

Backing Off From Infinity: Performance Bounds via Concentration of Spectral Measure for Random MIMO Channels

Yuxin Chen, *Student Member, IEEE*, Andrea J. Goldsmith, *Fellow, IEEE*, and Yonina C. Eldar, *Fellow, IEEE*

Abstract—The performance analysis of random vector channels, particularly multiple-input-multiple-output (MIMO) channels, has largely been established in the asymptotic regime of large channel dimensions, due to the analytical intractability of characterizing the exact distribution of the objective performance metrics. This paper exposes a new nonasymptotic framework that allows the characterization of many canonical MIMO system performance metrics to within a narrow interval under finite channel dimensionality, provided that these metrics can be expressed as a separable function of the singular values of the matrix. The effectiveness of our framework is illustrated through two canonical examples. In particular, we characterize the mutual information and power offset of random MIMO channels, as well as the minimum mean squared estimation error of MIMO channel inputs from the channel outputs. Our results lead to simple, informative, and reasonably accurate control of various performance metrics in the finite-dimensional regime, as corroborated by the numerical simulations. Our analysis framework is established via the concentration of spectral measure phenomenon for random matrices uncovered by Guionnet and Zeitouni, which arises in a variety of random matrix ensembles irrespective of the precise distributions of the matrix entries.

Index Terms—MIMO, massive MIMO, confidence interval, concentration of spectral measure, random matrix theory, non-asymptotic analysis, mutual information, MMSE.

I. INTRODUCTION

THE past decade has witnessed an explosion of developments in multi-dimensional vector channels [1], particularly multiple-input-multiple-output (MIMO) channels. The exploitation of multiple (possibly correlated) dimensions provides various benefits in wireless communication and signal processing systems, including channel capacity gain, improved energy efficiency, and enhanced robustness against noise and channel variation.

Manuscript received February 8, 2014; revised October 24, 2014; accepted October 25, 2014. Date of publication October 28, 2014; date of current version December 22, 2014. This work was supported in part by the NSF under Grant CCF-0939370 and Grant CIS-1320628, in part by the AFOSR within the MURI Programme under Grant FA9550-09-1-0643, and in part by the BSF Transformative Science under Grant 2010505.

Y. Chen is with the Department of Electrical Engineering and the Department of Statistics, Stanford University, Stanford, CA 94305 USA (e-mail: yxchen@stanford.edu).

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@ee.stanford.edu).

Y. C. Eldar is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Communicated by A. Lozano, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2014.2365497

Although many fundamental MIMO system performance metrics can be evaluated via the precise spectral distributions of finite-dimensional MIMO channels (e.g. channel capacity [2], [3], minimum mean square error (MMSE) estimates of vector channel inputs from the channel outputs [4], power offset [5], sampled capacity loss [6]), this approach often results in prohibitive analytical and computational complexity in characterizing the probability distributions and confidence intervals of these MIMO system metrics. In order to obtain more informative analytical insights into the MIMO system performance metrics, a large number of works (see [5], [7]–[21]) present more explicit expressions for these performance metrics with the aid of random matrix theory. Interestingly, when the number of input and output dimensions grow, many of the MIMO system metrics taking the form of linear spectral statistics converge to deterministic limits, due to various limiting laws and universality properties of (asymptotically) large random matrices [22]. In fact, the spectrum (i.e. singular-value distribution) of a random channel matrix \mathbf{H} tends to stabilize when the channel dimension grows, and the limiting distribution is often universal in the sense that it is independent of the precise distributions of the entries of \mathbf{H} .

These asymptotic results are well suited for massive MIMO communication systems. However, the limiting regime falls short in providing a full picture of the phenomena arising in most practical systems which, in general, have moderate dimensionality. While the asymptotic convergence rates of many canonical MIMO system performance metrics have been investigated as well, how large the channel dimension must be largely depends on the realization of the growing matrix sequences. In this paper, we propose an alternative method via concentration of measure to evaluate many canonical MIMO system performance metrics for finite-dimension channels, assuming that the target performance metrics can be transformed into linear spectral statistics of the MIMO channel matrix. Moreover, we show that the metrics fall within narrow intervals with high (or even overwhelming) probability in the nonasymptotic regime.

A. Related Work

Random matrix theory is one of the central topics in probability theory with many connections to wireless communications and signal processing. Several random matrix ensembles, such as Gaussian unitary ensembles, admit exact

characterization of their eigenvalue distributions [23] under any channel dimensionality. These spectral distributions associated with the finite-dimensional MIMO channels can then be used to compute analytically the distributions and confidence intervals of the MIMO system performance metrics (e.g. mutual information of MIMO fading channels [24]–[28]). However, the computational intractability of integrating a large-dimensional function over a finite-dimensional MIMO spectral distribution precludes concise and informative capacity expressions even in moderate-sized problems. For this reason, theoretical analysis based on precise eigenvalue characterization is generally limited to small-dimensional vector channels.

In comparison, one of the central topics in modern random matrix theory is to derive limiting distributions for the eigenvalues of random matrix ensembles of interest, which often turns out to be simple and informative. Several pertinent examples include the semi-circle law for symmetric Wigner matrices [29], the circular law for i.i.d. matrix ensembles [30], [31], and the Marchenko–Pastur law [32] for rectangular random matrices. One remarkable feature of these asymptotic laws is the *universality phenomenon*, whereby the limiting spectral distributions are often indifferent to the precise distribution of each matrix entry. This phenomenon allows theoretical analysis to accommodate a broad class of random matrix families beyond Gaussian ensembles. See [22] for a beautiful and self-contained exposition of these limiting results.

The simplicity and universality of these asymptotic laws have inspired a large body of work in characterizing the asymptotic performance limits of random vector channels. For instance, the limiting results for i.i.d. Gaussian ensembles have been applied in early work [2] to establish the linear increase of MIMO capacity with the number of antennas. This approach was then extended to accommodate separable correlation fading models [8], [12], [14]–[16], [33], [34] with general non-Gaussian distributions (particularly Rayleigh and Ricean fading [5], [13]). These models admit analytically-friendly solutions, and have become the cornerstone for many results for random MIMO channels. In addition, several works have characterized the second-order asymptotics (often in the form of central limit theorems) for mutual information [16], [35], [36], information density [37], second-order coding rates [38], [39], diversity-multiplexing tradeoff [40], etc., which uncover the asymptotic convergence rates of various MIMO system performance metrics under a large family of random matrix ensembles. The limiting tail of the distribution (or large deviation) of the mutual information has also been determined in the asymptotic regime [41]. In addition to information theoretic metrics, other signal processing and statistical metrics like MMSE [7], [42], [43], multiuser efficiency for CDMA systems [10], optical capacity [44], covariance matrix and principal components [45], canonical correlations [46], and likelihood ratio test statistics [46], have also been investigated via asymptotic random matrix theory. While these asymptotic laws have been primarily applied to performance metrics in the form of linear spectral statistics, more general performance metrics can

be approximated using the delicate method of “deterministic equivalents” (see [47], [48]).

A recent trend in statistics is to move from asymptotic laws towards non-asymptotic analysis of random matrices [49], [50], which aims at revealing statistical effects of a moderate-to-large number of components, assuming a sufficient amount of independence among them. One prominent effect in this context is the *concentration of spectral measure phenomenon* [51], [52], which indicates that many separable functions of a matrix’s singular values (called *linear spectral statistics* [53]) can be shown to fall within a narrow interval with high probability even in the moderate-dimensional regime. This phenomenon has been investigated in various fields such as high-dimensional statistics [49], statistical learning [54], and compressed sensing [50].

Inspired by the success of the measure concentration methods in the statistics literature, our recent work [6] develops a non-asymptotic approach to quantify the capacity of multi-band channels under random sub-sampling strategies, which to our knowledge is the first to exploit the concentration of spectral measure phenomenon to analyze random MIMO channels. In general, the concentration of measure phenomena are much less widely recognized and used in the communication community than in the statistics and signal processing community. Recent emergence of massive MIMO technologies [55]–[58], which uses a large number of antennas to obtain both capacity gain and improved radiated energy efficiency, provides a compelling application of these methods to characterize system performance. Other network/distributed MIMO systems (see [59], [60]) also require analyzing large-dimensional random vector channels. It is our aim here to develop a general framework that promises new insights into the performance of these emerging technologies under moderate-to-large channel dimensionality.

B. Contributions

We develop a general non-asymptotic framework for deriving performance bounds of random vector channels or any general MIMO system, based on the powerful concentration of spectral measure phenomenon of random matrices as revealed by Guionnet and Zeitouni [51]. Specifically, we introduce a general recipe that can be used to assess various MIMO system performance metrics to within vanishingly small confidence intervals, assuming that the objective metrics can be transformed into linear spectral statistics associated with the MIMO channel. Our framework and associated results can accommodate a large class of probability distributions for random channel matrices, including those with bounded support, a large class of sub-Gaussian measures, and heavy-tailed distributions. To broaden the range of metrics that we can accurately evaluate, we also develop a general concentration of spectral measure inequality for the cases where only the *exponential mean* (instead of the mean) of the objective metrics can be computed.

We demonstrate the effectiveness of our approach through two illustrative examples: (1) mutual information of random vector channels under equal power allocation; (2) MMSE in

estimating signals transmitted over random MIMO channels. These examples allow concise and, informative characterizations even in the presence of moderate-to-high SNR and moderate-to-large channel dimensionality. In contrast to a large body of prior works that focus on first-order limits or asymptotic convergence rate of the target performance metrics, we are able to derive full characterization of these metrics in the non-asymptotic regime. Specifically, we obtain narrow confidence intervals with precise order and reasonably accurate pre-constants, which do not rely on careful choice of the growing matrix sequence. Our numerical simulations also corroborate that our theoretical predictions are reasonably accurate in the finite-dimensional regime.

C. Organization and Notation

The rest of the paper is organized as follows. Section II introduces several families of probability measures investigated in this paper. We present a general framework characterizing the concentration of spectral measure phenomenon in Section III. In Section IV, we illustrate the general framework using a small sample of canonical examples. Finally, Section V concludes the paper with a summary of our findings and a discussion of several future directions.

For convenience of presentation, we let \mathbb{C} denote the set of complex numbers. For any function $g(x)$, the Lipschitz norm of $g(\cdot)$ is defined as

$$\|g\|_{\mathcal{L}} := \sup_{x \neq y} \left| \frac{g(x) - g(y)}{x - y} \right|. \quad (1)$$

We let $\lambda_i(\mathbf{A})$, $\lambda_{\max}(\mathbf{A})$, and $\lambda_{\min}(\mathbf{A})$ represent the i th largest eigenvalue, the largest eigenvalue, and the smallest eigenvalue of a Hermitian matrix \mathbf{A} , respectively, and use $\|\cdot\|$ to denote the operator norm (or spectral norm). The set $\{1, 2, \dots, n\}$ is denoted as $[n]$, and we write $\binom{[n]}{k}$ for the set of all k -element subsets of $[n]$. For any set s , we use $\text{card}(s)$ to represent the cardinality of s . Also, for any two index sets s and t , we denote by $\mathbf{A}_{s,t}$ the submatrix of \mathbf{A} containing the rows at indices in s and columns at indices in t . We use $x \in a + [b, c]$ to indicate that x lies within the interval $[a + b, a + c]$. Finally, the standard notation $f(n) = \mathcal{O}(g(n))$ means that there exists a constant $c > 0$ such that $f(n) \leq cg(n)$; $f(n) = \Theta(g(n))$ indicates that there are universal constants $c_1, c_2 > 0$ such that $c_1g(n) \leq f(n) \leq c_2g(n)$; and $f(n) = \Omega(g(n))$ indicates that there are universal constants $c > 0$ such that $f(n) \leq cg(n)$. Throughout this paper, we use $\log(\cdot)$ to represent the natural logarithm. Our notation is summarized in Table I.

II. THREE FAMILIES OF PROBABILITY MEASURES

In this section, we define precisely three classes of probability measures with different properties of the tails, which lay the foundation of the analysis in this paper.

Definition 1 (Bounded Distribution): We say that a random variable $X \in \mathbb{C}$ is bounded by D if

$$\mathbb{P}(|X| \leq D) = 1. \quad (2)$$

The class of probability distributions that have bounded support subsumes as special cases a broad class of distributions

TABLE I
SUMMARY OF NOTATION AND PARAMETERS

$\log(\cdot)$	natural logarithm
κ	$\kappa = 1$ for the real-valued case and $\kappa = 2$ for the complex-valued case
ν_{ij}, ν	the standard deviation of \mathbf{M}_{ij} , and the maximum standard deviation $\nu := \max_{i,j} \nu_{ij}$
c_{ls}	logarithmic Sobolev constant as defined in (3)
$\ \cdot\ _{\mathcal{L}}$	Lipschitz constant of a function
D	maximum absolute value in the compact support of a bounded measure
τ_c, σ_c	truncation threshold and the associated standard deviation defined in (7) and (8), respectively
ρ	spectral radius of the deterministic matrix \mathbf{R}
$[n]$	$[n] := \{1, 2, \dots, n\}$
$\binom{[n]}{k}$	set of all k -element subsets of $[n]$

encountered in practice (e.g. the Bernoulli distribution and uniform distribution). Another class of probability measures that possess light tails (i.e. sub-Gaussian tails) is defined as follows.

Definition 2 (Logarithmic Sobolev Measure): We say that a random variable $X \in \mathbb{C}$ with cumulative probability distribution (CDF) $F(\cdot)$ satisfies the logarithmic Sobolev inequality (LSI) with uniform constant c_{ls} if, for any differentiable function h , one has

$$\int h^2(x) \log \left(\frac{h^2(x)}{\int h^2(z) dF(z)} \right) dF(x) \leq 2c_{\text{ls}} \int |h'(x)|^2 dF(x). \quad (3)$$

Remark 1: One of the most popular techniques in demonstrating measure concentration is the “entropy method” (see, [61], [62]), which hinges upon establishing inequalities of the form

$$\text{KL}(\mathbb{P}_X^t \|\mathbb{P}_X) \leq O(t^2) \quad (4)$$

for some measure \mathbb{P}_X and its tilted measure \mathbb{P}_X^t , where $\text{KL}(\cdot\|\cdot)$ denotes the Kullback–Leibler divergence. See [62, Chapter 3] for detailed definitions and derivation. It turns out that for a probability measure satisfying the LSI, one can pick the function $h(\cdot)$ in (3) in some appropriate fashion to yield the entropy-type inequality (4). In fact, the LSI has been well recognized as one of the most fundamental criteria to demonstrate exponentially sharp concentration for various metrics.

Remark 2: While the measure satisfying the LSI necessarily exhibits sub-Gaussian tails (see [61], [62]), many concentration results under log-Sobolev measures cannot be extended to general sub-Gaussian distributions (e.g. for bounded measures the concentration results have only been shown for convex functions instead of general Lipschitz functions).

A number of measures satisfying the LSI have been discussed in the expository paper [63]. One of the most important examples is the standard Gaussian distribution, which satisfies the LSI with logarithmic Sobolev constant $c_{\text{ls}} = 1$. In many situations, the probability measures obeying the LSI exhibit very sharp concentration of spectral measure phenomenon (typically sharper than general bounded measures).

While prior works on measure concentration focus primarily on measures with bounded support or measures with sub-Gaussian tails, it is also of interest to accommodate a more general class of distributions (e.g. heavy-tailed distributions). For this purpose, we introduce sub-exponential distributions and heavy-tailed distributions as follows.

Definition 3 (Sub-Exponential Distribution): A random variable X is said to be sub-exponentially distributed with parameter λ if it satisfies

$$\mathbb{P}(|X| > x) \leq c_0 e^{-\lambda x}, \quad \forall x > 0 \quad (5)$$

for some absolute constant $c_0 > 0$.

Definition 4 (Heavy-Tailed Distribution): A random variable X is said to be heavy-tailed distributed if

$$\lim_{x \rightarrow \infty} e^{-\lambda x} \mathbb{P}(|X| > x) = \infty, \quad \forall \lambda > 0. \quad (6)$$

That said, sub-exponential (resp. heavy-tailed) distributions are those probability measures whose tails are lighter (resp. heavier) than exponential distributions. Commonly encountered heavy-tailed distributions include log-normal and power-law distributions.

To facilitate analysis, for both sub-exponential and heavy-tailed distributions, we define the following two quantities τ_c and σ_c with respect to X for some sequence $c(n)$ such that

$$\mathbb{P}(|X| > \tau_c) \leq \frac{1}{mn^{c(n)+1}}, \quad (7)$$

and

$$\sigma_c = \sqrt{\mathbb{E}[|X|^2 \mathbf{1}_{\{|X| < \tau_c\}}]}. \quad (8)$$

In short, τ_c represents a truncation threshold such that the truncated X coincides with the true X with high probability, and σ_c denotes the standard deviation of the truncated X . The idea is to study $X \mathbf{1}_{\{|X| < \tau_c\}}$ instead of X in the analysis, since the truncated value is bounded and exhibits light tails. For instance, if X is exponentially distributed such that $\mathbb{P}(|X| > x) = e^{-x}$ and $c(n) = \log n$, then this yields that

$$\tau_c = \log^2 n + \log(mn).$$

Typically, one would like to pick $c(n)$ such that τ_c becomes a small function of m and n obeying $\sigma_c^2 \approx \mathbb{E}[|X|^2]$.

III. CONCENTRATION OF SPECTRAL MEASURE IN LARGE RANDOM VECTOR CHANNELS

We present a general mathematical framework that facilitates the analysis of random vector channels. The proposed approach is established upon the concentration of spectral measure phenomenon derived by Guionnet and Zeitouni [51], which is a consequence of Talagrand's concentration inequalities [64]. While the adaptation of such general concentration results to our settings requires moderate mathematical effort, it leads to a very effective framework to assess the fluctuation of various MIMO system performance metrics. We will provide a few canonical examples in Section IV to illustrate the power of this framework.

Consider a random matrix $\mathbf{M} = [\mathbf{M}_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m}$, where \mathbf{M}_{ij} 's are assumed to be independently distributed.

