C (bit/s), the maximum proportions for different priorities be γ_1, γ_2, γ_3, and γ_4, and the number of active SFUs be n.

For bandwidth-guaranteed services T-CONT 1 and T-CONT 2, each active SFU first receives a basic bandwidth $\text{BW}_{i,k}$, proportionally divided according to the service proportion and the number of active SFUs.

$$\text{BW}_{i,1} = \frac{L_f \cdot C \cdot \gamma_1}{n} \tag{42}$$

$$\text{BW}_{i,2} = \frac{L_f \cdot C \cdot \gamma_2}{n} \tag{43}$$

For the highest-priority T-CONT 1 services, a "fixed allocation and excess grant" mechanism is adopted. If its reported demand $R_{i,1}$ exceeds the basic bandwidth, the MFU grants additional bandwidth to ensure the timely transmission of latency-sensitive traffic. Thus, the actual bandwidth $G_{i,1}$ allocated to each active SFU for T-CONT 1 is:

$$G_{i,1} = \min\left\{\max\{BW_{i,1}, R_{i,1}\}, \frac{L_f \cdot C - \sum_{i=1}^{n} BW_{i,2}}{n}\right\} \tag{44}$$

where $\dfrac{L_f \cdot C - \sum_{i=1}^{n} BW_{i,2}}{n}$ represents the upper allocation bound after deducting the basic bandwidth of T-CONT 2, ensuring that all active SFUs of T-CONT 1 obtain transmission opportunities. T-CONT 2 follows a similar allocation logic, but its actual granted bandwidth is limited by the allocation result of T-CONT 1:

$$G_{i,2} = \min\left\{\max\{\text{BW}_{i,2}, R_{i,2}\}, \frac{L_f \cdot C - \sum_{i=1}^{n} G_{i,1}}{n}\right\} \tag{45}$$

After allocating bandwidth to T-CONT 1 and T-CONT 2, the MFU distributes the remaining bandwidth to T-CONT 3 and T-CONT 4 in sequence. The granted bandwidth is subject to both the services' bandwidth cap and the residual link capacity. For T-CONT 3, the allocation is

$$\begin{cases} \text{BW}_{i,3} = \dfrac{L_f \cdot C \cdot \gamma_3}{n} \\ G_{i,3} = \min\left\{\text{BW}_{i,3}, R_{i,3}, L_f \cdot C - \sum_{i=1}^{n}\left(G_{i,1} + G_{i,2}\right) - \sum_{1}^{i-1} G_{i,3}\right\} \end{cases} \tag{46}$$

Similarly, for T-CONT 4, the allocation is

$$\begin{cases} \text{BW}_{i,4} = \dfrac{L_f \cdot C \cdot \gamma_4}{n} \\ G_{i,4} = \min\left\{\text{BW}_{i,4}, R_{i,4}, L_f \cdot C - \sum_{i=1}^{n}\left(G_{i,1} + G_{i,2} + G_{i,3}\right) - \sum_{1}^{i-1} G_{i,4}\right\} \end{cases} \tag{47}$$

Through this mechanism, the highest-priority services receive rigid guarantees, the second-highest priority services ob-

tain conditional guarantees, and the lowest-priority services are scheduled only when surplus bandwidth is available.

After completing the bandwidth allocation, the MFU proceeds to dynamically adjust the buffer thresholds for each SFU. This adjustment process adheres to two fundamental principles. First, the threshold should be corrected based on the deviation between the services' measured latency T_k^{measured} and its target latency deadline D_k. Second, after adjustment, the threshold of a high-priority queue must remain no greater than that of a low-priority queue to maintain the priority order. If the measured latency of a certain priority class exceeds its requirement, that is

$$\Delta_k = T_k^{\text{measured}} - D_k > 0 \tag{48},$$

and then in the next scheduling cycle, its buffer threshold is decreased:

$$Th_{i,k}^{\text{new}} = Th_{i,k}^{\text{old}} - \eta_k \cdot \Delta_k \tag{49},$$

where η_k is the step size for decreasing the threshold of priority k. To prevent the violation of priority order, a correction is applied:

$$Th_{i,k}^{\text{new}} = \min\left\{Th_{i,k}^{\text{new}}, Th_{i,k+1}^{\text{new}}\right\} \tag{50}.$$

Conversely, when $\Delta_k < 0$ and this condition holds for n_Δ consecutive cycles, it indicates that the threshold is overly conservative. In this case, the threshold can be moderately increased:

$$Th_{i,k}^{\text{new}} = \min\left\{Th_k^{\max}, Th_{i,k}^{\text{old}} - \beta_k \cdot \Delta_k\right\} \tag{51},$$

where $Th_k^{\max}$ is the maximum allowable buffer threshold for priority k, and β_k is the step size for increasing the threshold.

In summary, the MFU's energy-saving bandwidth allocation follows three principles. First, it allocates bandwidth among active SFUs in order of service priority, meeting latency requirements of different services. Second, within the same priority class, it prioritizes the scheduling of SFUs with higher Wi-Fi side latency to improve end-to-end latency performance. Third, it dynamically adjusts buffer thresholds based on service latency feedback, achieving a balance between QoS assurance and energy saving.

5 Simulation Results and Analyses

This section evaluates the performance of the proposed MBES using a simulation environment built on the Java platform. The MBES is compared with two other baseline strategies: the no energy saving (NES) scheme and the single-threshold buffer energy saving (SBES) scheme. The simulation scenario is configured as a home network, with average system latency and energy consumption as the primary performance metrics. This evaluation aims to assess the effectiveness of energy saving by the different strategies under service performance constraints, thereby validating the efficiency of MBES.

5.1 Parameter Setting

Given the small scale of FTTR deployments in residential settings, the number of SFUs is limited and each SFU typically serves only a few STAs. In our simulations, each SFU is therefore assumed to serve two STAs, while the total number of SFUs varies from 2 to 5. Each STA generates a constant aggregate traffic load of 120 Mbit/s, apportioned equally across four priority services—VO, VI, BE, and BK—at 30 Mbit/s per service. Traffic arrivals follow a Poisson process with an average flow duration of 1 s, and packet sizes are fixed at 1 500 bytes. On the wireless side, we model IEEE 802.11ax and use the enhanced distributed channel access (EDCA) mechanism. Detailed PHY/MAC parameter values are listed in Table 1.

