

On Noncoherent Multiple-Antenna Rayleigh Block-Fading Channels at Finite Blocklength

Chao Qi and Tobias Koch, *Senior Member, IEEE*

Abstract—This paper investigates the maximum coding rate at which data can be transmitted over a noncoherent, multiple-input, multiple-output (MIMO) Rayleigh block-fading channel using an error-correcting code of a given blocklength with a block-error probability not exceeding a given value. A high-SNR normal approximation is derived that becomes accurate as the signal-to-noise ratio (SNR) and the number of coherence intervals over which we code tend to infinity. The obtained normal approximation complements the nonasymptotic bounds that have appeared in the literature, but whose evaluation is computationally demanding. It further lays the theoretical foundation for an analytical analysis of the fundamental tradeoff between diversity, multiplexing, and channel-estimation cost at finite blocklength and finite SNR.

Index Terms—Channel dispersion, finite blocklength, high SNR, MIMO, normal approximation, Rayleigh fading, wireless communications

I. INTRODUCTION

There exists an increasing interest in the transmission of short packets in wireless communications [1]. For example, the vast majority of wireless connections in the next generations of cellular systems will most likely be originated by autonomous machines and devices, which predominantly exchange short packets. It is also expected that enhanced mobile-broadband services will be complemented by new services that target systems requiring reliable real-time communication with stringent requirements on latency and reliability. While *capacity* and *outage capacity* provide accurate benchmarks for the throughput achievable in wireless communication systems when the package length is not restricted, for short-package wireless communications, a more refined analysis of the maximum coding rate as a function of the blocklength is needed. Such an analysis is provided in this paper.

Let $R^*(n, \epsilon)$ denote the maximum coding rate at which data can be transmitted using an error-correcting code of blocklength n with a block-error probability no larger than ϵ . Hayashi [2] and Polyanskiy, Poor, and Verdú [3] showed

Tobias Koch has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 714161), from the Spanish Ministerio de Ciencia, Innovación y Universidades under Grants PID2024-159557OB-C21 (MICIU/AEI/10.13039/501100011033 and ERDF/UE), PID2020-116683GB-C21 (AEI/10.13039/501100011033), TEC2016-78434-C3-3-R (AEI/FEDER, EU), and RYC-2014-16332, and from the Comunidad de Madrid under Grant IDEA-CM (TEC-2024/COM-89). The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Los Angeles, CA, June 2020.

Chao Qi is with School of Computer and Software Engineering, Xihua University, 610039, Chengdu, China (email: chaoqi@mail.xhu.edu.cn).

Tobias Koch is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911, Leganés, Spain and also with the Gregorio Marañón Health Research Institute, 28007, Madrid, Spain (email: tkoch@ing.uc3m.es).

that, for various channels with a positive capacity C , $R^*(n, \epsilon)$ can be tightly approximated as

$$R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \quad (1)$$

where V is the so-called channel dispersion, $Q^{-1}(\cdot)$ denotes the inverse of the Q-function $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$, and $\mathcal{O}(\log n/n)$ comprises terms that decay no slower than $\log n/n$. The approximation that follows by ignoring the $\mathcal{O}(\log n/n)$ term is sometimes referred to as *normal approximation*. The normal approximation has been established as a benchmark for short error-correcting codes; see, e.g., [4], [5]. It further serves as a proxy for the maximum coding rate in the analysis and optimization of communication systems that exchange short packets and has appeared in numerous papers on short-packet wireless communications; see, e.g., [6]–[15].

However, many of these works consider the normal approximation of the Gaussian channel, which may fail to capture the effects of key parameters in wireless communication systems, such as coherence time, diversity, or multiplexing gain.

To address this shortcoming, the work of Polyanskiy, Poor, and Verdú has been generalized to several wireless communication channels [16]–[27]. In particular, the channel dispersion of coherent fading channels, where the receiver has perfect knowledge of the realizations of the fading coefficients, was obtained by Polyanskiy and Verdú for the single-antenna case [16], and by Collins and Polyanskiy for the multiple-input single-output (MISO) [17] and the multiple-input multiple-output (MIMO) case [18], [19]. The case where the number of transmit and receive antennas grows with the blocklength was considered in [20]. When both the transmitter and the receiver have perfect knowledge of the realization of the fading coefficients and the transmitter satisfies a long-term power constraint, the channel dispersion of single-antenna quasi-static fading channels was obtained by Yang *et al.* [21]. In the noncoherent setting, the channel dispersion is only known in the quasi-static case, where it is zero [22]. For general noncoherent Rayleigh block-fading channels, nonasymptotic bounds on the maximum coding rate were presented in [23], [24]. Saddlepoint approximations that accurately approximate these bounds in the single-antenna case with a reduced computational cost were given in [25], [26]. However, a closed-form expression of the channel dispersion for general noncoherent Rayleigh block-fading channels is still unknown. Obtaining such an expression is difficult because the capacity-achieving input distribution is unknown, so the standard approach of obtaining expressions of the form (1), which consists of first evaluating nonasymptotic upper and lower bounds on $R^*(n, \epsilon)$ for the capacity-achieving input and output distributions and

then analyzing these bounds in the limit as $n \rightarrow \infty$, cannot be followed. Fortunately, the asymptotic behavior of the capacity of such channels at high signal-to-noise ratio (SNR) is well understood [28], [29]. This fact was exploited in [27] to derive a high-SNR normal approximation of $R^*(n, \epsilon)$ for noncoherent single-antenna Rayleigh block-fading channels.

In this paper, we generalize [27] to the MIMO case. In particular, we present an expression of $R^*(n, \epsilon)$ similar to (1) for noncoherent MIMO Rayleigh block-fading channels. By deriving asymptotically-tight approximations of the capacity and the channel dispersion at high SNR, we obtain a high-SNR normal approximation of $R^*(n, \epsilon)$, which complements the existing nonasymptotic bounds.

The rest of the paper is organized as follows. Section II introduces the system model. Section III presents and discusses the main result of this paper: a high-SNR normal approximation for noncoherent MIMO block-fading channels. The proof of the main result is given in Section IV. The paper concludes with a summary and discussion of our results in Section V. Some of the proofs are deferred to the appendices.

Notation

Upper-case letters such as X denote scalar random variables and their realizations are written in lower case, e.g., x . We use boldface upper-case letters to denote random matrices, e.g., \mathbf{X} , and upper-case letters of a special font for their realizations, e.g., \mathbf{X} . The distribution of a circularly-symmetric complex Gaussian random variable with mean μ and variance σ^2 is denoted by $\mathcal{CN}(\mu, \sigma^2)$. The gamma distribution of shape parameter α and rate parameter λ is denoted by $\Gamma(\alpha, \lambda)$. We use $\mathbb{E}[\cdot]$ and $\text{Var}(\cdot)$ to denote expectation and variance, respectively. The symbol “ $\stackrel{d}{=}$ ” indicates equivalence in distribution.

We write $(\cdot)^\mathsf{H}$, $(\cdot)^*$, and $(\cdot)^\mathsf{T}$ to denote Hermitian transpose, complex conjugation, and transpose, respectively, and $\text{tr}(\cdot)$ and $\det(\cdot)$ to denote the trace and the determinant, respectively. The $(a \times a)$ -dimensional identity matrix is written as \mathbf{I}_a , and $\text{diag}\{d_1, \dots, d_a\}$ denotes an $(a \times a)$ -dimensional diagonal matrix with entries d_1, \dots, d_a . The diagonal submatrix containing the first i columns and rows of a matrix \mathbf{X} is denoted as $[\mathbf{X}]_{(1:i)^2}$. For any matrix \mathbf{X} , $\|\mathbf{X}\|_\mathsf{F}$ denotes the Frobenius norm.

Throughout the paper, $\log(\cdot)$ denotes the natural logarithm function, $[a]^+$ stands for $\max\{a, 0\}$, and $\mathbb{1}\{\cdot\}$ denotes the indicator function. We shall further use the following gamma and digamma functions:

$$\Gamma(x) \triangleq \int_0^\infty t^{x-1} e^{-t} dt, \quad \text{gamma function} \quad (2a)$$

$$\Gamma_m(x) \triangleq \pi^{m(m-1)/2} \prod_{k=1}^m \Gamma(x - k + 1), \quad \text{complex multivariate gamma function} \quad (2b)$$

$$\Psi(x) \triangleq \frac{d}{dx} \log \Gamma(x), \quad \text{digamma function} \quad (2c)$$

$$\Psi'(x) \triangleq \frac{d}{dx} \Psi(x), \quad \text{derivative of } \Psi(\cdot). \quad (2d)$$

Last but not least, we denote by $\overline{\lim}$ the *limit superior* and by $\underline{\lim}$ the *limit inferior*.

II. SYSTEM MODEL

We consider a Rayleigh block-fading channel with n_t transmit antennas, n_r receive antennas, and coherence interval T . For this channel, within the ℓ -th coherence interval, the channel input-output relation is given by

$$\mathbf{Y}_\ell = \mathbf{X}_\ell \mathbf{H}_\ell + \mathbf{W}_\ell \quad (3)$$

where \mathbf{X}_ℓ is the complex-valued, $(T \times n_t)$ -dimensional, transmitted matrix; \mathbf{Y}_ℓ is the complex-valued, $(T \times n_r)$ -dimensional, received matrix; \mathbf{H}_ℓ is the $(n_t \times n_r)$ -dimensional fading matrix with independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ entries; \mathbf{W}_ℓ is the $(T \times n_r)$ -dimensional, additive noise matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries. We assume that \mathbf{H}_ℓ and \mathbf{W}_ℓ are independent and take on independent realizations over successive coherence intervals. We further assume that the joint law of $(\mathbf{H}_\ell, \mathbf{W}_\ell)$ does not depend on \mathbf{X}_ℓ . We consider a noncoherent setting where transmitter and receiver know the statistics of the fading matrix \mathbf{H}_ℓ , but do not have knowledge of its realization.

We assume that $n_t \leq n_r$ and $T \geq n_t + n_r$. The assumption $n_t \leq n_r$ incurs no loss in capacity at high SNR. Indeed, Zheng and Tse [28] showed that the capacity of the noncoherent Rayleigh block-fading channel can be expressed as

$$C(\rho) = n_* \left(1 - \frac{n_*}{T} \right) \log \rho + \mathcal{O}_\rho(1) \quad (4)$$

where $n_* \triangleq \min\{n_t, n_r, \lfloor T/2 \rfloor\}$ and $\mathcal{O}_\rho(1)$ summarizes terms that are bounded in the SNR ρ . This implies that, for a given coherence time T and number of receive antennas n_r , the capacity pre-log is maximized by using $n_t = \min\{n_r, \lfloor T/2 \rfloor\}$ transmit antennas. Thus, using more than n_r transmit antennas does not increase the high-SNR capacity. The assumption $T \geq n_t + n_r$ is reasonable for slow-fading channels when the number of antennas is moderate. Under this assumption, an input distribution referred to as *unitary space-time modulation (USTM)* achieves a lower bound on the capacity that is asymptotically tight as the SNR tends to infinity [28], [29]. When $T < n_t + n_r$, USTM inputs are no longer optimal and an input distribution called *beta-variate space-time modulation (BSTM)* should be considered instead [29]. Analyzing the maximum coding rate for this input distribution requires a different analysis that is beyond the scope of this paper.

We shall restrict ourselves to codes of blocklength $n = LT$, where L denotes the number of blocks of coherence interval T a codeword spans. An $(L, T, M, \epsilon, \rho)$ code consists of:

- (1) An encoder $f: \{1, \dots, M\} \rightarrow (\mathbb{C}^{T \times n_t})^L$ that maps a message Φ , which is uniformly distributed on $\{1, \dots, M\}$, to a codeword $\mathbf{X}^L = [\mathbf{X}_1, \dots, \mathbf{X}_L]$. The codewords are assumed to satisfy the power constraint¹

$$\|\mathbf{X}_\ell\|_\mathsf{F}^2 \leq T\rho, \quad \ell = 1, \dots, L, \text{ almost surely.} \quad (5)$$

¹In the information theory literature, it is more common to impose a power constraint per codeword \mathbf{X}^L . Such a constraint would allow for a more flexible power allocation across coherence intervals and may give rise to a larger maximum coding rate. However, practical systems typically require a per-coherence-interval constraint to avoid large peak-to-average-power ratios. Furthermore, allocating equal power to each coherence interval is optimal with respect to channel capacity, in which case the per-codeword power constraint specializes to the per-coherence-interval constraint. It therefore seems plausible that both power constraints give rise to the same maximum coding rate for L sufficiently large, where the channel-capacity term dominates.

Since the variances of \mathbf{H}_ℓ and \mathbf{W}_ℓ are normalized, ρ can be interpreted as the average SNR at the receiver.

(2) A decoder $g: (\mathbb{C}^{T \times n_t})^L \rightarrow \{1, \dots, M\}$ satisfying the maximum error probability constraint

$$\max_{1 \leq \phi \leq M} \Pr[g(\mathbf{Y}^L) \neq \Phi | \Phi = \phi] \leq \epsilon \quad (6)$$

where $\mathbf{Y}^L = [\mathbf{Y}_1, \dots, \mathbf{Y}_L]$ is the channel output induced by the codeword $\mathbf{X}^L = f(\phi)$, according to (3).

The maximum coding rate is defined as

$$R^*(L, T, \epsilon, \rho) \triangleq \sup \left\{ \frac{\log M}{LT} : \exists (L, T, M, \epsilon, \rho) \text{ code} \right\}. \quad (7)$$

III. MAIN RESULTS

The main result of this paper is a high-SNR normal approximation of $R^*(L, T, \epsilon, \rho)$ presented in Section III-A. In Section III-B, we discuss the accuracy of this approximation by means of numerical results. In Section III-C, the normal approximation is applied to study the optimal number of active transmit antennas.

A. High-SNR Normal Approximation

Theorem 1: Assume that $T \geq n_r + n_t$, $n_r \geq n_t$, and $0 < \epsilon < \frac{1}{2}$. Then, at high SNR,

$$R^*(L, T, \epsilon, \rho) = \tilde{I}(T, \rho) - \sqrt{\frac{\tilde{V}(T) + K_{\tilde{V}}(T, \rho)}{L}} Q^{-1}(\epsilon) + K_{\tilde{I}}(T, \rho) + K_L(L, T, \rho) \quad (8)$$

where

$$\begin{aligned} \tilde{I}(T, \rho) &\triangleq n_t \left(1 - \frac{n_t}{T} \right) \log \frac{\rho}{n_t} + n_t \left(1 - \frac{n_t}{T} \right) \log \frac{T}{e} \\ &+ \frac{1}{T} \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \left(1 - \frac{n_t}{T} \right) \sum_{i=0}^{n_t-1} \Psi(n_r - i) \end{aligned} \quad (9)$$

$$\tilde{V}(T) \triangleq \frac{n_t}{T} \left(1 - \frac{n_t}{T} \right) + \left(1 - \frac{n_t}{T} \right)^2 \sum_{i=0}^{n_t-1} \Psi'(n_r - i). \quad (10)$$

In (8), $K_{\tilde{I}}(T, \rho)$ and $K_{\tilde{V}}(T, \rho)$ are functions of T and ρ that satisfy

$$\lim_{\rho \rightarrow \infty} K_\xi(T, \rho) = 0, \quad \xi \in \{\tilde{I}, \tilde{V}\} \quad (11)$$

and $K_L(L, T, \rho)$ is a function of L , T , and ρ that satisfies

$$\sup_{\rho \geq \rho_0} |K_L(L, T, \rho)| \leq A \frac{\log L}{L}, \quad L \geq L_0 \quad (12)$$

for some A , L_0 , and ρ_0 independent of L and ρ .

Proof: See Sec. IV. \blacksquare

The quantity $\tilde{I}(T, \rho)$ is an asymptotically-tight lower bound on the capacity $C(\rho)$ of the noncoherent MIMO Rayleigh block-fading channel [29]. Similarly, $\tilde{V}(T)$ can be viewed as a high-SNR approximation of the channel dispersion. By ignoring the error terms $K_{\tilde{I}}(T, \rho)$, $K_{\tilde{V}}(T, \rho)$, and $K_L(L, T, \rho)$ in (8), we obtain the high-SNR normal approximation

$$R^*(L, T, \epsilon, \rho) \approx \tilde{I}(T, \rho) - \sqrt{\frac{\tilde{V}(T)}{L}} Q^{-1}(\epsilon). \quad (13)$$

Observe that $\tilde{I}(T, \rho)$ and $\tilde{V}(T)$ depend on n_r only via the digamma function $\Psi(n_r - i)$ and its derivative $\Psi'(n_r - i)$, respectively. On the domain of positive integers, the digamma function is monotonically increasing, and its derivative is monotonically decreasing. As a consequence, the approximation (13) is monotonically increasing in n_r . As we shall observe in Sections III-B and III-C, the dependence of (13) on n_t is more intricate. Intuitively, increasing the number of transmit antennas achieves a higher multiplexing gain $\min\{n_t, n_r\}$, but it also requires the estimation of more channel coefficients.

For comparison, at high SNR, the capacity $C_c(\rho)$ of the coherent MIMO Rayleigh block-fading channel satisfies [30]

$$\lim_{\rho \rightarrow \infty} \left\{ C_c(\rho) - n_t \log \frac{\rho}{n_t} - \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \right\} = 0. \quad (14)$$

Furthermore, the channel dispersion $V_c(\rho)$ converges to [19]

$$\lim_{\rho \rightarrow \infty} V_c(\rho) = \frac{n_t}{T} + \text{Var}(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})). \quad (15)$$

Applying [31, Lemma A.2] (see also Lemma 7 below), the sum of the digamma functions $\Psi(n_r - i)$ in (9) can be identified as $\mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})]$, and the sum of the derivatives of these functions in (10) can be identified as $\text{Var}(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}))$. The high-SNR capacity and dispersion of the noncoherent block-fading channel can thus be written as

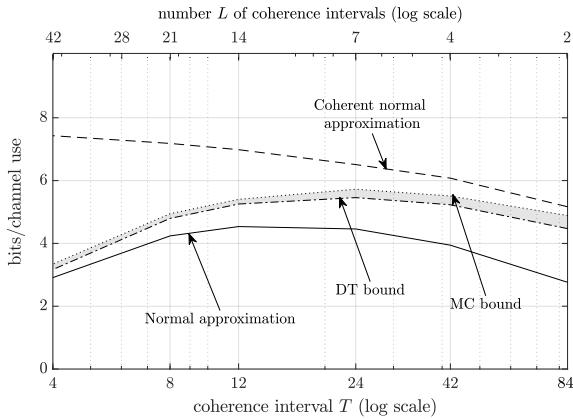
$$\begin{aligned} \tilde{I}(T, \rho) &= n_t \left(1 - \frac{n_t}{T} \right) \log \frac{\rho}{n_t} + n_t \left(1 - \frac{n_t}{T} \right) \log \frac{T}{e} \\ &+ \left(1 - \frac{n_t}{T} \right) \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] + \frac{1}{T} \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} \end{aligned} \quad (16)$$

and

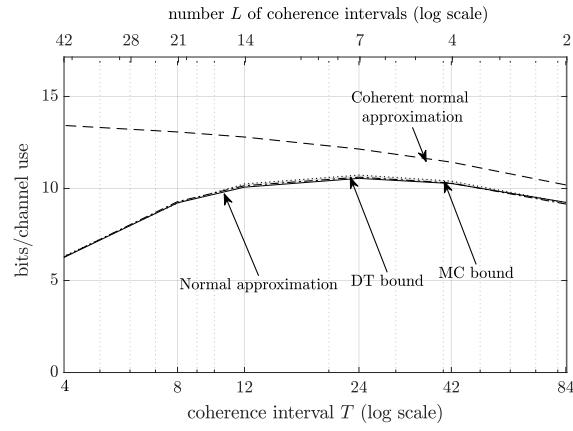
$$\tilde{V}(T) = \frac{n_t}{T} \left(1 - \frac{n_t}{T} \right) + \left(1 - \frac{n_t}{T} \right)^2 \text{Var}(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})). \quad (17)$$

Observe that, up to SNR-independent terms, $\tilde{I}(T, \rho)$ is given by $(1 - n_t/T)$ times the high-SNR approximation of $C_c(\rho)$. Similarly, it can be shown that $\tilde{V}(T)$ corresponds to the high-SNR channel dispersion one obtains by transmitting in each coherence block one pilot symbol per transmit antenna to estimate the fading coefficient, and by then transmitting $T - n_t$ symbols over a coherent fading channel. This suggests the heuristic that one should transmit one pilot symbol per transmit antenna and coherence block followed by coherent transmission. However, this heuristic may be misleading, since the SNR-independent terms of $\tilde{I}(T, \rho)$ may be significant, and since it is *prima facie* unclear whether one pilot symbol per transmit antenna and coherence block suffices to obtain a fading estimate of sufficient accuracy. For the single-antenna case, an analysis of the maximum coding rate achievable with pilot-assisted transmission can be found in [32].

Theorem 1 provides an asymptotic analysis of the maximum coding rate $R^*(L, T, \epsilon, \rho)$ as L tends to infinity for a fixed coherence interval T . Indeed, the error terms in (8) vanish as $L \rightarrow \infty$ and $\rho \rightarrow \infty$, provided that T is held fixed. This corresponds to the ergodic case where codewords span independent coherence intervals. The complementary case where $T \rightarrow \infty$ as L is held fixed was considered



(a) $n = 168$, $\epsilon = 10^{-3}$, $\rho = 15$ dB, and $n_t = n_r = 2$.



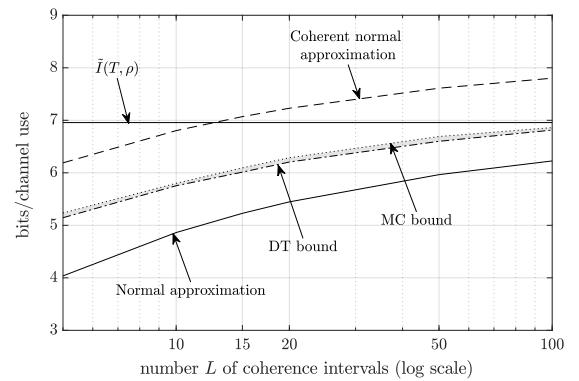
(b) $n = 168$, $\epsilon = 10^{-3}$, $\rho = 25$ dB, and $n_t = n_r = 2$.

Fig. 1. $R^*(L, T, \epsilon, \rho)$ as a function of T for a fixed $n = TL$.

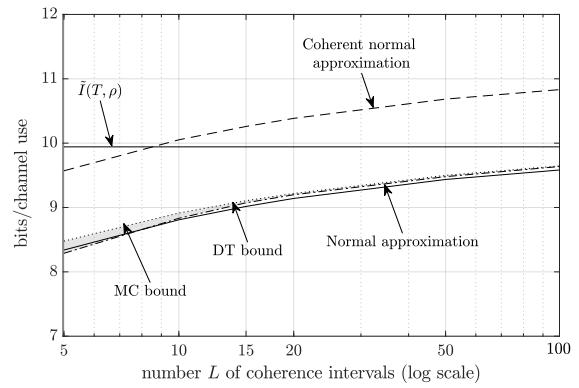
in [22]. Specifically, [22] considers a quasi-static multiple-antenna fading channel and analyzes the convergence of the maximum coding rate to the outage capacity in the limit as $T \rightarrow \infty$. By replacing the fading matrix \mathbb{H} in [22] by a block-diagonal matrix, one obtains a channel model that is equivalent to the block-fading channel (3). The corresponding results in [22, Ths. 3 & 9] demonstrate that the ϵ -dispersion is zero, irrespective of whether channel-state information is available at the transmitter or receiver. This implies that the maximum coding rate converges quickly to the outage capacity as $T \rightarrow \infty$. Consequently, in this regime, the outage capacity is a good proxy for the maximum coding rate. In general, both analyses (ergodic and quasi-static) characterize the maximum coding rate in different regimes and complement each other.

B. Numerical Results

We next provide numerical examples that illustrate the accuracy of the high-SNR normal approximation in Theorem 1. In the following figures, we depict the high-SNR normal approximation (8), the normal approximation of the coherent MIMO Rayleigh block-fading channel obtained in [19], a nonasymptotic (in ρ and L) lower bound on $R^*(L, T, \epsilon, \rho)$ that is based on the dependence testing (DT) bound [3, Th. 22], and a nonasymptotic upper bound that is based on the meta converse (MC) bound [3, Th. 31], both specialized to



(a) $T = 24$, $\epsilon = 10^{-3}$, $n_t = 2$, $n_r = 2$, and $\rho = 15$ dB.



(b) $T = 24$, $\epsilon = 10^{-3}$, $n_t = 2$, $n_r = 4$, and $\rho = 15$ dB.

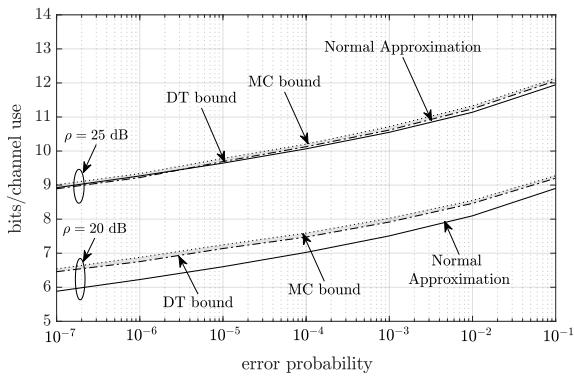
Fig. 2. $R^*(L, T, \epsilon, \rho)$ as a function of n for a fixed T .

the noncoherent MIMO Rayleigh block-fading channel in [24]. The nonasymptotic bounds are evaluated using the communication toolbox SPECTRE [33]. While these bounds are very accurate, their computational complexity grows exponentially in the blocklength and can be substantial. For the noncoherent single-antenna block-fading channel, a discussion of the computational complexity of these bounds and approximations thereof can be found in [26, Sec. VIII-A]. In the figures, the shaded area indicates the area where $R^*(L, T, \epsilon, \rho)$ lies.

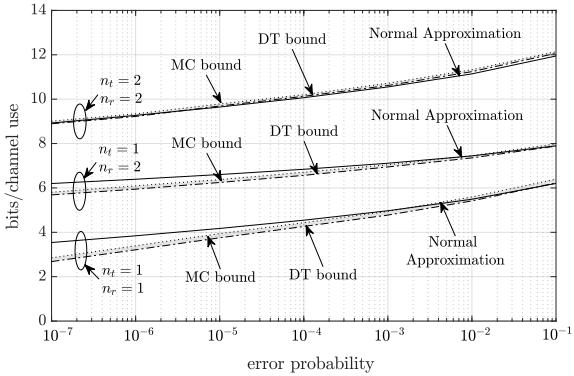
In Figs. 1a and 1b, we show $R^*(L, T, \epsilon, \rho)$ as a function of T for a fixed blocklength² $n = LT = 168$ and SNR values $\rho = 15$ dB and $\rho = 25$ dB. Observe that the high-SNR normal approximation (8) accurately describes the maximum coding rate for $\rho = 25$ dB but is loose for $\rho = 15$ dB unless L is large. Furthermore, the normal approximation of the coherent setting is strictly larger than that of the noncoherent setting, and it becomes more accurate as T increases, indicating that the cost for estimating the channel decreases with T .

In Figs. 2a and 2b, we show $R^*(L, T, \epsilon, \rho)$ as a function of the blocklength $n = LT$ for $T = 24$ and $\rho = 15$ dB, and for $n_t = n_r = 2$, and $n_t = 2, n_r = 4$, respectively. For comparison, we further depict $\tilde{I}(T, \rho)$ (9), which is an asymptotically tight lower bound on capacity. Observe that the normal approximation (8) becomes more accurate as L and n_r increase, and for $n_r = 4$, it is very close to the nonasymptotic bounds for the considered values of L . Moreover, the gap

²Thus, L is inversely proportional to T .



(a) $n_t = 2$, $n_r = 2$, $T = 24$, and $L = 7$.



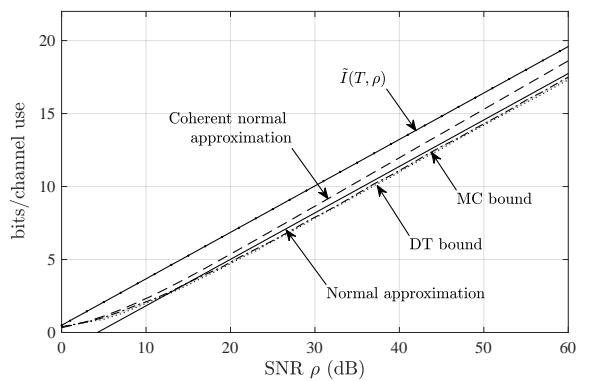
(b) $\rho = 25$ dB, $T = 24$, and $L = 7$.

Fig. 3. $R^*(L, T, \epsilon, \rho)$ as a function of ϵ .

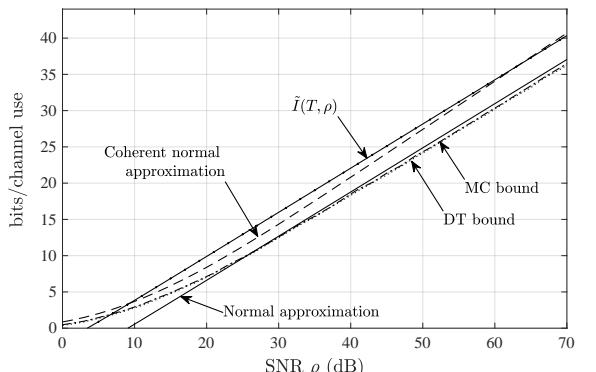
between the coherent and the noncoherent normal approximations appears to be independent of L . This agrees with the intuition that the cost for estimating the channel is determined by the coherence interval T .

In Figs. 3a and 3b, we study $R^*(L, T, \epsilon, \rho)$ as a function of ϵ for $T = 24$ and $L = 7$. Specifically, Fig. 3a plots $R^*(L, T, \epsilon, \rho)$ for $n_t = 2, n_r = 2$, and two different SNR values 20 dB and 25 dB. Fig. 3b plots $R^*(L, T, \epsilon, \rho)$ for 25 dB and three different numbers of antennas $(n_t, n_r) = (1, 1), (1, 2)$, and $(2, 2)$. Observe that the accuracy of normal approximation (8) increases as the SNR value and the number of antennas become larger, and for $\rho = 25$ dB and $(n_t, n_r) = (2, 2)$ it is very close to the nonasymptotic bounds over the entire range of error probabilities considered.

In Figs. 4a and 4b, we study $R^*(L, T, \epsilon, \rho)$ as a function of the SNR for $T = 24$, $L = 7$, $\epsilon = 10^{-5}$, and $(n_t, n_r) = (1, 2)$ and $(2, 2)$, respectively. For comparison, we also show the high-SNR approximation $\tilde{I}(T, \rho)$ of channel capacity. First observe that the DT lower bound on $R^*(L, T, \epsilon, \rho)$ (which is based on USTM channel inputs) is close to the MC upper bound (which is valid for any input distribution satisfying the power constraint (5)). Thus, USTM channel inputs are nearly capacity-achieving for all SNR values considered. Further observe that the SNR range over which the normal approximation (8) is accurate depends on the number of antennas. Specifically, when $n_t = 1$ and $n_r = 2$, the normal approximation is accurate for SNR values above $\rho = 12$ dB, whereas when $n_t = 2$ and $n_r = 2$, it is accurate for



(a) $T = 24$, $L = 7$, $\epsilon = 10^{-5}$, $n_t = 1$, and $n_r = 2$.



(b) $T = 24$, $L = 7$, $\epsilon = 10^{-5}$, $n_t = 2$, and $n_r = 2$.

Fig. 4. $R^*(L, T, \epsilon, \rho)$ as a function of ρ .

SNR values above $\rho = 23$ dB. As expected, the normal approximation of the coherent channel is strictly larger than both the high-SNR normal approximation of the noncoherent channel and the nonasymptotic bounds. However, its gap to the nonasymptotic bounds diminishes as ρ becomes small. Intuitively, this is because, as ρ decreases, knowledge of the fading coefficients becomes less important.

