MIMO Block-Fading Channels with Mismatched CSI

A. Taufiq Asyhari and Albert Guillén i Fàbregas

Abstract

We study transmission over multiple-input multiple-output (MIMO) block-fading channels with imperfect channel state information (CSI) at both the transmitter and receiver. Specifically, based on mismatched decoding theory for a fixed channel realization, we investigate the largest achievable rates with independent and identically distributed inputs and a nearest neighbor decoder. We then study the corresponding information outage probability in the high signal-to-noise ratio (SNR) regime and analyze the interplay between estimation error variances at the transmitter and at the receiver to determine the optimal outage exponent, defined as the high-SNR slope of the outage probability plotted in a logarithmic-logarithmic scale against the SNR. We demonstrate that despite operating with imperfect CSI, power adaptation can offer substantial gains in terms of outage exponent.

Index Terms

Block fading, diversity, generalized mutual information, imperfect channel state information, MIMO, mismatched decoding, multiple antenna, nearest neighbor decoding, outage probability, outage exponent, power adaptation.

I. INTRODUCTION

The block-fading channel [1] is a commonly used model to study delay-limited transmission over slowly-varying fading channels. In such a channel, each codeword spans only a finite number of fading blocks. The important feature of this model is that the channel remains constant during a block (which consists of several symbols) and varies from block to block according to a certain probability distribution. Despite being a very simple model, it is a useful model that captures the non-ergodic nature of delay-limited transmission.

The information-outage probability [1] [2] is the fundamental performance indicator for block-fading channels undergoing non-ergodic fading. A key aspect that affects the outage performance is the availability of channel state information (CSI). A large body of works on block-fading channels assume perfect CSI [3]–[9]. Perfect CSI at the receiver (CSIR) allows for coherent decoding. Under perfect CSIR, nearest neighbor decoding is optimal in the sense that it minimizes the error probability [4] [10]. If CSI at the transmitter (CSIT) is available, the information-outage performance can be improved by adaptive transmission, which in some cases leads to zero outage probability (see [5]–[9], [11]–[14] and references therein).

Different CSI setups for block-fading channels have been studied in the literature. At the transmitter end, CSIT for power adaptation can be causal or non-causal depending on the system constraints. Full non-causal CSIT [6], [8], [9], [15] is inspired by block-fading models that resemble instantaneous parallel channels such as in orthogonal frequency division multiplexing (OFDM) systems. In this setup, estimates of all fading coefficients are available prior to transmission. Causal CSIT is motivated by block-fading channels with time-domain transmission. In this setup, considered in [7], [16], [17] with perfect CSIT, only past fading estimates may be available at the transmitter. In some systems, a number of future fading estimates may be available at the transmitter prior to transmission, effectively implementing predictive-CSIT power allocation [18]. At the receiver end, given the nature of block transmission, the decoder has to wait for the estimates of all fading coefficients that affect the current codeword. References [3]–[9], [15]–[17] have studied block-fading channels with perfect CSIR.

The perfect CSI assumption implies that the nearest neighbor decoder at the receiver and the power controller at the transmitter have access to the exact fading coefficients. This is too optimistic and impossible to guarantee in practical systems. In this paper, we drop the perfect CSI

assumption and study the reliability of a communication system that employs nearest neighbor decoding and power adaptation with imperfect CSI. Furthermore, we model the imperfect CSIT as a noisy version of the actual fading and the imperfect CSIR as a noisy version of a parameter that captures both the adapted power level and fading. This CSIR model is particularly closer to practice as compared to the setups used in the vast majority of previous works, e.g., [4]–[8], [15], [19]–[21], where CSIR is only characterized by fading coefficients. This is so because in practice not only the fading but also the data power is unknown *a priori*, especially when the power is continually adapted to channel conditions.

This paper proposes a unified framework for studying mismatched CSI at both communicating terminals in block-fading channels. In particular, we study the generalized mutual information (GMI) [22] [23] of the nearest neighbor decoder. The GMI is an achievable rate when a fixed decoding rule—which is not necessarily the channel likelihood—is employed [24]. In our context, the GMI is the maximum rate for which under a fixed fading, data power and their estimates, the error probability of independent and identically distributed (i.i.d.) codebooks is guaranteed to vanish as the codeword length tends to infinity. As the fading, data power and their estimates vary from one codeword to another, the generalized outage probability—the probability that the instantaneous GMI is less than the target coding rate R—provides an error probability benchmark of a good code for sufficiently large block length.

We focus the analysis in the high signal-to-noise ratio (SNR) regime. In particular, we study the outage exponent defined by

$$d \triangleq \lim_{P \to \infty} -\frac{\log P_{\text{gout}}(R)}{\log P} \tag{1}$$

where $P_{\text{gout}}(R)$ is the generalized outage probability and P is the average data SNR. For full non-causal CSIT power allocation, we obtain the outage exponent $d_{\text{icsi}}^{\text{f}}$ as

$$d_{\text{icsi}}^{\text{f}} = d_{\text{csir}}^{\text{u}} \left(\min(1, \hat{d}_e) + d_{\text{csir}}^{\text{u}} \tilde{d}_e \right) \tag{2}$$

where the CSIT and CSIR error variances are parametrized as $P^{-\tilde{d}_e}$ and $P^{-(\hat{d}_e - 1)}$, respectively, and $d_{\text{csir}}^{\text{u}}$ is the perfect CSIR outage exponent with uniform power allocation across all fading blocks and transmit antennas. This relationship holds for both Gaussian and discrete inputs. This result generalizes all previous outage exponent results in [4], [6], [8], [15], [21] and suggests that

power adaptation based on imperfect CSIT offers performance improvements irrespectively how poor the CSIR quality is. The significance of employing full non-causal CSIT power adaptation is well illustrated by the outage exponent in (2). While having CSIR only provides gains that scale linearly with $d_{\text{csir}}^{\text{u}}$, i.e., $\min(1, \hat{d}_e)d_{\text{csir}}^{\text{u}}$, investing efforts in obtaining minimally accurate CSIT ($\tilde{d}_e > 0$) yields gains that scale quadratically with $d_{\text{csir}}^{\text{u}}$, i.e., $(d_{\text{csir}}^{\text{u}})^2 \tilde{d}_e$. The region of mismatched CSIR exponent $\hat{d}_e \geq 1$ corresponds to the "nearly perfect" CSIR criterion as the outage exponent (2) coincides with the perfect-CSIR outage exponent [6] [8]. This criterion strengthens a similar criterion in our earlier work [21] that considered uniform power allocation. As shown in the paper, for causal and predictive CSIT, in addition to mismatched CSIT and CSIR exponents $\tilde{d}_e$, $\hat{d}_e$, the delay and the prediction parameters also affect the outage exponent. In principle, full non-causal, causal- and predictive-CSIT power adaptation reduce communication outages through inverting the channel effects and controlling the CSIR quality for decoding purposes.

The rest of the paper is organized as follows. Section II introduces the system model. Section III provides the relevant information-theoretic background. Section IV presents our main results on outage exponents. Discussions and connections with previous results in the literature are provided in Section V. Section VI summarizes the important points of the paper and provides some practical implications for the results. Proofs of the main results can be found in the appendices.

Notation: Scalar, vector and matrix variables are characterized with normal (non-boldfaced), boldfaced lowercase, and sans serif uppercase letters, respectively. I_m denotes the $m \times m$ identity matrix. The symbols †, $\| \cdot \|$ and $\| \cdot \|_F$ represent the conjugate transpose, the Euclidean norm of a vector and the Frobenius norm of a matrix, respectively. Random variables are denoted by uppercase letters X and realizations by lowercase letters x; random vectors are denoted by boldfaced uppercase letters $\boldsymbol{X}$ and realizations by boldfaced lowercase letters $\boldsymbol{x}$; random matrices are denoted by blackboard upper case letters $\mathbb{X}$ and realizations by sans serif uppercase letters 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 [25]. 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, and $[x]_+ \triangleq \max(0, x)$.
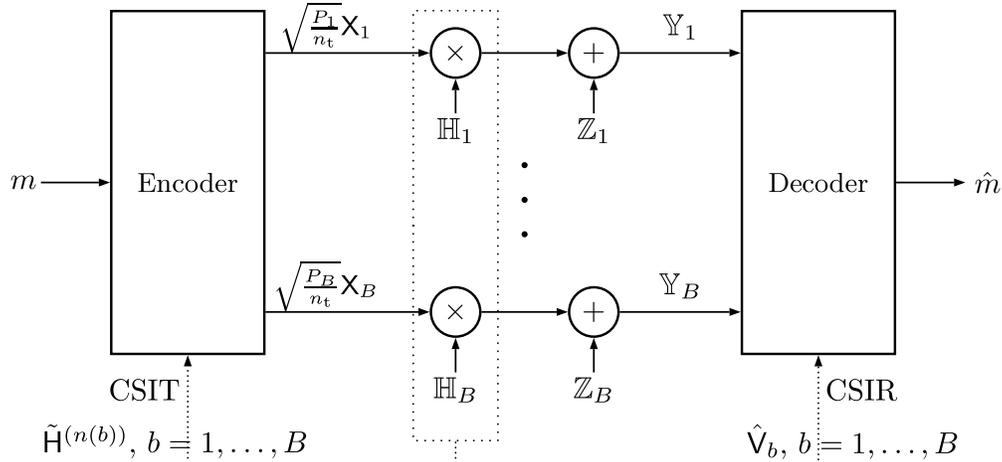
Fig. 1. System model for MIMO block-fading channels with imperfect CSI at both terminals.

II. System Model

The system model is depicted in Fig. 1. We consider a MIMO block-fading channel with n_t transmit antennas, n_r receive antennas and B fading blocks per codeword. The output of the channel for block b is an $n_\mathrm{r} \times J$-dimensional random matrix

$$\mathbb{Y}_b = \mathbb{H}_b \mathsf{P}_b^{\frac{1}{2}} \mathsf{X}_b + \mathbb{Z}_b, \quad b = 1, \dots, B \tag{3}$$

where $\mathbb{Z}_b$ is the $n_\mathrm{r} \times J$-dimensional random noise matrix and $\mathsf{X}_b \in \mathcal{X}^{n_\mathrm{t} \times J}$ is the transmitted signal matrix; J and $\mathcal{X} \subseteq \mathbb{C}$ denote the channel block length and the signal constellation, respectively. We assume that the entries of $\mathbb{Z}_b$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. The $(n_\mathrm{r} \times n_\mathrm{t})$-dimensional random matrix $\mathbb{H}_b$ denotes the fading for block b and is assumed to be i.i.d. from block to block. Furthermore, the $(n_\mathrm{r} \times n_\mathrm{t})$ entries of $\mathbb{H}_b$ are i.i.d. zero-mean unit-variance complex Gaussian random variables. It follows that the magnitude for each entry of $\mathbb{H}_b$ is Rayleigh distributed.

A codeword that represents a message $m \in \{1, \dots, 2^{BJR}\}$ to be transmitted is denoted by $\mathsf{X}(m) = \{\mathsf{X}_1(m), \dots, \mathsf{X}_B(m)\}$ where R is the target coding rate. The BJ entries of X are drawn i.i.d. from a probability distribution over $\mathcal{X}^{n_\mathrm{t}}$. Herein we focus on zero-mean unit-variance Gaussian inputs and equiprobable discrete inputs of size $|\mathcal{X}| = 2^M$. We further assume that the coding rate R is a fixed positive constant; hence the *multiplexing gain* [25] is zero. Finally, we

assume that the constellations have average unit energy.

We focus our attention on an additive-noise CSI model. This model accurately represents pilot-aided channel estimation that exploits channel reciprocity [26] [27] and utilizes pilot symbols at both the forward and reverse link of a time-division duplex (TDD) system.

At the transmitter, the CSIT is given by the noisy fading matrix

$$\text{CSIT} \quad \tilde{\mathbb{H}}_b = \mathbb{H}_b + \tilde{\mathbb{E}}_b \tag{4}$$

where $\tilde{\mathbb{E}}_b$ is the CSIT noise matrix whose entries are assumed to be i.i.d. complex Gaussian random variables with zero mean and variance given by $\tilde{\varepsilon}^2 = P^{-\tilde{d}_e}$, and where P is the average data SNR. This model considers the CSIT quality that improves with the average transmission power P for the forward link of the TDD system. The CSIT quality is typically proportional to the reciprocal of the power for the reverse link transmission. Parameter $\tilde{d}_e > 0$, denoting the CSIT-error diversity, is thus incorporated to reflect the difference between the forward and reverse link powers and to gain further insights on the impacts of CSIT quality in the system performance.

The CSIT acquired by the transmitter is used to perform power allocation across fading blocks. The power matrix $\mathsf{P}_b \in \mathbb{R}^{n_\mathrm{t} \times n_\mathrm{t}}$ is a diagonal matrix. Power allocation at block b (6) is performed after knowing the noisy CSIT matrix

$$\tilde{\mathsf{H}}^{(n(b))} = \left[\tilde{\mathsf{H}}_1, \ldots, \tilde{\mathsf{H}}_{n(b)} \right] \tag{5}$$

where $n(b)$ is the number of fading blocks used for power adaptation. In particular, we use a scaled identity power matrix

$$\mathsf{P}_b\left(\tilde{\mathsf{H}}^{(n(b))} \right) = \frac{P_b\left(\tilde{\mathsf{H}}^{(n(b))} \right)}{n_\mathrm{t}} \mathsf{I}_{n_\mathrm{t}} \tag{6}$$

where $P_b(\tilde{\mathsf{H}}^{(n(b))})$ denotes the scalar power coefficient. According to [6], [8], [21, App. D], [25], the scaled identity power matrix is optimum when it comes to outage exponent (1). Depending on $n(b)$ we have the following cases.

- *Full non-causal-CSIT* power allocation if $n(b) = B$ for all $b = 1, \ldots, B$. Fading estimates for the whole B blocks in a codeword are available at the transmitter prior to transmission. This setup is relevant for orthogonal multi-carrier transmission such as OFDM, where data

transmission occurs in the frequency domain and full information about all fading estimates is used for power allocation.

- *Causal-CSIT* power allocation if $n(b) = b - \tau_{\mathrm{d}}$ with a fixed delay $\tau_{\mathrm{d}} > 0$ for any $b = 1, \ldots, B$. This corresponds to CSIT being limited only to the past imperfect fading estimates due to the delay τ_{d}. This setup is motivated by block-fading channels with time-domain transmission for which only past fading estimates may be available at the transmitter.

- *Predictive-CSIT* power allocation if $n(b) = b + \tau_{\mathrm{f}}$ with a fixed $\tau_{\mathrm{f}} \geq 0$ (indicating the number of predicted fading blocks) for any $b = 1, \ldots, B$. This corresponds to CSIT including past, current and a number of predicted future fading estimates. This setup is also relevant for instantaneous parallel transmission such as in OFDM where (possibly) not all fading blocks are used for power allocation. More specifically, for each fading block $b = 1, \ldots, B$, only $(n(b) = b + \tau_{\mathrm{f}})$ fading matrices are used for allocating power at block b.

For the above power allocation schemes, the corresponding long-term average power constraint is given by

$$\mathbb{E}\left[\frac{1}{B}\sum_{b=1}^{B} \mathrm{tr}\left(\mathsf{P}_b\big(\tilde{\mathbb{H}}^{(n(b))}\big)\right)\right] \leq P. \tag{7}$$

At the receiver, two CSI elements, namely the adapted power levels and fading coefficients, are crucial for reliable decoding. As both components are not available *a priori*, the receiver needs to estimate them. Instead of estimating both components separately, the receiver can choose to estimate both variables in a single entity, namely $\mathbb{V}_b = \mathbb{H}_b \mathsf{P}_b^{\frac{1}{2}}$. This give rises to a noisy CSIR model

$$\mathrm{CSIR} \quad \hat{\mathbb{V}}_b = \mathbb{V}_b + \hat{\mathbb{E}}_b \tag{8}$$

where $\hat{\mathbb{E}}_b$ is the CSIR noise matrix whose entries are independent of $\tilde{\mathbb{E}}_b$ and assumed to be i.i.d. complex Gaussian random variables with zero mean and variance given by $\hat{\varepsilon}^2 = P^{-(\hat{d}_e - 1)}$. Parameter $\hat{d}_e > 0$ denotes the CSIR-error diversity and indicates the quality improvement factor of the CSIR. Furthermore, the exponent $\hat{d}_e - 1$ can be related to the average power used for channel estimation (See Section V-A for the details).

This CSIR closely resembles setups used in pilot-aided channel estimation where instead of estimating fading alone, the receiver jointly estimates both power and fading. This generalizes

various models used in [6], [8], [13], [18], [21], [28].

