Outage Behavior of Integer Forcing With Random Unitary Pre-Processing

Elad Domanovitz and Uri Erez

Abstract—Integer forcing is an equalization scheme for the multiple-input multiple-output communication channel that has been demonstrated to allow operating close to capacity for "most" channels. In this work, the measure of "bad" channels is quantified by considering a compound channel setting where the transmitter communicates over a fixed channel but knows only its mutual information. The transmitter encodes the data into independent streams, all taken from the same linear code. The coded streams are transmitted after applying a unitary transformation. At the receiver side, integer-forcing equalization is applied, followed by standard single-stream decoding. Considering pre-processing matrices drawn from a random ensemble, outage corresponds to the event that the target rate exceeds the achievable rate of integer forcing for a given channel matrix. For the case of the circular unitary ensemble, an explicit universal bound on the outage probability for a given target rate is derived that holds for any channel in the compound class. The derived bound depends only on the gap-to-capacity and the number of transmit antennas. The results are also applied to obtain universal bounds on the gap-to-capacity of multiple-antenna closed-loop multicast, achievable via linear pre-processed integer forcing.

I. INTRODUCTION

The Multiple-Input Multiple-Output (MIMO) Gaussian channel is central to modern communication and has been extensively studied over the past several decades. Nonetheless, while the capacity limits, under different assumptions on the availability of channel state information, are well understood, the design of low-complexity communication schemes that approach these limits still poses challenges in some scenarios.

For a static channel and a point-to-point closed-loop setting, capacity may be approached without much difficulty by employing an architecture that decouples coding and modulation. That is, one may use "off-the-shelf" codes in conjunction with linear pre- and post-processing based on matrix decompositions. For instance, one may use the singular-value decomposition (SVD) to transform the channel into parallel scalar additive white Gaussian noise (AWGN) channels, over which standard codes may be employed [1]. Alternatively, standard scalar codes may be used in conjunction with the QR decomposition and successive interference cancellation (SIC), see, e.g., [2]. Coding for MIMO channels in an ergodic fading environment is more involved but has also been successfully addressed. See, e.g., [3].

In contrast, we address the problem of coding over a compound MIMO channel. More specifically, the focus of this paper is on static (and frequency-flat) MIMO channels where the transmitter only knows (or may only utilize its knowledge of) the mutual information of the channel.

The design of a practical coding scheme for such a compound MIMO channel scenario was addressed in [4] where an architecture employing space-time linear pre-processing (that is independent of the channel) at the transmitter side and integer-forcing (IF) equalization at the receiver side was proposed. It was shown that such an architecture *universally* achieves the MIMO capacity up to a constant gap, provided the space-time pre-processing satisfies the non-vanishing determinant (NVD) criterion [5]. While this result is encouraging as it points to the robustness of the IF scheme, the derived gap is very large and calls for further work.

In the present work, we study the performance of IF where random unitary linear pre-processing is performed over the spatial dimension only. Rather than aiming at guaranteeing successful transmission, we study the outage probability of the scheme.[1] We focus attention to pre-processing matrices drawn from the isotropic (circular) unitary ensemble as this ensures that all channels having the same singular values, will have the same outage probability.

It is worth noting that the random pre-processing operation serves the purpose of quantifying the measure of "bad" channels for IF receivers. We further note that the receiver considered is the standard one of [8] but whereas the results of [8] deal with distributed transmit antennas (i.e., with no encoding across the transmit antennas), in the present paper joint (unitary) pre-processing is assumed. While this may appear to preclude a distributed setting, some important statistical scenarios are in fact covered by the model. Specifically, in the case of a channel matrix whose entries are i.i.d. Gaussian random variables, the random unitary transformation assumed in the analysis to follow is in reality performed by nature, as discussed in Section III-C2 below.

The outage probability in the considered setting thus corresponds to a "scheme outage".[2] Namely, it is the probability that a random linear pre-processing matrix results in an effective channel for which the rate achievable with an IF receiver is smaller than the target rate. In order to provide universal performance guarantees, we study the *worst-case*

---

[1]This approach is similar to that taken in [6] and [7] with respect to other transmission schemes.

[2]We use the term "scheme outage" as opposed to "channel outage". Specifically, in the considered setting, the channel is known to have sufficient mutual information to support the chosen transmission rate.

outage probability with respect to all possible singular value combinations corresponding to a given mutual information. Thus, the guaranteed performance does not depend on channel statistics.

We note that the performance of a coding scheme over the compound channel is a strong measure of its robustness. Clearly, performance guarantees for the compound channel immediately translate to guarantees for a statistical channel model (as explained in the next section). To the best of our knowledge, IF is the first practical scheme for which (provable) universal bounds are known for the MIMO channel.

We begin by empirically observing that space-only linear pre-processing IF (P-IF) has much better performance in terms of worst-case outage probability when compared to linear equalization. We then derive an explicit bound on the performance of P-IF that depends only on the number of transmit antennas and the gap-to-capacity, where moderate gaps suffice to guarantee a small outage probability.

As another example of an application of the results, we use the probabilistic method to obtain guarantees on the number of users that can be supported in closed-loop MIMO multicast (guaranteeing no outage occurs) as a function of the gap-to-capacity, when using linear pre-processed IF.

The paper is organized as follows. Section II defines the channel model of interest and formulates the problem described above. Section III provides background on the integer-forcing receiver as well as its use in conjunction with linear pre-processing. Section IV derives a universal upper bound for the outage probability of randomly linear pre-processed IF over the compound MIMO channel; tighter bounds for the specific case of two transmit antennas for a receiver employing a successive interference cancellation (SIC) variant of IF are also derived. Section V describes the application of the derived bounds to a close-loop MIMO multicast setting.

II. Channel Model and Problem Formulation

A point-to-point (complex) MIMO channel is considered where the transmitter is equipped with N_t antennas and the receiver is equipped with N_r antennas. Thus, the channel is described by the relation[3]

$$\boldsymbol{y}_c = \mathbf{H}_c \boldsymbol{x}_c + \boldsymbol{z}_c, \tag{1}$$

where $\mathbf{x}_c \in \mathbb{C}^{N_t}$ is the channel input vector, $\mathbf{y}_c \in \mathbb{C}^{N_r}$ is the channel output vector, $\mathbf{H}_c$ is an $N_r \times N_t$ complex channel matrix, and $\boldsymbol{z}_c$ is an additive noise vector of i.i.d. unit variance circularly-symmetric complex Gaussian random variables. The input vector $\boldsymbol{x}_c$ is subject to the power constraint[4]

$$\mathbb{E}(\boldsymbol{x}_c^H \boldsymbol{x}_c) \leq N_t \cdot \mathsf{SNR}. \tag{2}$$

We assume that the channel is fixed throughout the whole transmission of a codeword.

For a given input covariance matrix $\mathbf{Q}_c$, satisfying the power constraint $\text{Tr}(\mathbf{Q}_c) \leq N_t \mathsf{SNR}$, the mutual information of the

channel (1) is maximized by a Gaussian input, and is given by

$$C = \log \det \left(\mathbf{I}_{N_r \times N_r} + \mathbf{H}_c \mathbf{Q}_c \mathbf{H}_c^H \right) \tag{3}$$

When it comes to designing transmission strategies, without loss of generality we may assume that $\mathbf{Q} = \mathbf{I}$ (isotropic transmission). Namely, we may "absorb" $\mathbf{Q}$ into the channel matrix by replacing $\mathbf{H}_c$ in (3) with $\bar{\mathbf{H}}_c = \mathbf{H}_c \mathbf{Q}^{1/2}$ (and with abuse of notation, we omit the bar). Similarly, we may set $\mathsf{SNR} = 1$. Hence, we may rewrite (3) as

$$C = \log \det \left(\mathbf{I} + \mathbf{H}_c \mathbf{H}_c^H \right). \tag{4}$$

We define the set

$$\mathbb{H}(C) = \left\{ \mathbf{H}_c \in \mathbb{C}^{N_r \times N_t} : \log \det \left(\mathbf{I} + \mathbf{H}_c \mathbf{H}_c^H \right) = C \right\} \tag{5}$$

of all channel matrices having the same WI mutual information C.

The corresponding compound channel model is defined by (1) with the channel matrix $\mathbf{H}_c$ arbitrarily chosen from the set $\mathbb{H}(C)$. The matrix $\mathbf{H}_c$ that was chosen by nature is revealed to the receiver, but not to the transmitter. Clearly, the capacity of this compound channel is C, and is achieved with an isotropic Gaussian input.

Employing the IF receiver allows approaching C for "most" but not all matrices $\mathbf{H}_c \in \mathbb{H}(C)$. We quantify the measure of the set of bad channel matrices by considering outage events, i.e., those events where integer forcing fails even though the channel has sufficient mutual information. More broadly, for a given coding scheme, denote the achievable rate for a given channel matrix $\mathbf{H}_c$ as $R_{\text{scheme}}(\mathbf{H}_c)$. Then, given a target rate $R < C$ and a channel $\mathbf{H}_c \in \mathbb{H}(C)$, the scheme is in outage when $R_{\text{scheme}}(\mathbf{H}_c) < R$. For the case of integer forcing, the explicit expression for $R_{\text{IF}}(\mathbf{H}_c)$ is recalled in Section III-A.

Since applying a linear pre-processing matrix $\mathbf{P}_c$ results in an effective channel $\mathbf{H}_c \cdot \mathbf{P}_c$, it follows that the achievable rate of a transmission scheme over this channel is $R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c)$. When $\mathbf{P}_c$ is drawn at random, the latter rate is also random. The worst-case (WC) scheme outage probability is defined in turn as

$$P_{\text{out,scheme}}^{\text{WC}}(C, R) = \sup_{\mathbf{H}_c \in \mathbb{H}(C)} \Pr \left(R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c) < R \right), \tag{6}$$

where the probability is over the ensemble of linear pre-processing matrices. The goal of this paper is to quantify the tradeoff between the transmission rate R and the outage probability $P_{\text{out,IF}}^{\text{WC}}(C, R)$ as defined in (6).

Remark 1. *Assume that $\mathbf{H}_c$ is modeled as having a probability distribution over the compound class $\mathbb{H}(C)$. In such a case, the outage probability is given by*

$$\mathbb{E}_{\mathbf{H}_c} \left[\Pr \left(R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c) < R \mid \mathbf{H}_c \right) \right] \tag{7}$$

where the probability is with respect to $\mathbf{P}_c$ (given $\mathbf{H}_c$) and the expectation is with respect to the distribution over $\mathbb{H}(C)$.

Note that in (6), we take the supremum over the entire compound class rather than averaging over a given distribution.

---

[3]We denote all complex variables with c to distinguish them from their real-valued representation.

[4]We denote by $[\cdot]^T$, transpose of a vector/matrix and by $[\cdot]^H$, the Hermitian transpose of a vector/matrix.
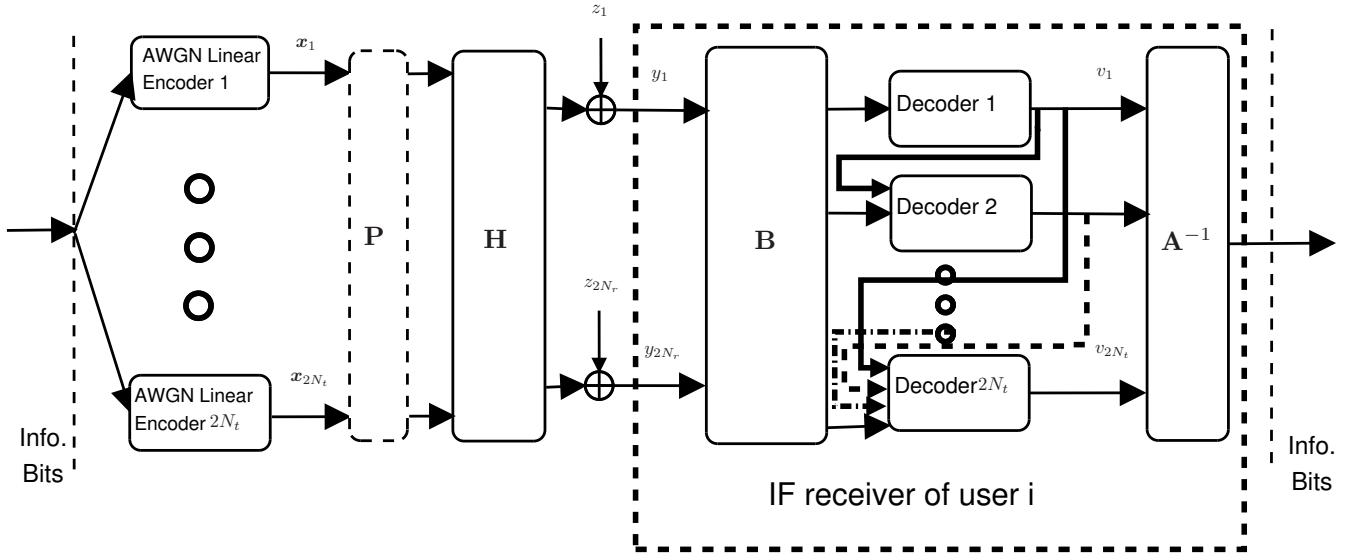
Fig. 1. Linear pre-processed-IF / Linear pre-processed IF-SIC scheme. Feedback in the receiver is active only when SIC is used.

