To be published in Journal of Optical Communications and Networking:

Title: ERON: An Energy-Efficient and Elastic RF-Optical Architecture for mm-Wave 5G Radio Access Networks

Authors: Hongbo Lu,Roberto Proietti,Gengchen Liu,Xiaoliang Chen,S. J. Ben Yoo

Accepted: 13 April 20

Posted 15 April 20

DOI: https://doi.org/10.1364/JOCN.390428

ERON: An Energy-Efficient and Elastic RF-Optical Architecture for mm-Wave 5G Radio Access Networks

HONGBO LU[1], ROBERTO PROIETTI[1], GENGCHEN LIU[1], XIAOLIANG CHEN[1], AND S.J.BEN YOO[1]

[1] Department of Electrical and Computer Engineering, University of California Davis, Davis, CA, 95616 USA
[*] Corresponding author: sbyoo@ucdavis.edu

Compiled March 31, 2020

This paper proposes an Elastic RF-Optical Networking(ERON) architecture solution for millimeter-Wave 5G radio access networks. The ERON architecture achieves energy efficiency and throughput elasticity by utilizing photonic-enhanced multi-beam mmW spatial multiplexing capability at the radio units. The centralization of the hardware resources and the converged management of the RF and optical resources in the data units offers high resources pooling gain. A first numerical study on the energy efficiency of an ERON's photonic-enabled mmW 5G system reveals that ERON is 5x more energy-efficient than both conventional digital and hybrid RF beamforming implementations. We also conducted a user mobility-aware network resources study whose results show a 10-dB network resource pooling gain when compared to classic RAN implementations.

http://dx.doi.org/10.1364/ao.XX.XXXXXX

1. INTRODUCTION

The new generation of mobile networking, known as 5G, has proposed three promising features: the enhanced Mobile Broad Band, the Ultra-Reliable Low Latency Communications, and the massive Machine Type Communications for the Internet of Things[1]. These substantial improvements for the mobile connections exert unprecedented technological challenges to both wireless networks and fiber-connected radio access networks (RANs). The entire network architecture is yearning for an overhaul in every aspect of its design: from the radio unit signal generation to the 5G transport network signal transmission schemes and the control unit spectrum selection algorithms.

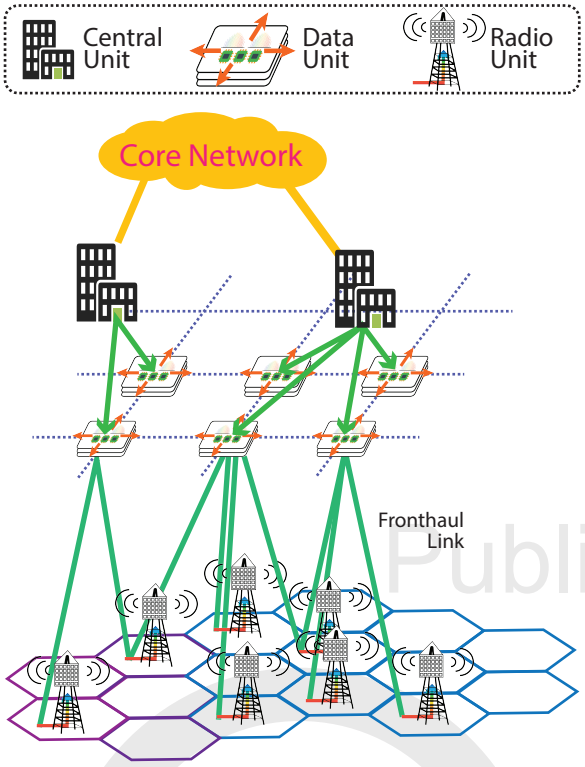
The convergence of photonic and RF technologies is a promising solution to 5G networks. Photonic signal processing technologies have been contributing to the mmW 5G research related to the mmW generation with Radio-over-Fiber(RoF) scheme. Ref [2] summarized their work on Radio-over-Fiber(RoF) for mmW communication systems, a prime spectrum proposed in the 5G NR FR2[3]. Microwave photonics is an active research area, which applies photonic technologies to microwave applications[4–6].

On the other hand, the 5G radio access network (RAN) mainly focuses on digital signal transmission as a communication pipeline between the data units (DU) and the radio units (RU). Topics such as fronthaul-network optimization and base unit placement studies are some of the popular field. Reference [7] studied the trade-off between the number of baseband unit (BBU) hotels, the BBU-hotel placement, and the network-capacity utilization in wavelength-division multiplexing (WDM) aggregation networks. Reference [8] proposed a WDM fronthaul with a partitioned asymmetric arrayed waveguide grating router and distributed feedback Bragg (DFB) directly modulated lasers (DMLs). Reference [9] investigated the joint photonic-radio resource provisioning in a centralized radio access network (C-RAN) for the hybrid sub-array antennas. All these works tried to offer solutions to 5G RAN, but their scope solely focuses on the optical signal transmission technologies such as the Common Public Radio Interface(CPRI) and evolved CPRI(eCPRI) C-RAN architecture[10].

A joint study uniting the photonic signal processing for mmW 5G and the design of the mmW 5G RAN can break a new ground and possibility to an energy-efficient and throughput-elastic architecture.

In this paper, we present the Elastic RF-Optical Network (ERON) architecture with superior energy efficiency and resource utilization flexibility. The improvements in energy and resource usage benefit from the combination of the photonic signal processing technologies and the converged mmW-optical signal orchestration algorithm. This architecture study based on our previous works[11–13] describes the ERON architecture and demonstrates its energy and resource efficiency with numerical analysis. This new architecture provides a solution to the mmW power consumption challenge[14], the RAN fronthaul link capacity crunch challenge[15], and the surging resource requirement challenge[16].

Figure 1 shows the overall architecture of the proposed ERON system. It consists of three components: the radio unit (RU), the data unit (DU), and the central unit (CU). We list the main differ-



	C-RAN	ERON
Fronthaul Link	Digital Signal trannsmission	Radio-over-Fiber transmission
Radio Unit	RF-based beamforming	Photonic-enabled beamforming
Data Unit	Centralization varies according to different options	High-level centralization RF-optical co-orchestration

Fig. 1. The high-level schematics of the ERON structure. The green lines are the fronthaul links connecting the data units and the radio units. The length of the fronthaul link is up to 20 km. The brown lines are the midhaul link connecting the central units and the data units. The length of the midhaul link is up to 80km with 400-ZR optical modules[17].

ence between the proposed ERON architecture and the C-RAN scheme. The fronthaul link connects the RUs and the DUs with analog Radio-over-Fiber (RoF) technology which facilitates the photonic signal processing technologies and maximizes energy efficiency in the RU. The DUs are designed with high-level centralization and RF-optical co-orchestration for efficient resource utilization. In this paper, we emphasize the design on the fronthaul link and its connecting units: RUs and DUs.

Figure 1 is a high-level illustration of the ERON architecture where RUs are drawn with the same coverage and design for illustration clarity. In practical deployments, the RUs have different coverage and performance to accommodate the deployment requirement to achieve optimal throughput, coverage, energy efficiency, and latency.

The design of the RUs prioritizes energy efficiency and throughput elasticity, at the same time maintains its structural simplicity. To achieve high energy efficiency, we adopt photonic signal processing for mmW beam forming and steering. In this study, we conduct a first numerical analysis of the energy efficiency for the proposed photonic integrated circuit solution. We derive a photonic-to-mmW energy consumption model for the proposed RU and benchmark its energy efficiency with the state-of-art RF schemes.

The design of the DUs focuses on resource centralization and flexible resource orchestration. The DU centralizes all base-band hardware resources including the RoF signal generation, radio resource scheduler, and user mobility information collections, facilitating hardware and software management and optimizing operational complexity. The centralized design provides significant hardware pooling gain and improves latency performance. The RoF links connect the RUs with the DUs and the DUs connect to the CUs in a similar fashion to the current C-RAN design. We analyze the impact of the RoF technology on RAN latency and simulate the user mobility-aware environment to study the gain from centralization pooling.

The study of C-RAN pooling gain has been an important topic in recent years. [18–20] investigated the adaptive bandwidth allocation algorithms to accommodate the large traffic demand for the CPRI-based fronthaul link. [8, 21] proposed to use an Arrayed Waveguide Grating Router (AWGR) as a routing photonic device to improve the fronthaul bandwidth utilization. In this study, we emphasize the novelties of our pooling gain study from two perspectives:

1) The ERON pooling gain study is based on a analog RoF (aRoF) fronthaul design instead of a CPRI-based design. The allocation of RoF and BBU resources in the DU directly addresses the wireless connection requests.

2) We include the user mobility behavior to the network simulation which is important in the RF-optical converged network scenario. As the RU coverage gets smaller in the mmW scheme, connection requests coming from the neighboring RUs become prevalent. Statistical traffic models cannot accurately emulate this change.

The advantages of energy efficiency and pooling gain are codependent. The energy-efficiency of RUs makes mmW beamforming system practical. The beamforming capability amplifies the pooling gain with SDM-aware network resource allocations.

The remainder of this paper elaborates on the design rationale and details of the ERON architecture with supporting numerical analyses. Section 2 analyzes the three challenges on the 5G RAN designs: energy , throughput, and resources. Section 3 describes the design of the ERON architecture and explains the principle of multi-beam generation. Section 4 analyzes the power consumption of the ERON RU beamformer and compares its performance against RF digital, analog, and hybrid beamformer. Section 5 analyzes the hardware resource pooling gain of the proposed architecture with user equipment(UE) mobility-aware simulation using a converged RF-optical resource allocation algorithm. Last, Section 6 concludes this paper.

We use the following letter to represent some physical parameters:

- N: Number of the antenna elements
- U: Number of the spatial beams
- Bw: mmW frequency window bandwidth
- C: System throughput
- λ: Carrier frequency
- $\#bits$: Number of bits in SerDes circuits

2. CHALLENGES FOR 5G RAN

mmW technology can achieve the 5G capacity goal of 10-Gbps peak speed. The 2-GHz spectrum window available at mmW carrier is 10x larger than the carrier used in the 4G-LTE system. However, the adaptation of mmW for 5G wireless access network introduces three challenges in the design of the 5G RAN: the surge of power consumption, the increasing hardware resource requirement, and the volatility of the traffic demand.

A. Power Consumption

The power consumption surge in the adaptation of mmW carrier is driven by the three factors: power increase to compensate for 1) the extra free space propagation loss; 2) the penetration loss; 3) the extra noise due to a larger spectrum window.

A.1. Free Space Propagation Loss and Penetration loss

The free space propagation loss (FSPL) and the penetration loss outside the Kirchhoff area follow the Friis transmission formula[22]:

$$FSPL = G_{tx} + G_{rx} + 20 \cdot \log_{10} \lambda - n \cdot 10 \cdot \log_{10}(4\pi d) \quad \textbf{(1)}$$

where the G_{tx} and G_{rx} are the antenna gains for the transmitter and receiver, respectively. d is the propagation distance. n refers to the path loss exponent, which is equal to 2 in free space and can range from 2 to 6 different scenarios due to the penetration loss.

The penetration loss of mmW is higher than the radio wave as the carrier frequency increases. The average penetration losses for brick and tinted glass are about 28 dB and 40 dB, with the losses for drywall and clear glass, which are relatively low [23]. The non-light-of-sight (NLOS) loss can be so high that the communication channel cannot be established[24].

The carrier migration from 2-GHz carrier in 4G system to mmW carrier requires extra 20-dB higher transceiver link power according to the equation 1.

A.2. Noise issue

The noise source in the radio signal detection is the thermal noise. The thermal noise follows[25]:

$$P_{thermal} = k_B T \delta f \quad \textbf{(2)}$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and δf is the bandwidth. The 10-times increase of the bandwidth alleviates the noise level by 10 dB. Therefore, higher signal power is necessary to compensate for the noise increase.

