Shannon Meets Nyquist: Capacity of Sampled Gaussian Channels

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

Abstract—We explore two fundamental questions at the intersection of sampling theory and information theory: how channel capacity is affected by sampling below the channel's Nyquist rate, and what sub-Nyquist sampling strategy should be employed to maximize capacity. In particular, we derive the capacity of sampled analog channels for three prevalent sampling strategies: sampling with filtering, sampling with filter banks, and sampling with modulation and filter banks. These sampling mechanisms subsume most nonuniform sampling techniques applied in practice. Our analyses illuminate interesting connections between undersampled channels and multiple-input multiple-output channels. The optimal sampling structures are shown to extract out the frequencies with the highest SNR from each aliased frequency set, while suppressing aliasing and out-of-band noise. We also highlight connections between undersampled channel capacity and minimum mean-squared error (MSE) estimation from sampled data. In particular, we show that the filters maximizing capacity and the ones minimizing MSE are equivalent under both filtering and filter-bank sampling strategies. These results demonstrate the effect upon channel capacity of sub-Nyquist sampling techniques, and characterize the tradeoff between information rate and sampling rate.

Index Terms—Channel capacity, sampled analog channels, sampling rate, sub-Nyquist sampling.

I. INTRODUCTION

THE capacity of continuous-time Gaussian channels and the corresponding capacity-achieving water-filling power allocation strategy over frequency are well-known [1], and provide much insight and performance targets for practical communication system design. These results implicitly assume sampling above the Nyquist rate at the receiver end. However, chan-

nels that are not bandlimited have an infinite Nyquist rate and hence cannot be sampled at that rate. Moreover, hardware and power limitations often preclude sampling at the Nyquist rate associated with the channel bandwidth, especially for wideband communication systems. This gives rise to several natural questions at the intersection of sampling theory and information theory, which we will explore in this paper: 1) how much information, in the Shannon sense, can be conveyed through undersampled analog channels; and 2) under a sub-Nyquist sampling-rate constraint, which sampling structures should be chosen in order to maximize information rate.

A. Related Work

The derivation of the capacity of linear time-invariant (LTI) channels was pioneered by Shannon [2]. Making use of the asymptotic spectral properties of Toeplitz operators [3], this capacity result established the optimality of a water-filling power allocation based on signal-to-noise ratio (SNR) across the frequency domain [1]. Similar results for discrete-time Gaussian channels have also been derived using Fourier analysis [4]. On the other hand, the Shannon–Nyquist sampling theorem, which dictates that channel capacity is preserved when the received signal is sampled at or above the Nyquist rate, has frequently been used to transform analog channels into their discrete counterparts (see, e.g., [5] and [6]). For instance, this paradigm of discretization was employed by Medard to bound the maximum mutual information in time-varying channels [7]. However, all of these works focus on analog channel capacity sampled at or above the Nyquist rate, and do not account for the effect upon capacity of reduced-rate sampling.

The Nyquist rate is the sampling rate required for perfect reconstruction of bandlimited analog signals or, more generally, the class of signals lying in shift-invariant subspaces. Various sampling methods at this sampling rate for bandlimited functions have been proposed. One example is recurrent nonuniform sampling proposed by Yen [8], which samples the signal in such a way that all sample points are divided into blocks where each block contains N points and has a recurrent period. Another example is generalized multibranch sampling first analyzed by Papoulis [9], in which the input is sampled through M linear systems. For perfect recovery, these methods require sampling at an aggregate rate above the Nyquist rate.

In practice, however, the Nyquist rate may be excessive for perfect reconstruction of signals that possess certain structures. For example, consider multiband signals, whose spectral content resides continuously within several subbands over a wide spectrum, as might occur in a cognitive radio system. If the spectral support is known *a priori*, then the sampling rate require-

ment for perfect recovery is the sum of the subband bandwidths, termed the *Landau rate* [10]. One type of sampling mechanism that can reconstruct multiband signals sampled at the Landau rate is a filter bank followed by sampling, studied in [11] and [12]. The basic sampling paradigm of these works is to apply a bank of prefilters to the input, each followed by a uniform sampler.

When the channel or signal structure is unknown, sub-Nyquist sampling approaches have been recently developed to exploit the structure of various classes of input signals, such as multiband signals [13]. In particular, sampling with modulation and filter banks is very effective for signal reconstruction, where the key step is to scramble spectral contents from different subbands through the modulation operation. Examples include the modulated wideband converter proposed by Mishali *et al.* [13], [14]. In fact, modulation and filter-bank sampling represents a very general class of realizable nonuniform sampling techniques applied in practice.

Most of the above sampling theoretic work aims at finding optimal sampling methods that admit perfect recovery of a class of analog signals from *noiseless* samples. There has also been work on minimum reconstruction error from *noisy* samples based on certain statistical measures (e.g., mean squared error (MSE)). Another line of work pioneered by Berger *et al.* [15]–[18] investigated joint optimization of the transmitted pulse shape and receiver prefiltering in pulse amplitude modulation over a sub-sampled analog channel. In this paper, the optimal receiver prefilter that minimizes the MSE between the original signal and the reconstructed signal is shown to prevent aliasing. However, this study does not consider optimal sampling techniques based on capacity as a metric. The optimal filters derived in [15] and [16] are used to determine an SNR metric which, in turn, is used to approximate sampled channel capacity based on the formula for capacity of bandlimited AWGN channels. However, this approximation does not correspond to the precise channel capacity we derive herein, nor is the capacity of more general undersampled analog channels considered.

The tradeoff between capacity and hardware complexity has been studied in another line of work on sampling precision [19], [20]. These works demonstrate that due to quantization, oversampling can be beneficial in increasing achievable data rates. The focus of these works is on the effect of oversampling upon capacity loss due to quantization error, rather than the effect of quantization-free subsampling upon channel capacity.

B. Contribution

In this paper, we explore sampled Gaussian channels with the following three classes of sampling mechanisms: 1) a filter followed by sampling: the analog channel output is prefiltered by an LTI filter followed by an ideal uniform sampler (see Fig. 2); 2) filter banks followed by sampling: the analog channel output is passed through a bank of LTI filters, each followed by an ideal uniform sampler (see Fig. 3); and 3) modulation and filter banks followed by sampling: the channel output is passed through M branches, where each branch is prefiltered by an LTI filter, modulated by different modulation sequences, passed through another LTI filter, and then sampled uniformly. Our main contributions are summarized as follows.

1) *Filtering followed by sampling:* We derive the sampled channel capacity in the presence of both white and colored noise. Due to aliasing, the sampled channel can be represented as a multiple-input single-output (MISO) Gaussian channel in the spectral domain, while the optimal input effectively performs maximum ratio combining (MRC). The optimal prefilter is derived and shown to extract out the frequency with the highest SNR while suppressing signals from all other frequencies and hence preventing aliasing. This prefilter also minimizes the MSE between the original signal and the reconstructed signal, illuminating a connection between capacity and MMSE estimation.

2) *Filter banks followed by sampling:* A closed-form expression for sampled channel capacity is derived, along with analysis that relates it to a multiple-input multiple-output (MIMO) Gaussian channel. We also derive optimal filter banks that maximize capacity. The M filters select the M frequencies with highest SNRs and zero out signals from all other frequencies. This alias-suppressing strategy is also shown to minimize the MSE between the original and reconstructed signals. This mechanism often achieves larger sampled channel capacity than a single filter followed by sampling if the channel is nonmonotonic, and it achieves the analog capacity of multiband channels at the Landau rate if the number of branches is appropriately chosen.

3) *Modulation and filter banks followed by sampling:* For modulation sequences that are periodic with period T_q, we derive the sampled channel capacity and show its connection to a general MIMO Gaussian channel in the frequency domain. For sampling following a single branch of modulation and filtering, we provide an algorithm to identify the optimal modulation sequence for piecewise flat channels when T_q is an integer multiple of the sampling period. We also show that the optimal single-branch mechanism is equivalent to an optimal filter bank with each branch sampled at a period T_q.

One interesting fact we discover for all these techniques is the nonmonotonicity of capacity with sampling rate, which indicates that at certain sampling rates, channel degrees of freedom are lost. Thus, more sophisticated sampling techniques are needed to maximize achievable data rates at sub-Nyquist sampling rates in order to preserve all channel degrees of freedom.

C. Organization

The remainder of this paper is organized as follows. In Section II, we describe the problem formulation of sampled analog channels. The capacity results for three classes of sampling strategies are presented in Sections III–V. In each section, we analyze and interpret the main theorems based on Fourier analysis and MIMO channel capacity, and identify sampling structures that maximize capacity. The connection between the capacity-maximizing samplers and the MMSE samplers is provided in Section VI. Proofs of the main theorems are provided in the appendices, and the notation is summarized in Table I.

$\mathcal{L}_1$	set of measurable functions f such that $\int \vert f\vert\, d\mu < \infty$
$\mathbb{S}_+$	set of positive semidefinite matrices
$h(t), H(f)$	impulse response, and frequency response of the analog channel
$s_i(t), S_i(f)$	impulse response, and frequency response of the ith post-modulation filter
$p_i(t), P_i(f)$	impulse response, and frequency response of the ith pre-modulation filter
$\mathcal{S}_\eta(f), \mathcal{S}_x(f)$	power spectral density of the noise $\eta(t)$ and the stationary input signal $x(t)$
M	number of prefilters
f_s, T_s	aggregate sampling rate, and the corresponding sampling interval ($T_s = 1/f_s$)
$q_i(t)$	modulating sequence in the ith channel
T_q	period of the modulating sequence $q_i(t)$
$\Vert\cdot\Vert_{\mathrm{F}}, \Vert\cdot\Vert_2$	Frobenius norm, ℓ_2 norm
$[x]^+, \log^+ x$	$\max\{x, 0\}, \max\{\log x, 0\}$

II. PRELIMINARIES: CAPACITY OF UNDERSAMPLED CHANNELS

A. Capacity Definition

We consider the continuous-time additive Gaussian channel (see [1, Ch. 8]), where the channel is modeled as an LTI filter with impulse response $h(t)$ and frequency response $H(f) = \int_{-\infty}^{\infty} h(t)\exp(-j2\pi ft)dt$. The transmit signal $x(t)$ is time-constrained to the interval $(0, T)$. The analog channel output is given as

$$r(t) = h(t) * x(t) + \eta(t) \tag{1}$$

and is observed over[1] $(0, T)$, where $\eta(t)$ is stationary zero-mean Gaussian noise. We assume throughout the paper that perfect channel state information, i.e., perfect knowledge of $h(t)$, is known at both the transmitter and the receiver. The analog channel capacity is defined as [1, Sec. 8.1],

$$C = \lim_{T \to \infty} \frac{1}{T} \sup I\left(\{x(t)\}_{t=0}^T; \{r(t)\}_{t=0}^T\right)$$

where the supremum is over all input distributions subject to an average power constraint $\mathbb{E}(\frac{1}{T}\int_0^T |x(\tau)|^2 d\tau) \le P$. Since any given analog channel can be converted to a countable number of independent parallel discrete channels by a Karhunen–Loève decomposition, the capacity metric quantifies the maximum mutual information between the input and output of these discrete channels. If we denote $[x]^+ = \max\{x, 0\}$ and $\log^+ x = \max\{0, \log x\}$, then the analog channel capacity is given as follows.

Theorem 1 [1, Th. 8.5.1]: Consider an analog channel with power constraint P and noise power spectral density (PSD) $\mathcal{S}_\eta(f)$. Assume that $|H(f)|^2 / \mathcal{S}_\eta(f)$ is bounded and integrable, and that either $\int_{-\infty}^{\infty} \mathcal{S}_\eta(f)df < \infty$ or that $\mathcal{S}_\eta(f)$ is white. Then, the analog channel capacity is given by

$$C = \frac{1}{2}\int_{-\infty}^{\infty} \log^+\left(\nu \frac{|H(f)|^2}{\mathcal{S}_\eta(f)}\right) df \tag{2}$$

[1]We impose the assumption that both the transmit signal and the observed signal are constrained to finite time intervals to allow for a rigorous definition of channel capacity. In particular, as per Gallager's analysis [1, Ch. 8], we first calculate the capacity for finite time intervals and then take the limit of the interval to infinity.

where ν satisfies

$$\int_{-\infty}^{\infty}\left[\nu - \frac{\mathcal{S}_\eta(f)}{|H(f)|^2}\right]^+ df = P. \tag{3}$$

For a channel whose bandwidth lies in $[-B, B]$, if we remove the noise outside the channel bandwidth via prefiltering and sample the output at a rate $f \ge 2B$, then we can perfectly recover all information conveyed within the channel bandwidth, which allows (2) to be achieved without sampling loss. For this reason, we will use the terminology *Nyquist-rate channel capacity* for the analog channel capacity (2), which is commensurate with sampling at or above the Nyquist rate of the received signal after optimized prefiltering.

Under sub-Nyquist sampling, the capacity depends on the sampling mechanism and its sampling rate. Specifically, the channel output $r(t)$ is now passed through the receiver's analog front end, which may include a filter, a bank of M filters, or a bank of preprocessors consisting of filters and modulation modules, yielding a collection of analog outputs $\{y_i(t) : 1 \le i \le M\}$. We assume that the analog outputs are observed over the time interval $(0, T)$ and then passed through ideal uniform samplers, yielding a set of digital sequences $\{y_i[n] : n \in \mathbb{Z}, 1 \le i \le M\}$, as illustrated in Fig. 1. Here, each branch is uniformly sampled at a sampling rate of f_s/M samples per second.

Define $\mathbf{y}[n] = [y_1[n], \dots, y_M[n]]$, and denote by $I(\{x(t)\}_{t=0}^T; \{\mathbf{y}[n]\}_{t=0}^T)$ the mutual information between the input $x(t)$ on the interval $0 < t \le T$ and the samples $\{y[n]\}$ observed on the interval $0 < t \le T$. We pose the problem of finding the capacity $C(f_s)$ of sampled channels as quantifying the maximum mutual information in the limit as $T \to \infty$. The sampled channel capacity can then be expressed as

$$C(f_s) = \lim_{T \to \infty} \frac{1}{T}\sup I\left(\{x(t)\}_{t=0}^T; \{\mathbf{y}[n]\}_{t=0}^T\right)$$

where the supremum is over all possible input distributions subject to an average power constraint $\mathbb{E}(\frac{1}{T}\int_0^T |x(\tau)|^2 d\tau) \le P$. We restrict the transmit signal $x(t)$ to be continuous with bounded variance (i.e., $\sup_t \mathbb{E}|x(t)|^2 < \infty$), and restrict the probability measure of $x(t)$ to be uniformly continuous. This restriction simplifies some mathematical analysis, while still encompassing most practical signals of interest.[2]

B. Sampling Mechanisms

In this section, we describe three classes of sampling strategies with increasing complexity. In particular, we start from sampling following a single filter and extend our results to incorporate filter banks and modulation banks.

