# Programmable Millimeter-Wave MIMO Radios with Real-Time Baseband Processing

Zhenzhou Qi, Zhihui Gao, Chung-Hsuan Tung, Tingjun Chen
Department of Electrical and Computer Engineering, Duke University
{zhenzhou.qi,zhihui.gao,chunghsuan.tung,tingjun.chen}@duke.edu

## ABSTRACT

Baseband processing is one of the most time-consuming and computationally expensive tasks in radio access networks (RANs), which is typically realized in dedicated hardware. The concept of virtualizing the RAN functions by moving their computation to edge data centers can significantly reduce the deployment cost and enable more flexible use of the network resources. Recent studies have focused on software-based baseband processing for large-scale sub-6 GHz MIMO systems, while 5G also embraces the millimeter-wave (mmWave) frequency bands to achieve further improved data rates leveraging the widely available spectrum. Therefore, it is important to build a platform for the experimental investigation of software-based baseband processing for mmWave MIMO systems. In this paper, we implement programmable mmWave MIMO radios equipped with real-time baseband processing capability, leveraging the open-access PAWR COSMOS testbed. We first develop Agora-UHD, which enables UHD-based software-defined radios (SDRs) to interface with Agora, an open-source software realization of real-time massive MIMO baseband processing. Next, we integrate Agora-UHD with the USRP SDRs and IBM 28 GHz phased array antenna module (PAAM) subsystem boards deployed in the PAWR COSMOS testbed. We demonstrate a 2×2 28 GHz polarization MIMO link with a bandwidth of 122.88 MHz, and show that it can meet the real-time processing deadline of 0.375 ms (3 transmission time intervals for numerology 3 in 5G NR FR2) using only 8 CPU cores. The source code of Agora-UHD and its integration with the programmable 28 GHz radios in the COSMOS testbed with example tutorials are made publicly available.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*WiNTECH '23 , October 6, 2023, Madrid, Spain*
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0340-9/23/10…$15.00
https://doi.org/10.1145/3615453.3616521

## CCS CONCEPTS

• **Networks → Network experimentation**; **Wireless access points, base stations and infrastructure**.

## KEYWORDS

Millimeter-wave communication; software-defined radios; virtualized radio access networks; wireless experimentation

**ACM Reference Format:**
Zhenzhou Qi, Zhihui Gao, Chung-Hsuan Tung, Tingjun Chen. 2023. Programmable Millimeter-Wave MIMO Radios with Real-Time Baseband Processing. In *The 17th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization 2023 (WiNTECH '23 ), October 6, 2023, Madrid, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3615453.3616521

## 1 INTRODUCTION

Emerging applications such as augmented reality, autonomous vehicles, and smart cities impose demanding requirements on the throughput and latency of next-generation mobile networks. In addition to the multiple-input and multiple-output (MIMO) technique that has been widely deployed in wireless networks, 5G New Radio (NR) [3] has embraced radios operating in millimeter-wave (mmWave) frequency bands [8, 27]. mmWave systems with large signal bandwidth and different numbers of RF chains can introduce significant challenges in the efficient processing of baseband signals, or *baseband processing*, which is one of the most time-consuming and computationally expensive tasks in radio access networks (RANs). Although current RANs employ dedicated hardware such as FPGAs and ASICs for baseband processing, virtualizing the RAN functions by moving their computation to edge data centers can significantly reduce the deployment cost and enable more flexible use of the network resources [26].

Recent works have focused on realizing software-based baseband processing for large-scale multi-user multiple-input multiple-output (MU-MIMO) systems in the sub-6 GHz frequency band. For example, Agora [6], as a pure software-based real-time baseband processing framework designed for a single many-core server, supports BS with up to 64 antennas and is demonstrated using a massive MIMO hardware platform. Hydra [9] expands Agora and focuses on distributing the baseband processing across many servers,

and is evaluated using a server cluster. As 5G has embraced the mmWave frequency bands in FR2 to further improve data rated leveraging the widely available spectrum, it is critical to examine if existing software-based frameworks for baseband processing, which are designed for sub-6 GHz (massive) MIMO networks, are compatible with emerging mmWave specifications, such as those specified by 5G NR frequency range 2 (FR2) [3]. In addition, a research platform equipped with programmable mmWave radios for system-level experimental evaluation of these frameworks is desired.

In this paper, *we present the design and implementation of programmable mmWave MIMO radios equipped with real-time baseband processing capabilities.* Specifically, on the hardware side, we leverage the IBM 28 GHz 64-element dual-polarized phased array antenna module (PAAM) subsystem boards in the PAWR COSMOS testbed [4, 10, 20, 21], which are connected to the USRP software-defined radios (SDRs) to form programmable 28 GHz MIMO radios. On the software side, we implement Agora-UHD, which builds on top of and enables software-defined radios (SDRs) driven by the USRP hardware driver (UHD) to interface with Agora [2] that is originally designed to support only the Iris radios [6].

We perform extensive experiments to evaluate the performance of the prototyped programmable 28 GHz radios based on the 5G NR FR2 Physical (PHY) layer with numerology 3 ($\mu = 3$), a sampling rate of 122.88 MHz, and a transmission time interval (TTI) of 0.125 ms. We consider two settings of the 28 GHz link consisting of two programmable 28 GHz radios that emulate a base station (BS) and user equipment (UE): (*i*) 1×1 single-input single-output (SISO) settings with single-polarization (H-pol *or* V-pol), and (*ii*) 2×2 MIMO settings with dual-polarization (H-pol *and* V-pol). We show that the 2×2 28 GHz polarization MIMO link can achieve an average error vector magnitude (EVM) of 10.5% and 7.9% for 16QAM and 64QAM modulation with a link signal-to-noise ratio (SNR) of 25 dB and 30 dB, respectively. The achieved bit error rate (BER) is below 1e-5 with a low-density parity check (LDPC) code rate of 1/3. In addition, we show that the developed system can meet the real-time processing deadline of 0.375 ms (3 TTIs) using only 8 CPU cores on a single many-core server, with two MIMO streams and full traffic load. The source code of Agora-UHD, its integration with the programmable 28 GHz radios in COSMOS, and the corresponding example tutorials are made publicly available [2, 5].

