Achievable Rates of Hybrid Architectures with Few-Bit ADC Receivers

Jianhua Mo*, Ahmed Alkhateeb*, Shadi Abu-Surra† and Robert W. Heath, Jr.*
*Department of ECE, The University of Texas at Austin, Austin, TX 78712
†Samsung Research America-Dallas, Richardson, TX 75082
Email: {jhmo, aalkhateeb, rheath}@utexas.edu, shadi.as@samsung.com

***Abstract*—Massive multiple-input multiple-output (MIMO) will likely be a key component of future wireless systems. This is thanks to its large multiplexing and beamforming gains which are essential for both mmWave and low frequency systems. Hybrid analog/digital architecture and receivers with low-resolution analog-to-digital converters (ADCs) are two solutions that were proposed to reduce the cost and power consumption. The work done so far on these two architectures, though, represent two extreme cases in which either small number of RF chains with full-resolution ADCs or low-resolution ADC with a number of RF chains equal to the number of antennas is assumed. In this paper, the generalized hybrid architecture with a small number of RF chains and arbitrary resolution ADCs is considered. For this architecture, the spectral efficiency is analyzed and the achievable rates with channel inversion and SVD based transmission methods are derived. The achievable rate is comparable to that achieved by full-precision ADC receiver at low and medium SNR.**

I. Introduction

Large-scale MIMO will play a central role in next-generation wireless systems. Thanks to its large multiplexing and beamforming gains, massive MIMO is important for both low-frequency and mmWave systems. [1], [2]. Unfortunately, the high hardware cost and power consumption of mixed-signal components makes the fully-digital precoding solution, that allocates an RF chain per antenna, difficult to realize in practice [3], [4]. To overcome this challenge, new architectures that relax the requirement of associating an RF chain per antenna need to be developed [5]. Hybrid analog/digital architectures [6], [7], and 1-bit ADC receivers [8] are two promising solutions that gained much interest in the last few years. Those two solutions, though, represent two extreme cases in terms of the number of bits and RF chains. It is interesting, therefore, to explore the generalization of those two architectures, which is the objective of this paper.

To overcome the limitations on the RF chains, and to support multi-stream multiplexing, hybrid analog/digital architectures, that divide the precoding processing between analog and digital domains, were proposed for both mmWave and low-frequency massive MIMO systems [6], [9]–[12]. Hybrid architectures employ a number of radio frequency (RF) chains much less than the number of antennas, and rely on the extra RF beamforming processing that is normally implemented using networks of phase shifters [6], [9], [10]. Considering general MIMO systems, hybrid precoding for diversity and multiplexing gain were investigated in [9], and for interference management in [10]. These solutions, however, did not make use of the special large MIMO characteristics in the design. For mmWave massive MIMO systems, the sparse nature of the channels was exploited in [6] to derive matching pursuit based solutions for the hybrid precoders. Hybrid precoding was also shown to achieve a near-optimal performance in low-frequency massive MIMO systems when the deployed number of RF chains is large enough compared to the number of users and multiplexed streams [11], [12]. The main objective of the hybrid precoding algorithms in [6], [9], [11]–[13] was to achieve an spectral efficiency close to that obtained by fully-digital solutions. The hybrid architectures adopted in [6], [9], [11]–[13], though still assume that receive RF chains include high-resolution ADCs, which consume high power [3].

An alternative to high resolution ADCs is to live with ultra low resolution ADCs (1-4 bits), which reduces power consumption and hardware cost. In [8], [14]–[23], receiver architectures where the received signal at each antenna is directly quantized by low resolution ADCs without any analog combining is considered. At present, the exact capacity of quantized multiple-input multiple-output (MIMO) channel is generally unknown, except for the simple multiple-input single-output (MISO) channel and some special cases, such as in the low or high SNR regime [8], [16], [24]. Transmitting independent QAM signals [14]–[16] or Gaussian signals [18], [20], [21] from each antenna nearly achieves the capacity at low SNR, but is far from the optimal at high SNR. The case with channel state information at the transmitter (CSIT) was studied in our previous work [8], [25] where two methods were proposed to design the input constellation to maximize the channel capacity. It was shown that the proposed methods achieve much larger rate than QAM signaling, especially at high SNR. In addition, there is also a lot of interests to use 1-bit ADCs for the massive MIMO receiver where a large number of ADCs are needed [22], [23], [26]. In this architecture, though, the number of RF chains was assumed to be equal to the number of antennas, which means the hardware cost is still high, and no gain was made of the

This work was done in part when the first author interned with Samsung Research America-Dallas. The authors at the University of Texas at Austin are supported in part by the National Science Foundation under Grant No. 1319556 and No. NSF-CCF-1527079.

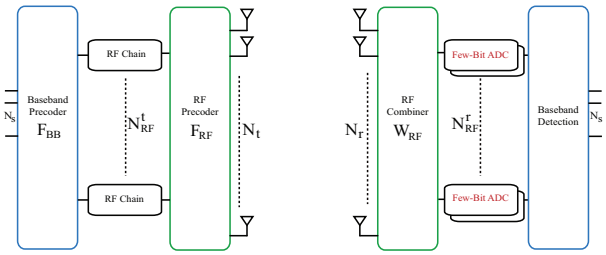

Fig. 1. A MIMO system with hybrid precoding and few-bit ADCs. The transmitter (receiver) has $N_{\rm t}(N_{\rm r})$ antennas and $N^{\rm t}_{\rm RF}(N^{\rm r}_{\rm RF})$ RF chains. The receiver has only few-bit low resolution ADCs.

possible processing in the RF domain.

The hybrid architecture and 1-bit ADC receiver architecture studied in the literature represent two extreme points in terms of the number of ADC bits and RF chains, namely the hybrid architecture employed a small number of RF chains but with high resolution ADCs and the 1-bit ADC receivers assumed a number of RF chains equal the number of antennas. In this paper, we investigate the generalized hybrid architecture with few-bit ADC receivers and draw important conclusions about its energy-rate trade-off.

The contributions of this paper can be summarized as follows. For the transceiver architecture with hybrid precoding and low resolution ADCs, we propose two transmission methods and derive their achievable rates in closed form. We also derive an upper bound of the channel capacity for one-bit quantization. The bound could be achieved by the proposed two transmission methods under certain conditions. In the simulations, we show that the proposed architecture with few-bit ADC receiver can achieve a performance comparable to that obtained with fully-digital or hybrid architecture with full-precision ADC receiver in the low-to-medium SNR range, which is of a special importance for mmWave communications.

