Precoded Integer-Forcing Universally Achieves the MIMO Capacity to Within a Constant Gap

Or Ordentlich and Uri Erez, Member, IEEE

Abstract—An open-loop single-user multiple-input multiple-output communication scheme is considered where a transmitter, equipped with multiple antennas, encodes the data into independent streams all taken from the same linear code. The coded streams are then linearly precoded using the encoding matrix of a perfect linear dispersion space-time code. At the receiver side, integer-forcing equalization is applied, followed by standard single-stream decoding. It is shown that this communication architecture achieves the capacity of any Gaussian multiple-input multiple-output channel up to a gap that depends only on the number of transmit antennas.

I. INTRODUCTION

The Gaussian Multiple-Input Multiple-Output (MIMO) channel has been the focus of extensive research efforts since the pioneering works of Foschini [1], Foschini and Gans [2], and Telatar [3]. While the capacity of the channel, under various assumptions on channel state information, is easy to derive, practical schemes that allow to approach capacity are only known in the two extremes: when the channel is either ergodic [4], with or without channel state information at the transmitter, or static where the transmitter knows the channel matrix [3]. In contrast, this paper considers a static scenario where the receiver has full channel knowledge whereas the transmitter either has no knowledge of the channel, or alternatively knows only its capacity.

More specifically, a single-user complex MIMO channel

\[ y = Hx + z \]  

(1)

with \( M \) transmit and \( N \) receive antennas is considered. The input vector \( x \) is subject to the power constraint

\[ \mathbb{E}(x^H x) \leq M \cdot \text{SNR}, \]

and the additive noise \( z \) is a vector of i.i.d. circularly symmetric complex Gaussian entries with zero mean and unit variance.

The mutual information of this channel is maximized by a circularly symmetric complex Gaussian input [3] with covariance matrix \( Q \) satisfying \( \text{trace}(Q) \leq M \cdot \text{SNR} \), and is given by

\[ C = \max_{Q>0, \text{trace}(Q) \leq M \cdot \text{SNR}} \log \det (I + QH^H H). \]  

(2)

The choice of \( Q \) that maximizes (2) is determined by the water-filling solution. When the matrix \( H \) is known at both transmission ends, i.e., in a closed-loop scenario, this mutual information is the capacity of the channel and may closely be approached using the singular-value decomposition in conjunction with standard scalar codes designed for an additive white Gaussian noise (AWGN) channel. Often, the sub-optimal choice \( Q = \text{SNR} \cdot I \) is used, resulting in the white-input (WI) mutual information

\[ C_{\text{WI}} = \log \det (I + \text{SNR} H^H H). \]

For all channel matrices \( H \) and all values of \( \text{SNR} \), the WI mutual information loses less than \( M \log M \) bits w.r.t. capacity [5], [6]. Thus, a scheme that performs a constant gap from \( C_{\text{WI}} \) also performs a constant gap from the closed-loop capacity \( C \). In this paper, an open-loop setting is considered. In this case, transmitting a white input is a natural choice and \( C_{\text{WI}} \) will serve as a benchmark in the sequel.

While the theoretical performance limits of open-loop communication over a Gaussian MIMO channel are well understood, unlike for closed-loop transmission, much is still lacking when it comes to practical schemes that are able to approach these limits. Such a scheme is known for the \( 1 \times 2 \) MISO channel where Alamouti modulation offers an optimal solution. More generally, modulation via orthogonal space-time block “codes” allows to approach the WI mutual information using scalar AWGN coding and decoding in the limit of small rate, where the mutual information is governed solely through the Frobenius norm of the channel matrix.

Beyond the low rate regime, the multiple degrees of freedom offered by the channel need to be utilized in order to approach capacity. For this reason, despite considerable work and progress, the problem of designing a practical open-loop scheme that simultaneously approaches \( C_{\text{WI}} \) for all channels \( H \) with the same white-input mutual information remains unsolved. As a consequence, less demanding benchmarks became widely accepted in the literature. First, since statistical modeling of a wireless communication link is often available, one may be content with guaranteeing good performance only for channel realizations that have a “high” probability. Further, to simplify analysis and design, the asymptotic criterion of the diversity-multiplexing tradeoff (DMT) [6] has broadly been adopted.

Unfortunately, statistical characterizations, and the DMT criterion in particular, offer only a coarse figure of merit for assessing schemes. Specifically, assuming an i.i.d. fading model with a continuous distribution on the channel coefficients precludes the possibility of having an entire row in the channel matrix nulled out. As a result, DMT optimality...
of a scheme for a MIMO channel with $M$ transmit and $N$ receive antennas does not imply its optimality for a different number of receive antennas. Thus, the DMT framework is inadequate for analyzing communication scenarios with degrees-of-freedom mismatch, i.e., when the transmitter does not know in advance the number of receive antennas, or alternatively, has to simultaneously transmit (the same message) to several users with a different number of receive antennas.

In [7], Tavildar and Vishwanath introduced the notion of approximately universal space-time codes and derived a necessary and sufficient criterion for a code to be approximately universal. This criterion is closely related to the nonvanishing determinant criterion and is met by several known coding schemes [8]–[10]. Roughly speaking, approximate-universality guarantees that a scheme is DMT optimal for any statistical channel model. The criterion derived in [7] ensures that the minimum distance at the receiver scales appropriately with $C_{W1}$ regardless of the exact realization of $H$, which, in turn, guarantees DMT optimality. Thus, the problem of finding coding schemes that are DMT optimal regardless of the channel statistics is now solved.

Approximately universal schemes still suffer, however, from the asymptotic nature of the DMT criterion. Essentially, the approximate universality of a scheme guarantees that if the white-input mutual information of the MIMO channel is $C_{W1}$, the scheme’s error probability at a certain rate $R$ scales roughly as $Q(\sqrt{2^{R-C_{W1}}})$, for large $C_{W1}$. This is the same error probability behavior as that of uncoded transmission over a single-input single-output (SISO) AWGN channel with capacity $C_{W1}$. This may suffice when $C_{W1}$ is large enough and moderate error probabilities are required, but does not provide performance guarantees for finite values of $C_{W1}$.

While designing a practical communication scheme that universally approaches $C_{W1}$ is still out of reach, in the present work we take a step in this direction. Namely, a practical communication architecture that achieves the capacity of any MIMO channel up to a constant gap, that depends only on the number of transmit antennas, is studied. Such traditional information-theoretic performance guarantee is substantially stronger than approximate universality. In the considered scheme, which we term precoded integer-forcing, the transmitter encodes the data into independent streams via the same linear code. The coded streams are then linearly precoded using the generating matrix of a space-time code from the class of perfect codes [9]–[12], which are approximately universal. At the receiver side, integer-forcing (IF) equalization [13] is applied.

An IF receiver attempts to decode a full-rank set of linear combinations of the transmitted streams with integer-valued coefficients. Once these equations are decoded, they can be solved for the transmitted streams. The receiver’s front end consists of a linear equalization matrix that transforms the MIMO channel into a set of SISO sub-channels, each corresponding to a different linear combination, with an effective SNR that depends on the integer coefficients of this linear combination. The performance of the scheme is dictated by the worst effective SNR, over all sub-channels.

Precoded IF may be viewed as an extension of linear dispersion space-time “codes”. In such “codes”, uncoded QAM symbols are linearly modulated over space and time. This is done by linearly precoding the QAM symbols using a precoding matrix $P$. For precoded IF, the same precoding matrix $P$ is applied to the transmitted streams in each sub-channel. The performance of linear dispersion space-time “codes” is dictated by $d_{\text{min}}$, the minimum distance in the received constellation, whereas the performance of precoded IF is determined by the effective signal-to-noise ratio $\text{SNR}_{\text{eff}}$. A key result we derive is that the two quantities are closely related. Namely, minimum distance guarantees for precoded QAM symbols translate to guarantees on the effective SNR for precoded IF, when the same precoding matrix is used.

The design of precoding matrices for uncoded QAM, that guarantee an appropriate growth of $d_{\text{min}}$ as a function of $C_{W1}$, has been extensively studied over the last decade. A remarkable family of such matrices are the generating matrices of perfect linear dispersion space-time codes, which are approximately universal [9], [10]. As a consequence of
the tight connection between $d_{\text{min}}$ and $\text{SNR}_{\text{eff}}$, when such precoding matrices are used for precoded IF, $\text{SNR}_{\text{eff}}$ also grows appropriately with $C_{\text{W1}}$. Consequently, precoded IF achieves rates within a constant gap from $C_{\text{W1}}$, and hence also from the capacity, of any Gaussian MIMO channel.

Integer-forcing equalization essentially reduces to lattice-reduction (LR) in the case of uncoded transmission. Lattice-reduction aided receivers for perfect space-time modulated QAM constellations were considered in the literature, and were shown to be DMT optimal \[13\]. The key difference is that while the latter approach involves uncoded transmission and symbol-by-symbol detection, the proposed architecture uses linearly coded streams and the detection phase is replaced with equalization and decoding. This in turn, leads to performance guarantees that are valid at any (fixed) transmission rate.

The rest of the paper is outlined as follows. Section II gives an overview on IF equalization and analyzes its performance under various assumptions. In Section IV, several properties of perfect linear dispersion space-time codes are recalled and a lower bound on their worst-case minimum distance is derived, setting the grounds for the main result. The proof that precoded IF achieves the capacity of any MIMO channel to within a constant gap is given in Section V. As an example of the advantages of the proposed approach, low-complexity constructions of rateless coding schemes, which are based on precoded IF, are derived in Section VI. Concluding remarks appear in Section VII.