Since the average is always smaller than the supremum, it follows that (6) universally upper bounds (7). That is, the bound holds for any distribution over $\mathbb{H}(C)$.

Remark 2. *Assume that $\mathbf{H}_c$ is modeled as having any probability distribution (not restricted to the compound class $\mathbb{H}(C)$). In such a case, the outage probability can be expressed as*

$$\mathbb{E}\left[\mathbb{E}_{\mathbf{H}_c}\left[\Pr\left(R_{\text{scheme}}(\mathbf{H}_c \cdot \mathbf{P}_c) < R \mid \mathbf{H}_c\right) \,\Big|\, C\right]\right] \quad (8)$$

where again, the probability is with respect to $\mathbf{P}_c$, the inner expectation is with respect to the marginal distribution of $\mathbf{H}_c$ given the WI-MI C, while the outer one is with respect to C. Thus, in order to bound the scheme outage probability, it suffices to know only the distribution of the WI-MI mutual information C. Namely, from (6) we have that the outage probability will be no greater than

$$\mathbb{E}\left[P_{\text{out,scheme}}^{\text{WC}}(C, R)\right] \quad (9)$$

where the expectation is over C.

III. INTEGER-FORCING BACKGROUND

A. Single-User Integer-Forcing Equalization

In [8], a receiver architecture scheme named "integer forcing" was proposed which we briefly recall. For our purposes, it will suffice to only state the achievable rates of IF and a high-level operational description of its elements. The reader is referred to [8] for the derivation, details and proofs, and further to [9], [10] and references therein for implementation considerations.

We follow the derivation of [8] and describe integer forcing over the reals. Channel model (1) can be expressed via its real-valued representation as

$$\underbrace{\begin{bmatrix} \text{Re}(\mathbf{y}_c) \\ \text{Im}(\mathbf{y}_c) \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \text{Re}(\mathbf{H}_c) & -\text{Im}(\mathbf{H}_c) \\ \text{Im}(\mathbf{H}_c) & \text{Re}(\mathbf{H}_c) \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} \text{Re}(\mathbf{x}_c) \\ \text{Im}(\mathbf{x}_c) \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} \text{Re}(\mathbf{z}_c) \\ \text{Im}(\mathbf{z}_c) \end{bmatrix}}_{\mathbf{z}}.$$

$$(10)$$

This real-valued representation is used in the sequel to derive performance bounds for the complex channel $\mathbf{H}_c$. Note that the dimensions of $\mathbf{H}$ are $2N_r \times 2N_t$.

It is assumed that information bits are fed into $2N_t$ encoders, each of which uses the same *linear* code that is designed for an AWGN channel.[5] The latter produces $2N_t$ channel inputs (for example, x_m for the m'th antenna).[6] At the receiver, a *linear equalization* matrix $\mathbf{B} \in \mathbb{R}^{2N_t \times 2N_r}$ is applied. It is easiest to understand IF by first describing its zero-forcing variant. In this case, $\mathbf{B}$ is designed so that the resulting equivalent channel $\mathbf{A} = \mathbf{BH}$ is such that $\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t}$ is a full-rank integer matrix. In a practical implementation, it may be necessary for the matrix to be full-rank over a finite field $\mathbb{Z}_p$ (where p is prime) over which the code is defined. Nonetheless, by taking p large enough, it suffices for $\mathbf{A}$ to be invertible over the reals (See Appendix A, Lemma 2 in [11] and also [10]). This ensures that the output of the channel (without noise) after applying a modulo operation is a valid codeword.

Each of the equalized streams is next passed to a standard (up to the additional element of a modulo operation) AWGN decoder which tries to decode a linear combination of codewords, whose coefficients correspond to a row of $\mathbf{A}$. Finally, after the noise is removed, the original messages are recovered

---

[5]The effect of the chosen code on the overall performance of IF is discussed in [4].

[6]For simplicity of notation the time index is suppressed as the block length plays no role in our description. Of course, to approach capacity, one needs to use a long block.

3

by applying the inverse of $\mathbf{A}$. Thus, for IF equalization to be successful, decoding over all $2N_t$ subchannels should be successful and the worst subchannel constitutes a bottleneck. The operation of the receiver is depicted in Figure 1 (where at this stage the linear pre-processing matrix can be considered as the identity matrix, i.e., $\mathbf{P} = \mathbf{I}$).

When using minimum mean square error (MMSE) equalization, rather than zero-forcing, the linear equalizer takes the form

$$\mathbf{B} = \mathbf{A}\mathbf{H}^T \left(\mathbf{I} + \mathbf{H}\mathbf{H}^T\right)^{-1}, \tag{11}$$

and the input to the m'th decoder is

$$\boldsymbol{y}_{\text{eff,m}} = \boldsymbol{v}_m + \boldsymbol{z}_{\text{eff,m}} \tag{12}$$

where

$$\boldsymbol{z}_{\text{eff,m}} = (\boldsymbol{b}_m^T \mathbf{H} - \boldsymbol{a}_m^T)\boldsymbol{x} + \boldsymbol{b}_m^T \boldsymbol{z}. \tag{13}$$

Here, $\mathbf{a}_m^T$ and $\mathbf{b}_m^T$ are the m'th row of $\mathbf{A}$ and $\mathbf{B}$ respectively. We can define the effective SNR at the m'th subchannel as

$$\text{SNR}_{\text{eff}}(\mathbf{a}_m) = \left(\boldsymbol{a}_m^T(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{a}_m\right)^{-1}, \tag{14}$$

and the effective rate that can be achieved at the m'th subchannel as

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) &= \frac{1}{2}\log\left(\text{SNR}_{\text{eff}}(\mathbf{a}_m)\right) \\ &= -\frac{1}{2}\log\left(\boldsymbol{a}_m^T(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{a}_m\right). \end{aligned} \tag{15}$$

Note that the rate expression (15) is negative when $\text{SNR}_{\text{eff}}(\mathbf{a}_m) < 1$. Hence, the achievable rate should be understood as the maximum between (15) and zero.

By Theorem 3 in [8], transmission with IF equalization can achieve any rate satisfying $R < R_{\text{IF}}(\mathbf{H})$ where

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}) &= \max_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \min_{m=1,\ldots,2N_t} 2N_t \cdot R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) \\ &= \max_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \min_{m=1,\ldots,2N_t} 2N_t \frac{1}{2}\log\left(\text{SNR}_{\text{eff}}(\mathbf{a}_m)\right) \\ &= \max_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \min_{m=1,\ldots,2N_t} N_t \log\left(\frac{1}{\boldsymbol{a}_m^T(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{a}_m}\right) \\ &= N_t \log\left(\min_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \max_{m=1,\ldots,2N_t}\left(\boldsymbol{a}_m^T(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{a}_m\right)\right)^{-1} \\ &= -N_t \log\left(\min_{\substack{\mathbf{A} \in \mathbb{Z}^{2N_t \times 2N_t} \\ \det \mathbf{A} \neq 0}} \max_{m=1,\ldots,2N_t}\left(\boldsymbol{a}_m^T(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{a}_m\right)\right). \end{aligned} \tag{16}$$

The achievable rate of IF may also be described via the successive minima of a lattice associated with the channel matrix as we now recall. Any channel can be described via its SVD

$$\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T. \tag{17}$$

Using (17), the following decomposition is readily obtained

$$(\mathbf{I} + \mathbf{H}^T\mathbf{H})^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{V}^T, \tag{18}$$

where $\mathbf{D} = \mathbf{I} + \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$. It follows that (15) may be rewritten as (Theorem 4 in [8])

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}; \mathbf{a}_m) &= -\frac{1}{2}\log\left(\|\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}_m\|^2\right) \\ &\triangleq R_{\text{IF}}^{\mathbf{a}_m}(\mathbf{D}, \mathbf{V}). \end{aligned} \tag{19}$$

Let Λ be the lattice spanned by $\mathbf{G} = \mathbf{D}^{-1/2}\mathbf{V}^T$. Recall the definition of successive minima.

Definition 1. *Let $\Lambda(\mathbf{G})$ be a lattice spanned by the full-rank matrix $\mathbf{G} \in \mathbb{R}^{K \times K}$. For $k = 1, \ldots, K$, we define the k'th successive minimum as*

$$\lambda_k(\mathbf{G}) \triangleq \inf\{r : \dim(\text{span}(\Lambda(\mathbf{G}) \cap \mathcal{B}_K(r))) \geq k\} \tag{20}$$

where $\mathcal{B}_K(r) = \{\mathbf{x} \in \mathbb{R}^K : \|\mathbf{x}\| \leq r\}$ is the closed ball of radius r around $\mathbf{0}$. In words, the k'th successive minimum of a lattice is the minimal radius of a ball centered around $\mathbf{0}$ that contains k linearly independent lattice points.

Thus, the maximal rate achievable with integer-forcing equalization (16) may be written as

$$\begin{aligned} R_{\text{IF}}(\mathbf{H}) &= R_{\text{IF}}(\mathbf{D}, \mathbf{V}) \\ &= -N_t \log\left(\lambda_{2N_t}^2(\Lambda)\right) \\ &= N_t \log\left(\frac{1}{\lambda_{2N_t}^2(\Lambda)}\right). \end{aligned} \tag{21}$$

B. Integer-Forcing Equalization With Successive Interference Cancellation

We also consider a version of the IF incorporating successive interference cancellation. We will refer to it as IF-SIC.[7] We state only the achievable rates of IF-SIC and an operational description of its elements. The reader is referred to [12] for the derivation, details and proofs.

For a given choice of integer matrix $\mathbf{A}$, let $\mathbf{L}$ be defined by the following Cholesky decomposition

$$\begin{aligned} \mathbf{A}\left(\mathbf{I} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T &= \mathbf{A}\mathbf{V}\mathbf{D}^{-1}\mathbf{V}^T\mathbf{A}^T \\ &= \mathbf{L}\mathbf{L}^T. \end{aligned} \tag{22}$$

Denoting by $\ell_{m,m}$ the diagonal entries of $\mathbf{L}$, IF-SIC can achieve (see [12]) any rate satisfying $R < R_{\text{IF-SIC}}(\mathbf{H})$ where

$$R_{\text{IF-SIC}}(\mathbf{H}) = 2N_t \frac{1}{2} \max_{\mathbf{A}} \min_{m=1,\ldots,2N_t} \log\left(\frac{1}{\ell_{m,m}^2}\right) \tag{23}$$

$$= R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}), \tag{24}$$

and the maximization is over all full-rank $2N_t \times 2N_t$ integer matrices.[8]

---

[7]We note that IF-SIC may in general allow using different rates per stream as stated in Theorem 5 in [10]. We nevertheless assume throughout that all streams are encoded via an identical linear code and hence have the same rate.

[8]We note that since we choose to work with equal-rate streams, the constraints on the achievable rate tuples of IF with SIC, as stated in Theorem 2 of [12], play no role in the present work.