II. System Model

We propose a new architecture of MIMO system with hybrid analog/digital precoding and combining with few-bit ADCs, as shown in Fig. 1. The transmitter and receiver are equipped with $N_{\rm t}$ and $N_{\rm r}$ antennas, respectively. The transmitter is assumed to have $N^{\rm t}_{\rm RF}$ RF chains, while the receiver employs $N^{\rm r}_{\rm RF}$ RF chains with few-bit (1-4 bits) ADCs. Further, the number of antennas and RF chains are assumed to satisfy $(N^{\rm t}_{\rm RF} \le N_{\rm t}, N^{\rm r}_{\rm RF} \le N_{\rm r})$. The transmitter and receiver communicate via $N_{\rm s}$ data streams, with $N_{\rm s} \le \min\left(N^{\rm t}_{\rm RF}, N^{\rm r}_{\rm RF}\right)$.

Compared to the fully-digital architecture where the receiver has $N_{\rm r}$ pairs of high resolution ADCs, the proposed receiver architecture contains only $N^{\rm r}_{\rm RF}$ pairs of few-bit ADCs, which greatly reduces both the hardware cost and power consumption.

Assuming a narrowband channel and perfect synchronization, the signal at the receive antenna is

$$\mathbf{y} \;=\; \mathbf{H}\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{s}$ is the digital baseband signal with the covariance $\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \frac{P_{\rm t}}{N_{\rm s}}\mathbf{I}$ where $P_{\rm t}$ is the transmission power, $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I})$ is the white Gaussian noise, $\mathbf{F}_{\rm RF} \in \mathbb{C}^{N_{\rm t} \times N^{\rm t}_{\rm RF}}$ and $\mathbf{F}_{\rm BB} \in \mathbb{C}^{N^{\rm t}_{\rm RF} \times N_{\rm s}}$ is the frequency band analog and baseband digital precoder, respectively.

After the analog combining and low-resolution quantization, the received signal is

$$\mathbf{r} \;=\; \mathcal{Q}\left(\mathbf{W}^*_{\rm RF}\mathbf{H}\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\mathbf{s} + \mathbf{W}^*_{\rm RF}\mathbf{n}\right), \tag{2}$$

where $\mathbf{W}_{\rm RF} \in \mathbb{C}^{N_{\rm r} \times N^{\rm r}_{\rm RF}}$ is the analog combiner and $\mathcal{Q}()$ is the quantization function which applies to component-wise and separately to the real and imaginary parts. The effective noise $\widetilde{\mathbf{n}} \triangleq \mathbf{W}^*_{\rm RF}\mathbf{n}$ has covariance $\mathbf{W}^*_{\rm RF}\mathbf{W}_{\rm RF}$.

With known CSI at the transmitter, the capacity of this channel is

$$\begin{aligned} C \;&=\; \max_{\substack{\mathbf{F}_{\rm BB},\mathbf{F}_{\rm RF},\mathbf{W}_{\rm RF},\\ p(\mathbf{s}),\,\mathcal{Q}()}} I(\mathbf{s};\mathbf{r}|\mathbf{H}) \\ &=\; \max_{\substack{\mathbf{F}_{\rm BB},\mathbf{F}_{\rm RF},\mathbf{W}_{\rm RF},\\ p(\mathbf{s}),\,\mathcal{Q}()}} \int_{\mathbf{s}} \sum_{\mathbf{r}} \Pr(\mathbf{s})\Pr(\mathbf{r}|\mathbf{s}) \log_2 \frac{\Pr(\mathbf{r}|\mathbf{s})}{\Pr(\mathbf{r})} {\rm d}\mathbf{s} \end{aligned}$$

where $p(\mathbf{s})$ represents the distribution of $\mathbf{s}$ and $\Pr(\mathbf{r}|\mathbf{s})$ is the transitional probability. Note that the maximization is also over the quantization function $\mathcal{Q}()$, for example, thresholds of the ADCs [27]. If the simple uniform quantization is assumed, then the stepsize Δ is the only parameter in $\mathcal{Q}()$.

The capacity of this quantized hybrid precoding channel is unknown. In this paper, we will propose two transmission methods and analyze their achievable rates.

III. Problem Formulation

Since analog precoding and combining are implemented by analog phase shifters, each element of $\mathbf{F}_{\rm RF}$ and $\mathbf{W}_{\rm RF}$ is limited to have same norm. The optimization problem is to maximize the mutual information between $\mathbf{s}$ and $\mathbf{r}$ as follows.

$$\begin{aligned} \text{P1}: \quad \max_{\substack{\mathbf{F}_{\rm BB},\mathbf{F}_{\rm RF},\mathbf{W}_{\rm RF},\\ p(\mathbf{s}),\,\mathcal{Q}()}} \quad & I(\mathbf{s};\mathbf{r}|\mathbf{H}) && (3) \\ \text{s.t.} \quad & |[\mathbf{F}_{\rm RF}]_{mn}| = \frac{1}{\sqrt{N_{\rm t}}},\ \forall m,n, && (4) \\ & |[\mathbf{W}_{\rm RF}]_{mn}| = \frac{1}{\sqrt{N_{\rm r}}},\ \forall m,n, && (5) \\ & \|\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\|^2_F \le N_{\rm s}, && (6) \end{aligned}$$

where (6) is due to the transmission power constraint, i.e., $\|\mathbf{F}_{\rm RF}\mathbf{F}_{\rm BB}\|^2 \le P_{\rm t}$.

It is very difficult, if not impossible to solve the problem P1. First, it is hard to optimize the mutual information over so many parameters. Second, the quantization function $\mathcal{Q}()$ is nonlinear and also related to the input distribution $p(\mathbf{s})$. Third, the equality constraints in (4) and (5) is non-convex and hard to deal with.

Throughout the paper, it is assumed that $\mathbf{W}^*_{\rm RF}\mathbf{W}_{\rm RF} = \mathbf{I}$ and therefore the effective noise is still white Gaussian noise. This assumption simplifies the problem and the similar idea appeared in [28]. In addition, the analog precoder $\mathbf{F}_{\rm RF}$ is also assumed to satisfy $\mathbf{F}^*_{\rm RF}\mathbf{F}_{\rm RF} = \mathbf{I}$. Under this assumption, the coupled power constraint (6) involving the digital and analog

precoding become a simple constraint on the digital precoder $\mathbf{F}_{\mathrm{BB}}$.

