A Case for Digital Beamforming at mmWave

Sourjya Dutta, *Student Member, IEEE*, C. Nicolas Barati, *Member, IEEE*, David Ramirez, *Member, IEEE*, Aditya Dhananjay, James F. Buckwalter, *Senior Member, IEEE*, and Sundeep Rangan, *Fellow, IEEE*

***Abstract*— Due to the heavy reliance of millimeter-wave (mmWave) wireless systems on directional links, beamforming (BF) with high-dimensional arrays is essential for cellular systems in these frequencies. Thus, performing the array processing in a power-efficient manner is a fundamental challenge. Analog and hybrid BF require few analog-to-digital and digital-to-analog converters (ADCs and DACs), but can only communicate in a small number of directions at a time, limiting directional search, spatial multiplexing, and control signaling. Digital BF enables flexible spatial processing but must be operated at a low quantization resolution to stay within reasonable power levels. This decrease in quantizer resolution distorts the received as well as the transmitted signal. To assess the effect of coarse quantization at the receiver, this paper presents a system level analytic framework based on a simple additive quantization noise model (AQNM). The analysis verified through extensive simulations reveals that at moderate resolutions (3-4 bits per ADC), there is negligible loss in downlink cellular capacity from quantization. In essence, the low resolution ADCs limit the high SNR, where cellular systems typically do not operate. For the transmitter, it is shown that DACs with 4 or more bits of resolution can support high order modulations, and do not violate the adjacent carrier leakage limit set by 3rd Generation Partnership Project (3GPP) New Radio (NR) standards for cellular operations. In fact, our findings suggest that low resolution digital BF architectures can be a power-efficient alternative to analog or hybrid BF for both transmitters and receivers at millimeter-wave.**

***Index Terms*— Millimeter wave, 5G cellular, Low resolution quantizers, digital beamforming.**

I. Introduction

THE need for more bandwidth, driven by ever higher demand, has brought millimeter wave (mmWave) communication into the spotlight as an enabling technology for the 5th generation (5G) wireless communication systems. By offering large blocks of contiguous spectrum, mmWave presents a unique opportunity to overcome the bandwidth crunch problem faced by lower frequency bands [1]. At mmWave frequencies, high isotropic path loss necessitates the reliance on antenna arrays with large number of elements. These arrays overcome the path loss by high directional gains through beamforming (BF). Thus, a transmitter–receiver (Tx–Rx) pair uses large number of antennas to focus energy in a particular direction to boost link quality. A key challenge for large antenna arrays, and the motivation of our work, is to find an architecture capable of high-dimensional array processing in a power-efficient manner.

Most current mmWave designs use *analog* [2] or *hybrid beamforming* [3]. In these cases, beamforming is performed in radio frequency (RF) or at an intermediate frequency (IF) through a bank of phase shifters (PSs) – one per antenna element as shown in Fig. 1. This architecture reduces the power consumption by using only one pair of analog to digital converters (ADC) and digital to analog converters (DAC) at the Rx and Tx, respectively, per digital stream. While analog and hybrid beamforming are power efficient, they are only capable of transmitting in one or a few directions at a given time [4]. This essentially limits their multiplexing capabilities. In contrast, in *fully digital architectures* [5]–[8], shown in Fig. 2, beamforming is performed in baseband. Each RF chain has a pair of ADCs at the Rx and DACs at the Tx enabling the transceiver to simultaneously direct beams in theoretically infinite directions at a given time. But, for wide-band systems high precision ADCs and DACs can be very power hungry. To be energy efficient, fully digital beamformers need to use converters with a small number of bits of resolution (typically 4 or less) [9]. In this work, we seek to characterize the trade-off between power efficiency, quantizer resolution, and multiplexing capabilities of fully digital BF at mmWave.

A. Signal Processing With Low Resolution Quantizers

For communication systems, the degradation due to low resolution converters can be viewed as the introduction of *quantization noise* in the signal. Low resolution converters can be simply viewed as a noise source and the introduction of this additional noise has the effect of lowering the achievable link capacity. Studies on point-to-point links [7], [9], [10] have demonstrated that operations over wide band channels with low resolution ADCs can achieve sufficient spectral

Manuscript received January 23, 2019; revised June 6, 2019 and September 3, 2019; accepted October 13, 2019. Date of publication October 25, 2019; date of current version February 11, 2020. The work of S. Dutta, A. Dhananjay, and S. Rangan were supported in part by NSF under Grant 1936332, Grant 1824434, Grant 1833666, and Grant 1564142, in part by NYU Wireless and its industrial affiliates, in part by the National Institute of Standards and Technology (U.S. Department of Commerce) (NIST) under Grant 70NANB17H166, and in part by Semiconductor Research Corporation (SRC). The associate editor coordinating the review of this article and approving it for publication was I. Guvenc. *(Corresponding author: Sourjya Dutta.)*

S. Dutta, A. Dhananjay, and S. Rangan are with NYU Wireless, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: sdutta@nyu.edu; aditya@courant.nyu.edu; srangan@nyu.edu).

C. N. Barati is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: nicobarati@rice.edu).

D. Ramirez is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: dard@princeton.edu).

J. F. Buckwalter is with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: buckwalter@ece.ucsb.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2019.2948329

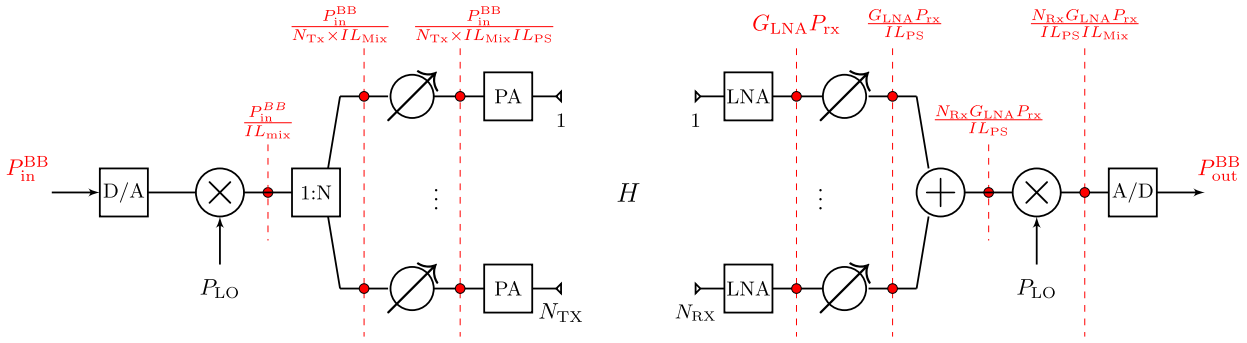


Fig. 1. Analog beamforming based transmitter (left) and receiver (right) use a bank of phase shifter to perform beamforming in the RF domain. This architecture uses just one pair of A/D or D/A at the baseband.

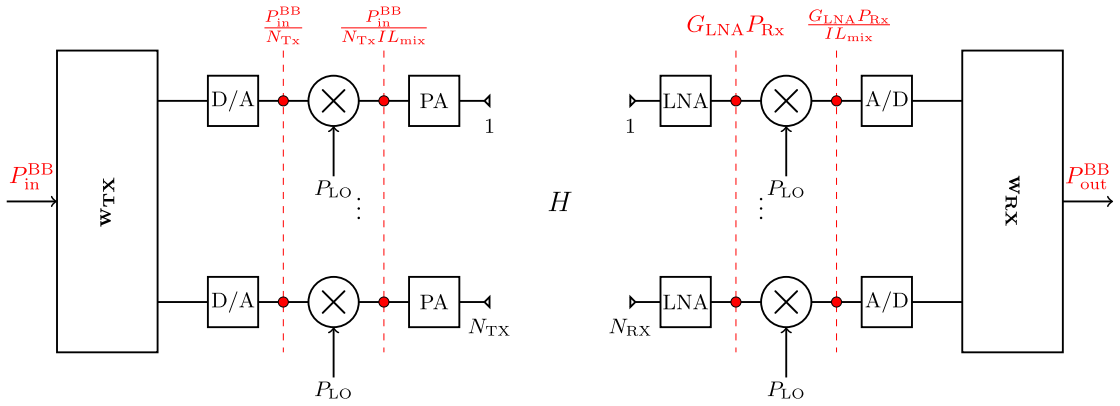


Fig. 2. Fully digital transmitter (left) and receiver (right) use a pair of DACs and ADCs per RF stream. Beamforming is performed in the baseband digital domain.

efficiency. Even with a single bit of precision, as shown in [7] and [10], wide band multi-antenna systems can achieve considerably high spectral efficiency when perfect channel information is assumed. Further, MIMO channel estimation for wide band systems has been recently studied under the low resolution limit in [11] and [12]. Moreover, the information theoretic work in [13] shows that for point-to-point systems, digital beamforming offers higher rates than analog for a given power budget. The authors in [14], similarly, shows that low resolution digital BF can achieve higher rates than hybrid beamforming with similar or even lower power consumption.

While these works consider low-resolution quantization in the data plane, there has also been considerable work for *control plane* procedures. For example, [15]–[17] consider low-resolution digital BF for directional cell search in initial access. It is shown that since fully digital front-ends can scan all directions simultaneously, they can offer a ten-fold reduction in control plane latency relative to analog BF design under typical parameters. Additionally, frequency division multiple access (FDMA) scheduling is feasible with digital BF and enables very efficient transmission of short data and control packets [18]. Since control plane signaling typically operates at low SNRs, these control plane procedures can be performed with low-resolution quantization with minimal loss.

B. Contributions

The purpose of this paper is to understand quantization and other front-end design choices impact two other key aspects: (1) capacity in multi-user, multi-antenna systems; and (2) transmission waveform quality, particularly out-of-band emissions. As we will see below, both of these performance metrics are fundamentally limited by the dynamic range of the transmitter. Hence, low resolution quantization may pose significant challenges. The contributions of this paper are threefold: (a)

1) Firstly, we provide a simple method to characterize the power consumption of wide band mmWave transceiver front ends (FEs) under different beamforming assumptions. Coupled with a detailed assessment of state-of-the-art transceiver circuits, we provide a comprehensive analysis of the power requirements by mmWave FEs. The model shows that transceiver employing low resolution DACs and ADCs at mmWave are indeed power-efficient. More interestingly, the power analysis shows that the power drawn by the mixers form the local oscillator (LO) is the dominant component of power consumed by fully digital FEs.
2) Next, we propose a novel analytical framework to assess the effect of coarse quantization from the perspective of physical and medium access control (PHY-MAC) layer design for multi-antenna multi-user systems. Our analysis is based a simple but accurate additive quantization noise model (AQNM). Based on the analysis and extensive link layer and cellular simulations we show that at the Rx, $3-4$ bits of resolution is sufficient for wide-band mmWave applications.

3) Finally, we study the effects of low resolution DACs on mmWave transmissions. Wireless transmissions have to adhere to strict limits set by regulatory bodies. In this work we show that 4 or more bits of DAC resolution is sufficient to conform to the required transmitted signal characteristics set by the 3GPP new radio (NR) specifications.

The rest of this paper is organized as follows. We introduce a model for the assessment of the power consumption of the mmWave front ends in Sec. II. In Sec. III, we analyze the effect of low resolution ADCs on mmWave reception. For this, we present the AQNM and discuss on the effects quantization noise under practical operating conditions. Next, in Sec. IV, we detail the effects of low resolution quantization on the transmitted signal. In Sec. V we validate the proposed AQNM and study its effect on system rate. Further, using extensive simulation we also determine the resolution needed for the DACs at the mmWave transmitters. Sec. VI concludes the paper. This work was presented in part in [19].

II. Power Consumption in mmWave Radio Circuits

Power efficient fully digital beamformers will have to rely on low resolution converters DACs and ADCs. To better understand the effect of decreasing the precision of the converters, in this section we model the power consumption of transceiver front ends (FE) at mmWave frequencies for analog, hybrid and fully digital beamforming.