## 2 RELATED WORK

**mmWave SDRs and hardware platforms**. Recent efforts have been devoted to developing mmWave SDRs and platforms supporting experimentation in the PHY and higher networking layers. For example, X60 [22] is a 60 GHz platform with a 12-element Tx/Rx phased array. M-Cube [30] is a 60 GHz MIMO SDR built on top of commodity 802.11ad radios. mm-FLEX [17] devotes to the wideband supports up to 2 GHz bandwidth. In addition, MIMORPG [16] is a dual-band SDR platform supporting MIMO operation in both the sub-6 GHz and 60 GHz frequency bands. Besides the 60 GHz frequency used in 802.11ad, mMobile [14] contributes to addressing the mobility effect for 5G NR FR2 systems.

**mmWave networking and communication**. Analog beamforming can be employed to improve the link SNR, especially for mmWave frequencies with higher signal attenuation. Beyond the conjugate beamforming along the direct path, other algorithms have recently been proposed. For example, mmReliable [13] jointly leverages multiple paths for higher SNR and reliability on the link; NulliFi [18] generates the wide nullings to minimize the interference from the side lobes. In a large-scale mmWave environment with dense multiple Tx-to-Rx links, BounceNet [15] mitigates the interference by employing the reflected paths; CoBF [29] further reduces the overhead of building links by detecting the radio mobility.

**MIMO baseband processing**. Despite the promising higher data rate the MIMO configuration brings, software-based MIMO baseband processing has been considered as one of the key challenges for virtualized RANs, due to its heavy computation demand. Existing works [6, 9, 28] have shown that pure software-based solutions are feasible by leveraging modern processors and carefully designed parallelism in the processing pipeline. For example, Agora [2] is an open-source framework supporting baseband processing for massive MIMO systems with up to 64 antennas using a single many-core server. The performance of Agora has been experimentally demonstrated using the Iris SDRs [6]. Hydra [9], a successive work from Agora with a distributed solution, utilizes emulated fronthaul traffic with another cluster of servers for system evaluation. However, these baseband processing frameworks have not yet been evaluated, using simulation or real radio hardware, for mmWave frequency bands.

This work demonstrates a 28 GHz 2×2 polarization MIMO link with real-time baseband processing using a single many-core server, via the integration of programmable mmWave frontends in the PAWR COSMOS testbed with an open-source C++ framework for baseband processing.

## 3 PRELIMINARIES

### 3.1 Wireless PHY and Virtualized RANs

**Orthogonal frequency-division multiplexing (OFDM)**. Modern communication systems, such as cellular and Wi-Fi, employ an OFDM-based PHY layer, which embeds data on overlapping but orthogonal subcarriers to improve spectrum efficiency. OFDM leverages the FFT/IFFT operations to convert signals between the time and frequency domains during

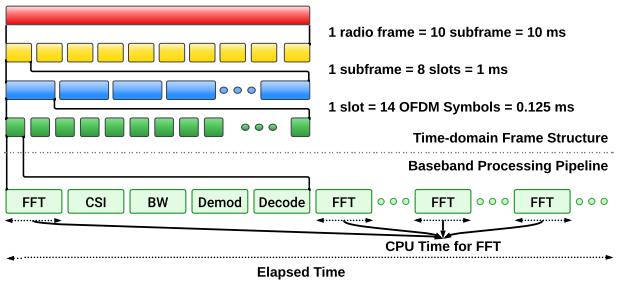



**Figure 1: Time-domain frame structure for 5G NR FR2 with numerology 3 ($\mu = 3$) and the baseband processing pipeline.**

demodulation/modulation, with the FFT size equal to the number of subcarriers in the frequency domain.

**Multiple-input multiple-output (MIMO) systems**. MIMO technology uses multiple transmitter and receiver antennas to simultaneously transfer multiple data streams leveraging spatial diversity and multiplexing. In a multi-user MIMO (MU-MIMO) system with an $M$-antenna BS and $K$ single-antenna UEs ($M \geq K$), the BS performs precoding of the $K$ data streams, each destined for a UE, onto the $M$ BS antennas, using the channel state information (CSI) between all pairs of BS and UE antennas [6, 24].

**Real-time baseband processing and virtualized RANs**. In conventional RANs, baseband processing usually takes place in dedicated hardware (e.g., ASICs) as part of the BS infrastructure. It refers to the process of converting information bits to the time-domain I/Q samples (for the Tx) and vice versa (for the Rx). The recent trend of virtualized RANs aims to softwarize and migrate (partial) baseband processing functions to the edge utilizing general-purpose hardware (e.g., CPUs), which has the benefits of reducing the deployment cost and enabling more flexible resource allocation. One key challenge in (virtualized) RANs is to meet the processing latency requirement, i.e., a PHY layer should be able to finish processing of the incoming/outgoing traffic in real-time, which is usually defined based on the transmission time interval (TTI). For example, Hydra [9] sets a processing latency of 2.5 ms (2.5 TTI for 1 ms slot) at the 99.99-th percentile.