Consequently, the optimization problem P1 is reformulated as

$$\mathrm{P2}: \quad \max_{\substack{\mathbf{F}_{\mathrm{BB}},\mathbf{F}_{\mathrm{RF}},\mathbf{W}_{\mathrm{RF}},\\ p(\mathbf{s}),\mathcal{Q}()}} \quad I(\mathbf{s};\mathbf{r}|\mathbf{H}) \tag{7}$$

$$s.t. \quad |[\mathbf{F}_{\mathrm{RF}}]_{mn}| = \frac{1}{\sqrt{N_{\mathrm{t}}}}, \quad \forall m,n, \tag{8}$$

$$|[\mathbf{W}_{\mathrm{RF}}]_{mn}| = \frac{1}{\sqrt{N_{\mathrm{r}}}}, \quad \forall m,n, \tag{9}$$

$$\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}} = \mathbf{I}, \tag{10}$$

$$\mathbf{W}_{\mathrm{RF}}^{*}\mathbf{W}_{\mathrm{RF}} = \mathbf{I}, \tag{11}$$

$$\|\mathbf{F}_{\mathrm{BB}}\|_F^2 \leq N_{\mathrm{s}}. \tag{12}$$

Although the Problem P2 is simpler than P1, it is still hard to solve it. In this paper, we develop two transmission strategies, including the precoding techniques, input signal distribution and the quantization design. We will investigate their available rates and show that their performance are close to optimum in certain cases.

IV. Upper Bound of the Achievable Rate

In this section, we provide upper bounds of the achievable rate for one-bit quantization. For multi-bit quantization, the upper bounds are unknown and left for future work.

Proposition 1. *An upper bound of the achievable rate with hybrid precoding and one-bit quantization is*

$$R^{\mathrm{1bit,ub}} = 2N_{\mathrm{RF}}^{\mathrm{r}}\left(1-\mathcal{H}\left(Q\left(\sqrt{\frac{P_{\mathrm{t}}\nu_1^2}{N_{\mathrm{RF}}^{\mathrm{r}}}}\right)\right)\right), \tag{13}$$

where $\mathcal{H}(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function, $Q(\cdot)$ is tail probability of the standard normal distribution and ν_1 is the maximum singular value of the effective channel matrix $\mathbf{G} \triangleq \mathbf{W}_{\mathrm{RF}}^{*}\mathbf{H}\mathbf{F}_{\mathrm{RF}}$.

Proof: The proof is similar to that of [8, Proposition 4] and omitted here. ∎

At low SNR, this upper bound is approximated as

$$R^{\mathrm{1bit,ub}} = \frac{2}{\pi}\frac{P_{\mathrm{t}}\nu_1^2}{\ln 2} + o(P_{\mathrm{t}}), \tag{14}$$

following the facts $Q(t) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}}t + o(t^2)$ and $\mathcal{H}(\frac{1}{2}+t) = 1 - \frac{2}{\ln 2}t^2 + o(t^2)$. So the bound increases linearly with the power.

Note that the upper bound given in (13) is related to the choice of analog combiner $\mathbf{W}_{\mathrm{RF}}$ and precoder $\mathbf{F}_{\mathrm{RF}}$.

V. Achievable Rate with Channel Inversion Based Transmission

In this section, we propose channel inversion based transmission. For this transmission method, this is no interference between streams at the receiver and each stream is quantized separately. Therefore, the exact achievable rate of this method can be found.

A. Channel Inversion Based Precoding Algorithm

For digital precoding design, we propose to use channel inversion precoding assuming that $N_{\mathrm{RF}}^{\mathrm{t}} \geq N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{s}}$. If there are more RF chains at the receiver than that at the transmitter, i.e., $N_{\mathrm{RF}}^{\mathrm{r}} > N_{\mathrm{RF}}^{\mathrm{t}}$, then only $N_{\mathrm{RF}}^{\mathrm{t}}$ out of $N_{\mathrm{RF}}^{\mathrm{r}}$ RF chains are used at the receiver. The digital precoder is

$$\mathbf{F}_{\mathrm{BB}} = \sqrt{\frac{N_{\mathrm{s}}}{\beta}}\mathbf{G}^{*}\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1} \tag{15}$$

where

$$\beta = \mathrm{tr}\left\{\mathbf{G}^{*}\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\mathbf{G}\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}}\right\} \tag{16}$$

such that the power constraint (6) is satisfied. As it is assumed that $\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}} = \mathbf{I}$, β is simplified to be

$$\beta = \mathrm{tr}\left\{\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\right\}. \tag{17}$$

Since there is no interference between streams because of channel inversion precoding, each stream of data could be detected separately. The received signal is

$$\mathbf{r} = \mathcal{Q}\left(\mathbf{W}_{\mathrm{RF}}^{*}\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{RF}}^{*}\mathbf{n}\right) \tag{18}$$

$$= \mathcal{Q}\left(\sqrt{\frac{N_{\mathrm{s}}}{\mathrm{tr}\left\{\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\right\}}}\mathbf{s} + \mathbf{W}_{\mathrm{RF}}^{*}\mathbf{n}\right). \tag{19}$$

The channel is converted to $2N_{\mathrm{s}}$ parallel sub channels, each of which is a quantized real-valued single-input single-output (SISO) channel. The SNR of each sub-channel is given by

$$\mathrm{SNR}_{\mathrm{CI}} = \frac{P_{\mathrm{t}}}{\mathrm{tr}\left\{\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\right\}}. \tag{20}$$

Maximizing the SNR is equivalent to maximizing the following term

$$\eta\left(\mathbf{G}\right) \triangleq \left(\mathrm{tr}\left\{\left(\mathbf{G}\mathbf{G}^{*}\right)^{-1}\right\}\right)^{-1} \tag{21}$$

$$= \left(\frac{1}{\nu_1^2} + \frac{1}{\nu_2^2} + \cdots + \frac{1}{\nu_{N_{\mathrm{s}}}^2}\right)^{-1}, \tag{22}$$

where $\nu_1, \nu_2, \cdots, \nu_{N_{\mathrm{s}}}$ are the singular values of the effective channel $\mathbf{G}$ in descending order. Therefore, $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{RF}}$ should be chosen to maximize the harmonic mean of the squared singular values of $\mathbf{G}$, or equivalently the harmonic mean of the eigenvalues of $\mathbf{G}\mathbf{G}^{*}$.

In this paper, we assume that the $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ consist of columns from the DFT matrices $\mathbf{D}_{N_{\mathrm{t}}}$ and $\mathbf{D}_{N_{\mathrm{r}}}$, respectively. Note that each entry of the DFT matrix has same norm, which satisfies the constraint imposed by analog precoding. In addition, the DFT matrix has orthogonal columns and therefore the selected $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ satisfy $\mathbf{F}_{\mathrm{RF}}^{*}\mathbf{F}_{\mathrm{RF}} = \mathbf{I}$ and $\mathbf{W}_{\mathrm{RF}}^{*}\mathbf{W}_{\mathrm{RF}} = \mathbf{I}$. The algorithm is summarized in Algorithm 1.

