# Nearest Neighbor Decoding in MIMO Block-Fading Channels With Imperfect CSIR

A. Taufiq Asyhari, *Student Member, IEEE*, and Albert Guillén i Fàbregas, *Senior Member, IEEE*

***Abstract*—This paper studies communication outages in multiple-input multiple-output (MIMO) block-fading channels with imperfect channel state information at the receiver (CSIR). Using mismatched decoding error exponents, we prove the achievability of the generalized outage probability, the probability that the generalized mutual information (GMI) is less than the data rate, and show that this probability is the fundamental limit for independent and identically distributed (i.i.d.) codebooks. Then, using nearest neighbor decoding, we study the generalized outage probability in the high signal-to-noise ratio (SNR) regime for random codes with Gaussian and discrete signal constellations. In particular, we study the SNR exponent, which is defined as the high-SNR slope of the error probability curve on a logarithmic-logarithmic scale. We show that the maximum achievable SNR exponent of the imperfect CSIR case is given by the SNR exponent of the perfect CSIR case times the minimum of one and the channel estimation error diversity. Random codes with Gaussian constellations achieve the optimal SNR exponent with finite block length as long as the block length is larger than a threshold. On the other hand, random codes with discrete constellations achieve the optimal SNR exponent with block length growing with the logarithm of the SNR. The results hold for many fading distributions, including Rayleigh, Rician, Nakagami-$m$, Nakagami-$q$ and Weibull as well as for optical wireless scintillation distributions such as lognormal-Rice and gamma-gamma.**

***Index Terms*—Block-fading channels, channel state information, diversity, error exponent, generalized mutual information (GMI), imperfect channel state information, MIMO, multiple antenna, nearest neighbor decoding, outage probability, SNR exponent.**

## I. Introduction

In recent years, the demands for fast wireless communication access have increased dramatically. While a large number of delay-limited applications requiring high data rate have successfully been deployed for wired media, the same development remains a challenge for wireless communications. One of the main challenges is fading, i.e., the fluctuations in the received signal strength due to mobility and multiple path propagation [1].

Understanding the fundamental limits of delay-limited transmission is therefore critical to design efficient codes for such applications. The block-fading channel is a relevant model to study the transmission of delay-limited applications over slowly-varying fading. Within a block-fading period, the channel gain remains constant, varying from block to block according to some underlying distribution. Modern communication systems utilising frequency hopping schemes such as Global System for Mobile Communications (GSM) and multi-carrier modulation such as Orthogonal-Frequency Division Multiplexing (OFDM) are practical examples that are well modelled by the block-fading channel.

Reliable communication over block-fading channels has traditionally been studied under the assumption of perfect channel state information at the receiver (CSIR) [2]–[7]. Since only a finite number of fading realizations are spanned in a single codeword, block-fading channel is not information stable [8]. The input-output mutual information between the transmitted and the received codewords is varying in a random manner and the channel is considered to be *non-ergodic*. For most fading distributions, the Shannon capacity is strictly zero [6]. Based on error exponent considerations, Malkämaki and Leib [5] showed that the outage probability, i.e., the probability that the mutual information is smaller than the target transmission rate [2], [3], is the natural fundamental limit of the channel. References [4]–[6] showed that optimal codes for the block-fading channel should be maximum-distance separable (MDS) on a blockwise basis, i.e., achieving the Singleton bound on the blockwise Hamming distance of the code with equality. Families of blockwise MDS codes based on convolutional codes [5], [6], turbo codes [9], low-density parity check (LDPC) codes [10] and Reed-Solomon codes [4] have also been proposed in the literature. Following the footsteps of [11], references [6], [7] proved that the outage probability as a function of the signal-to-noise ratio (SNR) decays linearly in a log-log scale with a slope given by the so-called outage diversity or SNR exponent. This SNR exponent is precisely given by the Singleton bound and shown to be achieved by random codes constructed over discrete signal constellations, providing the optimal tradeoff between the SNR exponent (diversity) and the target data rate [6]. On the other hand, random codes with Gaussian constellations always achieve the maximum diversity offered by the channel.

For perfect CSIR, nearest neighbor decoding yields minimal error probability [5], [12]. As such, nearest neighbor decoding is commonly used in practical wireless systems. The perfect CSIR

Manuscript received March 22, 2010; revised April 29, 2011; accepted July 14, 2011. This work was supported by the Yousef Jameel Scholarship. The material in this paper was presented in part at the 2009 IEEE Information Theory Workshop, Sicily, Italy, and the 2010 International Symposium on Information Theory and its Applications, Oct. 2010. Date of current version February 29, 2012.

A. T. Asyhari is with the Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: taufiq-a@ieee.org).

A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), The Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. He is also with the Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

Communicated by G. Taricco, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2011.2173719

assumption, however, implies that the channel estimator provides the decoder with a perfectly accurate channel estimate. This assumption is too optimistic and very difficult to guarantee in practice. The time-varying fading characteristics imply that the receiver is likely to be unable to keep track of the channel fluctuations precisely. This makes nearest neighbor decoding inherently suboptimal.

This paper studies nearest neighbor decoding for transmission over multiple-input multiple-output (MIMO) block-fading channels when the receiver is unable to obtain the perfect CSIR. We approach the problem by studying the achievable rates for nearest neighbor decoder with imperfect CSIR. In particular, for fixed channel and channel estimate realizations, we study the achievability of the generalized mutual information (GMI) [13], [14]. The GMI is generally known to be one of the achievable rates when a fixed decoding rule -which is not necessarily matched to the channel- is employed [15], [16]. In our case, it characterizes the maximum communication rate under fixed fading and fading estimate realizations, below which the average error probability is guaranteed to vanish as the codeword length tends to infinity. We further analyze the GMI by studying its important properties. Due to the time-varying fading and its corresponding estimate, we introduce the generalized outage probability, the probability that the GMI is less than the target rate, as an achievable error probability performance for MIMO block-fading channels. By achievability, we mean that random codes are able to achieve this performance but it does not mean that there are no codes that perform better than the generalized outage probability. However, as shown in [16]–[18], independent and identically distributed (i.i.d.) codebooks have a GMI converse, i.e., no rate larger than the GMI can be transmitted with vanishing error probability for i.i.d. codebooks. Thus, for i.i.d. codebooks, the GMI is the largest achievable rate and the generalized outage probability becomes the fundamental limit for block-fading channels with mismatched CSIR.

We are particularly interested in the high-SNR slope of the error probability (plotted in a log-log curve) captured by the following SNR exponent (diversity gain)

$$d \triangleq \lim_{\mathsf{snr}\to\infty} -\frac{\log P_e(\mathsf{snr})}{\log \mathsf{snr}} \tag{1}$$

where $P_e(\mathsf{snr})$ denotes the probability of decoding error as a function of the SNR, $\mathsf{snr}$. We characterize this SNR exponent using the generalized outage probability for the converse and generalized Gallager exponents [13], [14] for the achievability. Our main results show that the generalized outage SNR exponent with imperfect CSIR $d_{\rm icsir}$ is given by the simple formula

$$d_{\rm icsir} = \min(1, d_e) \times d_{\rm csir} \tag{2}$$

where $d_{\rm csir}$ is the perfect-CSIR outage SNR exponent and $d_e$ is the channel estimation error diversity, measuring the decay of the channel estimation error variance as a function of the SNR. Interestingly, this relationship holds for both Gaussian and discrete-input codebooks. We further show that the above SNR exponent can be achieved by Gaussian random codes with finite block length. The minimum block length achieving $d_{\rm icsir}$ is determined by a channel parameter representing the fading distribution, the number of transmit antennas and the number of receive antennas. On the other hand, discrete-input random codes are only able to achieve $d_{\rm icsir}$ with block length growing with the logarithm of the SNR. These results validate the use of perfect-CSIR code designs and provide guidelines for designing reliable channel estimators that achieve the optimal SNR exponents. The theorem also generalizes the results in [6], [11], [19] for MIMO block-fading channels with imperfect CSIR. Note that the perfect CSIR condition is easily recovered by letting $d_e$ to infinity. Finally, the results are applicable to a wide-range of fading distributions including Rayleigh, Rician, Nakagami-$m$, Nakagami-$q$ and Weibull fading models; it can also be directly extended to optical wireless scintillation channels based on log-normal-Rice and gamma-gamma distributions [20]–[22].

There are several theoretical implications that may be extracted from this paper. First, with the SNR exponent as the performance benchmark, we give a meaningful channel estimation design criterion. As long as we can construct channel estimation schemes with decay in channel estimation error variance $\mathsf{snr}^{-1}$ or faster, the perfect CSIR exponent is achievable with mismatched CSIR. Second, this criterion is valid for code design based on i.i.d. Gaussian and discrete inputs over MIMO block-fading channels. Third, the results validate the use of perfect CSIR code design under imperfect CSIR as long as we are able to construct a reliable channel estimator with this criterion.

Throughout the paper, the following notation is used. Scalar, vector and matrix variables are characterized with normal (non-boldfaced), boldfaced lowercase, and boldfaced uppercase letters, respectively. $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix and the symbols $^\dagger$ and $\|\cdot\|_F$ represent the conjugate transpose and the Frobenius norm for a matrix (or equivalently the Euclidean norm for a vector). Random variables are denoted by uppercase letters $X$ and realizations by lowercase letters $x$; random vectors are denoted by lowercase sans serif letters $\mathsf{x}$ and realizations by boldfaced lowercase letters $\boldsymbol{x}$; random matrices are denoted by uppercase sans serif letters $\mathsf{X}$ and realizations by boldfaced uppercase letters $\boldsymbol{X}$. The exponential equality $f(x) \doteq x^d$ indicates that $\lim_{x\to\infty} \frac{\log f(x)}{\log x} = d$ as defined in [11]. The exponential inequalities $\dot{\leq}$ and $\dot{\geq}$ are similarly defined. The symbols $\succeq$, $\preceq$, $\succ$ and $\prec$ describe componentwise inequality $\geq$, $\leq$, $>$ and $<$. Expectation is denoted by $\mathbb{E}[\cdot]$. Sets are denoted by calligraphic fonts with the complement denoted by superscript $^c$. The indicator function is defined by $\mathbb{1}\{\cdot\}$; $\lfloor x \rfloor (\lceil x \rceil)$ denotes the largest (smallest) integer smaller (larger) than or equal to $x$, while $[x]_+ = \max(0, x)$.

The rest of the paper is organized as follows. Section II introduces the channel, imperfect CSIR and fading models. Section III reviews basic material on error exponents for the block-fading channel. Section IV discusses the achievability of the generalized outage probability using mismatched decoding arguments. Section V establishes our main theorem stated in (2), discusses the main findings, shows numerical evidence and provides important remarks. Section VI discusses some important observations on the results and extension to optical wireless scintillation channels. Section VII summarizes the important points of the paper. Proofs of theorems, propositions and lemmas can be found in the appendices.

TABLE I
PDF FOR DIFFERENT FADING DISTRIBUTION

| Fading type | p d f | $w_0$ | $\tau$ | $w_1$ | $w_2$ | $\varphi$ |
|---|---|---|---|---|---|---|
| Rayleigh | $\frac{1}{\pi\Omega}e^{-\frac{\lvert h\rvert^2}{\Omega}}$ | $\frac{1}{\pi\Omega}$ | $0$ | $\frac{1}{\Omega}$ | $0$ | $2$ |
| Rician | $\frac{1}{\pi\Omega}e^{-\frac{\lvert h-\mu\rvert^2}{\Omega}}$ | $\frac{1}{\pi\Omega}$ | $0$ | $\frac{1}{\Omega}$ | $\mu$ | $2$ |
| Nakagami-$m$ | $\frac{m^m\lvert h\rvert^{2m-2}}{\pi\Omega^m\Gamma(m)}e^{-\frac{m\lvert h\rvert^2}{\Omega}}$ | $\frac{m^m}{\pi\Omega^m\Gamma(m)}$ | $2m-2$ | $\frac{m}{\Omega}$ | $0$ | $2$ |
| Weibull | $\frac{\eta\Omega^{-\eta}}{2\pi}\lvert h\rvert^{\eta-2}e^{-\left(\frac{\lvert h\rvert}{\Omega}\right)^\eta}$ | $\frac{\eta\Omega^{-\eta}}{2\pi}$ | $\eta-2$ | $\frac{1}{\Omega^\eta}$ | $0$ | $\eta$ |
| Nakagami-$q$ | $\frac{1+q^2}{2\pi q\Omega}I_0\left(\frac{(1-q^4)\lvert h\rvert^2}{4q^2\Omega}\right)e^{-\frac{(1+q^2)^2}{4q^2\Omega}\lvert h\rvert^2}$ | see footnote[1] | | | | |

[1]Note that we have the bounds: $1 \le I_0(x) = \sum_{i=0}^{\infty}\frac{(x/2)^{2i}}{i!i!} \le \left(\sum_{i=0}^{\infty}\frac{(x/2)^i}{i!}\right)^2 = e^x$, for $x \ge 0$ as in [23], [24].

## II. System Model

Consider transmission over a MIMO block-fading channel with $n_{\rm t}$ transmit antennas, $n_{\rm r}$ receive antennas and $B$ fading blocks corrupted by additive white Gaussian noise (AWGN) and affected by an i.i.d. channel fading matrix, $\boldsymbol{H}_b \in \mathbb{C}^{n_{\rm r}\times n_{\rm t}}$, $b = 1,\ldots,B$. The input-output relationship of the channel is

$$\boldsymbol{Y}_b = \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b\boldsymbol{X}_b + \boldsymbol{Z}_b, \quad b = 1,\ldots,B \tag{3}$$

where $\boldsymbol{Y}_b \in \mathbb{C}^{n_{\rm r}\times L}$, $\boldsymbol{X}_b \in \mathbb{C}^{n_{\rm t}\times L}$, $\boldsymbol{Z}_b \in \mathbb{C}^{n_{\rm r}\times L}$ are the received, transmitted and noise signal matrices corresponding to block $b$; $L$ denotes the channel block length. We assume that the entries of $\boldsymbol{Z}_b$ are i.i.d. complex circularly symmetric Gaussian random variables with zero mean and unit variance.

At the transmitter end, we consider coding schemes of fixed rate $R$ and length $N = BL$, whose codewords are defined as $\boldsymbol{X} \triangleq [\boldsymbol{X}_1,\ldots,\boldsymbol{X}_B] \in \mathcal{X}^{n_{\rm t}\times BL}$, where $\mathcal{X}$ denotes the signal constellation of size $|\mathcal{X}| = 2^M$. By fixed rate, we mean that the coding rate is a positive constant and independent of the SNR; thus the multiplexing gain is zero [11]. The multiplexing gain was introduced in [11] and defined as

$$k \triangleq \lim_{\mathsf{snr}\to\infty}\frac{R(\mathsf{snr})}{\log\mathsf{snr}}. \tag{4}$$

The multiplexing gain characterizes the high-SNR linear gain of the coding rate with respect to the logarithm of the SNR. A nonzero multiplexing gain is only possible with an input constellation that has continuous probability distribution (such as Gaussian input) or an input constellation that has discrete probability distribution but with alphabet size $|\mathcal{X}|$ increasing with the SNR. Since many practical systems employ coding schemes with fixed code rate and finite alphabet size, the assumption of zero multiplexing gain is highly relevant in practice. Furthermore, codewords are assumed to satisfy the average input power constraint $\frac{1}{BL}\mathbb{E}\left[\|\mathsf{X}\|_F^2\right] \le n_{\rm t}$. Herein we assume that no channel state information at the transmitter (CSIT) is available; thus, the uniform power allocation over all fading blocks and transmit antennas is used.

The entries of $\boldsymbol{H}_b$ are i.i.d. random variables that are drawn according to a certain probability distribution. We use the general fading model of [23], [24]. Let $h_{b,r,t}$ be the channel coefficient for fading block $b$, receive antenna $r$ and transmit antenna $t$. Then, the probability density function (pdf) of $H_{b,r,t}$ is given by

$$p_{H_{b,r,t}}(h) = w_0|h|^\tau e^{-w_1|h-w_2|^\varphi} \tag{5}$$

where $w_0 > 0$, $\tau \in \mathbb{R}$, $w_1 > 0$, $w_2 \in \mathbb{C}$ and $\varphi \ge 1$ are constants (finite and SNR independent). This model subsumes a number of widely used fading distributions such as Rayleigh, Rician, Nakagami-$m$ ($m \ge 1$), Nakagami-$q$ ($0 < q \le 1$) and Weibull ($\eta \ge 2$) as tabulated in Table I. For Rayleigh and Rician fading channels, the above pdf represents the pdf of the complex Gaussian random variable with independent real and imaginary parts. For Nakagami-$m$, Weibull and Nakagami-$q$ fading channels, the above pdf is derived assuming that magnitude and phase are independently distributed, and that the phase is uniformly distributed over $[0, 2\pi)$. Furthermore, we assume that the average fading gain is normalized to one, i.e., $\mathbb{E}\left[|H|^2\right] = 1$.

At the receiver side, when perfect CSIR is assumed, nearest neighbor decoding is optimal in minimizing the word error probability. However, practical systems employ channel estimators that yield accurate yet imperfect channel estimates. We model the channel estimate as

$$\hat{\boldsymbol{H}}_b = \boldsymbol{H}_b + \boldsymbol{E}_b, \quad b = 1,\ldots,B \tag{6}$$

where $\hat{\boldsymbol{H}}_b, \boldsymbol{E}_b \in \mathbb{C}^{n_{\rm r}\times n_{\rm t}}$ are the noisy channel estimate and the channel estimation error, respectively. In particular, the entries of $\boldsymbol{E}_b$ are assumed to be independent of the entries of $\boldsymbol{H}_b$ and to have an i.i.d. complex circularly symmetric Gaussian distribution with zero mean and variance equal to

$$\sigma_e^2 = \mathsf{snr}^{-d_e}, \quad \text{with } d_e > 0. \tag{7}$$

Thus, we have assumed a family of channel estimation schemes for which the CSIR noise variance is a decreasing function of the SNR. This model is widely used in pilot-based channel estimation for which the error variance is proportional to the reciprocal of the pilot SNR [25], [26]. We generalize this reciprocal of the pilot SNR with the parameter $d_e$ which denotes the channel estimation error diversity.

A nearest neighbor decoder is used to infer the transmitted message. Due to its optimality under perfect CSIR and its simplicity, this decoder is widely used in practice even when perfect CSIR cannot be guaranteed. With imperfect CSIR, the decoder treats the imperfect channel estimate as if it were perfect. Assuming a memoryless channel, it performs decoding by calculating the following metric:

$$Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\boldsymbol{y}_{b,\ell}|\boldsymbol{x},\hat{\boldsymbol{H}}_b)=\frac{1}{\pi^{n_\mathrm{r}}}e^{-\|\boldsymbol{y}_{b,\ell}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\|_F^2} \tag{8}$$

for each channel use, i.e., for $b=1,\dots,B$, $\ell=1,\dots,L$. The decision is made at the end of $BL$ channel uses for a single codeword.

Recall that $h_{b,r,t}$, $\hat{h}_{b,r,t}$ and $e_{b,r,t}$ are the elements of $\boldsymbol{H}_b$, $\hat{\boldsymbol{H}}_b$ and $\boldsymbol{E}_b$ at row $r, r=1,\dots,n_\mathrm{r}$, and column $t$, $t=1,\dots,n_\mathrm{t}$, respectively. We define $\alpha_{b,r,t}\triangleq-\frac{\log|h_{b,r,t}|^2}{\log\mathsf{snr}}$, $\hat{\alpha}_{b,r,t}\triangleq-\frac{\log|\hat{h}_{b,r,t}|^2}{\log\mathsf{snr}}$ and $\theta_{b,r,t}\triangleq-\frac{\log|e_{b,r,t}|^2}{\log\mathsf{snr}}$. Then, $\boldsymbol{A}_b$, $\hat{\boldsymbol{A}}_b$ and $\boldsymbol{\Theta}_b$ are $n_\mathrm{r}\times n_\mathrm{t}$ matrices whose element at row $r$ and column $t$ is given by $\alpha_{b,r,t}$, $\hat{\alpha}_{b,r,t}$ and $\theta_{b,r,t}$, respectively, for all $r=1,\dots,n_\mathrm{r}$ and $t=1,\dots,n_\mathrm{t}$. We use this change of random variables to analyze the communication performance in the high-SNR regime.

## III. Mutual Information and Outage Probability

It has been shown in [5], [12] that the average error probability for the ensemble of random codes of rate $R$ and a fixed channel realization $\mathsf{H}=\boldsymbol{H}$, input distribution $P_\mathsf{x}(\boldsymbol{x})$ over alphabet $\mathcal{X}$ and perfect CSIR is given by

$$\bar{P}_{e,\mathrm{ave}}(\boldsymbol{H})\le 2^{-NE_r(R,\boldsymbol{H})} \tag{9}$$

where

$$E_r(R,\boldsymbol{H})=\sup_{0\le\rho\le1}\frac{1}{B}\sum_{b=1}^{B}E_0(\rho,\boldsymbol{H}_b)-\rho R \tag{10}$$

is the error exponent for channel realization $\boldsymbol{H}$ and

$$E_0(\rho,\boldsymbol{H}_b)=-\log_2\mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{P_{\mathsf{y}|\mathsf{x},\mathsf{H}}(\mathsf{y}|\mathsf{x}',\mathsf{H}_b)}{P_{\mathsf{y}|\mathsf{x},\mathsf{H}}(\mathsf{y}|\mathsf{x},\mathsf{H}_b)}\right)^{\frac{1}{1+\rho}}\middle|\mathsf{x},\mathsf{y},\mathsf{H}_b\right]\right)^{\rho}\middle|\mathsf{H}_b=\boldsymbol{H}_b\right] \tag{11}$$

is the Gallager function for a given fading realization $\boldsymbol{H}_b$ [12]. Note that the inner expectation is taken over $\mathsf{x}'$, while the outer expectation is taken over $(\mathsf{x},\mathsf{y})$ for a fixed fading realization $\boldsymbol{H}_b$. Remark that the upper bound in (9) corresponds to the average over all codebooks whose entries have been generated i.i.d. (*i.i.d. codebooks*). The ensemble-average in (9) implies that there exists at least one code in the ensemble whose average error probability is bounded as $P_{e,\mathrm{ave}}(\boldsymbol{H})\le 2^{-NE_r(R,\boldsymbol{H})}$ [12]. Then, the average error probability for that code averaged over all fading states is

$$P_{e,\mathrm{ave}}\le\mathbb{E}\left[2^{-NE_r(R,\mathsf{H})}\right]. \tag{12}$$

Basic error exponent results show that $E_r(R,\boldsymbol{H})$ is positive only when $R\le I(\boldsymbol{H})-\varepsilon$, where $I(\boldsymbol{H})$ is the input-output mutual information and $\varepsilon>0$ is a small number, and zero otherwise. The instantaneous mutual information for block-fading channels is easily expressed as

$$I(\boldsymbol{H})=\frac{1}{B}\sum_{b=1}^{B}I^{\mathrm{awgn}}\left(\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b\right) \tag{13}$$

where $I^{\mathrm{awgn}}(\boldsymbol{\Psi})$ is the mutual information of an AWGN MIMO channel for a given channel matrix $\boldsymbol{\Psi}$. In particular, assuming uniform power distribution[1] the mutual information is given by

$$I^{\mathrm{awgn}}\left(\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b\right)=\log_2\det\left(\boldsymbol{I}_{n_\mathrm{r}}+\frac{\mathsf{snr}}{n_\mathrm{t}}\boldsymbol{H}_b\boldsymbol{H}_b^\dagger\right) \tag{14}$$

for Gaussian inputs and

$$I^{\mathrm{awgn}}\left(\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b\right)=Mn_\mathrm{t}-\mathbb{E}\left[\log_2\sum_{\boldsymbol{x}'\in\mathcal{X}^{n_\mathrm{t}}}e^{-\left\|\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b(\mathsf{x}-\boldsymbol{x}')+\mathsf{z}\right\|_F^2+\|\mathsf{z}\|_F^2}\right] \tag{15}$$

for discrete inputs[2], where $\det(\cdot)$ is the matrix determinant operator [27] and $M\triangleq\log_2|\mathcal{X}|$. To the best of our knowledge there is no closed form expression for the expectation term in the mutual information for discrete inputs. However, this expectation can be computed efficiently using Gauss-Hermite quadratures [28] for systems of small size; for larger sizes the above expectation needs to be evaluated using Monte Carlo methods.

Using Arimoto's converse [29] it is possible to show that the mutual information with an optimized[3] input distribution over alphabet $\mathcal{X}$, $I^\star(\boldsymbol{H})$, is the largest rate that can be reliably transmitted. There is no rate larger than $I^\star(\boldsymbol{H})$ having a vanishing error probability. This converse is strong in the sense that for rates larger than $I^\star(\boldsymbol{H})$, the error probability tends to one for sufficiently large block length. This converse is also the "dual" to the Gallager theorem for rates above the mutual information, i.e., for a fixed channel realization $\boldsymbol{H}$ the average error probability of any coding scheme constructed from the alphabet $\mathcal{X}$ is lower-bounded by [29]

$$P_{e,\mathrm{ave}}(\boldsymbol{H})\ge 1-2^{-NE_r'(R,\boldsymbol{H})} \tag{16}$$

where

$$E_r'(R,\boldsymbol{H})=\sup_{-1\le\rho\le0}\inf_{P_\mathsf{x}}\frac{1}{B}\sum_{b=1}^{B}E_0(\rho,\boldsymbol{H}_b)-\rho R \tag{17}$$

[1]This assumption is mainly due to the unavailability of channel state information at the transmitter (CSIT). Hence, the reasonable way to allocate the power across all transmit antennas and all fading blocks is by distributing the power equally. Note that this assumption is optimal at high SNR [11].

[2]This expression assumes equiprobable inputs. The results of the paper remain unchanged for an arbitrary distribution over $\mathcal{X}^{n_\mathrm{t}}$, since for high $\mathsf{snr}$ the equiprobable distribution is optimal.

[3]Herein optimality refers to an input distribution over alphabet $\mathcal{X}$ that maximizes mutual information $I(\boldsymbol{H})$. The maximized mutual information $I^\star(\boldsymbol{H})$ is usually called as capacity for a given input alphabet $\mathcal{X}$ and channel $\boldsymbol{H}$.

is of the same form as (10) except for the range of $\rho$ in the supremum and for the infimum of the probability measure $P_{\mathsf{x}}$ over alphabet $\mathcal{X}$. With channel realization being random, over all fading coefficients, the average error probability becomes

$$P_{e,\text{ave}} \geq 1 - \mathbb{E}\left[2^{-NE'_r(R,\mathsf{H})}\right]. \tag{18}$$

It has been shown in [29] that $E'_r(R,\boldsymbol{H})$ is positive whenever $R > I^\star(\boldsymbol{H})$ and zero otherwise. Let $P^\star_{\mathsf{x}}(\boldsymbol{x})$ be the input distribution over alphabet $\mathcal{X}$ that achieves $I^\star(\boldsymbol{H})$. Suppose that $P^\star_{\mathsf{x}}(\boldsymbol{x})$ is used to evaluate the Gallager error exponent (10). Then, it follows from [5], [12], [29] that $I(\boldsymbol{H})$ is equal to $I^\star(\boldsymbol{H})$, and for sufficiently large $N$, (12) and (18) converge and we obtain that [5]

$$P_{e,\text{ave}} \cong \inf_{\varepsilon>0} \Pr\{I(\mathsf{H}) - \varepsilon < R\} \tag{19}$$

$$= \Pr\{I^\star(\mathsf{H}) < R\} \triangleq P_{\text{out}}(R) \tag{20}$$

which is the *information outage probability* [3]. The above results show the convergence of the random coding achievability and converse to the outage probability as the block length increases to infinity. These results also imply that the outage probability is the natural fundamental limit for block-fading channels.

One has to note that the convergence in (20) holds when the capacity-achieving distribution for a given alphabet $\mathcal{X}$, $P^\star_{\mathsf{x}}(\boldsymbol{x})$ is used to construct codebooks. For a fixed input distribution $P_{\mathsf{x}}(\boldsymbol{x})$, the probability in (19) only characterizes random coding achievability bound to the average error probability; it does not imply anything on the converse bound for any given code. For continuous alphabet, it is well known that i.i.d. Gaussian inputs achieve the capacity for AWGN channels. On the other hand, for discrete inputs with alphabet size $|\mathcal{X}| = 2^M$, the capacity-achieving distribution for AWGN channels depends on the operating SNR. For high SNR, equiprobable distribution over $\mathcal{X}$ is optimal.

*Definition 1 (Outage Diversity):* The outage diversity or the outage SNR exponent, $d_{\text{csir}}$, is defined as the high-SNR slope of the outage probability curve, as a function of the SNR, in log-log scale plot when the receiver has access to perfect CSIR

$$d_{\text{csir}} \triangleq \lim_{\mathsf{snr}\to\infty} -\frac{\log P_{\text{out}}(R)}{\log \mathsf{snr}}. \tag{21}$$

Suppose that $d_{\text{csir}}$ is a finite real-valued quantity, which is the case for many fading distributions. Then, we have the following lemma on the strict lower bound to the error probability at high SNR [11].

*Lemma 1 (Converse Outage Bound):* For any coding scheme with any arbitrary length, the average error probability at high SNR is lower-bounded as

$$P_{e,\text{ave}}(\mathsf{snr}) \,\dot{\geq}\, \mathsf{snr}^{-d_{\text{csir}}}. \tag{22}$$

*Proof:* This lemma has been proven using Fano's inequality in [11]. The claim in [11] works for finite length codes when the multiplexing gain is nonzero. For fixed-rate transmission (zero multiplexing gain), it will no longer work unless we assume that the block length grows with $\log \mathsf{snr}$. Here, we provide an alternative proof to this lemma based on Arimoto's converse. Recall $E'_r(R,\boldsymbol{H})$ is positive if and only if $R > I^\star(\boldsymbol{H})$ and zero otherwise. When the error exponent $E'_r(R,\boldsymbol{H}) = 0$, then $1 - 2^{-NE'_r(R,\boldsymbol{H})} = 0$. Therefore, the error probability can be lower-bounded as

$$P_{e,\text{ave}} \geq \mathbb{E}\left[1 - 2^{-NE'_r(R,\mathsf{H})}\right] \tag{23}$$

$$= \int_{\{I^\star(\boldsymbol{H})<R\}} \left(1 - 2^{-NE'_r(R,\boldsymbol{H})}\right) p_{\mathsf{H}}(\boldsymbol{H}) d\boldsymbol{H} \tag{24}$$

$$= \int_{\{I^\star(\boldsymbol{H})<R\}} p_{\mathsf{H}}(\boldsymbol{H}) d\boldsymbol{H} - \int_{\{I^\star(\boldsymbol{H})<R\}} 2^{-NE'_r(R,\boldsymbol{H})} p_{\mathsf{H}}(\boldsymbol{H}) d\boldsymbol{H} \tag{25}$$

$$= P_{\text{out}}(R) - \int_{\{I^\star(\boldsymbol{H})<R\}} 2^{-NE'_r(R,\boldsymbol{H})} p_{\mathsf{H}}(\boldsymbol{H}) d\boldsymbol{H} \tag{26}$$

$$\doteq P_{\text{out}}(R). \tag{27}$$

Note that for the set $\{I^\star(\boldsymbol{H}) < R\}$, we have that

$$\int_{\{I^\star(\boldsymbol{H})<R\}} 2^{-NE'_r(R,\boldsymbol{H})} p_{\mathsf{H}}(\boldsymbol{H}) d\boldsymbol{H} < P_{\text{out}}(R) \tag{28}$$

with strict inequality due to $2^{-NE'_r(R,\boldsymbol{H})} < 1$. Since the right-hand side (RHS) of Arimoto's converse is always bounded by zero, as the SNR tends to infinity, the integral term of $2^{-NE'_r(R,\boldsymbol{H})}$ decays, as a function of the SNR, at the rate faster than or equal to the decaying rate of $P_{\text{out}}(R)$. Thus, the last exponential equation follows accordingly. ■

Equation (20) shows that the converse is tight with the achievability result suggesting that the optimal SNR exponent for large block length is given by the outage diversity. On the other hand, the lower bound of the error probability in Lemma 1 implies that the outage diversity provides an upper bound for the SNR exponent of any coding scheme constructed from the alphabet $\mathcal{X}$ with a fixed block length $L$. The error probability performance of good codes typically improves as the block length increases. Thus, the outage probability, which characterizes the dominating error event for large code length provides an upper bound to the SNR exponent of the error probability for fixed block-length codes. Furthermore, the proof in Lemma 1 is more general than the proof based on Fano's inequality in [11] in the sense that the proof is well applicable for any zero or nonzero multiplexing gain coding schemes and fixed block length.