4

We describe the operation of the IF-SIC receiver, adopting the nomenclature of [12]. We note that we describe the MMSE-GDFE version of IF-SIC, as given in Appendix A of [12], rather than its noise-prediction variant. First, calculate:

1) The optimal integer matrix $\mathbf{A}$, i.e., the matrix maximizing (23).
2) The covariance matrix (22) of the effective noise (see (14)) that arises when using the equalization matrix $\mathbf{B}$ as given in (11).
3) The optimal SIC matrix $\mathbf{S}$ as:

$$\mathbf{S} = \mathrm{diag}(\ell_{11}, ..., \ell_{MM}) \cdot \mathbf{L}^{-1}. \quad (25)$$

4) The optimal combined linear front-end processing matrix:

$$\begin{aligned} \widetilde{\mathbf{B}} &= \mathbf{SB} \\ &= \mathbf{SAH}^T \left(\mathbf{I} + \mathbf{HH}^T \right)^{-1}. \quad (26) \end{aligned}$$

The operation of the receiver is depicted in Figure 1 where the feedback depicted in the receiver is now active, and where now $\mathbf{B}$ is to be understood as $\widetilde{\mathbf{B}}$. Note that this change of linear post-processing is essential to guarantee that the resulting noise variance is minimized. The outputs of decoders $1, \ldots, m-1$ are multiplied by $S_{m,1}, \ldots, S_{m,m-1}$, respectively, and are then subtracted from the input to decoder m, thereby performing SIC.

C. Linear Pre-Processed Integer Forcing

1) Motivating example: Performance comparison of linear and IF receivers: As a motivating example, following Remark 1, we compare the performance of linear MMSE and IF equalizers over an ensemble of channels defined over $\mathbb{H}(C = 8)$. Specifically, we consider a "normalized" 2×2 Rayleigh fading ensemble, where the capacity is fixed to $C = 8$ bits. The ensemble is generated by drawing a channel matrix with i.i.d. circularly symmetric complex Gaussian entries and then scaling the matrix (multiplying it by a value that we find by numerical search) such that the mutual information equals 8 bits.

Figure 2 depicts the probability density function of the rate achieved for this ensemble when using linear MMSE and IF receivers. Since linear MMSE equalization is a special case of IF (setting $\mathbf{A} = \mathbf{I}$), as expected IF displays improved performance.

The real strength of IF lies however in the behavior of the "tail". For conventional linear equalizers, bad channels correspond to ill-conditioned matrices. An extreme case is the following channel

$$\mathbf{H}_{c,\mathrm{WORST}} = \sqrt{2^8 - 1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \quad (27)$$

In this case, the data stream sent from the second antenna is completely lost when transmitted over the channel. Clearly, in this example, no receiver (including maximum likelihood) will be able to recover the lost data stream and thus the achievable rate of both linear and IF equalization is also zero.
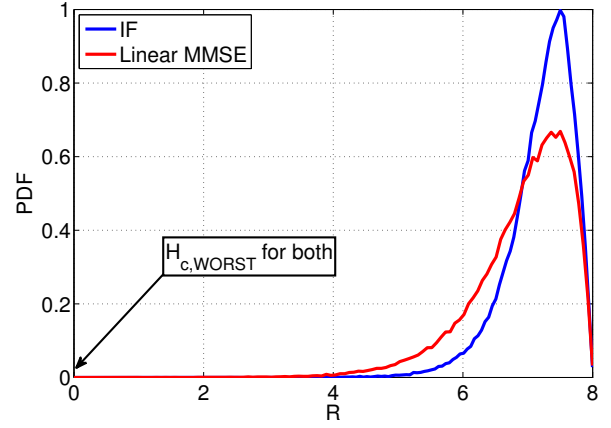

Fig. 2. Approximate probability density functions (based on Monte Carlo simulation) of the rates achievable with the linear MMSE receivers over a Rayleigh 2×2 MIMO channel normalized to $C = 8$ bits.

Consider now the channel $\mathbf{H}_{c,\mathrm{WORST}} \cdot \mathbf{P}_c$ where $\mathbf{P}_c$ is a unitary matrix. As the singular values remain unchanged, it is clear that the channel remains ill-conditioned and hence a linear receiver will still suffer from poor performance. On the other hand, the IF receiver performs well even over ill-conditioned MIMO channels, and in fact, the performance of the IF receiver for the channel (27) is good for "most" pre-processing matrices as illustrated next.

2) Linear pre-processing ensemble and resulting performance: The transmission scheme we analyze consists of applying a unitary pre-processing matrix at the transmitter and IF equalization (either with or without SIC) at the receiver, as depicted in Figure 1. Applying linear pre-processing may be viewed as generating a "virtual" channel $\widetilde{\mathbf{H}}_c = \mathbf{H}_c\mathbf{P}_c$ over which transmission takes place. We restrict ourselves to unitary linear pre-processing matrices in order to keep the transmission power unchanged.

Throughout this paper, we assume that the linear pre-processing matrix $\mathbf{P}_c$ is drawn from what is referred to as the "circular unitary ensemble" (CUE). The ensemble is defined by the unique distribution on unitary matrices that is invariant under left and right unitary transformations (Theorem 8.3 in [13]).[9]

Remark 3. *While in general the transformation $\mathbf{P}_c$ implies joint processing at the encoders, we note that in some natural statistical scenarios, including that of an i.i.d. Rayleigh fading environment, the random transformation is actually performed by nature. In such settings, our analysis holds even in a setup where the transmitters are distributed.*[10]

[9]An explanation on how to generate matrices belonging to the CUE can be found in [14].

[10]Specifically, note that the second order statistics of a random matrix with uncorrelated entries of equal variance remains unchanged by a unitary transformation applied at either side. As the distribution of a Gaussian matrix with i.i.d. entries is unaffected by such a transformation, it is readily seen that the (right and left) singular vectors of such a matrix follow the circular unitary distribution.

5

Figure 3 compares the achievable rates of the linear MMSE and IF receivers over the singular channel (27), when applying random CUE pre-processing. As can be seen, the achievable rate of IF is high for most pre-processing matrices, achieving a large fraction of C with high probability.
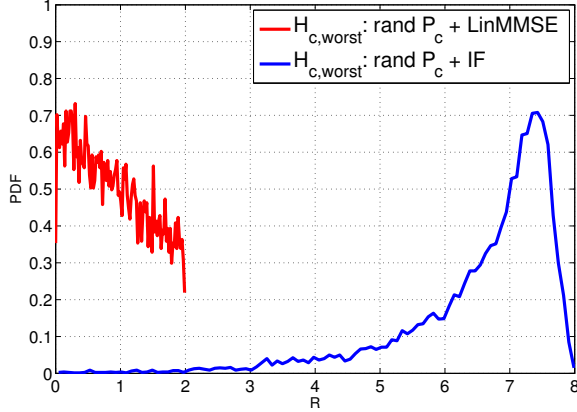

Fig. 3. Approximate probability density functions (based on Monte Carlo simulation) of the rates achievable with the linear MMSE and IF receivers over the channel (27), when applying a random linear pre-processing matrix drawn from the CUE.

3) Properties of CUE pre-processing: The SVD of the effective channel resulting from pre-processing is given by

$$\mathbf{H}_c \mathbf{P}_c = \mathbf{U}_c \mathbf{\Sigma}_c \mathbf{V}_c^H \mathbf{P}_c. \tag{28}$$

Since $\mathbf{V}_c^H \mathbf{P}_c$ is equal in distribution to $\mathbf{P}_c$, for the sake of computing outage probabilities, we may simply assume that $\mathbf{V}_c^H$ (and also $\mathbf{V}_c$) is drawn from the CUE.

We note that the eigenvalue decomposition of the equivalent real channel can be written as

$$(\mathbf{I} + \mathbf{H}^T \mathbf{H})^{-1} = \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T, \tag{29}$$

where

$$\mathbf{V} = \begin{bmatrix} \mathrm{Re}(\mathbf{V}_c) & -\mathrm{Im}(\mathbf{V}_c) \\ \mathrm{Im}(\mathbf{V}_c) & \mathrm{Re}(\mathbf{V}_c) \end{bmatrix}. \tag{30}$$

and

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_c & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_c \end{bmatrix}. \tag{31}$$

Further, the rates of IF, with or without SIC, for such a channel come in pairs.

Denoting the gap-to-capacity by ΔC, we may therefore rewrite the worst-case IF outage probability as defined in (6) as

$$P_{\mathrm{out,IF}}^{\mathrm{WC}}(C, C - \Delta C) = \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr\left(R_{\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C\right) \tag{32}$$

where $\mathbb{D}(C)$ is the set of all diagonal matrices $\mathbf{D}$, with diagonal elements appearing in pairs, such that $\det(\mathbf{D}) = 2^C$.

Another property we use in the sequel is the following.

Denote by $d_{c,i}$ the diagonal entries of $\mathbf{D}_c$. Then

$$
\begin{aligned}
2^C &= 2^{\log \det\left(\mathbf{I} + \mathbf{H}_c^H \mathbf{H}_c\right)} \\
&= \det\left(\mathbf{V}_c \mathbf{D}_c \mathbf{V}_c^H\right) \\
&= \prod_{i=1}^{N_t} d_{c,i}.
\end{aligned} \tag{33}
$$

Denoting by d_i the diagonal entries of $\mathbf{D}$, we similarly have

$$
\begin{aligned}
2^C &= 2^{\frac{1}{2} \log \det\left(\mathbf{I} + \mathbf{H}^T \mathbf{H}\right)} \\
&= \sqrt{\det\left(\mathbf{V} \cdot \mathbf{D} \cdot \mathbf{V}^T\right)} \\
&= \prod_{i=1}^{2N_t} \sqrt{d_i}.
\end{aligned} \tag{34}
$$

From (31) we observe that since the singular values of the real channel come in pairs, we have

$$\prod_{i=1}^{2N_t} \sqrt{d_i} = \prod_{i=1}^{N_t} d_{c,i} = 2^C. \tag{35}$$

We denote $d_{\min} = \min_i d_i$ and $d_{\max} = \max_i d_i$.

The following lemma will prove useful in characterizing the performance of CUE pre-processed IF. It relates the outage probability of CUE pre-processed IF to that arising when the pre-processing is performed using the circular real ensemble (CRE).[11]

Lemma 1. *Let $\mathbf{O}$ be a real $2N_t \times 2N_t$ matrix drawn from the CRE. Further, let $\mathbf{a}$ be a $2N_t \times 1$ vector of integers. When applying a random complex linear pre-processing matrix $\mathbf{V}_c$ which is drawn from the CUE (inducing a real-valued orthogonal pre-processing matrix $\mathbf{V}$), we have that $\left\| \mathbf{D}^{1/2} \mathbf{V} \mathbf{a} \right\|$ and $\left\| \mathbf{D}^{1/2} \mathbf{O} \mathbf{a} \right\|$ are equal in distribution.*

Proof. See Appendix A. □

IV. BOUND ON THE OUTAGE PROBABILITY OF CUE PRE-PROCESSED INTEGER-FORCING

A. Derivation of Upper Bounds

Define the dual lattice Λ^* which is spanned by the matrix

$$(\mathbf{G}^T)^{-1} = \mathbf{D}^{1/2} \mathbf{V}^T. \tag{36}$$

Recall that the rate of IF is given by (21). Now, the successive minima of Λ and Λ^* are related by (Theorem 2.4 in [15])

$$\lambda_1(\Lambda^*)^2 \lambda_{2N_t}(\Lambda)^2 \leq \frac{2N_t + 3}{4} \gamma_{2N_t}^{*2}, \tag{37}$$

[11]The CRE is defined analogously to the CUE for the case of real orthonormal matrices. That is, the ensemble is defined by the unique distribution on orthonormal matrices that is invariant under left and right orthonormal transformations

6

where γ_{2N_t} is Hermite's constant.[12] Therefore, we may express the achievable rates of IF via the dual lattice as follows

$$R_{\text{IF}}(\mathbf{D}, \mathbf{V}) > N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\frac{2N_t+3}{4}\gamma_{2N_t}^{*2}} \right). \tag{38}$$

Hermite's constant is known only for dimensions $1 - 8$ and 24. Since it has been never proved that γ_{2N_t} is monotonically increasing, we define

$$\gamma_{2N_t}^* = \max \left\{ \gamma_i : 1 \leq i \leq 2N_t \right\}. \tag{39}$$

