On the Minimax Capacity Loss Under Sub-Nyquist Universal Sampling

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

***Abstract*— This paper investigates the information rate loss in analog channels, when the sampler is designed to operate independent of the instantaneous channel occupancy. Specifically, a multiband linear time-invariant Gaussian channel under universal sub-Nyquist sampling is considered. The entire channel bandwidth is divided into n subbands of equal bandwidth. At each time, only k constant-gain subbands are active, where the instantaneous subband occupancy is not known at the receiver and the sampler. We study the information loss through an information rate loss metric, that is, the gap of achievable rates caused by the lack of instantaneous subband occupancy information. We characterize the minimax information rate loss for the sub-Nyquist regime, provided that the number n of subbands and the SNR are both large. The minimax limits depend almost solely on the band sparsity factor and the undersampling factor, modulo some residual terms that vanish as n and SNR grow. Our results highlight the power of randomized sampling methods (i.e., the samplers that consist of random periodic modulation and low-pass filters), which are able to approach the minimax information rate loss with exponentially high probability.**

***Index Terms*— Channel capacity, minimax sampling, non-asymptotic random matrix, log-determinant, concentration of spectral measure.**

I. Introduction

THE maximum rate of information that can be conveyed through a continuous-time communication channel is dependent on the sampling technique employed at the receiver end. In some cutting-edge communication systems, hardware and cost limitations often preclude sampling at or above the Nyquist rate, which presents a major bottleneck in transferring wideband and energy-efficient receiver design paradigms from theory to practice. Understanding the effects upon capacity of sub-Nyquist sampling is thus crucial in circumventing this bottleneck.

Manuscript received April 28, 2013; revised May 7, 2015; accepted December 17, 2016. Date of publication April 18, 2017; date of current version May 18, 2017. This work was supported in part by the NSF under Grant CCF-0939370 and Grant CIS-1320628, in part by the AFOSR through MURI under Grant FA9550-12-1-0215, and in part by BSF Transformative Science under Grant 2010505. This paper was presented at the 2013 IEEE International Symposium on Information Theory. *(Corresponding author: Yuxin Chen.)*

Y. Chen is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: yuxin.chen@princeton.edu).

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

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

Communicated by Y.-H. Kim, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2017.2695541

In many practical scenarios, the occupancy of the communication channel varies over time. An ideal adaptive sampler can be dynamically optimized relative to these channel variations. Nevertheless, in most practical systems, the samplers and the analog-to-digital converters are static and are designed independent of which subbands are active at any given time. This has no effect if the sampling rate employed is commensurate with the maximum bandwidth (or the Nyquist rate) of the channel. However, in the sub-Nyquist regime, the sampler design significantly impacts the information rate achievable over a given channel. As was shown in [1], the capacity-maximizing sub-Nyquist sampling mechanism depends on knowledge of the channel spectrum. When the subbands available for communications are time-varying and a universal (static) sub-Nyquist sampler is used, *capacity loss* is typically incurred and our work characterizes this loss.

In the present paper, we consider a linear time-invariant (LTI) Gaussian channel with known channel gain, whereby the entire channel bandwidth is divided into n subbands of equal bandwidth. At each timeframe, only a subset of k subbands are active for transmission, but the spectral occupancy information is not available at either the receiver or the sampler. The goal is to explore universal (channel-independent) design of a sub-Nyquist sampling system that is robust vis-a-vis the uncertainty of instantaneous channel occupancy. In particular, we aim to understand the resulting loss of information rates between sampling with and without subband occupancy information in some minimax sense (as will be detailed in Section II-C), and design a sub-Nyquist sampling system under which the information rate loss can be uniformly controlled and optimized over all possible channel support.

A. Related Work

In various scenarios, sampling above the Nyquist rate is not necessary for preserving signal information in the sense that it generates a discrete-time sufficient statistic, provided that certain signal structures are appropriately exploited [2], [3]. Take multiband signals for example, that reside within several subbands over a wide spectrum. If the spectral support is known, then the sampling rate necessary for perfect signal reconstruction is the spectral occupancy, termed the *Landau rate* [4]. Such signals admit perfect recovery when sampled at rates approaching the Landau rate, assuming appropriately chosen sampling sets (e.g. [5], [6]). Inspired by recent "compressed sensing" [7]–[9] ideas, spectrum-blind sub-Nyquist samplers have also been developed for multiband signals [10]. Two of the most widely used modules employed in the sampler designs are filter banks and periodic modulation [10]–[13]. These sampling-theoretic works, however, were not based on capacity as a metric in the sampler design.

On the other hand, the Shannon-Nyquist sampling theorem has frequently been invoked to investigate the capacity of analog waveform channels (e.g. [14], [15]). The effects upon capacity of oversampling have been investigated as well in the presence of quantization [16], [17]. However, none of these works considered the effect of undersampling upon capacity. Another recent line of work [18] investigated the tradeoff between sparse coding and subsampling in AWGN channels, but did not consider capacity-achieving input distributions.

Our recent work [1], [19] established a new framework for investigating the capacity of LTI Gaussian channels under a broad class of sub-Nyquist sampling strategies, including filter-bank and modulation-bank sampling and, more generally, time-preserving sampling. We demonstrated that sampling with a filter bank is sufficient to approach maximum capacity, assuming that perfect channel state information (CSI) is available at both the receiver and the transmitter. In many practical scenarios, however, the active frequency set available for communications might be changing over time, like in cognitive radio networks where the spectral subbands available to cognitive users are varying over time. To the best of our knowledge, no prior work has investigated, from a capacity perspective, a channel-blind sub-Nyquist sampling paradigm in the absence of subband occupancy information.

Finally, the effect of undersampling has been explored from a source coding perspective as well. For instance, the fundamental rate-distortion function of Gaussian sources has been determined under sub-Nyquist sampling with filtering [20], revealing that the alias suppressing sampler design achieves the optimal rate distortion function. For the case where the input source signals are sparse, the recent work [21] characterized the rate-distortion function under independent and memoryless random sampling. The main results and techniques presented herein might potentially extend to these source coding settings to quantify the rate loss caused by spectral-blind sampling design.

B. Main Contributions

We consider a frequency-flat multiband channel model and the class of sampling systems with filter banks and periodic modulation. For this model, our main contributions are summarized as follows.

- We derive a lower bound (Theorem 2) on the minimax sampled information rate loss (defined in Section II) incurred due to the lack of channel occupancy information, under super-Nyquist universal sampling. This minimax lower limit depends almost only on the band sparsity factor and the undersampling factor, modulo some residual terms that vanish when SNR and n increase.
- We characterize in Theorem 3 the sampled information rate loss under a class of sampling systems with periodic modulation and low-pass filters with passband $[0, W/n]$, when the Fourier coefficients of the modulation waveforms are generated in an i.i.d. Gaussian fashion (termed *Gaussian sampling*). We demonstrate that with exponentially high probability, the resulting sampled information rate loss matches the lower bound given in Theorem 2 uniformly over all possible subband occupancy. This implies that random sampling strategies are minimax-optimal in terms of a universal sampling design.
- The power of random sampling arises due to sharp concentration of spectral measures of large random matrices [23]. To establish Theorem 3, we derive measure concentration of several log-determinant functions for i.i.d. Gaussian ensembles, which might be of independent interest for other works involving log-determinant metrics.

C. Organization

The remainder of this paper is organized as follows. In Section II we introduce our system model of multiband Gaussian channels. A metric called sampled information rate loss, and a minimax sampler, are defined with respect to sampled channel capacity. We then determine in Section III the minimax information rate loss. Specifically, we develop lower bounds on the minimax information rate loss in Section III-A. The achievability is treated in Section III-B. Besides, we derive measure concentration of several log-determinant functions in Section IV-C. Section V-A summarizes the key observation and implications from our results. Section VI closes the paper with a short summary of our findings and potential future directions.

D. Notation

Denote by $\mathcal{H}(\beta) := -\beta \log \beta - (1-\beta)\log(1-\beta)$ the binary entropy function. The standard notation $f(n) = \mathcal{O}(g(n))$ means there exists a constant $c > 0$ such that $|f(n)| \leq c|g(n)|$, $f(n) = \Theta(g(n))$ means there exist constants $c_1, c_2 > 0$ such that $c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$, $f(n) = \omega(g(n))$ means that $\lim_{n\to\infty} \frac{g(n)}{f(n)} = 0$, and $f(n) = o(g(n))$ indicates that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. For a matrix $\boldsymbol{A}$, we use $\boldsymbol{A}_{i*}$ and $\boldsymbol{A}_{*i}$ to denote the ith row and ith column of $\boldsymbol{A}$, respectively. We let $[n]$ denote the set $\{1, 2, \cdots, n\}$, and write $\binom{[n]}{k}$ for the set of all k-element subsets of $\{1, 2, \cdots, n\}$. We also use $\mathrm{card}(\mathcal{A})$ to denote the cardinality of a set $\mathcal{A}$. Let $\boldsymbol{W}$ be a $p \times p$ random matrix that can be expressed as $\boldsymbol{W} = \Sigma_{i=1}^{n} \boldsymbol{Z}_i \boldsymbol{Z}_i^{\top}$, where $\boldsymbol{Z}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ are jointly independent Gaussian vectors. Then $\boldsymbol{W}$ is said to have a central Wishart distribution with n degrees of freedom and scale matrix $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{W} \sim \mathcal{W}_p(n, \boldsymbol{\Sigma})$. Our notation is summarized in Table I.

II. Problem Formulation and Preliminaries

A. Compound Multiband Channel

We consider a multiband Gaussian channel of total bandwidth W, and it is divided into n continuous subbands[1] each of

[1]Note that in practice, n is typically a large number. For instance, the number of subcarriers ranges from 128 to 2048 in LTE [24], [25].

TABLE I
SUMMARY OF NOTATION AND PARAMETERS

$\mathcal{H}(x)$	binary entropy function, i.e. $\mathcal{H}(x) = -x\log x - (1-x)\log(1-x)$
$h(t)$,$H(f)$	impulse response, and frequency response of the LTI analog channel
$\mathcal{S}_\eta(f)$	power spectral density of the additive Gaussian noise $\eta(t)$
f_s, T_s	aggregate sampling rate ($f_\mathrm{s} = \frac{m}{n}W$), and the corresponding sampling interval ($T_\mathrm{s} = 1/f_\mathrm{s}$)
W, W_0	channel bandwidth, size of instantaneous channel support
n, m, k	number of subbands, number of sampling branches, number of subbands being simultaneously active
$\alpha = m/n$	undersampling factor
$\beta = k/n$	sparsity factor
$\boldsymbol{Q}$, $\boldsymbol{Q}^\mathrm{w}$	sampling matrix, whitened sampling matrix ($\boldsymbol{Q}^\mathrm{w} = (\boldsymbol{Q}\boldsymbol{Q}^*)^{-\frac{1}{2}}\boldsymbol{Q}$)
$L_{\boldsymbol{s}}^{\boldsymbol{Q}}$	information rate loss associated with a sampling matrix $\boldsymbol{Q}$ given state $\boldsymbol{s}$
$\boldsymbol{A}_{i*}$, $\boldsymbol{A}_{*i}$	ith row of $\boldsymbol{A}$, ith column of $\boldsymbol{A}$
card $(\mathcal{A})$	cardinality of a set $\mathcal{A}$
$[n]$	$[n] := \{1, 2, \cdots, n\}$
$\binom{[n]}{k}$	set of all k-element subsets of $[n]$
$\mathcal{W}_p(n, \boldsymbol{\Sigma})$	p-dimensional central Wishart distribution with n degrees of freedom and scale matrix $\boldsymbol{\Sigma}$

bandwidth W/n. A state $\boldsymbol{s} \in \binom{[n]}{k}$ is generated, which dictates the channel support. For ease of presentation, the present work focuses on the frequency-flat channel model, which suffices to capture the essence of our findings. Specifically, given a state $\boldsymbol{s}$, the channel is assumed to be an LTI filter with impulse response $h_{\boldsymbol{s}}(t)$ and frequency response

$$H_{\boldsymbol{s}}(f) = \begin{cases} H, & \text{if } f \text{ lies within subbands at indices from } \boldsymbol{s}, \\ 0, & \text{else.} \end{cases} \tag{1}$$

A transmit signal $x(t)$ with a power constraint P is passed through this multiband channel, yielding a channel output

$$r_{\boldsymbol{s}}(t) = h_{\boldsymbol{s}}(t) * x(t) + \eta(t), \tag{2}$$

where $\eta(t)$ is stationary zero-mean Gaussian noise with power spectral density $\mathcal{S}_\eta(f) \equiv 1$. It is assumed throughout that the knowledge of H and $\mathcal{S}_\eta$ are available at both the transmitter and the receiver, while the state $\boldsymbol{s}$ is known only at the transmitter. The results derived herein can be extended to more general frequency selective channels with optimal power control. These extensions are described in more details in Section V-B and derived in [26].

B. Sampled Channel Capacity

We aim to design a sampler that operates below the Nyquist rate (i.e. the channel bandwidth W). In particular, the present work focuses on the class of filter-bank and modulation-bank sampling systems, which subsumes the most widely used sampling mechanisms in practice.

1) Sampling System and Channel Capacity: We consider the class of sampling systems that consist of a combination of filter banks and periodic modulation, as illustrated in Fig. 1(a). Specifically, the sampling system comprises m branches, where at the ith branch, the channel output is passed through a pre-modulation LTI filter $F_i(f)$, modulated by a periodic waveform $q_i(t)$ of period $T_q = (W/n)^{-1}$, and then passed through a post-modulation LTI filter $S_i(f)$ followed by uniform sampling at rate W/n. The aggregate sampling rate is $f_\mathrm{s} = \frac{m}{n}W$. When specialized to the frequency-flat channels, it is natural to concentrate on the case where $F_i(f)$ and $S_i(f)$ are both flat[2] within each subband $\left[\frac{lW}{n}, \frac{(l+1)W}{n}\right)$ $(l \in \mathbb{Z})$.

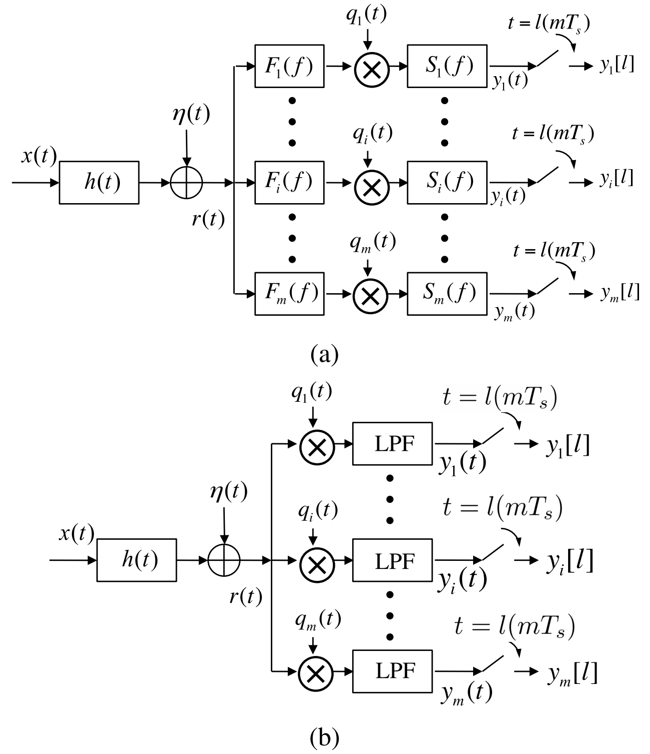

Fig. 1. (a) Sampling with modulation and filter banks: the channel output $r(t)$ is passed through m branches, each consisting of a pre-modulation filter, a periodic modulator and a post-modulation filter followed by a uniform sampler with sampling rate W/n. (b) Sampling with a bank of modulators and low-pass filters: the channel output is passed through m branches, each consisting of a modulator with modulation waveform $q_i(t)$ and a low-pass filter of pass band [0, W/n] followed by a uniform sampler at rate W/n.