RF Front End: The RF front end (RFFE) refers to the circuitry between the antenna and the baseband data converters (DACs or ADCs). As shown in Fig.1 and Fig.2, this includes the power amplifiers (PAs) or low noise amplifiers (LNAs), mixers, PSs, combiners and splitters. At the Tx, consider that the total power delivered by the base-band circuit is $P_{\text{in}}^{\text{BB}}$. The mixers, splitters, and PSs are considered to be passive devices which introduce insertion loss (IL) but do not draw any power. From Fig. 1 we see that for analog and K-stream hybrid beamformer, the power of input signal at the PA is given as

$$\begin{aligned} P_{\text{in,ana}}^{\text{PA}} &= P_{\text{in,hyb}}^{\text{PA}} \\ &= P_{\text{in}}^{\text{BB}} - 10\log_{10}(N_{\text{Tx}}) - IL_{\text{PS}} - IL_{\text{mix}} \text{ (dBm)}, \end{aligned} \quad (1)$$

where, IL_{PS} and IL_{mix} is the IL due to the PS and mixer respectively, and $10\log_{10}(N_{\text{Tx}})$ is the loss in signal power due to the $1 : N_{\text{Tx}}$ power splitter. Similarly, from Fig. 2, for the fully digital beamformer we can write,

$$P_{\text{in,dig}}^{\text{PA}} = P_{\text{in}}^{\text{BB}} - 10\log_{10}(N_{\text{Tx}}) - IL_{\text{mix}} \text{ (dBm)}. \quad (2)$$

As a reference for the design of hybrid beamformers, the reader is directed to the work in [18]. To transmit an output power $P_{\text{out}}^{\text{RF}}$ dBm, the D.C power drawn by the PA is

$$P_{\text{DC,BF}}^{\text{PA}} = \frac{1}{\eta_{\text{PAE}}}\left(10^{0.1P_{\text{out}}^{\text{RF}}} - 10^{0.1P_{\text{in,BF}}^{\text{PA}}}\right) \text{ mW}, \quad (3)$$

where η_{PAE} is the power added efficiency of the PA and $P_{\text{in,BF}}$ is the input power for a given BF architecture. Note that, given an effective isotropic radiated power (EIRP), for an N_{Tx} antenna system, $P_{\text{out}}^{\text{RF}} = \text{EIRP} - 20\log(N_{\text{Tx}})$. The total power draw of the Tx RFFE hence is,

$$P_{\text{Tx}}^{\text{RFFE}} = N_{\text{Tx}}P_{\text{DC,ana}}^{PA} + N_s P_{\text{LO}}, \quad (4)$$

where P_{LO} is the power drawn by the mixer from the local oscillator (LO) in linear scale, and N_s is the number of baseband streams; $N_s = 1$ for analog BF, $N_s = K$ for K-stream hybrid BF and $N_s = N_{\text{Tx}}$ for fully digital BF. Based on (4), the power consumption of the Tx RFFE is reported in Table I where the calculations consider an EIRP $= 30$ dBm[1], $IL_{\text{PS}} = 10$ dB [21], $IL_{\text{mix}} = 6$ dB, $P_{\text{LO}} = 10$ dBm in dB scale [22] and $\eta_{\text{PAE}} = 20\%$ [2].

At the Rx, LNAs are characterized by their figure of merit (FoM) which relates the gain (G_{LNA}) and the noise figure (N_{LNA}) to the D.C. power draw ($P_{\text{dc}}^{\text{LNA}}$) as [23],

$$P_{\text{dc}}^{\text{LNA}} = \frac{G_{\text{LNA}}}{\text{FoM}(N_{\text{LNA}} - 1)} \quad (5)$$

in linear scale. The total RFFE power consumption at the Rx is thus,

$$P_{\text{Rx}}^{\text{RFFE}} = N_{\text{Rx}}P_{\text{dc}}^{\text{LNA}} + N_s P_{\text{LO}}, \quad (6)$$

where N_s is the number of baseband streams at the Rx. Now, for the digital BF, if the LNA gain is selected as $G_{\text{LNA,dig}}$, then for analog/hybrid BF the required LNA gain required will be $(G_{\text{LNA,dig}} + IL_{\text{PS}})$ which compensates for the IL due to the RF PSs. In Table. I, we report the Rx RFFE power consumption given $G_{\text{LNA,dig}} = 10$ dB, $IL_{\text{PS}} = 10$ dB, $N_{\text{LNA}} = 3.2$ dB and a LNA FoM $= 8.12$ mW^{-1} as reported in [24].

Gain Control at the Rx: Given a fixed Tx EIRP, the power received at the Rx is $P_{\text{Rx}}(d) = \text{EIRP}_{\text{Tx}} - \text{PL}(d)$, where $\text{PL}(d)$ is the path loss for a Tx-Rx separation of d meters. To maintain a constant baseband power of $P_{\text{out}}^{\text{BB}}$, the variable gain amplifier (VGA) at the input of the ADC needs a gain range from $0 - G_{\text{max}}^{VGA}$ dB. Noting that G_{LNA} is adjusted to compensate for IL_{PS} for analog/hybrid BF, to drive a total baseband power of $P_{\text{out}}^{\text{BB}}$, the VGA gain range required is,

$$\begin{aligned} G_{\text{max}}^{VGA} = P_{\text{out}}^{\text{BB}} - 10\log(N_{\text{Rx}}) + IL_{\text{mix}} - (G_{\text{LNA}} - IL_{\text{PS}}) \\ - P_{\text{Rx}}(d = d_{\text{cell}}) \end{aligned} \quad (7)$$

where d_{cell} is the radius of the cell. For a down-link (DL) transmission with $\text{EIRP}_{\text{Tx}} = 43$ dBm; at the cell edge, $d_{\text{cell}} = 100$ m, $P_{\text{Rx}}(d = d_{\text{cell}}) = -87$ dBm for a mmWave non-line of sight channel [25]. Assuming similar values of IL as on the Tx RFFE and considering $G_{\text{LNA,dig}} = (G_{\text{LNA}} - IL_{\text{PS}}) = 10$ dB, to maintain $P_{\text{out}}^{\text{BB}} = 10$ dBm we require a gain range of $G_{\text{max}}^{VGA} = 82$ dB.

The figure of merit of a VGA (FoM_{VGA}) is defined by [26] as,

$$\text{FoM}_{\text{VGA}} = \frac{\text{G}_{\text{max}}^{\text{VGA}} \times \text{f}_{\text{BW}}}{\text{P}_{\text{dc}}^{\text{VGA}} \times \text{A}_{\text{chip}}}, \quad (8)$$

where, $G_{\text{max}}^{\text{VGA}}$ is in dB, the bandwidth f_{BW} is in GHz, the D.C power draw $P_{\text{dc}}^{\text{VGA}}$ in mW, and the VGA active area A_{chip}

[1]For handheld UEs operating in the mmWave bands the EIRP can be between 22.4 and 43 dBm [20].

TABLE I

POWER CONSUMPTION (IN MW) FOR EACH COMPONENT IN THE RF CHAIN FOR VARIOUS RECEIVER AND TRANSMITTER ARCHITECTURES WITH 16 TX AND RX ANTENNAS. PS, COMBINERS AND MIXERS ARE CONSIDERED TO BE PASSIVE CIRCUITS. THE MIXERS DRAW POWER FOR THE LO WHICH IS ACCOUNTED ABOVE. THE DACS AND THE ADCS OPERATE AT $f_s = 1$ GHZ. THE TX POWER IS 30 DBM EIRP

Rx Front End Power Consumption [mW]						
BF Arch.	LNA	LO	VGA	ADC (8 bits)	ADC (4 bits)	Total
Analog	197.9	10	1.55	33.3	–	242.75
Hybrid ($K = 2$)	197.9	20	3.11	66.6	–	287.61
Digital (High res.)	19.8	160	24.85	532.8	–	737.45
Digital (Low res.)	19.8	160	24.85	–	33.3	237.95
Tx Front End Power Consumption [mW]						
BF Arch.	PA	LO	LPF	DAC (8 bits)	DAC (4 bits)	Total
Analog	311.2	10	0.52	34.4	–	356.12
Hybrid ($K = 2$)	311.2	20	1.04	69.2	–	401.44
Digital (High res.)	299.9	160	8.32	553.6	–	1021.82
Digital (Low res.)	299.9	160	8.32	–	34.4	502.62

is in mm^2. The FoM reported by [26] for a 90-nm CMOS process with an active area of 0.01mm^2 is 5280. Considering the same active area, we report the power drawn by the VGA(s) for the three beamforming architectures in Table I.

DAC and ADC: For wide-band wireless applications the data converters, DACs and the ADCs, are considered to be the most power hungry elements. The power consumed by an ADC or a DAC (P_{conv}) is a linear function of the sampling frequency (f_s) and grows exponentially with the number of bits of resolution (n) as

$$P_{\text{conv}} = \text{FoM} \times f_s \times 2^n, \tag{9}$$

where FoM is the figure of merit of the converter. As mmWave systems are envisioned for ultra wide-band applications, the sampling frequencies are in the order of 1 GHz. In analog or hybrid beamforming the use of one or a few pairs of converters limit the power consumption. For fully digital systems, a reduction of n is hence the only way to reduce the power consumption.

Contrary to the assumption made in [27], we observe that both DACs and the ADCs are equally power hungry. For instance the 4-bit Flash based ADC designed in [28] has a FoM $= 65$ fJ/conv, while a state of the art DAC proposed in [29] has a FoM $= 67.6$ fJ/conv. Thus, a pair of 8-bit ADC consumes 33.28 mW of power at $f_s = 1$ GHz while a pair of 4-bit ADC will consume just 2.08 mW. At this sampling rate, a pair of 8-bit DAC consumes 34.6 mW of power, nearly same as that of the ADC. On the other hand, a pair of 4-bit DAC will consume just 2.16 mW at $f_s = 1$ GHz. Similar trends can be observed in more recent works [30], [31]. Hence, as shown in Table I, the use of low resolution converters can considerably decrease the power consumed by the fully digital transceiver.

Filtering at the Tx: The output of the DACs will require analog low pass filters (LPF) to reject spectral images, and maintain out of band emission limits as discussed in Sec. IV. In this work we assume the use of active switched capacitor filters. For an m-th order active LPF with cutoff frequency at f_c, the FoM is the power consumed per pole per Hertz [32], and is given as

$$\text{FoM}_{\text{LPF}} = P_{\text{dc}}^{\text{LPF}}/(m \times f_c). \tag{10}$$

For wide band LPFs, based on [32], [33], we consider the FoM $= 1.3$ mW/GHz. For mmWave beamformers, as discussed in Sec. V-C, we can use a first order LPF with $f_c = 400$ MHz, each of these filters will thus consume as power of $P_{\text{dc}}^{\text{LPF}} = 0.52$ mW. Depending on the BF architecture, the total power drawn by the LPF is equal to $N_s P_{\text{dc}}^{\text{LPF}}$, with $N_s = 1$ for analog BF.

From Table I, we see that both at the Tx as well at the Rx, low resolution quantizers can considerably reduce the power consumption of the front-end circuitry. This reduction in power, though, comes at the cost of increased quantization noise in the system. We will analyze the effect of coarse quantization in the sequel.

III. MMWAVE RECEPTION WITH LOW RESOLUTION ADCS

In this section, we study the effect of low resolution quantization at the receiver. Without loss of generality, we study the down-link (DL) channel with the UE receiver employing low bit ADCs. We start by discussing the network in which mmWave cellular transceivers operate, and the possible beam formed multiple access strategies. Next, based on a simple additive quantization noise model to capture the effect of low precision ADCs we present an analytical framework to quantify the effect of the quantization noise on the performance of mmWave receivers.