Massive MIMO with large phased array to enable multi-beam spatial division multiplexing access (SDMA) is a promising solution to the power consumption surge. Assume the output power of each power amplifier for a single mmW beam has P_{ele} dBm power, the user equipment with an isotropic antenna receives a signal power equal to:

$$P_{UE} = P_{ele} + 20 \cdot \log_{10} N - FSLP - P_{pen} - P_{thermal} \quad \textbf{(3)}$$

where N is the number of array elements. FSLP is the result from equation 1, and P_{pen} is the extra penetration loss. $P_{thermal}$ is the thermal noise from equation 2

We show that the massive array antenna and the multiple spatial beam capability is necessary to support the 5G eMBB. Assuming the parameters listed in Table 1, we can calculate the required power per element to achieve the 10-Gbps throughput as a function of the number of array elements, as shown in Figure 2.

Table 1. Link budget estimation parameters

Carrier Frequency (GHz)	28
Bandwidth (MHz)	850
Propagation Distance (Meter)	500
FSLP (dB)	115.4
Penetration Loss (dB)	12.7
Temperature (K)	300
Power Budget Margin(dB)	5
Rx Antenna Gain(dBi)	0

Figure 2 shows the energy efficiency using the spatial domain to boost link throughput. The four lines in Figure 2 are the mmW system power requirements with 1, 2, 4, and 8 spatial beams. The 1-beam scenario is impractical for implementation because of its overwhelming power consumption. It demands each element with 25-dBm power even with 1024 elements, whose total power can reach over 1,000 W. The spatial multiplexing multi-beam schemes, on the other hand, are energy efficient to achieve the same throughput with much lower power requirement per antenna element. It requires 7dBm, -2dBm, and -8dBm for 2-beam, 4-beam, and 8-beam scene respectively to achieve 10-Gbps throughput with 1024 antenna elements.

The reason behind the improved energy efficiency of the SDMA scheme relates to how the throughput is composed of. In wireless communications, there are three dimensions to increase the throughput: the type of the modulation formats, the width of the spectrum window, and the number of the spatial channels. The modulation format transmits the information by the number of bits encoded on a constellation and its information rate relates to $logN$. The spectrum window and the spatial channels transmit the information by the spectrum width NHz and the number of the spatial beams N, both of which contribute to the throughput linearly. The linearity throughput increase using spatial channels overwhelms the logarithm one using modulation formats. The adaptation of beamforming completes the critical piece to use additional spatial channels in mmW communication. As the result shown in Figure 2, the future wireless system should migrate toward larger spectrum window such as mmW and multiple beams for SDMA as its primary approach to increase the overall capacity.

Therefore, the capability of the multi-beam SDMA scheme is fundamental to tackle the power consumption issue. We design the ERON architecture with excellent energy efficiency performance comparing to other multi-beam SDMA implementations. We discuss the energy efficiency numerical analysis in section 4.

B. Resource Requirement

The mmW 5G system requires more hardware resources comparing to the 4G-LTE system. Such increase stems from two aspects: 1) the increased deployment density; 2) the increased number of RF and BBU units.

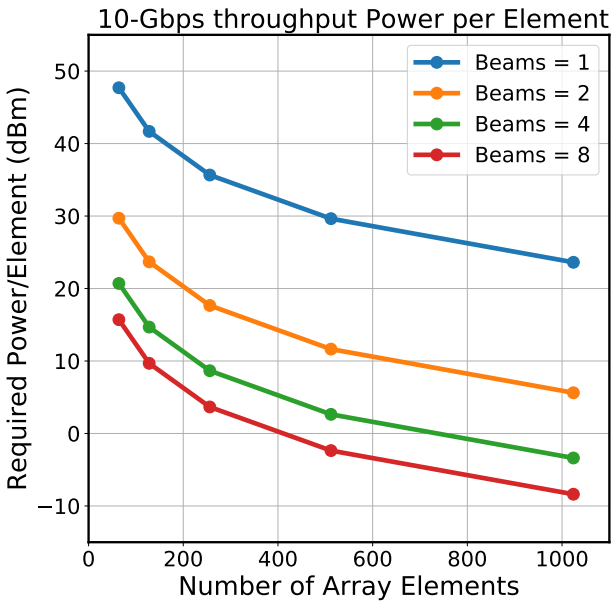


Fig. 2. Required output power per antenna array element to achieve 10-Gbps throughput with 850-MHz available of mmW spectrum. The blue, orange, green, and red line represents 1, 2, 4, and 8 spatial mmW beams, respectively.

The increased RU deployment density is inevitable as the mmW carrier makes the wireless environment more susceptible to penetration loss and shading effects. A region which is covered by a macro base station in the 4G system now requires dozens of micro and femto base stations to provide reliable mmW connections. Assuming the same transceiver structure as the 4G system, the migration from current 4G carrier to mmW 5G carrier increases the number of base stations by 10 times as the coverage of a base station is one tenth of the 4G system. In addition to the shrinkage of a single base station coverage, the higher penetration loss requires even more base stations comparing to the number of the 4G system.

The second reason for such a resource demand surge is the implementation of the massive MIMO solution to compensate for the increased signal loss in the mmW carrier. It is widely acknowledged that digital beamforming implementation will increase the number of digital-to-analog converters enormously [26]. Even with analog beamforming implementation, the RU needs to accommodate multiple sets transceiver for the multiple spatial-beam generations. Regardless of the implementation the RAN designer chooses, the mmW 5G requires multiple times of components for the massive MIMO scheme.

Here we show a back-of-the-envelope calculation. We assume the 4G system with 2.1 GHz carrier and a single antenna, and the mmW 5G system with 28 GHz carrier with a digital-implemented 64-element array antenna. The mmW 5G system requires 9 base stations to cover the same area of the 4G system. Each base station requires 64 sets of baseband unit due to the digital implementation. In total, the mmW 5G system requires an over 500-fold increase for the coverage of a single 4G base station. Furthermore, the hardware utilization rate will be low even the large investment of the base stations is possible.

Therefore, a centralized scheme and flexible resource orchestration capability are essential to make a practical 5G mmW RAN. We designed the ERON architecture with efficient resource allocation capability and high hardware resource pooling gain comparing to the conventional methods. With the capability of SDM at mmW band, the centralization scheme allows the RAN to explore better network resource allocation methods. The system can fully utilize the spatial-temporal-spectral coding to facilitate the users' connection requests. In Section 5, we discuss the resource pooling gain quantitative analysis with location-aware resource allocation method.

C. Traffic Demand Volatility

The traffic demand volatility is defined as the change of the traffic throughput as a function of the time. The traffic demand volatility is high in the 5G mmW scenario as the coverage of RU shrinks and the user traffic demand increases. The traffic demand volatility is higher in the 5G mmW. It challenges the 5G mmW RAN design to meet the high ratio of the peak and average traffic volume. This phenomenon attributes to two factors: 1) the traffic patterns of the emerging applications; 2) the shrinking size of the RU coverage. We illustrate an example with network mobility simulation and analyze its impacts on this part.

We build a user mobility-aware network traffic simulation to show the traffic demand volatility. The simulation process works as the following. At the initial state of the simulation, there are N UEs within each RU. Each UE generates service requests following a Poisson distribution with a mean arrival rate equal to λ_1. The services duration follow an exponential distribution with a mean time interval equal to λ_2. The overall traffic load is $\lambda_1 \cdot \lambda_2$ Erlangs. We assume the UE mobility follows the Random Way Point (RWP) process to emulate the UE movement across time[27]. The principle of RWP is described as follows:

1. Each UE randomly selects a location (uniform distribution) in the simulation field as the destination.
2. The UE selects a velocity which is uniformly distributed in $(0, v_{max})$) and starts moving to the destination.
3. Move toward the destination until the destination is reached.
4. If the destination is reached, the UE stays for a time T (following an exponential distribution)
5. If T reaches 0, Repeated steps 1)-4).

Because the mobility and the traffic requests are both time-dependent, we run each simulation of a period of a 27.7-hour period (10^5 seconds). In the simulation, the mean pause duration is 200 seconds. The mobility velocity ranges from 1 m/s to 15 m/s. The simulation field is a $5km$ by $5km$ square area with 400 RUs uniformly distributed. Each RU covers a 250-meter by 250-meter square area.

Figure 3 shows the simulation result of the traffic throughput of individual RUs. Figure 3 (a) shows the simulation grid. We select three RU areas to observe the traffic demand variation as a function of time.

Figure 3 (b) shows the changes of the traffic demand as a function of time for the RU at A(4,5), B(8,8), and C(16,10). High volatility of the throughput is the common property of these three RUs. We observed the throughput fluctuations in each RUs ranged from 5 to 20 Gbps, 10 to 32 Gbps, and 5 to 15 Gbps for RU A, B, and C, respectively.

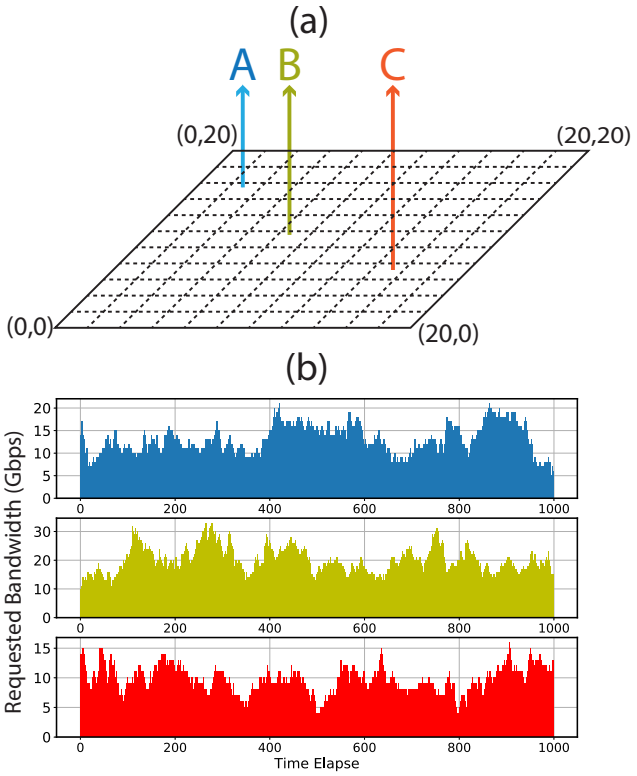


Fig. 3. Bandwidth requested from selected RUs over time in a user mobility-aware network traffic simulation.

The high volatility can be compensated by grouping RUs to maintain a statistical constancy. This result from the numerical mobility simulation analysis is consistent with the common belief that centralized control is essential to achieving efficient utilization of network resources under the ultra-dense 5G deployment scheme. As the cell size shrinks under the ultra-dense deployment scheme, the peak to average ratio of the throughput requirement continues to increase. The traditional RAN deployment scheme demands too much hardware resources to accommodate the peak traffic scenario. We designed two features to tackle this issue: 1) the throughput-elastic capability in the radio unit to satisfy the high peak to average ratio of the traffic demand pattern. We explain its principle in Section 3; 1) a centralized architecture with a converged RF-optical resource allocation algorithm, more detailed discussion in Section 5;

D. *Latency Requirement*

The ultra-low latency feature has been studied in several research works [10, 28, 29] and is an active research topic in future RAN designs. In particular, the C-RAN architecture has received much attentions to investigate its potential to reduce the latency in the RAN. Figure 4 (a) shows the latency composition in the scope of a wireless access system with CPRI-based C-RAN [10]. The fronthaul link is a digital link to connect the DU and the RU. The latency consists of the user processing delay, air propagation delay, RU processing delay, RU fronthaul processing delay, fiber propagation delay, DU fronthaul processing delay, and baseband processing delay. The latency of the RAN refers to the delay between the RU and DU. According to [29], the fronthaul latency requirement ranges from $100\mu s$ to $2ms$ depending on the option of functional splitting.