Since the modulation waveform $q_i(t)$ is periodic, its Fourier transform can be represented by a weighted δ-train, namely,

$$\mathcal{F}(q_i(t)) = \sum_{l=-\infty}^{\infty} \hat{q}_{i,l}\,\delta(f + lW/n) \tag{3}$$

for some sequence $\{\hat{q}_{i,l}\}_{l\in\mathbb{Z}}$. This modulation operation scrambles the spectral content of the channel input $X(f)$. As can be seen, the signal after post-modulation filtering (i.e. $y_i(t)$ in Fig. 1(a)) has Fourier response

$$\sum_{l=0}^{n} \hat{q}_{i,l} S_i(f)\, F_i\left(f + l\frac{W}{n}\right) X\left(f + l\frac{W}{n}\right), \quad \forall f. \tag{4}$$

Due to aliasing, the final sampling output (i.e. $y_i[n]$ in Fig. 1(a)) is tantamount to a signal of Fourier

[2]Our main results and analytical tools can be extended to more general frequency-varying periodic sampling systems without difficulty. Interested readers are referred to [26] for details.

responses

$$\sum_{\tau=-\infty}^{\infty}\sum_{l=0}^{n-1}\hat{q}_{i,l}S_i\left(f+\tau\frac{W}{n}\right)F_i\left(f+(\tau+l)\frac{W}{n}\right) \cdot X\left(f+(\tau+l)\frac{W}{n}\right),\quad 0\le f<\frac{W}{n}. \tag{5}$$

Since $F_i(f)$, $S_i(f)$ are piecewise flat, one can write (5) as

$$\sum_{l=1}^{n}\boldsymbol{Q}_{i,l}X\left(f+(l-1)\frac{W}{n}\right),\quad 0\le f<\frac{W}{n} \tag{6}$$

for some sequence $\left\{\boldsymbol{Q}_{i,l}\right\}_{1\le i\le m,1\le l\le n}$. As a result, one can use an $m\times n$ matrix $\boldsymbol{Q}=\left[\boldsymbol{Q}_{i,l}\right]_{1\le i\le m,1\le l\le n}$, to represent the sampling system, termed a *sampling coefficient matrix*.

On the other hand, for any given $\boldsymbol{Q}\in\mathbb{C}^{m\times n}$, there exists a sampling system such that the Fourier response of its sampled output obeys (6). This can be realized via the m-branch sampling system illustrated in Fig. 1(b). In the ith branch, the channel output is modulated by a periodic waveform $q_i(t)$ with Fourier response

$$\mathcal{F}\left(q_i(t)\right)=\sum_{l=1}^{n}\boldsymbol{Q}_{i,l}\delta\left(f+(l-1)W/n\right),$$

passed through a low-pass filter with pass band $[0, W/n]$, and then uniformly sampled at rate W/n. In the current paper, a sampling system within this class is said to be *Gaussian sampling* if the entries of $\boldsymbol{Q}$ are i.i.d. Gaussian random variables. It turns out that Gaussian sampling structures suffice to achieve overall robustness in terms of sampled information rate loss, as will be seen in Section III.

C. Universal Sampling

As was shown in [1], the optimal sampling mechanism for a given LTI channel with perfect CSI extracts out the frequency set with the highest SNR and hence suppresses aliasing. Such an alias-suppressing sampler may result in a very low capacity for some channel support. In this paper, we desire a sampler that operates independent of the instantaneous subband occupancy, and our objective is to design a single linear sampling system that incurs minimal information rate loss across all possible channel occupancy. In particular, the information rate loss we consider is the gap between the capacity under sampling with and without spectral occupancy (i.e. Fig. 2(a) vs. Fig. 2(b)).

1) Sampled Information Rate Loss: For notational convenience, define the *undersampling factor* and the *sparsity factor* as

$$\begin{cases}\alpha & := m/n,\\ \beta & := k/n,\end{cases} \tag{7}$$

respectively. It will be assumed throughout that $\alpha,\beta\in(0,1)$ are some constants independent of n.

Our prior work [1] reveals that for any given state $\boldsymbol{s}$ and sampling rate $f_{\mathrm{s}}=\alpha W$, the capacity under channel-optimized sampling is given by

$$\begin{aligned}C_{\boldsymbol{s}}&=\frac{W}{2n}\min\{k,m\}\log\left(1+\frac{P}{\min\{\alpha,\beta\}W}\frac{|H|^2}{\mathcal{S}_\eta}\right)\\&=\frac{W}{2}\min\{\alpha,\beta\}\log\left(1+\mathsf{SNR}\right),\end{aligned} \tag{8}$$

where we set

$$\mathsf{SNR}:=\frac{P}{\min\{\alpha,\beta\}W}\frac{|H|^2}{\mathcal{S}_\eta}. \tag{9}$$

In addition, the channel capacity under the aforementioned filter-bank and modulation-bank sampling has also been derived [1, Th. 5]. When specialized to the frequency-flat channel model under uniform power allocation, the achievable rate at a given state $\boldsymbol{s}$ without subband occupancy information is given by

$$\begin{aligned}C_{\boldsymbol{s}}^{\boldsymbol{Q}}&=\frac{W}{2n}\log\det\left(\boldsymbol{I}_m+\mathsf{SNR}\cdot\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-\frac{1}{2}}\boldsymbol{Q}_{\boldsymbol{s}}\boldsymbol{Q}_{\boldsymbol{s}}^*\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-\frac{1}{2}}\right)\\&:=\frac{W}{2n}\log\det\left(\boldsymbol{I}_m+\mathsf{SNR}\cdot\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\right).\end{aligned} \tag{10}$$

Here, we let $\boldsymbol{A}_{\boldsymbol{s}}$ represent the submatrix of $\boldsymbol{A}$ consisting of the columns at indices of $\boldsymbol{s}$, and $\boldsymbol{Q}^{\mathrm{w}}:=\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-\frac{1}{2}}\boldsymbol{Q}$ the prewhitened sampling coefficient matrix. Note that $\boldsymbol{Q}^{\mathrm{w}}\boldsymbol{Q}^{\mathrm{w}*}:=\boldsymbol{I}_m$.

For any sampling system with a sampling coefficient matrix $\boldsymbol{Q}$, we define the *sampled information rate loss* for each state $\boldsymbol{s}$ as

$$L_{\boldsymbol{s}}^{\boldsymbol{Q}}:=C_{\boldsymbol{s}}-C_{\boldsymbol{s}}^{\boldsymbol{Q}}. \tag{11}$$

This metric quantifies the information rate loss of universal sampling due to the lack of subband occupancy information, i.e. the gap of achievable rates under the sampler in Fig. 2(a) relative to the sampler in Fig. 2(b).

2) Minimax Sampler: We aim to design a sampler that minimizes the loss function in some overall sense. Let L represent the *minimax information rate loss*, that is,

$$L:=\inf_{\boldsymbol{Q}}\max_{\boldsymbol{s}\in\binom{[n]}{k}}L_{\boldsymbol{s}}^{\boldsymbol{Q}}. \tag{12}$$

A sampling system associated with a sampling coefficient matrix $\boldsymbol{M}$ is then called a *minimax sampler* if it satisfies

$$\max_{\boldsymbol{s}\in\binom{[n]}{k}}L_{\boldsymbol{s}}^{\boldsymbol{M}}=L. \tag{13}$$

The minimax criterion is of interest for designing a sampler robust against all possible channel occupancy situations, that is, we expect the resulting sampled channel capacity to be within a minimal gap relative to maximum capacity in a uniform manner. Note that the minimax sampler is in general different from the one that maximizes the lowest capacity among all states (worst-case capacity). While the latter guarantees an optimal worst-case capacity that can be achieved regardless of which channel is realized, it may result in significant information rate loss in many other states, as illustrated in Fig. 3.

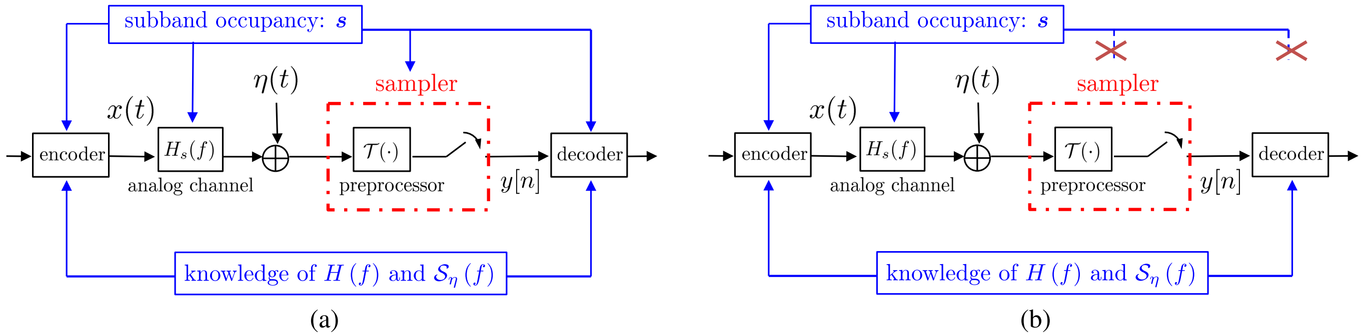

Fig. 2. Sampling with subband occupancy information vs. sampling without subband occupancy information (universal sampling).

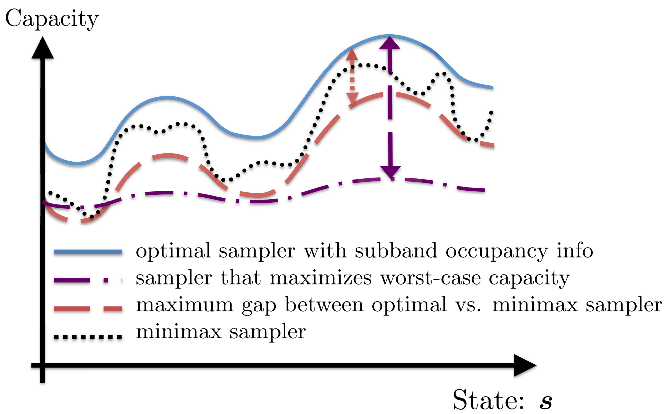

Fig. 3. Minimax sampler vs. the sampler that maximizes worst-case capacity. The blue solid line represents the capacity under channel-optimized sampling with subband occupancy information, the black dotted line represents the capacity achieved by minimax sampler, the orange dashed illustrates the maximum capacity minus the minimax information rate loss, while the purple dashed line corresponds to maximum worst-case capacity.

III. Minimax Sampled Information Rate Loss

The minimax sampled information rate loss problem boils down to minimizing $\max_{\boldsymbol{s}} L_{\boldsymbol{s}}^{\boldsymbol{Q}}$ over all sampling coefficient functions $\boldsymbol{Q}$. In general, this problem is non-convex in $\boldsymbol{Q}$, and hence it is computationally intractable to find the optimal sampler by solving a numerical optimization program. Fortunately, for the entire sub-Nyquist regime, the minimax sampled information rate loss can be quantified reasonably well at moderate-to-high SNR, and can be well approached by a sampler generated in a random fashion.

Our main results are summarized in the following theorem.

Theorem 1: Suppose that $0 < \alpha, \beta < 1$. *Define*

$$\Lambda := \min\{\beta, \alpha\} \log\left(1 + \frac{1}{\mathsf{SNR}}\right); \tag{14}$$

$$\Psi_1 := \min\left\{\left\lceil \frac{\beta}{\alpha - \beta} \right\rceil \frac{1}{\mathsf{SNR}}, (1+\beta)\frac{1}{\sqrt{\mathsf{SNR}}}\right\}; \tag{15}$$

$$\Psi_2 := \min\left\{\left\lceil \frac{1-\alpha}{\beta - \alpha} \right\rceil \frac{1}{\mathsf{SNR}}, (1+\alpha)\frac{1}{\sqrt{\mathsf{SNR}}}\right\}. \tag{16}$$

(a) If $\beta < \alpha$ *and* $\alpha + \beta < 1$, *then*

$$L = \frac{W}{2}\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Delta_1\right\}; \tag{17}$$

(b) If $\beta > \alpha$, *then*

$$L = \frac{W}{2}\left\{\mathcal{H}(\alpha) - \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) + \Delta_2\right\}; \tag{18}$$

(c) If $\beta = \alpha$ *or if* $\beta < \alpha$ *and* $\alpha + \beta \geq 1$, *then*

$$L = \frac{W}{2}\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Delta_3\right\}. \tag{19}$$

Here, $\Delta_1 \sim \Delta_3$ *are some residual terms obeying*

$$\Lambda - \Psi_1 - \frac{\log(n+1)}{n} \leq \Delta_1 \leq \Lambda + \frac{c_1 \log n}{n^{1/3}},$$
$$\Lambda - \Psi_2 - \frac{\log(n+1)}{n} \leq \Delta_2 \leq \Lambda + \frac{c_2 \log n}{n^{1/3}},$$
$$\Lambda - \Psi_1 - \frac{\log(n+1)}{n} \leq \Delta_3 \leq \Lambda + \frac{c_3 \mathsf{SNR}^{\frac{1}{3}} \log n}{n^{1/3}},$$

and $c_1 \sim c_3$ *are some universal constants independent of* n *and* SNR.

Remark 1: Note that $\mathcal{H}(\cdot)$ *denotes the binary entropy function. Its appearance is due to the fact that it is a tight estimate of the rate function of binomial coefficients.*

Theorem 1 provides a tight characterization of the minimax sampled information rate loss relative to the capacity under channel-optimized sampling. The minimax limits per unit bandwidth are given by

$$L \approx \begin{cases} \frac{1}{2}\mathcal{H}(\beta) - \frac{1}{2}\alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right), & \text{if } \alpha \geq \beta, \\ \frac{1}{2}\mathcal{H}(\alpha) - \frac{1}{2}\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right), & \text{if } \alpha \leq \beta, \end{cases} \tag{20}$$

modulo some residual terms. For $3/4$ of the sub-Nyquist regime, the residuals are at most of order $\mathcal{O}\left(\frac{\log n}{n^{1/3}} + \frac{1}{\mathsf{SNR}}\right)$, which are negligible at high SNR and when the number n of subbands is large. For another $1/4$ of the sub-Nyquist regime, our results are tight to within a gap $\mathcal{O}\left(\frac{\mathsf{SNR}^{1/3} \log n}{n^{1/3}} + \frac{1}{\mathsf{SNR}}\right)$, which will vanish if[3] $\frac{\mathsf{SNR}}{n} = o(1)$. For the special Landau-rate sampling case (i.e. $\alpha = \beta$), our bounds are accurate up to some gap $\mathcal{O}\left(\frac{\mathsf{SNR}^{1/3} \log n}{n^{1/3}} + \frac{1}{\sqrt{\mathsf{SNR}}}\right)$. We remark, however,

[3] This is a practically common situation. For instance, in the LTE communication systems, the median-to-high SNR for urban macrocells is typically between $10 \sim 20$dB, while the number of sub-carriers is around $128 \sim 2048$ [27, Ch. 26].

that the factor $\frac{\mathsf{SNR}^{1/3}\log n}{n^{1/3}}$ is not an optimal order and might be refined by other techniques.

The proof of Theorem 1 involves the verification of two parts: a converse part that provides a lower bound on the minimax sampled information rate loss, and an achievability part that provides a sampling scheme to approach this bound. As we show, the class of sampling systems with random periodic modulation followed by low-pass filters, as illustrated in Fig. 1(b), is sufficient to approach the minimax sampled loss.

A. Lower Bound on the Minimax Information Rate Loss

We need to demonstrate that the minimax sampled information rate loss under any channel-independent sampler cannot be lower than (17)-(19) in respective regimes. This is given by the theorem below.