II. PERFORMANCE OF THE INTEGER-FORCING SCHEME

Integer-forcing equalization is a low-complexity architecture for the MIMO channel, which was proposed by Zhan et al. \[13\]. The key idea underlying IF is to first decode integral linear combinations of the signals transmitted by all antennas, and then, after the noise is removed, invert those linear combinations to recover the individual transmitted signals. This is made possible by transmitting codewords from the same linear/lattice code from all $M$ transmit antennas, leveraging the property that linear codes are closed under (modulo) linear combinations with integer-valued coefficients.

In this section we review and extend some of the results of \[13\] and \[15\] in a way that is suitable for our purposes.

A. Nested Lattice Codes

Let $\Lambda_c \subset \Lambda_f$ be a pair of $n$-dimensional nested lattices (see \[16\] for a more thorough treatment of lattice definitions and properties). The lattice $\Lambda_c$ is referred to as the coarse lattice and $\Lambda_f$ as the fine lattice. Denote by $V_c$ the Voronoi region of $\Lambda_c$, and define the second moment of $\Lambda_c$ as $\sigma^2(\Lambda_c) \triangleq \frac{1}{n \text{Vol}(V_c)} \int_{u \in V_c} \|u\|^2 du$, where $\text{Vol}(V_c)$ is the volume of $V_c$. A nested lattice codebook $C = \Lambda_f \cap V_c$, with rate $R = \frac{1}{n} \log |\Lambda_f \cap V_c|$ bits/channel use is associated with the nested lattice pair. The codebook is scaled such that $\sigma^2(\Lambda_c) = \text{SNR}/2$.

Example 1: We give three examples of common structures of nested lattice codebooks. See Figure 2 for illustration. More examples can be found in \[17\].

- **Uncoded transmission**: The simplest nested lattice codebook is an uncoded one, where the fine lattice $\Lambda_f$ is the integer lattice $\mathbb{Z}$ whereas the coarse lattice is $\Lambda_c = q\mathbb{Z}$ for some integer $q > 1$. The Voronoi region in this case is $V_c = [-q/2, q/2]$ and the obtained nested lattice codebook $C$ consists of all integers in the interval $[-q/2, q/2]$. The rate of this codebook is $R = \log q$ bits/channel use.
- **$q$-ary linear code without shaping**: A more sophisticated, yet reasonable to implement, nested lattice codebook can be obtained by lifting a $q$-ary linear code with block length $n$ to Euclidean space using Construction A \[18\], \[19\], and taking the resulting lattice as $\Lambda_f$. The coarse lattice is taken as $\Lambda_c = q\mathbb{Z}^n$, as in the uncoded case. The obtained nested lattice codebook $C$ is therefore simply the $q$-ary linear code coupled with a PAM constellation.
- **“Good” nested lattice pair of high dimension**: A third option is to use a pair of lattices of high dimension where the fine lattice is “good” for coding over an AWGN channel, whereas the coarse lattice is “good” for mean squared error quantization (see \[16\] for precise definitions of “goodness”). The obtained nested lattice codebook admits a relatively simple performance analysis, that yields closed-form rate expressions. However, implementing such a codebook is more complicated (although some progress in this direction was made in \[20\]).

The performance improvement obtained by using such a codebook w.r.t. a $q$-ary linear code without shaping is bounded from above by $1/2 \log(2\pi e/12)$ bits per real dimension, provided that the $q$-ary linear code performs well over an AWGN channel.

B. Description of the IF scheme

In the IF scheme, the information bits to be transmitted are partitioned into $2M$ streams, labeled $\{1_{\text{Re}}, 1_{\text{Im}}, \ldots, M_{\text{Re}}, M_{\text{Im}}\}$. Each of the $2M$ streams is encoded by the nested lattice code $C$, producing $2M$ row vectors, each in $C \subset \mathbb{R}^{1 \times n}$. In particular, the stream $m_{\text{Re}}$, consisting of $nR$ information bits, is mapped to a lattice point $t_{m_{\text{Re}}} \in C$. Then, a random dither $d_{m_{\text{Re}}} \in \mathbb{R}^{1 \times n}$ uniformly distributed over $V_c$ and statistically independent of $t_{m_{\text{Re}}}$, known to both the transmitter and the receiver, is used to produce the signal $x_{m_{\text{Re}}} = [t_{m_{\text{Re}}} - d_{m_{\text{Re}}} \mod \Lambda_c]$. The signal $x_{m_{\text{Re}}}$ is uniformly distributed over $V_c$ and is statistically independent of $t_{m_{\text{Re}}}$ due to the Crypto Lemma \[16\] (Lemma 1). It follows that $\frac{1}{n} \mathbb{E}[\|x_{m_{\text{Re}}}\|^2] = \sigma^2(\Lambda_c) = \frac{\text{SNR}}{2}$.

A similar procedure is used to construct the signal $x_{m_{\text{Im}}}$. The $m$th antenna transmits the signal $x_m = x_{m_{\text{Re}}} + i x_{m_{\text{Im}}} \in \mathbb{C}^{1 \times n}$ over $n$ consecutive channel uses. Thus, the total transmission rate is $R_{\text{IF}} = 2MR$ bits/channel use.
Let $X \triangleq [x_1^T \cdots x_M^T]^T \in \mathbb{C}^{M \times n}$. The received signal is

$$Y = HX + Z,$$

where $Z \in \mathbb{C}^{N \times n}$ is a vector with i.i.d. circularly symmetric complex Gaussian entries. Letting the subscripts $\text{Re}$ and $\text{Im}$ denote the real and imaginary parts of a matrix, respectively, the channel can be expressed by its real-valued representation

$$\begin{bmatrix} Y_{\text{Re}} \\ Y_{\text{Im}} \end{bmatrix} = \begin{bmatrix} H_{\text{Re}} & -H_{\text{Im}} \\ H_{\text{Im}} & H_{\text{Re}} \end{bmatrix} \begin{bmatrix} X_{\text{Re}} \\ X_{\text{Im}} \end{bmatrix} + \begin{bmatrix} Z_{\text{Re}} \\ Z_{\text{Im}} \end{bmatrix},$$

which will be written as

$$\tilde{Y} = \tilde{H} \tilde{X} + \tilde{Z}$$

for notational compactness. Let

$$\tilde{T} \triangleq [t_{1,i}^T \cdots t_{M,i}^T]^T$$

be a $2M \times n$ real-valued matrix whose rows consist of the lattice points corresponding to the $2M$ bit streams, and

$$\tilde{D} \triangleq [d_{1,i}^T \cdots d_{M,i}^T]^T$$

be a $2M \times n$ real-valued matrix whose rows correspond to the $2M$ different dither vectors.

The IF receiver chooses an equalizing matrix $B \in \mathbb{R}^{2M \times 2N}$ and a full-rank target integer-valued matrix $A \in \mathbb{Z}^{2M \times 2M}$, and computes

$$\tilde{Y}_{\text{eff}} = \begin{bmatrix} B \tilde{Y} + \tilde{A} \tilde{D} \\ \tilde{A} \tilde{T} + (B \tilde{H} - A)\tilde{X} + B\tilde{Z} \end{bmatrix} \mod \Lambda_c$$

is a $2M \times n$ real-valued matrix with each row being a codeword in $\mathcal{C}$ owing to the linearity of the code,

$$Z_{\text{eff}} \triangleq (B \tilde{H} - A)\tilde{X} + B\tilde{Z}$$

is additive noise statistically independent of $V$ (as $\tilde{X}$, as well as $\tilde{Z}$ are statistically independent of $T$), and the notation $\mod \Lambda_c$ is to be understood as reducing each row of the obtained matrix modulo the coarse lattice. Each row of $\tilde{Y}_{\text{eff}}$ is the modulo sum of a codeword and effective noise. Thus, the IF receiver transforms the original MIMO channel into a set of $2M$ point-to-point modulo-additive sub-channels

$$\tilde{y}_{\text{eff},k} = [v_k + z_{\text{eff},k}] \mod \Lambda_c, \quad k = 1, \ldots, 2M.$$ \hspace{1cm} (6)

The additive noise vectors $z_{\text{eff},1}, \ldots, z_{\text{eff},2M}$ are not statistically independent. Therefore, strictly speaking, the $2M$ effective channels $\tilde{y}_{\text{eff},1}, \ldots, \tilde{y}_{\text{eff},2M}$ are not parallel. However, the IF decoder ignores the correlation between the noise vectors and decodes the output of each sub-channel separately. If the decoding is successful over all $2M$ sub-channels, the receiver has access to $V$, from which it can recover the matrix $T$ by solving the (modulo) set of equations (5).

Let $a_k^T$ and $b_k^T$ be the $k$th rows of $A$ and $B$, respectively, and define the effective variance of $z_{\text{eff},k}$ as

$$\sigma_{\text{eff},k}^2 \triangleq \frac{1}{n} E \|z_{\text{eff},k}\|^2 = \frac{1}{n} E \left\| (b_k^T \tilde{H} - a_k^T)\tilde{X} + b_k^T \tilde{Z} \right\|^2 = \frac{\text{SNR}}{2} \left\| (b_k^T \tilde{H} - a_k^T) \right\|^2 + \frac{1}{2} \|b_k^T\|^2.$$

A natural criterion for choosing the equalizing matrix $B$ and the target integer-valued matrix $A$ is to minimize the effective variance

\[ \rho_{\text{eff}}(B, A) \triangleq \frac{1}{n} E \|z_{\text{eff}}\|^2 = \frac{1}{n} E \left\| (b_k^T \tilde{H} - a_k^T)\tilde{X} + b_k^T \tilde{Z} \right\|^2 = \frac{\text{SNR}}{2} \left\| (b_k^T \tilde{H} - a_k^T) \right\|^2 + \frac{1}{2} \|b_k^T\|^2. \]

Some improvement can be obtained by exploiting these correlations. Yet, we do not pursue this possibility in the present paper.

\[ \text{In [22] it is shown that it suffices that } A \text{ is invertible over } \mathbb{R} \text{ in order to recover } T \text{ from } V. \]
noise variances. It turns out [23] that for a given matrix $A$, the optimal choice of $B$ under this criterion is

$$B^{\text{opt}} = A\tilde{H}^T \left( \frac{1}{\text{SNR}} I + \tilde{H}\tilde{H}^T \right)^{-1},$$

which results in the effective variances

$$\sigma^2_{\text{eff},k} = \frac{\text{SNR}}{2} a_k^T \left( I + \text{SNR}\tilde{H}\tilde{H}^T \right)^{-1} a_k,$$

for $k = 1, \ldots, 2M$.