In Fig. 5, we plot the error probability as a function of the SNR for a fixed rate $R = 4$ and for $T = 24$, $L = 7$, and the numbers of antennas $(n_t, n_r) = (1, 1), (1, 2), (1, 4)$, and $(2, 4)$. As in the previous figures, the plotted DT bounds are based on USTM channel inputs, whereas the plotted MC bounds indicate the smallest probability of error that can be achieved by channel inputs satisfying the power constraint (5). Observe that the error-probability curves become steeper as the number of transmit and receive antennas increases. This is consistent with the diversity-multiplexing tradeoff (DMT) proposed by Zheng and Tse [34]. Indeed, let the *multiplexing gain* r and the *diversity gain* d be defined as

$$r \triangleq \lim_{\rho \rightarrow \infty} \frac{R^*(L, T, \epsilon, \rho)}{\log \rho} \quad (18a)$$

$$d \triangleq -\lim_{\rho \rightarrow \infty} \frac{\log \epsilon}{\log \rho}. \quad (18b)$$

Intuitively, if a system has a diversity gain of d , then the corresponding error probability decays as ρ^{-d} with the SNR ρ . A steeper error probability curve thus corresponds to a higher diversity gain. For the noncoherent Rayleigh block-

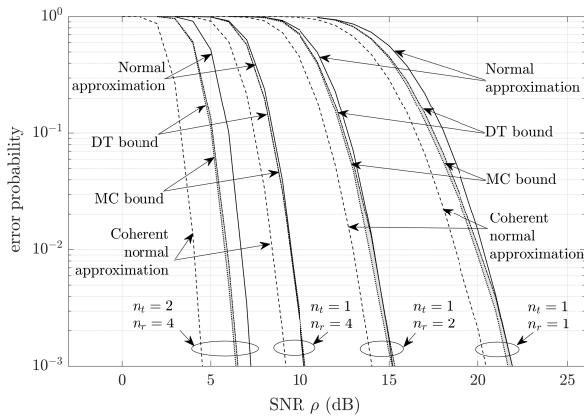


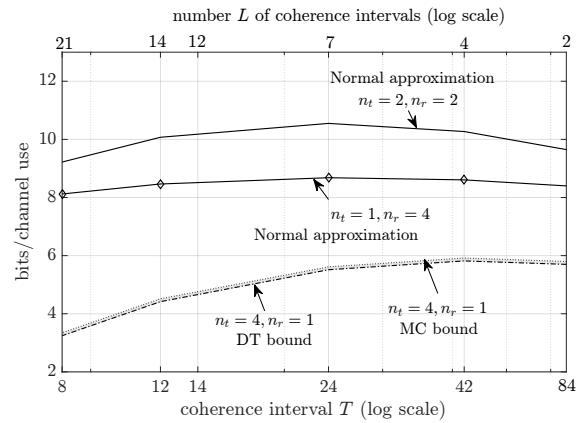
Fig. 5. The minimum error probability ϵ as a function of the SNR for $R = 4$, $T = 24$, and $L = 7$.

fading channel, the DMT is given by the piecewise linear curve joining the points [35]

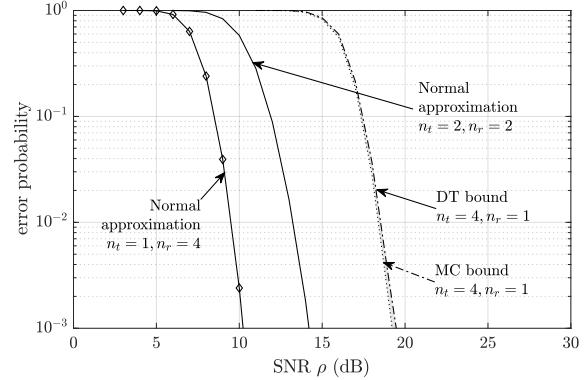
$$\left(r = \left(1 - \frac{n_*}{T} \right) k, d = L(n_t - k)(n_r - k) \right) \quad (19)$$

for $k = 0, \dots, \min(n_t, n_r)$, where n_* is as in (4). If the multiplexing gain r (or, equivalently, k) is held fixed, then (19) implies that the diversity gain grows with the number of transmit and receive antennas. This is the setting considered in Fig. 5, where we assume a fixed rate, and hence also a fixed multiplexing gain. Our observation that the slope of ϵ becomes steeper as the number of transmit and receive antennas increases is thus supported theoretically by the DMT. Further observe that the normal approximation (8) is accurate when the number of transmit antennas is $n_t = 1$, but it is overly-pessimistic when $n_t = 2$. In contrast, the coherent normal approximation is overly-optimistic for all parameters considered in this figure. Last but not least, observe that the error probability decreases significantly as the number of antennas increases. This demonstrates the benefit of multiple antennas at the transmitter and receiver at short blocklengths.

Finally, we study the impact of antenna allocation on the maximum coding rate and the error probability when the total number of antennas is equal to 4. To this end, we plot in Fig. 6a the maximum coding rate $R^*(L, T, \epsilon, \rho)$ as a function of T for a fixed blocklength $n = 168$ for $\rho = 25$ dB, error probability $\epsilon = 10^{-3}$, and the numbers of antennas $(n_t, n_r) = (1, 4)$, $(2, 2)$, and $(4, 1)$. Similarly, we plot in Fig. 6b the minimum error probability ϵ as a function of the SNR for $R = 4$, $T = 24$, $L = 7$, and the same numbers of antennas. For the cases $(n_t, n_r) = (1, 4)$ and $(n_t, n_r) = (2, 2)$, we plot the normal approximation (8), whereas for the case $(n_t, n_r) = (4, 1)$ we plot the DT and the MC bounds, as the case $n_t > n_r$ is not covered by Theorem 1. Observe that, for a fixed error probability, the coding rate is maximized by allocating the same number of antennas to the transmitter and receiver. This corresponds to the case where the transmit and receive antennas are used to maximize spatial multiplexing. In contrast, for a fixed rate, the error probability is minimized by maximizing the number of receive antennas. Intuitively, all cases exhibit the same diversity order $n_t n_r$, but the smallest



(a) $R^*(L, T, \epsilon, \rho)$ as a function of T for a fixed $TL = 168$, $\epsilon = 10^{-3}$, and $\rho = 25$ dB.



(b) Error probability ϵ as a function of the SNR for $R = 4$, $T = 24$, and $L = 7$.

Fig. 6. Maximum coding rate and minimum error probability for $(n_t, n_r) = (1, 4)$, $(2, 2)$, and $(4, 1)$.

number of transmit antennas results in the smallest cost for estimating the fading matrix \mathbf{H}_ℓ . For example, estimating \mathbf{H}_ℓ by means of pilot symbols would require one pilot symbol per coherence interval and transmit antenna.

It is well-known that normal approximations are accurate for sufficiently large blocklengths and moderate error probabilities; see also the discussion in [26, Sec. VIII-B]. The same applies to the high-SNR normal approximation presented in Theorem 1. Specifically, comparing the high-SNR normal approximation (8) against the nonasymptotic bounds, we observe that, for $T = 24$ and $L \geq 7$, the normal approximation deviates from the true value of $R^*(L, T, \epsilon, \rho)$ by less than 5% for SNR values $\rho \geq 25$ dB and error probabilities $\epsilon \geq 10^{-5}$, and for SNR values $\rho \geq 21$ dB and error probabilities $\epsilon \geq 10^{-3}$, irrespective of the number of transmit and receive antennas.

C. Optimal Number of Active Transmit Antennas

In [28], Zheng and Tse showed that, at high SNR, the channel capacity of the noncoherent Rayleigh block-fading channel behaves as

$$C(\rho) = n_* \left(1 - \frac{n_*}{T} \right) \log \rho + \mathcal{O}_\rho(1) \quad (20)$$

where n_* is as in (4). If $T \geq n_t + n_r$ and $n_t \leq n_r$, then $n_* = n_t$ and the pre-log factor $n_t(1 - n_t/T)$ is monotonically

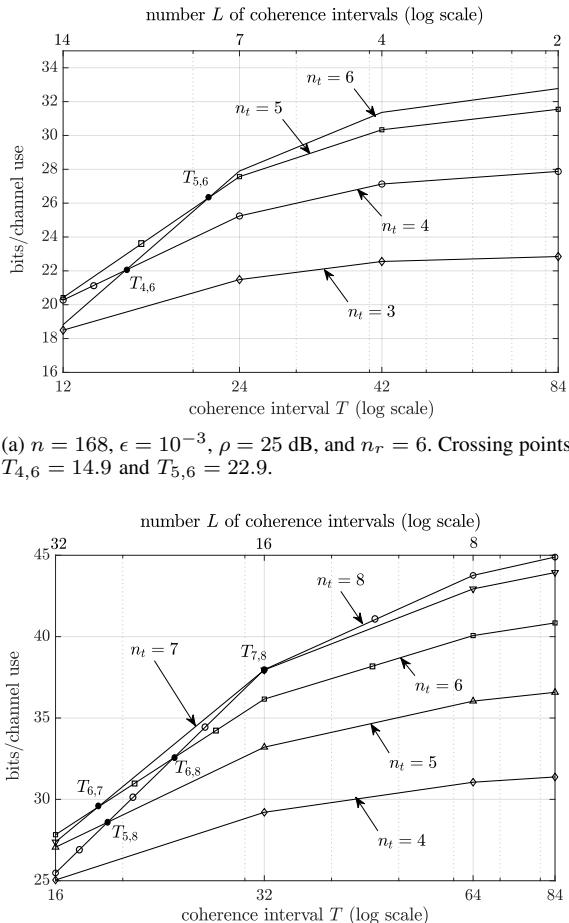


Fig. 7. High-SNR normal approximation as a function of T for a fixed $n = TL$ and varying numbers of transmit antennas.

increasing in n_t . In this case, it is optimal to use all available transmit antennas. However, it is *prima facie* unclear whether the same is true at finite blocklength. To gain some insights on this question, we analyze the behavior of the high-SNR normal approximation (8) as a function of the number of transmit antennas n_t for a given number of receive antennas n_r .

In Figs. 7a and 7b, we plot the high-SNR normal approximation as a function of T for a fixed blocklength $n = LT$, a given number of receive antennas, and varying numbers of transmit antennas. We consider the cases $\rho = 25$ dB, $\epsilon = 10^{-3}$, $n = 168$, and $n_r = 6$ (Fig. 7a), and $\rho = 25$ dB, $\epsilon = 10^{-5}$, $n = 512$, and $n_r = 8$ (Fig. 7b). Comparing the high-SNR normal approximation against the nonasymptotic bounds, it can be observed that, for the considered parameters, the normal approximation deviates from the true value of $R^*(L, T, \epsilon, \rho)$ by less than 5.2%, hence it is a good proxy for the maximum coding rate. We further indicate the values of the coherence interval where two lines cross. Specifically, we use the notation T_{n_1, n_2} to indicate the crossing point of the curves for $n_t = n_1$ and $n_t = n_2$. Observe that, in contrast to the pre-log factor of the high-SNR asymptotic capacity, the number of transmit antennas that maximizes the high-SNR normal approximation is not necessarily equal to the maximum

value $n_t = n_r$ and depends on the coherence interval T . For example, when $n = 168$ and $n_r = 6$, the maximum number of available transmit antennas should only be used when $T \geq 23$; cf. Fig. 7a. Similarly, when $n = 512$ and $n_r = 8$, using all available transmit antennas is suboptimal when $T < 32$; cf. Fig. 7b. Finally, the optimal number of active transmit antennas is not necessarily monotonically increasing in T . As can be observed from Fig. 7b, when $n = 512$ and $n_r = 8$, using $n_t = 7$ transmit antennas is optimal when $18 \leq T \leq 32$.

In summary, while the high-SNR asymptotic capacity suggests that, at large blocklengths and high SNR, it is optimal to use all available transmit antennas, in general the maximum coding rate has a more intricate dependence on the number of active transmit antennas. The high-SNR normal approximation presented in Theorem 1 allows us to unveil this dependence without having to resort to nonasymptotic bounds that need to be evaluated numerically at a high computation cost.

IV. PROOF OF THEOREM 1

To prove Theorem 1, we derive in Section IV-A a lower bound on $R^*(L, T, \epsilon, \rho)$ and in Section IV-B an upper bound on $R^*(L, T, \epsilon, \rho)$. Since both bounds are equal to (8) up to error terms that have the same behavior as $K_{\bar{I}}(T, \rho)$, $K_{\bar{V}}(T, \rho)$, and $K_L(L, T, \rho)$ (cf. (43) and (125)), we conclude that $R^*(L, T, \epsilon, \rho)$ is given by (8).

A. Lower Bound on $R^*(L, T, \epsilon, \rho)$

To derive a lower bound on $R^*(L, T, \epsilon, \rho)$, we evaluate the DT lower bound [3, Th. 22] for an USTM input distribution. Such an input can be written as $\mathbf{X}^L = \sqrt{T\rho/n_t} \mathbf{U}^L$, where $\mathbf{U}^L = [\mathbf{U}_1, \dots, \mathbf{U}_L]$ is a sequence of i.i.d., isotropically distributed, random matrices satisfying $\mathbf{U}_\ell^\mathsf{H} \mathbf{U}_\ell = \mathbf{I}_{n_t}$, $\ell = 1, \dots, L$. We define the *information density* between the random vectors \mathbf{X}^L and \mathbf{Y}^L as

$$i(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \log \frac{f_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L|\mathbf{X}^L)}{f_{\mathbf{Y}^L}(\mathbf{Y}^L)} \quad (21)$$

where $f_{\mathbf{Y}^L|\mathbf{X}^L}$ denotes the conditional probability density function (pdf) of the outputs \mathbf{Y}^L of the channel (3) given the inputs \mathbf{X}^L , and $f_{\mathbf{Y}^L}$ is the output pdf induced by the input distribution and the channel law. When the input distribution is USTM, the channel outputs \mathbf{Y}^L are i.i.d., $(T \times n_r)$ -dimensional random matrices whose joint pdf is given by

$$f_{\mathbf{Y}^L}(\mathbf{Y}^L) = \prod_{\ell=1}^L f_{\mathbf{Y}}^{(U)}(\mathbf{Y}_\ell), \quad \mathbf{Y}^L = [\mathbf{Y}_1, \dots, \mathbf{Y}_L] \quad (22)$$

where $f_{\mathbf{Y}}^{(U)}$ denotes the pdf of the outputs of the channel (3) induced by USTM channel inputs. Since the channel is blockwise memoryless, the information density can then be expressed as

$$i(\mathbf{X}^L; \mathbf{Y}^L) = \sum_{\ell=1}^L i_\ell^{(U)}(\mathbf{X}_\ell; \mathbf{Y}_\ell) \quad (23)$$

where the *USTM information density* $i_\ell^{(U)}(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ is defined as

$$i_\ell^{(U)}(\mathbf{X}_\ell; \mathbf{Y}_\ell) \triangleq \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_\ell|\mathbf{X}_\ell)}{f_{\mathbf{Y}}^{(U)}(\mathbf{Y}_\ell)} \quad (24)$$

and $f_{\mathbf{Y}|\mathbf{X}}$ denotes the conditional pdf of the channel output \mathbf{Y}_ℓ given the corresponding channel input \mathbf{X}_ℓ .

We next note that the cumulative distribution function $\Pr[i(\mathbf{x}^L; \mathbf{Y}^L) \leq \alpha]$ for (23) and \mathbf{Y}^L distributed according to (22) does not depend on \mathbf{x}^L . This follows from the isotropy of USTM inputs combined with the unitary invariance of the output distribution [24, App. A]. Furthermore, USTM inputs satisfy the power constraint (5) with probability one. Under these assumptions, the DT bound can be expressed as follows:

Lemma 1 (Dependence-Testing Bound): Fix an arbitrary distribution on \mathbf{X}^L . Assume that the cumulative distribution function $\Pr[i(\mathbf{x}^L; \mathbf{Y}^L) \leq \alpha]$ does not depend on \mathbf{x}^L when \mathbf{Y}^L is distributed according to $f_{\mathbf{Y}^L}$, and that \mathbf{X}^L satisfies the power constraint (5) with probability one. Then, there exists a code of blocklength $n = LT$ with M codewords whose maximum error probability satisfies

$$\epsilon \leq \mathbb{E} [\exp \{-[i(\mathbf{X}^L; \mathbf{Y}^L) - \log(M-1)]^+\}]. \quad (25)$$

Proof: The lemma follows by optimizing the bound in [3, Th. 23] for the case where $\Pr[i(\mathbf{x}^L; \mathbf{Y}^L) \leq \alpha]$ does not depend on \mathbf{x}^L , as in the proof of [3, Th. 22]. ■

Distinguishing between the cases where $i(\mathbf{X}^L; \mathbf{Y}^L) > \log(M-1)$ and $i(\mathbf{X}^L; \mathbf{Y}^L) \leq \log(M-1)$, the inequality (25) in Lemma 1 can be written as

$$\epsilon \leq (M-1) \mathbb{E} [e^{-i(\mathbf{X}^L; \mathbf{Y}^L)} \mathbb{1} \{i(\mathbf{X}^L; \mathbf{Y}^L) > \log(M-1)\}] + \Pr[i(\mathbf{X}^L; \mathbf{Y}^L) \leq \log(M-1)]. \quad (26)$$

Consequently, if we can find an M such that the right-hand side (RHS) of (26) is upper-bounded by ϵ , then there exists an $(L, T, M, \epsilon, \rho)$ code with these parameters. By the definition of the maximum coding rate (7), we then obtain that $R^*(L, T, \epsilon, \rho) \geq \frac{\log M}{LT}$.

To find such an M , we closely follow [3, Eqs. (258)–(267)]. To this end, we will need the following auxiliary results. In the following, we shall omit the subscript ℓ where it is immaterial. We further define $I(T, \rho) \triangleq \mathbb{E}[i^{(U)}(\mathbf{X}; \mathbf{Y})]$ and $\underline{U}(T, \rho) \triangleq \text{Var}(i^{(U)}(\mathbf{X}; \mathbf{Y}))$. Recall that these quantities, as well as the third moment $\mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3]$, are computed for USTM channel inputs.

Lemma 2: At high SNR, $\underline{U}(T, \rho)$ can be approximated as

$$\underline{U}(T, \rho) = T^2 \tilde{V}(T) + K_{\underline{U}}(T, \rho) \quad (27)$$

where $\tilde{V}(T)$ is defined in (10) and $K_{\underline{U}}(T, \rho)$ is a function of T and ρ that satisfies

$$\lim_{\rho \rightarrow \infty} K_{\underline{U}}(T, \rho) = 0. \quad (28)$$

Proof: See Appendix I. ■

Lemma 3: There exists a sufficiently large ρ_0 such that

$$\sup_{\rho \geq \rho_0} \underline{U}(T, \rho) < \infty \quad (29a)$$

$$\sup_{\rho \geq \rho_0} \mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3] < \infty. \quad (29b)$$

Proof: See Appendix II. ■

Lemma 2 implies that there exists a sufficiently large ρ_0 such that

$$\underline{U}(T, \rho) \geq \frac{T^2}{2} \tilde{V}(T), \quad \rho \geq \rho_0. \quad (30)$$

Furthermore, by Lemma 3, there exist a sufficiently large ρ_0 and an $\underline{S}(T)$ that is independent of ρ such that

$$\mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3] \leq \underline{S}(T), \quad \rho \geq \rho_0. \quad (31)$$

Combining (30) and (31), it then follows that

$$6\mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3] \leq \frac{6\underline{S}(T)}{[\frac{T^2}{2} \tilde{V}(T)]^{3/2}} \triangleq \underline{B}(T). \quad (32)$$

for $\rho \geq \rho_0$. The first term on the RHS of (26) can be upper-bounded using the following lemma.

Lemma 4 (Polyanskiy-Poor-Verdú'10): Let Z_1, \dots, Z_L be independent random variables, $\sigma^2 = \sum_{\ell=1}^L \text{Var}(Z_\ell)$, and $T = \sum_{\ell=1}^L \mathbb{E}[|Z_\ell - \mathbb{E}[Z_\ell]|^3]$. Then, for any A ,

$$\begin{aligned} \mathbb{E} \left[\exp \left\{ - \sum_{\ell=1}^L Z_\ell \right\} \mathbb{1} \left\{ \sum_{\ell=1}^L Z_\ell > A \right\} \right] \\ \leq 2 \left(\frac{\log 2}{\sqrt{2\pi}} + \frac{12T}{\sigma^2} \right) \frac{1}{\sigma} \exp\{-A\}. \end{aligned} \quad (33)$$

Proof: See [3, Lemma 47]. ■

Applying Lemma 4 to the first term on the RHS of (26), we obtain that

$$\begin{aligned} (M-1) \mathbb{E} [e^{-i(\mathbf{X}^L; \mathbf{Y}^L)} \mathbb{1} \{i(\mathbf{X}^L; \mathbf{Y}^L) > \log(M-1)\}] \\ \leq 2 \left(\frac{\log 2}{\sqrt{2\pi}} + 2\underline{B}(T) \right) \frac{1}{\sqrt{L}}. \end{aligned} \quad (34)$$

We next use (32) together with the Berry-Esseen theorem to upper-bound the second term on the RHS of (26):

Lemma 5 (Berry-Esseen Theorem): Let Z_1, \dots, Z_L be independent random variables with $\mu_\ell = \mathbb{E}[Z_\ell]$, $\sigma^2 = \sum_{\ell=1}^L \text{Var}(Z_\ell)$, and $T = \sum_{\ell=1}^L \mathbb{E}[|Z_\ell - \mathbb{E}[Z_\ell]|^3]$. Then, for any $\lambda \in \mathbb{R}$,

$$\left| \mathbb{P} \left[\sum_{\ell=1}^L (Z_\ell - \mu_\ell) \geq \lambda \sigma \right] - Q(\lambda) \right| \leq \frac{6T}{\sigma^3}. \quad (35)$$

Proof: See [36, Ch. XVI.5]. ■

Let

$$\tau = Q^{-1} \left(\epsilon - \left(\frac{2 \log 2}{\sqrt{2\pi}} + 5\underline{B}(T) \right) \frac{1}{\sqrt{L}} \right) \quad (36a)$$

$$\log(M-1) = LI(T, \rho) - \tau \sqrt{L\underline{U}(T, \rho)}. \quad (36b)$$

Then, by setting in (35) $Z_\ell = i^{(U)}(\mathbf{X}_\ell; \mathbf{Y}_\ell)$, $\lambda \triangleq \tau$, $T \triangleq L\mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3]$, and $\sigma \triangleq \sqrt{L\underline{U}(T, \rho)}$, the Berry-Esseen theorem yields that

$$\begin{aligned} \Pr[i(\mathbf{X}^L; \mathbf{Y}^L) \leq \log(M-1)] \\ \leq Q(\tau) + \frac{6\mathbb{E}[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3]}{\sqrt{L}(\underline{U}(T, \rho))^{3/2}} \\ \leq \epsilon - 2 \left(\frac{\log 2}{\sqrt{2\pi}} + 2\underline{B}(T) \right) \frac{1}{\sqrt{L}} \end{aligned} \quad (37)$$

where the second step follows from (32) and the definition of τ . It can be checked that the sum of the RHSs of (37) and

³Note that T and σ defined in (35) are different from the definitions in this paper.

(34) is equal to ϵ . Consequently, there exists an $(L, T, M, \epsilon, \rho)$ code with M given in (36b) and we have

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\geq \frac{\log(M-1)}{LT} \\ &= \frac{I(T, \rho)}{T} - \sqrt{\frac{\underline{U}(T, \rho)}{LT^2}}\tau. \end{aligned} \quad (38)$$

A Taylor-series expansion of τ around ϵ yields that

$$|\tau - Q^{-1}(\epsilon)| \leq \frac{A}{\sqrt{L}}, \quad L \geq L_0 \quad (39)$$

for some constants A and L_0 that are independent of L and ρ . Furthermore, Lemma 2 shows that $\underline{U}(T, \rho)$ is uniformly bounded in $\rho \geq \rho_0$. Thus, combining (38) and (39), we obtain

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\geq \frac{I(T, \rho)}{T} - \sqrt{\frac{\underline{U}(T, \rho)}{LT^2}}Q^{-1}(\epsilon) + K_{DT}(L, T, \rho) \end{aligned} \quad (40)$$

where $K_{DT}(L, T, \rho)$ is a function of L, T , and ρ that satisfies

$$\sup_{\rho \geq \rho_0} |K_{DT}(L, T, \rho)| \leq \frac{A}{L}, \quad L \geq L_0 \quad (41)$$

for some constants (A, L_0, ρ_0) that are independent of (L, ρ) .

The expressions $I(T, \rho)/T$ and $\underline{U}(T, \rho)/T^2$ can be approximated at high SNR as follows:

$$\frac{I(T, \rho)}{T} = \tilde{I}(T, \rho) + \tilde{K}_I(T, \rho) \quad (42a)$$

$$\frac{\underline{U}(T, \rho)}{T^2} = \tilde{V}(T) + \tilde{K}_{\underline{U}}(T, \rho) \quad (42b)$$

where $\tilde{I}(T, \rho)$ and $\tilde{V}(T)$ are given in (9) and (10), and where $\tilde{K}_\xi(T, \rho), \xi \in \{I, \underline{U}\}$ are functions of T and ρ that both satisfy $\lim_{\rho \rightarrow \infty} \tilde{K}_\xi(T, \rho) = 0$. Here, (42a) follows from [29, Eq. (49)] and (42b) follows from Lemma 2. Combining (42a) and (42b) with (40), we obtain that

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\geq \tilde{I}(T, \rho) - \sqrt{\frac{\tilde{V}(T) + \tilde{K}_{\underline{U}}(T, \rho)}{L}}Q^{-1}(\epsilon) \\ &\quad + \tilde{K}_I(T, \rho) + K_{DT}(L, T, \rho). \end{aligned} \quad (43)$$

Note that $\tilde{K}_\xi(T, \rho), \xi \in \{I, \underline{U}\}$ satisfy (11) and $K_{DT}(L, T, \rho)$ satisfies (12).

B. Upper Bound on $R^*(L, T, \epsilon, \rho)$

Our upper bound on $R^*(L, T, \epsilon, \rho)$ is based on the MC bound:

Lemma 6 (Meta-Converse Bound): The cardinality of a codebook \mathcal{C} with codewords $(\mathbf{X}_1, \dots, \mathbf{X}_L)$ belonging to a set \mathcal{F} and maximum error probability not exceeding ϵ satisfies

$$\log |\mathcal{C}| \leq \sup_{\mathbf{X}^L \in \mathcal{F}} \log \frac{1}{\beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L})} \quad (44)$$

where $\beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L})$ denotes the minimum probability of error of a binary hypothesis test under hypothesis $q_{\mathbf{Y}^L}$ if the probability of error under hypothesis $f_{\mathbf{Y}^L|\mathbf{X}^L}$ does not exceed ϵ [3, Eq. (100)], and $q_{\mathbf{Y}^L}$ is an auxiliary pdf of \mathbf{Y}^L .

Proof: See [3, Th. 31]. \blacksquare

Characterizing $\beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L})$ asymptotically as $L \rightarrow \infty$ and $\rho \rightarrow \infty$ is intractable. To sidestep this problem, we use [3, Eq. (106)] to lower-bound

$$\begin{aligned} \beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L}) &\geq \Pr \left[\frac{f_{\mathbf{Y}^L|\mathbf{X}^L}(\mathbf{Y}^L|\mathbf{X}^L)}{q_{\mathbf{Y}^L}(\mathbf{Y}^L)} < \xi(\mathbf{X}^L) \mid \mathbf{X}^L = \mathbf{X}^L \right] - \epsilon \end{aligned} \quad (45)$$

for an arbitrary $\xi(\mathbf{X}^L) > 0$ which may depend on \mathbf{X}^L . Together with (44), this gives rise to the weakened MC bound

$$\begin{aligned} M &\leq \sup_{\mathbf{X}^L \in \mathcal{F}} \left\{ \log \xi(\mathbf{X}^L) \right. \\ &\quad \left. - \log \left(\Pr[j(\mathbf{X}^L; \mathbf{Y}^L) < \log \xi(\mathbf{X}^L) \mid \mathbf{X}^L = \mathbf{X}^L] - \epsilon \right) \right\} \end{aligned} \quad (46)$$

where

$$j(\mathbf{X}^L; \mathbf{Y}^L) \triangleq \sum_{\ell=1}^L \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_\ell|\mathbf{X}_\ell)}{q_{\mathbf{Y}}(\mathbf{Y}_\ell)} \quad (47)$$

is the so-called *mismatched information density*.⁴

In (44), we choose the auxiliary pdf $q_{\mathbf{Y}}$ that was chosen in [29, Sec. IV-A] to derive an upper bound on the channel capacity of noncoherent MIMO Rayleigh block-fading channels. That is, $q_{\mathbf{Y}^L} = q_{\mathbf{Y}} \times \dots \times q_{\mathbf{Y}}$ is a product distribution with

$$\begin{aligned} q_{\mathbf{Y}}(\mathbf{Y}) &= \frac{1}{\mu^{n_t n_r} \pi^{n_r T}} \frac{\Gamma_{n_t}(T)}{\Gamma_{n_t}(n_t)} \frac{e^{-\sum_{i=1}^{n_t} \frac{\sigma_i^2}{\mu}}}{\left(\prod_{i=1}^{n_t} \sigma_i^2 \right)^{T-2n_r+n_t}} \times \\ &\quad \times \frac{e^{-\sum_{i=n_t+1}^{n_r} \frac{\sigma_i^2}{\mu}}}{\prod_{i=1}^{n_t} \prod_{j=n_t+1}^{n_r} (\sigma_i^2 - \sigma_j^2)^2} \end{aligned} \quad (48)$$

where $\sigma_1 \geq \dots \geq \sigma_{n_r}$ are the ordered nonzero singular values of \mathbf{Y} and $\mu \triangleq T\rho/n_t$.

We next note that, for every $T \times T$ unitary matrix \mathbf{A} and every \mathbf{X} and \mathbf{Y} , [24, Eqs. (54) & (55)]

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_\ell|\mathbf{A}^H \mathbf{X}_\ell) = f_{\mathbf{Y}|\mathbf{X}}(\mathbf{A}\mathbf{Y}_\ell|\mathbf{X}_\ell) \quad (49a)$$

$$q_{\mathbf{Y}}(\mathbf{A}\mathbf{Y}) = q_{\mathbf{Y}}(\mathbf{Y}). \quad (49b)$$

It then follows from [37, Prop. 19] that $\beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L})$ does not change if we multiply the channel input \mathbf{X}_ℓ by a unitary matrix \mathbf{A}_ℓ . We further note that

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_\ell|\mathbf{X}_\ell) = \frac{e^{-\text{tr}(\mathbf{Y}_\ell^H (\mathbf{I}_T + \mathbf{X}_\ell \mathbf{X}_\ell^H)^{-1} \mathbf{Y}_\ell)}}{\pi^{n_r T} \det(\mathbf{I}_T + \mathbf{X}_\ell \mathbf{X}_\ell^H)^{n_r}} \quad (50)$$

and, hence, also $\beta_{1-\epsilon}(\mathbf{X}^L, q_{\mathbf{Y}^L})$ depends on \mathbf{X}_ℓ only via the product $\mathbf{X}_\ell \mathbf{X}_\ell^H$. By expressing \mathbf{X}_ℓ in terms of its singular value decomposition (SVD)

$$\mathbf{X}_\ell = \mathbf{U}_\ell \Sigma_\ell \mathbf{V}_\ell^H \quad (51)$$

(where \mathbf{U}_ℓ and \mathbf{V}_ℓ are unitary matrices of dimensions $T \times T$ and $n_t \times n_t$, respectively, and Σ_ℓ is a $T \times n_t$ rectangular diagonal matrix), and by setting $\mathbf{A}_\ell = \mathbf{U}_\ell$, we thus obtain that we can restrict ourselves without loss of optimality to channel inputs \mathbf{X}_ℓ that are given by a $T \times n_t$ rectangular diagonal matrix with diagonal entries $d_{1,\ell}, \dots, d_{n_t,\ell}$. Intuitively, $d_{i,\ell}^2$ denotes the power at transmit antenna $i = 1, \dots, n_t$ and coherence

⁴We use the word “mismatched” to indicate that the output distribution $q_{\mathbf{Y}}$ is not the one induced by the input distribution and the channel.

interval $\ell = 1, \dots, L$. Let $D_\ell \triangleq \text{diag}\{d_{1,\ell}, \dots, d_{n_t,\ell}\}$ and $\alpha_\ell \triangleq \text{tr}(D_\ell^2)/T$. Thus, D_ℓ is the diagonal square matrix that is composed of the first n_t columns and rows of X_ℓ , i.e., $D_\ell = [X_\ell]_{(1:n_t)^2}$ (recall that $[X_\ell]_{(1:n_t)^2}$ denotes the diagonal submatrix that is composed of the first n_t columns and rows of X_ℓ ; see also the notation subsection at the end of Section I). By the power constraint (5), we have for every ℓ that $0 \leq \alpha_\ell \leq \rho$.

Since directly maximizing the weakened MC bound over X^L is challenging, we wish to first derive an asymptotic expansion of the conditional cumulative distribution function of $j(X^L; Y^L)$ by resorting to the Berry-Esseen theorem (Lemma 5). Maximizing the resulting expansion over X^L is then feasible. However, this requires the absolute third central moment of $j(X^L; Y^L)$ divided by its variance to the power of $3/2$ to be bounded in $X^L \in \mathcal{F}$. Unfortunately, this is not the case if the norm of X^L is small, since the variance of $j(X^L; Y^L)$ vanishes as X^L tends to zero. To avoid this issue, we follow the idea in [19] and separate the codebook \mathcal{C} into two sub-codebooks \mathcal{C}_1 and \mathcal{C}_2 , where \mathcal{C}_1 contains all the codewords for which $d_{i,\ell}^2 > \bar{\delta}\rho$, $i = 1, \dots, n_t$ (with $\bar{\delta}$ defined in (54)) in at least half of the coherence intervals, and where \mathcal{C}_2 contains the remaining codewords. Clearly, if the maximum error probabilities of \mathcal{C}_1 and \mathcal{C}_2 are ϵ , then the maximum error probability of \mathcal{C} cannot be smaller than ϵ . For codewords in \mathcal{C}_1 , the variance of $j(X^L; Y^L)$ is bounded away from zero, so we can apply the Berry-Esseen theorem. For codewords in \mathcal{C}_2 , we apply Chebyshev's inequality, which is less precise but applies to all values of X^L . We then show that, as $L \rightarrow \infty$ and $\rho \rightarrow \infty$, the cardinality of the entire codebook \mathcal{C} is dominated by the cardinality of \mathcal{C}_1 , so $R^*(L, T, \epsilon, \rho)$ is asymptotically upper-bounded by the upper bound on the maximum coding rate of \mathcal{C}_1 and a less precise analysis for codewords in \mathcal{C}_2 is unproblematic.

In the following, we provide a detailed proof of the upper bound. We begin by dividing the codebook \mathcal{C} into the sub-codebooks \mathcal{C}_1 and \mathcal{C}_2 and upper-bound the cardinalities of both codebooks using the weakened MC bound (46)

$$\log |\mathcal{C}_i| \leq \sup_{X^L \in \mathcal{C}_i} \left\{ \log \xi(D^L) - \log \left(\Pr[j(X^L; Y^L) < \log \xi(D^L) \mid X^L = X^L] - \epsilon \right) \right\} \quad (52)$$

for $i = 1, 2$, where (with a slight abuse of notation) we replaced $\xi(X^L)$ in (46) by $\xi(D^L)$. Thus, $\xi(D^L) > 0$ is an arbitrary threshold which may depend on D^L and which we shall define later.

