Minimax Universal Sampling for Compound Multiband Channels

Yuxin Chen
EE, Stanford University
Email: yxchen@stanford.edu

Andrea J. Goldsmith
EE, Stanford University
Email: andrea@wsl.stanford.edu

Yonina C. Eldar
EE, Technion
Email: yonina@ee.technion.ac.il

***Abstract*—This paper considers the capacity of sub-sampled analog channels when the sampler is designed to operate independent of the instantaneous channel realization, and investigates sampling methods that minimize the worst-case (minimax) sampled capacity loss due to channel-independent (universal) sampling design. Specifically, a compound multiband channel with unknown subband occupancy is considered, when perfect channel side information is available to both the receiver and the transmitter. We restrict our attention to a general class of periodic sub-Nyquist samplers, which subsumes as special cases sampling with modulation and filter banks. Our results demonstrate that under both Landau-rate and super-Landau-rate sampling, the minimax sampled capacity loss due to universal design depends only on the band sparsity ratio and the undersampling factor, modulo a residual term that vanishes at high signal-to-noise ratio. We quantify the capacity loss under sampling with periodic modulation and low-pass filters, when the Fourier coefficients of the modulation waveforms are randomly generated (called random sampling). Our results highlight the power of random sampling methods, which achieve minimax sampled capacity loss uniformly across all channel realizations and are thus optimal in a universal design sense.**

***Index Terms*—Sub-Nyquist sampling, minimaxity, universal sampling, channel capacity, non-asymptotic random matrices**

I. Introduction

The maximum rate of information that can be conveyed through an analog channel largely depends on the sampling technique and rate employed at the receiver end. In wideband communication systems, hardware and cost limitations often preclude sampling at or above the Nyquist rate, which becomes a major bottleneck in designing energy-efficient receiver paradigms. Understanding the effects upon capacity of sub-Nyquist sampling is crucial in circumventing this bottleneck.

In practice, the receiver hardware and, in particular, its sampling mechanism is typically static and hence designed based on a family of possible channel realizations. During operation, the actual channel realization will vary over this class of channels, and the sampler thus operates independently of instantaneous channel side information (CSI). This has no effect if the sampling rate is equal to or above the Nyquist rate of the channel family. However, in the sub-Nyquist sampling rate regime, the sampler design significantly impacts the information rate achievable over different channel realizations. As was shown in [1], the capacity-maximizing sub-Nyquist sampling mechanism for a given linear time-invariant (LTI) channel depends on specific channel realizations. In time-varying channels, *sampled capacity loss* relative to the Nyquist-rate capacity is necessarily incurred due to channel-independent (universal) sub-Nyquist sampler design. Moreover, it turns out that the capacity-optimizing sampler for a given channel structure might result in very low data rates for other channel realizations.

In this paper, we explore universal design of a sub-Nyquist sampling system that is robust against the uncertainty and variation of instantaneous channel realizations, based on sampled capacity loss as a metric. In particular, we investigate the fundamental limit of sampled capacity loss in some overall sense (as detailed in Section II-C), and design a sub-Nyquist sampling system for which the capacity loss can be uniformly controlled and optimized over all possible channel realizations.

A. Related Work and Motivation

In various situations, sampling at the Nyquist rate is not necessary for preserving signal information if certain signal structures are appropriately exploited [2]. For example, consider multiband signals 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 (e.g. [3]), termed the *Landau rate*. Inspired by compressed sensing, spectrum-blind sub-Nyquist samplers have also been developed, e.g. [4]. These 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 used to investigate analog channel capacity (e.g. [5]), under the premise that sampling, when it is performed at the Nyquist rate, preserves information. The effects upon capacity of oversampling with quantization have also been studied [6]. Our recent work [1], [7] established a new framework for investigating the capacity of LTI channels under sub-Nyquist sampling, including filter-bank and modulation-bank sampling and, more generally, time-preserving sampling. We showed that periodic sampling or, more simply, sampling with a filter bank, is sufficient to maximize capacity among all sampling structures under a given sampling rate constraint, assuming perfect CSI at both the receiver and the transmitter.

Practical communication systems, however, involve time-varying channels, e.g. wireless fading channels. Time-varying channels are often modeled as a channel with state [8], where the channel variation is captured by a state, or, more simply, a compound channel [9] where the channel realization lies within a set of possible channels. One class of compound channel models concerns multiband channels, where the instantaneous active frequency support resides within several continuous intervals, spread over a wide spectrum. This model naturally arises in several wideband communication systems, including time division multiple access systems and cognitive radio networks. However, to the best of our knowledge, no prior work has investigated, from a capacity perspective, a universal sub-Nyquist sampling paradigm that is robust to channel variations in the above channel models.