We use $\mathbf{M}_{ij}^{\text{R}}$ and $\mathbf{M}_{ij}^{\text{I}}$ to represent respectively the real and imaginary parts of \mathbf{M}_{ij} , which are also generated independently. Note, however, that \mathbf{M}_{ij} 's are *not* necessarily drawn from identical distributions. Set the numerical value

$$\kappa := \begin{cases} 1, & \text{in the real-valued case,} \\ 2, & \text{in the complex-valued case.} \end{cases} \quad (9)$$

We further assume that all entries have matching two moments such that for all i and j :

$$\mathbb{E}[\mathbf{M}_{ij}] = 0, \quad \mathbf{Var}(\mathbf{M}_{ij}^{\text{R}}) = v_{ij}^2, \quad (10)$$

$$\mathbf{Var}(\mathbf{M}_{ij}^{\text{I}}) = \begin{cases} v_{ij}^2, & \text{in the complex-valued case,} \\ 0, & \text{in the real-valued case,} \end{cases} \quad (11)$$

where v_{ij}^2 ($v_{ij} \geq 0$) are uniformly bounded by

$$\max_{i,j} |v_{ij}| \leq v. \quad (12)$$

In this paper, we focus on the class of MIMO system performance metrics that can be transformed into additively separable functions of the eigenvalues of the random channel matrix (called *linear spectral statistics* [53]), i.e. the type of metrics taking the form

$$\sum_{i=1}^n f\left(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right) \quad (13)$$

for some function $f(\cdot)$ and some deterministic matrix $\mathbf{R} \in \mathbb{C}^{m \times m}$, where $\lambda_i(\mathbf{X})$ denotes the i th eigenvalue of a matrix \mathbf{X} . As can be seen, many vector channel performance metrics (e.g. MIMO mutual information, MMSE, sampled channel capacity loss) can all be transformed into certain forms of linear spectral statistics. We will characterize in Proposition 1 and Theorem 1 the measure concentration of (13), which quantifies the fluctuation of (13) in finite-dimensional random vector channels.

A. Concentration Inequalities for the Spectral Measure

Many linear spectral statistics of random matrices sharply concentrate within a narrow interval indifferent to the precise entry distributions. Somewhat surprisingly, such spectral measure concentration is shared by a very large class of random matrix ensembles. We provide formal quantitative illustration of such concentration of spectral measure phenomena below, which follows from [51, Corollary 1.8]. This behavior, while not widely used in communication and signal processing, is a general result in probability theory derived from the Talagrand concentration inequality (see [64] for details).

Before proceeding to the concentration results, we define

$$f_0(\mathbf{M}) := \frac{1}{n} \sum_{i=1}^{\min(m,n)} f\left(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right) \quad (14)$$

for the sake of notational simplicity.

Proposition 1: Consider a random matrix $\mathbf{M} = [\mathbf{M}_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m}$ satisfying (10) and (12). Suppose that $\mathbf{R} \in \mathbb{C}^{m \times m}$ is any given matrix and $\|\mathbf{R}\| \leq \rho$. Consider a function $f(x)$ such that $g(x) := f(x^2)$ ($x \in \mathbb{R}^+$) is a real-valued Lipschitz function with Lipschitz constant $\|g\|_{\mathcal{L}}$

as defined in (1). Let κ and ν be defined in (9) and (12), respectively.

(a) *Bounded Measure*: If \mathbf{M}_{ij} 's are bounded by D and $g(\cdot)$ is convex, then for any $\beta > 8\sqrt{\pi}$,

$$|f_0(\mathbf{M}) - \mathbb{E}[f_0(\mathbf{M})]| \leq \frac{\beta D \rho \nu \|g\|_{\mathcal{L}}}{n} \quad (15)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{8\kappa}\right)$.

(b) *Logarithmic Sobolev Measure*: If the measure of \mathbf{M}_{ij} satisfies the LSI with a uniformly bounded constant c_{ls} , then for any $\beta > 0$,

$$|f_0(\mathbf{M}) - \mathbb{E}[f_0(\mathbf{M})]| \leq \frac{\beta \sqrt{c_{\text{ls}}} \rho \nu \|g\|_{\mathcal{L}}}{n} \quad (16)$$

with probability exceeding $1 - 2 \exp\left(-\frac{\beta^2}{\kappa}\right)$.

(c) *Sub-Exponential and Heavy-Tailed Measure*: Suppose that \mathbf{M}_{ij} 's are independently drawn from either sub-exponential distributions or heavy-tailed distributions and that their distributions are symmetric about 0, with τ_c and σ_c defined respectively in (7) and (8) with respect to \mathbf{M}_{ij} for some sequence $c(n)$. If $g(\cdot)$ is convex, then

$$\left|f_0(\mathbf{M}) - \mathbb{E}\left[f_0\left(\tilde{\mathbf{M}}\right)\right]\right| \leq \frac{2\kappa \sqrt{c(n)} \log n \tau_c \sigma_c \rho \|g\|_{\mathcal{L}}}{n} \quad (17)$$

with probability exceeding $1 - 5n^{-c(n)}$, where $\tilde{\mathbf{M}}$ is defined such that $\tilde{\mathbf{M}}_{ij} := \mathbf{M}_{ij} \mathbf{1}_{\{|\mathbf{M}_{ij}| < \tau_c\}}$.

Proof: See Appendix A. \square

Remark 3: By setting $\beta = \sqrt{\log n}$ in Proposition 1(a) and (b), one can see that under either measures of bounded support or measures obeying the LSI,

$$\left|f_0(\mathbf{M}) - \mathbb{E}\left[f_0\left(\tilde{\mathbf{M}}\right)\right]\right| = \mathcal{O}\left(\frac{\sqrt{\log n}}{n}\right) \quad (18)$$

with high probability. In contrast, under heavy-tailed measures,

$$\left|f_0(\mathbf{M}) - \mathbb{E}\left[f_0\left(\tilde{\mathbf{M}}\right)\right]\right| = \mathcal{O}\left(\frac{\sqrt{c(n)} \tau_c^2 \sigma_c^2 \log n}{n}\right) \quad (19)$$

with high probability, where $\sqrt{c(n)} \tau_c \sigma_c$ is typically a growing function in n . As a result, the concentration under sub-Gaussian distributions is sharper than that under heavy-tailed measures by a factor of $\sqrt{c(n)} \tau_c \sigma_c$.

Remark 4: The bounds derived in Proposition 1 scale linearly with ν , which is the maximum standard deviation of the entries of \mathbf{M} . This allows us to assess the concentration phenomena for matrices with entries that have non-uniform variance.

Proposition 1(a) and 1(b) assert that for both measures of bounded support and a large class of sub-Gaussian distributions, many separable functions of the spectra of random matrices exhibit sharp concentration, assuming a sufficient amount of independence between the entries. More remarkably, the tails behave at worst like a Gaussian random variable with well-controlled variance. Note that for a bounded measure, we require the objective metric of the form (13) to satisfy certain convexity conditions in order to guarantee concentration. In contrast, the fluctuation of general

Lipschitz functions can be well controlled for logarithmic Sobolev measures. This agrees with the prevailing wisdom that the standard Gaussian measure (which satisfies the LSI) often exhibits sharper concentration than general bounded distributions (e.g. Bernoulli measures).

Proposition 1(c) demonstrates that spectral measure concentration arises even when the tail distributions of \mathbf{M}_{ij} are much heavier than standard Gaussian random variables, although it might not be as sharp as for sub-Gaussian measures. This remarkable feature comes at a price, namely, the deviation of the objective metrics is much less controlled than for sub-Gaussian distributions. However, this degree of concentration might still suffice for most practical purposes. Note that the concentration result for heavy-tailed distributions is stated in terms of the truncated version $\tilde{\mathbf{M}}$. The nice feature of the truncated $\tilde{\mathbf{M}}$ is that its entries are all bounded (and hence sub-Gaussian), which can often be quantified or estimated in a more convenient fashion. Finally, we remark that the concentration depends on the choice of the sequence $c(n)$, which in turn affects the size of τ_c and σ_c . We will illustrate the resulting size of confidence intervals in Section III-C via several examples.

B. Approximation of Expected Empirical Distribution

Although Proposition 1 ensures sharp measure concentration of various linear spectral statistics, a more precise characterization requires evaluating the mean value of the target metric (13) (i.e. $\mathbb{E}[f_0(\mathbf{M})]$). While limiting laws often admit simple asymptotic characterization of this mean value for a general class of metrics, whether the convergence rate can be quantified often needs to be studied on a case-by-case basis. In fact, this has become an extensively researched topic in mathematics (see [65]–[68]). In this subsection, we develop an approximation result that allows the expected value of a broader class of metrics to be well approximated by concise and informative expressions.

Recall that

$$f_0(\mathbf{M}) := \frac{1}{n} \sum_{i=1}^{\min(m,n)} f\left(\lambda_i\left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right) \quad (20)$$

We consider a large class of situations where the exponential mean of the target metric (13)

$$\begin{aligned} \mathcal{E}(f) &:= \mathbb{E}\left[\exp(n f_0(\mathbf{M}))\right] \\ &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{\min\{m,n\}} f\left(\lambda_i\left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right)\right)\right] \end{aligned} \quad (21)$$

(instead of $\mathbb{E}[f_0(\mathbf{M})]$) can be approximated in a reasonably accurate manner. This is particularly relevant when $f(\cdot)$ is a logarithmic function. For example, this applies to log-determinant functions

$$f_0(\mathbf{M}) = \frac{1}{n} \log \det\left(\mathbf{I} + \frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right),$$

which are of significant interest in various applications such as wireless communications [69], multivariate hypothesis testing [70], etc. While $\mathbb{E}[\det(\mathbf{I} + \frac{1}{n} \mathbf{M} \mathbf{M}^*)]$ often admits a

TABLE II
SUMMARY OF PARAMETERS OF THEOREM 1

$c_{\rho,f,D} := \log \left(1 + \sqrt{8\kappa\pi} D \rho \nu \ g\ _{\mathcal{L}} e^{\frac{8\pi}{\kappa} + 2\kappa D^2 \rho^2 \nu^2 \ g\ _{\mathcal{L}}^2} \right)$
$c_{\rho,f,c_{1s}} := \log \left(1 + \sqrt{4\kappa\pi} c_{1s} \rho \nu \ g\ _{\mathcal{L}} e^{\frac{\kappa}{4} c_{1s} \rho^2 \nu^2 \ g\ _{\mathcal{L}}^2} \right)$
$c_{\rho,f,\tau_c,\sigma_c} := \log \left(1 + \sqrt{8\kappa\pi} \tau_c \sigma_c \rho \ g\ _{\mathcal{L}} e^{\frac{8\pi}{\kappa} + 2\kappa \tau_c^2 \sigma_c^2 \rho^2 \ g\ _{\mathcal{L}}^2} \right)$

simple distribution-free expression, $\mathbb{E} \left[\log \det \left(\mathbf{I} + \frac{1}{n} \mathbf{M} \mathbf{M}^* \right) \right]$ is highly dependent on precise distributions of the entries of the matrix.

One might already notice that, by Jensen's inequality, $\log \mathcal{E}(f)$ is larger than the mean objective metric $\mathbb{E}[f_0(\mathbf{M})]$. Nevertheless, in many situations, these two quantities differ by only a vanishingly small gap, which is formally demonstrated in the following lemma.

Lemma 1: (a) Sub-Exponential Tail: Suppose that

$$\mathbb{P} \left(|f_0(\mathbf{M}) - \mathbb{E}[f_0(\mathbf{M})]| > \frac{y}{n} \right) \leq c_1 \exp(-c_2 y)$$

for some values $c_1 > 0$ and $c_2 > 1$, then

$$\frac{1}{n} \log \mathcal{E}(f) - \frac{\log \left(1 + \frac{c_1}{c_2 - 1} \right)}{n} \leq \mathbb{E}[f_0(\mathbf{M})] \leq \frac{1}{n} \log \mathcal{E}(f) \quad (22)$$

with $\mathcal{E}(f)$ defined in (21).

(b) Sub-Gaussian Tail: Suppose that

$$\mathbb{P} \left(|f_0(\mathbf{M}) - \mathbb{E}[f_0(\mathbf{M})]| > \frac{y}{n} \right) \leq c_1 \exp(-c_2 y^2)$$

for some values $c_1 > 0$ and $c_2 > 0$, then

$$\begin{aligned} \frac{1}{n} \log \mathcal{E}(f) &\geq \mathbb{E}[f_0(\mathbf{M})] \\ &\geq \frac{1}{n} \log \mathcal{E}(f) - \frac{\log \left(1 + \sqrt{\frac{\pi c_2^2}{c_2}} \exp \left(\frac{1}{4c_2} \right) \right)}{n} \end{aligned} \quad (23)$$

with $\mathcal{E}(f)$ defined in (21).

Proof: See Appendix B. \square

Remark 5: For measures with sub-exponential tails, if $c_2 < 1$, the concentration is not decaying sufficiently fast and is unable to ensure that $\mathcal{E}(f) = \mathbb{E}[e^{nf_0(\mathbf{M})}]$ exists.

In short, Lemma 1 asserts that if $f_0(\mathbf{M})$ possesses a sub-exponential or a sub-Gaussian tail, then $\mathbb{E}[f_0(\mathbf{M})]$ can be approximated by $\frac{1}{n} \log \mathcal{E}(f)$ in a reasonably tight manner, namely, within a gap no worse than $\mathcal{O}\left(\frac{1}{n}\right)$. Since Proposition 1 implies sub-Gaussian tails for various measures, we immediately arrive at the following concentration results that concern a large class of sub-Gaussian and heavy-tailed measures.

Theorem 1: Let $c_{\rho,f,D}$, $c_{\rho,f,c_{1s}}$, and $c_{\rho,f,\tau_c,\sigma_c}$ be numerical values defined in Table II, and set

$$\mu_{\rho,g,A} := \nu \rho \|g\|_{\mathcal{L}} A. \quad (24)$$

(1) Bounded Measure: Under the assumptions of Proposition 1(a), we have

$$\begin{cases} f_0(\mathbf{M}) \leq \frac{1}{n} \log \mathcal{E}(f) + \frac{\beta \mu_{\rho,g,D}}{n}, \\ f_0(\mathbf{M}) \geq \frac{1}{n} \log \mathcal{E}(f) - \frac{\beta \mu_{\rho,g,D}}{n} - \frac{c_{\rho,f,D}}{n}, \end{cases} \quad (25)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{8\kappa}\right)$.

(2) Logarithmic Sobolev Measure: Under the assumptions of Proposition 1(b), we have

$$\begin{cases} f_0(\mathbf{M}) \leq \frac{1}{n} \log \mathcal{E}(f) + \frac{\beta \mu_{\rho,g,\sqrt{c_{1s}}}}{n}, \\ f_0(\mathbf{M}) \geq \frac{1}{n} \log \mathcal{E}(f) - \frac{\beta \mu_{\rho,g,\sqrt{c_{1s}}}}{n} - \frac{c_{\rho,f,c_{1s}}}{n}, \end{cases} \quad (26)$$

with probability at least $1 - 2 \exp\left(-\frac{\beta^2}{\kappa}\right)$.

(3) Heavy-Tailed Distribution: Under the assumptions of Proposition 1(c), we have

$$\begin{cases} f_0(\mathbf{M}) \leq \frac{1}{n} \log \mathcal{E}_{\tilde{\mathbf{M}}}(f) + \frac{\mu_{\rho,g,\zeta}}{n}, \\ f_0(\mathbf{M}) \geq \frac{1}{n} \log \mathcal{E}_{\tilde{\mathbf{M}}}(f) - \frac{\mu_{\rho,g,\zeta}}{n} - \frac{c_{\rho,f,\tau_c,\sigma_c}}{n}, \end{cases} \quad (27)$$

with probability exceeding $1 - 5n^{-c(n)}$, where

$$\begin{cases} \zeta := \tau_c \sigma_c \sqrt{8\kappa c(n) \log n}; \\ \mathcal{E}_{\tilde{\mathbf{M}}}(f) := \mathbb{E} \left[\exp \left(n f_0 \left(\tilde{\mathbf{M}} \right) \right) \right]. \end{cases}$$

Proof: See Appendix C. \square

While Lemma 1 focuses on sub-exponential and sub-Gaussian measures, we are able to extend the concentration phenomenon around $\frac{1}{n} \log \mathcal{E}(f)$ to a much larger class of distributions including heavy-tailed measures through certain truncation arguments.

To get a more informative understanding of Theorem 1, consider the case where $\|g\|_{\mathcal{L}}$, τ_c , D , c_{1s} , ρ are all constants. One can see that with probability exceeding $1 - \epsilon$ for any small constant $\epsilon > 0$,

$$f_0(\mathbf{M}) = \frac{1}{n} \log \mathcal{E}(f) + \mathcal{O}\left(\frac{1}{n}\right)$$

holds for measures of bounded support and measures satisfying the LSI. In comparison, for symmetric power-law measures satisfying $\mathbb{P}(|M_{ij}| \geq x) \leq x^{-\lambda}$ for some $\lambda > 0$, the function τ_c is typically a function of the form n^δ for some constant $\delta > 0$. In this case, with probability at least $1 - \epsilon$, one has

$$f_0(\mathbf{M}) = \frac{1}{n} \log \mathcal{E}_{\tilde{\mathbf{M}}}(f) + \mathcal{O}\left(\frac{1}{n^{1-\delta}}\right)$$

for heavy-tailed measures, where the uncertainty cannot be controlled as well as for bounded measures or measures obeying the LSI.

In the scenario where $\mathcal{E}(f)$ can be computed, Theorem 1 presents a full characterization of the confidence interval of the objective metrics taking the form of linear spectral statistics. In fact, in many applications, $\mathcal{E}(f)$ (rather than $\mathbb{E}[f_0]$) can be precisely computed for a general class of probability distributions beyond the Gaussian measure, which allows for accurate characterization of the concentration of the objective metrics via Theorem 1, as illustrated in Section IV.

C. Confidence Interval

In this subsection, we demonstrate that the sharp spectral measure concentration phenomenon allows us to estimate the mean of the target metric in terms of narrow confidence intervals.

Specifically, suppose that we have obtained the value of the objective metric $f_0(\mathbf{M}) = \frac{1}{n} \sum_i f(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right))$ for

a given realization \mathbf{M} . The goal is find an interval (called $(1 - \alpha_0)$ confidence interval)

$$[l(\mathbf{M}), u(\mathbf{M})]$$

such that

$$\mathbb{P}(\mathbb{E}[f_0] \in [l(\mathbf{M}), u(\mathbf{M})]) \geq 1 - \alpha_0 \quad (28)$$

for some constant $\alpha_0 \in (0, 1)$.

Consider the assumptions of Proposition 1, i.e. \mathbf{M}_{ij} 's are independently generated satisfying $\mathbb{E}[\mathbf{M}_{ij}] = 0$ and $\mathbb{E}[|\mathbf{M}_{ij}|^2] = v_{ij}^2 \leq v^2$. An immediate consequence of Proposition 1 is stated as follows.

- **(Bounded Measure)** If \mathbf{M}_{ij} 's are bounded by D and $g(\cdot)$ is convex, then

$$\left[f_0(\mathbf{M}) \pm \frac{\sqrt{8\kappa} D \rho v \|g\|_{\mathcal{L}} \sqrt{\log \frac{4}{\alpha_0}}}{n} \right] \quad (29)$$

is an $(1 - \alpha_0)$ confidence interval for $\mathbb{E}[f_0(\mathbf{M})]$.