For the optical link, key parameters are configured according to the G.fin protocol standard[9]. The downlink and uplink capacities are set to 10 Gbit/s and 2.5 Gbit/s, respectively, and the physical fiber length between the MFU and SFUs is 10 m. The bandwidth allocation period is 125 μs. In the MFU scheduling strategy, bandwidth assignment considers both real-time SFU requests and predefined service-level constraints. To reflect differentiated service priorities, the maximum proportions of the total uplink bandwidth that T-CONT 1 – 4 queues can occupy are set to 0.2, 0.5, 0.3, and 0.1, respectively, as shown in Table 2. Importantly, their sum is not required to equal 1, allowing the MFU to flexibly perform statistical multiplexing based on real-time traffic demand. Two baseline schemes are implemented for comparison. In NES, all SFUs remain continuously active, and the MFU applies the classic GIANT algorithm[30], simplified by setting the service interval of all services to 1 to guarantee allocation in every cycle. In SBES, each SFU employs a single global buffer threshold to trigger its optical transmitter; once active, it follows the same allocation procedure as in NES. The MFU-side bandwidth scheduling parameters used

Table 1. Simulation scenario, traffic, and EDCA parameter settings

Scenario Settings			
Number of SFUs	2/3/4/5	Number of Stations per SFU	2
Traffic Settings			
VO data rate	30 Mbit/s	VI data rate	30 Mbit/s
BE data rate	30 Mbit/s	BK data rate	30 Mbit/s
Packet size	1 500 B	Traffic duration	1 s
EDCA Settings			
Access category	CWmin	CWmax	AIFSN
VO	3	7	2
VI	7	15	3
BE	15	1 023	6
BK	15	1 023	9

AIFSN: arbitration inter-frame space number
BE: best effort
BK: background
CWmax: maximum contention window
CWmin: minimum contention window
EDCA: enhanced distributed channel access
SFU: subordinate fiber unit
VI: video
VO: voice

Table 2. MFU bandwidth scheduling parameter settings

Scenario Settings			
MFU-SFU distance	10 m	Total uplink bandwidth	2.5 Gbit/s
SFU uplink rate	2.5 Gbit/s	Total downlink bandwidth	10 Gbit/s
Proposed Scheme (MBES)			
DBA Type	MBES	Allocation period and frame length	125 μs
T1 total bandwidth ratio	0.2	T2 total bandwidth ratio	0.5
T3 total bandwidth ratio	0.3	T4 total bandwidth ratio	0.1
Baseline Scheme (NES/SBES)			
DBA Type	GIANT	Allocation period and frame length	125 μs
T1 Total Bandwidth Ratio	0.2	T2 total bandwidth ratio	0.5
T3 Total Bandwidth Ratio	0.3	T4 total bandwidth ratio	0.1
SI_{max}	1	SI_{min}	1

DBA: dynamic bandwidth allocation
MFU: main fiber unit
MBES: multi-threshold buffer energy saving
NES: no energy saving
SBES: single-threshold buffer energy saving
SFU: subordinate fiber unit

Table 3. Parameter settings for the power consumption model

Parameter	Value	Parameter	Value	Parameter	Value
P_{base}	1	P_{Tx}	1.2	P_{Rx}	0.8
P_{RF_base}	0.6	P_{max}	6	α_D	0.2
D	5 m	σ	1 dB	G_A	3 dBi
M	5 GHz	W_k(5G)	15 dB	j_k	0
P_L(5G)	23 dBm	R^m_{max}(5G)	2 402 Mbit/s	RSSI	−40 dBm
P_{M_max}	10	α_{M_Base}	0.4	$\alpha_{M\_Dynamic}$	0.6
$T_{upoverhead}$	125 μs	$T_{downoverhead}$	125 μs		

in the simulations are summarized in Table 2.

For the energy consumption analysis, the power parameters of the SFU and MFU functional modules are listed in Table 3. These values are normalized based on data from actual products[31 – 32]. The SFU's base power consumption is set to 1 unit, while its optical transmitter and receiver consume 1.2 and 0.8 units, respectively. The wake-up overhead for the optical transceiver is 125 μs[33]. The SFU's maximum power consumption is capped at 6 units, with dynamic power consumption accounting for up to 20% (1.2 units). The MFU's maximum power consumption is set to 10 units, with the base and maximum dynamic power consumptions accounting for 40% and 60%, respectively; it is assumed to operate at full power throughout the simulation. On the Wi-Fi side, each RF module has a base power consumption of 0.6 units. The dynamic transmission power of the Wi-Fi module is calculated according to link parameters. According to the Wi-Fi 6 standard, the simulation models Wi-Fi transmission in the 5 GHz band with a 160 MHz channel bandwidth, using modulation and coding scheme (MCS) index 11. The SFU is assumed to support 2×2 MIMO with a maximum transmission rate of 2 402 Mbit/s, and the STA parameters are configured to match. To determine the RF module's transmission power, the initial target RSSI at the STA is set to −40 dBm[34]. The actual transmission power under the EIRP constraint and the resulting received RSSI are calculated using the channel parameters in Table 3 and Eqs. (5) – (11), which determine the final dynamic power consumption of the Wi-Fi module.

Under the MBES strategy, the buffer thresholds and latency requirements of different priority services directly influence the on/off timing of SFU transmitters, thereby affecting both energy efficiency and service latency. Table 4 summarizes the latency requirements, buffer thresholds, adjustment step sizes, and parameter settings used for the SBES comparison scheme. For high-priority services, the initial buffer thresholds are determined based on traffic arrival rates and latency requirements and are dynamically adjusted during operation according to the parameters in Table 4. For comparison, the buffer threshold in the SBES scheme is set to the sum of all priority-specific thresholds used in the multi-threshold strategy.

5.2 Result Analysis

This subsection analyzes the latency and energy consumption of the proposed MBES scheme. In the result figures, the legends NES, MBES, and SBES represent the respective schemes, while VO, VI, BE, and BK denote voice, video, best-effort, and background services, respectively.