A. Network Model

We consider a wireless network with N base stations (BS) each with $N_{\text{BS}}^{\text{ant}}$ antennas. Each BS serves a multiplicity of UEs each with $N_{\text{UE}}^{\text{ant}}$ antennas. The BSs and UEs operate over mmWave frequencies i.e., $f_c > 10$ GHz, where f_c is the carrier frequency. Down-link (DL) and uplink (UL) transmissions use the same channel using time division duplexing (TDD). To mitigate the high isotropic path loss at mmWave frequencies, the BSs and UEs employ digital beamforming with low resolution DACs and ADCs both at the Tx and the Rx.

DL Transmission: In this system, a BS transmits single streams of data to $K = |\mathcal{K}(t)|$ associated UEs, where $K \leq \min(K_0, N_{\text{BS}}^{\text{beam}})$, $\mathcal{K}(t)$ is the set of scheduled users at time instant t, K_0 is the total number of UEs associated with the BS, and $N_{\text{BS}}^{\text{beam}}$ is the number of beams supported by the BS. Under a fixed power budget of P watts, the transmitted signal is given by

$$\mathbf{x}(t) = \sum\nolimits_{k \in \mathcal{K}(t)} \rho \mathbf{v}_k s_k(t), \tag{11}$$

where, $\rho = \sqrt{P/K}$, $\mathbf{v}_k \in \mathbb{C}^{N_{\text{BS}}^{\text{ant}} \times 1}$ is the transmit side long-term beamforming vector between the BS and the k-th user. Throughout this work we assume long-term beamforming [25] where BF vectors are computed based on the channel covariance matrices. Without loss of generality, we assume $P = 1$, and through out this work we considered transmit power is split equally among all beams by the BS[2]. At each time instant t the BS schedules $\mathcal{K}(t)$ UEs which can be multiplexed in time, in frequency or spatially as discussed in Sec. III-B.

DL Reception: The signal received at the k-th UE, *before* digital beamforiming is applied is given as,

$$\mathbf{y}_k(t) = \rho \mathbf{H}_k \mathbf{v}_k s_k(t) + \sum_{\jmath \in \mathcal{K}(t), \jmath \neq k} \rho \mathbf{H}_k \mathbf{v}_\jmath s_\jmath(t) + \mathbf{z}_k + \mathbf{n}_k, \tag{12}$$

where $\mathbf{H}_k \in \mathbb{C}^{N_{\text{UE}}^{\text{ant}} \times N_{\text{BS}}^{\text{ant}}}$ represents the channel matrix between user k and the BS, $\mathbf{z}_k \in \mathbb{C}^{N_{\text{UE}}^{\text{ant}} \times 1}$ represents the inter-cell interference, and $\mathbf{n}_k \in \mathbb{C}^{N_{\text{UE}}^{\text{ant}} \times 1}$ represents the receiver noise. Following the work in [25], the channel matrix for a BS-UE pair can be given as,

$$\mathbf{H}_k = \frac{1}{\sqrt{L}} \sum_{m=1}^{M} \sum_{\ell=1}^{L} g_{ml} \mathbf{u}_{\text{rx}}(\theta_{ml}^{\text{rx}}, \phi_{ml}^{\text{rx}}) \mathbf{u}_{\text{tx}}^{\mathbf{H}}(\theta_{ml}^{\text{tx}}, \phi_{ml}^{\text{tx}}), \tag{13}$$

where $m = 1, \ldots M$ is the cluster index, $\ell = 1, \ldots L$ is the sub-path index within the cluster, g_{ml} is the small-scale fading gain on the ℓ-th sub-path of the m-th cluster, $(\theta_{ml}^{\text{rx}}, \phi_{ml}^{\text{rx}})$ are the zenith and azimuth angle of arrivals of the subpaths respectively, and $(\theta_{ml}^{\text{tx}}, \phi_{ml}^{\text{tx}})$ are the zenith and azimuth angles of departures of the sub-paths respectively. Here, $\mathbf{u}_{\text{tx}}$ and $\mathbf{u}_{\text{rx}}$ are the spatial signatures of the Tx and Rx arrays. Further, from (12), note that the second term on the right hand side accounts for the *intra-cell interference* (ICI). The receiver noise is assumed to be zero mean i.i.d. Gaussian with covariance matrix given as $\sigma_n^2 \, I_{N_{\text{UE}}^{\text{ant}}}$. Similarly, the inter-cell interference $\mathbf{z}_k$ is also be assumed to be i.i.d Gaussian with the covariance matrix $\sigma_z^2 \, I_{N_{\text{UE}}^{\text{ant}}}$.

B. DL Multiple Access

MmWave BSs employing fully digital beamforming are capable of transmitting to multiple directions at the same time. Unlike conventional analog/hybrid architectures this allows the multiplexing of a large number of UEs, spatially or in frequency, leading to an increase in the degrees of freedom of the system which enables low latency transmissions. The multiple access techniques available for fully digital BS transmissions are as follows.

a) TDMA: Time division multiple access (TDMA) based DL will have $|\mathcal{K}(t)| = 1$, i.e., only one UE can be scheduled for transmission at any given time instance and will have access to the entire bandwidth. As user allocations are orthogonal in time, the ICI is zero. But, TDMA can potentially lead to wastage of allocated bandwidth especially for UEs with bursty or low rate traffic. Sophisticated scheduling and frame design, such as [18], is required to intelligently exploit the large bandwidth available at mmWave.

b) OFDMA: A key attraction of digital beamforming for data transmission at mmWave frequencies is that it enables orthogonal frequency division multiple access (OFDMA). In OFDMA systems, the BS allocates chunks of the total bandwidth, physical resource blocks (PRBs), to multiple UEs at each scheduling instance. The scheduler at BS j allocates a variable portion of the bandwidth $B_{\jmath,k} \geq 0$ to $k = 1, 2 \ldots K_0$ UEs based on their data requirement, channel condition, and scheduling priority. Due to the use of orthogonal channels for transmission, OFDMA-based DL signals do not encounter ICI. Moreover, unlike TDMA, more users get access to the channel at each time slot leading to greater utilization of mmWave bands, and faster transmission of small packets. This is especially attractive for low latency communication.

c) SDMA: Fully digital beamforming has the potential to support space division multiple access (SDMA) as the BS Tx can transmit to multiple users on separate spatial streams[3]. This increases the available degrees of freedom K folds, where K is the number of streams. In this case $|\mathcal{K}(t)| = K(t)$, the "optimal" number of streams that can be supported at time t. A key challenge for SDMA based transmission is mitigating the effect of the ICI. To minimize ICI, the BS scheduler must carefully select users or beamforming patterns or both, limiting the number of beams over which transmissions occur at any scheduling instance.

C. Link Layer AQNM Model

1) Effective SINR: We first derive an analytical model for the effective SINR under the finite quantization limit for a multi-antenna receiver. For this purpose, we use a slightly modified version of the additive quantization noise model (AQNM) presented in [34]. In our AQNM, shown in Fig. 3, the effect of finite uniform quantization of a scalar input y is represented as a constant gain plus an additive white Gaussian noise. Furthermore, [34] showed that if an input complex sample y is modeled as a random variable, then the quantizer output y_q is given as

$$y_q = Q(y) = (1 - \alpha) y + w_q, \quad \mathbb{E}|w_q|^2 = \alpha(1 - \alpha)\mathbb{E}|y|^2, \tag{14}$$

where $Q(\cdot)$ denotes the quantization operation, and w_q represents quantization errors uncorrelated with y and approximated

[2]We consider that the variability in link quality can be overcome using adaptive modulation and coding like in 4G LTE. Beams having equal power allocation also avoids the near far effect which can be critical for low resolution Rx; this simplifies system design. Although, there may be further gains from power allocations and can assessed in a future work.

[3]Although the point-to-point mmWave link is low rank, additional degrees of freedom can be achieved by multi-user MIMO

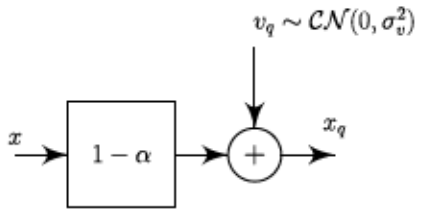


Fig. 3. Additive quantization noise model for low resolution quantizer. The parameter α is the inverse coding gain and models the quantizer resolution (e.g., $\alpha = 0$ implies infinite resolution).

as a complex Gaussian. The parameter $\alpha \in [0, 1]$ is the *inverse coding gain* of the quantizer, i.e., $\alpha = 0$ implies infinite quantizer resolution. The parameter α is assumed to depend on the resolution of the quantizer and is independent of the input distribution.

We now extend our model to a multi-antenna receiver model. For the received signal in (12), each component $y_k^i(t)$ of the received signal $\mathbf{y}_k(t)$ is independently quantized by an ADC before an appropriate receiver-side beamforming is applied. Thus, from (12) and (14), the quantized received vector is given as

$$\begin{aligned}\hat{\mathbf{y}}_k(t) &= \mathbf{Q}(\mathbf{y}_k(t)) = (1-\alpha)\rho\mathbf{H}_k\mathbf{v}_k s_k(t) \\ &\quad + (1-\alpha)\sum_{\jmath\in\mathcal{K}(t),\jmath\neq k}\rho\mathbf{H}_k\mathbf{v}_\jmath s_\jmath(t) \\ &\quad + (1-\alpha)\mathbf{z}_k + (1-\alpha)\mathbf{n}_k + \mathbf{w}. \end{aligned} \tag{15}$$

The vector $\mathbf{w}$ denotes the additive quantization noise with covariance $\sigma_w^2\mathbf{I}_{N_{\mathrm{UE}}^{\mathrm{ant}}}$. We assume that the quantization errors across antennas are uncorrelated. The effect of correlated quantization noise across antennas have been recently analyzed in [35] for single input multiple output (SIMO) Rx, and for multi-antenna Tx in [36]. However, here we argue that at the cellular receivers the input power to the quantizer is controlled by the *automatic gain control* (AGC) loop which brings the power of the input to the quantizer within the dynamic range of the ADC. Further, the quantizers (ADC) can be designed to minimize the mean squared error (MSE) for an input Gaussian distribution [37]. For an OFDM system [4] with large number of subcarriers, i.e., wide band transmission, the scaled input samples to the quantizer can be considered Gaussian [38]. This, as also pointed in [36], makes the assumption of uncorrelated quantization noise across the multiple RF chains valid.

Hence, from (15), the average per component energy to the input $\mathbf{y}$ of the quantizer is

$$\frac{1}{N_{\mathrm{UE}}}\mathbb{E}\|\mathbf{y}\|^2 = \rho^2\ E_k + \rho^2\sum_{\jmath\in\mathcal{K}(t),\jmath\neq k} E_\jmath + \sigma_{\mathbf{z}}^2 + \sigma_n^2, \tag{16}$$

where $E_\jmath = (1/N_{\mathrm{UE}}^{\mathrm{ant}})\mathbb{E}\|\mathbf{H}_k\mathbf{v}_\jmath s_\jmath\|^2\ \forall j$ is the average received symbol energy per antenna for each stream $j \in \mathcal{K}$. From (14), the quantization noise variance is

$$\sigma_w^2 = \alpha(1-\alpha)\left[\rho^2\ E_k + \rho^2\sum_{\jmath\in\mathcal{K}(t),\jmath\neq k} E_\jmath + \sigma_z^2 + \sigma_n^2\right], \tag{17}$$

[4]For any other modulation the input distribution needs to be determined for the quantizer to minimize the MSE.

where we consider that at any given time (t) the inter-cell interference is independent of the transmitted signal, i.e., $\mathbb{E}\left[z_k^H(\mathbf{H}_k\mathbf{v}_\jmath s_\jmath)\right] = 0,\ \forall j \in \mathcal{K}(t)$, and symbols transmitted to different users are independent, i.e. $\mathbb{E}[s_i^* s_\jmath] = 0$. Hence, for point to point links

$$\mathbb{E}\left[(\mathbf{H}_k\mathbf{v}_k s_k)^H\mathbf{H}_\jmath\mathbf{v}_\jmath s_\jmath\right] = 0;\quad \forall j \in \mathcal{K}(t),\ j \neq k. \tag{18}$$

After applying a receiver-side beamforming vector $\mathbf{u_k}$, the channel between the UE and the BS is an effective SISO channel. Define the Rx side BF gain on signal in stream $j \in \mathcal{K}(t)$ as

$$G_\jmath := \mathbb{E}|\mathbf{u}_k^*\mathbf{H}_k\mathbf{v_j}s_\jmath|^2/E_\jmath, \tag{19}$$

which is the ratio of post-beamforming signal energy to the received signal energy per antenna. We note that in (19) if the transmit beamforming vectors are chosen such that $\mathbf{H}_k\mathbf{v_j} = \mathbf{0},\ \forall j \neq k, j \in \mathcal{K}(t)$, then the ICI at user k is zero, but this requires careful beam planning, and scheduling and may not, under certain channel conditions, be even possible. On the other hand, to remove ICI one can set $s_\jmath = 0$, $\forall j \neq k, j \in \mathcal{K}$, which is achieved by orthogonal transmissions schemes (TDMA and OFDMA). Although the latter approach for mitigating ICI simplifies scheduling, there is a loss in the available degrees of freedom.