Define the effective signal-to-noise ratio (SNR) at the $k$th sub-channel as

$$\text{SNR}_{\text{eff},k} \triangleq \frac{\sigma^2(\Lambda_c)}{\sigma^2_{\text{eff},k}},$$

$$= \frac{\text{SNR}}{2} a_k^T \left( I + \text{SNR}\tilde{H}\tilde{H}^T \right)^{-1} a_k,$$

$$= \left( a_k^T \left( I + \text{SNR}\tilde{H}\tilde{H}^T \right)^{-1} a_k \right)^{-1}, \quad (7)$$

and let

$$\text{SNR}_{\text{eff}} \triangleq \min_{k=1,\ldots,2M} \text{SNR}_{\text{eff},k}. \quad (8)$$

For IF equalization to be successful, decoding over all $2M$ sub-channels should be correct. Therefore, the worst sub-channel constitutes a bottleneck. For this reason, the total performance of the receiver is dictated by $\text{SNR}_{\text{eff}}$.

C. Achievable rates for IF

When the codebook $\mathcal{C}$ is constructed from a good pair of nested lattices (see Example 1), the distribution of the effective noise at each sub-channel $k$, which is a linear combination of an AWGN and $2M$ dither vectors, approaches that of an AWGN with zero mean and variance $\sigma^2_{\text{eff},k}$ [23]. Good nested lattice codebooks can achieve any rate satisfying

$$R < \frac{1}{2} \log \left( \text{SNR}_{\text{eff},k} \right) \quad (9)$$

over a mod $-\Lambda_c$ AWGN channel with signal-to-noise ratio $\text{SNR}_{\text{eff},k}$ [16], [23]. Since $v_k$ is a codeword from a good nested lattice code and $z_{\text{eff},k}$ approaches an AWGN in distribution, $v_k$ can be decoded [13], [23] from $\tilde{y}_{\text{eff},k}$ as long as the rate of the codebook $\mathcal{C}$ satisfies (9). It follows that as long as

$$R < \frac{1}{2} \log \left( \text{SNR}_{\text{eff}} \right),$$

all sub-channels $k = 1, \ldots, 2M$ can decode their linear combinations $v_k$ without error, and therefore IF equalization can achieve any rate satisfying

$$R_{\text{IF}} < 2M \frac{1}{2} \log \left( \text{SNR}_{\text{eff}} \right)$$

$$= M \log \left( \text{SNR}_{\text{eff}} \right). \quad (10)$$

As mentioned in Example 1 good nested lattice codebooks can be difficult to implement in practice. A more appealing alternative may be to use a $q$-ary linear code without shaping. In this case, the effective noise $z_{\text{eff},k}$ at each sub-channel is a linear combination of an AWGN and $2M$ random dithers uniformly distributed over the Voronoi region of a 1-D integer lattice. This effective noise is i.i.d. (in contrast to the case where a higher-dimensional coarse lattice is used where $z_{\text{eff},k}$ has memory). It was shown in [24] Remark 3] that, for a prime
When a specific $q$-ary linear code (such as an LDPC code or a turbo code) is used, the achievable rate is further degraded by $2M$ times the code’s gap-to-capacity at the target error probability.

Finally, consider the case of uncoded transmission. In this case, $\Lambda_f = \gamma \mathbb{Z}$ and $\Lambda_c = \gamma q \mathbb{Z}$, where $\gamma = \sqrt{\text{SNR}/(2q^2)}$ is chosen such as to meet the power constraint, and $q > 1$ is an integer (see Example 1). The performance of uncoded transmission with IF equalization followed by a simple slicer is characterized by the following lemma.

**Lemma 1:** The error probability of the IF receiver with uncoded transmission rate $R_{\text{IF}}$ is upper bounded by

$$P_{e,\text{IF-uncoded}} \leq 4M \exp \left\{ -\frac{3}{2} \frac{1}{2} (M \log(\text{SNR}_{\text{eff}}) - R_{\text{IF}}) \right\}.$$  \hfill (12)

**Proof:** See Appendix A

**Remark 1:** Integer-forcing equalization with uncoded transmission is quite similar to the extensively studied lattice-reduction-aided linear decoders framework [14], [25], [26]. However, two subtle differences should be pointed out. First, under the framework of LR-linear decoding, the target integer-valued matrix $\mathbf{A}$ has to be unimodular, i.e., it has to satisfy $\det(\mathbf{A}) = 1$, whereas in IF equalization $\mathbf{A}$ is only required to be full-rank. Second, the use of the dithers in IF equalization results in statistical independence between $\mathbf{v}_k$ and $\mathbf{z}_{\text{eff},k}$ at each of the $2M$ sub-channels. This allows for an exact rigorous analysis of the error probability, which is seemingly difficult under the LR framework.

**D. Bounding the Effective SNR for an optimal choice of $\mathbf{A}$**

In this subsection, a lower bound on $\text{SNR}_{\text{eff}}$ for an optimal choice of $\mathbf{A}$, is obtained.

The target integer-valued matrix $\mathbf{A}$ should be chosen such as to maximize $\text{SNR}_{\text{eff}}$. Using (7) and (9), this criterion translates to choosing

$$\mathbf{A}^{\text{opt}} = \arg\min_{\mathbf{A} \in \mathbb{Z}^{2M \times 2M}} \max_{k=1, \ldots, 2M} \frac{\mathbf{a}_k^T (\mathbf{I} + \text{SNR}\mathbf{H}^T \mathbf{H})^{-1} \mathbf{a}_k}{\det(\mathbf{A}) \neq 0}.$$  

The matrix $(\mathbf{I} + \text{SNR}\mathbf{H}^T \mathbf{H})^{-1}$ is symmetric and positive definite, and therefore it admits a Cholesky decomposition

$$(\mathbf{I} + \text{SNR}\mathbf{H}^T \mathbf{H})^{-1} = \mathbf{L} \mathbf{L}^T,$$  \hfill (13)

where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries. With this notation the optimization criterion becomes

$$\mathbf{A}^{\text{opt}} = \arg\min_{\mathbf{A} \in \mathbb{Z}^{2M \times 2M}} \max_{k=1, \ldots, 2M} \frac{\|\mathbf{L}^T \mathbf{a}_k\|^2}{\det(\mathbf{A}) \neq 0}.$$  

Denote by $\Lambda(\mathbf{L}^T)$ the $2M$ dimensional lattice spanned by the matrix $\mathbf{L}^T$, i.e.,

$$\Lambda(\mathbf{L}^T) = \{ \mathbf{L}^T \mathbf{a} : \mathbf{a} \in \mathbb{Z}^{2M} \}.$$  

It follows that $\mathbf{A}^{\text{opt}}$ should consist of the set of $2M$ linearly independent integer-valued vectors that result in the shortest set of linearly independent lattice vectors in $\Lambda(\mathbf{L}^T)$.

**Definition 1 (Successive minima):** 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

$$\lambda_k(\mathbf{G}) \triangleq \inf \{ r : \dim (\text{span} (\Lambda(\mathbf{G}) \cap \mathcal{B}(0,r))) \geq k \},$$

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

With the above definition of successive minima, the effective signal-to-noise ratio, when the optimal integer-valued matrix $\mathbf{A}^{\text{opt}}$ is used, can be written as

$$\text{SNR}_{\text{eff}} = \frac{1}{\lambda_{2M}^2(\mathbf{L}^T)}.$$  \hfill (14)

Bounding the value of the $2M$th successive minimum of a lattice is seemingly difficult. Fortunately, a transference theorem by Banaszczyk [27] relates the $2M$th successive minimum of a lattice to the first successive minimum of its dual lattice. Following the derivation from [13] Proof of Theorem 4], we proceed to bound $\text{SNR}_{\text{eff}}$ using this relation.

**Definition 2 (Dual lattice):** For a lattice $\Lambda(\mathbf{G})$ with a generating full-rank matrix $\mathbf{G} \in \mathbb{R}^{2M \times 2M}$ the dual lattice is defined by

$$\Lambda^*(\mathbf{G}) \triangleq \Lambda (\mathbf{G}^{-1})^{-1} = \{ (\mathbf{G}^{-1})^{-1} \mathbf{a} : \mathbf{a} \in \mathbb{Z}^{2M} \}.$$  

**Theorem 1 (Banaszczyk [27] Theorem 2.1):** Let $\Lambda(\mathbf{G})$ be a lattice with a full-rank generating matrix $\mathbf{G} \in \mathbb{R}^{K \times K}$ and let $\Lambda^*(\mathbf{G}) = \Lambda (\mathbf{G}^{-1})^{-1}$ be its dual lattice. The successive minima of $\Lambda(\mathbf{G})$ and $\Lambda^*(\mathbf{G})$ satisfy the following inequality

$$\lambda_k(\mathbf{G}) \lambda_{K-k+1}(\mathbf{G}^{-1})^{-1} < K, \quad \forall k = 1, 2, \ldots, K.$$  

**Proof:** See [27]

The following theorem gives a lower bound for $\text{SNR}_{\text{eff}}$.