The tightest known bound for Hermite's constant, as derived in [17], is

$$\gamma_{2N_t} \leq \left(\frac{2}{\pi} \right) \Gamma \left(2 + N_t \right)^{1/N_t}. \tag{40}$$

Since this is an increasing function of N_t, it follows that $\gamma_{2N_t}^*$ is smaller than the r.h.s. of (40).[13] Combining the latter with the exact values of the Hermite constant for dimensions for which it is known, we obtain

$$R_{\text{IF}}(\mathbf{D}, \mathbf{V}) \geq N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\alpha(N_t)} \right), \tag{41}$$

where

$$\alpha(N_t) = \begin{cases} \frac{2N_t+3}{4}\gamma_{2N_t}^2, & N_t = 2,3,4,12 \\ \frac{2N_t+3}{4}\left(\frac{2}{\pi}\Gamma\left(2+N_t\right)^{1/N_t} \right)^2, & \text{otherwise} \end{cases}. \tag{42}$$

Hence,

$$\begin{aligned} &\Pr\left(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right) \\ &\leq \Pr\left(N_t \log \left(\frac{\lambda_1^2(\Lambda^*)}{\alpha(N_t)} \right) < C - \Delta C \right) \\ &= \Pr\left(\lambda_1^2(\Lambda^*) < 2^{\frac{C-\Delta C}{N_t}}\alpha(N_t) \right). \end{aligned} \tag{43}$$

The next lemma provides an upper bound on the outage probability as a function of the gap-to-capacity ΔC, the capacity C, and $d_{\min}$ (as well as the number of transmit antennas). Denote

$$\mathbb{A}(\beta, d) = \left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\frac{\beta}{d}} \right\}. \tag{44}$$

Lemma 2. *For any $N_r \times N_t$ complex channel with white-input mutual information C, i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued orthogonal pre-processing matrix $\mathbf{V}$), we have*

$$\Pr\left(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right)$$

$$\leq \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}}\alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1}2^C\frac{2}{\sqrt{d_{\min}}}}, \tag{45}$$

where

$$\beta = 2^{\frac{C-\Delta C}{N_t}}\alpha(N_t). \tag{46}$$

Remark 4. *The summation in (45) may be greater than 1 for certain values of C and ΔC. Obviously, one may take the minimum between this lemma and 1 when bounding the outage probability.*

Proof. For some $\beta > 0$, let us upper bound the probability $\Pr\left(\lambda_1^2(\Lambda^*) < \beta\right)$ or equivalently $\Pr\left(\lambda_1(\Lambda^*) < \sqrt{\beta}\right)$. Noting that the event $\left\{\lambda_1(\Lambda^*) < \sqrt{\beta}\right\}$ is equivalent to the event

$$\bigcup_{\mathbf{a} \in \mathbb{Z}^{2N_t}\backslash\{\mathbf{0}\}} \left\{ \|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\| < \sqrt{\beta} \right\} \tag{47}$$

and applying the union bound gives

$$\begin{aligned} \Pr\left(\lambda_1(\Lambda^*) < \sqrt{\beta} \right) &< \sum_{\mathbf{a} \in \mathbb{Z}^{2N_t}} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\| < \sqrt{\beta} \right) \\ &= \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{V}^T\mathbf{a}\| < \sqrt{\beta} \right), \end{aligned} \tag{48}$$

where the equality in (48) follows since whenever $\|\mathbf{a}\| \cdot \sqrt{d_{\min}} \geq \sqrt{\beta}$, we have that $\Pr\left(\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\| < \sqrt{\beta} \right) = 0$.

Let $\mathcal{S}$ denote the unit sphere of dimension $2N_t$, i.e.,

$$\mathcal{S} = \left\{ (x_1, x_2, \cdots, x_{2N_t}) : x_1^2 + x_2^2 + \cdots + x_{2N_t}^2 = 1 \right\}. \tag{49}$$

By Lemma 1

$$\Pr\left(\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\| < \sqrt{\beta} \right) = \Pr\left(\|\mathbf{D}^{1/2}\mathbf{O}\mathbf{a}\| < \sqrt{\beta} \right). \tag{50}$$

Let $\mathbf{o}_{\|\mathbf{a}\|} \sim \text{Unif}(\mathcal{S} \cdot \|\mathbf{a}\|)$, and note that $\mathbf{O}\mathbf{a}$ is equal in distribution to $\mathbf{o}_{\|\mathbf{a}\|}$. It follows that

$$\Pr\left(\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\| < \sqrt{\beta} \right) = \Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right). \tag{51}$$

Now the probability appearing on the r.h.s. of (51) has a simple geometric interpretation. Define an ellipsoid with axes $x_i = \sqrt{d_i} \cdot \|\mathbf{a}\|$ and denote its surface area by $L(x_1, x_2, ..., x_{2N_t})$. Then, the r.h.s. of (51) is the ratio of the part of the surface area of an ellipsoid that lies inside a sphere of radius $\sqrt{\beta}$ (denoted by $\text{CAP}_{\text{ell}}(x_1, x_2, ..., x_{2N_t})$) and the total surface area of the ellipsoid. This is illustrated in Figure 4 for the case of two real dimensions. We may rewrite (51) as

$$\begin{aligned} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) &= \frac{|\mathbf{D}^{1/2}\mathcal{S} \cdot \|\mathbf{a}\| \cap \sqrt{\beta}\mathcal{S}|}{|\mathbf{D}^{1/2}\|\mathbf{a}\|\mathcal{S}|} \\ &= \frac{\text{CAP}_{\text{ell}}(x_1, x_2, ..., x_{2N_t})}{L(x_1, x_2, ..., x_{2N_t})}. \end{aligned} \tag{52}$$

Neither the numerator nor the denominator of (52) has a closed form expression. In order to upper bound this ratio, we upper bound the numerator $\text{CAP}_{\text{ell}}(x_1, x_2, ..., x_{2N_t})$ and lower
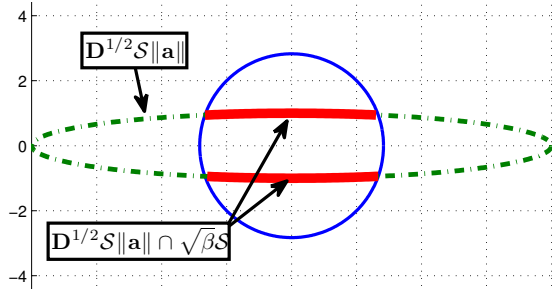
Fig. 4. Illustration of the geometric objects appearing in (52).

Fig. 5. Comparison of empirical results, Lemma 2 and Theorem 1 for $N_r \times 2$ channels, for various values of mutual information.

bound the denominator (the surface area of the ellipsoid). Using inequality (4.3) in [18] (see also inequality (57) and historical account in [19]), we have

$$
\begin{aligned}
L(x_1, ..., x_{2N_t}) &> \text{Vol}(\mathcal{B}_{2N_t}(1)) \|\mathbf{a}\|^{2N_t} \prod_{i=1}^{2N_t} \sqrt{d_i} \sum_{i=1}^{2N_t} \frac{1}{\|\mathbf{a}\|\sqrt{d_i}} \\
&\geq \text{Vol}(\mathcal{B}_{2N_t}(1)) \|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}} \\
&\triangleq \underline{L}(x_1, x_2, ..., x_{2N_t}),
\end{aligned} \tag{53}
$$

where $\mathcal{B}_{2N_t}(1)$ is a unit ball of dimension $2N_t$, and

$$
\text{Vol}(\mathcal{B}_{2N_t}(1)) = \frac{\pi^{N_t}}{\Gamma(1 + N_t)} \tag{54}
$$

is its volume.

As an upper bound for the numerator, we take the entire surface area of a sphere of radius $\sqrt{\beta}$, which is given by

$$
A_{2N_t}(\sqrt{\beta}) = 2N_t \frac{\pi^{N_t}}{\Gamma(1 + N_t)} \sqrt{\beta}^{2N_t-1}. \tag{55}
$$

We thus have

$$
\begin{aligned}
\text{CAP}_{\text{ell}}(x_1, x_2, ..., x_{2N_t}) &\leq A_{2N_t}(\sqrt{\beta}) \\
&\triangleq \overline{\text{CAP}_{\text{ell}}}(\sqrt{\beta}).
\end{aligned} \tag{56}
$$

We may therefore bound (52) by

$$
\Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) \leq \frac{\overline{\text{CAP}_{\text{ell}}}(\sqrt{\beta})}{\underline{L}(x_1, x_2, ..., x_{2N_t})}. \tag{57}
$$

Substituting (53), (56) into (57) yields

$$
\begin{aligned}
&\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right) \\
&< \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \frac{\pi^{N_t}}{\Gamma(1+N_t)} \sqrt{\beta}^{2N_t-1}}{\frac{\pi^{N_t}}{\Gamma(1+N_t)} \|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}} \\
&= \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t-1}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}.
\end{aligned} \tag{58}
$$

Substituting $\beta = 2^{\frac{C-\Delta C}{N_t}} \alpha(N_t)$, we finally arrive at

$$
\begin{aligned}
&\Pr\left(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right) \\
&\leq \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \Pr\left(\|\mathbf{D}^{1/2}\mathbf{o}_{\|\mathbf{a}\|}\| < \sqrt{\beta} \right)
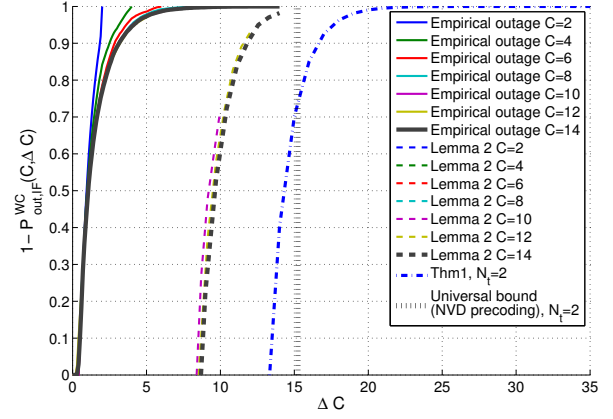\end{aligned}
$$

$$
< \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t) \right)^{N_t-1/2}}{\|\mathbf{a}\|^{2N_t-1} 2^C \frac{2}{\sqrt{d_{\min}}}}. \tag{59}
$$

$\square$

The bound of Lemma 2 is depicted in Figure 5. Rather than plotting the outage probability, its complement is depicted, i.e., we depict the cumulative distribution function of the event that the rate is achieved by IF. For given C and ΔC, Lemma 2 was numerically calculated over a grid of singular values. For each such vector of singular values, summation was performed over all $\mathbf{a} \in \mathbb{A}(\beta, d_{\min})$. The worst-case outage probability over all vectors of singular values from the grid is presented.

In addition, empirical (Monte Carlo) results are also plotted. For each vector of singular values, a large number of random unitary matrices were drawn and the outage probability was calculated. The integer matrix was derived using the LLL algorithm.[14] The worst case outage probability over all tested (i.e., those belonging to the grid) singular values is presented.

As a further reference, the figure also depicts the universal guaranteed gap-to-capacity derived in [4], which for the case of $N_t = 2$, amounts to $\Delta C = 15.24$ bits [4].[15]

While Lemma 2 provides an explicit bound on the outage probability, in order to calculate it, one needs to go over all diagonal matrices in $\mathbb{D}(C)$ and for each diagonal matrix, sum over all the relevant integer vectors in $\mathbb{A}(\beta, d_{\min})$. Hence, the bound can be evaluated only for moderate capacities and for a small number of transmit antennas. The following theorem provides (a looser) simple closed-form bound. Furthermore, this bound does not depend on capacity but rather only on the number of transmit antennas and the gap-to-capacity.

---

[14] Advanced techniques are known (see, e.g., [20] and [21]) that can be used to further improve the empirical results.

