Integer Forcing Analog-to-Digital Conversion for Massive MIMO Systems

Luis G. Ordóñez, Iñaki Estella Aguerri, and Maxime Guillaud
Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies
Email: {luis.ordonez, inaki.estella, maxime.guillaud}@huawei.com

***Abstract*—The large number of analog received signals to be processed in massive MIMO systems renders conventional high-precision analog-to-digital conversion (ADC) not feasible in terms of cost- and power-efficiency. In this paper we present a novel ADC architecture inspired by the principles of integer-forcing source coding, which jointly reconstructs the received signals in the digital domain. This scheme effectively exploits the correlation among the received signals in order to increase the achievable uplink transmission rates with respect to conventional ADC units in the low-resolution regime.**

I. Introduction

Establishing a communication link between multiple transmitters and a digital multi-antenna receiver requires an interface between the analog received signals and the digital signal processing (DSP) unit. To that end, the analog signals received at the multiple antennas are sampled and quantized using an analog-to-digital conversion (ADC) unit. When the received signals are sampled at the Nyquist rate or above and quantized with high precision, the effects of the ADC unit are negligible. However, high-precision ADC may not be feasible in terms of cost- and power-efficiency [1] in communication systems using a large number of receive antennas and a large bandwidth [2].

For instance, massive multiple-input multiple-output (MIMO) technology uses antenna arrays with the number of antenna elements being some orders of magnitude larger than current state-of-the-art MIMO technology [3], [4]. Converting all antenna outputs with high precision results in unacceptable costs and power consumption and, additionally, increases the throughput requirements in the fronthaul link connecting the ADC unit to the DSP unit. The previous problems are considered to be, in fact, among the key bottlenecks for realizing future 5G wireless systems' targets [2], [5].

Since the power consumption of ADC scales roughly exponentially with the number of bits [1], a simple solution follows from substituting high-precision quantizers (e.g. 10-15 bits) with low-precision quantizers (e.g. 3-5 bits), or, even 1-bit quantizers. This direction has been extensively explored in the massive MIMO literature [5]–[11]. For instance, the mixed-ADC architecture in [9]–[11] substitutes a large fraction of the high-precision quantizers by 1-bit quantizers. The authors of [9] show that in the large-system limit rates close to the infinite-precision (unquantized) capacity can be supported using only a relatively small number of high-precision quantizers. However, a significant performance degradation has been reported under realistic operating conditions [5]. Particularly important is the impact of large-scale fading and imperfect power control, which results in received power imbalances among the different users and renders the communication with the weakest users unfeasible.

In this paper we propose to circumvent these limitations by introducing a novel ADC architecture inspired by the principles of integer-forcing (IF) source coding in [12]. Conventional ADC units fail to exploit any correlation that may exist between the analog signals received at the different antennas, since each signal is independently quantized based on its marginal statistics and it is digitally reconstructed uniquely based on the output bits from the corresponding quantizer. To the contrary, the considered IF-ADC unit maps the signal received at each antenna to a quantization codeword and subsequently reconstructs each channel output by recovering linear combinations of possibly all the quantization codewords from all the quantizers in the ADC unit. This way, the IF-ADC strategy exploits the correlation between the analog received signals, which results in non-negligible improvements in the uplink (UL) achievable rates, particularly, when low-accuracy quantizers are used.

The IF-ADC architecture or, more exactly, (i) the specific analog pre-processing of the channel outputs before quantization and (ii) the joint digital reconstruction are inspired by the results of [12], where the authors present a source-coding scheme for distributed lossy compression of correlated sources with symmetric rate and distortion in the reconstruction of each source. The possibility of using the source-coding scheme in [12] for ADC was already suggested by the authors. However, our analysis is different from [12] in the sense that we are not interested in reconstructing the antenna outputs with some distortion requirements, but in maximizing the UL achievable sum-rate. This requires to adapt the behavior of the IF-ADC unit to the spatial correlation of the received signals.

II. ADC-Constrained System Model

We consider a single-cell uplink (UL) wireless system, in which K single-antenna user equipments (UEs) communicate with a base station (BS) with M antennas (see Fig. 1). The BS processing can be summarized as follows. First, the M analog received signals are converted by an ADC unit of limited resolution which provides a digital reconstruction of the M channel outputs. These digital signals are then fed into a decoder in order to recover the K transmitted messages. Next, we describe each one of these blocks in more detail.

A. Transmitter and Channel Model

The k-th UE wishes to transmit a rate R_k message $\omega_k \in \mathcal{W}_k = \{1, \ldots, 2^{nR_k}\}$ to the BS and uses an encoder $f_k : \mathcal{W}_k \to \mathbb{R}^{1\times n}$ to map the message into a length n channel input sequence $\mathbf{x}_k = \big[x_k(1), \ldots, x_k(n)\big] = f_k(\omega_k)$. We restrict the transmit codebook to be drawn from the Gaussian ensemble, so that each codeword $\mathbf{x}_k$ is a sequence of n independent identically distributed (i.i.d.) zero-mean real random variables, i.e., $x_k(t) \sim \mathcal{N}(0, P_k)$, $t = 1, \ldots, n$, where P_k denotes the transmit power constraint of the k-th UE.