**Theorem 2:** Consider the complex MIMO channel $\mathbf{y} = \mathbf{Hx} + \mathbf{z}$ with $M$ transmit antennas and $N$ receive antennas, power constraint $\mathbb{E}(\mathbf{x} \mathbf{x}^H) \leq M \cdot \text{SNR}$, and additive noise $\mathbf{z}$ with i.i.d. circularly symmetric complex Gaussian
entries with zero mean and unit variance. The effective signal-to-noise ratio when integer-forcing equalization is applied is lower bounded by

$$\text{SNR}_{\text{eff}} > \frac{1}{4M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} a^\dagger (I + \text{SNR} H^\dagger H) a. \quad (15)$$

**Proof:** Let $H$ be the real-valued representation of the channel $H$, as in (3), and let $L$ and $L^T$ be as in (13). From (14) we have

$$\text{SNR}_{\text{eff}} = \frac{1}{\lambda^2_{2M}(L^T)}.$$ 

The dual lattice of $\Lambda(L^T)$ is $\Lambda(L^{-1})$. Thus, Theorem 1 gives

$$\frac{1}{\lambda^2_{2M}(L^T)} > \frac{1}{(2M)^2} \lambda^2_1(L^{-1})$$

and therefore

$$\text{SNR}_{\text{eff}} > \frac{1}{(2M)^2} \lambda^2_1(L^{-1})$$

$\equiv \frac{1}{M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} \|L^{-1}a\|^2$

$\equiv \frac{1}{M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} a^T(LL^T)^{-1}a$

$\equiv \frac{1}{M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} a^T \left( I + \text{SNR} H^T H \right) a. \quad (16)$

where (16) follows from (13). Since the matrix $(I + \text{SNR} H^T H) \in \mathbb{R}^{2M \times 2M}$ is the real-valued representation of the complex matrix $(I + \text{SNR} H^T H) \in \mathbb{C}^{M \times M}$, (16) can be written in complex form as (15). 

---

**E. Relation between the effective SNR and the minimum distance for uncoded QAM**

A basic communication scheme for the MIMO channel is transmitting independent uncoded QAM symbols from each antenna. In this case, the error probability strongly depends on the minimum distance at the receiver. For a positive integer $L$, we define

$$d_{\text{min}}(H, L) \equiv \min_{a \in \text{QAM}^{2L}} \|Ha\|, \quad (17)$$

where

$$\text{QAM}(L) \equiv \{-L, -L + 1, \ldots, L - 1, L\}$$

and $\text{QAM}(L)$ is an $M$-dimensional vector whose components all belong to $\text{QAM}(L)$. Note that if $L$ is an even integer, $d_{\text{min}}(H, L)$ is the minimum distance at the receiver when each antenna transmits symbols from a QAM($L/2$) constellation. This is true since

$$\min_{x_1, x_2 \in \text{QAM}(L/2)} \|Hx_1 - Hx_2\| = \min_{x \in \text{QAM}(L)} \|Hx\|.$$ 

In the IF scheme there is no assumption that QAM symbols are transmitted. Rather, each antenna transmits codewords taken from a linear codebook. Nevertheless, we show that the performance of the IF receiver over the channel $H$ can be tightly related to those of a hypothetical uncoded QAM system over the same channel. See Figure 1. Namely, $\text{SNR}_{\text{eff}}$ is closely related to $d_{\text{min}}(H, L)$. This relation is formalized in the next key lemma, which is a simple consequence of Theorem 2.

**Lemma 2 (Relation between SNR_{eff} and d_{min}):** Consider the complex MIMO channel $y = Hx + z$ with $M$ transmit antennas and $N$ receive antennas, power constraint $\mathbb{E}(x^H x) \leq M \cdot \text{SNR}$, and additive noise $z$ with i.i.d. circularly symmetric complex Gaussian entries with zero mean and unit variance. The effective signal-to-noise ratio when integer-forcing equalization is applied is lower bounded by

$$\text{SNR}_{\text{eff}} > \frac{1}{4L^2} \min_{L = 1, 2, \ldots} (L^2 + \text{SNR} d_{\text{min}}^2(H, L)), \quad (18)$$

where $d_{\text{min}}^2(H, L)$ is defined in (17).

**Proof:** The bound from Theorem 2 can be written as

$$\text{SNR}_{\text{eff}} > \frac{1}{M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} \|a\|^2 + \text{SNR} \|Ha\|^2. \quad (19)$$

Let

$$\rho(a) \equiv \max_{m = 1, \ldots, M} \max \{a_m \text{real}, a_m \text{imag}\},$$

i.e., $\rho(a)$ is the maximum absolute value of all real and imaginary components of $a$. With this notation, (19) is equivalent to

$$\text{SNR}_{\text{eff}} > \frac{1}{M^2} \min_{L = 1, 2, \ldots} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} \|a\|^2 + \text{SNR} \|Ha\|^2$$

$$\geq \frac{1}{M^2} \min_{L = 1, 2, \ldots} \left( L^2 + \text{SNR} d_{\text{min}}^2(H, L) \right),$$

as desired.

**Remark 2:** In the transmission scheme described above each antenna transmits an independent stream. Therefore, the bounds from Theorem 2 and Lemma 2 continue to hold true for multiple access (MAC) channels with $M$ users equipped with a single transmit antenna and a receiver equipped with $N$ receive antennas, where the gains from the $m$th transmit antenna to the receiver are given by the $m$th column of $H$ and each user is subject to the power constraint $\mathbb{E}(\|x_m\|^2) \leq \text{SNR}$.

**Remark 3:** For real-valued $N \times M$ MIMO channels $y = Hx + z$ with power constraint $\mathbb{E}(x^H x) \leq M \cdot \text{SNR}$, and $z \sim \mathcal{N}(0, I)$ the bound from Theorem 2 becomes

$$\text{SNR}_{\text{eff}} > \frac{1}{M^2} \min_{a \in \mathbb{Z}_M^{+} + i \mathbb{Z}_M^{+}} a^T (I + \text{SNR} H^T H) a,$$

and the bound from Lemma 2 becomes

$$\text{SNR}_{\text{eff}} > \frac{1}{M^2} \min_{L = 1, 2, \ldots} \left( L^2 + \text{SNR} d_{\text{min}}^2(H, L) \right), \quad (20)$$

where

$$d_{\text{min}}(H, L) \equiv \min_{a \in \text{PAM}(L)} \|Ha\|,$$

$$\text{PAM}(L) \equiv \{-L, -L + 1, \ldots, L - 1, L\}.$$
Remark 4: The bound from Lemma 2 exhibits a Diophantine tradeoff, i.e., it depends on how small the norm $\|Ha\|^2$ can be made as a function of the largest component in the integer-valued vector $a$. The typical behavior of this minimal norm is the subject of several results in the metrical theory of Diophantine approximation, see e.g. [28]–[30]. Using these results one can easily prove that for almost all real-valued MIMO channels (w.r.t. Lebesgue measure), IF equalization achieves the optimal number of degrees-of-freedom (DoF), i.e., that

$$\lim_{SNR \to \infty} \frac{R_{IF}(SNR)}{\frac{1}{2} \log(SNR)} = \min(M, N).$$

Standard linear equalizers, such as the zero-forcing equalizer, or the MMSE equalizer, fail to achieve the optimal number of DoF when $N < M$ (in fact, when $N < M$, they achieve zero DoF). Thus, the fact that IF equalization achieves the full DoF is remarkable. For almost every real-valued MIMO-MAC channel with $M$ users equipped with a single transmit antenna and a receiver equipped with $N$ receive antennas, each user can achieve $\min(M, N)/M$ DoF using IF equalization. This extends Theorem 6 of [31], which only covered the case of $N = 1$.

The performance of IF equalization over Rayleigh fading channels was studied in [13] and it was shown that when $N \geq M$ the IF equalizer achieves the optimal receive DMT (corresponding to transmission of independent streams from each antenna). However, in order to approach capacity for all channel realizations, transmitting independent streams from each antenna is not sufficient.

III. PRECODED INTEGER-FORCING

Clearly, there are instances of MIMO channels for which the lower bound $d_{\text{min}}(\bar{H}, L)$ on $\text{SNR}_{\text{eff}}$ does not increase with the WI mutual information. For example, consider a channel $\bar{H}$ where one of the $NM$ entries equals $h$ whereas all other gains are zero. For such a channel $C_{W} = \log(1 + |h|^2\text{SNR})$, yet $\text{SNR}_{\text{eff}} = 1$ (and the bound (15) only gives $\text{SNR}_{\text{eff}} > 1/(4M^2)$). Thus, it is evident that IF equalization alone can perform arbitrarily far from $C_{W}$. This problem can be overcome by transmitting linear combinations of multiple streams from each antenna. More precisely, instead of transmitting $2M$ linearly coded streams, one from the in-phase component and one from the quadrature component of each antenna, over $n$ channel uses, $2nM$ linearly coded streams are precoded by a unitary matrix and transmitted over $nT$ channel uses.