Nearest neighbor decoding 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 is not available. The decoder treats the imperfect CSIR as if it were perfect. The decoder first computes the following metric for a given $\mathsf{Y} = [\mathsf{Y}_1, \ldots, \mathsf{Y}_B]$ and imperfect CSIR $\hat{\mathbb{V}} = \hat{\mathsf{V}} = \left[\hat{\mathsf{V}}_1, \ldots, \hat{\mathsf{V}}_B\right]$,

$$Q\left(\mathsf{Y}, \hat{\mathsf{V}}, \mathsf{X}(m)\right) \propto \exp\left(-\sum_{b=1}^{B} \left\|\mathsf{Y}_b - \hat{\mathsf{V}}\mathsf{X}_b(m)\right\|_F^2\right) \tag{9}$$

and then outputs

$$\hat{m} = \arg\max_{m \in \{1, \ldots, 2^{BJR}\}} Q\left(\mathsf{Y}, \hat{\mathsf{V}}, \mathsf{X}(m)\right). \tag{10}$$

III. INFORMATION-THEORETIC PRELIMINARIES

For the nearest neighbor rule in (10), and for a given fading H, its corresponding CSIR $\hat{\mathsf{V}}$ and the power matrix P, the average error probability—averaged over the ensemble of random codebooks whose entries are generated i.i.d. with input distribution $P_{\boldsymbol{X}}(\boldsymbol{x})$ (*i.i.d. codebooks*)—can be upper-bounded using the generalized Gallager bound [22] (see also [21] [23])

$$\bar{P}_e\left(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}\right) \leq 2^{-BJE_r^Q\left(R, \mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}\right)} \tag{11}$$

where the mismatched decoding error exponent is

$$E_r^Q\left(R, \mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}\right) = \sup_{\substack{s > 0 \\ 0 \leq \rho \leq 1}} \frac{1}{B}\sum_{b=1}^{B} E_0^Q\left(s, \rho; \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\left(\tilde{\mathsf{H}}^{(n(b))}\right)\right) - \rho R \tag{12}$$

and

$$E_0^Q\left(s, \rho; \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\left(\tilde{\mathsf{H}}^{(n(b))}\right)\right) =$$

$$-\log_2 \mathbb{E}\left[\left(\mathbb{E}\left[\left(\frac{Q\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}'\right)}{Q\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}\right)}\right)^s \middle| \boldsymbol{X}, \boldsymbol{Y}, \mathbb{H}_b, \hat{\mathbb{V}}_b, \tilde{\mathbb{H}}^{(n(b))}\right]\right)^\rho \middle| \mathbb{H}_b = \mathsf{H}_b, \hat{\mathbb{V}}_b = \hat{\mathsf{V}}_b, \tilde{\mathbb{H}}^{(n(b))} = \tilde{\mathsf{H}}^{(n(b))}\right] \tag{13}$$

is the generalized Gallager function [22].

For sufficiently large block length J, the error behavior of (11) depends on the positivity of $E_r^Q(R, \mathsf{H}, \hat{\mathsf{V}}, \mathsf{P})$. It has been shown in [21] [22] that $E_r^Q(R, \mathsf{H}, \hat{\mathsf{V}}, \mathsf{P})$ is strictly positive for $R \leq I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}) - \varepsilon$ for some $\varepsilon > 0$ and zero otherwise, where

$$I^{\mathrm{gmi}}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}) = \sup_{s>0} \frac{1}{B} \sum_{b=1}^{B} I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b) \tag{14}$$

is the generalized mutual information (GMI) for fading H, CSIR $\hat{\mathsf{V}}$ and power level P, and

$$I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b) = \mathbb{E}\left[\log_2 \frac{Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}\right)}{\mathbb{E}\left[Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}'\right) \middle| \boldsymbol{Y}, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b \right]} \middle| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b \right]. \tag{15}$$

As the fading, CSIR and power level vary from block to block, the generalized outage probability, defined by

$$P_{\mathrm{gout}}(R) \triangleq \Pr\left\{ I^{\mathrm{gmi}}(\mathbb{H}, \hat{\mathbb{V}}, \mathsf{P}) < R \right\}, \tag{16}$$

characterizes the Gallager bound on the RHS of (11) as the block length J tends to infinity. Furthermore, this random coding characterization implies that in general, $P_{\mathrm{gout}}(R)$ is only an achievable error probability, but may not be the smallest error probability due to the absence of a general converse theorem. However, for i.i.d. codebooks, it has been shown in [21] based on the results in [29]–[31] that $P_{\mathrm{gout}}(R)$ is the smallest possible error probability for block-fading channels. Therefore, for i.i.d. codebooks with sufficiently large block length, $P_{\mathrm{gout}}(R)$ is the natural fundamental limit of the channel.

We are interested in characterizing the behavior of $P_{\mathrm{gout}}(R)$ at high SNR. One important figure of merit is the outage diversity or outage exponent defined as in (1), i.e.,

$$d \triangleq \lim_{P \to \infty} -\frac{\log P_{\mathrm{gout}}(R)}{\log P}. \tag{17}$$

In words, the d is the high-SNR slope of $P_{\mathrm{gout}}(R)$ plotted in a logarithmic-logarithmic scale against the SNR. Our previous work [21] showed that with uniform power allocation the outage exponent is a function of the perfect-CSIR outage exponent and the quality of the imperfect

CSIR $\hat{d}_e$ as[1]

$$d_{\text{icsir}}^{\text{u}} = \min(1, \hat{d}_e) \times d_{\text{csir}}^{\text{u}} \tag{18}$$

where superscript $^{\text{u}}$ denotes the uniform power allocation, the subscript $_{\text{csir}}$ denotes perfect CSIR. From [5] [32], we have that

$$d_{\text{csir}}^{\text{u}} = \begin{cases} Bn_{\text{t}}n_{\text{r}}, & \text{for Gaussian inputs} \\ n_{\text{r}}\left(1 + \left\lfloor B\left(n_{\text{t}} - \frac{R}{M}\right)\right\rfloor\right), & \text{for discrete inputs} \end{cases} \tag{19}$$

where $M \triangleq \log_2|\mathcal{X}|$. The result in (18) implies that if the variance of the fading estimation error at the receiver decays faster than or equal to $1/P$, the perfect-CSIR diversity is achievable. Otherwise, the imperfect-CSIR diversity is smaller than the perfect-CSIR diversity.

If CSIT is available, then the transmitter can adapt its transmission power to minimize the generalized outage probability. The underlying idea is that in a very bad channel realization, power can be saved and used when channel conditions improve. References [9] [33] showed that if perfect CSI is available at both terminals, then zero outage is possible, implying that the delay-limited capacity [34] is positive. References [6] [8] extended the results to perfect CSIR and imperfect full non-causal-CSIT setup. In this case, the outage exponent is given by

$$d_{\text{icsit}}^{\text{f}} = d_{\text{csir}}^{\text{u}}\left(1 + d_{\text{csir}}^{\text{u}}\tilde{d}_e\right) \tag{20}$$

where the superscript $^{\text{f}}$ is used to indicate results with full non-causal CSIT. Assuming perfect CSIR, reference [18] considered cases where imperfect causal or predictive CSIT is available. In those cases, the outage exponent is given as a function of the CSIT-error diversity $\tilde{d}_e$ and the CSIT delay τ_{d} or the CSIT prediction parameter τ_{f}.

In this work, we find the outage exponents with imperfect CSI at both ends using nearest neighbor decoding and power allocation. In particular, the power allocation algorithm is given

---

[1]For uniform power allocation where power is not adapted to CSIT, it is reasonable to assume that the power level is perfectly known at the receiver.

by the solution to the following optimization problem

$$\text{minimize} \quad P_{\text{gout}}(R)$$

$$\text{subject to} \quad \mathbb{E}\left[\frac{1}{B}\sum_{b=1}^{B}\text{tr}\left(\mathsf{P}_b\left(\tilde{\mathbb{H}}^{(n(b))}\right)\right)\right] \leq P \quad (21)$$

$$\text{diag}\left(\mathsf{P}_b\left(\tilde{\mathbb{H}}^{(n(b))}\right)\right) \succeq \mathbf{0}, \quad b = 1,\ldots,B.$$

Solving the above optimization problem can be difficult in general. Given our CSIT and CSIR models, finding the minimum-outage solution can be difficult since $P_{\text{gout}}(R)$ depends on the probability distribution of the GMI, which is difficult to evaluate analytically. Nevertheless, we will see that despite this difficulty, studying the behavior of the optimal solution at high SNR is possible.

IV. OUTAGE EXPONENTS

The solution to the power allocation in (21) depends on whether the CSIT is full non-causal, causal or predictive. Therefore, we will separately study the outage exponent for each type of CSIT.

A. Full Non-Causal-CSIT Power Allocation

Theorem 1 (Full Non-Causal CSIT): For full non-causal CSIT (where $n(b) = B$ in (5)), the outage exponent $d_{\text{icsi}}^{\text{f}}$ of MIMO block-fading channels with n_{t} transmit antennas, n_{r} receive antennas, B fading blocks, CSIT-error diversity $\tilde{d}_e$ and CSIR-error diversity $\hat{d}_e$ for Gaussian and discrete constellations is

$$d_{\text{icsi}}^{\text{f}} = d_{\text{csir}}^{\text{u}}\left(\min(1,\hat{d}_e) + d_{\text{csir}}^{\text{u}}\tilde{d}_e\right) \quad (22)$$

where $d_{\text{csir}}^{\text{u}}$ is the perfect-CSIR outage exponent with uniform power allocation, given in (19).

Proof: See Appendix B. ∎

The result in Theorem 1 highlights the roles of CSIT and CSIR qualities in determining the outage exponent with full non-causal CSIT power allocation. The mismatched-CSI outage exponent is completely determined by $d_{\text{csir}}^{\text{u}}$, the CSIT-error diversity $\tilde{d}_e$ and the CSIR-error diversity $\hat{d}_e$.

Fig. 2 illustrates the resulting outage exponent as a function of $\hat{d}_e$ and $\tilde{d}_e$. For a given CSIR quality (fixed $\hat{d}_e$), we observe that power adaptation with an acceptable CSIT quality ($\tilde{d}_e > 0$)
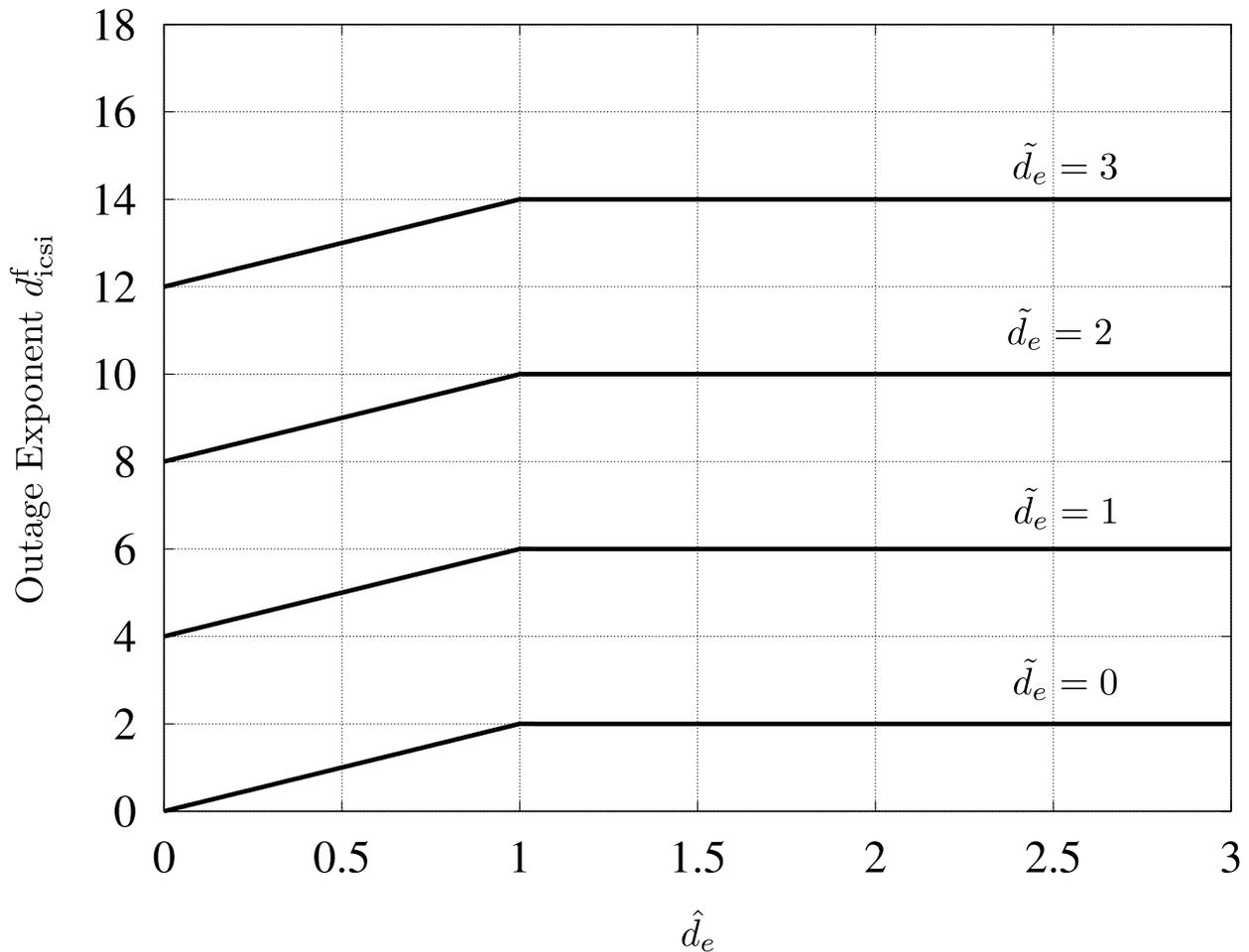
Fig. 2. Interplay of the CSIT-error diversity $\tilde{d}_e$, the CSIR-error diversity $\hat{d}_e$ and the outage exponents for *full non-causal CSIT* power allocation in 2×1 MIMO channels with Gaussian inputs.

will always be beneficial in improving the outage exponent. By comparing the result in Theorem 1 with equation (18), we see that the additional diversity gain due to power control is given by

$$d_{\text{icsi}}^{\text{f}} - d_{\text{icsir}}^{\text{u}} = (d_{\text{csir}}^{\text{u}})^2 \, \tilde{d}_e \tag{23}$$

which is linearly proportional to the CSIT quality.

For a given CSIT quality (fixed $\tilde{d}_e$), the outage exponent scales linearly with the CSIR-error diversity $\hat{d}_e$ up to the saturating point $\hat{d}_e = 1$. When $\hat{d}_e < 1$, unreliable CSIR affects the outage exponent and its impacts are particularly significant when the CSIT quality is poor (small $\tilde{d}_e$). Furthermore, when $\hat{d}_e < 1$ and $\tilde{d}_e = 0$, the mismatched CSIR dominates outage events as indicated by the outage exponent that is directly proportional to the CSIR-error diversity $\hat{d}_e$. This

observation is consistent with the explanation in [21] where CSIR errors introduce supplementary outage events (in addition to deep fades) and poor CSIR contributes to a smaller outage diversity. If $\hat{d}_e \geq 1$, then, irrespective how large $\hat{d}_e$, we achieve a diversity

$$d_{\text{icsi}}^{\text{f}} = d_{\text{csir}}^{\text{u}} \left(1 + d_{\text{csir}}^{\text{u}} \tilde{d}_e \right) \tag{24}$$

which coincides with the perfect-CSIR outage diversity (20), cf. [6] [8].

The aforementioned roles of $\tilde{d}_e$ and $\hat{d}_e$ on the outage exponent can be intuitively explained as follows. As discussed in [21, Sec. V], imperfect CSIR promotes additional outage events in addition to deep fades, yielding the CSIR-error diversity $\hat{d}_e$ in the outage exponent expression (18). The role of CSIT is to facilitate power adaptation in order to reduce the number of outage events, not only due to deep fades, but also due to erroneous CSIR. This benefit is well demonstrated in Fig. 2 where for any given CSIR quality, power adaptation is able to improve the outage exponent. How significant the improvement is depends on the CSIT quality.

The result in Theorem 1 is consistent with previous results in the literature. In particular, we recover the mismatched-CSIT perfect-CSIR outage exponent in [6] [8] by letting $\hat{d}_e \uparrow \infty$, and the no-CSIT mismatched-CSIR outage exponent in [21] by letting $\tilde{d}_e \downarrow 0$. The generality of Theorem 1 allows to strengthen the insights gained from the earlier works. One of the main conclusions in [6] [8] is that under perfect CSIR, where the exact fading and data power information is supposedly given at the receiver, power control—even with noisy CSIT—provides improvement on the outage performance. As demonstrated in Theorem 1, this insight continues to hold even when the CSIR that captures both fading and power level is noisy. This analysis suggests that even with unideal settings where CSI may not be perfectly known, the use of power adaptation for MIMO transmission is beneficial.

B. Causal-CSIT Power Allocation

Theorem 2 (Causal CSIT): Consider a MIMO block-fading channel with n_{t} transmit antennas, n_{r} receive antennas, B fading blocks, CSIT-error diversity $\tilde{d}_e$ and CSIR-error diversity $\hat{d}_e$. For causal CSIT with delay $\tau_{\text{d}} > 0$ (where $n(b) = b - \tau_{\text{d}}$ in (5)), the outage exponent $d_{\text{icsi}}^{\text{c}}$ for Gaussian inputs is given by

$$d_{\text{icsi}}^{\text{c}} = n_{\text{t}} n_{\text{r}} \sum_{b=1}^{B} \upsilon_b \tag{25}$$

where

$$
\upsilon_b \triangleq \begin{cases} \min\left(1, \hat{d}_e\right), & \text{for } b = 1, \ldots, \tau_\mathrm{d} \\ \min\left(1, \hat{d}_e\right) + n_\mathrm{t} n_\mathrm{r} \sum_{b'=1}^{b-\tau_\mathrm{d}} \min\left(\upsilon_{b'}, \tilde{d}_e\right), & \text{for } b = \tau_\mathrm{d} + 1, \ldots, B. \end{cases} \tag{26}
$$

On the other hand, the outage exponent $d_\mathrm{icsi}^\mathrm{c}$ for discrete inputs is given by

$$
d_\mathrm{icsi}^\mathrm{c} = n_\mathrm{t} n_\mathrm{r} \sum_{b=1}^{\hat{b}} \vartheta_b + n_\mathrm{r} \left(d^\ddagger - \hat{b} n_\mathrm{t}\right) \vartheta_{b+1} \tag{27}
$$

where

$$
d^\ddagger \triangleq B n_\mathrm{t} - \left\lceil \frac{BR}{M} \right\rceil + 1 \tag{28}
$$

$$
\hat{b} \triangleq \max_{\left\{b : \, b n_\mathrm{t} \le d^\ddagger\right\}} b \tag{29}
$$

$$
\vartheta_b \triangleq \begin{cases} \min\left(1, \hat{d}_e\right) & b = 1, \ldots, \min(\tau_\mathrm{d}, \hat{b} + 1) \\ \min\left(1, \hat{d}_e\right) + n_\mathrm{r} n_\mathrm{t} \sum_{b'=1}^{b-\tau_\mathrm{d}} \min(\vartheta_{b'}, \tilde{d}_e), & b = \min(\tau_\mathrm{d}, \hat{b} + 1) + 1, \ldots, \hat{b} + 1. \end{cases} \tag{30}
$$

Proof: See Appendix C. ∎

Causal CSIT corresponds to the case of having access to CSIT at previous fading blocks up to a certain delay. Theorem 2 captures the effects of the delay parameter τ_d and its interplay with the CSI quality on the outage exponent. The trade-off among the outage exponent, delay parameter τ_d, the CSIT quality indicator $\tilde{d}_e$ and the CSIR quality indicator $\hat{d}_e$ is described by the recursive functions of υ_b in (26) and ϑ_b in (30).