Fig. 5 illustrates how the average system latency of the FTTR network varies with the number of SFUs. Under the NES scheme, in which SFUs actively report in every cycle and the MFU bandwidth is sufficient, latency for all service priorities remains stable, with BK service latency staying below 5 ms. In contrast, the energy-saving schemes introduce additional latency because SFUs enter a sleep state when their buffer thresholds are not reached, delaying reporting and increasing optical-side latency. Among these, the MBES scheme demonstrates superior performance over SBES: by assigning lower thresholds to high-priority services, MBES triggers earlier reporting, enabling

Table 4. Buffer threshold parameter settings

Parameter	VO	VI	BE	BK
Latency requirement/ms	2	10	30	50
Initial buffer threshold/kB	3	30	60	100
Step size (decrease/increase)/(kB/ms)	2/1	2/1	2/1	2/1
Consecutive cycle threshold (cycle)	10	5	5	5
Maximum buffer threshold/kB	100	200	500	1 000
Single threshold buffer/kB	196			
Physical buffer constraint /MB	32			

BE: best effort BK: background VI: video VO: voice

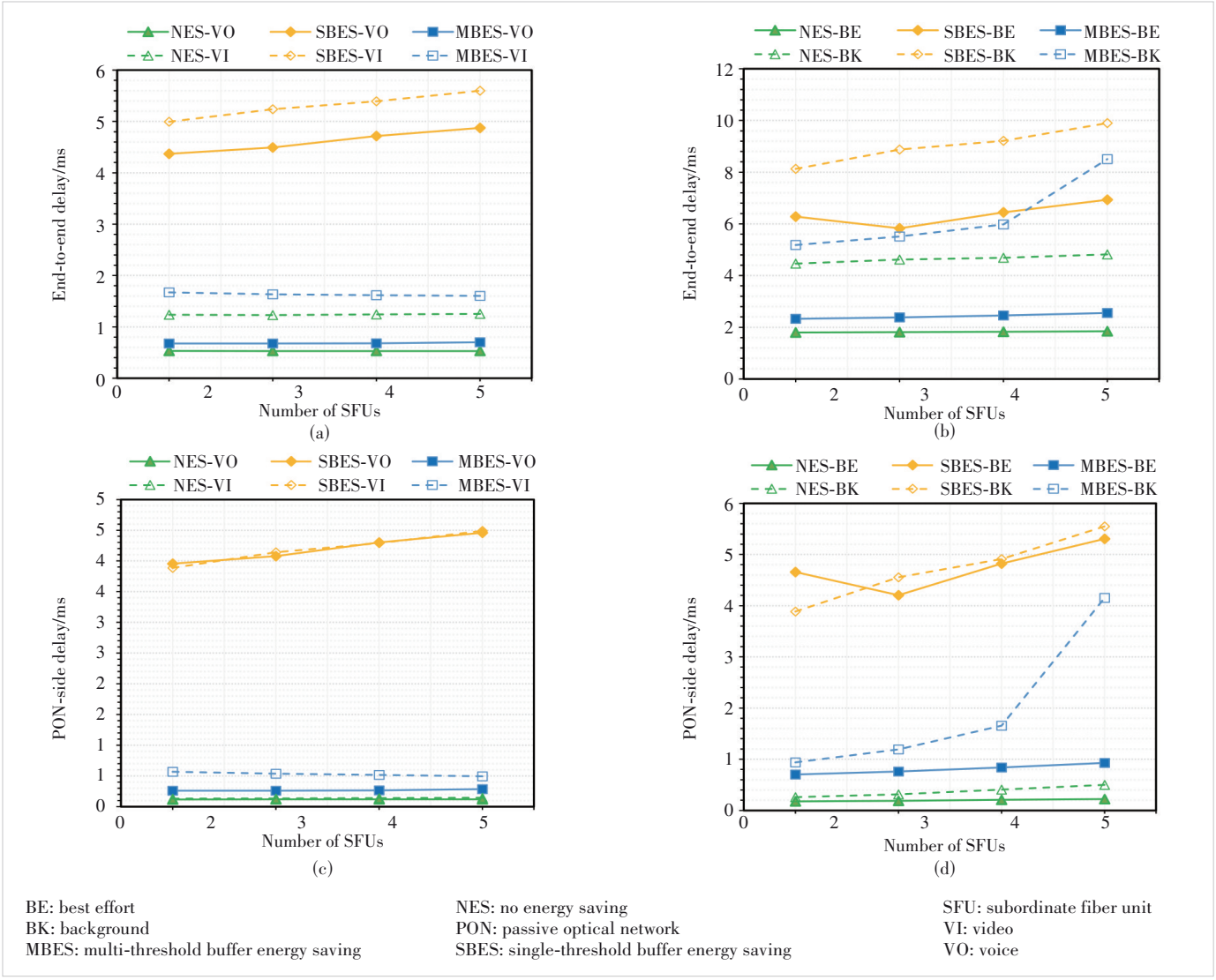

Figure 5. Average system latency under the NES, SBES, and MBES schemes (STA=2): (a) end-to-end latency of high-priority services, (b) end-to-end latency of low-priority services, (c) optical-side latency of high-priority services, and (d) optical-side latency of low-priority services

the concurrent transmission of lower-priority traffic and effectively reducing overall latency. Conversely, SBES uses a single unified buffer threshold for all services, causing SFUs to report less frequently, accumulate larger data bursts, and ultimately experience higher latency. The results in Figs. 5c and 5d show that under the energy-saving schemes, the optical-side latency of BK services increases most significantly, and this effect becomes more pronounced as the number of SFUs increases. The impact is especially severe in the SBES scheme, where BK traffic must wait for high-priority transmissions and can remain buffered for extended periods when bandwidth is insufficient. The MBES scheme alleviates this problem because frequent reporting triggered by high-priority services increases the scheduling opportunities for BK traffic. Additionally, under the SBES scheme, the optical-side latency of VO services is comparable to, and in some cases slightly higher than, that of VI services. This is because, with a unified threshold, VO traffic often enters the MFU buffer and is reported and transmitted together with VI traffic, reducing the latency gap between the two. As the number of SFUs increases, system bandwidth allocation becomes more constrained, amplifying the reporting delays introduced by energy-saving schemes. The MBES strategy achieves a balance between energy savings and latency by applying differentiated buffer thresholds, whereas the SBES strategy, with its longer reporting cycles, not only increases queuing delays for low-priority services but also weakens the scheduling advantage of high-priority services. These findings highlight that setting appropriate buffer thresholds and dynamically adjusting them based on service priorities is essential for maintaining service performance while optimizing energy efficiency in FTTR systems.