Domanovitz et al. [15] proposed to combine IF equalization with linear precoding. The idea is to transform the $N \times M$ complex MIMO channel (1) into an aggregate $NT \times MT$ complex MIMO channel and then apply IF equalization to the aggregate channel. The transformation is done using a unitary precoding matrix $P \in C^{MT \times MT}$. Specifically, let $x \in C^{MT \times 1}$ be the input vector to the aggregate channel. This vector is multiplied by $P$ to form the vector $\hat{x} = Px \in C^{MT \times 1}$ which is transmitted over the channel (1) during $T$ consecutive channel uses. Let

$$\hat{H} = I_T \otimes H = \begin{bmatrix} H & 0 & \ldots & 0 \\ 0 & H & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & H \end{bmatrix},$$

where $\otimes$ denotes the Kronecker product. The output of the aggregate channel is obtained by stacking $T$ consecutive outputs of the channel (1) one below the other and is given by

$$\hat{y} = \hat{H}P \hat{x} + \bar{z} = \bar{H}x + \bar{z},$$

where $\bar{H} \triangleq \hat{H}P = (I_T \otimes H)P \in C^{NT \times MT}$ is the aggregate channel matrix, and $\bar{z} \in C^{NT \times 1}$ is a vector of i.i.d. circularly symmetric complex Gaussian entries. See Figure 5.

A remaining major challenge is how to choose the precoding matrix $P$ (recall that an open-loop scenario is considered, and hence, the choice of $P$ cannot depend on $\bar{H}$). As observed in Section II-C, the performance of the IF equalizer is dictated by $\text{SNR}_{\text{eff}}$. Thus, in order to obtain achievable rates that are comparable to the WI mutual information, $\text{SNR}_{\text{eff}}$ must scale appropriately with $C_{W}$. The precoding matrix $P$ should therefore be chosen such as to guarantee this property for all channel matrices with the same WI mutual information.

Lemma 2 indicates that for the aggregate channel $\text{SNR}_{\text{eff}}$ increases with $d_{\text{min}}(\bar{H}, L)$, where

$$d_{\text{min}}(\bar{H}, L) = \min_{a \in \text{QAM}^{MT}(L)} \|\bar{H}Pa\|.$$

Thus, the precoding matrix $P$ should be chosen such as to guarantee that $d_{\text{min}}(\bar{H}, L)$ increases appropriately with $C_{W}$. This boils down to the problem of designing precoding matrices for transmitting QAM symbols over an unknown MIMO channel with the aim of maximizing the received minimum distance. This problem was extensively studied during the past decade, under the framework of linear dispersion space-time coding, and unitary precoding matrices that satisfy the aforementioned criterion were found. Therefore, the same matrices that proved so useful for space-time coding are also useful for precoded integer-forcing. A major difference, however, between the two is that while for linear dispersion space-time coding the precoding matrix $P$ is applied to uncoded QAM symbols, in precoded integer-forcing it is applied to coded streams. This in turn, yields an achievable rate characterization which is not available using linear dispersion space-time coding. In particular, very different asymptotics can be analyzed. Rather than fixing the block length and taking $\text{SNR}$ to infinity, as usually done in the space-time coding literature, here, we fix the channel and take the block length to infinity, as in the traditional information-theoretic framework.

In [15] the performance of IF equalization with the golden code’s [12] precoding matrix was numerically evaluated in a $2 \times 2$ MIMO Rayleigh fading environment. The scheme’s outage probability was found to be relatively close to that achieved by white i.i.d. Gaussian codewords. Here, we prove that, in fact, precoded IF equalization, where the precoding
matrix generates a perfect linear dispersion space-time code, achieves rates within a constant gap from the WI mutual information of any MIMO channel.

The aim of the next section is to lower bound \( d_{\text{min}}(\mathbf{H}, L) \) as a function of \( C_{\text{WI}} \) for precoding matrices \( \mathbf{P} \) that generate perfect linear dispersion space-time codes. This lower bound will be instrumental in proving that precoded IF universally attains the MIMO capacity to within a constant gap.

IV. LINEAR DISPERSION SPACE-TIME CODES

Before deriving the lower bound on \( d_{\text{min}}(\mathbf{H}, L) \) some necessary background on space-time codes is given.

An \( M \times T \) space-time (ST) code \( C_{\text{ST}} \) for the channel (1) with rate \( R \) is a set of \( [C_{\text{ST}}] = 2^{RT} \) complex matrices of dimensions \( M \times T \). The codebook \( C_{\text{ST}} \) has to satisfy the average power constraint\(^5\)

\[
\frac{1}{2^{RT}} \sum_{\mathbf{X} \in C_{\text{ST}}} ||\mathbf{X}||_F^2 \leq MT \cdot \text{SNR}.
\]

When the ST code \( C_{\text{ST}} \) is used, a code matrix \( \mathbf{X} \in C_{\text{ST}} \) is transmitted column by column over \( T \) consecutive channel uses, such that the \( T \) channel outputs can be expressed as

\[
\mathbf{Y} = \mathbf{H} \mathbf{X} + \mathbf{Z},
\]

where each column of the matrices \( \mathbf{Y}, \mathbf{Z} \in \mathbb{C}^{N \times T} \) represents the channel output and additive noise, respectively, at one of the \( T \) channel uses.

An ST code \( C_{\text{ST}} \) is said to be a linear dispersion ST code \(^6\) over the constellation \( \mathcal{S} \) if every code matrix \( \mathbf{X} \in C_{\text{ST}} \) can be uniquely decomposed as

\[
\mathbf{X} = \sum_{k=1}^{K} s_k \mathbf{F}_k, \quad s_k \in \mathcal{S},
\]

where \( \mathcal{S} \) is some constellation and the matrices \( \mathbf{F}_k \in \mathbb{C}^{M \times T} \) are fixed and independent of the constellation symbols \( s_k \). Denoting by \( \vec{\mathbf{X}}(\mathbf{X}) \) the vector obtained by stacking the columns of \( \mathbf{X} \) one below the other, and letting \( \mathbf{s} = [s_1 \cdots s_K]^T \) gives

\[
\vec{\mathbf{X}}(\mathbf{X}) = \mathbf{P} \mathbf{s},
\]

where

\[
\mathbf{P} = [\vec{\mathbf{X}}(\mathbf{F}_1), \vec{\mathbf{X}}(\mathbf{F}_2), \cdots, \vec{\mathbf{X}}(\mathbf{F}_K)]
\]

is the code’s \( MT \times K \) generating matrix. A linear dispersion ST code is full-rate if \( K = MT \). In the sequel, linear dispersion ST codes over a QAM\((L) \) constellation, defined in \(^{18}\), will play a key role. The linear dispersion ST code obtained by using the infinite constellation \( \text{QAM}(\infty) = \mathbb{Z} + i\mathbb{Z} \) is referred to as \( C_{\text{ST}}^\infty \), and, after vectorization, is in fact a complex lattice with generating matrix \( \mathbf{P} \). Since the QAM\((L) \) constellation is a subset of \( \mathbb{Z} + i\mathbb{Z} \) it follows that for any finite \( L \) the QAM\((L) \) based code \( C_{\text{ST}}^\infty \) is a subset of \( C_{\text{ST}}^\infty \).

An important class of linear dispersion ST codes is that of perfect codes \(^9\), \(^{10}\) which is defined next.

Definition 3: An \( M \times M \) linear dispersion ST code over a QAM constellation is called perfect if

1) It is full-rate;
2) It satisfies the nonvanishing determinant criterion

\[
d_{\text{min}}(C_{\text{ST}}^\infty) \triangleq \min_{\mathbf{X} \in C_{\text{ST}}^\infty} |\det(\mathbf{X})|^2 > 0;
\]
3) The code's generating matrix is unitary, i.e., \( P^\dagger P = I \).

Note that this definition is slightly different than the one used in [9], [10], where instead of condition [3] it is required that the energy of the codeword corresponding to the information symbols \( s \) will have the same energy as \( \|s\|^2 \), and that all the coded symbols in all \( T \) time slots will have the same average energy.

In [9], perfect linear dispersion ST codes were found for \( M = 2, 3, 4 \) and 6, whereas in [10] perfect linear dispersion ST codes were obtained for any positive integer \( M \). The constructions in [9], [10] are based on cyclic division algebras, and result in a unitary generating matrix \( P \). Thus, for any positive integer \( M \), there exist codes that satisfy the requirements of Definition [3].

Remark 5: Often, perfect linear dispersion space-time codes are defined over a HEX constellation [2], rather than a QAM one. While [9] provides constructions of QAM-based perfect codes only for \( M = 2, 4 \) whereas for \( M = 3, 6 \) the described perfect codes are over HEX constellations, in [10] QAM-based perfect codes are obtained for all dimensions. We remark that for a HEX-based perfect code, the obtained vectorized ST codewords are described as \( P's \), where \( s' \) is a vector of HEX symbols. While the same codewords cannot be described as \( P's \) where \( s \) is a vector of QAM symbols and \( P \) is a (different) generating matrix, their real-valued representation can be described by \( P's \) where \( P' \) is a real-valued matrix (which is not the expansion of a complex-valued matrix) and \( s \) is the real valued representation of a vector with QAM symbols. Although the matrix \( P' \) is not unitary (over the reals), it can be easily shown that \( \|P'x\|^2 \leq (3/2)\|x\|^2 \) for any real-valued vector \( x \) of appropriate dimensions. Thus, a generating matrix of a HEX based perfect code can be manipulated to form a QAM-based code with only a slight increase in the transmission power. The obtained code will not be a linear dispersion code over the complex-filed, but it will be a linear dispersion code over the reals.

The approximate universality of an ST code over the MIMO channel was studied in [7]. This property refers to an ST code being optimal in terms of DMT regardless of the fading statistics of \( H \). A sufficient and necessary condition for an ST code to be approximately universal was derived in [7]. This condition is closely related to the nonvanishing determinant criterion and is satisfied by perfect linear dispersion ST codes. The next theorem is a simple extension of [7] Theorem 3.1. The notation \( |x|^+ = \max(x, 0) \) is used.