Due to the non-negativity of υ_b, ϑ_b, $b = 1, \ldots, B$ and $\tilde{d}_e$, we can show for both Gaussian and discrete inputs that

$$
\upsilon_b \ge \min(1, \hat{d}_e), \qquad b = 1, \ldots B, \tag{31}
$$

$$
\vartheta_b \ge \min(1, \hat{d}_e), \qquad b = 1, \ldots, \min(\tau_\mathrm{d}, \hat{b} + 1) \tag{32}
$$

with equality if $\tilde{d}_e = 0$ or

$$
\tau_\mathrm{d} \ge \tau_\mathrm{th} = \begin{cases} B & \text{for Gaussian inputs} \\ \hat{b} + 1 & \text{for discrete inputs.} \end{cases} \tag{33}
$$

Both cases $\tilde{d}_e = 0$ and $\tau_{\rm d} \geq \tau_{\rm th}$ yield the same outage exponent as uniform power allocation. For $\tilde{d}_e = 0$, this means that power adaptation is not effective due to poor CSIT quality. The case $\tau_{\rm d} \geq \tau_{\rm th}$ corresponds to ineffective power adaptation due to a large CSIT delay. This later case is similar to the result in [7]. Indeed, the result in [7] is an instance of Theorem 2 with infinite $\hat{d}_e$.

Similarly to the full non-causal-CSIT case, causal-CSIT power allocation improves the outage exponent by mitigating communication outages due to fading and noisy CSIR. For any given CSIR quality (fixed $\hat{d}_e$), the improvement is possible provided that the CSIT quality is acceptable ($\tilde{d}_e > 0$) and the CSIT delay is not too long ($\tau_{\rm d} < \tau_{\rm th}$).

Despite a fewer number of CSIT blocks used for power adaptatation as compared to the full non-causal-CSIT case, it is still possible to improve the communication performance using causal CSIT. The intuitive explanation is as follows. As a codeword spans B fading blocks, any part of the codeword may experience severe fading or CSIR errors in several initial blocks. Power adaptation in the subsequent blocks—where information about previous fading estimates is available—can be employed to ensure that the remaining blocks will be less affected by fading and imperfect CSIR. Due to a fewer number of CSIT blocks for power adaptation, the outage exponent for the causal-CSIT case is smaller than that for the full non-causal-CSIT case.

With perfect CSI at both the transmitter and receiver ($\hat{d}_e \uparrow \infty$ and $\tilde{d}_e \uparrow \infty$), the outage exponent with causal CSIT is always finite for any value of $\tau_{\rm d} > 0$. This means that at high SNR, the slope of the generalized outage probability with respect to $\log P$ is finite and zero outage is not possible with finite SNR. It thus follows that the delay-limited capacity [34] is zero.

C. Predictive-CSIT Power Allocation

Theorem 3 (Predictive CSIT): Consider a MIMO block-fading channel with $n_{\rm t}$ transmit antennas, $n_{\rm r}$ receive antennas, B fading blocks, CSIT-error diversity $\tilde{d}_e$ and CSIR-error diversity $\hat{d}_e$. For predictive CSIT (where $n(b) = b + \tau_{\rm f}$ in (5), $\tau_{\rm f} \geq 0$), the outage exponent $d_{\rm icsi}^{\rm p}$ for Gaussian inputs is given by

$$d_{\rm icsi}^{\rm p} = n_{\rm t} n_{\rm r} \sum_{b=1}^{B} \Big(\min(1, \hat{d}_e) + n_{\rm t} n_{\rm r} \min\left(B, b + \tau_{\rm f}\right) \tilde{d}_e \Big). \tag{34}$$

On the other hand, the outage exponent $d_{\text{icsi}}^{\text{p}}$ for discrete inputs is given by

$$d_{\text{icsi}}^{\text{p}} = n_{\text{t}} n_{\text{r}} \sum_{b=1}^{\hat{b}} \eta_b + n_{\text{r}} \left(d^{\ddagger} - \hat{b} n_{\text{t}} \right) \eta_{b+1} \tag{35}$$

where $d^{\ddagger}$ and $\hat{b}$ are defined in (28) and (29), respectively, and

$$\eta_b \triangleq \begin{cases} \min(1, \hat{d}_e) + n_{\text{t}} n_{\text{r}} \left(b + \tau_{\text{f}} \right) \tilde{d}_e, & b + \tau_{\text{f}} \leq \hat{b} \\ \min(1, \hat{d}_e) + n_{\text{r}} d^{\ddagger} \tilde{d}_e, & b + \tau_{\text{f}} > \hat{b}. \end{cases} \tag{36}$$

Proof: See Appendix D. ∎

Theorem 3 characterizes the interplay of the prediction parameter τ_{f}, the CSI qualities indicated by $\tilde{d}_e$ and $\hat{d}_e$, and the outage exponent. We can see by comparing (34) and (35) with (18) that for any prediction parameter $\tau_{\text{f}} \geq 0$ and $\tilde{d}_e > 0$, predictive-CSIT power adaptation is capable of improving the outage exponent. In particular, if the number of predicted fading blocks as indicated by τ_{f} satisfies

$$\tau_{\text{f}} \geq \begin{cases} B - 1, & \text{for Gaussian inputs} \\ \hat{b}, & \text{for discrete inputs,} \end{cases} \tag{37}$$

then the outage exponent obtained with predictive CSIT is the same as that obtained with full non-causal CSIT.

Predictive CSIT reflects the case when the number of CSIT blocks for power adaptation is "in between" the causal CSIT and full-non-causal CSIT. The outage exponent for predictive CSIT is higher than that for causal CSIT because more number of fading blocks are used for adapting the power. Whether causal CSIT can improve the outage exponent depends on how long the CSIT delay is, whereas predictive CSIT will always offer additional outage gains. Incorporating the prediction parameter τ_{f} reveals that in order to achieve the maximum outage gain, it may not be necessary to have full non-causal CSIT information prior transmission. In fact, for discrete inputs, we only need to have CSIT access up to $\hat{b}$ future blocks.

As mentioned in the system model, predictive CSIT is mostly relevant to OFDM-type of data transmission where τ_{f} can be flexibly specified depending on the complexity. For time-domain transmission, predicting the future channel realizations, especially when the fading process is independent, seems to be unrealistic.

From Theorems 1, 2 and 3, we note that irrespective of full non-causal, causal or predictive CSIT power allocation, the condition $\hat{d}_e \geq 1$ corresponds to a "nearly perfect" CSIR criterion as the outage exponent achieved under imperfect CSIR is identical to that achieved under perfect CSIR ($\hat{d}_e \uparrow \infty$). This extends a similar criterion in [21] where uniform power allocation is considered.

V. DISCUSSION

A. Pilot-Assisted Channel Estimation

The CSI models in (4) and (8) are an abstraction of pilot-based channel estimation for which two-way pilot transmissions are used for estimating relevant channel information at both terminals. Channel estimations exploit the slow nature of the fading process and channel reciprocity in TDD systems. The channel remains constant for block b and thus, the two-way pilot transmissions for estimating $\mathbb{H}_b$ at the transmitter and $(\mathsf{P}_b, \mathbb{H}_b)$ at the receiver occur within the block b. For orthogonal pilot design [20] [35], where orthogonal vectors are used to estimate the $(n_\mathrm{r} \times n_\mathrm{t})$ entries of the channel matrix at block b, $b = 1, \ldots, B$, these transmissions require $(n_\mathrm{t} + n_\mathrm{r})$ channel uses and are done prior transmitting the data for block b. Since the transmitter has access to noisy fading coefficients up to block b, for block-fading channels with time-domain transmission, only causal-CSIT power allocation with delay $\tau_\mathrm{d} > 0$ or predictive-CSIT power allocation with $\tau_\mathrm{f} = 0$ is realistic. On the other hand, for block fading channels with frequency-domain transmission (such as multi-carrier transmission and OFDM), the full non-causal CSIT assumption is of practical relevance.

The CSI to be estimated at the transmitter and receiver is not equivalent. At the transmitter side, CSIT consists of estimates of the fading matrix for power adaptation. At the receiver side, CSIR comprises estimates of the fading matrix and data power that are required by the nearest neighbor decoder (10).

Suppose that orthogonal pilots [20] [35] are employed and for each training symbol only one antenna is active at a time. This requires n_t time instants to transmit pilot symbols from the transmitter and n_r time instants to transmit pilot symbols from the receiver. Assume that at block b, the receiver initiates transmission of a scaled unity pilot symbol with power $P^{\tilde{d}_e}$, which will be used to estimate the fading matrix at the transmitter. When the receive antenna r is active in

sending the pilot symbol, the transmitter obtains an observation at the transmit antenna t as

$$\tilde{Y}_{b,r,t} = \sqrt{P^{\tilde{d}_e}} H_{b,r,t} + \tilde{Z}_{b,r,t} \tag{38}$$

where $\tilde{Z}_{b,r,t}$ is zero-mean unit-variance complex-Gaussian noise at the transmitter. Dividing (38) by $\sqrt{P^{\tilde{d}_e}}$ leads to the CSIT model in (4). Upon computing the data power based on the obtained CSIT, the data power information is embedded into pilot symbols that will be sent to the receiver. Depending on the pilot power constraint, the transmitter scales the data power by $\kappa n_{\rm t}$ where $\kappa \geq 0$ is a predefined constant. Recalling the power structure in (6), the resulting pilot power for each transmit antenna is effectively given by $\kappa P_b(\tilde{\mathsf{H}}^{(n(b))})$. With this strategy, the pilot symbols satisfy an average power constraint

$$\mathbb{E}\left[\frac{1}{B} \sum_{b=1}^{B} \kappa P_b\left(\tilde{\mathbb{H}}^{(n(b))} \right) \right] \leq \kappa P. \tag{39}$$

Further assuming that the pilot symbols are simply power scaling, we have that when the transmit antenna t emits a pilot symbol, the receiver observes at the receive antenna r

$$\hat{Y}_{b,r,t} = H_{b,r,t} \sqrt{\kappa P_b\left(\tilde{\mathsf{H}}^{(n(b))} \right)} + \hat{Z}_{b,r,t} \tag{40}$$

where $\hat{Z}_{b,r,t}$ is zero-mean unit-variance complex-Gaussian noise at the receiver and is independent from the noise at the transmitter $\tilde{Z}_{b,r,t}$. Dividing (40) by $\sqrt{\kappa}$ and setting $\kappa = P^{\hat{d}_e - 1}$, $\hat{d}_e > 0$ to give various CSIR qualities lead to the CSIR model in (8) that captures both fading and power estimates at the receiver. The parameter κ is a constant and it is thus reasonable to be known at both terminals prior to transmission. By substituting $\kappa = P^{\hat{d}_e - 1}$ in (39), we observe that $\hat{d}_e$ characterizes the exponent of the average pilot power constraint, i.e. $P^{\hat{d}_e}$. A special case $\hat{d}_e = 1$ corresponds to allocating the same amount of power to the data symbols and the pilot symbols at the transmitter, which in turn yields the same outage exponent as the perfect-CSIR case [13], [28], [36]. Even though equal pilot and data powers at the transmitter may justify perfect CSIR analysis of the outage exponent, the generalized outage probability can still improve if the pilot power is higher than the data power as shown in [21].

The aforementioned pilot-based estimation is precisely maximum-likelihood (ML) channel estimation [19]. In the limit of $J \to \infty$, the pilot fraction $\frac{n_{\rm t} + n_{\rm r}}{J}$ vanishes and hence, the pilot insertion does not affect the outage exponents in Theorems 1, 2 and 3.

We provide a characterization on different CSIR qualities based on the model that jointly estimates fading and data power at the receiver. This model assumes that the list of codewords are available at the receiver, but the data power $\mathsf{P}_b(\tilde{\mathbb{H}}^{(n(b))})$ and fading $\mathbb{H}_b$ are not known *a priori*. Compared to the models used in the vast majority of previous works (see, e.g., [4]–[8], [15], [19]–[21]) where the CSIR corresponds to fading realizations alone, this model is more realistic, especially when the data power is dynamically adapted to the channel conditions. As the receiver needs to obtain $\mathsf{P}_b(\tilde{\mathbb{H}}^{(n(b))})$ for decoding, we need to embed this information into the pilot symbols.

Incorporating data power information into the pilot symbols turns out to be a good design approach. This insight can be confirmed by comparing the results in this paper with our earlier results [37] [38]. In those works, the CSIR is the fading information and pilot symbols from the transmitter (for channel estimation at the receiver) have a fixed power $P^{\hat{d}_e}$, which in turn yields CSIR

$$\hat{H}_{b,r,t} = H_{b,r,t} + W_{b,r,t}. \tag{41}$$

Herein $W_{b,r,t}$ is zero-mean variance $P^{-\hat{d}_e}$ complex Gaussian, representing the CSIR noise. Even after assuming the availability of data power information at the receiver (provided by a genie), the outage exponent for full non-causal CSIT achieved with this fixed pilot power is given by [37] [38]

$$d_{\text{icsi}}^{\text{f}} = d_{\text{csir}}^{\text{u}} \times \min\left(\hat{d}_e,\ 1 + d_{\text{csir}}^{\text{u}}\tilde{d}_e\right), \tag{42}$$

which is generally smaller than the outage exponent in Theorem 1. The main weakness of the fixed pilot power for CSIR estimation is that the CSIR quality is not being adapted to the current power used for data transmission. When the data power is high, the fixed pilot power cannot adjust the CSIR quality accordingly. On the other hand, when the data power is low, the fixed pilot power implies that excessive resources may have been spent without translating to outage gains. The extreme case when the data power is zero[2], the pilot symbols are still transmitted with a positive power, yielding unnecessary spending of resources.

We can further gain insights from the preceding explanation. If we employ separate estimation

---

[2]This occurs when for a given power constraint, we cannot invert the channel effects to mitigate outages and the best strategy is just to accept outages.

of data power and fading at the receiver, how much the outage exponent can we expect? If the fading is also estimated using the pilot power $P^{\hat{d}_e}$, then the outage exponent in (42) serves as the maximum outage diversity that can be achieved. This is so because the outage exponent in (42) assumes imperfect fading estimation at the receiver and is achieved when the data power is perfectly known at the receiver. Any kind of imperfect estimation of data power will degrade the performance. This confirms that separate estimation of data power and fading is inferior to joint estimation of both parameters.

B. Comments on Achievable Rates

The technique used to derive the outage exponent is based on the GMI, which is the largest achievable rate for i.i.d. generated codebooks [24], [29], [30] and nearest neighbor decoding. Therefore, the result in Theorems 1, 2 and 3 is the optimal diversity for i.i.d. codebooks (Gaussian or discrete) and a nearest neighbor decoder. An improved achievable rate (LM rate) can be obtained with codewords satisfying a good cost constraint [24] [29]. The main difficulty of using LM rate is the optimization over all possible cost functions, which in general cannot be solved analytically.