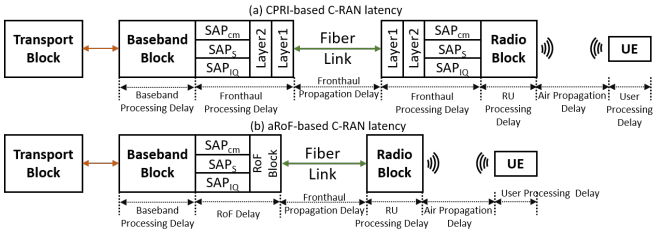


Fig. 4. The latency components of (a) CPRI-based C-RAN system; (b) aRoF-based C-RAN system. SAP: Service Access Point.

The analog-RoF (aRoF) based C-RAN system has received a renaissance in the mmW 5G era [30–33] as the bandwidth requirement in the fronthaul link using digital transport exploded [29]. The aRoF-based C-RAN is a variation of the C-RAN. The latency composition of aRoF-based C-RAN is shown in the Figure 4 (b). With the adoption of aRoF link in the fronthaul, the system eliminates the delay components for the digital communication link, such as the layer 1 and layer 2 on both the RU and DU side. The function of the RU simplifies and is only responsible for the optical to electrical signal conversion and mmW transmission, shown as the radio block in Figure 4 (b). Furthermore, it requires a single stack of Service Access Point (SAP) as the aRoF link, because it eliminates the signal to bit-sequence conversion happening on both sides of a digital transport link.

Therefore, an aRoF-based C-RAN structure like in ERON can improve the latency performance compared to a CPRI-based C-RAN system, as also reported in [28] where the authors show how the aRoF-based C-RAN can meet the 5G latency requirement.

As the ERON architecture proposed in this paper adopts an aRoF fronthaul link to facilitate the signal photonic processing at the RUs, it can achieve a comparable latency performance as the aRoF C-RAN in [28].

3. ERON ARCHITECTURE

We designed the ERON architecture to meet the 5G RAN requirements discussed above. Figure 1 shows the overall structure of the proposed architecture. The ERON architecture evolves the functional splitting C-RAN[10] and elastic optical networking[34], incorporates the analog Radio-over-Fiber (RoF), photonic signal processing and mmW beamforming technologies. The fronthaul links connect the RUs and the DUs and the midhaul links connect the DUs and the CUs. The RUs are responsible for establishing the mmW links with the UEs. The DUs are interconnected with each other. The CUs covers a large area of DUs. For the fronthaul link, we adopt the RoF technology to transmit the signals and allocate all hardware inside the DUs. We called it beyond option-8 functional splitting. Hardware resources are centralized which helps ease the hardware resource issue and reduce the deployment cost[35, 36].

In the following subsections, we describe the design of the RU, DU, and CU, respectively.

A. Radio Unit Design

The RU design employs photonic signal processing technology to achieve energy efficiency and to facilitate the use of SDMA with massive MIMO large phased array antennas. The RUs locate at the edge of the RAN and interface to the mobile UE

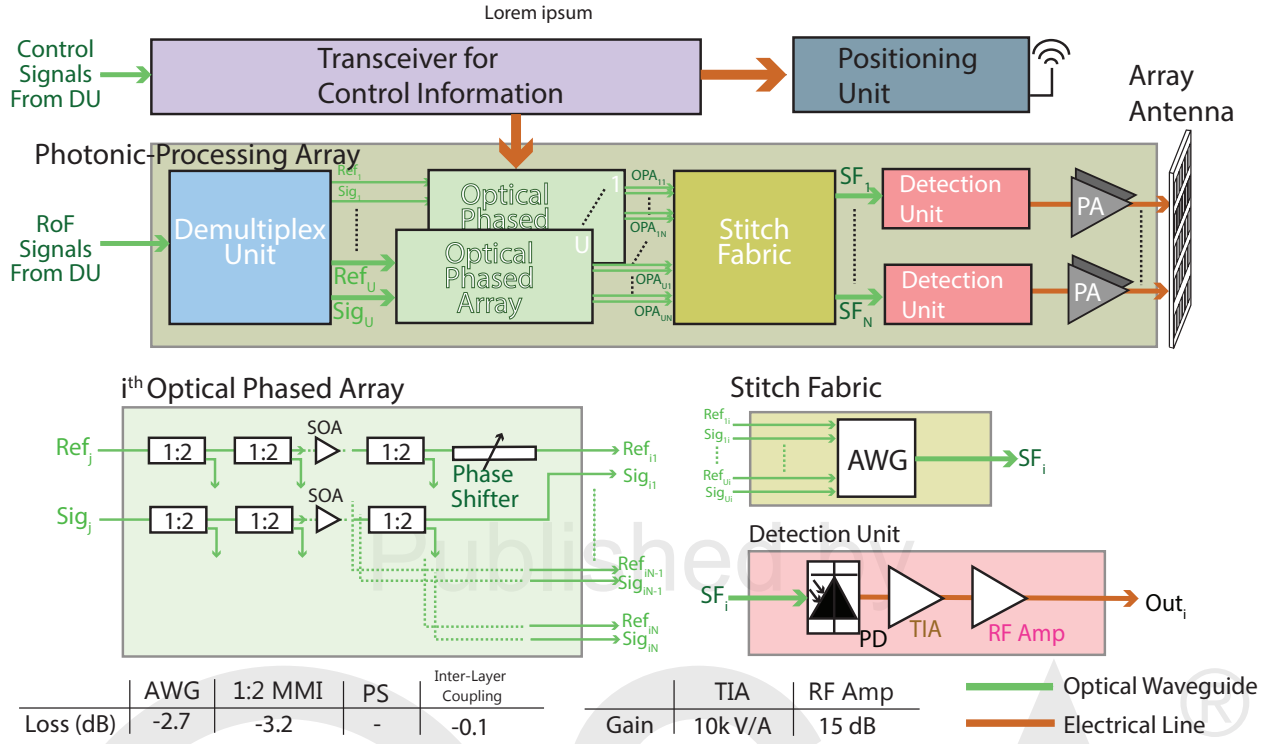


Fig. 5. The schematics of the RU physical structure. AWG: Arrayed waveguide grating; 1:2 MMI: Multi-mode interference coupler; PD: Photodetector; TIA: Transimpedance amplifier. SOA: Semiconductor optical amplifier. RF amp: RF amplifier. PA: Power amplifier.

with multi-user MIMO connections. It converts RoF signals to the mmW signal and delivers the mmW signal to the UEs through SDMA.

Figure 5 shows the RU structure. It consists of four components. At the top of the Figure 5, a control information transponder is responsible for communicating control information with its DU, such as phased array weight values, UE positions, etc. A positioning unit using sub-6-GHz frequency is connected to the control information transponder, shown on the right of the control information transponder in Figure 5. The positioning unit senses the UE mobility information to assist phased array weight synthesis. A photonic processing array (PPA) executes the photonic phased array generation on the RoF signals based on the instruction from the control information transponder. A patch-phased array antenna transmits the mmW radiation.

A.1. Photonic Processing Array

The PPA is responsible for the phased array generation. The PPAs are at the center of the RU and it enables the energy-efficient multi-user mmW beam generation. It minimizes the usage of mmW RF circuits, simplifies the RU structure and lowers power consumption as RF circuits are prone to losses in the mmW spectrum.

Figure 5 shows the PPA structure. The incoming RoF signals first enter an optical de-multiplex unit to separate the user signal wavelengths and their reference wavelengths. The de-multiplex unit separates the RoF signal into U wavelength pairs of the signal wavelength Sig_i and reference wavelength Ref_i, where $i \in (0, U)$, U is the number of the mmW beams. Each Signal/Reference pair is the input of an optical phased array circuit, which imposes a set of phased array weights on the reference wavelength signals. Each optical phased array circuit has N pairs of outputs, and each pair OPA_{ij} consists of a signal wavelength and a reference wavelength, where $i \in (0, U)$, and $j \in (0, N)$, N is the number of the antenna array element. The stitching fabric consists of an array of arrayed waveguide gratings(AWG). It takes the outputs of the optical phased arrays, OPA_{1i} to OPA_{Ui} where $i \in (0, N)$, as the input to each AWG, and generates output SF_i. Finally, the detection unit takes the SF_i in and converts the RoF signals into mmW signals. A trans-impedance amplifier(TIA) transfers the current signal to a voltage signal with $50 - \Omega$ load before an RF amplifier regulates the mmW signal and feeds it to the power amplifier.

The array antenna in the RU consists of N antenna elements. It transmits the mmW signal processed by the PPA. There is a power amplifier in front of each antenna element to amplify the input signal to the required signal power level.

A.2. Positioning Unit

The positioning unit uses sub-6-GHz radio frequency to carry out the two functions: 1) Emit positioning signals to help UE find its geolocation in the cellular system; 2) Receive UE location information periodically for cellular mobility management and mmW beam delivery. The geo-positioning signals enable the UE to calculate its location. The positioning calculation algorithms have been widely studied in cellular networks [37]. The UE informs the RAN on its location periodically:

$$T_{update} = \frac{FWHM \cdot d_{res}}{V_{Max}} \tag{4}$$

where $FWHM$ is the radius value of the antenna beam main lobe, d_{res} is the antenna resolution distance, and V_{Max} is the maximum speed allowed in the RU coverage area.

The UE positioning information solves two issues: First, for the mmW beam delivery, it directs the DU and RU to deliver the mmW beam to the target location through the cookbook look-up method [38]. It no longer requires beam scanning for an accurate beamforming location, which increases latency and power consumption; Second, for RAN management, it can reduce RAN operation latency by proactive user mobility management.

A.3. Control Information Transponder

Lastly, an optical transponder communicates the control and management information between the RU and the DU such as data plane timing, user beam weights, etc. The optical transponder can use commercially available PON transceivers working on a different wavelength window to the RoF signals, e.g. RoF signals are on the C-band and control information on the O-band.

A.4. Multi-beam Generation for Throughput Volitality

Finally, we explain the principle of the multi-beam generation in the ERON system, discuss its impact on tackling throughput volatility and its limitation.

The RU in the ERON architecture uses multi-beam generation to enable SDMA as discussed in section 2. The mmW circuit implementations for mmW beamforming have significant limitations[39]: the digital beamforming and hybrid beamforming methods can provide mmW multi-beam forming, but they may consume huge power which we will discuss in the next subsection. The analog beamforming is simpler and has acceptable power consumption, but only generates one data stream beam per circuit[26].

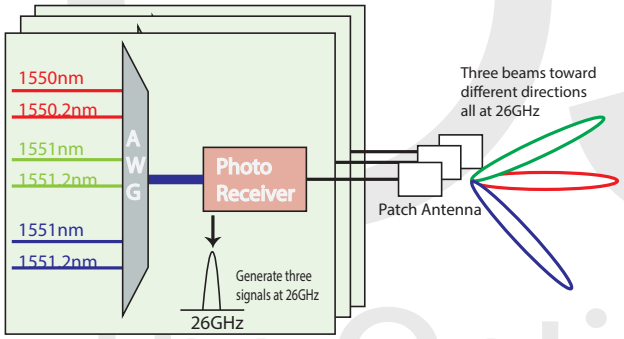


Fig. 6. Illustration of a three-mmW-beam generation using the ERON architecture.