1) Filtering Followed by Sampling: Ideal uniform sampling is performed by sampling the analog signal uniformly at a rate $f_s = T_s^{-1}$, where T_s denotes the sampling interval. In order to avoid aliasing, suppress out-of-band noise, and compensate for linear distortion of practical sampling devices, a prefilter is

[2]Note that this condition is not necessary for our main theorems. An alternative proof based on correlation functions is provided in [21], which does not require this condition.
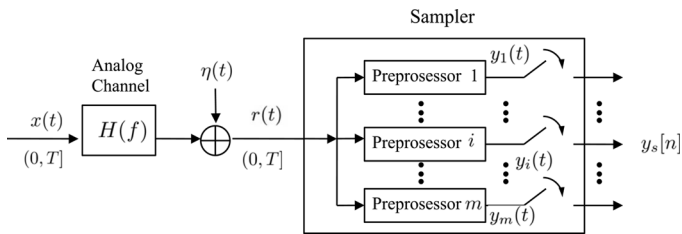
Fig. 1. Sampled Gaussian channel. The channel output $r(t)$ is passed through ranches of the receiver analog front end to yield analog outputs $\{y_i(t) : 1 \leq i \leq M\}$; each $y_i(t)$ is observed over $(0, T]$ and uniformly sampled at a rate f_s/M to yield the sampled sequence $y_i[n]$. The preprocessor can be a filter, or combination of a filter and a modulator.
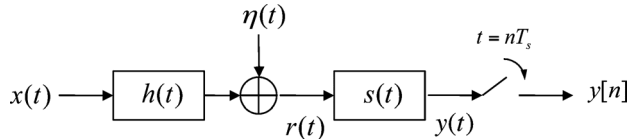

Fig. 2. Filtering followed by sampling: the analog channel output $r(t)$ is linearly filtered prior to ideal uniform sampling.

often added prior to the ideal uniform sampler [22]. Our sampling process thus includes a general analog prefilter, as illustrated in Fig. 2. Specifically, before sampling, we prefilter the received signal with an LTI filter that has impulse response $s(t)$ and frequency response $S(f)$, where we assume that $h(t)$ and $s(t)$ are both bounded and continuous. The filtered output is observed over $(0, T]$ and can be written as

$$y(t) = s(t) * (h(t) * x(t) + \eta(t)), \quad t \in (0, T]. \quad (4)$$

We then sample $y(t)$ using an ideal uniform sampler, leading to the sampled sequence

$$y[n] = y(nT_s).$$

2) Sampling Following Filter Banks: Sampling following a single filter often falls short of exploiting channel structure. In particular, although Nyquist-rate uniform sampling preserves information for bandlimited signals, for multiband signals, it does not ensure perfect reconstruction at the Landau rate (i.e., the total widths of spectral support). That is because uniform sampling at sub-Nyquist rates may suppress information by collapsing subbands, resulting in fewer degrees of freedom. This motivates us to investigate certain nonuniform sampling mechanisms. We begin by considering a popular class of nonuniform sampling mechanisms, where the received signal is preprocessed by a bank of filters. Most practical nonuniform sampling techniques [9], [11], [12] fall under filter-bank sampling and modulation-bank sampling (as described in Section II-B3). Note that the filters may introduce delays, so that this approach subsumes that of a filter bank with different sampling times at each branch.

In this sampling strategy, we replace the single prefilter in Fig. 2 by a bank of M analog filters each followed by ideal sampling at rate f_s/M, as illustrated in Fig. 3. We denote by $s_i(t)$ and $S_i(f)$ the impulse response and frequency response of the ith LTI filter, respectively. The filtered analog output in the ith branch prior to sampling is then given as

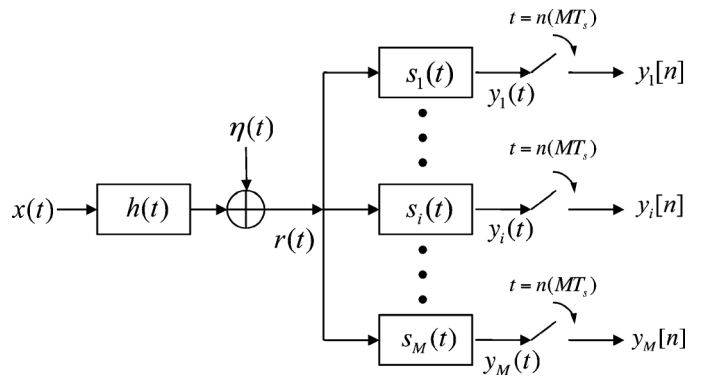$$y_i(t) = (h(t) * s_i(t)) * x(t) + s_i(t) * \eta(t), \quad t \in (0, T]. \quad (5)$$

Fig. 3. A filter bank followed by sampling: the received analog signal $r(t)$ is passed through M branches. In the ith branch, the signal $r(t)$ is passed through an LTI prefilter with impulse response $s_i(t)$, and then sampled uniformly at a rate f_s/M.
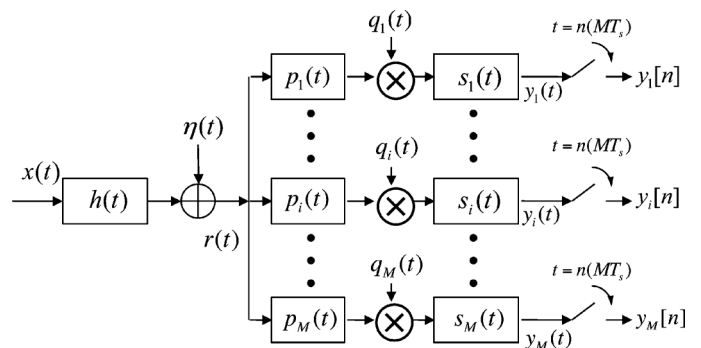

Fig. 4. Modulation and filter banks followed by sampling: in each branch, the received signal is prefiltered by an LTI filter with impulse response $p_i(t)$, modulated by a periodic waveform $q_i(t)$, filtered by another LTI filter with impulse response $s_i(t)$, and then sampled at a rate f_s/M.

These filtered signals are then sampled uniformly to yield

$$y_i[n] \triangleq y_i(nMT_s) \quad \text{and} \quad \mathbf{y}[n] \triangleq [y_1[n], y_2[n], \ldots, y_M[n]]$$

where $T_s = f_s^{-1}$.

3) Modulation and Filter Banks Followed by Sampling: We generalize the filter-bank sampling strategy by adding an additional filter bank and a modulation bank, which includes as special cases a broad class of nonuniform sampling methods that are applied in both theory and practice. Specifically, the sampling system with sampling rate f_s comprises M branches. In the ith branch, the received signal $r(t)$ is prefiltered by an LTI filter with impulse response $p_i(t)$ and frequency response $P_i(f)$, modulated by a periodic waveform $q_i(t)$ of period T_q, filtered by another LTI filter with impulse response $s_i(t)$ and frequency response $S_i(f)$, and then sampled uniformly at a rate $f_s/M = (MT_s)^{-1}$, as illustrated in Fig. 4. The first prefilter $P_i(f)$ will be useful in removing out-of-band noise, while the periodic waveforms scramble spectral contents from different aliased sets, thus bringing in more design flexibility that may potentially lead to better exploitation of channel structure. By taking advantage of random modulation sequences to achieve incoherence among different branches, this sampling mechanism has proven useful for subsampling multiband signals [13].

In the ith branch, the analog signal after postmodulation filtering prior to sampling can be written as

$$y_i(t) = s_i(t) * (q_i(t) \cdot [p_i(t) * (h(t) * x(t) + \eta(t))]) \quad (6)$$

resulting in the digital sequence of samples

$$y_i[n] = y_i(nMT_s) \quad \text{and} \quad \mathbf{y}[n] = [y_1[n], \ldots, y_M[n]]^T.$$

III. A FILTER FOLLOWED BY SAMPLING

A. Main Results

The sampled channel capacity under sampling with filtering is stated in the following theorem.

Theorem 2: Consider the system shown in Fig. 2, where $\eta(t)$ is Gaussian noise with PSD $\mathcal{S}_\eta(f)$. Assume that $h(t), s(t), S(f)\sqrt{\mathcal{S}_\eta(f)}$ are all continuous, bounded, and absolutely Riemann integrable. Additionally, suppose that $h_\eta(t) := \mathcal{F}^{-1}\left(\frac{H(f)}{\sqrt{\mathcal{S}_\eta(f)}}\right)$ satisfies $h_\eta(t) = o(t^{-\epsilon})$ for some constant[3] $\epsilon > 1$. The capacity $C(f_s)$ of the sampled channel with a power constraint P is then given parametrically as

$$C(f_s) = \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \frac{1}{2} \log^+ (\nu \gamma^{\mathrm{s}}(f)) \, df \qquad (7)$$

where ν satisfies

$$\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} [\nu - 1/\gamma^{\mathrm{s}}(f)]^+ \, df = P. \qquad (8)$$

Here, we denote

$$\gamma^{\mathrm{s}}(f) := \frac{\sum_{l \in \mathbb{Z}} |H(f - lf_s)S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2 \mathcal{S}_\eta(f - lf_s)}.$$

As expected, applying the prefilter modifies the channel gain and colors the noise accordingly. The color of the noise is reflected in the denominator term of the corresponding SNR in (7) at each $f \in [-f_s/2, f_s/2]$ within the sampling bandwidth. The channel and prefilter response leads to an equivalent frequency-selective channel, and the ideal uniform sampling that follows generates a folded version of the nonsampled channel capacity. Specifically, this capacity expression differs from the analog capacity given in Theorem 1 in that the SNR in the sampled scenario is $\gamma^{\mathrm{s}}(f)$ in contrast to $|H(f)|^2/\mathcal{S}_\eta(f)$ for the nonsampled scenario. Water filling over $1/\gamma^{\mathrm{s}}(f)$ determines the optimal power allocation.

B. Approximate Analysis

Rather than providing here a rigorous proof of Theorem 2, we first develop an approximate analysis by relating the aliased channel to MISO channels, which allows for a communication theoretic interpretation. The rigorous analysis, which is deferred to Appendix A, makes use of a discretization argument and asymptotic spectral properties of Toeplitz matrices.

[3]This condition is used in Appendix A as a sufficient condition to guarantee asymptotic properties of Toeplitz matrices. A similar condition will be used in Theorems 4 and 6.
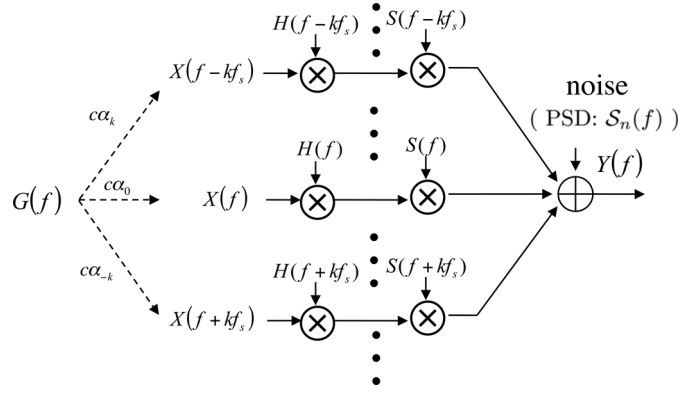

Fig. 5. Equivalent MISO Gaussian channel for a given $f \in [-f_s/2, f_s/2]$ under filtering followed by sampling. The additive noise has PSD $\mathcal{S}_n(f) = \sum_{l \in \mathbb{Z}} \mathcal{S}_\eta(f - lf_s) |S(f - lf_s)|^2$.

Consider first the equivalence between the sampled channel and a MISO channel at a single frequency $f \in [-f_s/2, f_s/2]$. As part of the approximation, we suppose the Fourier transform $X(f)$ of the transmitted signal exists.[4] The Fourier transform of the sampled signal at any $f \in [-f_s/2, f_s/2]$ is given by

$$\frac{1}{T_s} \sum_{k \in \mathbb{Z}} H(f - kf_s) S(f - kf_s) X(f - kf_s) \qquad (9)$$

due to aliasing. The summing operation allows us to treat the aliased channel at each f within the sampling bandwidth as a separate MISO channel with countably many input branches and a single-output branch, as illustrated in Fig. 5.

By assumption, the noise has spectral density $\mathcal{S}_\eta(f)$, so that the filtered noise has PSD $\mathcal{S}_\eta(f)|S(f)|^2$. The PSD of the sampled noise sequence at $f \in [-f_s/2, f_s/2]$ is then given by $\sum_{l \in \mathbb{Z}} \mathcal{S}_\eta(f - lf_s) |S(f - lf_s)|^2$. If we term $\{f - lf_s \mid l \in \mathbb{Z}\}$ the aliased frequency set for f, then the amount of power allocated to $X(f - lf_s)$ should "match" the corresponding channel gain within each aliased set in order to achieve capacity. Specifically, denote by $G(f)$ the transmitted signal for every $f \in [-f_s/2, f_s/2]$. This signal is multiplied by a constant gain $c\alpha_l(l \in \mathbb{Z})$, and sent through the lth input branch, i.e.,

$$X(f - lf_s) = c\alpha_l G(f), \quad \forall l \in \mathbb{Z} \qquad (10)$$

where c is a normalizing constant, and

$$\alpha_l = \frac{H^*(f - lf_s) S^*(f - lf_s)}{\sum_l |H(f - lf_s)S(f - lf_s)|^2}.$$

The resulting SNR can be expressed as the sum of SNRs (as shown in [23]) at each branch. Since the sampling operation combines signal components at frequencies from each aliased set $\{f - lf_s \mid l \in \mathbb{Z}\}$, it is equivalent to having a set of parallel MISO channels, each indexed by some $f \in [-f_s/2, f_s/2]$. The water-filling strategy is optimal in allocating power among the set of parallel channels, which yields the parametric (8) and completes our approximate analysis.

[4]The Fourier transform of the input signal typically does not exist since the input may have infinite energy.

C. Proof Sketch

Since the Fourier transform is not well-defined for signals with infinite energy, there exist technical flaws lurking in the approximate treatment of the previous subsection. The key step to circumvent these issues is to explore the asymptotic properties of Toeplitz matrices/operators. This approach was used by Gallager [1, Th. 8.5.1] to prove the analog channel capacity theorem. Under uniform sampling, however, the sampled channel no longer acts as a Toeplitz operator, but instead becomes a block-Toeplitz operator. Since conventional approaches [1, Ch. 8.4] do not accommodate for block-Toeplitz matrices, a new analysis framework is needed. We provide here a roadmap of our analysis framework and defer the complete proof to Appendix A.

1) Discrete Approximation: The channel response and the filter response are both assumed to be continuous, which motivates us to use a discrete-time approximation in order to transform the continuous-time operator into its discrete counterpart. We discretize a time domain process by pointwise sampling with period Δ, e.g., $h(t)$ is transformed into $\{h[n]\}$ by setting $h[n] = h(n\Delta)$. For any given T, this allows us to use a finite-dimensional matrix to approximate the continuous-time block-Toeplitz operator. Then, due to the continuity assumption, an exact capacity expression can be obtained by letting Δ go to zero.

2) Spectral Properties of Block-Toeplitz Matrices: After discretization, the input–output relation is similar to a MIMO discrete-time system. Applying MIMO channel capacity results leads to the capacity for a given T and Δ. The channel capacity is then obtained by taking T to infinity and Δ to zero, which can be related to the channel matrix's spectrum using Toeplitz theory. Since the filtered noise is nonwhite and correlated across time, we need to whiten it first. This, however, destroys the Toeplitz properties of the original system matrix. In order to apply established results in Toeplitz theory, we make use of the concept of asymptotic equivalence [24] that builds connections between Toeplitz matrices and non-Toeplitz matrices. This allows us to relate the capacity limit with spectral properties of the channel and filter response.

D. Optimal Prefilters

1) Derivation of Optimal Prefilters: Since different prefilters lead to different channel capacities, a natural question is how to choose $S(f)$ to maximize capacity. The optimizing prefilter is given in the following theorem.

Theorem 3: Consider the system shown in Fig. 2 and define

$$\gamma_l(f) := \frac{|H(f - lf_s)|^2}{\mathcal{S}_\eta(f - lf_s)}$$

for any integer l. Suppose that in each aliased set $\{f - lf_s \mid l \in \mathbb{Z}\}$, there exists k such that

$$\gamma_k(f) = \sup_{l \in \mathbb{Z}} \gamma_l(f).$$

Then, the capacity in (7) is maximized by the filter with frequency response

$$S(f - kf_s) = \begin{cases} 1, & \text{if } \gamma_k(f) = \sup_{l \in \mathbb{Z}} \gamma_l(f) \\ 0, & \text{otherwise,} \end{cases} \qquad (11)$$

for any $f \in [-f_s/2, f_s/2]$.

Proof: It can be observed from (7) that the frequency response $S(f)$ at any f can only affect the SNR at f mod f_s, indicating that we can optimize for frequencies f_1 and f_2 $\left(f_1 \neq f_2; f_1, f_2 \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]\right)$ separately. Specifically, the SNR at each f in the aliased channel is given by

$$\gamma^{\mathrm{s}}(f) = \sum_{l \in \mathbb{Z}} \gamma_l(f) \lambda_l(f)$$

where

$$\lambda_l(f) = \frac{\mathcal{S}_\eta(f - lf_s) |S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2 \mathcal{S}_\eta(f - lf_s)}$$

and $\sum_l \lambda_l(f) = 1$. That said, $\gamma^{\mathrm{s}}(f)$ is a convex combination of $\{\gamma_l, l \in \mathbb{Z}\}$ and is thus upper bounded by $\sup_{l \in \mathbb{Z}} \gamma_l$. This bound can be attained by the filter given in (11). ∎

The optimal prefilter puts all its mass in those frequencies with the highest SNR within each aliased set $\{f - lf_s \mid l \in \mathbb{Z}\}$. Even if the optimal prefilter does not exist, we can find a prefilter that achieves an information rate arbitrarily close to the maximum capacity once $\sup_{l \in \mathbb{Z}} \gamma_l(f)$ exists. The existence of the supremum is guaranteed under mild conditions, e.g., when $\gamma_l(f)$ is bounded.