Theorem 2: (1) If $\beta \leq \alpha \leq 1$, then

$$L \geq \frac{W}{2}\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Lambda - \frac{\log(n+1)}{n} - \min\left\{\left\lceil\frac{\beta}{\alpha-\beta}\right\rceil \log\left(1+\frac{1}{\mathsf{SNR}}\right), (1+\beta)\log\left(1+\frac{1}{\mathsf{SNR}^{\frac{1}{2}}}\right)\right\}\right\}. \tag{21}$$

(2) If $0 < \alpha < \beta$, then

$$L \geq \frac{W}{2}\left\{\mathcal{H}(\alpha) - \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) + \Lambda - \frac{\log(n+1)}{n} - \min\left\{\left\lceil\frac{1-\alpha}{\beta-\alpha}\right\rceil \log\left(1+\frac{1}{\mathsf{SNR}}\right), (1+\alpha)\log\left(1+\frac{1}{\mathsf{SNR}^{\frac{1}{2}}}\right)\right\}\right\} \tag{22}$$

Here, Λ is defined in (14).

For the entire sub-Nyquist regime, the lower bounds we derive are tantamount to some constants dependent only on α and β, except for some residual terms that vanish when the number n of subbands and the SNR tend to infinity. More precisely, when $\alpha \neq \beta$, one has $\Lambda, \Psi = \mathcal{O}\left(\frac{1}{\mathsf{SNR}}\right)$, and hence these residuals are at most the order of $\mathcal{O}\left(\frac{1}{\mathsf{SNR}} + \frac{\log n}{n}\right)$. In contrast, in the Landau-rate regime ($\alpha = \beta$), the residual term is bounded in magnitude by $\frac{2}{\sqrt{\mathsf{SNR}}} + \frac{\log(n+1)}{n}$. In fact, the term $\mathcal{O}\left(\frac{\log n}{n}\right)$ arises when using the entropy function to approximate the rate of binomial coefficients, while an additional approximation loss $\mathcal{O}\left(\frac{1}{\mathsf{SNR}}\right)$ occurs when employing $\log\mathsf{SNR}$ to approximate $\log(1+\mathsf{SNR})$.

B. Achievability

In general, it is computationally intractable to find a deterministic solution to approach the minimax limits by solving a numerical optimization program. Fortunately, when n and SNR are both large, simple random sampling strategies suffice in approaching the minimax information rate loss limits uniformly under all channel occupancy. The achievability result is formally stated in the following theorem.

Theorem 3: Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ be a Gaussian matrix such that $\boldsymbol{M}_{ij}$'s are independently drawn from $\mathcal{N}(0,1)$. Then with probability exceeding $1 - C\exp(-n)$, the following holds:

(a) If $\beta < \alpha$ and $\alpha + \beta < 1$, then

$$\max_{s\in\binom{[n]}{k}} L_{s}^{\boldsymbol{M}} \leq \frac{W}{2}\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Lambda + \frac{c_1\log n}{n^{1/3}}\right\}; \tag{23}$$

(b) If $\alpha < \beta$, then

$$\max_{s\in\binom{[n]}{k}} L_{s}^{\boldsymbol{M}} \leq \frac{W}{2}\left\{\mathcal{H}(\alpha) - \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) + \Lambda + \frac{c_2\log n}{n^{1/3}}\right\}; \tag{24}$$

(c) If $\beta = \alpha$ or if $\beta < \alpha$ and $\alpha + \beta \geq 1$, then

$$\max_{s\in\binom{[n]}{k}} L_{s}^{\boldsymbol{M}} \leq \frac{W}{2}\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Lambda + \frac{c_3\mathsf{SNR}^{\frac{1}{3}}\log n}{n^{1/3}}\right\}. \tag{25}$$

Here, $C, c_1 \sim c_3 > 0$ are some universal constants independent of SNR *and n, and Λ is given in (14).*

Theorem 3 indicates that Gaussian sampling approaches the minimax information rate loss (which is about $\frac{1}{2}\mathcal{H}(\beta) - \frac{1}{2}\alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right)$ per Hertz) to within a small gap. In fact, with exponentially high probability, the sampled information rate loss is almost equivalent to the minimax limit uniformly across all states $\boldsymbol{s} \in \binom{[n]}{k}$, as will be shown later.

IV. Equivalent Algebraic Problems

Our main results in Section III can be established by investigating equivalent algebraic problems. Observe that the information rate loss can be expressed as

$$\begin{aligned} L_{\boldsymbol{s}}^{\boldsymbol{Q}} &= -C_{\boldsymbol{s}}^{\boldsymbol{Q}} + C_{\boldsymbol{s}} \\ &= -\frac{W}{2n}\log\det\left(\boldsymbol{I}_m + \mathsf{SNR}\cdot\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\right) \\ &\quad + \frac{W\min\{k,m\}}{2n}\left\{\log\mathsf{SNR} + \log\left(1+\frac{1}{\mathsf{SNR}}\right)\right\} \end{aligned} \tag{26}$$

$$= \begin{cases} \frac{W}{2}\left\{-\frac{1}{n}\log\det\left(\frac{1}{\mathsf{SNR}}\boldsymbol{I}_k + \boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\right) + \Lambda\right\}, & \text{if } k \leq m \\ \frac{W}{2}\left\{-\frac{1}{n}\log\det\left(\frac{1}{\mathsf{SNR}}\boldsymbol{I}_m + \boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\right) + \Lambda\right\}, & \text{if } k \geq m \end{cases} \tag{27}$$

where $\Lambda = \mathcal{O}\left(\frac{1}{\mathsf{SNR}}\right)$ is defined in (14). This identity makes $\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\right)$ a quantity of interest. In the sequel, we provide tight bounds on this quantity, which in turn establish Theorems 2-3. The proofs of these results rely heavily on *non-asymptotic* (random) matrix theory. In particular, the proofs for the achievability bounds are established based on measure concentration of log-determinant functions, which will be provided at the end of this section.

A. Upper Bound on Log Determinants

Recall that $\boldsymbol{Q}^{\mathrm{w}}\boldsymbol{Q}^{\mathrm{w}*} = \boldsymbol{I}$, and that $\boldsymbol{B_s}$ represents the $m \times k$ submatrix of $\boldsymbol{B}$ with columns coming from the index set $\boldsymbol{s}$. The following theorem investigates the properties of $\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B_s}\right)$ for any $m \times n$ matrix $\boldsymbol{B}$ that has orthonormal rows.

Theorem 4: Consider any $\epsilon > 0$. Let $\boldsymbol{B}$ be any $m \times n$ matrix that satisfies $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}_m$.

(1) If $\beta \le \alpha \le 1$, then

$$\begin{aligned}\min_{\boldsymbol{s}\in\binom{[n]}{k}} & \log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B_s}\right)\\ &\le \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta) + \frac{\log(n+1)}{n}\\ &\quad + \min\left\{(1+\beta)\log\left(1+\sqrt{\epsilon}\right), \left\lceil\frac{\beta}{\alpha-\beta}\right\rceil\log(1+\epsilon)\right\}. \end{aligned} \tag{28}$$

(2) If $\alpha \le \beta \le 1$, then

$$\begin{aligned}\min_{\boldsymbol{s}\in\binom{[n]}{k}} & \log\det\left(\epsilon\boldsymbol{I}_m + \boldsymbol{B_s}\boldsymbol{B}_{\boldsymbol{s}}^*\right)\\ &\le \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) - \mathcal{H}(\alpha) + \frac{\log(n+1)}{n}\\ &\quad + \min\left\{(1+\alpha)\log\left(1+\sqrt{\epsilon}\right), \left\lceil\frac{(1-\alpha)\alpha}{\beta-\alpha}\right\rceil\log(1+\epsilon)\right\}. \end{aligned} \tag{29}$$

Theorem 4 together with (27) suggests that

$$\frac{L}{W/2} \ge \begin{cases}\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{O}\left(\frac{1}{\mathsf{SNR}}\right) - \frac{\log(n+1)}{n}, & \text{if } \alpha \ge \beta,\\ \mathcal{H}(\alpha) - \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) - \mathcal{O}\left(\frac{1}{\mathsf{SNR}}\right) - \frac{\log(n+1)}{n}, & \text{if } \alpha < \beta,\end{cases}$$

which completes the proof of Theorem 2.

One of the key ingredients in establishing Theorem 4 is to demonstrate that the sum

$$\sum_{\boldsymbol{s}\in\binom{[n]}{k}} \det\left(\epsilon\boldsymbol{I} + \boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B_s}\right) \tag{30}$$

is a *constant* independent of the matrix $\boldsymbol{B}$, as long as $\boldsymbol{B}$ has orthonormal rows. Consequently, in order to maximize $\min_{\boldsymbol{s}}\det\left(\epsilon\boldsymbol{I} + \boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B_s}\right)$, one would wish to find a matrix $\boldsymbol{B}$ such that $\det\left(\epsilon\boldsymbol{I} + \boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B_s}\right)$ are almost identical over all $\boldsymbol{s}$. When translated to the language of channel capacity, this observation suggests that an ideal minimax sampling method should be able to achieve (almost) equivalent information rate loss uniformly over all states $\boldsymbol{s}$, for which random sampling becomes a natural candidate due to sharp concentration of measures.

B. Achievability Under Gaussian Ensembles

When it comes to the achievability part, the major step is to quantify $\log\det(\epsilon\boldsymbol{I} + (\boldsymbol{M}\boldsymbol{M}^\top)^{-1}\boldsymbol{M_s}\boldsymbol{M}_{\boldsymbol{s}}^\top)$ for every $\boldsymbol{s} \in \binom{[n]}{k}$. Interestingly, this quantity can be uniformly controlled due to the concentration of spectral measure of random matrices [23]. This is stated in the following theorem, which demonstrates the optimality of Gaussian sampling mechanisms.

Theorem 5: Consider any $\epsilon > 0$. Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ be an i.i.d. random matrix satisfying $\boldsymbol{M}_{ij} \sim \mathcal{N}(0,1)$.

(a) If $\beta < \alpha$ and $\alpha + \beta < 1$, then with probability at least $1 - C\exp(-n)$,

$$\begin{aligned}\min_{\boldsymbol{s}\in\binom{[n]}{k}} & \frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M}_{\boldsymbol{s}}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)\\ &\ge -\mathcal{H}(\beta) + \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \frac{c_1\log n}{n^{1/3}}. \end{aligned} \tag{31}$$

(b) If $\alpha < \beta \le 1$, then with probability at least $1 - 9\exp(-2n)$,

$$\begin{aligned}\min_{\boldsymbol{s}\in\binom{[n]}{k}} & \frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_m + (\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\boldsymbol{M_s}\boldsymbol{M}_{\boldsymbol{s}}^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\right)\\ &\ge \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) - \mathcal{H}(\alpha) - \frac{c_2\log n}{n^{1/3}}. \end{aligned} \tag{32}$$

(c) If $\beta = \alpha$ or if $\beta < \alpha$ and $\alpha+\beta \ge 1$, then with probability exceeding $1 - 9\exp(-2n)$,

$$\begin{aligned}\min_{\boldsymbol{s}\in\binom{[n]}{k}} & \frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M}_{\boldsymbol{s}}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)\\ &\ge \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta) - \frac{c_3\log n}{(\epsilon n)^{1/3}}. \end{aligned} \tag{33}$$

Here, $c_1, c_2, c_3, C > 0$ are some universal constants independent of n and ϵ.

Putting Theorem 5 and Equation (27) together implies that $\forall \boldsymbol{s} \in \binom{[n]}{k}$,

$$\frac{L_{\boldsymbol{s}}^{\boldsymbol{M}}}{W/2} \le \begin{cases}\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Lambda + \mathcal{O}\left(\frac{\log n}{n^{\frac{1}{3}}}\right), & \text{if } \beta < \alpha \text{ and } \alpha+\beta < 1\\ \mathcal{H}(\alpha) - \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) + \Lambda + \mathcal{O}\left(\frac{\log n}{n^{\frac{1}{3}}}\right), & \text{if } \beta > \alpha\\ \mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + \Lambda + \mathcal{O}\left(\frac{\mathsf{SNR}^{\frac{1}{3}}\log n}{n^{\frac{1}{3}}}\right), & \text{if } \beta = \alpha \text{or if } \beta < \alpha \text{ and } \alpha+\beta \ge 1\end{cases}$$

with exponentially high probability, which establishes Theorem 3. The above achievability bounds are established via the concentration of spectral measure of large random matrices.

C. Measure Concentration of Log-Determinant Functions for Random Matrices

As mentioned above, the key machinery in establishing the achievability bounds is to evaluate certain log-determinant functions. In fact, many limiting results for i.i.d. Gaussian ensembles have been derived when studying MIMO fading channels (e.g., [28]–[31]), which focus on the first-order limits instead of the convergence rate. Furthermore, the second-order asymptotics and the large deviation for mutual information have also been studied (e.g., [32], [33]) in the asymptotic regime of large n. Most of these results focus on a special case of the log-determinant function (i.e. $\log\det(\epsilon \boldsymbol{I} + \frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top)$) and suppose that n scales independent of ϵ. On the other hand, the concentration of some log-determinant functions has been studied in the random matrix literature as a key step in establishing universal laws for linear spectral statistics (e.g. [34, Proposition 48]). However, these bounds are only shown to hold with overwhelming probability (i.e. with probability $1 - e^{-\omega(\log n)}$), which are not sharp enough for our purpose. As a result, we provide sharper measure concentration results of log-determinants in this subsection.

One important class of log-determinant functions takes the form of $\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)$. The concentration of such functions for i.i.d. rectangular Gaussian matrices is characterized in the following lemmas.

Lemma 1: Suppose that $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a random matrix whose entries are independent standard Gaussian random variables. Assume that $0 < \alpha < 1$. Then for any $\delta > 0$ and any $\tau > 0$,

$$\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) < \delta\right\}}{n} < \frac{\alpha}{1-\alpha-\frac{1}{n}}\delta + \frac{4\sqrt{\alpha\tau}}{\sqrt{n\delta}} \tag{34}$$

holds with probability exceeding $1 - 2\exp(-\tau n)$.

Proof: See Appendix C. ∎

Lemma 2: Suppose that $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a random matrix with independent entries satisfying $\boldsymbol{A}_{ij} \sim \mathcal{N}(0,1)$. Assume that $0 < \alpha < 1$.

(1) For any $\tau > 0$ and any $n > \max\left\{\frac{2}{1-\sqrt{\alpha}}, \frac{2}{\tau}, 7\right\}$,

$$\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) \le (1-\alpha)\log\frac{1}{1-\alpha} - \alpha + \frac{2\log n}{n} + \frac{5\sqrt{\alpha}}{\left(1-\sqrt{\alpha}-\frac{2}{n}\right)}\frac{\tau}{\sqrt{n}} \tag{35}$$

with probability exceeding $1 - 2\exp\left(-2\tau^2 n\right)$.

(2) For any $n > \max\left\{\frac{6.414}{1-\alpha}\cdot e^{\frac{\tau^2}{1-\alpha}}, \left(\frac{\alpha}{1-\alpha-\frac{1}{n}} + 4\sqrt{\alpha}\tau\right)^3\right\}$ and any $\tau > 0$,

$$\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) \ge (1-\alpha)\log\frac{1}{1-\alpha} - \alpha - \frac{\left(\frac{2}{1-\alpha-\frac{1}{n}} + 10\tau\right)\log n}{n^{1/3}} \tag{36}$$

with probability exceeding $1 - 7\exp(-\tau^2 n)$.

Proof: See Appendix D. ∎

The last log-determinant function considered here takes the form of $\log\det\left(\epsilon\boldsymbol{I} + \boldsymbol{A}^\top\boldsymbol{B}^{-1}\boldsymbol{A}\right)$ for some independent random matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, as stated in the following lemma.