- **(Logarithmic Sobolev Measure)** If the measure of \mathbf{M}_{ij} satisfies the LSI with a uniform constant c_{ls} , then

$$\left[f_0(\mathbf{M}) \pm \frac{\sqrt{\kappa c_{\text{ls}}} \rho v \|g\|_{\mathcal{L}} \sqrt{\log \frac{2}{\alpha_0}}}{n} \right] \quad (30)$$

is an $(1 - \alpha_0)$ confidence interval for $\mathbb{E}[f_0(\mathbf{M})]$.

- **(Sub-Exponential Measure)** If the measure of \mathbf{M}_{ij} is symmetric about 0 and $\mathbb{P}(|\mathbf{M}_{ij}| > x) \leq e^{-\lambda x}$ for some constant $\lambda > 0$, then $c(n) = \frac{\log(\frac{5}{\alpha_0})}{\log n}$ and $\tau_c = \frac{1}{\lambda} \log\left(\frac{5mn}{\alpha_0}\right)$, indicating that

$$\left[f_0(\mathbf{M}) \pm \frac{\sqrt{4\kappa \log\left(\frac{5}{\alpha_0}\right)} \rho v \|g\|_{\mathcal{L}} \log\left(\frac{5mn}{\alpha_0}\right)}{\lambda n} \right] \quad (31)$$

is an $(1 - \alpha_0)$ confidence interval for $\mathbb{E}[f_0(\tilde{\mathbf{M}})]$.

- **(Power-Law Measure)** If the measure of \mathbf{M}_{ij} is symmetric about 0 and $\mathbb{P}(|\mathbf{M}_{ij}| > x) \leq x^{-\lambda}$ for some constant $\lambda > 0$, then $c(n) = \frac{\log(\frac{5}{\alpha_0})}{\log n}$ and $\tau_c = \left(\frac{5mn}{\alpha_0}\right)^{\frac{1}{\lambda}}$, indicating that

$$\left[f_0(\mathbf{M}) \pm \frac{\left(\sqrt{4\kappa \log\left(\frac{5}{\alpha_0}\right)} \left(\frac{5}{\alpha_0}\right)^{\frac{1}{\lambda}} \rho v \|g\|_{\mathcal{L}}\right) (mn)^{\frac{1}{\lambda}}}{n} \right]$$

is an $(1 - \alpha_0)$ confidence interval for $\mathbb{E}[f_0(\tilde{\mathbf{M}})]$.

One can see from the above examples that the spans of the confidence intervals under power-law distributions are much less controlled than that under sub-Gaussian measure. Depending on the power-law decay exponent λ , the typical deviation can be as large as $\mathcal{O}\left(n^{-1+\frac{2}{\lambda}}\right)$ as compared to $\mathcal{O}\left(\frac{1}{n}\right)$ under various sub-Gaussian measures.

When D , c_{ls} , ρ , $\|g\|_{\mathcal{L}}$, and α_0 are all constants, the widths of the above confidence intervals decay with n , which is negligible for many metrics of interest.

D. A General Recipe for Applying

Proposition 1 and Theorem 1

For pedagogical reasons, we provide here a general recipe regarding how to apply Proposition 1 and Theorem 1 to evaluate the fluctuation of system performance metrics in random MIMO channels with channel matrix \mathbf{H} .

- 1) Transform the performance metric into a linear spectral statistic, i.e. write the metric in the form $f_0(\mathbf{M}) = \frac{1}{n} \sum_{i=1}^n f(\lambda_i(\frac{1}{n} \mathbf{H} \mathbf{R} \mathbf{R}^* \mathbf{H}^*))$ for some function $f(\cdot)$ and some deterministic matrix \mathbf{R} .
- 2) For measures satisfying the LSI, it suffices to calculate the Lipschitz constant of $g(x) := f(x^2)$. For both measures of bounded support and heavy-tailed distributions, since the function $g(x) := f(x^2)$ is non-convex in general, one typically needs to convexify $g(x)$ first. In particular, one might want to identify two reasonably tight approximation $g_1(x)$ and $g_2(x)$ of $g(x)$ such that: 1) $g_1(x)$ and $g_2(x)$ are both convex (or concave); 2) $g_1(x) \leq g(x) \leq g_2(x)$.
- 3) Apply Proposition 1 and / or Theorem 1 on either $g(x)$ (for measures satisfying the LSI) or on $g_1(x)$ and $g_2(x)$ (for bounded measures or heavy-tailed measures) to obtain concentration bounds.

This recipe will be used to establish the canonical examples provided in Section IV.

IV. SOME CANONICAL EXAMPLES

In this section, we apply our general analysis framework developed in Section III to a few canonical examples that arise in wireless communications and signal processing. Rather than making each example as general as possible, we present only simple settings that admit concise expressions from measure concentration. We emphasize these simple illustrative examples in order to demonstrate the effectiveness of our general treatment.

A. Mutual Information and Power Offset of Random MIMO Channels

Consider the following MIMO channel

$$\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{z}, \quad (32)$$

where $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ denotes the channel matrix, $\mathbf{x} \in \mathbb{R}^{n_t}$ represents the transmit signal, and $\mathbf{y} \in \mathbb{R}^{n_r}$ is the received signal. We denote by $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ the additive Gaussian noise. Note that (32) allows modeling of a large class of random vector channels (e.g. MIMO-OFDM channels [71], CDMA systems [18], undersampled channels [72]) beyond multiple antenna channels. For instance, in unfaded direct-sequence CDMA systems, the columns of \mathbf{H} can represent random spreading sequences; see [73, Sec. 3.1.1] for details.

The total power is assumed to be P , independent of n_t and n_r , and the signal-to-noise ratio (SNR) is denoted by

$$\text{SNR} := \frac{P}{\sigma^2}. \quad (33)$$

In addition, we denote the degrees of freedom as

$$n := \min(n_t, n_r), \quad (34)$$

and use $\alpha > 0$ to represent the ratio

$$\alpha := \frac{n_t}{n_r}. \quad (35)$$

We suppose throughout that α is a *universal constant* that does not scale with n_t and n_r .

Consider the simple channel model where $\{\mathbf{H}_{ij} : 1 \leq i \leq n_r, 1 \leq j \leq n_t\}$ are independently distributed. Suppose that channel state information (CSI) is available to both the transmitter and the receiver. When equal power allocation is adopted at all transmit antennas, it is well known that the mutual information $C(\mathbf{H}, \text{SNR})$ of the MIMO channel (32) under equal power allocation is [2]

$$C(\mathbf{H}, \text{SNR}) = \log \det \left(\mathbf{I} + \text{SNR} \cdot \frac{1}{n_t} \mathbf{H} \mathbf{H}^* \right), \quad (36)$$

which depends only on the eigenvalue distribution of $\mathbf{H} \mathbf{H}^*$. In the presence of asymptotically high SNR and channel dimensions, it is well known that if \mathbf{H}_{ij} 's are independent with zero mean and unit variance, then, almost surely,

$$\begin{aligned} & \lim_{\text{SNR} \rightarrow \infty} \lim_{n_r \rightarrow \infty} \left(\frac{C(\mathbf{H}, \text{SNR})}{n_r} - \min\{\alpha, 1\} \cdot \log \text{SNR} \right) \\ &= \begin{cases} -1 + (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right), & \text{if } \alpha \geq 1, \\ -\alpha \log(\alpha e) + (1 - \alpha) \log \left(\frac{1}{1 - \alpha} \right), & \text{if } \alpha < 1, \end{cases} \end{aligned} \quad (37)$$

which is independent of the precise entry distribution of \mathbf{H} (see [11]). The method of deterministic equivalents has also been used to obtain good approximations under finite channel dimensions [47, Ch. 6]. In contrast, our framework characterizes the concentration of mutual information in the presence of finite SNR and finite n_r with reasonably tight confidence intervals. Interestingly, the MIMO mutual information is well-controlled within a narrow interval, as formally stated in the following theorem.

Theorem 2: Assume perfect CSI at both the transmitter and the receiver, and equal power allocation at all transmit antennas. Suppose that $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ where \mathbf{H}_{ij} 's are independent random variables satisfying $\mathbb{E}[\mathbf{H}_{ij}] = 0$ and $\mathbb{E}[|\mathbf{H}_{ij}|^2] = 1$. Set $n := \min\{n_r, n_t\}$.

(a) If \mathbf{H}_{ij} 's are bounded by D , then for any $\beta > 8\sqrt{\pi}$,

$$\begin{aligned} & \frac{C(\mathbf{H}, \text{SNR})}{n_r} \\ & \in \begin{cases} \log \text{SNR} + \left[\frac{1}{n_r} \log \mathcal{R} \left(\frac{2}{e \text{SNR}}, n_r, n_t \right) + \frac{\beta r_{\text{bd}}^{\text{lb},+}}{n_r}, \right. \\ \quad \left. \frac{1}{n_r} \log \mathcal{R} \left(\frac{e}{2 \text{SNR}}, n_r, n_t \right) + \frac{\beta r_{\text{bd}}^{\text{ub},+}}{n_r} \right], & \text{if } \alpha \geq 1 \\ \alpha \log \frac{\text{SNR}}{\alpha} + \left[\frac{1}{n_r} \log \mathcal{R} \left(\frac{2\alpha}{e \text{SNR}}, n_t, n_r \right) + \frac{\beta r_{\text{bd}}^{\text{lb},-}}{n_r}, \right. \\ \quad \left. \frac{1}{n_r} \log \mathcal{R} \left(\frac{e\alpha}{2 \text{SNR}}, n_t, n_r \right) + \frac{\beta r_{\text{bd}}^{\text{ub},-}}{n_r} \right], & \text{if } \alpha < 1 \end{cases} \end{aligned} \quad (38)$$

with probability exceeding $1 - 8 \exp\left(-\frac{\beta^2}{8}\right)$.

(b) If \mathbf{H}_{ij} 's satisfy the LSI with respect to a uniform constant c_{ls} , then for any $\beta > 0$,

$$\begin{aligned} & \frac{C(\mathbf{H}, \text{SNR})}{n_r} \\ & \in \begin{cases} \log \text{SNR} + \frac{1}{n_r} \log \mathcal{R} \left(\frac{1}{\text{SNR}}, n_r, n_t \right) \\ \quad + \frac{\beta}{n_r} \left[r_{\text{ls}}^{\text{lb},+}, r_{\text{ls}}^{\text{ub},+} \right], & \text{if } \alpha \geq 1 \\ \alpha \log \frac{\text{SNR}}{\alpha} + \frac{1}{n_r} \log \mathcal{R} \left(\frac{\alpha}{\text{SNR}}, n_t, n_r \right) \\ \quad + \frac{\beta}{n_r} \left[r_{\text{ls}}^{\text{lb},-}, r_{\text{ls}}^{\text{ub},-} \right], & \text{if } \alpha < 1 \end{cases} \end{aligned} \quad (39)$$

with probability exceeding $1 - 4 \exp(-\beta^2)$.

(c) Suppose that \mathbf{H}_{ij} 's are independently drawn from either sub-exponential distributions or heavy-tailed distributions and that the distributions are symmetric about 0. Let $\tau_c(n)$ be defined as in (7) with respect to \mathbf{H}_{ij} 's for some sequence $c(n)$. Then,

$$\begin{aligned} & \frac{C(\mathbf{H}, \text{SNR})}{n_r} \\ & \in \begin{cases} \log \text{SNR} + \left[\frac{1}{n_r} \log \mathcal{R} \left(\frac{2}{e \sigma_c^2 \text{SNR}}, n_r, n_t \right) + \frac{r_{\text{ht}}^{\text{lb},+}}{n_r}, \right. \\ \quad \left. \frac{1}{n_r} \log \mathcal{R} \left(\frac{e}{2 \sigma_c^2 \text{SNR}}, n_r, n_t \right) + \frac{r_{\text{ht}}^{\text{ub},+}}{n_r} \right] \\ \quad + 2 \log \sigma_c, & \text{if } \alpha \geq 1 \\ \alpha \log \frac{\text{SNR}}{\alpha} + \left[\frac{1}{n_r} \log \mathcal{R} \left(\frac{2\alpha}{e \sigma_c^2 \text{SNR}}, n_r, n_t \right) + \frac{r_{\text{ht}}^{\text{lb},-}}{n_r}, \right. \\ \quad \left. \frac{1}{n_r} \log \mathcal{R} \left(\frac{e\alpha}{2 \sigma_c^2 \text{SNR}}, n_t, n_r \right) + \frac{r_{\text{ht}}^{\text{ub},-}}{n_r}, \right] \\ \quad + 2 \log \sigma_c, & \text{if } \alpha < 1 \end{cases} \end{aligned} \quad (40)$$

with probability exceeding $1 - \frac{10}{n^{c(n)}}$.

Here,

$$\mathcal{R}(\epsilon, n, m) := \sum_{i=0}^n \binom{n}{i} \frac{\epsilon^{n-i} m^{-i} m!}{(m-i)!}, \quad (41)$$

and the residual terms are provided in Table III.

Proof: See Appendix E. \square

Remark 6: In the regime where $\beta = \Theta(1)$, one can easily see that the magnitudes of all the residual terms $r_{\text{bd}}^{\text{ub},+}$, $r_{\text{bd}}^{\text{lb},+}$, $r_{\text{bd}}^{\text{ub},-}$, $r_{\text{bd}}^{\text{lb},-}$, $r_{\text{ls}}^{\text{ub},+}$, $r_{\text{ls}}^{\text{lb},+}$, $r_{\text{ls}}^{\text{ub},-}$, $r_{\text{ls}}^{\text{lb},-}$ do not scale with n .

TABLE III
SUMMARY OF PARAMETERS OF THEOREM 2 AND COROLLARY 1

Theorem 2	$r_{\text{bd}}^{\text{lb},+} := -\frac{D\sqrt{\text{SNR}}}{\sqrt{\alpha}} - \frac{\log\left\{1 + \sqrt{\frac{8\pi D^2 \text{SNR}}{\alpha}} e^{8\pi + \frac{2D^2 \text{SNR}}{\alpha}}\right\}}{\beta}$	$r_{\text{bd}}^{\text{ub},+} := \frac{D\sqrt{\text{SNR}}}{\sqrt{\frac{\alpha}{2}}}$
	$r_{\text{bd}}^{\text{lb},-} := -D\sqrt{\text{SNR}} - \frac{\log\left\{1 + \sqrt{8\pi D^2 \text{SNR}} e^{8\pi + 2D^2 \text{SNR}}\right\}}{\beta}$	$r_{\text{bd}}^{\text{ub},-} := \frac{D\sqrt{\text{SNR}}}{\sqrt{\frac{\alpha}{2}}}$
	$r_{\text{ls}}^{\text{lb},+} := -\frac{\sqrt{c_{\text{ls}} \text{SNR}}}{\sqrt{\alpha}} - \frac{\log\left(1 + \sqrt{\frac{4\pi c_{\text{ls}} \text{SNR}}{\alpha}} e^{\frac{c_{\text{ls}} \text{SNR}}{4\alpha}}\right)}{\beta}$	$r_{\text{ls}}^{\text{ub},+} := \frac{\sqrt{c_{\text{ls}} \text{SNR}}}{\sqrt{\alpha}}$
	$r_{\text{ls}}^{\text{lb},-} := -\sqrt{c_{\text{ls}}} \sqrt{\text{SNR}} - \frac{\log\left(1 + \sqrt{4\pi c_{\text{ls}} \text{SNR}} e^{\frac{c_{\text{ls}} \text{SNR}}{4}}\right)}{\beta}$	$r_{\text{ls}}^{\text{ub},-} := \sqrt{c_{\text{ls}}} \sqrt{\text{SNR}}$
	$r_{\text{ht}}^{\text{lb},+} := -\frac{\tau_c \sigma_c \sqrt{8c(n)} \log n \sqrt{\text{SNR}}}{\sqrt{\alpha}} - \log\left(1 + \sqrt{\frac{8\pi \tau_c^2 \sigma_c^2 \text{SNR}}{\alpha}} e^{8\pi + \frac{2\tau_c^2 \sigma_c^2 \text{SNR}}{\alpha}}\right)$	$r_{\text{ht}}^{\text{ub},+} := \frac{\tau_c \sigma_c \sqrt{8c(n)} \log n \sqrt{\text{SNR}}}{\sqrt{\frac{\alpha}{2}}}$
	$r_{\text{ht}}^{\text{lb},-} := -\tau_c \sigma_c \sqrt{8c(n)} \log n \sqrt{\text{SNR}} - \log\left(1 + \sqrt{8\pi \tau_c^2 \sigma_c^2 \text{SNR}} e^{4\pi + 2\tau_c^2 \sigma_c^2 \text{SNR}}\right)$	$r_{\text{ht}}^{\text{ub},-} := \frac{\tau_c \sigma_c \sqrt{8c(n)} \log n \sqrt{\text{SNR}}}{\sqrt{\frac{\alpha}{2}}}$
Corollary 1	$\gamma_{\text{bd}}^{\text{ub},+} := \frac{1.5 \log(en_r)}{n_r} + 4\sqrt{\frac{2\alpha}{e\text{SNR}}} \log \sqrt{\frac{e\text{SNR}}{2\alpha}}$	$\gamma_{\text{bd}}^{\text{lb},+} := \frac{\frac{1}{2} \log n_r - \log \frac{n_t + 1}{2\pi}}{n_r}$
	$\gamma_{\text{bd}}^{\text{ub},-} := \frac{1.5\alpha \log(en_t)}{n_r} + 4\alpha^2 \sqrt{\frac{2}{e\text{SNR}}} \log \sqrt{\frac{e\text{SNR}}{2}}$	$\gamma_{\text{bd}}^{\text{lb},-} := \frac{\alpha\left(\frac{1}{2} \log n_t - \log \frac{n_r + 1}{2\pi}\right)}{n_r}$
	$\gamma_{\text{ls}}^{\text{ub},+} := \frac{1.5 \log(en_r)}{n_r} + 4\sqrt{\frac{\alpha}{\text{SNR}}} \log \sqrt{\frac{\text{SNR}}{\alpha}}$	$\gamma_{\text{ls}}^{\text{lb},+} := \frac{\frac{1}{2} \log n_r - \log \frac{n_t + 1}{2\pi}}{n_r}$
	$\gamma_{\text{ls}}^{\text{ub},-} := \frac{1.5\alpha \log(en_t)}{n_r} + 4\frac{\alpha^2}{\sqrt{\text{SNR}}} \log \sqrt{\text{SNR}}$	$\gamma_{\text{ls}}^{\text{lb},-} := \frac{\alpha\left(\frac{1}{2} \log n_t - \log \frac{n_r + 1}{2\pi}\right)}{n_r}$
	$\gamma_{\text{ht}}^{\text{ub},+} := \frac{1.5 \log(en_r)}{n_r} + 4\sqrt{\frac{2\alpha}{e\sigma_c^2 \text{SNR}}} \log \sqrt{\frac{e\sigma_c^2 \text{SNR}}{2\alpha}}$	$\gamma_{\text{ht}}^{\text{lb},+} := \frac{\frac{1}{2} \log n_r - \log \frac{n_t + 1}{2\pi}}{n_r}$
	$\gamma_{\text{ht}}^{\text{ub},-} := \frac{1.5\alpha \log(en_t)}{n_r} + 4\alpha^2 \sqrt{\frac{2}{e\sigma_c^2 \text{SNR}}} \log \sqrt{\frac{e\sigma_c^2 \text{SNR}}{2}}$	$\gamma_{\text{ht}}^{\text{lb},-} := \frac{\alpha\left(\frac{1}{2} \log n_t - \log \frac{n_r + 1}{2\pi}\right)}{n_r}$

The above theorem relies on the expression $\mathcal{R}(\epsilon, m, n)$. In fact, this function is exactly equal to $\mathbb{E}\left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right]$ in a *distribution-free* manner, as revealed by the following lemma.