The ERON RU can generate multiple mmW beams by taking advantage of the large optical wavelength window and wavelength orthogonality. A pair of the RoF signal RoF_1 with signals at λ_{sig1} and λ_{ref1}, can generate one spatial mmW beam $beam_1$ after PPA processing. Similarly, another pair of the RoF signals RoF_2 with signals at λ_{sig2} and λ_{ref2}, can generate one spatial mmW beam $beam_2$ at the same frequency. As long as the spectrum distance between RoF_1 and RoF_2 is larger than the mmW carrier frequency, there is no interference between the two mmW signals generated by these two sets of RoF signals. Figure 6 shows an example of a three-beam generation using the multiple pairs of RoF signal.

With the ITU-U DWDM 100-GHz channel spacing[40], there can be 80 spatial beams in one single RU. If the RoF signals are placed in the middle of each DWDM channel, it can guarantee no interference of the generated mmW signals.

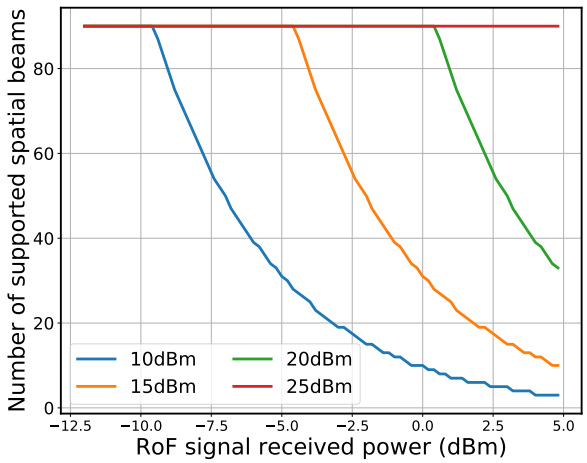


Fig. 7. Number of spatial beams as a function of received optical power of a pair of RoF signal for PD saturation power of 10dBm (blue), 15dBm (brown), 20dBm (green), and 25dBm (red).

This PPA-enhanced multiple mmW beam generation enables throughput elasticity in the ERON architecture. The number of RoF signals at the input of the PPA determines the number of mmW beams, and the number of the mmW beams determines the system throughput in the RU. Elastic management of the RoF signals and optical resources can, therefore, enable the mobile system throughput elasticity. The PPA can scale up the number of beams. Under the saturation power limit of the photodetectors (PDs), the PPA can realize dozens of mmW beams without introducing extra complexity in the design by scaling up the multiplexing range of the AWG.

The RU can accommodate different performance requirements including coverage, throughput, and latency. The network designer can adapt the RU phased array size to the target coverage of the RU, perform throughput elasticity with multi-beam SDM, and shorten the latency by optimizing the fronthaul link topology.

A.5. RU Scalability

We define the scalability of RUs as the number of users which can be supported by a single RU. A radio access network with mmW beamforming capability can use spatial-temporal-spectral coding (SDM, TDM, FDM) to boost the number of users a single RU can support. As the scalability of the FDM and the TDM scheme have been well studied in previous generations of the cellular network designs [41], we focus on the scalability analysis of the SDM scheme.

The ERON system implements multiple spatial beams through multiple sets of RoF signals. The beam number determines the scalability of the SDM scheme. The number of spatial beams relies on two factors: the saturation power of the PD and the number of available RoF pairs: 90 pairs of RoF signal available at C band with DWDM 50GHz spacing. Figure 7 shows the number of spatial beams as a function of the received RoF signal power. When the received RoF signal power is below -10 dBm, the number of the spatial beams is limited by the available pairs of RoF signals. As the received power increases, the number begins to be confined by the Pd saturation power.

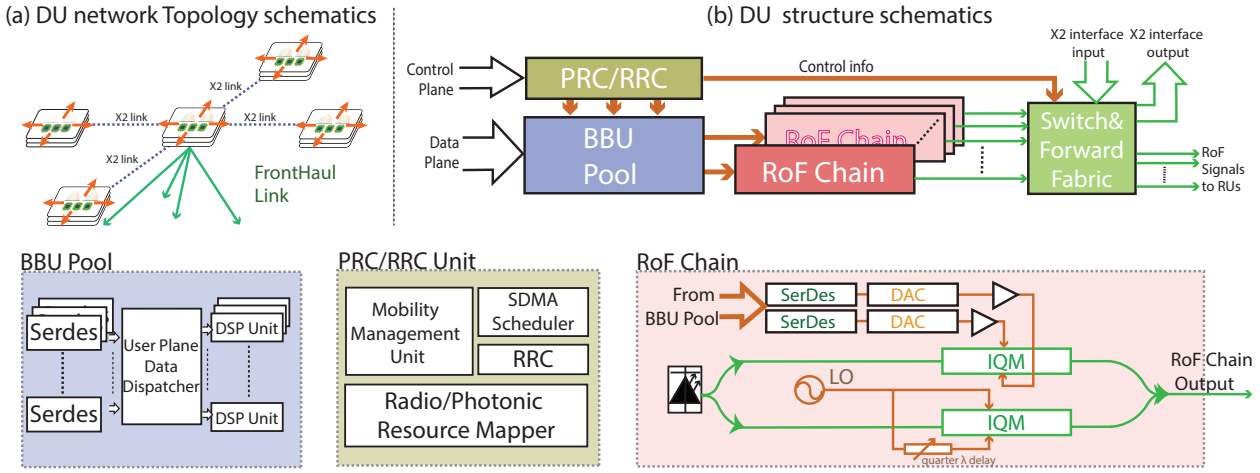


Fig. 8. The DU physical structure. (a) DU network topology; (b) the schematics of the DU structure. PRC: Photonic Resource Control; RRC: Radio Resource Control; BBU: Base-band Unit; SDMA: Spatial-Division Multiplexing Access. DAC: Digital-to-Analog convertor. IQM: IQ Modulator. LO: Local oscillator.

The higher the saturation power, the more spatial beams it can support. With a PD of 25-dBm saturation power, the system can support 90 spatial beams with received RoF signal power up to 5 dBm.

B. Data Unit Design

The DU design evolves the C-RAN architecture. It transmits analog RoF signals to the RUs through the fronthaul link. Such design minimizes the processing delay to reduce the latency on the fronthaul link and alleviate the need for a large number of spectrum resources to transmit the data from the DU to the RU. The DU converts the data plane information to the RoF signals following the control plane instructions and sends them to the RUs. The DU interconnects to neighboring DUs through *X2* interface[42] to conduct handover operations. The DU includes the BBUs, the RoF chains, the photonic resource control(PRC) unit, and a switch&forward fabric. It allows the ERON to centralize all data units which are important in saving hardware resources in the dense mmW 5G deployment scheme. Furthermore, RoF eliminates several parts in the protocol stacks which can minimize the latency in the RAN.

The DU converts the data information from the CU to the RoF signals, sends the analog RoF signals to the RUs, and manages the radio access scheme in both SDMA and TDMA approaches.

The DUs interconnect to their neighboring DUs (see Figure 8 (a)) to handle the mobility management, such as the handover operation, coordinated multiple point (CoMP) operations, etc. Figure 8 (a) shows a mesh-interconnected topology. Each DU has four interconnect ports to its neighboring DUs. The interconnection links utilize the legacy LTE *X2 Interface* to support backward compatibility.

Figure 8 (b) shows the unit structure of the proposed DU. It consists of four parts: a photonic-resource control/radio resource control (PRC/RRC) unit for processing the control plane information, a BBU pool for handling the data plane base-band information, a RoF chain pool for modulating base-band information to optical domain, and a switch & forward fabric for transporting the RoF signals to the RUs.

B.1. BBU Pool

The BBU pool first takes in the data information to a set of SerDes interfaces. The SerDes circuits convert the data plane serial information to multi-bit parallel forms, so the data dispatcher can process the data with large throughput and low latency. The DSP units process the data for channel coding, equalization, FFT, and etc., before sending the data to the RoF chains for modulations.

B.2. RoF Pool

The RoF pool is a collection of RoF chains for optical RoF signal generations. Each RoF chain consists of an electrical part and an optical part. In the electrical part, the SerDes circuits take the data plane information from the BBU pool according, then two DACs convert the data information into the base-band I/Q signals. The base-band signals are the inputs to the optical IQM. In the optical part, the output of a laser source is split to the optical signal modulation and the optical reference modulation. The two are recombined before forwarded to the switch&forward fabric.

B.3. Switch&Forward Fabric

The Switch&Forward fabric consists of optical wavelength switching components. Several kinds of optical devices can perform the function of optical wavelength switching with different principles, such as AWGR[43], LCOS[44], MEMS[45], etc. The inputs of the switch&forward fabric take the output signals from the RoF pool and the input from the *X2* interfaces. The outputs of the unit connect to the RUs and other DUs through *X2* interfaces.

B.4. PRC unit

The PRC unit consists of a mobility management unit, an SDMA scheduler, and a radio/photonic resource mapper. It is responsible for the connection scheduling between the RUs and the UEs, and the wavelength management of the RoF signals. It takes in the control plane information from the CU and calculates the photonic resource allocation in association with the mobility information collected from the UE.

B.5. DU Scalability

We define the scalability of the DU as the number of RUs which can be supported by a single DU. We discuss the DU scalability in two fronthaul connection scenarios: 1) dedicated fiber link between each RU and the DU; 2) shared fiber link between RUs and the DU. Figure 9 shows the illustration of both schemes.

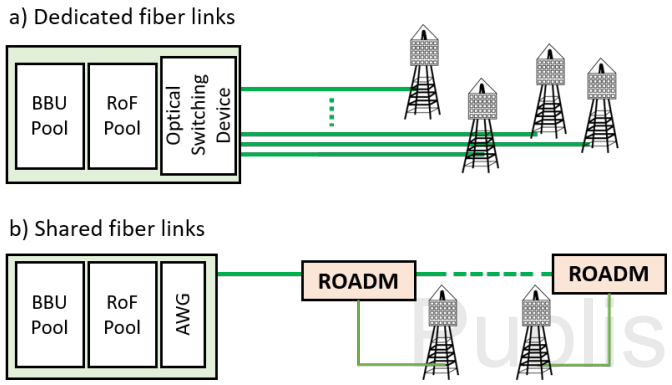


Fig. 9. Illustration of the fiber connection schemes between DU and RUs

The first scalability scenario is for a dedicated fiber link scheme. Each RU has a dedicated fiber link to its DU. The number of the RUs supported by a single DU can scale up if the DU can deploy enough optical switching devices to increase the number of switching ports. The number of switching ports can scale up by using a cascaded design of optical switching devices. A large number of switching ports is an active research area [46, 47].

The second scalability scenario is the shared fiber link scheme. A cluster of RUs connect to the DU through a shared fiber link. The DU sends out the RoF signals for the RUs on the shared fiber link. Each RU collects its RoF signals through a reconfigurable ROADM. The maximum number of RUs supported is limited by the number of RoF signals in the optical band. Assume the mmW carrier frequency is 28 GHz, 90 pairs of RoF signal are available with ROADMs of DWDM 50 GHz spacing at C band. With each RU gets allocated at least one spatial beam, there are 90 RUs supported by a single shared fiber link for the DU.

We also point out that the two scenarios do not exclude each other. In practical deployment scenarios, network designer can combine the design of the dedicated fiber link scheme with the shared fiber link scheme to achieve the optimal trade-off for the scalability and deployment cost.