Fig. 6 compares the total FTTR system energy consumption and the average SFU energy consumption of the NES, MBES,

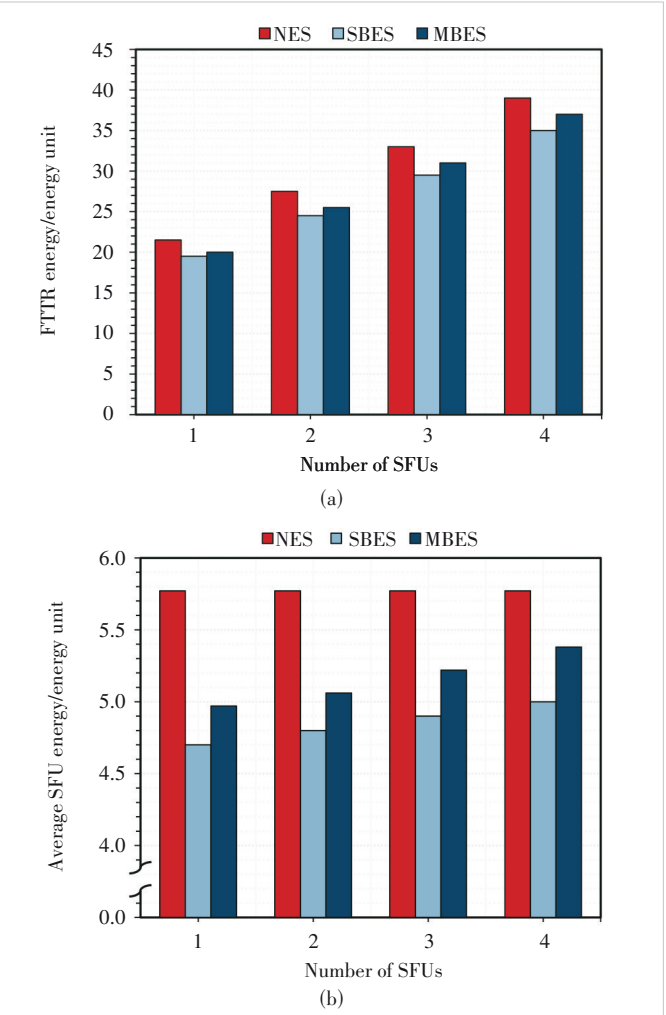

Figure 6. System energy consumption under the NES, SBES, and MBES schemes (STA=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

and SBES schemes as the number of SFUs increases from 1 to 5. System energy consumption is calculated according to Eq. (25) and expressed in power units. As shown in Fig. 6a, the NES scheme exhibits the highest total energy consumption. The SBES scheme achieves the lowest consumption, reducing energy use by approximately 9.50% to 10.43% compared with NES. The MBES scheme consumes slightly more than SBES but still provides savings of about 4.58% to 7.28% compared with NES. Fig. 6b further shows that the average SFU energy consumption follows a similar trend. Consumption is the highest under the NES scheme and the lowest under the SBES scheme, achieving savings of 13.02% to 17.76% compared to NES. Although slightly higher than SBES, the MBES scheme also provides substantial energy savings of 6.17% to 13.62% compared with NES. Furthermore, with the uplink rate fixed at 2.5 Gbit/s, the energy-saving effect tends to diminish when the number of SFUs increases to five, rather than improving continuously. This occurs because adding more SFUs reduces the available bandwidth per SFU. Once an SFU reaches its buffer threshold and triggers a report, it must activate its optical transmitter more frequently and use additional transmission cycles to send the buffered data, thereby lowering the overall energy-saving efficiency.

In the FTTR system, the SFU optical transmitter can operate at a higher uplink rate of 10 Gbit/s in addition to 2.5 Gbit/s. Therefore, the simulation also evaluates the performance of the proposed energy-saving strategy at this higher rate. For the 10 Gbit/s system, the SFU optical transmitter power consump-

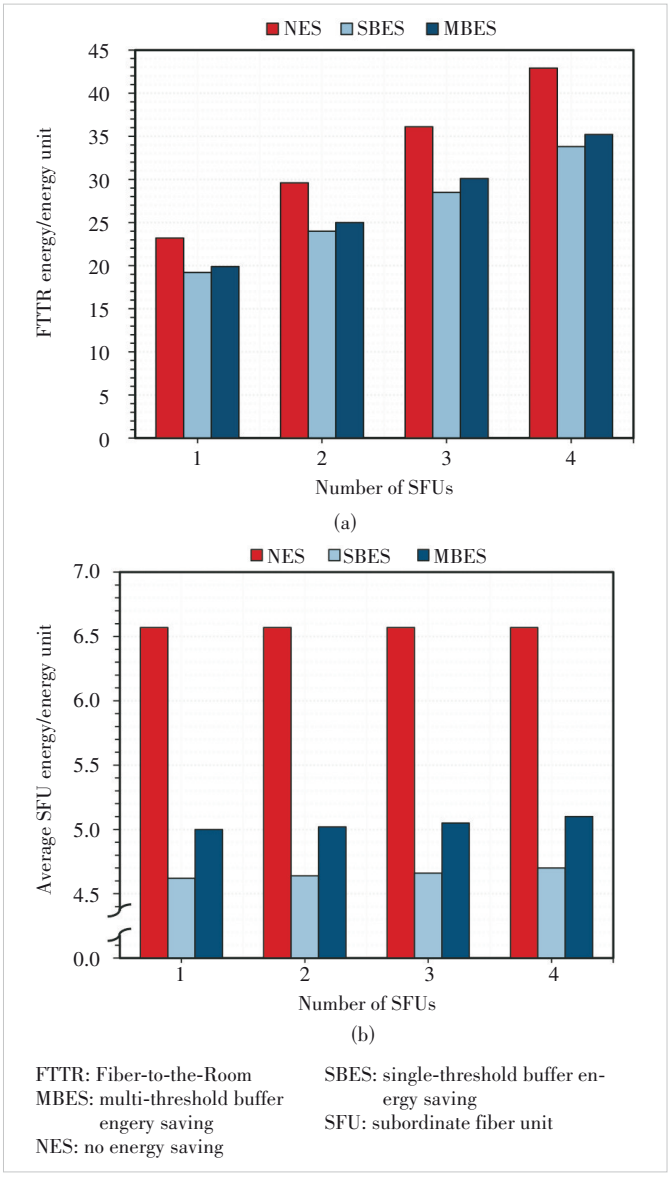

Figure 7. System energy consumption with a 10 Gbit/s uplink rate under the NES, SBES, and MBES schemes (STA=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

tion is assumed to be 2 units, with a maximum of 6.8 units. All other power parameters remain consistent with those listed in Table 3. Fig. 7 presents the simulation results at 10 Gbit/s. With the substantial increase in transmitter power, the energy-saving effects of both MBES and SBES are more pronounced than at 2.5 Gbit/s. Furthermore, the advantages of both schemes increase as the number of SFUs grows. Specifically, for the total system energy consumption, the maximum savings achieved by SBES and MBES relative to NES reach 21.41% and 17.75%, respectively, compared with 10.43% and 7.28% at 2.5 Gbit/s. For average SFU energy consumption, the maximum savings are 29.11% (SBES) and 23.30% (MBES), substantially higher than the 17.76% and 13.62% observed at 2.5 Gbit/s. These results demonstrate that the energy-saving potential of buffer-based schemes increases significantly with higher SFU optical transmitter power. This finding also highlights the effectiveness of the proposed strategy for next-generation FTTR systems, which are expected to adopt PONs with even higher data rates, such as 50G PON.