Lemma 2: Consider any random matrix $\mathbf{M} \in \mathbb{C}^{n \times m}$ such that \mathbf{M}_{ij} 's are independently generated satisfying

$$\mathbb{E}[\mathbf{M}_{ij}] = 0, \quad \text{and} \quad \mathbb{E}\left[|\mathbf{M}_{ij}|^2\right] = 1. \quad (42)$$

Then one has

$$\mathbb{E}\left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right] = \sum_{i=0}^n \binom{n}{i} \frac{\epsilon^{n-i} m^{-i} m!}{(m-i)!} \quad (43)$$

Proof: This lemma improves upon known results under Gaussian random ensembles by generalizing them to a very general class of random ensembles. See Appendix D for the detailed proof. \square

To get a more quantitative assessment of the concentration intervals, we plot the 95% confidence interval of the MIMO mutual information for a few cases in Fig. 1 when the channel matrix \mathbf{H} is an i.i.d. Gaussian random matrix. The expected value of the capacity is adopted from the mean of 3000 Monte Carlo trials. Our theoretical predictions of the deviation bounds are compared against the simulation results consisting of Monte Carlo trials. One can see from the plots that our theoretical predictions are fairly accurate even for small channel dimensions, which corroborates the power of concentration of measure techniques.

At moderate-to-high SNR, the function $\mathcal{R}(\epsilon, n, m)$ admits a simple approximation. This observation leads to a concise and informative characterization of the mutual information of the MIMO channel, as stated in the following corollary.

Corollary 1: Suppose that $\text{SNR} > \frac{2}{\epsilon} \max\{a, 1\} \cdot \max\{\epsilon^2 a^3, 4a\}$, and set $n := \min\{n_t, n_r\}$.

(a) Consider the assumptions of Theorem 2(a). For any $\beta > 8\sqrt{\pi}$,

$C(\mathbf{H}, \text{SNR})$

$$n_r \in \begin{cases} \log \frac{\text{SNR}}{e} + (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1}\right) \\ + \left[\gamma_{\text{bd}}^{\text{lb},+} + \frac{\beta r_{\text{bd}}^{\text{lb},+}}{n_r}, \gamma_{\text{bd}}^{\text{ub},+} + \frac{\beta r_{\text{bd}}^{\text{ub},+}}{n_r} \right], & \text{if } \alpha \geq 1, \\ \alpha \log \frac{\text{SNR}}{\alpha e} + (1 - \alpha) \log \left(\frac{1}{1 - \alpha}\right) \\ + \left[\gamma_{\text{bd}}^{\text{lb},-} + \frac{\beta r_{\text{bd}}^{\text{lb},-}}{n_r}, \gamma_{\text{bd}}^{\text{ub},-} + \frac{\beta r_{\text{bd}}^{\text{ub},-}}{n_r} \right], & \text{if } \alpha < 1, \end{cases} \quad (44)$$

with probability exceeding $1 - 8 \exp\left(-\frac{\beta^2}{8}\right)$.

(b) Consider the assumptions of Theorem 2(b). For any $\beta > 0$,

$C(\mathbf{H}, \text{SNR})$

$$n_r \in \begin{cases} \log \frac{\text{SNR}}{e} + (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1}\right) \\ + \left[\gamma_{\text{ls}}^{\text{lb},+} + \frac{\beta r_{\text{ls}}^{\text{lb},+}}{n_r}, \gamma_{\text{ls}}^{\text{ub},+} + \frac{\beta r_{\text{ls}}^{\text{ub},+}}{n_r} \right], & \text{if } \alpha \geq 1, \\ \alpha \log \frac{\text{SNR}}{\alpha e} + (1 - \alpha) \log \left(\frac{1}{1 - \alpha}\right) \\ + \left[\gamma_{\text{ls}}^{\text{lb},-} + \frac{\beta r_{\text{ls}}^{\text{lb},-}}{n_r}, \gamma_{\text{ls}}^{\text{ub},-} + \frac{\beta r_{\text{ls}}^{\text{ub},-}}{n_r} \right], & \text{if } \alpha < 1, \end{cases} \quad (45)$$

with probability exceeding $1 - 4 \exp(-\beta^2)$.

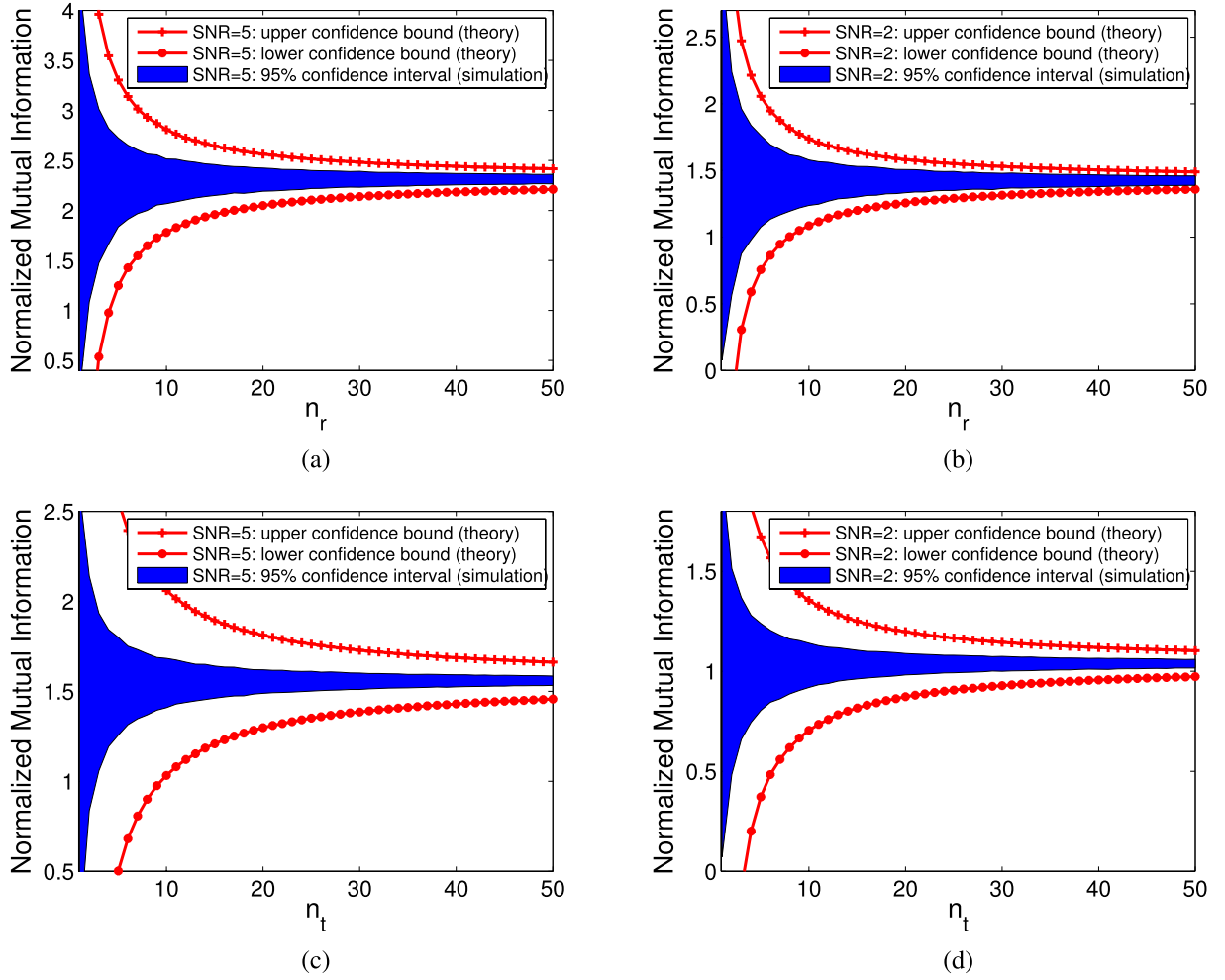


Fig. 1. 95%-confidence interval for MIMO mutual information $\frac{C(\mathbf{H}, \text{SNR})}{n_r}$ when \mathbf{H} is drawn from i.i.d. standard Gaussian ensemble. From Table III, the sizes of the upper and lower confidence spans are given by $\frac{\beta\sqrt{\text{SNR}}}{\sqrt{\alpha n_r}}$ ($\alpha > 1$) and $\frac{\beta\sqrt{\text{SNR}}}{n_r}$ ($\alpha < 1$), respectively, where $\beta = \sqrt{\log \frac{8}{(1-5\%)}}$. The expected value of $\frac{C(\mathbf{H}, \text{SNR})}{n_r}$ is adopted from the mean of 3000 Monte Carlo trials. The theoretical confidence bounds are compared against the simulation results (with 3000 Monte Carlo trials). The results are shown for (a) $\alpha = 2$, SNR = 5; (b) $\alpha = 2$, SNR = 2; (c) $\alpha = 1/2$, SNR = 5; and (d) $\alpha = 1/2$, SNR = 2.

(c) Consider the assumptions of Theorem 2(c). Then, one has

$\gamma_{\text{ls}}^{\text{ub},+}$, and $\gamma_{\text{ls}}^{\text{lb},+}$ are no worse than the order

$$\mathcal{O}\left(\frac{\log n}{n} + \frac{\log \text{SNR}}{\sqrt{\text{SNR}}}\right), \quad (47)$$

which vanish as SNR and channel dimensions grow.

In fact, Corollary 1 is established based on the simple approximation of $\mathcal{R}(\epsilon, n, m) := \mathbb{E}[\det(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*)]$. At moderate-to-high SNR, this function can be approximated reasonably well through much simpler expressions, as stated in the following lemma.

Lemma 3: Consider a random matrix $\mathbf{M} = (\mathbf{M}_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ such that $\alpha := m/n \geq 1$. Suppose that \mathbf{M}_{ij} 's are independent satisfying $\mathbb{E}[\mathbf{M}_{ij}] = 0$ and $\mathbb{E}[|\mathbf{M}_{ij}|^2] = 1$. If $\frac{4}{n} < \epsilon < \min\{\frac{1}{e^2 \alpha^3}, \frac{1}{4\alpha}\}$, then one has

$$\begin{aligned} & \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\ &= (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right) - 1 + r_E, \end{aligned} \quad (48)$$

$C(\mathbf{H}, \text{SNR})$

n_r

$$\in \begin{cases} \log \frac{\text{SNR}}{e} + (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right) + \log \sigma_c + \\ \left[\gamma_{\text{ht}}^{\text{lb},+} + \frac{r_{\text{ht}}^{\text{lb},+}}{n_r}, \gamma_{\text{ht}}^{\text{ub},+} + \frac{r_{\text{ht}}^{\text{ub},+}}{n_r} \right], & \text{if } \alpha \geq 1, \\ \alpha \log \frac{\text{SNR}}{\alpha e} + (1 - \alpha) \log \left(\frac{1}{1 - \alpha} \right) + \log \sigma_c + \\ \left[\gamma_{\text{ht}}^{\text{lb},-} + \frac{r_{\text{ht}}^{\text{lb},-}}{n_r}, \gamma_{\text{ht}}^{\text{ub},-} + \frac{r_{\text{ht}}^{\text{ub},-}}{n_r} \right], & \text{if } \alpha < 1, \end{cases} \quad (46)$$

with probability at least $1 - 10n^{-c(n)}$.

Here, the residual terms are formally provided in Table III.

Proof: See Appendix F. \square

Remark 7: One can see that the magnitudes of all these extra residual terms $\gamma_{\text{bd}}^{\text{ub},-}$, $\gamma_{\text{bd}}^{\text{lb},-}$, $\gamma_{\text{bd}}^{\text{ub},+}$, $\gamma_{\text{bd}}^{\text{lb},+}$, $\gamma_{\text{ls}}^{\text{ub},-}$, $\gamma_{\text{ls}}^{\text{lb},-}$,

where the residual term r_E satisfies

$$r_E \in \left[\frac{\frac{1}{2} \log n - \log \frac{1}{2\pi} (m+1)}{n}, \frac{1.5 \log(en)}{n} + 2\sqrt{\alpha\epsilon} \log \frac{1}{\alpha\epsilon} \right].$$

Proof: Appendix F. \square

Combining Theorem 2 and Lemma 3 immediately establishes Corollary 1.

1) *Implications of Theorem 2 and Corollary 1: Some implications of Corollary 1 are listed as follows.*

- 1) When we set $\beta = \Theta(\sqrt{\log n})$, Corollary 1 implies that in the presence of moderate-to-high SNR and moderate-to-large channel dimensionality $n = \min\{n_t, n_r\}$, the information rate per receive antenna behaves like

$$\frac{C(\mathbf{H}, \text{SNR})}{n_r} = \begin{cases} \log \frac{\text{SNR}}{e} + (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right) \\ + \mathcal{O} \left(\frac{\text{SNR} + \sqrt{\log n}}{n} + \frac{\log \text{SNR}}{\sqrt{\text{SNR}}} \right), & \text{if } \alpha \geq 1 \\ \alpha \log \frac{\text{SNR}}{ae} + (1 - \alpha) \log \left(\frac{1}{1 - \alpha} \right) \\ + \mathcal{O} \left(\frac{\text{SNR} + \sqrt{\log n}}{n} + \frac{\log \text{SNR}}{\sqrt{\text{SNR}}} \right), & \text{if } \alpha < 1 \end{cases} \quad (49)$$

with high probability. That said, for a fairly large regime, the mutual information falls within a narrow range. In fact, the mutual information depends almost only on SNR and the ratio $\alpha = \frac{n_t}{n_r}$, and is independent of the precise number of antennas n_r and n_t , except for some vanishingly small residual terms. The first-order term of the expression (49) coincides with existing *asymptotic* results (see [11]). This in turn validates our concentration of measure approach.

- 2) Theorem 2 indicates that the size of the typical confidence interval for $\frac{C(\mathbf{H}, \text{SNR})}{n_r}$ decays with n at a rate not exceeding $\mathcal{O}(\frac{1}{n})$ (by picking $\beta = \Theta(1)$) under measures of bounded support or measures satisfying the LSI. Note that it has been established (see [35, Th. 2 and Proposition 2]) that the asymptotic standard deviation of $\frac{C(\mathbf{H}, \text{SNR})}{n_r}$ scales as $\Theta(\frac{1}{n})$. This reveals that the concentration of measure approach we adopt is able to obtain the confidence interval with optimal size. Recent works by Li, McKay and Chen [74], [75] derive polynomial expansion for each of the moments (e.g. mean, standard deviation, and skewness) of MIMO mutual information, which can also be used to approximate the distributions. In contrast, our results are able to characterize any concentration interval in a simpler and more informative manner.
- 3) Define the *power offset* for any channel dimension as follows

$$\mathcal{L}(\mathbf{H}, \text{SNR}) := \log \text{SNR} - \frac{C(\mathbf{H}, \text{SNR})}{\min\{n_r, n_t\}}. \quad (50)$$

One can see that this converges to the notion of *high-SNR power offset* investigated by Lozano et. al. [5], [12] in the limiting regime (i.e. when $\text{SNR} \rightarrow \infty$). Our

results reveal the fluctuation of $\mathcal{L}(\mathbf{H}, \text{SNR})$ such that

$$\mathcal{L}(\mathbf{H}, \text{SNR}) = \begin{cases} (\alpha - 1) \log \left(\frac{\alpha - 1}{\alpha} \right) + 1 \\ + \mathcal{O} \left(\frac{\text{SNR} + \sqrt{\log n}}{n} + \frac{\log \text{SNR}}{\sqrt{\text{SNR}}} \right), & \text{if } \alpha \geq 1 \\ \frac{1 - \alpha}{\alpha} \log(1 - \alpha) + \log(e\alpha) \\ + \mathcal{O} \left(\frac{\text{SNR} + \sqrt{\log n}}{n} + \frac{\log \text{SNR}}{\sqrt{\text{SNR}}} \right), & \text{if } \alpha < 1 \end{cases} \quad (51)$$

with high probability. The first-order term of $\mathcal{L}(\mathbf{H}, \text{SNR})$ agrees with the known results on asymptotic power offset (see [5, Proposition 2] [12, Equation (84)]) when $n \rightarrow \infty$ and $\text{SNR} \rightarrow \infty$. Our results distinguish from existing results in the sense that we can accurately accommodate a much larger regime beyond the one with asymptotically large SNR and channel dimensions.

- 4) The same mutual information and power offset values are shared among a fairly general class of distributions even in the non-asymptotic regime. Moreover, heavy-tailed distributions lead to well-controlled information rate and power-offset as well, although the spectral measure concentration is often less sharp than for sub-Gaussian distributions (which ensure exponentially sharp concentration).
- 5) Finally, we note that Corollary 1 not only characterizes the order of the residual term, but also provides reasonably accurate characterization of a narrow confidence interval such that the mutual information and the power offset lies within it, with high probability. All pre-constants are explicitly provided, resulting in a full assessment of the mutual information and the power offset. Our results do not rely on careful choice of growing matrix sequences, which are typically required in those works based on asymptotic laws.

2) *Connection to the General Framework:* We now illustrate how the proof of Theorem 2 follows from the general framework presented in Section III.

- 1) Note that the mutual information can be alternatively expressed as

$$\begin{aligned} \frac{1}{n_r} C(\mathbf{H}, \text{SNR}) &= \frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} \log \left(1 + \text{SNR} \cdot \lambda_i \left(\frac{1}{n_t} \mathbf{H} \mathbf{H}^* \right) \right) \\ &:= \frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} \left\{ \log \text{SNR} + f \left(\lambda_i \left(\frac{1}{n_t} \mathbf{H} \mathbf{H}^* \right) \right) \right\}, \end{aligned}$$

where $f(x) := \log \left(\frac{1}{\text{SNR}} + x \right)$ on the domain $x \geq 0$. As a result, the function $g(x) := f(x^2) = \log \left(\frac{1}{\text{SNR}} + x^2 \right)$ has Lipschitz norm $\|g\|_{\mathcal{L}} \leq \text{SNR}^{1/2}$.