2) Interpretations: Recall that $S(f)$ is applied after the noise is added. One distinguishing feature in the subsampled channel is the noninvertibility of the prefiltering operation, i.e., we cannot recover the analog channel output from sub-Nyquist samples. As shown above, the aliased SNR is a convex combination of SNRs at all aliased branches, indicating that $S(f)$ plays the role of "*weighting*" different branches. As in MRC, those frequencies with larger SNRs should be given larger weight, while those that suffer from poor channel gains should be suppressed.

The problem of finding optimal prefilters corresponds to *joint* optimization over all input and filter responses. Looking at the equivalent aliased channel for a given frequency $f \in [-f_s/2, f_s/2]$ as illustrated in Fig. 5, we have full control over both $X(f)$ and $S(f)$. Although MRC at the transmitter side maximizes the combiner SNR for a MISO channel [23], it turns out to be suboptimal for our joint optimization problem. Rather, the optimal solution is to perform selection combining [23] by setting $S(f - lf_s)$ to one for some $l = l_0$, as well as noise suppression by setting $S(f - lf_s)$ to zero for all other ls. In fact, setting $S(f)$ to zero precludes the undesired effects of noise from low SNR frequencies, which is crucial in maximizing data rate.

Another interesting observation is that optimal prefiltering equivalently generates an *alias-free* channel. After passing through an optimal prefilter, all frequencies modulo f_s except the one with the highest SNR are removed, and hence, the optimal prefilter suppresses aliasing and out-of-band noise.
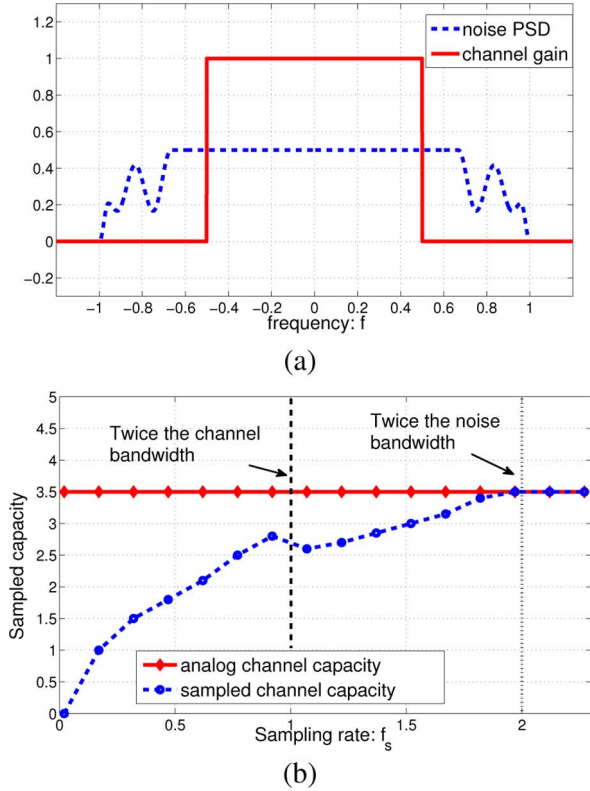
Fig. 6. Capacity of sampled additive Gaussian noise channel under ideal uniform sampling without filtering. (a) The channel gain and the PSD of the noise. (b) Sampled channel capacity versus analog channel capacity under a power constraint $P = 5$.

This alias-suppressing phenomena, while different from many sub-Nyquist works that advocate mixing instead of alias suppressing [13], arises from the fact that we have control over the input shape.

E. Numerical Examples

1) Additive Gaussian Noise Channel Without Prefiltering: The first example we consider is the additive Gaussian noise channel. The channel gain is flat within the channel bandwidth $B = 0.5$, i.e., $H(f) = 1$ if $f \in [-B, B]$ and $H(f) = 0$ otherwise. The noise is modeled as a measurable and stationary Gaussian process with the PSD plotted in Fig. 6(a). This is the noise model adopted by Lapidoth in [25] to approximate white noise, which avoids the infinite variance of the standard model for unfiltered white noise. We employ ideal pointwise sampling without filtering.

Since the noise bandwidth is larger than the channel bandwidth, ideal uniform sampling without prefiltering does not allow analog capacity to be achieved when sampling at a rate equal to twice the channel bandwidth, i.e., the Nyquist rate. Increasing the sampling rate above twice the channel bandwidth (but below the noise bandwidth) spreads the total noise power over a larger sampling bandwidth, reducing the noise density at each frequency. This allows the sampled capacity to continue increasing when sampling above the Nyquist rate, as illustrated in Fig. 6(b). It can be seen that the capacity does not increase monotonically with the sampling rate. We will discuss this phenomena in more detail in Section III-E3.

2) Optimally Filtered Channel: In general, the frequency response of the optimal prefilter is discontinuous, which may

be hard to realize in practice. However, for certain classes of channel models, the prefilter has a smooth frequency response. One example of this channel class is a *monotone channel*, whose channel response obeys $|H(f_1)|^2 / \mathcal{S}_\eta(f_1) \geq |H(f_2)|^2 / \mathcal{S}_\eta(f_2)$ for any $f_1 > f_2$. Theorem 3 implies that the optimizing prefilter for a monotone channel reduces to a low-pass filter with cutoff frequency $f_s/2$.

For nonmonotone channels, the optimal prefilter may not be a low-pass filter, as illustrated in Fig. 7. Fig. 7(b) shows the optimal filter for the channel given in Fig. 7(a) with $f_s = 0.4 f_{\mathrm{NYQ}}$, which is no longer a low-pass filter.

3) Capacity Nonmonotonicity: When the channel is not monotone, a somewhat counterintuitive fact arises: the channel capacity $C(f_s)$ is not necessarily a nondecreasing function of the sampling rate f_s. This occurs, for example, in multiband channels as illustrated in Fig. 8. Here, the Fourier transform of the channel response is concentrated in two subintervals within the overall channel bandwidth. Specifically, the entire channel bandwidth is contained in $[-0.5, 0.5]$ with Nyquist rate $f_{\mathrm{NYQ}} = 1$, and that the channel frequency response is given by

$$H(f) = \begin{cases} 1, & \text{if } |f| \in \left[\frac{1}{10}, \frac{1}{5}\right] \bigcup \left[\frac{2}{5}, \frac{1}{2}\right] \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

If this channel is sampled at a rate $f_s = \frac{3}{5} f_{\mathrm{NYQ}}$, then aliasing occurs and leads to an aliased channel with one subband (and hence one degree of freedom). However, if sampling is performed at a rate $f_s = \frac{2}{5} f_{\mathrm{NYQ}}$, it can be easily verified that the two subbands remain nonoverlapping in the aliased channel, resulting in two degrees of freedom.

The tradeoff curve between capacity and sampling rate with an optimal prefilter is plotted in Fig. 8(b). This curve indicates that increasing the sampling rate may not necessarily increase capacity for certain channel structures. In other words, a single filter followed by sampling largely constrains our ability to exploit channel and signal structures. This is not the case for more general sampling structures, as we show in the next section.
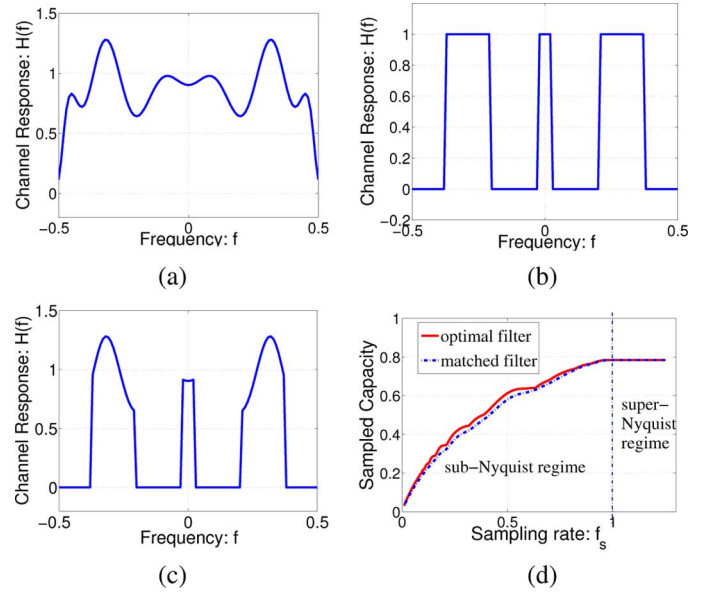

Fig. 7. Capacity of optimally filtered channel: (a) frequency response of the original channel; (b) optimal prefilter associated with this channel for sampling rate 0.4; (c) optimally filtered channel response with sampling rate 0.4; (d) capacity versus sampling rate for the optimal prefilter and for the matched filter.
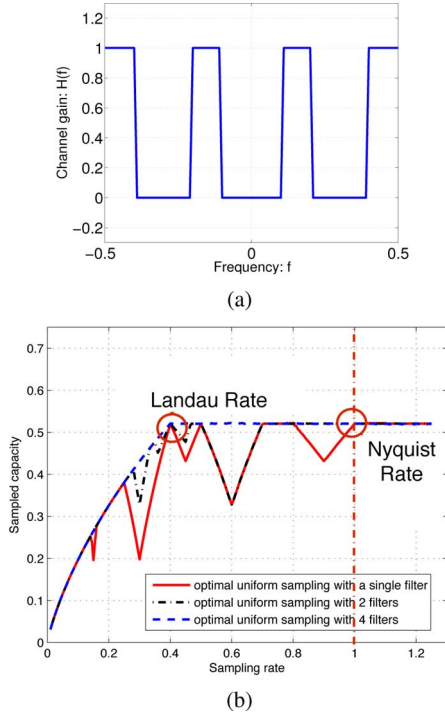
Fig. 8. Sampled channel capacity for a multiband channel under filter-bank sampling. (a) Channel gain of the multiband channel. The power constraint is $P = 10$, and the noise power is $\sigma_\eta^2 = 1$. (b) Sampled channel capacity for a single filter followed by sampling and for a filter bank followed by sampling for a bank of two filters and of four filters.

IV. A BANK OF FILTERS FOLLOWED BY SAMPLING

A. Main Results

We now treat filter-bank sampling, in which the channel output is filtered and sampled through M multiple branches as illustrated in Fig. 3.

In order to state our capacity results, we introduce two matrices $\mathbf{F}_s$ and $\mathbf{F}_h$ defined in the Fourier domain. Here, $\mathbf{F}_s$ is an infinite matrix of m rows and infinitely many columns and $\mathbf{F}_h$ is a diagonal infinite matrix such that for every $i(1 \leq i \leq k)$ and every integer l:

$$(\mathbf{F}_s(f))_{i,l} = S_i\left(f - \frac{lf_s}{M}\right)\sqrt{\mathcal{S}_\eta\left(f - \frac{lf_s}{M}\right)}$$

$$(\mathbf{F}_h(f))_{l,l} = H\left(f - \frac{lf_s}{M}\right)\bigg/\sqrt{\mathcal{S}_\eta\left(f - \frac{lf_s}{M}\right)}.$$

Theorem 4: Consider the system shown in Fig. 3. Assume that $h(t)$ and $s_i(t)$ $(1 \leq i \leq M)$ are all continuous, bounded and absolutely Riemann integrable. Additionally, assume that $h_\eta(t) := \mathcal{F}^{-1}\left(\frac{H(f)}{\sqrt{\mathcal{S}_\eta(f)}}\right)$ satisfies $h_\eta(t) = o\left(t^{-\epsilon}\right)$ for some constant $\epsilon > 1$, and that $\mathbf{F}_s$ is right-invertible for every f. Define $\tilde{\mathbf{F}}_s \triangleq (\mathbf{F}_s \mathbf{F}_s^*)^{-\frac{1}{2}} \mathbf{F}_s$. The capacity $C(f_s)$ of the sampled channel with a power constraint P is given as

$$C(f_s) = \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \frac{1}{2} \sum_{i=1}^{M} \log^+\left(\nu\lambda_i\left(\tilde{\mathbf{F}}_s\mathbf{F}_h\mathbf{F}_h^*\tilde{\mathbf{F}}_s^*\right)\right) df$$

where

$$\int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \sum_{i=1}^{M} \left[\nu - \frac{1}{\lambda_i\left(\tilde{\mathbf{F}}_s\mathbf{F}_h\mathbf{F}_h^*\tilde{\mathbf{F}}_s^*\right)}\right]^+ df = P.$$

Here, $\lambda_i\left(\tilde{\mathbf{F}}_s\mathbf{F}_h\mathbf{F}_h^*\tilde{\mathbf{F}}_s^*\right)$ denotes the ith largest eigenvalue of $\tilde{\mathbf{F}}_s\mathbf{F}_h\mathbf{F}_h^*\tilde{\mathbf{F}}_s^*$.

Remark 1: We can express this capacity in a more traditional MIMO capacity form as

$$C(f_s) = \max_{\{\mathbf{Q}(f)\}\in\mathcal{Q}} \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \frac{1}{2} \log\det\left(\mathbf{I}_M + \tilde{\mathbf{F}}_s\mathbf{F}_h\mathbf{Q}\mathbf{F}_h^*\tilde{\mathbf{F}}_s^*\right) df \tag{13}$$

where $\tilde{\mathbf{F}}_s \triangleq (\mathbf{F}_s\mathbf{F}_s^*)^{-\frac{1}{2}} \mathbf{F}_s$ and

$$\mathcal{Q} = \left\{\mathbf{Q}(f) : |f| \leq \frac{f_s}{2M}, \mathbf{Q}(f) \in \mathbb{S}_+; \right.$$

$$\left. \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \mathrm{Tr}\left(\mathbf{Q}(f)\right) df = P\right\}.$$

The optimal $\{\mathbf{Q}(f)\}$ corresponds to a water-filling power allocation strategy based on the singular values of the equivalent channel matrix $\tilde{\mathbf{F}}_s\mathbf{F}_h$, where $\mathbf{F}_h$ is associated with the original channel and $\tilde{\mathbf{F}}_s$ arises from prefiltering and noise whitening. For each $f \in [-f_s/2M, f_s/2M]$, the integrand in (13) can be interpreted as a MIMO capacity formula. We have M receive branches and can still optimize the transmitted signals $\left\{X\left(f - \frac{lf_s}{M}\right) | l \in \mathbb{Z}\right\}$ at a countable number of input branches, but this time we have M receive branches. The channel capacity is achieved when the transmit signals are designed to decouple this MIMO channel into M parallel channels (and hence M degrees of freedom), each associated with one of its singular directions.

B. Approximate Analysis

The sampled analog channel under filter-bank sampling can be studied through its connection with MIMO Gaussian channels (see Fig. 9). Consider first a single frequency $f \in [-f_s/2M, f_s/2M]$. Since we employ a bank of filters each followed by an ideal uniform sampler, the equivalent channel has M receive branches, each corresponding to one branch of filtered sampling at rate f_s/M. The noise received in the ith branch is zero-mean Gaussian with spectral density

$$\sum_{l\in\mathbb{Z}} \left|S_i\left(f - \frac{lf_s}{M}\right)\right|^2 \mathcal{S}_\eta\left(f - \frac{lf_s}{M}\right), \quad f \in \left[-\frac{f_s}{2M}, \frac{f_s}{2M}\right]$$

indicating the mutual correlation of noise at different branches. The received noise vector can be whitened by multiplying $\mathbf{Y}(f) = [\ldots, Y(f), Y(f - f_s), \ldots]^T$ by an $M \times M$ whitening matrix $(\mathbf{F}_s(f)\mathbf{F}_s^*(f))^{-\frac{1}{2}}$. Since the whitening operation is invertible, it preserves capacity. After whitening, the channel of Fig. 9 at frequency f has the following channel matrix:

$$(\mathbf{F}_s(f)\mathbf{F}_s^*(f))^{-\frac{1}{2}} \mathbf{F}_s(f)\mathbf{F}_h(f) = \tilde{\mathbf{F}}_s(f)\mathbf{F}_h(f). \tag{14}$$
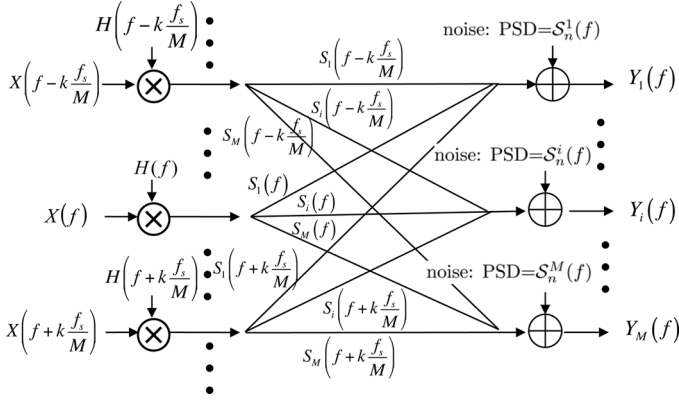
Fig. 9. Equivalent MIMO Gaussian channel for a frequency $f \in [-f_s/2M, f_s/2M]$ under sampling with a bank of M filters. Here, $\mathcal{S}_n^i(f) = \sum_{l \in \mathbb{Z}} \mathcal{S}_n(f - lf_s/M)|S_i(f - lf_s/M)|^2$.