Several works in the literature studied a similar outage exponent problem, but used different information rates as a departure point [13], [28], [36]. For simplicity and for the sake of comparison, we consider a single-input single-output (SISO) quasi-static channel ($B = 1$). References [13], [28], [36] assumed Gaussian inputs and linear minimum mean-squared error (LMMSE) channel estimation at the receiver, where the estimate $\hat{H}$ is related to the actual fading H as

$$H = \hat{H} + \hat{E} \tag{43}$$

and where $\hat{E}$ is the scalar fading estimation error having zero mean and variance $P^{-\hat{d}_e}$. Thus, from (3) and (43) we can write the input-output relationship as

$$\boldsymbol{Y} = \sqrt{P}\hat{H}\boldsymbol{x} + \sqrt{P}\left(H - \hat{H}\right)\boldsymbol{x} + \boldsymbol{Z} \tag{44}$$

where $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are the random received and noise vectors, respectively, which take values on $\mathbb{C}^J$; $\boldsymbol{x}$ is the J-dimensional input vector; P is the transmission power. Since every realization of $\hat{H}$ is known at the receiver, the argument in [13], [28], [36] is that one can treat the term

$\sqrt{P}(H - \hat{H})\boldsymbol{x}$ as an additional noise term. It was further argued in [28] [36] that by modeling the signal-dependent noise

$$\boldsymbol{Z}' = \sqrt{P}\left(H - \hat{H}\right)\boldsymbol{x} + \boldsymbol{Z} \tag{45}$$

as a zero-mean Gaussian noise with i.i.d. entries and each having variance $1 + P|H - \hat{H}|^2$, one can obtain a rate that is claimed to be a lower bound to the instantaneous mutual information as [28]

$$\underline{I}(H, \hat{H}, P) = \log_2\left(1 + \frac{P|\hat{H}|^2}{1 + P|H - \hat{H}|^2}\right). \tag{46}$$

The above expression leads to an outage exponent that is obtained by studying

$$\Pr\left\{\underline{I}(H, \hat{H}, P) < R\right\} = \Pr\left\{\log_2\left(1 + \frac{P|\hat{H}|^2}{1 + P|H - \hat{H}|^2}\right) < R\right\}. \tag{47}$$

Interestingly, following the steps used in [21, App. D] for $B = 1$, the GMI can be lower-bounded as

$$I^{\mathrm{gmi}}(H, \hat{H}, P) \geq \log_2\left(1 + \frac{P|\hat{H}|^2}{1 + P|H - \hat{H}|^2}\right) - \frac{1}{\log 2}. \tag{48}$$

In the high-SNR regime, the constant difference between (46) and the right-hand side (RHS) of (48) does not affect the outage exponents. Thus, it is not surprising that for the case under consideration, our results are identical to the results in [28] [36].[3]

Rate (46) seems to be easier to evaluate than the GMI. However, there are some technical problems associated with the derivation of (46), which we explain in the following.

- To the best of our knowledge, there is no explicit proof on the achievability of $\underline{I}(H, \hat{H}, P)$ for fixed fading and fading estimate realizations. The argument to derive (46) follows from [39], where LMMSE channel estimation is used at the receiver to derive a lower bound to the blockwise-ergodic capacity. In this blockwise-ergodic setup, the block length J is finite, and the capacity expression is obtained via coding over infinitely many blocks, where the estimate $\hat{H}$ and the error $(H - \hat{H})$ have uncorrelated statistics over these many blocks. A lower bound to the blockwise-ergodic capacity can then be obtained using the steps in [40,

---

[3]Although LMMSE channel estimation is slightly different to our CSIR model that resembles maximum-likelihood (ML) estimation, the variances of the fading estimation error for both estimations have a similar high-SNR behavior [19]. For the purpose of comparing our approach with the approach in [13], [28], [36], we use a constant data power P. It can be shown that using this constant power, the difference between the LMMSE and ML estimations does not affect the outage exponent.
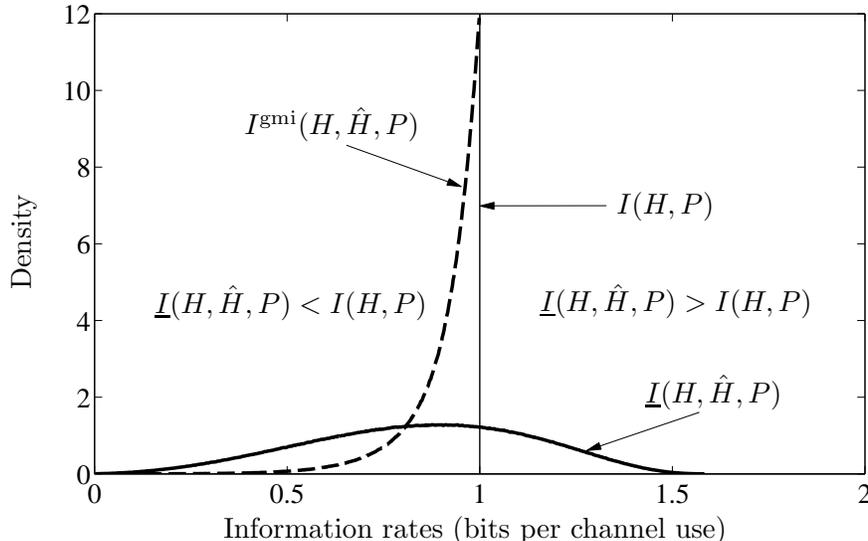
Fig. 3. Comparison of the densities of the GMI and and the lower bound (46) with fading realization $H = 1$, transmission power $P = 1$ (unit power) and CSIR-error variance $\hat{\varepsilon}^2 = 0.1$.

Sec. III] via averaging over all states of fading and its corresponding estimate.

It is not clear whether the technique in [39] can directly be applied to non-ergodic fading channels. As opposed to coding over infinitely many blocks, in a quasi-static channel, coding is performed for only one block and the block length J is taken to infinity to recover the information outage probability [4], [25]. During a single block, both fading H and fading estimate $\hat{H}$ are constant. Hence, rate (46) may not be an accurate lower bound to the instantaneous mutual information for the block of interest as both H and $\hat{H}$ are constant within a single block alone. It therefore follows that for fixed fading and fading estimate realizations, there is no guarantee that transmitting codeword at rate $R = \underline{I}(H, \hat{H}, P) - \varepsilon$ for any $\varepsilon > 0$ has a vanishing error probability as the block length J tends to infinity. This is in contrast with $I^{\mathrm{gmi}}(H, \hat{H}, P)$ for which the achievability has been proven in [22].

- For some H and $\hat{H}$, we may find $\underline{I}(H, \hat{H}, P)$ that is larger than the perfect-CSIR mutual information

$$I(H, P) = \log_2 \left(1 + P|H|^2 \right). \tag{49}$$

We illustrate this in Fig. 3 where we assume power $P = 1$ and fading realization $H = 1$,

and we use estimation (43) to compute the density of $I^{\mathrm{gmi}}(H, \hat{H}, P)$ and $\underline{I}(H, \hat{H}, P)$.[4] For a given $H = 1$, the probability that $\underline{I}(H, \hat{H}, P)$ is greater than $I(H, P)$ is non zero, which implies that rate (46) violates the data-processing inequality. This result indirectly disproves the achievability of $\underline{I}(H, \hat{H}, P)$, in contrast to $I^{\mathrm{gmi}}(H, \hat{H}, P)$, which is always smaller than $I(H, P)$ as shown in [22], [23], [41].

- It is not clear whether modeling $\mathbf{Z}'$ in (45) as a signal-independent Gaussian noise would still yield the correct exponent for discrete inputs.

C. Comments on Continuous Input Distributions

We have considered Gaussian inputs for our outage exponent analysis as motivated by the fact that Gaussian inputs are optimal (i.e., capacity-achieving) for the channel in (3) when perfect CSIR is available. However, when only noisy CSIR is known, Gaussian inputs may no longer be optimal.

We can show using (15) that the outage exponent for Gaussian inputs is a lower bound to the outage exponent for some input distributions satisfying certain conditions. We first assume that the input vector is i.i.d. over all transmit antennas and all channel uses and is such that $\mathbb{E}\left[|X|^2\right] = 1$. The expression in (15) (in natural-base logarithm) can be decomposed into two terms as follows

$$
\begin{aligned}
I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b) = {} & \mathbb{E}\left[\log Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}\right) \middle| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right] \\
& - \mathbb{E}\left[\log \mathbb{E}\left[Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}'\right) \middle| \boldsymbol{Y}, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right] \middle| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right]. \quad (50)
\end{aligned}
$$

---

[4]The setup of $H = 1$ is for illustration purpose. By continuity argument, the phenomenon of $\underline{I}(H, \hat{H}, P) > I(H, P)$ can also be observed for some ranges of $H \in \mathbb{C}$. Over all possible H, there is a non-zero probability that $\underline{I}(H, \hat{H}, P) > I(H, P)$.

Evaluating the first term of the RHS of (50) yields

$$\mathbb{E}\left[\log Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}\right)\Big| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right]$$

$$= -s\mathbb{E}\left[\left\|\left(\mathsf{H}_b\mathsf{P}_b^{\frac{1}{2}} - \hat{\mathsf{V}}_b\right)\boldsymbol{X} + \boldsymbol{Z}\right\|^2\Big| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right] \tag{51}$$

$$= -s\left(n_{\mathrm{r}} + \mathbb{E}\left[\left\|\hat{\mathsf{E}}_b\boldsymbol{X}\right\|^2\Big| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right]\right) \tag{52}$$

$$\geq -s\left(n_{\mathrm{r}} + \mathbb{E}\left[\left\|\hat{\mathsf{E}}_b\right\|_F^2\|\boldsymbol{X}\|^2\Big| \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right]\right) \tag{53}$$

$$= -s\left(n_{\mathrm{r}} + n_{\mathrm{t}}\left\|\hat{\mathsf{E}}_b\right\|_F^2\right) \tag{54}$$

where the inequality is due to the property $\|\mathsf{A}\boldsymbol{x}\|_F^2 \leq \|\mathsf{A}\|_F^2 \cdot \|\boldsymbol{x}\|^2$ [42, Sec. 5.6]. The first expectation in the second term of the RHS of (50) can be evaluated as follows

$$\mathbb{E}\left[Q^s\left(\boldsymbol{Y}, \hat{\mathsf{V}}_b, \boldsymbol{X}'\right)\Big| \boldsymbol{Y}, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b\right] = \int_{\boldsymbol{x}'} P_{\boldsymbol{X}}(\boldsymbol{x}')e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{V}}_b\boldsymbol{x}'\right\|^2}d\boldsymbol{x}'. \tag{55}$$

Then, if the input probability density function (PDF) can be bounded as

$$P_{\boldsymbol{X}}(\boldsymbol{x}) \leq \frac{C}{\pi^{n_{\mathrm{t}}}}e^{-\|\boldsymbol{x}\|^2}, \qquad \boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{t}}} \tag{56}$$

for some constant $C > 0$, independent of the SNR, then the above expectation can be bounded as

$$\int_{\boldsymbol{x}'} P_{\boldsymbol{X}}(\boldsymbol{x}')e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{V}}_b\boldsymbol{x}'\right\|^2}d\boldsymbol{x}'$$

$$\leq C\int_{\boldsymbol{x}'} \frac{1}{\pi^{n_{\mathrm{t}}}}e^{-\|\boldsymbol{x}'\|^2}e^{-s\left\|\boldsymbol{Y}-\hat{\mathsf{V}}_b\boldsymbol{x}'\right\|^2}d\boldsymbol{x}' \tag{57}$$

$$= \frac{C}{\det\left(\mathsf{I}_{n_{\mathrm{r}}} + s\hat{\mathsf{V}}_b\hat{\mathsf{V}}_b^\dagger\right)}\exp\left(-s\boldsymbol{Y}^\dagger\left(\mathsf{I}_{n_{\mathrm{r}}} + s\hat{\mathsf{V}}_b\hat{\mathsf{V}}_b^\dagger\right)^{-1}\boldsymbol{Y}\right). \tag{58}$$

It follows that with $s > 0$, $I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b)$ can be lower-bounded as

$$I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b)$$

$$\geq \log\det\left(\mathsf{I}_{n_{\mathrm{r}}} + s\hat{\mathsf{V}}_b\hat{\mathsf{V}}_b^\dagger\right) - \log C - s\left(n_{\mathrm{r}} + n_{\mathrm{t}}\left\|\hat{\mathsf{E}}_b\right\|_F^2\right)$$

$$+ \mathbb{E}\left[s\boldsymbol{Y}^\dagger\left(\mathsf{I}_{n_{\mathrm{r}}} + s\hat{\mathsf{V}}_b\hat{\mathsf{V}}_b^\dagger\right)^{-1}\boldsymbol{Y}\right]. \tag{59}$$

The RHS of (59) is similar to $I_b^{\mathrm{gmi}}(s, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b)$ for Gaussian inputs, except for the extra terms $-\log C$ and n_{t} in $n_{\mathrm{t}} \|\hat{\mathsf{E}}_b\|_F^2$ (n_{t} is replaced by 1 for Gaussian inputs). However, since those terms do not depend on the SNR, they do not affect the outage exponent. Then, noting that the outage exponent for Gaussian inputs derived using GMI upper and lower bounds is tight (as given in Theorems 1, 2 and 3), it follows that for any input distribution meeting the condition (56), the outage exponent is lower-bounded by the outage exponent for Gaussian inputs. It is not yet clear whether this lower bound is tight because solving the GMI upper bound for input distributions such that (56) holds remains a challenge.

VI. CONCLUSION

We have studied the effects of imperfect CSI on the performance of data transmission over MIMO block-fading channels. In particular, we derived the outage exponent as a function of CSIR and CSIT noise variances, $\hat{\varepsilon}^2 = P^{-(\hat{d}_e - 1)}$ and $\tilde{\varepsilon}^2 = P^{-\tilde{d}_e}$, where P is the average data SNR. We have demonstrated that even in the non-ideal situations where CSI at the transmitter and receiver is noisy, power control across fading blocks is still capable of offering outage performance improvements over uniform power allocation.

The results shed new light into the design of pilot-assisted channel estimation in block-fading channels. Specifically, we have highlighted the importance of embedding the data power information into pilot symbols as a good design approach. This is essential to ensure that power adaptation is not only able to invert the channel effects up to a certain extent, but also able to control the accuracy of CSIR via the transmitted pilot power. These two features in turn contribute in reducing the number of communication outages. For full non-causal CSIT, this reduction is indicated by the outage exponent that is a function of the CSI qualities measured by $\hat{d}_e$ and $\tilde{d}_e$, and the perfect-CSIR outage exponent with uniform power allocation $d_{\mathrm{csir}}^{\mathrm{u}}$. While having CSIR provides outage gains that are linear in $d_{\mathrm{csir}}^{\mathrm{u}}$, investing resources to obtain CSIT yields outage gains that scale quadratically with $d_{\mathrm{csir}}^{\mathrm{u}}$. For causal and predictive CSIT, the outage exponent does not only depend on the CSIT-error diversity $\tilde{d}_e$ and the CSIR-error diversity $\hat{d}_e$, but also the CSIT delay τ_{d} or the CSIT prediction parameter τ_{f}. The delay τ_{d} crucially determines whether causal-CSIT power allocation can provide outage exponent improvements over uniform power allocation. Incorporating prediction parameter τ_{f} reveals that the outage exponent for full non-causal CSIT can be achieved by predictive-CSIT power control provided that τ_{f} is sufficiently

large.

The outage exponents derived in this paper are the optimal code diversity when using i.i.d. codebooks (Gaussian or discrete) and a nearest neighbor decoder. In order to obtain a potentially better outage exponent, one should consider non-i.i.d. codebooks or different decoding strategies.

APPENDIX A
HIGH-SNR ASYMPTOTIC ANALYSIS

A. Change of Random Variables for High-SNR Evaluation

To facilitate the analysis, we first introduce the normalized CSIT matrix

$$\bar{\mathbb{H}}_b \triangleq \frac{1}{\tilde{\varepsilon}} \tilde{\mathbb{H}}_b. \tag{60}$$

We can observe from the CSIT (4) that given $\mathbb{H}_b = \mathsf{H}_b$, $\bar{\mathbb{H}}_b$ is a complex Gaussian random matrix with mean $\frac{1}{\tilde{\varepsilon}}\mathsf{H}_b$ and entries having a unit variance. We further observe from CSI equations (4) and (8), and (60) that the generalized outage probability (16) can directly be expressed in terms of $\mathbb{H}_b$, $\hat{\mathbb{E}}_b$ and $\bar{\mathbb{H}}_b$.

Magnitude-Squared Notation		Phase Notation	
Matrix	Entry (r, t)	Matrix	Entry (r, t)
Γ_b	$\gamma_{b,r,t} \triangleq |h_{b,r,t}|^2$	Φ_b^{H}	$\phi_{b,r,t}^h \triangleq \angle h_{b,r,t}$
$\hat{\Xi}_b$	$\hat{\xi}_{b,r,t} \triangleq |\hat{e}_{b,r,t}|^2$	$\Phi_b^{\hat{\mathsf{E}}}$	$\phi_{b,r,t}^{\hat{e}} \triangleq \angle \hat{e}_{b,r,t}$
$\tilde{\Gamma}_b$	$\tilde{\gamma}_{b,r,t} \triangleq |\tilde{h}_{b,r,t}|^2$	$\Phi_b^{\tilde{\mathsf{H}}}$	$\phi_{b,r,t}^{\tilde{h}} \triangleq \angle \tilde{h}_{b,r,t}$
$\bar{\Gamma}_b$	$\bar{\gamma}_{b,r,t} \triangleq |\bar{h}_{b,r,t}|^2$	$\Phi_b^{\bar{\mathsf{H}}}$	$\phi_{b,r,t}^{\bar{h}} \triangleq \angle \bar{h}_{b,r,t}$

TABLE I
DEFINITION OF MAGNITUDE-SQUARED AND PHASE VARIABLES.