Let $h_{m,k}$ be the channel coefficient between the m-th BS antenna and the k-th UE, then the power-normalized channel output at the m-th BS antenna can be expressed as[1]

$$y_m(t) = \frac{1}{\sigma_m}\Big(\sum_{k=1}^{K} h_{m,k} x_k(t) + z_m(t)\Big), \quad m = 1, \ldots, M \tag{1}$$

where $z_m(t) \sim \mathcal{N}(0, N_0)$ is the t-th sample of the length n i.i.d. additive Gaussian noise sequence at the m-th antenna and

$$\sigma_m^2 = \sum_{k=1}^{K} h_{m,k}^2 P_k + N_0. \tag{2}$$

[1]For ease of exposition, we restrict to real-valued signals and channels and, furthermore, we let the ADC unit to have unit-power inputs.

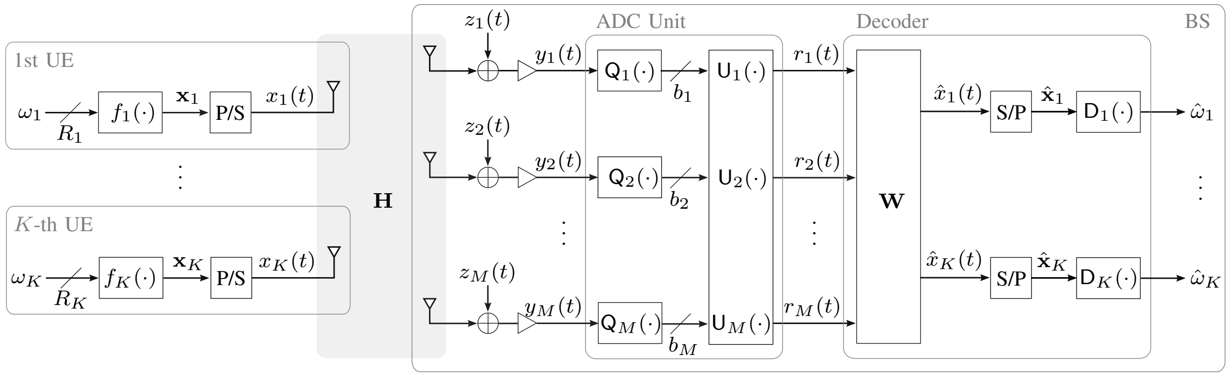

Fig. 1. ADC-Constrained System Model.

B. ADC Unit Model

The channel outputs described by (1) are processed with a limited-resolution ADC unit working on a sample-by-sample basis, which is modeled by the quantization mappings $\mathsf{Q}_m : \mathbb{R} \to \{1,\dots,2^{b_m}\}$ and the (possibly joint) reconstruction mappings $\mathsf{U}_m : \{1,\dots,2^{b_1}\} \times \cdots \times \{1,\dots,2^{b_M}\} \to \mathbb{R}$, for $m = 1,\dots,M$, as depicted in Fig. 1.

The ADC unit is equipped with M quantizers, each connected to one of the M receive antennas. The m-th quantizer has a resolution of b_m-bits and maps sample-by-sample the channel output at the m-th antenna to a quantization index, $\iota_m(t) \in \{1,\dots,2^{b_m}\}$, corresponding to the one of the 2^{b_m} codewords in the quantization codebook $\mathcal{Q}_m$. Thus, applying the m-th quantizer to the m-th received signal generates the length n sequence $\boldsymbol{\iota}_m = [\iota_m(1),\dots,\iota_m(n)]$ with

$$\iota_m(t) = \mathsf{Q}_m\big(y_m(t)\big). \tag{3}$$

The sequence of bits from all M quantizers is used to reconstruct sample-by-sample the received signal. In particular, the t-th sample of the m-th entry of the reconstruction vector $\mathbf{r}(t) = [r_1(t),\dots,r_M(t)]^{\mathrm{T}}$ is obtained (possibly) using the M quantization indices $\boldsymbol{\iota}(t) = [\iota_1(t),\dots,\iota_M(t)]^{\mathrm{T}}$ as

$$r_m(t) = \mathsf{U}_m\big(\boldsymbol{\iota}(t)\big). \tag{4}$$

The previous general ADC unit model is particularized for the conventional ADC and the proposed integer-forcing ADC architectures in Sec. IV and V, respectively.

C. Decoder Model

After the ADC unit, the reconstructed signal sequence $\mathbf{r}(t)$ for $t = 1,\dots,n$ is provided to the decoder, which recovers the K transmitted messages. Here we restrict to a nearest neighbor (NN) decoding structure, which captures the behavior of the decoding strategies used in practical systems and works as follows. The reconstructed channel output $\mathbf{r}(t)$ is first filtered using a linear combiner $\mathbf{w}_k \in \mathbb{R}^{M\times 1}$ to estimate the channel inputs as

$$\hat{x}_k(t) = \mathbf{w}_k^{\mathrm{T}}\mathbf{r}(t), \qquad k = 1,\dots,K, \quad t = 1,\dots,n. \tag{5}$$

Then, the corresponding message ω_k is decoded using a NN-decoding rule, i.e., upon observing $\hat{\mathbf{x}}_k = [\hat{x}_k(1),\dots,\hat{x}_k(n)] \in \mathbb{R}^{1\times n}$, the decoder computes $\mathbf{x}_k = f_k(\omega_k)$, for all possible messages $\omega_k \in \mathcal{W}_k$, and the distance metric

$$D_k(\hat{\mathbf{x}}_k, \omega_k) = \frac{1}{n}\sum_{t=1}^{n}\big|\hat{x}_k(t) - a_k x_k(t)\big|^2 \tag{6}$$

with $a_k \in \mathbb{R}$ selected to optimize the decoding performance. Finally, $\hat{\omega}_k = \mathsf{D}_k(\hat{\mathbf{x}}_k) = \arg\min_{\omega_k\in\mathcal{W}_k} D_k(\hat{\mathbf{x}}_k,\omega_k)$ is selected as the message minimizing the distance metric in (6).