Observe that, if there is no quantization error ($\alpha = 0$), the beamformed SINR of user k is

$$\begin{aligned}\gamma_k^{\mathrm{BF}} &:= \frac{\rho^2\mathbb{E}|\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_k s_k|^2}{\sigma_n^2 + \sigma_z^2 + \rho^2\sum_{\jmath\neq k}\mathbb{E}|\mathbf{u}_k^H\mathbf{H}_\jmath\mathbf{v}_\jmath s_\jmath|^2} \\ &= \frac{\rho^2 G_k E_k}{\sigma_n^2 + \sigma_z^2 + \rho^2\sum_{\jmath\neq k} G_\jmath E_\jmath} = \frac{\gamma_k'}{1 + \sum_{\jmath\neq k}\gamma_\jmath'}, \end{aligned} \tag{20}$$

where $\gamma_\jmath' = \rho^2\ G_\jmath E_\jmath/(\sigma_n^2 + \sigma_z^2), \forall j \in \mathcal{K}$. Now, with finite resolution quantization, using the AQNM, the signal after beamforming is given by

$$\begin{aligned}y_k^{\mathrm{Q,BF}} := \mathbf{u}_k^H\hat{\mathbf{y}}_k &= \ \rho(1-\alpha)\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_k s_k \\ &\quad + \rho(1-\alpha)\sum\nolimits_{\jmath\neq k}\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_\jmath s_\jmath \\ &\quad + (1-\alpha)\mathbf{u}_k^H\mathbf{z} + (1-\alpha)\mathbf{u}_k^H\mathbf{n} + \mathbf{u}_k^H\mathbf{w}. \end{aligned} \tag{21}$$

Without loss of generality, assuming $\|\mathbf{u}_k\| = 1$ the mean beamformed received signal energy is

$$E_k^{\mathrm{BF}} = (1-\alpha)^2\rho^2\mathbb{E}|\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_k s_k|^2 = \rho^2(1-\alpha)^2 G_k E_k, \tag{22}$$

while the average noise plus interference energy is

$$W_k^{BF} = (1-\alpha)^2\left(\sigma_z^2 + \sigma_n^2 + \rho^2\sum\nolimits_{\jmath\neq k} G_\jmath E_\jmath\right) + \sigma_v^2. \tag{23}$$

Finally, combining (17), (20), (22), and (23), the SINR after beamforming is

$$\begin{aligned}\gamma_k^{\mathrm{Q,BF}} &= \frac{E_k^{\mathrm{BF}}}{W_k^{\mathrm{BF}}} \\ &= \frac{(1-\alpha)\gamma_k'}{1 + (1-\alpha)\sum_{\jmath\neq k}\gamma_\jmath' + \alpha\left(\frac{\gamma_k'}{G_k} + \sum_{\jmath\neq k}\frac{\gamma_j'}{G_j}\right)}. \end{aligned} \tag{24}$$

2) Orthogonal Transmission: With TDMA or OFDMA, (24) simplifies as

$$\gamma_k^{\mathrm{Q,BF}} = \frac{(1-\alpha)\gamma_k^{\mathrm{BF}}}{1+(\alpha/G_k)\gamma_k^{\mathrm{BF}}}, \tag{25}$$

where we use the fact that $\gamma_k^{\mathrm{BF}} = \gamma_k'$ in the absence of ICI. Using (25), we can qualitatively understand the system-level effects of quantization by looking at the following two regimes.

a) Low SNR: In the low-SNR (or SINR) regime, γ_k^{BF} is small, hence

$$\gamma_k^{\mathrm{Q,BF}} \approx (1-\alpha)\gamma_k^{\mathrm{BF}}, \tag{26}$$

i.e., the SINR is decreased only by a factor $1-\alpha$. We show in Sec. V that at moderate quantization levels this has very little impact on system performance.

b) High SNR: In this regime as $\gamma_k^{\mathrm{BF}} \to \infty$

$$\gamma_k^{\mathrm{Q,BF}} \to \frac{G_k(1-\alpha)}{\alpha}. \tag{27}$$

Thus, the effect of quantization is to saturate the maximum SINR i.e., the effect of finite quantization is critical only at high SNR (or SINR). In Sec. V we show that even for $3-4$ bits of resolution, this effect at the high SINR limit is not significant for cellular systems.

3) SDMA Transmission: Comparing (24) with (25) we note that for SDMA transmission the effect of quantization noise is further enhanced by the presence of ICI. In (24), we note that

$$G_j = \frac{\mathbb{E}\|\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_j s_j\|^2}{E_j} = N_{\mathrm{UE}}^{\mathrm{ant}}\frac{\mathbb{E}\|\mathbf{u}_k^H\mathbf{H}_k\mathbf{v}_j s_j\|^2}{\mathbb{E}\|\mathbf{H}_k\mathbf{v}_j s_j\|^2}. \tag{28}$$

Under the assumption that $\|s_j\|^2 = 1$, using the Kronecker model [39] we obtain

$$\mathbb{E}\|\mathbf{u}_k^*\mathbf{H}_k\mathbf{v}_j s_j\|^2 = \frac{1}{N_{\mathrm{UE}}^{\mathrm{ant}}}\left(\mathbf{u}_k^H\mathbf{Q}_k^{\mathrm{rx}}\mathbf{u}_k\right)\left(\mathbf{v}_j^H\mathbf{Q}_k^{\mathrm{tx}}\mathbf{v}_j\right),$$

where $\mathbf{Q}_k^{\mathrm{rx}} = \mathbb{E}|\mathbf{H}_k\mathbf{H}_k^H|$, and $\mathbf{Q}_k^{\mathrm{tx}} = \mathbb{E}|\mathbf{H}_k^H\mathbf{H}_k|$ are the receive and transmit covariance matrices respectively for the channel between the BS and user k, and similarly,

$$\mathbb{E}\|\mathbf{H}_k\mathbf{v}_j s_j\|^2 = \left(\mathbf{v}_j^H\mathbf{Q}_k^{\mathrm{tx}}\mathbf{v}_j\right). \tag{29}$$

Thus we can rewrite (28) as

$$G_j = \mathbf{u}_k^H\mathbf{Q}_k^{\mathrm{rx}}\mathbf{u}_k = G_k,\ \forall j \in \mathcal{K}, \tag{30}$$

which implies that any signal received by the k-th UE from the associated BS will have the same receiver-side beamforming gain. The presence of ICI in SDMA systems essentially limits the maximum achievable SINR. From (29), we note that although ICI can be eliminated by selecting transmit beamforming vectors such that $\left(\mathbf{v}_j^H\mathbf{Q}_k^{\mathrm{tx}}\mathbf{v}_j\right) \to 0$, $\forall j \neq k$, this is a hard problem in practice and in many cases a solution may not exist. Thus, in our analysis we assume that ICI is always present for SDMA based systems.

For this, we simplify (20) by representing the ICI term as

$$\sum\nolimits_{j\neq k}\gamma_j' = \psi\gamma_k', \quad \psi \geq 0, \tag{31}$$

where $\frac{1}{\psi}$ is the signal to interference ratio (SIR), and as $\gamma_k' \to \infty$, $\gamma_k^{\mathrm{BF}} \to \frac{1}{\psi}$. Observe that, for SDMA systems even in the absence of quantization noise the SINR saturates due to ICI.

Effect of low resolution quantization: Using (30), and (31) in (24) we can express the SINR in the presence of quantization noise and ICI as

$$\gamma_k^{\mathrm{Q,BF}} = \frac{(1-\alpha)\gamma_k'}{1+(1-\alpha)\psi\gamma_k' + (\psi+1)\frac{\alpha}{G_k}\gamma_k'}. \tag{32}$$

Furthur as in (25) we can rewrite (32) as

$$\gamma_k^{\mathrm{Q,BF}} = (1-\alpha\beta)\gamma_k^{\mathrm{BF}}, \tag{33}$$

where

$$\beta = \frac{1+(\psi+1)\gamma_k'/G_k}{1+(1-\alpha)\psi\gamma_k' + \alpha(\psi+1)\gamma_k'/G_k}. \tag{34}$$

Finally, from (34) we note that $\beta < \frac{1}{\alpha}$ is always satisfied under the AQNM (note that $\beta > \frac{1}{\alpha} \Leftrightarrow \alpha > 1$, which is not admissible). It follows from (34) that for $\beta < 1$ we must have $G_k > 1 + \frac{1}{\psi}$, and hence,

$$\gamma_k^{\mathrm{Q,BF}} > (1-\alpha)\gamma_k^{\mathrm{BF}},\ \text{if } G_k > 1 + \frac{1}{\psi}. \tag{35}$$

We note that when ψ is *large*, i.e., when ICI dominates, it is easy to satisfy the inequality in (35). SINR degradation in this regime is dominated by the large ICI and *not* quantization noise. On the other hand, as $\psi \to 0$ the analysis for the orthogonal transmissions hold.

D. Summary

To summarize our analysis in this section, for multi-user communications with long-term beamforming and low resolution ADCs, we state the following:

1) For orthogonal transmission (i.e. FDMA or TDMA) in low SNR/SINR regime there is very little or effectively no loss due to low resolution quantization.
2) For orthogonal transmission in high SNR or SINR regimes there exists a saturation of the effective SNR or SINR due to quantizer resolution.
3) For transmission schemes where orthogonality within the cell is not always guaranteed (e.g., SDMA), the degradation in SINR is dominated by the ICI.

In Sec. V we will validate these claims through extensive simulations.

IV. Signal Transmission With Low Resolution DAC

The previous section analyzed quantization noise at the Rx. In this section, we look at the Tx. Low resolution quantization at the Tx can result in quantization noise being present both in-band and out-of-band. For cellular communication systems, regulatory bodies like 3GPP specify signal characteristics that all cellular transmitters need to comply by. These are (a) output power, (b) the adjacent carrier leakage ratio (ACLR), and (c) transmitted signal quality specified by the error vector magnitude (EVM). The output power is not affected by the addition of quantization noise but the EVM of the Tx increases due to the addition of this noise source. Moreover, the quantization noise is not band limited as opposed to the signal and contributes to an increase in the power leaked

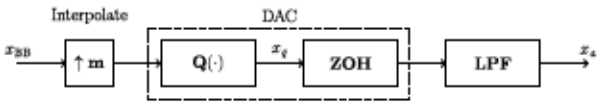


Fig. 4. High level model of a digital to analog converter at baseband. The DAC is clocked at f_s where $f_s = mf_{\mathrm{BB}}$.

into the adjacent carrier. The out-of-band distortion is most constraining as, for cellular communications systems, 3GPP specifies signal characteristics that all commercial transmitters must comply with. In this section, we show the effects of low resolution DACs on the transmitted signal, and in the process determine the precision required by the DACs in the mmWave transmitter to meet 3GPP regulations.