**Theorem 3:** Let \( C_{\infty}^{ST} \) be an \( M \times M \) perfect linear dispersion ST code over a QAM\((\infty)\) constellation with \( \delta_{\text{min}}(C_{\infty}^{ST}) = \min_{0 \neq x \in \mathbb{C}^M} |\det(x)|^2 > 0 \), and let \( C^{ST} \) be its subcode over a QAM\((L)\) constellation. Then, for all channel matrices \( H \) with corresponding WI mutual information \( C_W = \log \det(I + SNR H H^\dagger) \) and all \( 0 \neq X \in C^{ST} \)

\[
\text{SNR}||HX||^2_F \geq \left[ \delta_{\text{min}}(C_{\infty}^{ST}) \right]^+ \frac{2 \delta_{\text{min}}(C_{\infty}^{ST})}{4 M^2} - 2M^2 L^2.
\]

**Proof:** The proof closely follows that of [7] Theorem 3.1, and is given in Appendix [B].

Let \( H = I_M \otimes H \), as in (20). The next simple corollary of Theorem [3] will be used in Section [V] to prove the main result of this paper.

**Corollary 1:** Let \( P \in \mathbb{C}^{M^2 \times M^2} \) be a generating matrix of a perfect \( M \times M \) QAM based linear dispersion ST code \( C_{\infty}^{ST} \) with \( \delta_{\text{min}}(C_{\infty}^{ST}) = \min_{0 \neq x \in \mathbb{C}^M} |\det(x)|^2 > 0 \). Then, for all channel matrices \( H \) with corresponding white input mutual information \( C_W = \log \det(I + SNR H H^\dagger) \)

\[
\text{SNR}_{\text{eff}}(H, P, L) \geq \left[ \delta_{\text{min}}(C_{\infty}^{ST}) \right]^+ \frac{2 \delta_{\text{min}}(C_{\infty}^{ST})}{4 M^2} - 2M^2 L^2.
\]

**Proof:** Consider the subcode \( C^{ST} \) of \( C_{\infty}^{ST} \), defined over a QAM\((L)\) constellation. Then, for any \( a \in \text{QAM}^2 \) there exist a code matrix \( X \in \text{QST} \) such that

\[
\text{vec}(X) = Pa.
\]

Now,

\[
\text{SNR}||HPa||^2 = \text{SNR}||H \text{ vec}(X)||^2 = \text{SNR}||HX||^2_F \geq \left[ \delta_{\text{min}}(C_{\infty}^{ST}) \right]^+ \frac{2 \delta_{\text{min}}(C_{\infty}^{ST})}{4 M^2} - 2M^2 L^2
\]

where the last inequality follows from Theorem [3]. It follows that

\[
\text{SNR}_{\text{eff}}^2(H, P, L) = \min_{a \in \text{QAM}^2} \text{SNR}||HPa||^2 \geq \left[ \delta_{\text{min}}(C_{\infty}^{ST}) \right]^+ \frac{2 \delta_{\text{min}}(C_{\infty}^{ST})}{4 M^2} - 2M^2 L^2.
\]

**V. MAIN RESULT**

The next theorem lower bounds the effective signal-to-noise ratio of precoded IF equalization, where the precoding matrix generates a perfect linear dispersion ST code. The obtained bound depends on the channel matrix \( H \) only through its corresponding WI mutual information.

**Theorem 4:** Consider the aggregate MIMO channel

\[
\bar{y} = HPx + z
\]

where \( H = I_M \otimes H \in \mathbb{C}^{NM \times M^2} \), and \( P \in \mathbb{C}^{M^2 \times M^2} \) is a generating matrix of a perfect \( M \times M \) QAM based linear dispersion ST code \( C_{\infty}^{ST} \) with \( \delta_{\text{min}}(C_{\infty}^{ST}) = \min_{0 \neq x \in \mathbb{C}^M} |\det(x)|^2 > 0 \). Then, applying IF equalization to the aggregate channel yields

\[
\text{SNR}_{\text{eff}} > \frac{1}{8 M^2} \delta_{\text{min}}(C_{\infty}^{ST}) \frac{2 \delta_{\text{min}}(C_{\infty}^{ST})}{4 M^2},
\]

for all channel matrices \( H \) with corresponding WI mutual information \( C_W = \log \det(I + SNR H H^\dagger) \).

**Proof:** Applying Lemma [2] to the aggregate \( NM \times M^2 \) channel matrix \( H = HP \) gives

\[
\text{SNR}_{\text{eff}} > \frac{1}{4 M^4} \min_{L=1,2,...} \left( L^2 + \text{SNR}_{\text{eff}}^2(H, L) \right).
\]

(23)
Using Corollary 1 this is bounded by
\[
\text{SNR}_{\text{eff}} > \frac{1}{4M^2} \min_{L=1,2,\ldots} \left( L^2 + \left[ \delta_{\min}(C_{\infty}^{\text{ST}}) \left( \frac{32}{2\text{SNR}^2} + 2M^2L^2 \right) \right]^{1/2} \right)
\]
\[
\geq \frac{1}{4M^2} \min_{L=1,2,\ldots} \left( L^2 + \left[ \delta_{\min}(C_{\infty}^{\text{ST}}) \left( \frac{32}{2\text{SNR}^2} + 2M^2 \right) \right]^{1/2} \right)
\]
\[
\geq \frac{1}{8M^6} \delta_{\min}(C_{\infty}^{\text{ST}}) \left( \frac{32}{2\text{SNR}^2} \right)^{1/2}
\]
as desired.

The next theorem is the main result of this paper.

**Theorem 5:** Let \( P \in \mathbb{C}^{M \times M} \) be a generating matrix of a perfect \( M \times M \) QAM based linear dispersion ST code \( C_{\infty}^{\text{ST}} \) with \( \delta_{\min}(C_{\infty}^{\text{ST}}) = \min_{0 \neq X \in C_{\infty}^{\text{ST}}} |\det(X)|^2 > 0 \). For all channel matrices \( H \) with \( M \) transmit antennas and an arbitrary number of receive antennas, precoded integer-forcing with the precoding matrix \( P \) achieves any rate satisfying
\[
R_{\text{P-IF}} < C_{W_1} - \Gamma \left( \delta_{\min}(C_{\infty}^{\text{ST}}), M \right),
\]
where \( C_{W_1} = \log |\det(I + \text{SNR}H^*H)| \), and
\[
\Gamma \left( \delta_{\min}(C_{\infty}^{\text{ST}}), M \right) \triangleq \log \frac{1}{\delta_{\min}(C_{\infty}^{\text{ST}})} + 3M \log(2M^2) \tag{24}
\]

**Proof:** In precoded IF the matrix \( P \) is used as a precoding matrix that transforms the original \( N \times M \) MIMO channel [1] to the aggregate \( NM \times M^2 \) MIMO channel
\[
\tilde{y} = HPx + z
\]
\[
= Hx + z, \tag{25}
\]
as described in Section III and then IF equalization is applied to the aggregate channel. Assuming a “good” nested lattice codebook is used to encode all \( 2M^2 \) streams transmitted over the aggregate channel, by [10], IF equalization can achieve any rate satisfying
\[
R_{\text{IF-aggregate}} < M^2 \log(\text{SNR}_{\text{eff}}).
\]
Using Theorem 4 it follows that any rate satisfying
\[
R_{\text{IF-aggregate}} < M^2 \log \left( \frac{1}{8M^6} \delta_{\min}(C_{\infty}^{\text{ST}}) \left( \frac{32}{2\text{SNR}^2} \right)^{1/2} \right)
\]
\[
= MC_{W_1} - M \log \frac{1}{\delta_{\min}(C_{\infty}^{\text{ST}})} - M^2 \log(8M^6)
\]
is achievable over the aggregate channel.

Since each channel use of the aggregate channel (25) corresponds to \( M \) channel uses of the original channel (1), the communication rate should be normalized by a factor of \( 1/M \). Thus, \( R_{\text{P-IF}} = R_{\text{IF-aggregate}}/M \), and the theorem follows.

**Remark 6:** The gap between \( C_{W_1} \) and the closed-loop capacity is bounded by a constant number of bits. Therefore, Theorem 5 implies that precoded IF universally achieves the closed-loop capacity of any Gaussian MIMO channel to within a constant gap that depends only on the number of transmit antennas.

**Example 2:** The golden-code [12] is a QAM-based perfect \( 2 \times 2 \) linear dispersion space time code, with \( \delta_{\min}(C_{\infty}^{\text{ST}}) = 1/5 \). Thus, for a MIMO channel with \( M = 2 \) transmit antennas its generating matrix \( P \in \mathbb{C}^{2 \times 4} \) can be used for precoded integer-forcing. Theorem 5 implies that with this choice of \( P \), precoded integer-forcing achieves \( C_{W_1} \) to within a gap of \( (1/5,2) = 20.32 \) bits, which translates to a gap of 5.08 bits per real dimension. In fact, using a slightly more careful analysis it can be shown that, with this choice of \( P \), precoded integer-forcing achieves \( C_{W_1} \) to within 15.24 bits, i.e., 3.81 bits per real dimension.

While these constants may seem quite large, one has to keep in mind that this is a worst-case bound, whereas for the typical case, under common statistical assumptions such as Rayleigh fading, the gap-to-capacity obtained by precoded IF is considerably smaller, as seen in the numerical results of [15].