III. Sum-Rate in ADC-Constrained Systems

Given a channel matrix $\mathbf{H} = [h_{m,k}] \in \mathbb{R}^{M\times K}$, a transmit power set $P_1,\dots,P_K$, a per-antenna additive noise power N_0, and an ADC unit structure $(\mathsf{Q}_m,\mathsf{U}_m)_{m=1}^M$ with a resolution of $(b_m)_{m=1}^M$ bits, we say that the sum-rate $R = \sum_{k=1}^K R_k$ is achievable by the ADC-constrained system in Sec. II if, for any $\epsilon > 0$ and a sufficiently long block length n, the message of the k-th UE can be decoded at a rate of at least R_k with vanishing probability of decoding error for all K UEs, i.e., $\mathsf{Pr}\big((\hat{\omega}_1,\dots,\hat{\omega}_K) \neq (\omega_1,\dots,\omega_K)\big) \le \epsilon$.

Under NN-decoding, the highest sum-rate, for which the probability of error averaged over the ensemble of Gaussian codebooks converges to zero as the block length $n \to \infty$, is given by the generalized mutual information (GMI) [13]–[15] introduced in the following lemma.

Lemma 1 ([9, Prop. 1 and 3])**.** *Consider the the ADC-constrained system model in Sec.* II*. The GMI of the k-th UE is*

$$I_{\mathsf{GMI},k} = \frac{1}{2}\log\Big(1 + \frac{\xi_k}{1-\xi_k}\Big), \text{ with} \tag{7}$$

$$\xi_k = \frac{1}{P_k}\boldsymbol{\Sigma}_k^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_k \tag{8}$$

where $\boldsymbol{\Sigma}_k = \mathrm{E}\{\mathsf{r}\mathsf{x}_k\}$ and $\boldsymbol{\Sigma} = \mathrm{E}\{\mathsf{r}\mathsf{r}^{\mathrm{T}}\}$, with r distributed as the i.i.d. sequence $\mathbf{r}(t)$ given $y_1(t),\dots,y_M(t)$, $t = 1,\dots,n$, and x_k distributed as the i.i.d. sequence $x_k(t)$, $t = 1,\dots,n$. The expression in (8) *assumes that the linear combiner $\mathbf{w}_k$ and the decoding parameter a_k are optimally chosen as $\mathbf{w}_k = \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_k$ and $a_k = \xi_k$.*

In consequence, the maximum achievable sum-rate for the ADC-constrained system model is obtained as

$$R = \sum_{k=1}^{K} R_k = \sum_{k=1}^{K} I_{\mathsf{GMI},k} \tag{9}$$

where $I_{\mathsf{GMI},k}$ denotes the GMI of k-th UE as given in Lem. 1. Next we use the expression in (9) to evaluate and optimize the performance of the conventional ADC and the proposed IF-ADC units.

IV. Conventional ADC Unit

In this section we derive the maximum sum-rate achievable with conventional ADC (C-ADC) units, which are composed of M independent quantization-reconstruction chains as depicted in Fig. 2.(a). That is, the m-th received signal is quantized sample-by-sample[2] and independently of the other signals, and reconstructed based only on the output of the m-th quantizer.

[2]Since practical quantizers operate on a sample-by-sample basis, henceforth we omit for the sake of notation the channel use index t.

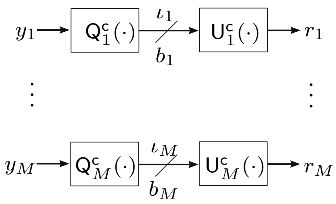

Fig. 2.(a) Coventional ADC Unit

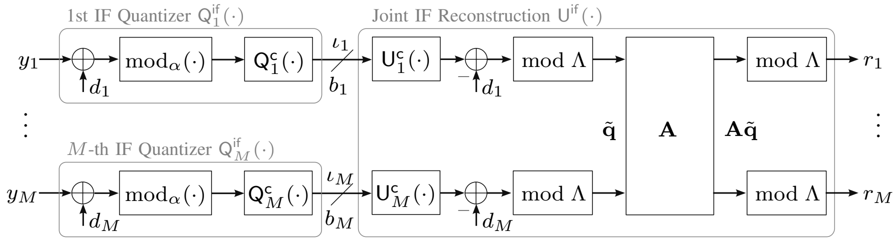

Fig. 2.(b) Integer Forcing ADC Unit.

We assume that the m-th quantizer is a mid-rise uniform quantizer with b_m bits, input range $[-\alpha/2, \alpha/2]$, quantization step-size $\alpha_m = 2^{-b_m}\alpha$, and reconstruction step-size β_m, so that

$$\iota_m = \mathsf{Q}^{\mathsf{c}}_m(y_m) = \max\big(\min\big(\lfloor y_m/\alpha_m \rfloor, 2^{b_m-1}-1\big), -2^{b_m-1}\big) + 2^{b_m-1}+1 \quad (10)$$

$$r_m = \mathsf{U}^{\mathsf{c}}_m(\iota_m) = 2\beta_m\big(\iota_m - 2^{b_m-1} - 1/2\big) \quad (11)$$

with $\lfloor\cdot\rfloor$ denoting the floor function. From Lem. 1, the sum-rate in the next theorem follows.