We recall existing results on the optimal SNR exponents for both Gaussian and discrete inputs with perfect CSIR.

*Lemma 2:* Consider transmission over a MIMO block-fading channel at fixed rate $R$ with fading model parameters $w_0$, $w_1$, $w_2, \tau$ and $\varphi$ as described in (5) and perfect CSIR using Gaussian constellation and discrete signal constellation of size $2^M$. Then,

$$P_{\text{out}}(R) = \Pr\{I^\star(\mathsf{H}) < R\} \doteq \mathsf{snr}^{-d_{\text{csir}}} \tag{29}$$

where

$$d_{\text{csir}} = \begin{cases} \left(1 + \frac{\tau}{2}\right) B n_{\text{t}} n_{\text{r}} & \text{for Gaussian inputs} \\ \left(1 + \frac{\tau}{2}\right) d_B(R) & \text{for discrete inputs} \end{cases} \tag{30}$$

and

$$d_B(R) = n_{\rm r}\left(1 + \left\lfloor B\left(n_{\rm t} - \frac{R}{M}\right)\right\rfloor\right) \tag{31}$$

is the Singleton bound. The exponent $d_{\rm csir}$ is achieved by random codes for both Gaussian and equiprobable discrete inputs.

*Proof:* For Gaussian inputs, the proof is outlined in [23] by setting zero multiplexing gain. The proof for discrete inputs extending the results of [30] to the general fading model in (5) is outlined in Appendix A. ■

The results in Lemma 2 show the interplay among the system and the channel parameters in determining the optimal SNR exponents. For any positive target rate, Gaussian inputs achieve the maximum diversity. On the other hand, the diversity achieved by the discrete input has a tradeoff with the target rate given by the Singleton bound. Note that

$$\lim_{M\to\infty}\left\lfloor B\left(n_{\rm t} - \frac{R}{M}\right)\right\rfloor = Bn_{\rm t} - 1 \tag{32}$$

which implies that sufficiently large constellations can always achieve maximum diversity. The diversity characterization for discrete inputs provides a benchmark for the error performance of practical codes. A good practical code must have an SNR exponent (1) that achieves the Singleton bound.

## IV. Gallager Exponents, GMI and Generalized Outage

When the nearest neighbor decoder has no perfect knowledge of $\boldsymbol{H}_b$ but rather has access to the noisy estimates $\hat{\boldsymbol{H}}_b$ for all $b = 1,\ldots,B$, then the decoder is mismatched [13]–[16]. This problem is generally encountered for a wide-range of communication systems, where the only way to obtain CSIR is via a channel estimator, inducing accurate yet imperfect channel coefficients. In this situation, the decoder treats the channel estimate $\hat{\boldsymbol{H}}_b$ as if it was the true channel.

Following the same steps outlined in Section III, we can upper-bound the error probability of the ensemble of random codes as [13] (see also [14] for a detailed derivation)

$$\bar{P}_{e,\rm ave}(\hat{\boldsymbol{H}}) \le 2^{-NE_r^Q(R,\hat{\boldsymbol{H}})} \tag{33}$$

where now the mismatched decoding error exponent is

$$E_r^Q(R,\hat{\boldsymbol{H}}) = \sup_{\substack{s>0\\ 0\le\rho\le 1}} \frac{1}{B}\sum_{b=1}^{B} E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) - \rho R \tag{34}$$

and

$$E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) = -\log_2 \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{Q_{\mathsf{y|x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y|x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^s \middle| \mathsf{x},\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\right)^\rho \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right] \tag{35}$$

is the generalized Gallager function for a given fading realization $\boldsymbol{H}_b$ and estimation error $\boldsymbol{E}_b$ [13].

*Remark 1:* The use of the decoder metric $Q_{\mathsf{y|x},\hat{\mathsf{H}}}$ here is not only restricted to the distance metric with noisy channel estimate in (8). The random coding upper bound (33) holds for any positive decoding metric.

*Proposition 1 (Concavity of $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$):* For a fixed input distribution, the generalized Gallager function, $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$, is a concave function of $s$ for $s > 0$ and of $\rho$ for $0 \le \rho \le 1$.

*Proof:* See Appendix B. ■

Since $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ is concave in $\rho$ for $0\le\rho\le 1$, the maximum slope of $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ with respect to $\rho$ occurs at $\rho = 0$. The maximization over $s$ results in a maximum slope equal to the generalized mutual information (GMI) [13], [15], given by

$$I^{\rm gmi}(\hat{\boldsymbol{H}}) = \sup_{s>0} \frac{1}{B}\sum_{b=1}^{B} I_b^{\rm gmi}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) \tag{36}$$

where

$$I_b^{\rm gmi}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) = \mathbb{E}\left[\log_2 \frac{Q^s_{\mathsf{y|x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}{\mathbb{E}\left[Q^s_{\mathsf{y|x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]}\middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right]. \tag{37}$$

Note that if $\hat{\boldsymbol{H}} = \boldsymbol{H}$, then $I^{\rm gmi}(\hat{\boldsymbol{H}})$ is equal to $I(\boldsymbol{H})$. The evaluation of (37) gives

$$I_b^{\rm gmi}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) = \log_2\det\hat{\Sigma}_{\boldsymbol{y}} - \frac{s}{\log 2}\left(n_{\rm r} + \frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b - \boldsymbol{H}_b\|_F^2\right) + \frac{\mathbb{E}\left[s\mathsf{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}\middle|\mathsf{H}_b = \boldsymbol{H}_b,\mathsf{E}_b = \boldsymbol{E}_b\right]}{\log 2} \tag{38}$$

for Gaussian inputs, where

$$\hat{\Sigma}_{\boldsymbol{y}} \triangleq \boldsymbol{I}_{n_{\rm r}} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger\frac{\mathsf{snr}}{n_{\rm t}} \tag{39}$$

and

$$I_b^{\rm gmi}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) = Mn_{\rm t} - \mathbb{E}\left[\log_2\sum_{\boldsymbol{x}'\in\mathcal{X}^{n_{\rm t}}}\left(e^{-s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}(\boldsymbol{H}_b\mathsf{x}-\hat{\boldsymbol{H}}_b\boldsymbol{x}')+\mathsf{z}\right\|_F^2} \cdot e^{s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}(\boldsymbol{H}_b-\hat{\boldsymbol{H}}_b)\mathsf{x}+\mathsf{z}\right\|_F^2}\right)\right] \tag{40}$$

for discrete inputs.

*Proposition 2 (Nonnegativity of the GMI):* For the decoding metric $Q_{\mathsf{y|x},\hat{\mathsf{H}}}$ in (8), $I^{\rm gmi}(\hat{\boldsymbol{H}})$ is always nonnegative, i.e.,

$$I^{\rm gmi}(\hat{\boldsymbol{H}}) \ge 0. \tag{41}$$

*Proof:* See Appendix B. ■

*Proposition 3 (GMI Upper Bound):* The GMI is upper-bounded as

$$I^{\text{gmi}}(\hat{\boldsymbol{H}}) \leq \frac{1}{B}\sum_{b=1}^{B} \sup_{s_b>0} I_b^{\text{gmi}}\left(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s_b\right). \quad (42)$$

Suppose that the supremum on the RHS of (42) is achieved for some $s_b > 0$. Let $s_b^*$ be

$$s_b^* = \arg\sup_{s_b>0} I_b^{\text{gmi}}\left(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s_b\right). \quad (43)$$

Equality in (42) occurs if $s_1^* = s_b^*, b = 1, \ldots, B$.

*Proof:* See Appendix B. ■

The maximum slope analysis on the concave function $E_0^Q(s, \rho, \hat{\boldsymbol{H}}_b)$ (due to Proposition 1) shows that the exponent $E_r^Q(R, \hat{\boldsymbol{H}})$ is only positive whenever $R \leq I^{\text{gmi}}(\hat{\boldsymbol{H}}) - \varepsilon$, and zero otherwise, proving the achievability of $I^{\text{gmi}}(\hat{\boldsymbol{H}})$. Then, following the same argument as the one from [12], there exists at least one code in the ensemble such that the average error probability -averaged over all fading and its corresponding estimate states- is bounded as

$$P_{e,\text{ave}} \leq \mathbb{E}\left[2^{-NE_r^Q(R,\hat{\mathsf{H}})}\right] \quad (44)$$

which, for large $N$ becomes

$$P_{e,\text{ave}} \leq \inf_{\varepsilon>0} \Pr\{I^{\text{gmi}}(\hat{\mathsf{H}}) - \varepsilon < R\} \quad (45)$$

$$= \Pr\{I^{\text{gmi}}(\hat{\mathsf{H}}) < R\} \triangleq P_{\text{gout}}(R) \quad (46)$$

the *generalized outage probability*.

The above analysis shows the achievability of $P_{\text{gout}}(R)$ which indicates that for large $N$ one may find codes whose error probability approaches $P_{\text{gout}}(R)$. Unfortunately, there are no generally tight converse results for mismatched decoding [15] which implies that one might be able to find codes whose error probability for large $N$ might be lower than $P_{\text{gout}}(R)$. However, as shown in [16]–[18], a converse exists for i.i.d. codebooks, i.e., no rate larger than the GMI can be transmitted with vanishing error probability.

*Proposition 4 (Generalized Outage Converse):* For i.i.d. codebooks with sufficiently large block length, we have that

$$P_{\text{out}}(R) \leq P_{\text{gout}}(R) \leq P_{e,\text{ave}}. \quad (47)$$

*Proof:* The inequality of $P_{e,\text{ave}} \geq P_{\text{gout}}(R)$ comes from the GMI converse in [18] for i.i.d. codebooks. Furthermore, due to the data-processing inequality for error exponents $E_r^Q(R, \hat{\boldsymbol{H}}) \leq E_r(R, \boldsymbol{H})$ [13], [14], we obtain that $I^{\text{gmi}}(\hat{\boldsymbol{H}}) \leq I(\boldsymbol{H})$ [15], and hence $P_{\text{gout}}(R) \geq P_{\text{out}}(R)$. ■

From the above proposition, we say that the generalized outage probability is the fundamental limit for i.i.d. codebooks. With perfect CSIR, the nearest neighbor decoder yields the minimal average error probability. Thus, with imperfect CSIR, the average error probability can never be smaller than that with perfect CSIR. This implies that the converse outage bound in Lemma 1 also applies to the nearest neighbor decoder with mismatched CSIR. Then, as the SNR tends to infinity, the decay in average error probability for this mismatched decoder as a function of the SNR can never be larger than $d_{\text{csir}}$.

## V. Generalized Outage Diversity

In this section, we describe the behavior of the generalized outage probability for codebooks that are generated from i.i.d. Gaussian and discrete inputs. In particular, we study the high-SNR regime characterized by the generalized outage diversity, which is defined as follows.

*Definition 2 (Generalized Outage Diversity):* The generalized outage diversity or the generalized outage SNR exponent, $d_{\text{icsir}}$, is defined by the high-SNR slope of the generalized outage probability curve in log-log scale plot when the receiver only knows the noisy CSIR

$$d_{\text{icsir}} \triangleq \lim_{\mathsf{snr}\to\infty} -\frac{\log P_{\text{gout}}(R)}{\log \mathsf{snr}}. \quad (48)$$

This definition can be viewed as the extension of the SNR exponent definition, $d_{\text{csir}}$ for perfect CSIR case, where now we have mismatched CSIR.

From the data-processing inequality summarized in Proposition 4, it is straightforward to show that the mismatched SNR exponent is upper-bounded as $d_{\text{icsir}} \leq d_{\text{csir}}$. However, this relationship does not provide an explicit characterization of the mismatched-CSIR SNR exponent $d_{\text{icsir}}$ as a function of the matched-CSIR SNR exponent $d_{\text{csir}}$. In the following, we provide a precise fundamental relationship of $d_{\text{icsir}}$ and $d_{\text{csir}}$. Furthermore, we also show the way to achieve the optimal SNR exponent using random codes of a given block length $L$, which is of practical interest.

The converse results, which use the large block length results on the generalized outage probability explained in Proposition 4, are summarized in the following theorem.

*Theorem 1 (Converse):* Consider the MIMO block-fading channel, imperfect CSIR and fading model described by (3), (6) and (5), respectively. Then, for high SNR the generalized outage probability using nearest neighbor decoding (8) behaves as

$$P_{\text{gout}}(R) = \Pr\left\{I^{\text{gmi}}(\hat{\mathsf{H}}) < R\right\} \doteq \mathsf{snr}^{-d_{\text{icsir}}} \quad (49)$$

where

$$d_{\text{icsir}} = \min(1, d_e) \times d_{\text{csir}} \quad (50)$$

is the generalized outage SNR exponent or the generalized outage diversity. This relationship holds for code constructions based on both i.i.d. Gaussian and discrete inputs.

*Proof:* We use bounding techniques to prove the result. The lower bound is derived by evaluating the GMI structure for each type of input distribution. On the other hand, the upper bound uses the simple upper bound in Proposition 3. See Appendices C (for discrete inputs) and D (for Gaussian inputs), respectively, for the details. ■

*Remark 2:* Standard proof methodologies to derive the upper bound to the SNR exponent for MIMO channels with perfect

CSIR employ a genie-aided receiver [30]. However, this approach fails for mismatched CSIR. In general, mismatched decoding introduces additional interference during the decoding process and this interference may not be reduced by having that genie-aided receiver.

The converse in Theorem 1 gives a strict upper bound to the decaying rate of the average error probability as a function of the SNR for sufficiently long codes. The results also provide a precise fundamental relationship between the perfect and imperfect CSIR SNR exponents. Suppose that a noisy channel estimator produces a Gaussian random estimation error with variance $\sigma_e^2 = \mathsf{snr}^{-d_e}$. Then, Theorem 1 shows that the imperfect-CSIR SNR exponent is a linear function of the perfect-CSIR SNR exponent with a linear scaling factor of $\min(1, d_e)$.

The intuition on the scaling factor $\min(1, d_e)$ is as follows. Channel estimation errors introduce supplementary outage events, adding to those due to deep fades [6], [11]. Therefore, the generalized outage set contains the perfect-CSIR outage set, and a generalized outage occurs when there is a deep fade, or when the channel estimation error is high.

The above analysis also shows that the phases of the fading and of the channel estimation error play no role in determining the SNR exponents for both Gaussian and discrete signal constellations. However, as shown in the proof (Appendix C), it seems that the phases affect high-SNR outage events for discrete signal constellations; the exact effect depends on the configuration of the specific signal constellation.

For large block length, as shown in (46) and (47), it is sufficient to study $P_{\rm gout}(R)$ to characterize the error probability for i.i.d. codebooks. However, practical wireless communications typically operate with a fixed and finite block length. Herein, we present the results of achievable random coding SNR exponents for a given block length $L$. In this context, we use the generalized Gallager exponents [13], [14] and evaluate the lower-bound to the SNR exponents for any arbitrary length.

To prove the general theorem, we first state the SNR exponent achieved by random codes with Gaussian constellations for any $\tau$ in the fading model (5). We will then provide a tighter block length threshold for $\tau = 0$.

*Theorem 2 (Achievability—Gaussian Inputs):* Consider the MIMO block-fading channel (3) with fading distribution in (5) and data rate growing with the logarithm of the SNR at multiplexing gain $k \geq 0$ as defined in (4). In the presence of mismatched CSIR (6), there exists a Gaussian random code whose average error probability is upper-bounded as

$$P_{e,\rm ave}(\mathsf{snr}) \dot{\leq} \mathsf{snr}^{-d_G^\ell(k)} \quad (51)$$

where

$$\begin{aligned} d_G^\ell(k) = \inf_{\mathcal{A}_G^c \cap \{\boldsymbol{A} \succeq 0, \boldsymbol{\Theta} \succeq d_e \times \mathbf{1}\}} \Bigg\{ & \left(1 + \frac{\tau}{2}\right) \sum_{b=1}^{B} \sum_{r=1}^{n_{\rm r}} \sum_{t=1}^{n_{\rm t}} \alpha_{b,r,t} \\ & + \sum_{b=1}^{B} \sum_{r=1}^{n_{\rm r}} \sum_{t=1}^{n_{\rm t}} (\theta_{b,r,t} - d_e) \\ & + L \left( \sum_{b=1}^{B} [\min(1, \theta_{\min}) - \alpha_{b,\min}]_+ - Bk \right) \Bigg\} \end{aligned} \quad (52)$$

and $\mathcal{A}_G^c$ is the complement of the set

$$\mathcal{A}_G \triangleq \Bigg\{ \boldsymbol{A}, \boldsymbol{\Theta} \in \mathbb{R}^{B \times n_{\rm r} \times n_{\rm t}} : \sum_{b=1}^{B} [\min(1, \theta_{\min}) - \alpha_{b,\min}]_+ \leq Bk \Bigg\} \quad (53)$$

where

$$\theta_{\min} \triangleq \min \{\theta_{1,1,1}, \ldots, \theta_{b,r,t}, \ldots, \theta_{B,n_{\rm r},n_{\rm t}}\} \quad (54)$$

$$\alpha_{b,\min} \triangleq \min \{\alpha_{b,1,1}, \ldots, \alpha_{b,r,t}, \ldots, \alpha_{b,n_r,n_t}\}. \quad (55)$$

The function $d_G^\ell(k)$ can be computed explicitly for any block length $L \geq 1$.

*Proof:* See Appendix E. ■

*Corollary 1:* Following Theorem 2, the SNR exponent lower bound for Gaussian input with zero multiplexing gain is given by

$$d_G^\ell(k=0) = \begin{cases} d_{\rm icsir} & \text{for } L \geq \left\lceil \left(1 + \frac{\tau}{2}\right) n_{\rm t} n_{\rm r} \right\rceil \\ BL \min(1, d_e) & \text{otherwise.} \end{cases} \quad (56)$$

*Proof:* The proof is obtained by solving the infimum of the variables in Theorem 2 for a given length $L$. The optimizer of $\boldsymbol{\Theta}$ to $d_G^\ell(k)$ is given by $\boldsymbol{\Theta}^* = d_e \times \mathbf{1}$ because an increase in $\theta_{\min}$ increases $\alpha_{b,\min}$ in the constraint set and by definition $\theta_{b,r,t} \geq \theta_{\min}$. To find the $\alpha_{b,r,t}^*$ that gives the infimum of $d_G^\ell(k=0)$, let $\alpha_{b,r,t} \in [0, \min(1, d_e)]$ and $\alpha_{b,\min}^*$ be the infimum value that is tight in the constraint. Since $\alpha_{b,r,t} \geq \alpha_{b,\min}$, the infimum solution for the rest of $\alpha_{b,r,t}$ is given by $\alpha_{b,r,t}^* = \alpha_{b,\min}^*$. Using these, it is straightforward to prove (56). ■

For $\tau = 0$, we have the following proposition that provides a tighter achievability bound than Theorem 2.

*Proposition 5:* Consider the MIMO block-fading channel (3) with fading model (5) for $\tau = 0$, imperfect CSIR (6), data rate growing with the logarithm of the SNR at multiplexing gain $k \geq 0$ as defined in (4) and $n^* = \min(n_{\rm t}, n_{\rm r})$. Let $\hat{\boldsymbol{\lambda}}_b$ be a row vector consisting of the nonzero eigenvalues of $\hat{\boldsymbol{H}}_b \hat{\boldsymbol{H}}_b^\dagger$, with $0 < \hat{\lambda}_{b,1} \leq \cdots \leq \hat{\lambda}_{b,n*}$, and $\hat{\boldsymbol{\lambda}} \triangleq \{\hat{\boldsymbol{\lambda}}_1, \ldots, \hat{\boldsymbol{\lambda}}_b, \ldots, \hat{\boldsymbol{\lambda}}_B\}$. Define the variable $\hat{\beta}_{b,i}$ as $\hat{\beta}_{b,i} \triangleq -\frac{\log \hat{\lambda}_{b,i}}{\log \mathsf{snr}}$. There exists a Gaussian random code whose average error probability is upper-bounded as

$$P_{e,\rm ave}(\mathsf{snr}) \dot{\leq} \mathsf{snr}^{-d_G^\ell(k)} \quad (57)$$

where

$$\begin{aligned} d_G^\ell(k) = \inf_{\bar{\mathcal{A}}_G^c \cap \{\hat{\boldsymbol{\beta}} \succeq \mathbf{0}, \boldsymbol{\Theta} \succeq d_e \times \mathbf{1}\}} \Bigg\{ & \sum_{b=1}^{B} \sum_{i=1}^{n^*} (2i - 1 + |n_{\rm t} - n_{\rm r}|) \hat{\beta}_{b,i} \\ & + \sum_{b=1}^{B} \sum_{r=1}^{n_{\rm r}} \sum_{t=1}^{n_{\rm t}} (\theta_{b,r,t} - d_e) \\ & + L \left( \sum_{b=1}^{B} \sum_{i=1}^{n^*} [\min(1, \theta_{\min}) - \hat{\beta}_{b,i}]_+ - Bk \right) \Bigg\} \end{aligned} \quad (58)$$

and $\bar{\mathcal{A}}_G^c$ is the complement of the set

$$\bar{\mathcal{A}}_G \triangleq \left\{ \hat{\boldsymbol{\beta}} \in \mathbb{R}^{B\times n^*}, \boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm t}\times n_{\rm r}} : \sum_{b=1}^{B}\sum_{i=1}^{n^*}\Big[\min(1,\theta_{\min}) - \hat{\beta}_{b,i}\Big]_+ \le Bk \right\} \quad (59)$$

where

$$\theta_{\min} \triangleq \min\left\{\theta_{1,1,1},\ldots,\theta_{b,r,t},\ldots,\theta_{B,n_{\rm r},n_{\rm t}}\right\}. \quad (60)$$

The function $d_G^\ell(k)$ can be computed explicitly for any given block length $L \ge 1$.

*Proof:* See Appendix G. ■

The optimizer of $\boldsymbol{\Theta}$ in (58) is given by $\boldsymbol{\Theta}^* = d_e \times \mathbf{1}$ because an increase in $\theta_{\min}$ increases $\hat{\beta}_{b,i}$ in the objective function and the constraint set. To find $\hat{\beta}_{b,i}^*$ that gives the infimum of $d_G^\ell(k=0)$ in (58), let $\hat{\beta}_{b,i} \in [0, \min(1,d_e)]$. Following [11], for $L \ge n_{\rm t}+n_{\rm r}-1$, the infimum always occurs with $\sum_b^B\sum_i^{n^*}\left(\min(1,d_e)-\hat{\beta}_{b,i}^*\right) = Bk$. For $k=0$ (fixed coding rate), the length of $L \ge n_{\rm t}+n_{\rm r}-1$ leads to

$$d_G^\ell(k=0) = \min(1,d_e)Bn_{\rm t}n_{\rm r}. \quad (61)$$

Using similar steps as in [11], if $L < n_{\rm t}+n_{\rm r}-1$, then $d_G^\ell(k=0)$ is not tight to the $d_{\rm icsir}$ and the random coding achievable SNR exponents are given by solving the infimum of $d_G^\ell(k=0)$ in (58), which is strictly smaller than $d_{\rm icsir}$.

Therefore, for random codes with Gaussian constellations, the lower bound to the SNR exponent is $d_{\rm icsir}$ as long as the block length satisfies the following constraint

$$L \ge \begin{cases} n_{\rm t}+n_{\rm r}-1 & \text{for } \tau = 0 \\ \left\lceil\left(1+\frac{\tau}{2}\right)n_{\rm t}n_{\rm r}\right\rceil & \text{otherwise.}\end{cases} \quad (62)$$

The achievability of the random codes constructed over discrete alphabets of size $|\mathcal{X}| = 2^M$ for a given block length $L$ is summarized as follows.

*Theorem 3 (Achievability—Discrete Inputs):* Let the block length $L$ grow as $\lim_{\mathsf{snr}\to\infty}\frac{L(\mathsf{snr})}{\log\mathsf{snr}} = \omega$, $\omega \ge 0$. Then, there exists a random code constructed over discrete-input alphabet of size $|\mathcal{X}| = 2^M$ such that the average error probability with mismatched CSIR (6) is upper-bounded by

$$P_{e,{\rm ave}}(\mathsf{snr}) \dot{\le} \mathsf{snr}^{-d_{\mathcal{X}}^\ell(R)} \quad (63)$$

where $d_{\mathcal{X}}^\ell(R)$ is given in (64), as shown at the bottom of the page.

*Proof:* See Appendix H. ■

The assumption of the block length $L$ to grow with the SNR as $L(\mathsf{snr}) = \omega\log\mathsf{snr}$, $\omega \ge 0$ is to obtain a more precise characterization of what can be achieved using random codes with a discrete constellation of size $2^M$. The results show the interplay among the growth rate $\omega$, the cardinality of the constellation $2^M$ and the achievable SNR exponent with random coding.

*Remark 3:* By letting $\omega \to \infty$, we have that

$$\begin{aligned} d_{\mathcal{X}}^\ell(R) &= \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}\left\lceil B\left(n_{\rm t}-\frac{R}{M}\right)\right\rceil & (65)\\ &\le \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}\left(1+\left\lfloor B\left(n_{\rm t}-\frac{R}{M}\right)\right\rfloor\right) & (66)\\ &= d_{\rm icsir}. & (67)\end{aligned}$$

This implies that $d_{\mathcal{X}}^\ell(R)$ is equal to $d_{\rm icsir}$ only at the continuous points of the Singleton bound. Hence, random codes based on discrete constellations are only able to achieve $d_{\rm icsir}$ at the continuous points only for all possible $R$, with the block length growing as $\log\mathsf{snr}$ and the growth rate is very large. Herein we have also shown analytically that the block length is required to grow with $\log\mathsf{snr}$ at a certain growth rate $\omega > 0$ to obtain nonzero random coding SNR exponent. To achieve the optimal SNR exponent at the continuous points of the Singleton bound, the growth rate, $\omega$, should approach infinity, and the channel estimator should provide reliable estimates ($d_e \ge 1$). This is illustrated in Fig. 1.

Fig. 1 also explains the case of a finite-valued $\omega$. If $\omega$ is fixed and satisfies $\omega M\log 2 < \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}$, then the lower bound $d_{\mathcal{X}}^\ell(R)$ can never achieve the scaled Singleton bound for any positive rate $R$. On the other hand, if $\omega$ is fixed and satisfies $\omega M\log 2 \ge \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}$, then the lower bound $d_{\mathcal{X}}^\ell(R)$ can achieve the scaled Singleton bound for some values of $R$ (as indicated by the dashed line in Fig. 1); larger $\omega$ implies a larger range of values of $R$ for which $d_{\mathcal{X}}^\ell(R)$ is equal

$$d_{\mathcal{X}}^\ell(R) = \begin{cases} \omega BM\log 2\left(n_{\rm t}-\frac{R}{M}\right), & 0 \le \omega < \frac{\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}}{M\log 2} \\ \omega M\log 2\left(1+\left\lfloor\frac{BR}{M}\right\rfloor-\frac{BR}{M}\right) + \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}\left(\left\lceil B\left(n_{\rm t}-\frac{R}{M}\right)\right\rceil-1\right), & \frac{\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}}{M\log 2} \le \omega < \frac{1}{M\log 2}\times\left(\frac{\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}}{1+\left\lfloor\frac{BR}{M}\right\rfloor-\frac{BR}{M}}\right) \\ \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}\left\lceil B\left(n_{\rm t}-\frac{R}{M}\right)\right\rceil, & \omega \ge \frac{1}{M\log 2}\times\left(\frac{\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_{\rm r}}{1+\left\lfloor\frac{BR}{M}\right\rfloor-\frac{BR}{M}}\right)\end{cases} \quad (64)$$

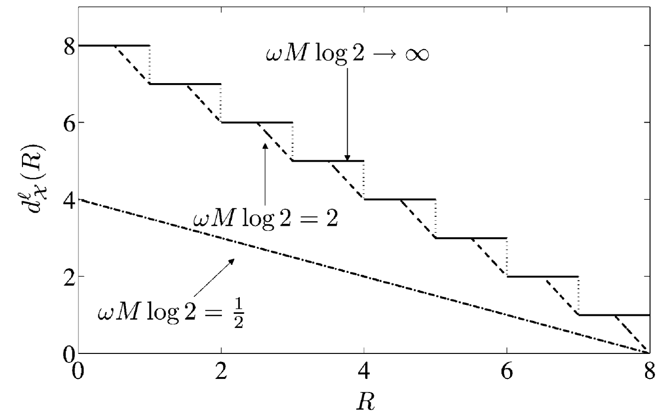


Fig. 1. Random coding SNR exponent lower bound for discrete signal codebooks as a function of target rate $R$ (in bits per channel use), $B = 4$, $n_{\rm t} = 2$, $n_{\rm r} = 2$, $\tau = 0$, $M = 4$ and $d_e = 0.5$.

to $d_{\rm icsir}$. As $\omega$ tends to infinity, we achieve $d_{\rm icsir}$ for all $R$ except at the discontinuity points of the Singleton bound (as shown by the solid line in Fig. 1).

## VI. Discussion

This section discusses some important insights obtained from our main results.

### A. Important Remarks

The following observations can be obtained from Theorems 1, 2, 3 and Proposition 5.

1) The optimal SNR exponent for any coding scheme can be obtained when $d_e \geq 1$. From Lemma 1, Proposition 4 and Theorem 1, the converse on the SNR exponent is strong since $P_{\rm gout}(R)$ has the same exponential decay in SNR as $P_{\rm out}(R)$. We need both a good channel estimation and a good code design to achieve the optimal SNR exponent.
2) The results emphasise the role of the channel estimation error SNR exponent $d_e$ for determining the generalized outage exponent. Even with noisy CSIR, we are still able to achieve the perfect CSIR SNR exponent provided that $d_e \geq 1$. If $d_e < 1$, the resulting SNR exponent scales linearly with $d_e$ and approaching zero for $d_e \downarrow 0$. Fig. 2 illustrates this effect in a discrete-input block-fading channel with $B = 4$, $\tau = 0$, $n_{\rm t} = 2$ and $n_{\rm r} = 2$.
3) The term $\min(1, d_e)$ appears naturally from the data-processing inequality. It highlights the importance of having channel estimators that can achieve error diversity $d_e \geq 1$. With $d_e \geq 1$, the error level is likely to be much less than the reciprocal of the SNR level as the SNR tends to infinity. In a block-fading set-up, this result provides a more precise characterization on the accuracy of the channel estimation at high SNR than [31].
4) The role of the channel estimation error diversity $d_e$ is governed by the channel estimation model. With a maximum-likelihood (ML) estimator, it can be shown that $d_e$ is proportional to the pilot power [25]. Larger pilot power implies larger $d_e$. Hence, the price for obtaining high outage diversity is in the pilot power which does not contain any information data. The bounding condition $\min(1, d_e)$ implies that the perfect-CSIR outage diversity can be achieved with $d_e = 1$. Note that although having larger $d_e$ for $d_e > 1$ shows no diversity improvement, it still leads to a better outage performance. As $d_e$ tends to infinity, the outage performance converges to that with perfect CSIR.
5) The outage diversity in Theorem 1 is valid for the general fading model described by (5). This fading model is used extensively in analysing the performance of radio-frequency (RF) wireless communications.
6) For a given $d_e \geq 1$, Gaussian random codes with finite block length can achieve the full diversity of $Bn_{\rm t}n_{\rm r}$ as long as the block length is larger than a threshold. On the other hand, discrete-alphabet random codes with finite block length cannot achieve the diversity given by the Singleton bound $d_B(R)$. In order for these random codes to achieve the Singleton bound, the block length needs to grow as $\omega \log \mathsf{snr}$ [6].