## 3.2 Agora: Real-time Baseband Processing

Agora [2, 6] is a real-time, open-sourced software for massive MIMO baseband processing on a single many-core server. A crucial aspect of Agora is its ability to leverage data parallelism across CPU cores to optimize computational power and performance. Agora utilizes the Intel Math Kernel Library (MKL) [11] for matrix operations, improving processing efficiency and speed. Moreover, for LDPC decoding, Agora relies on Intel FlexRAN [12]. Agora supports two operation modes, with the added flexibility of an optional Data Plane Development Kit (DPDK) [25] feature that can be enabled for

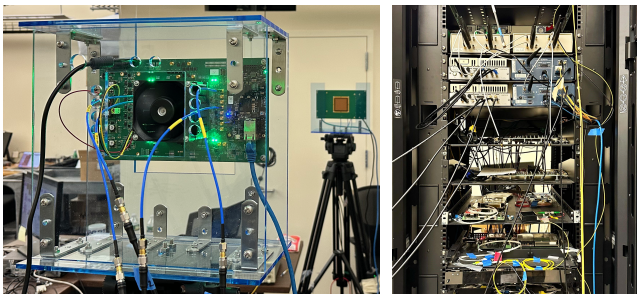


**Figure 2: Two IBM 28 GHz PAAM boards integrated in the COSMOS Sandbox 2 (sb2), which are connected to two USRP N310 SDRs to form a 28 GHz MIMO link. The USRP SDRs are connected to a compute server (not shown in the figure), which runs Agora-UHD for real-time baseband processing.**

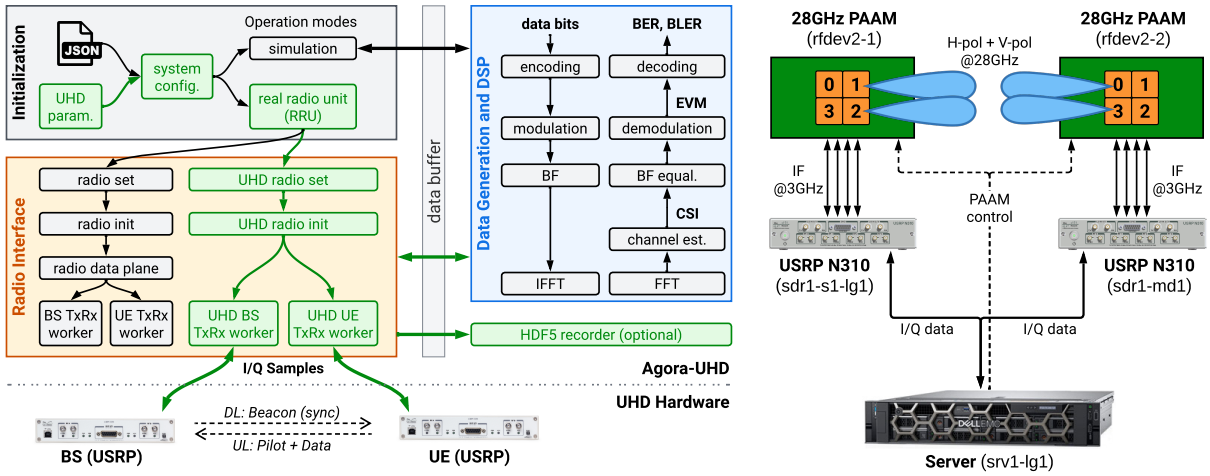
data streaming: In the *simulation mode*, Agora processes emulated RRU traffic generated by a host, while in the *RRU mode*, Agora processes I/Q baseband samples from the real radio hardware. Agora is also compatible with and can be accelerated by the Data Plane Development Kit (DPDK) library [25] in both modes. Agora supports the Iris radio hardware from Skylark Wireless using the SoapyIris driver [2], and has been integrated with the sub-6 GHz massive MIMO platform as part of the POWDER-RENEW testbed [1].

## 3.3 5G NR FR2 with Numerology 3 ($\mu = 3$)

5G NR includes two frequency ranges in 410–7,125 MHz (FR1) and 24.25–52.6 GHz (FR2). The *numerology*, also denoted as $\mu$, defines the subcarrier spacing (SCS) and corresponding time-domain length of the 5G OFDM-based PHY layer [3]. In particular, 3GPP Release 17 specifies 7 types of numerology ($\mu = 0, 1, \ldots, 6$) that correspond to a SCS of $2^{\mu} \cdot 15$ kHz. In this work, we consider the OFDM-based PHY layer configuration of 5G NR FR2 in numerology 3 ($\mu = 3$), with an SCS of 120 kHz, a sampling rate of 122.88 MHz, FFT size of 1,024, and 64 resource blocks (RBs), each with 12 subcarriers [12]. Under this configuration, each TTI (or slot) consisting of 14 OFDM symbols has a duration of 0.125 ms, as shown in Fig. 1. Note that different from previous work that focuses on real-time baseband processing for MU-MIMO systems in 5G NR FR1 [6, 9, 28], real-time processing for 5G NR FR2 systems with higher numerology presents different challenges due to the much shorter TTI duration that is inversely proportional to $2^{\mu}$.

## 3.4 Programmable mmWave Radios in the PAWR COSMOS Testbed

The PAWR COSMOS testbed [20] is an open-access platform to support advanced wireless, optical, and edge computing research. One key component of the COSMOS testbed is the mmWave frontends that are connected to SDRs, including




(a) (b)

**Figure 3: (a) Block diagram of the implemented Agora-UHD, which enables UHD-based SDRs to interface with Agora [6]. (b) System connectivity diagram of the 2×2 28 GHz dual-polarization MIMO link using the COSMOS Sandbox 2 (sb2), which includes two USRP N310 SDRs, two IBM 28 GHz dual-polarized PAAM subsystems boards, and a compute server (see also Fig. 2).**

the IBM 28 GHz 64-element dual-polarized phased array antenna module (PAAM) subsystem boards [4, 10, 21], as shown in Fig. 2. In particular, each PAAM contains four 16-element subarrays (4×4 subarray with 0.55 $\lambda$ element spacing), and supports 8 simultaneous 16-element beams or 2 simultaneous 64-element beams across the horizontal polarization (H-pol) and vertical polarization (V-pol). It also supports arbitrary beamforming with a per-element amplitude and phase resolution of approximately 0.5 dB and 4.9 degree, respectively. The PAAM board has IF interfaces at 3 GHz, which are connected to the USRP RF ports via SMP-to-SMA coaxial cables, and performs signal up-/down-conversion between 3 GHz and 28 GHz. Using a Python-based API, users can configure and control the PAAM board by sending API commands to the MicroZed FGPA plugged into the PAAM board via the Ethernet interface. More details about the measurements and control of the 28 GHz PAAM boards can be found in [4].