A. Model of the DAC

To model the effect of low resolution on the transmit signal, we first describe a model for the DAC. A DAC comprises of a quantizer $Q(\cdot)$, and the zero order hold (ZOH) circuit. An output analog low pass filter (LPF), as shown in Fig. 4, is used to attenuate the spectral images of the signal located at an interval of f_s, where f_s is the sampling rate of the DAC. The sampling rate must be at least $f_s = f_{\mathrm{BW}}$, where f_{BW} is the bandwidth of the input signal x_{BB}. In most designs, like [40], $f_s = m \times f_{\mathrm{BW}}$, with $m > 1$ so that the spectral images formed at the output of the DAC are spaced sufficiently apart in frequency. Thus the base band signal is interpolated, i.e., upsampled by m and filtered, before it is converted to analog. The interpolator not only relaxes the design of the analog LPF but also minimizes the distortion caused by the ZOH on the in-band component. Additionally, upsampling the signal by a factor of m also reduces the power of the quantization noise by the same factor, i.e., a $m = 2$ interpolation of the baseband signal will lead to lowering the quantization noise by 3 dB. This gain, although, comes at the cost of doubling the sampling rate f_s of the DAC which, from (9), doubles the power consumed. As pointed out in [40], high over sampling is not practical for wide-band systems as power consumption increases linearly with sampling rate.

B. Adjacent Carrier Leakage

From our analysis in Section III-C, quantization noise for low resolution converters can be modeled as white Gaussian noise. This implies that the quantization noise has a flat spectrum while the signal of interest is band limited. This is problematic for practical systems as this noise causes unwanted interference in the adjacent bands. For interoperability, cellular transmissions need to limit the amount of power leaked into the adjacent bands. This restriction is quantified by 3GPP using the *adjacent carrier leakage ratio* (ACLR) defined as,

$$\mathrm{ALCR} = 10\log_{10}\left(\frac{P_{\mathrm{in}}}{P_{\mathrm{ac}}}\right), \tag{36}$$

where P_{in} is the total power in the transmission channel, and P_{ac} is the total power accumulated over a given adjacent channel.

Well known effective techniques, like windowed overlap and add (WOLA) OFDM [41], are used to reduce the ACLR in LTE systems. These techniques focus on reducing the inter-modulation products and operate on the signal in the digital domain. Quantization noise due to finite resolution DACs, however, is introduced in these signals after digital processing. Thus techniques like WOLA have no effect on the quantization noise. A classical method of dealing with quantization noise is the use of $\Delta\Sigma$ feedback structures. Recent works on Rx beamforming like [42] have considered such circuits for low resolution receivers to "clean up" the in band signal. Although attractive, as pointed out in [43] and [44], such techniques rely on circuits that eliminate matching and timing errors, which increase the power consumption considerably. Moreover, $\Delta\Sigma$ modulators also require the DAC to operate with a high oversampling ratio, which further increases the power consumption.

Thus, the only practical option to control the ACL due to quantization noise is by imposing stricter restrictions on the analog LPF which may imply that higher order filters will be required when low resolution DACs are used at the transmitter. Thus, the filters will either take more space on the chip (when they are passive) or consume more power (for active CMOS filters). In Sec. V-C we perform extensive simulations to determine the order of the LPF that meets the ACLR requirements at the Tx. More importantly, in the sequel we show that for moderately low resolution DACs, no additional restrictions are imposed on the analog LPF.

C. Transmitted Signal Quality

As in [45], we quantify the transmitted signal quality by its EVM. Intuitively, the EVM captures the error in the modulated symbol produced due to Tx impairments. It is considered a key factor in determining the maximum modulation order a transmitter can faithfully support. The EVM, ϵ, for a Tx signal is given as

$$\epsilon^2 = \frac{\mathbb{E}_{t,f}|Z(t,f) - I(t,f)|^2}{\mathbb{E}_{t,f}|I(t,f)|^2}, \tag{37}$$

where $Z(t,f)$ and $I(t,f)$ are the received symbol and ideal modulation symbol respectively at time t and sub-carrier f. From (37) it is clear that lower the value of ϵ the cleaner is the transmitted signal, and ϵ must be small in order to support high order modulations as these are more sensitive to distortions.

The signal impairments introduced by the mmWave RFFE, including the LO phase noise, LO leakage, I-Q imbalance, etc. can be modeled as an AWGN noise source following the work in [22]. We represent the RF impairments as zero mean complex Gaussian random variable $n_{\mathrm{RF}} \sim \mathcal{CN}(0, \sigma^2_{\mathrm{RF}})$. Based on the AQNM in (14), we can re-write (37) as,

$$\epsilon^2 = \alpha^2 + \frac{\sigma^2_{\mathrm{RF}} + \sigma^2_{\mathrm{v}}}{\mathbb{E}|I(t,f)|^2}, \tag{38}$$

where σ^2_{v} is the variance of the quantization noise. Thus, the presence of quantization noise effectively limits the EVM from going to 0 even when $\sigma_{\mathrm{RF}} \to 0$. Therefore, low resolution quantizers essentially limits the maximum spectral

TABLE II
OFDM PARAMETERS FOR LINK LEVEL SIMULATIONS

Parameter	Value
Channel bandwidth ($f_{\text{BW}}^{\text{ch}}$)	400 MHz
FFT size (N_{fft})	4096
Subcarrier spacing	120 kHz
OFDM chip rate (f_{chip})	491.52 MHz
Subcarriers per PRB	12
Max. PRBs used ($N_{\text{PRB}}^{\max}$)	275
Symbol duration	10.67μs.

efficiency that can be achieved by limiting the highest modulation order that can be supported by a transmitter. This is crucial for the utilization of the large bandwidths available at mmWave frequencies.

In Sec. V-C, we perform extensive simulations to analyze the effect of low resolution DACs on the transmitted signal quality. We show that low resolution DACs can be used for mmWave transmitters under the 3GPP specified limits on ACLR and EVM. A theoretical analysis of the trade-off between the in-band signal quality and ACL is shown in [36]. In this work we look at the practical limits of low resolution DACs.

V. SIMULATIONS AND RESULTS

In this section we present our results obtained through link level and cellular simulations. Firstly, we verify the AQNM presented in Section III-C through a series of link layer OFDM simulations. Next, we use a multi-cell multi-user simulation at 28 GHz to study the effect of low resolution quantization and multiple access schemes on link quality and throughput. Finally, we investigate the effect of low resolution DACs on the transmitted signals.

A. Verification of the AQNM

To verify the proposed AQNM for low resolution converters, we use a link level OFDM simulator. The simulation parameters, given in Table II, are from the 3GPP NR standards [46]. We consider a wide band AWGN channel with a receiver structure shown in Fig. 5.

The ADC is modeled as a finite resolution scalar quantizer. The input to the quantizer are infinite precision discrete time samples. These are mapped to one of the finite number of quantization bins. The AGC ensures that the input to the quantizer has a unit variance. A digital finite impulse response (FIR) LPF is used at the output of the ADC to remove the out of band noise. Here, we note that practical transceivers will also employ analog LPFs before the ADC or the AGC to attenuate signals on the adjacent carriers. As we do not model adjacent carrier blocking, in this simulation we omit the analog filtering.

Orthogonal Transmission: Fig. 6 compares the effective SNR predicted by the AQNM with the simulated post-equalization SNR, for varying quantization levels (n). The value of α is computed assuming an optimal uniform n-bit quantizer [13]. From Fig. 6a, we see that the AQNM very

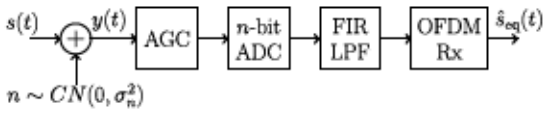


Fig. 5. Simulation model for Rx front end with low resolution ADC modeled as a n-bit scalar quantizer.

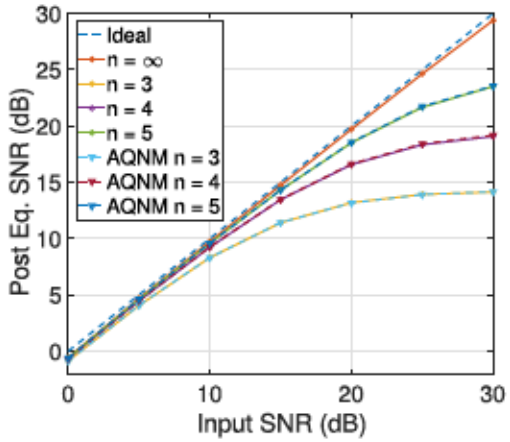
accurately predicts the signal degradation due to finite quantizer resolution. This shows that the linear AQNM accurately capture the effect of the non-linear low resolution quantization. Here we assume that the Tx uses an $n+2$-bit DAC, and hence, quantization noise added by the Tx is 6 dB lower than that at the Rx. We observe that finite quantization has the effect of saturating the effective SNR in the high SNR regime. On the other hand, with $n > 3$, and input SNR $\leq$ 15 dB, the effect of quantization noise on the system is negligible. In the case when both the DAC at the Tx and the ADC at the Rx have n-bits of resolution, the quantization noise power doubles. This is observed in Fig. 6b, where in the high SNR regime, the simulated curve is nearly 3 dB lower than the SNR predicted by (25).

More interestingly, comparing Fig. 6b and Fig. 6c, we observe the effect of *oversampling* on quantization noise. OFDM systems generally have an OFDM chip rate slightly higher than the signal bandwidth. For instance in 3GPP NR, for a 400 MHz channel the OFDM chip rate is 491.52 MHz. In the high SNR regime, in the presence of quantization noise, for a system with n-bit quantizers at both Tx and Rx, we can express the SNR as,

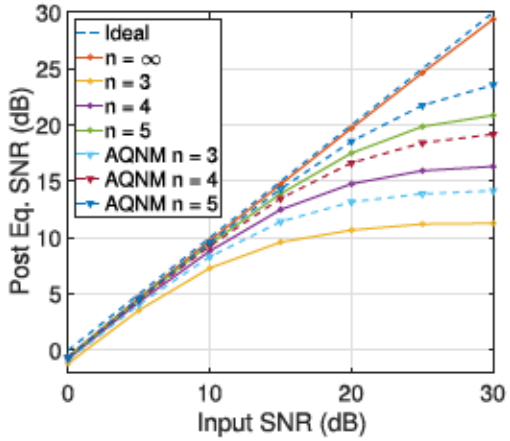
$$\begin{aligned}\text{SNR (lin.)} &= \frac{P_{\text{sig}}}{N_0 + 2\sigma_q^2/\text{OSR}};\\ \text{SNR (dB)} &\approx P_{\text{sig}}(\text{dB}) + 3\\ &\quad + 10\log_{10}(\sigma_q^2) + 10\log_{10}\left(\frac{N_{\text{fft}}}{N_{\text{sc}}}\right), \end{aligned} \tag{39}$$

where $\text{OSR} = \frac{N_{\text{fft}}}{N_{\text{sc}}}$ is the oversampling ratio. The effect of oversampling is demonstrated in Fig. 6b and Fig. 6c where by using only 200 PRBs as opposed to 274 we increase the OSR from 0.95 dB to 2.32 dB. This points to an interesting trade-off. When the system employs low resolution quantizers, under good channel conditions, oversampling by using a smaller part of the bandwidth can reduce the quantization noise at the cost of spectral efficiency but without an increase in power consumption. In fact, in the high SNR regime, reducing the allocated bandwidth can enable the use of high modulation and coding schemes.