[15] This upper bound on the gap-to-capacity is guaranteed for a different coding scheme than that considered in this paper, where space-time pre-processing is employed. Nevertheless, it serves as a useful benchmark.
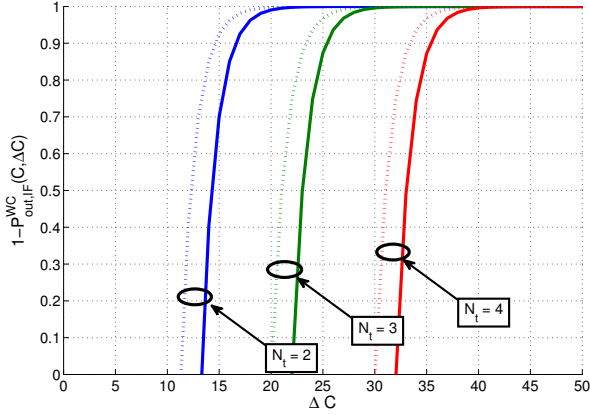
8

Fig. 6. The performance guaranteed by Theorem 1 for $N_t = 2, 3, 4$ transmit antennas. The dashed lines correspond to the improvement provided by Corollary 2.

Theorem 1. *For any $N_r \times N_t$ complex channel with WI mutual information C, and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued linear pre-processing matrix $\mathbf{V}$), we have*

$$P_{\text{out,IF}}^{\text{WC}}(C, \Delta C) \leq c(N_t) 2^{-\Delta C}, \tag{60}$$

where

$$c(N_t) = \left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)} \tag{61}$$

and

$$\alpha(N_t) = \frac{2N_t + 3}{4} \left(\frac{2}{\pi} \Gamma(2 + N_t)^{1/N_t}\right)^2. \tag{62}$$

Thus, $c(N_t)$ is a constant that depends only on N_t.

Proof. See Appendix B. ☐

This bound is also depicted in Figure 5 for the case of two transmit antennas. For other values of N_t, the bound is depicted in Figure 6 (solid lines). Recall again that, for $N_t = 2, 3, 4$, we use the actual values of $\gamma_4 = \sqrt{2}$, $\gamma_6 = \left(\frac{64}{3}\right)^{1/6}$ and $\gamma_8 = 2$, rather than the bound of [17]. We note that when the number of transmit antennas increases, the gap between the theorem and the empirical results grows mainly due to the penalty incurred in (37) from using the dual lattice.

B. Improved Upper Bounds

A close inspection of Theorem 1 reveals that there are two main sources for looseness in the bound that may be further tightened:

- *Union bound* - While there is an inherent loss in the union bound, some terms in the summation (48) may be completely dropped. See Corollary 1 below.
- *Dual lattice* - Bounding via the dual lattice induces a loss reflected in (37). This may be circumvented for the case of two transmit antennas, as accomplished (along with other improvements) in Theorem 2 below.

We first tighten the union bound. As expressed in (47), the event where the first of the successive minima is smaller than $\sqrt{\beta}$ is equivalent to going over all integer vectors and checking whether any of them meet the norm condition. However, going over *all* integer vectors is superfluous. In case that an integer vector $\mathbf{b} \in \mathbb{A}(\beta, \mathbf{d}_{\min})$ is an integer multiple of another integer vector $\mathbf{a} \in \mathbb{A}(\beta, \mathbf{d}_{\min})$, there is no need to count both of them. Rather, it suffices to include in the union bound only the event corresponding to $\mathbf{a}$.

It follows that one may replace the set $\mathbb{A}(\beta, d_{\min})$ appearing in the summation in (2) by a smaller set $\mathbb{B}(\beta, d_{\min})$ where

$$\mathbb{B}(\beta, d) = \left\{ \mathbf{a} : 0 < \|\mathbf{a}\| < \sqrt{\frac{\beta}{d}} \text{ and } \nexists \, 0 < c < 1 \text{ s.t. } c\mathbf{a} \in \mathbb{Z}^n \right\} \tag{63}$$

as described by the next corollary.

Corollary 1. *For any $N_r \times N_t$ complex channel and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued linear pre-processing matrix $\mathbf{V}$), we have*

$$\Pr\left(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C\right)$$
$$\leq \sum_{\mathbf{a} \in \mathbb{B}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C - \Delta C}{N_t}} \alpha(N_t)\right)^{N_t - 1/2}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}. \tag{64}$$

where $\beta = 2^{\frac{C - \Delta C}{N_t}} \alpha(N_t)$.

Another improvement may be obtained by noting that $\mathbf{D}$ and $\mathbf{V}$ are the real representations of complex matrices. Using the notations of (10), the integer vector $\mathbf{a}$ may be viewed as the real representation of the complex vector $\mathbf{a}_c$. Thus,

$$\|\mathbf{D}_c^{1/2} \mathbf{V}_c \mathbf{a}_c\| = \|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\|. \tag{65}$$

As multiplication of $\mathbf{a}_c$ by $\{-1, j, -j\}$ does not change the value of $\|\mathbf{D}_c^{1/2} \mathbf{V}_c \mathbf{a}_c\|$ (and equivalently, it does not change the value of $\|\mathbf{D}^{1/2} \mathbf{V} \mathbf{a}\|$), it suffices to include only one of these members of $\mathbb{A}(\beta, d_{\min})$ in the summation. Hence, a simple multiplicative improvement may be obtained.

Corollary 2. *For any $N_r \times N_t$ complex channel and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued linear pre-processing matrix $\mathbf{V}$), we have*

$$\Pr\left(R_{\text{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C\right)$$
$$\leq \frac{1}{4} \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \beta^{N_t - 1/2}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}, \tag{66}$$

where $\beta = 2^{\frac{C - \Delta C}{N_t}} \alpha(N_t)$.

While the improvement of Corollary 1 depends on β (and hence also on C), we may tighten Theorem 1 by invoking Corollary 2 as shown by the dashed lines in Figure 6. For a given value of C, we may combine the two corollaries. Figure 7 shows the different bounds on the outage probability for the case of an $N_r \times 2$ MIMO channel with $C = 14$, where both corollaries are utilized for tightening Lemma 2.

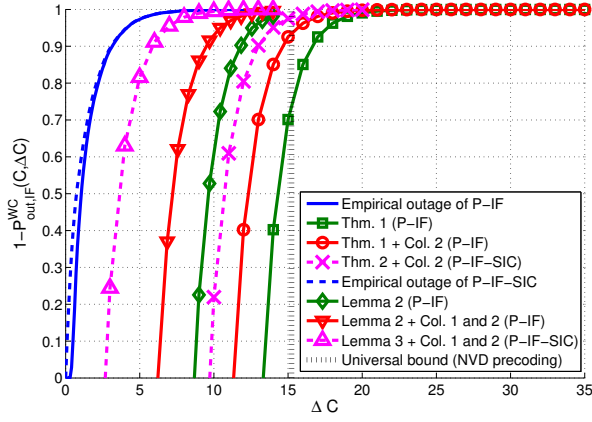As mentioned above, there is an additional significant loss

Fig. 7. Outage bounds for a $N_r \times 2$ channel with mutual information $C = 14$ bits.

Fig. 8. Worst-case performance comparison between IF-SIC and joint ML decoding, for the case of $N_t = 2$ transmit antennas.

due to using the dual lattice for deriving both Lemma 2 and Theorem 1. For the case of $N_t = 2$, this loss may be circumvented by analyzing the performance of IF-SIC. When using IF-SIC, (6) can be rewritten as

$$P_{\text{out,IF-SIC}}^{\text{WC}} (C, C - \Delta C)$$
$$= \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr \left(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right). \quad (67)$$

The next lemma provides a bound on the outage probability of IF-SIC.

Lemma 3. *For any $N_r \times 2$ complex channel with white-input mutual information $C > 1$, i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued linear pre-processing matrix $\mathbf{V}$), we have*

$$\Pr \left(R_{\text{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right)$$
$$< \sum_{\mathbf{a} \in \mathbb{A}(\beta, 1/d_{\max})} \frac{2\pi^2 2^{-3/4(C+\Delta C)}}{\pi^2 \frac{\|\mathbf{a}\|^3}{2^C} \sqrt{d_{\max}}}, \quad (68)$$

where $\mathbb{A}(\beta, 1/d_{\max})$ is defined in (44), $\beta = 2^{-1/2(C+\Delta C)}$ and for all $\Delta C > 1$.

Proof. See Appendix C. $\qquad\square$

Remark 5. *Lemma 3 can be further tightened using Corollary 1, i.e., by replacing $\mathbb{A}(\beta, 1/d_{\max})$ with $\mathbb{B}(\beta, 1/d_{\max})$.*

In a similar manner to the derivation of Theorem 1 using Lemma 2, for IF-SIC, Lemma 3 leads to the following theorem.

Theorem 2. *For any $N_r \times 2$ complex channel $\mathbf{H}_c$ with white-input mutual information $C > 1$, i.e., $\mathbf{D} \in \mathbb{D}(C)$, and for $\mathbf{V}_c$ drawn from the CUE (inducing a real-valued linear pre-processing matrix $\mathbf{V}$), we have*

$$P_{\text{out,IF-SIC}}^{\text{WC}} (C, \Delta C) \leq 81\pi^2 2^{-\Delta C}, \quad (69)$$

for all $\Delta C > 1$.
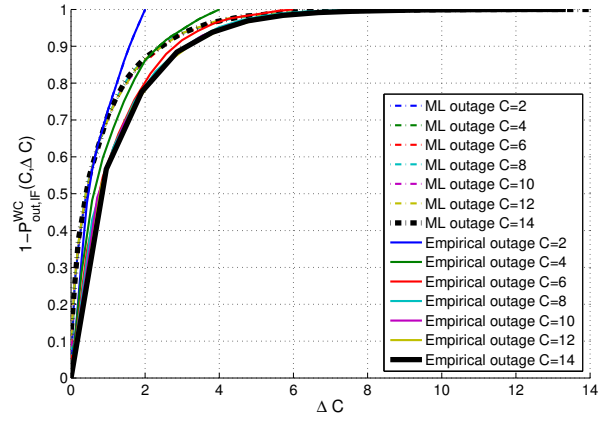
Proof. See Appendix C. $\qquad\square$

Figure 7 depicts the improved bounds of Lemma 3 (incorporating the improvements provided by Corollary 1 and Corollary 2) and Theorem 2 for a system employing IF-SIC with $N_t = 2$.

C. Lower Bound via Maximum-Likelihood Decoding

Beyond the upper bounds on performance derived thus far, it is natural to compare the performance attained by an IF receiver with that of an optimal maximum likelihood (ML) decoder for the same randomly linear pre-processed scheme but where each stream is coded using an independent Gaussian codebooks. This provides a lower bound on the outage probability of IF.

let $\mathbf{H}_S$ denote the submatrix of $\mathbf{H}_c \mathbf{P}_c$ formed by taking the columns with indices in $S \subseteq \{1, 2, ..., N_t\}$. For a joint ML decoder, the following is the maximal rate achievable [8] over the considered MIMO multiple-access channel:

$$R_{\text{JOINT}} = \min_{S \subseteq \{1,2,...,N_t\}} \frac{N_t}{|S|} \log \det \left(\mathbf{I}_{N_r} + \mathbf{H}_S \mathbf{H}_S^H \right). \quad (70)$$

Note that since $\mathbf{H}_S$ depends on the random linear preprocessing matrix $\mathbf{P}_c$, R_{JOINT} is a random variable.

Figure 8 provides a comparison between the empirical performance of IF-SIC and the empirical performance of the corresponding scheme with ML decoding, for the case of $N_t = 2$. In both cases, CUE pre-processing is applied. Thus, performance depends only on the singular values of the channel and hence the outage probability curves are the supremum of the outage probability over a grid of singular values.

As can be seen, the gap between IF and ML is quite small. This suggests that most of the loss with respect to the WI mutual information is due to the separate encoding of the data streams rather than the suboptimailty of the IF receiver.

10

V. Application : Universal Gap-To-Capacity for Multi-User Closed-Loop Multicast using P-IF

Closed-loop MIMO multicast is a scenario where a transmitter equipped with N_t transmit antennas wishes to send the same message to K users, where user i is equipped with N_i antennas.

Even though channel state information is available at both transmission ends, designing practical capacity-approaching schemes for closed-loop MIMO multicast with $K \geq 3$ users is challenging as detailed in [22]. The outage bound derived above suggests that pre-processed IF may be an attractive practical closed-loop MIMO multicast scheme, allowing to obtain a small gap-to-capacity with space-only pre-processing. Namely, we use the probabilistic method to establish the existence of a pre-processing matrix guaranteeing a desired gap-to-capacity.