Theorem 1 ([16])**.** *Consider the ADC-constrained system model from Sec.* II *with a C-ADC unit using M mid-rise uniform quantizers with bit-resolutions, and quantization and reconstruction step-sizes $(b_m, \alpha_m, \beta_m)_{m=1}^M$. The largest achievable sum-rate is given by the sum of the K individual GMIs in Lem.* 1 *with the $M \times M$ correlation matrix $\mathbf{\Sigma}$ and the $M \times 1$ cross-correlation vectors $\mathbf{\Sigma}_k$ defined as*

$$\big[\mathbf{\Sigma}\big]_{m,m} = 2\sum_{i=1}^{2^{b_m-1}} \big(i\beta_m\big)^2 \mathsf{Pr}_i(\alpha_m) \quad (12)$$

$$\big[\mathbf{\Sigma}\big]_{m,m'} = 2\sum_{i,j=1}^{2^{b_m-1}} (i\beta_m)(j\beta_{m'})\big(\mathsf{Pr}_{i,j}(\alpha_m, \alpha_{m'}) - \mathsf{Pr}_{i,j}(\alpha_m, -\alpha_{m'})\big) \quad (13)$$

$$\big[\mathbf{\Sigma}_k\big]_m = \frac{2h_{m,k}P_k}{\sqrt{2\pi}\sigma_m}\sum_{i=1}^{2^{b_m-1}} (i\beta_m)\big(e^{-\frac{((i-1)\alpha_m)^2}{2}} - e^{-\frac{(i\alpha_m)^2}{2}}\big) \quad (14)$$

for $m, m' = 1, \dots, M$ and $m \neq m'$, with

$$\mathsf{Pr}_i(\alpha_m) = Q\big((i-1)\alpha_m\big) - Q\big(i\alpha_m\big) \quad (15)$$

$$\mathsf{Pr}_{i,j}(\alpha_m, \alpha_{m'}) = Q\big((i-1)\alpha_m, (j-1)\alpha_{m'}, \rho_{m,m'}\big) + Q\big(i\alpha_m, j\alpha_{m'}, \rho_{m,m'}\big) - Q\big(i\alpha_m, (j-1)\alpha_{m'}, \rho_{m,m'}\big) - Q\big((i-1)\alpha_m, j\alpha_{m'}, \rho_{m,m'}\big) \quad (16)$$

where σ_m^2 is given in (2)*, $\rho_{m,m'} = (\sigma_m\sigma_{m'})^{-1}\sum_{k=1}^K h_{m,k}h_{m',k}P_k$, and we have used the uni- and bi-dimensional Gaussian Q-functions defined as* [17, eq. (4.1) and eq. (4.3)]

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{u^2}{2}}du \quad (17)$$

$$Q(x,y,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_x^\infty\int_y^\infty e^{-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}}dvdu. \quad (18)$$

Remark 1.1. *Thm.* 1 *provides a generalization of* [9, Prop. 6] *for uniform mid-rise quantizers of $b_m > 1$ bits.*

The sum-rate under the C-ADC unit given in Thm. 1 can be maximized with respect to the quantization step-size α_m and reconstruction step-size β_m of each C-ADC chain by minimizing the mean-squared error (MSE) of quantizing the zero-mean unit-variance Gaussian channel output y_m [15, Sec. V].

V. Integer-Forcing ADC Unit

In this section we describe the IF-ADC unit depicted in Fig. 2.(b), we derive its maximum achievable sum-rate, and we provide an algorithm to adapt the IF-ADC parameters. As opposed to the conventional ADC unit in Fig. 2.(a), the IF-ADC unit in Fig. 2.(b): (i) pre-processes (depending on parameter α) each analog channel output before independently quantizing them, and (ii) jointly reconstructs (using the reconstruction matrix $\mathbf{A}$) all received signals based on the outputs of the M quantizers. The IF-ADC parameters α and $\mathbf{A}$ need to be selected to maximize the achievable sum-rate.

(i) *IF Quantization*: We consider that the IF-ADC unit uses M mid-rise uniform quantizers as defined in Sec. IV, with the m-th quantizer having b_m bits, input range $[-\alpha/2, \alpha/2]$, quantization step-size $\alpha_m = 2^{-b_m}\alpha$ and reconstruction step-size $\beta_m = \alpha_m/2$. The m-th quantization mapping $\mathsf{Q}^{\mathsf{if}}_m(\cdot)$ operates as follows. First, the m-th analog channel output is processed before quantization as

$$\tilde{y}_m = \mathrm{mod}_\alpha\big(y_m + d_m\big) \quad (19)$$

where d_m is a random dither signal uniformly distributed in the range $\big[-\alpha_m/2, \alpha_m/2\big)$, and

$$\mathrm{mod}_\alpha(x) = x - \alpha\big(1 + \lfloor x/\alpha - 1/2 \rfloor\big) \quad (20)$$

denotes a symmetric modulo operation performed in the analog domain. Note that, the quantizer input signal $\tilde{y}_m$ always lies in the quantizer input range $[-\alpha/2, \alpha/2]$. Finally, the quantization index ι_m is obtained by applying the m-th symmetric uniform quantizer to analog pre-processed signal $\tilde{y}_m$, i.e., $\iota_m = \mathsf{Q}^{\mathsf{if}}_m(y_m) = \mathsf{Q}^{\mathsf{u}}_m(\tilde{y}_m)$.