Lemma 3: Suppose that $\beta < \alpha$ and $\alpha + \beta \le 1$. Let $\boldsymbol{A} = \mathbb{R}^{m\times k}$ be a random matrix whose entries are independent standard Gaussian random variables, and let $\boldsymbol{B} \sim \mathcal{W}_m(n-k, \boldsymbol{I}_m)$ be independent of $\boldsymbol{A}$. Then for any $\tau > 0$,

$$\begin{aligned}\frac{1}{n}\log\det&\left(\epsilon\boldsymbol{I}_k + \boldsymbol{A}^\top\boldsymbol{B}^{-1}\boldsymbol{A}\right)\\ &\ge -(\alpha-\beta)\log(\alpha-\beta) + \alpha\log\alpha - \beta\log(1-\alpha)\\ &\quad + (1-\alpha-\beta)\log\left(1-\frac{\beta}{1-\alpha}\right) - \frac{(c_8+c_9\tau)\log n}{n^{1/3}}\end{aligned} \tag{37}$$

with probability exceeding $1-9\exp\left(-\tau^2 n\right)$, where $c_8, c_9 > 0$ are some universal constants[4] independent of n and ϵ.

Proof: See Appendix E. ∎

As demonstrated in (27), the information rate loss is captured by some logarithmic function. The preceding concentration of measure results will prove useful in determining such a information rate loss metric.

V. Discussion

A. Implications of Main Results

In this subsection, we summarize several key insights from the main theorems.

(1) For the whole sub-Nyquist regime, the minimax information rate loss is captured by several binary entropy functions. When the number of subbands and the SNR are sufficiently large and $\frac{\mathsf{SNR}}{n} = o(1)$, the minimax limits depend almost only on the undersampling factor and the sparsity factor rather than (n, k, m), which are plotted in Fig. 4. It can be observed from the plot that when sampling above the Landau rate (but below the Nyquist rate), increasing the α/β ratio improves the capacity gap, and shrinks the locus. In contrast, when $\alpha < \beta$, increasing α/β results in a worse capacity gap. This implies that in the sub-Landau regime, the *relative information rate loss* is easier to control when α/β decreases, although the achievable channel capacity also shrinks. In fact, the information rate loss is the largest under Landau-rate sampling, as illustrated in Fig. 4. Since the capacity under channel-optimized sampling scales as $\Theta(W\log\mathsf{SNR})$, our results indicate that the ratio of the minimax information rate loss to the maximum capacity vanishes at a rate $\Theta(1/\log\mathsf{SNR})$.

(2) Note that under Landau-rate sampling (i.e. $\alpha = \beta$), the minimax loss is $\frac{1}{2}\mathcal{H}(\beta)$ (or $\frac{1}{2}\mathcal{H}(\alpha)$) modulo some residual terms. As a result, if we fix the channel sparsity factor and increase the sampling rate above the Landau rate, then the capacity benefit per unit bandwidth is captured by the term $\frac{1}{2}\alpha\mathcal{H}(\beta/\alpha)$. On the other hand, if we fix the sampling rate but increase the channel occupancy, then the capacity gain per Hertz one can harvest amounts to $\frac{1}{2}\beta\mathcal{H}(\alpha/\beta)$. For either case, if $\alpha \to 1$, the information rate loss per Hertz reduces to

$$\frac{1}{2}\mathcal{H}(\beta) - \frac{1}{2}\alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) = 0,$$

[4]More precisely, by setting $\zeta := \max\left\{\frac{\beta}{\alpha}, \frac{\alpha}{1-\beta}\right\}$, one can take $c_8 = \frac{3}{1-\zeta}$ and $c_9 = \frac{8(2-\sqrt{\zeta})}{1-\sqrt{\zeta}}$ for sufficiently large n.

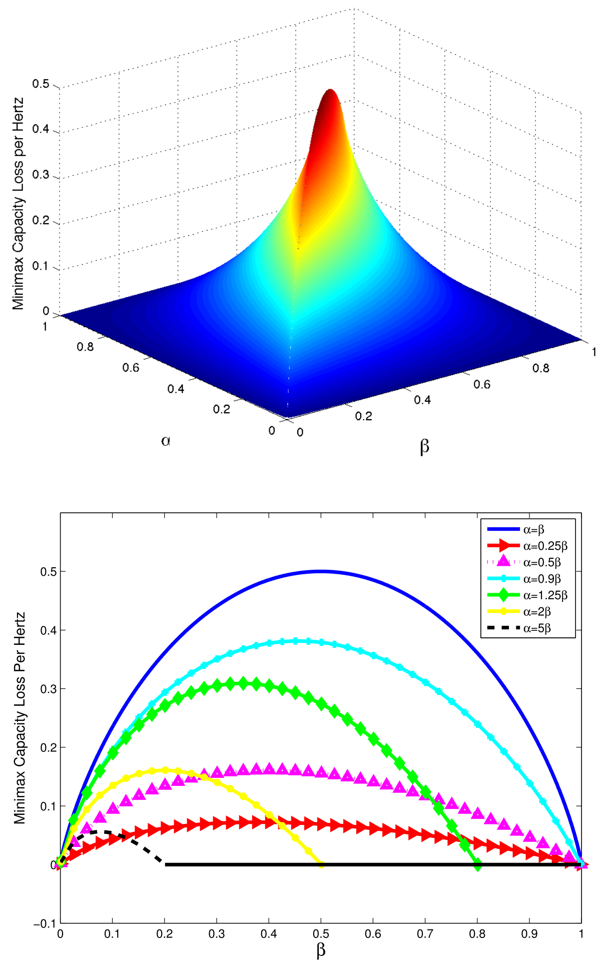

Fig. 4. The minimax loss per Hertz (without residual terms) vs. the sparsity factor β and the undersampling factor α.

meaning that there is effectively no information rate loss under Nyquist-rate sampling. This agrees with the fact that Nyquist-rate sampling is information preserving.

(3) The information rate loss incurred by Gaussian random sampling meets the minimax limit for Landau-rate sampling uniformly across all states $\boldsymbol{s}$, which reveals that with exponentially high probability, random sampling is optimal in terms of universal sampling design. This arises since the capacity achievable by random sampling exhibits very sharp measure concentration.

B. Extension

- **Universality of Random Sampling Schemes Beyond Gaussian Sampling**. While the main theorems provided in the present paper focus on Gaussian sampling, we remark that a much broader class of random sampling strategies are also minimax-optimal. This subsumes the class of sampling coefficient matrices $\boldsymbol{M}$ such that its entries are independent sub-Gaussian random variables with matching moments up to the second order. The *universality* phenomenon that arises in large random matrices (e.g. [35]) suggests that the minimaxity of random sampling matrices does not depend on the particular distribution of the coefficients, although they might affect the convergence rate to some degree. Interested readers are referred to [26] for derivation of these results and associated insights.
- **Beyond Frequency-Flat Channels and Uniform Power Allocation**. The present paper concentrates on the frequency-flat channel models for simplicity of presentation. Note, however, that the main results derived herein can be readily extended to more general frequency-varying channels. Specifically, suppose that the information rate loss metric is defined as the gap of achievable rates under universal sampling relative to channel-optimized sampling (both employing optimal power allocation). Then as long as the peak-to-average SNR
$$\frac{\sup_{f\in[0,W]} |H(f)|^2 / \mathcal{S}_\eta(f)}{\frac{1}{W}\int_0^W |H(f)|^2 / \mathcal{S}_\eta(f)}$$
is bounded, all results presented in this paper still hold, except for some additional gap on the order of $\mathcal{O}\left(\frac{1}{\mathsf{SNR}_{\min}}\right)$, where
$$\mathsf{SNR}_{\min} := \inf_{f\in[0,W]} \frac{P}{\min\{\alpha,\beta\}W} \frac{|H(f)|^2}{\mathcal{S}_\eta(f)}.$$
A proof of this result and derivation of the optimal power allocation is provided in [26].

VI. Conclusions

We have investigated minimax universal sampling design from a capacity perspective. In order to characterize the loss due to universal sub-Nyquist sampling design, we introduced the notion of sampled information rate loss relative to the capacity under channel-optimized sampling, and characterize overall robustness of the sampling design through the minimax information rate loss metric. Specifically, we have determined the minimax limit on the sampled information rate loss achievable by a class of channel-blind periodic sampling systems. This minimax limit turns out to be a constant that depends solely on the band sparsity factor and undersampling factor, modulo some residual term that vanishes as the SNR and the number of subbands grow. Our results demonstrate that with exponentially high probability, Gaussian random sampling is minimax-optimal in terms of a channel-blind sampler design.

It remains to study how to extend this framework to situations beyond compound multiband channels. Our notion of sampled information rate loss might be useful in studying the robustness for these scenarios. Our framework and results may also be appropriate for other channels with state where sparsity exists in other transform domains. In addition, when it comes to multiple access channels or random access channels [36], it would be interesting to see how to design a channel-blind sampler that is robust for the entire capacity region.

Appendix A
Proof of Theorem 4

Before proving the results, we first state two facts. Consider any $m \times m$ matrix $\boldsymbol{A}$, and list the eigenvalues of $\boldsymbol{A}$ as

$\lambda_1, \cdots, \lambda_m$. Define the characteristic polynomial of $\boldsymbol{A}$ as

$$\begin{aligned} p_{\boldsymbol{A}}(t) &= \det(t\boldsymbol{I} - \boldsymbol{A}) \\ &= t^m - S_1 t^{m-1} + \cdots + (-1)^m S_m, \end{aligned} \tag{38}$$

where S_l is the lth elementary symmetric function of $\lambda_1, \cdots, \lambda_m$ defined as follows:

$$S_l := \sum_{1 \le i_1 < \cdots < i_l \le m} \prod_{j=1}^{l} \lambda_{i_j}. \tag{39}$$

We also define $E_l(\boldsymbol{A})$ as the sum of determinants of all l-by-l principal minors of $\boldsymbol{A}$. According to [37, Th. 1.2.12], $S_l = E_l(\boldsymbol{A})$ holds for all $1 \le l \le m$, indicating that

$$\det(t\boldsymbol{I} + \boldsymbol{A}) = t^m + E_1(\boldsymbol{A}) t^{m-1} + \cdots + E_m(\boldsymbol{A}). \tag{40}$$

Another fact we will rely on is the entropy formula of binomial coefficients [38, Example 11.1.3], that is, for every $0 \le k \le n$,

$$\mathcal{H}\left(\frac{k}{n}\right) - \frac{\log(n+1)}{n} \le \frac{1}{n}\log\binom{n}{k} \le \mathcal{H}\left(\frac{k}{n}\right), \tag{41}$$

where $\mathcal{H}(x) := x\log\frac{1}{x} + (1-x)\log\frac{1}{1-x}$ stands for the entropy function.

Now we are in position to prove the theorem. For any $m \times n$ matrix $\boldsymbol{B}$ obeying $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}_m$, using the identity (40) we get

$$\begin{aligned} &\sum_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s \boldsymbol{B}_s^*\right) \\ &= \sum_{s \in \binom{[n]}{k}} \left\{ \epsilon^m + \sum_{l=1}^{m} \epsilon^{m-l} E_l\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right) \right\}, \qquad (42) \\ &= \epsilon^m \binom{n}{k} + \sum_{l=1}^{m} \epsilon^{m-l} \sum_{s \in \binom{[n]}{k}} E_l\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right). \qquad (43) \end{aligned}$$

Consider an index set $\boldsymbol{r} \in \binom{[m]}{l}$ with $l \le \min\{k, m\}$, and denote by $\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right)_{\boldsymbol{r}}$ the submatrix of $\boldsymbol{B}_s \boldsymbol{B}_s^*$ with rows and columns coming from the index set $\boldsymbol{r}$. It can be verified that

$$\det\left(\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right)_{\boldsymbol{r}}\right) = \det\left(\boldsymbol{B}_{\boldsymbol{r},s} \boldsymbol{B}_{\boldsymbol{r},s}^*\right) = \sum_{\tilde{\boldsymbol{r}} \in \binom{s}{l}} \det\left(\boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}} \boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}}^*\right),$$

where the last equality is a consequence of the Cauchy-Binet formula (e.g. [39]). Some algebraic manipulation yields that for any $l \le \min\{k, m\}$,

$$\begin{aligned} \sum_{s \in \binom{[n]}{k}} E_l\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right) &= \sum_{s \in \binom{[n]}{k}} \sum_{\boldsymbol{r} \in \binom{[m]}{l}} \det\left(\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right)_{\boldsymbol{r}}\right) \\ &= \sum_{s \in \binom{[n]}{k}} \sum_{\boldsymbol{r} \in \binom{[m]}{l}} \sum_{\tilde{\boldsymbol{r}} \in \binom{s}{l}} \det\left(\boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}} \boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}}^*\right) \\ &= \sum_{\boldsymbol{r} \in \binom{[m]}{l}} \sum_{\tilde{\boldsymbol{r}} \in \binom{[n]}{l}} \sum_{s: \tilde{\boldsymbol{r}} \subseteq s} \det\left(\boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}} \boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}}^*\right) \\ &\overset{(a)}{=} \sum_{\boldsymbol{r} \in \binom{[m]}{l}} \sum_{\tilde{\boldsymbol{r}} \in \binom{[n]}{l}} \binom{n-l}{k-l} \det\left(\boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}} \boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}}^*\right) \\ &\overset{(b)}{=} \sum_{\boldsymbol{r} \in \binom{[m]}{l}} \binom{n-l}{k-l} = \binom{n-l}{k-l}\binom{m}{l}, \qquad (44) \end{aligned}$$

where (a) follows since the number of k-combinations (out of $[n]$) containing $\tilde{\boldsymbol{r}}$ (an l-combination) is $\binom{n-l}{k-l}$, and (b) arises from the Cauchy-Binet formula together with $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}_m$, i.e.

$$\sum_{\tilde{\boldsymbol{r}} \in \binom{[n]}{l}} \det\left(\boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}} \boldsymbol{B}_{\boldsymbol{r},\tilde{\boldsymbol{r}}}^*\right) = \det\left(\boldsymbol{B}_{\boldsymbol{r},[n]} \boldsymbol{B}_{\boldsymbol{r},[n]}^*\right) = \det\left(\boldsymbol{I}_l\right) = 1.$$

(1) Suppose that $k \le m \le n$. The identity (43) reduces to

$$\begin{aligned} &\sum_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s \boldsymbol{B}_s^*\right) \\ &\quad = \epsilon^m \binom{n}{k} + \sum_{l=1}^{k} \epsilon^{m-l} \sum_{s \in \binom{[n]}{k}} E_l\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right), \end{aligned} \tag{45}$$

since any lth order ($l > k$) minor of $\boldsymbol{B}_s \boldsymbol{B}_s^*$ is rank deficient, i.e. $E_l\left(\boldsymbol{B}_s \boldsymbol{B}_s^*\right) = 0$. Substituting (44) into (45) yields

$$\sum_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s \boldsymbol{B}_s^*\right) = \sum_{l=0}^{k} \binom{n-l}{k-l}\binom{m}{l} \epsilon^{m-l}, \tag{46}$$

which further gives

$$\begin{aligned} \binom{n}{k} \min_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s^* \boldsymbol{B}_s\right) &\le \sum_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s^* \boldsymbol{B}_s\right) \\ &= \sum_{s \in \binom{[n]}{k}} \epsilon^{k-m} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s \boldsymbol{B}_s^*\right) \qquad (47) \\ &= \sum_{l=0}^{k} \binom{n-l}{k-l}\binom{m}{l} \epsilon^{k-l}. \qquad (48) \end{aligned}$$

The above expression allows us to derive an upper bound as

$$\begin{aligned} &\binom{n}{k} \min_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s^* \boldsymbol{B}_s\right) \\ &\le \sum_{l=0}^{k} \binom{n}{k-l}\binom{m}{l} \epsilon^{k-l} = \sum_{l=0}^{k} \binom{n}{l}\binom{m}{m-k+l} \epsilon^{l} \\ &= \binom{m}{k} \sum_{l=0}^{k} \binom{n}{l} \frac{\binom{m}{m-k+l}}{\binom{m}{k}} \epsilon^{l}. \qquad (49) \end{aligned}$$

Since the term $\binom{m}{m-k+l}/\binom{m}{k}$ can be bounded above by

$$\begin{aligned} \frac{\binom{m}{m-k+l}}{\binom{m}{k}} &= \frac{\frac{m!}{(m-k+l)!(k-l)!}}{\frac{m!}{k!(m-k)!}} = \frac{k!}{(k-l)!\frac{(m-k+l)!}{(m-k)!}} \\ &= \frac{\binom{k}{l}}{\binom{m-k+l}{l}} \le \binom{k}{l}, \end{aligned}$$

plugging this inequality into (49) we obtain

$$\begin{aligned} &\binom{n}{k} \min_{s \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s^* \boldsymbol{B}_s\right) \\ &\le \binom{m}{k} \sum_{l=0}^{k} \binom{n}{l}\binom{k}{l} \epsilon^{l} \le \binom{m}{k} \sum_{l=0}^{k} \binom{n+k}{2l} \left(\sqrt{\epsilon}\right)^{2l} \\ &\le \binom{m}{k} \left(1 + \sqrt{\epsilon}\right)^{n+k}, \qquad (50) \end{aligned}$$

where the last equality follows from the binomial theorem.