Note that although the proof of Theorem 5 assumed that a “good” nested lattice code was used, a similar result holds when a \( q \)-ary linear code without shaping is used. This follows from the fact that the performance of the latter is only degraded by the shaping loss of \( \log(2q/12) \) bits per antenna w.r.t. the former. Moreover, Theorem 4 can also be used to obtain an upper bound on the performance of precoded IF with uncoded transmission.

**Proposition 1:** For all channel matrices \( H \) with corresponding WI mutual information \( C_{W_1} = \log |\det(I + \text{SNR}H^*H)| \), \( M \) transmit antennas and an arbitrary number of receive antennas, the error probability of precoded IF with uncoded transmission is bounded by
\[
P_{e,\text{P-IF-uncoded}} \leq 4M^2 \exp \left\{-\frac{3}{2} \left( C_{W_1} - R_{\text{P-IF}} - \Gamma(\delta_{\min}(C_{\infty}^{\text{ST}}), M) \right) \right\}
\]
provided that the preceding matrix \( P \) generates an \( M \times M \) perfect linear dispersion ST code \( C_{\infty}^{\text{ST}} \) with \( \delta_{\min}(C_{\infty}^{\text{ST}}) = \min_{0 \neq X \in C_{\infty}^{\text{ST}}} |\det(X)|^2 > 0 \).

**Proof:** Using (12), the error probability of uncoded IF equalization over the aggregate channel (25) is bounded by
\[
P_{e,\text{P-IF-uncoded}} \leq 4M^2 \exp \left\{-\frac{3}{2} \left( C_{\infty}^{\text{ST}} - 2\text{SNR}^2 - M R_{\text{P-IF}} \right) \right\},
\]
where we have used the fact that the transmission rate over the aggregate channel is \( M \) times larger than the actual communication rate \( R_{\text{P-IF}} \). Now, replacing \( \text{SNR}_{\text{eff}} \) with its bound from Theorem 4 establishes the proposition.

**VI. APPLICATION: RATELESS CODING FOR MIMO CHANNELS VIA PRECODED INTEGER-FORCING**

A notable feature of precoded IF is that the scheme, as well as its performance guarantees, do not depend on the number of antennas at the receiver side. In this section, we exploit this

6Namely, the product of successive minima of a lattice and its dual lattice in Theorem 4 can be bounded using Proposition 3.3 from [13] instead of the result from [27]. The bound from [13] involves Hermite’s constant and gives better results than those obtained using [27] only when very small values of \( M \) are of interest.
A rateless code is defined as a family of codes which has the property that codewords of the higher rates codes are prefixes of those of the lower rate ones. A family of such codes is called perfect (not to be confused with perfect linear dispersion ST codes) if each of the codes in the family is capacity-achieving.

In this section, we show how precoded IF can be used for constructing a rateless code for the MIMO channel which is a constant number of bits from perfect, i.e., each of its subcodes achieve the MIMO capacity to within a constant number of bits. For sake of brevity, we only illustrate the scheme through an example rather than give a full description.

Assume the channel model is the one from (1), and the goal is to design two codes with rates R, and R/2, where the higher rate code is a prefix of the lower rate one. It is further required that for some predefined δ > 0 if the channel’s capacity C satisfies C > R + δ the high-rate (short) code can be decoded reliably, and if C > R/2 + δ the low-rate (long) code can be decoded reliably. This problem can be viewed as that of designing a code which is simultaneously good for the two channel matrices

\[
H_1 = \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad H_2 = \begin{bmatrix} H & 0 \\ 0 & H \end{bmatrix},
\]

since the effective channel \(H_2\) is obtained from twice as many channel uses as \(H_1\), which corresponds to a code twice as long. If \(H \in \mathbb{C}^{N \times M}\), then \(H_1, H_2 \in \mathbb{C}^{2N \times 2M}\). In the previous section, it was shown that precoded IF can simultaneously achieve the capacity of any MIMO channel to within a constant gap. In particular, it can simultaneously achieve the capacity of \(H_1\) and \(H_2\) to within a constant gap.

The rateless code is therefore constructed from \(4M^2\) complex streams of linear codewords (each consisting of one linear codeword in its quadrature component and one in its in-phase components). Each complex stream is of length \(n\) and carries \(nR/2M\) bits. These streams are then precoded using the matrix \(P \in \mathbb{C}^{4M^2 \times 4M^2}\) which generates a perfect \(2M \times 2M\) linear dispersion ST code. This results in a set of \(4M^2\) linear combinations of the coded streams. The linear combinations are then split into \(4M\) groups each containing \(M\) linear combinations, such that the first group consists of the first \(M\) linear combinations, the next group contains the next \(M\) linear combinations, and so on. The short code consists of the odd groups of linear combinations, whereas the long code consists of both odd and even groups of linear combinations. See Figure 6 for an illustration of the code construction.

The long code is transmitted during \(4Mn\) consecutive channel uses. At the receiver side, integer-forcing equalization is applied. The receiver, which knows the channel capacity, can decide whether the first \(2M^2\) linear combinations, corresponding to the first \(2Mn\) channel uses, suffice for correct decoding of the \(4M^2\) coded streams, or all \(4M^2\) linear combinations, corresponding to all \(4Mn\) channel uses, are needed. Theorem 5 implies that if the capacity is greater than \(R + \Gamma (\delta_{min}(C_{ST}), 2M)\) the short code can be decoded reliably, and if it is greater than \(R/2 + \Gamma (\delta_{min}(C_{\infty}), 2M)\) the long code can be decoded reliably.

Note that although we have only described the construction of a code that is compatible with two different rates, the aforementioned construction can be easily extended to any number of rates.

VII. DISCUSSION AND SUMMARY

The additive Gaussian noise MIMO channel in an open-loop scenario, where the receiver has complete channel state information whereas the transmitter has no channel state information was considered in this paper. It was shown that using linear precoding at the transmitter in conjunction with integer-forcing equalization at the receiver suffices to approach the closed-loop capacity of this channel to within a constant gap, depending only on the number of transmit antennas, regardless of the channel matrix \(H\). To the best of our knowledge, this is the first practical scheme that guarantees only an additive
loss w.r.t. capacity. Such a performance guarantee is much stronger than DMT optimality, which is at present the common benchmark for evaluating schemes. In particular, although our results are free from any statistical assumptions, they can be interpreted to obtain performance guarantees in a MIMO fading environment. Specifically, a scheme that achieves a constant gap from capacity is DMT optimal under any fading statistics, and achieves a constant gap from the outage capacity under any fading statistics.

IF equalization uses coded streams, and is therefore usually less suitable for fast fading environments. Nevertheless, we have also developed new upper bounds on an uncoded version of IF equalization, which is more adequate for fast fading. We note that while uncoded IF equalization is quite similar to lattice reduction aided decoding, to the best of our knowledge, the performance of the latter was never analyzed at such a fine scale.

Another appealing feature of the described scheme, inherited from the properties of its underlying perfect ST codes, is that it is independent of the number of receive antennas, and the performance guarantees obtained in this paper do not depend on the number of receive antennas as well. Hence, the scheme is not sensitive to a degrees-of-freedom mismatch.

ACKNOWLEDGMENT

Helpful discussions with Yair Yona, Elad Domanovitz, Barak Weiss and Bobak Nazer are greatly acknowledged.

APPENDIX A

PROOF OF LEMMA 1

The output of the $k$th sub-channel with uncoded transmission is

$$\tilde{y}_k = [v_k + z_{\text{eff},k}] \mod \gamma q Z,$$

where $v_k \in \gamma Z$. The estimate $\hat{v}_k$ is generated by applying a simple slicer (nearest-neighbor quantizer w.r.t. $\gamma Z$) to $\tilde{y}_k$, followed by $\mod \gamma q Z$ reduction. The detection error probability at the $k$th sub-channel is upper bounded by

$$P_{e,k} \triangleq \Pr (\hat{v}_k \neq v_k) \leq \Pr (|z_{\text{eff},k}| \geq \frac{\gamma}{2}).$$

In order to bound $P_{e,k}$, a simple lemma, which is based on [17] Theorem 7] is needed.

Lemma 3: Consider the random variable

$$z_{\text{eff}} = \sum_{\ell=1}^{L} \alpha_{\ell} z_{\ell} + \sum_{k=1}^{K} \beta_k d_k$$

where $\{z_{\ell}\}_{\ell=1}^{L}$ are i.i.d. Gaussian random variables with zero mean and some variance $\sigma_z^2$ and $\{d_{k}\}_{k=1}^{K}$ are i.i.d. random variables, statistically independent of $\{z_{\ell}\}_{\ell=1}^{L}$, uniformly distributed over the interval $[-\rho/2, \rho/2)$ for some $\rho > 0$. Let $\sigma_{\text{eff}}^2 \triangleq \mathbb{E}(z_{\text{eff}}^2)$. Then

$$\Pr(z_{\text{eff}} > \tau) = \Pr(z_{\text{eff}} < -\tau) \leq \exp \left(-\frac{\tau^2}{2\sigma_{\text{eff}}^2}\right).$$

Proof: The probability density function of $z_{\text{eff}}$ is symmetric around zero and hence

$$\Pr(z_{\text{eff}} \geq \tau) = \Pr(z_{\text{eff}} \leq -\tau).$$

Applying Chernoff’s bound gives (for $s > 0$)

$$\Pr(z_{\text{eff}} \geq \tau) \leq e^{-s\tau} \mathbb{E}(e^{s z_{\text{eff}}}) = e^{-s\tau} \mathbb{E}\left(e^{s \sum_{\ell=1}^{L} \alpha_{\ell} z_{\ell} + \sum_{k=1}^{K} \beta_k d_k}\right) = e^{-s\tau} \prod_{\ell=1}^{L} \mathbb{E}(e^{s \alpha_{\ell} z_{\ell}}) \prod_{k=1}^{K} \mathbb{E}(e^{s \beta_k d_k}).$$

Using the well-known expressions for the moment generating function of Gaussian and uniform random variables gives

$$\mathbb{E}(e^{s \alpha_{\ell} z_{\ell}}) = e^{s^2 \sigma_z^2/2},$$

$$\mathbb{E}(e^{s \beta_k d_k}) = \frac{\sinh(s \beta_k / \rho)}{s \beta_k / \rho} \leq e^{s^2 \sigma_{\text{eff}}^2/2},$$

where the last inequality follows from $\sinh(x)/x \leq \exp(x^2/6)$ (which can be obtained by simple Taylor expansion) [17]. It follows that

$$\Pr(z_{\text{eff}} \geq \tau) \leq e^{-s\tau} e^{s^2 \sigma_{\text{eff}}^2/2} (s^2 \sum_{\ell=1}^{L} \alpha_{\ell}^2 \sigma_z^2 + s \sum_{k=1}^{K} \beta_k^2 \sigma_{\text{eff}}^2).$$

Setting $s = \tau / \sigma_{\text{eff}}^2$ gives the desired result.