To define \mathcal{C}_1 and \mathcal{C}_2 mathematically, we first introduce the sets

$$\mathcal{D}_1 \triangleq \{D \in \mathcal{D}_{n_t} : d_i^2 > \bar{\delta}\rho, i = 1, \dots, n_t\}, \quad (53a)$$

$$\mathcal{D}_2 \triangleq \mathcal{D}_{n_t} \setminus \mathcal{D}_1, \quad (53b)$$

where \mathcal{D}_{n_t} denotes the set of $(n_t \times n_t)$ -dimensional diagonal matrices D with non-negative, real-valued entries that satisfy $\text{tr}(D^2) \leq T\rho$; and d_i denotes the i -th diagonal element of D . Further let

$$\bar{\delta} \triangleq \frac{T}{n_t} - \frac{T}{2n_r\sqrt{n_t}\sqrt{\mathbb{E}[\log \det(H_\ell H_\ell^H)] + 1}}. \quad (54)$$

This choice of $\bar{\delta}$ satisfies the power constraint $n_t\bar{\delta}\rho \leq T\rho$ and ensures that the Berry-Esseen ratio $\bar{B}(D^L, T, \rho)$ (see (70) below) is bounded. We further argue that $\bar{\delta} > 0$ by evaluating $\mathbb{E}[(\log \det(H_\ell H_\ell^H))^2]$ using the following lemma.

Lemma 7 (Wishart matrices): Let \mathbf{W} be a Wishart matrix with distribution $\mathcal{W}(m, n)$, $m \leq n$. Then

$$\mathbb{E}[\log \det \mathbf{W}] = \sum_{i=0}^{m-1} \Psi(n-i) \quad (55a)$$

$$\text{Var}(\log \det \mathbf{W}) = \sum_{i=0}^{m-1} \Psi'(n-i). \quad (55b)$$

Proof: See [31, Lemma A.2]. \blacksquare

Lemma 7 yields that

$$\begin{aligned} \mathbb{E}[(\log \det(H_\ell H_\ell^H))^2] \\ = \left(\sum_{i=0}^{n_t-1} \Psi(n_t-i) \right)^2 + \sum_{i=0}^{n_t-1} \Psi'(n_t-i). \end{aligned} \quad (56)$$

Since $n_t \leq n_r$ (by assumption) and $\Psi'(1) = \pi^2/6$ [38, Th. 2.11], it follows that $\bar{\delta} > 0$.

For a sequence of matrices $D^L = (D_1, \dots, D_L)$, let

$$L_{\mathcal{D}_1}(D^L) \triangleq \sum_{\ell=1}^L \mathbb{1}\{D_\ell \in \mathcal{D}_1\}. \quad (57)$$

Then, we define the two sub-codebooks \mathcal{C}_1 and \mathcal{C}_2 as

$$\mathcal{C}_1 \triangleq \{X^L \in \mathcal{C} : L_{\mathcal{D}_1}(D^L) \geq L/2, D_\ell = [X_\ell]_{(1:n_t)^2}\} \quad (58a)$$

$$\mathcal{C}_2 \triangleq \{X^L \in \mathcal{C} : L_{\mathcal{D}_1}(D^L) < L/2, D_\ell = [X_\ell]_{(1:n_t)^2}\}. \quad (58b)$$

In words, the sub-codebook \mathcal{C}_1 contains all codewords X^L for which the diagonal entries satisfy $d_{i,\ell}^2 > \bar{\delta}\rho$ for all i and at least half of the time instants ℓ . We next derive upper bounds on the cardinalities of \mathcal{C}_1 and \mathcal{C}_2 .

1) *Upper bound on $\log |\mathcal{C}_1|$:* An upper bound on $\log |\mathcal{C}_1|$ follows from the weakened MC bound (52). To obtain a tractable expression, we further upper-bound (52) by lower-bounding the conditional probability. Indeed, as argued above, we can assume without loss of optimality that X_ℓ is given by a $T \times n_t$ rectangular diagonal matrix with diagonal entries $d_{1,\ell}, \dots, d_{n_t,\ell}$. By substituting $q_Y(Y)$ in (48) and $f_{Y|X}(Y_\ell | X_\ell)$ in (50) into (47), we then obtain that

$$j(X^L; Y^L) = \sum_{\ell=1}^L j_\ell(X_\ell; Y_\ell) \quad (59)$$

where

$$\begin{aligned} j_\ell(X_\ell; Y_\ell) \\ = n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \frac{n_t}{T\rho} \sum_{i=1}^{n_t} \sigma_{i,\ell}^2 + \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 \\ - \text{tr}(Y_\ell^H (I_T + X_\ell X_\ell^H)^{-1} Y_\ell) - n_r \log \det(I_{n_t} + X_\ell X_\ell^H) \\ + (T - 2n_r + n_t) \log \left(\prod_{i=1}^{n_t} \sigma_{i,\ell}^2 \right) \\ + 2 \log \left(\prod_{i=1}^{n_t} \prod_{j=n_t+1}^{n_r} (\sigma_{i,\ell}^2 - \sigma_{j,\ell}^2) \right). \end{aligned} \quad (60)$$

In (60), $\sigma_\ell = (\sigma_{1,\ell}, \dots, \sigma_{n_r,\ell})$ are the ordered singular values of \mathbf{Y}_ℓ .

We show in Appendix III that $j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ can be upper-bounded by a term that, conditioned on $\mathbf{X}_\ell = \mathbf{X}_\ell$, has the same distribution as

$$\begin{aligned} & \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \\ & \triangleq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \sum_{i=1}^{n_t} \left(\frac{n_t + n_t d_{i,\ell}^2}{T\rho} - 1 \right) Z'_{i,\ell} \\ & - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} Z''_{i,\ell} + (T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \\ & + (T - n_t) \log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2)^{-1}) \end{aligned} \quad (61)$$

where $\mathbf{Z}'_\ell = \{Z'_{1,\ell}, \dots, Z'_{n_t,\ell}\}$ and $\mathbf{Z}''_\ell = \{Z''_{1,\ell}, \dots, Z''_{n_t,\ell}\}$ are both sequences of i.i.d. random variables, the former distributed according to the gamma distribution $\Gamma(n_r, 1)$ and satisfying $\sum_{i=1}^{n_t} Z'_{i,\ell} = \text{tr}(\mathbf{H}_\ell^\mathsf{H} \mathbf{H}_\ell)$, and the latter distributed according to the gamma distribution $\Gamma(T - n_t, 1)$; \mathbf{Q}_ℓ is an $(T - n_t) \times n_r$ random matrices with i.i.d. $\mathcal{CN}(0, 1)$ entries that are independent of \mathbf{H}_ℓ ; and $\lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)$ denotes the largest eigenvalue of $\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell$. Furthermore, the sequences $\{\mathbf{Z}'_1, \dots, \mathbf{Z}'_L\}$, $\{\mathbf{Z}''_1, \dots, \mathbf{Z}''_L\}$, and $\{\mathbf{Q}_1, \dots, \mathbf{Q}_L\}$ are all i.i.d., and $\{\mathbf{Z}''_1, \dots, \mathbf{Z}''_L\}$ and $\{\mathbf{H}^L, \mathbf{Z}'_1, \dots, \mathbf{Z}'_L\}$ are independent of each other. Since $\bar{j}_\ell(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$ depends on \mathbf{X}^L only through \mathbf{D}^L , it then follows that

$$\begin{aligned} & \Pr \left[\sum_{\ell=1}^L j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell) < \log \xi(\mathbf{D}^L) \mid \mathbf{X}^L = \mathbf{X}^L \right] \\ & \geq \Pr \left[\sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) < \log \xi(\mathbf{D}^L) \right]. \end{aligned} \quad (62)$$

For future reference, the expected value of $\bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$ is given by

$$\begin{aligned} & \bar{j}(\mathbf{D}_\ell, T, \rho) \\ & = n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \sum_{i=1}^{n_t} \left(\frac{n_t + n_t d_{i,\ell}^2}{T\rho} - 1 \right) n_r \\ & - \frac{T\rho - n_t}{T\rho} n_t (T - n_t) + (T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \\ & + (T - n_t) \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2)^{-1})] \end{aligned} \quad (63)$$

since $Z'_{i,\ell}$ and $Z''_{i,\ell}$ are gamma distributed with means $\mathbb{E}[Z'_{i,\ell}] = n_r$ and $\mathbb{E}[Z''_{i,\ell}] = T - n_t$, respectively.

Combining (62) with (52), we can upper-bound the cardinality of the sub-codebook \mathcal{C}_1 as

$$\begin{aligned} \log |\mathcal{C}_1| & \leq \sup_{\mathbf{X}^L \in \mathcal{C}_1} \left\{ \log \xi(\mathbf{D}^L) \right. \\ & \left. - \log \left(\Pr \left[\sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) < \log \xi(\mathbf{D}^L) \right] - \epsilon \right) \right\} \end{aligned} \quad (64)$$

for every $\xi : \mathcal{D}_{n_t}^L \rightarrow (0, \infty)$. We next perform an asymptotic analysis of (64) based on the Berry-Esseen theorem. To this end, we need the following auxiliary lemmas:

Lemma 8: Let

$$\bar{U}(\mathbf{D}, T, \rho) \triangleq \text{Var}(\bar{j}(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)) \quad (65a)$$

$$\bar{S}(\mathbf{D}, T, \rho) \triangleq \mathbb{E}[\bar{j}(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \bar{J}(\mathbf{D}, T, \rho)]^3 \quad (65b)$$

for $\mathbf{D} \in \mathcal{D}_{n_t}$. Then, we have for every $T \geq n_r + n_t$ and any positive ρ_0

$$\sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{U}(\mathbf{D}, T, \rho) < \infty \quad (66a)$$

$$\sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{S}(\mathbf{D}, T, \rho) < \infty. \quad (66b)$$

Proof: See Appendix IV. \blacksquare

Lemma 9: For every $\mathbf{D} \in \mathcal{D}_1$, we have

$$\bar{U}(\mathbf{D}, T, \rho) \geq (T - n_t) n_t + K_{\bar{U}}(\mathbf{D}, T, \rho) \quad (67)$$

where $K_{\bar{U}}(\mathbf{D}, T, \rho)$ is a function of \mathbf{D} , T , and ρ that satisfies

$$\lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathbf{D}, T, \rho)| = 0.$$

Proof: See Appendix V. \blacksquare

Lemma 8 implies that there exists an $\bar{S}(T)$ that only depends on T and satisfies

$$\bar{S}(\mathbf{D}, T, \rho) \leq \bar{S}(T), \quad \mathbf{D} \in \mathcal{D}_{n_t}, \rho > \rho_0. \quad (68)$$

Furthermore, for $\rho \geq \rho_0$ and sufficiently large ρ_0 , we have $|K_{\bar{U}}(\mathbf{D}, T, \rho)| \leq (T - n_t) n_t / 2$, $\mathbf{D} \in \mathcal{D}_1$, by the definition of $K_{\bar{U}}(\mathbf{D}, T, \rho)$. It follows from Lemma 9 that, for $\rho \geq \rho_0$ and $\mathbf{X}^L \in \mathcal{C}_1$,

$$\sum_{\ell=1}^L \bar{U}(\mathbf{D}_\ell, T, \rho) \geq \frac{L}{2} \frac{(T - n_t) n_t}{2} \quad (69)$$

since all codewords in \mathcal{C}_1 satisfy $L_{\mathcal{D}_1}(\mathbf{D}^L) \geq L/2$. Combining these two results, we can upper-bound the ratio

$$\bar{B}(\mathbf{D}^L, T, \rho) \triangleq \frac{6 \sum_{\ell=1}^L \bar{S}(\mathbf{D}_\ell, T, \rho)}{\left(\sum_{\ell=1}^L \bar{U}(\mathbf{D}_\ell, T, \rho) \right)^{3/2}} \quad (70)$$

for $\mathbf{X}^L \in \mathcal{C}_1$ and $\mathbf{D}_\ell = [\mathbf{X}_\ell]_{(1:n_t)^2}$ as

$$\bar{B}(\mathbf{D}^L, T, \rho) \leq \frac{6 L \bar{S}(T)}{\left(\frac{L(T - n_t) n_t}{4} \right)^{3/2}} \triangleq \frac{\bar{B}(T)}{\sqrt{L}}. \quad (71)$$

With this, we are ready to apply the Berry-Esseen theorem (Lemma 5) to the upper bound (64). Let

$$\log \xi(\mathbf{D}^L) = \sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, T, \rho) - \lambda \sqrt{\sum_{\ell=1}^L \bar{U}(\mathbf{D}_\ell, T, \rho)} \quad (72a)$$

$$\lambda = Q^{-1} \left(\epsilon + \frac{2 \bar{B}(T)}{\sqrt{L}} \right), \quad 0 < \epsilon < 1/2. \quad (72b)$$

It then follows from the Berry-Esseen theorem that

$$\Pr \left[\sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) < \log \xi(\mathbf{D}^L) \right] \geq \epsilon + \frac{\bar{B}(T)}{\sqrt{L}}. \quad (73)$$

Substituting (72) and (73) into the upper bound (64), we obtain for $L \geq L_0$ and a sufficiently large L_0 that

$$\begin{aligned} \frac{\log |\mathcal{C}_1|}{L} &\leq \sup_{X^L \in \mathcal{C}_1} \left\{ \sum_{\ell=1}^L \frac{\bar{J}(D_\ell, T, \rho)}{L} - \frac{\lambda}{L} \sqrt{\sum_{\ell=1}^L \bar{U}(D_\ell, T, \rho)} \right\} \\ &\quad - \frac{\log \bar{B}(T)}{L} + \frac{\log L}{2L} \\ &\leq \sup_{X^L \in \mathcal{C}_1} \frac{1}{L} \sum_{\ell=1}^L \left\{ \bar{J}(D_\ell, T, \rho) - \lambda \sqrt{\frac{\bar{U}(D_\ell, T, \rho)}{L}} \right\} \\ &\quad - \frac{\log \bar{B}(T)}{L} + \frac{\log L}{2L} \end{aligned} \quad (74)$$

where the second inequality follows because, by the assumption $0 < \epsilon < 1/2$, λ is positive for sufficiently large L so, applying Jensen's inequality, the square-root of the sum $\sum_{\ell}^L \bar{U}(D_\ell, T, \rho)$ can be lower-bounded by the sum of square-roots $\sum_{\ell}^L \sqrt{\bar{U}(D_\ell, T, \rho)}$ divided by \sqrt{L} .

Performing a Taylor-series expansion of λ around ϵ , we obtain

$$|\lambda - Q^{-1}(\epsilon)| \leq \frac{A}{\sqrt{L}}, \quad L \geq L_0 \quad (75)$$

for some constants A , L_0 , and ρ_0 that are independent of L and ρ . Further using that, by Lemma 8, $\bar{U}(D_\ell, T, \rho)$ is bounded in $D_\ell \in \mathcal{D}_{n_t}$ and $\rho \geq \rho_0$, and collecting terms of order $\log L/L$, we can then rewrite the upper bound (74) as

$$\frac{\log |\mathcal{C}_1|}{L} \leq \frac{1}{L} \sup_{X^L \in \mathcal{C}_1} \sum_{\ell=1}^L \bar{R}_{\text{NA}}(D_\ell, L, T, \rho) + K_1(L, T, \rho) \quad (76)$$

where

$$\bar{R}_{\text{NA}}(D, L, T, \rho) \triangleq \bar{J}(D, T, \rho) - \sqrt{\frac{\bar{U}(D, T, \rho)}{L}} Q^{-1}(\epsilon) \quad (77)$$

and where $K_1(L, T, \rho)$ is a function of L , T , and ρ that satisfies

$$\sup_{\rho \geq \rho_0} |K_1(L, T, \rho)| \leq A \frac{\log L}{L}, \quad L \geq L_0 \quad (78)$$

for some constants (A, L_0, ρ_0) that are independent of (L, ρ) .

Since X_ℓ satisfies the power constraint $\|X_\ell\|^2 \leq T\rho$ for every ℓ , and since $\bar{R}_{\text{NA}}(D_\ell, L, T, \rho)$ depends on ℓ only via X_ℓ , the supremum on the RHS of (76) can be written as

$$\begin{aligned} &\sup_{D^L \in \mathcal{D}_{n_t}^L : L_{\mathcal{D}_1}(D^L) \geq L/2} \left\{ L_{\mathcal{D}_1}(D^L) \sup_{D \in \mathcal{D}_1} \bar{R}_{\text{NA}}(D, L, T, \rho) \right. \\ &\quad \left. + (L - L_{\mathcal{D}_1}(D^L)) \sup_{D \in \mathcal{D}_2} \bar{R}_{\text{NA}}(D, L, T, \rho) \right\}. \end{aligned} \quad (79)$$

We next upper-bound (79) using the following lemmas:

Lemma 10: For every $D \in \mathcal{D}_{n_t}$ satisfying $\text{tr}(D^2) \geq T\rho(1 - \delta)$ for some $0 < \delta \leq \frac{1}{2}$, we have

$$\bar{U}(D, T, \rho) \geq \bar{U} \left(\sqrt{\frac{\rho}{\alpha}} D, T, \rho \right) - \Upsilon(T)\delta, \quad \rho \geq \rho_0 \quad (80)$$

where $\Upsilon(T)$ is a positive constant that only depends on T and ρ_0 is an arbitrary positive constant. \blacksquare

Proof: See Appendix VI. \blacksquare

Lemma 11: Assume that $T \geq n_t + n_r$. For $\rho \geq \rho_0$ and $L \geq L_0$ and sufficiently large ρ_0 and L_0 ,

$$\begin{aligned} &\sup_{D \in \mathcal{D}_1} \bar{R}_{\text{NA}}(D, L, T, \rho) \\ &\leq \sup_{\substack{D \in \mathcal{D}_{n_t} : \\ \frac{\text{tr}(D^2)}{T\rho} \geq 1 - \frac{K(T)}{L}}} \left\{ \bar{J}^*(D, T, \rho) - \sqrt{\frac{\bar{U}(D, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ &\quad + K_{\mathcal{D}_1}(T, \rho) \end{aligned} \quad (81)$$

for some constant $K(T) \geq 0$ that only depends on T . Here,

$$\begin{aligned} &\bar{J}^*(D, T, \rho) \\ &\triangleq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \left(\frac{T\alpha - T\rho}{T\rho} \right) n_r n_t \\ &\quad - n_t(T - n_t) + (T - n_t - n_r) \log \det(I_{n_t} + D^2) \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)] \end{aligned} \quad (82)$$

and $K_{\mathcal{D}_1}(T, \rho)$ is a nonnegative function of T and ρ that satisfies $\lim_{\rho \rightarrow \infty} K_{\mathcal{D}_1}(T, \rho) = 0$.

Proof: See Appendix VII. \blacksquare

Lemma 12: Maximized over $D \in \mathcal{D}_1$, the term $\bar{J}^*(D, T, \rho)$ is given by

$$\sup_{D \in \mathcal{D}_1} \bar{J}^*(D, T, \rho) = \bar{J}_{\mathcal{D}_1}(T, \rho), \quad \rho > 0 \quad (83)$$

where

$$\begin{aligned} \bar{J}_{\mathcal{D}_1}(T, \rho) &\triangleq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} - n_t(T - n_t) \\ &\quad + (T - n_t - n_r) \log \left(1 + \frac{T\rho}{n_t} \right)^{n_t} \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)] \end{aligned} \quad (84)$$

Proof: See Appendix VIII. \blacksquare

Lemma 13: For every $D \in \mathcal{D}_1$, we have

$$\bar{U} \left(\sqrt{\frac{\rho}{\alpha}} D, T, \rho \right) \geq T^2 \tilde{V}(T) + K_{\bar{U}}(D, T, \rho) \quad (85)$$

where $\tilde{V}(T)$ was defined in (10) and $K_{\bar{U}}(D, T, \rho)$ is a function of D , T , and ρ which satisfies

$$\lim_{\rho \rightarrow \infty} \sup_{D \in \mathcal{D}_1} |K_{\bar{U}}(D, T, \rho)| = 0. \quad (86)$$

Proof: See Appendix IX. \blacksquare

Lemma 14: Maximized over $D \in \mathcal{D}_2$, the term $\bar{J}(D, T, \rho)$ can be bounded as

$$\bar{J}_{\mathcal{D}_2}(T, \rho) \leq \sup_{D \in \mathcal{D}_2} \bar{J}(D, T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho) \quad (87)$$

for $\rho \geq \rho_0$ and a sufficiently large $\rho_0 > 0$, where

$$\begin{aligned} &\bar{J}_{\mathcal{D}_2}(T, \rho) \\ &\triangleq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} - n_t(T - n_t) \\ &\quad + (T - n_t - n_r) \log \left((1 + \bar{\delta}\rho) \left(1 + \frac{T - \bar{\delta}}{n_t - 1} \rho \right)^{n_t - 1} \right) \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)] \end{aligned} \quad (88)$$

and $K_{\mathcal{D}_2}(T, \rho)$ is a nonnegative correction term that only depends on T and ρ and satisfies

$$\lim_{\rho \rightarrow \infty} K_{\mathcal{D}_2}(T, \rho) = 0. \quad (89)$$

Proof: See Appendix X. \blacksquare

We are now ready to upper-bound (79). Indeed, the first supremum on the RHS of (79) can be upper-bounded as

$$\begin{aligned} & \sup_{\mathcal{D} \in \mathcal{D}_1} \bar{R}_{\text{NA}}(\mathcal{D}, L, T, \rho) \\ & \leq \sup_{\substack{\mathcal{D} \in \mathcal{D}_{n_t}: \\ \frac{\text{tr}(\mathcal{D}^2)}{T\rho} \geq 1 - \frac{K(T)}{L}}} \left\{ \bar{J}^*(\mathcal{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathcal{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ & \quad + K_{\mathcal{D}_1}(T, \rho) \\ & \leq \max \left\{ \sup_{\mathcal{D} \in \mathcal{D}_2} \bar{J}(\mathcal{D}, T, \rho), \right. \\ & \quad \left. \sup_{\substack{\mathcal{D} \in \mathcal{D}_1: \\ \frac{\text{tr}(\mathcal{D}^2)}{T\rho} \geq 1 - \frac{K(T)}{L}}} \left\{ \bar{J}^*(\mathcal{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathcal{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \right\} \\ & \quad + K_{\mathcal{D}_1}(T, \rho) \end{aligned} \quad (90)$$

where the first inequality follows from Lemma 11; the second inequality follows by writing the supremum over \mathcal{D}_{n_t} as the maximum of the supremum over \mathcal{D}_1 and the supremum over \mathcal{D}_2 (because $\mathcal{D}_{n_t} = \mathcal{D}_1 \cup \mathcal{D}_2$ and $\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$), and by upper-bounding the latter supremum by removing the trace constraint and by lower-bounding $\bar{U}(\mathcal{D}, T, \rho) \geq 0$.

The supremum over \mathcal{D}_1 in (90) can be upper-bounded as

$$\begin{aligned} & \sup_{\substack{\mathcal{D} \in \mathcal{D}_1: \\ \frac{\text{tr}(\mathcal{D}^2)}{T\rho} \geq 1 - \frac{K(T)}{L}}} \left\{ \bar{J}^*(\mathcal{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathcal{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ & \leq \sup_{\substack{\mathcal{D} \in \mathcal{D}_1: \\ \frac{\text{tr}(\mathcal{D}^2)}{T\rho} \geq 1 - \frac{K(T)}{L}}} \left\{ \bar{J}^*(\mathcal{D}, T, \rho) \right. \\ & \quad \left. - \sqrt{\frac{\bar{U}(\sqrt{\frac{\rho}{\alpha}}\mathcal{D}, T, \rho) - \Upsilon(T)\frac{K(T)}{L}}{L}} Q^{-1}(\epsilon) \right\} \\ & \leq \sup_{\mathcal{D} \in \mathcal{D}_1} \bar{J}^*(\mathcal{D}, T, \rho) \\ & \quad - \sqrt{\frac{T^2\tilde{V}(T) - K_{\bar{U}}(T, \rho) - \Upsilon(T)\frac{K(T)}{L}}{L}} Q^{-1}(\epsilon) \\ & \leq \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(L, T) \end{aligned} \quad (91)$$

where

$$\begin{aligned} & \bar{R}_{\mathcal{D}_1}(L, T, \rho) \\ & \triangleq \bar{J}_{\mathcal{D}_1}(T, \rho) - \sqrt{\frac{T^2\tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (92)$$

and

$$K_{\bar{U}}(T, \rho) \triangleq \sup_{\mathcal{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathcal{D}, T, \rho)| \quad (93a)$$

$$K_{\mathcal{D}_1}(L, T) \triangleq \frac{\sqrt{\Upsilon(T)K(T)}}{L} Q^{-1}(\epsilon). \quad (93b)$$

In (91), the first inequality follows from Lemma 10; the second inequality follows from Lemma 13 and by optimizing $\bar{J}^*(\mathcal{D}, T, \rho)$ and the constant $K_{\bar{U}}(\mathcal{D}, T, \rho)$ in (85) over $\mathcal{D} \in \mathcal{D}_1$; the third inequality follows from Lemma 12 and the inequality $\sqrt{x-y} \geq \sqrt{x} - \sqrt{y}$, $x \geq y$.

We next use Lemma 14 to upper-bound the supremum over \mathcal{D}_2 in (90) as

$$\sup_{\mathcal{D} \in \mathcal{D}_2} \bar{J}(\mathcal{D}, T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho) \quad (94)$$

for $\rho \geq \rho_0$ and a sufficiently large $\rho_0 > 0$. Comparing $\bar{J}_{\mathcal{D}_2}(T, \rho)$ with $\bar{J}_{\mathcal{D}_1}(T, \rho)$, we obtain that

$$\begin{aligned} & \bar{J}_{\mathcal{D}_1}(T, \rho) - \bar{J}_{\mathcal{D}_2}(T, \rho) \\ & = (T - n_t - n_r) \log \left(\frac{\left(1 + \frac{T\rho}{n_t}\right)^{n_t}}{(1 + \bar{\delta}\rho) \left(1 + \frac{T\rho - \bar{\delta}\rho}{n_t - 1}\right)^{n_t - 1}} \right) \\ & \triangleq \Theta(T, \rho). \end{aligned} \quad (95)$$

In the limit as ρ tends to infinity, $\Theta(T, \rho)$ tends to

$$\Theta(T) \triangleq (T - n_t - n_r) \log \left(\frac{\left(\frac{T}{n_t}\right)^{n_t}}{\bar{\delta} \left(\frac{T - \bar{\delta}}{n_t - 1}\right)^{n_t - 1}} \right) \quad (96)$$

which is strictly positive since, by the geometric-arithmetic mean inequality,

$$\begin{aligned} \bar{\delta} \left(\frac{T - \bar{\delta}}{n_t - 1} \right)^{n_t - 1} & < \left(\frac{\bar{\delta} + \frac{T - \bar{\delta}}{n_t - 1} + \dots + \frac{T - \bar{\delta}}{n_t - 1}}{n_t} \right)^{n_t} \\ & = \left(\frac{T}{n_t} \right)^{n_t}. \end{aligned} \quad (97)$$

By the definitions of $\bar{R}_{\mathcal{D}_1}(L, T, \rho)$ and $\Theta(T, \rho)$, we have that

$$\begin{aligned} & \sup_{\mathcal{D} \in \mathcal{D}_2} \bar{J}(\mathcal{D}, T, \rho) - \bar{R}_{\mathcal{D}_1}(L, T, \rho) - K_{\mathcal{D}_1}(T, \rho) \\ & \leq K_{\mathcal{D}_2}(T, \rho) - \Theta(T, \rho) \\ & \quad + \sqrt{\frac{T^2\tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}} Q^{-1}(\epsilon) - K_{\mathcal{D}_1}(T, \rho). \end{aligned} \quad (98)$$

Since the first, second, and last terms on the RHS of (98) only depend on ρ , and the second-to-last term vanishes as $L \rightarrow \infty$ uniformly in ρ , we conclude that the double limit of the RHS of (98) as $\rho \rightarrow \infty$ and $L \rightarrow \infty$ exists and is equal to $-\Theta(T) < 0$. It follows that there exist ρ_0 and L_0 such that

$$\sup_{\mathcal{D} \in \mathcal{D}_2} \bar{J}(\mathcal{D}, T, \rho) < \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(T, \rho) \quad (99)$$

for $\rho \geq \rho_0$ and $L \geq L_0$. Combining (99) and (91) with (90), we obtain that

$$\sup_{\mathcal{D} \in \mathcal{D}_1} \bar{R}_{\text{NA}}(\mathcal{D}, L, T, \rho) \leq \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(L, T, \rho) \quad (100)$$

for $\rho \geq \rho_0$ and $L \geq L_0$ and sufficiently large (ρ_0, L_0) , where

$$K_{\mathcal{D}_1}(L, T, \rho) \triangleq K_{\mathcal{D}_1}(T, \rho) + K_{\mathcal{D}_1}(L, T). \quad (101)$$

We next show that the second supremum on the RHS of (79) does not exceed the RHS of (100). It follows then that our upper bound on (79) is maximized for $L_{\mathcal{D}_1}(\mathcal{D}^L) = L$. To

prove this claim, we use again Lemma 14 together with the nonnegativity of $\bar{U}(\mathbf{D}, T, \rho)$ to upper-bound

$$\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{R}_{\mathcal{D}_1}(\mathbf{D}, L, T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho). \quad (102)$$

It then follows from (102) and (95) that

$$\begin{aligned} & \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(L, T, \rho) - \sup_{\mathbf{D} \in \mathcal{D}_2} \bar{R}_{\mathcal{D}_1}(\mathbf{D}, L, T, \rho) \\ & \geq \Theta(T, \rho) - \sqrt{\frac{T^2 \tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}} Q^{-1}(\epsilon) \\ & \quad + K_{\mathcal{D}_1}(L, T, \rho) - K_{\mathcal{D}_2}(T, \rho). \end{aligned} \quad (103)$$

Using similar arguments as the ones leading to (99), we can show that the double limit of the RHS of (103) as $\rho \rightarrow \infty$ and $L \rightarrow \infty$ exists and is equal to $\Theta(T)$. Consequently, there exist (ρ_0, L_0) such that, for $\rho \geq \rho_0$ and $L \geq L_0$, the RHS of (103) is strictly positive. Thus, for $\rho \geq \rho_0$ and $L \geq L_0$,

$$\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{R}_{\mathcal{D}_1}(\mathbf{D}, L, T, \rho) < \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(L, T, \rho). \quad (104)$$

Combining (100)–(104) with (74) and (79), we obtain that

$$\begin{aligned} \frac{\log |\mathcal{C}_1|}{L} & \leq \bar{R}_{\mathcal{D}_1}(L, T, \rho) + K_{\mathcal{D}_1}(L, T, \rho) - \frac{\log \bar{B}(T)}{L} + \frac{\log L}{2L} \\ & = \bar{J}_{\mathcal{D}_1}(T, \rho) - \sqrt{\frac{T^2 \tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}} Q^{-1}(\epsilon) \\ & \quad + K_{\mathcal{D}_1}(T, \rho) + K_{\mathcal{C}_1}(L, T) \\ & \triangleq \bar{R}_{\mathcal{C}_1}(L, T, \rho) \end{aligned} \quad (105)$$

where $K_{\mathcal{C}_1}(L, T) \triangleq K_{\mathcal{D}_1}(L, T) - \frac{\log \bar{B}(T)}{L} + \frac{\log L}{2L}$ is a function of L and T that satisfies

$$|K_{\mathcal{C}_1}(L, T)| \leq A \frac{\log L}{L}, \quad L \geq L_0 \quad (106)$$

for some constants A and L_0 that are independent of L and ρ . Recall that $\lim_{\rho \rightarrow \infty} K_{\mathcal{D}_1}(T, \rho) = \lim_{\rho \rightarrow \infty} K_{\bar{U}}(T, \rho) = 0$.