## 4 IMPLEMENTATION OF AGORA-UHD: ADDING UHD SUPPORT TO AGORA

We now present the implementation of Agora-UHD, an extension of Agora in the RRU mode to support UHD-based SDRs such as USRP. Fig. 3(a) shows the block diagram of Agora, where the green blocks are developed as part of Agora-UHD. We implement Agora-UHD in C++ as a userspace application, including 1,500 source lines of code added to the original Agora codebase. Agora-UHD is made open-source at [2].

**System configuration and UHD radio set/init**. Similar to Agora, Agora-UHD takes a `json` formatted file as input, which includes the configuration parameters of the PHY layer (e.g., signal bandwidth and frame schedule) and radio hardware (e.g., carrier frequency and RF channels/gains) used for program initialization. The UHD radio interface is implemented using the native `uhd::usrp`, `uhd::tx_streamer`, and `uhd::rx_streamer` class objects, which provide functions to initialize and tune the radio hardware, activate the Tx/Rx streamer, and stream I/Q samples between the host and USRPs. In addition, multiple USRP SDRs synchronized via an external clock (e.g., the OctoClock-G clock distribution module) can be used to form a multi-antenna BS. Compared to vanilla Agora, which treats each Iris radio as a separate object, Agora-UHD manages multiple radios via a single object based on the `uhd::usrp::multi_usrp` device class.

**Radio hardware time offset calibration**. Agora-UHD calibrates two types of time offset when integrated with USRP SDRs. The first type is the potential time offset between the USRPs serving as the BS and UE, which needs to be calibrated to ensure that the transmitted and received symbols align perfectly with the frame schedule. The second type is the inherent time offset between the FPGA timestamp specified by UHD to issue a Tx/Rx command and the actual timestamp of when the RF signal is transmitted/received by the radio frontend. Agora-UHD calibrates the first type of time offset by adding a calibrated offset to the UE's Tx timestamp with respect to that of the BS. For the second type of time offset, Agora-UHD adds a delay to the BS's Tx timestamp, which is empirically determined by measuring the number of I/Q samples to be delayed such that the signal transmitted by the radio frontend aligns precisely with the timestamp specified by UHD. Note that both calibrations only need to be performed once for a given radio sampling rate.

**UHD TxRx worker and real-time DSP**. We implement the BS and UE worker classes to support the Tx/Rx streaming between the server and the USRP SDRs that serve as the BS and UE, respectively. These worker classes handle the tasks and provide an interface to Agora's real-time DSP pipeline, as shown in Fig. 3(a). The real-time DSP pipeline includes multiple "doer" classes, and each handles a signal processing task, such as FFT, channel estimation, equalization, demodulation, and decoding. Taking the UL transmission as an example. Each UE captures beacon signals sent by the BS and synchronizes its timestamp with that of the BS. It then transmits a sequence of uplink pilot and data symbols following the predefined frame schedule. The BS then extracts the uplink CSI using the received uplink pilot and performs demodulation and decoding of the uplink data symbols. Overall, the DSP pipeline reports, in real-time, different metrics for measuring the quality of transmission of the MIMO system, including SNR, EVM, BER, and block error rate (BLER).

**Optional HDF5 file recording for post-processing and analysis**. In addition to the real-time DSP pipeline provided by Agora, we also update the optional hierarchical data format version 5 (HDF5) file recorder system, which uses a separate thread to store raw data from selected buffers, e.g., raw I/Q samples received by the BS. This optional HDF5 recorder, which is built on top of the open-source RENEW Sounder framework [7, 23], facilitates the visualization and post-processing of the I/Q data received by the BS. Via analyzing the data recorded in the HDF5 file, metrics for measuring the quality of transmission of MIMO systems, including SNR, EVM, BER, and BLER, can be extracted.

## 5 INTEGRATION OF AGORA-UHD WITH THE IBM 28 GHZ PAAM BOARDS

**USRP-PAAM setup in COSMOS**. We consider $2 \times 2$ polarization MIMO in 5G NR FR2, whose setup is depicted in Fig. 3(b), leveraging the dual-polarized antenna and beamforming ICs as part of the 28 GHz PAAM board. We prototype two 28 GHz MIMO radios, using the two USRP N310 SDRs (`sdr1-s1-lg1` and `sdr1-md1`) and two IBM 28 GHz PAAM boards (`rfdev2-1` and `rfdev2-2`) in the COSMOS sb2. In particular, two uplink and downlink channels are enabled between the BS and UE in the H-pol and V-pol, using two RF chains on each USRP and PAAM board. Note that the limited isolation between the H-pol and V-pol antennas can result in cross-polarization interference across the two MIMO streams [21]. We use a customized Python program built based on the PAAM API to configure each PAAM board, including the activation of individual ICs and frontend elements as well as the configuration of each element's amplitude/phase for beamforming. Due to the relatively short link distance of around 8 ft between the two PAAM boards

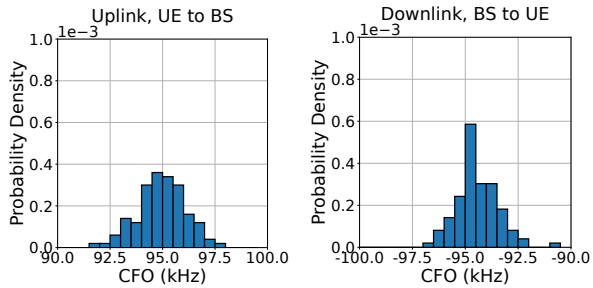



**Figure 4: Probability density distributions of the estimated CFO values between the two 28 GHz PAAM boards in the COSMOS sb2 on the uplink (UE→BS), and downlink (BS→UE).**

in the COSMOS sb2, we use a single IC to form 16-element Tx and Rx beams with −8 dB gain backoff on each element. Both USRP N310 SDRs are connected to the same compute server (`srv1-lg1`) via dual 10 GbE connections. The server, which is equipped with an Intel Xeon Gold 6226 CPU with 48 cores, runs Agora-UHD and controls both PAAM boards.