In addition, the identity (48) can also be written as

$$\sum_{s\in\binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I} + \boldsymbol{B}_s^*\boldsymbol{B}_s\right) = \sum_{l=0}^{k}\binom{n-k+l}{l}\binom{m}{k-l}\epsilon^l. \tag{51}$$

For any integer $K \geq n$, one can verify that $\forall 0 \leq l \leq k$,

$$\begin{aligned}\frac{\binom{n-k+l}{l}}{\binom{K}{l}}\frac{\binom{m}{k-l}}{\binom{m}{k}} &= \frac{\prod_{i=1}^{l}(n-k+i)}{\prod_{i=0}^{l-1}(K-i)}\cdot\frac{\prod_{i=1}^{l}(k-l+i)}{\prod_{i=0}^{l-1}(m-k+l-i)}\\ &\leq \frac{(n-k+l)^l}{(K-l+1)^l}\cdot\frac{k^l}{(m-k+1)^l}\\ &\leq \left(\frac{nk}{K(m-k)}\right)^l. \end{aligned}\tag{52}$$

Consequently, if $K \geq \frac{k}{m-k}n = \frac{\beta}{\alpha-\beta}n$, then

$$\binom{n-k+l}{l}\binom{m}{k-l} \leq \binom{K}{l}\binom{m}{k}, \tag{53}$$

which combined with (51) reveals that

$$\begin{aligned}&\binom{n}{k}\min_{s\in\binom{[n]}{k}}\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s^*\boldsymbol{B}_s\right)\\ &\quad\leq \binom{m}{k}\sum_{l=0}^{k}\binom{K}{l}\epsilon^l \leq \binom{m}{k}(1+\epsilon)^K. \end{aligned}\tag{54}$$

Set $K = \left\lceil\frac{\beta}{\alpha-\beta}\right\rceil n$. Putting the preceding bounds (50) and (54) together suggests that

$$\begin{aligned}&\min_{s\in\binom{[n]}{k}}\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s^*\boldsymbol{B}_s\right)\\ &\quad\leq \frac{1}{n}\log\binom{m}{k}-\frac{1}{n}\log\binom{n}{k}\\ &\qquad+\underbrace{\min\left\{\frac{n+k}{n}\log\left(1+\sqrt{\epsilon}\right),\frac{K}{n}\log(1+\epsilon)\right\}}_{:=\varrho_1(\epsilon)}. \end{aligned}$$

The entropy formula (41) then allows us to simplify:

$$\begin{aligned}&\min_{s\in\binom{[n]}{k}}\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s^*\boldsymbol{B}_s\right)\\ &\quad\leq \frac{m}{n}\mathcal{H}\left(\frac{k}{m}\right)-\mathcal{H}\left(\frac{k}{n}\right)+\frac{\log(n+1)}{n}+\varrho_1(\epsilon) \qquad (55)\\ &\quad= \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right)-\mathcal{H}(\beta)+\varrho_1(\epsilon)+\frac{\log(n+1)}{n} \qquad (56)\end{aligned}$$

as claimed.

(2) When $m \leq k$, the identity (43) combined with (44) leads to

$$\begin{aligned}&\binom{n}{k}\min_{s\in\binom{[n]}{k}}\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s\boldsymbol{B}_s^*\right) \leq \sum_{s\in\binom{[n]}{k}}\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s\boldsymbol{B}_s^*\right)\\ &\quad= \sum_{l=0}^{m}\binom{n-l}{k-l}\binom{m}{l}\epsilon^{m-l} = \sum_{l=0}^{m}\binom{n-m+l}{k-m+l}\binom{m}{l}\epsilon^l. \end{aligned}\tag{57}$$

Observe that

$$\begin{aligned}\frac{\binom{n-m+l}{k-m+l}}{\binom{n-m}{k-m}} &= \frac{\prod_{i=1}^{l}(n-m+i)}{\prod_{i=1}^{l}(k-m+i)} \leq \frac{\prod_{i=1}^{l}(n-m+i)}{l!}\\ &= \binom{n-m+l}{l} \leq \binom{n}{l}. \end{aligned}$$

This taken collectively with (57) suggests that

$$\begin{aligned}&\binom{n}{k}\min_{s\in\binom{[n]}{k}}\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s\boldsymbol{B}_s^*\right) \leq \binom{n-m}{k-m}\sum_{l=0}^{m}\binom{n}{l}\binom{m}{l}\epsilon^l\\ &\quad\leq \binom{n-m}{k-m}\sum_{l=0}^{m}\binom{n+m}{2l}\left(\sqrt{\epsilon}\right)^{2l}\\ &\quad\leq \binom{n-m}{k-m}\left(1+\sqrt{\epsilon}\right)^{n+m}. \end{aligned}\tag{58}$$

On the other hand, we claim that

$$\binom{n-m+l}{k-m+l}\binom{m}{l} \leq \binom{n-m}{k-m}\binom{K}{l}, \quad 0 \leq l \leq m \tag{59}$$

for some integer $K \geq m$. Putting this claim and (57) together leads to

$$\begin{aligned}\binom{n}{k}\min_{s\in\binom{[n]}{k}}\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s\boldsymbol{B}_s^*\right) &\leq \sum_{l=0}^{m}\binom{n-m}{k-m}\binom{K}{l}\epsilon^l\\ &\leq \binom{n-m}{k-m}(1+\epsilon)^K. \end{aligned}$$

This together with (58) implies that

$$\begin{aligned}&\min_{s\in\binom{[n]}{k}}\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{B}_s\boldsymbol{B}_s^*\right)\\ &\quad\leq \frac{1}{n}\log\binom{n-m}{k-m}-\frac{1}{n}\log\binom{n}{k}+\\ &\qquad\underbrace{\min\left\{\frac{K}{n}\log(1+\epsilon),\frac{n+m}{n}\log\left(1+\sqrt{\epsilon}\right)\right\}}_{:=\varrho_2(\epsilon)} \qquad (60)\\ &\quad\leq (1-\alpha)\mathcal{H}\left(\frac{\beta-\alpha}{1-\alpha}\right)-\mathcal{H}(\beta)+\frac{\log(n+1)}{n}+\varrho_2(\epsilon)\\ &\quad= -\mathcal{H}(\alpha)+\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right)+\frac{\log(n+1)}{n}+\varrho_2(\epsilon), \qquad (61)\end{aligned}$$

where the last inequality results from the fact $\log(1+\epsilon) \leq \epsilon$ as well as the following identity:

$$\begin{aligned}&\frac{1}{n}\log\frac{\binom{n-m}{k-m}}{\binom{n}{k}}\\ &= (1-\alpha)\mathcal{H}\left(\frac{\beta-\alpha}{1-\alpha}\right)-\mathcal{H}(\beta)\\ &= -(\beta-\alpha)\log\left(\frac{\beta-\alpha}{1-\alpha}\right)-(1-\beta)\log\left(\frac{1-\beta}{1-\alpha}\right)-\mathcal{H}(\beta)\\ &= -(\beta-\alpha)\log(\beta-\alpha)+(1-\alpha)\log(1-\alpha)+\beta\log\beta\\ &= -\mathcal{H}(\alpha)+\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right). \end{aligned}$$

Finally, it remains to establish the claim (59). In fact, when $K \geq m$, one has

$$\frac{\binom{n-m}{k-m}\binom{K}{l}}{\binom{n-m+l}{k-m+l}\binom{m}{l}} = \frac{\prod_{i=0}^{l-1}(K-i)}{\prod_{i=0}^{l-1}(m-i)} \cdot \frac{\prod_{i=1}^{l}(k-m+i)}{\prod_{i=1}^{l}(n-m+i)} \geq \left(\frac{K}{m}\right)^l \cdot \left(\frac{k-m}{n-m}\right)^l \geq 1,$$

provided that $K \geq \frac{n-m}{k-m}m = \frac{1-\alpha}{\beta-\alpha} \cdot m$, which justifies the claim (59). This taken collectively with (61) leads to

$$\begin{aligned}&\min_{s\in\binom{[n]}{k}} \frac{1}{n}\log\det\left(\epsilon \boldsymbol{I} + \boldsymbol{B_s}\boldsymbol{B_s}^*\right)\\ &\leq -\mathcal{H}(\alpha) + \beta\mathcal{H}\left(\frac{\alpha}{\beta}\right) + \frac{\log(n+1)}{n}\\ &\quad + \min\left\{\left\lceil\frac{1-\alpha}{\beta-\alpha}\right\rceil\log(1+\epsilon), (1+\alpha)\log\left(1+\sqrt{\epsilon}\right)\right\},\end{aligned}$$

concluding the proof.

Appendix B
Proof of Theorem 5

A. *Proof of Theorem 5(a)*

Our goal is to evaluate $\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top \left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)$ for some small $\epsilon > 0$. We first define two Wishart matrices

$$\Xi_{\backslash \boldsymbol{s}} := \frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top - \frac{1}{n}\boldsymbol{M_s}\boldsymbol{M_s}^\top; \tag{62}$$

$$\Xi_{\boldsymbol{s}} := \frac{1}{n}\boldsymbol{M_s}\boldsymbol{M_s}^\top. \tag{63}$$

Apparently, $\Xi_{\boldsymbol{s}} \sim \mathcal{W}_m\left(k, \frac{1}{n}\boldsymbol{I}_m\right)$ and $\Xi_{\backslash \boldsymbol{s}} \sim \mathcal{W}_m\left(n-k, \frac{1}{n}\boldsymbol{I}_m\right)$. When $1-\alpha > \beta$, i.e. $n-k > m$, the Wishart matrix $\Xi_{\backslash \boldsymbol{s}}$ is invertible with probability 1.

One difficulty in evaluating $\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top \left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)$ is that $\boldsymbol{M_s}$ and $\boldsymbol{M}\boldsymbol{M}^\top$ are not independent. This motivates us to decouple them first as follows

$$\begin{aligned}&\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)\\ &= \epsilon^{k-m}\det\left(\epsilon\boldsymbol{I}_m + \left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\frac{1}{n}\boldsymbol{M_s}\boldsymbol{M_s}^\top\right)\\ &= \epsilon^{k-m}\det\left(\epsilon\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top + \frac{1}{n}\boldsymbol{M_s}\boldsymbol{M_s}^\top\right)\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\\ &= \epsilon^{k-m}\det\left(\epsilon\,\Xi_{\backslash \boldsymbol{s}} + (1+\epsilon)\,\Xi_{\boldsymbol{s}}\right)\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\\ &= \epsilon^{k-m}\det\left(\epsilon\boldsymbol{I}_m + (1+\epsilon)\,\Xi_{\boldsymbol{s}}\Xi_{\backslash \boldsymbol{s}}^{-1}\right)\det\left(\Xi_{\backslash \boldsymbol{s}}\right)\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\\ &= \det\left(\epsilon\boldsymbol{I}_k + \frac{1+\epsilon}{n}\boldsymbol{M_s}^\top\Xi_{\backslash \boldsymbol{s}}^{-1}\boldsymbol{M_s}\right)\det\left(\Xi_{\backslash \boldsymbol{s}}\right)\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\end{aligned}$$

or, equivalently,

$$\begin{aligned}&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)\\ &= \frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + (1+\epsilon)\frac{1}{n}\boldsymbol{M_s}^\top\Xi_{\backslash \boldsymbol{s}}^{-1}\boldsymbol{M_s}\right)\\ &\quad + \frac{1}{n}\log\det\left(\Xi_{\backslash \boldsymbol{s}}\right) - \frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right).\end{aligned} \tag{64}$$

The point of developing this identity (64) is to decouple the left-hand side of (64) through 3 matrices $\boldsymbol{M_s}^\top\Xi_{\backslash \boldsymbol{s}}^{-1}\boldsymbol{M_s}$, $\Xi_{\backslash \boldsymbol{s}}$ and $\boldsymbol{M}\boldsymbol{M}^\top$. In particular, since $\boldsymbol{M_s}$ and $\Xi_{\backslash \boldsymbol{s}}$ are jointly independent, we can examine the concentration of measure for $\boldsymbol{M_s}$ and $\Xi_{\backslash \boldsymbol{s}}$ separately when evaluating $\boldsymbol{M_s}^\top\Xi_{\backslash \boldsymbol{s}}^{-1}\boldsymbol{M_s}$.

The second and third terms of (64) can be evaluated through Lemma 2. Specifically, Lemma 2 indicates that

$$\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right) \leq -(1-\alpha)\log(1-\alpha) - \alpha + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \tag{65}$$

with probability at least $1 - C_6\exp(-2n)$ for some constant $C_6 > 0$, and that for all $\boldsymbol{s} \in \binom{[n]}{k}$,

$$\begin{aligned}&\frac{1}{n}\log\det\left(\Xi_{\backslash \boldsymbol{s}}\right)\\ &= \frac{\log\det\left(\frac{n}{n-k}\Xi_{\backslash \boldsymbol{s}}\right)}{n} + \frac{\log\det\left(\frac{n-k}{n}\boldsymbol{I}\right)}{n}\\ &\geq (1-\beta)\left\{-\left(1-\frac{\alpha}{1-\beta}\right)\log\left(1-\frac{\alpha}{1-\beta}\right) - \frac{\alpha}{1-\beta}\right\}\\ &\qquad + \alpha\log(1-\beta) + \mathcal{O}\left(\frac{\log n}{n^{1/3}}\right) \qquad (66)\\ &\geq -(1-\alpha-\beta)\log\left(1-\frac{\alpha}{1-\beta}\right) - \alpha + \alpha\log(1-\beta)\\ &\qquad + \mathcal{O}\left(\frac{\log n}{n^{1/3}}\right). \qquad (67)\end{aligned}$$

hold simultaneously with probability exceeding $1 - C_9\exp(-2n)$.

Our main task then amounts to quantifying $\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top\Xi_{\backslash \boldsymbol{s}}^{-1}\boldsymbol{M_s}\right)$, which can be lower bounded via Lemma 3. This together with (65), (67) and (64) yields that

$$\begin{aligned}&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M_s}^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M_s}\right)\\ &\geq -(\alpha-\beta)\log(\alpha-\beta) + \alpha\log\alpha\\ &\qquad + (1-\alpha-\beta)\log\left(1-\frac{\beta}{1-\alpha}\right) - \beta\log(1-\alpha)\\ &\qquad - (1-\alpha-\beta)\log\left(1-\frac{\alpha}{1-\beta}\right) - \alpha + \alpha\log(1-\beta)\\ &\qquad + (1-\alpha)\log(1-\alpha) + \alpha - \mathcal{O}\left(\frac{\log n}{n^{1/3}}\right) \qquad (68)\\ &= \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta) - \mathcal{O}\left(\frac{\log n}{n^{1/3}}\right) \qquad (69)\end{aligned}$$

with probability exceeding $1 - C_9\exp(-2n)$ for some constants $C_9 > 0$.

Since there are at most $\binom{n}{k} \leq 2^n$ different states $\boldsymbol{s}$, applying the union bound over all states completes the proof.