Last, noting that

$$\left(\frac{1}{\nu_1^2} + \frac{1}{\nu_2^2} + \cdots + \frac{1}{\nu_{N_{\mathrm{s}}}^2}\right)^{-1} \geq \frac{\nu_{N_{\mathrm{s}}}^2}{N_{\mathrm{s}}}, \tag{24}$$

Algorithm 1 Channel Inversion Based Transmission Method

1) Analog precoding design:
   a) Compute the effective channel $\widehat{\mathbf{G}} = \mathbf{D}_{N_\mathrm{r}}^* \mathbf{H} \mathbf{D}_{N_\mathrm{t}}$.
   b) Use exhaustive search to find from $\widehat{\mathbf{G}}$ a $N_{\mathrm{RF}}^\mathrm{r} \times N_{\mathrm{RF}}^\mathrm{t}$ a submatrix $\widetilde{\mathbf{G}}$ maximizing the harmonic mean of the eigenvalues of $\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^*$.
   Denote the optimal selected row and column indices as $\mathcal{I} \subset \{1, 2, ..., N_\mathrm{r}\}$ and $\mathcal{J} \subset \{1, 2, ...N_\mathrm{t}\}$, respectively.
   c) The analog combiner $\overline{\mathbf{W}}_{\mathrm{RF}}$ is composed of the $\mathcal{I}$ columns of $\mathbf{D}_{N_\mathrm{r}}$ and the analog precoder $\overline{\mathbf{F}}_{\mathrm{RF}}$ is composed of the $\mathcal{J}$ columns of $\mathbf{D}_{N_\mathrm{t}}$.
2) Digital precoding design:
   a) Compute the optimal effective channel $\overline{\mathbf{G}} \triangleq \overline{\mathbf{W}}_{\mathrm{RF}}^* \mathbf{H} \overline{\mathbf{F}}_{\mathrm{RF}}$.
   b) Set the digital precoder $\mathbf{F}_{\mathrm{BB}}$ as

$$\mathbf{F}_{\mathrm{BB}} = \sqrt{\frac{N_\mathrm{s}}{\mathrm{tr}\left\{\left(\overline{\mathbf{G}}\,\overline{\mathbf{G}}^*\right)^{-1}\right\}}}\overline{\mathbf{G}}^* \left(\overline{\mathbf{G}}\,\overline{\mathbf{G}}^*\right)^{-1}. \quad (23)$$

3) Signaling: $\mathbf{s}$ is chosen to be 2^{2b}-QAM symbols.
4) Quantization: uniform quantization.

a lower bound of the SNR of the proposed precoding design is

$$\mathrm{SNR}_{\mathrm{lb}} = \frac{P_\mathrm{t} \nu_{N_\mathrm{s}}^2}{N_\mathrm{s}}. \quad (25)$$

B. Rate Analysis with One-Bit Quantization

In this subsection, we focus on the special case of one-bit quantization and derive the gap between the achievable rate and the upper bound given in Proposition 1.

If 1-bit ADCs are used at the receiver, the capacity of each real-valued SISO sub-channel is achieved by binary antipodal signaling and is given by [24]

$$1 - \mathcal{H}\left(Q\left(\sqrt{\mathrm{SNR}_{\mathrm{CI}}}\right)\right). \quad (26)$$

The total sum rate therefore is

$$R_{\mathrm{CI}}^{\mathrm{1bit}} = 2N_\mathrm{s}\left(1 - \mathcal{H}\left(Q\left(\sqrt{\mathrm{SNR}_{\mathrm{CI}}}\right)\right)\right). \quad (27)$$

Based on the SNR lower bound in (25), a lower bound of the achievable rate is

$$R_{\mathrm{CI}}^{\mathrm{1bit,lb}} = 2N_\mathrm{s}\left(1 - \mathcal{H}\left(Q\left(\sqrt{\frac{P_\mathrm{t}\nu_{N_\mathrm{s}}^2}{N_\mathrm{s}}}\right)\right)\right) \quad (28)$$

$$= 2N_\mathrm{s}\left(1 - \mathcal{H}\left(Q\left(\sqrt{\frac{P_\mathrm{t}\nu_1^2}{N_\mathrm{s}}\frac{\nu_{N_\mathrm{s}}^2}{\nu_1^2}}\right)\right)\right). \quad (29)$$

Comparing (29) and (13), we find that the power gap between $R^{\mathrm{1bit,ub}}$ and $R_{\mathrm{CI}}^{\mathrm{1bit,lb}}$ is $10\log_{10}\frac{\nu_1^2}{\nu_{N_\mathrm{s}}^2}$ dB. Therefore, we conclude that compared to optimal digital precoding (which is unknown), the power loss of the channel inversion precoding is at most $10\log_{10}\frac{\nu_1^2}{\nu_{N_\mathrm{s}}^2}$ dB. It also implies that for a well-conditioned effective channel, the power loss is small.

Furthermore, if there is only one RF chain at the receiver, i.e., $N_{\mathrm{RF}}^\mathrm{r} = N_\mathrm{s} = 1$, the achievable rate given in (27) and the upper bound in (13) are exactly same and it implies that the channel capacity is achieved by the proposed transmission method.

C. Rate Analysis with Few-Bit Quantization

For a b-bit quantizer, consider equiprobable, equispaced 2^b-PAM (pulse amplitude modulated) input, with quantizer thresholds chosen to be the mid-points of the input mass point locations. Although this combination of input and quantization is suboptimal, it is shown in [24] to be nearly optimal, especially at high SNR. In addition, this combination is actually optimal for one-bit quantization.

We next provide an example of two-bit quantization. The derivation with multi-bit quantization is similar. The set of input signals is $\mathcal{S} = \left\{-\frac{3\Delta}{2}, -\frac{\Delta}{2}, \frac{\Delta}{2}, \frac{3\Delta}{2}\right\}$ where Δ is the stepsize. The transition probability between the ADC input s and quantization output r can be simply computed. For example,

$$\Pr\left(r = -\frac{\Delta}{2}\middle| s = -\frac{3\Delta}{2}\right) = \Phi\left(\frac{3\Delta}{2\xi}\right) - \Phi\left(\frac{\Delta}{2\xi}\right) \quad (30)$$

where ξ^2 denotes the noise variance and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The transition probability matrix of higher resolution ADCs could be obtained similarly.