2) *Upper bound on $\log |\mathcal{C}_2|$:* An upper bound on $\log |\mathcal{C}_2|$ follows from the weakened MC bound (52) and (62):

$$\begin{aligned} \log |\mathcal{C}_2| & \leq \sup_{\mathbf{X}^L \in \mathcal{C}_2} \left\{ \log \xi(\mathbf{D}^L) \right. \\ & \left. - \log \left(\Pr \left[\sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) < \log \xi(\mathbf{D}^L) \right] - \epsilon \right) \right\}. \end{aligned} \quad (107)$$

We next note that, by Lemma 8, there exists an $\bar{U}(T)$ that only depends on T and satisfies

$$\bar{U}(\mathbf{D}, T, \rho) \leq \bar{U}(T), \quad \mathbf{D} \in \mathcal{D}_{nt}, \rho > \rho_0. \quad (108)$$

If we choose

$$\log \xi(\mathbf{D}^L) = \sum_{\ell=1}^L \bar{J}_\ell(\mathbf{D}_\ell, T, \rho) + \sqrt{\frac{2L\bar{U}(T)}{1-\epsilon}} \quad (109)$$

then Chebyshev's inequality yields that

$$\Pr \left[\sum_{\ell=1}^L \bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \geq \log \xi(\mathbf{D}^L) \right] \leq \frac{1-\epsilon}{2} \quad (110)$$

and we can upper-bound (107) as

$$\begin{aligned} \log |\mathcal{C}_2| & \leq \sup_{\mathbf{X}^L \in \mathcal{C}_2} \sum_{\ell=1}^L \bar{J}(\mathbf{D}_\ell, T, \rho) + \sqrt{\frac{2L\bar{U}(T)}{1-\epsilon}} \\ & \quad + \log 2 - \log(1-\epsilon). \end{aligned} \quad (111)$$

It remains to upper-bound the supremum on the RHS of (111). Similar to (79), using that $\bar{J}_\ell(\mathbf{D}_\ell, T, \rho)$ depends on ℓ only via \mathbf{D}_ℓ , we obtain that

$$\begin{aligned} & \sup_{\mathbf{X}^L \in \mathcal{C}_2} \sum_{\ell=1}^L \bar{J}(\mathbf{D}_\ell, T, \rho) \\ & = \sup_{\mathbf{D}^L \in \mathcal{D}_{nt}^L : L_{\mathcal{D}_1}(\mathbf{D}^L) < L/2} \left\{ L_{\mathcal{D}_1}(\mathbf{D}^L) \sup_{\mathbf{D} \in \mathcal{D}_1} \bar{J}_\ell(\mathbf{D}_\ell, T, \rho) \right. \\ & \quad \left. + (L - L_{\mathcal{D}_1}(\mathbf{D}^L)) \sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}_\ell(\mathbf{D}_\ell, T, \rho) \right\} \\ & \leq \sup_{\mathbf{D}^L : L_{\mathcal{D}_1}(\mathbf{D}^L) < L/2} \left\{ L_{\mathcal{D}_1}(\mathbf{D}^L) (\bar{J}_{\mathcal{D}_1}(T, \rho) + K_{\mathcal{D}_1}(T, \rho)) \right. \\ & \quad \left. + (L - L_{\mathcal{D}_1}(\mathbf{D}^L)) (\bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho)) \right\} \\ & \leq L (\bar{J}_{\mathcal{D}_1}(T, \rho) + K_{\mathcal{D}_1}(T, \rho) + K_{\mathcal{D}_2}(T, \rho)) - \frac{L}{2} \Theta(T, \rho) \end{aligned} \quad (112)$$

where $K_{\mathcal{D}_1}(T, \rho)$ and $K_{\mathcal{D}_2}(T, \rho)$ are functions of T and ρ that vanish as $\rho \rightarrow \infty$. Here, the first inequality follows from (269) in the proof of Lemma 11 and from Lemmas 12 and 14; the second inequality follows from the definition of $\Theta(T, \rho)$ in (95) and because $0 \leq L_{\mathcal{D}_1}(\mathbf{D}^L) < L/2$.

Substituting (112) into (111), we obtain the desired upper bound

$$\begin{aligned} \frac{\log |\mathcal{C}_2|}{L} & \leq \bar{J}_{\mathcal{D}_1}(T, \rho) - \frac{1}{2} \Theta(T, \rho) + K_{\mathcal{D}_1}(T, \rho) \\ & \quad + K_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{C}_2}(L, T) \\ & \triangleq \bar{R}_{\mathcal{C}_2}(L, T, \rho) \end{aligned} \quad (113)$$

where

$$K_{\mathcal{C}_2}(L, T) \triangleq \sqrt{\frac{2\bar{U}(T)}{L(1-\epsilon)}} - \frac{1}{L} \log \left(\frac{1-\epsilon}{2} \right) \quad (114)$$

is a function of L and T that satisfies

$$|K_{\mathcal{C}_2}(L, T)| \leq \frac{A}{L}, \quad L \geq L_0 \quad (115)$$

for some constants A and L_0 that are independent of (L, ρ) .

3) *Upper bound on the cardinality of \mathcal{C} :* As shown in the previous subsections, the cardinalities of \mathcal{C}_1 and \mathcal{C}_2 can be upper-bounded by $\bar{R}_{\mathcal{C}_1}(L, T, \rho)$ and $\bar{R}_{\mathcal{C}_2}(L, T, \rho)$ (cf. (105) and (113)). Therefore, the cardinality of \mathcal{C} , i.e., $|\mathcal{C}| = M$, can be upper-bounded as

$$\begin{aligned} \log M & \leq \log \left(e^{L\bar{R}_{\mathcal{C}_1}(L, T, \rho)} + e^{L\bar{R}_{\mathcal{C}_2}(L, T, \rho)} \right) \\ & = L\bar{R}_{\mathcal{C}_1}(L, T, \rho) \\ & \quad + \log \left(1 + e^{L(\bar{R}_{\mathcal{C}_2}(L, T, \rho) - \bar{R}_{\mathcal{C}_1}(L, T, \rho))} \right). \end{aligned} \quad (116)$$

By the asymptotic behaviors of $\Theta(T, \rho)$, $K_{\bar{U}}(T, \rho)$, and $K_{\mathcal{D}_2}(T, \rho)$ (cf. (96), (86), and (89)), we have that $\Theta(T, \rho) \geq \Theta(T)/2$, $T^2 \tilde{V}(T) - K_{\bar{U}}(T, \rho) \leq 2T^2 \tilde{V}(T)$, and $K_{\mathcal{D}_2}(T, \rho) \leq$

$\Theta(T)/8$ for $\rho \geq \rho_0$ and a sufficiently large ρ_0 . Thus, for $\rho \geq \rho_0$,

$$\begin{aligned} & \bar{R}_{\mathcal{C}_2}(L, T, \rho) - \bar{R}_{\mathcal{C}_1}(L, T, \rho) \\ &= -\frac{1}{2}\Theta(T, \rho) + \sqrt{\frac{T^2\tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}}Q^{-1}(\epsilon) \\ &+ K_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{C}_2}(L, T) - K_{\mathcal{C}_1}(L, T) \\ &\leq -\frac{\Theta(T)}{8} + \sqrt{\frac{2T^2\tilde{V}(T)}{L}}Q^{-1}(\epsilon) \\ &+ K_{\mathcal{C}_2}(L, T) - K_{\mathcal{C}_1}(L, T). \end{aligned} \quad (117)$$

Similarly, by the asymptotic behaviors of $K_{\mathcal{C}_1}(L, T)$ and $K_{\mathcal{C}_2}(L, T)$ (cf. (106) and (115)), we have that

$$\sqrt{\frac{2T^2\tilde{V}(T)}{L}}Q^{-1}(\epsilon) + K_{\mathcal{C}_2}(L, T) - K_{\mathcal{C}_1}(L, T) \leq \frac{\Theta(T)}{16} \quad (118)$$

for $L \geq L_0$ and a sufficiently large L_0 . Consequently,

$$\bar{R}_{\mathcal{C}_2}(L, T, \rho) - \bar{R}_{\mathcal{C}_1}(L, T, \rho) \leq -\frac{\Theta(T)}{16} \quad (119)$$

for $\rho \geq \rho_0$ and $L \geq L_0$. Together with (116), this allows us to upper-bound the cardinality of \mathcal{C} as

$$\frac{\log M}{L} \leq \bar{R}_{\mathcal{C}_1}(L, T, \rho) + \frac{1}{L} \log \left(1 + e^{-L \frac{\Theta(T)}{16}} \right) \quad (120)$$

for $\rho \geq \rho_0$ and $L \geq L_0$. The second term on the RHS of (120) decays super-exponentially in L . We thus conclude from (105) and (120) that

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &= \frac{\log M}{LT} \\ &\leq \frac{\bar{J}_{\mathcal{D}_1}(T, \rho)}{T} - \sqrt{\frac{\tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}}Q^{-1}(\epsilon) \\ &+ K_{\mathcal{D}_1}(T, \rho) + K_{\text{MC}}(L, T) \end{aligned} \quad (121)$$

where

$$K_{\text{MC}}(L, T) \triangleq K_{\mathcal{C}_1}(L, T) + \frac{1}{L} \log \left(1 + e^{-L \frac{\Theta(T)}{16}} \right) \quad (122)$$

is a function of L and T that satisfies

$$|K_{\text{MC}}(L, T)| \leq A \frac{\log L}{L}, \quad L \geq L_0 \quad (123)$$

for some constants A and L_0 that are independent of (L, ρ) .

Comparing (84) with (9), we note that

$$\frac{\bar{J}_{\mathcal{D}_1}(T, \rho)}{T} = \tilde{I}(T, \rho) + K_{\bar{J}}(T, \rho) \quad (124)$$

where $K_{\bar{J}}(T, \rho)$ is a function of T and ρ that vanishes as $\rho \rightarrow \infty$. We can thus write (121) as

$$\begin{aligned} R^*(L, T, \epsilon, \rho) &\leq \tilde{I}(T, \rho) + K_{\text{MC}}(T, \rho) \\ &- \sqrt{\frac{\tilde{V}(T) - K_{\bar{U}}(T, \rho)}{L}}Q^{-1}(\epsilon) + K_{\text{MC}}(L, T) \end{aligned} \quad (125)$$

where $K_{\text{MC}}(T, \rho) \triangleq K_{\mathcal{D}_1}(T, \rho) + K_{\bar{J}}(T, \rho)$. Note that $\tilde{K}_\xi(T, \rho)$, $\xi \in \{\text{MC}, \bar{U}\}$ satisfy (11), and $K_{\text{MC}}(L, T)$ satisfies (12).

V. CONCLUSION

We presented a high-SNR normal approximation for the maximum coding rate $R^*(L, T, \epsilon, \rho)$ achievable over noncoherent MIMO Rayleigh block-fading channels using an error-correcting code that spans L coherence intervals of length T , satisfies a per-block power constraint ρ , and achieves a probability of error not larger than ϵ . While the approximation was derived under the assumption that the number of coherence intervals L and the SNR ρ tend to infinity, numerical analyses suggest that it becomes accurate already for moderate SNR values and coherence intervals.

The obtained normal approximation complements the nonasymptotic bounds presented in [23], [24], whose evaluation is computationally demanding. Furthermore, it lays a theoretical foundation for analytical analyses that study the behavior of the maximum coding rate as a function of system parameters such as SNR, number of coherence intervals, number of antennas, or blocklength. *Inter alia*, it enables an analysis of the fundamental tradeoff between diversity, multiplexing, and channel-estimation cost at finite blocklength. In Section III-C, we applied the normal approximation to determine how many of the n_t available transmit antennas should be employed in order to optimize the maximal coding rate at finite blocklength. Indeed, there is a tradeoff between multiplexing gain and cost of estimating the fading coefficients, which both grow with the number of active transmit antennas. While this tradeoff is trivial with respect to channel capacity, the maximum coding rate has a more intricate dependence on the number of active transmit antennas at finite blocklength. Last but not least, our normal approximation may serve as a proxy for the maximum coding rate in the analysis of communication protocols for short-packet wireless communication systems, similarly to the normal approximations used in [6]–[15].

APPENDIX I PROOF OF LEMMA 2

To prove Lemma 2, we analyze in Appendix I-A the USTM information density $i^{(U)}(\mathbf{X}; \mathbf{Y})$ and express it in terms of the singular values of \mathbf{Y} .⁵ The asymptotic behavior of these singular values as $\rho \rightarrow \infty$ is studied in Appendix I-B. The convergence of the channel dispersion is proved in Appendix I-C. Some auxiliary derivations are deferred to Appendix I-D.

A. The USTM Information Density $i^{(U)}(\mathbf{X}; \mathbf{Y})$

As argued in [24, App. A], the conditional distribution of the USTM information density $i^{(U)}(\mathbf{X}; \mathbf{Y})$ given $\mathbf{X} = \sqrt{\frac{T\rho}{n_t}}\mathbf{U}$ does not depend on \mathbf{U} . Without loss of generality, we can thus set in the following $\mathbf{X} = \sqrt{\frac{T\rho}{n_t}}\mathbf{U}$ with

$$\mathbf{U} = \begin{bmatrix} \mathbf{I}_{n_t} \\ \mathbf{0}_{(T-n_t) \times n_t} \end{bmatrix}. \quad (126)$$

Consider the SVD $\mathbf{H} = \mathbf{U}_\mathbf{H} \mathbf{S}_\mathbf{H} \mathbf{V}_\mathbf{H}^\mathsf{H}$, where $\mathbf{U}_\mathbf{H}$ and $\mathbf{V}_\mathbf{H}$ are unitary matrices of dimensions $n_t \times n_t$ and $n_r \times n_r$, respectively, and $\mathbf{S}_\mathbf{H}$ is an $(n_t \times n_r)$ -dimensional rectangular diagonal

⁵Throughout this appendix, the subscript ℓ is immaterial and therefore omitted.

matrix containing the n_t singular values $\sigma_{\mathbf{H},1}, \dots, \sigma_{\mathbf{H},n_t}$ of \mathbf{H} . Then, \mathbf{XH} can be expressed as $\mathbf{XH} = a\mathbf{UU}_\mathbf{H}\mathbf{S}_\mathbf{H}\mathbf{V}_\mathbf{H}^\mathsf{H}$, where $a \triangleq \sqrt{T\rho/n_t}$. Using the unitary invariant property [39, Ch. 5.2] and the circular symmetry of the complex Gaussian matrix \mathbf{W} [40, App. A], [28, Lemma 16], conditioned on \mathbf{X} , the random matrix $\mathbf{Y} = \mathbf{XH} + \mathbf{W}$ has the same conditional distribution as $\tilde{\mathbf{Y}} = a\mathbf{UU}_\mathbf{H}\Sigma_\mathbf{H} + \mathbf{W}$, where $\Sigma_\mathbf{H}$ is an $(n_t \times n_t)$ -dimensional diagonal matrix containing the singular values $\sigma_{\mathbf{H},1}, \dots, \sigma_{\mathbf{H},n_t}$ [28, Lemma 16]. Hence,

$$\tilde{\mathbf{Y}} = \begin{bmatrix} a\tilde{\Sigma}_\mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} \quad (127)$$

where $\tilde{\Sigma}_\mathbf{H} \triangleq \mathbf{U}_\mathbf{H}\Sigma_\mathbf{H}$, and where $\mathbf{W}_{11}, \mathbf{W}_{12}, \mathbf{W}_{21}$, and \mathbf{W}_{22} are submatrices of dimensions $n_t \times n_t$, $n_t \times (n_r - n_t)$, $(T - n_t) \times n_t$, and $(T - n_t) \times (n_r - n_t)$, respectively, with i.i.d. $\mathcal{CN}(0, 1)$ entries that are independent of \mathbf{H} .

We next use that, for $\mathbf{X} = a\mathbf{U}$,

$$\begin{aligned} & \text{tr}(\tilde{\mathbf{Y}}^\mathsf{H}(\mathbf{I}_T + \mathbf{XX}^\mathsf{H})^{-1}\tilde{\mathbf{Y}}) \\ &= \frac{a^2}{1+a^2}\text{tr}(\mathbf{H}^\mathsf{H}\mathbf{H}) + \text{tr}(\mathbf{W}_{21}^\mathsf{H}\mathbf{W}_{21}) + \text{tr}(\mathbf{W}_{22}^\mathsf{H}\mathbf{W}_{22}) \\ &+ \frac{2a}{1+a^2}\text{Re}\left(\text{tr}(\tilde{\Sigma}_\mathbf{H}^\mathsf{H}\mathbf{W}_{11})\right) + \frac{1}{1+a^2}\text{tr}(\mathbf{W}_{11}^\mathsf{H}\mathbf{W}_{11}) \\ &+ \frac{1}{1+a^2}\text{tr}(\mathbf{W}_{12}^\mathsf{H}\mathbf{W}_{12}) \end{aligned} \quad (128)$$

where $\text{Re}(\cdot)$ denotes the part. Consequently, the conditional pdf of \mathbf{Y} given $\mathbf{X} = a\mathbf{U}$ satisfies

$$\begin{aligned} & f_{\mathbf{Y}|\mathbf{X}}(\tilde{\mathbf{Y}}|a\mathbf{U}) \\ &= \frac{e^{-\text{tr}\{\tilde{\mathbf{Y}}^\mathsf{H}(\mathbf{I}_T + a^2\mathbf{UU}^\mathsf{H})^{-1}\tilde{\mathbf{Y}}\}}}{\pi^{Tn_r} \det(\mathbf{I}_T + a^2\mathbf{UU}^\mathsf{H})^{n_r}} \\ &= \frac{e^{-\frac{a^2}{1+a^2}\text{tr}(\mathbf{H}^\mathsf{H}\mathbf{H}) - \frac{2a}{1+a^2}\text{Re}(\text{tr}(\tilde{\Sigma}_\mathbf{H}^\mathsf{H}\mathbf{W}_{11}))}}{\pi^{Tn_r}(1+a^2)^{n_t n_r}} \times \\ & \times e^{-\text{tr}(\mathbf{W}_{21}^\mathsf{H}\mathbf{W}_{21}) - \text{tr}(\mathbf{W}_{22}^\mathsf{H}\mathbf{W}_{22}) - \frac{\text{tr}(\mathbf{W}_{11}^\mathsf{H}\mathbf{W}_{11}) + \text{tr}(\mathbf{W}_{12}^\mathsf{H}\mathbf{W}_{12})}{1+a^2}}. \end{aligned} \quad (129)$$

We next consider the output pdf $f_{\mathbf{Y}}^{(U)}$ induced by USTM channel inputs and the channel (3). To this end, we denote the singular values of \mathbf{Y} arranged in decreasing order by $\sigma_1, \dots, \sigma_{n_r}$. Then, by expressing \mathbf{Y} in terms of its SVD $\mathbf{Y} = \mathbf{U}_\mathbf{Y}\Sigma_\mathbf{Y}\mathbf{V}_\mathbf{Y}^\mathsf{H}$, where $\mathbf{U}_\mathbf{Y}$ and $\mathbf{V}_\mathbf{Y}$ are (truncated) unitary matrices and $\Sigma_\mathbf{Y}$ is a diagonal matrix containing the singular values $\sigma_1, \dots, \sigma_{n_r}$, the pdf $f_{\mathbf{Y}}^{(U)}$ can be written as [29]

$$f_{\mathbf{Y}}^{(U)}(\mathbf{Y}) = \frac{f_{\mathbf{U}_\mathbf{Y}}(\mathbf{U}_\mathbf{Y})f_{\Sigma_\mathbf{Y}}(\sigma_1, \dots, \sigma_{n_r})f_{\mathbf{V}_\mathbf{Y}}(\mathbf{V}_\mathbf{Y})}{J(T, \Sigma_\mathbf{Y})} \quad (130)$$

for $\mathbf{Y} = \mathbf{U}_\mathbf{Y}\Sigma_\mathbf{Y}\mathbf{V}_\mathbf{Y}^\mathsf{H}$, where $\Sigma_\mathbf{Y}$ is a diagonal matrix containing the singular values of \mathbf{Y} . In (130), $J(T, \Sigma_\mathbf{Y})$ is the Jacobian of the SVD, i.e.,

$$J(T, \Sigma_\mathbf{Y}) = \prod_{1 \leq i < j \leq n_r} (\sigma_i^2 - \sigma_j^2)^2 \prod_{i=1}^{n_r} \sigma_i^{2(T-n_r)+1} \quad (131)$$

and the pdfs of $\mathbf{U}_\mathbf{Y}$ and $\mathbf{V}_\mathbf{Y}$ are given by $f_{\mathbf{U}_\mathbf{Y}}(\mathbf{U}_\mathbf{Y}) = \frac{\Gamma_{n_r}(T)}{\pi^{n_r(T-1)}}$ and $f_{\mathbf{V}_\mathbf{Y}}(\mathbf{V}_\mathbf{Y}) = \frac{\Gamma_{n_r}(n_r)}{2^{n_r}\pi^{n_r^2}}$. We next define

$$\tilde{\sigma}_i \triangleq \begin{cases} \sqrt{\frac{n_t}{T\rho}}\sigma_i, & i = 1, \dots, n_t \\ \sigma_i, & i = n_t + 1, \dots, n_r. \end{cases} \quad (132)$$

It follows that

$$f_{\Sigma_\mathbf{Y}}(\sigma_1, \dots, \sigma_{n_r}) = \left(\frac{T\rho}{n_t}\right)^{n_t/2} f_{\tilde{\sigma}}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r}) \quad (133)$$

and

$$J(T, \Sigma_\mathbf{Y}) = \bar{J}(T, \bar{\Sigma}_\mathbf{Y}) \left(\frac{n_t}{T\rho}\right)^{n_t(T+n_r-n_t-1/2)} \quad (134)$$

where $f_{\tilde{\sigma}}$ denotes the joint pdf of $\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r}$ and

$$\begin{aligned} \bar{J}(T, \bar{\Sigma}_\mathbf{Y}) &\triangleq \left(\prod_{i=1}^{n_r} \bar{\sigma}_i^{2(T-n_r)+1}\right) \left(\prod_{1 \leq i < j \leq n_t} (\bar{\sigma}_i^2 - \bar{\sigma}_j^2)^2\right) \times \\ &\times \left(\prod_{n_t+1 \leq i < j \leq n_r} (\bar{\sigma}_i^2 - \bar{\sigma}_j^2)^2\right) \left(\prod_{i=1}^{n_t} \prod_{j=n_t+1}^{n_r} \left(\bar{\sigma}_i^2 - \frac{n_t}{T\rho}\bar{\sigma}_j^2\right)^2\right). \end{aligned} \quad (135)$$

Finally, we denote by $\tilde{\sigma}_i$, $i = 1, \dots, n_r$ the random variables that are obtained by replacing in (132) the singular values σ_i by the singular values of $\tilde{\mathbf{Y}}$. Since $\tilde{\mathbf{Y}}$ has the same distribution as \mathbf{Y} , it follows that $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ have the same joint distribution as $(\bar{\sigma}_1, \dots, \bar{\sigma}_{n_r})$, and the joint pdf of $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$, denoted by $f_{\tilde{\sigma}}$, is equal to $f_{\tilde{\sigma}}$.

Combining (129)–(135), we obtain that the USTM information density $i^{(U)}(\mathbf{X}; \mathbf{Y})$ satisfies

$$\begin{aligned} i^{(U)}(\mathbf{X}; \mathbf{Y}) &\stackrel{d}{=} \log \frac{f_{\mathbf{Y}|\mathbf{X}}(\tilde{\mathbf{Y}}|a\mathbf{U})}{f_{\mathbf{Y}}^{(U)}(\tilde{\mathbf{Y}})} \\ &= n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\bar{J}(T, \bar{\Sigma}_\mathbf{Y})}{f_{\tilde{\sigma}}^{(U)}(\tilde{\Sigma}_\mathbf{Y})} \\ &- \Xi(\mathbf{H}, \mathbf{W}) - \Delta_\Xi(\rho, \mathbf{H}, \mathbf{W}) \end{aligned} \quad (136)$$

where

$$\begin{aligned} \Xi(\mathbf{H}, \mathbf{W}) &\triangleq \text{tr}(\mathbf{H}^\mathsf{H}\mathbf{H}) + \text{tr}(\mathbf{W}_{21}^\mathsf{H}\mathbf{W}_{21}) + \text{tr}(\mathbf{W}_{22}^\mathsf{H}\mathbf{W}_{22}) \\ &+ Tn_r \log \pi + \log \frac{\Gamma_{n_r}(T)\Gamma_{n_r}(n_r)}{2^{n_r}\pi^{n_r(T-1)+n_r^2}} \end{aligned} \quad (137a)$$

$$\begin{aligned} \Delta_\Xi(\rho, \mathbf{H}, \mathbf{W}) &\triangleq \frac{\text{tr}(\mathbf{W}_{11}^\mathsf{H}\mathbf{W}_{11}) + \text{tr}(\mathbf{W}_{12}^\mathsf{H}\mathbf{W}_{12}) - \text{tr}(\mathbf{H}^\mathsf{H}\mathbf{H})}{1+a^2} \\ &+ \frac{2a}{1+a^2}\text{Re}\left(\text{tr}(\tilde{\Sigma}_\mathbf{H}^\mathsf{H}\mathbf{W}_{11})\right) + n_t n_r \log\left(1 + \frac{1}{a^2}\right) \end{aligned} \quad (137b)$$

and where, with a slight abuse of notation, we denote by $\tilde{\Sigma}_\mathbf{Y}$ both the random variables $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ and the diagonal matrix that contains these random variables.

In order to characterize the convergence of the channel dispersion, we next analyze the asymptotic behavior of the RHS of (136). This, in turn, requires an asymptotic analysis of $\tilde{\Sigma}_\mathbf{Y}$ and of $f_{\tilde{\sigma}}$, which we shall carry out in the next subsection.

B. Asymptotics of $\tilde{\Sigma}_\mathbf{Y}$ and $f_{\tilde{\sigma}}$

Let $\varsigma_1, \dots, \varsigma_{n_t}$ be the ordered nonzero singular values of \mathbf{H} , and let $\varsigma_{n_t+1}, \dots, \varsigma_{n_r}$ be the ordered nonzero singular values of \mathbf{W}_{22} . The following lemmas concern the convergence of $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ to $(\varsigma_1, \dots, \varsigma_{n_r})$ and of $f_{\tilde{\sigma}}$ to f_ς (where f_ς denotes the joint pdf of $(\varsigma_1, \dots, \varsigma_{n_r})$).

Lemma 15 (Yang-Durisi-Riegler'13): The singular values $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ and $(\varsigma_1, \dots, \varsigma_{n_r})$ and their corresponding pdfs satisfy the following:

- 1) $\sum_{i=1}^{n_r} E[\tilde{\sigma}_i^2] \leq K_1$, where K_1 is a finite constant independent of ρ .
- 2) For any subset $\tilde{\Sigma} \subseteq \{\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r}\}$, the pdf of $\tilde{\Sigma}$ satisfies $f_{\tilde{\Sigma}}(\tilde{\Sigma}) \leq K_2$ for every $\tilde{\Sigma}$ and $\rho \geq \rho_0$, where K_2 and ρ_0 are finite constants independent of ρ .
- 3) $f_{\tilde{\sigma}}$ converges pointwise to f_{ς} as $\rho \rightarrow \infty$.

Proof: The boundedness of $\sum_{i=1}^{n_r} E[\tilde{\sigma}_i^2]$ and $f_{\tilde{\sigma}}$ (Parts 1) and 2) was shown in [29, App. B]. The boundedness of $f_{\tilde{\Sigma}}$ follows by marginalizing the pdf $f_{\tilde{\sigma}}$, characterized in [29, App. A], over $\{\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r}\} \setminus \tilde{\Sigma}$ and by bounding the corresponding integrals. Part 3) is [29, Lemma 12]. ■

Lemma 16 (Zheng-Tse'02): As $\rho \rightarrow \infty$, the singular values $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ converge almost surely to $(\varsigma_1, \dots, \varsigma_{n_r})$.

Proof: The proof of Lemma 16 follows the steps outlined in [28, App. F]. Indeed, the convergence of $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_r})$ to $(\varsigma_1, \dots, \varsigma_{n_r})$ is determined by the asymptotic behavior of the eigenvalues $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{n_r})$ of $\tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}}$.

We shall first consider the eigenvalues that are bounded in ρ . So let $\tilde{\lambda}$ be an eigenvalue of $\tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}}$ that satisfies $\overline{\lim}_{\rho \rightarrow \infty} \tilde{\lambda} < \infty$. It follows that the characteristic polynomial

$$f(\tilde{\lambda}) \triangleq \det(\tilde{\lambda}I_{n_r} - \tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}}) \quad (138)$$

is zero. Using (127), $f(\tilde{\lambda})$ can be written as

$$f(\tilde{\lambda}) = \det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{D} \end{pmatrix} \quad (139)$$

where

$$\mathbf{A} \triangleq \tilde{\lambda}I_{n_r} - (a\tilde{\Sigma}_{\mathbf{H}} + \mathbf{W}_{11})^H (a\tilde{\Sigma}_{\mathbf{H}} + \mathbf{W}_{11}) - \mathbf{W}_{21}^H \mathbf{W}_{21} \quad (140a)$$

$$\mathbf{B} \triangleq -(a\tilde{\Sigma}_{\mathbf{H}}^H \mathbf{W}_{12} + \mathbf{W}_{11}^H \mathbf{W}_{12} + \mathbf{W}_{21}^H \mathbf{W}_{22}) \quad (140b)$$

$$\mathbf{D} \triangleq \tilde{\lambda}I_{n_r - n_t} - (\mathbf{W}_{12}^H \mathbf{W}_{12} + \mathbf{W}_{22}^H \mathbf{W}_{22}). \quad (140c)$$

Next note that, since $\tilde{\lambda}$ is bounded, $\frac{1}{a^2} \mathbf{A}$ converges almost surely to $-\tilde{\Sigma}_{\mathbf{H}}^H \tilde{\Sigma}_{\mathbf{H}}$ as ρ (and hence also a) tends to infinity. Since $\det(\tilde{\Sigma}_{\mathbf{H}}) > 0$, it follows that $\underline{\lim}_{\rho \rightarrow \infty} |\det(\mathbf{A})| > 0$. We can therefore use the Schur complement to express the characteristic polynomial as

$$f(\tilde{\lambda}) = \det(\mathbf{A}) \det(\mathbf{D} - \mathbf{B}^H \mathbf{A}^{-1} \mathbf{B}). \quad (141)$$

Furthermore, $f(\tilde{\lambda}) = 0$ implies that $\det(\mathbf{D} - \mathbf{B}^H \mathbf{A}^{-1} \mathbf{B}) = 0$. It can be shown that, as $\rho \rightarrow \infty$,

$$\mathbf{B}^H \mathbf{A}^{-1} \mathbf{B} \rightarrow -\mathbf{W}_{12}^H \mathbf{W}_{12}, \quad \text{almost surely.} \quad (142)$$

It thus follows from the continuity of the determinant that $\det(\mathbf{D} - \mathbf{B}^H \mathbf{A}^{-1} \mathbf{B})$ converges almost surely to $\det(\tilde{\lambda}I_{n_r - n_t} - \mathbf{W}_{22}^H \mathbf{W}_{22})$. We conclude that $\tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}}$ has $n_r - n_t$ bounded eigenvalues $\tilde{\lambda}$ that converge almost surely to the eigenvalues of $\mathbf{W}_{22}^H \mathbf{W}_{22}$ as $\rho \rightarrow \infty$. Clearly, these are the $n_r - n_t$ smallest eigenvalues and correspond to the singular values $(\tilde{\sigma}_{n_t+1}, \dots, \tilde{\sigma}_{n_r})$, since the remaining eigenvalues are unbounded. It follows that $(\tilde{\sigma}_{n_t+1}, \dots, \tilde{\sigma}_{n_r})$ converge almost surely to the $n_r - n_t$ nonzero singular values of \mathbf{W}_{22} .

We next consider the unbounded eigenvalues of $\tilde{\mathbf{Y}}^H \tilde{\mathbf{Y}}$. Let $\tilde{\lambda} = a^2 \tilde{\sigma}^2$ for some $\tilde{\sigma} > 0$. Since $\tilde{\lambda}$ is unbounded, it follows that $\underline{\lim}_{\rho \rightarrow \infty} |\det(\mathbf{D})| > 0$. We can therefore use again the Schur complement to express the characteristic polynomial as

$$f(\tilde{\lambda}) = \det(\mathbf{D}) \det(\mathbf{A} - \mathbf{B} \mathbf{D}^{-1} \mathbf{B}^H). \quad (143)$$

Furthermore, $f(\tilde{\lambda}) = 0$ implies that $\det(\mathbf{A} - \mathbf{B} \mathbf{D}^{-1} \mathbf{B}^H) = 0$. It can be shown that, as $\rho \rightarrow \infty$,

$$\frac{1}{a^2} \mathbf{A} \rightarrow \tilde{\sigma} I_{n_t} - \tilde{\Sigma}_{\mathbf{H}}^H \tilde{\Sigma}_{\mathbf{H}}, \quad \text{almost surely} \quad (144a)$$

$$\frac{1}{a} \mathbf{B} \rightarrow -\tilde{\Sigma}_{\mathbf{H}}^H \mathbf{W}_{12}, \quad \text{almost surely} \quad (144b)$$

$$\mathbf{D}^{-1} \rightarrow 0, \quad \text{almost surely.} \quad (144c)$$

It thus follows that $\det(\frac{1}{a^2} \mathbf{A} - \frac{1}{a^2} \mathbf{B} \mathbf{A}^{-1} \mathbf{B}^H)$ converges almost surely to $\det(\tilde{\sigma}^2 I_{n_t} - \tilde{\Sigma}_{\mathbf{H}}^H \tilde{\Sigma}_{\mathbf{H}})$. Consequently, as $\rho \rightarrow \infty$, the scaled eigenvalue $\tilde{\sigma}^2$ converges almost surely to an eigenvalue of $\tilde{\Sigma}_{\mathbf{H}}^H \tilde{\Sigma}_{\mathbf{H}}$. It follows that $(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_t})$ converge almost surely to the n_t nonzero singular values of \mathbf{H} . ■

C. The Convergence of the Channel Dispersion

In Appendix I-D, we shall show that

$$\lim_{\rho \rightarrow \infty} \text{Var}(i^{(U)}(\mathbf{X}; \mathbf{Y}))^2 = \text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}))^2 \quad (145)$$

where

$$i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) \triangleq n_t(T - n_t) \log \frac{T\rho}{n_t} - \log f_{\varsigma}(\tilde{\Sigma}_{\varsigma}) + \log \tilde{J}(T, \tilde{\Sigma}_{\varsigma}) - \Xi(\mathbf{H}, \mathbf{W}) \quad (146)$$

and

$$\begin{aligned} \tilde{J}(T, \tilde{\Sigma}_{\varsigma}) \triangleq & \left(\prod_{i=1}^{n_r} \varsigma_i^{2(T-n_r)+1} \right) \left(\prod_{1 \leq i < j \leq n_t} (\varsigma_i^2 - \varsigma_j^2)^2 \right) \times \\ & \times \left(\prod_{n_t+1 \leq i < j \leq n_r} (\varsigma_i^2 - \varsigma_j^2)^2 \right) \left(\prod_{i=1}^{n_t} \varsigma_i^{4(n_r-n_t)} \right). \end{aligned} \quad (147)$$

For future reference, $I^*(T, \rho) \triangleq E[i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})]$ is given by

$$\begin{aligned} I^*(T, \rho) = & n_t(T - n_t) \log \frac{\rho}{n_t} + n_t(T - n_t) \log \frac{T}{e} \\ & + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + (T - n_t) \sum_{i=0}^{n_t-1} \Psi(n_r - i) \end{aligned} \quad (148)$$

Similar to (136), we denote by $\tilde{\Sigma}_{\varsigma}$ both the random variables $(\varsigma_1, \dots, \varsigma_{n_r})$ and a diagonal matrix containing these random variables. To compute the expected value in (148), we use Lemma 16 together with a change of variable to express the pdfs of $(\varsigma_1, \dots, \varsigma_{n_r})$ and $(\varsigma_{n_t+1}, \dots, \varsigma_{n_r})$ in terms of the pdfs of \mathbf{H} and \mathbf{W}_{22} , respectively. After some algebraic manipulations, we obtain

$$\begin{aligned} \log \tilde{J}(T, \tilde{\Sigma}_{\varsigma}) - \log f_{\varsigma}(\tilde{\Sigma}_{\varsigma}) - \Xi(\mathbf{H}, \mathbf{W}) = & \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} \\ & + (T - n_t) \log \det(\mathbf{H}^H \mathbf{H}) - \text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21}) \end{aligned} \quad (149)$$

whose expected value can be evaluated using that $\sum_{i=1}^{n_t} E[\log(\varsigma_i^2)] = \sum_{i=0}^{n_t-1} \Psi(n_r - i)$ (Lemma 7) and $E[\text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21})] = n_t(T - n_t)$. We next note that

$$\begin{aligned} \text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})) = & n_t(T - n_t) + (T - n_r)^2 \sum_{i=0}^{n_t-1} \Psi'(n_r - i) \end{aligned} \quad (150)$$

which is equal to (10). In (150), we used that \mathbf{H} and \mathbf{W}_{21} are independent, and that $\text{Var}(\log \det(\mathbf{H}^H \mathbf{H})) = \sum_{i=0}^{n_t-1} \Psi'(n_r - i)$ (Lemma 7). Lemma 2 follows therefore directly from (145).