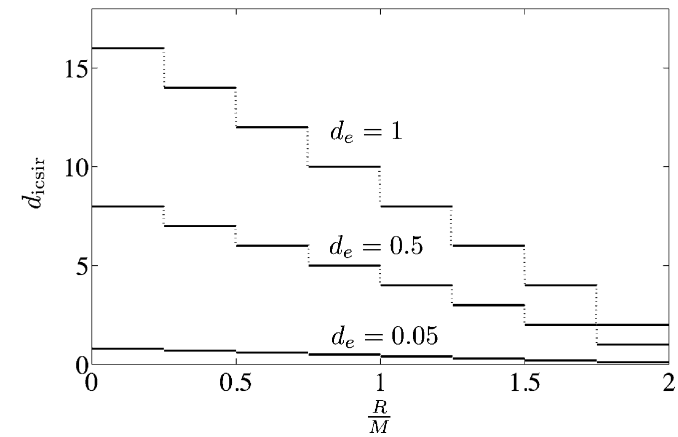


Fig. 2. Generalized outage SNR exponent for discrete-input block-fading channel, $B = 4$, $\tau = 0$ (Rayleigh, Rician and Nakagami-$q$ fading), $n_{\rm t} = 2$ and $n_{\rm r} = 2$.

Figs. 3 and 4 illustrate the generalized outage probability $P_{\rm gout}(R)$ in (46) for Gaussian and binary phase-shift keying (BPSK) inputs, respectively, over a MIMO Rayleigh block-fading channel with $n_{\rm t} = 2$ and $n_{\rm r} = 1$. The following parameters are specified: $B = 2$ and $R = 2$ bits/channel use for Gaussian input and $B = 2$, and $R = 1$ bits/channel use for BPSK input. The curves were generated as follows. Monte Carlo simulation was used to compute the number of outage events. Firstly, each entry of $\boldsymbol{H}_b$ and $\boldsymbol{E}_b$, $b = 1, \ldots, B$ were independently generated from zero-mean complex Gaussian distributions with variance one and $\sigma_e^2 = \mathsf{snr}^{-d_e}$, respectively. The values of $d_e = 0.5$, 1, and 2 were used for comparison with the perfect CSIR outage probability. Secondly, for a fixed $s > 0$, channel $\boldsymbol{H}_b$ and channel estimate $\hat{\boldsymbol{H}}_b = \boldsymbol{H}_b + \boldsymbol{E}_b$, $I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s)$, $b = 1, \ldots, B$ were computed for Gaussian inputs (38) and BPSK input (40). Note that for Gaussian inputs, the generalized outage diversity may not be derived directly from (38) particularly due to the term $\mathbb{E}\left[ s\mathsf{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y} \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b \right]$. However, this term can be evaluated numerically using the singular-value decomposition [27] as in Appendix D. On the other hand, for BPSK input, we compute the expectation in (40) using the Gauss-Hermite quadratures [28]. Thirdly, for a fixed $\boldsymbol{H}_b$ and $\hat{\boldsymbol{H}}_b$, the supremum

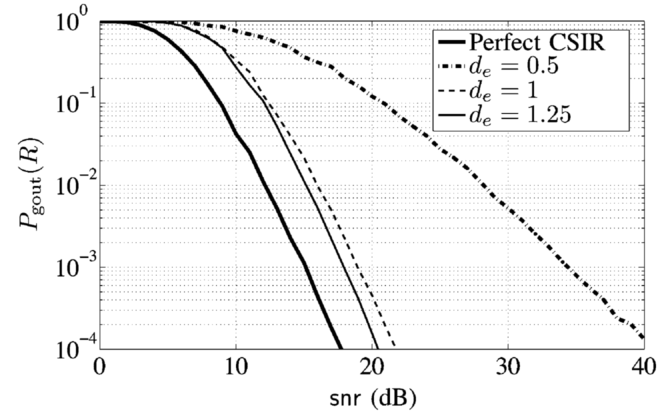


Fig. 3. Generalized outage probability for Gaussian-input MIMO Rayleigh block-fading channel with $B=2$, $R=2$, $n_{\rm t}=2$ and $n_{\rm r}=1$.

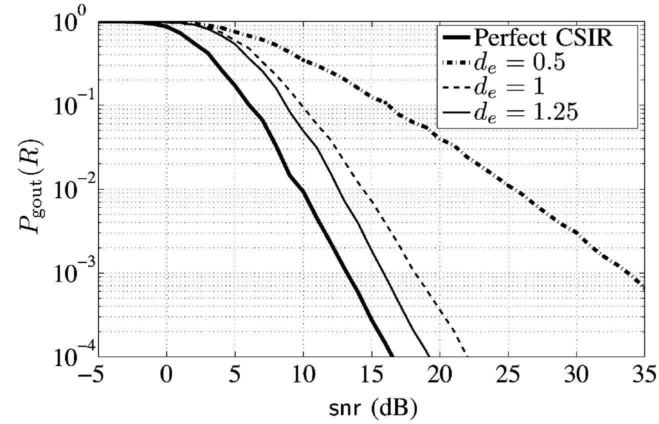


Fig. 4. Generalized outage probability for BPSK-input MIMO Rayleigh block-fading channel with $B=2$, $R=1$, $n_{\rm t}=2$ and $n_{\rm r}=1$.

over $s$ in the RHS of (36) was solved using standard convex optimization algorithm since the function

$$\frac{1}{B}\sum_{b=1}^{B} I_b^{\rm gmi}({\sf snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s) \tag{68}$$

is concave[4] in $s$ for $s>0$. Finally, an outage event was declared whenever $I^{\rm gmi}(\hat{\boldsymbol{H}})$ was less than $R$. $P_{\rm gout}(R)$ was given by the ratio of the number of outage events and the number of total transmissions. From the figures, we observe that $d_{\rm csir}$ is equal to 4 and 3 for Gaussian and BPSK inputs, respectively. As predicted by Theorem 1, the slope becomes steeper with increasing $d_e$, eventually becoming parallel to the perfect CSIR outage curve for $d_e \geq 1$. For $d_e > 1$, the slope does not increase as $d_e$ increases. However, we still observe the improvement in outage gain; the curves for $d_e = 1.25$ achieve the same $P_{\rm gout}(R)$ as the one for $d_e = 1$ at a lower SNR.

### *B. Remarks on the Diversity-Multiplexing Tradeoff (DMT) for the Gaussian Inputs*

In this paper, we focus on fixed-rate transmission such that the coding schemes have zero multiplexing gain [11]. The analysis in this paper can be extended to the nonzero multiplexing gain $(k>0)$ case. In this case, the data rate can be expressed as a function of the SNR, $R({\sf snr}) = k\log{\sf snr}$. A nonzero multiplexing gain is relevant for continuous inputs such as Gaussian inputs or discrete inputs with alphabet size increasing with the SNR. In the following, we provide some remarks on extending the results to the nonzero multiplexing gain for the Gaussian inputs.

The analysis of nonzero multiplexing gain is implicitly covered in Theorem 2 and Proposition 5. Both results provide a lower-bound on the optimal diversity-multiplexing tradeoff (DMT) with infinite block length. Note that from Theorem 2, one may obtain nonzero $d_G^\ell(k)$ for general fading parameter $\tau$ and for $L\to\infty$ as

$$d_G^\ell(k) = \left(1+\frac{\tau}{2}\right) B n_{\rm t} n_{\rm r}\left(\min(1,d_e)-k\right), \quad \text{for } 0\leq k\leq \min(1,d_e). \tag{69}$$

Note that this lower bound can be loose since the maximum multiplexing gain for a positive diversity is $\min(1,d_e)$ as compared to $\min(n_{\rm t},n_{\rm r})$ for the case of perfect CSIR [11]. From Proposition 5, one may obtain nonzero $d_G^\ell(k)$ for fading parameter $\tau=0$ and for $L\to\infty$ as the tradeoff with the multiplexing gain. Indeed, as shown in Appendix G, the lower-bound to the optimal DMT curve $d_G^\ell(k)$ for $L\to\infty$ is given by the piecewise-linear function connecting the points $(k, d_G^\ell(k))$, where

$$k = 0, \min(1,d_e), 2\min(1,d_e), \ldots, \min(n_{\rm t},n_{\rm r})\min(1,d_e) \tag{70}$$

$$d_G^\ell(k) = \min(1,d_e)\cdot B\left(n_{\rm t} - \frac{k}{\min(1,d_e)}\right)\cdot\left(n_{\rm r} - \frac{k}{\min(1,d_e)}\right). \tag{71}$$

Note that we have $d_{G,\max}^\ell = \min(1,d_e)Bn_{\rm t}n_{\rm r}$ and $k_{\max} = \min(n_{\rm t},n_{\rm r})\min(1,d_e)$. This lower bound is tight for $B=1$ and $d_e\geq 1$, which is the perfect-CSIR DMT [11].

There are several reasons why the above bounds may not be tight for mismatched CSIR. The first one is that we only evaluate the above bounds based on Gallager's lower bound to the error exponent (34) which for large $L$ yields an upper bound to the generalized outage probability as shown in Appendix E and Appendix G. Hence, these bounds are not an exact characterization of the generalized outage probability. The second one is that for the bound in Theorem 2, the tradeoff is derived by using the joint pdf of the entries of $\boldsymbol{H}_b$ and $\boldsymbol{E}_b$. Note that this leads to a further lower bound as shown in Appendix G. A tighter bound is obtained by considering the analysis using the joint pdf of the eigenvalues. However, this last approach has some technical difficulties particularly for $\tau\neq 0$ as also shown in Appendix G.

Using the upper bound for the GMI (Appendix D), an upper bound for the optimal DMT can be derived. This yields

$$d_{\rm icsir}^G(k) \leq \left(1+\frac{\tau}{2}\right) B n_{\rm t} n_{\rm r}\min\left(1-\frac{k}{\min(n_{\rm t},n_{\rm r})}, d_e\right), \quad \text{for } 0\leq k\leq\min(n_{\rm t},n_{\rm r}). \tag{72}$$

Note the upper bound is trivial for any $d_e \leq 1-\frac{k}{\min(n_{\rm t},n_{\rm r})}$ because this is identical to the result of zero multiplexing gain; the diversity with zero multiplexing gain is an upper bound to

[4]The concavity of $\frac{1}{B}\sum_{b=1}^{B} I_b^{\rm gmi}({\sf snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s)$ in $s$ is a consequence of $I_b^{\rm gmi}({\sf snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s)$ being concave in $s$ for $s>0$. The concavity of $I_b^{\rm gmi}({\sf snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s)$ can be shown using the same technique used to prove the concavity of $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ (Appendix B).

the optimal tradeoff. The upper bound (72) and the lower bound (69) are tight only for zero multiplexing gain.

### C. Extension to Optical Wireless Scintillation Distributions

In optical wireless scintillation channels, we mainly deal with the received signal intensities (or instantaneous power signals), and not complex input symbols or complex fading realizations; thus the use of real amplitude modulation such as pulse-position modulation (PPM) is common [20]–[22], [32]. This means that the channel phase is not being considered in the detection (noncoherent detection) and therefore only the real-part of the complex Gaussian noise affects the decision. However, the mutual information and the GMI expressions in (13) and (36) are valid for real-valued signals and real-valued fading responses as well (see also [20], [32] for the compact expression of single-input single-output (SISO) mutual information with PPM inputs). Notice that in our converse and achievability results, we have used the joint probability of $\mathsf{A}$ and $\Theta$. The parameter that distinguishes the resulting SNR exponents for different fading conditions is the channel parameter $\tau$ in the form of $1+\frac{\tau}{2}$. This form comes out naturally from the pdf after defining $\alpha_{b,r,t} = -\frac{\log |h_{b,r,t}|^2}{\log \mathsf{snr}}$. Thus, as long as after performing the change of random variables, we can express the pdf of the normalized fading gain for each channel matrix entry as

$$p_{A_{b,r,t}}(\alpha) \doteq \exp\left(-\left(1+\frac{\tau}{2}\right)\alpha \log \mathsf{snr}\right), \quad \text{for } \alpha \geq 0 \tag{73}$$

then our main results are valid for those fading distributions as well. Consequently, the results are valid for fading distributions used in optical wireless scintillation channels such as lognormal-Rice distribution for which $\left(1+\frac{\tau}{2}\right)=\frac{1}{2}$ and gamma-gamma distribution for which $\left(1+\frac{\tau}{2}\right)=\frac{1}{2}\min(a,b)$, where $a$, $b$ are the parameters of the individual gamma distributions [20]–[22].

## VII. Conclusion

We have examined the outage behavior of nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR. In particular, we have proved the achievability of the generalized outage probability using error exponents for mismatched decoding. Due to the data-processing inequality for error exponents and mismatched decoders, the generalized outage probability is larger than the outage probability of the perfect CSIR case. Using the GMI converse, we have shown the generalized outage probability as the fundamental limit for i.i.d. codebooks. We have further analyzed the generalized outage probability in the high-SNR regime for nearest neighbor decoding and we have derived the SNR exponents for both Gaussian and discrete inputs. We have shown that in both cases, the SNR exponent is given by the perfect CSIR SNR exponent scaled by the minimum of the channel estimation error diversity and one. Therefore, in order to achieve the highest possible SNR exponent, the channel estimator scheme should be designed in such a way so as to make the estimation error diversity equal to or larger than one. Furthermore, the optimal SNR exponent for Gaussian inputs can be achieved using Gaussian random codes with finite block length as long as the block length is greater than a threshold; this threshold depends on the fading distribution and the number of antennas. On the other hand, for discrete-constellation random codes, these SNR exponents are achievable using block length that grows very fast with $\log \mathsf{snr}$. The results obtained are well applicable for a general fading model subsuming Rayleigh, Rician, Nakagami-$m$, Nakagami-$q$ and Weibull distributions, as well as optical wireless channels with lognormal-Rice and gamma-gamma scintillation.

## Appendix A
## Proof of Discrete Input Perfect CSIR SNR Exponent

Recall the fading model in (5). We can bound the pdf as [5]

$$w_0|h|^\tau e^{-w_1(|h|+|w_2|)^\varphi} \leq w_0|h|^\tau e^{-w_1|h-w_2|^\varphi} \tag{74}$$
$$\leq w_0|h|^\tau e^{-w_1|(|h|-|w_2|)|^\varphi}. \tag{75}$$

Let $h=|h|e^{\imath\phi^h}$ and $\alpha = -\frac{\log|h|^2}{\log \mathsf{snr}}$ where $\imath=\sqrt{-1}$. The lower bound for the joint pdf of $A_{b,r,t}$ and $\Phi^H_{b,r,t}$ is given as

$$p_{A_{b,r,t},\Phi^H_{b,r,t}}(\alpha,\phi) \geq \frac{w_0}{2}\log\mathsf{snr}\cdot\mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\alpha}\cdot e^{-w_1\left(\mathsf{snr}^{-\frac{\alpha}{2}}+|w_2|\right)^\varphi}. \tag{76}$$

Remark that for $\alpha<0$, we can see from the exponential term that the above lower bound decays exponentially with the SNR to zero; for $\alpha \geq 0$, the exponential term converges to a constant as $\mathsf{snr}$ tends to infinity. As for the joint pdf upper bound, everything remains unchanged except for the exponential term. We can write the upper bound for the exponential term as follows:

$$e^{-w_1|(|h|-|w_2|)|^\varphi} = e^{-w_1\left|\mathsf{snr}^{-\frac{\alpha}{2}}-|w_2|\right|^\varphi}. \tag{77}$$

If $\alpha<0$, then the above term decays exponentially to zero as $\mathsf{snr}$ tends to infinity. On the other hand, if $\alpha\geq 0$, then the above term converges to a constant as $\mathsf{snr}$ tends to infinity. We therefore have that both upper and lower bounds behave similarly for high $\mathsf{snr}$. Let $\mathcal{O}_\mathcal{X}$ be the asymptotic outage set for the discrete constellation $\mathcal{X}$. This set has been characterized in [30] for perfect CSIR. We then have the outage probability for $n_\mathrm{r}\times n_\mathrm{t}$ MIMO channels with $B$ fading blocks

$$P^\mathcal{X}_{\text{out}}(R) \doteq \mathsf{snr}^{-d^\mathcal{X}_{\text{csir}}} \tag{78}$$
$$\doteq \int_{\mathcal{O}_\mathcal{X}\cap\{\boldsymbol{A}\succeq\boldsymbol{0}\}} \mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\sum_{b=1}^B\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\alpha_{b,r,t}} d\boldsymbol{A}d\boldsymbol{\Phi^H}. \tag{79}$$

Applying Varadhan's lemma [33] yields the following result:

$$d^\mathcal{X}_{\text{csir}} = \inf_{\mathcal{O}_\mathcal{X}\cap\{\boldsymbol{A}\succeq\boldsymbol{0}\}}\left(1+\frac{\tau}{2}\right)\sum_{b=1}^B\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\alpha_{b,r,t} \tag{80}$$

which is exactly the Rayleigh fading result [30] multiplied by $\left(1+\frac{\tau}{2}\right)$.

[5]Note that for any real nonnegative numbers $c$ and $d$ satisfying $c\geq d$, it follows that $c^\varphi \geq d^\varphi$ for any $\varphi>0$. Then, we have the triangle inequality $|h-w_2|^\varphi \leq (|h|+|w_2|)^\varphi$. Applying the reverse triangle inequality we also have $|h-w_2|\geq \left||h|-|w_2|\right| \geq 0$.

## APPENDIX B
## SECTION IV PROOFS

*1) Concavity of $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$:* Fix the input distribution $P_{\mathsf{x}}(\boldsymbol{x})$. We first define $0<g_0<g_1$, $0<\Upsilon<1$, $\rho=\Upsilon g_0+(1-\Upsilon)g_1$,

$$f(s,\boldsymbol{x},\boldsymbol{y}) \triangleq \mathbb{E}\left[\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^s\middle|\mathsf{x}=\boldsymbol{x},\mathsf{y}=\boldsymbol{y},\mathsf{H}_b,\mathsf{E}_b\right] \tag{81}$$

and

$$E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) = -\log\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^\rho\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]. \tag{82}$$

Then, we have that

$$\begin{aligned}&\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^\rho\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\&=\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^{\Upsilon g_0}f(s,\mathsf{x},\mathsf{y})^{(1-\Upsilon)g_1}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right].\end{aligned} \tag{83}$$

Using Hölder's inequality [34], we have that

$$\begin{aligned}&\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^{\Upsilon g_0}f(s,\mathsf{x},\mathsf{y})^{(1-\Upsilon)g_1}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\&\le\left(\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^{g_0}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\right)^{\Upsilon}\\&\quad\times\left(\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^{g_1}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\right)^{(1-\Upsilon)}.\end{aligned} \tag{84}$$

Taking the logarithm, which is a monotonously increasing function, on both sides yields

$$-E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)\le-\Upsilon E_0^Q(s,g_0,\hat{\boldsymbol{H}}_b)-(1-\Upsilon)E_0^Q(s,g_1,\hat{\boldsymbol{H}}_b) \tag{85}$$

which shows the concavity of the function $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ in $\rho$ for $\rho\ge0$.

Now, let $s=\Upsilon g_0+(1-\Upsilon)g_1$. Then

$$\begin{aligned}&f(s,\boldsymbol{x},\boldsymbol{y})\\&=\mathbb{E}\left[\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^{\Upsilon g_0+(1-\Upsilon)g_1}\middle|\mathsf{x}=\boldsymbol{x},\mathsf{y}=\boldsymbol{y},\mathsf{H}_b,\mathsf{E}_b\right] \qquad (86)\\&\le\left(\mathbb{E}\left[\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^{g_0}\middle|\mathsf{x}=\boldsymbol{x},\mathsf{y}=\boldsymbol{y},\mathsf{H}_b,\mathsf{E}_b\right]\right)^{\Upsilon}\\&\quad\times\left(\mathbb{E}\left[\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^{g_1}\middle|\mathsf{x}=\boldsymbol{x},\mathsf{y}=\boldsymbol{y},\mathsf{H}_b,\mathsf{E}_b\right]\right)^{1-\Upsilon} \qquad (87)\\&=f(g_0,\boldsymbol{x},\boldsymbol{y})^{\Upsilon}\times f(g_1,\boldsymbol{x},\boldsymbol{y})^{1-\Upsilon} \qquad (88)\end{aligned}$$

where the inequality is due to Hölder's inequality [34]. Evaluating the generalized Gallager function (82), we have that

$$\begin{aligned}&\mathbb{E}\left[f(s,\mathsf{x},\mathsf{y})^\rho\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\&\le\mathbb{E}\left[f(g_0,\mathsf{x},\mathsf{y})^{\rho\Upsilon}\times f(g_1,\mathsf{x},\mathsf{y})^{\rho(1-\Upsilon)}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right] \qquad (89)\\&\le\left(\mathbb{E}\left[f(g_0,\mathsf{x},\mathsf{y})^\rho\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\right)^{\Upsilon}\\&\quad\times\left(\mathbb{E}\left[f(g_1,\mathsf{x},\mathsf{y})^\rho\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\right)^{(1-\Upsilon)}. \qquad (90)\end{aligned}$$

Taking logarithm on both sides gives us

$$-E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)\le-\Upsilon E_0^Q(g_0,\rho,\hat{\boldsymbol{H}}_b)-(1-\Upsilon)E_0^Q(g_1,\rho,\hat{\boldsymbol{H}}_b) \tag{91}$$

which proves the concavity of $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ in $s$ for $s\ge0$.

*2) Nonnegativity of $I^{\mathrm{gmi}}(\hat{\boldsymbol{H}})$:* Let $s^*$ be the optimizing value of $s$ on the RHS of (36). Then, by substituting a specific value of $s$ to the RHS of (36), i.e., $s\downarrow0$, we have that

$$\begin{aligned}I^{\mathrm{gmi}}(\hat{\boldsymbol{H}})&=\sup_{s>0}\frac{1}{B}\sum_{b=1}^{B}I_b^{\mathrm{gmi}}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) \qquad (92)\\&=\frac{1}{B}\sum_{b=1}^{B}I_b^{\mathrm{gmi}}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s^*) \qquad (93)\\&\ge\lim_{s\downarrow0}\frac{1}{B}\sum_{b=1}^{B}I_b^{\mathrm{gmi}}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s) \qquad (94)\end{aligned}$$

due to the fact that $s^*$ always maximizes the RHS of (36). Let $s=\frac{1}{s'}$. Note that from (37) and (94)

$$\begin{aligned}&\lim_{s\downarrow0}I_b^{\mathrm{gmi}}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s)\\&=\lim_{s'\uparrow\infty}\Bigg\{\frac{1}{s'}\mathbb{E}\left[\log_2 Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\&\quad-\mathbb{E}\left[\log_2\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\Bigg\} \qquad (95)\\&=\lim_{s'\uparrow\infty}-\mathbb{E}\left[\log_2\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]. \qquad (96)\end{aligned}$$

Since the function $Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)$ in (8) can be bounded as $0<Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)<1$ for $s=\frac{1}{s'}>0$, we have that

$$\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]<1 \tag{97}$$

and

$$-\log_2\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]>0. \tag{98}$$

It follows from (96) that

$$\begin{aligned}&\lim_{s'\uparrow\infty}-\mathbb{E}\left[\log_2\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\&\ge\mathbb{E}\left[\lim_{s'\uparrow\infty}-\log_2\mathbb{E}\left[Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right] \qquad (99)\\&=\mathbb{E}\left[-\log_2\mathbb{E}\left[\lim_{s'\uparrow\infty}Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}^{\frac{1}{s'}}(\mathsf{y}|\mathsf{x}',\hat{\mathsf{H}}_b)\middle|\mathsf{y},\mathsf{H}_b,\mathsf{E}_b\right]\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]=0 \qquad (100)\end{aligned}$$

where the inequality (99) is obtained from applying Fatou's lemma [37], and the equality (100) is obtained from applying the dominated convergence theorem [35]. This proves Proposition 2. This property serves as a weak lower bound for $I^{\mathrm{gmi}}(\hat{\boldsymbol{H}})$.

*3) GMI Upper Bound:* We have that

$$\sup_{s>0}\frac{1}{B}\sum_{b=1}^{B} I_b^{\rm gmi}\left(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s\right) \le \frac{1}{B}\sum_{b=1}^{B}\sup_{s_b>0} I_b^{\rm gmi}\left(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s_b\right). \quad (101)$$

The left-hand side (LHS) supremum over $s$ is taken over all $B$ blocks. Thus, the optimizing $s$ does not necessarily maximize the value of $I_b^{\rm gmi}(\mathsf{snr},\boldsymbol{H}_b,\hat{\boldsymbol{H}}_b,s)$ for each block $b$. In this simple upper bound, however, the supremum is taken for each individual block giving a larger quantity in total. In general, $I^{\rm gmi}(\hat{\boldsymbol{H}})$ is equal to the upper bound when $s_b^*=s^*$ for all $b=1,\ldots,B$, where $s_b^*$ is the optimizing $s_b$ for the RHS of (101) and $s^*$ is the optimizing $s$ for the LHS of (101).

## Appendix C
## Proof of Discrete Input Mismatched CSIR SNR Exponent

We first state the following lemma. This lemma is general for both i.i.d. Gaussian and discrete inputs.

*Lemma 3:* Consider the MIMO block-fading channel (3) with mismatched CSIR (6) for the general fading model in (5) and the high-SNR generalized outage set denoted by $\mathcal{O}$ expressed in terms of the normalized fading gain matrix $\boldsymbol{A}$, fading phase matrix $\boldsymbol{\Phi^H}$, normalized error power matrix $\boldsymbol{\Theta}$ and error phase matrix $\boldsymbol{\Phi^E}$. Then, the generalized outage probability satisfies

$$P_{\rm gout}(R) \doteq \mathsf{snr}^{-d_{\rm icsir}} \quad (102)$$

$$\doteq \int_{\mathcal{O}} p_{\mathsf{A},\Phi^{\mathsf{H}}}(\boldsymbol{A},\boldsymbol{\Phi^H})p_{\Theta}(\boldsymbol{\Theta})p_{\Phi^{\mathsf{E}}}(\boldsymbol{\Phi^E})d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi^H}d\boldsymbol{\Phi^E} \quad (103)$$

$$\doteq \int_{\mathcal{O}} p_{\mathsf{A},\Phi^{\mathsf{H}}}(\boldsymbol{A},\boldsymbol{\Phi^H})p_{\Theta}(\boldsymbol{\Theta})d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi^H}d\boldsymbol{\Phi^E}. \quad (104)$$

For the fading model in (5), the generalized outage diversity is given by the solution of the following infimum:

$$d_{\rm icsir}=\inf_{\mathcal{O}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}\alpha_{b,r,t} + \sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}(\theta_{b,r,t}-d_e)\right\}. \quad (105)$$

*Proof:* The joint probability of $\mathsf{A}$, $\Theta$, $\Phi^{\mathsf{H}}$ and $\Phi^{\mathsf{E}}$ over the outage region can be written as in (103) because the random matrices $\Theta$ and $\Phi^{\mathsf{E}}$ are independent. Note that since each entry of estimation error phase matrix is uniformly distributed over $[0,2\pi)$, the density $p_{\Phi^{\mathsf{E}}}(\boldsymbol{\Phi^E})$ does not affect the exponential equality. Then, this lemma is obtained by evaluating the integral (103) over $\mathcal{O}$ and applying Varadhan's lemma [33]. The condition $\boldsymbol{A}\succeq\boldsymbol{0}$ is the same as that for perfect CSIR in Appendix A. On the other hand, the condition $\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}$ is derived as follows. Consider the entry of $\boldsymbol{\Theta}$ at block $b$, receive antenna $r$ and transmit antenna $t$. The pdf of $\theta$ is given by

$$p_{\Theta_{b,r,t}}(\theta)=\log\mathsf{snr}\cdot\mathsf{snr}^{d_e-\theta}\cdot\exp\left(-\mathsf{snr}^{d_e-\theta}\right). \quad (106)$$

From the above pdf, we can see that the interval of $\theta$ for which the pdf does not decay exponentially with the SNR to zero is given by $\theta\ge d_e$. The result follows by considering all entries of $\boldsymbol{\Theta}$. ■

We start the proof of the discrete-input SNR exponent with the proof for single-input single-output (SISO) channels. The proof is based on both upper and lower bounds on the GMI. The MIMO proof follows as a simple extension of the SISO proof.

*1) SISO Case:*

*GMI Lower Bound:* For the SISO channel, (40) becomes

$$I_b^{\rm gmi}(\mathsf{snr},h_b,\hat{h}_b,s) = M-\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\left[\log_2\sum_{x'\in\mathcal{X}}\left(e^{-s|\sqrt{\mathsf{snr}}(h_bx-\hat{h}_bx')+Z|^2}\cdot e^{s|\sqrt{\mathsf{snr}}(h_b-\hat{h}_b)x+Z|^2}\right)\right] \quad (107)$$

and the GMI is given by

$$I^{\rm gmi}(\hat{\boldsymbol{h}})=\sup_{s>0}\frac{1}{B}\sum_{b=1}^{B}I_b^{\rm gmi}(\mathsf{snr},h_b,\hat{h}_b,s). \quad (108)$$

For a given $s>0$, the sum of $I_b^{\rm gmi}(\mathsf{snr},h_b,\hat{h}_b,s)$ over all $b=1,\ldots,B$ yields

$$BM-\sum_{b=1}^{B}\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\left[\log_2\sum_{x'\in\mathcal{X}}\left(e^{-s|\sqrt{\mathsf{snr}}(h_bx-\hat{h}_bx')+Z|^2}\cdot e^{s|\sqrt{\mathsf{snr}}(h_b-\hat{h}_b)x+Z|^2}\right)\right]$$

$$=BM-\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\left[\sum_{b=1}^{B}\log_2\sum_{x'\in\mathcal{X}}\left(e^{-s|\sqrt{\mathsf{snr}}(h_bx-\hat{h}_bx')+Z|^2}\cdot e^{s|\sqrt{\mathsf{snr}}(h_b-\hat{h}_b)x+Z|^2}\right)\right]. \quad (109)$$

For any noise realization $z\in\mathbb{C}$ with $|z|<\infty$, we can bound the term inside the expectation as

$$0\le\sum_{b=1}^{B}\log_2\sum_{x'\in\mathcal{X}}\left(e^{-s|\sqrt{\mathsf{snr}}(h_bx-\hat{h}_bx')+z|^2}\cdot e^{s|\sqrt{\mathsf{snr}}(h_b-\hat{h}_b)x+z|^2}\right) \quad (110)$$

$$\le\sum_{b=1}^{B}\log_2\left(|\mathcal{X}|e^{s|\sqrt{\mathsf{snr}}(h_b-\hat{h}_b)x+z|^2}\right) \quad (111)$$

$$=\sum_{b=1}^{B}\frac{\log|\mathcal{X}|+s|z-\sqrt{\mathsf{snr}}e_bx|^2}{\log 2}. \quad (112)$$

We have the expectation over $Z$

$$\mathbb{E}\left[\sum_{b=1}^{B}\frac{\log|\mathcal{X}|+s|Z-\sqrt{\mathsf{snr}}e_bx|^2}{\log 2}\right] = \frac{B\log|\mathcal{X}|+s(B+\mathsf{snr}\sum_{b=1}^{B}|e_b|^2|x|^2)}{\log 2}. \quad (113)$$

Note that the discrete constellation size $|\mathcal{X}|$ and the energy $|x|^2$, $\forall x \in \mathcal{X}$ are assumed to be finite and independent of the SNR. Thus, to make sure that the RHS of (113) is finite, i.e.,

$$\frac{B \log |\mathcal{X}| + s(B + \mathsf{snr} \sum_{b=1}^{B} |e_b|^2 |x|^2)}{\log 2} < \infty \tag{114}$$

we can pick $s$ in the set $\mathcal{S} \subset \mathbb{R}^+$, which is defined as

$$\mathcal{S} \triangleq \left\{ s \in \mathbb{R}^+ : 0 < s \leq \frac{1}{B + \mathsf{snr} \sum_{b=1}^{B} |e_b|^2} \right\}. \tag{115}$$

Hence, for any $s \in \mathcal{S}$, we can apply the dominated convergence theorem [35], for which

$$\begin{aligned} &\lim_{\mathsf{snr}\to\infty} \mathbb{E}\left[ \log_2 \sum_{x' \in \mathcal{X}} \left( e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2} \right) \right] \\ &= \mathbb{E}\left[ \lim_{\mathsf{snr}\to\infty} \log_2 \sum_{x' \in \mathcal{X}} \left( e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2} \right) \right]. \end{aligned} \tag{116}$$

Replacing the supremum over $s > 0$ in (108) with the supremum over $s \in \mathcal{S}$ results in the lower bound to the GMI due to the suboptimal $s$. This suboptimality occurs because the optimizing $s$ that minimizes the expectation (109) can only be obtained from the set $\mathcal{S}$, not from the interval $s > 0$.