B. Proof of Theorem 5(b)

We first recognize that

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_m+(\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\boldsymbol{M}_s\boldsymbol{M}_s^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\right)\\
&=-\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)+\frac{1}{n}\log\det\left(\frac{\epsilon\boldsymbol{M}\boldsymbol{M}^\top+\boldsymbol{M}_s\boldsymbol{M}_s^\top}{n}\right)\\
&\geq-\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)+\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}_s\boldsymbol{M}_s^\top\right)\\
&=-\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)+\frac{\beta}{k}\log\det\left(\frac{1}{k}\boldsymbol{M}_s\boldsymbol{M}_s^\top\right)+\alpha\log\beta.
\end{aligned}\tag{70}$$

When $\alpha<\beta\leq 1$, Lemma 2 implies that

$$\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)\leq(1-\alpha)\log\frac{1}{1-\alpha}-\alpha+\frac{c_8}{\sqrt{n}}\tag{71}$$

and

$$\frac{1}{k}\log\det\left(\frac{1}{k}\boldsymbol{M}_s\boldsymbol{M}_s^\top\right)\geq\left(1-\frac{\alpha}{\beta}\right)\log\frac{1}{1-\frac{\alpha}{\beta}}-\frac{\alpha}{\beta}-\frac{c_9\log n}{n^{1/3}}\tag{72}$$

hold with probability exceeding $1-9\exp(-3n)$, where c_8 and c_9 are some positive universal constants. Putting the above three bounds together gives

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_m+(\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\boldsymbol{M}_s\boldsymbol{M}_s^\top(\boldsymbol{M}\boldsymbol{M}^\top)^{-\frac{1}{2}}\right)\\
&\quad\geq-(1-\alpha)\log\frac{1}{1-\alpha}+\alpha+(\beta-\alpha)\log\frac{1}{1-\frac{\alpha}{\beta}}\\
&\qquad\quad-\alpha-\frac{c_{10}\log n}{n^{1/3}}+\alpha\log\beta \qquad (73)\\
&\quad=-\mathcal{H}(\alpha)+\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right)-\frac{c_{10}\log n}{n^{1/3}} \qquad (74)
\end{aligned}$$

for some universal constant $c_{10}>0$, where the last identity follows since

$$\begin{aligned}
&-(1-\alpha)\log\frac{1}{1-\alpha}+(\beta-\alpha)\log\frac{1}{1-\frac{\alpha}{\beta}}+\alpha\log\beta\\
&\quad=-\mathcal{H}(\alpha)-\alpha\log\alpha+\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right)-\alpha\log\frac{\beta}{\alpha}+\alpha\log\beta\\
&\quad=-\mathcal{H}(\alpha)+\beta\mathcal{H}\left(\frac{\alpha}{\beta}\right).
\end{aligned}\tag{75}$$

Applying the union bound over all $\binom{n}{k}$ states concludes the proof.

C. Proof of Theorem 5(c)

Without loss of generality, consider first the case where $\boldsymbol{s}=\{n-k+1,\cdots,n\}$. The quantity under study can be rearranged as

$$\begin{aligned}
&\log\det\left(\epsilon\boldsymbol{I}_k+\boldsymbol{M}_s^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M}_s\right)\\
&\quad=\log\left\{\epsilon^{k-m}\det\left(\epsilon\boldsymbol{I}_m+\boldsymbol{M}_s\boldsymbol{M}_s^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\right)\right\}\\
&\quad=\log\det\left(\epsilon\boldsymbol{M}\boldsymbol{M}^\top+\boldsymbol{M}_s\boldsymbol{M}_s^\top\right)-\log\det\left(\boldsymbol{M}\boldsymbol{M}^\top\right)\\
&\qquad\quad+(k-m)\log\epsilon \qquad (76)\\
&\quad=\log\det\left(\frac{1}{n}\boldsymbol{M}\underbrace{\begin{bmatrix}\epsilon\boldsymbol{I}_{n-k}&\\&(1+\epsilon)\boldsymbol{I}_k\end{bmatrix}}_{:=\boldsymbol{D}_\epsilon}\boldsymbol{M}^\top\right)\\
&\qquad\quad-\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)+(k-m)\log\epsilon. \qquad (77)
\end{aligned}$$

The term $\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)$ can be controlled by Lemma 2. Thus, it amounts to derive a reasonably tight lower bound on the term $\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)$.

Fortunately, the concentration of spectral measure inequality [23, Corollary 1.8] can also be applied to control the quantity $\sum_{i=1}^n f\left(\lambda_i\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right)$ for a variety of functions $f(\cdot)$. Consider the auxiliary functions

$$f_{1,\delta}(x):=\begin{cases}\frac{2}{\sqrt{\delta}}\left(\sqrt{x}-\sqrt{\delta}\right)+\log\epsilon, & 0<x<\delta,\\ \log x, & x\geq\delta,\end{cases}\tag{78}$$

and

$$\det{}^{\delta}(\boldsymbol{X}):=\prod_{i=1}^m e^{f_{1,\delta}(\lambda_i(\boldsymbol{X}))}.\tag{79}$$

If we set

$$Z_\epsilon:=\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)-\mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right],$$

then using [23, Corollary 1.8] we get[5]

$$\mathbb{P}\left\{\left|\frac{1}{n}Z_\epsilon\right|>4\sqrt{\frac{\alpha(1+\epsilon)}{\delta}}\sqrt{\frac{\tau}{n}}\right\}<2\exp(-2\tau n)\tag{80}$$

for any $\tau>0$. Furthermore, repeating the same argument as in (119) and (123) we obtain

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right]\\
&\qquad\geq\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right]\\
&\qquad\quad-\frac{5\alpha(1+\epsilon)}{n\delta},\quad\forall\delta<(1+\epsilon)\alpha,
\end{aligned}\tag{81}$$

and for sufficiently large n,

$$\mathbb{P}\left\{\lambda_{\min}\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)<\frac{\epsilon}{n^{7/3}}\right\}<3e^{-3n}.\tag{82}$$

[5]This arises since the concentration of spectral measure results given in [23, Corollary 1.8] depend only on the spectral norm $\|\boldsymbol{D}_\epsilon\|$.

By setting $\delta = \epsilon^{2/3} n^{-1/3}$, one has, with probability exceeding $1 - 7\exp(-\tau n)$, that

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right) - \frac{1}{n}\log\det{}^\delta\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\\
&\geq \sum_{i:\lambda_i\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)<\delta} \left\{\log\lambda_{\min}\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right) - \log\delta\right\}\\
&\geq -\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{\epsilon}{n}\boldsymbol{M}\boldsymbol{M}^\top\right) < \delta\right\}}{n}\cdot\log\left(\frac{n^2}{\epsilon^{1/3}}\right) \qquad (83)\\
&\geq -\left(\frac{\alpha}{\left(1-\alpha-\frac{1}{n}\right)} + 4\sqrt{\epsilon\alpha\tau}\right)\cdot\frac{\log\left(\frac{n^2}{\epsilon^{1/3}}\right)}{(\epsilon n)^{1/3}}, \qquad (84)
\end{aligned}$$

where (83) follows since $\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top \succeq \epsilon\boldsymbol{M}\boldsymbol{M}^\top$, and (84) arises from Lemma 1. This combined with (80) and (119) suggests that for any $\epsilon \in \left(\frac{1}{n}, 1\right)$, there exist universal constants $c_1 \sim c_4 > 0$ such that

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\\
&\geq \frac{1}{n}\mathbb{E}\left[\log\det{}^\delta\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right] - \frac{\left(c_1 + c_2\sqrt{\tau}\right)\log\frac{n^2}{\epsilon^{\frac{1}{3}}}}{(\epsilon n)^{1/3}} \qquad (85)\\
&\geq \frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right] - \frac{\left(c_3 + c_4\sqrt{\tau}\right)\log n}{(\epsilon n)^{1/3}} \qquad (86)
\end{aligned}$$

with probability exceeding $1 - 7\exp(-\tau n)$.

It remains to develop a tight lower bound on $\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right]$. Applying Cauchy-Binet identity gives

$$\begin{aligned}
&\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right]\\
&= \sum_{\boldsymbol{r},\tilde{\boldsymbol{r}}\in\binom{[n]}{m}} \mathbb{E}\left[\det\left(\boldsymbol{M}_{[m],\boldsymbol{r}}\right)\det\left(\frac{1}{n}\left(\boldsymbol{D}_\epsilon\right)_{\boldsymbol{r},\tilde{\boldsymbol{r}}}\right)\det\left(\boldsymbol{M}_{[m],\tilde{\boldsymbol{r}}}\right)\right]\\
&= \sum_{\boldsymbol{r}\in\binom{[n]}{m}} \mathbb{E}\left[\det\left(\boldsymbol{M}_{[m],\boldsymbol{r}}\boldsymbol{M}_{[m],\boldsymbol{r}}^\top\right)\right]\det\left(\frac{1}{n}\left(\boldsymbol{D}_\epsilon\right)_{\boldsymbol{r},\tilde{\boldsymbol{r}}}\right) \qquad (87)\\
&= \mathbb{E}\left[\det\left(\boldsymbol{M}_{[m],\boldsymbol{r}}\boldsymbol{M}_{[m],\boldsymbol{r}}^\top\right)\right]\sum_{\boldsymbol{r}\in\binom{[n]}{m}}\frac{1}{n^m}\det\left(\left(\boldsymbol{D}_\epsilon\right)_{\boldsymbol{r},\tilde{\boldsymbol{r}}}\right)\\
&= \frac{m!}{n^m}\sum_{l=\max\{m-k,0\}}^{\min\{n-k,m\}}\varphi_{n,k,m}(l), \qquad (88)
\end{aligned}$$

where we define $\varphi_{n,k,m}(l) := \binom{n-k}{l}\binom{k}{m-l}\epsilon^l(1+\epsilon)^{m-l}$. In the above arguments, the identity (87) arises since $\det\left(\left(\boldsymbol{D}_\epsilon\right)_{\boldsymbol{r},\tilde{\boldsymbol{r}}}\right) \neq 0$ only if $\boldsymbol{r} = \tilde{\boldsymbol{r}}$. The identity (88) follows from the definition of $\boldsymbol{D}_\epsilon$ as well as the fact that $\mathbb{E}\left[\det\left(\boldsymbol{M}_{[m],\boldsymbol{r}}\boldsymbol{M}_{[m],\boldsymbol{r}}^\top\right)\right] = m!$ (e.g. [40, Th. 3.1]).

Since $m \geq k$, (88) further yields

$$\begin{aligned}
\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right] &\geq \frac{m!}{n^m}\varphi_{n,k,m}(m-k)\\
&= \frac{m!}{n^m}\binom{n-k}{m-k}\epsilon^{m-k}(1+\epsilon)^k.
\end{aligned}$$

As a result,

$$\begin{aligned}
&\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\right]\\
&\geq \frac{1}{n}\log\left\{\frac{m!}{n^m}\binom{n-k}{m-k}\epsilon^{m-k}\right\}\\
&= \frac{\log m!}{n} - \frac{m\log m}{n} + \frac{m\log\left(\frac{m}{n}\right) + \log\binom{n-k}{m-k} + \log\epsilon^{m-k}}{n}\\
&\geq -\alpha + \alpha\log\alpha + (1-\beta)\mathcal{H}\left(\frac{\alpha-\beta}{1-\beta}\right) + \frac{\log\epsilon^{m-k}}{n} - \frac{\log n}{n}.
\end{aligned}$$

Here, the last inequality makes use of the entropy formula (41) as well as the following inequality

$$\begin{aligned}
\frac{\log m!}{n} - \frac{m\log m}{n} &\geq \frac{\left(m+\frac{1}{2}\right)}{n}\log m - \frac{m}{n} - \frac{m}{n}\log m\\
&\geq -\alpha, \qquad (89)
\end{aligned}$$

a consequence of the Stirling approximation (105). Substituting it back into the concentration bound (86) we get

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)\\
&\quad\geq -\alpha + \alpha\log\alpha + (1-\beta)\mathcal{H}\left(\frac{\alpha-\beta}{1-\beta}\right) - \frac{\log n}{n}\\
&\qquad - \frac{\left(c_3 + c_4\sqrt{\tau}\right)\log n}{(\epsilon n)^{1/3}} + \frac{1}{n}\log\left(\epsilon^{m-k}\right) \qquad (90)
\end{aligned}$$

with probability at least $1 - 7\exp(-\tau n)$.

So far we have developed lower bounds on the term $\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right)$. The above bound taken collectively with (77) and Lemma 2 leads to

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k + \boldsymbol{M}_s^\top\left(\boldsymbol{M}\boldsymbol{M}^\top\right)^{-1}\boldsymbol{M}_s\right)\\
&= \frac{\log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{D}_\epsilon\boldsymbol{M}^\top\right) - \log\det\left(\frac{1}{n}\boldsymbol{M}\boldsymbol{M}^\top\right)}{n} + \frac{k-m}{n}\log\epsilon\\
&\geq -\alpha + \alpha\log\alpha + (1-\beta)\mathcal{H}\left(\frac{\alpha-\beta}{1-\beta}\right) - \frac{\left(c_3 + c_4\sqrt{\tau}\right)\log n}{(\epsilon n)^{1/3}}\\
&\qquad - \frac{\log n}{n} - \left\{(1-\alpha)\log\frac{1}{1-\alpha} - \alpha + \frac{2\log n}{n} + c_5\sqrt{\frac{\tau}{n}}\right\}\\
&\geq \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta) - \frac{\left(c_5 + c_6\sqrt{\tau}\right)\log n}{(\epsilon n)^{1/3}} \qquad (91)
\end{aligned}$$

with probability at least $1 - 9\exp(-\tau n)$, where $c_5, c_6 > 0$ are some universal constants. Here, the inequality (91) follows from the following identity:

$$\begin{aligned}
&-\alpha + \alpha\log\alpha + (1-\beta)\mathcal{H}\left(\frac{\alpha-\beta}{1-\beta}\right) - (1-\alpha)\log\frac{1}{1-\alpha} + \alpha\\
&\quad= -\alpha + \alpha\log\alpha - (\alpha-\beta)\log\left(\frac{\alpha-\beta}{1-\beta}\right) + \alpha\\
&\qquad - (1-\alpha)\log\left(\frac{1-\alpha}{1-\beta}\right) + (1-\alpha)\log(1-\alpha)\\
&\quad= \alpha\log\alpha - (\alpha-\beta)\log(\alpha-\beta) + (1-\beta)\log(1-\beta)\\
&\quad= \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta). \qquad (92)
\end{aligned}$$

The proof is then complete by applying the union bound and setting $\tau = 3$.