B. Contribution

We consider a compound multiband channel, whereby the channel bandwidth W is divided into n continuous subbands and only k subbands are active for transmission. We consider the class of periodic sampling (i.e. a system that consists of a periodic preprocessor and a recurrent sampling set, defined in Section II-B) with period n/W and sampling rate $f_{\mathrm{s}} = mW/n$. The system is termed Landau-rate sampling (super-Landau-rate sampling) if f_{s} is equal to (greater than) the spectral size of the instantaneous channel support.

We first derive, in Theorem 3, a fundamental lower limit on the largest sampled capacity loss incurred by any channel-independent sampler, under both Landau-rate sampling and super-Landau-rate sampling. Theorem 4 studies the capacity loss under a class of channel-independent sampling with periodic modulation (of period n/W) and low-pass filters, when the Fourier coefficients of the modulation waveforms are independently randomly generated (termed independent random sampling). We demonstrate that with exponentially high probability, the sampled capacity loss matches the lower bound of Theorem 3 uniformly over all realizations. This implies that independent random sampling achieves minimum worst-case (or minimax) capacity loss among all periodic sampling methods with period n/W. Furthermore, for a large class of super-Landau-rate samplers, we determine the capacity loss under independent random sampling when the Fourier coefficients of the modulation waveforms are i.i.d. Gaussian-distributed. With high probability, this sampling method achieves minimax capacity loss among all periodic sampling with period n/W.

II. Problem Formulation and Preliminaries

A. Compound Multiband Channel

We consider multiband channels with n subbands[1] and total bandwidth W as follows. Denote by $\binom{[n]}{k}$ the *set* of all k-combinations of $[n] := \{1, \cdots, n\}$. A state $\boldsymbol{s} \in \binom{[n]}{k}$ is generated, which dictates the channel support and realization. Specifically, given a state $\boldsymbol{s}$, the channel is an LTI filter with impulse response $h_{\boldsymbol{s}}(t)$ and frequency response $H_{\boldsymbol{s}}(f)$. It is assumed that there exists a function $H(f, \boldsymbol{s})$ such that for every f and $\boldsymbol{s}$, $H_{\boldsymbol{s}}(f)$ can be expressed as $H_{\boldsymbol{s}}(f) = H(f, \boldsymbol{s})\mathbf{1}_{\boldsymbol{s}}(f)$, where $\mathbf{1}_{\boldsymbol{s}}(f)$ denotes the indicator function that equals 1 when f lies in the set of active subbands given $\boldsymbol{s}$.

A transmit signal $x(t)$ with a power constraint P is passed through this multiband channel, which yields a channel output $r_{\boldsymbol{s}}(t) = h_{\boldsymbol{s}}(t) * x(t) + \eta(t)$, where $\eta(t)$ is stationary zero-mean Gaussian noise with power spectral density (PSD) $\mathcal{S}_\eta(f)$. Assume that both the transmitter and the receiver have perfect CSI (including $\boldsymbol{s}$, $H_{\boldsymbol{s}}(f)$, $\mathcal{S}_\eta(f)$).

The above model subsumes as special cases the following communication scenarios.

- **Time Division Multiple Access Model.** The channel is shared by a set of different users. At each time-frame, the receiver (e.g. the base station) allocates a subset of subbands to one designated sender over that timeframe.
- **White-Space Cognitive Radio Network.** In this setting, cognitive users exploit spectrum holes unoccupied by primary users for transmission. The spectrum hole locations available to cognitive users may change over time.

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

B. Sampled Channel Capacity

We aim to design a sampler that works at rates below the Nyquist rate (i.e. the channel bandwidth W). In particular, we consider periodic sampling as defined in [1], which is the most widely used sampling method in practice.

Definition 1 (Periodic Sampling). Consider a sampling system consisting of a preprocessor with an impulse response $q(t, \tau)$ followed by a sampling set $\Lambda = \{t_k \mid k \in \mathbb{Z}\}$. A linear sampling system is said to be periodic with period T_q and sampling rate f_s ($f_s T_q \in \mathbb{Z}$) if $\forall t, \tau \in \mathbb{R}$ and $\forall k \in \mathbb{Z}$,

$$q(t, \tau) = q(t + T_q, \tau + T_q), \qquad t_{k+f_s T_q} = t_k + T_q. \quad (1)$$