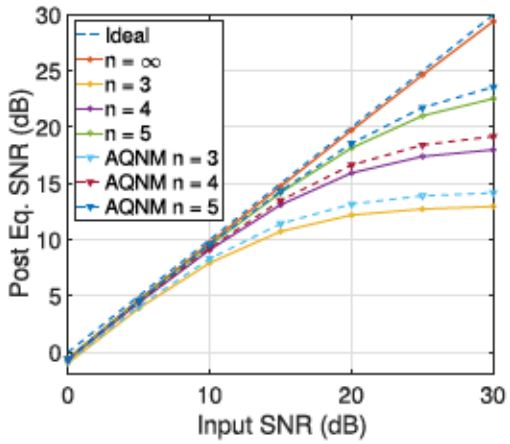
SDMA: Next we turn our attention to the effect of quantization noise on systems with in-band ICI. Such scenarios are interesting especially when large arrays are available at the Tx and multiple users can be scheduled on multiple transmit beams simultaneously with all the available bandwidth. In practical scenarios, transmitting on multiple orthogonal beams may not be possible due to the nature of the multi-user channel. In such cases, ICI becomes dominant in the system. In Fig. 7, we plot the post-equalization SINR versus the SIR ($1/\psi$) for a given SNR (γ_0). As predicted by (35), when the interference is high (low SIR), without any assumption on processing gains (G_k) at the receiver, there is



(a) $n_{\mathrm{DAC}} = n_{\mathrm{ADC}} + 2 = n + 2$; $N_{\mathrm{PRB}} = 274$



(b) $n_{\mathrm{ADC}} = n_{\mathrm{DAC}} = n$; $N_{\mathrm{PRB}} = 274$



(c) $n_{\mathrm{ADC}} = n_{\mathrm{DAC}} = n$; $N_{\mathrm{PRB}} = 200$

Fig. 6. Post-equalization SNR as a function of input SNR for varying quantization levels. Fig. (6a) shows the accuracy our proposed AQNM with link level OFDM simulation based on 3GPP NR specification under assumption of orthogonal transmission. Fig. (6b) and (6c) show the effect of oversampling on quantization noise.

effectively no degradation due to quantization noise. The effect of quantization only becomes perceivable when SIR is high as discussed in III-C.3. As shown in Fig. 7, in noise-limited scenarios ($\gamma_0 = 0$ dB), even at high SIR, the degradation due

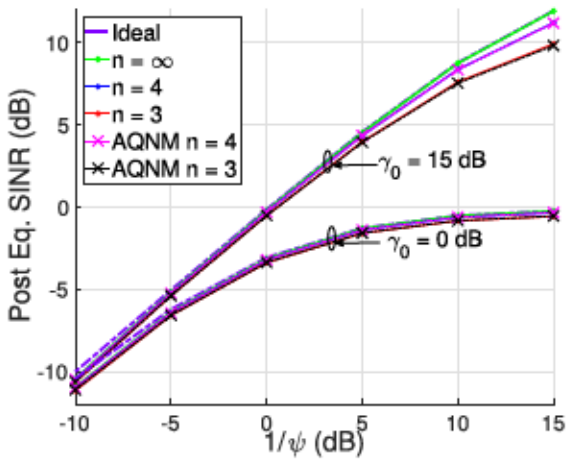


Fig. 7. Post-equalization SINR as a function SIR ($1/\psi$) with varying quantization levels for a given SNR (γ_0). Results validate the proposed AQNM using link level OFDM simulation based on 3GPP NR specification in the presence of ICI.

to quantization noise is less than 0.5 dB. Substantial loss in link quality is only observed when both SIR and SNR (γ_0) are high, as seen in Fig. 7 for $\gamma_0 = 15$ dB. In this case, receivers with 3-bit ADCs have a 2 dB loss in the SINR due to quantization noise, while for 4-bit ADCs the loss is less than 1 dB.

B. Multi-Cell Multi-User Simulations

We apply our link layer AQNM to understand the effect of low resolution quantization on the DL system capacity. We simulate a 1Km by 1Km area covered by hexagonal cells of radius 100 m. Each cell is assumed to serve on average 10 UEs which are randomly "dropped". We then compute a random path loss between the BS and the UEs based on the urban mmWave channel model presented in [25]. We simulate a DL transmission scenario where BSs transmit a single stream to every user. Both BSs and UEs are assumed to perform long-term digital beamforming [47] making use of the spatial second-order statistics of channel. For our simulations, we assume that the BSs always have data to send to every UE (full-buffer assumption). The relevant parameters for our simulations are summarized in Table III.

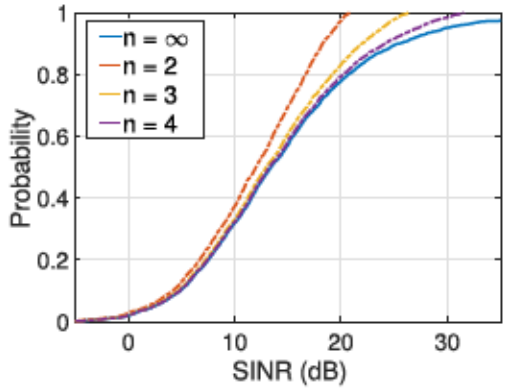
OFDMA: For OFDMA-based cellular systems, within one transmission time interval (TTI), each UE is assigned a non-overlapping part of the total bandwidth by the associated BS. Each link gets full beamforming gain but only uses a part of the total bandwidth. Orthogonalization in frequency eliminates ICI but also limits the maximum achievable rate. To study such a system in a practical setting, we employ a proportional fair scheduling algorithm for medium access control. At the T-th TTI, the k-th UE associated to BS j is assigned a weight,

$$\mu_{j,k}^{T} = \lambda_{j,k}^{T} \Big/ \sum_{t=0}^{T} r_t \tag{40}$$

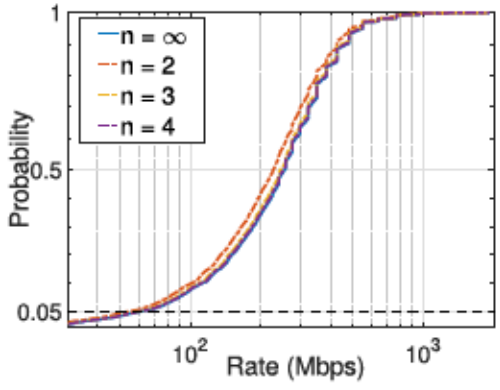
where r_t is the data that had been scheduled at the t-th TTI, and $\lambda_{j,k}^{T}$ is the spectral efficiency of the link at the T-th TTI. It is assumed that the BS does not have any information about

TABLE III
MULTI-CELL SIMULATION PARAMETERS

Parameter	Value
Cell radius	100 m
Carrier frequency	28 GHz
Pathloss model	[25]
DL bandwidth (B_{tot})	1 GHz
DL Tx power	35 dBm
Rx noise figure	8 dB
Max. spectral efficiency	7.4063 b/s/Hz
BS antenna array	8×8 uniform planar
UE antenna array	4×4 uniform planar
BF mode	Digital long-term, single stream
Transmission time interval	$125\mu s$
Control overhead	20%
Traffic Model	Full buffer



(a) CDF of SINR



(b) CDF of Rate

Fig. 8. Millimeter wave DL multi-user simulations showing the effect of low resolution ADCs on the link quality and achieved rates for an OFDMA system. For $n \geq 3$ the loss due to quantization becomes negligible since very few users operate at a sufficiently high SNR for quantization noise to have noticeable impact.

the quantization noise at the Rx. The weights are normalized as $\mu^T_{j,k} = \mu^T_{j,k} / \sum_k \mu^T_{j,k}$, and each UE k associated with the BS j is assigned a bandwidth of $B^T_{j,k} = \mu^T_{j,k} \times B_{\text{tot}}$.

In Fig. 8a we plot the cumulative distribution function of the DL SINR for systems with $n = 2, 3$ and 4 bits of quantization at the Rx. For comparison, we also plot the case when infinite ADC resolution is available ($n = \infty$) at the Rx. We observe that at low SINR, the deviation from the $n = \infty$ curve is minimal if any. On the other hand, at high SINR regimes we observe a "clipping" of the maximum achievable SINR. More specifically, the SINR penalty for 2 bit quantization is nearly 10 dB for the 90-th percentile UE. For a 3-bit Rx, the 90-th percentile UEs have less than 5 dB of loss in SINR. In the 50-th percentile on the other hand, this difference goes down to about 2.5 and 1 dB for 2 and 3-bits of resolutions respectively.

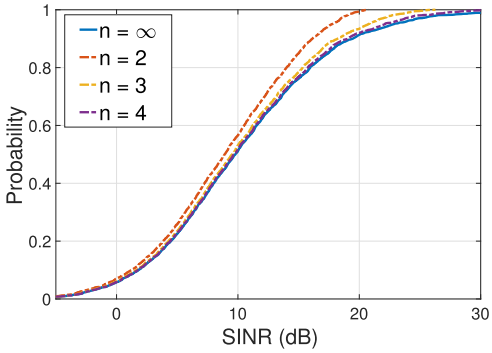
We next plot the achievable rates under various quantizer resolutions in Fig. 8b. Following the analysis in [25] and the link-layer model [48], we assume a 3 dB loss from Shannon capacity, and a 20% overhead. A maximum spectral efficiency of $\lambda_{\max} = 7.4063$ b/s/Hz is assumed based on the 256 QAM modulation scheme proposed in 3GPP NR standards [49]. For 3 or more bits of ADC resolution, there is no loss in the DL rates of the users. This is because very few users in the system will operate at high SINR, thus the clipping of SINR as observed in Fig. 8b has little effect on the average rate. Moreover, as rate is a logarithmic function of the SINR, increasing the SINR beyond a certain point produces diminishing increase in the rate, more so with a limit on the maximum spectral efficiency. Further, we observe that under a full-buffer assumption, TDMA and OFDMA will achieve the same rates with full buffer assumption. But, OFDMA is only possible through digital BF, and is more efficient for low latency transmission of short mission critical data packets. Analog or hybrid BF based systems have to rely on TDMA where transmission of short packets can be wasteful in terms of radio resource utilization and increase transmission latency [18].

SDMA: Following our discussion in Sec. III-B, for SDMA systems, each BS assigns the entire bandwidth to k users at any given TTI. The number of users scheduled at each TTI depends on the maximum number of simultaneous beams supported by the system $N^{\max}_{\text{beam}}$ and the multi-user channel condition. A simple scheduler based on proportional fair selection and sum rate maximization is used to demonstrate our results. At the T-th TTI, the j-th BS will select a group of UEs $\mathcal{K}(T)$ where $|\mathcal{K}(T)| \leq N^{\text{beams}}_{\text{BS}}$. The first UE ($k = 1$) is selected into the scheduled group $\mathcal{K}(T)$ such that,

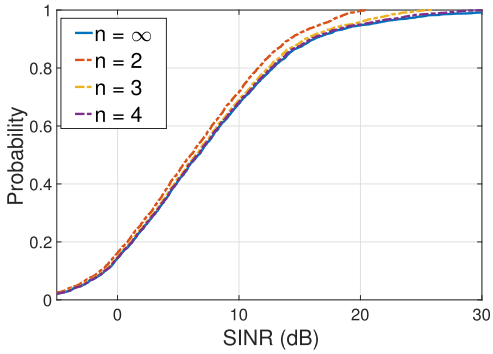
$$\mu^T_{j,1} = \max_k \mu^T_{j,k}, \tag{41}$$

where $\mu^T_{j,k}$ is computed using (40). Next, the j-th BS will admit users to the scheduled group if the achievable sum rate of the BS increases by that admission. The BS will stop admitting users to the scheduled group when either $N^{\text{beams}}_{\text{BS}}$ UEs are scheduled or given the associated UEs and their channel quality, no UE can be added to the group such that the sum-rate increases.