Substituting a specific value of $s$ in the interval $\left(0, \frac{1}{B + \mathsf{snr} \sum_{b=1}^{B} |e_b|^2}\right]$ further lower-bounds the GMI. As we will show later, the following choice of $s$:

$$\hat{s} = \frac{1}{B + \mathsf{snr}^{(1+\varepsilon)} \sum_{b=1}^{B} |e_b|^2}, \qquad \varepsilon > 0 \tag{117}$$

can give a tight GMI lower bound at high SNR by selecting an appropriate value of $\varepsilon$. At high SNR, the choice of $\varepsilon > 0$ allows for $\hat{s} \downarrow 0 (\varepsilon \uparrow \infty)$ and for $\hat{s} \to \frac{1}{B + \mathsf{snr} \sum_{b=1}^{B} |e_b|^2} (\varepsilon \downarrow 0)$.

Using the transformation of variables $\alpha_b = -\frac{\log |h_b|^2}{\log \mathsf{snr}}$ and $\theta_b = -\frac{\log |e_b|^2}{\log \mathsf{snr}}$, the exponential term in (107) with $s = \hat{s}$ becomes

$$e^{-\hat{s}|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + z|^2 + \hat{s}|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + z|^2} = e^{-\hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath \phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x'\right|^2 + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x\right|^2} \tag{118}$$

where $\phi_b^h$ and $\phi_b^e$ are the angles of $h_b$ and $e_b$, respectively. We have the following cases on the exponential convergence of (118) for all $B$ blocks as the SNR tends to infinity.

1) **Case 1:** $\alpha_b > 1$ for all $b = 1, \ldots, B$ (for any $\theta_b$). Note that under this condition, the perfect-CSIR mutual information $I(\boldsymbol{h})$ goes to zero as the SNR tends to infinity [6]. Due to the data-processing inequality (Proposition 4) and the nonnegativity property (Proposition 2), $I^{\mathrm{gmi}}(\hat{\boldsymbol{h}})$ also tends to zero as the SNR tends to infinity.
2) **Case 2:** $\alpha_b < 1$ and $\alpha_b < \theta_b$ for all $b = 1, \ldots, B$. From (118), we have the following dot equality for $x \neq x'$

$$\begin{aligned} &-\hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath \phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x'\right|^2 + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x\right|^2 \\ &\doteq -\hat{s} \times \mathsf{snr}^{1-\alpha_b}. \end{aligned} \tag{119}$$

The suboptimal $s$ is given by

$$\hat{s} = \frac{1}{B + \mathsf{snr}^{(1+\epsilon)} \sum_{b=1}^{B} |e_b|^2} \doteq \mathsf{snr}^{\min(0, \theta_{\min} - 1 - \varepsilon)} \tag{120}$$

where $\theta_{\min} = \min\{\theta_1, \ldots, \theta_B\}$. By exchanging the limit and the expectation as in (116), we can show that (118) tends to zero as the SNR tends to infinity if and only if $\alpha_b < \min(1, \theta_{\min} - \varepsilon)$ and $x \neq x'$. Otherwise, we can upper-bound (118) by one. This yields a further lower bound for the GMI. A tight lower bound is obtained by letting $\varepsilon \downarrow 0$.
3) **Case 3:** $\alpha_b < 1$ and $\alpha_b > \theta_b$ for all $b = 1, \ldots, B$. Then, the terms with $\mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x'$ and $\mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x$ in (118) dominate, and we have the following dot equality

$$\begin{aligned} &-\hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath \phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x'\right|^2 + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x\right|^2 \\ &\doteq -\hat{s}\left(|x'|^2 - |x|^2\right) \mathsf{snr}^{1-\theta_b} \end{aligned} \tag{121}$$

for $|x| \neq |x'|$ and

$$\begin{aligned} &-\hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath \phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x'\right|^2 + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath \phi_b^e} x\right|^2 \\ &\doteq -\hat{s} \cdot \mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}} |x|^2 \cdot \left(\cos(\phi_b^{eh}) - \cos(\phi_b^{eh} + \phi^{x'x})\right) \end{aligned} \tag{122}$$

for $|x| = |x'|$, $x \neq x'$, where $\phi_b^{eh} = \phi_b^e - \phi_b^h$, $\phi^{x'x} = \phi^{x'} - \phi^x$, and where $\phi^{x'}$ and $\phi^x$ are the angles of $x'$ and $x$, respectively. Note that using

$$\hat{s} = \frac{1}{B + \mathsf{snr}^{(1+\epsilon)} \sum_{b=1}^{B} |e_b|^2} \doteq \mathsf{snr}^{\min(0, \theta_{\min} - 1 - \varepsilon)} \tag{123}$$

with a strictly positive $\varepsilon$, both (121) and (122) always tend to zero as the SNR tends to infinity as a result of exchanging the limit and the expectation in (116). This makes (118) tend to one, and $I^{\mathrm{gmi}}(\hat{\boldsymbol{h}})$ is lower-bounded by zero.
4) **Case 4:** Without loss of generality, we group $\alpha_b$ and $\theta_b$ as follows
   - $\alpha_b > 1$ and $\alpha_b < \theta_b$, for $b = 1, \ldots, B_0$;
   - $\alpha_b < 1$ and $\alpha_b < \theta_b$, for $b = B_0 + 1, \ldots, B_1$;
   - $\alpha_b > 1$ and $\alpha_b > \theta_b$, for $b = B_1 + 1, \ldots, B_2$;
   - $\alpha_b < 1$ and $\alpha_b > \theta_b$, for $b = B_2 + 1, \ldots, B$.

Note that we select $s = \hat{s}$ that satisfies the exponential equality

$$\hat{s} \doteq \min\left(\mathsf{snr}^0, \mathsf{snr}^{\theta_{\min}-1-\varepsilon}\right) \tag{124}$$

where $\varepsilon$ is chosen such that

$$0 < \varepsilon < \theta_{\min} - \bar{\alpha}_{\max} \tag{125}$$

and where

$$\bar{\alpha}_{\max} = \max\left\{\alpha_b \middle| \alpha_b < \min(1, \theta_{\min}),\, b = 1, \ldots, B\right\}. \tag{126}$$

Then, the convergence of (118) can be explained as follows.

- For $b \in \{1, \ldots, B_0\}$, we have that $\alpha_b > 1$ and $\alpha_b < \theta_b$. Under this condition and after exchanging the limit and the expectation in (116), (118) tends to one for any $\hat{s} > 0$. It implies that $I_b^{\rm gmi}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to 0$, $b = 1, \ldots, B_0$ for any $\varepsilon > 0$.
- For $b \in \{B_0 + 1, \ldots, B_1\}$, we have that $\alpha_b < 1$ and $\alpha_b < \theta_b$. The dominating term in the exponent of (118) is given by $-\hat{s} \times \mathsf{snr}^{1-\alpha_b}$. Thus, for $b \in \{B_0 + 1, \ldots, B_1\}$ and $\alpha_b < \theta_{\min}$, exchanging the limit and the expectation in (116) yields the convergence of (118) to zero as the SNR tends to infinity. We then have that $I_b^{\rm gmi}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to M$. On the other hand, for $b \in \{B_0 + 1, \ldots, B_1\}$ and $\alpha_b \geq \theta_{\min}$, as the SNR tends to infinity, we observe the convergence of (118) to one as the SNR tends to infinity. This implies that $I_b^{\rm gmi}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to 0$.
- For $b \in \{B_1 + 1, \ldots, B_2\}$ and $\theta_b < 1$, we have the dot equality

$$\begin{aligned} - \hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath\phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath\phi_b^e} x'\right|^2 \\ + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath\phi_b^e} x\right|^2 \\ \doteq -\hat{s}\left(|x'|^2 - |x|^2\right)\mathsf{snr}^{1-\theta_b} \end{aligned} \tag{127}$$

for $|x| \neq |x'|$ and

$$\begin{aligned} - \hat{s}\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}} e^{\imath\phi_b^h}(x - x') + z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath\phi_b^e} x'\right|^2 \\ + \hat{s}\left|z - \mathsf{snr}^{\frac{1-\theta_b}{2}} e^{\imath\phi_b^e} x\right|^2 \\ \doteq \hat{s}|z|\mathsf{snr}^{\frac{1-\theta_b}{2}}\cos(\phi^z - \phi_b^e)|x| \\ \cdot\left(\cos\phi^{x'} - \cos\phi^x\right) \\ + \hat{s}|z|\mathsf{snr}^{\frac{1-\theta_b}{2}}\sin(\phi^z - \phi_b^e)|x| \\ \cdot\left(\sin\phi^{x'} - \sin\phi^x\right) \end{aligned} \tag{128}$$

for $|x| = |x'|$, $x \neq x'$, where $\phi^z$ is the angle of $z$. In this case, we cannot use the dominated convergence theorem [35] in (116) since there is a dependency on $z$. Instead, since the logarithm is a concave function of its argument, we first apply Jensen's inequality [36, Theorem 2.6.2] to the expectation in (107)

$$\begin{aligned} &\mathbb{E}\left[\log_2 \sum_{x'\in\mathcal{X}} \left(e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2}\right)\right] \\ &\leq \log_2 \mathbb{E}\left[\sum_{x'\in\mathcal{X}} \left(e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2}\right)\right] \end{aligned} \tag{129}$$

$$= \log_2 \sum_{x'\in\mathcal{X}} \mathbb{E}\left[\left(e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2}\right)\right]. \tag{130}$$

For a given $z \in \mathbb{C}$, $|z| < \infty$, we have the bounds

$$0 \leq \left(e^{-s|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + z|^2} \cdot e^{s|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + z|^2}\right) \tag{131}$$

$$\leq e^{s|z - \sqrt{\mathsf{snr}} e_b x|^2}. \tag{132}$$

Averaging over $Z$ yields

$$\mathbb{E}\left[e^{s|Z - \sqrt{\mathsf{snr}} e_b x|^2}\right] = \frac{1}{1-s} e^{\left(\frac{s^2}{1-s} + s\right)\mathsf{snr}|e_b|^2|x|^2} \tag{133}$$

where we have assumed $s < 1$ so that the the above expectation can be evaluated. Furthermore, using $s = \hat{s}$ in (117), the RHS of the above equation can be guaranteed to be finite. Thus, with $s = \hat{s}$, we can apply the dominated convergence theorem [35] as

$$\begin{aligned} &\lim_{\mathsf{snr}\to\infty} \mathbb{E}\left[\left(e^{-\hat{s}|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{\hat{s}|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2}\right)\right] \\ &= \mathbb{E}\left[\lim_{\mathsf{snr}\to\infty} \left(e^{-\hat{s}|\sqrt{\mathsf{snr}}(h_b x - \hat{h}_b x') + Z|^2} \cdot e^{\hat{s}|\sqrt{\mathsf{snr}}(h_b - \hat{h}_b)x + Z|^2}\right)\right]. \end{aligned} \tag{134}$$

For $|x| \neq |x'|$, using the relationship in (134) and $\varepsilon$ in (125), we observe that (127) tends to zero as the SNR tends to infinity and (118) tends to one. To evaluate (128), we first upper-bound

$$\begin{aligned} &\hat{s}|z|\mathsf{snr}^{\frac{1-\theta_b}{2}}\cos(\phi^z - \phi_b^e)|x| \cdot\left(\cos\phi^{x'} - \cos\phi^x\right) \\ &+ \hat{s}|z|\mathsf{snr}^{\frac{1-\theta_b}{2}}\sin(\phi^z - \phi_b^e)|x| \cdot\left(\sin\phi^{x'} - \sin\phi^x\right) \\ &\leq 4\hat{s}|z|\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|. \end{aligned} \tag{135}$$

Let $\nu = |z|$. Note that $\nu$ has the Rayleigh pdf

$$p(\nu) = 2\nu e^{-\nu^2}, \qquad \nu \geq 0. \tag{136}$$

Using the result in (130) and the upper bound in (135) for $|x| = |x'|$, we have that the expectation over $|Z|$ as

$$\begin{aligned}
&\mathbb{E}\left[e^{4\hat{s}|Z|\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|}\right] \\
&= \mathbb{E}\left[e^{4\hat{s}\nu\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|}\right] &(137)\\
&= \int_0^\infty e^{4\hat{s}\nu\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|}\cdot 2\nu e^{-\nu^2}d\nu &(138)\\
&= 1 + 2\sqrt{\pi}\hat{s}\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|e^{\left(4\hat{s}^2\mathsf{snr}^{1-\theta_b}|x|^2\right)} \\
&\qquad \cdot\left(1+\operatorname{erf}\left(2\hat{s}\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|\right)\right) &(139)\\
&\leq 1 + 4\sqrt{\pi}\hat{s}\mathsf{snr}^{\frac{1-\theta_b}{2}}|x|e^{\left(4\hat{s}^2\mathsf{snr}^{1-\theta_b}|x|^2\right)} &(140)
\end{aligned}$$

where $\operatorname{erf}(\cdot)$ is the error function [28]. The last inequality is due to the upper bound $\operatorname{erf}(a) \leq 1$. Note that for $\theta_b < 1$, we have

$$\begin{aligned}
\hat{s}\cdot\mathsf{snr}^{\frac{1-\theta_b}{2}} &\doteq \mathsf{snr}^{\theta_{\min}-1-\varepsilon}\cdot\mathsf{snr}^{\frac{1-\theta_b}{2}} &(141)\\
&\dot{\leq} \mathsf{snr}^{\theta_{\min}-1-\varepsilon}\cdot\mathsf{snr}^{1-\theta_b} &(142)\\
&\doteq \mathsf{snr}^{\theta_{\min}-\theta_b-\varepsilon}. &(143)
\end{aligned}$$

As $\theta_{\min} - \varepsilon$ is always less than $\theta_b$, the last dot equality implies that as the SNR tends to infinity, the upper bound in (140) tends to one. This provides an upper bound to the expectation over $Z$ in (130) at high SNR when $|x| = |x'|$, $\alpha_b > 1$ and $\theta_b < 1$. Complementing the result with the one for $|x| \neq |x'|$, $\alpha_b > 1$ and $\theta_b < 1$, we have that $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to 0$ when $\theta_b < 1$.

On the other hand, for $b \in \{B_1+1, \ldots, B_2\}$ and $\theta_b > 1$, we have that as the SNR tends to infinity, (118) tends to one for any $\hat{s} > 0$. It implies that $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to 0$ for any $\varepsilon > 0$.

- For $b \in \{B_2+1, \ldots, B\}$, we always have $\theta_b < 1$. It follows that for $|x| \neq |x'|$, we have the dot equality as in (121). On the other hand, for $|x| = |x'|$, $x \neq x'$, the dot equality follows from (122). Then, using $\varepsilon$ in (125) and exchanging the limit and the expectation in (116), we observe that for both $|x| \neq |x'|$ and $|x| = |x'|$, (118) tends to one as the SNR tends to infinity. Thus, we have that $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, \hat{s}) \to 0$ for all $b = B_2+1, \ldots, B$.

From **cases 1** to **4**, the generalized outage probability can be upper-bounded as follows:

$$\begin{aligned}
&P_{\text{gout}}(R) \\
&\doteq \mathsf{snr}^{-d_{\text{icsir}}^{\mathcal{X}}} &(144)\\
&\dot{\leq} \Pr\left\{\sum_{b=1}^{B}\mathbb{1}\Big\{\{\alpha_b \leq 1-\epsilon\}\cap\{\alpha_b \leq \theta_{\min}-\delta\}\Big\} < \frac{BR}{M}\right\} &(145)\\
&\doteq \int_{\underline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta}} p_{\boldsymbol{\alpha},\boldsymbol{\phi}^{\mathsf{h}}}(\boldsymbol{\alpha},\boldsymbol{\phi}^{\boldsymbol{h}})p_{\boldsymbol{\theta}}(\boldsymbol{\theta})p_{\boldsymbol{\phi}^{\mathsf{e}}}(\boldsymbol{\phi}^{\boldsymbol{e}})d\boldsymbol{\alpha}d\boldsymbol{\theta}d\boldsymbol{\phi}^{\boldsymbol{h}}d\boldsymbol{\phi}^{\boldsymbol{e}} &(146)
\end{aligned}$$

where we have defined

$$\underline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta} \triangleq \left\{\boldsymbol{\alpha},\boldsymbol{\theta} \in \mathbb{R}^B : \sum_{b=1}^{B}\mathbb{1}\Big\{\{\alpha_b \leq 1-\epsilon\}\cap\{\alpha_b \leq \theta_{\min}-\delta\}\Big\} < \frac{BR}{M}\right\} \tag{147}$$

for any $\epsilon, \delta > 0$. Applying the result in Lemma 3, we have that

$$d_{\text{icsir}}^{\mathcal{X}} \geq \inf_{\underline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta}\cap\{\boldsymbol{\alpha}\succeq\mathbf{0},\boldsymbol{\theta}\succeq d_e\times\mathbf{1}\}}\left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\alpha_b + \sum_{b=1}^{B}(\theta_b - d_{\text{e}})\right\}. \tag{148}$$

Following the steps used in [6], we can show that the values of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ achieving the infimum are given by

$$\begin{aligned}
\theta_b^* &= d_{\text{e}}, &&\text{for } b = 1,\ldots,B &(149)\\
\alpha_b^* &= \min(1-\epsilon, \theta_b^*-\delta), &&\text{for } b = 1,\ldots,B-b^* &(150)\\
\alpha_b^* &= 0, &&\text{for } b = B-b^*+1,\ldots,B &(151)
\end{aligned}$$

where $b^* \in \{0,\ldots,B-1\}$ is the unique integer satisfying $\frac{b^*}{B} < \frac{R}{M} \leq \frac{b^*+1}{B}$. As this is valid for any $\epsilon > 0$ and $\delta > 0$, the lower bound for $d_{\text{icsir}}^{\mathcal{X}}$ is tight if we let $\epsilon, \delta \downarrow 0$. This yields

$$d_{\text{icsir}}^{\mathcal{X}} \geq \min(1, d_{\text{e}}) \times \left(1+\frac{\tau}{2}\right)d_B(R) \tag{152}$$

where $d_B(R)$ is the Singleton bound [6]

$$d_B(R) = 1 + \left\lfloor B\left(1-\frac{R}{M}\right)\right\rfloor. \tag{153}$$

*GMI Upper Bound:* For a given $x, x' \in \mathcal{X}$, we define

$$f_{x,x'}(s_b,\mathsf{snr},h_b,e_b,z) \triangleq e^{-s_b\left|\sqrt{\mathsf{snr}}h_b(x-x')+z-\sqrt{\mathsf{snr}}e_b x'\right|^2+s_b\left|z-\sqrt{\mathsf{snr}}e_b x\right|^2} \tag{154}$$

and for a given $x \in \mathcal{X}$, we define

$$f_x(s_b,\mathsf{snr},h_b,e_b,z) \triangleq \log_2\sum_{x'\in\mathcal{X}} f_{x,x'}(s_b,\mathsf{snr},h_b,e_b,z). \tag{155}$$

Then, the GMI can be upper-bounded as (Proposition 3)

$$I^{\text{gmi}}(\hat{\boldsymbol{h}}) \leq \frac{1}{B}\sum_{b=1}^{B}\sup_{s_b>0} I_b^{\text{gmi}}\left(\mathsf{snr},h_b,\hat{h}_b,s_b\right) \tag{156}$$

where

$$\begin{aligned}
&I_b^{\text{gmi}}\left(\mathsf{snr},h_b,\hat{h}_b,s_b\right) \\
&= M - \frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big] &(157)\\
&= M - \mathbb{E}\left[\frac{1}{2^M}\sum_{x\in\mathcal{X}}f_x(s_b,\mathsf{snr},h_b,e_b,Z)\right]. &(158)
\end{aligned}$$

The configuration of the signal points in the constellation $\mathcal{X}$ plays an important role in the evaluation of the GMI upper bound. In particular, this relates to the energy level for each signal point $|x|^2$. Suppose that for a given constellation $\mathcal{X}$, we have $n$ energy levels. Denote $\mathcal{X}_{n'}$, $n'=1,\ldots,n$, as the subset of $\mathcal{X}$ corresponding to the $n'$-th energy level. Then, we can partition $\mathcal{X}$ into $n$ disjoint subsets $\mathcal{X}_{n'}$, $n'=1,\ldots,n$ such that

$$\mathcal{X}=\mathcal{X}_1\cup\cdots\cup\mathcal{X}_n. \tag{159}$$

Note that for each $n'$, $n'=1,\ldots,n$, all signal points in $\mathcal{X}_{n'}$ have the same energy.

The above partition represents all possible configurations of $\mathcal{X}$ with respect to the energy levels. For instance, with *fully*-non-equal-energy constellations, we have $|\mathcal{X}_{n'}|=1$, $n'=1,\ldots,n$. On the other hand, with *fully*-equal-energy constellations, we have $n=1$ and $|\mathcal{X}_1|=|\mathcal{X}|$.

The following asymptotic high-SNR analysis is based on the upper-bounding techniques from Proposition 3 and Fatou's lemma [37]. To this end, we use the change of variables from $|h_b|^2$ and $|e_b|^2$ to $\alpha_b$ and $\theta_b$ so that we can write

$$\begin{aligned}&e^{-s_b|\sqrt{\mathsf{snr}}h_b(x-x')+z-\sqrt{\mathsf{snr}}e_bx'|^2+s_b|z-\sqrt{\mathsf{snr}}e_bx|^2}\\&=e^{-s_b\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}}e^{\imath\phi_b^h}(x-x')+z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x'\right|^2}\\&\qquad\cdot e^{s_b\left|z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x\right|^2}.\end{aligned} \tag{160}$$

Similarly to the lower bound analysis, we expand the exponential term and consider the following cases.

1) **Case 1:** $\alpha_b>1$ Regardless of the value of $\theta_b$, the supremum of $I_b^{\mathrm{gmi}}(\mathsf{snr},h_b,\hat{h}_b,s_b)$ over $s_b>0$ in (156) tends to zero as it is upper-bounded by the perfect-CSIR mutual information for block $b$ [6].

2) **Case 2:** $\alpha_b<1$ and $\alpha_b<\theta_b$. The supremum in the RHS of (156) is equivalent to the following infimum

$$\inf_{s_b>0}\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big] \tag{161}$$

which can be lower-bounded as

$$\begin{aligned}&\inf_{s_b>0}\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big]\\&\geq\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[\inf_{s_b>0}f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big]\end{aligned} \tag{162}$$

by exchanging the infimum over $s_b$ and the expectation twice. Let $\bar{s}_b$ be the value of $s_b$ that gives the infimum in the RHS of (162). The choice of $\bar{s}_b$ depends on the behavior of the following term

$$e^{-s_b|\sqrt{\mathsf{snr}}h_b(x-x')+z-\sqrt{\mathsf{snr}}e_bx'|^2+s_b|z-\sqrt{\mathsf{snr}}e_bx|^2}. \tag{163}$$

It follows that if

$$|\sqrt{\mathsf{snr}}h_b(x-x')+z-\sqrt{\mathsf{snr}}e_bx'|^2>|z-\sqrt{\mathsf{snr}}e_bx|^2 \tag{164}$$

the solution of $\bar{s}_b$ is given by $\bar{s}_b\uparrow\infty$. Otherwise, the solution of $\bar{s}_b$ is given by $\bar{s}_b\downarrow 0$. Note that since at high SNR we have the exponential equality for $x\neq x'$

$$\begin{aligned}&-s_b\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}}e^{\imath\phi_b^h}(x-x')+z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x'\right|^2\\&\qquad+s_b\left|z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x\right|^2\\&\doteq -s_b\mathsf{snr}^{1-\alpha_b}\end{aligned} \tag{165}$$

for $\alpha_b<1$ and $\alpha_b<\theta_b$, it follows that in this case $\bar{s}_b\uparrow\infty$. Since $f_x(\bar{s}_b,\mathsf{snr},h_b,e_b,z)\geq 0$, we can apply Fatou's lemma [37] to the RHS of (162) as follows

$$\begin{aligned}&\lim_{\mathsf{snr}\to\infty}\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[f_x(\bar{s}_b,\mathsf{snr},h_b,e_b,Z)\Big]\\&\geq\frac{1}{2^M}\sum_{x\in\mathcal{X}}\mathbb{E}\Big[\lim_{\mathsf{snr}\to\infty}f_x(\bar{s}_b,\mathsf{snr},h_b,e_b,Z)\Big].\end{aligned} \tag{166}$$

This gives a further lower bound to the RHS of (162) at high SNR and yields an upper bound to $I^{\mathrm{gmi}}(\hat{\boldsymbol{h}})$. Using (165) and the limit in (166), we can show that (160) tends to zero for $x\neq x'$ as the SNR tends to infinity, and (160) is equal to one for $x=x'$. Thus, the supremum of $I_b^{\mathrm{gmi}}(\mathsf{snr},h_b,\hat{h}_b,s_b)$ over $s_b>0$ in (156) is upper-bounded by $M$ for $\alpha_b<1$ and $\alpha_b<\theta_b$.

3) **Case 3:** $\alpha_b<1$ and $\alpha_b>\theta_b$. The supremum in (156) is equivalent to the following infimum:

$$\inf_{s_b>0}\mathbb{E}\left[\frac{1}{2^M}\sum_{x\in\mathcal{X}}f_x(s_b,\mathsf{snr},h_b,e_b,Z)\right] \tag{167}$$

which can be lower-bounded as

$$\begin{aligned}&\inf_{s_b>0}\mathbb{E}\Big[\frac{1}{2^M}\sum_{x\in\mathcal{X}}f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big]\\&\geq\mathbb{E}\Big[\inf_{s_b>0}\frac{1}{2^M}\sum_{x\in\mathcal{X}}f_x(s_b,\mathsf{snr},h_b,e_b,Z)\Big]\end{aligned} \tag{168}$$

by exchanging the infimum over $s_b$ and the expectation. Let $\bar{s}_b$ be the value of $s_b$ that gives the infimum in the RHS of (168). Note that this $\bar{s}_b$ is different to the one in **case 2**. The dominating terms in the exponent of (160) can be shown to have the following exponential equality

$$\begin{aligned}&-s_b\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}}e^{\imath\phi_b^h}(x-x')+z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x'\right|^2\\&\qquad+s_b\left|z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x\right|^2\\&\doteq -s_b\left(|x'|^2-|x|^2\right)\mathsf{snr}^{1-\theta_b}\end{aligned} \tag{169}$$

for $|x|\neq|x'|$. On the other hand, for $|x|=|x'|$, $x\neq x'$, we have that

$$\begin{aligned}&-s_b\left|\mathsf{snr}^{\frac{1-\alpha_b}{2}}e^{\imath\phi_b^h}(x-x')+z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x'\right|^2\\&\qquad+s_b\left|z-\mathsf{snr}^{\frac{1-\theta_b}{2}}e^{\imath\phi_b^e}x\right|^2\\&\doteq -s_b\cdot\mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}}|x|^2\\&\qquad\cdot\left(\cos(\phi_b^{eh})-\cos(\phi_b^{eh}+\phi^{x'x})\right)\end{aligned} \tag{170}$$

with probability one since the constellation $\mathcal{X}$ is discrete and both $\phi_b^h$ and $\phi_b^e$ are uniformly distributed over $[0,2\pi)$. Hence, for a given $x \in \mathcal{X}$ and $x \neq x'$, we have that at high SNR

$$f_{x,x'}(s_b, \mathsf{snr}, h_b, e_b, z) = e^{-s_b\left(|x'|^2-|x|^2\right)\mathsf{snr}^{1-\theta_b}} \tag{171}$$

for $|x| \neq |x'|$, and

$$f_{x,x'}(s_b, \mathsf{snr}, h_b, e_b, z) = e^{-s_b \cdot \mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}}|x|^2\cdot\left(\cos(\phi_b^{eh})-\cos(\phi_b^{eh}+\phi^{x'x})\right)} \tag{172}$$

for $|x| = |x'|$. It follows that

$$f_x(s_b, \mathsf{snr}, h_b, e_b, z) = \log_2 \sum_{x'\in\mathcal{X}} f_{x,x'}(s_b, \mathsf{snr}, h_b, e_b, z). \tag{173}$$

Using the second order derivative of the log-sum-exp function $f_x(s_b, \mathsf{snr}, h_b, e_b, z)$ above, it can be shown that the function

$$\sum_{x\in\mathcal{X}} f_x(s_b, \mathsf{snr}, h_b, e_b, z) \tag{174}$$

is convex in $s_b$ for $s_b > 0$. To check whether the extreme point, which gives the global minimum to (174), exists for $s_b > 0$, we can simply find the derivative of (174) at $s_b = 0$ as shown in (175) and (176), at the bottom of the page.

Consider a pair of signal points $(x_1, x_2)$ having the same energy, i.e., $x_1, x_2 \in \mathcal{X}_{n'}$, $|\mathcal{X}_{n'}| \geq 2$, for $n' \in \{1, \ldots, n\}$, where $n$ is the number of energy levels in $\mathcal{X}$ as stated in (159). The contribution of the pair $(x_1, x_2)$ in the summations in (176) is given by

$$\begin{aligned}
&|x_1|^2\cos(\phi_b^{eh}) - |x_1|^2\cos(\phi_b^{eh}+\phi^{x_2x_1}) \\
&\quad + |x_2|^2\cos(\phi_b^{eh}) - |x_2|^2\cos(\phi_b^{eh}+\phi^{x_1x_2}) \\
&= |x_1|^2\Big(2\cos(\phi_b^{eh}) - \cos(\phi_b^{eh}+\phi^{x_2x_1}) \\
&\qquad\qquad - \cos(\phi_b^{eh}-\phi^{x_2x_1})\Big) \qquad (177)\\
&= |x_1|^2\Big(2\cos(\phi_b^{eh}) - 2\cos(\phi_b^{eh})\cos(\phi^{x_2x_1})\Big) \qquad (178)\\
&= \cos(\phi_b^{eh})|x_1|^2\Big(2 - 2\cos(\phi^{x_2x_1})\Big) \qquad (179)\\
&= \cos(\phi_b^{eh})|x_1|^2\Big(1 - \cos(\phi^{x_2x_1}) + 1 - \cos(\phi^{x_1x_2})\Big) \qquad (180)\\
&= \cos(\phi_b^{eh})|x_1|^2\big(1 - \cos(\phi^{x_2x_1})\big) \\
&\qquad + \cos(\phi_b^{eh})|x_2|^2\big(1 - \cos(\phi^{x_1x_2})\big) \qquad (181)
\end{aligned}$$

where we have used the equality $\phi^{x_1x_2} = -\phi^{x_2x_1}$ by definition of $\phi^{x_1x_2} = \phi^{x_1} - \phi^{x_2}$, the equality in energy $|x_1|^2 = |x_2|^2$, and the trigonometry identities

$$\cos(a+b) + \cos(a-b) = 2\cos(a)\cos(b) \tag{182}$$

and $\cos(-a) = \cos(a)$. Define

$$\mathcal{Q} \triangleq \{n' : |\mathcal{X}_{n'}| \geq 2, n' = 1, \ldots, n\}. \tag{183}$$

Using the result in (181), we can re-write the summations in (176) as

$$\begin{aligned}
&\sum_{x\in\mathcal{X}} \sum_{\substack{x'\neq x\\|x'|=|x|\\x'\in\mathcal{X}}} \left(|x|^2\cos(\phi_b^{eh}) - |x|^2\cos(\phi_b^{eh}+\phi^{x'x})\right) \\
&= \cos(\phi_b^{eh}) \sum_{n'\in\mathcal{Q}} \sum_{x\in\mathcal{X}_{n'}} \sum_{\substack{x'\neq x,\\x'\in\mathcal{X}_{n'}}} |x|^2(1-\cos(\phi^{x'x})) \qquad (184)
\end{aligned}$$

where we have incorporated all disjoint subsets $\mathcal{X}_{n'}, n' = 1, \ldots, n$ as given in (159) that satisfy $|\mathcal{X}_{n'}| \geq 2$.