Consider a periodic sampling system $\mathcal{P}$ with period $T_q = n/W$ and sampling rate $f_{\mathrm{s}} := mW/n$ for some integer m. A special case consists of sampling with a combination of filter banks and periodic modulation with period n/W, as shown in [7, Fig. 4]. The channel capacity in a sub-sampled LTI channel under periodic sampling has been derived in [1, Theorem 5]. Specifically, denote by $\boldsymbol{s}_i$ and f_i the ith smallest element in $\boldsymbol{s}$ and the lowest frequency of the ith subband, respectively, and define $\boldsymbol{H}_{\boldsymbol{s}}(f)$ as a $k \times k$ diagonal matrix obeying

$$(\boldsymbol{H}_{\boldsymbol{s}}(f))_{ii} = \frac{|H(f_{\boldsymbol{s}_i} + f, \boldsymbol{s})|}{\sqrt{\mathcal{S}_\eta(f_{\boldsymbol{s}_i} + f)}}.$$

Then the sampled channel capacity, when specialized to our setting, is given as follows. See [1] for more details.

Theorem 1 ([1]). *Consider a channel with total bandwidth W. Perfect CSI is known at both the transmitter and the receiver, and equal power allocation is employed over the active subbands. If a periodic sampler $\mathcal{P}$ with period n/W and sampling rate $f_s = \frac{m}{n}W$ is employed, then the sampled channel capacity at a given state $\boldsymbol{s}$ obeys*

$$C_{\boldsymbol{s}}^{\boldsymbol{Q}} = \int_0^{\frac{W}{n}} \frac{1}{2} \log\det \left(\boldsymbol{I}_m + \frac{P}{\beta W} \left(\boldsymbol{Q}(f)\boldsymbol{Q}^*(f)\right)^{-\frac{1}{2}} \boldsymbol{Q}_{\boldsymbol{s}}(f) \right.$$
$$\left. \cdot \boldsymbol{H}_{\boldsymbol{s}}^2(f) \boldsymbol{Q}_{\boldsymbol{s}}^*(f) \left(\boldsymbol{Q}(f)\boldsymbol{Q}^*(f)\right)^{-\frac{1}{2}} \right) \mathrm{d}f, \quad (2)$$

Here, for any given f, $\boldsymbol{Q}(f)$ is an $m \times n$ matrix that only depends on $\mathcal{P}$, and $\boldsymbol{Q}_{\boldsymbol{s}}(f)$ denotes the submatrix consisting of the columns of $\boldsymbol{Q}(f)$ at indices of $\boldsymbol{s}$.

In general, $\boldsymbol{Q}(f)$ is a function that varies with f. Unless otherwise specified, we call $\boldsymbol{Q}(\cdot)$ the *sampling coefficient function* with respect to the sampling system $\mathcal{P}$.

A special class of sampling systems concerns the ones whose $\boldsymbol{Q}(\cdot)$ are flat over $[0, f_{\mathrm{s}}/m]$, in which case we can use an $m \times n$ matrix to represent $\boldsymbol{Q}(\cdot)$, termed a *sampling coefficient matrix*. This class of sampling systems can be realized through an m-branch system shown in Fig. 1. At the ith branch, the channel output is modulated by a periodic waveform $q_i(t)$ of period n/W, passed through a low-pass filter with pass band $[0, f_{\mathrm{s}}/m]$, and then uniformly sampled at rate f_{s}/m, where the Fourier transform of $q_i(t)$ obeys $\mathcal{F}(q_i(t)) = \sum_{l=1}^n \boldsymbol{Q}_{i,l}\delta\left(f - l\frac{W}{n}\right)$. A sampling system within this class is said to be *(independent) random sampling* if the entries of $\boldsymbol{Q}$ are (independently) randomly generated, and is termed *Gaussian sampling* if the entries of $\boldsymbol{Q}$ are i.i.d. Gaussian-distributed.

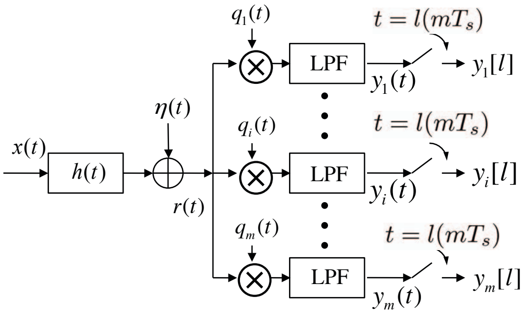


Figure 1. Sampling with a modulation bank and low-pass filters. The channel output $r(t)$ is passed through m branches, each consisting of a modulator with modulation waveform $q_i(t)$ and a low-pass filter of pass band $[0, f_s/m]$ followed by a uniform sampler with sampling rate f_s/m.