MIMO Gaussian channel capacity results [26] immediately imply that the capacity of the channel in Fig. 9 at any $f \in [-f_s/2M, f_s/2M]$ can be expressed as

$$\max_{\mathbf{Q}} \frac{1}{2} \log \det \left[\mathbf{I} + \tilde{\mathbf{F}}_s(f) \mathbf{F}_h(f) \mathbf{Q}(f) \mathbf{F}_h^*(f) \tilde{\mathbf{F}}_s^*(f) \right] \quad (15)$$

subject to the constraints that trace $(\mathbf{Q}(f)) \leq P(f)$ and $\mathbf{Q}(f) \in \mathbb{S}_+$, where $\mathbf{Q}(f)$ denotes the power allocation matrix. Performing water-filling power allocation across all parallel channels leads to our capacity expression.

C. Optimal Filter Bank

1) Derivation of Optimal Filter Banks: In general, $\log \det[\mathbf{I}_M + \tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{Q} \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*]$ is not perfectly determined by $\tilde{\mathbf{F}}_s(f)$ and $\mathbf{F}_h(f)$ at a single frequency f, but also depends on the water-level, since the optimal power allocation strategy relies on the power constraint P/σ_η^2 as well as $\mathbf{F}_s$ and $\mathbf{F}_h$ across all f. In other words, $\log \det[\mathbf{I}_M + \tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{Q} \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*]$ is a function of all singular values of $\tilde{\mathbf{F}}_s \mathbf{F}_h$ and the universal water-level associated with optimal power allocation. Given two sets of singular values, we cannot determine which set is preferable without accounting for the water-level, unless one set is elementwise larger than the other. That said, if there exists a prefilter that maximizes all singular values simultaneously, then this prefilter will be universally optimal regardless of the water-level. Fortunately, such optimal schemes exist, as we characterize in Theorem 5.

Since $\mathbf{F}_h(f)$ is a diagonal matrix, $\lambda_k(\mathbf{F}_h\mathbf{F}_h^*)$ denotes the kth largest entry of $\mathbf{F}_h\mathbf{F}_h^*$. The optimal filter bank can then be given as follows.

Theorem 5: Consider the system shown in Fig. 3. Suppose that for each aliased set $\left\{ f - \frac{if_s}{M} \mid i \in \mathbb{Z} \right\}$ and each k ($1 \leq k \leq M$), there exists an integer l such that $\frac{|H(f - \frac{lf_s}{M})|^2}{\mathcal{S}_n(f - \frac{lf_s}{M})}$ is equal to the kth largest element in $\left\{ \frac{|H(f - \frac{if_s}{M})|^2}{\mathcal{S}_n(f - \frac{if_s}{M})} \mid i \in \mathbb{Z} \right\}$. The capacity (13) under filter-bank sampling is then maximized by a bank of filters for which the frequency response of the kth filter is given by

$$S_k\left(f - \frac{lf_s}{M}\right) = \begin{cases} 1, & \text{if } \frac{|H(f - \frac{lf_s}{M})|^2}{\mathcal{S}_n(f - \frac{lf_s}{M})} = \lambda_k(\mathbf{F}_h(f)\mathbf{F}_h^*(f)); \\ 0, & \text{otherwise}, \end{cases} \quad (16)$$

for all $l \in \mathbb{Z}, 1 \leq k \leq M$ and $f \in \left[-\frac{f_s}{2M}, \frac{f_s}{2M} \right]$. The resulting maximum channel capacity is given by

$$C(f_s) = \frac{1}{2} \int_{-f_s/2M}^{f_s/2M} \sum_{k=1}^{M} \log^+ \left(\nu \cdot \lambda_k(\mathbf{F}_h\mathbf{F}_h^*) \right) df \quad (17)$$

where ν is chosen such that

$$\int_{-f_s/2M}^{f_s/2M} \sum_{k=1}^{M} \left[\nu - \frac{1}{\lambda_k(\mathbf{F}_h\mathbf{F}_h^*)} \right]_+ df = P. \quad (18)$$

Proof: See Appendix C. ∎

The choice of prefilters in (16) achieves the upper bounds on all singular values, and is hence universally optimal regardless of the water level. Since $\tilde{\mathbf{F}}_s$ has orthonormal rows, it acts as an orthogonal projection and outputs an M-dimensional subspace. The rows of the diagonal matrix $\mathbf{F}_h$ are orthogonal to each other. Therefore, the subspace closest to the channel space spanned by $\mathbf{F}_h$ corresponds to the M rows of $\mathbf{F}_h$ containing the highest channel gains out of the entire aliased frequency set $\left\{ f - \frac{lf_s}{M} \mid l \in \mathbb{Z} \right\}$. The maximum data rate is then achieved when the filter bank outputs M frequencies with the highest SNR among the set of frequencies equivalent modulo $\frac{f_s}{M}$ and suppresses noise from all other branches.

We note that if we consider the enlarged aliased set $\{f - lf_s/M \mid l \in \mathbb{Z}\}$, then the optimal filter bank is equivalent to generating an alias-free channel over the frequency interval $[-f_s/2M, f_s/2M]$. This again arises from the nature of the joint-optimization problem: since we are allowed to control the input shape and sampling jointly, we can adjust the input shape based on the channel structure in each branch, which turn out to be alias-suppressing.

D. Discussion and Numerical Examples

In a *monotone* channel, the optimal filter bank will sequentially crop out the M best frequency bands, each of bandwidth f_s/M. Concatenating all of these frequency bands results in a low-pass filter with cutoff frequency $f_s/2$, which is equivalent to single-branch sampling with an optimal filter. In other words, for monotone channels, using filter banks harvests no gain in capacity compared to a single branch with a filter followed by sampling.

For more general channels, the capacity is not necessarily a monotone function of f_s. Consider again the multiband channel where the channel response is concentrated in two subintervals, as illustrated in Fig. 8(a). As discussed above, sampling following a single filter only allows us to select the best single frequency with the highest SNR out of the set $\{f - lf_s \mid l \in \mathbb{Z}\}$, while sampling following filter banks allows us to select the best f out of the set $\left\{ f - l\frac{f_s}{M} \mid l \in \mathbb{Z} \right\}$. Consequently, the channel capacity with filter-bank sampling exceeds that of sampling with a single filter, but neither capacity is monotonically increasing in f_s. This is shown in Fig. 8(b). Specifically, we see in this figure that when we apply a bank of two filters prior to sampling, the capacity curve is still nonmonotonic but outperforms a single filter followed by sampling.

Another consequence of our results is that when the number of branches is optimally chosen, the Nyquist-rate channel capacity can be achieved by sampling at any rate above the Landau rate. In order to show this, we introduce the following notion of a channel permutation. We call $\tilde{H}(f)$ a *permutation* of a channel response $H(f)$ at rate f_s if, for any f,

$$\left\{ \frac{|\tilde{H}(f - lf_s)|^2}{S_\eta(f - lf_s)} \mid l \in \mathbb{Z} \right\} = \left\{ \frac{|H(f - lf_s)|^2}{S_\eta(f - lf_s)} \mid l \in \mathbb{Z} \right\}.$$

The following proposition characterizes a sufficient condition that allows the Nyquist-rate channel capacity to be achieved at any sampling rate above the Landau rate.

Proposition 1: If there exists a permutation $\tilde{H}(f)$ of $H(f)$ at rate $\frac{f_s}{M}$ such that the support of $\tilde{H}(f)$ is $[-f_L/2, f_L/2]$, then optimal sampling following a bank of M filters achieves Nyquist-rate capacity when $f_s \geq f_L$.

Examples of channels satisfying Proposition 1 include any multiband channel with N subbands among which K subbands have nonzero channel gain. For any $f_s \geq f_L = \frac{K}{N} f_{\mathrm{NYQ}}$, we are always able to permute the channel at rate f_s/K to generate a band-limited channel of spectral support size f_L. Hence, sampling above the Landau rate following K filters achieves the Nyquist-rate channel capacity. This is illustrated in Fig. 8(b) where sampling with a four-branch filter bank has a higher capacity than sampling with a single filter, and achieves the Nyquist-rate capacity whenever $f_s \geq \frac{2}{5} f_{\mathrm{NYQ}}$. The optimal filter-bank sampling for most general channels is identified in [21], where both the number of branches and per-branch sampling rate are allowed to vary.

V. Modulation and Filter Banks Followed by Sampling

A. Main Results

We now treat modulation and filter banks followed by sampling. Assume that $\tilde{T}_s := MT_s = \frac{b}{a} T_q$, where a and b are coprime integers, and that the Fourier transform of $q_i(t)$ is given as $\sum_l c_i^l \delta(f - lf_q)$. Before stating our theorem, we introduce the following two Fourier symbol matrices $\mathbf{F}^\eta$ and $\mathbf{F}^h$. The $aM \times \infty$-dimensional matrix $\mathbf{F}^\eta$ contains M submatrices with the αth submatrix given by an $a \times \infty$-dimensional matrix $\mathbf{F}_\alpha^\eta \mathbf{F}_\alpha^p$. Here, for any $v \in \mathbb{Z}, 1 \leq l \leq a$, and $1 \leq \alpha \leq M$, we have

$$(\mathbf{F}_\alpha^\eta)_{l,v} = (\mathbf{F}_\alpha^p)_{v,v} \left[\sum_u c_\alpha^u S_\alpha \left(-f + uf_q + v\frac{f_q}{b} \right) \right. $$
$$\left. \cdot \exp \left(-j2\pi l MT_s \left(f - uf_q - v\frac{f_q}{b} \right) \right) \right].$$

The matrices $\mathbf{F}_\alpha^p$ and $\mathbf{F}^h$ are infinite diagonal matrices such that for every integer l:

$$(\mathbf{F}_\alpha^p)_{l,l} = P_\alpha \left(-f + l\frac{f_q}{b} \right) \sqrt{S_\eta \left(-f + l\frac{f_q}{b} \right)}$$

$$(\mathbf{F}^h)_{l,l} = \frac{H \left(-f + l\frac{f_q}{b} \right)}{\sqrt{S_\eta \left(-f + l\frac{f_q}{b} \right)}}.$$

Theorem 6: Consider the system shown in Fig. 4. Assume that $h(t), p_i(t)$, and $s_i(t)$ $(1 \leq i \leq M)$ are all continuous, bounded and absolutely Riemann integrable, $\mathbf{F}^\eta$ is right invertible, and that the Fourier transform of $q_i(t)$ is given as $\sum_l c_i^l \delta(f - lf_q)$. Additionally, suppose that $h_\eta(t) := \mathcal{F}^{-1} \left(\frac{H(f)}{\sqrt{S_\eta(f)}} \right)$ satisfies $h_\eta(t) = o(t^{-\epsilon})$ for some constant $\epsilon > 1$. We further assume that $aMT_s = bT_q$, where a and b are coprime integers. The capacity $C(f_s)$ of the sampled channel with a power constraint P is given by

$$C(f_s) = \int_{-\frac{f_s}{2aM}}^{\frac{f_s}{2aM}} \frac{1}{2} \sum_{i=1}^{aM} \log^+ \left(\nu \lambda_i \left((\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}} \mathbf{F}^\eta \mathbf{F}^h \cdot \right. \right.$$
$$\left. \left. \mathbf{F}^{h*} \mathbf{F}^{\eta*} (\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}} \right) \right) df, \qquad (19)$$

where ν is chosen such that

$$P = \int_{-\frac{f_s}{2aM}}^{\frac{f_s}{2aM}} \sum_{i=1}^{aM} \left[\nu - \lambda_i^{-1} \left((\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}} \mathbf{F}^\eta \mathbf{F}^h \cdot \right. \right.$$
$$\left. \left. \mathbf{F}^{h*} \mathbf{F}^{\eta*} (\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}} \right) \right]^+ df.$$

Remark 2: The right invertibility of $\mathbf{F}^\eta$ ensures that the sampling method is nondegenerate, e.g., the modulation sequence cannot be zero.

The optimal ν corresponds to a water-filling power allocation strategy based on the singular values of the equivalent channel matrix $(\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}} \mathbf{F}^\eta \mathbf{F}^h$, where $(\mathbf{F}^\eta \mathbf{F}^{\eta*})^{-\frac{1}{2}}$ is due to noise prewhitening and $\mathbf{F}^\eta \mathbf{F}^h$ is the equivalent channel matrix after modulation and filtering. This result can again be interpreted by viewing (19) as the MIMO Gaussian channel capacity of the equivalent channel. We note that a closed-form capacity expression may be hard to obtain for general modulating sequences $q_i(t)$. This is because the multiplication operation corresponds to convolution in the frequency domain which does not preserve Toeplitz properties of the original operator associated with the channel filter. When $q_i(t)$ is periodic, however, it can be mapped to a spike train in the frequency domain, which preserves block Toeplitz properties, as described in more detail in Appendix D.

B. Approximate Analysis

The Fourier transform of the signal prior to modulation in the ith branch at a given frequency f can be expressed as $P_i(f)R(f)$, where $R(f) = H(f)X(f) + N(f)$. Multiplication of this premodulation signal with the modulation sequence $q_i(t) = \sum_l c_i^l \delta(f - lf_q)$ corresponds to convolution in the frequency domain.

Recall that $bT_q = aMT_s$ with integers a and b. We, therefore, divide all samples $\{y_i[k] \mid k \in \mathbb{Z}\}$ in the ith branch into a groups, where the lth $(0 \leq l < a)$ group contains $\{y_i[l + ka] \mid k \in \mathbb{Z}\}$. Hence, each group is equivalent to the samples obtained by sampling at rate $f_s/Ma = f_q/b$. The sampling system, when restricted to the output on each group of the sampling set, can be treated as LTI, thus justifying its equivalent representation in the spectral domain. Specifically, for the ith branch, we denote by

$$g_\eta^i(t, \tau) := \int s_i(t - \tau_1) q_i(\tau_1) p(\tau_1 - \tau) d\tau_1$$

the output response of the preprocessing system at time t due to an input impulse at time τ. We then introduce a new LTI impulse response $\tilde{g}_l^i(t)$ associated with the lth group such that $\tilde{g}_l^i(t) := g_\eta^i(l\tilde{T}_s, l\tilde{T}_s - t)$. It can easily be shown that when the same sampling set $\left\{ (l + ka)\tilde{T}_s \mid k \in \mathbb{Z} \right\}$ is employed, the preprocessing system associated with $g_\eta^i(t, \tau)$ results in the same sampled output as the one associated with $\tilde{g}_l^i(t)$. This allows us to treat the samples of each distinct group as the ones obtained by an LTI preprocessing system followed by uniform sampling.