**CFO calibration between two 28 GHz PAAM boards**. Since each 28 GHz PAAM board is driven by its own on-board phase lock loop (PLL), the CFO between two 28 GHz PAAM boards needs to be calibrated in order to establish a 28 GHz communication link. We implement a coarse CFO calibration scheme that estimates the CFO between two 28 GHz PAAM boards, which is then applied to the on-board PLL. The Zadoff-Chu sequence is selected due to its uniform power distribution in the time and frequency domain [19]. In particular, we use a Zadoff-Chu (ZC) sequence with a length of $L$ I/Q samples, denoted by $x_{\text{zc}}[k]\,(k = 1, \ldots, L)$, given by

$$x_{\text{zc}}[k] = \exp\left(-j \cdot \frac{\pi R k(k-1)}{L}\right),\ k = 1, \ldots, L, \tag{1}$$

where $R = 25$ is the root order of the ZC sequence. Given the radio sampling rate of $F_{\text{samp}}$, the CFO $\widehat{\Delta f}$ can be estimated by the auto-correlation on the two received copies of the ZC sequence with a delay of $L$ samples. Let $y_{\text{zc}}[k]\,(k = 1, \ldots, 2L)$ denote the received ZC sequence, $\widehat{\Delta f}$ is given by:

$$\widehat{\Delta f} = -\frac{F_{\text{samp}}}{2\pi L} \angle \left(\sum_{k=1}^{L} x_{\text{zc}}[k] \cdot y^*_{\text{zc}}[k+L]\right), \tag{2}$$

where $(\cdot)^*$ and $\angle(\cdot)$ denote the conjugation operation and phase of a complex number, respectively. Since the phase term in (2) can only be resolved within the range of $[-\pi, +\pi]$, the estimated value of CFO, $\widehat{\Delta f}$, is limited to $\left[-\frac{F_{\text{samp}}}{2L}, +\frac{F_{\text{samp}}}{2L}\right]$.

For the baseband sampling rate of $F_{\text{samp}} = 122.88$ MHz, we select a ZC sequence length of $L = 139$ so that the CFO between the 2 PAAM boards, which is around 100 kHz, can be calibrated using (2) with $\frac{F_{\text{samp}}}{2L} \approx 440$ kHz. Using the two 28 GHz PAAM boards in COSMOS, we repeat the CFO estimation process 100 times, and Fig. 4 shows the histogram of the estimated CFO values. Specifically, the mean CFO values

**Table 1: Modulation and coding schemes (MCSs) and the corresponding per-stream data rates (at a sampling rate of 122.88 MHz) used in the experimental evaluation.**

| Modulation | Code rate | Spectrum efficiency | Per-stream data rate (@122.88 MHz) |
|---|---|---|---|
| 16QAM | 1/3 | 1.99 bps/Hz | 245.51 Mbps |
| 16QAM | 2/3 | 2.67 bps/Hz | 327.72 Mbps |
| 64QAM | 1/2 | 2.00 bps/Hz | 245.76 Mbps |
| 64QAM | 2/3 | 3.00 bps/Hz | 368.64 Mbps |

on the uplink (UE→BS) and downlink (BS→UE) are 95.0 kHz and −93.0 kHz, respectively, which are almost symmetric. We then calculate the mean value of the estimated CFO value on the downlink and apply the CFO to the onboard PLL of the PAAM board, resulting in a residual CFO of around 1 kHz across the 28 GHz link, which is further calibrated during the equalization of the received signal.

## 6 EXPERIMENTAL EVALUATION

In this section, we present the experimental evaluation of the prototyped programmable 28 GHz MIMO radio using the COSMOS sb2. We focus on (*i*) the *quality of transmission*, and (*ii*) *real-time baseband processing* with three configurations:

(1) **SISO, H-pol**: 1×1 28 GHz link in H-polarization;

(2) **SISO, V-pol**: 1×1 28 GHz link in V-polarization;

(3) **MIMO, dual-pol**: 2×2 28 GHz link in H- and V-polarization.

We consider the 5G NR FR2 PHY layer in numerology 3 [12] and various modulation and coding schemes (MCSs) listed in Table 1. For each configuration with a given MCS and link SNR value, we collect results across a total number of 10,000 frames to extract the performance metrics of interest, which include SNR, EVM, BER, and the time required for baseband processing. The developed software, including Agora-UHD, and experiments are made publicly available at [2, 5].