We denote by $\mathbf{H}_{c,i}$ the $N_i \times N_t$ channel matrix corresponding to the ith user and by $\mathcal{H} = \{\mathbf{H}_{c,i}\}_{i=1}^{K}$ the set of channels. The received signal at user i is

$$\boldsymbol{y}_{c,i} = \mathbf{H}_{c,i}\boldsymbol{x}_c + \boldsymbol{z}_{c,i}. \tag{71}$$

We assume that channel state information (CSI) is available at both transmission ends.

The multicast capacity is defined as the capacity of the compound channel (71). It is attained by a Gaussian input vector, where the mutual information is maximized over all covariance matrices $\mathbf{Q}_c$ satisfying $\mathrm{Tr}(\mathbf{Q}_c) \leq N_t$:

$$C(\mathcal{H}) = \max_{\mathbf{Q}_c:\mathrm{Tr}(\mathbf{Q}_c)\leq N_t} \min_{\mathbf{H}_c\in\mathcal{H}} \log\det(\mathbf{I} + \mathbf{H}_c\mathbf{Q}_c\mathbf{H}_c^H). \tag{72}$$

We assume without loss of generality that the input covariance matrix is the identity matrix. We may do so since the covariance shaping matrix $\mathbf{Q}_c^{1/2}$ may be absorbed into the channel by defining the effective channel $\hat{\mathbf{H}}_{c,i} = \mathbf{H}_{c,i}\mathbf{Q}_c^{1/2}$. Thus,

$$C(\mathcal{H}) = \min_i \log\det(\mathbf{I} + \hat{\mathbf{H}}_{c,i}\hat{\mathbf{H}}_{c,i}^H). \tag{73}$$

In other words, after finding the optimal covariance matrix $\mathbf{Q}_c$, when it comes to the transmission scheme, it suffices to consider WI transmission over the effective channels $\hat{\mathbf{H}}_{c,i}$. With a slight abuse of notation, we use $\mathbf{H}_{c,i}$ to denote the effective channel, i.e., we drop the hat. We note that for each user i, there exists an $\alpha_i \geq 1$ such that

$$\mathbf{H}_{c,i} = \alpha_i\breve{\mathbf{H}}_{c,i}. \tag{74}$$

where

$$\breve{\mathcal{H}} = \{\breve{\mathbf{H}}\}_{i=1}^{K} \in \mathbb{H}(C(\mathcal{H})), \tag{75}$$

i.e., $\{\breve{\mathbf{H}}\}_{i=1}^{K}$ is contained in the (continuum) set of channels, having the same capacity $C(\mathcal{H})$. Further, α_i can be interpreted as excess SNR that user i enjoys, beyond the minimum it needs in the multicast setting. Since the achievable rate of IF is monotonically increasing in SNR, it follows that the achievable rates over the set of channels $\mathcal{H}$ can only be higher than over $\breve{\mathcal{H}}$, which we next lower bound.

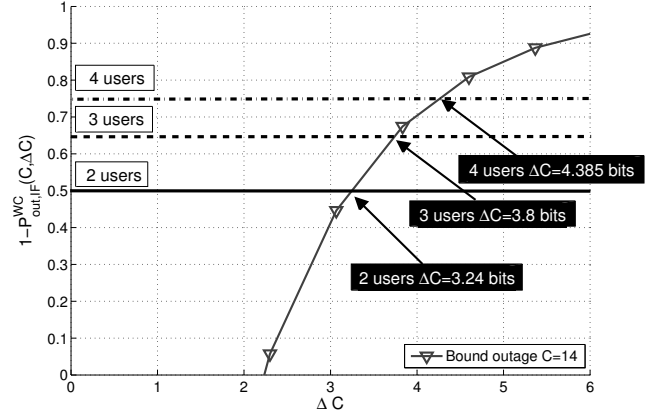Let us consider applying the random CUE pre-processing

Fig. 9. Guaranteed achievable rates with pre-processed IF for closed-loop MIMO multicast transmission with two transmit antennas and with 2,3,4 users.

scheme to the compound channel set $\breve{\mathcal{H}}$.[16] Define $A_i(R)$ as the event where the pre-processing matrix $\mathbf{P}_c$ is such that IF achieves a desired target R for user i

$$A_i(R) = \{\mathbf{P}_c : R_{\mathrm{IF}}(\mathbf{H}_{c,i} \cdot \mathbf{P}_c) \geq R\}. \tag{76}$$

We are interested in the probability of achieving the target rate for all users, i.e., $\mathrm{Pr}(\cap A_i(R))$. Note that

$$\mathrm{Pr}(\cap A_i(R)) = 1 - \mathrm{Pr}\left(\overline{\cap A_i(R)}\right) = 1 - \mathrm{Pr}\left(\cup \overline{A_i(R)}\right). \tag{77}$$

Applying the union bound, we get

$$\mathrm{Pr}\left(\cup \overline{A_i(R)}\right) \leq \sum \mathrm{Pr}\left(\overline{A_i(R)}\right) \tag{78}$$

and hence

$$\mathrm{Pr}(\cap A_i(R)) \geq 1 - \sum \mathrm{Pr}\left(\overline{A_i(R)}\right). \tag{79}$$

Define

$$\widebrece{A}_i(R) = \left\{\mathbf{P}_c : R_{\mathrm{IF-SIC}}(\breve{\mathbf{H}}_{c,i} \cdot \mathbf{P}_c) \geq R\right\}. \tag{80}$$

Since $\mathrm{Pr}\left(\widebrece{A}_i(R)\right)$ is the probability of achieving the target rate, whereas $P_{\mathrm{out,IF}}^{\mathrm{WC}}$ bounds the probability of the complement event, we have

$$\mathrm{Pr}(A_i(R)) \geq \mathrm{Pr}\left(\widebrece{A}_i(R)\right) \geq 1 - P_{\mathrm{out,IF}}^{\mathrm{WC}}(C(\mathcal{H}), C - R), \tag{81}$$

or equivalently,

$$\mathrm{Pr}\left(\overline{A_i(R)}\right) = 1 - \mathrm{Pr}(A_i(R)) \leq P_{\mathrm{out,IF}}^{\mathrm{WC}}(C(\mathcal{H}), C - R). \tag{82}$$

It follows that,

$$\mathrm{Pr}(\cap A_i(R)) \geq 1 - K P_{\mathrm{out,IF}}^{\mathrm{WC}}(C(\mathcal{H}), C - R). \tag{83}$$

---

[16]IF-SIC is used for $N_t = 2$ since it provides improved bounds.

11

This provides a means to obtain a guaranteed achievable transmission rate $R_{\mathrm{WC-CL}}(\mathcal{H})$ for closed-loop linear pre-processed IF. Namely, $R_{\mathrm{WC-CL}}(\mathcal{H})$ is the maximum rate for which

$$P_{\mathrm{out,IF}}^{\mathrm{WC}}\left(C(\mathcal{H}), C - R\right) \leq \frac{1}{K}. \tag{84}$$

Substituting (84) in (83) we get that for any $R < R_{\mathrm{WC-CL}}(\mathcal{H})$

$$\Pr\left(\cap A_i(R)\right) > 1 - K \cdot \frac{1}{K} = 0. \tag{85}$$

Thus, there must exist a linear pre-processing matrix $\mathbf{P}_c$ for which a target rate $R < R_{\mathrm{WC-CL}}(\mathcal{H})$ is achievable (via P-IF transmission) for the compound channel (71).

Figure 9 depicts the corresponding upper bounds on the gap-to-capacity for MIMO multicast with two transmit antennas, $K = 2, 3, 4$ users, and where $C(\mathcal{H}) = 14$ bits. For calculating the upper bound on the guaranteed gap-to-capacity, we use the tightest bound we have developed for $N_t = 2$ which is Corollary 1 of Lemma 3. We observe that

- For 2 users, a rate of 10.76 bits is guaranteed (gap of 3.24 bits to capacity).
- For 3 users, a rate of 10.2 bits is guaranteed (gap of 3.8 bits to capacity).
- For 4 users, a rate of 9.615 bits is guaranteed (gap of 4.385 to capacity).

VI. Conclusion

We obtained explicit universal bounds for the outage probability of a transmission scheme employing random unitary pre-processing at the transmitter side and integer-forcing equalization at the receiver side. These bounds provide meaningful performance guarantees for transmission over MIMO channels that depend only on the channel's mutual information and number of transmit antennas. Nonetheless, simulations suggest that there is still a considerable gap between the obtained bounds and the true (worst-case) outage probability of the examined scheme, calling for further work.

VII. Acknowledgement

Appendix A
Proof of Lemma 1

We start by expressing $\left\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\right\|$ equivalently in complex notation. We note that $\mathbf{a}$ (which is a vector of $2N_t$ real integers) can be viewed as the real representation of a complex vector $\mathbf{a}_c$ such that

$$\mathbf{a} = \begin{bmatrix} \mathrm{Re}(\mathbf{a}_c) \\ \mathrm{Im}(\mathbf{a}_c) \end{bmatrix}. \tag{86}$$

Obviously, $\|\mathbf{a}\| = \|\mathbf{a}_c\|$.

With this notation, and since $\mathbf{D}$ and $\mathbf{V}$ are the real representation of the complex matrices $\mathbf{D}_c$ and $\mathbf{V}_c$, it follows that

$$\left\|\mathbf{D}^{1/2}\mathbf{V}\mathbf{a}\right\| = \left\|\mathbf{D}_c^{1/2}\mathbf{V}_c\mathbf{a}_c\right\|. \tag{87}$$

Now since $\mathbf{V}_c$ is drawn from the CUE, the distribution of $\left\|\mathbf{D}_c^{1/2}\mathbf{V}_c\mathbf{a}_c\right\|$ is equal to that of $\left\|\mathbf{D}_c^{1/2}\mathbf{V}_c\begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T \|\mathbf{a}_c\|\right\|$. Note also that

$$\left\|\mathbf{D}_c^{1/2}\mathbf{V}_c\begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T \|\mathbf{a_c}\|\right\| =$$
$$\left\|\mathbf{D}_c^{1/2}\mathbf{V}_c\begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T\right\| \|\mathbf{a}\| =$$
$$\left\|\mathbf{D}_c^{1/2}\mathbf{v}_{c,1}\right\| \|\mathbf{a}\|. \tag{88}$$

where $\mathbf{v}_{c,1}$ is the first column of $\mathbf{V}_c$.

As described in [23], $\mathbf{v}_{c,1}$ is uniformly distributed over the surface of the complex unit sphere. Such a vector can be generated by taking a vector with zero-mean i.i.d. complex Gaussian components and scaling it by its norm. The components of such a vector can be expressed as

$$\mathbf{v}_{c,1} = \frac{G_{c,i}}{\sqrt{\sum_{i=1}^{N_t} |G_{c,i}|^2}}, \tag{89}$$

where $G_{c,i}$ are zero-mean i.i.d. complex circularly symmetric Gaussian random variables.

Similarly, a vector taken from a CRE matrix is uniformly distributed over the surface of the real unit sphere and it can be generated by taking a vector with zero-mean i.i.d. real Gaussian components and scaling it by its norm. The components of such a vector can be expressed as

$$\mathbf{o}_{r,1} = \frac{G_{r,i}}{\sqrt{\sum_{i=1}^{N_t} G_{r,i}^2}}, \tag{90}$$

where $G_{r,i}$ are zero-mean i.i.d. real Gaussian random variables.

We may rewrite (88) over the reals as

$$\left\|\mathbf{D}^{1/2}\begin{bmatrix} \mathrm{Re}(\mathbf{v}_{c,1}) \\ \mathrm{Im}(\mathbf{v}_{c,1}) \end{bmatrix}\right\| \|\mathbf{a}\|. \tag{91}$$

Now, since the real and imaginary part of the complex Gaussian components are i.i.d. real Gaussian random variables, it follows that the resulting $2N_t \times 1$ vector

$$\mathbf{o} = \begin{bmatrix} \mathrm{Re}(\mathbf{v}_{c,1}) \\ \mathrm{Im}(\mathbf{v}_{c,1}) \end{bmatrix} \tag{92}$$

is of the form of (90). Hence, it is uniformly distributed over the surface of the ($2N_t$-dimensional) real unit sphere and thus it can be interpreted as the first vector from a real matrix $\mathbf{O}$ drawn from CRE ensemble. Therefore

$$\left\|\mathbf{D}^{1/2}\mathbf{o}\right\| \|\mathbf{a}\| \tag{93}$$

which equals (88) has the same distribution as

$$\left\|\mathbf{D}^{1/2}\mathbf{O}\mathbf{a}\right\|. \tag{94}$$

12

It follows that $\left\|\mathbf{D}^{1/2}\mathbf{Va}\right\|$ and $\left\|\mathbf{D}^{1/2}\mathbf{Oa}\right\|$ have the same distribution.