Suppose the channel output $R(f)$ is passed through the LTI preprocessing system associated with the lth group of the ith branch, i.e., the one associated with $\tilde{g}_l^i(t)$. The Fourier transform of the output of this LTI system prior to uniform sampling, as marked in Fig. 10(b), can be written as

$$
\begin{aligned}
&\tilde{Y}_i^l(f) \\
&\triangleq P_i(f)R(f)\left(S_i(f)\exp\left(j2\pi f l\tilde{T}_s \right) * \sum_u c_i^u \delta\left(f - u f_q \right) \right) \\
&= P_i(f)R(f)\sum_u c_i^u S_i\left(f - u f_q \right)\exp\left(j2\pi l\tilde{T}_s\left(f - u f_q \right) \right).
\end{aligned}
$$

After uniform sampling at rate f_q/b, the Fourier transform of the samples in the lth group can be expressed as

$$
\begin{aligned}
Y_i^l(f) &= \sum_v \tilde{Y}_i^l\left(f - \frac{v f_q}{b} \right) \\
&= \sum_v P_i\left(f - \frac{v f_q}{b} \right) R\left(f - \frac{v f_q}{b} \right) \sum_u c_i^u \cdot \\
&\quad S_i\left(f - u f_q - \frac{v f_q}{b} \right) \exp\left(j2\pi l\tilde{T}_s\left(f - u f_q - \frac{v f_q}{b} \right) \right) \\
&= \sum_v A_{l,v}^i(f) P_i\left(f - v\frac{f_q}{b} \right) R\left(f - v\frac{f_q}{b} \right),
\end{aligned}
$$

where

$$
\begin{aligned}
A_{l,v}^i(f) &\triangleq \sum_u c_i^u S_i\left(f - u f_q - \frac{v f_q}{b} \right) \cdot \\
&\quad \exp\left(j2\pi l\tilde{T}_s\left(f - u f_q - \frac{v f_q}{b} \right) \right). \quad (20)
\end{aligned}
$$

Since the sampled outputs of the original sampling system are equivalent to the union of samples obtained by Ma LTI systems each followed by uniform sampling at rate f_q/b, we can transform the true sampling system into a MIMO Gaussian channel with an infinite number of input branches and finitely many output branches, as illustrated in Fig. 10. The well-known formula for the capacity of a MIMO channel can now be used to derive our capacity results.

We note that due to the convolution in the spectral domain, the frequency response of the sampled output at frequency f is a linear combination of frequency components $\{X(f)\}$ and $\{N(f)\}$ from several different aliased frequency sets. We define the *modulated aliased frequency set* as a generalization of the aliased set. Specifically, for each f, the modulated aliased set is
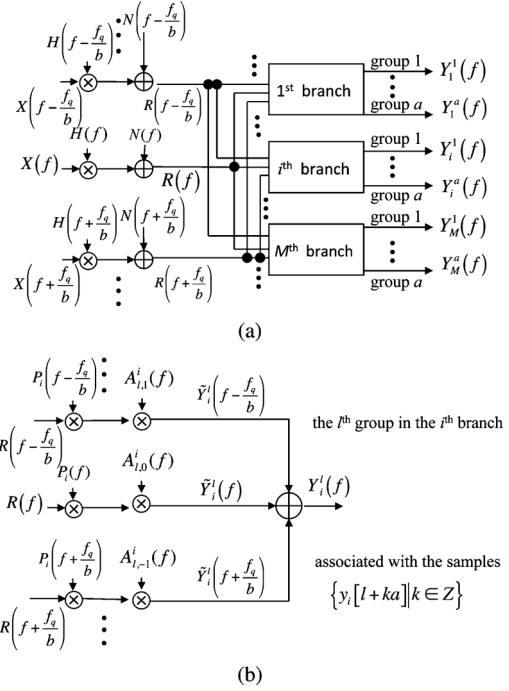

Fig. 10. Equivalent MIMO Gaussian channel for a given $f \in [0, f_q/b)$ under sampling with modulation banks and filter banks. (a) The overall MIMO representation, where each branch has a output each corresponding to a distinct group. (b) The MISO representation of the lth group in the ith branch, where $A_{l,v}^i(f)$ is defined in (20). This is associated with the set of samples $\{y_i[l + ka] \mid k \in \mathbb{Z}\}$.

given by[5] $\left\{ f - l f_q - k\tilde{f}_s \mid l, k \in \mathbb{Z} \right\}$. By our assumption that $f_q = \frac{b}{a}\tilde{f}_s$ with a and b being relatively prime, simple results in number theory imply that

$$
\begin{aligned}
\left\{ f_0 - l f_q - k\tilde{f}_s \mid l, k \in \mathbb{Z} \right\} &= \left\{ f_0 - l f_q/b \mid l \in \mathbb{Z} \right\} \\
&= \left\{ f_0 - l\tilde{f}_s/a \mid l \in \mathbb{Z} \right\}.
\end{aligned}
$$

In other words, for a given $f_0 \in [-f_q/2b, f_q/2b]$, the sampled output at f_0 depends on the input in the entire modulated aliased set. Since the sampling bandwidth at each branch is $\tilde{f}_s$, all outputs at frequencies $\left\{ f_0 - l f_q/b \mid l \in \mathbb{Z}; -\tilde{f}_s/2 \leq f_0 - l f_q/b \leq \tilde{f}_s/2 \right\}$ rely on the inputs in the same modulated aliased set. This can be treated as a Gaussian MIMO channel with a countable number of input branches at the frequency set $\left\{ f_0 - l\tilde{f}_s/a \mid l \in \mathbb{Z} \right\}$ and aM groups of output branches, each associated with one group of sample sequences in one branch. As an example, we illustrate in Fig. 10 the equivalent MIMO Gaussian channel under sampling following a single branch of modulation and filtering, when $S(f) = 0$ for all $f \notin [-f_s/2, f_s/2]$.

The effective frequencies of this frequency-selective MIMO Gaussian channel range from $-f_q/2b$ to $f_q/2b$, which gives us a set of parallel channels each representing a single frequency f. The water-filling power allocation strategy is then applied to achieve capacity.

---

[5]We note that although each modulated aliased set is countable, it may be a dense set when $f_q/\tilde{f}_s$ is irrational. Under the assumption in Theorem 6, however, the elements in the set have a minimum spacing of f_q/b.

A rigorous proof of Theorem 6 based on Toeplitz properties is provided in Appendix D.

C. An Upper Bound on Sampled Capacity

Following the same analysis of optimal filter-bank sampling developed in Section IV-C, we can derive an upper bound on the sampled channel capacity.

Corollary 1: Consider the system shown in Fig. 4. Suppose that for each aliased set $\{f - if_q/b \mid i \in \mathbb{Z}\}$ and each k ($1 \leq k \leq aM$), there exists an integer l such that $\frac{|H(f-lf_q/b)|^2}{\mathcal{S}_\eta(f-lf_q/b)}$ is equal to the kth largest element in $\left\{ \frac{|H(f-if_q/b)|^2}{\mathcal{S}_\eta(f-if_q/b)} \mid i \in \mathbb{Z} \right\}$. The capacity (19) under sampling following modulation and filter banks can be upper bounded by

$$C^u(f_s) \triangleq \frac{1}{2} \int_{-f_q/2b}^{f_q/2b} \sum_{k=1}^{aM} \log^+ \left(\nu \cdot \lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right) \right) df \qquad (21)$$

where ν is chosen such that

$$\int_{-f_q/2b}^{f_q/2b} \sum_{k=1}^{aM} \left[\nu - \frac{1}{\lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right)} \right]^+ df = P. \qquad (22)$$

Proof: By observing that $\left(\mathbf{F}^\eta \mathbf{F}^{\eta*} \right)^{-\frac{1}{2}} \mathbf{F}^\eta$ has orthonormal rows, we can derive the result using Proposition 4 in Appendix C. ∎

The upper bound of Corollary 1 coincides with the upper bound on sampled capacity under aM-branch filter-bank sampling. This basically implies that for a given sampling rate f_s, modulation and filter bank sampling does not outperform filter-bank sampling in maximizing sampled channel capacity. In other words, we can always achieve the same performance by adding more branches in filter-bank sampling.

Note, however, that this upper bound may not be tight, since we restrict our analysis to periodic modulation sequences. General modulation is not discussed here.

D. Single-Branch Sampling With Modulation and Filtering Versus Filter-Bank Sampling

Although the class of modulation and filter bank sampling does not provide capacity gain compared with filter-bank sampling, it may potentially provide implementation advantages, depending on the modulation period T_q. Specifically, modulation-bank sampling may achieve a larger capacity region than that achievable by filter-bank sampling with the same number of branches. We consider here two special cases of single-branch modulation sampling and investigate whether any hardware benefit can be harvested.

1) $f_s/M = f_q/a$ for some integer a: In this case, the modulated aliased set is $\{f - kf_s/M - lf_q \mid k, l \in \mathbb{Z}\} = \{f - kf_s/M \mid k \in \mathbb{Z}\}$, which is equivalent to the original aliased frequency set. That said, the sampled output $Y(f)$ is still a linear combination of $\{R(f - kf_s/M) \mid k \in \mathbb{Z}\}$. But since linear combinations of these components can be attained by simply adjusting the prefilter response $S(f)$, the modulation bank does not provide any further design degrees of freedom,
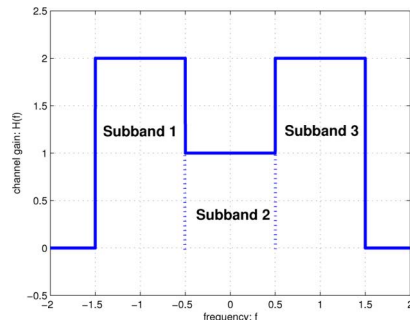

Fig. 11.   The channel gain of Example 1. The noise is of unit power spectral density.

and hence does not improve the capacity region achievable by sampling with a bank of M filters.

2) $f_s/M = bf_q$ for some integer b: In this case, the modulated aliased set is enlarged to $\{f - kf_s/M - lf_q \mid k, l \in \mathbb{Z}\} = \{f - lf_q \mid l \in \mathbb{Z}\}$, which may potentially provide implementation gain compared with filter-bank sampling with the same number of branches. We illustrate this in the following example.

Example 1: Suppose that the channel contains 3 subbands with channel gains as plotted in Fig. 11, and that the noise is of unit spectral density within these 3 subbands and 0 otherwise.

i) Let us first consider single-branch sampling with filtering with $f_s = 2$. As illustrated in Fig. 11, Subbands 1 and 3 are mixed together due to aliasing. According to Section III-D, the optimal prefilter without modulation would be a bandpass filter with passband $[-1.5, 0.5]$, resulting in a channel containing two subbands with respective channel gains 2 and 1.

ii) If we add a modulation sequence with period $T_q = 2T_s$, then the channel structure can be better exploited. Specifically, suppose that the modulation sequence obeys $c^0 = 1, c^3 = 1$, and $c^i = 0$ for all other i's, and that the postmodulation filter is a bandpass filter with passbands $[-1.5, -0.5] \cup [3.5, 4.5]$. We can see that this moves spectral contents of Subbands 1 and 3 to frequency bands $[-1.5, -0.5]$ and $[3.5, 4.5]$, respectively, which are alias-free. Therefore, we obtain a two-subband channel with respective channel gains both equal to 2, thus outperforming a single branch of sampling with filtering.

More generally, let us consider the following scenario. Suppose that the channel of bandwidth $W = \frac{2L}{K} f_s$ is equally divided into $2L$ subbands each of bandwidth $f_q = f_s/K$ for some integers K and L. The SNR $|H(f)|^2 / \mathcal{S}_\eta(f)$ within each subband is assumed to be flat. For instance, in the presence of white noise, if $f_q \ll B_c$ with B_c being the coherence bandwidth [23], the channel gain (and hence the SNR) is roughly equal across the subband. Algorithm 1 given below generates an alias-free sampled analog channel, which is achieved by moving the K subbands with the highest SNRs to alias-free locations. By Corollary 1, this algorithm determines an optimal sampling mechanism that maximizes capacity under a single branch of sampling with modulation and filtering. Specifically, take any $f \in [-f_q/2, f_q/2]$, and the algorithm works as follows.

Algorithm 1

---

1. **Initialize.** Find the K largest elements in $\left\{ \frac{|H(f-lf_q)|^2}{S_n(f-lf_q)} \mid l \in \mathbb{Z}, -L \le l \le L-1 \right\}$. Denote by $\{l_i \mid 1 \le i \le K\}$ the index set of these K elements such that $l_1 > l_2 > \cdots > l_K$. Set $L^* := \min \{k \mid k \in \mathbb{Z}, k \ge 2L, k \bmod K = 0\}$.

2.   For $i = 1 : K$

      Let $\alpha := i \cdot L^* + i - l_i$.

      Set $c^\alpha = 1$, and $S(f + \alpha f_p) = 1$.

Algorithm 1 first selects the K subbands with the highest SNR and then moves each of the selected subbands to a new location by appropriately setting $\{c^i\}$, which guarantees that 1) the movement does not corrupt any of the previously chosen locations; and 2) the contents in the newly chosen locations will be alias-free. The postmodulation filter is applied to suppress the frequency contents outside the set of newly chosen subband locations. One drawback of this algorithm is that we need to preserve as many as $2LK$ subbands in order to make it work.

The performance of Algorithm 1 is equivalent to the one using an optimal filter bank followed by sampling with sampling rate f_q at each branch. Hence, single-branch sampling effectively achieves the same performance as multibranch filter-bank sampling. This approach may be preferred since building multiple analog filters is often expensive (in terms of power consumption, size, or cost). We note, however, that for a given overall sampling rate, modulation-bank sampling does not outperform filter-bank sampling with an arbitrary number of branches. The result is formally stated as follows.

Proposition 2: Consider the setup in Theorem 6. For a given overall sampling rate f_s, sampling with M branches of optimal modulation and filter banks does not achieve higher sampled capacity compared to sampling with an optimal bank of aM filters.

Hence, the main advantage of applying a modulation bank is a hardware benefit, namely, using fewer branches and hence less analog circuitry to achieve the same capacity.

VI. CONNECTIONS BETWEEN CAPACITY AND MMSE

In Sections III-D and IV-C, we derived, respectively, the optimal prefilter and the optimal filter bank that maximize capacity. It turns out that such choices of sampling methods coincide with the optimal prefilter/filter bank that minimize the MSE between the Gaussian channel input and the signal reconstructed from sampling the channel output, as detailed below.

Consider the following sampling problem. Let $x(t)$ be a zero-mean wide-sense stationary (WSS) stochastic signal whose PSD $S_X(f)$ satisfies a power constraint[6] $\int_{-\infty}^{\infty} S_X(f)df = P$. This input is passed through a channel consisting of an LTI filter and additive stationary Gaussian noise. We sample the channel output using a filter bank at a fixed rate f_s/M in each branch, and recover a *linear* MMSE estimate $\hat{x}(t)$ of $x(t)$ from its samples in the sense of minimizing $\mathbb{E}(|x(t) - \hat{x}(t)|^2)$ for all sample

---
[6]We restrict our attention to WSS input signals. This restriction, while falling short of generality, allows us to derive sampling results in a simple way.

---

times. We propose to jointly optimize $x(t)$ and the sampling method. Specifically, our joint optimization problem can now be posed as follows: for which input process $x(t)$ and for which filter bank is the estimation error $\mathbb{E}(|x(t) - \hat{x}(t)|^2)$ minimized for all sample times.

It turns out that the optimal input and the optimal filter bank coincide with those maximizing channel capacity, which is captured in the following proposition.

Proposition 3: Suppose the channel input $x(t)$ is any WSS signal. For a given sampling system, let $\hat{x}(t)$ denote the optimal linear estimate of $x(t)$ from the digital sequence $\{\mathbf{y}[n]\}$. Then, the capacity-optimizing filter bank given in (16) and its corresponding optimal input $x(t)$ minimize the linear MSE $\mathbb{E}(|x(t) - \hat{x}(t)|^2)$ over all possible LTI filter banks for any sample time.

    Proof: See Appendix E. ∎

Proposition 3 implies that the input signal and the filter bank optimizing channel capacity also minimize the MSE between the original input signal and its reconstructed output. We note that if the samples $\{\mathbf{y}[n]\}$ and $x(t)$ are jointly Gaussian random variables, then the MMSE estimate $\hat{x}(t)$ for a given input process $x(t)$ is linear in $\{\mathbf{y}[n]\}$. That said, for Gaussian inputs passed through Gaussian channels, the capacity-maximizing filter bank also minimizes the MSE even if we take into account nonlinear estimation. Thus, under sampling with filter-banks for Gaussian channels, information theory reconciles with sampling theory through the SNR metric when determining optimal systems. Intuitively, high SNR typically leads to large capacity and small MSE.

Proposition 3 includes the optimal prefilter under single-prefilter sampling as a special case. We note that a similar MSE minimization problem was investigated decades ago with applications in PAM [17], [18]: a given random input $x(t)$ is prefiltered, corrupted by noise, uniformly sampled, and then postfiltered to yield a linear estimate $\hat{x}(t)$. The goal in that work was to minimize the MSE between $x(t)$ and $\hat{x}(t)$ over all prefiltering (or pulse shaping) and postfiltering mechanisms. While our problem differs from this PAM design problem by optimizing directly over the random input instead of the pulse shape, the two problems are similar in spirit and result in the same alias-suppressing filter. However, earlier work did not account for filter-bank sampling or make connections between minimizing MSE and maximizing capacity.