(ii) *IF Reconstruction*: The joint reconstruction mapping $\mathsf{U}^{\mathsf{if}}(\cdot)$ operates as follows. First, the quantization codewords corresponding to the M quantization indices are recovered using the C-ADC reconstruction mapping in (11) for $\beta_m = \alpha_m/2$:

$$q_m = \mathsf{U}^{\mathsf{c}}_m(\iota_m) = \alpha_m\big(\iota_m - 2^{b_m-1} - 1/2\big) \quad (21)$$

$$= \alpha_m\big(\lfloor \mathrm{mod}_\alpha\big(y_m + d_m\big)/\alpha_m \rfloor + 1/2\big). \quad (22)$$

Following [12], the resulting quantization codebooks can be described resorting to one-dimensional lattices [18] as $\mathcal{Q}_m = \Lambda_m \cap \mathcal{V}(\Lambda)$, where $\Lambda \subseteq \Lambda_m$ form M pairs of nested lattices with $\Lambda = \alpha \cdot \mathbb{Z}$ and $\Lambda_m = \alpha_m \cdot \mathbb{Z}$, and $\mathcal{V}(\Lambda) = [\alpha/2, \alpha/2)$ denotes the fundamental Voronoi region of Λ. Accordingly, the quantization codewords in (22) can be equivalently written as

$$q_m = \big[\mathsf{Q}_{\Lambda_m}\big(y_m + d_m - \alpha_m/2\big)\big] \bmod \Lambda + \alpha_m/2 \quad (23)$$

where $\mathsf{Q}_{\Lambda_m}(\cdot)$ denotes the lattice-quantization with respect to Λ_m and $[\cdot] \bmod \Lambda$ denotes lattice-modulo operation with respect to Λ. Then, the dither $\mathbf{d} = \big[d_1, \dots, d_M\big]^{\mathrm{T}}$ is removed from the M quantization codewords $\mathbf{q} = \big[q_1, \dots, q_M\big]^{\mathrm{T}}$ and the resulting vector is modulo-Λ reduced:

$$\tilde{\mathbf{q}} = \big[\mathbf{q} - \mathbf{d}\big] \bmod \Lambda = \big[\mathbf{y} - \mathbf{z}_{\mathsf{Q}}\big] \bmod \Lambda \quad (24)$$

where $\mathbf{z}_{\mathsf{Q}} = \left[z_{\mathsf{Q},1}, \ldots, z_{\mathsf{Q},M}\right]^{\mathrm{T}}$ denotes the quantization error with

$$z_{\mathsf{Q},m} = \left[y_m + d_m - \alpha_m/2\right] \bmod \Lambda_m \tag{25}$$

uniformly distributed in $\mathcal{V}(\Lambda_m) = [-\alpha_m/2, \alpha_m/2)$ and independent of y_m, due to the Crypto lemma [18, Lem. 4.1.1], so that

$$\mathrm{E}\{\mathbf{z}_{\mathsf{Q}}\mathbf{z}_{\mathsf{Q}}^{\mathrm{T}}\} = (\alpha^2/12)\mathbf{D}_{\boldsymbol{b}} = (\alpha^2/12)\mathrm{diag}(2^{-2b_1}, \cdots, 2^{-2b_M}). \tag{26}$$

Finally, given the full row-rank integer valued matrix $\mathbf{A} = \left[\mathbf{a}_1^{\mathrm{T}}, \ldots \mathbf{a}_M^{\mathrm{T}}\right] \in \mathbb{Z}^{M\times M}$, the reconstruction vector $\mathbf{r}$ is computed as

$$\mathbf{r} = \left[r_1, \ldots, r_M\right]^{\mathrm{T}} = \left[\mathbf{A}\tilde{\mathbf{q}}\right] \bmod \Lambda = \left[\mathbf{A}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})\right] \bmod \Lambda. \tag{27}$$

Observe that an error in the reconstruction of the m-th entry of $\mathbf{r}$, i.e. r_m, occurs if $\mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})$ lies outside $\mathcal{V}(\Lambda) = [-\alpha/2, \alpha/2)$. Thus, we have that $r_m = \mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})$ with probability $\mathsf{Pr}(\alpha, \mathbf{a}_m) = \mathsf{Pr}\big(|\mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})| \leq \alpha/2\big)$ and r_m behaves as random signal uniformly distributed in $\mathcal{V}(\Lambda)$ with probability $1 - \mathsf{Pr}(\alpha, \mathbf{a}_m)$.

Given the IF-ADC parameters, α and $\mathbf{A}$, the sum-rate in the next theorem follows.