D. Proof of (145)

By (136), we have that

$$\underline{U}(T, \rho) = \mathbb{E} \left[\left(n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\bar{J}(T, \tilde{\Sigma}_{\tilde{Y}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}})} - \Xi(\mathbf{H}, \mathbf{W}) - \Delta_{\Xi}(\rho, \mathbf{H}, \mathbf{W}) - I(T, \rho) \right)^2 \right]. \quad (151)$$

We next note that

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[(\Delta_{\Xi}(\rho, \mathbf{H}, \mathbf{W}))^2 \right] = 0 \quad (152)$$

from which we obtain that, for every $0 < \varepsilon < 1$, there exists a sufficiently large ρ_0 such that

$$\mathbb{E} \left[(\Delta_{\Xi}(\rho, \mathbf{H}, \mathbf{W}))^2 \right] \leq \varepsilon, \quad \rho \geq \rho_0. \quad (153)$$

Indeed, the absolute value of $\Delta_{\Xi}(\rho, \mathbf{H}, \mathbf{W})$ converges almost surely to zero as $\rho \rightarrow \infty$ and is bounded by

$$\begin{aligned} & |\text{tr}(\mathbf{W}_{11}^H \mathbf{W}_{11})| + |\text{tr}(\mathbf{W}_{12}^H \mathbf{W}_{12})| + |\text{tr}(\mathbf{H}^H \mathbf{H})| \\ & + 2 \left| \text{Re} \left(\text{tr}(\tilde{\Sigma}_{\mathbf{H}}^H \mathbf{W}_{11}) \right) \right| + n_t n_r \log \left(1 + \frac{n_t}{T \rho_0} \right) \end{aligned} \quad (154)$$

for $\rho \geq \rho_0$, which has a finite second moment. The claim (152) thus follows from the dominated convergence theorem. We next note that, by (42a), there exists a sufficiently large $\rho_0 > 0$ such that

$$|I(T, \rho) - I^*(T, \rho)| \leq \varepsilon, \quad \rho \geq \rho_0. \quad (155)$$

Defining

$$\hat{U}(T, \rho) \triangleq \mathbb{E} \left[\left(n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\bar{J}(T, \tilde{\Sigma}_{\tilde{Y}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}})} - \Xi(\mathbf{H}, \mathbf{W}) - I^*(T, \rho) \right)^2 \right] \quad (156)$$

we can then approximate $\underline{U}(T, \rho)$ by $\hat{U}(T, \rho)$ as

$$\sqrt{\hat{U}(T, \rho)} - 2\varepsilon \leq \sqrt{\underline{U}(T, \rho)} \leq \sqrt{\hat{U}(T, \rho)} + 2\varepsilon. \quad (157)$$

Consequently, the limit of $\underline{U}(T, \rho)$ as $\rho \rightarrow \infty$ follows from the limit of $\hat{U}(T, \rho)$ upon letting ε tend to zero from above.

To analyze $\hat{U}(T, \rho)$, we next define the set

$$\begin{aligned} \mathcal{B}_{\varepsilon} &\triangleq \left\{ \Sigma \in \mathcal{D}_{n_r}^+ : \varepsilon \leq \Sigma_{ii} \leq \frac{1}{\varepsilon}, i = 1, \dots, n_r; f_{\tilde{\sigma}}(\Sigma) \geq \varepsilon; \right. \\ &\quad \left. \text{and } \bar{J}(T, \Sigma) \geq \varepsilon \right\} \end{aligned} \quad (158)$$

where $\mathcal{D}_{n_r}^+$ denotes the set of $(n_r \times n_r)$ -dimensional diagonal matrices with non-negative, real-valued entries and Σ_{ii} denotes the i -th diagonal element of Σ . We shall study the asymptotic behavior of $\hat{U}(T, \rho)$ as $\rho \rightarrow \infty$ by dividing the expected value into two parts, depending on whether $\tilde{\Sigma}_{\tilde{Y}}$ lies in $\mathcal{B}_{\varepsilon}$, and by letting then ε tend to zero from above. To this end, we define

$$\underline{U}(T, \rho, \mathcal{S}) \triangleq \mathbb{E} \left[\left(n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\bar{J}(T, \tilde{\Sigma}_{\tilde{Y}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}})} - \Xi(\mathbf{H}, \mathbf{W}) - I^*(T, \rho) \right)^2 \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{S}\} \right] \quad (159)$$

and

$$\begin{aligned} \tilde{U}(T, \rho, \mathcal{S}) &\triangleq \mathbb{E} \left[\left(n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\tilde{J}(T, \tilde{\Sigma}_{\tilde{Y}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}})} \right. \right. \\ &\quad \left. \left. - \Xi(\mathbf{H}, \mathbf{W}) - I^*(T, \rho) \right)^2 \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{S}\} \right] \end{aligned} \quad (160)$$

for $\mathcal{S} \subseteq \mathcal{D}_{n_r}^+$. Consequently, $\hat{U}(T, \rho)$ can be written as

$$\hat{U}(T, \rho) = \underline{U}(T, \rho, \mathcal{B}_{\varepsilon}) + \tilde{U}(T, \rho, \mathcal{B}_{\varepsilon}^c). \quad (161)$$

In the following subsections, we analyze the terms on the RHS of (161) separately.

1) *Limit of $\underline{U}(T, \rho, \mathcal{B}_{\varepsilon})$:* By Lemma 15, $f_{\tilde{\sigma}}$ converges pointwise to $f_{\tilde{\sigma}}$. It then follows from Egoroff's theorem [41, Th. 2.5.5] that, for every $\varepsilon > 0$, there exists a set $\mathcal{A}_{\varepsilon}$ of probability less than ε such that $f_{\tilde{\sigma}}$ converges uniformly to $f_{\tilde{\sigma}}$ on $\tilde{\mathcal{B}}_{\varepsilon} \triangleq \mathcal{B}_{\varepsilon} \setminus \mathcal{A}_{\varepsilon}$.⁶ Thus, there exists a sufficiently large $\rho_0 > 0$ such that

$$|f_{\tilde{\sigma}}(\Sigma) - f_{\tilde{\sigma}}(\Sigma)| < \varepsilon^2, \quad \Sigma \in \tilde{\mathcal{B}}_{\varepsilon}, \rho \geq \rho_0 \quad (162)$$

in which case $|\log f_{\tilde{\sigma}}(\Sigma) - \log f_{\tilde{\sigma}}(\Sigma)| \leq \frac{\varepsilon}{1-\varepsilon}$. Similarly, there exists a sufficiently large $\rho_0 > 0$ such that

$$|\bar{J}(T, \Sigma) - \tilde{J}(T, \Sigma)| \leq \varepsilon^2, \quad \Sigma \in \tilde{\mathcal{B}}_{\varepsilon}, \rho \geq \rho_0 \quad (163)$$

and, hence, $|\log \bar{J}(T, \Sigma) - \log \tilde{J}(T, \Sigma)| \leq \frac{\varepsilon}{1-\varepsilon}$.

We next write $\hat{U}(T, \rho, \mathcal{B}_{\varepsilon})$ as

$$\hat{U}(T, \rho, \mathcal{B}_{\varepsilon}) = \underline{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon}) + \tilde{U}(T, \rho, \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon}). \quad (164)$$

Using (162)–(163), we can then approximate $\hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})$ as

$$\begin{aligned} \sqrt{\hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})} - \frac{2\varepsilon}{1-\varepsilon} &\leq \sqrt{\underline{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})} \\ &\leq \sqrt{\tilde{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})} + \frac{2\varepsilon}{1-\varepsilon} \end{aligned} \quad (165)$$

so the limit of $\hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})$ as $\rho \rightarrow \infty$ follows from the limit of $\tilde{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})$ upon letting ε tend to zero from above.

We next apply the dominated convergence theorem to evaluate the limit of $\hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon})$ as $\rho \rightarrow \infty$. Indeed, after algebraic manipulations similar to the ones that lead to (149), it can be shown that

$$\begin{aligned} &n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\tilde{J}(T, \tilde{\Sigma}_{\tilde{Y}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}})} - \Xi(\mathbf{H}, \mathbf{W}) - I^*(T, \rho) \\ &= (T - n_t) \sum_{i=1}^{n_t} \log \tilde{\sigma}_i^2 + \sum_{i=1}^{n_r} \tilde{\sigma}_i^2 - \text{tr}(\mathbf{H}^H \mathbf{H}) \\ &\quad - \text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21}) - \text{tr}(\mathbf{W}_{22}^H \mathbf{W}_{22}) + n_t(T - n_t) \\ &\quad - (T - n_t) \sum_{i=0}^{n_t-1} \Psi(n_r - i) \end{aligned} \quad (166)$$

which on the set

$$\tilde{\mathcal{B}}_{\varepsilon,1} \triangleq \left\{ \Sigma \in \mathcal{D}_{n_r}^+ : \varepsilon \leq \Sigma_{ii} \leq \frac{1}{\varepsilon}, i = 1, \dots, n_r \right\} \quad (167)$$

⁶Without loss of generality, we can assume that $\mathcal{A}_{\varepsilon}$ decreases monotonically to a set of probability zero, i.e., $\mathcal{A}_{\varepsilon_1} \supseteq \mathcal{A}_{\varepsilon_2}$ for $\varepsilon_1 \geq \varepsilon_2$. Indeed, if $f_{\tilde{\sigma}}$ converges uniformly to $f_{\tilde{\sigma}}$ on $\mathcal{B}_{\varepsilon} \setminus \mathcal{A}_{\varepsilon_1}$ and $\mathcal{B}_{\varepsilon} \setminus \mathcal{A}_{\varepsilon_2}$, then it also converges uniformly on $\mathcal{B}_{\varepsilon} \setminus \tilde{\mathcal{A}}_{\varepsilon_2}$, where $\tilde{\mathcal{A}}_{\varepsilon_2} = \mathcal{A}_{\varepsilon_1} \cap \mathcal{A}_{\varepsilon_2}$. The set $\tilde{\mathcal{A}}_{\varepsilon_2}$ satisfies $\mathcal{A}_{\varepsilon_1} \supseteq \tilde{\mathcal{A}}_{\varepsilon_2}$ and has probability less than ε_2 (because $\tilde{\mathcal{A}}_{\varepsilon_2} \subseteq \mathcal{A}_{\varepsilon_2}$), hence the claim follows.

is bounded by a random variable of finite second moment. Furthermore, by Lemma 16, the RHS of (166) converges almost surely to

$$(T - n_t) \sum_{i=1}^{n_t} \log \varsigma_i^2 - (T - n_t) \sum_{i=0}^{n_t-1} \Psi(n_r - i) - \text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21}) + n_t(T - n_t) \quad (168)$$

as $\rho \rightarrow \infty$, which is equal to $i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho)$ since $\sum_{i=1}^{n_t} \log \varsigma_i^2 = \log \det(\mathbf{H}^H \mathbf{H})$. Further using that, by (162)–(163),

$$\tilde{\mathcal{B}}_{\varepsilon,2} \subseteq \tilde{\mathcal{B}}_{\varepsilon} \subseteq \tilde{\mathcal{B}}_{\varepsilon,1} \quad (169)$$

where

$$\tilde{\mathcal{B}}_{\varepsilon,2} \triangleq \left\{ \Sigma \in \mathcal{D}_{n_r}^+ \setminus \mathcal{A}_{\varepsilon} : \varepsilon \leq \Sigma_{ii} \leq 1/\varepsilon, i = 1, \dots, n_r; f_{\varsigma}(\Sigma) \geq \varepsilon - \varepsilon^2; \text{ and } \tilde{J}(T, \Sigma) \geq \varepsilon - \varepsilon^2 \right\} \quad (170)$$

it follows from the dominated convergence theorem that

$$\begin{aligned} & \mathbb{E} \left[(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho))^2 \mathbf{1}\{\tilde{\Sigma}_{\varsigma} \in \tilde{\mathcal{B}}_{\varepsilon,2}\} \right] \\ & \leq \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon}) \\ & \leq \mathbb{E} \left[(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho))^2 \mathbf{1}\{\tilde{\Sigma}_{\varsigma} \in \tilde{\mathcal{B}}_{\varepsilon,1}\} \right]. \end{aligned} \quad (171)$$

Noting that both $\tilde{\mathcal{B}}_{\varepsilon,1}$ and $\tilde{\mathcal{B}}_{\varepsilon,2}$ increase monotonically to $\mathcal{D}_{n_r}^+$ as $\varepsilon \downarrow 0$, we then obtain from (165) and the monotone convergence theorem that

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho, \tilde{\mathcal{B}}_{\varepsilon}) = \text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})). \quad (172)$$

We conclude the analysis of $\hat{U}(T, \rho, \mathcal{B}_{\varepsilon})$ by showing that

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho, \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon}) = 0. \quad (173)$$

Indeed, by Part 2) of Lemma 15 and the definitions of $\mathcal{B}_{\varepsilon}$ and $\tilde{J}(T, \Sigma)$, we have that

$$|\log f_{\tilde{\sigma}}(\Sigma)| + |\log \tilde{J}(T, \Sigma)| \leq \kappa \log \frac{1}{\varepsilon}, \quad \Sigma \in \mathcal{B}_{\varepsilon} \quad (174)$$

for some κ that only depends on T , n_r , and n_t . Using that $(a+b)^2 \leq 2a^2 + 2b^2$, we then obtain that

$$\begin{aligned} & \left[n_t(T - n_t) \log \frac{T\rho}{n_t} + \log \frac{\tilde{J}(T, \tilde{\Sigma}_{\tilde{\mathbf{Y}}})}{f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}})} - \Xi(\mathbf{H}, \mathbf{W}) - I^*(T, \rho) \right]^2 \\ & \leq 2 \left(\Xi(\mathbf{H}, \mathbf{W}) + I^*(T, \rho) - n_t(T - n_t) \log \frac{T\rho}{n_t} \right)^2 \\ & \quad + 2\kappa^2 \log^2(\varepsilon) \end{aligned} \quad (175)$$

We can thus upper-bound $\hat{U}(T, \rho, \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon})$ as

$$\begin{aligned} & \hat{U}(T, \rho, \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon}) \\ & \leq 2\kappa^2 \log^2(\varepsilon) \Pr \left[\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon} \right] + 2\mathbb{E} \left[\left(\Xi(\mathbf{H}, \mathbf{W}) \right. \right. \\ & \quad \left. \left. + I^*(T, \rho) - n_t(T - n_t) \log \frac{T\rho}{n_t} \right)^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon}\} \right] \\ & \leq 2\sqrt{\mathbb{E} \left[\left(\Xi(\mathbf{H}, \mathbf{W}) + I^*(T, \rho) - n_t(T - n_t) \log \frac{T\rho}{n_t} \right)^4 \right]} \varepsilon \\ & \quad + 2\kappa^2 \log^2(\varepsilon) \varepsilon \end{aligned} \quad (176)$$

where we used the Cauchy-Schwarz inequality and that $\Pr[\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon} \cap \mathcal{A}_{\varepsilon}] \leq \Pr[\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{A}_{\varepsilon}] \leq \varepsilon$. The term $\Xi(\mathbf{H}, \mathbf{W}) + I^*(T, \rho) - n_t(T - n_t) \log(T\rho/n_t)$ is independent of ρ and has a finite fourth moment. Consequently, the RHS of (176) tends to zero as $\varepsilon \downarrow 0$, from which (173) follows. To conclude, (164), (172), and (173) demonstrate that

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho, \mathcal{B}_{\varepsilon}) = \text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})). \quad (177)$$

2) *Limit of $\hat{U}(T, \rho, \mathcal{B}_{\varepsilon}^c)$:* We shall show that

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho, \mathcal{B}_{\varepsilon}^c) = 0. \quad (178)$$

To this end, we use that $(a + b + c)^2 \leq 9a^2 + 9b^2 + 9c^2$ for any real numbers (a, b, c) to upper-bound $\hat{U}(T, \rho, \mathcal{B}_{\varepsilon}^c)$ as

$$\begin{aligned} & \hat{U}(T, \rho, \mathcal{B}_{\varepsilon}^c) \\ & \leq 9\mathbb{E} \left[\left(\Xi(\mathbf{H}, \mathbf{W}) + I^*(T, \rho) \right. \right. \\ & \quad \left. \left. - n_t(T - n_t) \log \frac{T\rho}{n_t} \right)^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c\} \right] \\ & \quad + 9\mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c\} \right] \\ & \quad + 9\mathbb{E} \left[(\log \tilde{J}(T, \tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c\} \right]. \end{aligned} \quad (179)$$

We next analyze each term on the RHS of (179) separately:

1) Using the Cauchy-Schwarz inequality, the first term on the RHS of (179) can be upper-bounded by the square-root of

$$\begin{aligned} & 81\mathbb{E} \left[\left(\Xi(\mathbf{H}, \mathbf{W}) + I^*(T, \rho) - n_t(T - n_t) \log \frac{T\rho}{n_t} \right)^4 \right] \times \\ & \quad \times \Pr[\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c]. \end{aligned} \quad (180)$$

The expected value in (180) is independent of ρ and finite. Furthermore, by (169), we have $\mathcal{B}_{\varepsilon}^c \subseteq \tilde{\mathcal{B}}_{\varepsilon,2}$ and $\tilde{\mathcal{B}}_{\varepsilon,2}$ is independent of ρ and increases monotonically to $\mathcal{D}_{n_r}^+$ as $\varepsilon \downarrow 0$. It thus follows from the continuity of measures that $\sup_{\rho > 0} \Pr[\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c] \rightarrow 0$ as $\varepsilon \downarrow 0$. Consequently, the first term on the RHS of (179) vanishes as $\rho \rightarrow \infty$ and then $\varepsilon \downarrow 0$.

2) The second expected value on the RHS of (179) can be divided into the two terms

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c\} \right] \\ & = \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c \cap \mathcal{A}_{\|\Sigma\|,r}\} \right] \\ & \quad + \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c \cap \mathcal{A}_{\|\Sigma\|,r}^c\} \right] \end{aligned} \quad (181)$$

where

$$\mathcal{A}_{\|\Sigma\|,r} \triangleq \left\{ \Sigma \in \mathcal{D}_{n_r}^+ : \|\Sigma\|_{\text{F}} \leq r \right\} \quad (182)$$

for some r that we let tend to infinity at the end of the proof.

By Part 2) of Lemma 15, $f_{\tilde{\sigma}} \leq K_2$. Consequently, $f_{\tilde{\sigma}} \log^2 f_{\tilde{\sigma}}$ is bounded by $\max\{4/e^2, K_2 \log^2(K_2)\}$ and

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{\mathbf{Y}}}))^2 \mathbf{1}\{\tilde{\Sigma}_{\tilde{\mathbf{Y}}} \in \mathcal{B}_{\varepsilon}^c \cap \mathcal{A}_{\|\Sigma\|,r}\} \right] \\ & = \int_{\mathcal{B}_{\varepsilon}^c \cap \mathcal{A}_{\|\Sigma\|,r}} f_{\tilde{\sigma}}(\Sigma) \log^2 f_{\tilde{\sigma}}(\Sigma) d\Sigma \\ & \leq \max\{4/e^2, K_2 \log^2(K_2)\} \lambda(\mathcal{B}_{\varepsilon}^c \cap \mathcal{A}_{\|\Sigma\|,r}) \end{aligned} \quad (183)$$

where $\lambda(\cdot)$ denotes the Lebesgue measure. By (169), we have $\mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r} \subseteq \tilde{\mathcal{B}}_{\varepsilon,2}^c \cap \mathcal{A}_{\|\Sigma\|,r}$ which, for a fixed r , is bounded, independent of ρ , and decreases to the empty set as $\varepsilon \downarrow 0$. It follows from the continuity of measures that the first expected value on the RHS of (181) vanishes as $\rho \rightarrow \infty$ and then $\varepsilon \downarrow 0$.

As for the second term on the RHS of (181), we further divide the set $\mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c$ into $\mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}$ and $\mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}^c$, where

$$\mathcal{A}_{f_{\tilde{\sigma}}} \triangleq \{\Sigma \in \mathcal{D}_{n_r}^+ : f_{\tilde{\sigma}}(\Sigma) \leq \|\Sigma\|_{\text{F}}^{-n_r-1}\}. \quad (184)$$

When $r > e^{2/(n_r+1)}$, we have that

$$f_{\tilde{\sigma}}(\Sigma) \log^2 f_{\tilde{\sigma}}(\Sigma) \leq (n_r + 1)^2 \frac{\log^2(\|\Sigma\|_{\text{F}})}{\|\Sigma\|_{\text{F}}^{n_r+1}} \quad (185)$$

for $\Sigma \in \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}$. Consequently,

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}}))^2 \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}} \} \right] \\ & \leq (n_r + 1)^2 \int_{\|\Sigma\|_{\text{F}} > r} \frac{\log^2(\|\Sigma\|_{\text{F}})}{\|\Sigma\|_{\text{F}}^{n_r+1}} d\Sigma \\ & \leq (n_r + 1)^2 \frac{5\pi^{\frac{n_r}{2}}}{\Gamma(\frac{n_r}{2})} \frac{\log^2(r)}{r} \end{aligned} \quad (186)$$

where we bounded the integral by performing a change of variable from Σ to $\|\Sigma\|_{\text{F}}$. Moreover, when $r > K_2^{1/(n_r+1)}$,

$$\log^2 f_{\tilde{\sigma}}(\Sigma) \leq (n_r + 1)^2 \log^2(\|\Sigma\|_{\text{F}}) \quad (187)$$

for $\Sigma \in \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}^c$. Since $x \mapsto \log^2(x)$ is a concave function for $x \geq e$, it thus follows from Jensen's inequality that, for $r > \max\{K_2^{1/(n_r+1)}, e\}$,

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}}))^2 \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}^c \} \right] \\ & \leq (n_r + 1)^2 \int_{\|\Sigma\|_{\text{F}} > r} f_{\tilde{\sigma}}(\Sigma) \log^2(\|\Sigma\|_{\text{F}}) d\Sigma \\ & \leq (n_r + 1)^2 \Pr[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} > r] \log^2 \mathbb{E}[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} \mid \|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} > r] \\ & \leq (n_r + 1)^2 \Pr[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} > r] \log^2 \left(\frac{\sqrt{K_1}}{\Pr[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} > r]} \right) \end{aligned} \quad (188)$$

where the last step follows by upper-bounding the conditional expectation by $\mathbb{E}[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}}] / \Pr[\|\tilde{\Sigma}_{\tilde{Y}}\|_{\text{F}} > r]$, and by then applying Part 1) of Lemma 15. Applying Chebyshev's inequality together with Part 1) of Lemma 15, and noting that the function $x \mapsto \log^2(x/a)$ is monotonically increasing for $x > ae^2$, this can be further upper-bounded by

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}}))^2 \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \cap \mathcal{A}_{f_{\tilde{\sigma}}}^c \} \right] \\ & \leq (n_r + 1)^2 \frac{K_1}{r^2} \log^2 \left(\frac{r^2}{\sqrt{K_1}} \right) \end{aligned} \quad (189)$$

for $r > \max\{K_2^{1/(n_r+1)}, e, \sqrt{K_1}e^2\}$. With (186) and (189), we can upper-bound the second expected value on the RHS of (181) as

$$\begin{aligned} & \mathbb{E} \left[(\log f_{\tilde{\sigma}}(\tilde{\Sigma}_{\tilde{Y}}))^2 \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \cap \mathcal{A}_{\|\Sigma\|,r}^c \} \right] \\ & \leq (n_r + 1)^2 \left(\frac{5\pi^{\frac{n_r}{2}}}{\Gamma(\frac{n_r}{2})} \frac{\log^2(r)}{r} + \frac{K_1}{r^2} \log^2 \left(\frac{r^2}{\sqrt{K_1}} \right) \right) \end{aligned} \quad (190)$$

which is independent of ρ and ε , and vanishes as $r \rightarrow \infty$. We thus obtain from (181), (183), and (190) that the second term on the RHS of (179) vanishes as we let first $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $r \rightarrow \infty$.

3) We upper-bound the third term on the RHS of (179) using the definition of $\bar{J}(T, \tilde{\Sigma}_{\tilde{Y}})$ in (135) and that

$$|a_1 + \dots + a_\mu|^\nu \leq c_{\mu,\nu} (|a_1|^\nu + \dots + |a_\mu|^\nu), \quad \mu, \nu \in \mathbb{N} \quad (191)$$

for some positive constant $c_{\mu,\nu}$ that only depends on μ and ν . It follows that

$$\begin{aligned} & (\log \bar{J}(T, \tilde{\Sigma}_{\tilde{Y}}))^2 \\ & \leq c_{\mu,\nu} (2(T - n_r) + 1)^2 \sum_{i=1}^{n_r} \log^2(\tilde{\sigma}_i) \\ & \quad + 4c_{\mu,\nu} \sum_{1 \leq i < j \leq n_t} \log^2(\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \\ & \quad + 4c_{\mu,\nu} \sum_{n_t < i < j \leq n_t} \log^2(\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \\ & \quad + 4c_{\mu,\nu} \sum_{i=1}^{n_t} \sum_{j=n_t+1}^{n_r} \log^2 \left(\tilde{\sigma}_i^2 - \frac{n_t}{T\rho} \tilde{\sigma}_j^2 \right). \end{aligned} \quad (192)$$

To show that the RHS of (179) vanishes as we first let $\rho \rightarrow \infty$ and then $\varepsilon \downarrow 0$, it thus suffices to show that

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \right] = 0 \quad (193a)$$

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \mathbb{E} \left[\log^2(\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \right] = 0 \quad (193b)$$

$$\lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \mathbb{E} \left[\log^2 \left(\tilde{\sigma}_i^2 - \frac{n_t}{T\rho} \tilde{\sigma}_j^2 \right) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \right] = 0 \quad (193c)$$

for the combinations of indices (i, j) indicated in (192).

To prove (193a), we bound the expected value as

$$\begin{aligned} & \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \right] \\ & \leq \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \mathbb{1} \left\{ \delta \leq \tilde{\sigma}_i \leq \frac{1}{\delta} \right\} \right] \\ & \quad + \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\sigma}_i < \delta \} \right] \\ & \quad + \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \left\{ \tilde{\sigma}_i > \frac{1}{\delta} \right\} \right] \end{aligned} \quad (194)$$

for some $\delta > 0$ that we shall let tend to zero from above at the end of the proof. The first term on the RHS of (194) can be upper-bounded as

$$\begin{aligned} & \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \} \mathbb{1} \left\{ \delta \leq \tilde{\sigma}_i \leq \frac{1}{\delta} \right\} \right] \\ & \leq \log^2(\delta) \Pr \left[\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c \right]. \end{aligned} \quad (195)$$

As noted before, $\sup_{\rho > 0} \Pr[\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c] \rightarrow 0$ as $\varepsilon \downarrow 0$ by the continuity of measures. Consequently, the first term on the RHS of (194) vanishes as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$.

For $\delta \leq 1$, the second term on the RHS of (194) can be bounded as

$$\begin{aligned} & \mathbb{E} \left[\log^2(\tilde{\sigma}_i) \mathbb{1} \{ \tilde{\sigma}_i < \delta \} \right] \\ & \leq \int_{\sigma < \delta, f_{\tilde{\sigma}_i}(\sigma) \leq \sigma^{-1/2}} f_{\tilde{\sigma}_i}(\sigma) \log^2(\sigma) d\sigma \end{aligned}$$

$$\begin{aligned}
 & + \int_{\substack{\sigma < \delta, \\ f_{\tilde{\sigma}_i}(\sigma) > \sigma^{-1/2}}} f_{\tilde{\sigma}_i}(\sigma) \log^2(\sigma) d\sigma \\
 & \leq \int_0^\delta \frac{\log^2(\sigma)}{\sqrt{\sigma}} d\sigma + 4 \int_0^\delta f_{\tilde{\sigma}_i}(\sigma) \log^2(f_{\tilde{\sigma}_i}(\sigma)) d\sigma \quad (196)
 \end{aligned}$$

where in the last step we bounded the first integral by upper-bounding $f_{\tilde{\sigma}_i}$ by $\sigma^{-1/2}$, and we bounded the second integral by noting that, for $\sigma < \delta < 1$, $\log^2(\sigma) = \log^2(1/\sigma)$ and by upper-bounding $1/\sigma$ by $f_{\tilde{\sigma}_i}^2$. The first integral on the RHS of (196) can be evaluated directly. For the second integral, we use that, by Part 2) of Lemma 15, $f_{\tilde{\sigma}_i} \log^2 f_{\tilde{\sigma}_i}$ is bounded by $\max\{4/e^2, K_2 \log^2(K_2)\}$. It follows that

$$\begin{aligned}
 & \mathbb{E}[\log^2(\tilde{\sigma}_i) \mathbb{1}\{\tilde{\sigma}_i < \delta\}] \\
 & \leq 40 \log^2(\delta) \sqrt{\delta} + \delta \max\{4/e^2, K_2 \log^2(K_2)\} \quad (197)
 \end{aligned}$$

which implies that the second term on the RHS of (194) vanishes as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$.

As for the third term on the RHS of (194), we recall that $x \mapsto \log^2(x)$ is a concave function for $x \geq e$. Consequently, by following the same steps as in (188)–(189), we obtain that, for $\delta \leq (\max\{e, \sqrt{K_1}e^2\})^{-1}$,

$$\mathbb{E}\left[\log^2(\tilde{\sigma}_i) \mathbb{1}\left\{\tilde{\sigma}_i > \frac{1}{\delta}\right\}\right] \leq K_1 \delta^2 \log^2\left(\sqrt{K_1} \delta^2\right). \quad (198)$$

Thus, the third term on the RHS of (194) also vanishes as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$.

Combining (194), (195), (197), and (198), we obtain (193a).

We next prove (193b). To this end, we use that $\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2 = (\tilde{\sigma}_i - \tilde{\sigma}_j)(\tilde{\sigma}_i + \tilde{\sigma}_j)$ and the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ to upper-bound the expected value in (193b) as

$$\begin{aligned}
 & \mathbb{E}\left[\log^2(\tilde{\sigma}_i^2 - \tilde{\sigma}_j^2) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right] \\
 & \leq 2\mathbb{E}\left[\log^2(\tilde{\sigma}_i - \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right] \\
 & \quad + 2\mathbb{E}\left[\log^2(\tilde{\sigma}_i + \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right]. \quad (199)
 \end{aligned}$$

We continue similarly to the proof of (193a). Indeed, we can bound the expected values on the RHS of (199) as

$$\begin{aligned}
 & \mathbb{E}[\log^2(\tilde{\sigma}_i - \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}] \\
 & \leq \mathbb{E}\left[\log^2(\tilde{\sigma}_i - \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\} \mathbb{1}\left\{\delta \leq \tilde{\sigma}_i - \tilde{\sigma}_j \leq \frac{1}{\delta}\right\}\right] \\
 & \quad + \mathbb{E}\left[\log^2(\tilde{\sigma}_i - \tilde{\sigma}_j) \mathbb{1}\{\tilde{\sigma}_i - \tilde{\sigma}_j < \delta\}\right] \\
 & \quad + \mathbb{E}\left[\log^2(\tilde{\sigma}_i - \tilde{\sigma}_j) \mathbb{1}\left\{\tilde{\sigma}_i - \tilde{\sigma}_j > \frac{1}{\delta}\right\}\right] \quad (200)
 \end{aligned}$$

and

$$\begin{aligned}
 & \mathbb{E}[\log^2(\tilde{\sigma}_i + \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}] \\
 & \leq \mathbb{E}\left[\log^2(\tilde{\sigma}_i + \tilde{\sigma}_j) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\} \mathbb{1}\left\{\delta \leq \tilde{\sigma}_i + \tilde{\sigma}_j \leq \frac{1}{\delta}\right\}\right] \\
 & \quad + \mathbb{E}\left[\log^2(\tilde{\sigma}_i + \tilde{\sigma}_j) \mathbb{1}\{\tilde{\sigma}_i + \tilde{\sigma}_j < \delta\}\right] \\
 & \quad + \mathbb{E}\left[\log^2(\tilde{\sigma}_i + \tilde{\sigma}_j) \mathbb{1}\left\{\tilde{\sigma}_i + \tilde{\sigma}_j > \frac{1}{\delta}\right\}\right]. \quad (201)
 \end{aligned}$$

The first terms on the RHSs of (200) and (201) are upper-bounded by $\log^2(\delta) \Pr[\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c]$, which tends to zero as $\varepsilon \downarrow 0$

by the continuity of measures. Hence, these terms vanish as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$.