We have from the last equation that the condition $1 - \cos(\phi^{x'x}) \geq 0$ is always true. Then, if $\cos(\phi_b^{eh}) \leq 0$, then the derivative in (176) is always nonnegative, which implies that the solution of $s_b$ that leads to the infimum in the RHS of (168) is given by $\bar{s}_b \downarrow 0$. By using $\bar{s}_b \downarrow 0$ and applying Fatou's lemma [37] to the RHS of (168)

$$\begin{aligned}
&\lim_{\mathsf{snr}\to\infty} \mathbb{E}\Big[\frac{1}{2^M}\sum_{x\in\mathcal{X}} f_x(\bar{s}_b, \mathsf{snr}, h_b, e_b, Z)\Big] \\
&\geq \mathbb{E}\Big[\lim_{\mathsf{snr}\to\infty} \frac{1}{2^M}\sum_{x\in\mathcal{X}} f_x(\bar{s}_b, \mathsf{snr}, h_b, e_b, Z)\Big] \qquad (185)
\end{aligned}$$

---

$$\begin{aligned}
&\frac{\partial}{\partial s_b}\left[\sum_{x\in\mathcal{X}}\log_2\left(1 + \sum_{\substack{x'\neq x\\|x'|\neq|x|\\x'\in\mathcal{X}}} e^{-s_b\mathsf{snr}^{1-\theta_b}\left(|x'|^2-|x|^2\right)} + \sum_{\substack{x'\neq x\\|x'|=|x|\\x'\in\mathcal{X}}} e^{-s_b\cdot\mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}}|x|^2\cdot\left(\cos(\phi_b^{eh})-\cos(\phi_b^{eh}+\phi^{x'x})\right)}\right)\right]\Bigg|_{s_b=0} \\
&= \frac{1}{|\mathcal{X}|\log 2}\sum_{x\in\mathcal{X}}\left(\mathsf{snr}^{1-\theta_b}\cdot\sum_{\substack{x'\neq x\\|x'|\neq|x|\\x'\in\mathcal{X}}}(|x|^2-|x'|^2) - \mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}}|x|^2\cdot\sum_{\substack{x'\neq x\\|x'|=|x|\\x'\in\mathcal{X}}}\left(\cos(\phi_b^{eh})-\cos(\phi_b^{eh}+\phi^{x'x})\right)\right) \qquad (175)\\
&= \frac{-\mathsf{snr}^{1-\frac{\alpha_b+\theta_b}{2}}}{|\mathcal{X}|\log 2}\sum_{x\in\mathcal{X}}\sum_{\substack{x'\neq x\\|x'|=|x|\\x'\in\mathcal{X}}}\left(|x|^2\cos(\phi_b^{eh}) - |x|^2\cos(\phi_b^{eh}+\phi^{x'x})\right) \qquad (176)
\end{aligned}$$

we have that the upper bound for the supremum of $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, s_b)$ over $s_b > 0$ in (156) tends to zero as the SNR tends to infinity.

On the other hand, if $\cos(\phi_b^{eh}) > 0$, then the derivative in (176) is always nonpositive. Thus, there is a possibility that there exists a positive number $s_b$ in the interval $s_b > 0$ that leads to the infimum in the RHS of (168). This also implies that the upper bound for the supremum of $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, s_b)$ over $s_b > 0$ in (156) is in the interval $[0, M]$. Herein we can derive a loose upper bound as follows. We first define the event $\Xi_b$ as

$$\Xi_b \triangleq \left\{\phi_b^h, \phi_b^e \in [0, 2\pi) : \cos\left(\phi_b^e - \phi_b^h\right) > 0\right\} \quad (186)$$

where we have explicitly written $\phi_b^{eh} = \phi_b^e - \phi_b^h$. A loose upper bound is then obtained by considering that when $\Xi_b$ occurs, the upper bound for the supremum of $I_b^{\text{gmi}}(\mathsf{snr}, h_b, \hat{h}_b, s_b)$ over $s_b > 0$ in (156) is given by $M$. This loose bound is sufficient to show that the upper bound for the SNR exponent is tight with the lower bound.

From **cases 1** to **3**, we can show that the generalized outage probability is lower-bounded as follows:

$$\begin{aligned} &P_{\text{gout}}(R) \\ &\doteq \mathsf{snr}^{-d_{\text{icsir}}^{\mathcal{X}}} && (187) \\ &\dot{\geq} \Pr\left\{\frac{1}{B}\sum_{b=1}^{B} M \cdot \mathbb{1}\left\{\mathcal{E}^{\epsilon,\delta}(\alpha_b, \theta_b, \Xi_b)\right\} < R\right\} && (188) \\ &\doteq \int_{\overline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta}} p_{\boldsymbol{\alpha},\boldsymbol{\phi}^h}(\boldsymbol{\alpha}, \boldsymbol{\phi}^{\boldsymbol{h}}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) p_{\boldsymbol{\phi}^e}(\boldsymbol{\phi}^{\boldsymbol{e}}) d\boldsymbol{\alpha} d\boldsymbol{\theta} d\boldsymbol{\phi}^{\boldsymbol{h}} d\boldsymbol{\phi}^{\boldsymbol{e}} && (189) \end{aligned}$$

where we have defined

$$\begin{aligned} &\mathcal{E}^{\epsilon,\delta}(\alpha_b, \theta_b, \Xi_b) \\ &\triangleq \Big\{\left\{\{\alpha_b \leq 1+\epsilon\} \cap \{\alpha_b \leq \theta_b + \delta\}\right\} \\ &\qquad \cup \left\{\{\alpha_b \leq 1+\epsilon\} \cap \{\alpha_b > \theta_b + \delta\} \cap \Xi_b\right\}\Big\} \end{aligned} \quad (190)$$

for $\epsilon, \delta > 0$, and

$$\overline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta} \triangleq \left\{\boldsymbol{\alpha}, \boldsymbol{\theta} \in \mathbb{R}^B : \sum_{b=1}^{B} \mathbb{1}\left\{\mathcal{E}^{\epsilon,\delta}(\alpha_b, \theta_b, \Xi_b)\right\} < \frac{BR}{M}\right\}. \quad (191)$$

From Lemma 3 and the set $\overline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta}$, we have that

$$d_{\text{icsir}}^{\mathcal{X}} \leq \inf_{\overline{\mathcal{O}}_{\mathcal{X}}^{\epsilon,\delta} \cap \{\boldsymbol{\alpha} \succeq \mathbf{0}, \boldsymbol{\theta} \succeq d_{\text{e}} \times \mathbf{1}\}} \left\{\left(1 + \frac{\tau}{2}\right)\sum_{b=1}^{B} \alpha_b + \sum_{b=1}^{B}(\theta_b - d_{\text{e}})\right\}. \quad (192)$$

Similarly to the GMI lower bound, it is not difficult to show that the values of $\theta_b$, $b = 1, \ldots, B$ achieving the infimum are given by $d_e$. To find the values of $\alpha_b$, $b = 1, \ldots, B$ that solve for the infimum, we need to see whether it is possible to find the values of $\phi_b^h$ and $\phi_b^e$ that do not belong to $\Xi_b$. Note that the following condition:

$$\frac{\pi}{2} \leq \phi_b^e - \phi_b^h \leq \frac{3\pi}{2} \quad (193)$$

implies that $\cos(\phi_b^e - \phi_b^h) \leq 0$. Thus, from (186) and (193), we can always find $\phi_b^h$ and $\phi_b^e$ that do not belong to $\Xi_b$. It then follows from [6] that the values of $\alpha_b$, $b = 1, \ldots, B$ achieving the infimum are given by

$$\begin{aligned} \alpha_b^* &= \min(1+\epsilon, \theta_b^* + \delta), && \text{for } b = 1, \ldots, B - b^* && (194) \\ \alpha_b^* &= 0, && \text{for } b = B - b^* + 1, \ldots, B && (195) \end{aligned}$$

where $b^* \in \{0, \ldots, B-1\}$ is the unique integer satisfying $\frac{b^*}{B} < \frac{R}{M} \leq \frac{b^*+1}{B}$.

Substituting the values of $\alpha_b$ and $\theta_b$, $b = 1, \ldots, B$ that achieve the infimum in the RHS of (192), we obtain the upper bound for the SNR exponent

$$d_{\text{icsir}}^{\mathcal{X}} \leq \min(1, d_e) \times \left(1 + \frac{\tau}{2}\right) d_B(R) \quad (196)$$

where we have let $\epsilon, \delta \downarrow 0$ to make the upper bound tight.

*2) MIMO Case:* Recall the function in (40)

$$\begin{aligned} &I_b^{\text{gmi}}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s) \\ &= Mn_{\text{t}} - \mathbb{E}\left[\log_2 \sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\text{t}}}} \left(e^{-s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\text{t}}}}(\boldsymbol{H}_b \mathsf{x} - \hat{\boldsymbol{H}}_b \boldsymbol{x}') + \mathsf{z}\right\|_F^2} \cdot e^{s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\text{t}}}}(\boldsymbol{H}_b - \hat{\boldsymbol{H}}_b)\mathsf{x} + \mathsf{z}\right\|_F^2}\right)\right] \end{aligned} \quad (197)$$

where the expectation is over $(\mathsf{x}, \mathsf{z})$. It follows that the GMI is given by

$$I^{\text{gmi}}(\hat{\boldsymbol{H}}) = \sup_{s>0} \frac{1}{B}\sum_{b=1}^{B} I_b^{\text{gmi}}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s). \quad (198)$$

Mimicking the analysis done for the SISO case, we have the GMI lower and upper bounds as follows.

*GMI Lower Bound:* Using the suboptimal $\hat{s} \in \mathcal{S}$ to apply the dominated convergence theorem [35], we have that

$$\begin{aligned} &-\hat{s}\left\|\sqrt{\frac{\mathsf{snr}}{n_{\text{t}}}}\boldsymbol{H}_b(\boldsymbol{x} - \boldsymbol{x}') + \boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\text{t}}}}\boldsymbol{E}_b \boldsymbol{x}'\right\|_F^2 + \hat{s}\left\|\boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\text{t}}}}\boldsymbol{E}_b \boldsymbol{x}\right\|_F^2 \\ &\doteq -\hat{s}\sum_{r=1}^{n_{\text{r}}}\left|\sum_{t=1}^{n_{\text{t}}} \mathsf{snr}^{\frac{1-\alpha_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^h}(x_t - x_t') + z_r - \sum_{t=1}^{n_{\text{t}}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t'\right|^2 \\ &\quad + \hat{s}\sum_{r=1}^{n_{\text{r}}}\left|z_r - \sum_{t=1}^{n_{\text{t}}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t\right|^2 \end{aligned} \quad (199)$$

where now $|e_b|^2$ in (117) changes to $\|\boldsymbol{E}_b\|_F^2$, and where $\phi_{b,r,t}^h$ and $\phi_{b,r,t}^e$ are the angles of $h_{b,r,t}$ and $e_{b,r,t}$, respectively. Similarly to what it is done in [30], define the following sets $\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$, $\underline{\mathcal{S}}_{b}^{(\epsilon,\delta)}$, and $\underline{\kappa}_b$ for $\epsilon$, $\delta > 0$ as

$$\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \Big\{t : \{\alpha_{b,r,t} \le 1-\epsilon\} \cap \{\alpha_{b,r,t} \le \theta_{\min} - \delta\},\ t = 1,\dots,n_{\rm t}\Big\} \quad (200)$$

$$\underline{\mathcal{S}}_{b}^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_{\rm r}} \underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \quad (201)$$

$$\underline{\kappa}_b \triangleq \left|\underline{\mathcal{S}}_{b}^{(\epsilon,\delta)}\right| \quad (202)$$

where now $\theta_{\min} \triangleq \min\{\theta_{1,1,1},\dots,\theta_{b,r,t},\dots,\theta_{B,n_{\rm r},n_{\rm t}}\}$. Note that $\hat{s}$ satisfies

$$\hat{s} \doteq \mathsf{snr}^{\min(0,\theta_{\min}-1-\varepsilon)} \quad (203)$$

where $\varepsilon$ is chosen such that

$$0 < \varepsilon < \theta_{\min} - \bar{\alpha}_{\max} \quad (204)$$

and where

$$\bar{\alpha}_{\max} = \max\Big\{\alpha_{b,r,t}\Big|\alpha_{b,r,t} < \min(1,\theta_{\min}),\ b = 1,\dots,B,\ r = 1,\dots,n_{\rm r},\ t = 1,\dots,n_{\rm t}\Big\}. \quad (205)$$

For $r = 1,\dots,n_{\rm r}$ and $x_t \neq x_t'$, if there exists $\alpha_{b,r,t}$ satisfying the constraint set $\underline{\mathcal{S}}_{b}^{(\epsilon,\delta)}$, then with $s = \hat{s}$, the exponential function inside the expectation in the RHS of (197) tends to zero as the SNR tends to infinity. Otherwise, the exponential function converges to one (as the lower bound implies). Therefore, we can write the following exponential equality for high SNR

$$-\hat{s}\left|\sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{\frac{1-\alpha_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^h}(x_t - x_t') + z_r - \sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t'\right|^2 + \hat{s}\left|z_r - \sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t\right|^2$$
$$\doteq -\hat{s}\left|\sum_{\substack{t\in\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}\\ x_t\neq x_t'}} \mathsf{snr}^{\frac{1-\alpha_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^h}(x_t - x_t') + z_r - \sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t'\right|^2 + \hat{s}\left|z_r - \sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{\frac{1-\theta_{b,r,t}}{2}} e^{\imath\phi_{b,r,t}^e} x_t\right|^2. \quad (206)$$

Let $s^*$ be the value of $s > 0$ that solves the supremum in the RHS of (198). Due to the use of suboptimal $s$, $\hat{s}$ in (206), we have the upper bound for the expectation over z at high SNR as follows

$$\lim_{\mathsf{snr}\to\infty} \mathbb{E}\left[\log_2 \sum_{\boldsymbol{x}'\in\mathcal{X}^{n_{\rm t}}} \left(e^{-s^*\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}')+\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2} \cdot e^{s^*\left\|\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)\right]$$
$$\le \mathbb{E}\left[\log_2 \sum_{\boldsymbol{x}'\in\mathcal{X}^{n_{\rm t}}} \mathbb{1}\left\{x_t \neq x_t', \forall t \in \underline{\mathcal{S}}_{b}^{(\epsilon,\delta)}\right\}\right] \quad (207)$$
$$= M(n_{\rm t} - \underline{\kappa}_b) \quad (208)$$

for all $\boldsymbol{x} \in \mathcal{X}^{n_{\rm t}}$. Thus

$$\lim_{\mathsf{snr}\to\infty} I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s^*) \ge M\underline{\kappa}_b \quad (209)$$

and $P_{\rm gout}(R)$ is upper-bounded as

$$P_{\rm gout}(R) \dot{\le} \Pr\left\{\frac{1}{B}\sum_{b=1}^{B} M\underline{\kappa}_b < R\right\}. \quad (210)$$

Define

$$\underline{\mathcal{O}}_{\mathcal{X}} \triangleq \left\{\boldsymbol{A},\boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm r}\times n_{\rm t}} : \sum_{b=1}^{B}\underline{\kappa}_b < \frac{BR}{M}\right\}. \quad (211)$$

It follows that applying the result in Lemma 3 yields the lower bound for the SNR exponent:

$$d_{\rm icsir}^{\mathcal{X}} \ge \inf_{\underline{\mathcal{O}}_{\mathcal{X}}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}} \left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}\alpha_{b,r,t} + \sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}(\theta_{b,r,t} - d_e)\right\}. \quad (212)$$

We can observe from (200) that the solution of $\theta_{b,r,t}$ for all $b = 1,\dots,B$, $r = 1,\dots,n_{\rm r}$ and $t = 1,\dots,n_{\rm t}$ to the above infimum is given by $d_e$. Following [30], it can be proved that the solution of the above infimum is given as

$$d_{\rm icsir}^{\mathcal{X}} \ge \min(1,d_e)\cdot\left(1+\frac{\tau}{2}\right)n_{\rm r}\left(1+\left\lfloor B\left(n_{\rm t} - \frac{R}{M}\right)\right\rfloor\right). \quad (213)$$

*GMI Upper Bound:* Similarly to the SISO analysis, the GMI upper bound is evaluated using Proposition 3 and Fatou's lemma [37]. The only difference with the GMI lower bound is in the definition of the sets

$$\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \Big\{t : \{\{\alpha_{b,r,t} \le 1+\epsilon\} \cap \{\alpha_{b,r,t} \le \theta_{b,r,t}+\delta\}\} \cup \{\{\alpha_{b,r,t} \le 1+\epsilon\} \cap \{\alpha_{b,r,t} > \theta_{b,r,t}+\delta\} \cap \Xi_{b,r,t}\},\ t = 1,\dots,n_{\rm t}\Big\} \quad (214)$$

$$\overline{\mathcal{S}}_{b}^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_{\rm r}} \overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \quad (215)$$

$$\overline{\kappa}_b \triangleq \left|\overline{\mathcal{S}}_{b}^{(\epsilon,\delta)}\right| \quad (216)$$

where

$$\Xi_{b,r,t} \triangleq \left\{\phi_{b,r,t}^{h}, \phi_{b,r,t}^{e} \in [0,2\pi) : \cos\left(\phi_{b,r,t}^{e} - \phi_{b,r,t}^{h}\right) > 0\right\}. \tag{217}$$

Following the same steps used in the SISO analysis, we can lower-bound the expectation over z as follows:

$$\lim_{\mathsf{snr}\to\infty} \mathbb{E}\left[\log_2 \sum_{\boldsymbol{x}'\in\mathcal{X}^{n_t}} \left(e^{-s^*\left\|\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}')+\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2} \cdot e^{s^*\left\|\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)\right]$$

$$\geq \mathbb{E}\left[\log_2 \sum_{\boldsymbol{x}'\in\mathcal{X}^{n_t}} \mathbb{1}\left\{x_t \neq x_t', \forall t \in \overline{\mathcal{S}}_b^{(\epsilon,\delta)}\right\}\right] \tag{218}$$

$$= M(n_t - \overline{\kappa}_b). \tag{219}$$

The generalized outage probability can be lower-bounded as

$$P_{\text{gout}}(R) \dot{\geq} \Pr\left\{\frac{1}{B}\sum_{b=1}^{B} M\overline{\kappa}_b < R\right\}. \tag{220}$$

Define

$$\overline{\mathcal{O}}_{\mathcal{X}} \triangleq \left\{\boldsymbol{A}, \boldsymbol{\Theta} \in \mathbb{R}^{B\times n_r\times n_t} : \sum_{b=1}^{B} \overline{\kappa}_b < \frac{BR}{M}\right\}. \tag{221}$$

Using $\overline{\mathcal{O}}_{\mathcal{X}}$ to apply the result in Lemma 3 and following the technique used for the GMI lower bound, the SNR exponent can be proved to be upper-bounded as

$$d_{\text{icsir}}^{\mathcal{X}} \leq \min(1, d_e) \cdot \left(1 + \frac{\tau}{2}\right) n_r \left(1 + \left\lfloor B\left(n_t - \frac{R}{M}\right)\right\rfloor\right). \tag{222}$$

This is because the infimum solution for $\theta_{b,r,t}$ in (214) is same as the solution for $\theta_{\min}$ in (200) (given by $d_e$), and because we can always find $\phi_{b,r,t}^{h}$ and $\phi_{b,r,t}^{e}$ that do not belong to $\Xi_{b,r,t}$. This completes the converse part of the main theorem for discrete inputs.

## Appendix D
## Proof of Gaussian Input Mismatched CSIR SNR Exponent

For i.i.d. Gaussian inputs, (38) can be written as (in natural base log)

$$I_b^{\text{gmi}}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s) = \log\det\left(\boldsymbol{I}_{n_r} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}\frac{\mathsf{snr}}{n_t}\right) - s\left(n_r + \frac{\mathsf{snr}}{n_t}\|\hat{\boldsymbol{H}}_b - \boldsymbol{H}_b\|_F^2\right) + \mathbb{E}\left[s\mathsf{y}^{\dagger}\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}\,\middle|\,\mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right] \tag{223}$$

where

$$\hat{\Sigma}_{\boldsymbol{y}} \triangleq \boldsymbol{I}_{n_r} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}\frac{\mathsf{snr}}{n_t}. \tag{224}$$

Herein we derive the lower and upper bounds to the above expression to prove the converse part for the Gaussian inputs.

*1) GMI Lower Bound:* It is not difficult to see that $\mathbb{E}\left[s\mathsf{y}^{\dagger}\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}\,\middle|\,\mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right]$ is nonnegative. Then, we have that

$$I_b^{\text{gmi}}(\mathsf{snr}, \hat{\boldsymbol{H}}_b, \boldsymbol{E}_b, s) \geq \log\det\left(\boldsymbol{I}_{n_r} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}\frac{\mathsf{snr}}{n_t}\right) - s\left(n_r + \frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2\right). \tag{225}$$

Without loss of generality, assume that $n_t \geq n_r$[6]. Let $\hat{\lambda}_{b,i}, i = 1, \ldots, n_r$ be the $i$-th eigenvalue of $\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}$. Then, the RHS of (225) can be converted into eigenvalues expression

$$I_b^{\text{gmi}}(\mathsf{snr}, \hat{\boldsymbol{\lambda}}_b, \boldsymbol{E}_b, s) \geq \log\prod_{i=1}^{n_r}\left(1 + s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_t}\right) - s\left(n_r + \frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2\right). \tag{226}$$

Note that we can loosen the bound by considering the following technique:

$$\prod_{i=1}^{n_r}\left(1 + s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_t}\right) = \left(1 + s\hat{\lambda}_{b,1}\frac{\mathsf{snr}}{n_t}\right)\left(1 + s\hat{\lambda}_{b,2}\frac{\mathsf{snr}}{n_t}\right)\ldots\left(1 + s\hat{\lambda}_{b,n_r}\frac{\mathsf{snr}}{n_t}\right) \tag{227}$$

$$\geq 1 + s\hat{\lambda}_{b,1}\frac{\mathsf{snr}}{n_t} + s\hat{\lambda}_{b,2}\frac{\mathsf{snr}}{n_t} + \ldots + s\hat{\lambda}_{b,n_r}\frac{\mathsf{snr}}{n_t} \tag{228}$$

$$= 1 + s\frac{\mathsf{snr}}{n_t}\sum_{i=1}^{n_r}\hat{\lambda}_{b,i} \tag{229}$$

$$= 1 + s\frac{\mathsf{snr}}{n_t}\|\hat{\boldsymbol{H}}_b\|_F^2 \tag{230}$$

$$= 1 + s\frac{\mathsf{snr}}{n_t}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}|\hat{h}_{b,r,t}|^2 \tag{231}$$

where the last two equations are due to the fact that $\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}$ is a positive semidefinite matrix, where the singular values are always zero or positive. Hence, it holds that [27]

$$\|\hat{\boldsymbol{H}}_b\|_F^2 = \sum_{i=1}^{n_r}\hat{\lambda}_{b,i} = \sum_{r=1}^{n_r}\sum_{t=1}^{n_t}|\hat{h}_{b,r,t}|^2. \tag{232}$$

We then have a compact representation of the lower bound to the GMI as

$$I^{\text{gmi}}(\hat{\boldsymbol{H}}) = \sup_{s>0}\frac{1}{B}\sum_{b=1}^{B} I_b^{\text{gmi}}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s) \tag{233}$$

$$\geq \sup_{s>0}\frac{1}{B}\sum_{b=1}^{B}\left\{\log\left(1 + s\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_t}\right) - s\left(n_r + \frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2\right)\right\}. \tag{234}$$

[6] If $n_t < n_r$, then it suffices to replace $(\boldsymbol{I}_{n_r} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}\frac{\mathsf{snr}}{n_t})$ with $(\boldsymbol{I}_{n_t} + s\hat{\boldsymbol{H}}_b^{\dagger}\hat{\boldsymbol{H}}_b\frac{\mathsf{snr}}{n_t})$.

The high-SNR optimizing $s$ is difficult to evaluate in a closed form due to the sum for all $B$ blocks involving the logarithm function. The suboptimal $s$ can be obtained using the following argument. For any $s > 0$, we can lower-bound

$$\sum_{b=1}^{B}\left\{\log\left(1+s\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_\mathrm{t}}\right)-s\left(n_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2\right)\right\}$$
$$\geq \sum_{b=1}^{B}\log\left(s\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_\mathrm{t}}\right)-s\left(Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2\right). \tag{235}$$

We continue the evaluation by performing the first-order derivative to the RHS of (235) with respect to $s$ and equating it to zero. From this step, we obtain a suboptimal $s$ with respect to (234) and this suboptimal $s$ is given as

$$\hat{s}=\frac{B}{Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2}. \tag{236}$$

Substituting this $\hat{s}$ into (234) gives

$$I^{\mathrm{gmi}}(\hat{\boldsymbol{H}})$$
$$\geq \frac{1}{B}\log\left(\prod_{b=1}^{B}e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_\mathrm{t}}\right)\right). \tag{237}$$

Note that from $\hat{h}_{b,r,t}=h_{b,r,t}+e_{b,r,t}$, we have the following relationship

$$|\hat{h}_{b,r,t}|^2=|h_{b,r,t}|^2+|e_{b,r,t}|^2+2|h_{b,r,t}||e_{b,r,t}|\cos(\phi_{b,r,t}^h-\phi_{b,r,t}^e). \tag{238}$$

Let $\hat{\alpha}_{b,r,t}=-\frac{\log|\hat{h}_{b,r,t}|^2}{\log\mathsf{snr}}$, $\alpha_{b,r,t}=-\frac{\log|h_{b,r,t}|^2}{\log\mathsf{snr}}$ and $\theta_{b,r,t}=-\frac{\log|e_{b,r,t}|^2}{\log\mathsf{snr}}$. Then, for any real positive number $\varsigma>0$, we have that for $\alpha_{b,r,t}\neq\theta_{b,r,t}$

$$\mathsf{snr}^{\varsigma}|\hat{h}_{b,r,t}|^2=\mathsf{snr}^{\varsigma-\hat{\alpha}_{b,r,t}}\doteq\mathsf{snr}^{\varsigma-\min(\alpha_{b,r,t},\theta_{b,r,t})}. \tag{239}$$

On the other hand, for $\alpha_{b,r,t}=\theta_{b,r,t}$, we have the following four cases.

- If $h_{b,r,t}=e_{b,r,t}$ and $h_{b,r,t}\neq 0$, we have that

$$\mathsf{snr}^{\varsigma}|\hat{h}_{b,r,t}|^2=\mathsf{snr}^{\varsigma-\hat{\alpha}_{b,r,t}} \tag{240}$$
$$=4\mathsf{snr}^{\varsigma}|h_{b,r,t}|^2 \tag{241}$$
$$\doteq\mathsf{snr}^{\varsigma-\alpha_{b,r,t}}\doteq\mathsf{snr}^{\varsigma-\theta_{b,r,t}}. \tag{242}$$

- If $h_{b,r,t}=e_{b,r,t}^*$ and $\cos(\phi_{b,r,t}^h)\neq 0$, where $e_{b,r,t}^*$ denotes the complex conjugate of $e_{b,r,t}$, we have that

$$\mathsf{snr}^{\varsigma}|\hat{h}_{b,r,t}|^2=\mathsf{snr}^{\varsigma-\hat{\alpha}_{b,r,t}} \tag{243}$$
$$=4\mathsf{snr}^{\varsigma}|h_{b,r,t}|^2\cos^2(\phi_{b,r,t}^h) \tag{244}$$
$$\doteq\mathsf{snr}^{\varsigma-\alpha_{b,r,t}}\doteq\mathsf{snr}^{\varsigma-\theta_{b,r,t}}. \tag{245}$$

- If $-h_{b,r,t}=e_{b,r,t}^*$ and $\sin(\phi_{b,r,t}^h)\neq 0$, where $e_{b,r,t}^*$ denotes the complex conjugate of $e_{b,r,t}$, we have that

$$\mathsf{snr}^{\varsigma}|\hat{h}_{b,r,t}|^2=\mathsf{snr}^{\varsigma-\hat{\alpha}_{b,r,t}} \tag{246}$$
$$=4\mathsf{snr}^{\varsigma}|h_{b,r,t}|^2\sin^2(\phi_{b,r,t}^h) \tag{247}$$
$$\doteq\mathsf{snr}^{\varsigma-\alpha_{b,r,t}}\doteq\mathsf{snr}^{\varsigma-\theta_{b,r,t}}. \tag{248}$$

- If $h_{b,r,t}=-e_{b,r,t}$, we have that

$$\mathsf{snr}^{\varsigma}|\hat{h}_{b,r,t}|^2=\mathsf{snr}^{\varsigma-\hat{\alpha}_{b,r,t}}=0. \tag{249}$$

Note that the condition for $h_{b,r,t}=-e_{b,r,t}$ also covers the condition for $h_{b,r,t}=e_{b,r,t}=0$, the condition for $h_{b,r,t}=e_{b,r,t}^*$ with $\cos(\phi_{b,r,t}^h)=0$, and the condition for $-h_{b,r,t}=e_{b,r,t}^*$ with $\sin(\phi_{b,r,t}^h)=0$.
Using the preceding relationships, we have that

$$e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_\mathrm{t}}\right)$$
$$=e^{-1}\left(1+\frac{B\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\mathsf{snr}^{-\hat{\alpha}_{b,r,t}}}{Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\mathsf{snr}^{-\theta_{b,r,t}}}\right) \tag{250}$$
$$\doteq\max\left(\mathsf{snr}^0,\frac{\mathsf{snr}^{1-\hat{\alpha}_{b,\min}}}{\max\left(\mathsf{snr}^0,\mathsf{snr}^{1-\theta_{\min}}\right)}\right) \tag{251}$$
$$\doteq\max\left(\mathsf{snr}^0,\min\left(\mathsf{snr}^0,\mathsf{snr}^{\theta_{\min}-1}\right)\times\mathsf{snr}^{1-\hat{\alpha}_{b,\min}}\right) \tag{252}$$

where

$$\theta_{b,\min}\triangleq\min\{\theta_{b,1,1},\ldots,\theta_{b,r,t},\ldots,\theta_{b,n_\mathrm{r},n_\mathrm{t}}\} \tag{253}$$
$$\theta_{\min}\triangleq\min\{\theta_{1,\min},\ldots,\theta_{b,\min},\ldots,\theta_{B,\min}\} \tag{254}$$
$$\hat{\alpha}_{b,\min}\triangleq\min\{\hat{\alpha}_{b,1,1},\ldots,\hat{\alpha}_{b,r,t},\ldots,\hat{\alpha}_{b,n_\mathrm{r},n_\mathrm{t}}\}. \tag{255}$$

Define $\alpha_{b,\min}$, $(r,t)_{\alpha_{b,\min}}$ and $(r,t)_{\theta_{b,\min}}$ as

$$\alpha_{b,\min}\triangleq\min\{\alpha_{b,1,1},\ldots,\alpha_{b,r,t},\ldots,\alpha_{b,n_\mathrm{r},n_\mathrm{t}}\} \tag{256}$$
$$(r,t)_{\alpha_{b,\min}}\triangleq\arg\min_{\substack{r=1,\ldots,n_\mathrm{r}\\t=1,\ldots,n_\mathrm{t}}}\alpha_{b,r,t} \tag{257}$$
$$(r,t)_{\theta_{b,\min}}\triangleq\arg\min_{\substack{r=1,\ldots,n_\mathrm{r}\\t=1,\ldots,n_\mathrm{t}}}\theta_{b,r,t}. \tag{258}$$

We have the following cases.