To evaluate the impact of the number of STAs on energy-saving performance and service latency, the number of SFUs is fixed at 2, while the number of STAs per SFU is increased from 1 to 4. Traffic and scheduling parameters are listed in Tables 1 and 2, respectively. Fig. 8 shows the change in average FTTR

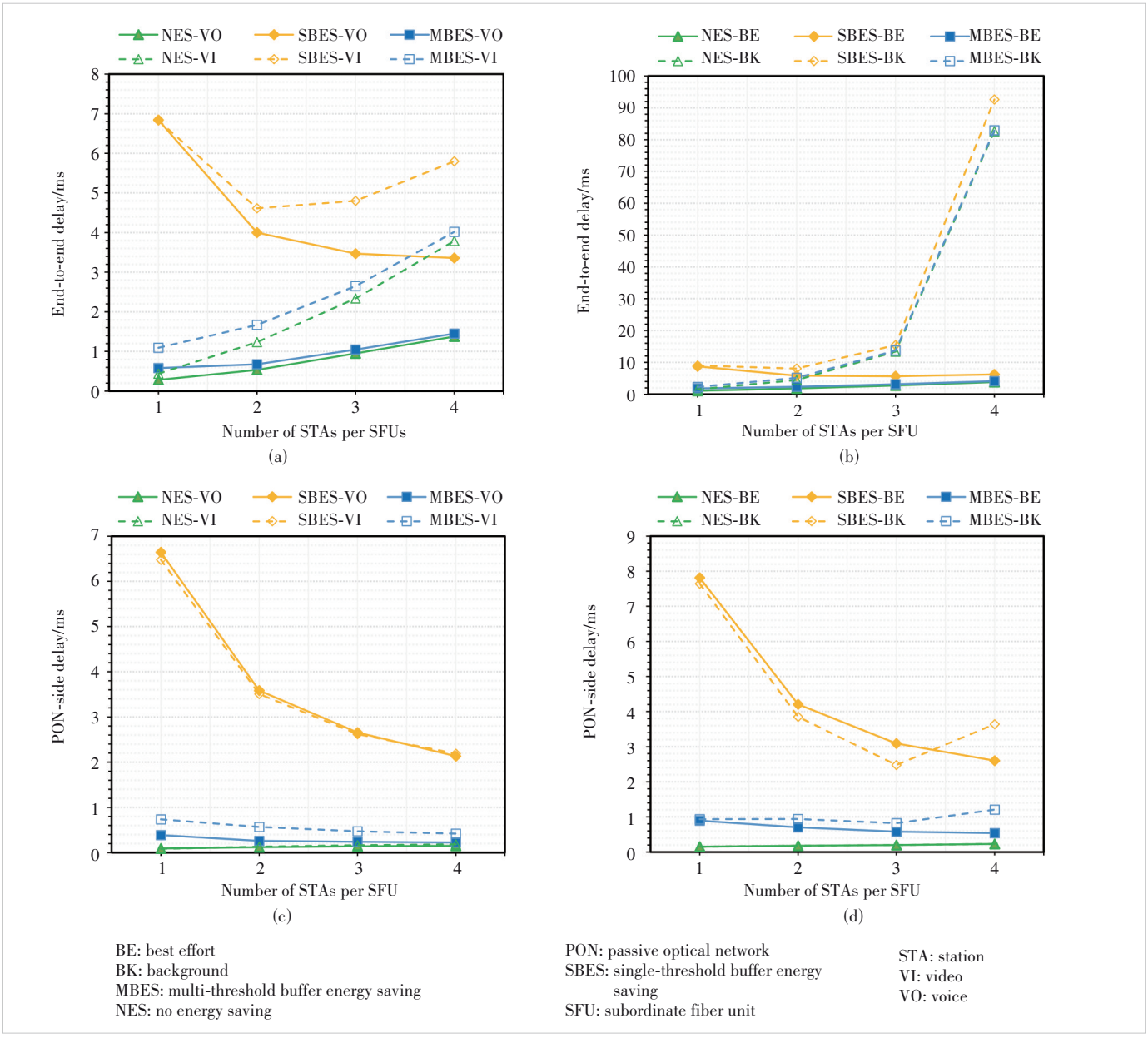

Figure 8. System latency under the NES, SBES, and MBES schemes (SFU=2): (a) end-to-end latency of high-priority services, (b) end-to-end latency of low-priority services, (c) optical-side latency of high-priority services, and (d) optical-side latency of low-priority services

system latency as the number of STAs increases. When two SFUs are deployed, end-to-end latency for all priority services rises with the number of STAs under both MBES and NES schemes. This is primarily due to increased contention on the Wi-Fi side, leading to higher EDCA-based access latency. Under NES, optical-side latency fluctuates only slightly because SFUs report every cycle and system bandwidth is sufficient. In MBES, more STAs increase the traffic arrival rate at the SFU, allowing buffers to reach thresholds faster, reducing SFU sleep time and optical-side latency. However, sleep duration is largely governed by the low thresholds of high-priority services, so SFUs do not enter long sleep periods even under low traffic loads. Consequently, optical-side latency fluctuations under MBES remain minor, and end-to-end latency is dominated by EDCA contention. As shown in Figs. 8c and 8d, although optical-side latency for BK services increases when the number of STAs reaches 4, end-to-end latency is still primarily determined by Wi-Fi access contention. Under SBES, with a small number of STAs, end-to-end latency is mainly affected by the queuing delay needed to fill the buffer. As the number of STAs increases, higher traffic volumes allow thresholds to be reached faster, reducing this delay. However, beyond a certain number of STAs, EDCA contention dominates, causing latency to rise again. For BK services, optical-side latency increases significantly when the number of STAs is 4, because BK traffic lacks guaranteed bandwidth; even after the buffer threshold is met, its scheduling is often postponed by high-priority traffic, leading to accumulated latency.