Appendix C
Proof of Lemma 1

Since the indicator function $\mathbf{1}_{[0,\delta]}(\cdot)$ entails discontinuous points, we consider instead an upper bound on $\mathbf{1}_{[0,\delta]}(\cdot)$ as

$$f_{2,\delta}(x) := \begin{cases} 1, & \text{if } 0 \le x \le \delta; \\ -x/\epsilon + 2, & \text{if } \delta < x \le 2\delta; \\ 0, & \text{else.} \end{cases} \tag{93}$$

Note that for any $\delta > 0$ and $x \ge 0$, one has

$$\frac{\delta}{x} - \left(-\frac{x}{\delta} + 2\right) = \frac{\delta}{x} + \frac{x}{\delta} - 2 \ge 2\sqrt{\frac{\delta}{x} \cdot \frac{x}{\delta}} - 2 = 0.$$

This together with the facts $\frac{\epsilon}{x} \ge 1$ $(0 \le x \le \epsilon)$ and $\frac{\epsilon}{x} \ge 0$ $(x \ge 0)$ indicates that

$$f_{2,\delta}(x) \le \frac{\delta}{x}, \quad \forall x \ge 0, \tag{94}$$

leading to the upper bound

$$\begin{aligned} \sum_{i=1}^{n} f_{2,\delta}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right) &\le \sum_{i=1}^{m} \frac{\delta}{\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)} \\ &= \delta \mathrm{tr}\left(\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)^{-1}\right). \end{aligned} \tag{95}$$

It then follows from the property of inverse Wishart matrices (e.g. [41, Th. 2.2.8]) that

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f_{2,\delta}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right)\right] \le \frac{\delta}{n}\mathbb{E}\left[\mathrm{tr}\left(\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)^{-1}\right)\right] \\ &= \frac{m}{n-m-1}\delta = \frac{\alpha}{1-\alpha-\frac{1}{n}}\delta. \end{aligned} \tag{96}$$

Clearly, the Lipschitz constant of the function

$$g_{2,\delta}(x) := f_{2,\delta}(x^2) = \begin{cases} 1, & \text{if } 0 \le x \le \sqrt{\delta}; \\ -x^2/\delta + 2, & \text{if } \sqrt{\delta} < x \le \sqrt{2\delta}\,; \\ 0, & \text{else} \end{cases}$$

is bounded above by $\sqrt{8/\delta}$. Applying [23, Corollary 1.8(b)] then yields the following: for any $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{[0,\delta]}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right) > \frac{\alpha}{1-\alpha-\frac{1}{n}}\delta + \varepsilon\right) \\ &\quad \le \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f_{2,\delta}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right) > \frac{\alpha}{1-\alpha-\frac{1}{n}}\delta + \varepsilon\right) \\ &\quad \le \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} f_{2,\delta}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right)\right. \\ &\qquad\qquad \left. > \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f_{2,\delta}\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right)\right] + \varepsilon\right) \\ &\quad \le 2\exp\left(-\frac{\delta}{16\alpha}\varepsilon^2 n^2\right). \end{aligned} \tag{97}$$

Put in other words, for any $\tau > 0$, one has

$$\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) < \delta\right\}}{n} < \frac{\alpha}{1-\alpha-\frac{1}{n}}\delta + \frac{4\sqrt{\alpha\tau}}{\sqrt{n\delta}} \tag{98}$$

with probability exceeding $1 - 2\exp(-\tau n)$, as claimed.

In particular, by setting $\delta = n^{-1/3}$, one has

$$\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) < \frac{1}{n^{1/3}}\right\}}{n} < \frac{\frac{\alpha}{1-\alpha-\frac{1}{n}} + 4\sqrt{\alpha\tau}}{n^{1/3}} \tag{99}$$

with probability at least $1 - 2\exp(-\tau n)$.

Appendix D
Proof of Lemma 2

Before proceeding to the proof of measure concentration, we first derive tight bounds on the quantity $\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right]$. First, the Cauchy-Binet formula indicates that

$$\mathbb{E}\left[\det\left(\boldsymbol{A}\boldsymbol{A}^\top\right)\right] = \sum_{\boldsymbol{s}\in\binom{[n]}{m}} \mathbb{E}\left[\det\left(\boldsymbol{A}_{\boldsymbol{s}}\boldsymbol{A}_{\boldsymbol{s}}^\top\right)\right], \tag{100}$$

where $\boldsymbol{s}$ ranges over all m-combinations of $\{1, \cdots, n\}$, and $\boldsymbol{A}_{\boldsymbol{s}}$ is the $m \times m$ minor of $\boldsymbol{A}$ whose columns come from the columns of $\boldsymbol{A}$ at indices from $\mathbf{s}$. It is well known that (e.g. [40, Th. 3.1]) for i.i.d. random ensembles with zero mean and unit variance, the determinant satisfies

$$\mathbb{E}\left[\det\left(\boldsymbol{A}_{\boldsymbol{s}}\boldsymbol{A}_{\boldsymbol{s}}^\top\right)\right] = m!, \tag{101}$$

which immediately leads to

$$\begin{aligned} \mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right] &= \frac{\sum_{\boldsymbol{s}\in\binom{[n]}{m}} \mathbb{E}\left[\det\left(\boldsymbol{A}_{\boldsymbol{s}}\boldsymbol{A}_{\boldsymbol{s}}^\top\right)\right]}{n^m} \\ &= \frac{m!}{n^m}\binom{n}{m}. \end{aligned} \tag{102}$$

Next, from the well-known Stirling inequality

$$\sqrt{2\pi}\, m^{m+\frac{1}{2}}e^{-m} \le m! \le e m^{m+\frac{1}{2}}e^{-m}, \tag{103}$$

one can obtain

$$\log(m!) \le \left(m + \frac{1}{2}\right)\log m - m + 1; \tag{104}$$

$$\log(m!) \ge \left(m + \frac{1}{2}\right)\log m - m + \frac{1}{2}\log(2\pi). \tag{105}$$

These together with the entropy formula (41) give rise to

$$\begin{aligned} &\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right] \\ &\quad \le -\frac{m}{n}\log n + \frac{\left(m+\frac{1}{2}\right)\log m}{n} - \frac{m}{n} + \frac{1}{n} + \mathcal{H}\left(\frac{m}{n}\right) \\ &\quad = (1-\alpha)\log\frac{1}{1-\alpha} - \alpha + \frac{\log\left(e^2 m\right)}{2n} \end{aligned} \tag{106}$$

and, similarly,

$$\begin{aligned} &\frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right] \\ &\qquad \ge (1-\alpha)\log\frac{1}{1-\alpha} - \alpha - \frac{\log(n+1)}{n}. \end{aligned} \tag{107}$$

(1) We are now in a position to establish the upper bound on $\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)$. Since $\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right) = \sum_{i=1}^{m}\log\left(\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top\right)\right)$ is a separable function on the spectrum of $\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^\top$, one strategy is to make use of the concentration of spectral measure results [23]. Note, however, that the function

$\log x$ does not satisfy the Lipschitz condition required therein. To resolve this issue, we define an auxiliary function

$$f_{1,\delta}(x) := \begin{cases} \frac{2}{\sqrt{\delta}}\left(\sqrt{x}-\sqrt{\delta}\right)+\log\epsilon, & 0<x<\delta, \\ \log x, & x\geq\delta, \end{cases} \tag{108}$$

as well as

$$\det{}^{\delta}(\boldsymbol{X}) := \prod_{i=1}^{m} e^{f_{1,\delta}(\lambda_i(\boldsymbol{X}))}. \tag{109}$$

Apparently, $f_{1,\delta}(x)\geq\log x$, and the Lipschitz constant of the *concave* function

$$g_{1,\delta}(x) := f_{1,\delta}\left(x^2\right) = \begin{cases} \frac{2}{\sqrt{\delta}}\left(x-\sqrt{\delta}\right)+\log\delta, & 0<x<\sqrt{\delta}, \\ 2\log x & x\geq\sqrt{\delta}, \end{cases}$$

is bounded above by $\frac{2}{\sqrt{\delta}}$. By definition,

$$\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) = \det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \tag{110}$$

holds in the event that $\left\{\lambda_{\min}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\geq\delta\right\}$.

The deviation of $\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$ can be controlled via the concentration of spectral measure inequalities. Specifically, if we set

$$Z := \log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) - \mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right], \tag{111}$$

then [23, Corollary 1.8] suggests that for any $\tau>0$,

$$\mathbb{P}\left(|Z|>\tau\right) \leq 2\exp\left(-\frac{\delta\tau^2}{8\alpha}\right) \tag{112}$$

or, equivalently,

$$\mathbb{P}\left\{\left|\frac{1}{n}Z\right|>4\sqrt{\frac{\alpha}{\delta}}\sqrt{\frac{\tau}{n}}\right\} < 2\exp\left(-2\tau n\right). \tag{113}$$

Since $\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\leq\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$, it amounts to derive a tight upper bound on $\mathbb{E}\left[\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]$. Note that the behavior of the least singular value of a rectangular Gaussian matrix has been largely studied in the random matrix literature (e.g. [42, Corollary 5.35]), which we cite as follows.

Lemma 4: Consider a Gaussian random matrix $\boldsymbol{M}\in\mathbb{R}^{m\times n}$ such that $\boldsymbol{M}_{ij}$'s are i.i.d. standard Gaussian random variables. For any constant $0<\xi<\sqrt{n}-\sqrt{m}$,

$$\sigma_{\min}\left(\boldsymbol{M}\boldsymbol{M}^{*}\right) > \left(\sqrt{n}-\sqrt{m}-\xi\right)^2 \tag{114}$$

with probability at least[6] $1-\exp\left(-\xi^2/2\right)$.

Lemma 4 indicates that if $n>\frac{2}{1-\sqrt{\alpha}}$, then the event

$$\lambda_{\min}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \geq \left(1-\sqrt{\alpha}-\frac{2}{n}\right)^2$$

occurs with probability at least $1-\exp\left(-\frac{2}{n}\right)$. Conditioned on this event, we have

$$\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) = \det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \tag{115}$$

[6]Note that this follows from [42, Proposition 5.34 and Th. 5.32] by observing that $\sigma_{\min}(\boldsymbol{M})$ is a 1-Lipschitz function.

holds for the numerical value $\delta=\left(1-\sqrt{\alpha}-\frac{2}{n}\right)^2$. This taken collectively with (113) (by setting $\tau=\frac{1}{4}$) yields that for any $n>\max\left\{\frac{2}{1-\sqrt{\alpha}},5\right\}$,

$$\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) = \det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$$
$$\text{and}\quad \det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) > \frac{1}{e^{\sqrt{\frac{4\alpha}{\delta n}}}}e^{\mathbb{E}\left[\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]}$$

hold simultaneously with probability at least $1-\exp\left(-\frac{2}{n}\right)-2\exp\left(-\frac{n}{2}\right)>1/n$. Since $\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$ is non-negative, taking expectation gives

$$\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \geq \frac{1}{n}\cdot\frac{1}{e^{\sqrt{\frac{4\alpha}{\delta n}}}}e^{\mathbb{E}\left[\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]}, \tag{116}$$

and therefore for any $n>\max\left\{\frac{2}{1-\sqrt{\alpha}},7\right\}$ and $\delta=\left(1-\sqrt{\alpha}-\frac{2}{n}\right)^2$,

$$\begin{aligned} &\frac{1}{n}\mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \\ &\leq \frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]+\frac{\log n}{n}+\frac{\sqrt{\frac{4\alpha}{\delta}}}{n^{1.5}} \\ &\leq (1-\alpha)\log\frac{1}{1-\alpha}-\alpha+\frac{\log\left(e^2m\right)+2\log n}{2n}+\frac{2\sqrt{\alpha}n^{-1.5}}{1-\sqrt{\alpha}-\frac{2}{n}} \\ &\leq (1-\alpha)\log\frac{1}{1-\alpha}-\alpha+\frac{2\log n}{n}+\frac{2\sqrt{\alpha}}{\left(1-\sqrt{\alpha}-\frac{2}{n}\right)}\frac{1}{n^{1.5}}, \end{aligned}$$

where the second inequality follows from (106). Putting this and (113) together gives that for any $n>\max\left\{\frac{2}{1-\sqrt{\alpha}},7,\frac{2}{\sqrt{\tau}}\right\}$,

$$\begin{aligned} \frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) &\leq \frac{4\sqrt{\alpha}}{\left(1-\sqrt{\alpha}-\frac{2}{n}\right)}\sqrt{\frac{\tau}{n}}-\alpha \\ &\quad +(1-\alpha)\log\frac{1}{1-\alpha}+\frac{2\log n}{n}+\frac{2\sqrt{\alpha}}{1-\sqrt{\alpha}-\frac{2}{n}}\frac{1}{n^{1.5}} \\ &< (1-\alpha)\log\frac{1}{1-\alpha}-\alpha+\frac{2\log n}{n} \\ &\quad +\frac{4\sqrt{\alpha}}{\left(1-\sqrt{\alpha}-\frac{2}{n}\right)\sqrt{n}}\left(\sqrt{\tau}+\frac{1}{2n}\right) \\ &\leq (1-\alpha)\log\frac{1}{1-\alpha}-\alpha+\frac{2\log n}{n}+\frac{5\sqrt{\alpha}}{1-\sqrt{\alpha}-\frac{2}{n}}\sqrt{\frac{\tau}{n}} \end{aligned}$$

with probability exceeding $1-2\exp\left(-2\tau n\right)$, as claimed.

(2) In order to derive a lower bound on $\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$ based on (113), one would first need to bound $\mathbb{E}\left[\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]$ from below. Observe the following consequence from (112):

$$\begin{aligned} \mathbb{E}\left[e^{Z}\right] &\leq \mathbb{E}\left[e^{|Z|}\right] \\ &= -e^{z}\mathbb{P}\left(|Z|>z\right)\Big|_{z=0}^{\infty}+\int_0^{\infty}e^{z}\mathbb{P}\left(|Z|>z\right)\mathrm{d}z \qquad (117) \\ &\leq 1+\int_0^{\infty}2\exp\left(z-\frac{\delta z^2}{8\alpha}\right)\mathrm{d}z \\ &< 1+4\sqrt{\frac{2\pi\alpha}{\delta}}\exp\left(\frac{2\alpha}{\delta}\right). \qquad (118) \end{aligned}$$

Taking the logarithm on both sides of (118) and plugging in the expression of Z yields

$$\log \mathbb{E}\left[\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \leq \mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] + \log\left[1+4\sqrt{\frac{2\pi\alpha}{\delta}}\exp\left(\frac{2\alpha}{\delta}\right)\right],$$

leading to

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \\
&\geq \frac{1}{n}\log\mathbb{E}\left[\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] - \frac{\log\left[1+4\sqrt{\frac{2\pi\alpha}{\delta}}\exp\left(\frac{2\alpha}{\delta}\right)\right]}{n} \\
&\geq \frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] - \frac{\log\left[1+4\sqrt{\frac{2\pi\alpha}{\delta}}\exp\left(\frac{2\alpha}{\delta}\right)\right]}{n} \\
&\geq \frac{1}{n}\log\mathbb{E}\left[\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] - \frac{5\alpha}{n\delta}. \qquad (119)
\end{aligned}$$

for any $\delta \leq \alpha$. Here, the last inequality follows from simple numerical inequality $1+4\sqrt{2\pi x}\exp(2x) \leq \exp(5x)$ for any $x \geq 1$. Consequently, for any $\delta \leq \alpha$

$$\frac{1}{n}\mathbb{E}\left[\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \geq (1-\alpha)\log\frac{1}{1-\alpha} - \alpha - \frac{\log(n+1)}{n} - \frac{5\alpha}{n\delta}. \qquad (120)$$

This together with (113) characterizes a lower bound on $\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$.