Let $H_{b,r,t}$, $\hat{E}_{b,r,t}$ and $\bar{H}_{b,r,t}$ be the entries of $\mathbb{H}_b$, $\hat{\mathbb{E}}_b$ and $\bar{\mathbb{H}}_b$ at row r and column t. It follows from (4) and (60) that conditioned on $H_{b,r,t} = h_{b,r,t}$, $\bar{H}_{b,r,t}$ is complex-Gaussian distributed with mean $\frac{1}{\tilde{\varepsilon}}h_{b,r,t}$ and variance one. For convenience, we define magnitude-squared variables and phase variables in Table I. From the definition in Table I, we note that $\gamma_{b,r,t}$ and $\hat{\xi}_{b,r,t}$ have the

Magnitude-Squared Matrix	Entry (r,t)	Normalized Matrix	Entry (r,t)
Γ_b	$\gamma_{b,r,t} \triangleq \|h_{b,r,t}\|^2$	A_b	$\alpha_{b,r,t} \triangleq -\frac{\log \gamma_{b,r,t}}{\log P}$
$\hat{\bar{\Xi}}_b$	$\hat{\xi}_{b,r,t} \triangleq \|\hat{e}_{b,r,t}\|^2$	$\hat{\Theta}_b$	$\hat{\theta}_{b,r,t} \triangleq -\frac{\log \hat{\xi}_{b,r,t}}{\log P}$
$\tilde{\Gamma}_b$	$\tilde{\gamma}_{b,r,t} \triangleq \|\tilde{h}_{b,r,t}\|^2$	$\tilde{\mathsf{A}}_b$	$\tilde{\alpha}_{b,r,t} \triangleq -\frac{\log \tilde{\gamma}_{b,r,t}}{\log P}$
$\bar{\Gamma}_b$	$\bar{\gamma}_{b,r,t} \triangleq \|\bar{h}_{b,r,t}\|^2$	$\bar{\mathsf{A}}_b$	$\bar{\alpha}_{b,r,t} \triangleq -\frac{\log \bar{\gamma}_{b,r,t}}{\log P}$

TABLE II

DEFINITION OF NORMALIZED MAGNITUDE-SQUARED VARIABLES.

exponential probability density functions (PDFs)

$$f(\gamma_{b,r,t}) = e^{-\gamma_{b,r,t}} \tag{61}$$

$$f\left(\hat{\xi}_{b,r,t}\right) = P^{\hat{d}_e-1} e^{-P^{\hat{d}_e-1}\hat{\xi}_{b,r,t}}. \tag{62}$$

Conditioned on $H_{b,r,t} = h_{b,r,t}$, $\bar{\gamma}_{b,r,t}$ has the non-central Chi-square PDF

$$f(\bar{\gamma}_{b,r,t}|\nu) = e^{-\bar{\gamma}_{b,r,t}-\nu} I_0\left(2\sqrt{\bar{\gamma}_{b,r,t}\nu}\right) \tag{63}$$

where $\nu = \frac{1}{\bar{\varepsilon}^2}|h_{b,r,t}|^2 = \frac{1}{\bar{\varepsilon}^2}\gamma_{b,r,t}$ is the non-centrality parameter and $I_0(\cdot)$ is the zeroth order modified Bessel function of the first kind.

For high-SNR analysis, we define transformed variables in Table II. It follows from (61)–(63) that we have the following PDFs

$$f(\alpha_{b,r,t}) = \log(P) P^{-\alpha_{b,r,t}} e^{-P^{-\alpha_{b,r,t}}}, \tag{64}$$

$$f(\hat{\theta}_{b,r,t}) = \log(P) P^{\hat{d}_e-1-\hat{\theta}_{b,r,t}} e^{-P^{\hat{d}_e-1-\hat{\theta}_{b,r,t}}}, \tag{65}$$

$$f(\bar{\alpha}_{b,r,t}|\alpha_{b,r,t}) = \log(P) P^{-\bar{\alpha}_{b,r,t}} e^{-P^{-\bar{\alpha}_{b,r,t}}-P^{\tilde{d}_e-\alpha_{b,r,t}}} I_0\left(2P^{\frac{\tilde{d}_e-\alpha_{b,r,t}-\bar{\alpha}_{b,r,t}}{2}}\right). \tag{66}$$

B. Asymptotic Power Allocation

We consider power allocation with a scaled identity precoding matrix (6)

$$\mathsf{P}_b\left(\tilde{\mathsf{H}}^{(n(b))}\right) = \frac{P_b\left(\tilde{\mathsf{H}}^{(n(b))}\right)}{n_{\mathsf{t}}} \mathsf{I}_{n_{\mathsf{t}}}, \qquad b = 1, \ldots, B. \tag{67}$$

One can show that power allocation with constraint $\mathbb{E}[P_b(\tilde{\mathbb{H}}^{(n(b))})] \leq BP$ for all $b = 1, \ldots, B$ results in an upper bound to the outage exponent; note that this violates the constraint (7). On the other hand, one can consider suboptimal power allocation with $\mathbb{E}[P_b(\tilde{\mathbb{H}}^{(n(b))})] \leq P$ to obtain a lower bound to the outage exponent. Let

$$\tilde{\Gamma}^{(n(b))} \triangleq \left[\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{n(b)}\right], \tag{68}$$

$$\tilde{\mathsf{A}}^{(n(b))} \triangleq \left[\tilde{\mathsf{A}}_1, \ldots, \tilde{\mathsf{A}}_{n(b)}\right], \tag{69}$$

$$\Phi^{\tilde{\mathsf{H}}(n(b))} \triangleq \left[\Phi_1^{\tilde{\mathsf{H}}}, \ldots, \Phi_{n(b)}^{\tilde{\mathsf{H}}}\right]. \tag{70}$$

Then, optimal power allocation satisfies

$$\mathbb{E}\left[P_b\left(\tilde{\mathbb{H}}^{(n(b))}\right)\right]$$
$$= \int_{\substack{\tilde{\Gamma}^{(n(b))} \in \mathbb{R}_+^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}, \\ \Phi^{\tilde{\mathsf{H}}(n(b))} \in [0,2\pi)^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}}} P_b\left(\tilde{\mathsf{H}}^{(n(b))}\right) f\left(\tilde{\Gamma}^{(n(b))}\right) f\left(\Phi^{\tilde{\mathsf{H}}(n(b))}\right) d\tilde{\Gamma}^{(n(b))} d\Phi^{\tilde{\mathsf{H}}(n(b))} \dot{\leq} P. \tag{71}$$

Let $P_b(\tilde{\mathsf{H}}^{(n(b))}) \doteq P^{\varpi_b}$. Using transformation in Table II, the constraint on the RHS of (71) can be expressed as

$$\int_{\substack{\tilde{\mathsf{A}}^{(n(b))} \in \mathbb{R}_+^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}, \\ \Phi^{\tilde{\mathsf{H}}(n(b))} \in [0,2\pi)^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}}} P^{\varpi_b} P^{-\sum_{b'=1}^{n(b)} \sum_{r=1}^{n_{\mathrm{r}}} \sum_{t=1}^{n_{\mathrm{t}}} \tilde{\alpha}_{b',r,t}} d\tilde{\mathsf{A}}^{(n(b))} d\Phi^{\tilde{\mathsf{H}}(n(b))} \dot{\leq} P. \tag{72}$$

Herein we have neglected the terms irrelevant to the outage exponent such as the phase as $f(\Phi^{\tilde{\mathsf{H}}(n(b))})$ is uniform over $[0, 2\pi)^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}$ and the interval of $\tilde{\alpha}_{b',r,t} < 0$ as its probability decays exponentially with P. Applying Varadhan's lemma [43] to (72) yields

$$\sup_{\substack{\tilde{\mathsf{A}}^{(n(b))} \in \mathbb{R}_+^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}, \\ \Phi^{\tilde{\mathsf{H}}(n(b))} \in [0,2\pi)^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}}} \left\{\varpi_b\left(\tilde{\mathsf{A}}^{(n(b))}, \Phi^{\tilde{\mathsf{H}}(n(b))}\right) - \sum_{b'=1}^{n(b)} \sum_{r=1}^{n_{\mathrm{r}}} \sum_{t=1}^{n_{\mathrm{t}}} \tilde{\alpha}_{b',r,t}\right\} \leq 1. \tag{73}$$

Noting that $\bar{\alpha}_{b,r,t} = \tilde{\alpha}_{b,r,t} - \tilde{d}_e$ as a consequence of (60), the supremum constraint can be expressed in terms of $\bar{\mathsf{A}}^{(n(b))}$ and its entries $\bar{\alpha}_{b,r,t}$ as

$$\sup_{\substack{\bar{\mathsf{A}}^{(n(b))} \in \mathbb{R}_+^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}, \\ \Phi^{\bar{\mathsf{H}}(n(b))} \in [0,2\pi)^{n(b) \cdot n_{\mathrm{r}} \cdot n_{\mathrm{t}}}}} \left\{\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}, \Phi^{\bar{\mathsf{H}}(n(b))}\right) - n(b) n_{\mathrm{r}} n_{\mathrm{t}} \tilde{d}_e - \sum_{b'=1}^{n(b)} \sum_{r=1}^{n_{\mathrm{r}}} \sum_{t=1}^{n_{\mathrm{t}}} \bar{\alpha}_{b',r,t}\right\} \leq 1. \tag{74}$$

The optimal power exponent $\varpi_b(\cdot, \cdot)$ minimizes $P_{\text{gout}}(R)$. Throughout the remaining parts of appendices, we will consider $\varpi_b(\cdot, \cdot)$ that depends on the magnitude but not the phase, i.e., $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$. We shall observe later in Appendices B, C and D that this allocation does not incur loss in the terms of outage exponent.

C. Asymptotic Outage Analysis

Let $\mathcal{O}_{\mathcal{X}}$ be the high-SNR outage set from an input alphabet $\mathcal{X}$. We have that

$$P_{\text{gout}}(R) = \Pr\left\{ I^{\text{gmi}}(\mathbb{H}, \hat{\mathbb{V}}, \mathsf{P}) < R \right\} \tag{75}$$

$$= \int_{\mathcal{O}_{\mathcal{X}}} \prod_{b=1}^{B} f\left(\bar{\mathsf{\Gamma}}_b, \mathsf{H}_b, \hat{\mathsf{E}}_b\right) d\bar{\mathsf{\Gamma}}_b d\mathsf{H}_b d\hat{\mathsf{E}}_b \tag{76}$$

$$= \int_{\mathcal{O}_{\mathcal{X}}} \prod_{b=1}^{B} f\left(\bar{\mathsf{\Gamma}}_b \Big| \mathsf{\Gamma}_b\right) f\left(\mathsf{\Gamma}_b\right) f\left(\Phi_b^{\mathsf{H}}\right) f\left(\hat{\bar{\Xi}}_b\right) f\left(\Phi_b^{\hat{\mathsf{E}}}\right) d\bar{\mathsf{\Gamma}}_b d\mathsf{\Gamma}_b d\hat{\bar{\Xi}}_b d\Phi_b^{\mathsf{H}} d\Phi_b^{\hat{\mathsf{E}}}. \tag{77}$$

By transforming $\bar{\mathsf{\Gamma}}_b$, $\mathsf{\Gamma}_b$ and $\hat{\bar{\Xi}}_b$ into $\bar{\mathsf{A}}_b$, A_b and $\hat{\Theta}_b$ as shown in Table II, we have that

$$P_{\text{gout}}(R)$$
$$\doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{b=1}^{B} \prod_{r=1}^{n_{\text{r}}} \prod_{t=1}^{n_{\text{t}}} f(\bar{\alpha}_{b,r,t}|\alpha_{b,r,t}) f(\alpha_{b,r,t}) f(\hat{\theta}_{b,r,t}) d\bar{\alpha}_{b,r,t} d\alpha_{b,r,t} d\hat{\theta}_{b,r,t} d\phi_{b,r,t}^{h} d\phi_{b,r,t}^{\hat{e}} \tag{78}$$

where the PDFs have been expressed in terms of the entries of the matrices. Herein the PDFs of the phases do not appear because $\phi_{b,r,t}^{h}$ and $\phi_{b,r,t}^{\hat{e}}$ are uniformly distributed over $[0, 2\pi)$ and hence do not affect the dot equality.

For perfect CSIR ($\hat{d}_e \uparrow \infty$), by following the steps in [6], we have that

$$P_{\text{gout}}(R) \doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{b=1}^{B} \prod_{r=1}^{n_{\text{r}}} \prod_{t=1}^{n_{\text{t}}} f(\bar{\alpha}_{b,r,t}|\alpha_{b,r,t}) f(\alpha_{b,r,t}) d\bar{\alpha}_{b,r,t} d\alpha_{b,r,t} \tag{79}$$

$$\doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{(b,r,t): -\tilde{d}_e \leq \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0} \left(P^{-\alpha_{b,r,t}} d\alpha_{b,r,t} \right)$$

$$\cdot \prod_{(b,r,t): \bar{\alpha}_{b,r,t} \geq 0, \alpha_{b,r,t} \geq \tilde{d}_e} \left(P^{-(\alpha_{b,r,t} + \bar{\alpha}_{b,r,t})} d\bar{\alpha}_{b,r,t} d\alpha_{b,r,t} \right). \tag{80}$$

By comparing the perfect-CSIR outage (79) with the imperfect-CSIR outage (78), we observe that the additional density term in the imperfect-CSIR outage (78) is due to $f(\hat{\theta}_{b,r,t})$. Thus,

evaluating (78) by using the joint PDF $\prod_{b,r,t} f(\bar{\alpha}_{b,r,t}|\alpha_{b,r,t})f(\alpha_{b,r,t})$ on the RHS of (80) and the PDF $f(\hat{\theta}_{b,r,t})$ on the RHS of (65) yields

$$P_{\text{gout}}(R) \doteq P^{d_{\text{icsi}}} \tag{81}$$

$$\doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{b=1}^{B} \prod_{r=1}^{n_{\mathrm{r}}} \prod_{t=1}^{n_{\mathrm{t}}} f(\bar{\alpha}_{b,r,t}|\alpha_{b,r,t})f(\alpha_{b,r,t})f(\hat{\theta}_{b,r,t})d\bar{\alpha}_{b,r,t}d\alpha_{b,r,t}d\hat{\theta}_{b,r,t}d\phi^{h}_{b,r,t}d\phi^{\hat{e}}_{b,r,t} \tag{82}$$

$$\doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t}=\alpha_{b,r,t}-\tilde{d}_e<0} \left(\log P \cdot e^{-P^{-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)}} \cdot P^{-\alpha_{b,r,t}-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)} \right.$$

$$\left. \cdot\; d\alpha_{b,r,t}d\hat{\theta}_{b,r,t}d\phi^{h}_{b,r,t}d\phi^{\hat{e}}_{b,r,t} \right)$$

$$\times \prod_{(b,r,t):\bar{\alpha}_{b,r,t}\geq 0,\alpha_{b,r,t}\geq \tilde{d}_e} \left(\log P \cdot e^{-P^{-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)}} \cdot P^{-\alpha_{b,r,t}-\bar{\alpha}_{b,r,t}-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)} \right.$$

$$\left. \cdot\; d\bar{\alpha}_{b,r,t}d\alpha_{b,r,t}d\hat{\theta}_{b,r,t}d\phi^{h}_{b,r,t}d\phi^{\hat{e}}_{b,r,t} \right) \tag{83}$$

$$\doteq \int_{\mathcal{O}_{\mathcal{X}}} \prod_{\substack{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t}=\alpha_{b,r,t}-\tilde{d}_e<0, \\ \hat{\theta}_{b,r,t}\geq \hat{d}_e-1}} \left(P^{-\alpha_{b,r,t}-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)}d\alpha_{b,r,t}d\hat{\theta}_{b,r,t}d\phi^{h}_{b,r,t}d\phi^{\hat{e}}_{b,r,t} \right)$$

$$\times \prod_{\substack{(b,r,t):\, \bar{\alpha}_{b,r,t}\geq 0,\alpha_{b,r,t}\geq \tilde{d}_e, \\ \hat{\theta}_{b,r,t}\geq \hat{d}_e-1}} \left(P^{-\alpha_{b,r,t}-\bar{\alpha}_{b,r,t}-(\hat{\theta}_{b,r,t}+1-\hat{d}_e)}d\bar{\alpha}_{b,r,t}d\alpha_{b,r,t}d\hat{\theta}_{b,r,t}d\phi^{h}_{b,r,t}d\phi^{\hat{e}}_{b,r,t} \right) \tag{84}$$

where the last dot equality follows from the proof of [21, Lemma 3]. Applying Varadhan's lemma [43] to the RHS of (84) yields

$$d_{\text{icsi}} = \inf_{\mathsf{A},\bar{\mathsf{A}},\hat{\Theta}\in\mathcal{O}_{\mathcal{X}}} \left\{ \sum_{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t}=\alpha_{b,r,t}-\tilde{d}_e<0,\, \hat{\theta}_{b,r,t}\geq \hat{d}_e-1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right.$$

$$\left. \sum_{(b,r,t):\, \bar{\alpha}_{b,r,t}\geq 0,\, \alpha_{b,r,t}\geq \tilde{d}_e,\, \hat{\theta}_{b,r,t}\geq \hat{d}_e-1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right\}. \tag{85}$$

D. GMI Bounds, Outage Sets and Outage Exponents

For a given input distribution, evaluating the exact GMI (14) and thus the exact $\mathcal{O}_{\mathcal{X}}$ analytically may be infeasible due to the optimization over $s > 0$ across B fading blocks. Therefore, non-trivial upper and lower bounds are relevant in the analysis.