### 6.1 EVM vs. Link SNR

EVM is a metric commonly used in wireless communications to quantify the quality of the received signal compared to the ideal or reference signal. Fig. 5 shows the measured root-mean-square EVM (in %) as a function of the link SNR (in dB) for 16QAM and 64QAM. Note that the 28 GHz link in COSMOS sb2 can achieve a maximum SNR of around 32 dB, and the required EVM threshold for 16QAM and 64QAM are 12.5% and 8%, respectively, as defined by 3GPP [3]. Note that the interference across H-pol and V-pol is included in the calculation of the link SNR given by:

$$\mathrm{SNR}_{\mathrm{dB}} = 10 \times \log_{10}\left(\frac{S+I+N}{N}\right), \tag{3}$$

where $S$, $I$, and $N$ denote the received signal strength in the desired polarization, the interference from the other polarization, and the Rx noise, respectively. The results show that

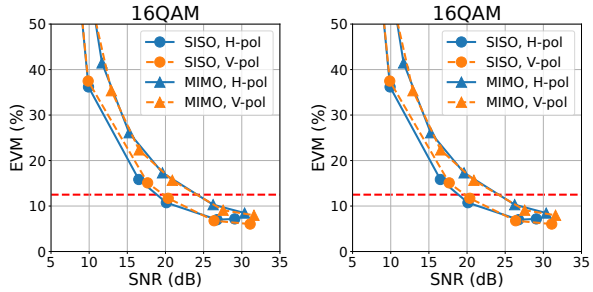



**Figure 5: EVM vs. link SNR in the SISO and MIMO configurations, with 16QAM and 64 QAM modulation. The red dashed lines indicate the minimum EVM requirement of 12.5% and 8% for 16QAM and 64QAM, respectively.**

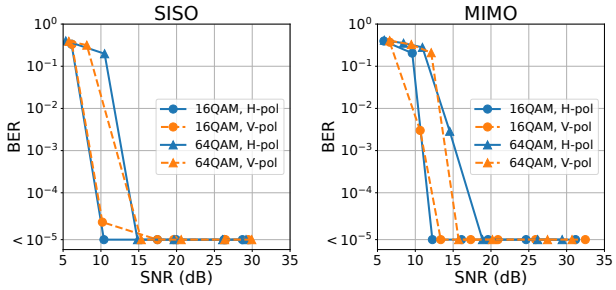



**Figure 6: BER vs. link SNR in the SISO and MIMO configurations, with 16QAM/64QAM and 1/3 LDPC code rate.**

for a SISO link with single-polarization (H-pol *or* V-pol) and 16QAM, a link SNR of 20 dB is required to obtain an EVM that is below the 12.5% threshold. For a dual-polarization (H-pol *and* v-pol) MIMO link, a higher link SNR of 25 dB is required, due to the finite isolation between the H- and V-polarization provided by the dual-polarized antenna arrays [21]. Similarly, for a SISO link with single-polarization and 64QAM, a link SNR of 25 dB is required to obtain an EVM that is below the 8% threshold. This requirement is increased to 30 dB for a dual-polarization MIMO link, a higher link SNR of 30 dB. Overall, the EVM in each polarization exhibits similar performance in both the SISO and MIMO settings.

### 6.2 BER vs. Link SNR

In addition to the EVM results, Agora-UHD also reports the BER of the link, which can be used to evaluate the end-to-end transmission quality over the 28 GHz link. For the LDPC decoding, we set the maximum number of LDPC iterations to be 5 during the iterative decoding processing. Fig. 6 shows the BER as a function of the link SNR in the SISO (single-polarization) and MIMO (dual-polarization) configurations, with an LDPC code rate of 1/3. Under such a code rate, both 16QAM and 64QAM in the SISO configuration with H-pol or V-pol can achieve high-quality transmissions with a BER of less than 1e-5 when the link SNR is higher than 15 dB. For the dual-polarization MIMO link, a higher SNR of greater than 20 dB is required to achieve the same level of BER.

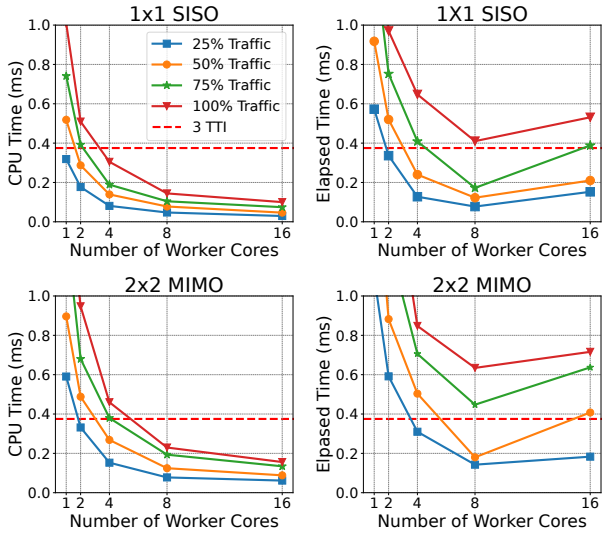



**Figure 7: CPU time and elapsed time for the 5G NR FR2 with numerology 3 ($\mu = 3$) with varying traffic loads.**

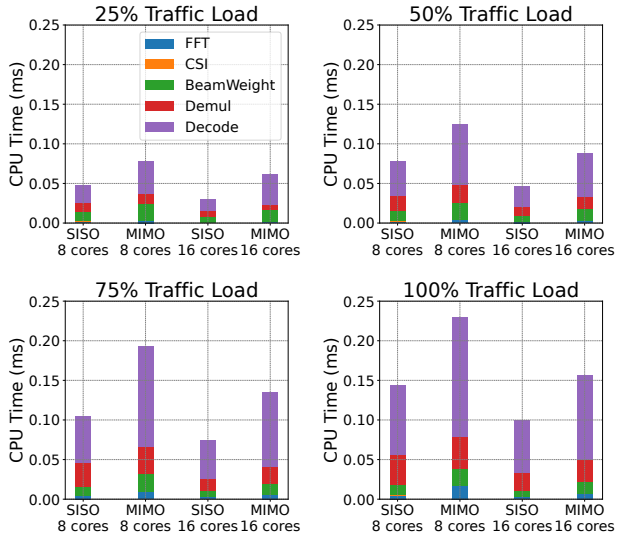



**Figure 8: Detailed decomposition of the CPU processing time with respect to each task in the SISO and MIMO configurations, with varying traffic loads and number of CPU cores.**

## 6.3 Real-time Baseband Processing

We consider two system metrics to evaluate the real-time baseband processing performance for the 28 GHz link: *CPU time* and *elapsed time*. In particular, CPU time represents the total time during which the CPU cores are actively engaged in processing a program's instructions, which correspond to the DSP tasks in Agora-UHD. This is an accurate measure of the computational resources consumed by a task, solely focusing on the period when the CPU is processing a task. The elapsed time, also known as wall-clock time or the "real" time, refers to the actual time taken from the start to the end of a computer process with respect to real-world time. This includes everything that happens during the execution of the computer process, including waiting for system resources to be available or dependent processes to finish execution. We define the elapsed time as the time from the server receiving the first packet to finishing decoding the last OFDM symbol in the corresponding slot (see Fig. 1).