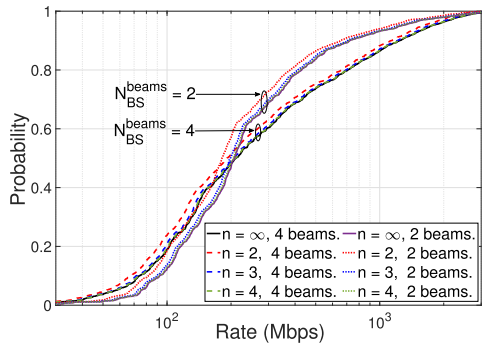
We present our results for a SDMA system with $N^{\text{beams}}_{\text{BS}} = 2$ and 4 in Fig. 9. We notice the effect of ICI on the system by comparing Fig. 9a with Fig. 8a. In the absence of quantization noise, we observe that with 2-stream SDMA around 5% of the users have a SINR less than 0 dB; with OFDMA less than 1% of the UEs have SINR < 0 dB. Also, the median SINR in the SDMA case is 10 dB for $N^{\text{beams}}_{\text{BS}} = 2$, and 7 dB for $N^{\text{beams}}_{\text{BS}} = 4$; a 3 and 6 dB additional loss compared to OFDMA. That is, due the presence of ICI in SDMA, UEs with poor channel conditions tend to perform



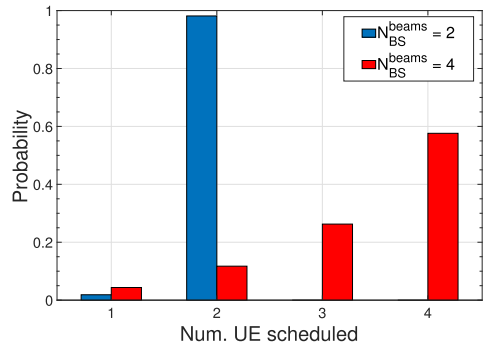
(a) CDF of SINR for $N_{\mathrm{BS}}^{\mathrm{beams}} = 2$



(b) CDF of SINR for $N_{\mathrm{BS}}^{\mathrm{beams}} = 4$



(c) CDF of rate



(d) Distribution of the number of UEs scheduled by each BS per TTI

Fig. 9. Millimeter wave DL multi-user simulations showing the effect of low resolution ADCs on the link quality and rates for SDMA with $N_{\mathrm{BS}}^{\mathrm{beams}} = \{2, 4\}$ spatial streams.

worse. In spite of the presence of ICI, like OFDMA, the effects of low resolution quantization is only noticeable in the high SINR regimes as evident from Figs. 9a and 9b. We note that due to ICI, the beamformed DL SINR rarely exceed 30 dB and hence, with $n = 4$ or more bits of resolution, there is no noticeable loss in link quality. Moreover, with multi-stream SDMA, rates greater than 1 Gbps can be achieved for the top 7% and 14% users with $N_{\mathrm{BS}}^{\mathrm{beams}} = 2$, and 4 respectively as shown in Fig. 9c. This is a considerable improvement over OFDMA where less than 1% of the users achieved rates exceeding 1 Gbps. More importantly, notice that there is a very little penalty in the DL rate due to quantization noise. Especially with $n = 3$ or more bits of resolution, the effect of quantization noise on the DL rate is negligible.

Interestingly, in Fig. 9d we observe that with $N_{\mathrm{BS}}^{\mathrm{beams}} = 2$, in more than 95% of the scheduling instances, the maximum possible number of beams are used. On the other hand when $N_{\mathrm{BS}}^{\mathrm{beams}} = 4$, we see that less than 60% of the time are all the beams are used. Based on this observation, we draw the following conclusions. To fully utilize the available spatial degrees of freedom offered by fully digital beamforming for data transmissions, sophisticated scheduling algorithms will be necessary. Yet, even when a large number of users are scheduled simultaneously, the effect of low resolution ADCs at the receiver is negligible for SDMA. Moreover, non-orthogonal multiple access (NOMA) schemes [50], [51] has recently gained considerable interest and can be investigated to better utilize the degrees of freedoms offered by the digital beamformer. Although, a study of NOMA considering a few bits of ADC resolution is not found in current literature and is an interesting avenue of future research.

C. Transmitter Characteristics

Finally, we look into the effects of a low resolution DAC on the transmitted signal quality. As discussed in Sec. IV, the quantization noise not only corrupts the transmitted signal but also increases the leakage into the adjacent channels. This is evident from the power spectral density (p.s.d) of the transmitted signal plotted in Fig. 10. The signal is transmitted over a $f_{\mathrm{BW}}^{\mathrm{ch}} = 400$ MHz channel around $f_c = 28$ GHz. The adjacent channels are $f_{\mathrm{BW}}^{\mathrm{ch}}$ wide and located around $f = f_c \pm n f_{\mathrm{BW}}^{\mathrm{ch}}$, $n = 1, 2, \ldots$. The DAC operates at a sampling frequency of $f_s = 2 \times f_{\mathrm{chip}} = 983.04$ MHz. The ACLR is measured over the maximum occupied bandwidth of $f_{\mathrm{BW}}^{\mathrm{meas}} = 396$ MHz.

In Fig. 10 we can note two important points. Firstly, the quantization noise considerably increases the leakage in the adjacent channels. For instance, the leakage is nearly 40 dB higher for $n = 4$ compared to the $n = \infty$. Secondly, the LPF serves two crucial purposes. Not only does it remove the spectral images introduced by the DAC, but it also attenuates the out-of-band quantization noise. In fact, with $n = \infty$, adjacent channel 2 in Fig. 10, has a very low ACLR due to the presence of the spectral image. In Fig. 11 we plot the ACLR on adjacent channel 1 and 2 versus the LPF order when the LPF is modeled as a Butterworth filter. For 5G mmWave systems, the 3GPP standards [20], [45] specify the

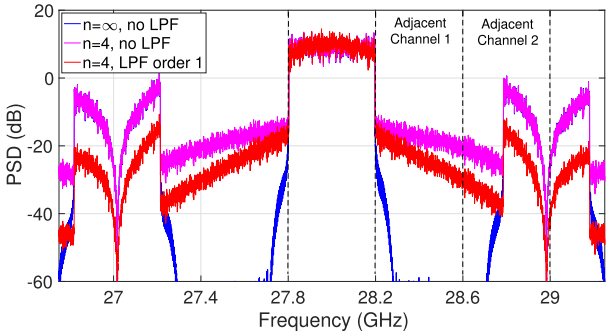
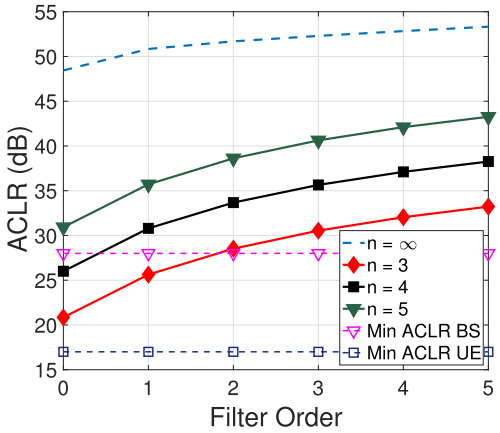
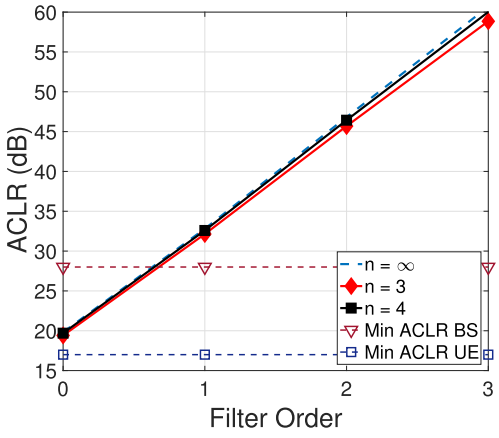


Fig. 10. Transmit p.s.d of a 400 MHz OFDM signal using a DAC sampled at $f_s = 983.04$ MHz centered at $f_c = 28$ GHz. Observe the effects of quantization noise on the adjacent channels centered at 28.4 and 28.8 GHz where a LPF of order 1 has considerable impact on leakage attenuation.



(a) ACLR on adjacent channel 1.



(b) ACLR on adjacent channel 2.

Fig. 11. ACLR versus filter order for Butterworth LPF measured over adjacent channels 1 and 2. Filter order 0 implies the absence of a LPF at the DAC output port. Dashed horizontal lines are the ACLR requirements of the BS and the UE transmitter specified by 3GPP NR specifications for mmWave frequencies (frequency range 2).

ACLR to be 28 dB and 17 dB for BSs and UEs respectively. From Fig. 11a we see that for $n \geq 4$, an order-1 Butterworth filter is sufficient to meet the ACLR requirements at the BS. Moreover, from Fig. 11b, we note that to achieve acceptable ACLR over adjacent channel 2, i.e., to have sufficient image rejection, the BS Tx requires at least an order-1 Butterworth LPF irrespective of the resolution of the DAC.

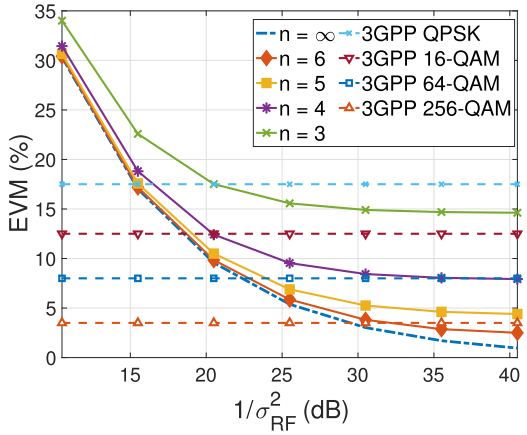


Fig. 12. EVM vs. RF impairment comparing the performance of finite resolution DACs with a first order Butterworth LPF at the output. Dashed lines represent the minimum EVM requirements for different modulation orders.

Thus we make a key observation that at the BS, to knock out spectral images, a Butterworth LPF of order-1 is necessary regardless of the DAC resolution. Furthermore, this order-1 LPF is also sufficient to attenuate the out of band quantization noise below the level specified in [45] with $n \geq 4$ bits of DAC resolution. Thus, for a BS Tx, the out of band emissions due to finite quantization can be sufficiently attenuated *without an increase in the hardware complexity* when compared to the infinite resolution case. Moreover, as shown in Sec. II, low order active analog filters consume very little power in the current state of the art. For UEs, the ACLR requirements [20] are met for $n \geq 3$ *without any assumption on filtering* for both adjacent channels 1 and 2. Hence, low resolution DACs can be used on UE FEs possibly without any analog LPF. For fully digital systems, this implies a saving in power or chip area for UEs.

To conclude our study on low resolution DACs, in Fig. 12 we plot the EVM versus $1/\sigma^2_{RF}$ representing the Tx signal to RF impairments ratio as discussed in Sec. IV-C when the Tx power is normalized to unity. From Fig. 12 we note that 4-bits of resolution are sufficient to support modulation orders up to 64-QAM which, as specified in [49], is the highest modulation order that UEs need to support. Thus, 4 bits of resolution is sufficient for DACs used in UEs. For the BS transmitter, using $n \leq 5$ implies that 256-QAM cannot be supported due to quantization noise. This does not violate specification, as 256-QAM is an optional feature in current NR standards, but limits the maximum achievable spectral efficiency on the DL. Thus, to fully exploit the large bandwidth at mmWave, the BS transmitter will need DACs with at least 6-bits of resolution. This is a feasible design choice for BSs as they operate with higher power budgets than UEs.