C. Central Unit Design

The CUs cover a group of DUs, which aggregates all the traffic outbound from the RAN to the core network, and allocate the inbound traffic to corresponding DUs. The CU is responsible for the control and management of the RAN and incorporates the Mobility Management Entity(MME), Serving Gateway(SWG), PGW(PDN Gateway), HSS(Home Subscriber Server), and other units located in the traditional LTE Evolved Packet Core(EPC) unit[42]. The function of the CU is more about software and protocols, and it is similar to the EPC in the LTE architecture. We use the legacy LTE EPC structure to work as the CU in the ERON architecture. This can make the ERON design backward compatible with the LTE networks. The control plane of the LTE protocol stacks can be embedded in the ERON structure, such as the *X2* interface which is responsible for the control signal exchanges among the DUs, the MME and the SGW units in the CUs. Most operations in the LTE system can retrofit to the proposed ERON architecture. This accommodation to the 4G control and management unit helps to leverage the legacy RAN knowledge to achieve better performance and wider applications, at the same time the new structure of the RU and DU proposed in the ERON architecture can promote mmW adaption in the 5G network.

4. PHOTONIC SIGNAL PROCESSING FOR ENERGY EFFICIENCY

A key motivation of the photonic signal processing technology for mmW beamforming used in the ERON RU is its energy efficiency. It is claimed that the photonic method is superior to the RF methods in terms of energy efficiency. However, there are no published numerical results on this topic which is extensively studied in RF research field[48]. A convincing numerical power consumption analysis for the photonic-enabled mmW beamforming is crucial to proceed toward a photonic solution for 5G mmW.

In this part, we conduct benchmark analyses with three power consumption models:

- Fully Digital Beamforming method, as Digital Array (DA)
- Fully-Connected Hybrid Beamforming method, as Fully-Connected Hybrid Array(FCHA)
- Photonic-Processing-enabled Beamforming method, as Photonic Processing Array (PPA)

The PPA method utilizes the proposed PPA described in Figure 5. Figure 10 shows the schematics for the implementation of DA and FCHA methods as described in [39].

The DA model includes the DSP units. We do not consider a DA model with DSP units in the data units (C-RAN option 8[10] due to its impractical bandwidth requirement on the fronthaul link. The DSP units generate the BBU data. The BBU data is fed into the RF chains. The number of RF chains equals the number of the antenna element N. Each RF chain consists of two SerDes units to convert BBU data to I/Q data, two DAC units to generate analog I/Q information, a mixer and a local oscillator to generate mmW signal, and an RF amplifier to boost the signal power to 5 dBm.

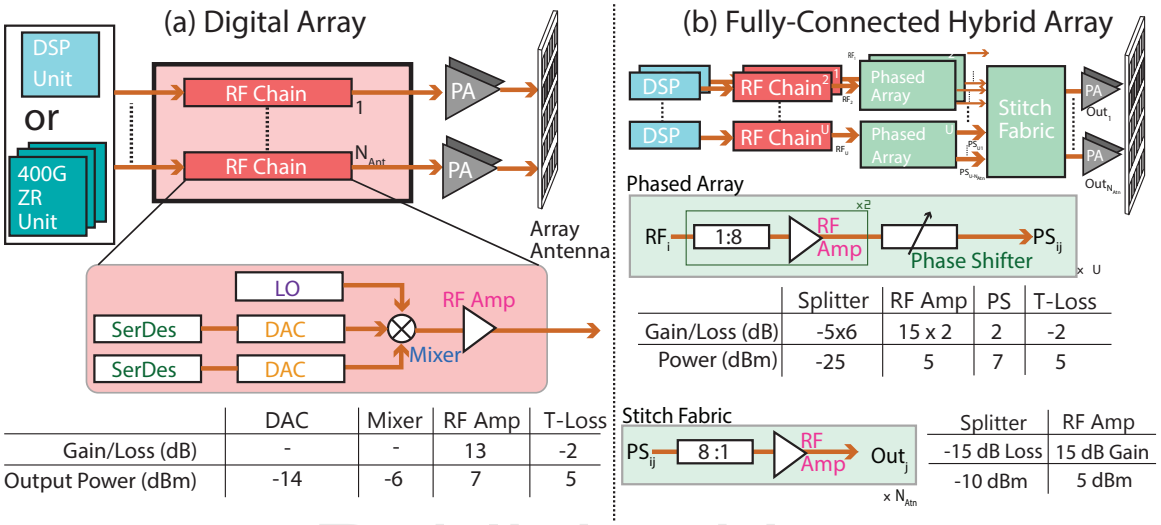
The FCHA model equips with U sets of DSP units, RF chains, and an mmW phased array, and U equals the number of the spatial beams. Then a stitching fabric combines the output of the phased arrays to the power amplifier accordingly. The stitching fabric in mmW consists of a chain of 2:1 Wilkinson combiner[49].

For all three models, a power amplifier amplifies the mmW signal to the required power level before feed the signals to the antenna elements.

A. Power Consumption Model

Table 2 lists the unit power consumption of the mmW components, the photonic components and the number required for the beamformer. Some units are listed with form of merits (FOM). The FOM is a parameter used when the power consumption of the component depends on several parameters of the RAN.

DSP Model: The DSP unit power consumption relates to the FFT operation size, bandwidth, and the number of spatial beams. Assume the FFT size is 4096 points, and each matrix element multiplication requires 6 fixed points operations:



	DAC	Mixer	RF Amp	T-Loss
Gain/Loss (dB)	-	-	13	-2
Output Power (dBm)	-14	-6	7	5

	Splitter	RF Amp	PS	T-Loss
Gain/Loss (dB)	-5x6	15 x 2	2	-2
Power (dBm)	-25	5	7	5

Splitter	RF Amp
-15 dB Loss	15 dB Gain
-10 dBm	5 dBm

Fig. 10. (a) Digital beamforming array implementation. (b) Fully-connected hybrid beamforming array implementation.

$$P_{dsp} = FOM_{DSP} \cdot Bw \cdot (6 \cdot U + 72 \cdot U) \quad (5)$$

SerDes Model: The SerDes circuits' power consumption relates to the number of bits for the DAC, the bandwidth, and the upsampling ratio. Assume the RAN requires 10 bits/symbol to maintain signal quality, the power consumption of a SerDes circuit unit is:

$$P_{SerDes} = FOM_{SerDes} \cdot Bw \cdot 2 \cdot \#bits \quad (6)$$

DAC Model: The DAC power consumption relates to the number of bits of the DAC and the baud rate. Assuming it follows the parameters used in the SerDes circuits calculation, the power consumption of the DAC unit is:

$$P_{DAC} = FOM_{DAC} \cdot BW \cdot 2 \cdot 2^{\#bits} \quad (7)$$

Placement of SOAs: We set the minimal optical output power of the PPA before the detection units to -10 dBm for both signal and reference wavelengths. Based on the minimal optical output power, we calculate the required number and the placement of SOAs in the PPA to compensate for the 1-by-2 MMI power splitting.

We use the III-V-on-silicon SOA proposed in [50] which provides 17.5-dBm on-chip output power, 28-dB small-signal gain, and 540-mW power consumption. The SOA can amplify -10-dBm input power to 15-dBm output. Assume the 1-by-2 MMI loss is 3.2 dB[51], and the AWG loss is 2.7 dB[52]. The SOA can support a 6-stage 1-by-2 MMI power splitting and AWG signal stitching with a 3-dB power margin, equivalent to 64 antenna elements. Hence, we design to place the SOA 6 stages before the phase shifters, and the number of SOAs on the PPA is $2 \cdot N/2^6$, where the factor of 2 counts for the signal and reference wavelengths. We assume the DU can produce signals with power up to 10 dBm.

For PPAs with less than 64 phase shifters, on-chip SOAs are not required, and the PPAs with 64, 256, 1024 phase shifters need 2, 8, and 32 on-chip SOAs, respectively.

Optical Phase Shifter Model: For the optical phase shifters, we assume the beam tuning frequency is less than 1 MHz and the power consumption is per bit per second[53]. The power consumption of the optical phase shifter unit is 0.3 mW.

Trans Impedance Amplifier: The RoF signal power and the TIA conversion gain determine the output mmW power from the PPA. The mmW signal current after photodetection follows:

$$I_{mmW} = R \cdot \sqrt{P_{Ref} \cdot P_{Sig}} cos(\omega_{mmW} \cdot t + \theta_{Sig}) \quad (8)$$

where R is the photodetector responsivity, P_{Ref} and P_{Sig} are the reference wavelength power and signal wavelength power, and θ_{Sig} is the modulated phase information.

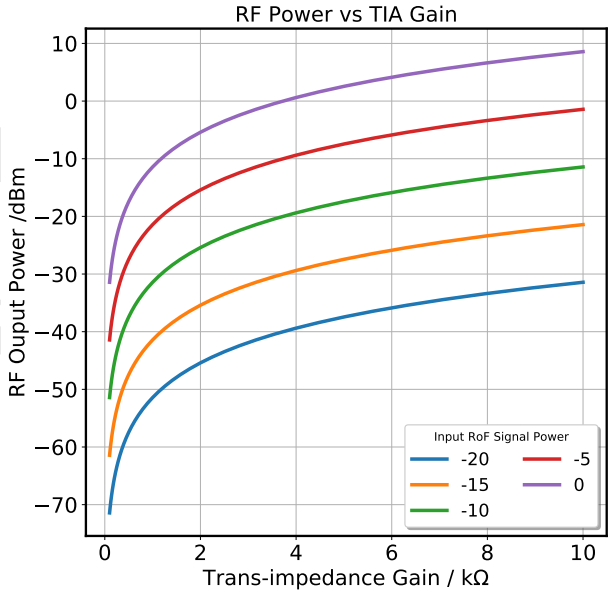


Fig. 11. The mmW output power as a function of the TIA conversion gain for input RoF signal from -20 to 0 dBm.

The I_{mmW} is maximized when P_{Ref} and P_{Sig} are equal. We calculated the converted mmW signal power as a function of the TIA conversion gain for input optical power from -20 dBm to 0 dBm for both P_{Ref} and P_{Sig}, shown in Figure 11. We labelled out several recent results on TIAs[60–65].

Table 2. Number of Component Rerquired for the mmW Beamformer

	Unit Power Consumption	Digital Array	Fully-Connected Hybrid Array	Photonic-aided Array
400G ZR	7 W [17]	$\lceil N \cdot Bw \cdot bits \cdot 2/400 \rceil$	-	-
DSP	13^{-9} mW/bit/s [54]	N	U	-
SerDes	10 mW/Gb/s [55]	$N \cdot 2$	$U \cdot 2$	-
DAC	0.08 pJ/conversion [56]	$N \cdot 2$	$U \cdot 2$	-
Mixer	10 mW [57]	N	U	-
LO	60 mW [57]	N	U	-
RF Amplifier	40 mW [58]	$M + \lceil \log_8 U \rceil \cdot N$	$U + \lceil \log_8 N \rceil \cdot N \cdot U + \lceil \log_8 U \rceil \cdot N$	N
Phase Shifter	10 mW [59]	-	$N \cdot U$	-
Optical PS	30 pJ/bit/s [53]	-	-	$N \cdot U$
PD+TIA	77 mW [60]	-	-	N
SOA	540 mW [50]	-	-	$N \cdot 2^{-6}$ if $N > 64$

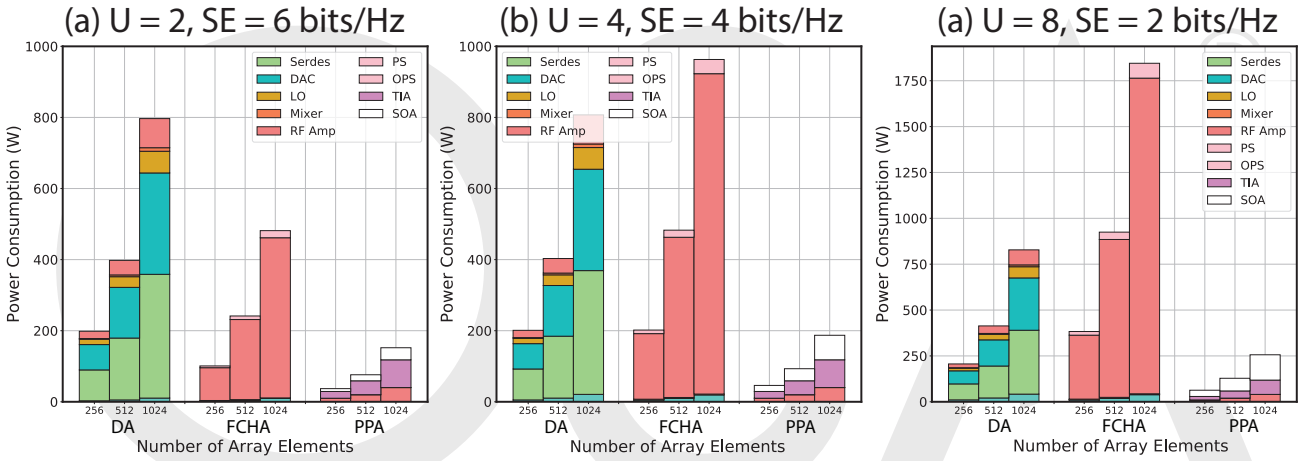


Fig. 12. Power consumption as a function of the number of antenna array elements. The three bars in each bar cluster represent(from left to right): DA model, FCHA model, and PPA model.