It remains to quantify the gap between $\log\det^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$ and $\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)$. On the one hand, the inequality (99) indicates that for any $n \geq \left(\frac{\alpha}{1-\alpha-\frac{1}{n}} + 4\sqrt{\alpha\tau}\right)^3$,

$$\begin{aligned}
&\mathbb{P}\left\{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}\right\} \\
&= \mathbb{P}\left\{\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}\right\}}{n} > 1\right\} \\
&\leq \mathbb{P}\left\{\frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}\right\}}{n} > \frac{\frac{\alpha}{1-\alpha-\frac{1}{n}}+4\sqrt{\alpha\tau}}{n^{1/3}}\right\} \\
&\leq 2e^{-\tau n}. \qquad (121)
\end{aligned}$$

On the other hand, it follows from [43, Th. 4.5] that for any $n \geq \frac{6.414}{1-\alpha}\cdot e^{\frac{\tau}{1-\alpha}}$,

$$\begin{aligned}
&\mathbb{P}\left\{\frac{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)}{\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)} > n^2\right\} \\
&= \mathbb{P}\left\{\frac{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)}{\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)} > \frac{n^2}{(n-m+1)^2}\cdot(n-m+1)^2\right\} \\
&\leq \frac{1}{\sqrt{2\pi}}\left(\frac{6.414}{(1-\alpha)n}\right)^{(1-\alpha)n} \leq \frac{1}{\sqrt{2\pi}}e^{-\tau n}. \qquad (122)
\end{aligned}$$

The above two probability bounds taken collectively imply that for any $n \geq \max\left\{\frac{6.414}{1-\alpha}\cdot e^{\frac{\tau}{1-\alpha}}, \left(\frac{\alpha}{1-\alpha-\frac{1}{n}}+4\sqrt{\alpha\tau}\right)^3\right\}$,

$$\begin{aligned}
&\mathbb{P}\left\{\lambda_{\min}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{7/3}}\right\} \leq \mathbb{P}\left\{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}\right\} \\
&\quad + \mathbb{P}\left\{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right) \geq \frac{1}{n^{1/3}} \text{ and } \lambda_{\min}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{7/3}}\right\} \\
&\leq \mathbb{P}\left\{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}\right\} + \mathbb{P}\left\{\frac{\lambda_{\max}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)}{\lambda_{\min}\left(\boldsymbol{A}\boldsymbol{A}^{\top}\right)} > n^2\right\} \\
&\leq 2e^{-\tau n} + \frac{1}{\sqrt{2\pi}}e^{-\tau n} < 3e^{-\tau n}. \qquad (123)
\end{aligned}$$

Consequently, by setting $\delta = n^{-1/3}$ and applying Lemma 1 one obtains

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \geq \frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \\
&\quad - \sum_{i:\lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{1/3}}} \log\frac{1}{n^{1/3}} - \log\lambda_{\min}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \\
&\geq \frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \\
&\quad - \frac{\mathrm{card}\left\{i \mid \lambda_i\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) < \frac{1}{n^{\frac{1}{3}}}\right\}\log\left(n^2\right)}{n} \\
&> \frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \\
&\quad - \left(\frac{2\alpha}{1-\alpha-\frac{1}{n}} + 8\sqrt{\alpha\tau}\right)\cdot\frac{\log n}{n^{1/3}}
\end{aligned}$$

with probability exceeding $1 - 5\exp(-\tau n)$. By making use of (113), one obtains that when $\delta = n^{-1/3}$,

$$\mathbb{P}\left\{\left|\frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) - \mathbb{E}\left[\frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right]\right| > \frac{\sqrt{8\tau\alpha}}{n^{1/3}}\right\} < 2\exp(-\tau n). \qquad (124)$$

Putting the above two bounds together implies that for any $n > \max\left\{\frac{6.414}{1-\alpha}\cdot e^{\frac{\tau}{1-\alpha}}, \left(\frac{\alpha}{1-\alpha-\frac{1}{n}}+4\sqrt{\alpha\tau}\right)^3\right\}$,

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) \\
&\geq \frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right) - \left(\frac{2\alpha}{1-\alpha-\frac{1}{n}} + 8\sqrt{\alpha\tau}\right)\frac{\log n}{n^{1/3}} \\
&\geq \mathbb{E}\left[\frac{1}{n}\log\det{}^{\delta}\left(\frac{1}{n}\boldsymbol{A}\boldsymbol{A}^{\top}\right)\right] \\
&\quad - \frac{\sqrt{8\tau\alpha}}{n^{\frac{1}{3}}} - \left(\frac{2\alpha}{1-\alpha-\frac{1}{n}} + 8\sqrt{\alpha\tau}\right)\frac{\log n}{n^{\frac{1}{3}}} \\
&> (1-\alpha)\log\frac{1}{1-\alpha} - \alpha - \frac{\log(n+1)}{n} - \frac{5\alpha}{n^{2/3}} \\
&\quad - \left(\frac{2\alpha}{1-\alpha-\frac{1}{n}} + 10\sqrt{\alpha\tau}\right)\cdot\frac{\log n}{n^{1/3}} \\
&> (1-\alpha)\log\frac{1}{1-\alpha} - \alpha - \left(\frac{2}{1-\alpha-\frac{1}{n}} + 10\sqrt{\tau}\right)\cdot\frac{\log n}{n^{1/3}}, \qquad (125)
\end{aligned}$$

with probability exceeding $1-7\exp(-\tau n)$. Here, (125) follows from (120), and the last inequality makes use of the fact that $\frac{5}{n^{2/3}}+\frac{\log(n+1)}{n}\leq\frac{2\log n}{n^{1/3}}$ for all $n>6$.

APPENDIX E
PROOF OF LEMMA 3

Suppose that the singular value decomposition of the real-valued $\boldsymbol{A}$ is given by $\boldsymbol{A}=\boldsymbol{U_A}\begin{bmatrix}\boldsymbol{\Sigma_A}\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{V_A}^\top$, where $\boldsymbol{\Sigma_A}$ is a diagonal matrix containing all k singular values of $\boldsymbol{A}$. One can then write

$$\begin{aligned}
&\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{A}^\top\boldsymbol{B}^{-1}\boldsymbol{A}\right)\\
&\quad=\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{V_A}\left[\boldsymbol{\Sigma_A}\boldsymbol{0}\right]\boldsymbol{U_A}^\top\boldsymbol{B}^{-1}\boldsymbol{U_A}\begin{bmatrix}\boldsymbol{\Sigma_A}\\ \boldsymbol{0}\end{bmatrix}\boldsymbol{V_A}^\top\right)\\
&\quad=\log\det\left(\epsilon\boldsymbol{I}+\boldsymbol{\Sigma_A}\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}\boldsymbol{\Sigma_A}\right)\\
&\quad\geq\log\det\left(\frac{1}{n}\boldsymbol{\Sigma_A}^2\right)-\log\det\left\{\frac{1}{n}\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}\right\}
\end{aligned}\tag{126}$$

where $\tilde{\boldsymbol{B}}=\boldsymbol{U_A}^\top\boldsymbol{B}\boldsymbol{U_A}\sim\mathcal{W}_m\left(n-k,\boldsymbol{U_A}^\top\boldsymbol{U_A}\right)=\mathcal{W}_m\left(n-k,\boldsymbol{I}_m\right)$ from the property of Wishart distribution. Here, $\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}$ denotes the leading $k\times k$ minor consisting of matrix elements of $\tilde{\boldsymbol{B}}^{-1}$ in rows and columns from 1 to k, which is independent of $\boldsymbol{A}$ by Gaussianality.

Note that $\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{\Sigma_A}^2\right)=\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}^\top\boldsymbol{A}\right)$. Then Lemma 2 implies that for any $\tau>0$ and sufficiently large n,

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{\Sigma_A}^2\right)=\frac{1}{n}\log\det\left(\frac{1}{n}\boldsymbol{A}^\top\boldsymbol{A}\right)\\
&\quad=\frac{1}{n}\log\det\left(\frac{m}{n}\boldsymbol{I}_k\right)+\frac{m}{n}\frac{1}{m}\log\det\left(\frac{1}{m}\boldsymbol{A}^\top\boldsymbol{A}\right)\\
&\quad\geq\beta\log\alpha+\alpha\left(-\left(1-\frac{\beta}{\alpha}\right)\log\left(1-\frac{\beta}{\alpha}\right)-\frac{\beta}{\alpha}\right)-\hat{\kappa}\\
&\quad\geq-(\alpha-\beta)\log\left(\frac{\alpha-\beta}{\alpha}\right)-\beta+\beta\log\alpha-\hat{\kappa}\\
&\quad=-(\alpha-\beta)\log(\alpha-\beta)-\beta+\alpha\log\alpha-\hat{\kappa},
\end{aligned}\tag{127}$$

with probability exceeding $1-7\exp\left(-\tau^2 n\right)$, where

$$\hat{\kappa}:=\frac{\left(\frac{2\alpha}{1-\frac{\beta}{\alpha}-\frac{1}{n}}+10\alpha\tau\right)\log m}{m^{1/3}}\leq\frac{\left(\frac{2}{1-\frac{\beta}{\alpha}-\frac{1}{n}}+10\tau\right)\log n}{n^{1/3}}.\tag{128}$$

On the other hand, it is well known (e.g. [41, Th. 2.3.3]) that for a Wishart matrix $\tilde{\boldsymbol{B}}\sim\mathcal{W}_m\left(n-k,\boldsymbol{I}_m\right)$, $\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}^{-1}$ also follows the Wishart distribution, that is,

$$\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}^{-1}\sim\mathcal{W}_k\left(n-m,\boldsymbol{I}_k\right).\tag{129}$$

By setting $\zeta:=\max\left\{\frac{\beta}{\alpha},\frac{\beta}{1-\alpha}\right\}$,then one can obtain from Lemma 2 that for sufficiently large n,

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\frac{1}{n}\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}^{-1}\right)\\
&=\frac{1}{n}\log\det\left(\frac{1}{n-m}\left(\tilde{\boldsymbol{B}}^{-1}\right)_{[k]}^{-1}\right)+\frac{1}{n}\log\det\left(\frac{n-m}{n}\boldsymbol{I}_k\right)\\
&\leq(1-\alpha)\left\{-\left(1-\frac{\beta}{1-\alpha}\right)\log\left(1-\frac{\beta}{1-\alpha}\right)-\frac{\beta}{1-\alpha}\right\}\\
&\qquad+\beta\log(1-\alpha)+\frac{2\log n}{n}+\frac{\frac{5\sqrt{\zeta}}{\left(1-\sqrt{\zeta}-\frac{2}{n}\right)}\tau}{\sqrt{n}}\\
&=-(1-\alpha-\beta)\log\left(1-\frac{\beta}{1-\alpha}\right)-\beta+\beta\log(1-\alpha)\\
&\qquad+\frac{2\log n}{n}+\frac{\frac{5\sqrt{\zeta}}{\left(1-\sqrt{\zeta}-\frac{2}{n}\right)}\tau}{\sqrt{n}}
\end{aligned}\tag{130}$$

holds with probability exceeding $1-2\exp\left(-2\tau^2 n\right)$.

Combining (126), (127) and (126) suggests that for any $\tau>0$ and sufficiently large n, one has

$$\begin{aligned}
&\frac{1}{n}\log\det\left(\epsilon\boldsymbol{I}_k+\boldsymbol{A}^\top\boldsymbol{B}^{-1}\boldsymbol{A}\right)\\
&\quad\geq-(\alpha-\beta)\log(\alpha-\beta)+\alpha\log\alpha-\beta\log(1-\alpha)\\
&\qquad+(1-\alpha-\beta)\log\frac{1-\alpha-\beta}{1-\alpha}\\
&\qquad-\frac{\left(\frac{2}{1-\zeta-\frac{1}{n}}+\frac{10-5\sqrt{\zeta}}{\left(1-\sqrt{\zeta}-\frac{2}{n}\right)}\tau\right)\log n}{n^{1/3}}
\end{aligned}$$

with probability exceeding $1-9\exp\left(-\tau^2 n\right)$.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Young-Han Kim and the anonymous reviewer for extensive and constructive comments that greatly improved the presentation and positioning of the paper. Y. Chen would like to thank Dr. Ping Xia for discussion of LTE.

REFERENCES

[1] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, “Channel capacity under sub-Nyquist nonuniform sampling,” *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4739–4756, Aug. 2014.
[2] P. L. Butzer and R. L. Stens, “Sampling theory for not necessarily band-limited functions: A historical overview,” *SIAM Rev.*, vol. 34, no. 1, pp. 40–53, 1992.
[3] M. Mishali and Y. C. Eldar, “Sub-Nyquist sampling,” *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 98–124, Nov. 2011.
[4] H. J. Landau, “Necessary density conditions for sampling and interpolation of certain entire functions,” *Acta Math.*, vol. 117, no. 1, pp. 37–52, 1967.
[5] C. Herley and P. W. Wong, “Minimum rate sampling and reconstruction of signals with arbitrary frequency support,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1555–1564, Jul. 1999.
[6] R. Venkataramani and Y. Bresler, “Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals,” *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2301–2313, Oct. 2001.
[7] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[8] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, vol. 95. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[10] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.
[11] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. 24, no. 11, pp. 652–654, Nov. 1977.
[12] Y.-P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 3, pp. 340–351, Mar. 1998.
[13] W. A. Gardner, "The spectral correlation theory of cyclostationary time-series," *Signal Process.*, vol. 11, no. 1, pp. 13–36, 1986.
[14] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
[15] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
[16] S. Shamai (Shitz), "Information rates by oversampling the sign of a bandlimited process," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1230–1236, Jul. 1994.
[17] T. Koch and A. Lapidoth. (Sep. 2010). "Increased capacity per unit-cost by oversampling." [Online]. Available: https://arxiv.org/abs/1008.5393
[18] M. Peleg and S. Shamai, "On sparse sensing of coded signals at Sub-Landau sampling rates," in *Proc. IEEE 27th Conv. Elect. Electron. Eng. Israel*, Nov. 2012, pp. 1–5.
[19] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: Capacity of sampled Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4889–4914, Aug. 2013.
[20] A. Kipnis, A. J. Goldsmith, T. Weissman, and Y. C. Eldar. (May 2014). "Distortion-rate function of sub-Nyquist sampled Gaussian sources." [Online]. Available: https://arxiv.org/abs/1405.5329
[21] V. P. Boda and P. Narayan, "Sampling rate distortion," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2014, pp. 3057–3061.
[22] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, "Backing off from infinity: Performance bounds via concentration of spectral measure for random MIMO channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 366–387, Jan. 2015.
[23] A. Guionnet and O. Zeitouni, "Concentration of the spectral measure-forlargematrices," *Electron. Commun. Probab.*, vol. 5, pp. 119–136, Jun. 2000.
[24] A. Ghosh, J. Zhang, R. Muhamed, and J. G. Andrews, *Fundamentals of LTE*. Upper Saddle River, New Jersey, USA: Pearson Edu., 2010.
[25] *LTE in a Nutshell: White Paper on Physical Layer*, Telesystem Innovations, 2010.
[26] Y. Chen, "Subsampling in information theory and data processing," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, Jan. 2015. [Online]. Available: http://web.stanford.edu/~yxchen/publications/thesis_YChen.pdf
[27] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolution*. Hoboken, NJ, USA: Wiley, 2009.
[28] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
[29] A. Lozano and A. M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.
[30] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.
[31] A. Tulino and S. Verdu, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*. Hanover, MA, USA: Now Publishers Inc., 2004.
[32] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.
[33] P. Kazakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, "Living at the edge: A large deviations approach to the outage MIMO capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1984–2007, Apr. 2011.
[34] T. Tao and V. Vu. (Jun. 2012). "Random matrices: Universality of local spectral statistics of non-Hermitian matrices." [Online]. Available: https://arxiv.org/abs/1206.1893
[35] T. Tao and V. Vu, "Random covariance matrices: Universality of local statistics of eigenvalues," *Ann. Probab.*, vol. 40, no. 3, pp. 1285–1315, 2012.
[36] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[37] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[38] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[39] T. Tao, "Topics in random matrix theory," in *Graduate Studies in Mathematics*. Providence, RI, USA: AMS, 2012.
[40] A. Edelman, "The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law," *J. Multivariate Anal.*, vol. 60, no. 2, pp. 203–232, Feb. 1997.
[41] Y. Fujikoshi, V. V. Ulyanov, and R. Shimizu, *Multivariate Statistics: High-Dimensional Large-Sample Approximations* (Probability and Statistics). Hoboken, NJ, USA: Wiley, 2010.
[42] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Compressed Sensing, Theory and Applications*. New York, NY, USA: Cambridge Univ. Press, 2012, pp. 210–268.
[43] Z. Chen and J. J. Dongarra, "Condition numbers of Gaussian random matrices," *SIAM J. Matrix Anal. Appl.*, vol. 27, no. 3, pp. 603–620, Jul. 2005.

Yuxin Chen (S'09) received the B.S. in Microelectronics with High Distinction from Tsinghua University in 2008, the M.S. in Electrical and Computer Engineering from the University of Texas at Austin in 2010, the M.S. in Statistics from Stanford University in 2013, and the Ph.D. in Electrical Engineering from Stanford University in 2015. He joins the Department of Electrical Engineering at Princeton University as an assistant professor in February 2017. His research interests include mathematical optimization, information theory, high-dimensional statistics, and statistical learning.

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

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

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

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

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