As for the remaining terms, we next note that the random variables $\alpha_{ij} \triangleq \tilde{\sigma}_i - \tilde{\sigma}_j$ and $\beta_{ij} \triangleq \tilde{\sigma}_i + \tilde{\sigma}_j$ have the joint pdf

$$f_{\alpha_{ij}, \beta_{ij}}(\alpha, \beta) = \frac{1}{2} f_{\tilde{\sigma}_i, \tilde{\sigma}_j}\left(\frac{\alpha + \beta}{2}, \frac{\beta - \alpha}{2}\right) \quad (202)$$

for $\alpha \geq 0$ and $\beta > \alpha$. By marginalizing over α or β , and by bounding the corresponding integral, it can be shown that both $f_{\alpha_{ij}}$ and $f_{\beta_{ij}}$ are bounded. Furthermore, by Part 1) of Lemma 15, the second moments of α_{ij} and β_{ij} are bounded, too. We can thus follow the steps (196)–(198) to show that the second and third terms on the RHSs of (200) and (201) also vanish as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$. Thus, (193b) follows.

Finally, to prove (193c), we can follow along the lines of the proof of (193b). Indeed, we have that

$$\begin{aligned}
 & \mathbb{E}\left[\log^2\left(\tilde{\sigma}_i^2 - \frac{n_t}{T\rho} \tilde{\sigma}_j^2\right) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right] \\
 & \leq 2\mathbb{E}\left[\log^2\left(\tilde{\sigma}_i - \frac{n_t}{T\rho} \tilde{\sigma}_j\right) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right] \\
 & \quad + 2\mathbb{E}\left[\log^2\left(\tilde{\sigma}_i + \frac{n_t}{T\rho} \tilde{\sigma}_j\right) \mathbb{1}\{\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c\}\right]. \quad (203)
 \end{aligned}$$

We can then bound the expected values depending on whether $\alpha_{\rho,ij} \triangleq \tilde{\sigma}_i - \frac{n_t}{T\rho} \tilde{\sigma}_j$ and $\beta_{\rho,ij} \triangleq \tilde{\sigma}_i + \frac{n_t}{T\rho} \tilde{\sigma}_j$ lie in one of the intervals $[\delta, 1/\delta]$, $(0, \delta)$, or $(1/\delta, \infty)$. In the first case, the expected values are bounded by $\log^2(\delta) \Pr[\tilde{\Sigma}_{\tilde{Y}} \in \mathcal{B}_\varepsilon^c]$, which vanishes as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$. For the latter two cases, we note that the marginal pdfs $f_{\alpha_{\rho,ij}}$ and $f_{\beta_{\rho,ij}}$ as well as the second moments of $\alpha_{\rho,ij}$ and $\beta_{\rho,ij}$ are bounded. We can thus follow the steps (196)–(198) to show that these expected values vanish as we first let $\rho \rightarrow \infty$, then $\varepsilon \downarrow 0$, and then $\delta \downarrow 0$. Thus, (193c) follows.

3) *Summary:* By (161), we can express $\hat{U}(T, \rho)$ as $\hat{U}(T, \rho, \mathcal{B}_\varepsilon) + \hat{U}(T, \rho, \mathcal{B}_\varepsilon^c)$. As shown in Appendix I-D1, the first term converges to $\text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}))$ as we first let $\rho \rightarrow \infty$ and then $\varepsilon \downarrow 0$; cf. (177). As shown in Appendix I-D2, the second term converges to zero as we first let $\rho \rightarrow \infty$ and then $\varepsilon \downarrow 0$; cf. (178). Since, by (157), $\hat{U}(T, \rho)$ converges to $\underline{U}(T, \rho)$ as we let ε tend to zero from above, we conclude that

$$\begin{aligned}
 \lim_{\rho \rightarrow \infty} \underline{U}(T, \rho) &= \lim_{\varepsilon \downarrow 0} \lim_{\rho \rightarrow \infty} \hat{U}(T, \rho) \\
 &= \text{Var}(i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})) \quad (204)
 \end{aligned}$$

which is (145).

APPENDIX II PROOF OF LEMMA 3

The result that $\underline{U}(T, \rho)$ is uniformly bounded in $\rho \geq \rho_0$ for a sufficiently large ρ_0 (that only depends on T) follows directly from Lemma 2. Indeed, according to this lemma, $\underline{U}(T, \rho)$ can be approximated as

$$\underline{U}(T, \rho) = T^2 \tilde{V}(T) + K_{\underline{U}}(T, \rho) \quad (205)$$

where $K_{\underline{U}}(T, \rho)$ is a function of T and ρ that satisfies $\lim_{\rho \rightarrow \infty} K_{\underline{U}}(T, \rho) = 0$. Since $\tilde{V}(T)$ is independent of ρ and finite, the claim thus follows.

We next show that $E[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3]$ is uniformly bounded in $\rho \geq \rho_0$ for sufficiently large ρ_0 . Indeed, by following similar steps as in the proof of Lemma 2, it can be shown that

$$\begin{aligned} \lim_{\rho \rightarrow \infty} E[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3] \\ = E[|i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho)|^3]. \end{aligned} \quad (206)$$

Furthermore, by the definitions of $i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W})$ and $I^*(T, \rho)$ in (146) and (148), we have that

$$\begin{aligned} E[|i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho)|^3] \\ = E\left[\left| (T - n_t)n_t - \text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21}) + (T - n_t) \log \left(\prod_{i=1}^{n_t} \varsigma_i^2 \right) \right|^3 \right. \\ \left. - (T - n_t) \mathbb{E} \left[\log \left(\prod_{i=1}^{n_t} \varsigma_i^2 \right) \right] \right|^3 \\ \leq c_{2,3} E[|\text{tr}(\mathbf{W}_{21}^H \mathbf{W}_{21})|^3] + c_{2,3} (T - n_t)^3 n_t^3 \\ + 2c_{2,3} |T - n_t|^3 E\left[\left| \log \left(\prod_{i=1}^{n_t} \varsigma_i^2 \right) \right|^3 \right] \end{aligned} \quad (207)$$

where we use (191) in the last step.

Note that each term on the RHS of (207) is independent of ρ and bounded. Indeed, the first term on the RHS of (207) is the third absolute moment of a gamma-distributed random variable with parameters $(T - n_t, 1)$. As for the second term on the RHS of (207), by the definition of $\varsigma_1, \dots, \varsigma_{n_t}$, we have

$$\left| \log \left(\prod_{i=1}^{n_t} \varsigma_i^2 \right) \right| = |\log \det(\mathbf{H}_\ell^H \mathbf{H}_\ell)|. \quad (208)$$

Since $\log \det(\mathbf{H}_\ell^H \mathbf{H}_\ell)$ is independent of ρ and its third absolute moment is finite, it follows that the second term on the RHS of (207) is bounded, too. We conclude that $E[|i^*(\rho, \varsigma, \mathbf{H}, \mathbf{W}) - I^*(T, \rho)|^3]$ is bounded, which together with (206) yields that

$$\sup_{\rho \geq \rho_0} E[|i^{(U)}(\mathbf{X}; \mathbf{Y}) - I(T, \rho)|^3] < \infty \quad (209)$$

for sufficiently large $\rho_0 > 0$. This proves Lemma 3.

APPENDIX III

MISMATCHED INFORMATION DENSITY UPPER BOUND

We show that, conditioned on $\mathbf{X}_\ell = \mathbf{X}_\ell$, the mismatched information density $j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ can be upper-bounded by a term that has the same distribution as $\bar{j}(\mathbf{D}_\ell, \mathbf{Z}_\ell, \mathbf{Z}_\ell'', \mathbf{H}_\ell, \mathbf{Q}_\ell)$ defined in (61). To this end, we first use that $\sum_{i=1}^{n_t} \sigma_{i,\ell}^2 = \text{tr}(\mathbf{Y}_\ell^H \mathbf{Y}_\ell) - \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2$ to write the third, fourth, and fifth terms on the RHS of (60) as

$$\begin{aligned} \frac{n_t}{T\rho} \sum_{i=1}^{n_t} \sigma_{i,\ell}^2 + \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 - \text{tr}(\mathbf{Y}_\ell^H (\mathbf{I}_T + \mathbf{X}_\ell \mathbf{X}_\ell^H)^{-1} \mathbf{Y}_\ell) \\ = \text{tr}(\mathbf{Y}_\ell^H \mathbf{Y}_\ell \mathbf{Y}_\ell) + \left(1 - \frac{n_t}{T\rho}\right) \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 \end{aligned} \quad (210)$$

where

$$\mathbf{Y}_\ell \triangleq \frac{n_t}{T\rho} \mathbf{I}_T - (\mathbf{I}_T + \mathbf{X}_\ell \mathbf{X}_\ell^H)^{-1}. \quad (211)$$

We next use that we can assume without loss of optimality that \mathbf{X}_ℓ is a $T \times n_t$ rectangular diagonal matrix with diagonal entries $d_{1,\ell}, \dots, d_{n_t,\ell}$, so the channel output \mathbf{Y}_ℓ can be expressed as

$$\mathbf{Y}_\ell = \begin{pmatrix} \mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell} \\ \mathbf{W}_{2,\ell} \end{pmatrix} \quad (212)$$

where $\mathbf{D}_\ell \triangleq \text{diag}\{d_{1,\ell}, \dots, d_{n_t,\ell}\}$, and $\mathbf{W}_{1,\ell}$ and $\mathbf{W}_{2,\ell}$ denote the first n_t rows and the last $(T - n_t)$ rows of \mathbf{W}_ℓ , respectively. It can then be shown that

$$\begin{aligned} \text{tr}(\mathbf{Y}_\ell^H \mathbf{Y}_\ell \mathbf{Y}_\ell) &= \text{tr}\left((\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})^H \mathbf{Y}_{1,\ell} (\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})\right) \\ &+ \left(\frac{n_t}{T\rho} - 1\right) \text{tr}(\mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell}) \end{aligned} \quad (213)$$

where the matrix $\mathbf{Y}_{1,\ell}$ contains the first n_t rows and columns of \mathbf{Y}_ℓ , i.e.,

$$\mathbf{Y}_{1,\ell} \triangleq \frac{n_t}{T\rho} \mathbf{I}_{n_t} - (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2)^{-1}. \quad (214)$$

We continue by upper-bounding $\sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2$ following the steps in [28, p. 377]. Specifically, we use that $\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell}$ has rank n_t , so we can find a semi-unitary $(n_r - n_t) \times n_r$ matrix \mathbf{Q}_0 (possibly depending on \mathbf{H}_ℓ and $\mathbf{W}_{1,\ell}$) such that $\mathbf{Q}_0 (\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})^H = 0$. Together with (212), this implies that $\text{tr}(\mathbf{Q}_0 \mathbf{Y}_\ell^H \mathbf{Y}_\ell \mathbf{Q}_0^H) = \text{tr}(\mathbf{Q}_0 \mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell} \mathbf{Q}_0^H)$. Furthermore, it follows from [39, Cor. 4.3.39] that

$$\begin{aligned} \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 &= \min_{\substack{\mathbf{Q} \in \mathbb{C}^{(n_r-n_t) \times n_r}: \\ \mathbf{Q} \mathbf{Q}^H = \mathbf{I}_{n_r-n_t}}} \text{tr}(\mathbf{Q} \mathbf{Y}_\ell^H \mathbf{Y}_\ell \mathbf{Q}^H) \\ &\leq \text{tr}(\tilde{\mathbf{Q}} \mathbf{Y}_\ell^H \mathbf{Y}_\ell \tilde{\mathbf{Q}}^H) \end{aligned} \quad (215)$$

for any $(n_r - n_t) \times n_r$ matrix $\tilde{\mathbf{Q}}$ satisfying $\tilde{\mathbf{Q}} \tilde{\mathbf{Q}}^H = \mathbf{I}_{n_r-n_t}$. By applying (215) with $\tilde{\mathbf{Q}} = \mathbf{Q}_0$, we thus obtain that

$$\sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 \leq \text{tr}(\mathbf{Q}_0 \mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell} \mathbf{Q}_0^H). \quad (216)$$

We next consider a unitary $n_r \times n_r$ matrix \mathbf{Q} that contains \mathbf{Q}_0 as a submatrix. Let \mathbf{Q}_1 denote the submatrix composed of the remaining rows of \mathbf{Q} so that

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_0 \\ \mathbf{Q}_1 \end{pmatrix}. \quad (217)$$

It then follows that

$$\begin{aligned} \text{tr}(\mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell}) \\ = \text{tr}(\mathbf{Q}_0 \mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell} \mathbf{Q}_0^H) + \text{tr}(\mathbf{Q}_1 \mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell} \mathbf{Q}_1^H). \end{aligned} \quad (218)$$

Combining (213), (216), and (218), we obtain

$$\begin{aligned} \text{tr}(\mathbf{Y}_\ell^H \mathbf{Y}_\ell \mathbf{Y}_\ell) + \left(1 - \frac{n_t}{T\rho}\right) \sum_{i=n_t+1}^{n_r} \sigma_{i,\ell}^2 \\ \leq \text{tr}\left((\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})^H \mathbf{Y}_{1,\ell} (\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})\right) \\ + \left(\frac{n_t}{T\rho} - 1\right) \text{tr}(\mathbf{Q}_1 \mathbf{W}_{2,\ell}^H \mathbf{W}_{2,\ell} \mathbf{Q}_1^H). \end{aligned} \quad (219)$$

Using that the matrices \mathbf{H}_ℓ , $\mathbf{W}_{1,\ell}$, and $\mathbf{W}_{2,\ell}$ have i.i.d. $\mathcal{CN}(0, 1)$ entries, both traces on the RHS of (219) have gamma distributions. Specifically, we have that, conditioned

on $\mathbf{X}_\ell = \mathbf{D}_\ell$, $\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell}$ has the same distribution as $(\mathbf{I}_{n_t} + \mathbf{D})^{1/2} \mathbf{H}_\ell$ [28, p. 377]. Furthermore, \mathbf{Q}_1 depends on \mathbf{Q}_0 (which in turn may depend on \mathbf{H}_ℓ and $\mathbf{W}_{1,\ell}$), but it is independent of $\mathbf{W}_{2,\ell}$. It can then be shown that

$$\begin{aligned} & \text{tr} \left((\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell})^\mathsf{H} \Upsilon_{1,\ell} (\mathbf{D}_\ell \mathbf{H}_\ell + \mathbf{W}_{1,\ell}) \right) \\ & \stackrel{d}{=} \frac{n_t}{T\rho} \text{tr} \left(\mathbf{H}^\mathsf{H} (\mathbf{I}_{n_t} + \mathbf{D}_\ell) \mathbf{H} \right) - \text{tr} (\mathbf{H}^\mathsf{H} \mathbf{H}) \\ & \stackrel{d}{=} \sum_{i=1}^{n_t} \left(\frac{n_t + n_t d_{i,\ell}^2}{T\rho} - 1 \right) Z'_{i,\ell} \end{aligned} \quad (220a)$$

$$\begin{aligned} & \left(\frac{n_t}{T\rho} - 1 \right) \text{tr} (\mathbf{Q}_1 \mathbf{W}_{2,\ell}^\mathsf{H} \mathbf{W}_{2,\ell} \mathbf{Q}_1^\mathsf{H}) \\ & \stackrel{d}{=} \frac{n_t - T\rho}{T\rho} \sum_{i=1}^{n_t} Z''_{i,\ell} \end{aligned} \quad (220b)$$

where $Z'_{i,\ell}$ and $Z''_{i,\ell}$ were defined after (61). In fact, it follows from (220a) that $Z'_{i,\ell}$ is given by the squared Euclidean norm of the i -th row of \mathbf{H}_ℓ . Thus, (210) can be upper-bounded by a term that, given $\mathbf{X}_\ell = \mathbf{X}_\ell$, has the same distribution as

$$\sum_{i=1}^{n_t} \left(\frac{n_t + n_t d_{i,\ell}^2}{T\rho} - 1 \right) Z'_{i,\ell} - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} Z''_{i,\ell}. \quad (221)$$

It remains to upper-bound the last two terms on the RHS of (60). To this end, we follow along the lines of [29, Sec. IV-A]. Indeed, we begin by upper-bounding

$$\begin{aligned} & \log \left(\prod_{i=1}^{n_t} \prod_{j=n_t+1}^{n_r} (\sigma_{i,\ell}^2 - \sigma_{j,\ell}^2) \right) \\ & \leq (n_r - n_t) \log \left(\prod_{i=1}^{n_t} \sigma_{i,\ell}^2 \right) \end{aligned} \quad (222)$$

using that $\sigma_{j,\ell}^2 \geq 0$, $j = n_t + 1, \dots, n_r$. To bound the RHS of (222) and the second-to-last term on the RHS of (60), we next use that, conditioned on $\mathbf{X}_\ell = \mathbf{D}_\ell$, $\mathbf{Y}_\ell^\mathsf{H} \mathbf{Y}_\ell$ has the same distribution as $\mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \mathbf{H}_\ell + \mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell$ [28, p. 377], where \mathbf{Q}_ℓ was defined after (61). It then follows from Weyl's theorem [39, Th. 4.3.1] that (cf. [29, Sec. IV-A])

$$\begin{aligned} & \log \left(\prod_{i=1}^{n_t} \sigma_{i,\ell}^2 \right) \\ & \leq \log \det ((\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t}) \\ & = \log \det (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \\ & \quad + \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2)^{-1}) \end{aligned} \quad (223)$$

where we used that $\mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \mathbf{H}_\ell$ and $(\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}$ have the same n_t nonzero eigenvalues [39, Th. 1.3.20].

Combining (221)–(223) with (61), we conclude that, conditioned on $\mathbf{X}_\ell = \mathbf{X}_\ell$, $j_\ell(\mathbf{X}_\ell; \mathbf{Y}_\ell)$ can be upper-bounded by a term that has the same distribution as

$$\begin{aligned} & n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \sum_{i=1}^{n_t} \left(\frac{n_t + n_t d_{i,\ell}^2}{T\rho} - 1 \right) Z'_{i,\ell} \\ & - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} Z''_{i,\ell} + (T - n_t - n_r) \log \det (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2) \\ & + (T - n_t) \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}_\ell^2)^{-1}) \end{aligned} \quad (224)$$

which is $\bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$.

APPENDIX IV PROOF OF LEMMA 8

A. Proof of (66a)

By the definitions of $\bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$ and $\bar{J}(\mathbf{D}_\ell, T, \rho)$ in (61) and (63), we have that

$$\begin{aligned} & \bar{U}(\mathbf{D}, T, \rho) \\ & = \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{n_t}{T\rho} + \frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) (Z'_{i,\ell} - n_r) \right. \right. \\ & \quad - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \\ & \quad + (T - n_t) \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1}) \\ & \quad \left. \left. - (T - n_t) \times \right. \right. \\ & \quad \times \mathbb{E} [\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \right)^2 \Bigg] \\ & \leq c_{3,2} \sum_{i=1}^{n_t} \left(\frac{n_t}{T\rho} + \frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right)^2 \text{Var}(Z'_{i,\ell}) \\ & \quad + c_{3,2} \left(\frac{T\rho - n_t}{T\rho} \right)^2 \sum_{i=1}^{n_t} \text{Var}(Z''_{i,\ell}) \\ & \quad + c_{3,2} (T - n_t)^2 \times \\ & \quad \times \mathbb{E} [\log^2 \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \\ & \leq c_{3,2} \left(\frac{n_t}{T\rho_0} + n_t \right)^2 n_r n_t + c_{3,2} \left(1 + \frac{n_t}{T\rho_0} \right)^2 n_t (T - n_t) \\ & \quad + c_{3,2} (T - n_t)^2 \times \\ & \quad \times \mathbb{E} [\log^2 \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \end{aligned} \quad (225)$$

where the second step follows from (191) and because $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$ for every random variable X ; the third step follows by bounding $d_{i,\ell}^2 \leq T\rho$, $\rho \geq \rho_0$, and $|T\rho - n_t|/(T\rho) \leq 1 + n_t/(T\rho_0)$, and by computing the variances of $Z'_{i,\ell}$ and $Z''_{i,\ell}$.

It remains to upper-bound the third term on the RHS of (225). To this end, we first note that

$$\begin{aligned} & \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) \\ & \leq \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1}) \\ & \leq \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t}) \\ & \leq n_t \log (1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)) \end{aligned} \quad (226)$$

where

$$\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) \triangleq \text{tr} (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell). \quad (227)$$

In (226), the first two inequalities follow because $\det(\mathbf{A} + \mathbf{B}) \geq \det(\mathbf{A})$ for every pair of positive definite matrices \mathbf{A} and \mathbf{B} [39, Th. 7.8.21] and because $(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1} \preceq \mathbf{I}_{n_t}$; the last inequality follows from Hadamard's inequality [39, Th. 7.8.1] and because every diagonal element of $\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}$ is upper-bounded by $\text{tr} (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})$. Therefore,

$$\begin{aligned} & \mathbb{E} [\log^2 \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \\ & \leq c_{2,2} \mathbb{E} [\log^2 \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \\ & \quad + c_{2,2} n_t^2 \mathbb{E} [\log^2 (1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell))]. \end{aligned} \quad (228)$$

By Lemma 7, the first expected value on the RHS of (228) is

$$\begin{aligned} \mathbb{E} [\log^2 \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \\ = \left(\sum_{i=0}^{n_t-1} \Psi(n_r - i) \right)^2 + \sum_{i=0}^{n_t-1} \Psi'(n_r - i). \end{aligned} \quad (229)$$

To bound the second expected value, we use that the function $x \mapsto \log^2(1+x)$ is concave for $x \geq e-1$. Consequently,

$$\begin{aligned} \mathbb{E} [\log^2(1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell))] \\ = \mathbb{E} [\log^2(1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)) \mathbb{1}\{\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) \leq 2\}] \\ + \mathbb{E} [\log^2(1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)) \mathbb{1}\{\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2\}] \\ \leq \Pr[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2] \log^2 \left(1 + \frac{\mathbb{E}[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)]}{\Pr[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2]} \right) \\ + \log^2(3) \\ \leq \sup_{0 < \delta \leq 1} \left\{ \delta \log^2 \left(1 + \frac{n_t n_r + (T - n_r) n_r}{\delta} \right) \right\} \\ + \log^2(3) \end{aligned} \quad (230)$$

where the second step follows by applying Jensen's inequality to the second expected value and by noting that

$$\mathbb{E}[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) \mid \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2] \leq \frac{\mathbb{E}[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)]}{\Pr[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2]};$$

the last step follows because $\mathbb{E}[\text{tr}(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] = n_t n_r$ and $\mathbb{E}[\lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)] \leq \mathbb{E}[\text{tr}(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)] = (T - n_t) n_r$, and by optimizing over $\delta = \Pr[\Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell) > 2]$.

Since $\delta \mapsto \delta \log^2(1 + (n_t n_r + (T - n_r) n_r)/\delta)$ is continuous on $0 < \delta \leq 1$ and vanishes as $\delta \downarrow 0$, (225)–(230) yield that

$$\sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{U}(\mathbf{D}, T, \rho) < \infty. \quad (231)$$

This is (66a).

B. Proof of (66b)

By the definitions of $\bar{j}(\mathbf{D}_\ell, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$ and $\bar{J}(\mathbf{D}_\ell, T, \rho)$ in (61) and (63), we have that

$$\begin{aligned} \bar{S}(\mathbf{D}, T, \rho) \\ = \mathbb{E} \left[\left| \sum_{i=1}^{n_t} \left(\frac{n_t}{T\rho} + \frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) (Z'_{i,\ell} - \mathbb{E}[Z'_{i,\ell}]) \right. \right. \\ \left. \left. - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} (Z''_{i,\ell} - \mathbb{E}[Z''_{i,\ell}]) \right. \right. \\ \left. \left. + (T - n_t) \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1}) \right. \right. \\ \left. \left. - (T - n_t) \times \right. \right. \\ \left. \left. \times \mathbb{E} [\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \right|^3 \right] \\ \leq 2c_{2+4n_t,3} \left(\sum_{i=1}^{n_t} \left| \frac{n_t}{T\rho} + \frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right|^3 \mathbb{E}[|Z'_{i,\ell}|^3] \right. \\ \left. + \left| \frac{T\rho - n_t}{T\rho} \right|^3 \sum_{i=1}^{n_t} \mathbb{E}[|Z''_{i,\ell}|^3] + (T - n_t)^3 \times \right. \\ \left. \times \mathbb{E} [\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \right)^3 \end{aligned}$$

$$\begin{aligned} &\leq 2c_{2+4n_t,3} n_t \left| \frac{n_t}{T\rho_0} + n_t \right|^3 \frac{\Gamma(n_r + 3)}{\Gamma(n_r)} \\ &+ 2c_{2+4n_t,3} n_t \left| 1 + \frac{n_t}{T\rho_0} \right|^3 \frac{\Gamma(T - n_t + 3)}{\Gamma(T - n_t + 3)} \\ &+ 2c_{2+4n_t,3} (T - n_t)^3 \times \\ &\quad \times \mathbb{E} [\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})]^3 \end{aligned} \quad (232)$$

where the second step follows from (191) and because $\mathbb{E}[X]^3 \leq \mathbb{E}[|X|^3]$ for every random variable X ; the third step follows because by bounding $d_{i,\ell}^2 \leq T\rho$, $\rho \geq \rho_0$, and $|T\rho - n_t|/(T\rho) \leq 1 + n_t/(T\rho_0)$, and by evaluating the third moment of the gamma-distributed random variable $Z''_{i,\ell}$.

It remains to upper-bound the third term on the RHS of (232). By (226) and (191), we have that

$$\begin{aligned} &\left| \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1}) \right|^3 \\ &\leq c_{2,3} |\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})|^3 \\ &\quad + c_{2,3} n_t^3 \log^3 (1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell)). \end{aligned} \quad (233)$$

The expected value of $|\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})|^3$ is finite. Moreover, the function $x \mapsto \log^3(1+x)$ is concave for $x \geq e^2 - 1$. By following similar steps as in (230), it can thus be shown that

$$\begin{aligned} &\mathbb{E} [\log^3(1 + \Lambda(\mathbf{H}_\ell, \mathbf{Q}_\ell))] \leq \log^3(8) \\ &\quad + \sup_{0 < \delta \leq 1} \left\{ \delta \log^3 \left(1 + \frac{n_t n_r + (T - n_r) n_r}{\delta} \right) \right\}. \end{aligned} \quad (234)$$

Since $\delta \mapsto \delta \log^3(1 + (n_t n_r + (T - n_r) n_r)/\delta)$ is continuous on $0 < \delta \leq 1$ and vanishes as $\delta \downarrow 0$, (232)–(234) yield that

$$\sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{S}(\mathbf{D}, T, \rho) < \infty. \quad (235)$$

This is (66b).

APPENDIX V

PROOF OF LEMMA 9

Recall that (cf. (61) and (63))

$$\begin{aligned} \bar{U}(\mathbf{D}, T, \rho) \\ = \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{n_t}{T\rho} + \frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) (Z'_{i,\ell} - \mathbb{E}[Z'_{i,\ell}]) \right. \right. \\ \left. \left. - \frac{T\rho - n_t}{T\rho} \sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \right. \right. \\ \left. \left. + (T - n_t) \log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1}) \right. \right. \\ \left. \left. - (T - n_t) \times \right. \right. \\ \left. \left. \times \mathbb{E} [\log \det (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \right)^2 \right]. \end{aligned} \quad (236)$$

To prove Lemma 9, we first show that

$$\bar{U}(\mathbf{D}, T, \rho) = \bar{U}^*(\mathbf{D}, T, \rho) + K_{\bar{U}}(\mathbf{D}, T, \rho) \quad (237)$$

where $K_{\bar{U}}(\mathbf{D}, T, \rho)$ is a function of \mathbf{D} , T , and ρ that satisfies

$$\lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathbf{D}, T, \rho)| = 0; \quad (238)$$

$\bar{U}^*(\mathbf{D}, T, \rho) \triangleq \text{Var}(\bar{j}^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell))$; and

$$\begin{aligned} \bar{j}^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell) \\ \triangleq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) Z'_{i,\ell} \\ - \sum_{i=1}^{n_t} Z''_{i,\ell} + (T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \\ + (T - n_t) \log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}). \end{aligned} \quad (239)$$

We prove (237) by bounding $(\bar{U}(\mathbf{D}, T, \rho) - \bar{U}^*(\mathbf{D}, T, \rho))^2$ as follows. For ease of exposition, let

$$\begin{aligned} \Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \\ \triangleq \bar{j}(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \bar{J}(\mathbf{D}, T, \rho) \end{aligned} \quad (240a)$$

$$\begin{aligned} \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell) \\ \triangleq \bar{j}^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell) - \bar{J}^*(\mathbf{D}, T, \rho). \end{aligned} \quad (240b)$$

It follows that

$$\begin{aligned} & (\bar{U}(\mathbf{D}, T, \rho) - \bar{U}^*(\mathbf{D}, T, \rho))^2 \\ &= \left(\mathbb{E}[(\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell)) \times \right. \\ & \quad \times (\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) + \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell))] \right)^2 \\ & \leq \mathbb{E}[(\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell))^2] \times \\ & \quad \times c_{2,2}(\bar{U}(\mathbf{D}, T, \rho) + \bar{U}^*(\mathbf{D}, T, \rho)) \end{aligned} \quad (241)$$

where the second step follows from the Cauchy-Schwarz inequality and from (191). By Lemma 8, $\bar{U}(\mathbf{D}, T, \rho)$ is bounded in ρ and \mathbf{D} . Similarly, it follows from (191) that

$$\begin{aligned} \bar{U}^*(\mathbf{D}, T, \rho) \\ \leq c_{3,2} \frac{(n_t T \alpha - T \rho)^2}{(T \rho)^2} \text{Var} \left(\sum_{i=1}^{n_t} Z'_{i,\ell} \right) + c_{3,2} \text{Var} \left(\sum_{i=1}^{n_t} Z''_{i,\ell} \right) \\ + c_{3,2} (T - n_t)^2 \text{Var}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \\ = c_{3,2} \frac{(n_t T \alpha - T \rho)^2}{(T \rho)^2} n_r n_t + c_{3,2} n_t (T - n_t) \\ + c_{3,2} (T - n_t)^2 \sum_{i=0}^{n_t-1} \Psi'(n_r - i) \end{aligned} \quad (242)$$

where we used that the variances of $Z'_{i,\ell}$, $Z''_{i,\ell}$, and $\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})$ are n_t , $(T - n_t)$, and $\sum_{i=0}^{n_t-1} \Psi'(n_r - i)$, respectively. Consequently, $\bar{U}^*(\mathbf{D}, T, \rho)$ is bounded in (ρ, \mathbf{D}) .

We thus obtain (237) by showing that the first expected value on the RHS of (241) vanishes as $\rho \rightarrow \infty$ uniformly in $\mathbf{D} \in \mathcal{D}_1$. To this end, we first note that

$$\begin{aligned} & \Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell) \\ &= \frac{n_t}{T\rho} \sum_{i=1}^{n_t} (Z'_{i,\ell} - n_r) + \frac{n_t}{T\rho} \sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \\ & \quad + (T - n_t) \log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \\ & \quad - (T - n_t) \times \\ & \quad \times \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right]. \end{aligned} \quad (243)$$

Using that $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$, we can thus upper-bound the first expected value on the RHS of (241) as

$$\begin{aligned} & \mathbb{E} \left[(\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell))^2 \right] \\ & \leq c_{3,2} \left(\left(\frac{n_t}{T\rho} \right)^2 n_t n_r + \left(\frac{n_t}{T\rho} \right)^2 n_t (T - n_t) + (T - n_t)^2 \times \right. \\ & \quad \times \left. \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right) \right] \right). \end{aligned} \quad (244)$$

Since $\det(\mathbf{A} + \mathbf{B}) \geq \det(\mathbf{A})$ for every pair of positive definite matrices \mathbf{A} and \mathbf{B} [39, Th. 7.8.21], the fraction inside the logarithm function is greater than or equal to one. Consequently, an upper bound on this fraction yields an upper bound on the squared logarithm. Using that $d_i^2 \geq \bar{\delta}\rho$, $i = 1, \dots, n_t$, for $\mathbf{D} \in \mathcal{D}_1$, we obtain $(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1} \leq (1 + \bar{\delta}\rho)^{-1} \mathbf{I}_{n_t}$, so for $\mathbf{D} \in \mathcal{D}_1$, we can further upper-bound (244) as

$$\begin{aligned} & \mathbb{E} \left[(\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) - \Delta^*(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell))^2 \right] \\ & \leq c_{3,2} \left(\left(\frac{n_t}{T\rho} \right)^2 n_t n_r + \left(\frac{n_t}{T\rho} \right)^2 n_t (T - n_t) + (T - n_t)^2 \times \right. \\ & \quad \times \left. \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(1 + \bar{\delta}\rho)^{-1} \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right) \right] \right). \end{aligned} \quad (245)$$

The RHS of (245) is independent of \mathbf{D} and vanishes by the dominated convergence theorem. Indeed, we have

$$\begin{aligned} & \lim_{\rho \rightarrow \infty} \log^2 \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(1 + \bar{\delta}\rho)^{-1} \mathbf{I}_{n_t}) \\ &= \log^2 \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) \end{aligned} \quad (246)$$

by the continuity of the logarithm function and the determinant. Furthermore, by (191),

$$\begin{aligned} & \log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(1 + \bar{\delta}\rho)^{-1} \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right) \\ & \leq c_{2,2} \log^2 \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) \\ & \quad + c_{2,2} \log^2 \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t}) \end{aligned} \quad (247)$$

because the numerator on the LHS of (247) is monotonically decreasing in ρ . The expected value of the RHS of (245) is finite (cf. (228)–(230)), hence the dominated convergence theorem applies and

$$\lim_{\rho \rightarrow \infty} \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(1 + \bar{\delta}\rho)^{-1} \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right) \right] = 0. \quad (248)$$

From (241)–(248), we obtain that

$$\lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} (\bar{U}(\mathbf{D}, T, \rho) - \bar{U}^*(\mathbf{D}, T, \rho))^2 = 0 \quad (249)$$

which is (237).