1) **Case 1:** $(r,t)_{\alpha_{b,\min}}\neq(r,t)_{\theta_{b,\min}}$. This refers to the case where the indexes $(r,t)$ for which the minimum occurs are different for $\alpha_{b,r,t}$ and $\theta_{b,r,t}$. Clearly, we have that

$$\frac{\mathsf{snr}}{n_\mathrm{t}}\|\hat{\boldsymbol{H}}_b\|_F^2\doteq\frac{\mathsf{snr}}{n_\mathrm{t}}\mathsf{snr}^{-\hat{\alpha}_{b,\min}}\doteq\mathsf{snr}^{1-\min(\alpha_{b,\min},\theta_{b,\min})}. \tag{259}$$

It follows that

$$e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_\mathrm{r}+\frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_\mathrm{t}}\right)$$
$$\doteq\max\left(\mathsf{snr}^0,\min\left(\mathsf{snr}^0,\mathsf{snr}^{\theta_{\min}-1}\right)\mathsf{snr}^{1-\hat{\alpha}_{b,\min}}\right) \tag{260}$$
$$\doteq\max\left(\mathsf{snr}^0,\min\left(\mathsf{snr}^0,\mathsf{snr}^{\theta_{\min}-1}\right)\times\mathsf{snr}^{1-\min(\alpha_{b,\min},\theta_{b,\min})}\right) \tag{261}$$
$$\doteq\max\left(\mathsf{snr}^0,\mathsf{snr}^{\min(1,\theta_{\min})-\min(\alpha_{b,\min},\theta_{b,\min})}\right). \tag{262}$$

2) **Case 2:** $(r,t)_{\alpha_{b,\min}}=(r,t)_{\theta_{b,\min}}$. This refers to the case where the indexes $(r,t)$ for which the minimum occurs are the same for both $\alpha_{b,r,t}$ and $\theta_{b,r,t}$.
   - **Case 2.1:** $\alpha_{b,\min}<\theta_{b,\min}$. We have that

$$\frac{\mathsf{snr}}{n_\mathrm{t}}\|\hat{\boldsymbol{H}}_b\|_F^2\doteq\frac{\mathsf{snr}}{n_\mathrm{t}}\mathsf{snr}^{-\hat{\alpha}_{b,\min}}\doteq\mathsf{snr}^{1-\alpha_{b,\min}}. \tag{263}$$

It follows that

$$e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_{\rm r}+\frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^B\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_{\rm t}}\right) \doteq \max\left(\mathsf{snr}^0,\mathsf{snr}^{\min(1,\theta_{\min})-\alpha_{b,\min}}\right). \quad (264)$$

- **Case 2.2:** $\alpha_{b,\min} > \theta_{b,\min}$. We have that

$$\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2 \doteq \frac{\mathsf{snr}}{n_{\rm t}}\mathsf{snr}^{-\hat{\alpha}_{b,\min}} \doteq \mathsf{snr}^{1-\theta_{b,\min}}. \quad (265)$$

If we have $\theta_{\min} < 1$, the high-SNR exponential equality can be evaluated as follows:

$$\begin{aligned} &e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_{\rm r}+\frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^B\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_{\rm t}}\right) \\ &\doteq \max\left(\mathsf{snr}^0,\mathsf{snr}^{\theta_{\min}-\theta_{b,\min}}\right) \quad (266)\\ &\doteq \mathsf{snr}^0 \quad (267)\end{aligned}$$

where the last dot equality follows from the condition $\theta_{\min} \le \theta_{b,\min}$. For $\theta_{\min} \ge 1$, we have that

$$\begin{aligned} &e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_{\rm r}+\frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^B\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_{\rm t}}\right) \\ &\doteq \max\left(\mathsf{snr}^0,\mathsf{snr}^{1-\theta_{b,\min}}\right) \quad (268)\\ &\doteq \mathsf{snr}^0 \quad (269)\end{aligned}$$

where the last dot equality is due to $\theta_{b,\min} \ge \theta_{\min}$.

- **Case 2.3:** $\alpha_{b,\min} = \theta_{b,\min}$. From (242), (245) and (248), if $h_{b,\min} = e_{b,\min}$, $h_{b,\min} \neq 0$ or $h_{b,\min} = e^*_{b,\min}$, $\cos(\phi^h_{b,\min}) \neq 0$ or $-h_{b,\min} = e^*_{b,\min}$, $\sin(\phi^h_{b,\min}) \neq 0$, then we observe the same convergence results as in **case 2.2**. Otherwise, we have from (249) that $\mathsf{snr}^{-\hat{\alpha}_{b,\min}} = 0$ and hence

$$e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_{\rm r}+\frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^B\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_{\rm t}}\right) \doteq \mathsf{snr}^0. \quad (270)$$

Note that the results in **cases 2.2 and 2.3** are identical.

Summarizing from the above two cases, we have that

$$e^{-1}\left(1+\frac{B\|\hat{\boldsymbol{H}}_b\|_F^2}{Bn_{\rm r}+\frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^B\|\boldsymbol{E}_b\|_F^2}\frac{\mathsf{snr}}{n_{\rm t}}\right) \doteq \mathsf{snr}^{[\min(1,\theta_{\min})-\alpha_{b,\min}]_+}. \quad (271)$$

Let the data rate $R(\mathsf{snr})$ satisfying the exponential equality $e^{R(\mathsf{snr})} \doteq \mathsf{snr}^k$, where $k$ is the multiplexing gain [11] ($k$ tends to zero for fixed rate transmission). Then, from (237) and (271), we can bound $P_{\rm gout}(R)$ as follows:

$$\begin{aligned} &P_{\rm gout}(R)\\ &= \Pr\{I^{\rm gmi}(\hat{\mathsf{H}}) < R(\mathsf{snr})\} \quad (272)\\ &\doteq \mathsf{snr}^{-d^G_{\rm icsir}} \quad (273)\\ &\dot{\le} \Pr\left\{\frac{1}{B}\sum_{b=1}^B[\min(1,\theta_{\min})-\alpha_{b,\min}]_+ < k\right\} \quad (274)\\ &\doteq \int_{\underline{\mathcal{O}}_G} p_{\mathsf{A},\Phi^{\mathsf{H}}}(\boldsymbol{A},\boldsymbol{\Phi}^{\boldsymbol{H}})p_{\Theta}(\boldsymbol{\Theta})p_{\Phi^{\mathsf{E}}}(\boldsymbol{\Phi}^{\boldsymbol{E}})d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{H}}d\boldsymbol{\Phi}^{\boldsymbol{E}} \quad (275)\end{aligned}$$

where we have defined

$$\underline{\mathcal{O}}_G \triangleq \left\{\boldsymbol{A},\boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm r}\times n_{\rm t}} : \sum_{b=1}^B[\min(1,\theta_{\min})-\alpha_{b,\min}]_+ < Bk\right\}. \quad (276)$$

Applying the result in Lemma 3, it is not difficult to show that

$$d^G_{\rm icsir} \ge \inf_{\underline{\mathcal{O}}_G\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^B\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}\alpha_{b,r,t} + \sum_{b=1}^B\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}(\theta_{b,r,t}-d_e)\right\}. \quad (277)$$

Since increasing $\theta_{\min}$ increases both the infimum function and the LHS of the constraint, the optimum $\theta_{\min}$ is given by $\theta^*_{\min} = d_e$. Since $\theta_{b,r,t} \ge \theta_{\min}$, the infimum solution is given by $\theta^*_{b,r,t} = d_e$ for all $b = 1,\ldots,B, r = 1,\ldots,n_{\rm r}$ and $t = 1,\ldots,n_{\rm t}$. On the other hand, the infimum solution of $\boldsymbol{A}$ is given by the intersection of the region defined by $\sum_{b=1}^B \alpha_{b,\min} > B(\min(1,d_e)-k)$ and the region defined by $\alpha_{b,\min} \in [0,\min(1,d_e)]$. Due to the fact that $\alpha_{b,r,t} \ge \alpha_{b,\min}$, for all $r = 1,\ldots,n_{\rm r}$ and $t = 1,\ldots,n_{\rm t}$, the solution to the above infimum is given by

$$d^G_{\rm icsir} \ge \left(1+\frac{\tau}{2}\right)Bn_{\rm t}n_{\rm r}\times(\min(1,d_e)-k) \quad (278)$$

for $k \in [0,\min(1,d_e)]$ and zero otherwise. For a fixed coding rate ($k = 0$), we have that

$$d^G_{\rm icsir} \ge \min(1,d_e)\times\left(1+\frac{\tau}{2}\right)Bn_{\rm t}n_{\rm r}. \quad (279)$$

*2) GMI Upper Bound:* The expectation $\mathbb{E}\left[s\mathsf{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]$ can be evaluated as follows:

$$\begin{aligned} &\mathbb{E}\left[s\mathsf{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]\\ &=\int_{\boldsymbol{x},\boldsymbol{y}} s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}P_{\mathsf{x}}(\boldsymbol{x})P_{\mathsf{y}|\mathsf{x},\mathsf{H}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{H}_b)d\boldsymbol{x}d\boldsymbol{y} \quad (280)\\ &=\int_{\boldsymbol{x},\boldsymbol{y}} s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}\cdot\frac{1}{\pi^{n_{\rm r}}}e^{-\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2}\cdot\frac{1}{\pi^{n_{\rm t}}}e^{-\|x\|_F^2}d\boldsymbol{x}d\boldsymbol{y} \quad (281)\\ &=\int_{\boldsymbol{y}} s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}\cdot\left(\int_{\boldsymbol{x}}\frac{1}{\pi^{n_{\rm r}}}e^{-\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2}\cdot\frac{1}{\pi^{n_{\rm t}}}e^{-\|x\|_F^2}d\boldsymbol{x}\right)d\boldsymbol{y} \quad (282)\\ &=\int_{\boldsymbol{y}}\frac{s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}}{\pi^{n_{\rm r}}\det\left(\frac{\mathsf{snr}}{n_{\rm t}}\boldsymbol{H}_b\boldsymbol{H}_b^\dagger+\boldsymbol{I}_{n_{\rm r}}\right)}e^{-\boldsymbol{y}^\dagger\Sigma_{\boldsymbol{y}}^{-1}\boldsymbol{y}}d\boldsymbol{y} \quad (283)\end{aligned}$$

where

$$\Sigma_{\boldsymbol{y}} = \boldsymbol{I}_{n_{\rm r}} + \frac{\mathsf{snr}}{n_{\rm t}}\boldsymbol{H}_b\boldsymbol{H}_b^\dagger \quad (284)$$

is a positive semi-definite matrix. Let $\boldsymbol{y} = \boldsymbol{Q}\tilde{\boldsymbol{y}}$, where $\boldsymbol{Q}$ is a unitary matrix diagonalising $\Sigma_{\boldsymbol{y}}^{-1}$. Then, $\tilde{\mathsf{y}}$ is a Gaussian random vector with zero mean and covariance matrix given by the diagonal matrix $\boldsymbol{Q}^\dagger \Sigma_{\boldsymbol{y}} \boldsymbol{Q}$. We have that [7]

$$\boldsymbol{y}^\dagger \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{y} = \tilde{\boldsymbol{y}}^\dagger \boldsymbol{Q}^\dagger \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{Q} \tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}^\dagger \Delta \tilde{\boldsymbol{y}} = \sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}} \quad (285)$$

where $\lambda_{b,i}$ is the $i$-th eigenvalue of $\boldsymbol{H}_b \boldsymbol{H}_b^\dagger$, and $\Delta$ is a diagonal matrix with diagonal elements given by $(1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i})^{-1}$, $i = 1,\ldots,n_\mathrm{r}$. Since $\hat{\Sigma}_{\boldsymbol{y}}^{-1}$ is a Hermitian matrix, we can apply the eigen-decomposition [27] such that

$$\hat{\Sigma}_{\boldsymbol{y}}^{-1} = \hat{\boldsymbol{Q}} \hat{\Delta} \hat{\boldsymbol{Q}}^\dagger \Leftrightarrow \hat{\Delta} = \hat{\boldsymbol{Q}}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \hat{\boldsymbol{Q}} \quad (286)$$

where $\hat{\boldsymbol{Q}}$ is another unitary matrix and $\hat{\Delta}$ is another diagonal matrix obtained by diagonalising $\hat{\Sigma}_{\boldsymbol{y}}^{-1}$. Let $\hat{\lambda}_{b,i}$ be the $i$-th eigenvalue of $\hat{\boldsymbol{H}}_b \hat{\boldsymbol{H}}_b^\dagger$, then the diagonal entries of $\hat{\Delta}$ are given by $(1+s\frac{\mathsf{snr}}{n_\mathrm{t}}\hat{\lambda}_{b,i})^{-1}$ for all $i = 1,\ldots,n_\mathrm{r}$. Applying this to $\boldsymbol{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \boldsymbol{y}$, we have that

$$\begin{aligned}\boldsymbol{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \boldsymbol{y} = \tilde{\boldsymbol{y}}^\dagger \boldsymbol{Q}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \boldsymbol{Q} \tilde{\boldsymbol{y}} &= \tilde{\boldsymbol{y}}^\dagger \boldsymbol{Q}^\dagger \hat{\boldsymbol{Q}} \hat{\Delta} \hat{\boldsymbol{Q}}^\dagger \boldsymbol{Q} \tilde{\boldsymbol{y}} && (287)\\ &= \tilde{\boldsymbol{y}}^\dagger \boldsymbol{V} \hat{\Delta} \boldsymbol{V}^\dagger \tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}^\dagger \hat{\Sigma}_{\tilde{\boldsymbol{y}}}^{-1} \tilde{\boldsymbol{y}} && (288)\end{aligned}$$

where $\boldsymbol{V} = \boldsymbol{Q}^\dagger \hat{\boldsymbol{Q}}$ is also a unitary matrix and $\hat{\Sigma}_{\tilde{\boldsymbol{y}}}^{-1}$ is a Hermitian matrix. Then, we have that

$$\begin{aligned}&\mathbb{E}\left[ s \mathsf{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \mathsf{y} \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b \right]\\ &= \frac{s}{\pi^{n_\mathrm{r}} \det\left(\frac{\mathsf{snr}}{n_\mathrm{t}} \boldsymbol{H}_b \boldsymbol{H}_b^\dagger + \boldsymbol{I}_{n_\mathrm{r}}\right)} \int_{\boldsymbol{y}} \boldsymbol{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \boldsymbol{y} e^{-\boldsymbol{y}^\dagger \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{y}} d\boldsymbol{y} && (289)\\ &= \frac{s}{\pi^{n_\mathrm{r}} \det\left(\frac{\mathsf{snr}}{n_\mathrm{t}} \boldsymbol{H}_b \boldsymbol{H}_b^\dagger + \boldsymbol{I}_{n_\mathrm{r}}\right)} \cdot \int_{\tilde{\boldsymbol{y}}} \left(\tilde{\boldsymbol{y}}^\dagger \boldsymbol{V} \hat{\Delta} \boldsymbol{V}^\dagger \tilde{\boldsymbol{y}}\right) e^{-\sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}}} d\tilde{\boldsymbol{y}}. && (290)\end{aligned}$$

[7]Without loss of generality, herein we assume $n_\mathrm{t} \geq n_\mathrm{r}$.

Let $\sigma_{i,j}$, $i = 1,\ldots,n_\mathrm{r}$, $j = 1,\ldots,n_\mathrm{r}$, be the entry of the matrix $\hat{\Sigma}_{\tilde{\boldsymbol{y}}}^{-1}$ at row $i$ and column $j$. Then, the integral in (290) can evaluated as in (291)–(294), as shown at the bottom of the page. Herein, $\tilde{y}_i^*$ and $v_{i,j}^*$ denote the complex conjugates of $\tilde{y}_i$ and $v_{i,j}$, respectively, and $\Re\{\cdot\}$ denotes the real part of a complex number. We have that from (291)

$$\sigma_{i,i} = \sum_{j=1}^{n_\mathrm{r}} \frac{|v_{i,j}|^2}{1+s\frac{\mathsf{snr}}{n_\mathrm{t}}\hat{\lambda}_{b,j}} \leq \sum_{j=1}^{n_\mathrm{r}} |v_{i,j}|^2 = 1 \quad (295)$$

where the inequality is because $\hat{\lambda}_{b,j}$ is nonnegative; the last equality is because $v_{i,j}$ is an element of a unitary matrix and the sum of $|v_{i,j}|^2$ over $j = 1,\ldots,n_\mathrm{r}$ is equal to one. Finally, we have that

$$\begin{aligned}&\mathbb{E}\left[ s \mathsf{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1} \mathsf{y} \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b \right]\\ &= s \sum_{i=1}^{n_\mathrm{r}} \sigma_{i,i} \left(1 + \frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}\right). && (296)\end{aligned}$$

Let $s^*$ be the optimizing $s$ that gives the supremum in the RHS of (36) for Gaussian inputs. We then apply Proposition 3.

$$\begin{aligned}&\int_{\tilde{\boldsymbol{y}}} \left(\tilde{\boldsymbol{y}}^\dagger \boldsymbol{V} \hat{\Delta} \boldsymbol{V}^\dagger \tilde{\boldsymbol{y}}\right) e^{-\sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}}} d\tilde{\boldsymbol{y}}\\ &= \int_{\tilde{\boldsymbol{y}}} [\tilde{y}_1^* \cdots \tilde{y}_{n_\mathrm{r}}^*]\\ &\quad \times \begin{pmatrix} v_{1,1} & \cdots & v_{1,n_\mathrm{r}} \\ \vdots & \ddots & \vdots \\ v_{n_\mathrm{r},1} & \cdots & v_{n_\mathrm{r},n_\mathrm{r}} \end{pmatrix} \begin{pmatrix} \frac{1}{1+s\frac{\mathsf{snr}}{n_\mathrm{t}}\hat{\lambda}_{b,1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{1+s\frac{\mathsf{snr}}{n_\mathrm{t}}\hat{\lambda}_{b,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{1+s\frac{\mathsf{snr}}{n_\mathrm{t}}\hat{\lambda}_{b,n_\mathrm{r}}} \end{pmatrix} \begin{pmatrix} v_{1,1}^* & \cdots & v_{n_\mathrm{r},1}^* \\ \vdots & \ddots & \vdots \\ v_{1,n_\mathrm{r}}^* & \cdots & v_{n_\mathrm{r},n_\mathrm{r}}^* \end{pmatrix} \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_{n_\mathrm{r}} \end{bmatrix}\\ &\quad \times e^{-\sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}}} d\tilde{\boldsymbol{y}} && (291)\\ &= \int_{\tilde{\boldsymbol{y}}} \left(\sum_{i=1}^{n_\mathrm{r}} \sigma_{i,i}|\tilde{y}_i|^2 + 2\sum_{i=1}^{n_\mathrm{r}} \sum_{j>i}^{n_\mathrm{r}} \Re\left(\sigma_{i,j}\tilde{y}_i^* \tilde{y}_j\right)\right) e^{-\sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}}} d\tilde{\boldsymbol{y}} && (292)\\ &= \sum_{i=1}^{n_\mathrm{r}} \int_{\tilde{\boldsymbol{y}}} \sigma_{i,i}|\tilde{y}_i|^2 e^{-\sum_{i=1}^{n_\mathrm{r}} \frac{|\tilde{y}_i|^2}{1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}}} d\tilde{\boldsymbol{y}} + 0 && (293)\\ &= \pi^{n_\mathrm{r}} \times \det\left(\frac{\mathsf{snr}}{n_\mathrm{t}} \boldsymbol{H}_b \boldsymbol{H}_b^\dagger + \boldsymbol{I}_{n_\mathrm{r}}\right) \times \sum_{i=1}^{n_\mathrm{r}} \sigma_{i,i}\left(1+\frac{\mathsf{snr}}{n_\mathrm{t}}\lambda_{b,i}\right) && (294)\end{aligned}$$

Since $s$, $\sigma_{i,i}$ and $\hat{\lambda}_{b,i}$ are all nonnegative, using (295) we have the upper bound to $I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s^*)$ as

$$\begin{aligned} & I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s^*) \\ & \le \sup_{s_b>0} \log\det\left(s_b \frac{\mathsf{snr}}{n_{\rm t}} \hat{\boldsymbol{H}}_b \hat{\boldsymbol{H}}_b^\dagger + \boldsymbol{I}_{n_{\rm r}}\right) \\ & \qquad - s_b\left(n_{\rm r} + \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) + s_b \sum_{i=1}^{n_{\rm r}}\left(1+\frac{\mathsf{snr}}{n_{\rm t}}\lambda_{b,i}\right) \quad (297) \\ & = \sup_{s_b>0} \log\prod_{i=1}^{n_{\rm r}}\left(1 + s_b\frac{\mathsf{snr}}{n_{\rm t}}\hat{\lambda}_{b,i}\right) - s_b\left(n_{\rm r} + \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) \\ & \qquad + s_b\left(n_{\rm r} + \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2\right) \quad (298) \\ & \le \sup_{s_b>0} n_{\rm r}\log\left(1 + s_b\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2\right) \\ & \qquad + s_b\left(\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2 - \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) \quad (299) \end{aligned}$$

where the last inequality is due to the fact that $\sum_i \hat{\lambda}_{b,i} = \|\hat{\boldsymbol{H}}_b\|_F^2$, thus each $\hat{\lambda}_{b,i}$ is upper-bounded by $\|\hat{\boldsymbol{H}}_b\|_F^2$. If $\mathsf{snr}\|\boldsymbol{H}_b\|_F^2$ is greater than or equal to $\mathsf{snr}\|\boldsymbol{E}_b\|_F^2$, then the supremum in the RHS of (299) is achieved with $s_b \uparrow \infty$ because the RHS of the last inequality is a strictly increasing function of $s_b$. However, using the data-processing inequality (Proposition 4), we can always bound $I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s^*)$ with the perfect-CSIR bound

$$\begin{aligned} I_b^{\rm gmi}(\mathsf{snr}, \boldsymbol{H}_b, \hat{\boldsymbol{H}}_b, s^*) & \le \log\det\left(\frac{\mathsf{snr}}{n_{\rm t}}\boldsymbol{H}_b\boldsymbol{H}_b^\dagger + \boldsymbol{I}_{n_{\rm r}}\right) \quad (300) \\ & \le n_{\rm r}\log\left(1 + \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2\right). \quad (301) \end{aligned}$$

On the other hand, if $\mathsf{snr}\|\boldsymbol{H}_b\|_F^2$ is less than $\mathsf{snr}\|\boldsymbol{E}_b\|_F^2$, the supremum is achieved with $s_b^*$ given by

$$s_b^* = \max\left(0, \frac{n_{\rm r}}{\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2 - \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2} - \frac{1}{\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2}\right). \quad (302)$$

The above $s_b^*$ is obtained from the solution of the first order derivative of

$$n_{\rm r}\log\left(1 + s_b\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2\right) + s_b\left(\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2 - \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) \quad (303)$$

with respect to $s_b$ when the derivative is equal to zero. The interval $s_b > 0$ yields the max function in the RHS of (302).

We continue the analysis by using the change of random variables as used in the GMI lower bound. The condition $\mathsf{snr}\|\boldsymbol{H}_b\|_F^2 \ge \mathsf{snr}\|\boldsymbol{E}_b\|_F^2$ for the perfect-CSIR bound implies that at high SNR, we have the following exponential inequality

$$\mathsf{snr}^{1-\alpha_{b,\min}} \dot{\ge} \mathsf{snr}^{1-\theta_{b,\min}} \Leftrightarrow \alpha_{b,\min} \le \theta_{b,\min}. \quad (304)$$

We then have the following asymptotic upper bound characterizations.

1) **Case 1:** $\alpha_{b,\min} \le \theta_{b,\min}$. From the perfect-CSIR bound (301), we have that

$$1 + \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2 \doteq \max\left(\mathsf{snr}^0, \mathsf{snr}^{1-\alpha_{b,\min}}\right). \quad (305)$$

2) **Case 2:** $\alpha_{b,\min} > \theta_{b,\min}$. If $\frac{1}{\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2}$ is greater than or equal to $\frac{n_{\rm r}}{\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2 - \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2}$, we have $s_b^* \downarrow 0$. From the RHS of (299), this yields

$$\exp\left(\frac{s_b^*\mathsf{snr}}{n_{\rm r}n_{\rm t}}\|\boldsymbol{H}_b\|_F^2 - \frac{s_b^*\mathsf{snr}}{n_{\rm r}n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) \times \left(1 + s_b^*\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2\right) \doteq \mathsf{snr}^0. \quad (306)$$

Otherwise, we have

$$s_b^* = \frac{n_{\rm r}}{\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2 - \frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{H}_b\|_F^2} - \frac{1}{\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2} \quad (307)$$

and this also yields

$$\exp\left(\frac{s_b^*\mathsf{snr}}{n_{\rm r}n_{\rm t}}\|\boldsymbol{H}_b\|_F^2 - \frac{s_b^*\mathsf{snr}}{n_{\rm r}n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\right) \times \left(1 + s_b^*\frac{\mathsf{snr}}{n_{\rm t}}\|\hat{\boldsymbol{H}}_b\|_F^2\right) \doteq \mathsf{snr}^0. \quad (308)$$

Let the data rate $R(\mathsf{snr})$ satisfying the exponential equality $e^{R(\mathsf{snr})} \doteq \mathsf{snr}^k$, where $k$ is the multiplexing gain [11] as used in the GMI lower bound. From the above cases, we have the bound for $P_{\rm gout}(R)$ as follows:

$$\begin{aligned} & P_{\rm gout}(R) \\ & \doteq \mathsf{snr}^{-d_{\rm icsir}^G} \quad (309) \\ & \dot{\ge} \Pr\left\{\frac{1}{B}\sum_{b=1}^{B} n_{\rm r}\left[1-\alpha_{b,\min}\right]_+ \cdot \mathbb{1}\{\alpha_{b,\min} \le \theta_{b,\min}\} < k\right\} \quad (310) \\ & \doteq \int_{\overline{\mathcal{O}}_G} p_{\mathsf{A},\Phi^{\mathsf{H}}}(\boldsymbol{A}, \boldsymbol{\Phi}^{\boldsymbol{H}}) p_{\Theta}(\boldsymbol{\Theta}) p_{\Phi^{\mathsf{E}}}(\boldsymbol{\Phi}^{\boldsymbol{E}}) d\boldsymbol{A} d\boldsymbol{\Theta} d\boldsymbol{\Phi}^{\boldsymbol{H}} d\boldsymbol{\Phi}^{\boldsymbol{E}} \quad (311) \end{aligned}$$

where we have defined

$$\overline{\mathcal{O}}_G \triangleq \left\{\boldsymbol{A}, \boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm r}\times n_{\rm t}} : \sum_{b=1}^{B}\left\{\left[1-\alpha_{b,\min}\right]_+ \cdot \mathbb{1}\{\alpha_{b,\min} \le \theta_{b,\min}\}\right\} < \frac{Bk}{n_{\rm r}}\right\}. \quad (312)$$

Thus, using the result in Lemma 3 to find the SNR exponent and following the same steps used for the GMI lower bound, it is not difficult to prove that

$$d_{\rm icsir}^G(k) \le \left(1 + \frac{\tau}{2}\right) B n_{\rm t} n_{\rm r} \min\left(1 - \frac{k}{n_{\rm r}}, d_e\right). \quad (313)$$

For fixed rate transmission ($k = 0$), we obtain

$$d_{\rm icsir}^G(k=0) \le \min(1, d_e)\left(1 + \frac{\tau}{2}\right) B n_{\rm t} n_{\rm r}. \quad (314)$$

This proves part of Theorem 1 on the converse for Gaussian inputs.

## APPENDIX E
## GAUSSIAN INPUTS ACHIEVABILITY PROOF

Recall that from (35), the generalized Gallager function for MIMO channels can be written as (in natural-base log)

$$E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) = -\log \mathbb{E}\left[\left(\int_{\boldsymbol{x}'} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\boldsymbol{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^s d\boldsymbol{x}'\right)^\rho \middle| \mathsf{H}_b=\boldsymbol{H}_b, \mathsf{E}_b=\boldsymbol{E}_b\right]. \quad (315)$$

Evaluating the inner expectation over $\mathsf{x}'$ for a given $\mathsf{y}=\boldsymbol{y}$, $\mathsf{x}=\boldsymbol{x}$, $\mathsf{H}_b=\boldsymbol{H}_b$ and $\mathsf{E}_b=\boldsymbol{E}_b$, we have that

$$\int_{\boldsymbol{x}'} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\boldsymbol{y}|\boldsymbol{x}',\hat{\boldsymbol{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\boldsymbol{y}|\boldsymbol{x},\hat{\boldsymbol{H}}_b)}\right)^s d\boldsymbol{x}' = e^{s\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2}\cdot\frac{e^{-s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}}}{\det\left(s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger\frac{\mathsf{snr}}{n_\mathrm{t}}+\boldsymbol{I}_{n_\mathrm{r}}\right)} \quad (316)$$

where

$$\hat{\Sigma}_{\boldsymbol{y}} = s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger\frac{\mathsf{snr}}{n_\mathrm{t}}+\boldsymbol{I}_{n_\mathrm{r}}. \quad (317)$$

Then, the expectation over $(\mathsf{x},\mathsf{y})$ is given as

$$\mathbb{E}\left[\left(\int_{\boldsymbol{x}'} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\boldsymbol{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^s d\boldsymbol{x}'\right)^\rho \middle| \mathsf{H}_b=\boldsymbol{H}_b, \mathsf{E}_b=\boldsymbol{E}_b\right] = \frac{\mathbb{E}\left[e^{\rho s\left\|\mathsf{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\mathsf{x}\right\|_F^2}\times e^{-\rho s\mathsf{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\mathsf{y}}\middle|\mathsf{H}_b=\boldsymbol{H}_b,\mathsf{E}_b=\boldsymbol{E}_b\right]}{\det\left(s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger\frac{\mathsf{snr}}{n_\mathrm{t}}+\boldsymbol{I}_{n_\mathrm{r}}\right)^\rho}. \quad (318)$$

We can evaluate the expectation operation as follows. For an arbitrary function $f(\boldsymbol{x},\boldsymbol{y})$, the expectation over $(\mathsf{x},\mathsf{y})$ is given by

$$\mathbb{E}\left[f(\mathsf{x},\mathsf{y})\right] = \int_{\boldsymbol{x},\boldsymbol{y}} f(\boldsymbol{x},\boldsymbol{y})\,P_{\mathsf{x}}(\boldsymbol{x})\,P_{\mathsf{y}|\mathsf{x}}(\boldsymbol{y}|\boldsymbol{x})d\boldsymbol{y}d\boldsymbol{x}. \quad (319)$$

We first apply the integration over $\boldsymbol{y}$

$$\int_{\boldsymbol{y}}\left(e^{\rho s\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2}\cdot e^{-\rho s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}}\right)\cdot\frac{1}{\pi^{n_\mathrm{r}}}e^{-\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2}d\boldsymbol{y}. \quad (320)$$

Using $\tilde{\boldsymbol{y}}=\hat{\boldsymbol{Q}}^\dagger\boldsymbol{y}$, we have that

$$\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y} = \tilde{\boldsymbol{y}}^\dagger\hat{\boldsymbol{Q}}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\hat{\boldsymbol{Q}}\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{y}}^\dagger\hat{\Delta}\tilde{\boldsymbol{y}} = \sum_{i=1}^{n_\mathrm{r}}\frac{|\tilde{y}_i|^2}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}} \quad (321)$$

where $\hat{\boldsymbol{Q}}$ is a unitary matrix similar to the one defined in Appendix D, and where we have assumed $n_\mathrm{t}\geq n_\mathrm{r}$ without loss of generality. Note that

$$\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2 = \left\|\hat{\boldsymbol{Q}}\tilde{\boldsymbol{y}}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2 \quad (322)$$

$$= \left\|\tilde{\boldsymbol{y}}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{Q}}^\dagger\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2 \quad (323)$$

because multiplication with the unitary matrix does not affect the Euclidean norm of a vector. Therefore, we have that

$$\int_{\boldsymbol{y}}\left(e^{\rho s\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2}\cdot e^{-\rho s\boldsymbol{y}^\dagger\hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}}\cdot\frac{1}{\pi^{n_\mathrm{r}}}e^{-\left\|\boldsymbol{y}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2}\right)d\boldsymbol{y}$$

$$= \int_{\tilde{\boldsymbol{y}}}\left(e^{\rho s\left\|\tilde{\boldsymbol{y}}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{Q}}^\dagger\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2}\cdot e^{-\rho s\tilde{\boldsymbol{y}}^\dagger\hat{\Delta}\tilde{\boldsymbol{y}}}\cdot\frac{1}{\pi^{n_\mathrm{r}}}e^{-\left\|\tilde{\boldsymbol{y}}-\sqrt{\frac{\mathsf{snr}}{n_\mathrm{t}}}\hat{\boldsymbol{Q}}^\dagger\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2}\right)d\tilde{\boldsymbol{y}} \quad (324)$$

$$\leq \frac{1}{\left(1-\frac{\rho s}{1-\rho s}\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2\right)^{n_\mathrm{t}}}\cdot\prod_{i=1}^{n_\mathrm{r}}\left(\frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\rho s)}\right) \quad (325)$$

$$= \left(\frac{1-\rho s}{1-\rho s-\rho s\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2}\right)^{n_\mathrm{t}}\cdot\prod_{i=1}^{n_\mathrm{r}}\left(\frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\rho s)}\right) \quad (326)$$

where the inequality in (325) is proved in Appendix F. Note that the result in (325) requires $\rho s<1$ and

$$s \leq \frac{u_1}{u_2+\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2} \quad (327)$$

–where $u_1$ and $u_2$ are some positive constants- so that the integral can be evaluated (see Appendix F). Following this, we have that

$$\mathbb{E}\left[\left(\int_{\boldsymbol{x}'} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\boldsymbol{x}',\hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}_b)}\right)^s d\boldsymbol{x}'\right)^\rho \middle| \mathsf{H}_b=\boldsymbol{H}_b, \mathsf{E}_b=\boldsymbol{E}_b\right] \leq \frac{1}{\det\left(s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger\frac{\mathsf{snr}}{n_\mathrm{t}}+\boldsymbol{I}_{n_\mathrm{r}}\right)^\rho}\cdot\left(\frac{1-\rho s}{1-\rho s-\rho s\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2}\right)^{n_\mathrm{t}}\cdot\prod_{i=1}^{n_\mathrm{r}}\left(\frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\rho s)}\right) \quad (328)$$

and from (315)

$$E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) \geq \left(\sum_{i=1}^{n_\mathrm{r}}(\rho-1)\log\left(1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}\right)\right)-n_\mathrm{t}\log(1-\rho s) + n_\mathrm{t}\log\left(1-\rho s-\rho s\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2\right) + \sum_{i=1}^{n_\mathrm{r}}\log\left(1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\rho s)\right). \quad (329)$$

Note that the random coding error exponent is given by

$$E_r^Q(R,\hat{\boldsymbol{H}}) = \sup_{\substack{s>0\\0\le\rho\le 1}} \frac{1}{B}\sum_{b=1}^{B} E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b) - \rho R. \tag{330}$$

The lower bound for $E_r^Q(R,\hat{\boldsymbol{H}})$ is obtained by replacing $E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)$ above with the RHS of (329) and given in (331), as shown at the bottom of the page. Note that from (331), we require $\rho s < 1$ and $\rho s + \rho s\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2 < 1$ for all $b=1,\ldots,B$ so that the logarithm functions are defined. Note that the following choices of $\rho = \underline{\rho} = 1$ and

$$s = \underline{s} = \frac{1}{n_\mathrm{r}\left(1 + \frac{\mathsf{snr}}{n_\mathrm{t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2\right)} \tag{332}$$

ensure that the logarithm functions in (331) are always defined. Since $\underline{\rho}\underline{s}$ and $\underline{\rho}\underline{s} + \underline{\rho}\underline{s}\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2$ are always bounded by some real-valued constants in the interval $[0,1]$, we have that for $\mathsf{snr} \ge 0$

$$\begin{aligned} &-n_\mathrm{t}\log(1-\underline{\rho}\underline{s}) + n_\mathrm{t}\log\left(1 - \underline{\rho}\underline{s} - \underline{\rho}\underline{s}\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2\right)\\ &\quad\ge n_\mathrm{t}\log\left(1-\frac{1}{n_\mathrm{r}}\right) \triangleq u_3 \end{aligned} \tag{333}$$

where $u_3 < 0$.