For each configuration, we execute the Agora-UHD on the 28 GHz link for over 10,000 frames to collect the performance metrics of interest with significant confidence. We then evaluate the maximum of both CPU time for each DSP task and elapsed time (see Fig. 1) to obtain a clear measure of the system performance under different traffic load conditions and with different numbers of CPU cores. Traffic load here refers to the proportion of the OFDM symbols in a TTI 1 that's being used to transmit data and occupy the computation resource from the server. Our test settings follow the 5G FR2 numerology 3 ($\mu = 3$) with the corresponding PHY layer parameters such as bandwidth, 3 TTI deadline, SCS, and the number of resource blocks (RB). As a case study, we showcase the performance of uplink transmission in Fig. 7.

Figs. 7 shows the CPU time for both 1×1 and 2×2 configurations with varying traffic load and number of CPU cores. The results show that under varying traffic loads, our system is capable of meeting the 3 TTI processing deadline using 8 cores. However, a discrepancy between the CPU time and the elapsed time can be observed. This is due to the inter-core communication overhead, which can be mitigated through optimized software and processing pipeline design. Fig. 8 plots the decomposed CPU time corresponding to each DSP task in the 1×1 SISO and 2×2 MIMO configurations. It can be observed that when the number of spatial streams increases from one to two, the CPU time required to complete each DSP task does not increase proportionately except for LDPC decoding, which is the most time-consuming task that can be offloaded to dedicated hardware as LDPC accelerators.

# 7 CONCLUSIONS

In this work, we presented the design and implementation of a 2×2 28 GHz dual-polarization MIMO link with real-time baseband processing via the integration of Agora-UHD with the programmable mmWave radios deployed in the open-access PAWR COSMOS testbed. Through extensive experiments, we demonstrated the reliability of the 2×2 28 GHz dual-polarization MIMO link in terms of its EVM and BER performance, and showed that its real-time baseband processing with full traffic load and a processing deadline of 0.375 ms (3 TTIs in 5G NR FR2 with numerology 3) can be satisfied using only 8 CPU cores on a commodity server. Future research directions include: (*i*) building larger-scale programmable MIMO radios in the mmWave frequency band, (*ii*) optimization of the baseband processing pipeline tailored

to the 5G NR FR2 PHY layer and traffic patterns, and (*iii*) integration of dynamic beam tracking and management schemes with the developed programmable mmWave MIMO radios.

## ACKNOWLEDGMENTS

This work was supported in part by NSF grants CNS-1827923, CNS-2211944, AST-2232458, and an IBM Academic Award. This work was also supported in part by the Center for Ubiquitous Connectivity (CUbiC), sponsored by Semiconductor Research Corporation (SRC) and Defense Advanced Research Projects Agency (DARPA) under the JUMP 2.0 program. We thank Arun Paidimarri, Bodhisatwa Sadhu, and Alberto Valdes-Garcia (IBM Research), Rahman Doost-Mohammady and Andrew Sedlmayr (Rice University), and Ivan Seskar (Rutgers University) for their contributions to this work.

## REFERENCES

[1] 2022. POWDER: Platform for Open Wireless Data-driven Experimental Research. https://powderwireless.net

[2] 2023. Agora: A complete software realization of real-time massive MIMO baseband processing. https://github.com/Agora-wireless/Agora/tree/develop

[3] 3GPP. 2020. *5G; NR; Base Station (BS) Radio Transmission and Reception.* Technical Specification (TS) 38.211. 3rd Generation Partnership Project (3GPP). https://www.etsi.org/deliver/etsi_ts/138100_138199/138104/16.04.00_60/ts_138104v160400p.pdf Version 16.4.0.

[4] Tingjun Chen, Prasanthi Maddala, Panagiotis Skrimponis, Jakub Kolodziejski, Xiaoxiong Gu, Arun Paidimarri, Sundeep Rangan, Gil Zussman, and Ivan Seskar. 2022. Programmable and open-access millimeter-wave radios in the PAWR COSMOS testbed. In *Proc. ACM WiNTECH'21.*

[5] COSMOS. 2023. COSMOS Tutorials. https://wiki.cosmos-lab.org/wiki/Tutorials/Wireless/mmwavePaamAgora

[6] Jian Ding, Rahman Doost-Mohammady, Anuj Kalia, and Lin Zhong. 2020. Agora: Real-time massive MIMO baseband processing in software. In *Proc. ACM CoNEXT'20.*

[7] Rahman Doost-Mohammady, Oscar Bejarano, Lin Zhong, Joseph R Cavallaro, Edward Knightly, Z Morley Mao, Wei Wayne Li, Xuemin Chen, and Ashutosh Sabharwal. 2018. RENEW: programmable and observable massive MIMO networks. In *Proc. IEEE ACSSC'18.*

[8] Jinfeng Du, Dmitry Chizhik, Reinaldo A Valenzuela, Rodolfo Feick, Guillermo Castro, Mauricio Rodriguez, Tingjun Chen, Manav Kohli, and Gil Zussman. 2020. Directional measurements in urban street canyons from macro rooftop sites at 28 GHz for 90% outdoor coverage. *IEEE Trans. Antennas Propag.* 69, 6 (2020), 3459–3469.