- 2) For measures satisfying the LSI, Theorem 2(b) immediately follows from Theorem 1(b). For measures of bounded support or heavy-tailed measures, one can introduce the following auxiliary function with respect

to some small ϵ :

$$g_\epsilon(x) := \begin{cases} \log(\epsilon + x^2), & \text{if } x \geq \sqrt{\epsilon}, \\ \frac{1}{\sqrt{\epsilon}}(x - \sqrt{\epsilon}) + \log(2\epsilon), & \text{if } 0 \leq x < \sqrt{\epsilon}, \end{cases}$$

which is obtained by convexifying $g(x)$. If we set $f_\epsilon(x) := g_\epsilon(\sqrt{x})$, then one can easily check that

$$f_{\text{SNR}^{-1}}(x) \leq f(x) \leq f_{\frac{\epsilon}{2}\text{SNR}^{-1}}(x).$$

Applying Theorem 1(a) allows us to derive the concentration on

$$\frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} f_{\text{SNR}^{-1}}\left(\lambda_i\left(\frac{1}{n_t}\mathbf{H}\mathbf{H}^*\right)\right)$$

and

$$\frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} f_{\frac{\epsilon}{2}\text{SNR}^{-1}}\left(\lambda_i\left(\frac{1}{n_t}\mathbf{H}\mathbf{H}^*\right)\right),$$

which in turn provide tight lower and upper bounds for $\frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} f\left(\lambda_i\left(\frac{1}{n_t}\mathbf{H}\mathbf{H}^*\right)\right)$.

- 3) Finally, the mean value $\mathbb{E}\left[\det\left(\mathbf{I} + \text{SNR} \cdot \frac{1}{n_t}\mathbf{H}\mathbf{H}^*\right)\right]$ admits a closed-form expression independent of precise entry distributions, as derived in Lemma 2.

B. MMSE Estimation in Random Vector Channels

Next, we show that our analysis framework can be applied in Bayesian inference problems that involve estimation of uncorrelated signals components. Consider a simple linear vector channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z},$$

where $\mathbf{H} \in \mathbb{C}^{n_r \times p}$ is a known matrix, $\mathbf{x} \in \mathbb{C}^p$ denotes an input signal with zero mean and covariance matrix $P\mathbf{I}_p$, and \mathbf{z} represents a zero-mean noise uncorrelated with \mathbf{x} , which has covariance $\sigma^2\mathbf{I}_n$. Here, we denote by p and n the input dimension and the sample size, respectively, which agrees with the conventional notation adopted in the statistics community. In this subsection, we assume that

$$\alpha := \frac{p}{n} \leq 1, \quad (52)$$

i.e. the sample size exceeds the dimension of the input signal. The goal is to estimate the channel input \mathbf{x} from the channel output \mathbf{y} with minimum ℓ_2 distortion.

The MMSE estimate of \mathbf{x} given \mathbf{y} can be written as [4]

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbb{E}[\mathbf{x} | \mathbf{y}] = \mathbb{E}[\mathbf{x}\mathbf{y}^*] (\mathbb{E}[\mathbf{y}\mathbf{y}^*])^{-1} \mathbf{y} \\ &= P\mathbf{H}^* (\sigma^2\mathbf{I}_n + P\mathbf{H}\mathbf{H}^*)^{-1} \mathbf{y}, \end{aligned} \quad (53)$$

and the resulting MMSE is given by

$$\begin{aligned} \text{MMSE}(\mathbf{H}, \text{SNR}) \\ = \text{tr}\left(P\mathbf{I}_p - P^2\mathbf{H}^* (\sigma^2\mathbf{I}_n + P\mathbf{H}\mathbf{H}^*)^{-1} \mathbf{H}\right). \end{aligned} \quad (54)$$

The normalized MMSE (NMMSE) can then be written as

$$\begin{aligned} \text{NMMSE}(\mathbf{H}, \text{SNR}) &:= \frac{\text{MMSE}(\mathbf{H}, \text{SNR})}{\mathbb{E}[\|\mathbf{x}\|^2]} \\ &= \text{tr}\left(\mathbf{I}_p - \mathbf{H}^* \left(\frac{1}{\text{SNR}}\mathbf{I}_n + \mathbf{H}\mathbf{H}^*\right)^{-1} \mathbf{H}\right), \end{aligned} \quad (55)$$

where

$$\text{SNR} := \frac{P}{\sigma^2}.$$

Using the concentration of measure technique, we can evaluate NMMSE(\mathbf{H}) to within a narrow interval with high probability, as stated in the following theorem.

Theorem 3: Suppose that $\mathbf{H} = \mathbf{A}\mathbf{M}$, where $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a deterministic matrix for some integer $m \geq p$, and $\mathbf{M} \in \mathbb{C}^{n \times p}$ is a random matrix such that \mathbf{M}_{ij} 's are independent random variables satisfying $\mathbb{E}[\mathbf{M}_{ij}] = 0$ and $\mathbb{E}[|\mathbf{M}_{ij}|^2] = \frac{1}{p}$.

(a) If $\sqrt{p}\mathbf{M}_{ij}$'s are bounded by D , then for any $\beta > 8\sqrt{\pi}$,

$$\begin{aligned} \frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} \in \left[\mathcal{M}\left(\frac{8}{9\text{SNR}}, \mathbf{H}\right) + \frac{\beta\tau_{\text{bd}}^{\text{lb}}}{p}, \right. \\ \left. \mathcal{M}\left(\frac{9}{8\text{SNR}}, \mathbf{H}\right) + \frac{\beta\tau_{\text{bd}}^{\text{ub}}}{p} \right] \end{aligned} \quad (56)$$

with probability exceeding $1 - 8\exp\left(-\frac{\beta^2}{16}\right)$.

(b) If $\sqrt{p}\mathbf{M}_{ij}$'s satisfy the LSI with respect to a uniform constant c_{ls} , then for any $\beta > 0$,

$$\frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} \in \mathcal{M}\left(\frac{1}{\text{SNR}}, \mathbf{H}\right) + \frac{\beta}{p} [\tau_{\text{ls}}^{\text{lb}}, \tau_{\text{ls}}^{\text{ub}}] \quad (57)$$

with probability exceeding $1 - 4\exp\left(-\frac{\beta^2}{2}\right)$.

(c) Suppose that $\sqrt{p}\mathbf{M}_{ij}$'s are independently drawn from either sub-exponential distributions or heavy-tailed distributions and that the distributions are symmetric about 0. Let τ_c be defined as in (7) with respect to $\sqrt{p}\mathbf{M}_{ij}$'s for some sequence $c(n)$. Then

$$\begin{aligned} \frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} \in \left[\mathcal{M}\left(\frac{8}{9\text{SNR}}, \mathbf{A}\tilde{\mathbf{M}}\right) + \frac{\tau_{\text{ht}}^{\text{lb}}}{p}, \right. \\ \left. \mathcal{M}\left(\frac{9}{8\text{SNR}}, \mathbf{A}\tilde{\mathbf{M}}\right) + \frac{\tau_{\text{ht}}^{\text{ub}}}{p} \right] \end{aligned} \quad (58)$$

with probability exceeding $1 - \frac{10}{p^{c(n)}}$, where $\tilde{\mathbf{M}}$ is defined such that $\tilde{\mathbf{M}}_{ij} := \mathbf{M}_{ij} \mathbf{1}_{\{\sqrt{p}|\mathbf{M}_{ij}| < \tau_c\}}$.

Here, the function $\mathcal{M}(\epsilon, \mathbf{H})$ is defined as

$$\mathcal{M}(\delta, \mathbf{H}) := \frac{1}{p} \mathbb{E}\left[\text{tr}\left((\delta\mathbf{I} + \mathbf{H}^*\mathbf{H})^{-1}\right)\right], \quad (59)$$

and the residual terms are formally defined in Table IV.

Proof: See Appendix G. \square

Theorem 3 ensures that the MMSE per signal component falls within a small interval with high probability. This concentration phenomenon again arises for a very large class of probability measures including bounded distributions,

TABLE IV
SUMMARY OF PARAMETERS OF THEOREM 3 AND COROLLARY 2

Theorem 3	$\tau_{\text{bd}}^{\text{lb}} := -\frac{2\sqrt{2}D\ \mathbf{A}\ \text{SNR}^{1.5}}{3\sqrt{3}}$	$\tau_{\text{bd}}^{\text{ub}} := \frac{3\sqrt{3}D\ \mathbf{A}\ \text{SNR}^{1.5}}{8}$
	$\tau_{\text{ls}}^{\text{lb}} := -\frac{3\sqrt{3}\sqrt{c_{\text{ls}}}\ \mathbf{A}\ \text{SNR}^{1.5}}{8}$	$\tau_{\text{ls}}^{\text{ub}} := -\frac{3\sqrt{3}\sqrt{c_{\text{ls}}}\ \mathbf{A}\ \text{SNR}^{1.5}}{8}$
	$\tau_{\text{ht}}^{\text{lb}} := -\frac{8\sqrt{2}\tau_c\sigma_c\ \mathbf{A}\ \sqrt{c(n)}\log n\text{SNR}^{1.5}}{3\sqrt{3}}$	$\tau_{\text{ht}}^{\text{ub}} := \frac{3\sqrt{3}\tau_c\sigma_c\ \mathbf{A}\ \sqrt{c(n)}\log n\text{SNR}^{1.5}}{2}$
Corollary 2	$\tau_{\text{g}}^{\text{ub}} := \frac{\alpha}{n(1-\alpha-\frac{1}{n})^2}$	$\tau_{\text{g}}^{\text{lb}} := -\frac{\alpha^2}{\text{SNR}(1-\alpha-\frac{3}{n})^3}$

sub-Gaussian distributions, and heavy-tailed distributions. Note that $\mathcal{M}(\frac{9}{8\text{SNR}}, \mathbf{H})$, $\mathcal{M}(\frac{8}{9\text{SNR}}, \mathbf{H})$, and $\mathcal{M}(\frac{1}{\text{SNR}}, \mathbf{H})$ are often very close to each other at moderate-to-high SNR (see the analysis of Corollary 2). The spans of the residual intervals for both bounded and logarithmic Sobolev measures do not exceed the order

$$\mathcal{O}\left(\frac{\beta\|\mathbf{A}\|\text{SNR}^{1.5}}{p}\right), \quad (60)$$

which is often negligible compared with the MMSE value per signal component.

In general, we are not aware of a closed-form expression for $\mathcal{M}(\delta, \mathbf{H})$ under various random matrix ensembles. If \mathbf{H} is drawn from a Gaussian ensemble, however, we are able to derive a simple expression and bound this value to within a narrow confidence interval with high probability, as follows.

Corollary 2: Suppose that $\mathbf{H}_{ij} \sim \mathcal{N}(0, \frac{1}{p})$ are independent complex-valued Gaussian random variables, and assume that $\alpha < 1$. Then

$$\frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} \in \frac{\alpha}{1-\alpha} + \left[\tau_{\text{g}}^{\text{lb}} + \frac{\beta\tau_{\text{ls}}^{\text{lb}}}{p}, \tau_{\text{g}}^{\text{ub}} + \frac{\beta\tau_{\text{ls}}^{\text{ub}}}{p} \right] \quad (61)$$

with probability exceeding $1 - 4\exp(-\frac{\beta^2}{2})$. Here, $\tau_{\text{g}}^{\text{lb}}$ and $\tau_{\text{g}}^{\text{ub}}$ are presented in Theorem 3, while $\tau_{\text{g}}^{\text{lb}}$ and $\tau_{\text{g}}^{\text{ub}}$ are formally defined in Table IV.

Proof: See Appendix G. \square

Corollary 2 implies that the MMSE per signal component behaves like

$$\frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} = \frac{\alpha}{1-\alpha} + \mathcal{O}\left(\frac{\text{SNR}^{1.5}}{n} + \frac{1}{n} + \frac{1}{\text{SNR}}\right)$$

with high probability (by setting $\beta = \Theta(1)$). Except for the residual term that vanishes in the presence of high SNR and large signal dimensionality n , the first order term depends only on the SNR and the oversampling ratio $\frac{1}{\alpha}$. Therefore, the normalized MMSE converges to an almost deterministic value even in the non-asymptotic regime. This illustrates the effectiveness of the proposed analysis approach.

1) *Connection to the General Framework:* We now demonstrate how the proof of Theorem 3 follows from the general template presented in Section III.

1) Observe that the MMSE can be alternatively expressed as

$$\begin{aligned} \frac{\text{NMMSE}(\mathbf{H}, \text{SNR})}{p} &= \frac{1}{p} \sum_{i=1}^{\min\{n,p\}} \frac{1}{\text{SNR}^{-1} + \lambda_i(\mathbf{H}\mathbf{H}^*)} \\ &:= \frac{1}{p} \sum_{i=1}^{\min\{n,p\}} f(\lambda_i(\mathbf{H}\mathbf{H}^*)), \end{aligned}$$

where $f(x) := \frac{1}{\text{SNR}^{-1} + x}$ ($x \geq 0$). Consequently, the function $g(x) := f(x^2) = \frac{1}{\text{SNR}^{-1} + x^2}$ has Lipschitz norm $\|g\|_{\mathcal{L}} \leq \frac{3\sqrt{3}}{8}\text{SNR}^{3/2}$.

2) For measures satisfying the LSI, Theorem 3(b) can be immediately obtained by Proposition 1(b). For measures of bounded support or heavy-tailed measures, one can introduce the following auxiliary function with respect to some small ϵ :

$$\tilde{g}_{\epsilon}(x) := \begin{cases} \frac{1}{\epsilon^2 + x^2}, & \text{if } x > \frac{1}{\sqrt{3}}\epsilon, \\ -\frac{3\sqrt{3}}{8}\left(x - \frac{\epsilon}{\sqrt{3}}\right) + \frac{3}{4\epsilon^2}, & \text{if } x \leq \frac{1}{\sqrt{3}}\epsilon, \end{cases}$$

which is obtained by convexifying $g(x)$. By setting $\tilde{f}_{\epsilon}(x) := \tilde{g}_{\epsilon}(\sqrt{x})$, one has

$$\tilde{f}_{\frac{3}{2\sqrt{2}}}\text{SNR}^{-1/2}(x) \leq f(x) \leq \tilde{f}_{\text{SNR}^{-1/2}}(x).$$

Applying Proposition 1(a) gives the concentration bounds on

$$\frac{1}{p} \sum_{i=1}^{\min\{p,n\}} \tilde{f}_{\frac{3}{2\sqrt{2}}}\text{SNR}^{-1/2}(\lambda_i(\mathbf{H}\mathbf{H}^*))$$

and

$$\frac{1}{p} \sum_{i=1}^{\min\{p,n\}} \tilde{f}_{\text{SNR}^{-1/2}}(\lambda_i(\mathbf{H}\mathbf{H}^*)),$$

which in turn provide tight sandwich bounds for $\frac{1}{p} \sum_{i=1}^{\min\{p,n\}} f(\lambda_i(\mathbf{H}\mathbf{H}^*))$.

V. CONCLUSIONS AND FUTURE DIRECTIONS

We have presented a general analysis framework that allows many canonical MIMO system performance metrics taking the form of linear spectral statistics to be assessed within a narrow concentration interval. Moreover, we can guarantee that the metric value falls within the derived interval with high probability, even under moderate channel dimensionality. We demonstrate the effectiveness of the proposed framework through two canonical metrics in wireless communications

and signal processing: mutual information and power offset of MIMO channels, and MMSE for Gaussian processes. For each of these examples, our approach allows a more refined and accurate characterization than those derived in previous work in the asymptotic dimensionality limit. While our examples here are presented for i.i.d. channel matrices, they can all be immediately extended to accommodate dependent random matrices using the same techniques. In other words, our analysis can be applied to characterize various performance metrics of wireless systems under correlated fading, as studied in [8].

Our work suggests a few future research directions for different matrix ensembles. Specifically, if the channel matrix \mathbf{H} cannot be expressed as a linear correlation model $\mathbf{M}\mathbf{A}$ for some known matrix \mathbf{A} and a random i.i.d. matrix \mathbf{M} , then it will be of great interest to see whether there is a systematic approach to analyze the concentration of spectral measure phenomenon. For instance, if \mathbf{H} consists of independent sub-Gaussian columns but the elements within each column are correlated (i.e. cannot be expressed as a weighted sum of i.i.d. random variables), then it remains to be seen whether generalized concentration of measure techniques (see [52]) can be applied to find confidence intervals for the system performance metrics. The spiked random matrix model [76], which also admits simple asymptotic characterization, is another popular candidate in modeling various system metrics. Another more challenging case is when the rows of \mathbf{H} are randomly drawn from a known set of structured candidates. An example is a random DFT ensemble where each row is independently drawn from a DFT matrix, which is frequently encountered in the compressed sensing literature (see [77]). Understanding the spectral concentration of such random matrices could be of great interest in investigating OFDM systems under random subsampling.

APPENDIX A PROOF OF PROPOSITION 1

Part (a) and (b) of Proposition 1 are immediate consequences of [51, Corollary 1.8] via simple algebraic manipulation. Here, we only provide the proof for heavy-tailed distributions.

Define a new matrix $\tilde{\mathbf{M}}$ as a truncated version of \mathbf{M} such that

$$\tilde{M}_{ij} = \begin{cases} M_{ij}, & \text{if } |M_{ij}| < \tau_c, \\ 0, & \text{otherwise.} \end{cases}$$

Note that the entries of $\tilde{\mathbf{M}}$ have zero mean and variance not exceeding σ_c^2 , and are uniformly bounded in magnitude by τ_c . Consequently, Theorem 1(a) asserts that for any $\beta > 8\sqrt{\pi}$, we have

$$\left| f_0(\tilde{\mathbf{M}}) - \mathbb{E} \left[f_0(\tilde{\mathbf{M}}) \right] \right| \leq \frac{\beta \tau_c \sigma_c \rho \|g\|_{\mathcal{L}}}{n} \quad (62)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{16}\right)$.

In addition, a simple union bound taken collectively with (7) yields

$$\begin{aligned} \mathbb{P}(\mathbf{M} \neq \tilde{\mathbf{M}}) &\leq \sum_{1 \leq i \leq n, 1 \leq j \leq m} \mathbb{P}\left(\frac{1}{v_{ij}} |M_{ij}| > \tau_c\right) \\ &\leq mn \mathbb{P}\left(\frac{1}{v_{ij}} |M_{ij}| > \tau_c\right) = \frac{1}{n^{c(n)}}. \end{aligned} \quad (63)$$

By setting $\beta = 4\sqrt{c(n) \log n}$, one has

$$4 \exp\left(-\frac{\beta^2}{16}\right) = \frac{4}{n^{c(n)}}.$$

This combined with (62) and the union bound implies that

$$\left| f_0(\mathbf{M}) - \mathbb{E} \left[f_0(\tilde{\mathbf{M}}) \right] \right| \leq \frac{4\sqrt{c(n) \log n} \tau_c \sigma_c \rho \|g\|_{\mathcal{L}}}{n} \quad (64)$$

holds with probability at least $1 - 5n^{-c(n)}$, as claimed.