Fig. 9 compares the total FTTR system energy consumption and the average SFU energy consumption under the NES, MBES, and SBES schemes as the number of STAs per SFU increases from 1 to 4, with the number of SFUs fixed at 2. The results show that as the number of STAs grows, higher traffic volumes cause the buffer thresholds in the energy-saving schemes to be reached more quickly. This prompts SFUs to activate their optical transmitters more frequently for data reporting and transmission, reducing sleep time and gradually diminishing the energy-saving effects of all three schemes. Specifically, for total system energy consumption, savings from MBES decrease from 9.12% to 4.65%, while SBES savings decrease from 10.33% to 7.39%. For average SFU energy consumption, MBES savings decline from 17.05% to 8.70%, and SBES savings decrease from 19.32% to 13.83%.

To evaluate the impact of buffer threshold settings on the energy-saving strategy, we compare the performance of MBES under dynamic and fixed thresholds. The scenario uses 2 SFUs, each connected to 2 STAs. Three MBES configurations are tested: dynamic thresholds (starting from low fixed values), fixed low thresholds (VO threshold set to 1 500 B, with other priorities scaled proportionally), and fixed high thresholds (VO threshold set to 20 KB, with other priorities unchanged). Traffic parameters, dynamic threshold adjustment parameters, and latency requirements are provided in Tables 1 and 4. As shown

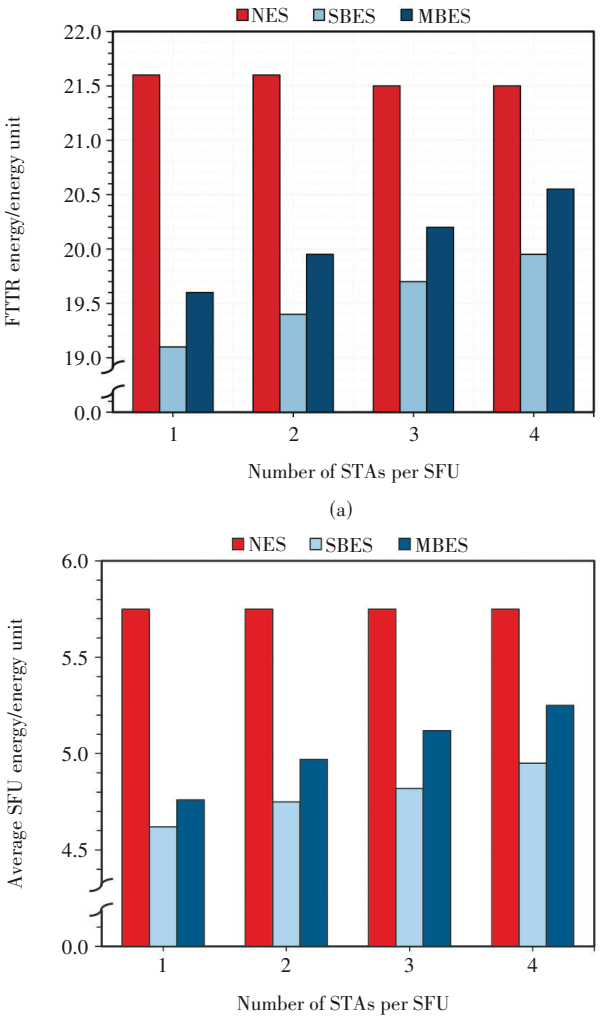

Figure 9. System energy consumption under the NES, SBES, and MBES schemes (SFU=2): (a) FTTR system energy consumption and (b) average SFU energy consumption

in Fig. 10a, the fixed-high-threshold scheme achieves the best energy savings, reducing average SFU consumption by 14% and total system consumption by 7.4%. However, the higher buffer threshold increases the buffer filling time, resulting in a significant reporting delay. Fig. 10b shows that VO service latency rises to 2.36 ms—an increase of 1.82 ms compared to the NES scheme—failing to meet the QoS requirement. In contrast, the fixed-low-threshold scheme ensures latency performance, with VO latency increasing by only 0.05 ms, but its energy savings are limited, achieving only 4.8% reduction in average SFU consumption and 2.5% reduction in total system consumption. Finally, the dynamic threshold scheme provides a better trade-

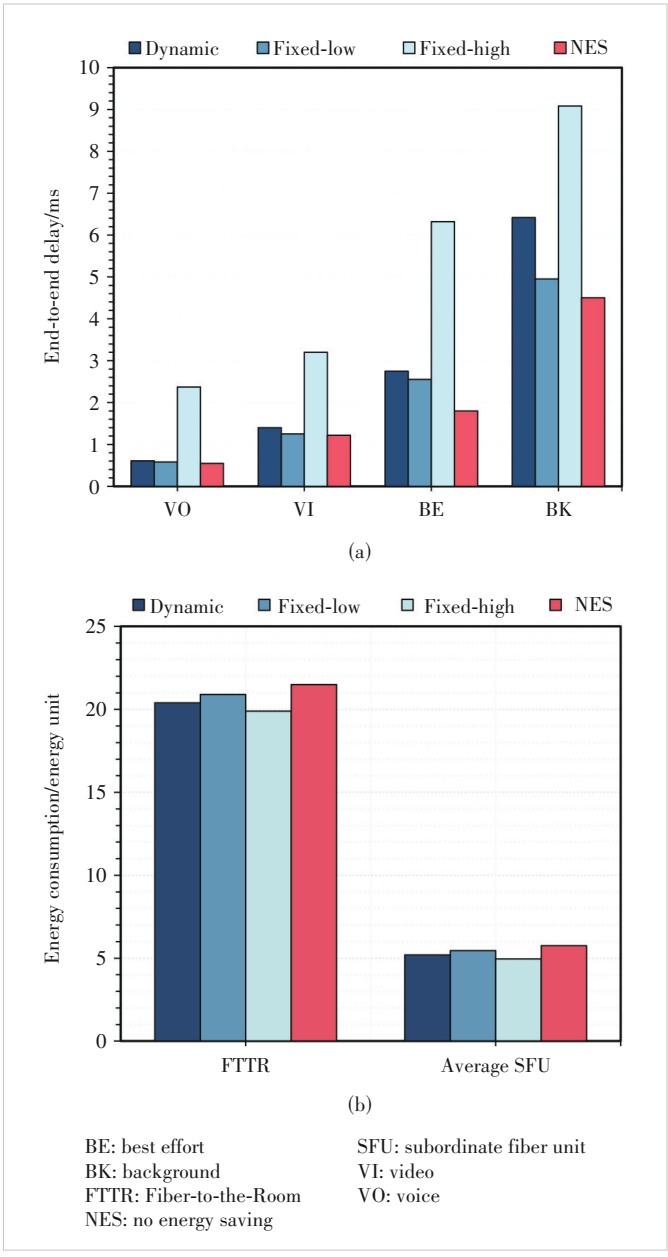

Figure 10. System latency and energy consumption under dynamic and fixed buffer thresholds: (a) latency comparison and (b) energy consumption comparison