A GMI upper bound can be obtained by exchanging the supremum and the average on the RHS of (14), i.e.,

$$I^{\mathrm{gmi}}(\mathsf{H},\hat{\mathsf{V}},\mathsf{P}) \le \frac{1}{B}\sum_{b=1}^{B}\sup_{s>0} I_b^{\mathrm{gmi}}(s,\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b). \tag{86}$$

By letting

$$I_b^{\overline{\mathrm{gmi}}}(\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b) \triangleq \sup_{s>0} I_b^{\mathrm{gmi}}(s,\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b), \tag{87}$$

we may represent the GMI upper bound on the RHS of (86) as

$$I^{\overline{\mathrm{gmi}}}(\mathsf{H},\hat{\mathsf{V}},\mathsf{P}) \triangleq \frac{1}{B}\sum_{b=1}^{B} I_b^{\overline{\mathrm{gmi}}}(\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b). \tag{88}$$

This upper bound leads to the exact GMI in a number of cases, e.g., when the optimizing s on the RHS of (86) does not depend on P_b, H_b and $\hat{\mathsf{V}}_b$ or when $B=1$ (quasi-static channel). Also, for SISO quasi-static channels with Gaussian inputs, the optimal value of s can be found analytically. Using $I^{\overline{\mathrm{gmi}}}(\mathsf{H},\hat{\mathsf{V}},\mathsf{P})$, we can then define an equivalent outage set with GMI upper bound as

$$\overline{\mathcal{O}}_{\mathcal{X}} \triangleq \left\{ \mathsf{H},\hat{\mathsf{V}},\mathsf{P} : I^{\overline{\mathrm{gmi}}}(\mathsf{H},\hat{\mathsf{V}},\mathsf{P}) < R \right\} \tag{89}$$

to find an upper bound to the outage exponent in (85).

An important feature of the GMI upper bound (86) is that the rate per block $I_b^{\overline{\mathrm{gmi}}}(\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b)$ is a non-decreasing function of the transmit power coefficient $\mathsf{P}_b(\tilde{\mathsf{H}}^{(n(b))})$ at high SNR [21, Apps. C and D]. It thus follows that using the maximum power exponent satisfying (74), i.e.,

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + n(b)n_{\mathrm{r}}n_{\mathrm{t}}\tilde{d}_e + \sum_{b'=1}^{n(b)}\sum_{r=1}^{n_{\mathrm{r}}}\sum_{t=1}^{n_{\mathrm{t}}} \bar{\alpha}_{b',r,t} \tag{90}$$

for $I_b^{\overline{\mathrm{gmi}}}(\mathsf{H}_b,\hat{\mathsf{V}}_b,\mathsf{P}_b)$ leads to a valid upper bound[5] to the outage exponent for a given supremum constraint (74).

A GMI lower bound can be obtained by choosing a particular s. As shown in [21, App. D]

---

[5]In general, the resulting upper bound may be loose with respect to the exact outage exponent. However, in this paper, we will show later that this upper bound is sufficient as it coincides with a lower bound.

a good choice is

$$\hat{s} = \frac{B}{Bn_\mathrm{r} + \sum_{b=1}^{B} \left\| \hat{\mathsf{E}}_b \right\|_F^2} = \frac{B}{Bn_\mathrm{r} + \sum_{b=1}^{B} \sum_{r=1}^{n_\mathrm{r}} \sum_{t=1}^{n_\mathrm{t}} |\hat{e}_{b,r,t}|^2}. \tag{91}$$

We thus have a lower bound

$$I^\mathrm{gmi}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}) \geq \frac{1}{B} \sum_{b=1}^{B} I_b^\mathrm{gmi}(\hat{s}, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b) \triangleq I^{\underline{\mathrm{gmi}}}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}). \tag{92}$$

Then, using $I^{\underline{\mathrm{gmi}}}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P})$, we can define an equivalent outage set with GMI lower bound as

$$\underline{\mathcal{O}}_\mathcal{X} \triangleq \left\{ \mathsf{H}, \hat{\mathsf{V}}, \mathsf{P} : I^{\underline{\mathrm{gmi}}}(\mathsf{H}, \hat{\mathsf{V}}, \mathsf{P}) < R \right\} \tag{93}$$

to find a lower bound to the outage exponent in (85). Note that the rate per block $I_b^\mathrm{gmi}(\hat{s}, \mathsf{H}_b, \hat{\mathsf{V}}_b, \mathsf{P}_b)$ depends not only on the CSIR error for the current block $\hat{\mathsf{E}}_b$, but also on the CSIR error for the other blocks $\hat{\mathsf{E}}_{b'}$ for all $b' \neq b$ due to the factor of $\hat{s}$.

APPENDIX B

FULL-CSIT POWER ALLOCATION

Initial steps to derive outage exponents have been outlined in Appendix A. The key parameters are the power exponent $\varpi_b(\cdot)$ satisfying the constraint (74), the outage-exponent expression in (85) and the technique to find the high-SNR outage bounds, cf. Appendix A-D.

For full non-causal CSIT, we have $n(b) = B$, $b = 1, \ldots, B$ and the maximum power exponent (90) becomes

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + Bn_\mathrm{r}n_\mathrm{t}\tilde{d}_e + \sum_{b'=1}^{B} \sum_{r=1}^{n_\mathrm{r}} \sum_{t=1}^{n_\mathrm{t}} \bar{\alpha}_{b',r,t}. \tag{94}$$

Using this maximum power exponent, we shall prove in the following that upper and lower bounds to the outage exponent are tight. For identification, superscript [f] will be used to indicate full non-causal-CSIT results.

Before presenting the details of the proof, we first note from [21, Apps. C and D] that it suffices to consider solving the outage exponent for discrete constellations of alphabet size $|\mathcal{X}| = 2^M$. The proof for Gaussian constellations with constant R independent of the SNR (such that the multiplexing gain tends to zero) follows along the same lines as that for discrete constellations

with a sufficiently large alphabet size such that $M \geq BR$. Thus, for the remaining part of this appendix, we shall focus on the outage set $\mathcal{O}_\mathcal{X}$ and its bounds for discrete constellations.

A. GMI Upper Bound

Replacing the outage set $\mathcal{O}_\mathcal{X}$ in the infimum of (85) with $\overline{\mathcal{O}}_\mathcal{X}$ in (89) yields an upper bound to the outage exponent

$$\bar{d}_{\text{icsi}}^{\text{f}} = \inf_{\mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \overline{\mathcal{O}}_\mathcal{X}} \left\{ \sum_{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0,\, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right.$$

$$\left. \sum_{(b,r,t):\, \bar{\alpha}_{b,r,t} \geq 0,\, \alpha_{b,r,t} \geq \hat{d}_e,\, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right\}. \tag{95}$$

Using $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ in (94) and following the derivation in [21, App. C], we can express $\overline{\mathcal{O}}_\mathcal{X}$ for discrete constellations of size $|\mathcal{X}| = 2^M$ as

$$\overline{\mathcal{O}}_\mathcal{X} = \left\{ \mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \mathbb{R}^{B n_r \times n_t} : \sum_{b=1}^{B} \bar{\kappa}_b < \frac{BR}{M} \right\} \tag{96}$$

where we have defined, for some $\epsilon, \delta > 0$, the following variables

$$\bar{\kappa}_b \triangleq \left| \overline{\mathcal{S}}_b^{(\epsilon,\delta)} \right|, \tag{97}$$

$$\overline{\mathcal{S}}_b^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_r} \overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}, \tag{98}$$

$$\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \left\{ t : \left\{ \{ \alpha_{b,r,t} \leq \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \epsilon \} \cap \{ \alpha_{b,r,t} \leq \hat{\theta}_{b,r,t} + \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \delta \} \right\} \right.$$

$$\left. \cup \left\{ \{ \alpha_{b,r,t} \leq \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \epsilon \} \cap \{ \alpha_{b,r,t} > \hat{\theta}_{b,r,t} + \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \delta \} \cap \mathcal{Q}_{b,r,t} \right\}, \right.$$

$$\left. t = 1, \ldots, n_t \right\}, \tag{99}$$

$$\mathcal{Q}_{b,r,t} \triangleq \left\{ \phi_{b,r,t}^h, \phi_{b,r,t}^{\hat{e}} \in [0, 2\pi) : \cos\left(\phi_{b,r,t}^{\hat{e}} - \phi_{b,r,t}^h \right) > 0 \right\}. \tag{100}$$

Increasing $\hat{\theta}_{b,r,t}$ increases both the objective function on the RHS of (95) and the threshold for $\alpha_{b,r,t}$ in (99). Hence, the infimum solutions for $\hat{\theta}_{b,r,t}$, $b = 1, \ldots, B$, $r = 1, \ldots, n_r$, $t = 1, \ldots, n_t$ on the RHS of (95) are given by $\hat{d}_e - 1$.

We next solve for $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ that attain the infimum on the RHS of (95). By noting that $\hat{\theta}_{b,r,t} = \hat{d}_e - 1$, assume without loss of generality that for each $r = 1, \ldots, n_{\mathrm{r}}$,

$$\alpha_{b,r,t} > \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(\epsilon, \hat{d}_e - 1 + \delta\right), \qquad bB + t > \frac{BR}{M} \tag{101}$$

if $\mathcal{Q}_{b,r,t}$ does not occur, and

$$\alpha_{b,r,t} > \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \epsilon, \qquad bB + t > \frac{BR}{M} \tag{102}$$

if $\mathcal{Q}_{b,r,t}$ occurs. Since the argument on the RHS of (95) is increasing with $\alpha_{b,r,t}$ and since with

$$\frac{\pi}{2} \le \phi_{b,r,t}^{\hat{e}} - \phi_{b,r,t}^{h} \le \frac{3\pi}{2}, \tag{103}$$

the event $\mathcal{Q}_{b,r,t}$ does not occur, it follows that the infimum is achieved with

$$\alpha_{b,r,t} = \begin{cases} \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(\epsilon, \hat{d}_e - 1 + \delta\right), & \text{if } bB + t > \frac{BR}{M} \\ 0, & \text{otherwise.} \end{cases} \tag{104}$$

Depending on the values of $\hat{d}_e$ and $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$, we have the following cases.

- **Case 1:** $\epsilon \le \hat{d}_e - 1 + \delta$

  We first note that for $bB + t \le \frac{BR}{M}$, the values of $\alpha_{b,r,t}$ attaining the infimum are given by zero and thus belong to $\{(b, r, t) : -\tilde{d}_e \le \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0\}$. This implies that the values for $\bar{\alpha}_{b,r,t}$, $bB + t \le \frac{BR}{M}$ attaining the infimum are given by $-\tilde{d}_e$.

  For (b'', r'', t'') and $b''B + t'' > \frac{BR}{M}$, if (b'', r'', t'') belongs to $\{(b, r, t) : -\tilde{d}_e \le \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0\}$, we have that

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \epsilon = 1 + \epsilon + (Bn_{\mathrm{r}}n_{\mathrm{t}} - 1)\tilde{d}_e + \alpha_{b'',r'',t''} + \sum_{(b',r,t)\ne(b'',r'',t'')} \bar{\alpha}_{b',r,t} \ge \alpha_{b'',r'',t''}. \tag{105}$$

  This in conjunction with equations (97)–(99) implies that the constraint in (96) can never be met. As such (b'', r'', t'') for $b''B + t'' > \frac{BR}{M}$ must belong to $\{(b, r, t) : \bar{\alpha}_{b,r,t} \ge 0, \alpha_{b,r,t} \ge \tilde{d}_e\}$. In this case, since $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ increases with $\bar{\alpha}_{b,r,t}$, the values of $\bar{\alpha}_{b,r,t}$ that solve the infimum in (95) are given by

$$\bar{\alpha}_{b,r,t} = \begin{cases} 0, & \text{for } bB + t > \frac{BR}{M} \\ -\tilde{d}_e, & \text{otherwise} \end{cases} \tag{106}$$

and results in

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + n_{\mathrm{r}}\left(1 + \left\lfloor B\left(n_{\mathrm{t}} - \frac{R}{M}\right)\right\rfloor\right)\tilde{d}_e. \tag{107}$$

Putting all relevant values of $\alpha_{b,r,t}$, $\bar{\alpha}_{b,r,t}$ and $\hat{\theta}_{b,r,t}$ for **Case 1** to the RHS of (95) yields

$$\bar{d}_{\mathrm{icsi}}^{\mathrm{f}} = d_B(R)\left(1 + d_B(R)\tilde{d}_e + \epsilon\right) \tag{108}$$

where we have defined the Singleton bound [44]

$$d_B(R) \triangleq n_{\mathrm{r}}\left(1 + \left\lfloor B\left(n_{\mathrm{t}} - \frac{R}{M}\right)\right\rfloor\right). \tag{109}$$

- **Case 2:** $\epsilon > \hat{d}_e - 1 + \delta$

  Similarly to **Case 1**, we first note that for $bB + t \leq \frac{BR}{M}$, the values of $\alpha_{b,r,t}$ attaining the infimum are given by zero. This implies that the values for $\bar{\alpha}_{b,r,t}$, $bB + t \leq \frac{BR}{M}$ attaining the infimum are given by $-\tilde{d}_e$.

  We continue by considering the case of $bB + t > \frac{BR}{M}$. For any (b'', r'', t'') such that $b''B + t'' > \frac{BR}{M}$, if (b'', r'', t'') belongs to $\{(b'', r'', t'') : -\tilde{d}_e \leq \bar{\alpha}_{b'',r'',t''} = \alpha_{b'',r'',t''} - \tilde{d}_e < 0\}$, we have that

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \hat{d}_e - 1 + \delta = \hat{d}_e + \delta + (Bn_{\mathrm{r}}n_{\mathrm{t}} - 1)\tilde{d}_e + \alpha_{b'',r'',t''} + \sum_{(b',r,t)\neq(b'',r'',t'')}\bar{\alpha}_{b',r,t} \tag{110}$$

$$\geq \alpha_{b'',r'',t''}. \tag{111}$$

Similarly to **Case 1**, this in conjunction with equations (97)–(99) implies that the constraint in (96) can never be met. As such, (b'', r'', t'') for $b''B + t'' > \frac{BR}{M}$ must belong to $\{(b'', r'', t'') : \bar{\alpha}_{b'',r'',t''} \geq 0, \alpha_{b'',r'',t''} \geq \tilde{d}_e\}$ and the values of $\bar{\alpha}_{b,r,t}$ that solve the infimum in (95) are therefore given by zero. Thus, the infimum in (95) can be attained with

$$\bar{\alpha}_{b,r,t} = \begin{cases} 0, & \text{for } bB + t > \frac{BR}{M} \\ -\tilde{d}_e, & \text{otherwise,} \end{cases} \tag{112}$$

which results in

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + d_B(R)\tilde{d}_e. \tag{113}$$

This yields

$$\bar{d}^{\text{f}}_{\text{icsi}} = d_B(R) \left(\hat{d}_e + d_B(R)\tilde{d}_e + \delta \right). \tag{114}$$

From **Cases 1** and **2**, by letting $\epsilon, \delta \downarrow 0$, we have an upper bound to the outage exponent

$$\bar{d}^{\text{f}}_{\text{icsi}} = d_B(R) \left(\min(1, \hat{d}_e) + d_B(R)\tilde{d}_e \right). \tag{115}$$

B. GMI Lower Bound

Replacing $\mathcal{O}_{\mathcal{X}}$ in (85) with $\underline{\mathcal{O}}_{\mathcal{X}}$ in (93) yields a lower bound $\underline{d}^{\text{f}}_{\text{icsi}} \leq d^{\text{f}}_{\text{icsi}}$ where

$$\underline{d}^{\text{f}}_{\text{icsi}} = \inf_{\mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \underline{\mathcal{O}}_{\mathcal{X}}} \left\{ \sum_{(b,r,t): \, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0, \, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right. $$
$$\left. \sum_{(b,r,t): \, \bar{\alpha}_{b,r,t} \geq 0, \alpha_{b,r,t} \geq \tilde{d}_e, \, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right\}. \tag{116}$$

In the following, we solve $\underline{d}^{\text{f}}_{\text{icsi}}$ using the same power exponent $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ used to derive the upper bound (cf. (94)), i.e.,

$$\varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) = 1 + Bn_{\text{r}}n_{\text{t}}\tilde{d}_e + \sum_{b'=1}^{B} \sum_{r=1}^{n_{\text{r}}} \sum_{t=1}^{n_{\text{t}}} \bar{\alpha}_{b',r,t}, \qquad b = 1, \ldots, B \tag{117}$$

and show that this exponent yields $\underline{d}^{\text{f}}_{\text{icsi}} = \bar{d}^{\text{f}}_{\text{icsi}}$. Following the derivation in [21, App. C], we can express $\underline{\mathcal{O}}_{\mathcal{X}}$ in (93) for discrete constellations of size $|\mathcal{X}| = 2^M$ as

$$\underline{\mathcal{O}}_{\mathcal{X}} = \left\{ \mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \mathbb{R}^{Bn_{\text{r}} \times n_{\text{t}}} : \sum_{b=1}^{B} \underline{\kappa}_b < \frac{BR}{M} \right\} \tag{118}$$

where we have defined, for some $\epsilon, \delta > 0$, the following variables

$$\underline{\kappa}_b \triangleq \left| \underline{\mathcal{S}}_b^{(\epsilon,\delta)} \right|, \tag{119}$$

$$\underline{\mathcal{S}}_b^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_{\text{r}}} \underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}, \tag{120}$$

$$\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \left\{ t : \{\alpha_{b,r,t} < \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) - \epsilon \} \cap \{\alpha_{b,r,t} < \hat{\theta}_{\min} + \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) - \delta \}, \ t = 1, \ldots, n_{\text{t}} \right\}, \tag{121}$$

$$\hat{\theta}_{\min} \triangleq \min\{\hat{\theta}_{1,1,1}, \ldots, \hat{\theta}_{B,n_{\text{r}},n_{\text{t}}}\}. \tag{122}$$

We observe from (116) and (121) that increasing $\hat{\theta}_{\min}$ increases both the objective function in (116) and the threshold for $\alpha_{b,r,t}$ in (121). Thus, the value of $\hat{\theta}_{\min}$ that solves the infimum (116) is given by $\hat{d}_e - 1$. Since for any b, r, t we have $\hat{\theta}_{b,r,t} \geq \hat{\theta}_{\min}$, the values of $\hat{\theta}_{b,r,t}$ that solve the infimum are also given by $\hat{d}_e - 1$ as they do not appear in $\underline{\mathcal{O}}_{\mathcal{X}}$.