APPENDIX B PROOF OF LEMMA 1

For notational simplicity, define

$$\begin{aligned} Y &:= n \{ f_0(\mathbf{M}) - \mathbb{E} [f_0(\mathbf{M})] \} \\ &= \sum_{i=1}^{\min\{m,n\}} \left\{ f\left(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right) \right. \\ &\quad \left. - \mathbb{E} \left[f\left(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right) \right] \right\} \end{aligned}$$

and

$$Z := n f_0(\mathbf{M}) = \sum_{i=1}^{\min\{m,n\}} f\left(\lambda_i \left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right),$$

and denote by $\mu_{|Y|}(y)$ the probability measure of $|Y|$.

- (1) Suppose that $\mathbb{P}(|Y| > y) \leq c_1 \exp(-c_2 y)$ holds for some universal constants $c_1 > 0$ and $c_2 > 1$. Applying integration by parts yields the following inequality

$$\begin{aligned} \mathbb{E} \left[e^Y \right] &\leq \mathbb{E} \left[e^{|Y|} \right] = \int_0^\infty e^{|Y|} d\mu_{|Y|}(y) \\ &= -e^y \mathbb{P}(|Y| > y) \Big|_0^\infty + \int_0^\infty e^y \mathbb{P}(|Y| > y) dy \\ &\leq 1 + c_1 \int_0^\infty \exp(-(c_2 - 1)y) dy \\ &= 1 + \frac{c_1}{c_2 - 1}. \end{aligned} \quad (65)$$

This gives rise to the following bound

$$\log \mathbb{E} \left[e^Z \right] - \log e^{\mathbb{E}[Z]} = \log \mathbb{E} \left[e^Y \right] \leq \log \left(1 + \frac{c_1}{c_2 - 1} \right).$$

Therefore, we can conclude

$$\frac{1}{n} \log \mathbb{E} \left[e^Z \right] - \frac{\log \left(1 + \frac{c_1}{c_2 - 1} \right)}{n} \leq \frac{1}{n} \mathbb{E} [Z] \leq \frac{1}{n} \log \mathbb{E} \left[e^Z \right], \quad (66)$$

where the last inequality follows from Jensen's inequality

$$\frac{1}{n} \mathbb{E} [Z] = \frac{1}{n} \log e^{\mathbb{E}[Z]} \leq \frac{1}{n} \log \mathbb{E} e^{[Z]}.$$

- (2) Similarly, if $\mathbb{P}(|Y| > y) \leq c_1 \exp(-c_2 y^2)$ holds for some universal constants $c_1 > 0$ and $c_2 > 0$, then one can bound

$$\begin{aligned} \mathbb{E}\left[e^Y\right] &\leq -e^y \mathbb{P}(|Y| > y)|_0^\infty + \int_0^\infty e^y \mathbb{P}(|Y| > y) dy \\ &\leq 1 + c_1 \int_0^\infty \exp(-c_2 y^2 + y) dy \\ &\leq 1 + c_1 \sqrt{\frac{\pi}{c_2}} \exp\left(\frac{1}{4c_2}\right) \\ &\quad \cdot \int_{-\infty}^\infty \sqrt{\frac{c_2}{\pi}} \exp\left(-c_2 y^2 + y - \frac{1}{4c_2}\right) dy \\ &\leq 1 + c_1 \sqrt{\frac{\pi}{c_2}} \exp\left(\frac{1}{4c_2}\right), \end{aligned} \quad (67)$$

where the last inequality follows since $\sqrt{\frac{c_2}{\pi}} \exp\left(-c_2 y^2 + y - \frac{1}{4c_2}\right)$ is the pdf of $\mathcal{N}\left(\frac{1}{2c_2}, \frac{2}{c_2}\right)$ and hence integrates to 1.

Applying the same argument as for the measure with sub-exponential tails then leads to

$$\begin{aligned} \frac{1}{n} \log \mathbb{E}\left[e^Z\right] &\geq \frac{1}{n} \mathbb{E}[Z] \\ &\geq \frac{1}{n} \log \mathbb{E}\left[e^Z\right] - \frac{\log\left(1 + \sqrt{\frac{c_1^2 \pi}{c_2}} \exp\left(\frac{1}{4c_2}\right)\right)}{n}. \end{aligned}$$

APPENDIX C PROOF OF THEOREM 1

For notational convenience, define

$$\begin{aligned} Z &:= \sum_{i=1}^{\min\{m,n\}} f\left(\lambda_i\left(\frac{1}{n} \mathbf{M} \mathbf{R} \mathbf{R}^* \mathbf{M}^*\right)\right), \\ Y &:= Z - \mathbb{E}[Z]. \end{aligned}$$

- (1) Under the assumptions of Proposition 1(a), the concentration result (15) implies that for any $y > 8\sqrt{\pi} \rho D \|g\|_{\mathcal{L}}$, one has

$$\mathbb{P}(|Y| > y) \leq 4 \exp\left(-\frac{y^2}{8\kappa D^2 \rho^2 v^2 \|g\|_{\mathcal{L}}^2}\right).$$

For $y \leq 8\sqrt{\pi} D \rho v \|g\|_{\mathcal{L}}$, we can employ the trivial bound $\mathbb{P}(|Y| > y) \leq 1$. These two bounds taken collectively lead to a universal bound such that for any $y \geq 0$, one has

$$\mathbb{P}(|Y| > y) \leq \exp\left(\frac{8\pi}{\kappa}\right) \exp\left(-\frac{y^2}{8\kappa D^2 \rho^2 v^2 \|g\|_{\mathcal{L}}^2}\right).$$

Lemma 1 then suggests

$$\frac{1}{n} \log \mathcal{E}(f) \geq \frac{1}{n} \mathbb{E}[Z] \geq \frac{\log \mathcal{E}(f)}{n} - \frac{\log\left(1 + \sqrt{8\kappa\pi} D \rho v \|g\|_{\mathcal{L}} e^{\frac{8\pi}{\kappa} + 2\kappa D^2 \rho^2 v^2 \|g\|_{\mathcal{L}}^2}\right)}{n}.$$

This together with (15) establishes (25).

- (2) Under the assumptions of Proposition 1(b), the concentration inequality (15) asserts that for any $y > 0$, we have

$$\mathbb{P}(|Y| > y) \leq 2 \exp\left(-\frac{y^2}{\kappa c_{1s} \rho^2 v^2 \|g\|_{\mathcal{L}}^2}\right).$$

Applying Lemma 1 then yields

$$\frac{1}{n} \log \mathcal{E}(f) \geq \frac{1}{n} \mathbb{E}[Z] \geq \frac{\log \mathcal{E}(f)}{n} - \frac{\log\left(1 + \sqrt{4\kappa\pi c_{1s} \rho^2 v^2 \|g\|_{\mathcal{L}}^2} e^{\frac{\kappa}{4} c_{1s} \rho^2 v^2 \|g\|_{\mathcal{L}}^2}\right)}{n}.$$

This combined with (16) establishes (26).

- (3) Under the assumptions of Proposition 1(c), define

$$\begin{aligned} \tilde{Z} &:= \sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{n} \tilde{\mathbf{M}} \mathbf{R} \mathbf{R}^* \tilde{\mathbf{M}}^*\right)\right), \\ \tilde{Y} &:= \tilde{Z} - \mathbb{E}[\tilde{Z}]. \end{aligned}$$

Combining the concentration result (62) with (25) gives rise to

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{n} \tilde{\mathbf{M}} \mathbf{R} \mathbf{R}^* \tilde{\mathbf{M}}^*\right)\right) \\ \leq \frac{1}{n} \log \mathcal{E}_{\tilde{\mathbf{M}}}(f) + \frac{\beta \tau_c \sigma_c \rho \|g\|_{\mathcal{L}}}{n}, \\ \frac{1}{n} \sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{n} \tilde{\mathbf{M}} \mathbf{R} \mathbf{R}^* \tilde{\mathbf{M}}^*\right)\right) \\ \geq \frac{1}{n} \log \mathcal{E}_{\tilde{\mathbf{M}}}(f) - \frac{\beta \tau_c \sigma_c \rho \|g\|_{\mathcal{L}}}{n} - \frac{c_{\rho, f, \tau_c, \sigma_c}}{n}, \end{cases} \quad (68)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{8\kappa}\right)$. We have shown in (63) that

$$\mathbb{P}(\mathbf{M} \neq \tilde{\mathbf{M}}) \leq \frac{1}{n^{c(n)}}$$

and $4 \exp\left(-\frac{\beta^2}{8\kappa}\right) = \frac{4}{n^{c(n)}}$ when $\beta = \sqrt{8\kappa c(n) \log n}$, concluding the proof via the union bound.

APPENDIX D PROOF OF LEMMA 2

For any $n \times n$ matrix \mathbf{A} with eigenvalues $\lambda_1, \dots, \lambda_n$, define the characteristic polynomial of \mathbf{A} as

$$\begin{aligned} p_{\mathbf{A}}(t) &= \det(t\mathbf{I} - \mathbf{A}) = t^n - S_1(\lambda_1, \dots, \lambda_n) t^{n-1} \\ &\quad + \dots + (-1)^n S_n(\lambda_1, \dots, \lambda_n), \end{aligned} \quad (69)$$

where $S_l(\lambda_1, \dots, \lambda_n)$ is the l th elementary symmetric polynomial defined as follows

$$S_l(\lambda_1, \dots, \lambda_n) := \sum_{1 \leq i_1 < \dots < i_l \leq n} \prod_{j=1}^l \lambda_{i_j}. \quad (70)$$

Let $E_l(\mathbf{A})$ represent the sum of determinants of all l -by- l principal minors of \mathbf{A} . It has been shown in [78, Th. 1.2.12] that

$$S_l(\lambda_1, \dots, \lambda_n) = E_l(\mathbf{A}), \quad 1 \leq l \leq n, \quad (71)$$

which follows that

$$\det(t\mathbf{I} + \mathbf{A}) = t^n + E_1(\mathbf{A}) t^{n-1} + \dots + E_n(\mathbf{A}). \quad (72)$$

On the other hand, for any principal minor $(\mathbf{M}\mathbf{M}^*)_{s,s}$ of $\mathbf{M}\mathbf{M}^*$ coming from the rows and columns at indices from s (where $s \subseteq [n]$ is an index set), then one can show that

$$\begin{aligned} \mathbb{E} \left[\det \left((\mathbf{M}\mathbf{M}^*)_{s,s} \right) \right] &= \mathbb{E} \left[\det \left(\mathbf{M}_{s,[m]} \mathbf{M}_{s,[m]}^* \right) \right] \\ &= \sum_{t: \text{card}(t)=\text{card}(s)} \mathbb{E} \left[\det \left(\mathbf{M}_{s,t} \mathbf{M}_{s,t}^* \right) \right] \\ &= \binom{m}{\text{card}(s)} \cdot \mathbb{E} \left[\det \left(\mathbf{M}_{s,s} \mathbf{M}_{s,s}^* \right) \right]. \end{aligned} \quad (73)$$

Consider now any i.i.d. random matrix $\mathbf{G} \in \mathbb{C}^{l \times l}$ such that $\mathbb{E}[\mathbf{G}_{ij}] = 0$ and $\mathbb{E}[|\mathbf{G}_{ij}|^2] = 1$. If we denote by \prod_l the permutation group of l elements, then the Leibniz formula for the determinant gives

$$\det(\mathbf{G}) = \sum_{\sigma \in \prod_l} \text{sgn}(\sigma) \prod_{i=1}^l \mathbf{G}_{i,\sigma(i)}.$$

Since \mathbf{G}_{ij} are jointly independent, we have

$$\begin{aligned} \mathbb{E} \left[\det(\mathbf{G}\mathbf{G}^*) \right] &= \mathbb{E} \left[|\det(\mathbf{G})|^2 \right] \\ &= \sum_{\sigma \in \prod_l} \mathbb{E} \left[\prod_{i=1}^l |\mathbf{G}_{i,\sigma(i)}|^2 \right] \\ &= \sum_{\sigma \in \prod_l} \prod_{i=1}^l \mathbb{E} \left[|\mathbf{G}_{i,\sigma(i)}|^2 \right] = l!, \end{aligned} \quad (74)$$

which is distribution-free.

So far, we are already able to derive the closed-form expression for $\mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \right]$ by combining (72), (73) and (74) via straightforward algebraic manipulation.

Alternatively, (72), (73) and (74) taken collectively indicate that $\mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \right]$ is independent of precise distribution of the entries \mathbf{M}_{ij} 's. As a result, we only need to evaluate $\mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \right]$ under i.i.d. Gaussian random matrices. To this end, one can simply cite the closed-form expression derived for Gaussian random matrices, which has been reported in [73, Th. 2.13]. This concludes the proof.

APPENDIX E PROOF OF THEOREM 2

When equal power allocation is adopted, the MIMO mutual information per receive antenna is given by

$$\begin{aligned} \frac{C(\mathbf{H}, \text{SNR})}{n_r} &= \frac{1}{n_r} \log \det \left(\mathbf{I}_{n_r} + \frac{\text{SNR}}{n_t} \mathbf{H}\mathbf{H}^* \right) \\ &= \frac{1}{n_r} \sum_{i=1}^{\min\{n_r, n_t\}} \log \left(\frac{1}{\text{SNR}} + \frac{1}{n_t} \lambda_i(\mathbf{H}\mathbf{H}^*) \right) \\ &\quad + \min\{1, \alpha\} \log \text{SNR}, \\ &= \begin{cases} \frac{1}{n_r} \log \det \left(\frac{1}{\text{SNR}} \mathbf{I}_{n_r} + \frac{1}{n_t} \mathbf{H}\mathbf{H}^* \right) + \log \text{SNR}, & \text{if } \alpha \geq 1 \\ \frac{\alpha}{n_t} \log \det \left(\frac{\alpha}{\text{SNR}} \mathbf{I}_{n_t} + \frac{1}{n_r} \mathbf{H}^* \mathbf{H} \right) + \alpha \log \frac{\text{SNR}}{\alpha}, & \text{if } \alpha < 1 \end{cases} \end{aligned} \quad (75)$$

where $\alpha = n_t/n_r$ is assumed to be an absolute constant.

The first term of the capacity expression (75) in both cases exhibits sharp measure concentration, as stated in the following lemma. This in turn completes the proof of Theorem 2.

Lemma 4: Suppose that $a := \frac{m}{n} \geq 1$ is an absolute constant. Consider a real-valued random matrix $\mathbf{M} = [\zeta_{ij}]_{1 \leq i \leq n, 1 \leq j \leq m}$, where ζ_{ij} 's are jointly independent with zero mean and unit variance.

(a) If ζ_{ij} 's are bounded by D , then for any $\beta > 8\sqrt{\pi}$, one has

$$\begin{cases} \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \leq \frac{1}{n} \log \mathcal{R} \left(\frac{\epsilon}{2}, n, m \right) + \frac{D}{\sqrt{\frac{\epsilon}{2} \epsilon a}} \frac{\beta}{n}, \\ \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \geq \frac{1}{n} \log \mathcal{R} \left(\frac{2}{\epsilon}, n, m \right) - \frac{D}{\sqrt{\epsilon a}} \frac{\beta}{n} \\ \quad - \frac{\log \left\{ 1 + \sqrt{\frac{8\pi D^2}{\epsilon a}} e^{8\pi + \frac{2D^2}{\epsilon a}} \right\}}{n}, \end{cases} \quad (76)$$

with probability exceeding $1 - 4 \exp \left(-\frac{\beta^2}{8} \right)$.

(b) If ζ_{ij} satisfies the LSI with respect to a uniform constant c_{1s} , then for any $\beta > 0$, one has

$$\begin{cases} \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \leq \frac{1}{n} \log \mathcal{R} \left(\epsilon, n, m \right) + \frac{\sqrt{c_{1s}}}{\sqrt{\epsilon a}} \frac{\beta}{n}, \\ \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \geq \frac{1}{n} \log \mathcal{R} \left(\epsilon, n, m \right) - \frac{\sqrt{c_{1s}}}{\sqrt{\epsilon a}} \frac{\beta}{n} \\ \quad - \frac{\log \left(1 + \sqrt{\frac{4\pi c_{1s}}{\epsilon a}} e^{\frac{c_{1s}}{4\epsilon a}} \right)}{n}, \end{cases} \quad (77)$$

with probability at least $1 - 2 \exp(-\beta^2)$.

(c) If ζ_{ij} 's are heavy-tailed distributed, then one has

$$\begin{cases} \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \leq \frac{1}{n} \log \mathcal{R} \left(\frac{\epsilon \epsilon_c}{2\sigma_c^2}, n, m \right) \\ \quad + 2 \log \sigma_c + \frac{\tau_c \sigma_c \sqrt{8c(n) \log n}}{\sqrt{\epsilon a n}}, \\ \frac{1}{n} \log \det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M}\mathbf{M}^* \right) \geq \frac{1}{n} \log \mathcal{R} \left(\frac{2\epsilon}{\epsilon \sigma_c^2}, n, m \right) \\ \quad + 2 \log \sigma_c - \frac{\tau_c \sigma_c \sqrt{8c(n) \log n}}{\sqrt{\epsilon a n}} - \frac{\log \left(1 + \sqrt{\frac{8\pi \tau_c^2 \sigma_c^2}{\epsilon a}} e^{8\pi + \frac{2\tau_c^2 \sigma_c^2}{\epsilon a}} \right)}{n}, \end{cases} \quad (78)$$

with probability exceeding $1 - \frac{5}{n^{c(n)}}$.

Here, the function $\mathcal{R}(\epsilon)$ is defined as

$$\mathcal{R}(\epsilon, n, m) := \sum_{i=0}^n \binom{n}{i} \frac{\epsilon^{n-i} m^{-i} m!}{(m-i)!}.$$

Proof of Lemma 4: Observe that the derivative of $g(x) := \log(\epsilon + x^2)$ is

$$g'(x) = \frac{2x}{\epsilon + x^2}, \quad x \geq 0,$$

which is bounded within the interval $\left[0, \frac{1}{\sqrt{\epsilon}}\right]$ when $x \geq 0$. Therefore, the Lipschitz norm of $g(x)$ satisfies

$$\|g\|_{\mathcal{L}} \leq \frac{1}{\sqrt{\epsilon}}.$$

The three class of probability measures are discussed separately as follows.