It turns out that this simple sampling structure is sufficient to achieve overall robustness among all periodic sampling systems of period $T_q = n/W$, as detailed in Section III.

C. Universal Sampling

As shown in [1], the optimal sampler for a given LTI channel extracts out a frequency set with the highest SNR and hence suppresses aliasing. Such an alias-suppressing sampler may achieve very low capacity for some channel realizations. In this paper, we desire a sampler that operates independent of the instantaneous CSI, and our objective is to design a single linear sampler that achieves to within a minimal gap of the Nyquist-rate capacity across all possible channel realizations.

1) Sampled Capacity Loss: Universal sub-Nyquist samplers incur capacity loss relative to channel-optimized samplers. For any $\boldsymbol{s}$, the Nyquist-rate capacity under equal power allocation and optimal power allocation can be respectively written as

$$C_{\boldsymbol{s}}^{P_{\mathrm{eq}}} = \int_0^{W/n} \frac{1}{2} \log\det\left(\boldsymbol{I}_k + \frac{P}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2(f)\right) \mathrm{d}f, \tag{3}$$

$$C_{\boldsymbol{s}}^{\mathrm{opt}} = \int_0^{W/n} \frac{1}{2} \log\det\left(\frac{\nu}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2(f)\right) \mathrm{d}f, \tag{4}$$

where ν is determined by a water-filling power allocation strategy. For any sampling coefficient function $\boldsymbol{Q}(\cdot)$ and any given state $\boldsymbol{s}$, we can then define the *sampled capacity loss* as

$$L_{\boldsymbol{s}}^{\boldsymbol{Q}} := C_{\boldsymbol{s}} - C_{\boldsymbol{s}}^{\boldsymbol{Q}}, \quad L_{\boldsymbol{s}}^{\boldsymbol{Q},\mathrm{opt}} := C_{\boldsymbol{s}}^{\mathrm{opt}} - C_{\boldsymbol{s}}^{\boldsymbol{Q}}.$$

These metrics quantify the capacity gaps relative to Nyquist-rate capacity due to universal sampling. When sampling is performed at or above the Landau rate but below the Nyquist rate, these gaps capture the rate loss due to channel-independent sampling relative to channel-optimized sampling.

For an $m \times n$ matrix $\boldsymbol{M}$, we denote by $L_{\boldsymbol{s}}^{\boldsymbol{M}}$ the sampled capacity loss with respect to a sampling coefficient function $\boldsymbol{Q}(f) \equiv \boldsymbol{M}$, which is flat across $[0, W/n]$.

2) Minimax sampler: Frequently used in the theory of statistics, minimaxity is a metric that seeks to minimize the loss function in some overall sense, defined as follows.

Definition 2. A sampling system associated with a sampling coefficient function $\boldsymbol{Q}^{\mathrm{m}}$ which minimizes the worst-case sampled capacity loss, that is, which satisfies

$$\max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q}^{\mathrm{m}}} = \inf_{\boldsymbol{Q}(\cdot)} \max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q}},$$

is called a minimax sampler with respect to $\binom{[n]}{k}$.

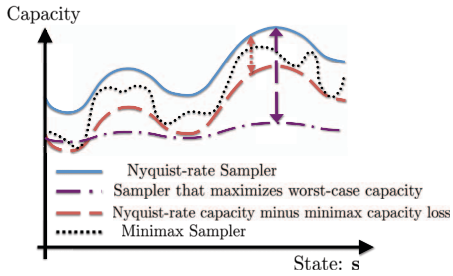


Figure 2. Minimax sub-Nyquist sampler v.s. the sub-Nyquist sampler that minimizes worst-case capacity, where the sampling is channel-independent.

The minimax criteria is of interest for designing a sampler robust to all possible channel states. It aims to control the rate loss across all states in a uniform manner, as illustrated in Fig. 2. 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 rate that can be achieved regardless of which channel is realized, it may result in significant sampled capacity loss in many states with large Nyquist-rate capacity, as illustrated in Fig. 2. In contrast, a desired minimax sampler controls the capacity loss for every single state $\boldsymbol{s}$, and allows for robustness over all channel states with universal sampling.

III. Main Results

In general, the minimax capacity loss problem is non-convex, and hence it is difficult to identify the optimal sampler. It turns out, however, that the minimax capacity loss can be approached by random sampling at moderate-to-high SNR.

Define the undersampling factor $\alpha := \frac{m}{n}$ and the sparsity ratio $\beta := \frac{k}{n}$. Our main results are summarized below.