[9] Junzhi Gong, Anuj Kalia, and Minlan Yu. 2023. Scalable distributed massive MIMO baseband processing. In *Proc. USENIX NSDI'23.*

[10] Xiaoxiong Gu, Arun Paidimarri, Bodhisatwa Sadhu, Christian Baks, Stanislav Lukashov, Mark Yeck, Young Kwark, Tingjun Chen, Gil Zussman, Ivan Seskar, et al. 2021. Development of a compact 28-GHz software-defined phased array for a city-scale wireless research testbed. In *Proc. IEEE IMS'21.*

[11] Intel. 2019. Math kernel library (MKL). https://software.intel.com/en-us/mkl

[12] Intel. 2022. intel/FlexRAN. https://github.com/intel/FlexRAN

[13] Ish Kumar Jain, Raghav Subbaraman, and Dinesh Bharadia. 2021. Two beams are better than one: Towards reliable and high throughput mmWave links. In *Proc. ACM SIGCOMM'21.*

[14] Ish Kumar Jain, Raghav Subbaraman, Tejas Harekrishna Sadarahalli, Xiangwei Shao, Hou-Wei Lin, and Dinesh Bharadia. 2020. mMobile: Building a mmWave testbed to evaluate and address mobility effects. In *Proc. ACM mmNets'20.*

[15] Suraj Jog, Jiaming Wang, Junfeng Guan, Thomas Moon, Haitham Hassanieh, and Romit Roy Choudhury. 2019. Many-to-Many Beam Alignment in Millimeter Wave Networks. In *Proc. USENIX NSDI'19.*

[16] Jesus Lacruz, Rafael Ortiz, and Joerg Widmer. 2021. A real-time experimentation platform for sub-6 GHz and millimeter-wave MIMO systems. In *Proc. ACM MobiSys'21.*

[17] Jesus Omar Lacruz, Dolores Garcia, Pablo Jiménez Mateo, Joan Palacios, and Joerg Widmer. 2020. mm-FLEX: an open platform for millimeter-wave mobile full-bandwidth experimentation. In *Proc. ACM MobiSys'20.*

[18] Sohrab Madani, Suraj Jog, Jesus O Lacruz, Joerg Widmer, and Haitham Hassanieh. 2021. Practical null steering in millimeter wave networks. In *Proc. USENIX NSDI'21.*

[19] Renaud-Alexandre Pitaval, Branislav M Popović, Peng Wang, and Fredrik Berggren. 2020. Overcoming 5G PRACH capacity shortfall: Supersets of Zadoff–Chu sequences with low-correlation zone. *IEEE Trans. Commun.* 68, 9 (2020), 5673–5688.

[20] Dipankar Raychaudhuri, Ivan Seskar, Gil Zussman, Thanasis Korakis, Dan Kilper, Tingjun Chen, Jakub Kolodziejski, Michael Sherman, Zoran Kostic, Xiaoxiong Gu, et al. 2020. Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless. In *Proc. ACM MobiCom'20.*

[21] Bodhisatwa Sadhu, Yahya Tousi, Joakim Hallin, Stefan Sahl, Scott K Reynolds, Örjan Renström, Kristoffer Sjögren, Olov Haapalahti, Nadav Mazor, Bo Bokinge, et al. 2017. A 28-GHz 32-element TRX phased-array IC with concurrent dual-polarized operation and orthogonal phase and gain control for 5G communications. *IEEE J. Solid-State Circuits* 52, 12 (2017), 3373–3391.

[22] Swetank Kumar Saha, Yasaman Ghasempour, Muhammad Kumail Haider, Tariq Siddiqui, Paulo De Melo, Neerad Somanchi, Luke Zakrajsek, Arjun Singh, Owen Torres, Daniel Uvaydov, et al. 2017. X60: A programmable testbed for wideband 60 GHz WLANs with phased arrays. In *Proc. ACM WiNTECH'17.*

[23] Clayton Shepard, Abeer Javed, and Lin Zhong. 2015. Control channel design for many-antenna MU-MIMO. In *Proc. ACM MobiCom'15.*

[24] Clayton Shepard, Hang Yu, Narendra Anand, Erran Li, Thomas Marzetta, Richard Yang, and Lin Zhong. 2012. Argos: Practical many-antenna base stations. In *Proc. ACM MobiCom'12.*

[25] The Linux Foundation. 2023. Data Plane Development Kit (DPDK). http://www.dpdk.org

[26] Prabhu Kaliyammal Thiruvasagam, Vinay Venkataram, Vivek Raja Ilangovan, Maneesha Perapalla, Rajisha Payyanur, Vishal Kumar, et al. 2023. Open RAN: Evolution of architecture, deployment aspects, and future directions. *arXiv preprint arXiv:2301.06713* (2023).

[27] Xiong Wang, Linghe Kong, Fanxin Kong, Fudong Qiu, Mingyu Xia, Shlomi Arnon, and Guihai Chen. 2018. Millimeter wave communication: A comprehensive survey. *IEEE Commun. Surveys Tuts.* 20, 3 (2018), 1616–1653.

[28] Qing Yang, Xiaoxiao Li, Hongyi Yao, Ji Fang, Kun Tan, Wenjun Hu, Jiansong Zhang, and Yongguang Zhang. 2013. BigStation: Enabling scalable real-time signal processing in large MU-MIMO systems. *ACM SIGCOMM CCR* 43, 4 (2013), 399–410.

[29] Ding Zhang, Panneer Selvam Santhalingam, Parth Pathak, and Zizhan Zheng. 2023. CoBF: Coordinated beamforming in dense mmWave networks. *Proc. ACM POMACS'23* 7, 2 (2023), 1–26.

[30] Renjie Zhao, Timothy Woodford, Teng Wei, Kun Qian, and Xinyu Zhang. 2020. M-Cube: A millimeter-wave massive MIMO software radio. In *Proc. ACM MobiCom'20.*