Theorem 2 ([16]). *Consider the ADC-constrained system model from Sec.* II *with an IF-ADC unit using M mid-rise uniform quantizers with bit-resolutions, quantization and reconstruction step-sizes $\big(b_m, 2^{-b_m}\alpha, 2^{-b_m-1}\alpha\big)_{m=1}^{M}$, and reconstruction matrix $\mathbf{A}$. The best achievable sum-rate is given by the sum of the K GMIs in Lem.* 1 *with the $M \times M$ correlation matrix $\mathbf{\Sigma}$ and the $M \times 1$ cross-correlation vectors $\mathbf{\Sigma}_k$ defined as*

$$\big[\mathbf{\Sigma}(\alpha, \mathbf{A})\big]_{m,m} = \big(\mathbf{a}_m^{\mathrm{T}}\big(\mathbf{\Sigma}_{\mathbf{yy}} + (\alpha^2/12)\mathbf{D}_{\boldsymbol{b}}\big)\mathbf{a}_m\big)\mathsf{Pr}(\alpha, \mathbf{a}_m) + (\alpha^2/12)\big(1 - \mathsf{Pr}(\alpha, \mathbf{a}_m)\big) \tag{28}$$

$$\big[\mathbf{\Sigma}(\alpha, \mathbf{A})\big]_{m,m'} = \big(\mathbf{a}_m^{\mathrm{T}}\big(\mathbf{\Sigma}_{\mathbf{yy}} + (\alpha^2/12)\mathbf{D}_{\boldsymbol{b}}\big)\mathbf{a}_{m'}\big)\big(\mathsf{Pr}(\alpha, \mathbf{a}_m, \mathbf{a}_{m'}) - \mathsf{Pr}(\alpha, \mathbf{a}_m, -\mathbf{a}_{m'})\big) \tag{29}$$

$$\big[\mathbf{\Sigma}_k(\alpha, \mathbf{A})\big]_{m} = \big(\mathbf{a}_m^{\mathrm{T}}\mathbf{\Sigma}_{\mathbf{y}x_k}\big)\mathsf{Pr}(\alpha, \mathbf{a}_m) \tag{30}$$

for $m \neq m' = 1, \ldots M$, with $\mathbf{\Sigma}_{\mathbf{yy}} = \mathrm{E}\{\mathbf{y}\mathbf{y}^{\mathrm{T}}\}$, $\mathbf{\Sigma}_{\mathbf{y}x_k} = \mathrm{E}\{\mathbf{y}x_k\}$, and[3]

$$\mathsf{Pr}(\alpha, \mathbf{a}_m) \approx 1 - 2Q\Big(\frac{\alpha}{2\sigma(\mathbf{a}_m)}\Big) \tag{31}$$

$$\mathsf{Pr}(\alpha, \mathbf{a}_m, \mathbf{a}_{m'}) \approx Q\Big(\frac{\alpha}{2\sigma(\mathbf{a}_m)}, \frac{\alpha}{2\sigma(\mathbf{a}_{m'})}, \rho(\mathbf{a}_m, \mathbf{a}_{m'})\Big) + Q\Big(-\frac{\alpha}{2\sigma(\mathbf{a}_m)}, -\frac{\alpha}{2\sigma(\mathbf{a}_{m'})}, \rho(\mathbf{a}_m, \mathbf{a}_{m'})\Big) \tag{32}$$

where we have used the Gaussian Q-functions in (17) *and* (18) *and*

$$\sigma^2(\mathbf{a}_m) = \mathbf{a}_m^{\mathrm{T}}(\mathbf{\Sigma}_{\mathbf{yy}} + (\alpha^2/12)\mathbf{D}_{\boldsymbol{b}})\mathbf{a}_m \tag{33}$$

$$\rho(\mathbf{a}_m, \mathbf{a}_{m'}) = \frac{1}{\sigma_m(\mathbf{a}_m)\sigma(\mathbf{a}_{m'})}\mathbf{a}_m^{\mathrm{T}}(\mathbf{\Sigma}_{\mathbf{yy}} + (\alpha^2/12)\mathbf{D}_{\boldsymbol{b}})\mathbf{a}_{m'}. \tag{34}$$

The IF-ADC architecture supports, then, a sum-rate of

$$R^{\mathsf{if}}(\alpha, \mathbf{A}) = -\frac{1}{2}\sum_{k=1}^{K}\log_2\big(1 - \mathbf{\Sigma}_k^{\mathrm{T}}(\alpha, \mathbf{A})\mathbf{\Sigma}^{-1}(\alpha, \mathbf{A})\mathbf{\Sigma}_k(\alpha, \mathbf{A})\big) \tag{35}$$

where the entries $\big[\mathbf{\Sigma}(\alpha, \mathbf{A})\big]_{m,m}$ and $\big[\mathbf{\Sigma}_k(\alpha, \mathbf{A})\big]_{m}$ take into account the the probability of the m-th reconstruction in (27) being successful, $\mathsf{Pr}(\alpha, \mathbf{a}_m)$, and the entries $\big[\mathbf{\Sigma}(\alpha, \mathbf{A})\big]_{m,m'}$, with $m \neq m'$, include the probability of both the m-th and m'-th reconstructions in (27) being successful $\mathsf{Pr}(\alpha, \mathbf{a}_m, \mathbf{a}_{m'}) - \mathsf{Pr}(\alpha, \mathbf{a}_m, -\mathbf{a}_{m'}) = \mathsf{Pr}\big(|\mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})| \leq \alpha/2, |\mathbf{a}_{m'}^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})| \leq \alpha/2\big)$. Unfortunately, the expressions in (28)-(30) do not lead to a tractable approach for obtaining the IF-ADC parameters α and $\mathbf{A}$, which maximize the sum-rate in (35). Here we propose to simplify the optimization problem as follows.