Figure 11 shows that the mmW output power is way below -10 dBm when the TIA convention gain is below $1k\Omega$. The $10 - k\Omega$ TIA conversion gain gives input optical power of -10-dBm with an over -10-dBm mmW output power. With the help of the RF amplifier after the TIA, the PPA can produce 5-dBm mmW signal power for the input of the power amplifier.

Power Amplifier: The power consumption of a power amplifier can be characterized using the PA efficiency number. Assume the PA is based on a Doherty PA architecture and the power-added efficiency is 0.3 [66] based on the SiGe CMOS, the saturated output power can reach over 20 dBm and the PA efficiency is 0.185 [67]. The power consumption of the power amplifiers is:

$$P_{PA} = P_{out}/\eta_{PA} \tag{9}$$

Where P_{out} is the total output power, η_{PA} is the PA efficiency.

B. Power Consumption Modelling Results

Figure 12 shows the power consumption as a function of the number of the antenna array elements with 2, 4, and 8 spatial beams to achieve over 10-Gbps throughput per RU. Each bar cluster consists of three bars representing the power consumption to implement a phased array with 256, 512, 1024 elements. The three-bar clusters, from left to right, are DA-DSP method, FCHA method, and PPA method, respectively. Note we exclude the power consumption of the power amplifier as it consumes the same amount of power in our models. With the model described in Section 4.A, the power consumption is 101 watts, 59.3 watts, and 29.9 watts for the 2, 4, and 8 spatial beams, respectively.

The PPA method shows energy efficiency advantages across all scenarios. We observe that the composition of the three methods is extremely different: the DA method is dominated by SerDes and DAC circuits; the FCHA method consumes most of the power on RF amplifier; the power consumption of the PPA method consists mainly of SOA and TIA.

The power consumption of the DA method is dominated by the SerDes circuits and the DACs. Its power consumption depends more on the number of the array elements, the resolution of the DACs, and the bandwidth. The impact of the number of beams is less significant as it requires only additional DSP

modules. Because the SerDes circuits and the DACs are the most power-consuming components (170 mW and 140 mW respectively in this case), the DA method is inevitably prone to high power consumption for very large array implementations.

The power consumption of the FCHA method is dominated by the RF amplifiers, and the RF amplifier power consumption is related to two factors: the number of the array elements and the number of the spatial beams. The increase of the large array size and the number of spatial beams is inevitable in mmW 5G RAN. Hence, the FCHA method is the most susceptible to power consumption problems. From Fig.6, we observe that for larger array elements such as 512 elements and 1024 elements and an 8-beam scenario, the FCHA method needs 1,000 W to compensate for the extra loss introduced by the Wilkinson splitters/combiners.

The power consumption of the PPA method consists of three parts: the TIAs, the RF amplifiers, and the SOAs. Although the number of the TIAs and the RF amplifiers increases as the number of the array elements, the total power consumption of the PPA method uses much less power than the DA method and the FCHA method. Thanks to the low-loss character of the photonic circuits, the required number of the SOAs is small even if each SOA consumes more than 500-mW power per unit.

We confirm the validity of the power consumption modeling method using commercial 5G-NR antenna unit data as a comparison. We found two commercial 5G-NR antenna units from the FCC filing: Samsung SFG-AA1100 and Samsung SFG-D1100[68]. They are both implemented using analog beamforming method with single data beam generation, which is FCHA with $U = 1$. Both antennas operate on 27.5-28.35 GHz, with 850 MHz spectrum window. The power consumption for SFG-AA1100 is 67.5 Watt for a 256-element antenna, and the one for SFG-D1100 is 12.5 Watt for a 32-element antenna. Our model shows the power consumption for the two is 45.45 Watt and 7.4 Watt, respectively. Considering we adopt the state-of-art results to calculate the power consumption, our modeling method is realistic and on the low-end of the actual mmW circuit power consumption.

The proposed PPA method achieves the best energy efficiency performance. The energy efficiency gap between the PPA method and the other two increases as the number of the array elements increases. Excluding the power amplifier power consumption, which is equal to all implementations, the PPA method consumes less than half of the power the other two methods require.

Another alternative method using RF implementation of multi spatial beam generation is called multi-chip method. The RU uses multiple chips, and each chip can only generate one beam with a dedicate array antenna, which is the FCHA method with $U = 1$. The use of multiple antennas to generate multiple beams eliminates the need for an RF stitch fabric. This method abandons the advantage of array antenna multi-beam generation, using more antenna components in exchange for better energy efficiency. By eliminating the stitching fabric which is a major source of power loss, the multi-chip FCHA method consumes less power than the single-chip FCHA method.

Finally, we compare the energy efficiency performance of the multi-chip FCHA method and the PPA method. Figure 13 shows that the PPA method consumes significantly lower power. We see that the more spatial beams used in the system, the higher the energy efficiency the PPA has in comparison to the multi-chip FCHA method. In the 64-element array case, the multi-chip method consumes 27 Watt of power for 4 spatial beams, while the PPA method consumes less than 25 W and can generate four times than the number of beams. The energy efficiency gap is bigger in the 256-element array case. Our proposed method can maintain less than 100 W power with the support of 16-spatial-beam generation. On the other hand, the multi-chip FCHA method consumes 200 W with a 4-spatial-beam generation, and the power consumption increases linearly as we continue to increase the number of spatial beams generated.

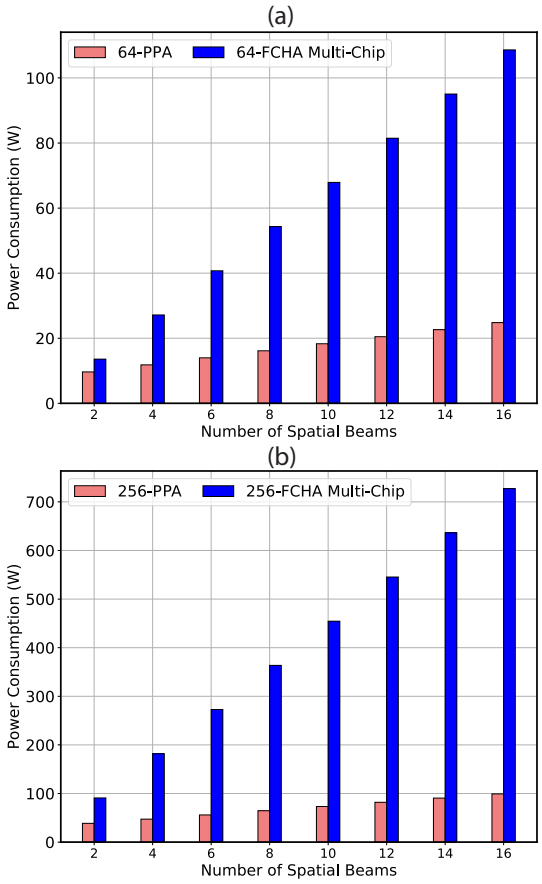


Fig. 13. The power consumption for PPA method and multi-chip method. (a) 64 element array. (b) 256 element array.

5. CENTRALIZED CONVERGED RESOURCE ORCHESTRATION FOR RESOURCE POOLING

We evaluate the network resource pooling performance of the ERON architecture with UE mobility-aware RF-optical converged simulations. The required network resource is set to meet the UE bandwidth requirement generated from the network traffic simulations.

We define the resource pooling gain as the ratio between the required network resource of the ERON system and the conventional RAN architecture which plans each RU with affluent resources to satisfy peak demands:

$$G_{pooling} = \frac{Resource_{RAN}}{Resource_{ERON}} \tag{10}$$

The ERON system allocates the network resources(the RoF signals, the BBU units, etc) according to the Algorithm 1. The RSMA algorithm in the ERON system consists of three parts: 1)

update user location information; 2) update the user traffic requirement information; 3) update RSMA of the user and update the RU resource usage database.

Algorithm 1. Resource allocation Algorithm in the mobility-aware network simulation

```
for Each user in all users do
    if v_i ≠ 0 then                                  ▷ Update user location
        user (i,x,y) ← user (i,x,y) + v_i
    if Req_i ≠ 0 then                                ▷ Update user BW request
        Bw_i ← Req_i
    else
        Release RSMA_i
    RU_i ← user (i,x,y)                              ▷ Find the corresponding RU
    if RSMA_i = None then                            ▷ Update RU RSMA
        RSMA_i ← new RSMA assignment from RU_i
    else
        if RSMA_i conflict in RU_i then
            RSMA_i ← new RSMA assignment from RU_i
```

The v_i is the mobility speed of the $user_i$, $user(i,x,y)$ is the $user_i$ location, Req_i is the traffic request remaining time of the $user_i$, Bw_i is the bandwidth requirement of the $user_i$ at the current time slot. $RSMA_i$ represents the assigned network resources for the mmW connection.

In this network simulation study (as also in [20, 69]) we assumed the simulated RUs with homogeneous coverage both for ERON and the conventional RAN architecture.

The UE mobility-aware RAN simulations here used the same parameters for Figure 3. The simulations run for a 27.7-hour (10^5 seconds) simulation time. The entire simulated field is 5km by 5km with each RU coverage of a 250-meter by 250-meter area. We assume the RU can adjust the transmitting power to perform the coverage and can coordinate with neighboring RUs to avoid signal interference using spatial-spectral coding. For the parameters of the mobile users, the population of the simulated mobile users is 4,000 with each RU of 10 initial users. The user follows the RWP mobility model described in Section 2. The mobility velocity follows a uniform distribution from $1m/s$ to $15m/s$. The mobile users generate traffic request independent of the mobility situation. The traffic request arrival rate follows the Poisson distribution with the mean arrival rate varying according to the traffic load parameter. The traffic load is set to vary from very light (5 Erlang) to extremely heavy (640 Erlang) to study the pooling gain performance under different network traffic scenarios. The service duration follows exponential distribution with a mean of 10 second. The simulation time granularity is set to 1 second. The traffic request, request time duration, user location, and user mobility information are updated at each time interval.

We calculated the 95% confidence interval for the simulation results. The simulation emulates over 10^5 time-slot samples to obtain enough data points. Although the combined impact of the user mobility and the traffic loads is complicated and the modeling of the pooling gain is still a state-of-art topic research topic, we use a normal distribution to model the confidence interval. The lower and upper limit are calculated by:

$$p \pm Z_{0.95} \cdot \sqrt{p(1-p)/N} \tag{11}$$

where p is the evaluated result from simulation, $Z_{0.95}$ is the number of standard deviations extending from the mean of a normal distribution required to contain 0.95 of the area, and N is the number of samples.