APPENDIX B
PROOF OF THEOREM 1

From Lemma 2, we have

$$\Pr(R_{\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C)$$

$$\leq \sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \left(2^{\frac{C-\Delta C}{N_t}} \alpha(N_t)\right)^{N_t - 1/2}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}, \quad (95)$$

where $\mathbb{A}$ and β are defined in (44) and (46). Reverting back to (58), this summation can be written as

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t - 1}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}$$

$$\leq \sum_{k=1}^{\left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor} \sum_{k \leq \|\mathbf{a}\| \leq k+1} \frac{2N_t \sqrt{\beta}^{2N_t - 1}}{k^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}. \quad (96)$$

Denoting $\eta = \eta(N_t, \beta, d_{\min}, C) = \frac{2N_t \sqrt{\beta}^{2N_t - 1}}{2^{C \frac{2}{\sqrt{d_{\min}}}}}$ and using Lemma 1 in [24], this may be further bounded as

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t - 1}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}$$

$$\leq \sum_{k=1}^{\left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor} \eta \mathrm{Vol}(\mathcal{B}_{2N_t}(1)) \times$$

$$\left[\frac{\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(\max\left(k - \frac{\sqrt{2N_t}}{2}, 0\right)\right)^{2N_t}}{k^{2N_t - 1}} \right]. \quad (97)$$

Further, the right hand side of (97) can be rewritten as

$$\eta \mathrm{Vol}(\mathcal{B}_{2N_t}(1)) \left[\underbrace{\sum_{k=1}^{\left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor} \frac{\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}}{k^{2N_t - 1}}}_{\mathrm{I}} + \right.$$

$$\left. \underbrace{\sum_{k=\left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor + 1}^{\left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor} \frac{\left[\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right]}{k^{2N_t - 1}}}_{\mathrm{II}} \right]. \quad (98)$$

We search for c_1 and c_2 (independent of k) such that

$$\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} \leq c_1 k^{2N_t - 1} \quad (99)$$

for $1 \leq k \leq \left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor$, and

$$\left[\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right] \leq c_2 k^{2N_t - 1} \quad (100)$$

for $k \geq 1$, since it will then follow that

$$I + II \leq \sum_{k=1}^{\left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor} \max(c_1, c_2)$$

$$= \left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor \eta \mathrm{Vol}(\mathcal{B}_{2N_t}(1)) \max(c_1, c_2). \quad (101)$$

First, we show that we may take

$$c_1 = \left(1 + \sqrt{2N_t}\right)^{2N_t}. \quad (102)$$

In other words, we need to establish that

$$\frac{\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}}{k^{2N_t - 1}} \leq \left(1 + \sqrt{2N_t}\right)^{2N_t} \quad (103)$$

holds for $1 \leq k \leq \left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor$. Since $k \geq 1$, we have

$$\frac{\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}}{k^{2N_t - 1}} \leq \left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}. \quad (104)$$

Now, for $k \leq \frac{\sqrt{2N_t}}{2}$, we have

$$\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} \leq \left(1 + \sqrt{2N_t}\right)^{2N_t}. \quad (105)$$

Hence, (99) indeed holds with $c_1 = \left(1 + \sqrt{2N_t}\right)^{2N_t}$.

Next, we show that we may take

$$c_2 = \left[\left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right]. \quad (106)$$

Thus, we need to show that for $k \geq 1$, the following holds

$$\frac{1}{k^{2N_t - 1}} \left[\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right]$$

$$\leq \left[\left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right]. \quad (107)$$

Using the binomial expansion

$$\frac{1}{k^{2N_t - 1}} \left[\left(k + 1 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(k - \frac{\sqrt{2N_t}}{2}\right)^{2N_t}\right]$$

$$= \sum_{i=0}^{2N_t} \binom{2N_t - i}{i} k^{1-i} \left[\left(1 + \frac{\sqrt{2N_t}}{2}\right)^{i} - \left(-\frac{\sqrt{2N_t}}{2}\right)^{i}\right]$$

$$= \sum_{i=1}^{2N_t} \binom{2N_t - i}{i} k^{1-i} \left[\left(1 + \frac{\sqrt{2N_t}}{2}\right)^{i} - \left(-\frac{\sqrt{2N_t}}{2}\right)^{i}\right] \quad (108)$$

13

$$\leq \sum_{i=1}^{2N_t} \binom{2N_t - i}{i} \left[\left(1 + \frac{\sqrt{2N_t}}{2}\right)^i - \left(-\frac{\sqrt{2N_t}}{2}\right)^i \right] \quad (109)$$

$$= \left[\left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2}\right)^{2N_t} \right], \quad (110)$$

where (109) follows since each of the summands in (108) is monotonically decreasing in k (when $k \geq 1$ and $i \geq 1$). Thus, (100) indeed holds when taking c_2 as defined in (106).

Substituting (102) and (106) into (98), we get

$$I + II$$

$$\leq \left[\sum_{k=1}^{\left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor} \left(1 + \sqrt{2N_t}\right)^{2N_t} \right.$$

$$\left. + \sum_{k=\left\lfloor \frac{\sqrt{2N_t}}{2} \right\rfloor + 1}^{\left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor} \left[\left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2}\right)^{2N_t} \right] \right].$$
$$(111)$$

Now, since for $N_t \geq 2$ we have

$$c_2 = \left[\left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} - \left(1 - \frac{\sqrt{2N_t}}{2}\right)^{2N_t} \right]$$

$$\leq \left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}$$

$$\leq \left(1 + \sqrt{2N_t}\right)^{2N_t}$$

$$= c_1, \quad (112)$$

it follows that

$$I + II \leq \left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor \left(1 + \sqrt{2N_t}\right)^{2N_t}. \quad (113)$$

Applying (113) to (97), we get

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, d_{\min})} \frac{2N_t \sqrt{\beta}^{2N_t - 1}}{\|\mathbf{a}\|^{2N_t - 1} 2^{C \frac{2}{\sqrt{d_{\min}}}}}$$

$$\leq \left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor \eta \mathrm{Vol}(\mathcal{B}_{2N_t}(1)) \cdot \left(1 + \sqrt{2N_t}\right)^{2N_t}$$

$$= \left\lfloor \sqrt{\frac{\beta}{d_{\min}}} \right\rfloor \frac{2 \left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \sqrt{\beta}^{2N_t - 1}}{2^{C \frac{2}{\sqrt{d_{\min}}}}} \mathrm{Vol}(\mathcal{B}_{2N_t}(1))$$

$$\leq \frac{2 \left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \sqrt{\beta}^{2N_t - 1}}{2^{C \frac{2}{\sqrt{d_{\min}}}}} \sqrt{\frac{\beta}{d_{\min}}} \mathrm{Vol}(\mathcal{B}_{2N_t}(1))$$

$$= \frac{\left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \beta^{N_t}}{2^C} \mathrm{Vol}(\mathcal{B}_{2N_t}(1)). \quad (114)$$

Further, setting $\beta = 2^{\frac{C - \Delta C}{N_t}} \alpha(N_t)$, we have

$$\Pr \left(R_{\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right)$$

$$\leq \frac{\left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \left(2^{\frac{C - \Delta C}{N_t}} \alpha(N_t)\right)^{N_t}}{2^C} \mathrm{Vol}(\mathcal{B}_{2N_t}(1))$$

$$= \frac{\left(1 + \sqrt{2N_t}\right)^{2N_t} N_t 2^{C - \Delta C} \alpha(N_t)^{N_t}}{2^C} \mathrm{Vol}(\mathcal{B}_{2N_t}(1))$$

$$= \left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)} 2^{-\Delta C}. \quad (115)$$

Hence,

$$\Pr \left(R_{\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right) \leq c(N_t) 2^{-\Delta C} \quad (116)$$

where

$$c(N_t) = \left(1 + \sqrt{2N_t}\right)^{2N_t} N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)} \quad (117)$$

is a constant that depends only on N_t. We note that (117) does not depend on $\mathbf{D}$ and hence it holds also for the supremum over $\mathbf{D}$. Recalling (32), we have

$$P_{\mathrm{out,IF}}^{\mathrm{WC}}(C, \Delta C) = \sup_{\mathbf{D} \in \mathbb{D}(C)} \Pr \left(R_{\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - \Delta C \right)$$

$$\leq c(N_t) 2^{-\Delta C} \quad (118)$$

which concludes the proof.

Remark 6. *It can be shown that the choice* $c_1 = \left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t}$ *also satisfies (99) and hence (117) can be further tightened to*

$$c(N_t) = \left(2 + \frac{\sqrt{2N_t}}{2}\right)^{2N_t} N_t \alpha(N_t)^{N_t} \frac{\pi^{N_t}}{\Gamma(N_t + 1)}. \quad (119)$$

Figure 6 uses this tighter bound.

APPENDIX C
TIGHTER BOUNDS FOR $N_r \times 2$ CHANNELS: PROOF OF LEMMA 3 AND THEOREM 2

We consider the performance of an IF-SIC receiver over $N_r \times 2$ channels. As mentioned in Section III-C, when using complex linear pre-processing matrices, the rates of both IF and IF-SIC come in pairs. Denote

$$R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) = R_{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) = R_{2,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) \quad (120)$$

and

$$R^{2,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) = R_{3,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) = R_{4,\mathrm{IF}}(\mathbf{D}, \mathbf{V}). \quad (121)$$

Similarly, denote

$$R^{1,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{1,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{2,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) \quad (122)$$

and

$$R^{2,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{3,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R_{4,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}). \quad (123)$$

From Section III.A in [12], we have

$$\sum_{i=1}^4 R_{i,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = C - \log |\det(\mathbf{A})|. \quad (124)$$

14

Furthermore, by Theorem 3 in [12], the optimal integer matrix $\mathbf{A}$ for IF-SIC is unimodular (i.e., has determinant 1 or -1). Hence,

$$\sum_{i=1}^{4} R_{i,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = C. \tag{125}$$

Since we use IF-SIC with equal rate per stream, we have the following

$$R_{\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = 4\min(R^{1,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}), R^{2,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V})). \tag{126}$$

When applying IF-SIC, one decodes first the equation with the *highest* SNR. Since for this equation SIC has no effect it follows that

$$R^{1,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}). \tag{127}$$

Substituting (127) into (125), we have

$$R^{2,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = \frac{C}{2} - R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}). \tag{128}$$

Now, from Theorem 3 in [11] (with equivalent four real dimensions), we have

$$2(R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) + R^{2,\mathrm{IF}}(\mathbf{D}, \mathbf{V})) > C - 4$$
$$R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) + R^{2,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) > \frac{C-4}{2}. \tag{129}$$

Since $R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) > R^{2,\mathrm{IF}}(\mathbf{D}, \mathbf{V})$, it follows that

$$R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) > \frac{C-4}{4}. \tag{130}$$

We conclude that

$$R_{\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) = 4\min\left(R^{1,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}), R^{2,\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V})\right)$$
$$= 4\min\left(R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}), \frac{C}{2} - R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V})\right)$$
$$> 4\min\left(\frac{C-4}{4}, \frac{C}{2} - R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V})\right)$$
$$= \min\left(C - 1, 2C - 4R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V})\right). \tag{131}$$