First, observe that all M entries of the reconstruction vector $\mathbf{r}$ in (27) are successful, i.e., $r_m = \mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})$, for $m = 1, \ldots, M$, with probability

$$\mathsf{Pr}\big(\alpha, \mathbf{A}\big) = 1 - \mathsf{Pr}\big(\max_m \big|\mathbf{a}_m^{\mathrm{T}}(\mathbf{y} - \mathbf{z}_{\mathsf{Q}})\big| > \alpha/2\big) \tag{36}$$

$$\geq 1 - 2\sum_{m=1}^{M}\exp\Big(-\frac{\alpha^2}{8\sigma^2(\mathbf{a}_m)}\Big) \tag{37}$$

where (37) comes from the analysis in [12]. Hence, we can conclude that a good approach to select the reconstruction matrix $\mathbf{A}$ is to maximize the lower bound on $\mathsf{Pr}(\mathbf{A}, \alpha)$ for some fixed $\alpha > 0$ as

$$\begin{aligned}\mathbf{A}^{\star} &= \arg\min_{\mathbf{A}\in\mathbb{Z}^{M\times M},\ \det(\mathbf{A})\neq 0}\ \max_m\ \sigma^2(\mathbf{a}_m)/\alpha^2 \\ &= \arg\min_{\mathbf{A}\in\mathbb{Z}^{M\times M},\ \det(\mathbf{A})\neq 0}\ \max_m\ \|\mathbf{\Gamma}(\alpha)^{\mathrm{T}}\mathbf{a}_m\|^2\end{aligned} \tag{38}$$

where $\mathbf{\Gamma}(\alpha)$ comes from the Cholesky decomposition: $\mathbf{\Gamma}(\alpha)\mathbf{\Gamma}(\alpha)^{\mathrm{T}} = ((12/\alpha^2)\mathbf{\Sigma}_{\mathbf{yy}} + \mathbf{D}_{\boldsymbol{b}})$. Although the problem in (38) is conjectured to be NP-hard, its solution can be approximated in polynomial time using lattice reduction algorithms, e.g., the LLL algorithm [19].

Once we have fixed the reconstruction matrix $\mathbf{A}$, we can maximize the sum-rate in (35) with respect to parameter $\alpha > 0$ and iterate between both optimization problems as shown in the following algorithm.

Algorithm 1 IF-ADC Parameter Optimization

initialization
  Set $i \leftarrow 0$ and choose $\alpha^{(0)} > 0$.
repeat
  $\mathbf{\Gamma}(\alpha^{(i)}) = \texttt{CHOLESKY}\big((12/(\alpha^{(i)})^2)\mathbf{\Sigma}_{\mathbf{yy}} + \mathbf{D}_{\boldsymbol{b}}\big)$
  $\mathbf{A}^{(i+1)} = \texttt{LLL}\big(\mathbf{\Gamma}(\alpha^{(i)})^{\mathrm{T}}\big)$.
  $\alpha^{(i+1)} = \arg\max_{\alpha>0} R^{\mathsf{if}}\big(\alpha, \mathbf{A}^{(i+1)}\big)$.
  $i \leftarrow i + 1$.
until convergence.

VI. Numerical Evaluation

In this section we evaluate the performance of the proposed IF-ADC strategy and compare it with respect to conventional ADCs through numerical simulations. We consider a single-cell UL wireless system, in which a BS with $M = 64$ antennas using ADC units with resolutions of $b = \{4, 5\}$ bits serves $K = 4$ single-antenna UEs with transmit power constraint P_k.

We compute the sum-rate supported by the considered ADC-constrained system averaged over 1000 uncorrelated Rayleigh channel realizations for (i) the C-ADC unit of Sec. IV using Thm. 1 with the optimum quantization step-sizes and reconstruction levels obtained by minimizing the MSE; and (ii) the IF-ADC unit of Sec. V using Thm. 2 with the IF-ADC parameters obtained through Algorithm 1. For reference, we also include the achievable sum-rate for (iii) an infinite-precision ADC unit obtained using Lem. 1 with the reconstructed channel outputs being equal the actual channel outputs. This serves as an upper bound on the achievable sum-rate under the system model in Sec. II.

[3]The approximations in (31) and (32) follow from assuming that the equivalent quantization error $\mathbf{A}\mathbf{z}_{\mathsf{Q}}$ is Gaussian distributed, which can be justified in the large system limit. The exact expressions are given in [16].