VII. CONCLUSION AND FUTURE WORK

We have characterized sampled channel capacity as a function of sampling rate for different sampling methods, thereby forming a new connection between sampling theory and information theory. We show how the capacity of a sampled analog channel is affected by reduced sampling rate and identify optimal sampling structures for several classes of sampling methods, which exploit structure in the sampling design. These results also indicate that capacity is not always monotonic in sampling rate, and illuminate an intriguing connection between MIMO channel capacity and capacity of undersampled analog channels. The capacity optimizing sampling structures are shown to extract the frequency components with highest SNRs from each aliased set, and hence suppress aliasing and

out-of-band noise. We also show that for Gaussian inputs over Gaussian channels, the optimal filter/filter bank also minimizes the MSE between the channel input and the reconstructed signal. Our work establishes a framework for using the information-theoretic metric of capacity to optimize sampling structures, offering a different angle from traditional design of sampling methods based on other performance metrics.

Our work uncovers additional questions at the intersection of sampling theory and information theory. For instance, an upper bound on sampled capacity under sampling rate constraints for more general nonuniform sampling methods would allow us to evaluate which sampling mechanisms are capacity-achieving for any channel. Moreover, for channels where there is a gap between achievable rates and the capacity upper bound, these results might provide insight into new sampling mechanisms that might close the gap to capacity. Investigation of capacity under more general nonuniform sampling techniques is an interesting topic that is studied in our companion paper [21]. In addition, the optimal sampling structure for time-varying channels will require different analysis than used in the time-invariant case. It is also interesting to investigate what sampling mechanisms are optimal for channels when the channel state is partially or fully unknown. A deeper understanding of how to exploit channel structure may also guide the design of sampling mechanisms for multiuser channels that require more sophisticated cooperation schemes among users and are impacted in a more complex way by subsampling.

APPENDIX A
PROOF OF THEOREM 2

We begin by an outline of the proof. A discretization argument is first used to approximate arbitrarily well the analog signals by discrete-time signals, which allows us to make use of the properties of Toeplitz matrices instead of the more general Toeplitz operators. By noise whitening, we effectively convert the sampled channel to a MIMO channel with i.i.d. noise for any finite time interval. Finally, the asymptotic properties of Toeplitz matrices are exploited in order to relate the eigenvalue distribution of the equivalent channel matrix with the Fourier representation of both channel filters and prefilters. The proofs of several auxiliary lemmas are deferred to Appendix F.

Instead of directly proving Theorem 2, we prove the theorem for a simpler scenario where the noise $\eta(t)$ is of unit spectral density. In this case, our goal is to prove that the capacity is equivalent to

$$C(f_s) = \frac{1}{2} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \log^+ \left(\nu \frac{\sum_{l \in \mathbb{Z}} |H(f - lf_s)S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2} \right) df$$

where the water level ν can be calculated through the following equation:

$$\int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} \left(\nu - \frac{\sum_l |S(f - lf_s)|^2}{\sum_l |H(f - lf_s)S(f - lf_s)|^2} \right)^+ df = P.$$

This capacity result under white noise can then be immediately extended to accommodate for colored noise. Suppose the additive noise is of PSD $\mathcal{S}_\eta(f)$. We can then split the channel

filter $H(f)$ into two parts with respective frequency response $H(f)/\sqrt{\mathcal{S}_\eta(f)}$ and $\sqrt{\mathcal{S}_\eta(f)}$. Equivalently, the channel input is passed through an LTI filter with frequency response $H(f)/\sqrt{\mathcal{S}_\eta(f)}$, contaminated by white noise, and then passed through a filter with transfer function $\sqrt{\mathcal{S}_\eta(f)}S(f)$ followed by an ideal sampler with rate f_s. This equivalent representation immediately leads to the capacity in the presence of colored noise by substituting corresponding terms into the capacity with white noise.

A) Channel Discretization and Diagonalization: Given that $h(t)$ is continuous and Riemann integrable, one approach to study the continuous-time problem is via reduction to an equivalent discrete-time problem [27, Ch. 16]. In this section, we describe the method of obtaining our equispace discretization approximations to the continuous-time problems, which will allow us to exploit the properties of block-Toeplitz matrices instead of the more complicated block-Toeplitz operators.

For notational simplicity, we define

$$g_{u,v} = \frac{1}{\Delta} \int_0^\Delta g\left(uT_s - v\Delta + \tau\right) d\tau$$

for any function $g(t)$. If $g(t)$ is a continuous function, then $\lim_{\Delta \to 0} g_{u,v} = g\left(uT_s - v\Delta\right)$, where v may be a function of Δ. We also define $\tilde{h}(t) := h(t) * s(t)$. Set $T = nT_s$ and $T_s = k\Delta$ for some integers n and k, and define

$$\tilde{\mathbf{h}}_i := \Delta \cdot \left[\tilde{h}_{i,0}, \tilde{h}_{i,1}, \ldots, \tilde{h}_{i,k-1}\right]$$

$$\mathbf{s}_i := \Delta \cdot \left[s_{i,0}, s_{i,1}, \ldots, s_{i,k-1}\right]$$

$$(\mathbf{x}^n)_i := \frac{1}{\Delta} \int_0^\Delta x\left(i\Delta + \tau\right) d\tau \; (0 \le i < nk)$$

$$(\boldsymbol{\eta})_i := \frac{1}{\Delta} \int_0^\Delta \boldsymbol{\eta}\left(i\Delta + \tau\right) d\tau \; (i \in \mathbb{Z}).$$

We also define

$$\tilde{\mathbf{H}}^n := \begin{bmatrix} \tilde{\mathbf{h}}_0 & \tilde{\mathbf{h}}_{-1} & \cdots & \tilde{\mathbf{h}}_{-n+1} \\ \tilde{\mathbf{h}}_1 & \tilde{\mathbf{h}}_0 & \cdots & \tilde{\mathbf{h}}_{-n+2} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{\mathbf{h}}_{n-1} & \tilde{\mathbf{h}}_{n-2} & \cdots & \tilde{\mathbf{h}}_0 \end{bmatrix}$$

$$\mathbf{S}^n := \begin{bmatrix} \cdots & \mathbf{s}_0 & \mathbf{s}_{-1} & \cdots \\ \cdots & \mathbf{s}_1 & \mathbf{s}_0 & \cdots \\ \cdots & \vdots & \vdots & \cdots \\ \cdots & \mathbf{s}_{n-1} & \mathbf{s}_{n-2} & \cdots \end{bmatrix}.$$

With these definitions, the original channel model can be approximated with the following discretized channel:

$$\mathbf{y}^n = \tilde{\mathbf{H}}^n \mathbf{x}^n + \mathbf{S}^n \boldsymbol{\eta}. \tag{23}$$

As can be seen, $\tilde{\mathbf{H}}^n$ is a fat *block Toeplitz* matrix. Moreover, $\mathbf{S}^n \mathbf{S}^{n*}$ is asymptotically equivalent to a Toeplitz matrix, as will be shown in Appendix A-C. We note that each element η_i is a zero-mean Gaussian variable with variance $\mathbb{E}(|\eta_i|^2) = 1/\Delta$. In addition, $\mathbb{E}(\eta_i \eta_l^*) = 0$ for any $i \ne l$, implying that $\boldsymbol{\eta}$ is an i.i.d. Gaussian vector. The filtered noise $\mathbf{S}^n \boldsymbol{\eta}$ is no longer i.i.d. Gaussian, which motivates us to whiten the noise first.

The prewhitening matrix is given by $\mathbf{S}_{\mathrm{w}}^n := (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}}$, which follows from the fact that

$$\mathbb{E} \mathbf{S}_{\mathrm{w}}^n \mathbf{S}^n \boldsymbol{\eta} (\mathbf{S}_{\mathrm{w}}^n \mathbf{S}^n \boldsymbol{\eta})^* = \mathbf{S}_{\mathrm{w}}^n \mathbf{S}^n \mathbb{E} (\boldsymbol{\eta} \boldsymbol{\eta}^*) \mathbf{S}^{n*} \mathbf{S}_{\mathrm{w}}^{n*}$$
$$= \frac{1}{\Delta} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \mathbf{S}^n \mathbf{S}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} = \frac{1}{\Delta} \mathbf{I}^n.$$

This basically implies that $\mathbf{S}_{\mathrm{w}}^n \mathbf{S}^n$ projects the i.i.d. Gaussian noise η onto an n-dimensional subspace, and that $\mathbf{S}_{\mathrm{w}}^n (\mathbf{S}^n \eta)$ is now n-dimensional i.i.d. Gaussian noise. Left-multiplication with this whitening matrix yields a new output

$$\tilde{\mathbf{y}}^n := (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \left(\tilde{\mathbf{H}}^n \mathbf{x}^n + \mathbf{S}^n \boldsymbol{\eta} \right)$$
$$= (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \mathbf{x}^n + \tilde{\boldsymbol{\eta}}^n.$$

Here, $\tilde{\eta}^n$ consists of independent zero-mean Gaussian elements with variance $1/\Delta$. Since the prewhitening operation $\mathbf{S}_{\mathrm{w}}^n$ is invertible, we have

$$I (\mathbf{x}^n; \tilde{\mathbf{y}}^n) = I (\mathbf{x}^n; \mathbf{y}^n). \tag{24}$$

In this paper, we will use $I_{k,\Delta} (\mathbf{x}^n; \mathbf{y}^n)$ and $I (\mathbf{x}^n; \mathbf{y}^n)$ interchangeably to denote the mutual information between the nk-dimensional vector $\mathbf{x}^n$ and $\mathbf{y}^n$.

Moreover, when $x(t)$ is of bounded variance (i.e., $\sup_t \mathbb{E} |x(t)|^2 \leq \infty$) and the additive noise is Gaussian, it has been shown [28] that the mutual information is weakly continuous in the input distribution. Therefore, $\lim_{k\to\infty} I_{k,\Delta} (\mathbf{x}^n; \mathbf{y}^n) \to I \left(\{x(t)\}_{t=0}^T ; \{\mathbf{y}[n]\}_{t=0}^T \right)$. As k increases, the discretized sequence becomes a finer approximation to the continuous-time signal. The uniform continuity of the probability measure of $x(t)$ and the continuity of mutual information immediately imply that $\lim_{n\to\infty} \frac{1}{nT_s} I_{k,\Delta} (\mathbf{x}^n; \mathbf{y}^n)$ converges uniformly in k. We also observe that for every given n, $\lim_{k\to\infty} I_{k,\Delta} (\mathbf{x}^n; \mathbf{y}^n)$ exists due to the continuity condition of the mutual information. Therefore, applying the Moore–Osgood theorem in real analysis allows us to exchange the order of limits.

Based on the above arguments, the capacity of the sampled analog channel can be expressed as the following limit:

$$C(f_s) = \lim_{k\to\infty} \lim_{n\to\infty} \frac{1}{nT_s} \sup_{p(x): \frac{1}{nk} \mathbb{E} \left(\|\mathbf{x}^n\|_2^2 \right) \leq P} I_{k,\Delta} (\mathbf{x}^n, \mathbf{y}^n)$$
$$= \lim_{k\to\infty} \lim_{n\to\infty} \frac{f_s}{n} \sup_{p(x): \frac{1}{nk} \mathbb{E} \left(\|\mathbf{x}^n\|_2^2 \right) \leq P} I_{k,\Delta} (\mathbf{x}^n, \tilde{\mathbf{y}}^n).$$

Note that it suffices to investigate the case where T is an integer multiple of T_s since $\lim_{T\to\infty} \frac{1}{T} \sup I (x(0,T]; \{y[n]\}) = \lim_{n\to\infty} \frac{1}{nT_s} \sup I (x(0, nT_s]; \{y[n]\})$.

B) Preliminaries on Toeplitz Matrices: Before proceeding to the proof of the theorem, we briefly introduce several basic definitions and properties related to Toeplitz matrices. Interested readers are referred to [3] and [24] for more details.

A Toeplitz matrix is an $n \times n$ matrix $\mathbf{T}^n$ where $(\mathbf{T}^n)_{k,l} = t_{k-l}$, which implies that a Toeplitz matrix $\mathbf{T}^n$ is uniquely defined by the sequence $\{t_k\}$. A special case of Toeplitz matrices is circulant matrices where every row of the matrix $\mathbf{C}^n$ is a right cyclic shift of the row above it. The Fourier

series (or symbol) with respect to the sequence of Toeplitz matrices $\{\mathbf{T}^n := [t_{k-l}; k, l = 0, 1, \ldots, n-1] : n \in \mathbb{Z}\}$ is given by

$$F(\omega) = \sum_{k=-\infty}^{+\infty} t_k \exp (jk\omega), \quad \omega \in [-\pi, \pi]. \tag{25}$$

Since the sequence $\{t_k\}$ uniquely determines $F(\omega)$ and vice versa, we denote by $\mathbf{T}^n(F)$ the Toeplitz matrix generated by F (and hence $\{t_k\}$). We also define a related circulant matrix $\mathbf{C}^n(F)$ with top row $(c_0^{(n)}, c_1^{(n)}, \ldots, c_{n-1}^{(n)})$, where

$$c_k^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} F \left(\frac{2\pi i}{n} \right) \exp \left(\frac{2\pi j i k}{n} \right). \tag{26}$$

One key concept in our proof is asymptotic equivalence, which is formally defined as follows [24].

Definition 1 (Asymptotic Equivalence): Two sequences of $n \times n$ matrices $\{\mathbf{A}^n\}$ and $\{\mathbf{B}^n\}$ are said to be asymptotically equivalent if
1) $\mathbf{A}^n$ and $\mathbf{B}^n$ are uniformly bounded, i.e., there exists a constant c independent of n such that

$$\|\mathbf{A}^n\|_2, \|\mathbf{B}^n\|_2 \leq c < \infty, \quad n = 1, 2, \cdots \tag{27}$$

2) $\lim_{n\to\infty} \frac{1}{\sqrt{n}} \|\mathbf{A}^n - \mathbf{B}^n\|_{\mathrm{F}} = 0.$
We will abbreviate asymptotic equivalence of $\{\mathbf{A}^n\}$ and $\{\mathbf{B}^n\}$ by $\mathbf{A}^n \sim \mathbf{B}^n$. Two important results regarding asymptotic equivalence are given in the following lemmas [24].

Lemma 1: Suppose $\mathbf{A}^n \sim \mathbf{B}^n$ with eigenvalues $\{\alpha_{n,k}\}$ and $\{\beta_{n,k}\}$, respectively. Let $g(x)$ be an arbitrary continuous function. Then, if the limits $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\alpha_{n,k})$ and $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\beta_{n,k})$ exist, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\alpha_{n,k}) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\beta_{n,k}). \tag{28}$$

Lemma 2:
a) Suppose a sequence of Toeplitz matrices $\mathbf{T}^n$ where $(\mathbf{T}^n)_{ij} = t_{i-j}$ satisfies that $\{t_i\}$ is absolutely summable. Suppose the Fourier series $F(\omega)$ related to $\mathbf{T}^n$ is positive and $\mathbf{T}^n$ is Hermitian. Then, we have

$$\mathbf{T}^n(F) \sim \mathbf{C}^n(F). \tag{29}$$

If we further assume that there exists a constant $\epsilon > 0$ such that $F(\omega) \geq \epsilon > 0$ for all $\omega \in [0, 2\pi]$, then we have

$$\mathbf{T}^n(F)^{-1} \sim \mathbf{C}^n(F)^{-1} = \mathbf{C}^n(1/F) \sim \mathbf{T}^n(1/F). \tag{30}$$

b) Suppose $\mathbf{A}^n \sim \mathbf{B}^n$ and $\mathbf{C}^n \sim \mathbf{D}^n$, then $\mathbf{A}^n \mathbf{C}^n \sim \mathbf{B}^n \mathbf{D}^n$.

Toeplitz or block Toeplitz matrices have well-known asymptotic spectral properties [3], [29]. The notion of asymptotic equivalence allows us to approximate non-Toeplitz matrices by Toeplitz matrices, which will be used in the next section to analyze the spectral properties of the channel matrix.

C) Capacity via Convergence of the Discrete Model: After channel discretization, we can calculate the capacity for each finite duration T using well-known MIMO Gaussian channel

capacity, which, however, depends on the spectrum of the truncated channel and may vary dramatically for different T. By our definition of capacity, we will pass T to infinity and see whether the finite-duration capacity converges, and if so, whether there is a closed-form expression for the limit. Fortunately, the beautiful asymptotic properties of block-Toeplitz matrices guarantees the existence of the limit and allows for a closed-form solution using the frequency response of $h(t)$ and $s(t)$.