Theorem 2. *Consider any sampling coefficient function $\boldsymbol{Q}(\cdot)$ with an undersampling factor α, and let the sparsity ratio be β. Define $\mathrm{SNR}_{\min} := \frac{P}{\beta W} \inf_{0\leq f\leq W, \boldsymbol{s}\in\binom{[n]}{k}} \frac{|H(f,\boldsymbol{s})|^2}{\mathcal{S}_\eta(f)}$, and let $\mathcal{H}(x) := -x\log x - (1-x)\log(1-x)$.*

(i) (Landau-rate sampling) If $\alpha = \beta$, then

$$\inf_{\boldsymbol{Q}} \max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q}} = \frac{W}{2}\left\{\mathcal{H}(\beta) + O\left(\frac{\log n}{n}\right) + \Delta\right\}, \tag{5}$$

(ii) (Super-Landau-rate sampling) Suppose that $\alpha > \beta$ and $1 - \alpha - \beta > 0$. Then

$$\inf_{\boldsymbol{Q}} \max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q}} = \frac{W\left\{\mathcal{H}(\beta) - \alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right) + O\left(\frac{\log^2 n}{\sqrt{n}}\right) + \tilde{\Delta}\right\}}{2}. \tag{6}$$

Here, the residual terms obey $\Delta, \tilde{\Delta} \in \left[-\frac{2}{\sqrt{\mathrm{SNR}_{\min}}}, \frac{\beta}{\mathrm{SNR}_{\min}}\right]$.

In addition, in both regimes, one has

$$0 \leq \inf_{\boldsymbol{Q}} \max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q},\mathrm{opt}} \leq \frac{\overline{A}}{\mathrm{SNR}_{\min}}$$

for some constant $\overline{A}$ determined by $\frac{H(f)}{\sqrt{\mathcal{S}_\eta(f)}}$ and β (see [10] for a detailed expression).

Theorem 2 characterizes the minimax sampled capacity loss under both Landau-rate and super-Landau-rate sampling. Note that the Landau-rate sampling regime in (i) is not a special case of the super-Landau-rate regime considered in (ii), since these results are given under the constraint $\beta+\alpha<1$.

The expressions (5) and (6) contain residual terms no larger than $O\left(\frac{\log^2 n}{\sqrt{n}}\right)+\frac{2}{\sqrt{\mathrm{SNR}_{\min}}}$, which vanishes for large n and high SNR. These fundamental minimax limits do not scale with the SNR and n except for a vanishing residual term.

Theorem 2 involves the verification of two parts: a converse part that provides a fundamental lower bound on the minimax sampled capacity loss, and an achievability part that provides a sampling scheme to achieve this bound. In fact, the sampling methods illustrated in Fig. 1 are sufficient to approach the minimax sampled capacity loss. Unless otherwise specified, we assume in our analysis that the noise has unit PSD[2] $\mathcal{S}_\eta(f)\equiv 1$.

A. The Converse

We need to show that the minimax capacity loss under any channel-independent sampler cannot be lower than (5) and (6). This is given by the following theorem, which takes into account the entire regime including the case in which $\alpha+\beta>1$.

Theorem 3. *Consider any Riemann-integrable sampling coefficient function $\boldsymbol{Q}(\cdot)$ with an undersampling factor $\alpha\geq\beta$. The minimax sampled capacity loss can be lower bounded by*

$$\inf_{\boldsymbol{Q}}\max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{Q}}\geq\frac{W\left\{\mathcal{H}\left(\beta\right)-\alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right)-\frac{2}{\sqrt{\mathrm{SNR}_{\min}}}-\frac{\log n}{n}\right\}}{2}. \tag{7}$$

For a given β, the bound is decreasing in α. While the active channel bandwidth is smaller than the total bandwidth, the noise (even though the SNR is large) is scattered over the entire bandwidth. Thus, none of the universal sub-Nyquist sampling strategies are information preserving, and increasing the sampling rate can always harvest capacity gain.

B. Achievability with Landau-rate Sampling

Consider the achievability under Landau-rate sampling ($\beta=\alpha$). A class of sampling methods that we can analyze concerns sampling with periodic modulation followed by low-pass filters or, more simply, independent random sampling. As n grows large, its capacity loss approaches (7) uniformly across all realizations. The results are stated in Theorem 4 after introducing a class of sub-Gaussian measure below.