The SNR of each sub-channel should be equal to the value given in (20). Hence,

$$\mathrm{SNR}_{\mathrm{CI}} = \frac{1}{2^{b+1}}\frac{\left(\Delta^2 + (3\Delta)^2 + \cdots + \left(\left(2^b - 1\right)\Delta\right)^2\right)}{\xi^2}$$

$$= \frac{1}{2^{b+1}}\frac{1}{3}2^{b-1}(2^{2b} - 1)\frac{\Delta^2}{\xi^2} \quad (31)$$

where (31) is from the fact that $1^2 + 3^2 + \cdots + (2n-1)^2 = \frac{1}{3}n(4n^2 - 1)$. Therefore, we have

$$\frac{\Delta}{\xi} = \sqrt{\frac{12\,\mathrm{SNR}_{\mathrm{CI}}}{2^{2b} - 1}}. \quad (32)$$

The achievable rate can therefore be computed as

$$R_{\mathrm{CI}}^{\mathrm{b\,bit}} = 2N_\mathrm{s}\sum_s\sum_r \Pr(s)\Pr(r|s)\log\frac{\Pr(r|s)}{\Pr(r)} \quad (33)$$

$$= 2N_\mathrm{s}\sum_s\sum_r \frac{1}{2^b}\Pr(r|s)\log\frac{\Pr(r|s)}{\Pr(r)} \quad (34)$$

where

$$\Pr(r) = \sum_s \Pr(s)\Pr(r|s) \quad (35)$$

$$= \frac{1}{2^b}\sum_s \Pr(r|s). \quad (36)$$

The achievable rate given in (34) is complicated and cannot provide any intuition. A simple lower bound of (34) can be

found by Fano's inequality [29, Section 2.10]. The conditional entropy is upper bounded by

$$\mathcal{H}(s|r) \le \mathcal{H}(P_e) + P_e \log(|\mathcal{S}| - 1), \tag{37}$$

where the error probability for 2^b-PAM signal is [30]

$$\begin{aligned} P_e &= 2\left(1 - \frac{1}{2^b}\right) Q\left(\frac{\Delta}{2\xi}\right) && (38)\\ &= 2\left(1 - \frac{1}{2^b}\right) Q\left(\sqrt{\frac{3\,\mathrm{SNR_{CI}}}{2^{2b}-1}}\right). && (39)\end{aligned}$$

Hence, the mutual information between s and r is

$$\begin{aligned} I(s;r) &= \mathcal{H}(s) - \mathcal{H}(s|r) && (40)\\ &\ge b - \mathcal{H}(P_e) - P_e \log\left(2^b - 1\right). && (41)\end{aligned}$$

A lower bound of the sum rate of $2N_\mathrm{s}$ sub-channels therefore is

$$R_{\mathrm{CI}}^{\mathrm{b\,bit,lb}} = 2N_\mathrm{s}\left(b - \mathcal{H}(P_e) - P_e \log\left(2^b - 1\right)\right) \tag{42}$$

From (42), we find that the as $\mathrm{SNR_{CI}}$ increases, P_e decreases to zero and $R_{\mathrm{CI}}^{\mathrm{b\,bit,lb}}$ converges to $2N_\mathrm{s}b$ bps/Hz. In addition, note that for the 1-bit case, (42) degrades to (27).

VI. Achievable Rate With Singular Value Decomposition Based Transmission

The channel inversion precoding generally works well at high SNR and has poor performance at low SNR. In the second transmission method, the singular value decomposition (SVD) digital precoding is adopted. Since the interference between each steams can not be completely eliminated before quantization as in the channel inversion case, the exact achievable rate is unknown. We therefore choose to apply the additive quantization noise model (AQNM) [18], [20], [31], which is accurate enough at low SNR, to find a lower bound of the achievable rate.

Applying the additive quantization noise model, the equivalent channel is

$$\mathbf{r} = (1-\rho_b)\left(\mathbf{G}\mathbf{F}_{\mathrm{BB}}\mathbf{s} + \mathbf{W}_{\mathrm{RF}}^*\mathbf{n}\right) + \mathbf{n}_\mathrm{Q}, \tag{43}$$

where ρ_b is the distortion factor for b-bit ADC (see Table 1 in [20] for the value of ρ_b) and $\mathbf{n}_\mathrm{Q}$ is the quantization noise with the variance $\rho_b(1-\rho_b)\mathrm{diag}\left\{\frac{P_\mathrm{t}}{N_\mathrm{s}}\mathbf{G}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^* + \mathbf{I}\right\}$.

Assuming Gaussian signaling $\mathbf{s}$, and the quantization noise is the worst case of Gaussian distributed, a lower bound of the achievable rate is obtained as

$$\begin{aligned} &R_{\mathrm{AQNM}}\\ &= \log_2\left|\mathbf{I} + (1-\rho_b)\frac{P_\mathrm{t}}{N_\mathrm{s}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\right.\\ &\quad\left.\left(\mathbf{I} + \rho_b\,\mathrm{diag}\left\{\frac{P_\mathrm{t}}{N_\mathrm{s}}\mathbf{G}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\right\}\right)^{-1}\mathbf{G}\mathbf{F}_{\mathrm{BB}}\right| && (44)\end{aligned}$$

At low SNR, the achievable rate is approximated as

$$\begin{aligned} R_{\mathrm{AQNM}} &\approx \log_2\left|\mathbf{I} + (1-\rho_b)\frac{P_\mathrm{t}}{N_\mathrm{s}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\mathbf{G}\mathbf{F}_{\mathrm{BB}}\right| && (45)\\ &= \frac{(1-\rho_b)P_\mathrm{t}}{N_\mathrm{s}\ln 2}\mathrm{tr}\left\{\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\mathbf{G}\mathbf{F}_{\mathrm{BB}}\right\} + o(P_\mathrm{t}) && (46)\end{aligned}$$

To maximize the term $\frac{1}{N_\mathrm{s}}\mathrm{tr}\left\{\mathbf{G}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\right\}$ under the constraint $\|\mathbf{F}_{\mathrm{BB}}\|_F^2 \le N_\mathrm{s}$, the optimal choice of $\mathbf{F}_{\mathrm{BB}}$ is the eigenmode beamforming, i.e.,

$$\mathbf{F}_{\mathrm{BB}} = \mathbf{v}_1, \tag{47}$$

where $\mathbf{v}_1$ is the right singular vector corresponding to the largest singular value of $\mathbf{G}$. The resulting rate is

$$R_{\mathrm{AQNM}} = \frac{(1-\rho_b)P_\mathrm{t}\nu_1^2}{\ln 2} + o(P_\mathrm{t}). \tag{48}$$

For one-bit quantization, $\rho_b = 1 - \frac{2}{\pi}$ and the rate is

$$R_{\mathrm{AQNM}} = \frac{2}{\pi}\frac{P_\mathrm{t}\nu_1^2}{\ln 2} + o(P_\mathrm{t}). \tag{49}$$

It is found that the upper bound of one-bit quantized channel at low SNR given in (14) is achieved by eigenmode beamforming.