To see this, we first construct a new channel whose capacity is easier to obtain, and will show that the new channel has asymptotically equivalent channel capacity as the original channel. As detailed below, each key matrix associated with the new channel is a Toeplitz matrix, whose spectrum can be well approximated in the asymptotic regime [24].

Consider the spectral properties of the Hermitian matrices $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ and $\mathbf{S}^n \mathbf{S}^{n*}$. We can see that

$$\left(\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}\right)_{ij} = \left(\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}\right)_{ji}^* = \sum_{t=-j+1}^{n-j} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^*. \quad (31)$$

Obviously, $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ is not a Toeplitz matrix. Instead of investigating the eigenvalue distribution of $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ directly, we look at a new *Hermitian Toeplitz* matrix $\hat{\mathbf{H}}^n$ associated with $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ such that for any $i \leq j$:

$$\left(\hat{\mathbf{H}}^n\right)_{ij} = \left(\hat{\mathbf{H}}^n\right)_{ij}^* = \sum_{t=-\infty}^{\infty} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^*. \quad (32)$$

Lemma 3: The above definition of $\hat{\mathbf{H}}^n$ implies that

$$\hat{\mathbf{H}}^n \sim \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}. \quad (33)$$

Proof: See Appendix F-A. ∎

On the other hand, for any $1 \leq i \leq j \leq n$, we have

$$(\mathbf{S}^n \mathbf{S}^{n*})_{ij} = (\mathbf{S}^n \mathbf{S}^{n*})_{ji}^* = \sum_{t=-\infty}^{\infty} \mathbf{s}_{j-i+t} \mathbf{s}_t^*. \quad (34)$$

Hence, the Hermitian matrix $\hat{\mathbf{S}}^n := \mathbf{S}^n \mathbf{S}^{n*}$ is still Toeplitz. However, the matrix of interest in the capacity will be $(\mathbf{S}^n \mathbf{S}^{n*})^{-1/2}$ instead. We, therefore, construct an asymptotically equivalent circulant matrix $\mathbf{C}^n$ as defined in (26), which will preserve the Toeplitz property when we take $(\mathbf{C}^n)^{-1/2}$ [24]. Formally speaking, $(\mathbf{S}^n \mathbf{S}^{n*})^{-1}$ can be related to $(\mathbf{C}^n)^{-1}$ as follows.

Lemma 4: If there exists some constant $\epsilon_s > 0$ such that for all $f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$,

$$\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2 \geq \epsilon_s > 0 \quad (35)$$

holds, then $(\mathbf{C}^n)^{-1} \sim (\mathbf{S}^n \mathbf{S}^{n*})^{-1}$.

Proof: See Appendix B. ∎

One of the most useful properties of a circulant matrix $\mathbf{C}^n$ is that its eigenvectors $\left\{ \mathbf{u}_c^{(m)} \right\}$ are

$$\mathbf{u}_c^{(m)} = \frac{1}{\sqrt{n}} \left(1, e^{-2\pi jm/n}, \ldots, e^{-2\pi j(n-1)m/n}\right). \quad (36)$$

Suppose the eigenvalue decomposition of $\mathbf{C}^n$ is given as

$$\mathbf{C}^n = \mathbf{U}_c \mathbf{\Lambda}_c \mathbf{U}_c^*. \quad (37)$$

where $\mathbf{U}_c$ is a Fourier coefficient matrix, and $\mathbf{\Lambda}_c$ is a diagonal matrix where each element in the diagonal is positive.

The concept of asymptotic equivalence allows us to explicitly relate our matrices of interest to both circulant matrices and Toeplitz matrices, whose asymptotic spectral densities have been well studied.

Lemma 5: For any continuous function $g(x)$, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(\lambda_i)$$
$$= T_s \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} g \left(\frac{\sum_{l \in \mathbb{Z}} |H(f - lf_s)S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2} \right) df,$$

where λ_i denotes the ith eigenvalue of $(\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}}$.

Proof: See Appendix F-C. ∎

We can now prove the capacity result. The standard capacity results for parallel channels [1, Th. 7.5.1] imply that the capacity of the discretized sampled analog channel is given by the parametric equations

$$C_T = \frac{1}{2T} \sum_i \log^+ (\nu \lambda_i), \quad (38)$$

$$\frac{Pnk}{1/\Delta} = \sum_i [\nu - 1/\lambda_i]^+, \quad (39)$$

where ν is the water level of the optimal power allocation over this discrete model, as can be calculated through (39). Since this capacity depends on the eigenvalues of $(\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}}$, then by Lemma 5, the convergence as $T \to \infty$ is guaranteed and the capacity $C = \lim_{T \to \infty} C_T$ can be expressed using $H(f)$ and $S(f)$. Specifically,

$$\lim_{T \to \infty} C_T(\nu) = \lim_{T \to \infty} \frac{1}{T} \sum_i \frac{1}{2} \log^+ [\nu \lambda_i]$$
$$= \frac{1}{2} \int_{-f_s/2}^{f_s/2} \log^+ \left(\nu \frac{\sum_{l \in \mathbb{Z}} |H(f - lf_s)S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2} \right) df.$$

Similarly, (39) can be transformed into

$$PT_s = \frac{Pk}{1/\Delta} = \frac{1}{n} \sum_i [\nu - 1/\lambda_i]^+$$
$$= T_s \int_{-f_s/2}^{f_s/2} \left[\nu - \frac{\sum_{l \in \mathbb{Z}} |S(f - lf_s)|^2}{\sum_{l \in \mathbb{Z}} |H(f - lf_s)S(f - lf_s)|^2} \right]^+ df,$$

which completes the proof.

APPENDIX B
PROOF OF THEOREM 4

We follow similar steps as in the proof of Theorem 2: we approximate the sampled channel using a discretized model

first, whiten the noise, and then find capacity of the equivalent channel matrix. Due to the use of filter banks, the equivalent channel matrix is no longer asymptotically equivalent to a Toeplitz matrix, but instead a block-Toeplitz matrix. This motivates us to exploit the asymptotic properties of block-Toeplitz matrices.

A) Channel Discretization and Diagonalization: Let $\tilde{T}_s = MT_s$, and suppose we have $T = n\tilde{T}_s$ and $\tilde{T}_s = k\Delta$ with integers n and k. Similarly, we can define

$$\tilde{h}_i(t) := h(t) * s_i(t),$$
$$\tilde{\mathbf{h}}_i^l = \left[\tilde{h}_i\left(l\tilde{T}_s\right), \tilde{h}_i\left(l\tilde{T}_s - \Delta\right), \ldots, \tilde{h}_i\left(l\tilde{T}_s - (k-1)\Delta\right)\right].$$

We introduce the following two matrices as

$$\tilde{\mathbf{H}}_i^n = \begin{bmatrix} \tilde{\mathbf{h}}_i^0 & \tilde{\mathbf{h}}_i^{-1} & \cdots & \tilde{\mathbf{h}}_i^{-n+1} \\ \tilde{\mathbf{h}}_i^1 & \tilde{\mathbf{h}}_i^0 & \cdots & \tilde{\mathbf{h}}_i^{-n+2} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{h}}_i^{n-1} & \tilde{\mathbf{h}}_i^{n-2} & \cdots & \tilde{\mathbf{h}}_i^0 \end{bmatrix}$$

$$\text{and} \quad \mathbf{S}_i^n = \begin{bmatrix} \cdots & \mathbf{s}_i^0 & \mathbf{s}_i^{-1} & \cdots \\ \cdots & \mathbf{s}_i^1 & \mathbf{s}_i^0 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \mathbf{s}_i^{n-1} & \mathbf{s}_i^{n-2} & \cdots \end{bmatrix}.$$

We also set

$$(\mathbf{x}^n)_i = \frac{1}{\Delta} \int_0^\Delta x\left(i\Delta + \tau\right) d\tau \ (0 \le i < nk)$$

and

$$(\boldsymbol{\eta})_i = \frac{1}{\Delta} \int_0^\Delta \boldsymbol{\eta}$$

$(i\Delta + \tau) d\tau \quad (i \in \mathbb{Z})$. Defining $\mathbf{y}^n = \left[y_1[0], \ldots, y_1[n-1], y_2[0], \ldots, y_2[n-1], \ldots, y_M[n-1] \right]^T$ leads to the discretized channel model

$$\mathbf{y}^n = \begin{bmatrix} \tilde{\mathbf{H}}_1^n \\ \tilde{\mathbf{H}}_2^n \\ \vdots \\ \tilde{\mathbf{H}}_M^n \end{bmatrix} \mathbf{x}^n + \begin{bmatrix} \mathbf{S}_1^n \\ \mathbf{S}_2^n \\ \vdots \\ \mathbf{S}_M^n \end{bmatrix} \boldsymbol{\eta}.$$

Whitening the noise gives us

$$\tilde{\mathbf{y}}^n = \left(\begin{bmatrix} \mathbf{S}_1^n \\ \mathbf{S}_2^n \\ \vdots \\ \mathbf{S}_M^n \end{bmatrix} \begin{bmatrix} \mathbf{S}_1^{n*} & \cdots & \mathbf{S}_M^{n*} \end{bmatrix} \right)^{-\frac{1}{2}} \begin{bmatrix} \tilde{\mathbf{H}}_1^n \\ \tilde{\mathbf{H}}_2^n \\ \vdots \\ \tilde{\mathbf{H}}_M^n \end{bmatrix} \mathbf{x}_n + \tilde{\boldsymbol{\eta}}$$

where $\tilde{\boldsymbol{\eta}}$ is i.i.d. Gaussian variable with variance $1/\Delta$. We can express capacity of the sampled analog channel under filter-bank sampling as the following limit:

$$C(f_s) = \lim_{k \to \infty} \lim_{n \to \infty} \frac{f_s}{Mn} \sup I\left(\mathbf{x}^n; \tilde{\mathbf{y}}^n\right).$$

Here, the supremum is taken over all distribution of $\mathbf{x}^n$ subject to a power constraint $\frac{1}{nk}\mathbb{E}\left(\|x_n\|_2^2\right) \le P$.

B) Capacity via Convergence of the Discrete Model: We can see that for any $1 \le u, v \le m$,

$$\mathbf{S}_u^n \mathbf{S}_v^{n*} = \tilde{\mathbf{S}}_{u,v}^n \tag{40}$$

where the Toeplitz matrix $\tilde{\mathbf{S}}_{u,v}^n$ is defined such that for any $1 \le i \le j \le n$

$$\left(\tilde{\mathbf{S}}_{u,v}^n\right)_{i,j} = \sum_{t=-\infty}^{\infty} \mathbf{s}_u^{j-i+t}\left(\mathbf{s}_v^t\right)^*. \tag{41}$$

Let $\mathbf{S}^n = [\mathbf{S}_1^{n*}, \mathbf{S}_2^{n*}, \ldots, \mathbf{S}_M^{n*}]^*$. Then, the Hermitian block Toeplitz matrix

$$\tilde{\mathbf{S}}^n := \begin{bmatrix} \tilde{\mathbf{S}}_{1,1}^n & \tilde{\mathbf{S}}_{1,2}^n & \cdots & \tilde{\mathbf{S}}_{1,M}^n \\ \tilde{\mathbf{S}}_{2,1}^n & \tilde{\mathbf{S}}_{2,2}^n & \cdots & \tilde{\mathbf{S}}_{2,M}^n \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{S}}_{M,1}^n & \tilde{\mathbf{S}}_{M,2}^n & \cdots & \tilde{\mathbf{S}}_{M,M}^n \end{bmatrix}$$

satisfies $\tilde{\mathbf{S}}^n = \mathbf{S}^n \mathbf{S}^{n*}$. Additionally, we define $\hat{\mathbf{H}}_{u,v}^n \ (1 \le u, v \le M)$, where

$$\left(\hat{\mathbf{H}}_{u,v}^n\right)_{i,j} = \sum_{t=-\infty}^{\infty} \tilde{\mathbf{h}}_u^{j-i+t}\left(\tilde{\mathbf{h}}_v^t\right)^* \tag{42}$$

and we let $\tilde{\mathbf{H}}^n = \left[\tilde{\mathbf{H}}_1^{n*}, \tilde{\mathbf{H}}_2^{n*}, \cdots, \tilde{\mathbf{H}}_M^{n*}\right]^*$. The block Toeplitz matrix

$$\hat{\mathbf{H}}^n := \begin{bmatrix} \hat{\mathbf{H}}_{1,1}^n & \hat{\mathbf{H}}_{1,2}^n & \cdots & \hat{\mathbf{H}}_{1,M}^n \\ \hat{\mathbf{H}}_{2,1}^n & \hat{\mathbf{H}}_{2,2}^n & \cdots & \hat{\mathbf{H}}_{2,M}^n \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\mathbf{H}}_{M,1}^n & \hat{\mathbf{H}}_{M,2}^n & \cdots & \hat{\mathbf{H}}_{M,M}^n \end{bmatrix}$$

satisfies

$$\lim_{n \to \infty} \frac{1}{\sqrt{nM}} \left\|\hat{\mathbf{H}}^n - \tilde{\mathbf{H}}^n\tilde{\mathbf{H}}^{n*}\right\|_{\mathrm{F}}$$
$$\le \lim_{n \to \infty} \frac{1}{\sqrt{M}} \sum_{1 \le u,v \le M} \frac{1}{\sqrt{n}} \left\|\hat{\mathbf{H}}_{u,v}^n - \tilde{\mathbf{H}}_u^n\tilde{\mathbf{H}}_v^{n*}\right\| = 0.$$
$$\implies \hat{\mathbf{H}}^n \sim \tilde{\mathbf{H}}^n\tilde{\mathbf{H}}^{n*}. \tag{43}$$

The $M \times M$ Fourier symbol matrix $\mathbf{F}_{\tilde{s}}(f)$ associated with $\tilde{\mathbf{S}}^n$ has elements $[\mathbf{F}_{\tilde{s}}(f)]_{u,v}$ given by

$$[\mathbf{F}_{\tilde{s}}(f)]_{u,v}$$
$$= \frac{\Delta^2}{\tilde{T}_s^2} \sum_{i=0}^{k-1} \left(\sum_{l_1} S_u\left(-f + l_1\tilde{f}_s\right) \exp\left(-j2\pi\left(f - l_1\tilde{f}_s\right)i\Delta\right)\right)$$
$$\left(\sum_{l_2} S_v\left(-f + l_2\tilde{f}_s\right) \exp\left(-j2\pi\left(f - l_2\tilde{f}_s\right)i\Delta\right)\right)^*$$
$$= \frac{\Delta^2}{\tilde{T}_s^2} \sum_{i=0}^{k-1} \left(\sum_{l_1,l_2} S_u\left(-f + l_1\tilde{f}_s\right) S_v^*\left(-f + l_2\tilde{f}_s\right)\right.$$
$$\left. \exp\left(-j2\pi(l_2 - l_1)\tilde{f}_s i\Delta\right)\right)$$
$$= \frac{\Delta}{\tilde{T}_s} \sum_{l \in \mathbb{Z}} S_u\left(-f + l\tilde{f}_s\right) S_v^*\left(-f + l\tilde{f}_s\right).$$

Denote by $\left\{ \mathbf{T}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right) \right\}$ the sequence of block Toeplitz matrices generated by $\mathbf{F}_{\tilde{s}}^{-1}(f)$, and denote by $\mathbf{T}_{l_1,l_2}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right)$ the (l_1, l_2) Toeplitz block of $\mathbf{T}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right)$. It can be verified that

$$\sum_{l_2=1}^{M} \mathbf{T}_{l_1,l_2}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right) \cdot \tilde{\mathbf{S}}_{l_2,l_3}^n \sim \mathbf{T}_n \left(\sum_{l_2=1}^{M} \left[\mathbf{F}_{\tilde{s}}^{-1} \right]_{l_1,l_2} \left[\mathbf{F}_{\tilde{s}} \right]_{l_2,l_3} \right)$$
$$= \mathbf{T}^n \left(\delta[l_1 - l_3] \right),$$

which immediately yields

$$\mathbf{T}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right) \tilde{\mathbf{S}}^n \sim \mathbf{I} \Longrightarrow \mathbf{T}^n \left(\mathbf{F}_{\tilde{s}}^{-1} \right) \sim \left(\tilde{\mathbf{S}}^n \right)^{-1} \sim \left(\mathbf{S}^n \mathbf{S}^{n*} \right)^{-1}.$$
(44)

Therefore, for any continuous function $g(x)$, [29, Th. 5.4] implies that

$$\lim_{n \to \infty} \frac{1}{nM} \sum_{i=1}^{nM} g \left(\lambda_i \left\{ \left(\mathbf{S}^n \mathbf{S}^{n*} \right)^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \left(\mathbf{S}^n \mathbf{S}^{n*} \right)^{-\frac{1}{2}} \right\} \right)$$
$$= \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \sum_{i=1}^{M} g \left(\lambda_i \left(\mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \mathbf{F}_{\tilde{h}} \mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \right) \right) df.$$

Denote $\mathbf{F}_s^{\ddagger} = \left(\mathbf{F}_s \mathbf{F}_s^* \right)^{-\frac{1}{2}} \mathbf{F}_s$; then, the capacity of parallel channels [1], which is achieved via water filling power allocation, yields

$$C(f_s)$$
$$= \lim_{n \to \infty} \frac{\sum_{i=1}^{nM} \log^+ \left(\nu \lambda_i \left\{ \left(\mathbf{S}^n \mathbf{S}^{n*} \right)^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \left(\mathbf{S}^n \mathbf{S}^{n*} \right)^{-\frac{1}{2}} \right\} \right)}{2nMT_s}$$
$$= \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \frac{1}{2} \sum_{i=1}^{M} \log^+ \left(\nu \lambda_i \left(\mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \mathbf{F}_{\tilde{h}} \mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \right) \right) df$$
$$= \frac{1}{2} \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \sum_{i=1}^{M} \log^+ \left(\nu \lambda_i \left(\mathbf{F}_s^{\ddagger} \mathbf{F}_h \mathbf{F}_h^* \mathbf{F}_s^{\ddagger *} \right) \right) df,$$

where

$$P = \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \sum_{i=1}^{M} \left[\nu - \frac{1}{\lambda_i \left(\mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \mathbf{F}_{\tilde{h}} \mathbf{F}_{\tilde{s}}^{-\frac{1}{2}} \right)} \right]^+ df$$
$$= \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \sum_{i=1}^{M} \left[\nu - \frac{1}{\lambda_i \mathbf{F}_s^{\ddagger} \mathbf{F}_h \mathbf{F}_h^* \mathbf{F}_s^{\ddagger *}} \right]^+ df.$$

This completes the proof.