Henceforth, we analyze the outage for $C > 1$ and target rates that are no greater than $C - 1$, so that the inequality $2C - 4R^{1,\mathrm{IF}}(\mathbf{D}, \mathbf{V}) < C - 1$ is satisfied. Thus, we consider gap-to-capacity values such that $\Delta C > 1$. Our goal is to bound

$$\Pr\left(R_{\mathrm{IF-SIC}}(\mathbf{D}, \mathbf{V}) < C - \Delta C\right)$$
$$= \Pr\left(2C - 4\frac{1}{2}\log\left(\frac{1}{\lambda_1^2(\Lambda)}\right) < C - \Delta C\right)$$
$$= \Pr\left(-2\log\left(\frac{1}{\lambda_1^2(\Lambda)}\right) < -(C + \Delta C)\right)$$
$$= \Pr\left(\lambda_1^2(\Lambda) < 2^{-1/2(C+\Delta C)}\right). \tag{132}$$

We are now ready to prove Lemma 3 and Theorem 2. Let $\beta = 2^{-1/2(C+\Delta C)}$. We wish to bound (132), or equivalently

$$\Pr\left(\lambda_1^2(\Lambda) < \beta\right) = \Pr\left(\lambda_1(\Lambda) < \sqrt{\beta}\right). \tag{133}$$

for a given matrix $\mathbf{D}$. Note that the event $\lambda_1(\Lambda) < \sqrt{\beta}$ is equivalent to the event

$$\bigcup_{\mathbf{a} \in \mathbb{Z}^{2\mathbf{N_t}} \setminus \{\mathbf{0}\}} ||\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}|| < \sqrt{\beta}. \tag{134}$$

Applying the union bound yields

$$\Pr\left(\lambda_1(\Lambda) < \sqrt{\beta}\right) < \sum_{\mathbf{a} \in \mathbb{Z}^{2N_t}} \Pr\left(||\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}|| < \sqrt{\beta}\right). \tag{135}$$

Note that if $\frac{||\mathbf{a}||}{\sqrt{d_{\max}}} > \sqrt{\beta}$, we have

$$\Pr\left(||\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}|| < \sqrt{\beta}\right) = 0. \tag{136}$$

Therefore, using (44), the set of relevant vectors $\mathbf{a}$ is

$$\mathbb{A}(\beta, 1/d_{\max}) = \left\{\mathbf{a} : 0 < ||\mathbf{a}|| < \sqrt{\beta d_{\max}}\right\}. \tag{137}$$

It follows from (135) and (136) that

$$\Pr\left(\lambda_1(\Lambda) < \sqrt{\beta}\right) < \sum_{\mathbf{a} \in \mathbb{A}(\beta, 1/d_{\max})} \Pr\left(||\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}|| < \sqrt{\beta}\right). \tag{138}$$

We now apply a similar derivation to that of Section IV. Applying Lemma 1, we have

$$\Pr\left(||\mathbf{D}^{-1/2}\mathbf{V}^T\mathbf{a}|| < \sqrt{\beta}\right) = \Pr\left(||\mathbf{D}^{-1/2}\mathbf{O}^T\mathbf{a}|| < \sqrt{\beta}\right). \tag{139}$$

where $\mathbf{O}$ is drawn from the CRE. Hence, we can apply the same geometric interpretation as in Section IV and interpret $\Pr\left(||\mathbf{D}^{-1/2}\mathbf{O}^T\mathbf{a}|| < \sqrt{\beta}\right)$ as the ratio of the surface area of the four-dimensional ellipsoid inside a ball with radius $\sqrt{\beta}$ and the surface area of this ellipsoid. The axes of this ellipsoid are defined as

$$x_i = \frac{||\mathbf{a}||}{\sqrt{d_i}}. \tag{140}$$

For the case of four real dimensions (52) can be written as

$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, 1/d_{\max})} \Pr\left(||\mathbf{D}^{-1/2}\mathbf{o}_{||\mathbf{a}||}|| < \sqrt{\beta}\right) =$$
$$\sum_{\mathbf{a} \in \mathbb{A}(\beta, 1/d_{\max})} \frac{\mathrm{CAP_{ell}}}{L(x_1, x_2, x_3, x_4)} \tag{141}$$

where

$$\mathrm{CAP_{ell}} < A_4(\sqrt{\beta}) = 4\frac{\pi^2}{2}\sqrt{\beta}^3 \triangleq \overline{\mathrm{CAP_{ell}}}, \tag{142}$$

and

$$L(x_1, x_2, x_3, x_4) > \frac{\pi^2}{2}\frac{||\mathbf{a}||^4}{\prod_{i=1}^{4}\sqrt{d_i}}\left(\frac{2\sqrt{d_{\min}}}{||\mathbf{a}||} + \frac{2\sqrt{d_{\max}}}{||\mathbf{a}||}\right)$$
$$> \pi^2\frac{||\mathbf{a}||^3}{2^C}\left(\sqrt{d_{\max}}\right) \triangleq \underline{L}(x_1, x_2, x_3, x_4). \tag{143}$$

15

Substituting (142) and (143) in (141), we obtain

$$\sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{\mathrm{CAP}_{\mathrm{ell}}}{L(x_1,x_2,x_3,x_4)} < \sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{\overline{\mathrm{CAP}_{\mathrm{ell}}}}{\underline{L}(x_1,x_2,x_3,x_4)}$$

$$= \sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{2\pi^2\sqrt{\beta}^3}{\pi^2\frac{\|\mathbf{a}\|^3}{2^C}\left(\sqrt{d_{\max}}\right)} \tag{144}$$

Recalling that $\beta = 2^{-1/2(C+\Delta C)}$, we get that for $\Delta C < 1$

$$\Pr\left(R_{\mathrm{IF-SIC}}(\mathbf{D},\mathbf{V}) < C - \Delta C\right)$$
$$< \sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{2\pi^2 2^{-3/4(C+\Delta C)}}{\pi^2\frac{\|\mathbf{a}\|^3}{2^C}\left(\sqrt{d_{\max}}\right)}, \tag{145}$$

which proves Lemma 3.

To establish Theorem 2, we follow the footsteps of the proof of Theorem 1 to obtain

$$\sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \Pr\left(\|\mathbf{D}^{-1/2}\mathbf{o}_{\|\mathbf{a}\|} < \sqrt{\beta}\right)$$

$$< \sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{2\pi^2\sqrt{\beta}^3}{\pi^2\frac{\|\mathbf{a}\|^3}{2^C}\left(\sqrt{d_{\max}}\right)}$$

$$= \sum_{\mathbf{a}\in\mathbb{A}(\beta,1/d_{\max})} \frac{2\sqrt{\beta}^3 2^C}{\|\mathbf{a}\|^3\left(\sqrt{d_{\max}}\right)}$$

$$< \sum_{k=1}^{\lfloor\sqrt{\beta d_{\max}}\rfloor}\sum_{k\leq\|\mathbf{a}\|\leq k+1} \frac{2\sqrt{\beta}^3 2^C}{k^3\left(\sqrt{d_{\max}}\right)}$$

$$= \sum_{k=1}^{\lfloor\sqrt{\beta d_{\max}}\rfloor} \frac{2\sqrt{\beta}^3 2^C}{k^3\left(\sqrt{d_{\max}}\right)}\frac{\pi^2}{2}\left[(k+2)^4-(k-1)^4\right]$$

$$< \sum_{k=1}^{\lfloor\sqrt{\beta d_{\max}}\rfloor} \frac{\sqrt{\beta}^3 2^C\pi^2}{k^3\left(\sqrt{d_{\max}}\right)}81k^3$$

$$= \sum_{k=1}^{\lfloor\sqrt{\beta d_{\max}}\rfloor} \frac{\sqrt{\beta}^3 2^C\pi^2 81}{\left(\sqrt{d_{\max}}\right)}$$

$$\leq \frac{\sqrt{\beta}^3 2^C\pi^2 81\sqrt{\beta d_{\max}}}{\left(\sqrt{d_{\max}}\right)}$$

$$= \beta^2 2^C\pi^2 81. \tag{146}$$

As (146) does not depend on $\mathbf{D}$, it follows that the bound holds also for the supremum over $\mathbf{D}$. Now since $\beta = 2^{-1/2(C+\Delta C)}$, we get

$$P_{\mathrm{out,IF-SIC}}^{\mathrm{WC}}(C,\Delta C) \leq 81\cdot\pi^2\cdot 2^C 2^{-C-\Delta C}$$
$$= 81\cdot\pi^2\cdot 2^{-\Delta C}. \tag{147}$$

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun*, vol. 10, pp. 585–598, Nov. 1999.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[3] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *Communications, IEEE Transactions on*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[4] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 323–340, Jan. 2015.

[5] P. Elia, K. Kumar, S. Pawar, P. Kumar, and H. feng Lu, "Space-time codes meeting the diversity-multiplexing gain tradeoff with low signalling complexity," *Information Theory, IEEE Transactions on*, vol. 52, no. 9, pp. 3869–3884, Sept. 2006.

[6] E. G. Larsson, "Constellation randomization (CoRa) for outage performance improvement on MIMO channels," in *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*, vol. 1, 2004, pp. 386–390.

[7] Y. Li, C. Georghiades, and G. Huang, "Transmit diversity over quasi-static fading channels using multiple antennas and random signal mapping," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1918–1926, Nov. 2003.

[8] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[9] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[10] O. Ordentlich, J. Zhan, U. Erez, M. Gastpar, and B. Nazer, "Practical code design for compute-and-forward," in *2011 IEEE International Symposium on Information Theory Proceedings*, July 2011, pp. 1876–1880.

[11] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *IEEE Transactions on Information Theory*, vol. 60, no. 6, pp. 3450–3482, June 2014.

[12] ——, "Successive integer-forcing and its sum-rate optimality," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2013, pp. 282–292.

[13] M. Mehta, *Random matrices and the statistical theory of energy level*. Academic Press, 1967.

[14] F. Mezzadri, "How to generate random matrices from the classical compact groups," *arXiv preprint math-ph/0609050*, 2006.

[15] J. C. Lagarias, H. W. Lenstra Jr, and C.-P. Schnorr, "Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice," *Combinatorica*, vol. 10, no. 4, pp. 333–348, 1990.

[16] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Mathematische Annalen*, vol. 296, no. 1, pp. 625–635, 1993.

[17] H. F. Blichfeldt, "The minimum value of quadratic forms, and the closest packing of spheres," *Mathematische Annalen*, vol. 101, no. 1, pp. 605–608, 1929.

[18] B. C. Carlson, "Some inequalities for hypergeometric functions," *Proceedings of the American Mathematical Society*, vol. 17, no. 1, pp. 32–39, 1966.

[19] G. J. Tee, "Surface area and capacity of ellipsoids in n dimensions," *New Zealand Journal of Mathematics*, vol. 34, pp. 165–198, 2005.

[20] A. Sakzad, J. Harshan, and E. Viterbo, "On complex LLL algorithm for integer forcing linear receivers," in *2013 Australian Communications Theory Workshop (AusCTW)*, Jan. 2013, pp. 13–17.

[21] R. F. Fischer, M. Cyran, and S. Stern, "Factorization approaches in lattice-reduction-aided and integer-forcing equalization," in *Proc. Int. Zurich Seminar on Communications, Zurich, Switzerland*, 2016.

[22] A. Khina, I. Livni, A. Hitron, and U. Erez, "Joint unitary triangularization for Gaussian multi-user MIMO networks," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2662–2692, May 2015.

[23] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2418–2433, Nov. 1999.

[24] O. Ordentlich and U. Erez, "A simple proof for the existence of "good" pairs of nested lattices," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4439–4453, Aug. 2016.

Elad Domanovitz received the B.Sc. degree (cum laude) and the M.Sc. degree in 2005 and 2011, respectively, in electrical engineering from Tel Aviv University, Israel. He is currently working toward the Ph.D. degree at Tel Aviv University.

16

Uri Erez (M'09) was born in Tel-Aviv, Israel, on October 27, 1971. He received the B.Sc. degree in mathematics and physics and the M.Sc. and Ph.D. degrees in electrical engineering from Tel-Aviv University in 1996, 1999, and 2003, respectively. During 2003-2004, he was a Postdoctoral Associate at the Signals, Information and Algorithms Laboratory at the Massachusetts Institute of Technology (MIT), Cambridge. Since 2005, he has been with the Department of Electrical Engineering-Systems at Tel-Aviv University. His research interests are in the general areas of information theory and digital communication. He served in the years 2009-2011 as Associate Editor for Coding Techniques for the IEEE TRANSACTIONS ON INFORMATION THEORY.