We next solve $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ that attain the infimum on the RHS of (116). To accomplish this, we compare $\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ in (121) with $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ in (99). There are two main differences between $\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ and $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$. Firstly, in the set $\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$, we have $\hat{\theta}_{\min}$ as the threshold for $\alpha_{b,r,t}$ instead of $\theta_{b,r,t}$ in the set $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$. However, since the value of $\hat{\theta}_{\min}$ that solves the infimum in (116) is also given by $\hat{d}_e - 1$ (same as the value of $\hat{\theta}_{b,r,t}$ that solves the infimum in (95)), this difference will not contrast the resulting infimum in (95) and (116). Secondly, we have an additional term in the set $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ that depends on the phases $\phi_{b,r,t}^{h}$ and $\phi_{b,r,t}^{\hat{e}}$, i.e.,

$$\left\{ \{\alpha_{b,r,t} \leq \varpi_b + \epsilon\} \cap \{\alpha_{b,r,t} > \hat{\theta}_{b,r,t} + \delta\} \cap \mathcal{Q}_{b,r,t} \right\} \tag{123}$$

where

$$\mathcal{Q}_{b,r,t} = \left\{ \phi_{b,r,t}^{h}, \phi_{b,r,t}^{\hat{e}} \in [0, 2\pi) : \cos\left(\phi_{b,r,t}^{\hat{e}} - \phi_{b,r,t}^{h}\right) > 0 \right\}. \tag{124}$$

The infimum solution in (95) is obtained when the event $\mathcal{Q}_{b,r,t}$ does not occur.

It follows that since the infimum solutions for both $\hat{\theta}_{\min}$ and $\hat{\theta}_{b,r,t}$ are identical and the set (123) is not active in solving the infimum in (95), the result for the infimum (116) has a similar form to that for (95), i.e.,

$$\underline{d}_{\text{icsi}}^{\text{f}} = d_B(R) \left(\min(1 - \epsilon, \hat{d}_e - \delta) + d_B(R)\tilde{d}_e \right) \tag{125}$$

where $d_B(R)$ has been defined in (109). By letting $\epsilon, \delta \downarrow 0$, combining (115) with (125) completes the proof of Theorem 1.

APPENDIX C

CAUSAL-CSIT POWER ALLOCATION

For causal CSIT, we have $n(b) = b - \tau_{\text{d}}$. The exponent of the optimal power allocation policy must satisfy the constraint (74). To analyze the outage exponent, we use the maximum power

exponent satisfying (74), i.e.,

$$\varpi_b\left(\bar{A}^{(n(b))}\right) = 1 + \sum_{b'=1}^{b-\tau_d} \sum_{r=1}^{n_r} \sum_{t=1}^{n_t} \left(\bar{\alpha}_{b',r,t} + \tilde{d}_e\right) \tag{126}$$

for both GMI upper and lower bounds in Appendix A-D. In the subsections that follow, we show that using $\varpi_b\left(\bar{A}^{(n(b))}\right)$ in (126), we obtain tight outage exponent bounds.

Similarly to Appendix B, we note that it suffices to consider solving the outage exponent for discrete inputs with alphabet size $|\mathcal{X}| = 2^M$. For identification, superscript ^{c} will be used to indicate causal-CSIT results.

A. GMI upper bound

An upper bound to the outage exponent with causal CSIT $\bar{d}_{\mathrm{icsi}}^c$ has a similar expression to the one with full non-causal CSIT (cf. equation (95)), i.e.,

$$\bar{d}_{\mathrm{icsi}}^c = \inf_{A,\bar{A},\hat{\Theta}\in\overline{\mathcal{O}}_{\mathcal{X}}} \left\{ \sum_{(b,r,t): -\tilde{d}_e\leq\bar{\alpha}_{b,r,t}=\alpha_{b,r,t}-\tilde{d}_e<0,\, \hat{\theta}_{b,r,t}\geq\hat{d}_e-1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right.$$
$$\left. \sum_{(b,r,t): \bar{\alpha}_{b,r,t}\geq 0,\, \alpha_{b,r,t}\geq\tilde{d}_e,\, \hat{\theta}_{b,r,t}\geq\hat{d}_e-1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right\}, \tag{127}$$

where $\overline{\mathcal{O}}_{\mathcal{X}}$ is now obtained using the power exponent $\varpi_b(\bar{A}^{(n(b))})$ in (126) instead of the one in (94). We can express $\overline{\mathcal{O}}_{\mathcal{X}}$ as

$$\overline{\mathcal{O}}_{\mathcal{X}} = \left\{ A,\bar{A},\hat{\Theta}\in\mathbb{R}^{Bn_r\times n_t} : \sum_{b=1}^{B} \overline{\kappa}_b < \frac{BR}{M} \right\} \tag{128}$$

where we have defined the following

$$\overline{\kappa}_b \triangleq \left| \overline{\mathcal{S}}_b^{(\epsilon,\delta)} \right|, \tag{129}$$

$$\overline{\mathcal{S}}_b^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_{\mathrm{r}}} \overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}, \tag{130}$$

$$\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \left\{ t : \left\{ \{\alpha_{b,r,t} \le \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \epsilon \} \cap \{\alpha_{b,r,t} \le \hat{\theta}_{b,r,t} + \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \delta \} \right\} \right.$$

$$\cup \left\{ \{\alpha_{b,r,t} \le \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \epsilon \} \cap \{\alpha_{b,r,t} > \hat{\theta}_{b,r,t} + \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \delta \} \cap \mathcal{Q}_{b,r,t} \right\},$$

$$\left. t = 1, \ldots, n_{\mathrm{t}} \right\}, \tag{131}$$

$$\mathcal{Q}_{b,r,t} \triangleq \left\{ \phi_{b,r,t}^h, \phi_{b,r,t}^{\hat{e}} \in [0, 2\pi) : \cos \left(\phi_{b,r,t}^{\hat{e}} - \phi_{b,r,t}^h \right) > 0 \right\} \tag{132}$$

for some $\epsilon, \delta > 0$.

To solve the infimum in (127), we first define the following

$$d^{\ddagger} \triangleq Bn_{\mathrm{t}} - \left\lceil \frac{BR}{M} \right\rceil + 1. \tag{133}$$

Following the same argument used in Appendix B-A, the infimum solutions for $\hat{\theta}_{b,r,t}$, for all b, r, t in (127) are given by $\hat{d}_e - 1$.

We next solve $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ attaining the infimum in (127). For each $r = 1, \ldots, n_{\mathrm{r}}$, assume, without loss of generality, the following conditions that make the constraint in $\overline{\mathcal{O}}_{\mathcal{X}}$ tight

$$\alpha_{b,r,t} > \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \min \left(\epsilon, \hat{d}_e - 1 + \delta \right), \left(\phi_{b,r,t}^h, \phi_{b,r,t}^{\hat{e}} \right) \notin \mathcal{Q}_{b,r,t}, \quad bn_{\mathrm{t}} + t \le d^{\ddagger}, \tag{134}$$

$$\alpha_{b,r,t} \le \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \min \left(\epsilon, \hat{d}_e - 1 + \delta \right), \quad bn_{\mathrm{t}} + t > d^{\ddagger}. \tag{135}$$

Thus, the infimum in (127) is achieved with $\alpha_{b,r,t}$ equal to

$$\alpha_{b,r,t} = \bar{\vartheta}_{b,r,t} = \begin{cases} \varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) + \min \left(\epsilon, \hat{d}_e - 1 + \delta \right), & \text{for } bn_{\mathrm{t}} + t \le d^{\ddagger} \\ 0, & \text{for } bn_{\mathrm{t}} + t > d^{\ddagger}. \end{cases} \tag{136}$$

For $b = 1, \ldots, \tau_{\mathrm{d}}$, we have $\varpi_b(\bar{\mathsf{A}}^{(n(b))}) = 1$ due to the causality of the CSIT. The exponent $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ sets a threshold for $\alpha_{b,r,t}$ in (131) (deep-fade threshold). Since increasing $\bar{\alpha}_{b',r,t}$, $b' = 1, \ldots, b - \tau_{\mathrm{d}}$ increases both $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ and the objective function in (127), it follows that

the solutions for $\bar{\alpha}_{b,r,t}$ that attain the infimum in (127) are given by

$$
\bar{\alpha}_{b,r,t} =
\begin{cases}
\bar{\vartheta}_{b,r,t} - \tilde{d}_e, & \text{if } \bar{\vartheta}_{b,r,t} < \tilde{d}_e \\
0, & \text{if } \bar{\vartheta}_{b,r,t} \geq \tilde{d}_e
\end{cases}
\tag{137}
$$

which can also be written as

$$
\bar{\alpha}_{b,r,t} = \min\left(\bar{\vartheta}_{b,r,t} - \tilde{d}_e, 0 \right).
\tag{138}
$$

Using these $\bar{\alpha}_{b,r,t}$, we have that for $b = \tau_{\mathrm{d}} + 1, \ldots, B$,

$$
\varpi_b \left(\bar{\mathsf{A}}^{(n(b))} \right) = 1 + \sum_{b'=1}^{b-\tau_{\mathrm{d}}} \sum_{r=1}^{n_{\mathrm{r}}} \sum_{t=1}^{n_{\mathrm{t}}} \min\left(\bar{\vartheta}_{b',r,t}, \tilde{d}_e \right).
\tag{139}
$$

Let

$$
\hat{b} = \max_{\left\{ b : \, bn_{\mathrm{t}} \leq d^{\ddagger} \right\}} b.
\tag{140}
$$

It follows from (136)–(139) that by letting $\epsilon, \delta \downarrow 0$, the infimum in (127) is given by

$$
\bar{d}_{\mathrm{icsi}}^{\mathrm{c}} = n_{\mathrm{t}} n_{\mathrm{r}} \sum_{b=1}^{\hat{b}} \bar{\vartheta}_b + n_{\mathrm{r}} \left(d^{\ddagger} - \hat{b} n_{\mathrm{t}} \right) \bar{\vartheta}_{\hat{b}+1}
\tag{141}
$$

where for $b = 1, \ldots, \min(\tau_{\mathrm{d}}, \hat{b} + 1)$

$$
\bar{\vartheta}_b = \min(1, \hat{d}_e)
\tag{142}
$$

and for $b = \min(\tau_{\mathrm{d}}, \hat{b} + 1) + 1, \ldots, \hat{b} + 1$

$$
\bar{\vartheta}_b = \min(1, \hat{d}_e) + n_{\mathrm{r}} n_{\mathrm{t}} \sum_{b'=1}^{b-\tau_{\mathrm{d}}} \min\left(\bar{\vartheta}_{b'}, \tilde{d}_e \right).
\tag{143}
$$

B. GMI Lower Bound

A lower bound to the outage exponent with causal CSIT $\underline{d}_{\mathrm{icsi}}^{\mathrm{c}}$ has an equivalent expression to the one with full non-causal CSIT (cf. equation (116)), i.e.,

$$
\underline{d}_{\mathrm{icsi}}^{\mathrm{c}} = \inf_{\mathsf{A},\bar{\mathsf{A}},\hat{\Theta}\in\underline{\mathcal{O}}_{\mathcal{X}}} \left\{ \sum_{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t}=\alpha_{b,r,t}-\tilde{d}_e<0,\, \hat{\theta}_{b,r,t}\geq \hat{d}_e-1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right.
$$
$$
\left. \sum_{(b,r,t):\, \bar{\alpha}_{b,r,t}\geq 0,\, \alpha_{b,r,t}\geq \tilde{d}_e,\, \hat{\theta}_{b,r,t}\geq \hat{d}_e-1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t}+1-\hat{d}_e\right) \right\} \tag{144}
$$

where now $\underline{\mathcal{O}}_{\mathcal{X}}$ is characterized by power exponent $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ satisfying the constraint (74) and accounting for $n(b) = b - \tau_{\mathrm{d}}$. Following the derivation in [21, App. D], we can express $\underline{\mathcal{O}}_{\mathcal{X}}$ for discrete constellations of size $|\mathcal{X}| = 2^M$ and causal-CSIT power allocation as

$$
\underline{\mathcal{O}}_{\mathcal{X}} = \left\{ \mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \mathbb{R}^{Bn_{\mathrm{r}}\times n_{\mathrm{t}}} : \sum_{b=1}^{B} \underline{\kappa}_b < \frac{BR}{M} \right\}. \tag{145}
$$

Here we have

$$
\underline{\kappa}_b \triangleq \left| \underline{\mathcal{S}}_b^{(\epsilon,\delta)} \right|, \tag{146}
$$

$$
\underline{\mathcal{S}}_b^{(\epsilon,\delta)} \triangleq \bigcup_{r=1}^{n_{\mathrm{r}}} \underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}, \tag{147}
$$

$$
\underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)} \triangleq \left\{ t : \{\alpha_{b,r,t} < \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) - \epsilon\} \cap \{\alpha_{b,r,t} < \hat{\theta}_{\min} + \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) - \delta\},\ t = 1,\ldots,n_{\mathrm{t}} \right\} \tag{148}
$$

where we have defined $\hat{\theta}_{\min} \triangleq \{\hat{\theta}_{1,1,1}, \ldots, \hat{\theta}_{B,n_{\mathrm{r}},n_{\mathrm{t}}}\}$.

Using maximum power exponent $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ in (126), a lower bound to the outage exponent can be obtained by solving $\hat{\theta}_{b,r,t}$, $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ that attain the infimum in (144). Since increasing $\hat{\theta}_{\min}$ increases the threshold for $\alpha_{b,r,t}$ in (148), the values of $\hat{\theta}_{\min}$ achieving the infimum (144) are given by $\hat{d}_e - 1$. Following the same argument in Appendix B-B, all other values of $\hat{\theta}_{b,r,t}$ attaining the infimum are also given by $\hat{d}_e - 1$.

We continue by solving $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ on the RHS of (144) by noting that $\hat{\theta}_{\min} = \hat{d}_e - 1$. For $r = 1, \ldots, n_{\mathrm{r}}$, assume, without loss of generality, the following conditions that make the

constraint $\underline{\mathcal{O}}_{\mathcal{X}}$ tight

$$\alpha_{b,r,t} > \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(-\epsilon, \hat{d}_e - 1 - \delta\right), \quad bn_\mathrm{t} + t \leq d^\ddagger, \tag{149}$$

$$\alpha_{b,r,t} \leq \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(-\epsilon, \hat{d}_e - 1 - \delta\right), \quad bn_\mathrm{t} + t > d^\ddagger \tag{150}$$

where $d^\ddagger$ has been defined in (133). We can then see that the infimum (144) can be achieved with $\alpha_{b,r,t}$ equal to

$$\alpha_{b,r,t} = \underline{\vartheta}_{b,r,t} = \begin{cases} \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(-\epsilon, \hat{d}_e - 1 - \delta\right), & \text{for } bn_\mathrm{t} + t \leq d^\ddagger \\ 0, & \text{for } bn_\mathrm{t} + t > d^\ddagger. \end{cases} \tag{151}$$

For $b = 1, \ldots, \tau_\mathrm{d}$, we have $\varpi_b(\bar{\mathsf{A}}^{(n(b))}) = 1$ and thus

$$\underline{\vartheta}_{b,r,t} = \min(1 - \epsilon, \ \hat{d}_e - \delta), \qquad b = 1, \ldots, \tau_\mathrm{d}. \tag{152}$$

We next evaluate the infimum solutions for $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ for all $b = 1, \ldots, B$ by analyzing $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$. The power exponent $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ sets a threshold for $\alpha_{b,r,t}$ in (148). Since increasing $\bar{\alpha}_{b',r,t}$, $b' = 1, \ldots, b - \tau_\mathrm{d}$ increases both $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ and the objective function in (144), it follows that the solutions for $\bar{\alpha}_{b',r,t}$ attaining the infimum in (144) are given by

$$\bar{\alpha}_{b',r,t} = \begin{cases} \underline{\vartheta}_{b',r,t} - \tilde{d}_e, & \text{if } \underline{\vartheta}_{b',r,t} < \tilde{d}_e \\ 0, & \text{if } \underline{\vartheta}_{b',r,t} \geq \tilde{d}_e \end{cases} \tag{153}$$

which can also be written as

$$\bar{\alpha}_{b',r,t} = \min\left(\underline{\vartheta}_{b',r,t} - \tilde{d}_e, 0\right). \tag{154}$$

Using these $\bar{\alpha}_{b',r,t}$, we have that for $b = \tau_\mathrm{d} + 1, \ldots, B$

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + \sum_{b'=1}^{b-\tau_\mathrm{d}} \sum_{r=1}^{n_\mathrm{r}} \sum_{t=1}^{n_\mathrm{t}} \min\left(\underline{\vartheta}_{b',r,t}, \tilde{d}_e\right). \tag{155}$$

Combining (155) and (152) with (151) provides complete solutions for $\alpha_{b,r,t}$ that attain the infimum in (144).