Note that choosing specific values of $\rho$ and $s$ further lower-bounds (331). We continue the analysis by following the same technique used in Appendix D. Since $E_r^Q(R,\hat{\boldsymbol{H}})$ can also be lower-bounded by 0 (i.e., $\rho=0$), we have that by substituting $s$ and $\rho$ to $\underline{\rho}$ and $\underline{s}$, the lower bounds in (334) and (335), as shown at the bottom of the page. Note that the inequality in (335) is due to the lower-bounding technique in (230). Define $\underline{E}_r^Q(R,\hat{\boldsymbol{H}})$ as the RHS of (335). Following the high-SNR analysis in the GMI lower bound (**cases 1** and **2**), we obtain the dot equality

$$e^{u_3}\cdot\left(1 + \underline{s}\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\underline{s})\right) \doteq \mathsf{snr}^{[\min(1,\theta_{\min})-\alpha_{b,\min}]_+}. \tag{336}$$

Recall the rate and multiplexing gain relationship $e^{R(\mathsf{snr})} \doteq \mathsf{snr}^k$ as in (4). It follows from $\underline{E}_r^Q(R,\hat{\boldsymbol{H}})$ in the RHS of (335), (336) and the dot equality $e^{R(\mathsf{snr})} \doteq \mathsf{snr}^k$ that for high SNR, if the following event:

$$\mathcal{A}_G = \left\{\boldsymbol{A},\boldsymbol{\Theta} \in \mathbb{R}^{B\times n_\mathrm{r}\times n_\mathrm{t}} : \sum_{b=1}^{B}[\min(1,\theta_{\min}) - \alpha_{b,\min}]_+ \le Bk\right\} \tag{337}$$

occurs, then $\underline{E}_r^Q(R,\hat{\boldsymbol{H}}) = 0$ and if the complementary event

$$\mathcal{A}_G^c = \left\{\boldsymbol{A},\boldsymbol{\Theta} \in \mathbb{R}^{B\times n_\mathrm{r}\times n_\mathrm{t}} : \sum_{b=1}^{B}[\min(1,\theta_{\min}) - \alpha_{b,\min}]_+ > Bk\right\} \tag{338}$$

occurs, then $\underline{E}_r^Q(R,\hat{\boldsymbol{H}}) > 0$. Therefore, the upper bound to the average error probability (for Gaussian random codes that satisfy the generalized Gallager bound) is given as

$$P_{e,\mathrm{ave}} \le \mathbb{E}\left[e^{-LB\underline{E}_r^Q(R,\hat{\mathsf{H}})}\right] \tag{339}$$

$$\begin{aligned} &\dot{\le} \int_{\mathcal{A}_G\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}} \Big(\mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\alpha_{b,r,t}}\\ &\qquad\times\mathsf{snr}^{\sum_{b=1}^{B}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}(d_e-\theta_{b,r,t})} d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{H}}d\boldsymbol{\Phi}^{\boldsymbol{E}}\Big)\\ &+\int_{\mathcal{A}_G^c\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}} \Big(\mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\alpha_{b,r,t}}\\ &\qquad\times\mathsf{snr}^{\sum_{b=1}^{B}\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}(d_e-\theta_{b,r,t})}\\ &\qquad\times\mathsf{snr}^{-L\left(\sum_{b=1}^{B}[\min(1,\theta_{\min})-\alpha_{b,\min}]_+-Bk\right)} d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{H}}d\boldsymbol{\Phi}^{\boldsymbol{E}}\Big) \end{aligned} \tag{340}$$

$$\doteq K_1\mathsf{snr}^{-d_1(k)} + K_2\mathsf{snr}^{-d_2(k)} \tag{341}$$

$$\doteq \mathsf{snr}^{-d_2(k)} \tag{342}$$

$$\begin{aligned} E_r^Q(R,\hat{\boldsymbol{H}}) \ge \sup_{\substack{s>0\\0\le\rho\le 1}} \frac{1}{B}\Bigg\{\sum_{b=1}^{B}\Bigg(\Bigg(\sum_{i=1}^{n_\mathrm{r}}(\rho-1)\log\left(1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}\right)\Bigg) - n_\mathrm{t}\log(1-\rho s) + n_\mathrm{t}\log\left(1-\rho s-\rho s\frac{\mathsf{snr}}{n_\mathrm{t}}\|\boldsymbol{E}_b\|_F^2\right)\\ + \sum_{i=1}^{n_\mathrm{r}}\log\left(1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\rho s)\right)\Bigg) - \rho Bk\log\mathsf{snr}\Bigg\} \end{aligned} \tag{331}$$

$$\begin{aligned} E_r^Q(R,\hat{\boldsymbol{H}}) &\ge \left[\frac{1}{B}\sum_{b=1}^{B}\left(u_3 + \sum_{i=1}^{n_\mathrm{r}}(\underline{\rho}-1)\log\left(1+\underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}\right) + \sum_{i=1}^{n_\mathrm{r}}\log\left(1+\underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\underline{\rho}\underline{s})\right)\right) - \underline{\rho}R\right]_+ \end{aligned} \tag{334}$$

$$\begin{aligned} &\ge \left[\frac{1}{B}\left\{\sum_{b=1}^{B}\log\left(e^{u_3}\cdot\left(1+\underline{s}\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_\mathrm{t}}(1-\underline{s})\right)\right) - BR\right\}\right]_+\\ &\triangleq \underline{E}_r^Q(R,\hat{\boldsymbol{H}}) \end{aligned} \tag{335}$$

where $K_1, K_2 \doteq \mathsf{snr}^0$, and where

$$d_1(k) = \left(1+\frac{\tau}{2}\right) B n_{\rm t} n_{\rm r} \times (\min(1,d_e)-k) \quad (343)$$

is the lower bound to the generalized outage SNR exponent achieved with infinite block length (notice that $\underline{\mathcal{O}}_G$ in Appendix D is similar to $\mathcal{A}_G$ except for the inequality $<$ which becomes $\le$ in $\mathcal{A}_G$) and

$$d_2(k) = \inf_{\mathcal{A}_G^c \cap \{\boldsymbol{A}\succeq \boldsymbol{0}, \boldsymbol{\Theta}\succeq d_e \times \boldsymbol{1}\}} \Bigg\{ \left(1+\frac{\tau}{2}\right) \sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} \alpha_{b,r,t} + \sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} (\theta_{b,r,t}-d_e) + L\left(\sum_{b=1}^{B}[\min(1,\theta_{\min})-\alpha_{b,\min}]_+ - Bk\right)\Bigg\}. \quad (344)$$

Since we need

$$L\left(\sum_{b=1}^{B}[\min(1,\theta_{\min})-\alpha_{b,\min}]_+ - Bk\right) > 0 \quad (345)$$

in the set $\mathcal{A}_G^c$ for $d_2(k)$, it is straightforward to deduce that $d_2(k) \le d_1(k)$. This follows from [11, Lemma 6]. Therefore, the SNR exponent lower bound for a given block length $L$ is given by $d_G^{\ell}(k) = d_2(k)$ in Theorem 2.

## Appendix F
## Proof of Inequality (325)

Basically, we want to evaluate the following expression:

$$\mathbb{E}\left[\int_{\boldsymbol{y}} \left( e^{\rho s \left\|\boldsymbol{y} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{H}}_b \mathsf{x}\right\|_F^2} \cdot e^{-\rho s \boldsymbol{y}^\dagger \hat{\Sigma}_{\boldsymbol{y}}^{-1}\boldsymbol{y}} \cdot \frac{1}{\pi^{n_{\rm r}}} e^{-\left\|\boldsymbol{y} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b \mathsf{x}\right\|_F^2} d\boldsymbol{y}\right) \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right]$$
$$= \mathbb{E}\left[\int_{\tilde{\boldsymbol{y}}} \left( e^{\rho s \left\|\tilde{\boldsymbol{y}} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger\hat{\boldsymbol{H}}_b \mathsf{x}\right\|_F^2} \cdot e^{-\rho s \tilde{\boldsymbol{y}}^\dagger \hat{\Delta}\tilde{\boldsymbol{y}}} \cdot \frac{1}{\pi^{n_{\rm r}}} e^{-\left\|\tilde{\boldsymbol{y}} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger\boldsymbol{H}_b \mathsf{x}\right\|_F^2} d\tilde{\boldsymbol{y}}\right) \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right] \quad (346)$$

where the above expectation is over $\mathsf{x}$. To simplify the process, we let $\boldsymbol{g} = \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger \hat{\boldsymbol{H}}_b \boldsymbol{x}$ and $\boldsymbol{c} = \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger \boldsymbol{H}_b \boldsymbol{x}$, $\boldsymbol{g}, \boldsymbol{c} \in \mathbb{C}^{n_{\rm r}\times 1}$. Then, expanding the argument in the exponential term for a given $\mathsf{x} = \boldsymbol{x}$, we have that

$$\rho s \left\|\tilde{\boldsymbol{y}} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger\hat{\boldsymbol{H}}_b \boldsymbol{x}\right\|_F^2 - \rho s \tilde{\boldsymbol{y}}^\dagger\hat{\Delta}\tilde{\boldsymbol{y}} - \left\|\tilde{\boldsymbol{y}} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger\boldsymbol{H}_b \boldsymbol{x}\right\|_F^2$$
$$= \rho s \sum_{i=1}^{n_{\rm r}} \left( |\tilde{y}_i - g_i|^2 - \frac{|\tilde{y}_i|^2}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}\right) - \sum_{i=1}^{n_{\rm r}} |\tilde{y}_i - c_i|^2. \quad (347)$$

Since $\tilde{y}_i$ is independent for $i=1,\ldots,n_{\rm r}$, we can integrate for one value of $i$ and generalize the result. By basic integration, we can easily obtain that

$$\int_{\tilde{y}_i} \frac{1}{\pi} e^{-|\tilde{y}_i - c_i|^2 - \rho s \frac{|\tilde{y}_i|^2}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}} + \rho s |\tilde{y}_i - g_i|^2} d\tilde{y}_i$$
$$= \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)} \cdot e^{-|c_i|^2 + \rho s |g_i|^2 + |\rho s g_i - c_i|^2 \cdot \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)}}. \quad (348)$$

Evaluating the integral for all $\tilde{y}_i$, $i=1,\ldots,n_{\rm r}$ yields

$$\left(\prod_{i=1}^{n_{\rm r}} \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)}\right) \cdot \exp\left( -\sum_{i=1}^{n_{\rm r}}|c_i|^2 + \rho s \sum_{i=1}^{n_{\rm r}}|g_i|^2 + \sum_{i=1}^{n_{\rm r}}\left(|\rho s g_i - c_i|^2 \cdot \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)}\right)\right). \quad (349)$$

Note that

$$\sum_{i=1}^{n_{\rm r}} |c_i|^2 = \left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger \boldsymbol{H}_b\boldsymbol{x}\right\|_F^2 = \frac{\mathsf{snr}}{n_{\rm t}}\left\|\boldsymbol{H}_b\boldsymbol{x}\right\|_F^2 \quad (350)$$
$$\sum_{i=1}^{n_{\rm r}} |g_i|^2 = \left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger \hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2 = \frac{\mathsf{snr}}{n_{\rm t}}\left\|\hat{\boldsymbol{H}}_b\boldsymbol{x}\right\|_F^2 \quad (351)$$

because $\hat{\boldsymbol{Q}}$ is a unitary matrix that does not change the Euclidean norm of a vector. This removes the difficulty of obtaining the exact expression for $\hat{\boldsymbol{Q}}$. On the other hand, the last term is difficult to evaluate as the summation involves the variable $\hat{\lambda}_{b,i}$. Herein we have to impose an additional condition such that the last term in (349) can be evaluated. Suppose that we restrict $\rho s < 1$ with strict inequality for $0 \le \rho \le 1$ and $s > 0$. Then, we have the bound for $\hat{\lambda}_{b,i} \ge 0$ and $\mathsf{snr} \ge 0$

$$1 \le \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)} \le \frac{1}{1-\rho s}. \quad (352)$$

Hence, we can upper-bound the last term in (349) as follows:

$$\sum_{i=1}^{n_{\rm r}} |\rho s g_i - c_i|^2 \times \frac{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}}{1+s\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\rho s)}$$
$$\le \frac{1}{1-\rho s}\sum_{i=1}^{n_{\rm r}} |\rho s g_i - c_i|^2 \quad (353)$$
$$= \frac{1}{1-\rho s}\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\hat{\boldsymbol{Q}}^\dagger(\rho s\hat{\boldsymbol{H}}_b\boldsymbol{x} - \boldsymbol{H}_b\boldsymbol{x})\right\|_F^2 \quad (354)$$
$$= \frac{1}{1-\rho s}\frac{\mathsf{snr}}{n_{\rm t}}\left\|\rho s\hat{\boldsymbol{H}}_b\boldsymbol{x} - \boldsymbol{H}_b\boldsymbol{x}\right\|_F^2 \quad (355)$$

where the last equality is due to the unitary matrix $\hat{\boldsymbol{Q}}^\dagger$ that does not affect the Euclidean norm of a vector. This removes the dependency on $\hat{\boldsymbol{Q}}^\dagger$. From the exponential term in (349) and by combining (350) and (351) and (355), we have the following upper bound to the expectation over x:

$$\mathbb{E}\left[\exp\left(-\frac{\mathsf{snr}}{n_t}\left(\|\boldsymbol{H}_b\mathsf{x}\|_F^2-\rho s\|\hat{\boldsymbol{H}}_b\mathsf{x}\|_F^2 - \frac{1}{1-\rho s}\|(\rho s\hat{\boldsymbol{H}}_b-\boldsymbol{H}_b)\mathsf{x}\|_F^2\right)\right)\right]$$

$$=\int_{\boldsymbol{x}}\exp\left(-\frac{\mathsf{snr}}{n_t}\left(\|\boldsymbol{H}_b\boldsymbol{x}\|_F^2-\rho s\|\hat{\boldsymbol{H}}_b\boldsymbol{x}\|_F^2 - \frac{1}{1-\rho s}\|(\rho s\hat{\boldsymbol{H}}_b-\boldsymbol{H}_b)\boldsymbol{x}\|_F^2\right)\right)\times\frac{1}{\pi^{n_t}}e^{-\|\boldsymbol{x}\|_F^2}d\boldsymbol{x} \quad (356)$$

$$=\frac{1}{\pi^{n_t}}\int_{\boldsymbol{x}}\exp\left(\frac{\rho s}{1-\rho s}\frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\boldsymbol{x}\|_F^2-\|\boldsymbol{x}\|_F^2\right)d\boldsymbol{x} \quad (357)$$

$$\le\frac{1}{\pi^{n_t}}\int_{\boldsymbol{x}}\exp\left(\frac{\rho s}{1-\rho s}\frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2\|\boldsymbol{x}\|_F^2-\|\boldsymbol{x}\|_F^2\right)d\boldsymbol{x} \quad (358)$$

$$=\frac{1}{\left(1-\frac{\rho s}{1-\rho s}\frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2\right)^{n_t}} \quad (359)$$

where the last inequality is due to $\|\boldsymbol{E}_b\boldsymbol{x}\|_F^2 \le \|\boldsymbol{E}_b\|_F^2\|\boldsymbol{x}\|_F^2$. Note that the above function is integrable if $\frac{\rho s}{1-\rho s}\frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2<1$ and as the SNR increases, there exist positive constants $u_1$ and $u_2$ for which $s\le\frac{u_1}{u_2+\frac{\mathsf{snr}}{n_t}\|\boldsymbol{E}_b\|_F^2}$ guarantees that the function is integrable. Thus, the choice of $s$ in (236) needs to modified such that this condition is satisfied. However, since we just need to modify the constants, i.e., $u_1$ and $u_2$, which are SNR independent, it does not affect the exponential equality in the high-SNR regime.

## Appendix G
## Tighter Achievability Bound for Gaussian Inputs

From (334) and (335) (Appendix E) with $\underline{\rho}=1$, we can rewrite the lower bound for the mismatched decoding error exponent as follows:

$$E_r^Q(R,\hat{\boldsymbol{H}}) \ge \left[\frac{1}{B}\sum_{b=1}^{B}\left(u_3+\sum_{i=1}^{n_r}\log\left(1+\underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_t}(1-\underline{s})\right)\right)-R\right]_+ \quad (360)$$

$$\ge\left[\frac{1}{B}\sum_{b=1}^{B}\log\left(e^{u_3}\cdot\left(1+\underline{s}\|\hat{\boldsymbol{H}}_b\|_F^2\frac{\mathsf{snr}}{n_t}(1-\underline{s})\right)\right)-R\right]_+ \quad (361)$$

where $u_3<0$ and

$$\underline{s}=\frac{1}{n_r\left(1+\frac{\mathsf{snr}}{n_t}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2\right)}. \quad (362)$$

We have used the last inequality above to derive the block length threshold for Gaussian input as shown in Theorem 2. The main benefit of using this is that the results are general for the fading model in (5) since we can easily evaluate the mismatched decoding error exponent in terms of $\boldsymbol{A}$ and $\boldsymbol{\Theta}$. However, using the last inequality implies looser achievability bound and the resulting block length threshold may not be tight.

A tighter bound is obtained by using the inequality (360). This requires the joint density function of $\hat{\boldsymbol{\lambda}}_b$ and the entries of $\mathsf{E}_b$. Note that for a given $\mathsf{E}$, $\hat{\mathsf{H}}$ has the same distribution and covariance as $\mathsf{H}$ but with the mean shifted by $\mathsf{E}$. Conditioned on $\mathsf{E}=\boldsymbol{E}$, the mean of $\hat{\mathsf{H}}$ is given by $\boldsymbol{E}$. From (5), the conditional distribution of each channel estimation entry, $\hat{H}_{b,r,t}$, is given by

$$p_{\hat{H}_{b,r,t}|E_{b,r,t}}(\hat{h}|e)=w_0|\hat{h}-e|^\tau e^{-w_1|\hat{h}-e-w_2|^\varphi}. \quad (363)$$

The characterization of the above pdf is difficult when $\tau\neq 0$. This can be explained as follows. As the SNR tends to infinity, the near zero behavior determines the dominating term in the pdf [11], [23]. Note that for $\tau\neq0$, the near zero behavior of the pdf is determined by the values of $\hat{h}, e$ in $|\hat{h}-e|^\tau$. This behavior does not only depend on $|\hat{h}|^\tau$ but also $|e|^\tau$ and the angles of $\hat{h}$ and $e$. These interplaying variables make the high-SNR behavior of the pdf intractable. On the other hand, when $\tau=0$, the variable $e$ only affects the exponential term, which for many cases tends to decay exponentially or converges to a constant for high SNR (see also [23], [24]).

Consider $\tau=0$ and assume $n_t\ge n_r$, we perform a change of random variables from the matrix entries in $\hat{\boldsymbol{H}}_b$ to its eigenvalues $\hat{\lambda}_{b,i}$ for all $i=1,\dots,n_r$. Since the entries of the channel matrix are assumed to be i.i.d., the pdf of $\hat{\mathsf{H}}_b$ for a given $\mathsf{E}_b$ is given by

$$p_{\hat{\mathsf{H}}_b|\mathsf{E}_b}\left(\hat{\boldsymbol{H}}_b\middle|\boldsymbol{E}_b\right)=\prod_{r=1}^{n_r}\prod_{t=1}^{n_t}w_0e^{-w_1|\hat{h}_{b,r,t}-e_{b,r,t}-w_2|^\varphi}. \quad (364)$$

Using the singular value decomposition of $\hat{\boldsymbol{H}}_b=\boldsymbol{UDS}$ and the eigenvalue decomposition in the form of $\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^\dagger=\boldsymbol{J}\Sigma\boldsymbol{J}^\dagger$ [27], random matrices results [38], [39] provide the joint distribution of the ordered eigenvalues in the following form [23], [24]:

$$p_{\hat{\boldsymbol{\lambda}}_b|\mathsf{E}_b}(\hat{\boldsymbol{\lambda}}_b|\boldsymbol{E}_b)=C_{n,m}\prod_{i=1}^{n_r}\hat{\lambda}_{b,i}^{n_t-n_r}\cdot\prod_{i<j}(\hat{\lambda}_{b,i}-\hat{\lambda}_{b,j})^2 \cdot\int_{\mathcal{V}_{n_r,n_r}}\int_{\mathcal{V}_{n_r,n_t}}p_{\hat{\mathsf{H}}_b|\mathsf{E}_b}(\boldsymbol{UDS}|\boldsymbol{E}_b)d\boldsymbol{S}d\boldsymbol{U} \quad (365)$$

where $C_{n,m}$ is the normalizing constant and $\mathcal{V}_{n_r,n_r}$ and $\mathcal{V}_{n_r,n_t}$ are the complex Stiefel manifolds [38], [39]. The definition of complex Stiefel manifold is as follows. Suppose that $\boldsymbol{G}$ is an $n\times m$ matrix $(n\ge m)$ with orthonormal columns so that $\boldsymbol{G}^\dagger\boldsymbol{G}=\boldsymbol{I}_m$. The set of all such matrices $\boldsymbol{G}$ is called the Stiefel manifold, defined by [38], [39]

$$\mathcal{V}_{m,n}=\left\{\boldsymbol{G}\in\mathbb{C}^{n\times m}:\boldsymbol{G}^\dagger\boldsymbol{G}=\boldsymbol{I}_m\right\}. \quad (366)$$

Note that $\Sigma=\mathrm{diag}[\hat{\lambda}_{b,1},\dots,\hat{\lambda}_{b,n_r}]$ with $\hat{\lambda}_{b,1}\le\cdots\le\hat{\lambda}_{b,n_r}$ and $\boldsymbol{D}=\mathrm{diag}[\hat{\lambda}_{b,1}^{\frac{1}{2}},\dots,\hat{\lambda}_{b,n_r}^{\frac{1}{2}}]$.

Let $\hat{\beta}_{b,i} = -\frac{\log \hat{\lambda}_{b,i}}{\log \mathsf{snr}}$. Using this change of variable, (364) and (365), we can write the above pdf for $\tau = 0$ as done in [24]

$$p_{\hat{\boldsymbol{\beta}}_b|\mathsf{E}_b}\left(\hat{\boldsymbol{\beta}}_b\middle|\boldsymbol{E}_b\right) = C_{n,m}\cdot(\log\mathsf{snr})^{n_\mathrm{r}} \cdot\prod_{i=1}^{n_\mathrm{r}}\mathsf{snr}^{-(n_\mathrm{t}-n_\mathrm{r}+1)\hat{\beta}_{b,i}}\cdot\prod_{i<j}\left(\mathsf{snr}^{-\hat{\beta}_{b,i}}-\mathsf{snr}^{-\hat{\beta}_{b,j}}\right)^2 \cdot\left(\int_{\mathcal{V}_{n_\mathrm{r},n_\mathrm{r}}}\int_{\mathcal{V}_{n_\mathrm{r},n_\mathrm{t}}} w_0^{n_\mathrm{r}n_\mathrm{t}}e^{-w_1\left(\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_\varphi\right)^\varphi}d\boldsymbol{S}d\boldsymbol{U}\right) \tag{367}$$

where $\boldsymbol{W}_2$ is an $n_\mathrm{r}\times n_\mathrm{t}$ matrix with all elements equal to $w_2$ and $\|\cdot\|_\varphi$ is the $\varphi$-norm [8]. As we deal with the achievability bound, it is sufficient to find a tight upper bound for the pdf. Note that since $\mathbb{C}^{n_\mathrm{r}\times n_\mathrm{t}}$ is a finite-dimensional complex space, all norms on $\mathbb{C}^{n_\mathrm{r}\times n_\mathrm{t}}$ are equivalent[9] [40]. Thus, we can find a real positive number $u_4 > 0$ such that the term in the exponent can be lower-bounded as

$$\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_\varphi \geq u_4\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_F. \tag{368}$$

Applying the backward triangle inequality for the matrix norm, we have that

$$\begin{aligned}\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_F &\geq \left|\|\hat{\boldsymbol{H}}_b\|_F-\|\boldsymbol{E}_b+\boldsymbol{W}_2\|_F\right| &(369)\\ &= \left|\sqrt{\sum_{i=1}^{n_\mathrm{r}}\hat{\lambda}_{b,i}}-\sqrt{\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}|e_{b,r,t}+w_2|^2}\right| &(370)\\ &= \left|\sqrt{\sum_{i=1}^{n_\mathrm{r}}\mathsf{snr}^{-\hat{\beta}_{b,i}}}-\sqrt{\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\left|\mathsf{snr}^{\frac{-\theta_{b,r,t}}{2}}e^{\iota\phi^e_{b,r,t}}+w_2\right|^2}\right|. &(371)\end{aligned}$$

Since $\varphi \geq 1$ by definition in (5), we can lower-bound (367) using

$$\begin{aligned}\left(\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_\varphi\right)^\varphi &\geq u_4^\varphi\left(\|\hat{\boldsymbol{H}}_b-\boldsymbol{E}_b-\boldsymbol{W}_2\|_F\right)^\varphi &(372)\\ &\geq u_4^\varphi\left|\|\hat{\boldsymbol{H}}_b\|_F-\|\boldsymbol{E}_b+\boldsymbol{W}_2\|_F\right|^\varphi. &(373)\end{aligned}$$

This follows from the monotonicity of the function $f(u) = u^\varphi$ over the interval $u > 0$. Remark that the conditional pdf in (365) is conditioned on $\boldsymbol{E}_b$. We can write the joint density function of $\hat{\boldsymbol{\lambda}}_b$, $\boldsymbol{E}_b$ as follows

$$p_{\hat{\boldsymbol{\lambda}}_b,\mathsf{E}_b}\left(\hat{\boldsymbol{\lambda}}_b,\boldsymbol{E}_b\right) = p_{\hat{\boldsymbol{\lambda}}_b|\mathsf{E}_b}\left(\hat{\boldsymbol{\lambda}}_b\middle|\boldsymbol{E}_b\right)p_{\mathsf{E}_b}(\boldsymbol{E}_b). \tag{374}$$

[8]The $p$-norm of $n$ elements matrix/vector with $p \geq 1$ is defined as $\|x\|_p \triangleq \left(\sum_{i=1}^n|x_i|^p\right)^{\frac{1}{p}}$ [27]. This complies with the constraint on $\varphi \geq 1$ in (5). For $p = 2$, the norm is called Frobenius norm for a matrix or Euclidean norm for a vector. For a matrix, this $p$-norm denotes the entry-wise norm, i.e., treating the matrix as a vector. We denote the Frobenius norm for a matrix and Euclidean norm for a vector as $\|\cdot\|_F$ instead of $\|\cdot\|_2$ because $\|\cdot\|_2$ is normally used to denote the induced matrix-norms [27], which are different from the entry-wise norms.

[9]The equivalence of norms can be explained as follows. Given a finite-dimensional space $\mathbb{C}^{m\times n}$ and a matrix $\boldsymbol{X}\in\mathbb{C}^{m\times n}$, there exists positive real numbers $P$ and $S$ independent of $\boldsymbol{X}$ such that $P\|\boldsymbol{X}\|_{p'} \leq \|\boldsymbol{X}\|_p \leq S\|\boldsymbol{X}\|_{p'}$ [40].

The density $p_{\hat{\boldsymbol{\lambda}}_b|\mathsf{E}_b}(\hat{\boldsymbol{\lambda}}_b|\boldsymbol{E}_b)p_{\mathsf{E}_b}(\boldsymbol{E}_b)$ can be further expanded as

$$p_{\hat{\boldsymbol{\lambda}}_b|\mathsf{E}_b}\left(\hat{\boldsymbol{\lambda}}_b\middle|\boldsymbol{E}_b\right)p_{\mathsf{E}_b}(\boldsymbol{E}_b) = p_{\hat{\boldsymbol{\lambda}}_b|\{E_{b,r,t}\}}\left(\hat{\boldsymbol{\lambda}}_b\middle|\{e_{b,r,t}\}\right)\prod_{r=1}^{n_\mathrm{r}}\prod_{t=1}^{n_\mathrm{t}}p_{E_{b,r,t}}(e_{b,r,t}) \tag{375}$$

where $\{e_{b,r,t}\}$ denotes the collection of $e_{b,r,t}$ for all $r = 1,\ldots,n_\mathrm{r}$ and $t = 1,\ldots,n_\mathrm{t}$. The equality in (375) holds since the matrix $\boldsymbol{E}_b$ can be completely expressed in terms of its entries $e_{b,r,t}$, $r = 1,\ldots,n_\mathrm{r}$, $t = 1,\ldots,n_\mathrm{t}$. Note that the entries of the random matrix $\mathsf{E}_b$ are i.i.d. random variables and for each entry, the phase of $E_{b,r,t}$, $\Phi^E_{b,r,t}$, is independent from its magnitude $|E_{b,r,t}|$ and uniformly distributed over $[0, 2\pi)$. Hence, applying the transformation of the variables $\hat{\lambda}_{b,i}$ and $|e_{b,r,t}|^2$ to $\hat{\beta}_{b,i} = -\frac{\log\hat{\lambda}_{b,i}}{\log\mathsf{snr}}$ and $\theta_{b,r,t} = -\frac{\log|e_{b,r,t}|^2}{\log\mathsf{snr}}$, we have the joint pdf of $\hat{\boldsymbol{\beta}}_b$, $\Theta_{b,r,t}$ and $\Phi^E_{b,r,t}$, $r = 1,\ldots,n_\mathrm{r}$, $t = 1,\ldots,n_\mathrm{t}$ as follows:

$$p_{\hat{\boldsymbol{\beta}}_b,\{\Theta_{b,r,t}\},\{\Phi^E_{b,r,t}\}}\left(\hat{\boldsymbol{\beta}}_b,\{\theta_{b,r,t}\},\{\phi^e_{b,r,t}\}\right) = p_{\hat{\boldsymbol{\beta}}_b|\{\Theta_{b,r,t}\},\{\Phi^E_{b,r,t}\}}\left(\hat{\boldsymbol{\beta}}_b\middle|\{\theta_{b,r,t}\},\{\phi^e_{b,r,t}\}\right)\cdot\prod_{r=1}^{n_\mathrm{r}}\prod_{t=1}^{n_\mathrm{t}}p_{\Theta_{b,r,t}}(\theta_{b,r,t})p_{\Phi^E_{b,r,t}}(\phi^e_{b,r,t}). \tag{376}$$

We continue the analysis from (375) and (376). Note that the term $p_{\hat{\boldsymbol{\beta}}_b|\{\Theta_{b,r,t}\},\{\Phi^E_{b,r,t}\}}(\hat{\boldsymbol{\beta}}_b|\{\theta_{b,r,t}\},\{\phi^e_{b,r,t}\})$ in (376) can be further upper-bounded using (373). We then group the exponential terms as follows:

$$\exp\left(-w_1u_4^\varphi\left|-\sqrt{\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\left|\mathsf{snr}^{\frac{-\theta_{b,r,t}}{2}}e^{\iota\phi^e_{b,r,t}}+w_2\right|^2}+\sqrt{\sum_{i=1}^{n_\mathrm{r}}\mathsf{snr}^{-\hat{\beta}_{b,t}}}\right|^\varphi-\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\mathsf{snr}^{(d_e-\theta_{b,r,t})}\right). \tag{377}$$

As the SNR increases, the behavior is dominated by the smallest values of $\hat{\beta}_{b,i}$, $i = 1,\ldots,n_\mathrm{r}$ and $\theta_{b,r,t}$, $r = 1,\ldots,n_\mathrm{r}$, $t = \ldots,n_\mathrm{t}$. Since the eigenvalues $\hat{\lambda}_{b,1},\ldots,\hat{\lambda}_{b,n_\mathrm{t}}$ are ordered in a nondecreasing order, the dominating terms are indicated by $\hat{\beta}_{b,n_\mathrm{r}}$ and $\theta_{b,\min}$. We have the following observations.