Definition 3. A measure ν satisfies the logarithmic Sobolev inequality[3] with constant c_{LS} if, for any differentiable g,

$$\int g^2\log\left(g^2/\int g^2\mathrm{d}\nu\right)\mathrm{d}\nu\leq 2c_{\mathrm{LS}}\int\left|g'\right|^2\mathrm{d}\nu.$$

Theorem 4. *Let $\boldsymbol{M}=\left(\zeta_{ij}\right)_{1\leq i\leq k,1\leq j\leq n}$ be a real-valued random matrix in which ζ_{ij}'s are jointly independent with zero mean and unit variance. In addition, ζ_{ij} satisfies one of the following conditions:*

(a) ζ_{ij} is almost surely bounded by a constant D;

(b) Its probability measure satisfies the logarithmic Sobolev inequality with a constant c_{LS}.

Then there exist some constants $c,C>0$ such that

$$\max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{M}}\leq\frac{W}{2}\left(\mathcal{H}\left(\beta\right)+\frac{5\log k}{n}+\frac{\beta}{\mathrm{SNR}_{\min}}\right) \tag{8}$$

with probability exceeding $1-C\exp\left(-cn\right)$.

Theorem 4 demonstrates that independent random sampling achieves minimax capacity loss among all periodic sampling methods with period n/W. A broad class of sub-Gaussian ensembles, as long as the entries are jointly independent with matching moments, suffices to generate minimax samplers.

C. Achievability with Super-Landau-Rate Sampling

Consider the super-Landau-rate regime where $\beta<\alpha<1$ and $\beta+\alpha<1$. The achievability result is stated as follows.

Theorem 5. *Let $\boldsymbol{M}=\left(\zeta_{ij}\right)_{1\leq i\leq m,1\leq j\leq n}$ be a random matrix in which ζ_{ij} are i.i.d. drawn from $\mathcal{N}\left(0,1\right)$. Suppose that $\beta=k/n$ and $\alpha=m/n$ are constants satisfying $\alpha+\beta<1$ and $\beta<\alpha$. Then there exist some constants $c,C>0$ such that*

$$\max_{\boldsymbol{s}\in\binom{[n]}{k}} L_{\boldsymbol{s}}^{\boldsymbol{M}}\leq\frac{W\left[\mathcal{H}(\beta)-\alpha\mathcal{H}\left(\frac{\beta}{\alpha}\right)+O\left(\frac{\log^2 n}{\sqrt{n}}\right)+\frac{\beta}{\mathrm{SNR}_{\min}}\right]}{2}$$

with probability exceeding $1-C\exp\left(-cn\right)$.

Theorem 5 indicates that i.i.d. Gaussian sampling achieves the minimax capacity loss (6) with a vanishingly small gap. In contrast to Theorem 4, we restrict our attention to Gaussian ensembles, which suffices for the proof of Theorem 2.

D. Equivalent Algebraic Problems

Our results can be converted to three equivalent algebraic problems without difficulty, which we state in this subsection. The analysis of these algebraic problems are mathematically involved, and interested readers are referred to [10] for details.

Recall that $\frac{P}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2\succeq\mathrm{SNR}_{\min}\boldsymbol{I}_k$. Define $\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}:=\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-\frac{1}{2}}\boldsymbol{Q}_{\boldsymbol{s}}$. Simple manipulations (detailed in [10]) yield

$$L_{\boldsymbol{s}}^{\boldsymbol{Q}}=-\int_0^{\frac{W}{n}}\frac{\log\det\left(\frac{\beta W}{P}\boldsymbol{H}_{\boldsymbol{s}}^{-2}+\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\right)\mathrm{d}f}{2}+\frac{\beta W\Delta_{\boldsymbol{s}}}{2},$$

where the second equality uses the fact that

$$\begin{aligned}0&\leq\frac{1}{k}\log\det\left(\boldsymbol{I}_k+\frac{P}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2\right)-\frac{1}{k}\log\det\left(\frac{P}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2\right)\\&=\frac{1}{k}\sum_{i=1}^k\log\left(1+\frac{1}{\lambda_i\left(\frac{P}{\beta W}\boldsymbol{H}_{\boldsymbol{s}}^2\right)}\right)\leq\frac{1}{\mathrm{SNR}_{\min}}.\end{aligned} \tag{9}$$

In the high-SNR regime (where $\frac{\beta W}{P}\boldsymbol{H}_{\boldsymbol{s}}^{-2}\ll\boldsymbol{I}_k$), (**??**) makes $\det\left(\epsilon\boldsymbol{I}_k+\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}*}\boldsymbol{Q}_{\boldsymbol{s}}^{\mathrm{w}}\right)$ an object of interest (for some small ϵ).