For higher SNR, the optimal $\mathbf{F}_{\mathrm{BB}}$ maximizing the rate R_{AQNM} in (44) is unknown [17]. We therefore use the conventional SVD precoding and waterfilling power allocation [17], [21]. The baseband digital prcoding is

$$\mathbf{F}_{\mathrm{BB}} = \mathbf{V}\,\mathrm{diag}\left\{\sqrt{\mathbf{p}}\right\}, \tag{50}$$

where $\mathbf{V}$ is obtained from the singular value decomposition of the matrix $\mathbf{G}$, i.e., $\mathbf{G} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ and $\mathbf{p}$ denotes the power allocation factor obtained from the conventional waterfilling method.

For the analog precoding and combining, we use exhaustive search over DFT codebooks. Since it is hard to maximize R_{AQNM} given in (44), we propose to maximize the rate assuming that ADCs have full-precision and the digital precoder satisfies $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{RF}}^* = \mathbf{I}$. As a result, the analog precoding is chosen to maximize the rate expression given in (51). The transmission method is summarized in Algorithm 2.

Last, we show why the AQNM model is not accurate enough to model the quantization channel at high SNR. At high SNR, R_{AQNM} converges as follows.

$$\begin{aligned} &R_{\mathrm{AQNM}}\\ &\approx \log_2\left|\mathbf{I} + \frac{1-\rho_b}{\rho_b}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\left(\mathrm{diag}\left\{\mathbf{G}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}^*\mathbf{G}^*\right\}\right)^{-1}\mathbf{G}\mathbf{F}_{\mathrm{BB}}\right|\\ &= \log_2\left|\mathbf{I} + \frac{1-\rho_b}{\rho_b}\mathbf{A}^*\mathrm{diag}\left\{\frac{1}{||\mathbf{a}_i||}\right\}\mathrm{diag}\left\{\frac{1}{||\mathbf{a}_i||}\right\}\mathbf{A}\right|\\ &= \log_2\left|\mathbf{I} + \frac{1-\rho_b}{\rho_b}\widetilde{\mathbf{A}}^*\widetilde{\mathbf{A}}\right|\\ &= \sum_{i=1}^{N_\mathrm{s}}\log_2\left(1 + \frac{1-\rho_b}{\rho_b}\lambda_i\left(\widetilde{\mathbf{A}}^*\widetilde{\mathbf{A}}\right)\right)\end{aligned}$$

where $\mathbf{A} \triangleq \mathbf{G}\mathbf{F}_{\mathrm{BB}}$ is a Hermitian matrix, $\mathbf{a}_i$ is the i-th column of $\mathbf{A}$, and $\widetilde{\mathbf{A}}$ is obtained by normalizing each each row of $\mathbf{A}$.

Since each row of $\widetilde{\mathbf{A}}$ has unit norm, then

$$\sum_{i=1}^{N_\mathrm{s}}\lambda_i\left(\widetilde{\mathbf{A}}^*\widetilde{\mathbf{A}}\right) = \mathrm{tr}\left(\widetilde{\mathbf{A}}^*\widetilde{\mathbf{A}}\right) = N_\mathrm{s}. \tag{52}$$

Algorithm 2 SVD Based Transmission Method

1) Analog precoding design:
   a) Compute the effective channel $\widehat{\mathbf{G}} = \mathbf{D}_{N_\mathrm{r}}^{*}\mathbf{H}\mathbf{D}_{N_\mathrm{t}}$.
   b) Use exhaustive search to find a $N_{\mathrm{RF}}^{\mathrm{r}} \times N_{\mathrm{RF}}^{\mathrm{t}}$ a submatrix $\widetilde{\mathbf{G}}$ from $\widehat{\mathbf{G}}$ maximizing the expression
   $$\log_2 \left| \mathbf{I} + \widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^{*} \right|. \tag{51}$$
   Denote the optimal selected row and column indices as $\mathcal{I} \subset \{1, 2, ..., N_\mathrm{r}\}$ and $\mathcal{J} \subset \{1, 2, ...N_\mathrm{t}\}$, respectively.
   c) The analog combiner $\overline{\mathbf{W}}_{\mathrm{RF}}$ is composed of the $\mathcal{I}$ columns of $\mathbf{D}_{N_\mathrm{r}}$ and the analog precoder $\overline{\mathbf{F}}_{\mathrm{RF}}$ is composed of the $\mathcal{J}$ columns of $\mathbf{D}_{N_\mathrm{t}}$.
2) Digital precoding design:
   a) Compute the optimal effective channel $\overline{\mathbf{G}} \triangleq \overline{\mathbf{W}}_{\mathrm{RF}}^{*}\mathbf{H}\overline{\mathbf{F}}_{\mathrm{RF}}$.
   b) Set the digital precoder $\mathbf{F}_{\mathrm{BB}}$ by SVD of $\overline{\mathbf{G}}$ and waterfilling method.
3) Signaling: $\mathbf{s}$ is chosen to follow Gaussian signaling.
4) Quantization: The thresholds of ADC are determined by Max-Lloyd algorithm [32], [33] which minimizes the MSE of Gaussian distributed input.

Therefore, we have

$$R_{\mathrm{AQNM}} \overset{(a)}{\leq} N_\mathrm{s} \log_2 \left(1 + \frac{1-\rho_b}{\rho_b}\right) \tag{53}$$

$$= N_\mathrm{s} \log_2 \frac{1}{\rho_b}, \tag{54}$$

where (a) follows from that $\log_2(1+x)$ is concave in x. When the ADC resolution b is large ($b \geq 3$), the distortion factor ρ_b could be approximated as [34]

$$\rho_b \approx \frac{\pi\sqrt{3}}{2} 2^{-2b}. \tag{55}$$

As a result,

$$R_{\mathrm{AQNM}} \leq 2N_\mathrm{s}b - N_\mathrm{s} \log_2 \frac{\pi\sqrt{3}}{2} \tag{56}$$

$$\approx 2N_\mathrm{s}b - 1.44N_\mathrm{s}. \tag{57}$$

As we know, the achievable rate of quantized MIMO channel is upper bounded by $2N_\mathrm{s}b$ bps/Hz and the channel inversion method can achieve the bound at high enough SNR as shown in (42). Therefore, the AQNM is not an accurate model at high SNR. The reason is threefold. First, the input signal $\mathbf{s}$ is assumed to follow suboptimal Gaussian distribution. Second, the quantization noise is assumed to be the worst-case Gaussian noise. Third, the Max-Loyd quantizer minimizing the MSE is not necessarily optimum for maximizing the channel capacity.