VI. Conclusion

Fully digital beamforming at mmWave requires the use of low resolution converters to keep the power consumption of the front ends reasonable. The gain in spatial multiplexing offered by the fully digital architecture is achieved at a cost of signal degradation due to coarse quantization. In this paper we have determined how many bits of resolution is required for power efficient communications over wide band mmWave

channels. We show that at the Rx, the loss due $3-4$ bits of ADC resolution is negligible for practical cellular deployment scenarios. Additionally, we show that mmWave receivers with $3-4$ bits of ADCs precision can achieve multi-Gbps rates when SDMA scheduling is used for medium access. For transmitters, low resolution DACs, with 4 or more bits of precision, meet the 3GPP transmission regulations on ACL without any additional hardware costs both at the BS and and the UE. Further, we show that the EVM required for the transmission of 64-QAM is met by 4-bit DACs while 6-bits of DAC resolution is required to support 256-QAM. This implies that 4-bits of resolution is sufficient at the UE Tx while 6-bits of resolution will be required for BS DACs if the 256-QAM is to be supported. Thus, low resolution fully digital beamforming can be used both at the receiver and the transmitter of wide band mmWave cellular equipments.

Indeed, beamforming at the digital baseband is more accurate and power-efficient than RF beamforming. Digital beamformers have higher beam resolution, and can simultaneously resolve multiple beams. This is possible as fast and numerically efficient algorithms are well suited for digital circuitry. These operations at the digital baseband are more power efficient and have considerably lower errors compared to computations using RF components. Nevertheless, a careful assessment of the cost and power consumption of the baseband processor warrants a careful assessment for practical implementation especially for multi-antenna systems. Further, communication protocols for mmWave cellular systems have been conventionally designed with the assumption of analog/hybrid beamforming. We believe that digital beamforming can considerably simplify and enhance both control and data plane signaling. A careful consideration of the potential and practicality of fully digital beamforming, thus, will be key for ultra-reliable low latency communications over the mmWave bands.

REFERENCES

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] B. Sadhu *et al.*, "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*, vol. 52, no. 12, pp. 3373–3391, Dec. 2017.

[3] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[4] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[5] X. Zhang, A. F. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

[6] J. H. C. van den Heuvel, J.-P. M. G. Linnartz, P. G. M. Baltus, and D. Cabric, "Full MIMO spatial filtering approach for dynamic range reduction in wideband cognitive radios," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 11, pp. 2761–2773, Nov. 2012.

[7] J. Mo and R. W. Heath, Jr., "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[9] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.

[10] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[11] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[12] A. Mezghani and A. L. Swindlehurst, "Blind estimation of sparse broadband massive MIMO channels with ideal and one-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2972–2983, Jun. 2018.

[13] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 191–198.

[14] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath, Jr., "Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2274–2287, Apr. 2017.

[15] C. N. Barati *et al.*, "Directional cell discovery in millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.

[16] C. N. Barati *et al.*, "Initial access in millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7926–7940, Dec. 2016.

[17] M. Giordani *et al.*, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, Sep. 2018.

[18] S. Dutta, M. Mezzavilla, R. Ford, M. Zhang, S. Rangan, and M. Zorzi, "Frame structure design and analysis for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1508–1522, Mar. 2017.

[19] S. Dutta, C. N. Barati, A. Dhananjay, and S. Rangan, "5G millimeter wave cellular system capacity with fully digital beamforming," in *Proc. 51st Asilomar Conf. Signals, Syst. Comput.*, Oct. 2017, pp. 1224–1228.

[20] *User Equipment (UE) Radio Transmission and Reception: Part 2: Range 2 Standalone, Release 15*, document TS 38.101-2 2018.

[21] R. Garg and A. S. Natarajan, "A 28-GHz low-power phased-array receiver front-end with 360 ° RTPS phase shift range," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4703–4714, Nov. 2017.

[22] A. K. Gupta and J. F. Buckwalter, "Linearity considerations for low-EVM, millimeter-wave direct-conversion modulators," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 10, pp. 3272–3285, Oct. 2012.

[23] I. Song *et al.*, "A simple figure of merit of RF MOSFET for low-noise amplifier design," *IEEE Electron Device Lett.*, vol. 29, no. 12, pp. 1380–1382, Dec. 2008.

[24] Z. Chen, H. Gao, D. Leenaerts, D. Milosevic, and P. Baltus, "A 29–37 GHz BiCMOS low-noise amplifier with 28.5 dB peak gain and 3.1–4.1 dB NF," in *Proc. IEEE RFIC*, Jun. 2018, pp. 288–291.

[25] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[26] Y. Wang, B. Afshar, L. Ye, V. C. Gaudet, and A. M. Niknejad, "Design of a low power, inductorless wideband variable-gain amplifier for high-speed receiver systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 4, pp. 696–707, Apr. 2012.

[27] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[28] B. Nasri, S. P. Sebastian, K. D. You, R. Ranjithkumar, and D. Shahrjerdi, "A 700 μW 1 GS/s 4-bit folding-flash ADC in 65 nm CMOS for wideband wireless communications," in *Proc. Int. Symp. Circuit Syst.*, May 2017, pp. 1–4.

[29] E. Olieman, A. J. Annema, and B. Nauta, "An interleaved full Nyquist high-speed DAC technique," *IEEE J. Solid-State Circuits*, vol. 50, no. 3, pp. 704–713, Mar. 2015.

[30] Juanda, W. Shu, and J. S. Chang, "A calibration-free/DEM-free 8-bit 2.4-GS/s single-core digital-to-analog converter with a distributed biasing scheme," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 26, no. 11, pp. 2299–2309, Nov. 2018. doi: 10.1109/TVLSI.2018.2850919.

[31] S.-N. Kim, M.-R. Kim, B.-R.-S. Sung, H.-W. Kang, M.-H. Cho, and S.-T. Ryu, "A SUC-based full-binary 6-bit 3.1-GS/s 17.7-mW current-steering DAC in 0.038 mm^2," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 2, pp. 794–798, Feb. 2016.

[32] F. Houfaf, M. Egot, A. Kaiser, A. Cathelin, and B. Nauta, "A 65 nm CMOS 1-to-10 GHz tunable continuous-time low-pass filter for high-data-rate communications," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2012, pp. 362–364.
[33] Y. Chen, P. Mak, S. D'Amico, L. Zhang, H. Qian, and Y. Wang, "A single-branch third-order pole–zero low-pass filter with 0.014- mm^2 die size and 0.8-kHz (1.25-nW) to 0.94-GHz (3.99-mW) bandwidth–power scalability," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 60, no. 11, pp. 761–765, Nov. 2013.
[34] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 618–632, Dec. 2007.
[35] B. Li, N. Liang, and W. Zhang, "On transmission model for massive MIMO under low-resolution output quantization," in *Proc. IEEE Veh. Technol. Conf. Spring (VTC-Spring)*, Sydney, NSW, Australia, Jun. 2017, pp. 1–5.
[36] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Linear precoding with low-resolution DACs for massive MU-MIMO-OFDM downlink," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1595–1609, Mar. 2019.
[37] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. New York, NY, USA: Springer, 2012.
[38] T. Araujo and R. Dinis, "On the accuracy of the Gaussian approximation for the evaluation of nonlinear effects in OFDM signals," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 346–351, Feb. 2012.
[39] J.-P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.
[40] M. S. Alavi, R. B. Staszewski, L. C. N. de Vreede, and J. R. Long, "A wideband 2×13-bit all-digital I/Q RF-DAC," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 732–752, Apr. 2014.
[41] *Waveform Candidates*, document 3GPP R1-162199, Qualcomm Incorporated, Apr. 2016.
[42] D. S. Palguna, D. J. Love, T. A. Thomas, and A. Ghosh, "Millimeter wave receiver design using low precision quantization and parallel $\Delta\Sigma$ architecture," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6556–6569, Oct. 2016.
[43] M. Mehrpoo, M. Hashemi, Y. Shen, L. C. N. de Vreede, and M. S. Alavi, "A wideband linear I/Q-interleaving DDRM," *IEEE J. Solid-State Circuits*, vol. 53, no. 5, pp. 1361–1373, May 2018.
[44] E. Roverato *et al.*, "All-digital RF transmitter in 28 nm CMOS with programmable RX-band noise shaping," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2017, pp. 222–223.
[45] *Base Station (BS) Radio Transmission and Reception, Release 15*, document TS 38.104, 3GPP, 2018.
[46] *Physical Channels and Modulation, Release 15*, document TS 38.211, 2018.
[47] A. Lozano, "Long-term transmit beamforming for wireless multicasting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, vol. 3, Apr. 2007, pp. III-417–III-420.
[48] P. Mogensen *et al.*, "LTE capacity compared to the Shannon bound," in *Proc. IEEE 65th Veh. Technol. Conf.*, Apr. 2007, pp. 1234–1238.
[49] *Physical Layer Procedure for Data, Release 15*, document TS 38.214, 3GPP, 2018.
[50] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
[51] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 537–552, Jan. 2016.

Sourjya Dutta (S'16) received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology (NIT), Durgapur, India, in 2012. He is currently pursuing the Ph.D. degree with the Tandon School of Engineering, New York University, Brooklyn, NY, USA. His research interests are in wireless communications, signal processing, and communication theory.

C. Nicolas Barati (S'14–M'19) received the Engineering Diploma degree in computer, telecommunications and network engineering from the University of Thessaly, Volos, Greece, in 2011, and the Ph.D. degree in electrical engineering from NYU Tandon School of Engineering, Brooklyn, NY, USA, in 2018. He is currently a Post-Doctoral Research Associate with Rice University, Houston, TX, USA. His research interests include wireless communications, massive mimo, mmWave, and wireless networking.

David Ramirez (S'11–M'12) received the B.S. degree (Hons.) in engineering physics from Tecnologico de Monterrey (ITESM) and the M.S. and Ph.D. degrees in electrical and computer engineering from Rice University. He is currently a Post-Doctoral Research Associate with Princeton University. He is also a Candidate Member of the *Sistema Nacional de Investigadores* in Mexico. He held a post-doctoral appointment with New York University. His research interests are in wireless networks, communication theory, and optimization.

Aditya Dhananjay received the Ph.D. degree from the Courant Institute of Mathematical Sciences, NYU. He was involved in mesh radio routing and resource allocation protocols, data communication over cellular voice channels, low-cost wireless rural connectivity, OFDM equalization, and phase noise mitigation in mmWave networks. He currently holds a post-doctoral position with NYU Wireless. He has developed and supervised much of the mmWave experimental work at the center. He has authored several refereed articles (including at SIGCOMM and MobiCom). He holds one patent and two provisional patents in the millimeter wave space.

James F. Buckwalter (S'01–M'06–SM'13) received the Ph.D. degree in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 2006.

From 1999 to 2000, he was a Research Scientist with Telcordia Technologies. In Summer 2004, he was with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. In July 2006, he joined the Faculty of the University of California at San Diego (UCSD) as an Assistant Professor and was promoted as an Associate Professor in 2012. He is currently a Professor of electrical and computer engineering with the University of California at Santa Barbara (UCSB), Santa Barbara, CA, USA.

Dr. Buckwalter was a recipient of a 2004 IBM Ph.D. Fellowship, the 2007 Defense Advanced Research Projects Agency (DARPA) Young Faculty Award, the 2011 NSF CAREER Award, and the 2015 IEEE MTT-S Young Engineer Award.

Sundeep Rangan (S'94–M'98–SM'13–F'16) received the B.A.Sc. degree from the University of Waterloo, Canada, and the M.Sc. and Ph.D. degrees from the University of California at Berkeley, Berkeley, all in electrical engineering.

He has held post-doctoral appointments with the University of Michigan, Ann Arbor, and Bell Labs. In 2000, he co-founded (with four others) Flarion Technologies, a spin-off of Bell Labs, which developed Flash OFDM, the first cellular OFDM data system and pre-cursor to 4G cellular systems including LTE and WiMAX. In 2006, Flarion was acquired by Qualcomm Technologies. He was the Director of engineering with Qualcomm, where he was involved in OFDM infrastructure products. He is currently a Professor of electrical and computer engineering with New York University and an Associate Director with NYU Wireless.