Note that $\left(\boldsymbol{Q}\boldsymbol{Q}^*\right)^{-\frac{1}{2}}\boldsymbol{Q}$ has orthonormal rows. The following theorem studies $\det\left(\epsilon\boldsymbol{I}_k+\boldsymbol{B}_{\boldsymbol{s}}^*\boldsymbol{B}_{\boldsymbol{s}}\right)$ for any $\boldsymbol{B}$ that has orthonormal rows, which is the basis for proving Theorem 3.

[2]Note that this incurs no loss of generality since we can always include a noise-whitening LTI filter at the first stage of the sampling system.

[3]A probability measure obeying the logarithmic Sobolev inequality possesses sub-Gaussian tails. See [10] for references. In particular, the standard Gaussian measure satisfies this inequality with constant $c_{\mathrm{LS}}=1$.

Theorem 6. *(1) Consider any $m \times n$ matrix $\boldsymbol{B}$ obeying $\boldsymbol{B}\boldsymbol{B}^* = \boldsymbol{I}_m$, and denote by $\boldsymbol{B}_{\boldsymbol{s}}$ the $m \times k$ submatrix of $\boldsymbol{B}$ with columns coming from the index set $\boldsymbol{s}$. Then for any $\epsilon \geq 0$,*

$$\binom{m}{k} \leq \sum_{\boldsymbol{s} \in \binom{[n]}{k}} \det\left(\epsilon \boldsymbol{I}_m + \boldsymbol{B}_{\boldsymbol{s}}^* \boldsymbol{B}_{\boldsymbol{s}}\right) \leq \binom{m}{k} \left(1+\sqrt{\epsilon}\right)^{n+k}.$$

(2) For any positive integer p, suppose that $\boldsymbol{B}_1, \cdots, \boldsymbol{B}_p$ are all $m \times n$ matrices such that $\boldsymbol{B}_i \boldsymbol{B}_i^* = \boldsymbol{I}_m$. Then,

$$\begin{aligned} &\max_{\boldsymbol{B}_1,\cdots,\boldsymbol{B}_p} \min_{\boldsymbol{s} \in \binom{[n]}{k}} \frac{1}{np} \sum_{i=1}^{p} \log\det\left(\epsilon \boldsymbol{I}_k + \left(\boldsymbol{B}_i\right)_{\boldsymbol{s}}^* \left(\boldsymbol{B}_i\right)_{\boldsymbol{s}}\right) \\ &\leq \alpha \mathcal{H}\left(\frac{\beta}{\alpha}\right) - \mathcal{H}(\beta) + 2\sqrt{\epsilon} + \frac{\log(n+1)}{n}. \end{aligned} \tag{10}$$

When it comes to the achievability part, the major step is to quantify $\det(\epsilon \boldsymbol{I}_k + (\boldsymbol{M}\boldsymbol{M}^T)^{-1}\boldsymbol{M}_{\boldsymbol{s}}\boldsymbol{M}_{\boldsymbol{s}}^T)$ for every $\boldsymbol{s}$. Interestingly, this quantity can be uniformly bounded for random ensembles. This is stated in the following theorem, which demonstrates the achievability for Landau-rate sampling.

Theorem 7. *Let $\boldsymbol{M}$ be a $k \times n$ real-valued random matrix. Under the conditions of Theorem 4, one has*

$$\begin{aligned} &\min_{\boldsymbol{s} \in \binom{[n]}{k}} \frac{1}{n} \log\det\left(\epsilon \boldsymbol{I}_k + \left(\boldsymbol{M}\boldsymbol{M}^T\right)^{-1} \boldsymbol{M}_{\boldsymbol{s}}\boldsymbol{M}_{\boldsymbol{s}}^T\right) \\ &\geq -\mathcal{H}(\beta) - (5\log k)/n \end{aligned}$$

with probability exceeding $1 - C\exp(-cn)$.

Instead of studying a large class of sub-Gaussian random ensembles[4], the following theorem focuses on i.i.d. Gaussian matrices, which establishes the optimality of Gaussian random sampling for the super-Landau regime.

Theorem 8. *Let $\boldsymbol{M}$ be an $m \times n$ real-valued Gaussian matrix. Under the conditions of Theorem 5, one has*

$$\begin{aligned} &\min_{\boldsymbol{s} \in \binom{[n]}{k}} \frac{1}{n} \log\det\left(\epsilon \boldsymbol{I}_k + \boldsymbol{M}_{\boldsymbol{s}}^T \left(\boldsymbol{M}\boldsymbol{M}^T\right)^{-1} \boldsymbol{M}_{\boldsymbol{s}}\right) \\ &\geq -\mathcal{H}(\beta) + \alpha\mathcal{H}(\beta/\alpha) + O\left(\left(\log^2 n\right)/\sqrt{n}\right) \end{aligned}$$

with probability at least $1 - C\exp(-cn)$.