VII. Simulation Results

We evaluate the performance of proposed methods in a mmWave MIMO channel with large antenna arrays and limited number of transmit and receive RF chains. According to the experiment results reported in [35], [36], the number of clusters tends to be lower in the mmWave band compared with lower frequencies. The mmWave channel will mostly consists of the line-of-sight (LOS) and a few NLOS clusters. In the simulations, the wireless channel is assumed to have 4 clusters, each of which consists of 5 rays and the angle spread is 7.5 degrees. The results are obtained by averaging over 100 channel realizations. The transmitter is assumed to have $N_\mathrm{t} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_\mathrm{r} = 8$ antennas.

Figs. 2-3 show the achievable rates when the receiver has 1 and 4 RF chains, respectively. In Fig. 2, it is seen that when there is only one RF chains at the transmitter, the channel

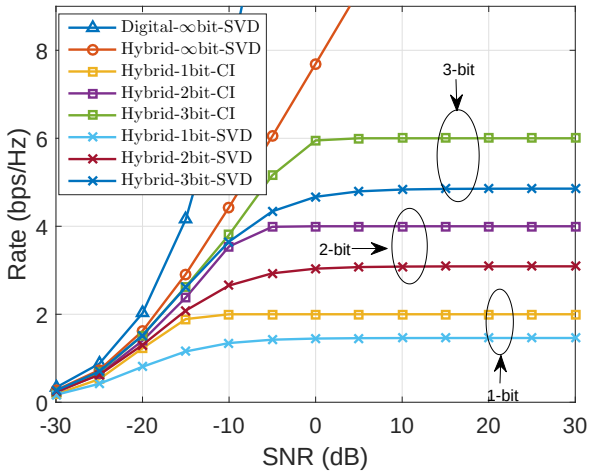

Fig. 2. This figure shows rates versus SNR of different transceiver architectures and transmission methods. The transmitter is assumed to have $N_\mathrm{t} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_\mathrm{r} = 8$ antennas and $N_{\mathrm{RF}}^{\mathrm{r}} = 1$ RF chains.

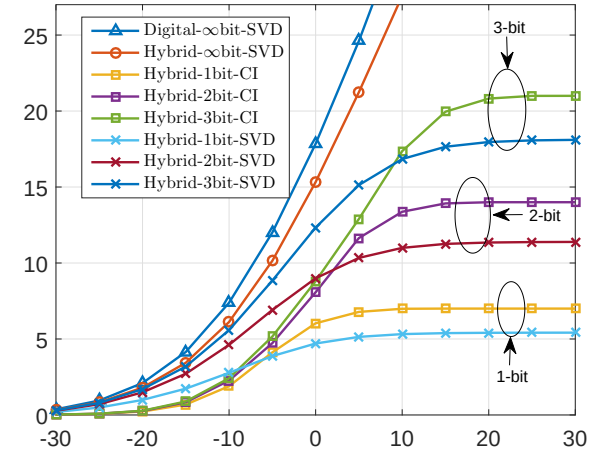

Fig. 3. This figure shows rates versus SNR of different transceiver architectures and transmission methods. The transmitter is assumed to have $N_\mathrm{t} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_\mathrm{r} = 8$ antennas and $N_{\mathrm{RF}}^{\mathrm{r}} = 4$ RF chains.

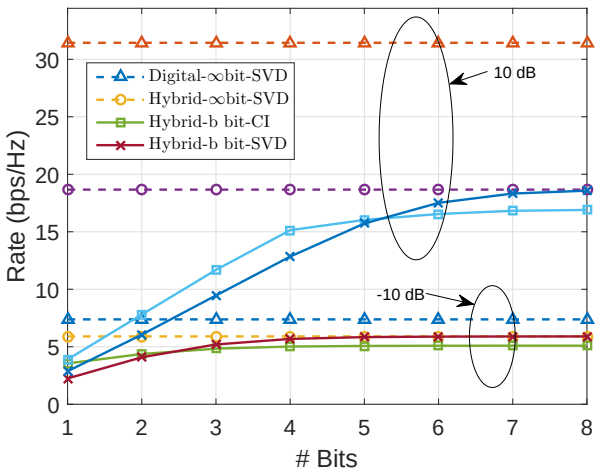

Fig. 4. This figure shows rates versus the ADC resolution of for different transceiver architectures and transmission methods. The transmitter is assumed to have $N_{\mathrm{t}} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_{\mathrm{r}} = 8$ antennas and $N_{\mathrm{RF}}^{\mathrm{r}} = 2$ RF chains.

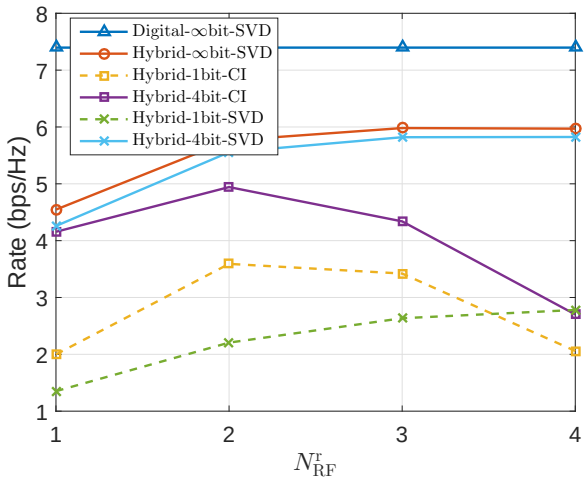

Fig. 5. This figure shows rates versus the number of receive RF chains for different transceiver architectures and transmission methods. The transmitter is assumed to have $N_{\mathrm{t}} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_{\mathrm{r}} = 8$ antennas, and the SNR is -10 dB.

inversion and SVD method has similar performance at low SNR. At high SNR, however, the channel inversion method achieves the rate $2b$ bps/Hz while the rate of SVD method saturates to $\log_2 \frac{1}{\rho_b}$ which is about 1.46, 3.09, 4.86 bps/Hz for $b = 1, 2, 3$, respectively. In Fig. 3, another case when $N_{\mathrm{RF}}^{\mathrm{r}} = 4$ is shown. It is found that although at high SNR the channel inversion method achieves larger rate than SVD method, its performance at low SNR is much worse. The reason is that since the channel has only 4 clusters, the fourth largest singular value ν_4 is small and the power loss $\log_2 \frac{\nu_1^2}{\nu_4^2}$ is large. Last, Figs. 2-3 show that if the maximum rate of two methods is chosen, the gap between the performance of few-bit ADC receiver and ∞-bit ADC receiver (hybrid or digital) is small at low and medium SNR. The gap between the curve of "Digital-∞bit-SVD" and "Hybrid-∞bit-SVD" represents the loss due to limited number of RF chains while the gap between the curve "Hybrid-∞bit" and the proposed algorithm is the loss due to low resolution ADCs. We can see the loss due to low resolution ADCs is small at low and medium SNRs. For example, the gap between the curve "Hybrid-∞bit" and "Hybrid-3bit-CI" is less than 3 dB when the SNR is less than 0 dB in Fig. 2.