The shaded areas in Figure 14 show the upper and lower limit of the 95% confidence interval. The results show the simulations are reliable.

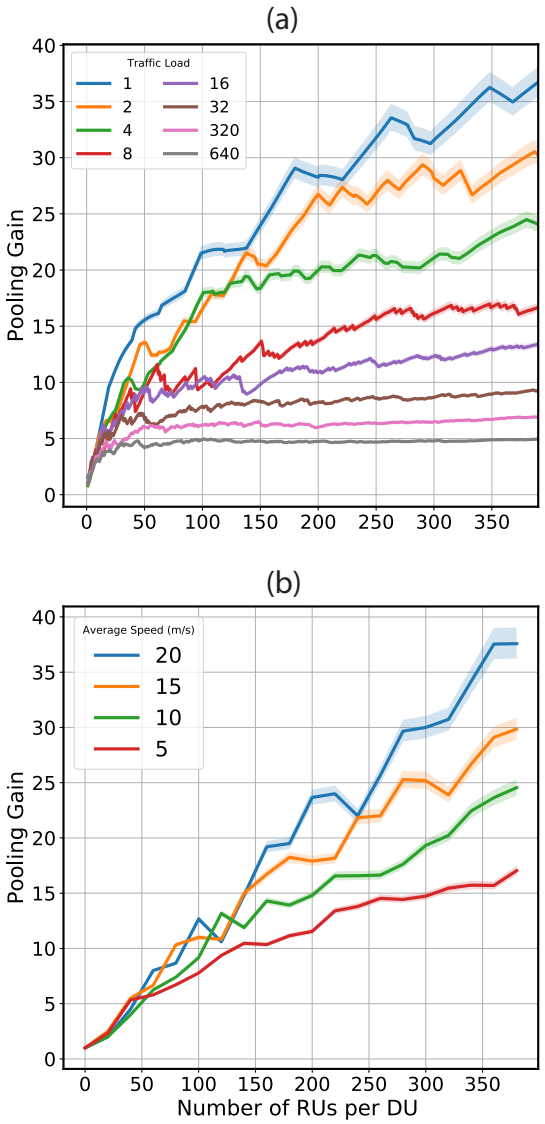


Fig. 14. Resource pool gain results in user mobility-aware simulation as a function of the number of RUs per DU under (a) different traffic loads (b) different mobility value.

Figure 14 (a) and (b) show the resource pooling gain results as a function of different number of RUs per DU.

Figure 14 (a) shows the pooling gain as a function of the DU coverage size. We observed large resource pooling gain along with different pooling sizes and across different traffic loads. The pooling size effect is consistent with the analysis mentioned above, as the pooling can mitigate the RU throughput volatility. The effect of traffic load over ERON resource pooling gain shows a meaningful guideline on the design of the DU coverage. Under a lighter traffic load, the resource pooling gain continues to increase as the DU coverage of the RUs enlarges. As the traffic

load increases, we observe a tap-out effect of enlarging DU coverage, which is expected as the resource pooling gain stems from the statistical multiplexing. Even with the tap-out effect under extremely heavy traffic load, the resource pooling gain value can be as high as 5 times. The traffic load over mobility simulation provides a tool to determine the DU coverage for the ERON system.

Figure 14 (b) shows the resource pooling gain under different UE mobility scenarios. The mobility value is defined as the average moving speed of the UEs. We observed that as the mobility value increases, the resource pooling gain increases correspondingly. UEs with high mobility change their locations more frequently and require more hand-over operations and inter-RU coordination. The ERON centralized scheme optimizes the scenario by orchestrating network resources across multiple RUs. In addition, ERON architecture can achieve seamless hand-over operation by utilizing the optical switching technology. It does not require extra networking resources in comparison to traditional radio access and transport networks.

6. CONCLUSION

We proposed the ERON architecture as a solution to the mmW 5G RAN. The ERON architecture takes advantage of the photonic technologies to facilitate the roll-out of mmW mobile communications. We show numerical results on the better energy efficiency with PPAs for mmW beam forming comparing with traditional RF implementations. The proposed scheme utilizes optical wavelength orthogonality to offer elastic and flexible throughput changes. The centralization of all control/management units and data units retains all the benefits from the proposal of C-RAN while providing higher pooling gain to handle traffic demand volatility due to the UEs' mobility.

From photonic signal processing for energy-efficient mmW beam generation to smart photonic resource planning and provisioning for the mobile SDMA scheme, the photonic research community can provide key technologies for the mmW 5G deployment that go beyond the simple use of optical transmission technologies for serving big data pipes in the front-haul communication links. The ERON architecture is a promising solution to the mmW 5G RAN, and further photonic research innovations can contribute to more innovations for the mobile communication scenario.

FUNDING

This work was supported in part by National Science Foundation (NSF) Grant 1619173 and NSF Grant 1302719.

REFERENCES

1. M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, I. Chih-Lin *et al.*, "Millimeter wave communications for future mobile networks," IEEE J. on Sel. Areas Commun. **35**, 1909–1935 (2017).
2. X. Li, J. Yu, and G.-K. Chang, "Photonics-assisted technologies for extreme broadband 5g wireless communications," J. Light. Technol. (2019).
3. 3GPP, "Ts 38.101-2: Nr; user equipment (ue) radio transmission and reception; part 2: Range 2 standalone," Release 15 (2017).
4. D. Marpaung, J. Yao, and J. Capmany, "Integrated microwave photonics," Nat. photonics **13**, 80 (2019).
5. C. Porzi, G. Serafino, M. Sans, F. Falconi, V. Sorianello, S. Pinna, J. E. Mitchell, M. Romagnoli, A. Bogoni, and P. Ghelfi, "Photonic integrated microwave phase shifter up to the mm-wave band with fast response time in silicon-on-insulator technology," J. Light. Technol. **36**, 4494–4500 (2018).
6. Y. Liu, A. R. Wichman, B. Isaac, J. Kalkavage, E. J. Adles, T. R. Clark, and J. Klamkin, "Ultra-low-loss silicon nitride optical beamforming network for wideband wireless applications," IEEE J. Sel. Top. Quantum Electron. **24**, 1–10 (2018).
7. F. Musumeci, C. Bellanzon, N. Carapellese, M. Tornatore, A. Pattavina, and S. Gosselin, "Optimal bbu placement for 5g c-ran deployment over wdm aggregation networks," J. Light. Technol. **34**, 1963–1970 (2015).
8. K. Zhang, Q. Zhuge, H. Xin, H. He, W. Hu, and D. V. Plant, "Low-cost wdm fronthaul enabled by partitioned asymmetric awgr with simultaneous flexible transceiver assignment and chirp management," IEEE/OSA J. Opt. Commun. Netw. **9**, 876–888 (2017).
9. J. Zhang, Y. Xiao, D. Song, L. Bai, and Y. Ji, "Joint wavelength, antenna, and radio resource block allocation for massive mimo enabled beamforming in a twdm-pon based fronthaul," J. Light. Technol. **37**, 1396–1407 (2019).
10. A. Pizzinat, P. Chanclou, F. Saliou, and T. Diallo, "Things you should know about fronthaul," J. Light. Technol. **33**, 1077–1083 (2015).
11. R. Proietti, H. Lu, G. Liu, A. Castro, M. Shamsabardeh, and S. J. B. Yoo, "Experimental demonstration of elastic rf-optical networking (eron) for 5g mm-wave systems," in *2017 European Conference on Optical Communication (ECOC),* (IEEE, 2017), pp. 1–3.
12. H. Lu, G. Liu, R. Proietti, V. Squitieri, K. Zhang, A. Castro, Q. J. Gu, Z. Ding, and S. J. B. Yoo, "mmwave beamforming using photonic signal processing for future 5g mobile systems," in *Optical Fiber Communication Conference,* (Optical Society of America, 2018), pp. M4J–3.
13. H. Lu, Y. Zhang, Y.-C. Ling, G. Liu, R. Proietti, and S. J. B. Yoo, "Experimental demonstration of mmwave multi-beam forming by sin photonic integrated circuits for elastic rf-optical networking," in *Optical Fiber Communication Conference,* (Optical Society of America, 2019), pp. Th1F–3.
14. P. Sagazio, S. Callender, W. Shin, O. Orhan, S. Pellerano, and C. Hull, "Architecture and circuit choices for 5g millimeter-wave beamforming transceivers," IEEE Commun. Mag. **56**, 186–192 (2018).
15. C. Chang, N. Nikaein, R. Knopp, T. Spyropoulos, and S. S. Kumar, "Flexcran: A flexible functional split framework over ethernet fronthaul in cloud-ran," in *2017 IEEE International Conference on Communications (ICC),* (2017), pp. 1–7.
16. E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," IEEE Commun. Mag. **52**, 186–195 (2014).
17. J. Cheng, C. Cole, J. Kahn, E. Maniloff, D. P. an d B. Thomsen, I. Tomkos, W. Way, and X. Zhou, "Workshop: Will coherent optics become a reality for intra-data center applications?" in *Optical Networking and Communication Conference (OFC),* (2019).
18. L. Velasco, A. Castro, A. Asensio, M. Ruiz, G. Liu, C. Qin, R. Proietti, and S. J. B. Yoo, "Meeting the requirements to deploy cloud ran over optical networks," IEEE/OSA J. Opt. Commun. Netw. **9**, B22–B32 (2017).
19. S. Das and M. Ruffini, "A variable rate fronthaul scheme for cloud radio access networks," J. Light. Technol. **37**, 3153–3165 (2019).
20. M. R. Raza, M. Fiorani, A. Rostami, P. Öhlen, L. Wosinska, and P. Monti, "Demonstration of dynamic resource sharing benefits in an optical c-ran," IEEE/OSA J. Opt. Commun. Netw. **8**, 621–632 (2016).
21. W. Hu, L. Yi, H. He, X. Yang, Z. Li, M. Bi, K. Zhang, H. Xin, Y. Liu, and W. Du, "Soft-stacked pon for soft c-ran," IEEE/OSA J. Opt. Commun. Netw. **8**, B12–B20 (2016).
22. H. T. Friis, "A note on a simple transmission formula," proc. IRE **34**, 254–256 (1946).
23. A. Ghosh, "The 5g mmwave radio revolution," Microw. J. **59** (2016).
24. W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5g cellular communications: Theoretical feasibility and prototype results," IEEE communications magazine **52**, 106–113 (2014).
25. J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication* (Springer Science & Business Media, 2012).
26. T. S. Rappaport, R. W. Heath Jr, R. C. Daniels, and J. N. Murdock,