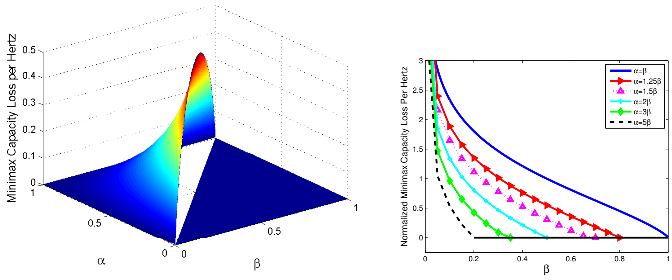
The proofs of Theorems 6-8 (which we omit here due to space limitations) rely heavily on *non-asymptotic* (random) matrix theory. Interested readers are referred to [10] for details.

IV. Implications and Discussion

Under both Landau-rate and super-Landau-rate sampling, the minimax sampled capacity loss depends almost entirely on β and α. Some implications are as follows.

1) The sampled capacity loss per unit Hertz is illustrated in Fig. 3(a). For instance, when $\alpha = \beta$, we suffer from the largest capacity loss when half of the bandwidth is active. The capacity loss vanishes when $\alpha \to 1$, since Nyquist-rate sampling is information preserving. The capacity loss divided by β is plotted in Fig. 2(b), which provides a normalized view of the capacity loss. It can be seen that the normalized loss decreases monotonically with β, indicating that the loss is more severe in sparse channels.
2) The sampled capacity loss incurred by independent (resp. Gaussian) random sampling meets the fundamental minimax limit for Landau-rate (resp. super-Landau-rate) sampling, which reveals that random sampling is optimal in terms of universal sampling design. The capacity achievable by random sampling exhibits very sharp concentration around the minimax limit uniformly across all states $\boldsymbol{s} \in \binom{[n]}{k}$.
3) For the Landau-rate regime, the optimal sampling matrix does not need to be i.i.d. generated due to the universality properties of random matrices. A large class of sub-Gaussian measure (and the mixture of them) suffices to generate minimax sampling systems, provided that all entries are jointly independent.

[4]The proof argument for Landau-rate sampling cannot be readily carried over to super-Landau regime since $\boldsymbol{M}_{\boldsymbol{s}}$ is now a tall matrix, and hence we cannot separate $\boldsymbol{M}_{\boldsymbol{s}}$ and $\boldsymbol{M}\boldsymbol{M}^*$ easily.



(a) (b)

Figure 3. (a) The minimax sampled capacity loss per unit Hertz v.s. the sparsity ratio β and the undersampling factor α, when $\beta \leq \alpha \leq 1$. (b) The minimax sampled capacity loss per Hertz divided by β v.s. the sparsity ratio β.

Acknowledgment

This work was supported by the NSF Center for Science of Information, IFC, and BSF Transformative Science Grant 2010505.

References

[1] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Channel capacity under sub-Nyquist nonuniform sampling," *under revision, IEEE Trans on Info Theory*, http://arxiv.org/abs/1204.6049.
[2] P. L. Butzer and R. L. Stens, "Sampling theory for not necessarily band-limited functions: A historical overview," *SIAM Review*, vol. 34, no. 1, pp. 40–53, 1992.
[3] C. Herley and P. W. Wong, "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1555 –1564, Jul. 1999.
[4] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375 –391, Apr. 2010.
[5] M. Medard, "The effect upon channel capacity in wireless comms of perfect and imperfect knowledge of the channel," *IEEE Trans. on Info Theory*, vol. 46, no. 3, pp. 933 –946, May. 2000.
[6] S. Shamai, "Information rates by oversampling the sign of a bandlimited process," *IEEE Transactions on Info Theory*, vol. 40, no. 4, pp. 1230 –1236, Jul 1994.
[7] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: capacity of sampled Gaussian channels," *accepted to IEEE Trans. on Information Theory*. [Online]. Available: http://arxiv.org/abs/1109.5415
[8] A. E. Gamal and Y. H. Kim, *Network Information theory*. Cambridge University Press, 2011.
[9] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Annals of Math. Stats.*, vol. 30, pp. 1229–1241, 1959.
[10] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, "Minimax capacity loss under sub-Nyquist universal sampling," *submitted to IEEE Trans on Information Theory*, April 2013. [Online]. Available: Arxiv:1304.7751