1) For the terms in the modulus $|\cdot|^\varphi$, if $\hat{\beta}_{b,n_\mathrm{r}} \geq 0$ and $\theta_{b,\min} \geq 0$, those terms are converging to some constant $|w_2'|^\varphi \geq 0$ as the SNR increases and the convergence of the exponential term is determined by

$$\begin{aligned}&-w_1u_4^\varphi|w_2'|^\varphi-\sum_{r=1}^{n_\mathrm{r}}\sum_{t=1}^{n_\mathrm{t}}\mathsf{snr}^{(d_e-\theta_{b,r,t})}\\ &\doteq -\mathsf{snr}^0-\mathsf{snr}^{d_e-\theta_{b,\min}}.\end{aligned} \tag{378}$$

If $\theta_{b,\min} < d_e$, then $\mathsf{snr}^{(d_e-\theta_{b,\min})}$ dominates and this makes the overall pdf upper bound decay exponentially to zero as the SNR increases. If $\theta_{b,\min} \geq d_e$, then the constant $w_1u_4^\varphi|w_2'|^\varphi$ dominates and eventually the exponential function converges to an SNR independent constant which can be neglected in the pdf upper bound for the asymptotic analysis.

2) If either $\hat{\beta}_{b,n_{\rm r}} < 0$ or $\theta_{b,\min} < 0$, then the exponential convergence can be explained in the following cases.

- If $\hat{\beta}_{b,n_{\rm r}} < \theta_{b,\min}$, then the following dominates the exponent

$$-w_1 u_4^{\varphi}\left(\sqrt{\sum_{i=1}^{n_{\rm r}} \mathsf{snr}^{-\hat{\beta}_{b,i}}}\right)^{\varphi} - \sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{(d_e-\theta_{b,r,t})}$$
$$\doteq -\mathsf{snr}^{-\frac{\varphi}{2}\hat{\beta}_{b,n_{\rm r}}} - \mathsf{snr}^{d_e-\theta_{b,\min}} \tag{379}$$
$$\doteq -\mathsf{snr}^{\max(-\frac{\varphi}{2}\hat{\beta}_{b,n_{\rm r}}, d_e-\theta_{b,\min})}. \tag{380}$$

Since $\hat{\beta}_{b,n_{\rm r}} < 0$, it can be seen that the exponential function always makes the pdf upper bound decay exponentially to zero as the SNR increases.

- If $\hat{\beta}_{b,n_{\rm r}} > \theta_{b,\min}$, with probability one, the dominating exponent for high SNR is given by

$$-w_1 u_4^{\varphi}\left(\sqrt{\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} \left|\mathsf{snr}^{\frac{-\theta_{b,r,t}}{2}} e^{\iota\phi^e_{b,r,t}} + w_2\right|^2}\right)^{\varphi} - \sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} \mathsf{snr}^{(d_e-\theta_{b,r,t})}$$
$$\doteq -\mathsf{snr}^{-\frac{\varphi}{2}\theta_{b,\min}} - \mathsf{snr}^{d_e-\theta_{b,\min}} \tag{381}$$
$$\doteq -\mathsf{snr}^{\max(-\frac{\varphi}{2}\theta_{b,\min},\, d_e-\theta_{b,\min})}. \tag{382}$$

Since $\theta_{b,\min}$ is less than zero, it can be seen that the exponential function always makes the pdf upper bound decay exponentially to zero as the SNR increases.

3) Note that we have $\hat{\beta}_{b,1} \geq \cdots \geq \hat{\beta}_{b,n_{\rm r}}$ and for any $r = 1,\ldots,n_{\rm t}, t=1,\ldots,n_{\rm t}, \theta_{b,r,t} \geq \theta_{b,\min}$.

Hence, from the above observations, we require that $\hat{\boldsymbol{\beta}}_b \succeq \mathbf{0}$ and $\boldsymbol{\Theta}_b \succeq d_e \times \mathbf{1}$, $b=1,\ldots,B$ so that the pdf upper bound does not decay exponentially to zero as the SNR tends to infinity.

We continue the analysis by evaluating the lower bound for $E_r^Q(R,\hat{\boldsymbol{H}})$ in (360)

$$E_r^Q(R,\hat{\boldsymbol{H}}) \geq \left[\frac{1}{B}\sum_{b=1}^{B}\left(u_3 + \sum_{i=1}^{n_{\rm r}} \log\left(1 + \underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\underline{s})\right)\right) - R\right]_+ \tag{383}$$
$$= \left[\frac{1}{B}\log\left(\prod_{b=1}^{B}\prod_{i=1}^{n_{\rm r}} e^{\frac{u_3}{n_{\rm r}}}\left(1 + \underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\underline{s})\right)\right) - R\right]_+$$
$$\triangleq E_r'^Q(R,\hat{\boldsymbol{H}}) \tag{384}$$

where

$$\underline{s} = \frac{1}{n_{\rm r}\left(1 + \frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2\right)} \tag{385}$$
$$= \frac{1}{n_{\rm r}\left(1 + \frac{\mathsf{snr}}{n_{\rm t}}\sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}|e_{b,r,t}|^2\right)}, \tag{386}$$

and where we have defined $E_r'^Q(R,\hat{\boldsymbol{H}})$ as the RHS of (384). Using the change of variables from $\hat{\lambda}_{b,i}$ and $|e_{b,r,t}|^2$ to $\hat{\beta}_{b,i}$ and $\theta_{b,r,t}$, we can show the following high-SNR exponential equality

$$e^{\frac{u_3}{n_{\rm r}}}\left(1 + \underline{s}\hat{\lambda}_{b,i}\frac{\mathsf{snr}}{n_{\rm t}}(1-\underline{s})\right) \doteq \mathsf{snr}^{[\min(1,\theta_{\min})-\hat{\beta}_{b,i}]_+} \tag{387}$$

where

$$\theta_{\min} = \min\left\{\theta_{1,1,1},\ldots,\theta_{b,r,t},\ldots,\theta_{B,n_{\rm r},n_{\rm t}}\right\}. \tag{388}$$

It follows from $E_r'^Q(R,\hat{\boldsymbol{H}})$ in the RHS of (384), (387) and the rate and multiplexing gain relationship $e^{R(\mathsf{snr})} \doteq \mathsf{snr}^k$ that for high SNR, if the following event:

$$\bar{\mathcal{A}}_G = \left\{\hat{\boldsymbol{\beta}} \in \mathbb{R}^{B\times n_{\rm r}}, \boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm r}\times n_{\rm t}} : \sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}}\left[\min(1,\theta_{\min}) - \hat{\beta}_{b,i}\right]_+ \leq Bk\right\} \tag{389}$$

occurs, then $E_r'^Q(R,\hat{\boldsymbol{H}}_b) = 0$ and otherwise if

$$\bar{\mathcal{A}}_G^c = \left\{\hat{\boldsymbol{\beta}} \in \mathbb{R}^{B\times n_{\rm r}}, \boldsymbol{\Theta} \in \mathbb{R}^{B\times n_{\rm r}\times n_{\rm t}} : \sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}}\left[\min(1,\theta_{\min}) - \hat{\beta}_{b,i}\right]_+ > Bk\right\} \tag{390}$$

occurs, then $E_r'^Q(R,\hat{\boldsymbol{H}}_b) > 0$. Therefore, for the fading model with $\tau = 0$, the upper bound to the average error probability is given as

$$P_{e,\rm ave} \leq \mathbb{E}\left[e^{-LBE_r^Q(R,\hat{\mathsf{H}})}\right] \tag{391}$$
$$\dot{\leq} \int_{\bar{\mathcal{A}}_G\cap\{\hat{\boldsymbol{\beta}}\succeq\mathbf{0},\boldsymbol{\Theta}\succeq d_e\times\mathbf{1}\}}\left(\mathsf{snr}^{-\sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}}(2i-1+n_{\rm t}-n_{\rm r})\hat{\beta}_{b,i}} \times \mathsf{snr}^{\sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}(d_e-\theta_{b,r,t})} d\hat{\boldsymbol{\beta}}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{E}}\right)$$
$$+ \int_{\bar{\mathcal{A}}_G^c\cap\{\hat{\boldsymbol{\beta}}\succeq\mathbf{0},\boldsymbol{\Theta}\succeq d_e\times\mathbf{1}\}}\left(\mathsf{snr}^{-\sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}}(2i-1+n_{\rm t}-n_{\rm r})\hat{\beta}_{b,i}} \times \mathsf{snr}^{\sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}}(d_e-\theta_{b,r,t})} \times \mathsf{snr}^{-L\left(\sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}}[\min(1,\theta_{\min})-\hat{\beta}_{b,i}]_+ - Bk\right)} d\hat{\boldsymbol{\beta}}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{E}}\right) \tag{392}$$
$$\doteq G_1\mathsf{snr}^{-d_1(k)} + G_2\mathsf{snr}^{-d_2(k)} \tag{393}$$
$$\doteq \mathsf{snr}^{-d_2(k)} \tag{394}$$

where $G_1, G_2 \doteq \mathsf{snr}^0$, and $d_1(k)$ is the generalized outage SNR exponent achieved with infinite block length. Note that to find the solution of $d_1(k)$, we follow the same approach of finding the optimal DMT in [11]. The lower-bound to the optimal DMT curve $d_1(k)$ is given by the piecewise-linear function connecting the points $(k, d_1(k))$, where

$$k = 0, \min(1,d_e), 2\min(1,d_e), \ldots, n_{\rm r}\min(1,d_e) \tag{395}$$
$$d_1(k) = \min(1,d_e)\cdot B\left(n_{\rm t} - \frac{k}{\min(1,d_e)}\right) \cdot\left(n_{\rm r} - \frac{k}{\min(1,d_e)}\right). \tag{396}$$

Note that we have $d_{1,\max} = \min(1, d_e)Bn_{\rm t}n_{\rm r}$ and $k_{\max} = \min(1, d_e)n_{\rm r}$. On the other hand, $d_2(k)$ is given as

$$d_2(k) = \inf_{\bar{\mathcal{A}}_G^c \cap \{\hat{\boldsymbol{\beta}} \succeq \mathbf{0}, \boldsymbol{\Theta} \succeq d_e \times \mathbf{1}\}} \Bigg\{ \sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}} (2i - 1 + n_{\rm t} - n_{\rm r})\hat{\beta}_{b,i} + \sum_{b=1}^{B}\sum_{r=1}^{n_{\rm r}}\sum_{t=1}^{n_{\rm t}} (\theta_{b,r,t} - d_e) + L\left(\sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}} [\min(1, \theta_{\min}) - \hat{\beta}_{b,i}]_+ - Bk\right) \Bigg\}. \tag{397}$$

Since we need

$$L\left(\sum_{b=1}^{B}\sum_{i=1}^{n_{\rm r}} [\min(1, \theta_{\min}) - \hat{\beta}_{b,i}]_+ - Bk\right) > 0 \tag{398}$$

for $d_2(k)$, it is straightforward to deduce that $d_2(k) \le d_1(k)$. This follows from [11, Lemma 6]. Thus, $d_2(k)$ leads to $d_G^{\ell}(k)$ in the proposition. Note that we just need to replace $(\boldsymbol{I}_{n_{\rm r}} + s\hat{\boldsymbol{H}}_b\hat{\boldsymbol{H}}_b^{\dagger}\frac{\mathsf{snr}}{n_{\rm t}})$ with $(\boldsymbol{I}_{n_{\rm t}} + s\hat{\boldsymbol{H}}_b^{\dagger}\hat{\boldsymbol{H}}_b\frac{\mathsf{snr}}{n_{\rm t}})$ in the analysis if $n_{\rm t} < n_{\rm r}$.

## Appendix H
## Discrete Inputs Achievability Proof

We use the generalized Gallager upper bound to derive the achievability by isolating the channel block length and the random coding exponent. Recall $E_0^Q(s, \rho, \hat{\boldsymbol{H}}_b)$ in (35) written in different form here

$$E_0^Q(s, \rho, \hat{\boldsymbol{H}}_b) = -\log_2 \mathbb{E}\left[\left(\sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\boldsymbol{x}', \hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}, \hat{\mathsf{H}}_b)}\right)^s\right)^{\rho} \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right]. \tag{399}$$

For a given $\mathsf{y} = \boldsymbol{y}, \mathsf{x} = \boldsymbol{x}, \mathsf{H}_b = \boldsymbol{H}_b$, and $\mathsf{E}_b = \boldsymbol{E}_b$, inserting the decoding metric (8) and evaluating the expectation over $\mathsf{x}'$, we have that

$$\sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\boldsymbol{y}|\boldsymbol{x}', \hat{\boldsymbol{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\boldsymbol{y}|\boldsymbol{x}, \hat{\boldsymbol{H}}_b)}\right)^s = 2^{-Mn_{\rm t}} \sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} \left(e^{-\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}') + \boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2} \cdot e^{\left\|\boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)^s. \tag{400}$$

Substituting (400) to the RHS of (399), we obtain

$$-\log_2 \mathbb{E}\left[\left(\sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} P_{\mathsf{x}}(\boldsymbol{x}')\left(\frac{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\boldsymbol{x}', \hat{\mathsf{H}}_b)}{Q_{\mathsf{y}|\mathsf{x},\hat{\mathsf{H}}}(\mathsf{y}|\mathsf{x}, \hat{\mathsf{H}}_b)}\right)^s\right)^{\rho} \middle| \mathsf{H}_b = \boldsymbol{H}_b, \mathsf{E}_b = \boldsymbol{E}_b\right]$$
$$= -\log_2 \sum_{\boldsymbol{x} \in \mathcal{X}^{n_{\rm t}}} \mathbb{E}\left[\left(\sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} \left(e^{-s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}') + \mathsf{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2} \cdot e^{s\left\|\mathsf{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)\right)^{\rho}\right] + (1+\rho)Mn_{\rm t}. \tag{401}$$

Note that

$$1 \le \left(\sum_{\boldsymbol{x}' \in \mathcal{X}^{n_{\rm t}}} \left(e^{-s\left\|\sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}') + \boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2} \cdot e^{s\left\|\boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)\right)^{\rho} \tag{402}$$

$$\le |\mathcal{X}^{n_{\rm t}}|^{\rho} e^{\rho s\left\|\boldsymbol{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}. \tag{403}$$

We have the expectation over z

$$\mathbb{E}\left[|\mathcal{X}^{n_{\rm t}}|^{\rho} e^{\rho s\left\|\mathsf{z} - \sqrt{\frac{\mathsf{snr}}{n_{\rm t}}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right]$$
$$= \frac{|\mathcal{X}^{n_{\rm t}}|^{\rho}}{(1-\rho s)^{n_{\rm r}}} e^{\left(\frac{\rho^2 s^2}{1-\rho s} + \rho s\right)\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\boldsymbol{x}\|_F^2} \tag{404}$$
$$\le \frac{|\mathcal{X}^{n_{\rm t}}|^{\rho}}{(1-\rho s)^{n_{\rm r}}} e^{\left(\frac{\rho^2 s^2}{1-\rho s} + \rho s\right)\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\|\boldsymbol{x}\|_F^2} \tag{405}$$

where we have assumed $\rho s < 1$ so that the expectation can be evaluated, and where we have used the Frobenius norm property $\|\boldsymbol{E}_b\boldsymbol{x}\|_F^2 \le \|\boldsymbol{E}_b\|_F^2\|\boldsymbol{x}\|_F^2$ in the last inequality. Since the signal energy $\|\boldsymbol{x}\|_F^2$, $\boldsymbol{x} \in \mathcal{X}^{n_{\rm t}}$ is finite, the condition

$$\frac{|\mathcal{X}^{n_{\rm t}}|^{\rho}}{(1-\rho s)^{n_{\rm r}}} e^{\left(\frac{\rho^2 s^2}{1-\rho s} + \rho s\right)\frac{\mathsf{snr}}{n_{\rm t}}\|\boldsymbol{E}_b\|_F^2\|\boldsymbol{x}\|_F^2} < \infty \tag{406}$$

can be satisfied by choosing the optimal solution of $s$ over

$$\mathcal{S} = \left\{s \in \mathbb{R}^+ : 0 < s \le \frac{1}{B + \mathsf{snr}\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2}\right\}. \tag{407}$$

The choice of $s \in \mathcal{S}$ leads to a lower bound to the mismatched decoding error exponent in (34). The term with $\sum_{b=1}^{B}\|\boldsymbol{E}_b\|_F^2$ is needed since we have $B$ blocks. As (406) can be satisfied with $s \in \mathcal{S}$, the dominated convergence theorem [35] can be applied here. Using the same argument as in the converse analysis, we can apply the point-wise limit in the expectation. Let $s^*$ be the solution of $s$ that solves the supremum in the RHS of

(34). Then, using a similar argument to the one used in the generalized outage evaluation (Appendix C), we can conclude the following expectation over z

$$\lim_{\mathsf{snr}\to\infty}\mathbb{E}\left[\left(\sum_{\boldsymbol{x}'\in\mathcal{X}^{n_t}}\left(e^{-s^*\left\|\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{H}_b(\boldsymbol{x}-\boldsymbol{x}')+\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{E}_b\boldsymbol{x}'\right\|_F^2}\cdot e^{s^*\left\|\mathsf{z}-\sqrt{\frac{\mathsf{snr}}{n_t}}\boldsymbol{E}_b\boldsymbol{x}\right\|_F^2}\right)\right)^{\rho}\right]$$

$$\leq\mathbb{E}\left[\left(\sum_{\boldsymbol{x}'\in\mathcal{X}^{n_t}}\mathbb{1}\left\{x_t\neq x_t',\forall t\in\underline{\mathcal{S}}_b^{(\epsilon,\delta)}\right\}\right)^{\rho}\right]\tag{408}$$

$$=2^{\rho M(n_t-\underline{\kappa}_b)}\tag{409}$$

where $\underline{\mathcal{S}}_b^{(\epsilon,\delta)}$ and $\underline{\kappa}_b$ have the same definition as in the converse analysis (Appendix C). Consequently, we have at high SNR

$$E_0^Q(s^*,\rho,\hat{\boldsymbol{H}}_b)\geq\rho M\underline{\kappa}_b\tag{410}$$

and the random coding error exponent $E_r^Q(R,\hat{\boldsymbol{H}}_b)$ can be bounded as follows:

$$E_r^Q(R,\hat{\boldsymbol{H}}_b)=\sup_{\substack{s>0\\0\leq\rho\leq1}}\frac{1}{B}\sum_{b=1}^{B}E_0^Q(s,\rho,\hat{\boldsymbol{H}}_b)-\rho R\tag{411}$$

$$\geq\sup_{0\leq\rho\leq1}\rho M\left(\frac{1}{B}\sum_{b=1}^{B}\underline{\kappa}_b-\frac{R}{M}\right).\tag{412}$$

Define $\Lambda$ and two mutually exclusive sets as follows:

$$\Lambda\triangleq\sum_{b=1}^{B}\underline{\kappa}_b-\frac{BR}{M}\tag{413}$$

$$\mathcal{A}_{\mathcal{X}}\triangleq\left\{\boldsymbol{A},\,\boldsymbol{\Theta}\in\mathbb{R}^{B\times n_r\times n_t}:\sum_{b=1}^{B}\underline{\kappa}_b>\frac{BR}{M}\right\}\tag{414}$$

$$\mathcal{A}_{\mathcal{X}}^{c}\triangleq\left\{\boldsymbol{A},\,\boldsymbol{\Theta}\in\mathbb{R}^{B\times n_r\times n_t}:\sum_{b=1}^{B}\underline{\kappa}_b\leq\frac{BR}{M}\right\}.\tag{415}$$

Note that $\rho$ that solves the supremum in the RHS of (412) is given by $\rho^*=1$ if $\Lambda>0$ and $\rho^*=0$ if $\Lambda\leq0$. Then, we can evaluate the upper bound to the average error probability as follows:

$$P_{e,\text{ave}}\leq\mathbb{E}\left[2^{-LBE_r^Q(R,\hat{\mathsf{H}})}\right]\tag{416}$$

$$\dot{\leq}\int_{\mathcal{A}_{\mathcal{X}}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left(\mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}\alpha_{b,r,t}}\times\mathsf{snr}^{-\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}(\theta_{b,r,t}-d_e)}\times2^{-LM\Lambda}d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{H}}d\boldsymbol{\Phi}^{\boldsymbol{E}}\right)$$
$$+\int_{\mathcal{A}_{\mathcal{X}}^{c}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left(\mathsf{snr}^{-\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}\alpha_{b,r,t}}\times\mathsf{snr}^{-\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}(\theta_{b,r,t}-d_e)}d\boldsymbol{A}d\boldsymbol{\Theta}d\boldsymbol{\Phi}^{\boldsymbol{H}}d\boldsymbol{\Phi}^{\boldsymbol{E}}\right)\tag{417}$$

$$\doteq C_1\mathsf{snr}^{-d_1}+C_2\mathsf{snr}^{-d_2}\tag{418}$$

$$\doteq\mathsf{snr}^{-\min(d_1,d_2)}\tag{419}$$

where $C_1,C_2\doteq\mathsf{snr}^0$. It is easy to see that

$$d_2=\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_r\left\lceil B\left(n_t-\frac{R}{M}\right)\right\rceil\tag{420}$$

is equivalent to the scaled Singleton bound [6] up to the discontinuity points. This is exactly the same as (213) when replacing $<$ in the outage set with $\leq$. On the other hand, following the same steps as in the converse analysis, we arrive to the following result for $d_1$:

$$d_1=\inf_{\mathcal{A}_{\mathcal{X}}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\,\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}\alpha_{b,r,t}+\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}(\theta_{b,r,t}-d_e)+\frac{LM\Lambda\log2}{\log\mathsf{snr}}\right\}.\tag{421}$$

If both $M$ and $L$ are not growing with $\log\mathsf{snr}$, it is clearly seen that $d_1=0$ as the SNR tends to infinity. Assume that $M$ is fixed and $L(\mathsf{snr})=\omega\log\mathsf{snr}$, $\omega\geq0$. Hence, we can write $d_1$ as

$$d_1=\inf_{\mathcal{A}_{\mathcal{X}}\cap\{\boldsymbol{A}\succeq\boldsymbol{0},\,\boldsymbol{\Theta}\succeq d_e\times\boldsymbol{1}\}}\left\{\left(1+\frac{\tau}{2}\right)\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}\alpha_{b,r,t}+\sum_{b=1}^{B}\sum_{r=1}^{n_r}\sum_{t=1}^{n_t}(\theta_{b,r,t}-d_e)+\omega M\Lambda\log2\right\}.\tag{422}$$

By letting $\epsilon,\delta\downarrow0$ to achieve a tight SNR exponent lower bound, the optimal solution of $\boldsymbol{\Theta}$ is given by $\boldsymbol{\Theta}^*=d_e\times\boldsymbol{1}$ and for $\boldsymbol{A}$ is given by evaluating the intersection of $\boldsymbol{A}\succeq\boldsymbol{0}$ and $\mathcal{A}_{\mathcal{X}}$. This yields the following solution of $d_1$

$$d_1=\inf_{1+\left\lfloor\frac{BR}{M}\right\rfloor\leq K\leq Bn_t}d_1(K)\tag{423}$$

where

$$d_1(K)=\omega M\log2\left(K-\frac{BR}{M}\right)+\min(1,d_e)\times\left(1+\frac{\tau}{2}\right)n_r(Bn_t-K).\tag{424}$$

Note that the derivative of $d_1(K)$ with respect to $K$ is given by

$$\frac{\partial d_1(K)}{\partial K}=\omega M\log2-\min(1,d_e)\left(1+\frac{\tau}{2}\right)n_r.\tag{425}$$

It follows that the value of $K$ that solves the infimum in (423) is given by

$$K^*=Bn_t\tag{426}$$

if $\omega M\log2<\min(1,d_e)\left(1+\frac{\tau}{2}\right)n_r$, and

$$K^*=1+\left\lfloor\frac{BR}{M}\right\rfloor\tag{427}$$

if $\omega M\log2\geq\min(1,d_e)\left(1+\frac{\tau}{2}\right)n_r$.

We are interested in the interval of $\omega$ for which $d_1\geq d_2$ as this is the point where the SNR exponent of discrete-input random codes is tight with the generalized outage diversity up to the discontinuity points of the Singleton bound [6]. Note that from (420), (424), (426) and (427), we deduce that $d_1\geq d_2$ is only possible with $\omega M\log2\geq\min(1,d_e)\left(1+\frac{\tau}{2}\right)n_r$. This implies that $K^*=1+\left\lfloor\frac{BR}{M}\right\rfloor$ and $d_1=d_1(K^*)$. It follows that by comparing $d_1(K^*)$ and $d_2$, we obtain the following threshold on $\omega$ for which $d_1(K^*)\geq d_2$

$$\omega\geq\frac{1}{M\log2}\cdot\frac{\min(1,d_e)\cdot\left(1+\frac{\tau}{2}\right)n_r}{1+\left\lfloor\frac{BR}{M}\right\rfloor-\frac{BR}{M}}.\tag{428}$$

With this $\omega$, we can achieve the scaled Singleton bound diversity up to the discontinuity points. Furthermore, using $K^*$ in (426) and (427), a complete characterization of the achievable SNR exponent with discrete-input random codes can be obtained and it is given in (64).

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the associate editor for their constructive comments that helped in improving the paper.

## REFERENCES

[1] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.
[2] E. Biglieri, J. G. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
[3] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.
[4] R. Knopp and P. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.
[5] E. Malkamäki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, Mar. 1999.
[6] A. Guillén i Fàbregas and G. Caire, "Coded modulation in the block-fading channel: Coding theorems and code construction," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 91–114, Jan. 2006.
[7] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "A tight lower bound to the outage probability of discrete-input block-fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4314–4322, Nov. 2007.
[8] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
[9] J. Boutros, E. C. Strinati, and A. Guillén i Fàbregas, "Turbo code design for block fading channels," in *Proc. 42nd Annual Allerton Conference on Communication, Control and Computing*, Allerton, IL, Sep. 2004.
[10] J. Boutros, A. Guillén i Fàbregas, E. Biglieri, and G. Zémor, "Low-density parity-check codes for nonergodic block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4286–4300, Sep. 2010.
[11] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[13] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEÜ Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, 1993.
[14] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. and Trends on Commun. and Inf. Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
[15] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
[16] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
[17] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.
[18] H. Weingarten, Y. Steinberg, and S. Shamai, "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
[19] K. D. Nguyen, "Adaptive Transmission for Block-Fading Channels," Ph.D. dissertation, Inst. for Telecommun. Res., Univ. South Australia, Adelaide, 2009.
[20] N. Letzepis and A. Guillén i Fàbregas, "Outage probability of the Gaussian MIMO free space optical channel with PPM," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3682–3690, Dec. 2009.
[21] N. Letzepis, K. D. Nguyen, A. Guillén i Fàbregas, and W. Cowley, "Outage analysis of the hybrid free-space optical and radio-frequency channel," *IEEE J. Select. Areas Commun.*, vol. 27, no. 9, pp. 1709–1719, Dec. 2009.
[22] N. Letzepis and A. Guillén i Fàbregas, "Outage probability of the free space optical channel with doubly stochastic scintillation," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 2899–2902, Oct. 2009.
[23] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1549–1557, Apr. 2007.
[24] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," in *Proc. Conf. Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2006.
[25] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.
[26] E. Biglieri, *Coding for Wireless Channels*. New York: Springer, 2005.
[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985.
[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.
[29] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 357–359, May 1973.
[30] K. D. Nguyen, L. K. Rasmussen, A. Guillén i Fàbregas, and N. Letzepis, "MIMO ARQ with multibit feedback: Outage analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, Feb. 2012.
[31] A. Lapidoth and S. Shamai, "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
[32] S. Dolinar, D. Divsalar, J. Hamkins, and F. Pollara, Capacity of Pulse-Position Modulation (PPM) on Gaussian and Webb channels, JPL TMO Progress Report 42-142, Aug. 2000. [Online]. Available: lasers.jpl.nasa.gov/PAPERS/OSA/142h.pdf.
[33] A. Dembo and O. Zeitouni, "Large deviations techniques and applications," in *Applic. of Math.*, 2nd ed. New York: Springer-Verlag, 1998, vol. 38.
[34] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. Cambridge, U.K.: Cambridge University Press, 1934.
[35] R. Durrett, *Probability: Theory and Examples*, 2nd ed. Belmont, CA: Duxbury, 1996.
[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.
[37] H. L. Royden, *Real Analysis*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[38] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[39] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," Ph.D. dissertation, MIT, Cambridge, MA, 1989.
[40] N. Bourbaki, "Topological vector spaces: Chapters 1–5," in *Elements of Mathematics*, 1st ed. New York: Springer-Verlag, 1987.

**A. Taufiq Asyhari** (S'09) received his B.Eng. (Hons.) degree from Nanyang Technological University in 2007. Since 2008, he has been pursuing his Ph.D. at the Department of Engineering, University of Cambridge, where he is also affiliated to Trinity Hall. He held a visiting appointment at the National Chiao Tung University, Hsinchu, Taiwan from September–October 2010. His research interests are in the area of information theory and communication theory, and their applications to wireless communications.

**Albert Guillén i Fàbregas** (S'01–M'05–SM'09) was born in Barcelona, Catalunya, Spain, in 1974. He received the Telecommunication Engineering degree and the Electronics Engineering degree from the Universitat Politècnica de Catalunya, and the Politecnico di Torino, Torino, Italy, respectively, in 1999, and the Ph.D. degree in communication systems from Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2004.

He is Research Professor of the Institució Catalana de Recerca i Estudis Avançats (ICREA) hosted at the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain, and a Reader (on leave) at the Department of Engineering, University of Cambridge, Cambridge, United Kingdom, where he is also a Fellow of Trinity Hall. He has held positions at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom and the University of South Australia. He has held visiting appointments at Centrum Wiskunde and Informatica, Ecole Nationale Supérieure des Télécommunications, Texas A&M University at Qatar, Universitat Pompeu Fabra, and the University of South Australia. His research interests are in communication theory, information theory, coding theory, digital modulation, and signal processing techniques with wireless applications.

Dr. Guillén i Fàbregas is a recipient of the Starting Grant of the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference (EUSIPCO), the 2004 Nokia Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a pre-doctoral Research Fellowship of the Spanish Ministry of Education to join ESA. He has been an Editor of the IEEE Transactions on Wireless Communications. He is also a Senior Member of IEEE (Information Theory and Communications Societies).