Recall that $d^\ddagger$ and $\hat{b}$ are defined in (133) and (140), respectively. It follows from (151)–(155)

that the infimum in (144) is given by

$$\underline{d}^{\text{c}}_{\text{icsi}} = n_{\text{t}} n_{\text{r}} \sum_{b=1}^{\hat{b}} \underline{\vartheta}_b + n_{\text{r}} \left(d^{\ddagger} - \hat{b} n_{\text{t}} \right) \underline{\vartheta}_{\hat{b}+1} \tag{156}$$

where for $b = 1, \ldots, \min(\tau_{\text{d}}, \hat{b} + 1)$

$$\underline{\vartheta}_b = \min(1 - \epsilon, \hat{d}_e - \delta) \tag{157}$$

and for $b = \min(\tau_{\text{d}}, \hat{b} + 1) + 1, \ldots, \hat{b} + 1$

$$\underline{\vartheta}_b = \min(1 - \epsilon, \hat{d}_e - \delta) + n_{\text{r}} n_{\text{t}} \sum_{b'=1}^{b-\tau_{\text{d}}} \min\left(\underline{\vartheta}_{b'}, \tilde{d}_e \right). \tag{158}$$

By letting $\epsilon, \delta \downarrow 0$, we can see that the lower bound (156) coincides with the upper bound (141), completing the proof of Theorem 2.

APPENDIX D
PREDICTIVE-CSIT POWER ALLOCATION

For predictive CSIT, we have $n(b) = b + \tau_{\text{f}}$. The exponent of the optimal power allocation must satisfy (74), i.e.,

$$\sup_{\substack{\bar{\text{A}}^{(n(b))} \in \mathbb{R}_{+}^{n(b) \cdot n_{\text{r}} \cdot n_{\text{t}}}, \\ \Phi^{\bar{\text{H}}^{(n(b))}} \in [0,2\pi)^{n(b) \cdot n_{\text{r}} \cdot n_{\text{t}}}}} \left\{ \varpi_b \left(\bar{\text{A}}^{(n(b))} \right) - \sum_{b'=1}^{\min(B,b+\tau_{\text{f}})} \sum_{r=1}^{n_{\text{r}}} \sum_{t=1}^{n_{\text{t}}} (\bar{\alpha}_{b',r,t} + \tilde{d}_e) \right\} \leq 1. \tag{159}$$

Here we have incorporated $\min(B, b+\tau_{\text{f}})$ to limit only fading estimates for the current codeword as fading matrices beyond the current codeword do not affect the current transmission.

Similarly to Appendix B, we note that it suffices to consider solving the outage exponent for discrete inputs with alphabet size $|\mathcal{X}| = 2^M$. In the following, superscript p is used to indicate predictive-CSIT results.

A. GMI Upper Bound

An upper bound to the outage exponent with predictive CSIT $\bar{d}^{\mathrm{p}}_{\mathrm{icsi}}$ has a similar expression to the one with full non-causal CSIT (cf. equation (95)), i.e.,

$$
\bar{d}^{\mathrm{p}}_{\mathrm{icsi}} = \inf_{\mathsf{A},\bar{\mathsf{A}},\hat{\Theta} \in \overline{\mathcal{O}}_{\mathcal{X}}} \left\{ \sum_{(b,r,t):\, -\tilde{d}_e \leq \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0,\, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right.
$$

$$
\left. \sum_{(b,r,t):\, \bar{\alpha}_{b,r,t} \geq 0,\, \alpha_{b,r,t} \geq \tilde{d}_e,\, \hat{\theta}_{b,r,t} \geq \hat{d}_e - 1} \alpha_{b,r,t} + \bar{\alpha}_{b,r,t} + \left(\hat{\theta}_{b,r,t} + 1 - \hat{d}_e \right) \right\} \quad (160)
$$

where the set $\overline{\mathcal{O}}_{\mathcal{X}}$ is now obtained using power exponent $\varpi_b(\cdot)$ that satisfies the constraint (159). As argued in Appendix A-D, to derive an upper bound to the outage exponent based on $\overline{\mathcal{O}}_{\mathcal{X}}$, we use the maximum power exponent satisfying the constraint (159), namely

$$
\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + \sum_{b'=1}^{\min(B,b+\tau_{\mathrm{f}})} \sum_{r=1}^{n_{\mathrm{r}}} \sum_{t=1}^{n_{\mathrm{t}}} \left(\bar{\alpha}_{b',r,t} + \tilde{d}_e \right). \quad (161)
$$

For discrete constellations of size $|\mathcal{X}| = 2^M$, we can express $\overline{\mathcal{O}}_{\mathcal{X}}$ with predictive-CSIT power exponent as

$$
\overline{\mathcal{O}}_{\mathcal{X}} = \left\{ \mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \mathbb{R}^{Bn_{\mathrm{r}} \times n_{\mathrm{t}}} : \sum_{b=1}^{B} \overline{\kappa}_b < \frac{BR}{M} \right\} \quad (162)
$$

where $\overline{\kappa}_b = |\overline{\mathcal{S}}_b^{(\epsilon,\delta)}|$ and $\overline{\mathcal{S}}_b^{(\epsilon,\delta)} = \bigcup_{r=1}^{n_{\mathrm{r}}} \overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ are all defined in (97) and (98) but with $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ given in (161).

We next evaluate the solutions for $\hat{\theta}_{b,r,t}$, $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ that achieve the infimum in (160). Following the same argument in Appendix B, the infimum solutions for $\hat{\theta}_{b,r,t}$ are given by $\hat{d}_e - 1$. As $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ is non-decreasing with b, without loss of generality, for each $r = 1, \ldots, n_{\mathrm{r}}$, assume the following conditions[6]

$$
\alpha_{b,r,t} > \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(\epsilon, \hat{d}_e - 1 + \delta\right), \quad \left(\phi^h_{b,r,t}, \phi^{\hat{e}}_{b,r,t}\right) \notin \mathcal{Q}_{b,r,t}, \quad bn_{\mathrm{t}} + t \leq d^{\ddagger}, \quad (163)
$$

$$
\alpha_{b,r,t} \leq \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(\epsilon, \hat{d}_e - 1 + \delta\right), \quad bn_{\mathrm{t}} + t > d^{\ddagger} \quad (164)
$$

that satisfy $\overline{\mathcal{O}}_{\mathcal{X}}$ with a tight inequality in the constraint. Herein we have recalled the definition

---

[6]Noting that $\overline{\kappa}_b$, $\overline{\mathcal{S}}_b^{(\epsilon,\delta)}$ and $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ are similarly defined to the ones in (97)–(99) but with $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ given in (161).

of $d^\ddagger$ in (133), namely

$$d^\ddagger \triangleq Bn_\mathrm{t} - \left\lceil \frac{BR}{M} \right\rceil + 1. \tag{165}$$

Then, as $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ is non-decreasing with b, the infimum in (160) is achieved with

$$\alpha_{b,r,t} = \bar{\eta}_{b,r,t} = \begin{cases} \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \min\left(\epsilon, \hat{d}_e - 1 + \delta\right), & \text{for } bn_\mathrm{t} + t \le d^\ddagger \\ 0, & \text{for } bn_\mathrm{t} + t > d^\ddagger. \end{cases} \tag{166}$$

From these $\alpha_{b,r,t}$ values, we can determine the solutions for $\bar{\alpha}_{b,r,t}$ by looking at either the set $\{(b,r,t) : -\tilde{d}_e \le \bar{\alpha}_{b,r,t} = \alpha_{b,r,t} - \tilde{d}_e < 0\}$ or $\{(b,r,t) : \bar{\alpha}_{b,r,t} \ge 0, \alpha_{b,r,t} \ge \tilde{d}_e\}$. Since increasing $\bar{\alpha}_{b',r,t}$, $b' = 1, \ldots, \min(B, b + \tau_\mathrm{f})$ increases both $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ and the objective function in (144), it follows that the solutions for $\bar{\alpha}_{b',r,t}$ attaining the infimum in (144) are given by

$$\bar{\alpha}_{b',r,t} = \begin{cases} \bar{\eta}_{b',r,t} - \tilde{d}_e, & \text{if } \bar{\eta}_{b',r,t} < \tilde{d}_e \\ 0, & \text{if } \bar{\eta}_{b',r,t} \ge \tilde{d}_e \end{cases} \tag{167}$$

which can be written as

$$\bar{\alpha}_{b',r,t} = \min\left(\bar{\eta}_{b',r,t} - \tilde{d}_e, 0\right). \tag{168}$$

When $\bar{\eta}_{b,r,t} = 0$, we have $\bar{\alpha}_{b,r,t} = -\tilde{d}_e$; when $\bar{\eta}_{b,r,t} = \varpi_b(\bar{\mathsf{A}}^{(n(b))}) + \min(\epsilon, \hat{d}_e - 1 + \delta)$, we have $\bar{\alpha}_{b,r,t} = 0$ to achieve the infimum in (160).[7] Thus, as a consequence of (166), the infimum solutions for $\bar{\alpha}_{b,r,t}$ are given by

$$\bar{\alpha}_{b,r,t} = \begin{cases} 0, & \text{for } bn_\mathrm{t} + t \le d^\ddagger \\ -\tilde{d}_e, & \text{for } bn_\mathrm{t} + t > d^\ddagger. \end{cases} \tag{169}$$

The values of $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ when $\bar{\alpha}_{b,r,t}$ attaining the infimum in (160) are then given by

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = \begin{cases} 1 + n_\mathrm{t} n_\mathrm{r} \left(b + \tau_\mathrm{f}\right) \tilde{d}_e, & b + \tau_\mathrm{f} \le \hat{b} \\ 1 + n_\mathrm{r} d^\ddagger \tilde{d}_e, & b + \tau_\mathrm{f} > \hat{b} \end{cases} \tag{170}$$

where $\hat{b}$ has been defined in (140), namely

$$\hat{b} = \max_{\{b : bn_\mathrm{t} \le d^\ddagger\}} b. \tag{171}$$

---

[7]If $\alpha_{b,r,t} = \varpi_b(\bar{\mathsf{A}}^{(n(b))}) + \min(\epsilon, \hat{d}_e - 1 + \delta)$, then $\bar{\alpha}_{b,r,t}$ must belong to $\{(b,r,t) : \bar{\alpha}_{b,r,t} \ge 0, \alpha_{b,r,t} \ge \tilde{d}_e, \hat{\theta}_{b,r,t} \ge \hat{d}_e\}$ since otherwise there is no solution for the infimum in (160).

It follows from (166) (by letting $\epsilon, \delta \downarrow 0$) and from (173) that the infimum solutions for $\alpha_{b,r,t}$ are given by

$$\alpha_{b,r,t} = \begin{cases} \bar{\eta}_b, & \text{for } bn_{\text{t}} + t \leq d^{\ddagger} \\ 0, & \text{for } bn_{\text{t}} + t > d^{\ddagger}. \end{cases} \tag{172}$$

where we have defined $\bar{\eta}_b$ as

$$\bar{\eta}_b \triangleq \begin{cases} \min(1, \hat{d}_e) + n_{\text{t}} n_{\text{r}} \left(b + \tau_{\text{f}}\right) \tilde{d}_e, & b + \tau_{\text{f}} \leq \hat{b} \\ \min(1, \hat{d}_e) + n_{\text{r}} d^{\ddagger} \tilde{d}_e, & b + \tau_{\text{f}} > \hat{b}. \end{cases} \tag{173}$$

We further observe from (169) and (160) that the sum of $\bar{\alpha}_{b,r,t}$ contributing to the infimum in (160) is zero. Thus, combining the solutions for $\alpha_{b,r,t}$ and $\bar{\alpha}_{b,r,t}$ in the preceding evaluation yields the infimum in (160) to be

$$\bar{d}_{\text{icsi}}^{\text{P}} = n_{\text{t}} n_{\text{r}} \sum_{b=1}^{\hat{b}} \bar{\eta}_b + n_{\text{r}} \left(d^{\ddagger} - \hat{b} n_{\text{t}}\right) \bar{\eta}_{\hat{b}+1}. \tag{174}$$

B. GMI Lower Bound

Following the same explanation as in the causal-CSIT case (Appendix C), the power exponent to prove a tight lower bound to the outage exponent is identical to the one used to prove the upper bound, i.e.,

$$\varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) = 1 + \sum_{b'=1}^{\min(B, b+\tau_{\text{f}})} \sum_{r=1}^{n_{\text{r}}} \sum_{t=1}^{n_{\text{t}}} (\bar{\alpha}_{b',r,t} + \tilde{d}_e). \tag{175}$$

Using this predictive-CSIT power exponent, we can then express the set $\underline{\mathcal{O}}_{\mathcal{X}}$ used to analyze a lower bound to the outage exponent as

$$\underline{\mathcal{O}}_{\mathcal{X}} = \left\{ \mathsf{A}, \bar{\mathsf{A}}, \hat{\Theta} \in \mathbb{R}^{Bn_{\text{r}} \times n_{\text{t}}} : \sum_{b=1}^{B} \underline{\kappa}_b < \frac{BR}{M} \right\} \tag{176}$$

where $\underline{\kappa}_b = |\underline{\mathcal{S}}_b^{(\epsilon,\delta)}|$ and $\underline{\mathcal{S}}_b^{(\epsilon,\delta)} = \bigcup_{r=1}^{n_{\text{r}}} \underline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$ are all defined in (119) and (120) but with $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ replaced by $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ in (175).

The evaluation of the lower bound follows along the same line as that of the upper bound (Appendix D-A). We first compare the high-SNR sets $\overline{\mathcal{O}}_{\mathcal{X}}$ and $\underline{\mathcal{O}}_{\mathcal{X}}$ corresponding to GMI upper and lower bounds, respectively. For the GMI upper bound (Appendix D-A), equation (163)

implies that the additional term in $\overline{\mathcal{S}}_{b,r}^{(\epsilon,\delta)}$, i.e.,

$$\left\{ \{\alpha_{b,r,t} \leq \varpi_b\left(\bar{\mathsf{A}}^{(n(b))}\right) + \epsilon\} \cap \{\alpha_{b,r,t} > \hat{\theta}_{b,r,t} + \delta\} \cap \mathcal{Q}_{b,r,t} \right\} \qquad (177)$$

does not affect the solution for the infimum. Hence, by not considering this additional term and by letting $\epsilon, \delta \downarrow 0$, it is straightforward to show that $\overline{\mathcal{O}}_{\mathcal{X}}$ derived using $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ in (161) and $\underline{\mathcal{O}}_{\mathcal{X}}$ derived using $\varpi_b(\bar{\mathsf{A}}^{(n(b))})$ in (175) tend to be identical. It can then be shown that the resulting lower bound to the outage exponent coincides to the upper bound (174). The proof for the lower bound to the outage exponent is not reproduced here for the sake of compactness.

REFERENCES

[1] 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.

[2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[3] E. Malkamäki and H. Leib, "Coded diversity on block-fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 771–781, Mar. 1999.

[4] 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.

[5] 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, pp. 765–779, Feb. 2012.

[6] T. T. Kim, K. D. Nguyen, and A. Guillén i Fàbregas, "Coded modulation with mismatched CSIT over MIMO block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5631–5640, Nov. 2010.

[7] K. D. Nguyen, N. Letzepis, A. Guillén i Fàbregas, and L. K. Rasmussen, "Outage diversity of MIMO block-fading channels with causal channel state information," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, Jun. 2010.

[8] T. T. Kim and G. Caire, "Diversity gains of power control with noisy CSIT in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1618–1626, Apr. 2009.

[9] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley, 2006.

[11] K. D. Nguyen, "Adaptive transmission for block-fading channels," Ph.D. dissertation, University of South Australia, 2009.

[12] T. T. Kim, G. Caire, and M. Skoglund, "Decode-and-forward relaying with quantized channel state feedback: An outage exponent analysis," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4548–4564, Oct. 2008.

[13] V. Aggarwal and A. Sabharwal, "Power-controlled feedback and training for two-way MIMO channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3310–3331, Jul. 2010.

[14] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Power allocation for block-fading channels with arbitrary input constellations," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2514–2523, May 2009.

[15] ——, "Outage exponents of block-fading channels with power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2373–2381, May 2010.

[16] J. Chen and K.-K. Wong, "Communication with causal CSI and controlled information outage," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2221 – 2229, May 2009.

[17] R. Negi and J. M. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2478 – 2494, Sep. 2002.

[18] K. D. Nguyen, N. Letzepis, A. Guillén i Fàbregas, and L. K. Rasmussen, "Causal/predictive channel state information in block-fading channels," *Submitted to IEEE Trans. Inf. Theory*, Jul. 2010.

[19] 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.

[20] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sep. 1999.

[21] A. T. Asyhari and A. Guillén i Fàbregas, "Nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1483–1517, Mar. 2012.

[22] G. Kaplan and S. Shamai (Shitz), "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.

[23] S. Shamai (Shitz) and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[24] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[25] 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.

[26] R. Knopp and G. Caire, "Power control and beamforming for systems with multiple transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 638–648, Oct. 2002.

[27] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. Eighth Int. Symp. Signal Process. and Its Applicat.*, Sydney, Australia, Aug. 2005.

[28] X. J. Zhang, Y. Gong, and K. B. Letaief, "Power control and channel training for MIMO channels: A DMT perspective," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2080–2089, Jul. 2011.

[29] A. Ganti, A. Lapidoth, and İ. 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.

[30] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[31] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "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.

[32] A. Chuang, A. Guillén i Fàbregas, L. K. Rasmussen, and I. B. Collings, "Optimal throughput-diversity-delay tradeoff in MIMO ARQ block-fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3968–3986, Sep. 2008.

[33] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[34] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels–Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[35] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[36] V. Aggarwal and A. Sabharwal, "Bits about the channel: Multiround protocols for two-way fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3352–3370, Jun. 2011.

[37] A. T. Asyhari and A. Guillén i Fàbregas, "Mismatched CSI outage exponents of block-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.–Aug. 2011.

[38] ——, "GMI and mismatched-CSI outage exponents in MIMO block-fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013.

[39] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[40] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[41] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Commun. and Inf. Theory*, vol. 5, no. 1-2, pp. 1–153, 2008.

[42] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[43] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[44] R. Knopp and P. A. Humblet, "On coding for block fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.