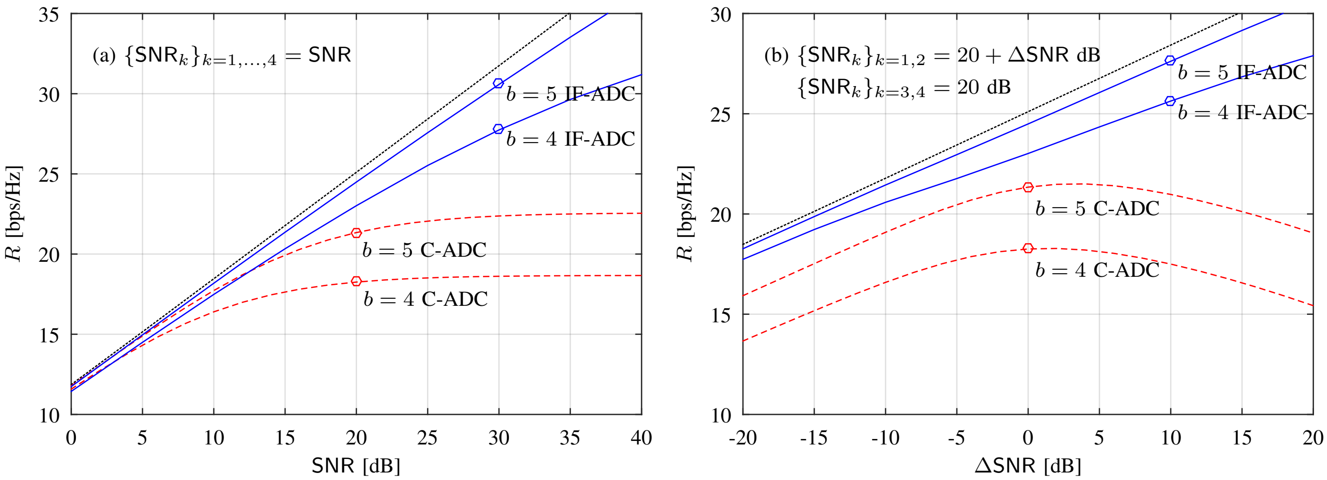

Fig. 3. Average achievable sum-rate R for $M = 64$, $K = 4$, and $b = \{4, 5\}$ bits using the C-ADC unit (solid blue), the proposed IF-ADC unit (dashed red), and the infinite-precision ADC unit (dotted black) as (a) a function of the received SNR, and (b) as a function of the received SNR imbalance $\Delta\mathsf{SNR}$.

In particular, Fig. 3.(a) shows the average achievable sum-rate R as function of the per-antenna average received SNR, $\mathsf{SNR}_k = P_k \mathrm{E}\{\|\mathbf{h}_k\|^2\}/(MN_0) = P_k/N_0$, when all $K = 4$ users transmit with equal power, i.e., $P_k = P$. As expected, the IF-ADC unit supports higher sum-rates than conventional ADCs when low-accuracy quantization is adopted. This gap can be substantially high in the medium- to high-SNR regime.

Fig. 3.(b) shows the average achievable sum-rate R as function of the per-antenna average received SNR imbalance $\Delta\mathsf{SNR}$, when users 1 and 2 are received with $\mathsf{SNR}_k = 20 + \Delta\mathsf{SNR}$ dB and users 3 and 4 are received with $\mathsf{SNR}_k = 20$ dB. Observe that the sum-rate achievable by the C-ADC unit decreases, as the received power imbalance $\Delta\mathsf{SNR}$ slightly deviates from 0 dB. The quantization step-sizes and reconstruction levels in C-ADC units are adapted to the marginal statistics of the signal received at each antenna, which is a noisy combination of the signals transmitted by the different users. In consequence, users received with higher powers are better adapted to the quantization parameters. In contrast, the proposed IF-ADC unit can better deal with received power imbalances, as this information is captured by the IF-ADC parameter adaptation procedure in Alg. 1. This is of high relevance and one of the main benefits of the IF-ADC architecture, since received power imbalances are always present in wireless communication systems, even when some kind of power control strategies are implemented.

VII. Conclusion

In this paper we have presented a novel ADC architecture inspired by the principles of integer-forcing source coding. The considered IF-ADC unit pre-processes the channel outputs before quantization and jointly reconstructs the received signals in the digital domain by jointly processing all quantization indices. This way, the correlation among the received signals is exploited and the IF-ADC strategy achieves higher UL sum-rates than conventional ADC units and better deals with receive power imbalance among the different users.

References

[1] B. Murmann, “ADC performance survey 1997-2016.” [Online]. Available: http://web.stanford.edu/ murmann/adcsurvey.html.

[2] J. Liu and H. Minn, “The death of 5G part 2: Will analog be the death of massive MIMO?” ComSoc Technology News Editorial, http://www.comsoc.org/ctn, Jun. 2015.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, “Massive MIMO with low-resolution ADCs,” *CoRR*, vol. abs/1602.01139, Feb. 2016. [Online]. Available: http://arxiv.org/abs/1602.01139.

[6] L. Fan, S. Jin, C. K. Wen, and H. Zhang, “Uplink achievable rate for massive MIMO systems with low-resolution ADC,” *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[7] C. Studer and G. Durisi, “Quantized massive MU-MIMO-OFDM uplink,” *IEEE Trans.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.

[8] J. Zhang, L. Dai, S. Sun, and Z. Wang, “On the spectral efficiency of massive MIMO systems with low-resolution ADCs,” *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, May 2016.

[9] N. Liang and W. Zhang, “Mixed-ADC massive MIMO,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[10] T. Zhang, C. K. Wen, S. Jin, and T. Jiang, “Mixed-ADC massive MIMO detectors: Performance analysis and design optimization,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.

[11] W. Tan, S. Jin, C. K. Wen, and Y. Jing, “Spectral efficiency of mixed-ADC receivers for massive MIMO systems,” *IEEE Access*, Aug. 2016.

[12] O. Ordentlich and U. Erez, “Integer-forcing source coding,” *CoRR*, vol. abs/1308.6552, 2013. [Online]. Available: http://arxiv.org/abs/1308.6552.

[13] A. Lapidoth and S. Shamai, “Fading channels: how perfect need “perfect side information” be?” *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[14] A. Ganti, A. Lapidoth, and E. Telatar, “Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[15] W. Zhang, “A general framework for transmission with transceiver distortion and some applications,” *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 384–399, Feb. 2012.

[16] L. G. Ordóñez , I. Estella, and M. Guillaud, “Integer forcing analog-to-digital conversion for massive MIMO systems,” *in Preparation*.

[17] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels*, 2nd ed., ser. Wiley series in telecommunications and signal processing. New York: Wiley, 2002.

[18] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. New York, NY, USA: Cambridge University Press, 2014.

[19] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, “Lattice reduction: A survey with applications in wireless communications,” *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.