We next lower-bound $\bar{U}^*(\mathbf{D}, T, \rho)$ for $\mathbf{D} \in \mathcal{D}_1$. Indeed,

$$\begin{aligned}
 & \bar{U}^*(\mathbf{D}, T, \rho) \\
 &= \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) (Z'_{i,\ell} - n_r) - \sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \right. \right. \\
 &\quad + (T - n_t) \log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) \\
 &\quad \left. \left. - (T - n_t) \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \right)^2 \right] \\
 &= \text{Var} \left(\sum_{i=1}^{n_t} Z''_{i,\ell} \right) + \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\rho} - 1 \right) (Z'_{i,\ell} - n_r) \right. \right. \\
 &\quad \left. \left. + (T - n_t) (\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}) - \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})]) \right)^2 \right] \\
 &\geq \mathbb{E} \left[\left(\sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \right)^2 \right] \\
 &= (T - n_t) n_t
 \end{aligned} \tag{250}$$

where the second step follows because $\{Z''_{1,1}, \dots, Z''_{n_t,L}\}$ and $\{\mathbf{H}^L, Z'_{1,1}, \dots, Z'_{n_t,L}\}$ are independent, and the last step follows because $\{Z''_{1,1}, \dots, Z''_{n_t,L}\}$ are i.i.d. random variables with variance $(T - n_t)$.

Combining (250) with (237), we obtain that

$$\bar{U}(\mathbf{D}, T, \rho) \geq (T - n_t) n_t + K_{\bar{U}}(\mathbf{D}, T, \rho) \tag{251}$$

which proves Lemma 9.

APPENDIX VI PROOF OF LEMMA 10

Recall that $\bar{U}(\mathbf{D}, T, \rho) = \mathbb{E}[\Delta^2(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)]$, where $\Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)$ was defined in (240a). Further let

$$\begin{aligned}
 \Delta_-(x, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) &\triangleq \Delta(\sqrt{x}\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \\
 &\quad - \Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)
 \end{aligned} \tag{252a}$$

$$\begin{aligned}
 \Delta_+(x, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) &\triangleq \Delta(\sqrt{x}\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \\
 &\quad + \Delta(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)
 \end{aligned} \tag{252b}$$

for $x > 0$. Using that

$$\begin{aligned}
 & \Delta^2 \left(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) - \Delta^2(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell) \\
 &= \Delta_- \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \Delta_+ \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right)
 \end{aligned} \tag{253}$$

it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
 & \bar{U} \left(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, T, \rho \right) - \bar{U}(\mathbf{D}, T, \rho) \\
 &\leq \sqrt{\mathbb{E} \left[\Delta_-^2 \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \right]} \times \\
 &\quad \times \sqrt{\mathbb{E} \left[\Delta_+^2 \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \right]} \tag{254}
 \end{aligned}$$

We next note that

$$\begin{aligned}
 & \sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \mathbb{E} \left[\Delta_+^2 \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \right] \\
 &\leq 2 \sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{U} \left(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, T, \rho \right) \\
 &\quad + 2 \sup_{\rho \geq \rho_0} \sup_{\mathbf{D} \in \mathcal{D}_{n_t}} \bar{U}(\mathbf{D}, T, \rho) \\
 &\leq 4 \bar{U}(T)
 \end{aligned} \tag{255}$$

for some constant $\bar{U}(T)$ that only depends on T and any $\rho_0 \geq 0$. Here, the first inequality follows from (252b), the inequality $(a_1 + a_2)^2 \leq 2a_1^2 + 2a_2^2$, and because $\bar{U}(\mathbf{D}, T, \rho) = \mathbb{E}[\Delta^2(\mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell)]$; the second inequality follows from Lemma 8. We further have that

$$\begin{aligned}
 & \mathbb{E} \left[\Delta_+^2 \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \right] \\
 &= \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{\rho}{\alpha} - 1 \right) \frac{n_t d_{i,\ell}^2}{T\rho} (Z'_{i,\ell} - n_r) \right. \right. \\
 &\quad + (T - n_t) \log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \\
 &\quad \left. \left. - (T - n_t) \times \right. \right. \\
 &\quad \times \left. \left. \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \right] \right)^2 \right].
 \end{aligned} \tag{256}$$

Consequently, using the inequalities $\mathbb{E}[(X - \mathbb{E}[X])^2] \leq \mathbb{E}[X^2]$ and $(a_1 + a_2)^2 \leq 2a_1^2 + 2a_2^2$, we obtain

$$\begin{aligned}
 & \mathbb{E} \left[\Delta_+^2 \left(\frac{\rho}{\alpha}, \mathbf{D}, \mathbf{Z}'_\ell, \mathbf{Z}''_\ell, \mathbf{H}_\ell, \mathbf{Q}_\ell \right) \right] \\
 &\leq 2 \left(\frac{\rho}{\alpha} - 1 \right)^2 \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \frac{n_t d_{i,\ell}^2}{T\rho} (Z'_{i,\ell} - n_r) \right)^2 \right] \\
 &\quad + 2(T - n_t)^2 \times \\
 &\quad \times \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \right) \right] \\
 &\leq 2 \left(\frac{\rho}{\alpha} - 1 \right)^2 n_t^3 n_r + 2(T - n_t)^2 \times \\
 &\quad \times \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}(\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \right) \right] \\
 &\leq 2 n_t^2 \left(\frac{\rho}{\alpha} - 1 \right)^2 (n_t n_r + (T - n_t)^2)
 \end{aligned} \tag{257}$$

where the second step follows by bounding $d_{i,\ell}^2 \leq T\rho$ and because $Z'_{1,\ell}, \dots, Z'_{n_t,\ell}$ are i.i.d. with variance n_r ; and the third step follows by upper-bounding the second term as shown below (see (259)–(263)).

Combining (254)–(257), and using that, by the lemma's assumptions, we have $T\alpha \geq T\rho(1 - \delta)$ and $0 < \delta < 1/2$, so $(\rho/\alpha - 1)^2 \leq \delta^2/(1 - \delta)^2 \leq 4\delta^2$, we obtain that $\bar{U}(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, T, \rho) - \bar{U}(\mathbf{D}, T, \rho)$ is upper-bounded by

$$\bar{U} \left(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, T, \rho \right) - \bar{U}(\mathbf{D}, T, \rho) \leq \delta \Upsilon(T) \tag{258}$$

where $\Upsilon(T) \triangleq 4n_t \sqrt{2\bar{U}(T)(n_t n_r + (T - n_t)^2)}$.

It remains to prove the last step in (257). To this end, we first note that, since $\frac{\rho}{\alpha} \geq 1$,

$$\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \geq 1. \quad (259)$$

An upper bound on the expected value in the second-to-last line in (257) thus follows by upper-bounding the logarithm inside the expected value. We have

$$\begin{aligned} \mathbb{E} \left[\log^2 \left(\frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \frac{\rho}{\alpha} \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} (\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})} \right) \right] \\ = \mathbb{E} \left[\log^2 \det \left(\mathbf{I}_{n_t} + \mathbf{A}_\ell^{-1} \left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2 \right) \right] \end{aligned} \quad (260)$$

where

$$\mathbf{A}_\ell \triangleq (\mathbf{I}_{n_t} + \mathbf{D}^2) + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})^{-1}. \quad (261)$$

This follows by writing $\frac{\rho}{\alpha} \mathbf{D}^2 = \mathbf{D}^2 + \left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2$ and by multiplying the matrices inside the determinants in the numerator and denominator by $\mathbf{A}_\ell^{-1} (\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})^{-1}$.⁷

Using that $\log \det(\mathbf{A}) \leq \text{tr}(\mathbf{A} - \mathbf{I}_{n_t})$ for every $n_t \times n_t$ positive-definite matrix \mathbf{A} , (260) can be upper-bounded as

$$\begin{aligned} \mathbb{E} \left[\log^2 \det \left(\mathbf{I}_{n_t} + \mathbf{A}_\ell^{-1} \left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2 \right) \right] \\ \leq \mathbb{E} \left[\left(\text{tr} \left(\mathbf{A}_\ell^{-1} \left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2 \right) \right)^2 \right] \\ \leq \mathbb{E} \left[\left(\sum_{k=1}^{n_t} \lambda_k(\mathbf{A}_\ell^{-1}) \lambda_k \left(\left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2 \right) \right)^2 \right] \\ \leq \left(\sum_{k=1}^{n_t} \frac{1}{\lambda_k(\mathbf{I}_{n_t} + \mathbf{D}^2)} \lambda_k \left(\left(\frac{\rho}{\alpha} - 1 \right) \mathbf{D}^2 \right) \right)^2 \\ = \left(\frac{\rho}{\alpha} - 1 \right)^2 \left(\sum_{k=1}^{n_t} \frac{d_k^2}{1 + d_k^2} \right)^2 \\ \leq \left(\frac{\rho}{\alpha} - 1 \right)^2 n_t^2 \end{aligned} \quad (262)$$

where $\lambda_k(\mathbf{A})$ denotes the k -th largest eigenvalue of the matrix \mathbf{A} . Here, the second step is by von Neumann's trace inequality [39, Th. 8.7.6]; the third step follows because, by Weyl's theorem [39, Th. 4.3.1],

$$\lambda_k(\mathbf{A}_\ell^{-1}) = \frac{1}{\lambda_k(\mathbf{A}_\ell)} \leq \frac{1}{\lambda_k(\mathbf{I}_{n_t} + \mathbf{D}^2)}. \quad (263)$$

Combining (260) and (262) yields (257) and concludes the proof of Lemma 10.

APPENDIX VII PROOF OF LEMMA 11

We begin by noting that, for every $\mathbf{D} \in \mathcal{D}_{n_t}$,

$$\bar{J}(\mathbf{D}, T, \rho) = \bar{J}^*(\mathbf{D}, T, \rho) + K_{\mathcal{D}_1}(\mathbf{D}, T, \rho) \quad (264)$$

⁷The matrix $\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H}$ has a complex Wishart distribution so, for $n_r \geq n_t$, it is invertible with probability one [42, Th. 3.1.4].

where $\bar{J}^*(\mathbf{D}, T, \rho)$ is defined in (82) and

$$\begin{aligned} K_{\mathcal{D}_1}(\mathbf{D}, T) &\triangleq \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) + (T - n_t) \times \\ &\times \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right]. \end{aligned} \quad (265)$$

Since $\det(\mathbf{A} + \mathbf{B}) \geq \det(\mathbf{A})$ for every pair of positive definite matrices \mathbf{A} and \mathbf{B} [39, Th. 7.8.21], we have $K_{\mathcal{D}_1}(\mathbf{D}, T) \geq 0$. Furthermore, for $\mathbf{D} \in \mathcal{D}_1$, we have $d_i^2 \geq \bar{\rho}$, $i = 1, \dots, n_t$. This implies that $(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1} \leq (1 + \bar{\rho})^{-1} \mathbf{I}_{n_t}$. Consequently,

$$\begin{aligned} K_{\mathcal{D}_1}(\mathbf{D}, T) \\ \leq \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) + (T - n_t) \times \\ \times \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (1 + \bar{\rho})^{-1} \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right] \\ \triangleq K_{\mathcal{D}_1}(T, \rho), \quad \mathbf{D} \in \mathcal{D}_1. \end{aligned} \quad (266)$$

We further have that

$$\begin{aligned} &|\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (1 + \bar{\rho})^{-1} \mathbf{I}_{n_t})| \\ &\leq |\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})| \\ &\quad + \log(1 + \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})) \end{aligned} \quad (267)$$

whose expected value is finite (cf. (226)–(230)).⁸ By the continuity of the logarithm function and the determinant, and by the dominated convergence theorem, we thus have that

$$\begin{aligned} \lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) (1 + \bar{\rho})^{-1} \mathbf{I}_{n_t})] \\ = \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})]. \end{aligned} \quad (268)$$

Combining (268) with (266) yields then that

$$\bar{J}(\mathbf{D}, T, \rho) \leq \bar{J}^*(\mathbf{D}, T, \rho) + K_{\mathcal{D}_1}(T, \rho) \quad (269)$$

for a nonnegative constant $K_{\mathcal{D}_1}(T, \rho)$ that only depends on T and ρ and that satisfies $\lim_{\rho \rightarrow \infty} K_{\mathcal{D}_1}(T, \rho) = 0$. Since $K_{\mathcal{D}_1}(T, \rho)$ does not depend on $\mathbf{D} \in \mathcal{D}_1$, this gives

$$\begin{aligned} &\sup_{\mathbf{D} \in \mathcal{D}_1} \left\{ \bar{J}(\mathbf{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathbf{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ &\leq \sup_{\mathbf{D} \in \mathcal{D}_1} \left\{ \bar{J}^*(\mathbf{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathbf{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ &\quad + K_{\mathcal{D}_1}(T, \rho). \end{aligned} \quad (270)$$

We next show that, for $\rho \geq \rho_0$, $L \geq L_0$ and sufficiently large ρ_0 and L_0 , the supremum over $\mathbf{D} \in \mathcal{D}_1$ can be replaced by a supremum over all matrices $\mathbf{D} \in \mathcal{D}_{n_t}$ that satisfy $\text{tr}(\mathbf{D}^2) \geq T\rho(1 - K(T)/L)$. To this end, we show first that, for $\rho \geq \rho^{(1)}$ and a sufficiently large $\rho^{(1)}$, we can assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \frac{K^{(1)}(T, \rho^{(1)})}{\sqrt{L}} \right) \quad (271)$$

for a constant $K^{(1)}(T, \rho^{(1)})$ that depends on T and $\rho^{(1)}$. We then gradually improve this bound on the trace until we obtain the desired result.

⁸Equations (229)–(230) demonstrate that the expected values of the squared logarithms are finite. This implies that the expected values of the absolute values of the logarithms are finite, too, since by Jensen's inequality $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}$ for any random variable X .

A. Lower Bound on $\text{tr}(\mathbf{D}^2)$: First Round

Recall that $\alpha = \text{tr}(\mathbf{D}^2)/T$, and define $\bar{\mathbf{D}} \triangleq \mathbf{D}/\sqrt{\alpha}$. Clearly, $\text{tr}(\bar{\mathbf{D}}^2) = T$ and, for $\mathbf{D} \in \mathcal{D}_1$, $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$. We then compare

$$\bar{J}^*(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho) - \sqrt{\frac{\bar{U}(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho)}{L}} Q^{-1}(\epsilon) \quad (272)$$

with

$$\bar{J}^*(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho) - \sqrt{\frac{\bar{U}(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho)}{L}} Q^{-1}(\epsilon). \quad (273)$$

Intuitively, every α for which (272) is larger than (273) is suboptimal and can be discarded without loss of optimality, since multiplying the matrix \mathbf{D} by $\sqrt{\rho/\alpha}$ would increase the difference (273).

To formalize this argument, we define the function

$$\begin{aligned} f_{L,\rho}(\alpha, \bar{\mathbf{D}}) &\triangleq \bar{J}^*(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho) - \sqrt{\frac{\bar{U}(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho)}{L}} Q^{-1}(\epsilon) \\ &\quad - \bar{J}^*(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho) + \sqrt{\frac{\bar{U}(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho)}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (274)$$

and analyze for which values of $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$ it is nonnegative. Indeed, Lemma 8 implies that there exists an $\bar{U}(T)$ that depends on T and an arbitrary $\rho_0 > 0$ such that

$$0 \leq \bar{U}(\mathbf{D}, T, \rho) \leq \bar{U}(T), \quad \mathbf{D} \in \mathcal{D}_{n_t}, \rho \geq \rho_0. \quad (275)$$

Consequently, when $T = n_t + n_r$, the function $f_{L,\rho}(\cdot)$ can be lower-bounded as

$$\begin{aligned} f_{L,\rho}(\alpha, \bar{\mathbf{D}}) &= \frac{T\rho - T\alpha}{T\rho} n_t n_t \\ &\quad - \left(\sqrt{\frac{\bar{U}(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho)}{L}} - \sqrt{\frac{\bar{U}(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho)}{L}} \right) Q^{-1}(\epsilon) \\ &\geq \frac{T\rho - T\alpha}{T\rho} n_t n_t - \sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (276)$$

which is positive if

$$\alpha < \rho \left(1 - \frac{\sqrt{\bar{U}(T)} Q^{-1}(\epsilon)}{\sqrt{L}} \right). \quad (277)$$

It follows that we can assume without loss of optimality that $\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - K^{(1)}(T, \rho_0)/\sqrt{L} \right)$, where

$$K^{(1)}(T, \rho_0) = \sqrt{\bar{U}(T)} \frac{Q^{-1}(\epsilon)}{n_t n_r}. \quad (278)$$

Similarly, when $T > n_t + n_r$, the function $f_{L,\rho}(\cdot)$ can be lower-bounded as

$$\begin{aligned} f_{L,\rho}(\alpha, \bar{\mathbf{D}}) &= \frac{T\rho - T\alpha}{T\rho} n_t n_t + (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{\mathbf{D}}^2)}{\det(\mathbf{I}_{n_t} + \alpha \bar{\mathbf{D}}^2)} \\ &\quad - \left(\sqrt{\frac{\bar{U}(\sqrt{\rho} \bar{\mathbf{D}}, T, \rho)}{L}} - \sqrt{\frac{\bar{U}(\sqrt{\alpha} \bar{\mathbf{D}}, T, \rho)}{L}} \right) Q^{-1}(\epsilon) \end{aligned}$$

$$\begin{aligned} &\geq (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{\mathbf{D}}^2)}{\det(\mathbf{I}_{n_t} + \alpha \bar{\mathbf{D}}^2)} - \sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon) \\ &\triangleq \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{\mathbf{D}}). \end{aligned} \quad (279)$$

Using that $\bar{\mathbf{D}}$ is a diagonal matrix, the first derivative of $\underline{f}_{L,\rho}^{(1)}(\cdot)$ with respect to α can be computed as

$$\frac{\partial}{\partial \alpha} \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{\mathbf{D}}) = -(T - n_t - n_r) \sum_{i=1}^{n_t} \frac{\bar{d}_i^2}{1 + \alpha \bar{d}_i^2} \quad (280)$$

where \bar{d}_i denotes the i -th diagonal element of $\bar{\mathbf{D}}$. Furthermore, by definition, we have $\bar{d}_i = d_i / \sqrt{\alpha}$, so for every $\mathbf{D} \in \mathcal{D}_1$

$$\bar{d}_i^2 \geq \frac{\bar{\delta} \rho}{\alpha} \geq \bar{\delta}, \quad i = 1, \dots, n_t \quad (281)$$

since, for such \mathbf{D} , we have that $d_i^2 \geq \bar{\delta} \rho$ and $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$. It follows that

$$\frac{\partial}{\partial \alpha} \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{\mathbf{D}}) \leq -(T - n_t - n_r) n_t \frac{\bar{\delta}}{1 + \alpha \bar{\delta}} \quad (282)$$

which is strictly negative and bounded away from zero for every $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$. Thus, $\underline{f}_{L,\rho}^{(1)}(\cdot)$ is a strictly decreasing function of $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$.

We next note that

$$\underline{f}_{L,\rho}^{(1)}(\rho, \bar{\mathbf{D}}) = -\sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon) \quad (283)$$

and

$$\begin{aligned} &\underline{f}_{L,\rho}^{(1)}\left(\frac{n_t \bar{\delta} \rho}{T}, \bar{\mathbf{D}}\right) \\ &= (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{\mathbf{D}}^2)}{\det(\mathbf{I}_{n_t} + \frac{n_t \bar{\delta} \rho}{T} \bar{\mathbf{D}}^2)} - \sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon) \\ &\geq (T - n_t - n_r) n_t \log \frac{1 + \rho \bar{\delta}}{1 + \frac{n_t \bar{\delta}^2}{T} \rho} - \sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (284)$$

where the inequality follows because the first term is monotonically increasing in the diagonal elements of $\bar{\mathbf{D}}$ and $\bar{d}_i^2 \geq \bar{\delta}$, $i = 1, \dots, n_t$. The first term on the RHS of (284) depends on ρ but not on L , and it converges to

$$(T - n_t - n_r) n_t \log \frac{T}{n_t \bar{\delta}} \quad (285)$$

as $\rho \rightarrow \infty$. Since $\bar{\delta}$, as defined in (54), is strictly smaller than T/n_t , this expression is strictly positive. Similarly, the second term on the RHS of (284) depends on L but not on ρ . Thus, we can find a $\rho^{(1)}$ such that, for $\rho \geq \rho^{(1)}$ and $T > n_t + n_r$,

$$(T - n_t - n_r) n_t \log \frac{1 + \rho \bar{\delta}}{1 + \frac{n_t \bar{\delta}^2}{T} \rho} \geq \frac{1}{2} \log \frac{T}{n_t \bar{\delta}}. \quad (286)$$

This in turn implies that

$$\underline{f}_{L,\rho}^{(1)}\left(\frac{n_t \bar{\delta} \rho}{T}, \bar{\mathbf{D}}\right) \geq \frac{1}{4} \log \frac{T}{n_t \bar{\delta}} > 0 \quad (287)$$

for $\rho \geq \rho^{(1)}$ and

$$L \geq L^{(1)} \triangleq 16 \bar{U}(T) \left(\frac{Q^{-1}(\epsilon)}{\log \frac{T}{n_t \bar{\delta}}} \right)^2. \quad (288)$$

We conclude that, when $T > n_t + n_r$, $\rho \geq \rho^{(1)}$, and $L \geq L^{(1)}$, $\alpha \mapsto \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{D})$ is a strictly decreasing function that is positive at $\alpha = n_t \bar{\delta} \rho / T$ and negative at $\alpha = \rho$. We can thus find an $\alpha_0^{(1)}(\bar{D}, L, \rho)$ such that $\underline{f}_{L,\rho}^{(1)}(\alpha, \bar{D}) > 0$ for $\alpha < \alpha_0^{(1)}(\bar{D}, L, \rho)$ and $\underline{f}_{L,\rho}^{(1)}(\alpha, \bar{D}) \leq 0$ for $\alpha \geq \alpha_0^{(1)}(\bar{D}, L, \rho)$. Defining

$$\delta_0^{(1)}(\bar{D}, L, \rho) \triangleq 1 - \frac{\alpha_0^{(1)}(\bar{D}, L, \rho)}{\rho} \quad (289)$$

this implies that we can assume without loss of optimality that $\alpha \geq \rho(1 - \delta_0^{(1)}(\bar{D}, L, \rho))$. We next show that, for $\rho \geq \rho^{(1)}$ and $L \geq L^{(1)}$, we have that $\delta_0^{(1)}(\bar{D}, L, \rho) \leq K^{(1)}(T, \rho^{(1)})/\sqrt{L}$ for a constant $K^{(1)}(T, \rho_0)$ that depends on T and $\rho^{(1)}$. It then follows that, for such ρ and L , we can assume without loss of optimality that $\text{tr}(\mathbf{D}^2) \geq T\rho(1 - K^{(1)}(T, \rho_0)/\sqrt{L})$.

To upper-bound $\delta_0^{(1)}(\bar{D}, L, \rho)$, we apply the mean value theorem [43, Th. 5.10] to express $\underline{f}_{L,\rho}^{(1)}(\cdot)$ as

$$\begin{aligned} \underline{f}_{L,\rho}^{(1)}(\rho, \bar{D}) &= \underline{f}_{L,\rho}^{(1)}(\rho, \bar{D}) - \underline{f}_{L,\rho}^{(1)}(\alpha_0^{(1)}(\bar{D}, L, \rho), \bar{D}) \\ &= \int_{\alpha_0^{(1)}(\bar{D}, L, \rho)}^{\rho} \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{D}) d\alpha \\ &= \rho \delta_0^{(1)}(\bar{D}, L, \rho) \dot{\underline{f}}_{L,\rho}^{(1)}(\tilde{\alpha}, \bar{D}) \end{aligned} \quad (290)$$

for some $\tilde{\alpha}$ between $\alpha_0^{(1)}(\bar{D}, L, \rho)$ and ρ . Here, we denote by $\dot{\underline{f}}_{L,\rho}^{(1)}(\cdot)$ the derivative of $\alpha \mapsto \underline{f}_{L,\rho}^{(1)}(\alpha, \bar{D})$ with respect to α . In (290), the first step follows because, by definition, $\underline{f}_{L,\rho}^{(1)}(\alpha_0^{(1)}(\bar{D}, L, \rho), \bar{D}) = 0$; the last step follows by the definition of $\delta_0^{(1)}(\bar{D}, L, \rho)$. Together with (280)–(283), this yields that, for every $\rho \geq \rho^{(1)}$,

$$\begin{aligned} \delta_0^{(1)}(\bar{D}, L, \rho) &= \frac{\underline{f}_{L,\rho}^{(1)}(\rho, \bar{D})}{\rho \dot{\underline{f}}_{L,\rho}^{(1)}(\tilde{\alpha}, \bar{D})} \\ &= \frac{\sqrt{\frac{\bar{U}(T)}{L}} Q^{-1}(\epsilon)}{\rho (T - n_t - n_r) \sum_{i=1}^{n_t} \frac{\bar{d}_i^2}{1 + \tilde{\alpha} \bar{d}_i^2}} \\ &\leq \frac{K^{(1)}(T, \rho^{(1)})}{\sqrt{L}} \end{aligned} \quad (291)$$

where

$$K^{(1)}(T, \rho^{(1)}) \triangleq \left(1 + \frac{1}{\bar{\delta} \rho^{(1)}}\right) \frac{\sqrt{\bar{U}(T)} Q^{-1}(\epsilon)}{(T - n_t - n_r) n_t}. \quad (292)$$

In (291), the third step follows by bounding $\bar{d}_i^2 \geq \bar{\delta}$, $\tilde{\alpha} \leq \rho$, and $\rho \geq \rho^{(1)}$.

B. Lower Bound on $\text{tr}(\mathbf{D}^2)$: k -th Round

Suppose that, for $\rho \geq \rho^{(k-1)}$ and $L \geq L^{(k-1)}$ (for some $\rho^{(k-1)}$ and $L^{(k-1)}$), we can assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \delta^{(k-1)}\right). \quad (293)$$

We next show that, in this case, we can also assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \sqrt{\frac{K(T)\delta^{(k-1)}}{L}}\right) \quad (294)$$

for $\rho \geq \rho_0$, $L \geq L_0$, and sufficiently large ρ_0 and L_0 that depend on $(T, \rho^{(1)}, L^{(1)})$ but not on (ρ, L, k) , where $K(T)$ is a constant that depends on T . Indeed, using Lemma 10, we can lower-bound $\bar{U}(\sqrt{\alpha}\bar{D}, T, \rho)$ as

$$\bar{U}(\sqrt{\alpha}\bar{D}, T, \rho) \geq \bar{U}(\sqrt{\rho}\bar{D}, T, \rho) - \Upsilon(T)\delta^{(k-1)} \quad (295)$$

where $\Upsilon(T)$ is a constant that depends on T . Together with the bound $\sqrt{a-b} \geq \sqrt{a} - \sqrt{b}$, $a \geq b$, this yields that

$$\sqrt{\frac{\bar{U}(\sqrt{\alpha}\bar{D}, T, \rho)}{L}} \geq \sqrt{\frac{\bar{U}(\sqrt{\rho}\bar{D}, T, \rho)}{L}} - \sqrt{\frac{\Upsilon(T)\delta^{(k-1)}}{L}}. \quad (296)$$

Note that the condition $\bar{U}(\sqrt{\rho}\bar{D}, T, \rho) \geq \Upsilon(T)\delta^{(k-1)}$, required to apply the inequality $\sqrt{a-b} \geq \sqrt{a} - \sqrt{b}$, is satisfied for sufficiently large ρ and L . Indeed, by Lemma 13, we have that $\bar{U}(\sqrt{\rho}\bar{D}, T, \rho) \geq T^2 \tilde{V}(T)/2$ for $\rho \geq \rho'$ and some sufficiently large ρ' , where $\tilde{V}(T)$ was defined in (10). Furthermore, $\delta^{(1)} = K^{(1)}(T, \rho^{(1)})/\sqrt{L}$ and, as we shall show in (314)–(317) below, we have $\delta^{(k)} \leq \delta^{(k-1)}$, $k = 2, 3, \dots$ for $L \geq L^\circ$ with L° defined in (316). We thus obtain that $\bar{U}(\sqrt{\rho}\bar{D}, T, \rho) \geq \Upsilon(T)\delta^{(k-1)}$, $k = 2, 3, \dots$ when $\rho \geq \rho'$ and

$$L \geq L' \triangleq \max \left\{ L^\circ, \left(\frac{2\Upsilon(T)K^{(1)}(T, \rho^{(1)})}{T^2 \tilde{V}(T)} \right)^2 \right\}. \quad (297)$$

We next use (296) to lower-bound $f_{L,\rho}(\cdot)$. Indeed, when $T = n_t + n_r$, we have

$$\begin{aligned} f_{L,\rho}(\alpha, \bar{D}) &= \frac{T\rho - T\alpha}{T\rho} n_t n_t \\ &\quad - \left(\sqrt{\frac{\bar{U}(\sqrt{\rho}\bar{D}, T, \rho)}{L}} - \sqrt{\frac{\bar{U}(\sqrt{\alpha}\bar{D}, T, \rho)}{L}} \right) Q^{-1}(\epsilon) \\ &\geq \frac{T\rho - T\alpha}{T\rho} n_t n_t - \sqrt{\frac{\Upsilon(T)\delta^{(k-1)}}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (298)$$

which is positive if

$$\alpha < \rho \left(1 - \sqrt{\frac{\left(\frac{Q^{-1}(\epsilon)}{n_t n_r}\right)^2 \Upsilon(T)\delta^{(k-1)}}{L}}\right). \quad (299)$$

It follows that we can assume without loss of optimality that (294) holds for $L_0 = L'$ and

$$K(T) = \Upsilon(T) \left(\frac{Q^{-1}(\epsilon)}{n_t n_r} \right)^2. \quad (300)$$

We next consider the case where $T > n_t + n_r$. In this case, we can lower-bound $f_{L,\rho}(\cdot)$ as

$$\begin{aligned} f_{L,\rho}(\alpha, \bar{D}) &= \frac{T\rho - T\alpha}{T\rho} n_t n_t + (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{D}^2)}{\det(\mathbf{I}_{n_t} + \alpha \bar{D}^2)} \\ &\quad - \left(\sqrt{\frac{\bar{U}(\sqrt{\rho}\bar{D}, T, \rho)}{L}} - \sqrt{\frac{\bar{U}(\sqrt{\alpha}\bar{D}, T, \rho)}{L}} \right) Q^{-1}(\epsilon) \end{aligned}$$

$$\begin{aligned} &\geq (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{\mathbf{D}}^2)}{\det(\mathbf{I}_{n_t} + \alpha \bar{\mathbf{D}}^2)} \\ &\quad - \sqrt{\frac{\Upsilon(T) \delta^{(k-1)}}{L}} Q^{-1}(\epsilon) \\ &\triangleq \underline{f}_{L,\rho}^{(k)}(\alpha, \bar{\mathbf{D}}). \end{aligned} \quad (301)$$

Since $\underline{f}_{L,\rho}^{(k)}(\cdot)$ differs from $\underline{f}_{L,\rho}^{(1)}(\cdot)$ only in terms that are independent of α , it follows that

$$\frac{\partial}{\partial \alpha} \underline{f}_{L,\rho}^{(k)}(\alpha, \bar{\mathbf{D}}) = \dot{\underline{f}}_{L,\rho}^{(1)}(\alpha, \bar{\mathbf{D}}). \quad (302)$$

Consequently, $\alpha \mapsto \underline{f}_{L,\rho}^{(k)}(\alpha)$ is a strictly decreasing function on $n_t \bar{\delta} \rho / T \leq \alpha \leq \rho$. Furthermore,

$$\underline{f}_{L,\rho}^{(k)}(\rho, \bar{\mathbf{D}}) = -\sqrt{\frac{\Upsilon(T) \delta^{(k-1)}}{L}} Q^{-1}(\epsilon) \quad (303)$$

and

$$\begin{aligned} &\underline{f}_{L,\rho}^{(k)}\left(\frac{n_t \bar{\delta} \rho}{T}, \bar{\mathbf{D}}\right) \\ &= (T - n_t - n_r) \log \frac{\det(\mathbf{I}_{n_t} + \rho \bar{\mathbf{D}}^2)}{\det(\mathbf{I}_{n_t} + \frac{n_t \bar{\delta} \rho}{T} \bar{\mathbf{D}}^2)} \\ &\quad - \sqrt{\frac{\Upsilon(T) \delta^{(k-1)}}{L}} Q^{-1}(\epsilon) \\ &\geq (T - n_t - n_r) n_t \log \frac{1 + \rho \bar{\delta}}{1 + \frac{n_t \bar{\delta}^2}{T} \rho} \\ &\quad - \sqrt{\frac{\Upsilon(T) \delta^{(k-1)}}{L}} Q^{-1}(\epsilon) \end{aligned} \quad (304)$$

similar to (284). Following the same steps as in (285)–(287), we can lower-bound the function $\underline{f}_{L,\rho}^{(k)}(\cdot)$ as

$$\underline{f}_{L,\rho}^{(k)}\left(\frac{n_t \bar{\delta} \rho}{T}, \bar{\mathbf{D}}\right) \geq \frac{1}{4} \log \frac{T}{n_t \bar{\delta}} > 0 \quad (305)$$

for $\rho \geq \rho^{(1)}$ and

$$L \geq L^{(k)} \triangleq 16 \Upsilon(T) \delta^{(k-1)} \left(\frac{Q^{-1}(\epsilon)}{\log \frac{T}{n_t \bar{\delta}}} \right)^2, \quad k = 2, 3, \dots \quad (306)$$

Since $\delta^{(1)} = K^{(1)}(T, \rho^{(1)}) / \sqrt{L}$, we have that

$$\frac{L^{(2)}}{L^{(1)}} = \frac{\Upsilon(T) K^{(1)}(T, \rho^{(1)})}{\sqrt{L} \bar{U}(T)} \quad (307)$$

and

$$\frac{L^{(k)}}{L^{(k-1)}} = \frac{\delta^{(k-1)}}{\delta^{(k-2)}}, \quad k = 3, 4, \dots \quad (308)$$

The former ratio is less than or equal to 1 if

$$L \geq L'' \triangleq \left(\frac{\Upsilon(T) K^{(1)}(T, \rho^{(1)})}{\bar{U}(T)} \right)^2. \quad (309)$$

The latter ratio is less than or equal to 1 since $\delta^{(k)} \leq \delta^{(k-1)}$, $k = 2, 3, \dots$ for $L \geq L^\circ$, as we shall show in (314)–(317). Comparing (297) with (309), we further note that $L'' \leq L'$ since $\tilde{V}(T) T^2 / 2 \leq \bar{U}(T)$ for $\rho \geq \rho'$. Consequently,

$$\underline{f}_{L,\rho}\left(\frac{n_t \bar{\delta} \rho}{T}, \bar{\mathbf{D}}\right) \geq \frac{1}{4} \log \frac{T}{n_t \bar{\delta}} > 0, \quad k = 2, 3, \dots \quad (310)$$

for $\rho \geq \max\{\rho^{(1)}, \rho'\}$ and $L \geq \max\{L^{(1)}, L'\}$.