off between energy and latency. By starting with low thresholds and gradually increasing them for services that consistently meet latency requirements, this approach reduces transmitter activation frequency and improves energy efficiency. Simulation results indicate that dynamic thresholds increase VO latency by only 0.083 ms while reducing average SFU consumption by 9.7% and total system consumption by 5.1%. Thus, the dynamic threshold scheme, through its adaptive adjustment mechanism, overcomes the limitations of fixed-threshold approaches and achieves co-optimization of latency assurance and energy savings.

6 Conclusions

This paper proposes an MBES scheme to address the increasing energy consumption in FTTR networks. By leveraging the architectural features of FTTR, the scheme enables efficient coordination between the MFU's centralized scheduling and the SFU's differentiated buffering mechanism, reducing system energy consumption while maintaining multi-level QoS. Specifically, at the SFU side, independent and dynamically adjustable buffer thresholds are assigned to different priority services, providing differentiated handling for latency-sensitive traffic. At the MFU side, the algorithm incorporates cooperative scheduling and dynamic threshold adjustment, taking Wi-Fi access latency into account during bandwidth allocation and using real-time performance feedback to form a closed-loop, cross-domain adaptive energy-saving mechanism.

Simulation results validate the effectiveness of the proposed MBES scheme. Compared with NES and SBES, MBES has achieved superior performance in meeting the strict latency requirements of high-priority services, such as VO and VI, while significantly reducing system energy consumption. It also provides better control over average system latency than SBES and maintains stable QoS guarantees under high-load, multi-user scenarios. In terms of energy savings, MBES has demonstrated considerable potential, particularly in future 10 Gbit/s scenarios with higher-power optical modules, achieving savings of up to 17.75%. Furthermore, for the MBES scheme, comparisons between dynamic and fixed threshold configurations show that the dynamic adjustment mechanism delivers a superior balance between latency and energy consumption, adaptively optimizing the performance-efficiency trade-off as traffic conditions vary.

References

[1] MAROTTA A, VALCARENGHI L, KONDEPU K, et al. Fiber to the room challenges and opportunities [C]//The 16th International Conference on COMmunication Systems & NETworkS (COMSNETS). IEEE, 2024: 1139 – 1142. DOI: 10.1109/COMSNETS59351.2024.10427248
[2] ETSI. Fifth Generation Fixed Network (F5G); F5G generation definition release #1 [S/OL]. (2020-12-10) [2025-09-10]. https://www. etsi. org/deliver/etsi_gr/F5G/001_099/001/01.01.01_60/gr_F5G001v010101p.pdf
[3] ZHANG D C, ZHU J L, LIU X, et al. Fiber-to-the-room: a key technology for F5G and beyond [J]. Journal of optical communications and networking, 2023, 15(9): D1 – D9. DOI: 10.1364/JOCN.485070
[4] Broadband Development Alliance. White paper on fiber to the room (FTTR) [R/OL]. [2025-09-10]. http://www.chinabda.cn/article/252789
[5] CAI J H, SHEN G X, LI J, et al. Is fiber-to-the-room (FTTR) green? Modeling and analysis of power and energy consumption [J]. IEEE transactions on green communications and networking, 2025, 9(2): 522 – 535. DOI: 10.1109/TGCN.2024.3447874
[6] ITU-T. G.984.3: gigabit-capable passive optical networks (G-PON): transmission convergence layer specification [S/OL]. [2025-09-10]. https://www.itu.int/rec/T-REC-G.984.3/en
[7] ITU-T. G. 987.3: 10-Gigabit-capable passive optical networks (XG-PON): transmission convergence (TC) layer specification [S/OL]. [2025-09-10]. https://www.itu.int/rec/T-REC-G.987.3/en
[8] ITU-T. G.9804.2: higher speed passive optical networks—common transmis-