(a) Consider first the measure uniformly bounded by D . In order to apply Theorem 1, we need to first convexify the objective metric. Define a function $g_\epsilon(x)$ such that

$$g_\epsilon(x) := \begin{cases} \log(\epsilon + x^2), & \text{if } x \geq \sqrt{\epsilon}, \\ \frac{1}{\sqrt{\epsilon}}(x - \sqrt{\epsilon}) + \log(2\epsilon), & \text{if } 0 \leq x < \sqrt{\epsilon}. \end{cases}$$

Apparently, $g_\epsilon(x)$ is a concave function at $x \geq 0$, and its Lipschitz constant can be bounded above by

$$\|g_\epsilon\|_{\mathcal{L}} \leq \frac{1}{\sqrt{\epsilon}}.$$

By setting $f_\epsilon(x) := g_\epsilon(\sqrt{x})$, one can easily verify that

$$\log\left(\frac{2}{e}\epsilon + x\right) \leq f_\epsilon(x) \leq \log(\epsilon + x),$$

and hence

$$\begin{aligned} \frac{1}{n} \log \det\left(\frac{2}{e}\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) &\leq \frac{1}{n} \sum_{i=1}^n \left[f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) \right] \\ &\leq \frac{1}{n} \log \det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right). \end{aligned} \quad (79)$$

One desired feature of $f_\epsilon(x)$ is that $g_\epsilon(x) = f_\epsilon(x^2)$ is concave with finite Lipschitz norm bounded above by $\frac{1}{\sqrt{\epsilon}}$, and is sandwiched between two functions whose *exponential means* are computable.

By assumption, each entry of \mathbf{M} is bounded by D , and $\rho = \sqrt{\frac{n}{m}} = \frac{1}{\sqrt{\alpha}}$. Theorem 1 taken collectively with (79) suggests that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) &\leq \frac{1}{n} \log \mathbb{E} \left[\prod_{i=1}^n \exp\left\{f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right)\right\} \right] + \frac{D}{\sqrt{\epsilon\alpha}} \frac{\beta}{n} \\ &\leq \frac{1}{n} \log \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right] + \frac{D}{\sqrt{\epsilon\alpha}} \frac{\beta}{n} \end{aligned} \quad (80)$$

and

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) &\geq \frac{1}{n} \log \mathbb{E} \left[\prod_{i=1}^n \exp\left\{f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right)\right\} \right] - \frac{D}{\sqrt{\epsilon\alpha}} \frac{\beta}{n} \\ &\quad - \frac{\log\left\{1 + \sqrt{\frac{8\pi D^2}{\epsilon\alpha}} e^{8\pi + \frac{2D^2}{\epsilon\alpha}}\right\}}{n} \\ &\geq \frac{1}{n} \log \mathbb{E} \left[\det\left(\frac{2}{e}\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right] - \frac{D}{\sqrt{\epsilon\alpha}} \frac{\beta}{n} \\ &\quad - \frac{\log\left\{1 + \sqrt{\frac{8\pi D^2}{\epsilon\alpha}} e^{8\pi + \frac{2D^2}{\epsilon\alpha}}\right\}}{n} \end{aligned} \quad (81)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{8}\right)$.

To complete the argument for Part (a), we observe that (79) also implies

$$\begin{aligned} \sum_{i=1}^n f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) &\leq \log \det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \\ &\leq \sum_{i=1}^n f_{\frac{\epsilon}{2}}\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right). \end{aligned}$$

Substituting this into (80) and (81) and making use of the identity given in Lemma 2 complete the proof for Part (a).

(b) If the measure of \mathbf{M}_{ij} satisfies LSI with uniformly bounded constant c_{ls} , then $g(x) := \log(\epsilon + x^2)$ is Lipschitz with Lipschitz bound $\frac{1}{\sqrt{\epsilon}}$. Applying Theorem 1 yields

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) \leq \frac{1}{n} \log \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right] \\ \quad + \frac{\sqrt{c_{\text{ls}}}}{\sqrt{\epsilon\alpha}} \frac{\beta}{n}, \\ \frac{1}{n} \sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) \geq \frac{1}{n} \log \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right] \\ \quad - \frac{\sqrt{c_{\text{ls}}}}{\sqrt{\epsilon\alpha}} \frac{\beta}{n} - \frac{\log\left(1 + \sqrt{\frac{4\pi c_{\text{ls}}}{\epsilon\alpha}} e^{\frac{c_{\text{ls}}}{4\epsilon\alpha}}\right)}{n}, \end{cases} \quad (82)$$

with probability exceeding $1 - 2 \exp(-\beta^2)$.

(c) If the measure of \mathbf{M}_{ij} is heavy-tailed, then we again need to use the sandwich bound between $f_\epsilon(x)$ and $f(x)$. Theorem 1 indicates that

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) \leq \frac{1}{n} \log \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \tilde{\mathbf{M}} \tilde{\mathbf{M}}^*\right) \right] \\ \quad + \frac{\tau_c \sigma_c \sqrt{8c(n) \log n}}{\sqrt{\epsilon\alpha n}}, \\ \frac{1}{n} \sum_{i=1}^n f_\epsilon\left(\lambda_i\left(\frac{1}{m} \mathbf{M} \mathbf{M}^*\right)\right) \geq \frac{1}{n} \log \mathbb{E} \left[\det\left(\frac{2}{e}\epsilon \mathbf{I} + \frac{1}{m} \tilde{\mathbf{M}} \tilde{\mathbf{M}}^*\right) \right] \\ \quad - \frac{\tau_c \sigma_c \sqrt{8c(n) \log n}}{\sqrt{\epsilon\alpha n}} \\ \quad - \frac{\log\left(1 + \sqrt{\frac{8\pi \tau_c^2 \sigma_c^2}{\epsilon\alpha}} e^{8\pi + \frac{2\tau_c^2 \sigma_c^2}{\epsilon\alpha}}\right)}{n}, \end{cases} \quad (83)$$

with probability exceeding $1 - \frac{5}{n^{c(n)}}$. The only other difference with the proof of Part (a) is that the entries of $\tilde{\mathbf{M}}$ has variance σ_c^2 , and hence

$$\begin{aligned} \frac{1}{n} \log \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \tilde{\mathbf{M}} \tilde{\mathbf{M}}^*\right) \right] &= 2 \log \sigma_c + \frac{1}{n} \log \mathbb{E} \left[\det\left(\frac{\epsilon}{\sigma_c^2} \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right]. \end{aligned}$$

Finally, the proof is complete by applying Lemma 2 on $\mathbb{E}[\det(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*)]$.

APPENDIX F PROOF OF LEMMA 3

Using the results derived in Lemma 2, one can bound

$$\begin{aligned} \mathbb{E} \left[\det\left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^*\right) \right] &= m^{-n} \sum_{i=0}^n \binom{n}{i} \frac{m!}{(m-i)!} \epsilon^{n-i} m^{n-i} \\ &= m! m^{-n} \sum_{i=0}^n \binom{n}{i} \frac{\epsilon^i m^i}{(m-n+i)!} \\ &\leq m! m^{-n} (n+1) \max_{i:0 \leq i \leq n} \binom{n}{i} \frac{\epsilon^i m^i}{(m-n+i)!}. \end{aligned}$$

Denote by i_{\max} the index of the largest term as follows

$$i_{\max} := \arg \max_{i:0 \leq i \leq n} \binom{n}{i} \frac{\epsilon^i m^i}{(m-n+i)!}.$$

Suppose for now that $i_{\max} = \delta n$ for some numerical value δ , then we can bound

$$\begin{aligned}
& \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\
& \leq \frac{\log m!}{n} + \frac{\log(n+1)}{n} + \frac{\log \binom{n}{\delta n}}{n} + \delta \log \epsilon \\
& \quad - (1-\delta) \log m - \frac{\log(m-n+\delta n)!}{n} \\
& = \frac{\log \frac{m!}{(m-n+\delta n)!(n-\delta n)!}}{n} + \frac{\log(n+1)}{n} + \frac{\log \binom{n}{\delta n}}{n} \\
& \quad + \frac{\log(n-\delta n)!}{n} - (1-\delta) \log m + \delta \log \epsilon \\
& = \frac{\alpha \log \binom{m}{n-\delta n}}{m} + \frac{\log(n+1)}{n} + \frac{\log \binom{n}{\delta n}}{n} + \frac{\log(n-\delta n)!}{n} \\
& \quad - (1-\delta) \log m + \delta \log \epsilon. \tag{84}
\end{aligned}$$

It follows immediately from the well-known entropy formula (see [79, Example 11.1.3]) that

$$\mathcal{H} \left(\frac{k}{n} \right) - \frac{\log(n+1)}{n} \leq \frac{1}{n} \log \binom{n}{k} \leq \mathcal{H} \left(\frac{k}{n} \right), \tag{85}$$

where $\mathcal{H}(x) := x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$ denotes the binary entropy function. Also, the famous Stirling approximation bounds asserts that

$$\begin{aligned}
& \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \leq n! \leq e n^{n+\frac{1}{2}} e^{-n}, \\
\iff & \frac{2 + \log n}{2n} + \log n - 1 \geq \frac{1}{n} \log n! \\
& \geq \frac{\log(2\pi) + \log n}{2n} + \log n - 1. \tag{86}
\end{aligned}$$

Substituting (85) and (86) into (84) leads to

$$\begin{aligned}
& \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\
& \leq \alpha \mathcal{H} \left(\frac{(1-\delta)n}{m} \right) + \frac{\log(n+1)}{n} + \mathcal{H}(\delta) \\
& \quad + \frac{\log e + \frac{1}{2} \log[(1-\delta)n]}{n} + (1-\delta) \log[(1-\delta)n] \\
& \quad - (1-\delta) \log e - (1-\delta) \log m + \delta \log \epsilon \\
& = \alpha \mathcal{H} \left(\frac{1-\delta}{\alpha} \right) + \frac{\log(n+1) + 1 + \frac{1}{2} \log[(1-\delta)n]}{n} \\
& \quad + \mathcal{H}(\delta) + (1-\delta) \log \frac{1}{\alpha} + (1-\delta) \log(1-\delta) \\
& \quad - (1-\delta) + \delta \log \epsilon.
\end{aligned}$$

Making use of the inequality

$$\log(n+1) + 1 + \frac{1}{2} \log n \leq 1.5 \log(en),$$

one obtains

$$\begin{aligned}
& \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\
& \leq \alpha \mathcal{H} \left(\frac{1-\delta}{\alpha} \right) + \frac{1.5 \log(en)}{n} + \delta \log \frac{1}{\delta} \\
& \quad - (1-\delta) \log(e\alpha) + \delta \log \epsilon \\
& = \alpha \mathcal{H} \left(\frac{1}{\alpha} \right) - \log(\alpha e) + \frac{1.5 \log(en)}{n} + \delta \log \frac{1}{\delta} \\
& \quad + \delta \log(e\alpha) + \delta \log \epsilon + \alpha \left[\mathcal{H} \left(\frac{1-\delta}{\alpha} \right) - \mathcal{H} \left(\frac{1}{\alpha} \right) \right] \\
& \leq (\alpha-1) \log \left(\frac{\alpha}{\alpha-1} \right) - 1 + \frac{1.5 \log(en)}{n} + \delta \log \frac{1}{\delta} \\
& \quad + \delta \log(e\alpha) + \delta \log \epsilon + \mathcal{H}(\delta), \tag{87}
\end{aligned}$$

where the last inequality follows from the following identity on binary entropy function [80]: for any p, q obeying $0 \leq p, q \leq 1$ and $0 < pq < 1$, one has

$$\begin{aligned}
& \mathcal{H}(p) - \mathcal{H}(pq) = (1-pq) \mathcal{H} \left(\frac{p-pq}{1-pq} \right) - p \mathcal{H}(q), \\
\Rightarrow & \mathcal{H} \left(\frac{1-\delta}{\alpha} \right) - \mathcal{H} \left(\frac{1}{\alpha} \right) \leq \frac{1}{\alpha} \mathcal{H}(1-\delta) = \frac{1}{\alpha} \mathcal{H}(\delta).
\end{aligned}$$

It remains to estimate the index i_{\max} or, equivalently, the value δ . If we define

$$s_\epsilon(i) := \binom{n}{i} \frac{\epsilon^i m^i}{(m-n+i)!} \quad \text{and} \quad r_\epsilon(i) := \frac{s_\epsilon(i+1)}{s_\epsilon(i)},$$

then one can compute

$$r_\epsilon(i) = \frac{\frac{n!}{(i+1)!(n-i-1)!} \frac{\epsilon^{i+1} m^{i+1}}{(m-n+i+1)!}}{\frac{n!}{i!(n-i)!} \frac{\epsilon^i m^i}{(m-n+i)!}} = \frac{\epsilon m(n-i)}{(i+1)(m-n+i+1)},$$

which is a decreasing function in i . Suppose that $n\epsilon > \max\{4, 2(1 - \frac{1}{\alpha} + \frac{1}{\alpha n})\}$. By setting $r_\epsilon(x) = 1$, we can derive a valid positive solution x_0 as follows

$$\begin{aligned}
x_0 & = \frac{-(m+n+2+\epsilon m)}{2} \\
& \quad + \frac{\sqrt{(m+n+2+\epsilon m)^2 - 4(m-n+1) + 4\epsilon mn}}{2}.
\end{aligned}$$

Simple algebraic manipulation yields

$$\begin{aligned}
x_0 & < \frac{-(m+n+2+\epsilon m) + (m+n+2+\epsilon m) + 2\sqrt{\epsilon mn}}{2} \\
& \leq \sqrt{\epsilon an},
\end{aligned}$$

and

$$\begin{aligned}
x_0 & > \frac{-(m+n+2+\epsilon m) + \sqrt{(m+n+2+\epsilon m)^2 + 2\epsilon mn}}{2} \\
& = \frac{\epsilon mn}{(m+n+2+\epsilon m) + \sqrt{(m+n+2+\epsilon m)^2 + 2\epsilon mn}} \\
& = \frac{\alpha \epsilon n}{\alpha + 1 + \frac{2}{n} + \alpha \epsilon + \sqrt{(\alpha + 1 + \frac{2}{n} + \alpha \epsilon)^2 + 2\epsilon \alpha}} \\
& \geq \frac{\alpha \epsilon n}{2(\alpha + 1 + \frac{2}{n} + \alpha) + \sqrt{2\alpha} \epsilon n} > \frac{\alpha}{5(\alpha + 1)} \epsilon n.
\end{aligned}$$

Therefore, we can conclude that $\delta \in \left[\frac{\alpha}{5(1+\alpha)}\epsilon, \sqrt{\alpha\epsilon} \right]$. Assume that $\epsilon < \frac{1}{e^2\alpha}$. Substituting this into (87) then leads to

$$\begin{aligned} & \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\ & \leq (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right) - 1 + \mathcal{R}_{n,\epsilon}, \end{aligned}$$

where $\mathcal{R}_{n,\epsilon}$ denotes the residual term

$$\mathcal{R}_{n,\epsilon} := \frac{1.5 \log(en)}{n} + \sqrt{\alpha\epsilon} \left(\log \frac{1}{\sqrt{\alpha\epsilon}} + \log(e\alpha) \right) + \mathcal{H}(\sqrt{\alpha\epsilon}).$$

In particular, if we further have $\epsilon < \min \left\{ \frac{1}{e^2\alpha^3}, \frac{1}{4\alpha} \right\}$, then one has $\log ae < \log \frac{1}{\sqrt{\alpha\epsilon}}$ and

$$\mathcal{H}(\sqrt{\alpha\epsilon}) \leq 2\sqrt{\alpha\epsilon} \log \frac{1}{\sqrt{\alpha\epsilon}},$$

which in turn leads to

$$\mathcal{R}_{n,\epsilon} \leq \frac{1.5 \log(en)}{n} + 4\sqrt{\alpha\epsilon} \log \frac{1}{\sqrt{\alpha\epsilon}}.$$

On the other hand, the lower bound on $\frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right]$ can be easily obtained through the following argument

$$\mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \geq \mathbb{E} \left[\det \left(\frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] = \frac{m!}{(m-n)!m^n},$$

and hence

$$\begin{aligned} & \frac{1}{n} \log \mathbb{E} \left[\det \left(\epsilon \mathbf{I} + \frac{1}{m} \mathbf{M} \mathbf{M}^* \right) \right] \\ & \geq \frac{1}{n} \log \binom{m}{n} + \frac{\log n!}{n} - \log m \\ & \geq \alpha \mathcal{H} \left(\frac{1}{\alpha} \right) - \frac{\log(m+1)}{n} + \frac{\log(2\pi) + \frac{1}{2} \log n}{n} \\ & \quad + \log \left(\frac{n}{m} \right) - 1 \tag{88} \\ & \geq (\alpha - 1) \log \left(\frac{\alpha}{\alpha - 1} \right) - 1 - \frac{\log \frac{1}{2\pi} (m+1) - \frac{1}{2} \log n}{n}, \tag{89} \end{aligned}$$

where (88) makes use of the bounds (85) and (86). This concludes the proof.

APPENDIX G

PROOF OF THEOREM 3 AND COROLLARY 2

Proof of Theorem 3: Suppose that the singular value decomposition of \mathbf{H} can be expressed by $\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^*$ where $\mathbf{\Lambda}$ is a diagonal matrix consist of all singular values of \mathbf{H} . Then the

MMSE per input component can be expressed as

NMMSE (\mathbf{H} , SNR)

$$\begin{aligned} & \frac{1}{p} \operatorname{tr} \left(\mathbf{V} \mathbf{\Lambda} \mathbf{U}^* \left(\frac{1}{\text{SNR}} \mathbf{I} + \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^* \right)^{-1} \mathbf{U} \mathbf{\Lambda} \mathbf{V}^* \right) \\ & = 1 - \frac{1}{p} \operatorname{tr} \left(\mathbf{\Lambda} \left(\frac{1}{\text{SNR}} \mathbf{I} + \mathbf{\Lambda}^2 \right)^{-1} \mathbf{\Lambda} \right) \\ & = \frac{1}{p} \sum_{i=1}^p \frac{1}{\text{SNR}^{-1} + \lambda_i(\mathbf{H}^* \mathbf{H})}. \end{aligned}$$

(i) Note that $g_\epsilon(x) := \frac{1}{\epsilon^2 + x^2}$ is not a convex function. In order to apply Proposition 1, we define a convexified variant $\tilde{g}_\epsilon(x)$ of $g_\epsilon(x)$. Specifically, we set

$$\tilde{g}_\epsilon(x) := \begin{cases} \frac{1}{\epsilon^2 + x^2}, & \text{if } x > \frac{1}{\sqrt{3}}\epsilon, \\ -\frac{3\sqrt{3}}{8\epsilon^3} \left(x - \frac{\epsilon}{\sqrt{3}} \right) + \frac{3}{4\epsilon^2}, & \text{if } x \leq \frac{1}{\sqrt{3}}\epsilon, \end{cases}$$

which satisfies

$$\|\tilde{g}_\epsilon(x)\|_{\mathcal{L}} \leq \frac{3\sqrt{3}}{8\epsilon^3}.$$

One can easily check that

$$g_{\frac{3}{2\sqrt{2}}\epsilon}(x) \leq \tilde{g}_{\frac{3}{2\sqrt{2}}\epsilon}(x) \leq g_\epsilon(x) \leq \tilde{g}_\epsilon(x) \leq g_{\frac{2\sqrt{2}}{3}\epsilon}(x).$$

This implies that we can sandwich the target function as follows

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^* \mathbf{H})} \\ & \leq \frac{1}{p} \sum_{i=1}^p \left(\tilde{g}_\epsilon \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right) - \mathbb{E} \left[\tilde{g}_\epsilon \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right) \right] \right) \\ & \quad + \mathbb{E} \left[g_{\frac{2\sqrt{2}}{3}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right) \right], \tag{90} \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^* \mathbf{H})} \tag{91} \\ & \geq \frac{1}{p} \sum_{i=1}^p \left(\tilde{g}_{\frac{3}{2\sqrt{2}}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right) \right. \\ & \quad \left. - \mathbb{E} \left[\tilde{g}_{\frac{3}{2\sqrt{2}}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right) \right] \right) \\ & \quad + \mathbb{E} g_{\frac{3}{2\sqrt{2}}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^* \mathbf{H})} \right). \end{aligned}$$

Recall that

$$\mathbf{H}^* \mathbf{H} = \mathbf{M}^* \mathbf{A}^* \mathbf{A} \mathbf{M}, \tag{92}$$

where the entries of \mathbf{M} are independently generated with variance $\frac{1}{n}$. Since $\tilde{g}_\epsilon(x)$ is a convex function in x , applying Proposition 1 yields the following results.