Fig. 4 shows the achievable with respect to the ADC resolution. First, the rates of the finite-bit ADC receiver increase with resolution. Second, with multi-bit ADCs (4-bit when SNR = -10 dB and 8-bit when SNR = 10 dB), the SVD method achieves the performance similar to that of hybrid architecture with ∞-bit ADCs. This implies that high resolution ADC does not provide much rate improvement compared to the few-bit ADC when the SNR is low. Third, when the ADC resolution is low, channel inversion method is better than the SVD method while with high resolution quantization, the SVD method is better. This is reasonable since with high resolution ADC, the channel is close to be a unquantized channel and in a unquantized channel, SVD method is optimal.

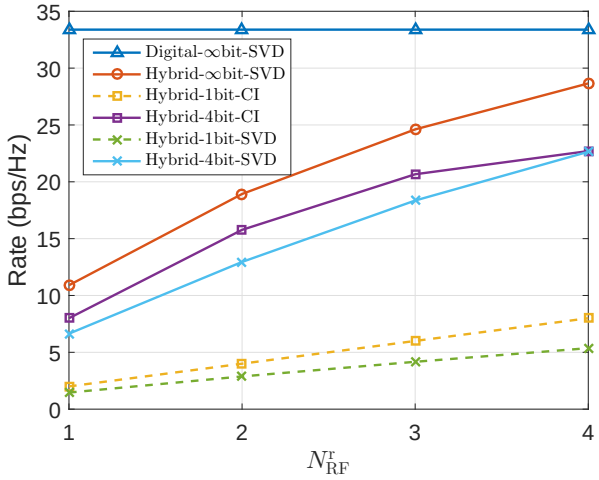

Fig. 6. This figure shows rates versus the number of receive RF chains for different transceiver architectures and transmission methods. The transmitter is assumed to have $N_{\mathrm{t}} = 64$ antennas and $N_{\mathrm{RF}}^{\mathrm{t}} = 8$ RF chains, while the receiver employs $N_{\mathrm{r}} = 8$ antennas, and the SNR is 10 dB.

Figs. 5 and 6 present the achievable rates versus the number of RF chains at the receiver. First, we find that the rate of SVD method always increases with SNR. Second, at low SNR (-10 dB), the channel inversion method achieves the largest rate when $N_{\mathrm{RF}}^{\mathrm{t}} = 2$. This means at low SNR, it is better to turn off some RF chains and transmit fewer number of steams. Note that the power consumption also decreases by turning off some RF chains. Third, we find that compared to fully digital architecture where $N_{\mathrm{RF}}^{\mathrm{t}} = N_{\mathrm{t}} = 64$ and $N_{\mathrm{RF}}^{\mathrm{r}} = N_{\mathrm{r}} = 8$, the hybrid architecture with limited number of RF chains ($N_{\mathrm{RF}}^{\mathrm{t}} = 8$ and $N_{\mathrm{RF}}^{\mathrm{r}} = 2$) and low resolution ADCs (4-bit) incurs about 20-30% spectral efficiency loss. As only a few of

low resolution ADCs are used, however, the energy efficiency of the proposed receiver is much higher than the fully digital architecture.

VIII. Conclusion

In this paper, we investigated the achievable spectral efficiency and energy-rate trade-off of a generalized hybrid architecture with few-bit ADC receivers. We considered channel inversion and SVD precoding based transmission methods and derived their achievable rates. The transmission methods include three elements: the design of analog and digital precoding, the choice of the transmit signal distribution and the setup of quantizer. We found when the ADC resolution is low and the SNR is high, that the channel inversion method is better than SVD method. Simulation results showed that at the low and medium SNR, the proposed architecture and precoding can achieve a comparable rate to the full-precision ADC solution.

References

[1] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] B. Murmann, "ADC performance survey 1997-2015," 2015. [Online]. Available: http://www.stanford.edu/ murmann/adcsurvey.html

[4] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, "Study and prototyping of practically large-scale mmWave antenna systems for 5G cellular devices," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 63–69, September 2014.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-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, December 2014.

[6] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath Jr, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[7] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

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

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

[10] V. Venkateswaran and A. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[11] T. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2014, pp. 4066–4071.

[12] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, Dec 2014.

[13] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4131–4143, 2010.

[14] B. Murray and I. Collings, "AGC and quantization effects in a zero-forcing MIMO wireless system," in *IEEE 63rd Vehicular Technology Conference*, vol. 4, May 2006, pp. 1802–1806.

[15] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proceedings of the 2006 international conference on Wireless communications and mobile computing*, ser. IWCMC '06. New York, NY, USA: ACM, 2006, pp. 1387–1392.

[16] A. Mezghani and J. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *IEEE International Symposium on Information Theory*, 2007, pp. 1286–1289.

[17] A. Mezghani, M.-S. Khoufi, and J. A. Nossek, "A modified MMSE receiver for quantized MIMO systems," *Proc. ITG/IEEE WSA, Vienna, Austria*, 2007.

[18] A. Mezghani and J. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.

[19] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 957–961.

[20] Q. Bai and J. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 1, pp. 3–14, 2015.

[21] 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. of Information Theory and Applications (ITA) Workshop*, 2015.

[22] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, April 2015.

[23] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," *arXiv preprint arXiv:1504.04540*, 2015.

[24] 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, 2009.

[25] J. Mo and R. W. Heath Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Information Theory and Applications Workshop (ITA), 2014*, Feb 2014, pp. 1–5.

[26] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *ArXiv e-prints*, Apr. 2014.

[27] U. Kamilov, V. Goyal, and S. Rangan, "Message-passing de-quantization with applications to compressed sensing," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6270–6281, Dec 2012.

[28] A. Molisch and X. Zhang, "FFT-based hybrid antenna selection schemes for spatially correlated MIMO channels," *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 36–38, Jan 2004.

[29] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[30] J. G. Proakis, "Digital communications." *McGraw-Hill, New York*, 2008.

[31] A. Fletcher, S. Rangan, V. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," vol. 1, no. 4, pp. 618–632, Dec 2007.

[32] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.

[33] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.

[34] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Springer Science & Business Media, 2012, vol. 159.

[35] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Select. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[36] T. Rappaport, G. Maccartney, M. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.