sion convergence layer specification [S/OL]. [2025-09-10]. https://www.itu.int/rec/T-REC-G.9804.2/en
[9] ITU-T. High speed fibre-based in-premises transceivers-data link layer [S/OL]. (2024-07-05) [2025-09-10]. https://www.itu.int/rec/T-REC-G.9942-202407-I/en
[10] YAN Y, WONG S W, VALCARENGHI L, et al. Energy management mechanism for Ethernet passive optical networks (EPONs) [C]//IEEE International Conference on Communications. IEEE, 2010: 1 – 5. DOI: 10.1109/ICC.2010.5502659
[11] SHI L, MUKHERJEE B, LEE S S. Energy-efficient PON with sleep-mode ONU: progress, challenges, and solutions [J]. IEEE network, 2012, 26(2): 36 – 41. DOI: 10.1109/MNET.2012.6172273
[12] ZHANG J J, ANSARI N. Toward energy-efficient 1G-EPON and 10G-EPON with sleep-aware MAC control and scheduling [J]. IEEE communications magazine, 2011, 49(2): s33 – s38. DOI: 10.1109/MCOM.2011.5706311
[13] BUTT R A, AKHUNZADA A, FAHEEM M, et al. Enhanced energy savings with adaptive watchful sleep mode for next generation passive optical network [J]. Energies, 2022, 15(5): 1639. DOI: 10.3390/en15051639
[14] ZIN A M, IDRUS S M, RAMLI A, et al. Performance evaluation of XG-PON with DBA based-watchful sleep mode [C]//The 7th International Conference on Photonics (ICP). IEEE, 2018: 1 – 3. DOI: 10.1109/ICP.2018.8533168
[15] LAMBERT S, LANNOO B, DIXIT A, et al. Energy efficiency analysis of high speed triple-play services in next-generation PON deployments [J]. Computer networks, 2015, 78: 68 – 82. DOI: 10.1016/j.comnet.2014.10.037
[16] DHAINI A R, HO P H, SHEN G X, et al. Energy efficiency in TDMA-based next-generation passive optical access networks [J]. IEEE/ACM transactions on networking, 2014, 22(3): 850 – 863. DOI: 10.1109/TNET.2013.2259596
[17] SHAH NEWAZ S H, CUEVAS A, LEE G M, et al. Evaluating energy efficiency of ONUs having multiple power levels in TDM-PONs [J]. IEEE communications letters, 2013, 17(6): 1248 – 1251. DOI: 10.1109/LCOMM.2013.043013.122648
[18] ZIN A M, IDRUS S M, ISMAIL N A, et al. Determination of optimized sleep interval for 10 gigabit-passive optical network using learning intelligence [J]. International journal of electrical and computer engineering (IJECE), 2022, 12(3): 2663. DOI: 10.11591/ijece.v12i3.pp2663-2671
[19] IEEE. IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements part 11: wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN (802.11ax-2021) [S/OL]. [2025-09-10]. https://ieeexplore.ieee.org/document/9442429
[20] LORINCZ J, CAPONE A, BEGUŠIĆ D. Heuristic algorithms for optimization of energy consumption in wireless access networks [J]. KSII transactions on Internet and information systems, 2011: 626 – 648. DOI: 10.3837/tiis.2011.04.001
[21] PACK S, CHOI Y. An adaptive power saving mechanism in IEEE 802.11 wireless IP networks [J]. Journal of communications and networks, 2005, 7(2): 126 – 134. DOI: 10.1109/JCN.2005.6387860
[22] GARROPPO R G, NENCIONI G, PROCISSI G, et al. The impact of the access point power model on the energy-efficient management of infrastructured wireless LANs [J]. Computer networks, 2016, 94: 99 – 111. DOI: 10.1016/j.comnet.2015.11.018
[23] SILVA P, ALMEIDA N T, CAMPOS R. A comprehensive study on enterprise Wi-Fi access points power consumption [J]. IEEE access, 2019, 7: 96841 – 96867. DOI: 10.1109/ACCESS.2019.2928754
[24] DEMBÉLÉ H, TOHME E, COLIN R. Assessing and modeling the energy consumption of PoE-powered WiFi access point [J]. IEEE access, 2023, 11: 74796 – 74804. DOI: 10.1109/ACCESS.2023.3295689
[25] NISHIYAMA H, TOGASHI K, KAWAMOTO Y, et al. A cooperative ONU sleep method for reducing latency and energy consumption of STA in smart-FiWi networks [J]. IEEE transactions on parallel and distributed systems, 2015, 26(10): 2621 – 2629. DOI: 10.1109/TPDS.2014.2360405
[26] CHOWDHURY P, TORNATORE M, SARKAR S, et al. Building a green wireless-optical broadband access network (WOBAN) [J]. Journal of lightwave technology, 2010, 28(16): 2219 – 2229. DOI: 10.1109/JLT.2010.2044369
[27] HAN P C, GUO L, LIU Y J, et al. Joint wireless and optical power states scheduling for green multi-radio fiber-wireless access network [J]. Journal of lightwave technology, 2016, 34(11): 2610 – 2623. DOI: 10.1109/JLT.2016.2529644
[28] SCHÜTZ G, CORREIA N. Design of QoS-aware energy-efficient fiber-wireless access networks [J]. Journal of optical communications and networking, 2012, 4(8): 586 – 594. DOI: 10.1364/JOCN.4.000586
[29] ISLAM M M, FUNABIKI N, SAHA M, et al. An improvement of throughput measurement minimization method for access-point transmission power minimization in wireless local-area network [C]//International Conference on Consumer Electronics (ICCE-TW). IEEE, 2019: 1 – 2. DOI: 10.1109/ICCE-TW46550.2019.8991715
[30] LELIGOU H C, LINARDAKIS C, KANONAKIS K, et al. Efficient medium arbitration of FSAN-compliant GPONs [J]. International journal of communication systems, 2006, 19(5): 603 – 617. DOI: 10.1002/dac.761
[31] Huawei Technologies Co., Ltd. Huawei OptiXstar V173 Datasheet 01 [DB/OL]. (2023-05-15)[2025-09-10]. https://e.huawei.com/cn/products/optical-terminal/optixstar-v173
[32] Huawei Technologies Co., Ltd. Huawei OptiXstar K153 Datasheet 01 [DB/OL]. [2025-09-10]. https://e.huawei.com/cn/products/optical-terminal/optixstar-k153
[33] DHAINI A R, HO P H, SHEN G X. Toward green next-generation passive optical networks [J]. IEEE communications magazine, 2011, 49(11): 94 – 101. DOI: 10.1109/MCOM.2011.6069715
[34] ZHANG J, LIU J, XIANG L, et al. Full-link AoI analysis of uplink transmission in next-generation FTTR WLANs [C]//Proceedings of IEEE 97th Vehicular Technology Conference (VTC2023-Spring). IEEE, 2023: 1 – 7. DOI: 10.1109/VTC2023-Spring57618.2023.10200633

Biographies

CAI Jinhan received his BE degree from Nanjing University of Science and Technology Zijin College, China in 2021 and MS degree in information and communication engineering from Soochow University, China in 2024, where he is currently pursuing his PhD degree with the School of Electronic and Information Engineering. His research interests include passive optical networks and fiber-wireless access networks.

ZAN Mingyuan received his BE degree from Nanjing University of Information Science and Technology, China in 2022 and MS degree in information and communication engineering from Soochow University, China in 2025. His research interests include passive optical networks and fiber-wireless access networks.

SHEN Gangxiang (shengx@suda.edu.cn) received his BE degree from Zhejiang University, China, MS degree from Nanyang Technological University, Singapore, and PhD degree from The University of Alberta, Canada in 2006. He is a Distinguished Professor and the Dean of the School of Electronic and Information Engineering at Soochow University, China. He previously worked as a Lead Engineer with Ciena, USA, and an ARC Postdoctoral Fellow with The University of Melbourne, Australia. His research interests include spectrum-efficient, green optical networks and integrated optical-wireless systems. He has authored over 300 peer-reviewed papers and received the Izaak Walton Killam Memorial Scholarship and the NSERC Fellowship. He was recognized as a Highly Cited Chinese Researcher by Elsevier (2014 – 2024) and is an Associate Editor of the *IEEE/OPTICA Journal of Lightwave Technology*. He is an OSA Fellow and was an IEEE Distinguished Lecturer of ComSoc.