- 1) If $\sqrt{p}\mathbf{M}_{ij}$'s are bounded by D , then for any $\beta > 8\sqrt{\pi}$, following simple bound we have

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^*\mathbf{H})} \\ & \leq \frac{3\sqrt{3}D\|\mathbf{A}\|\beta}{8\epsilon^3} \frac{\beta}{p} + \frac{1}{p} \sum_{i=1}^p \mathbb{E} \left[g_{\frac{2\sqrt{2}}{3}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^*\mathbf{H})} \right) \right], \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^*\mathbf{H})} \\ & \geq \left(\frac{2\sqrt{2}D\|\mathbf{A}\|}{3\sqrt{3}\epsilon^3} \right) \frac{\beta}{p} - \frac{1}{p} \sum_{i=1}^p \mathbb{E} \left[g_{\frac{3}{2\sqrt{2}}\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^*\mathbf{H})} \right) \right] \end{aligned}$$

with probability exceeding $1 - 8 \exp\left(-\frac{\beta^2}{16}\right)$.

- 2) If the measure of $\sqrt{p}\mathbf{M}_{ij}$ satisfies the LSI with a uniformly bounded constant c_{ls} , then for any $\beta > 0$, one has

$$\begin{aligned} & \left| \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^*\mathbf{H})} - \mathbb{E} \left[g_{\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^*\mathbf{H})} \right) \right] \right| \\ & \leq \left(\frac{3\sqrt{3}\sqrt{c_{\text{ls}}}\|\mathbf{A}\|}{8\epsilon^3} \right) \frac{\beta}{p} \end{aligned} \quad (93)$$

with probability exceeding $1 - 4 \exp\left(-\frac{\beta^2}{2}\right)$. Here, we have made use of the fact that $\|g_{\epsilon}\|_{\mathcal{L}} \leq \frac{3\sqrt{3}}{8\epsilon^3}$.

- 3) Suppose that $\sqrt{p}\mathbf{M}_{ij}$'s are independently drawn from symmetric heavy-tailed distributions. Proposition 1 suggests that

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^*\mathbf{H})} \\ & \leq \frac{3\sqrt{3}\tau_c\sigma_c\|\mathbf{A}\|\sqrt{c(n)\log n}}{2\epsilon^3 p} \\ & \quad + \frac{1}{p} \sum_{i=1}^p \mathbb{E} \left[g_{\frac{2\sqrt{2}}{3}\epsilon} \left(\sqrt{\lambda_i(\tilde{\mathbf{M}}^*\mathbf{A}^*\tilde{\mathbf{M}})} \right) \right], \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \frac{1}{\epsilon^2 + \lambda_i(\mathbf{H}^*\mathbf{H})} \\ & \geq \frac{8\sqrt{2}\tau_c\sigma_c\|\mathbf{A}\|\sqrt{c(n)\log n}}{3\sqrt{3}\epsilon^3 p} \\ & \quad - \frac{1}{p} \sum_{i=1}^p \mathbb{E} \left[g_{\frac{3}{2\sqrt{2}}\epsilon} \left(\sqrt{\lambda_i(\tilde{\mathbf{M}}^*\mathbf{A}^*\tilde{\mathbf{M}})} \right) \right] \end{aligned}$$

with probability exceeding $1 - \frac{10}{nc(n)}$. Here, $\tilde{\mathbf{M}}$ is a truncated copy of \mathbf{M} such that $\tilde{\mathbf{M}}_{ij} = \mathbf{M}_{ij} \mathbf{1}_{\{|\mathbf{H}_{ij}| \leq \tau_c\}}$.

This completes the proof.

Proof of Corollary 2: In the moderate-to-high SNR regime (i.e. when all of $\lambda_i(\mathbf{H}^*\mathbf{H})$ are large), one can apply the

$$\begin{aligned} & \frac{1}{p} \sum_{i=1}^p \left(\frac{1}{\lambda_i(\mathbf{H}^*\mathbf{H})} - \frac{\epsilon^2}{\lambda_i^2(\mathbf{H}^*\mathbf{H})} \right) \\ & \leq \frac{1}{p} \sum_{i=1}^p \mathbb{E} \left[g_{\epsilon} \left(\sqrt{\lambda_i(\mathbf{H}^*\mathbf{H})} \right) \right] \\ & \leq \frac{1}{p} \sum_{i=1}^p \frac{1}{\lambda_i(\mathbf{H}^*\mathbf{H})}. \end{aligned}$$

This immediately leads to

$$\begin{aligned} & \frac{1}{p} \text{tr} \left((\mathbf{H}^*\mathbf{H})^{-1} \right) - \frac{\epsilon^2}{p} \text{tr} \left((\mathbf{H}^*\mathbf{H})^{-2} \right) \\ & \leq \frac{1}{p} \text{tr} \left(\left(\epsilon^2 \mathbf{I}_p + \mathbf{H}^*\mathbf{H} \right)^{-1} \right) \\ & \leq \frac{1}{p} \text{tr} \left((\mathbf{H}^*\mathbf{H})^{-1} \right). \end{aligned} \quad (94)$$

When $\mathbf{H}_{ij} \sim \mathcal{N}\left(0, \frac{1}{p}\right)$, applying [70, Th. 2.2.8] gives

$$\text{tr} \left((\mathbf{H}^*\mathbf{H})^{-1} \right) = \frac{p^2}{(n-p-1)} = \frac{ap}{1-\alpha-\frac{1}{n}}. \quad (95)$$

Similarly, making use of [70, Th. 2.2.8] leads to

$$\begin{aligned} \text{tr} \left((\mathbf{M}^*\mathbf{M})^{-2} \right) & = p^2 \frac{p(n-p-2) + p + p^2}{(n-p)(n-p-1)(n-p-3)} \\ & = \frac{\left(1 - \frac{1}{n}\right) \alpha^2 p}{(1-\alpha)\left(1-\alpha-\frac{1}{n}\right)\left(1-\alpha-\frac{3}{n}\right)}. \end{aligned} \quad (96)$$

Substituting (95) and (96) into (94) yields

$$\begin{aligned} \frac{\alpha}{1-\alpha} - \frac{1}{\text{SNR}} \frac{\alpha^2}{\left(1-\alpha-\frac{3}{n}\right)^3} & \leq \frac{1}{p} \text{tr} \left(\left(\epsilon^2 \mathbf{I}_p + \mathbf{H}^*\mathbf{H} \right)^{-1} \right) \\ & \leq \frac{\alpha}{1-\alpha-\frac{1}{n}}, \end{aligned} \quad (97)$$

thus concluding the proof. \square

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for extensive and constructive comments that greatly improved the paper.

REFERENCES

- [1] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [3] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
- [4] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.
- [5] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.
- [6] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, submitted to *IEEE Transactions on Information Theory*.
- [7] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

- [8] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
- [9] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1245–1248, Aug. 2000.
- [10] S. V. Hanly and D. N. C. Tse, "Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1328–1351, May 2001.
- [11] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.
- [12] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.
- [13] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.
- [14] R. Müller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2002.
- [15] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: Asymptotic evaluation under correlated fading," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 829–838, Jun. 2003.
- [16] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
- [17] A. Lozano, "Capacity-approaching rate function for layered multi-antenna architectures," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 616–620, Jul. 2003.
- [18] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.
- [19] A. L. Moustakas and S. H. Simon, "On the outage capacity of correlated multiple-path MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3887–3903, Nov. 2007.
- [20] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5737–5753, Sep. 2011.
- [21] G. Taricco and E. Riegler, "On the ergodic capacity of correlated Rician fading MIMO channels with interference," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4123–4137, Jul. 2011.
- [22] T. Tao, *Topics in Random Matrix Theory* (Graduate Studies in Mathematics). Providence, RI, USA: AMS, 2012.
- [23] M. L. Mehta, *Random Matrices*, vol. 142. Amsterdam, The Netherlands: Elsevier, 2004.
- [24] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.
- [25] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.
- [26] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.
- [27] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Complex random matrices and Rayleigh channel capacity," *Commun. Inf. Syst.*, vol. 3, no. 2, pp. 119–138, 2003.
- [28] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Complex random matrices and Rician channel capacity," *Problems Inf. Transmiss.*, vol. 41, no. 1, pp. 1–22, 2005.
- [29] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Ann. Math.*, vol. 67, no. 2, pp. 325–327, 1958.
- [30] T. Tao, V. Vu, and M. Krishnapur, "Random matrices: Universality of ESDs and the circular law," *Ann. Probab.*, vol. 38, no. 5, pp. 2023–2065, 2010.
- [31] T. Tao and V. Vu, "Random matrices: The circular law," *Commun. Contemp. Math.*, vol. 10, no. 2, pp. 261–307, 2008.
- [32] V. A. Marchenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Matematicheskii Sbornik*, vol. 114, no. 4, pp. 507–536, 1967.
- [33] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.
- [34] R. Couillet and W. Hachem. (2013). "Analysis of the limiting spectral measure of large random matrices of the separable covariance type." [Online]. Available: <http://arxiv.org/abs/1310.8094>
- [35] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
- [36] G. Taricco, "Asymptotic mutual information statistics of separately correlated Rician fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3490–3504, Aug. 2008.
- [37] J. Hoydis, R. Couillet, P. Piantanida, and M. Debbah, "A random matrix approach to the finite blocklength regime of MIMO fading channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 2181–2185.
- [38] J. Hoydis, R. Couillet, and P. Piantanida. (2013). "The second-order coding rate of the MIMO Rayleigh block-fading channel." [Online]. Available: <http://arxiv.org/abs/1303.3400>
- [39] J. Hoydis, R. Couillet, and P. Piantanida, "Bounds on the second-order coding rate of the MIMO Rayleigh block-fading channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1526–1530.
- [40] S. Loyka and G. Levin, "Finite-SNR diversity-multiplexing tradeoff via asymptotic analysis of large MIMO systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4781–4792, Oct. 2010.
- [41] P. Kazakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, "Living at the edge: A large deviations approach to the outage MIMO capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1984–2007, Apr. 2011.
- [42] Y.-C. Liang, G. Pan, and Z. D. Bai, "Asymptotic performance of MMSE receivers for large systems using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4173–4190, Nov. 2007.
- [43] R. Couillet and M. Debbah, "Signal processing in large systems: A new paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24–39, Jan. 2013.
- [44] A. Karadimitrakis, A. L. Moustakas, and P. Vivo. (2013). "Outage capacity for the optical MIMO channel." [Online]. Available: <http://arxiv.org/abs/1302.0614>
- [45] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295–327, 2001.
- [46] I. M. Johnstone. (2006). "High dimensional statistical inference and random matrices." [Online]. Available: <http://arxiv.org/abs/math/0611589>
- [47] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [48] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated mimo multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.
- [49] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *AMS Math Challenges Lecture*. Stanford, CA, USA: Stanford Univ., 2000, pp. 1–32.
- [50] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing, Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 210–268.
- [51] A. Guionnet and O. Zeitouni, "Concentration of the spectral measure for large matrices," *Electron. Commun. Probab.*, vol. 5, pp. 119–136, 2000.
- [52] N. El Karoui, "Concentration of measure and spectra of random matrices: Applications to correlation matrices, elliptical distributions and beyond," *Ann. Appl. Probab.*, vol. 19, no. 6, pp. 2362–2405, 2009.
- [53] Z. D. Bai and J. W. Silverstein, "CLT for linear spectral statistics of large-dimensional sample covariance matrices," *Ann. Probab.*, vol. 32, no. 1A, pp. 553–605, 2004.
- [54] P. Massart and J. Picard, *Concentration Inequalities and Model Selection* (Lecture Notes in Mathematics). New York, NY, USA: Springer-Verlag, 2007.
- [55] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprasad, "Achieving 'massive MIMO' spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.
- [56] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [57] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2013.
- [58] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

- [59] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
- [60] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.
- [61] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. London, U.K.: Oxford Univ. Press, 2013.
- [62] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications and coding," in *Foundations and Trends in Communications and Information Theory*, vol. 10. Norwell, MA, USA: NOW Publishers, 2013.
- [63] M. Ledoux, "Concentration of measure and logarithmic Sobolev inequalities," in *Séminaire de Probabilités XXXIII*. Berlin, Germany: Springer-Verlag, 1999, pp. 120–216.
- [64] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Pub. Math. Inst. Hautes Etudes Sci.*, vol. 81, no. 1, pp. 73–205, 1995.
- [65] Z. D. Bai, "Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices," *Ann. Probab.*, vol. 21, no. 2, pp. 649–672, 1993.
- [66] Z. D. Bai, B. Miao, and J.-F. Yao, "Convergence rates of spectral distributions of large sample covariance matrices," *SIAM J. Matrix Anal. Appl.*, vol. 25, no. 1, pp. 105–127, 2003.
- [67] F. Götze and A. N. Tikhomirov, "The rate of convergence of spectra of sample covariance matrices," *Theory Probab. Appl.*, vol. 54, no. 1, pp. 129–140, 2010.
- [68] F. Götze and A. Tikhomirov, "The rate of convergence for spectra of GUE and LUE matrix ensembles," *Central Eur. J. Math.*, vol. 3, no. 4, pp. 666–704, 2005.
- [69] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
- [70] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu, *Multivariate Statistics: High-Dimensional and Large-Sample Approximations* (Probability and Statistics). Hoboken, NJ, USA: Wiley, 2010.
- [71] M. Chiani, A. Conti, M. Mazzotti, E. Paolini, and A. Zanella, "Outage capacity for MIMO-OFDM systems in block fading channels," in *Proc. Asilomar Conf. Signals, Syst., Comput. (ASILOMAR)*, Nov. 2011, pp. 779–783.
- [72] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: Capacity of sampled Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4889–4914, Aug. 2013.
- [73] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications* (Foundations and Trends in Communications and Information Theory). Hanover, MA, USA: Now Publishers, 2004.
- [74] S. Li, M. R. McKay, and Y. Chen, "On the distribution of MIMO mutual information: An in-depth Painlevé-based characterization," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5271–5296, Sep. 2013.
- [75] Y. Chen and M. R. McKay, "Coulumb fluid, Painlevé transcendents, and the information theory of MIMO systems," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4594–4634, Jul. 2012.
- [76] D. Passelier, M. R. McKay, and Y. Chen. (2014). "Asymptotic linear spectral statistics for spiked Hermitian random matrix models." [Online]. Available: <http://arxiv.org/abs/1402.6419>
- [77] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [78] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [79] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2012.
- [80] S. Verdú, *Information Theory*, to be published. [Online]. Available: <http://blogs.princeton.edu/blogit/>

Yuxin Chen (S'09) received the B.S. in Microelectronics with High Distinction from Tsinghua University in 2008, the M.S. in Electrical and Computer Engineering from the University of Texas at Austin in 2010, and the M.S. in Statistics from Stanford University in 2013. He is currently a Ph.D. candidate in the Department of Electrical Engineering at Stanford University. His research interests include information theory, compressed sensing, network science and high-dimensional statistics.

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) is the Stephen Harris professor in the School of Engineering and a professor of Electrical Engineering at Stanford University. She was previously on the faculty of Electrical Engineering at Caltech. Her research interests are in information theory and communication theory, and their application to wireless communications and related fields. She co-founded and serves as Chief Scientist of Acclera, Inc., and previously co-founded and served as CTO of Quantenna Communications, Inc. She has also held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Dr. Goldsmith is a Fellow of Stanford, and she has received several awards for her work, including the IEEE Communications Society and Information Theory Society joint paper award, the IEEE Communications Society Best Tutorial Paper Award, the National Academy of Engineering Gilbreth Lecture Award, the IEEE ComSoc Communications Theory Technical Achievement Award, the IEEE ComSoc Wireless Communications Technical Achievement Award, the Alfred P. Sloan Fellowship, and the *Silicon Valley/San Jose Business Journal's* Women of Influence Award. She is author of the book *Wireless Communications* and co-author of the books *MIMO Wireless Communications* and *Principles of Cognitive Radio*, all published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley.

Dr. Goldsmith has served on the Steering Committee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and as editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, the *Journal on Foundations and Trends in Communications and Information Theory* and in *Networks*, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the *IEEE Wireless Communications Magazine*. She participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She has also been a Distinguished Lecturer for both societies, served as President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory society, and chaired the Emerging Technology Committee of the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award, served as Chair of Stanfords Faculty Senate in 2009 and currently serves on its Faculty Senate and on its Budget Group.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering both from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently a Professor in the Department of Electrical Engineering at the Technion-Israel Institute of Technology, Haifa and holds the The Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in electrical engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, in 2007 the Henry Taub Prize for Excellence in Research, in 2008 the Hershel Rich Innovation Award, the Award for Women with Distinguished Contributions, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion Outstanding Lecture Award, in 2009 the Technion's Award for Excellence in Teaching, in 2010 the Michael Bruno Memorial Award from the Rothschild Foundation, and in 2011 the Weizmann Prize for Exact Sciences. In 2012 she was elected to the Young Israel Academy of Science and to the Israel Committee for Higher Education. In 2013 she received the Technion's Award for Excellence in Teaching, the Hershel Rich Innovation Award, and the IEEE Signal Processing Technical Achievement Award. She received several best paper awards together with her research students and colleagues. She received several best paper awards together with her research students and colleagues. She is the Editor in Chief of *Foundations and Trends in Signal Processing*. In the past, she was a Signal Processing Society Distinguished Lecturer, member of the IEEE Signal Processing Theory and Methods and Bio Imaging Signal Processing technical committees, and served as an associate editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the *EURASIP Journal of Signal Processing*, the *SIAM Journal on Matrix Analysis and Applications*, and the *SIAM Journal on Imaging Sciences*.