Millimeter wave wireless communications (Pearson Education, 2015).
27. C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks," IEEE Transactions on mobile computing **2**, 257–269 (2003).
28. C. Ranaweera, E. Wong, A. Nirmalathas, C. Jayasundara, and C. Lim, "5g c-ran with optical fronthaul: An analysis from a deployment perspective," J. Light. Technol. **36**, 2059–2068 (2018).
29. B. Skubic, M. Fiorani, S. Tombaz, A. Furuskär, J. Mårtensson, and P. Monti, "Optical transport solutions for 5g fixed wireless access [invited]," IEEE/OSA J. Opt. Commun. Netw. **9**, D10–D18 (2017).
30. S. Noor, P. Assimakopoulos, and N. J. Gomes, "A flexible subcarrier multiplexing system with analog transport and digital processing for 5g (and beyond) fronthaul," J. Light. Technol. **37**, 3689–3700 (2019).
31. N. Argyris, G. Giannoulis, K. Kanta, N. Iliadis, C. Vagionas, S. Papaioannou, G. Kalfas, D. Apostolopoulos, C. Caillaud, H. Debrégeas, N. Pleros, and H. Avramopoulos, "A 5g mmwave fiber-wireless ifof analog mobile fronthaul link with up to 24-gb/s multiband wireless capacity," J. Light. Technol. **37**, 2883–2891 (2019).
32. R. Zhang, F. Lu, M. Xu, S. Liu, P. Peng, S. Shen, J. He, H. J. Cho, Q. Zhou, S. Yao, and G. Chang, "An ultra-reliable mmw/fso a-rof system based on coordinated mapping and combining technique for 5g and beyond mobile fronthaul," J. Light. Technol. **36**, 4952–4959 (2018).
33. J. Brenes, T. D. Lagkas, D. Klonidis, R. Munoz, S. Rommel, G. Landi, I. Tafur Monroy, E. Grivas, E. Pikasis, G. Bernini, J. M. Fabrega, and R. Vilalta, "Network slicing architecture for sdm and analog-radio-over-fiber-based 5g fronthaul networks," IEEE/OSA J. Opt. Commun. Netw. **12**, B33–B43 (2020).
34. O. Gerstel, M. Jinno, A. Lord, and S. J. B. Yoo, "Elastic optical networking: a new dawn for the optical layer?" IEEE Commun. Mag. **50**, s12–s20 (2012).
35. T. Pfeiffer, "Next generation mobile fronthaul and midhaul architectures," J. Opt. Commun. Netw. **7**, B38–B45 (2015).
36. X. Liu and F. Effenberger, "Emerging optical access network technologies for 5g wireless," J. Opt. Commun. Netw. **8**, B70–B79 (2016).
37. P. Deng and P. Fan, "An aoa assisted toa positioning system," in *WCC 2000-ICCT 2000. 2000 International Conference on Communication Technology Proceedings (Cat. No. 00EX420),* vol. 2 (IEEE, 2000), pp. 1501–1504.
38. D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," in *IEEE International Conference on Communications, 2003. ICC'03.*, vol. 4 (IEEE, 2003), pp. 2618–2622.
39. S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "Mimo for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?" IEEE Commun. Mag. **52**, 110–121 (2014).
40. ITU-T, "Spectral grids for wdm applications: Dwdm frequency grid," G.694.1, Ser. G: Transm. Syst. Media, Digit. Syst. Networks (2018).
41. E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for mobile broadband* (Academic press, 2013).
42. C. Cox, *An introduction to LTE: LTE, LTE-advanced, SAE and 4G mobile communications* (John Wiley & Sons, 2012).
43. C.-T. Lea, "A scalable awgr-based optical switch," J. lightwave technology **33**, 4612–4621 (2015).
44. G. Lazarev, A. Hermerschmidt, S. Krüger, and S. Osten, "Lcos spatial light modulators: trends and applications," Opt. Imaging Metrol. Adv. Technol. pp. 1–29 (2012).
45. T.-W. Yeow, K. E. Law, and A. Goldenberg, "Mems optical switches," IEEE Commun. magazine **39**, 158–163 (2001).
46. M. Iwama, M. Takahashi, M. Kimura, Y. Uchida, J. Hasegawa, R. Kawahara, and N. Kagi, "Lcos-based flexible grid 1×40 wavelength selective switch using planar lightwave circuit as spot size converter," in *Optical Fiber Communication Conference,* (Optical Society of America, 2015), p. Tu3A.8.
47. R. Matsumoto, T. Kodama, K. Morita, N. Wada, and K. ichi Kitayama, "Scalable two- and three-dimensional optical labels generated by 128-port encoder/decoder for optical packet switching," Opt. Express **23**, 25747–25761 (2015).
48. H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," arXiv preprint arXiv:1807.07201 (2018).
49. K. Kibaroglu, M. Sayginer, T. Phelps, and G. M. Rebeiz, "A 64-element 28-ghz phased-array transceiver with 52-dbm eirp and 8–12-gb/s 5g link at 300 meters without any calibration," IEEE Transactions on Microw. Theory Tech. **66**, 5796–5811 (2018).
50. K. V. Gasse, R. Wang, and G. Roelkens, "27 db gain iii/v-on-silicon semiconductor optical amplifier with 17 dbm output power," Opt. Express **27**, 293–302 (2019).
51. K. Shang, S. Pathak, G. Liu, S. Feng, S. Li, W. Lai, and S. J. B. Yoo, "Silicon nitride tri-layer vertical y-junction and 3d couplers with arbitrary splitting ratio for photonic integrated circuits," Opt. express **25**, 10474–10483 (2017).
52. K. Shang, S. Pathak, C. Qin, and S. J. B. Yoo, "Low-loss compact silicon nitride arrayed waveguide gratings for photonic integrated circuits," IEEE Photonics J. **9**, 1–5 (2017).
53. K. Kajikawa, T. Tabei, and H. Sunami, "An infrared silicon optical modulator of metal–oxide–semiconductor capacitor based on accumulation-carrier absorption," Jpn. J. Appl. Phys. **48**, 04C107 (2009).
54. F.-L. Yuan and D. Marković, "A 13.1 gops/mw 16-core processor for software-defined radios in 40nm cmos," in *2014 Symposium on VLSI Circuits Digest of Technical Papers,* (IEEE, 2014), pp. 1–2.
55. D. Cui, H. Zhang, N. Huang, A. Nazemi, B. Catli, H. G. Rhew, B. Zhang, A. Momtaz, and J. Cao, "3.2 a 320mw 32gb/s 8b adc-based pam-4 analog front-end with programmable gain control and analog peaking in 28nm cmos," in *2016 IEEE International Solid-State Circuits Conference (ISSCC),* (2016), pp. 58–59.
56. A. Nazemi, K. Hu, B. Catli, D. Cui, U. Singh, T. He, Z. Huang, B. Zhang, A. Momtaz, and J. Cao, "3.4 a 36gb/s pam4 transmitter using an 8b 18gs/s dac in 28nm cmos," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers,* (2015), pp. 1–3.
57. R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober, "Linear massive mimo precoders in the presence of phase noise—a large-scale analysis," IEEE Transactions on Veh. Technol. **65**, 3057–3071 (2016).
58. S. Zihir, O. D. Gurbuz, A. Karroy, S. Raman, and G. M. Rebeiz, "A 60 ghz single-chip 256-element wafer-scale phased array with eirp of 45 dbm using sub-reticle stitching," in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* (2015), pp. 23–26.
59. S. Mondal, R. Singh, A. I. Hussein, and J. Paramesh, "A 25-30 ghz 8-antenna 2-stream hybrid beamforming receiver for mimo communication," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* (2017), pp. 112–115.
60. Z. Xuan, R. Ding, Y. Liu, T. Baehr-Jones, M. Hochberg, and F. Aflatouni, "A low-power hybrid-integrated 40-gb/s optical receiver in silicon," IEEE Transactions on Microw. Theory Tech. **66**, 589–595 (2018).
61. S. G. Kim, C. Hong, Y. S. Eo, J. Kim, and S. M. Park, "A 40-ghz mirrored-cascode differential transimpedance amplifier in 65-nm cmos," IEEE J. Solid-State Circuits (2018).
62. G. Kalogerakis, T. Moran, T. Nguyen, and G. Denoyer, "A quad 25gb/s 270mw tia in 0.13 μm bicmos with< 0.15 db crosstalk penalty," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers,* (IEEE, 2013), pp. 116–117.
63. T. Takemoto, H. Yamashita, T. Yazaki, N. Chujo, Y. Lee, and Y. Matsuoka, "A 4$\times$ 25-to-28gb/s 4.9 mw/gb/s- 9.7 dbm high-sensitivity optical receiver based on 65nm cmos for board-to-board interconnects," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers,* (IEEE, 2013), pp. 118–119.
64. C. Knochenhauer, S. Hauptmann, J. C. Scheytt, and F. Ellinger, "A jitter-optimized differential 40-gbit/s transimpedance amplifier in sige bicmos," IEEE Transactions on Microw. Theory Tech. **58**, 2538–2548 (2010).
65. J. Kim and J. F. Buckwalter, "A 40-gb/s optical transceiver front-end in 45 nm soi cmos," IEEE J. Solid-State Circuits **47**, 615–626 (2012).
66. A. Sarkar, F. Aryanfar, and B. A. Floyd, "A 28-ghz sige bicmos pa with 32% efficiency and 23-dbm output power," IEEE J. Solid-State Circuits **52**, 1680–1686 (2017).

67. A. Grebennikov, *RF and microwave power amplifier design* (McGraw-Hill New-York, 2005).
68. Samsung, *Samsung AA100AC Installation-Manual* (FCC Filing documents, 2018).
69. R. I. Tinini, D. M. Batista, G. B. Figueiredo, M. Tornatore, and B. Mukherjee, "Low-latency and energy-efficient bbu placement and vpon formation in virtualized cloud-fog ran," IEEE/OSA J. Opt. Commun. Netw. **11**, B37–B48 (2019).

AUTHOR BIOGRAPHIES

Hongbo Lu received the B.S. degree in communication engineering from Fudan University, Shanghai, China, in 2011, and the M.S. degree in electrical engineering from Tokyo University, Tokyo, Japan, in 2014. He is currently working toward the Ph.D. degree in electrical and computer engineering, University of California, Davis, CA, USA. His research interest focuses on coherent optical communication, radio access networks, and optical wireless converged networks.

Roberto Proietti received the M.S. degree in telecommunications engineering from the University of Pisa, Pisa, Italy, in 2004, and the Ph.D. degree in electrical engineering from Scuola Superiore Sant Anna, Pisa, Italy, in 2009. He is a Project Scientist with the Next Generation Networking Systems Laboratory, University of California, Davis, CA, USA. His research interests include optical switching technologies and architectures for supercomputing and data center applications, high-spectral-efficiency coherent transmission systems, and elastic optical networking.

Gengchen Liu received the B.S. degree in electrical engineering from Central Michigan University, Mount Pleasant, MI, USA, in 2015. He is currently working toward the Ph.D. degree in electrical and computer engineering, University of California, Davis, CA, USA. His research interest focuses on high-speed optical transceiver design, digital signal processing, and machine learning.

Xiaoliang Chen received the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2016. He is currently a Research Scholar with the University of California, Davis, CA, USA. He has published more than 40 papers on the journals and conferences of IEEE and OSA. His research interests include optical networks, network resilience, software-defined networking, and machine learning. He is an Associate Editor for Springer Telecommunication Systems Journal and a TPC member of the IEEE ICNC 2017 and ICNC 2018.

S.J.Ben Yoo (F'07) received the B.S., M.S., and Ph.D. degrees from Stanford University, Stanford, CA, USA, in 1984, 1986, and 1991, respectively.

He is a Distinguished Professor with the University of California, Davis (UC Davis), Davis, CA, USA. Prior to joining UC Davis in 1999, he was a Senior Research Scientist with Bellcore, leading technical efforts in integrated photonics, optical networking, and systems integration. His research at UC Davis includes 2D/3D photonic integration for future computing, cognitive networks, communication, imaging, and navigation systems, micro/nano systems integration, and the future internet. His research activities at Bellcore included the next-generation internet, reconfigurable multiwavelength optical networks (MONET), wavelength interchanging cross connects, wavelength converters, vertical-cavity lasers, and high-speed modulators. He led the MONET testbed experimentation efforts, and participated in ATD/MONET systems integration and a number of standardization activities. Prior to joining Bellcore in 1991, he conducted research on nonlinear optical processes in quantum wells, a fourwave-mixing study of relaxation mechanisms in dye molecules, and ultrafast diffusion-driven photodetectors at Stanford University.

Prof. Yoo is a Fellow of OSA and NIAC, and a recipient of the DARPA Award for Sustained Excellence (1997), the Bellcore CEO Award (1998), the Mid-Career Research Faculty Award (2004 UC Davis), and the Senior Research Faculty Award (2011 UC Davis).