We conclude that, when $T > n_t + n_r$, $\rho \geq \max\{\rho^{(1)}, \rho'\}$, and $L \geq \max\{L^{(1)}, L'\}$, $\alpha \mapsto \underline{f}_{L,\rho}^{(k)}(\alpha, \bar{\mathbf{D}})$ is a strictly decreasing function that is positive at $\alpha = n_t \bar{\delta} \rho / T$ and negative at $\alpha = \rho$. We can thus find an $\alpha_0^{(k)}(\bar{\mathbf{D}}, L, \rho)$ such that $\underline{f}_{L,\rho}^{(k)}(\alpha, \bar{\mathbf{D}}) > 0$ for $\alpha_0^{(k)}(\bar{\mathbf{D}}, L, \rho)$ and $\underline{f}_{L,\rho}^{(k)}(\alpha, \bar{\mathbf{D}}) \leq 0$ for $\alpha \geq \alpha_0^{(k)}(\bar{\mathbf{D}}, L, \rho)$. Defining

$$\delta_0^{(k)}(\bar{\mathbf{D}}, L, \rho) \triangleq 1 - \frac{\alpha_0^{(k)}(\bar{\mathbf{D}}, L, \rho)}{\rho} \quad (311)$$

this implies that we can assume without loss of optimality that $\alpha \geq \rho(1 - \delta_0^{(k)}(\bar{\mathbf{D}}, L, \rho))$. Following the steps in (290)–(291), and using (302), it can be shown that, for $\rho \geq \max\{\rho^{(1)}, \rho'\}$ and some $\tilde{\alpha}$ between $n_t \bar{\delta} \rho / T$ and ρ ,

$$\begin{aligned} \delta_0^{(k)}(\bar{\mathbf{D}}, L, \rho) &= \frac{\underline{f}_{L,\rho}^{(k)}(\rho, \bar{\mathbf{D}})}{\rho \dot{\underline{f}}_{L,\rho}^{(1)}(\tilde{\alpha}, \bar{\mathbf{D}})} \\ &= \frac{\sqrt{\frac{\Upsilon(T) \delta^{(k-1)}}{L}} Q^{-1}(\epsilon)}{\rho (T - n_t - n_r) \sum_{i=1}^{n_t} \frac{\bar{d}_i^2}{1 + \tilde{\alpha} \bar{d}_i^2}} \\ &\leq \sqrt{\frac{K(T) \delta^{(k-1)}}{L}} \end{aligned} \quad (312)$$

for

$$K(T) = \Upsilon(T) \left(1 + \frac{1}{\bar{\delta} \rho^{(1)}} \right)^2 \left(\frac{Q^{-1}(\epsilon)}{(T - n_t - n_r) n_t} \right)^2. \quad (313)$$

Thus, we can assume without loss of optimality that (294) holds with $\rho_0 = \max\{\rho^{(1)}, \rho'\}$ and $L_0 = \max\{L^{(1)}, L'\}$.

It remains to show that $\delta^{(k)} \leq \delta^{(k-1)}$, $k = 2, 3, \dots$ Indeed, by (294),

$$\delta^{(k)} = \sqrt{\frac{K(T) \delta^{(k-1)}}{L}}, \quad k = 2, 3, \dots \quad (314)$$

Since $\delta^{(1)} = K^{(1)}(T, \rho^{(1)}) / \sqrt{L}$, we obtain for $k = 2$ that

$$\frac{\delta^{(2)}}{\delta^{(1)}} = \sqrt{\frac{K(T)}{K^{(1)}(T, \rho^{(1)}) L^{1/2}}} \quad (315)$$

which is less than or equal to 1 if

$$L \geq L^\circ \triangleq \left(\frac{K(T)}{K^{(1)}(T, \rho^{(1)})} \right)^2. \quad (316)$$

Furthermore, if $\delta^{(k-1)} \leq \delta^{(k-2)}$, then we also have $\delta^{(k)} \leq \delta^{(k-1)}$, since the square-root function is a monotonically increasing:

$$\delta^{(k)} = \sqrt{\frac{K(T) \delta^{(k-1)}}{L}} \leq \sqrt{\frac{K(T) \delta^{(k-2)}}{L}} = \delta^{(k-1)}. \quad (317)$$

We conclude that $\delta^{(k)} \leq \delta^{(k-1)}$, $k = 2, 3, \dots$ for $L \geq L^\circ$.

C. Lower Bound on $\text{tr}(\mathbf{D}^2)$: Limit as $k \rightarrow \infty$

In the previous section, we have demonstrated that, for $\rho \geq \rho_0$ and $L \geq L_0$, we can assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \sqrt{\frac{K(T)\delta^{(k-1)}}{L}}\right), \quad k = 1, 2, \dots \quad (318)$$

where

$$\delta^{(1)} = \frac{K^{(1)}(T, \rho^{(1)})}{\sqrt{L}} \quad (319a)$$

$$\delta^{(k)} = \sqrt{\frac{K(T)\delta^{(k-1)}}{L}}, \quad k = 2, 3, \dots \quad (319b)$$

Since ρ_0 and L_0 do not depend on k , this implies that we can assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \sqrt{\frac{K(T)\delta}{L}}\right), \quad \rho \geq \rho_0, L \geq L_0 \quad (320)$$

where $\delta \triangleq \lim_{k \rightarrow \infty} \delta^{(k)}$ is the limit of the sequence $\{\delta^{(k)}\}$, which exists since the sequence is nonnegative and decreasing. We next determine δ . By the continuity of the square-root function, we obtain from (319b) that δ must satisfy

$$\delta = \sqrt{\frac{K(T)\delta}{L}}. \quad (321)$$

This equation has the two solutions

$$\delta = 0 \quad \text{and} \quad \delta = \frac{K(T)}{L}. \quad (322)$$

It follows that $\delta \leq K(T)/L$, which together with (320) demonstrates that, for $\rho \geq \rho_0$ and $L \geq L_0$, we can assume without loss of optimality that

$$\text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \frac{K(T)}{L}\right). \quad (323)$$

Applied to (270), this in turn demonstrates that

$$\begin{aligned} & \sup_{\mathbf{D} \in \mathcal{D}_1} \left\{ \bar{J}(\mathbf{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathbf{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ & \leq \sup_{\substack{\mathbf{D} \in \mathcal{D}_{n_t}: \\ \text{tr}(\mathbf{D}^2) \geq T\rho \left(1 - \frac{K(T)}{L}\right)}} \left\{ \bar{J}^*(\mathbf{D}, T, \rho) - \sqrt{\frac{\bar{U}(\mathbf{D}, T, \rho)}{L}} Q^{-1}(\epsilon) \right\} \\ & \quad + K_{\mathcal{D}_1}(T, \rho), \quad \rho \geq \rho_0, L \geq L_0 \end{aligned} \quad (324)$$

which is Lemma 11.

APPENDIX VIII PROOF OF LEMMA 12

To maximize $\bar{J}^*(\mathbf{D}, T, \rho)$ over $\mathbf{D} \in \mathcal{D}_1$, we note that the term in $\bar{J}^*(\mathbf{D}, T, \rho)$ that depends on \mathbf{D} is $(T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2)$. By the concavity of the logarithm function and Jensen's inequality, we have for all diagonal matrices \mathbf{D} satisfying $\text{tr}(\mathbf{D}^2) \leq T\rho$

$$\log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \leq n_t \log \left(1 + \frac{T\rho}{n_t}\right) \quad (325)$$

which holds with equality if $d_i^2 = T\rho/n_t$, $i = 1, \dots, n_t$. Since $\bar{\delta}$, as defined in (54), does not exceed T/n_t , this choice of d_i^2

satisfies $d_i^2 \geq \bar{\delta}\rho$ and, hence, the corresponding \mathbf{D} lies in \mathcal{D}_1 . Thus,

$$\sup_{\mathbf{D} \in \mathcal{D}_1} \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) = n_t \log \left(1 + \frac{T\rho}{n_t}\right) \quad (326)$$

which together with the definition of $\bar{J}^*(\mathbf{D}, T, \rho)$ in (82) proves Lemma 12.

APPENDIX IX PROOF OF LEMMA 13

By replacing in (237) \mathbf{D} by $\sqrt{\rho/\alpha}\mathbf{D}$, we obtain that

$$\begin{aligned} & \bar{U}\left(\sqrt{\rho/\alpha}\mathbf{D}, T, \rho\right) \\ & = \bar{U}^*\left(\sqrt{\rho/\alpha}\mathbf{D}, T, \rho\right) + K_{\bar{U}}\left(\sqrt{\rho/\alpha}\mathbf{D}, T, \rho\right) \end{aligned} \quad (327)$$

where $K_{\bar{U}}(\mathbf{D}, T, \rho)$ is a function of \mathbf{D} , T , and ρ that satisfies $\lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathbf{D}, T, \rho)| = 0$.

We next evaluate $\bar{U}^*(\tilde{\mathbf{D}}, T, \rho)$ for $\tilde{\mathbf{D}} = \sqrt{\rho/\alpha}\mathbf{D}$. Note that the trace of $\tilde{\mathbf{D}}$ is equal to $T\rho$, so the corresponding $\tilde{\alpha} = \text{tr}(\tilde{\mathbf{D}}^2)/T$ is equal to ρ . It follows that

$$\begin{aligned} & \bar{U}^*\left(\sqrt{\frac{\rho}{\alpha}}\mathbf{D}, T, \rho\right) \\ & = \mathbb{E} \left[\left(\sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\alpha} - 1 \right) (Z'_{i,\ell} - n_r) - \sum_{i=1}^{n_t} (Z''_{i,\ell} - (T - n_t)) \right. \right. \\ & \quad \left. \left. + (T - n_t) \left(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H) - \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)] \right) \right)^2 \right] \\ & = \text{Var} \left(\sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\alpha} - 1 \right) Z'_{i,\ell} \right) + \text{Var} \left(\sum_{i=1}^{n_t} Z''_{i,\ell} \right) \\ & \quad + (T - n_t)^2 \text{Var}(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)) \\ & \quad + 2(T - n_t) \times \\ & \quad \times \sum_{i=1}^{n_t} \left(\frac{n_t d_{i,\ell}^2}{T\alpha} - 1 \right) \mathbb{E}[(Z'_{i,\ell} - n_r) \log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)] \\ & \geq \text{Var} \left(\sum_{i=1}^{n_t} Z''_{i,\ell} \right) + (T - n_t)^2 \text{Var}(\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)) \\ & = n_t(T - n_t) + (T - n_t)^2 \sum_{i=0}^{n_t-1} \Psi'(n_r - i) \end{aligned} \quad (328)$$

where the second step follows because $\{Z''_{1,1}, \dots, Z''_{n_t,L}\}$ and $\{\mathbf{H}^L, Z'_{1,1}, \dots, Z'_{n_t,L}\}$ are independent, so the term depending on $\{Z''_{1,1}, \dots, Z''_{n_t,L}\}$ is uncorrelated with the other terms; the third step follows because the first term is nonnegative and the last term is zero, since $\mathbb{E}[(Z'_{i,\ell} - n_r) \log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)]$ does not depend on i and the sum over i is zero because $d_{1,\ell}^2 + \dots + d_{n_t,\ell}^2 = T\alpha$; the fourth step follows because the variance of $\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^H)$ is $\sum_{i=0}^{n_t-1} \Psi'(n_r - i)$ (Lemma 7).

The RHS of (328) is equal to $T^2 \tilde{V}(T)$ (with $\tilde{V}(T)$ given in (10)), so

$$\bar{U}\left(\sqrt{\frac{\rho}{\alpha}}\mathbf{D}, T, \rho\right) \geq T^2 \tilde{V}(T) + K_{\bar{U}}(\mathbf{D}, T, \rho) \quad (329)$$

where, with a slight abuse of notation, we replace $K_{\bar{U}}(\sqrt{\rho/\alpha}\mathbf{D}, T, \rho)$ by $K_{\bar{U}}(\mathbf{D}, T, \rho)$. Since the set of matrices

$\tilde{\mathcal{D}}_1 \triangleq \{\tilde{\mathbf{D}} \in \mathcal{D}_{n_t} : \tilde{\mathbf{D}} = \sqrt{\rho/\alpha} \mathbf{D}, \mathbf{D} \in \mathcal{D}_1\}$ is contained in \mathcal{D}_1 , we have that

$$\sup_{\mathbf{D} \in \mathcal{D}_1} \left| K_{\bar{U}} \left(\sqrt{\frac{\rho}{\alpha}} \mathbf{D}, T, \rho \right) \right| \leq \sup_{\mathbf{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathbf{D}, T, \rho)| \quad (330)$$

so the constant $K_{\bar{U}}(\mathbf{D}, T, \rho)$ in (329) satisfies $\lim_{\rho \rightarrow \infty} \sup_{\mathbf{D} \in \mathcal{D}_1} |K_{\bar{U}}(\mathbf{D}, T, \rho)| = 0$. This proves the lemma.

APPENDIX X PROOF OF LEMMA 14

We begin with (264) and optimize over $\mathbf{D} \in \mathcal{D}_2$. A lower bound on $\bar{J}(\mathbf{D}, T, \rho)$ follows by noting that $K_{\mathcal{D}_1}(\mathbf{D}, T) \geq 0$, so for every $\mathbf{D} \in \mathcal{D}_{n_t}$

$$\bar{J}(\mathbf{D}, T, \rho) \geq \bar{J}^*(\mathbf{D}, T, \rho). \quad (331)$$

To obtain an upper bound, and to optimize $\bar{J}^*(\mathbf{D}, T, \rho)$ over $\mathbf{D} \in \mathcal{D}_2$, we have to distinguish between the different cases of d_i^2 , $i = 1, \dots, n_t$ that can occur for $\mathbf{D} \in \mathcal{D}_2$. Indeed, without loss of optimality, assume that the diagonal elements of \mathbf{D} are ordered so that $d_1 \leq d_2 \leq \dots \leq d_{n_t}$. Then, for every $\mathbf{D} \in \mathcal{D}_2$, we have that

$$d_i^2 \leq \bar{\delta}\rho, \quad i = 1, \dots, n_k \quad (332a)$$

$$d_i^2 > \bar{\delta}\rho, \quad i = n_k + 1, \dots, n_t \quad (332b)$$

for some $n_k = 1, \dots, n_t$. We need to distinguish between the following two cases:

- 1) d_i tends to infinity as $\rho \rightarrow \infty$ for all $i = 1, \dots, n_t$;
- 2) there exists a finite $\bar{\rho}$ such that $d_i \leq \bar{\rho}$ for $i = 1, \dots, n_s$ and some $n_s = 1, \dots, n_k$.

We analyze the former case in Appendix X-A and the latter case in Appendix X-B. We then combine both cases in Appendix X-C.

A. Unbounded d_i

When $\lim_{\rho \rightarrow \infty} d_i = \infty$ for every $i = 1, \dots, n_t$, we can find a $\xi(\rho)$ satisfying $\lim_{\rho \rightarrow \infty} \xi(\rho) = \infty$ such that $d_i^2 \geq \xi(\rho)$, $i = 1, \dots, n_t$. It follows from (264) and (265) that

$$\begin{aligned} \bar{J}(\mathbf{D}, T, \rho) &= \bar{J}^*(\mathbf{D}, T, \rho) + \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) + (T - n_t) \times \\ &\quad \times \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right] \\ &\leq \bar{J}^*(\mathbf{D}, T, \rho) + \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) + (T - n_t) \times \\ &\quad \times \mathbb{E} \left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(1 + \xi(\rho))^{-1} \mathbf{I}_{n_t})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})} \right] \\ &\triangleq \bar{J}^*(\mathbf{D}, T, \rho) + K_{\mathcal{D}_2}(T, \rho) \end{aligned} \quad (333)$$

where the last step should be viewed as the definition of $K_{\mathcal{D}_2}(T, \rho)$. Consequently,

$$\bar{J}(\mathbf{D}, T, \rho) \leq \bar{J}^*(\mathbf{D}, T, \rho) + K_{\mathcal{D}_2}(T, \rho) \quad (334)$$

and, by the dominated convergence theorem, $\lim_{\rho \rightarrow \infty} K_{\mathcal{D}_2}(T, \rho) = 0$.

We next optimize $\bar{J}^*(\mathbf{D}, T, \rho)$ over $\mathbf{D} \in \mathcal{D}_2$. To this end, we recall that the only term in $\bar{J}^*(\mathbf{D}, T, \rho)$ that depends on \mathbf{D} is $(T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2)$. Defining $\sum_{i=1}^{n_k} d_i^2 \triangleq \bar{d}_{n_k}$, and using the concavity of the logarithm function and Jensen's inequality, we have that, for all diagonal matrices \mathbf{D} in \mathcal{D}_2 ,

$$\begin{aligned} &\log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \\ &= \sum_{i=1}^{n_k} \log(1 + d_i^2) + \sum_{i=n_k+1}^{n_t} \log(1 + d_i^2) \\ &\leq n_k \log \left(1 + \frac{\bar{d}_{n_k}}{n_k} \right) + (n_t - n_k) \log \left(1 + \frac{T\rho - \bar{d}_{n_k}}{n_t - n_k} \right) \\ &\leq \log(1 + \bar{\delta}\rho) + (n_t - 1) \log \left(1 + \frac{T\rho - \bar{\delta}\rho}{n_t - 1} \right) \end{aligned} \quad (335)$$

where the second step follows from Jensen's inequality and because $\sum_{i=1}^{n_t} d_i^2 \leq T\rho$, so $\sum_{i=n_k+1}^{n_t} d_i^2 \leq T\rho - \bar{d}_{n_k}$; the third step follows because the expression is monotonically increasing in $\bar{d}_{n_k} \leq n_k \bar{\delta}\rho$ and monotonically decreasing in $n_k = 1, \dots, n_t$.

The inequality (335) holds with equality if $d_1^2 = \bar{\delta}\rho$ and $d_i^2 = (T\rho - \bar{\delta}\rho)/(n_t - 1)$, $i = 2, \dots, n_t$. It follows that

$$\begin{aligned} &\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}^*(\mathbf{D}, T, \rho) \\ &= n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} - n_t(T - n_t) \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})] \\ &\quad + (T - n_t - n_r) \log \left((1 + \bar{\delta}\rho) \left(1 + \frac{T - \bar{\delta}\rho}{n_t - 1} \right)^{n_t - 1} \right) \\ &\triangleq \bar{J}_{\mathcal{D}_2}(T, \rho). \end{aligned} \quad (336)$$

We conclude from (334) and (336) that, when d_i tends to infinity as $\rho \rightarrow \infty$ for every $i = 1, \dots, n_t$, we have that

$$\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}(\mathbf{D}, T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho). \quad (337)$$

B. Bounded d_i

For every $\mathbf{D} \in \mathcal{D}_{n_t}$, the term $\bar{J}(\mathbf{D}, T, \rho)$ can be upper-bounded as

$$\begin{aligned} &\bar{J}(\mathbf{D}, T, \rho) \\ &= n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} + \left(\frac{T\alpha - T\rho}{T\rho} \right) n_r n_t \\ &\quad - n_t(T - n_t) + \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) \\ &\quad + (T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell)(\mathbf{I}_{n_t} + \mathbf{D}^2)^{-1})] \\ &\leq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} - n_t(T - n_t) \\ &\quad + \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) \\ &\quad + (T - n_t - n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \\ &\quad + (T - n_t) \mathbb{E} [\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})]. \end{aligned} \quad (338)$$

To optimize the RHS of (338) over \mathbf{D} , we note that the only term that depends on \mathbf{D} is again $(T - n_t -$

$n_r) \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2)$. Following similar steps as in (335), it can be shown that, for all matrices \mathbf{D} satisfying

$$d_i^2 \leq \bar{\rho}, \quad i = 1, \dots, n_s \quad (339a)$$

$$d_i^2 \leq \bar{\delta}\rho, \quad i = n_s + 1, \dots, n_k \quad (339b)$$

$$d_i^2 > \bar{\delta}\rho, \quad i = n_k + 1, \dots, n_t \quad (339c)$$

for some $n_s = 1, \dots, n_k$, $n_k = 1, \dots, n_t$, and $\rho \geq \bar{\rho}/\bar{\delta}$, we have that

$$\begin{aligned} & \log \det(\mathbf{I}_{n_t} + \mathbf{D}^2) \\ & \leq n_s \log(1 + \bar{\rho}) + (n_k - n_s) \log(1 + \bar{\delta}\rho) \\ & \quad + (n_t - n_k) \log\left(1 + \frac{T\rho - n_s\bar{\rho} - (n_k - n_s)\bar{\delta}\rho}{n_t - n_k}\right) \\ & \leq \log(1 + \bar{\rho}) + (n_t - 1) \log\left(1 + \frac{T\rho - \bar{\rho}}{n_t - 1}\right) \end{aligned} \quad (340)$$

where the second inequality follows because, for $\rho \geq \bar{\rho}/\bar{\delta}$, the expression is monotonically decreasing in n_s and n_k . Thus, for all matrices \mathbf{D} satisfying (339) and $\rho \geq \bar{\rho}/\bar{\delta}$,

$$\begin{aligned} & \bar{J}(\mathbf{D}, T, \rho) \\ & \leq n_t n_r \log \frac{T\rho}{n_t} + \log \frac{\Gamma_{n_t}(n_t)}{\Gamma_{n_t}(T)} - n_t(T - n_t) \\ & \quad + \frac{n_t}{T\rho} (n_t n_r + n_t(T - n_t)) \\ & \quad + (T - n_t - n_r) \log\left((1 + \bar{\rho})\left(1 + \frac{T\rho - \bar{\rho}}{n_t - 1}\right)^{n_t - 1}\right) \\ & \quad + (T - n_t) \mathbb{E}[\log \det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})] \\ & \triangleq \bar{J}'_{\mathcal{D}_2}(T, \rho). \end{aligned} \quad (341)$$

C. Combining both cases

We obtain from (331) and (336) that

$$\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}(\mathbf{D}, T, \rho) \geq \bar{J}_{\mathcal{D}_2}(T, \rho). \quad (342)$$

Similarly, we obtain from (337) and (341) that, for $\rho \geq \bar{\rho}/\bar{\delta}$,

$$\begin{aligned} & \sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}(\mathbf{D}, T, \rho) \\ & \leq \max\{\bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho), \bar{J}'_{\mathcal{D}_2}(T, \rho)\}. \end{aligned} \quad (343)$$

We next show that, for sufficiently large ρ , $\bar{J}'_{\mathcal{D}_2}(T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho)$. Hence, for such ρ ,

$$\sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}(\mathbf{D}, T, \rho) \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho). \quad (344)$$

Indeed,

$$\begin{aligned} & \bar{J}_{\mathcal{D}_2}(T, \rho) - \bar{J}'_{\mathcal{D}_2}(T, \rho) \\ & = \log\left((1 + \bar{\delta}\rho)\left(1 + \frac{T - \bar{\delta}}{n_t - 1}\rho\right)^{n_t - 1}\right) \\ & \quad - \log\left((1 + \bar{\rho})\left(1 + \frac{T\rho - \bar{\rho}}{n_t - 1}\right)^{n_t - 1}\right) \\ & \quad + (T - n_t) \mathbb{E}\left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}\right] \end{aligned}$$

$$\begin{aligned} & = \log\left(\frac{1 + \bar{\delta}\rho}{1 + \bar{\rho}}\right) + (n_t - 1) \log\left(\frac{n_t - 1 + T\rho - \bar{\delta}\rho}{n_t - 1 + T\rho - \bar{\rho}}\right) \\ & \quad + (T - n_t) \mathbb{E}\left[\log \frac{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H})}{\det(\mathbf{H}_\ell \mathbf{H}_\ell^\mathsf{H} + \lambda_1(\mathbf{Q}_\ell^\mathsf{H} \mathbf{Q}_\ell) \mathbf{I}_{n_t})}\right]. \end{aligned} \quad (345)$$

The first term on the RHS of (345) tends to infinity as $\rho \rightarrow \infty$. The remaining terms are bounded in $\rho \geq \bar{\rho}/\bar{\delta}$. Thus, for $\rho \geq \rho_0$ and a sufficiently large ρ_0 , the RHS of (345) is strictly positive. Since $K_{\mathcal{D}_2}(T, \rho)$ vanishes as $\rho \rightarrow \infty$, (344) follows.

Combining (344) with (342), we conclude that

$$\begin{aligned} \bar{J}_{\mathcal{D}_2}(T, \rho) & \leq \sup_{\mathbf{D} \in \mathcal{D}_2} \bar{J}(\mathbf{D}, T, \rho) \\ & \leq \bar{J}_{\mathcal{D}_2}(T, \rho) + K_{\mathcal{D}_2}(T, \rho) \end{aligned} \quad (346)$$

for a sufficiently large ρ_0 , and $\rho \geq \rho_0$, which is Lemma 14.

ACKNOWLEDGMENT

Fruitful discussions with W. Yang are gratefully acknowledged.

REFERENCES

- [1] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.
- [2] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [4] M. Shirvanimoghaddam, M. S. Mohammadi, R. Abbas, A. Minja, C. Yue, B. Matuz, G. Han, Z. Lin, W. Liu, Y. Li, S. Johnson, and B. Vucetic, "Short block-length codes for ultra-reliable low latency communications," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 130–137, Feb. 2019.
- [5] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," *Physical Communication*, vol. 34, pp. 66–79, Jun. 2019.
- [6] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823–2835, Aug. 2015.
- [7] ———, "Wireless energy and information transmission using feedback: Infinite and finite block-length analysis," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5304–5318, Dec. 2016.
- [8] L. Zhang and Y.-C. Liang, "Average throughput analysis and optimization in cooperative IoT networks with short packet communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11549–11562, Dec. 2018.
- [9] C. Li, N. Yang, and S. Yan, "Optimal transmission of short-packet communications in multiple-input single-output systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7199–7203, Jul. 2019.
- [10] N. H. Mahmood, O. A. López, H. Alves, and M. Latvala-Aho, "A predictive interference management algorithm for URLLC in beyond 5G networks," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 995–999, Mar. 2021.
- [11] A. Munari and F. Clazzer, "Spectral coexistence of QoS-constrained and IoT traffic in satellite systems," *Sensors*, vol. 21, no. 14, Jul. 2021.
- [12] J. Feng, H. Q. Ngo, and M. Matthaiou, "(Non)-coherent MU-MIMO block fading channels with finite blocklength and linear processing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2441–2461, Apr. 2023.
- [13] M. Xie, J. Gong, X. Jia, Q. Wang, and X. Ma, "Age and energy analysis for l th best relay enabled cooperative status update systems with short packet communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6294–6308, May 2023.
- [14] C. Zheng, F.-C. Zheng, J. Luo, P. Zhu, X. You, and D. Feng, "Differential modulation for short packet transmission in URLLC," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7230–7245, Jul. 2024.

- [15] L. Li, W. Chen, P. Popovski, and K. B. Letaief, "Reliability-latency-rate tradeoff in low-latency communications with finite-blocklength coding," *IEEE Trans. Inf. Theory*, vol. 71, no. 1, pp. 360–389, Jan. 2025.
- [16] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, St. Petersburg, Russia, Jul. 31–Aug. 5, 2011, pp. 2959–2963.
- [17] A. Collins and Y. Polyanskiy, "Orthogonal designs optimize achievable dispersion for coherent MISO channels," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Honolulu, HI, USA, Jun. 29–Jul. 4, 2014, pp. 2524–2528.
- [18] ———, "Dispersion of the coherent MIMO block-fading channel," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 10–15, 2016, pp. 1068–1072.
- [19] ———, "Coherent multiple-antenna block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 380–405, Jan. 2019.
- [20] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO quasi-static Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6591–6622, Dec. 2015.
- [21] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, "Optimum power control at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4598–4615, Nov. 2015.
- [22] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, 2014.
- [23] ———, "Diversity versus channel knowledge at finite block-length," in *Proc. IEEE Inf. Theory Workshops (ITW)*, Lausanne, Switzerland, Sep. 3–7, 2012, pp. 572–576.
- [24] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.
- [25] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vazquez-Vilar, "Saddlepoint approximations for noncoherent single-antenna Rayleigh block-fading channels," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Paris, France, Jul. 7–12 2019, pp. 612–616.
- [26] ———, "Saddlepoint approximations for short-packet wireless communications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4831–4846, Jul. 2021.
- [27] A. Lancho, T. Koch, and G. Durisi, "On single-antenna Rayleigh block-fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 496–519, Jan. 2019.
- [28] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
- [29] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.
- [30] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecom.*, vol. 10, no. 6, pp. 585–595, 1999.
- [31] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.
- [32] J. Östman, G. Durisi, E. G. Ström, M. C. Coşkun, and G. Liva, "Short packets over block-memoryless fading channels: Pilot-assisted or noncoherent transmission?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1521–1536, Feb. 2019.
- [33] Spectre, *SPECTRE: short packet communication toolbox*, 2015, GitHub repository, [Online]. Available: <https://github.com/yp-mit/spectre>.
- [34] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [35] ———, "The diversity-multiplexing tradeoff for non-coherent multiple antenna channels," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, USA, 2002, pp. 1011–1020.
- [36] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: John Wiley & Sons, 1971, vol. II.
- [37] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.
- [38] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Foundations and Trends in Communications and Information Theory, 2004, vol. 1, no. 1.
- [39] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [40] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [41] R. B. Ash and C. A. Doléans-Dade, *Probability and Measure Theory*, 2nd ed. Elsevier/Academic Press, 2000.
- [42] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, 2nd ed. John Wiley & Sons, 2005.
- [43] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill Book Co., 1976.

Chao Qi received the B.S. and Ph.D. degrees in information security from Southwest Jiaotong University, China, in 2009 and 2017, respectively. From April 2018 until February 2023, she was a Postdoc with the Universidad Carlos III de Madrid, Spain. She is currently an Associate Professor with School of Computer and Software Engineering, Xihua University, China. Her research interests are in physical security and information theory.

Tobias Koch (Senior Member IEEE) received the M.Sc. (Hons.) and Ph.D. degrees in electrical engineering from ETH Zurich, Switzerland, in 2004 and 2009, respectively. From June 2010 until May 2012, he was a Marie Curie Intra-European Research Fellow with the University of Cambridge, UK. He was also a Research Intern at Bell Labs, Murray Hill, NJ, USA, in 2004, and the Universitat Pompeu Fabra, Barcelona, Spain, in 2007. He joined the Universidad Carlos III de Madrid, Spain, in 2012, where he is currently an Associate Professor. His research interests are in digital communication theory and information theory.

Dr. Koch received a Starting Grant from the European Research Council (ERC), a Ramón y Cajal Research Fellowship, a Marie Curie Intra-European Fellowship, a Marie Curie Career Integration Grant, and a Fellowship for Prospective Researchers from the Swiss National Science Foundation. He further received a medal of the 2018 Young Researchers Award "Agustín de Betancourt y Molina" by the Royal Academy of Engineering of Spain. He served as the Vice Chair for the Spain Chapter of the IEEE Information Theory Society from 2013 to 2016, and he served as the Chapter's Chair from 2020 to 2023. He is an Associate Editor for Communications for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2020 to 2026. He is also a member of the Board of Governors of the IEEE Information Theory Society from 2025 to 2027.