Now, using Lemma 3, the probability of detection error at the $k$th sub-channel can be bounded as

$$P_{e,k} \leq \Pr \left(|z_{\text{eff},k}| \geq \frac{\gamma}{2}\right) \leq 2 \exp \left(-\frac{\gamma^2}{8\sigma_{\text{eff}}^2}\right) = 2 \exp \left(-\frac{3 \text{SNR}_{\text{eff},k}}{4q^2 \sigma_{\text{eff}}^2}\right) = 2 \exp \left(-\frac{3}{2} \frac{\text{SNR}_{\text{eff},k}}{q^2}\right),$$

where the definition of $\text{SNR}_{\text{eff},k}$ was used in the last equality. Using the fact that $q = 2^R$ and that $\text{SNR}_{\text{eff},k} \geq \text{SNR}_{\text{eff}}$ for all $k = 1, \ldots, 2M$, the detection error probability at each of the $2M$ sub-channels can be further bounded as

$$P_e \leq 2 \exp \left(-\frac{3}{2} \frac{\text{SNR}_{\text{eff}}}{q^2} \left(\frac{1}{2} \log(\text{SNR}_{\text{eff}}) - R}\right)\right).$$

Since the IF equalizer makes an error only if a detection error occurred in at least one of the $2M$ sub-channels, and since the total transmission rate is $R_{\text{IF}} = 2MR$, the total error probability of the IF equalizer with uncoded transmission rate $R_{\text{IF}}$ is bounded by

$$P_{e,\text{IF-uncoded}} \leq 4M \exp \left(-\frac{3}{2} \frac{\text{SNR}_{\text{eff}}}{q^2} \left(\frac{1}{2} \log(\text{SNR}_{\text{eff}}) - R_{\text{IF}}\right)\right) = 4M \exp \left(-\frac{3}{2} \frac{\text{SNR}_{\text{eff}}}{q^2} \left(M \log(\text{SNR}_{\text{eff}}) - R_{\text{IF}}\right)\right).$$
APPENDIX B

PROOF OF THEOREM

Consider some arbitrary \( 0 \neq X \in C^{ST} \) and let
\[
H = U_1 \Psi V_1^\dagger \quad \text{and} \quad X = U_2 \Delta V_2^\dagger
\]
be the singular value decompositions (SVD) of \( H \) and \( X \), respectively. With this notation
\[
\text{SNR}\|HX\|_F^2 = \text{SNR}\|\Psi V_1^\dagger U_2 \Delta \|_F^2.
\]

(S27)

Suppose the (absolute) singular values are ordered by increasing value in \( \Lambda \) and by decreasing value in \( \Psi \):
\[
\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_M\},
\Psi = \text{diag}\{\psi_1, \ldots, \psi_{m_1}, 0, \ldots, 0\},
\]
where \( m_1 \equiv \min\{M, N\} \). In order to establish the desired result one has to find the channel \( H \) with corresponding WI mutual information \( C_{WI} \) that minimizes (S27). The rotation matrix \( V_1 \) that minimizes (S27) is \( V_1 = U_2 \) which aligns the weaker singular values of the channel matrix with the stronger singular values of the code matrix (S24). Thus, the problem of finding the worst channel matrix \( H \) w.r.t. the codeword \( X \) reduces to the optimization problem
\[
\min_{\psi_1, \ldots, \psi_{m_1}} \text{SNR} \sum_{m=1}^{m_1} |\psi_m|^2 |\lambda_m|^2
\]
subject to \( \sum_{m=1}^{m_1} \log(1 + |\psi_m|^2 \text{SNR}) = C_{WI} \). (S28)

A lower bound on the solution of the minimization problem (S28) can be obtained by replacing \( m_1 \) with \( M \geq m_1 \), which increases (or does not change) the optimization space and results in
\[
\min_{\psi_1, \ldots, \psi_M} \text{SNR} \sum_{m=1}^{M} |\psi_m|^2 |\lambda_m|^2
\]
subject to \( \sum_{m=1}^{M} \log(1 + |\psi_m|^2 \text{SNR}) = C_{WI} \). (S29)

The solution to (S29) is given by standard water-filling (S7)
\[
\text{SNR}\|HX\|_F^2 \geq \sum_{m=1}^{M} \left[ \frac{1}{\lambda} - |\lambda_m|^2 \right]^+, \quad (S30)
\]
where \( \lambda \) satisfies
\[
\sum_{m=1}^{M} \left[ \log \left( \frac{1}{\lambda|\lambda_m|^2} \right) \right]^+ = C_{WI}. \quad (S31)
\]

Without loss of generality we may assume that \( 2M^2L^2 \leq \delta_{\min}(C_{\infty}^{ST}) \frac{2C_{\infty}}{2M} \) as otherwise the theorem is trivial. With this assumption, we next show that the \([ \cdot ]^+\) operation in (S31) is not needed, and hence its solution is given by
\[
\frac{1}{\lambda} = |\lambda_1 \cdots \lambda_M|^{\frac{2C_{\infty}}{2M}}. \quad (S32)
\]
Too see this, one has to show that with \( 1/\lambda \) as above the inequality \( 1/\lambda \geq |\lambda_m|^2 \) holds for all \( m = 1, \cdots, M \). First recall that \( X \) is a codeword from a perfect linear dispersion ST code over an QAM\((L)\) constellation. Let \( P \) be the generating matrix of the code \( C^{ST} \). Thus, \( \text{vec}(X) = Ps \) for some vector \( s \) whose \( M^2 \) components all belong to the QAM\((L)\) constellation. This implies that
\[
\sum_{m=1}^{M} |\lambda_m|^2 = \|X\|_F^2
\]
\[
= \|\text{vec}(X)\|^2
\]
\[
= \|Ps\|^2
\]
\[
= \|s\|^2 \quad (S33)
\]
\[
\leq 2M^2L^2, \quad (S34)
\]
where (S33) follows from the fact that \( P \) is unitary. In particular, (S34) implies that
\[
|\lambda_m|^2 \leq 2M^2L^2
\]
for all \( m = 1, \ldots, M \). Since by definition
\[
|\lambda_1 \cdots \lambda_M|^2 = |\text{det}(X)|^2 \geq \delta_{\min}(C_{\infty}^{ST}),
\]
we have for all \( m = 1, \ldots, M \)
\[
|\lambda_m|^2 \leq \delta_{\min}(C_{\infty}^{ST}) \frac{2M^2}{2M}
\]
\[
= \frac{1}{\lambda}. \quad (S35)
\]
Thus, (S32) indeed solves (S31).

Substituting (S32) into (S30) gives
\[
\text{SNR}\|HX\|_F^2 \geq \left[ M |\lambda_1 \cdots \lambda_M|^{\frac{2C_{\infty}}{2M}} - \sum_{m=1}^{M} |\lambda_m|^2 \right]^+
\]
\[
\geq \left[ M \delta_{\min}(C_{\infty}^{ST}) \frac{2C_{\infty}}{2M} - 2M^2L^2 \right]^+
\]
\[
\geq \left[ \delta_{\min}(C_{\infty}^{ST}) \frac{2C_{\infty}}{2M} - 2M^2L^2 \right]^+. \quad (S36)
\]
as desired.

REFERENCES

[1] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” Bell Labs Technical Journal, vol. 1, no. 2, pp. 41–59, Summer 1996.
[2] G. Foschini and M. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” Wireless Personal Commun., March 1998.
[3] E. Telatar, “Capacity of multi-antenna Gaussian channels,” European Transactions on Telecommunications, vol. 10, no. 6, pp. 585–595, November - December 1999.
[4] B. Hochwald and S. ten Brink, “Achieving near-capacity on a multiple-antenna channel,” IEEE Transactions on Communications, vol. 51, no. 3, pp. 389–399, March 2003.
[5] T. Philosof and R. Zamir, “The capacity loss of uncorrelated equi-power Gaussian input over MIMO channel,” in Proceedings of the 23rd IEEE Convention of Electrical and Electronics Engineering in Israel (IEEEI 2004), Eilat, Israel, Sept., pp. 227–229.
[6] L. Zheng and D. Tse, “Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels,” IEEE Transactions on Information Theory, vol. 49, no. 5, pp. 1073–1096, may 2003.
[7] S. Tavildar and P. Viswanath, “Approximately universal codes over slow-fading channels,” IEEE Transactions on Information Theory, vol. 52, no. 7, pp. 3233–3258, July 2006.