APPENDIX C
PROOF OF THEOREM 5

Theorem 5 immediately follows from the following proposition.

Proposition 4: The kth largest eigenvalue λ_k of the positive semidefinite matrix $\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*$ is bounded by

$$0 \leq \lambda_k \leq \lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right), \quad 1 \leq k \leq M. \tag{45}$$

These upper bounds can be attained simultaneously by the filter (16).

Proof: Recall that at a given f, $\mathbf{F}_h$ is an infinite diagonal matrix satisfying $(\mathbf{F}_h)_{l,l} = H \left(f - \frac{l f_s}{M} \right)$ for all $l \in \mathbb{Z}$, and that $\tilde{\mathbf{F}}_s = \left(\mathbf{F}_s \mathbf{F}_s^* \right)^{-\frac{1}{2}} \mathbf{F}_s$. Hence, $\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*$ is an $M \times M$ dimensional matrix. We observe that

$$\tilde{\mathbf{F}}_s \left(\tilde{\mathbf{F}}_s \right)^* = \left(\mathbf{F}_s \mathbf{F}_s^* \right)^{-\frac{1}{2}} \mathbf{F}_s \mathbf{F}_s^* \left(\mathbf{F}_s \mathbf{F}_s^* \right)^{-\frac{1}{2}} = \mathbf{I} \tag{46}$$

which indicates that the rows of $\tilde{\mathbf{F}}_s$ are orthonormal. Hence, the operator norm of $\tilde{\mathbf{F}}_s$ is no larger than 1, which leads to

$$\lambda_1 \left(\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \right) = \left\| \tilde{\mathbf{F}}_s \mathbf{F}_h \right\|_2^2 \leq \left\| \mathbf{F}_h \right\|_2^2 = \lambda_1 \left(\mathbf{F}_h \mathbf{F}_h^* \right).$$

Denote by $\{ \mathbf{e}_k, \ k \geq 1 \}$ the standard basis where $\mathbf{e}_k$ is a vector with a 1 in the kth coordinate and 0 otherwise. We introduce the index set $\{ i_1, i_2, \ldots, i_M \}$ such that $\mathbf{e}_{i_k}$ $(1 \leq k \leq M)$ is the eigenvector associated with the kth largest eigenvalues of the diagonal matrix $\mathbf{F}_h \mathbf{F}_h^*$.

Suppose that $\mathbf{v}_k$ is the eigenvector associated with the kth largest eigenvalue λ_k of $\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*$, and denote by $\left(\tilde{\mathbf{F}}_s \right)_k$ the kth *column* of $\tilde{\mathbf{F}}_s$. Since $\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^*$ is Hermitian positive semidefinite, its eigendecomposition yields an orthogonal basis of eigenvectors. Observe that $\{ \mathbf{v}_1, \ldots, \mathbf{v}_k \}$ spans a k-dimensional space and that $\left\{ \left(\tilde{\mathbf{F}}_s \right)_j, 1 \leq j \leq k - 1 \right\}$ spans a subspace of dimension no more than $k - 1$. For any $k \geq 2$, there exists k scalars $a_1, \ldots, a_k$ such that

$$\sum_{i=1}^{k} a_i \mathbf{v}_i \perp \left\{ \left(\tilde{\mathbf{F}}_s \right)_{i_j}, 1 \leq j \leq k - 1 \right\} \text{ and } \sum_{i=1}^{k} a_i \mathbf{v}_i \neq 0. \tag{47}$$

This allows us to define the following unit vector:

$$\tilde{\mathbf{v}}_k \overset{\triangle}{=} \sum_{i=1}^{k} \frac{a_i}{\sqrt{\sum_{j=1}^{k} |a_j|^2}} \mathbf{v}_i \tag{48}$$

which is orthogonal to $\left\{ \left(\tilde{\mathbf{F}}_s \right)_j, 1 \leq j \leq k - 1 \right\}$. We observe that

$$\left\| \tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \tilde{\mathbf{v}}_k \right\|_2^2 = \left\| \sum_{i=1}^{k} \frac{a_i}{\sqrt{\sum_{j=1}^{k} |a_j|^2}} \tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \mathbf{v}_i \right\|_2^2$$
$$= \left\| \sum_{i=1}^{k} \frac{a_i \lambda_i}{\sqrt{\sum_{j=1}^{k} |a_j|^2}} \mathbf{v}_i \right\|_2^2$$
$$= \sum_{i=1}^{k} \frac{\lambda_i^2 |a_i|^2}{\sum_{j=1}^{k} |a_j|^2} \geq \lambda_k^2. \tag{49}$$

Define $\mathbf{u}_k := \tilde{\mathbf{F}}_s^* \tilde{\mathbf{v}}_k$. From (47), we can see that $(\mathbf{u}_k)_i = \left\langle \left(\tilde{\mathbf{F}}_s \right)_i, \tilde{\mathbf{v}}_i \right\rangle = 0$ holds for all $i \in \{i_1, i_2, \ldots, i_{k-1}\}$. In other words, $\mathbf{u}_k \perp \{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{k-1}}\}$. This further implies that

$$\lambda_k^2 \leq \left\| \tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \tilde{\mathbf{v}}_k \right\|_2^2 \leq \left\| \tilde{\mathbf{F}}_s \right\|_2^2 \left\| \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \tilde{\mathbf{v}}_k \right\|_2^2$$

$$\leq \| \mathbf{F}_h \mathbf{F}_h^* \mathbf{u}_k \|_2^2 \tag{50}$$

$$\leq \sup_{\mathbf{x} \perp \text{span}\{\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_{k-1}}\}} \| \mathbf{F}_h \mathbf{F}_h^* \mathbf{x} \|_2^2 \tag{51}$$

$$= \lambda_k^2 \left(\mathbf{F}_h \mathbf{F}_h^* \right) \tag{52}$$

by observing that $\mathbf{F}_h \mathbf{F}_h^*$ is a diagonal matrix.

Setting

$$S_k \left(f - \frac{l f_s}{M} \right)$$
$$= \begin{cases} 1, & \text{if } \left| H \left(f - \frac{l f_s}{M} \right) \right|^2 = \lambda_k \left(\mathbf{F}_h(f) \mathbf{F}_h^*(f) \right) \\ 0, & \text{otherwise,} \end{cases}$$

yields $\tilde{\mathbf{F}}_s = \mathbf{F}_s$, and hence, $\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s$ is a diagonal matrix such that

$$\left(\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \right)_{k,k} = \lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right). \tag{53}$$

Apparently, this choice of $S_k(f)$ allows the upper bounds

$$\lambda_k \left(\tilde{\mathbf{F}}_s \mathbf{F}_h \mathbf{F}_h^* \tilde{\mathbf{F}}_s^* \right) = \lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right), \quad \forall 1 \leq k \leq M \tag{54}$$

to be attained simultaneously. ∎

By extracting out the M frequencies with the highest SNR from each aliased set $\{f - l f_s / M \mid l \in \mathbb{Z}\}$, we achieve $\lambda_k = \lambda_k \left(\mathbf{F}_h \mathbf{F}_h^* \right)$, thus achieving the maximum capacity.

APPENDIX D
PROOF OF THEOREM 6

Following similar steps as in the proof of Theorem 4, we approximately convert the sampled channel into its discrete counterpart and calculate the capacity of the discretized channel model after noise whitening. We note that the impulse response of the sampled channel is no longer LTI due to the use of modulation banks. But the periodicity assumption of the modulation sequences allows us to treat the channel matrix as blockwise LTI, which provides a way to exploit the properties of block-Toeplitz matrices.

Again, we give a proof for the scenario where noise is white Gaussian with unit spectral density. The capacity expression in the presence of colored noise can immediately be derived by replacing $P_i(f)$ with $P_i(f)\sqrt{\mathcal{S}_\eta(f)}$ and $H(f)$ with $H(f)/\sqrt{\mathcal{S}_\eta(f)}$.

In the ith branch, the noise component at time t is given by

$$s_i(t) * (q_i(t) \cdot (p_i(t) * \eta(t)))$$
$$= \int_{\tau_1} d\tau_1 s_i(t - \tau_1) \int_{\tau_2} q_i(\tau_1) p_i(\tau_1 - \tau_2) \eta(\tau_2) d\tau_2$$
$$= \int_{\tau_2} \left(\int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau_2) d\tau_1 \right) \eta(\tau_2) d\tau_2$$
$$= \int_{\tau_2} g_i^\eta(t, \tau_2) \eta(\tau_2) d\tau_2,$$

where $g_i^\eta(t, \tau_2) \stackrel{\triangle}{=} \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau_2) d\tau_1$.

Let $\tilde{T}_s = M T_s$. Our assumption $b T_q = a \tilde{T}_s$ immediately leads to

$$g_i^\eta \left(t + a \tilde{T}_s, \tau + b T_q \right)$$
$$= \int_{\tau_1} s_i \left(t + a \tilde{T}_s - \tau_1 \right) q_i(\tau_1) p_i \left(\tau_1 - \tau - a \tilde{T}_s \right) d\tau_1$$
$$= \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1 + b T_q) p_i(\tau_1 - \tau) d\tau_1$$
$$= \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau) d\tau_1 = g_i^\eta(t, \tau),$$

implying that $g_i^\eta(t, \tau)$ is a block-Toeplitz function.

Similarly, the signal component

$$s_i(t) * (q_i(t) \cdot (p_i(t) * h(t) * x(t))) = \int_{\tau_2} g_i^h(t, \tau_2) x(\tau_2) d\tau_2$$

where

$$g_i^h(t, \tau_2) \stackrel{\triangle}{=} \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) \int_{\tau_3} p_i(\tau_1 - \tau_2 - \tau_3) h(\tau_3) d\tau_3 d\tau_1$$

which also satisfies the block-Toeplitz property $g_i^h \left(t + a \tilde{T}_s, \tau + b T_q \right) = g_i^h(t, \tau)$.

Suppose that $T = n \tilde{T}_s$ and $\tilde{T}_s = k \Delta$ hold for some integers n and k. We can introduce two matrices $\mathbf{G}_i^\eta$ and $\mathbf{G}_i^h$ such that $\forall m \in \mathbb{Z}, 0 \leq l < n$

$$\begin{cases} (\mathbf{G}_i^\eta)_{l,m} = g_i^\eta \left(l \tilde{T}_s, m \Delta \right) \\ (\mathbf{G}_i^h)_{l,m} = g_i^h \left(l \tilde{T}_s, m \Delta \right). \end{cases}$$

Setting $\mathbf{y}_i^n = [y_i[0], y_i[1], \ldots, y_i[n-1]]^T$ leads to similar discretized approximation as in the proof of Theorem 2:

$$\mathbf{y}_i^n = \mathbf{G}_i^h \mathbf{x}^n + \mathbf{G}_i^\eta \boldsymbol{\eta}. \tag{55}$$

Here, $\boldsymbol{\eta}$ is a i.i.d. zero-mean Gaussian vector where each entry is of variance $1/\Delta$.

Hence, $\mathbf{G}_i^h$ and $\mathbf{G}_i^\eta$ are block Toeplitz matrices satisfying $(\mathbf{G}_i^h)_{l+a, m+ak} = (\mathbf{G}_i^h)_{l,m}$ and $(\mathbf{G}_i^\eta)_{l+a, m+ak} = (\mathbf{G}_i^\eta)_{l,m}$.

Using the same definition of $\mathbf{x}^n$ and $\boldsymbol{\eta}$ as in Appendix B, we can express the system equation as

$$\mathbf{y}^n = \begin{bmatrix} \mathbf{G}_1^h \\ \mathbf{G}_2^h \\ \vdots \\ \mathbf{G}_M^h \end{bmatrix} \mathbf{x}^n + \begin{bmatrix} \mathbf{G}_1^\eta \\ \mathbf{G}_2^\eta \\ \vdots \\ \mathbf{G}_M^\eta \end{bmatrix} \boldsymbol{\eta}. \tag{56}$$

Whitening the noise component yields

$$\tilde{\mathbf{y}}_n = \left(\begin{bmatrix} \mathbf{G}_1^\eta \\ \mathbf{G}_2^\eta \\ \vdots \\ \mathbf{G}_M^\eta \end{bmatrix} \begin{bmatrix} \mathbf{G}_1^\eta \\ \mathbf{G}_2^\eta \\ \vdots \\ \mathbf{G}_M^\eta \end{bmatrix}^* \right)^{-\frac{1}{2}} \begin{bmatrix} \mathbf{G}_1^h \\ \mathbf{G}_2^h \\ \vdots \\ \mathbf{G}_M^h \end{bmatrix} \mathbf{x}_n + \tilde{\eta} \tag{57}$$

where $\tilde{\eta}$ is i.i.d. Gaussian noise with variance $1/\Delta$.

In order to calculate the capacity limit, we need to investigate the Fourier symbols associated with these block Toeplitz matrices.

Lemma 6: At a given frequency f, the Fourier symbol with respect to $\mathbf{G}_\alpha^\eta \mathbf{G}_\beta^\eta$ is given by $ak\mathbf{F}_\alpha^\eta \mathbf{F}_\beta^p \mathbf{F}_\beta^{p*} \mathbf{F}_\beta^{\eta*}$, and the Fourier symbol with respect to $\mathbf{G}_\alpha^h \mathbf{G}_\beta^h$ is given by $ak\mathbf{F}_\alpha^\eta \mathbf{F}_\alpha^p \mathbf{F}^h \mathbf{F}^{h*} \mathbf{F}_\beta^{p*} \mathbf{F}_\beta^{\eta*}$. Here, for any (l, v) such that $v \in \mathbb{Z}$ and $1 \le l \le a$, we have

$$(\mathbf{F}_\alpha^\eta)_{l,v} = \sum_u c_\alpha^u S_\alpha \left(-f + u f_q + v \frac{f_q}{b} \right) \cdot \\ \exp\left(-j 2\pi l \frac{T_s}{M} \left(f - u f_q - v \frac{f_q}{b} \right) \right).$$

Also, $\mathbf{F}_\alpha^p$ and $\mathbf{F}^h$ are infinite diagonal matrices such that for all $l \in \mathbb{Z}$

$$\begin{cases} (\mathbf{F}_\alpha^p)_{l,l} &= P_\alpha \left(-f + l\frac{f_q}{b} \right) \\ (\mathbf{F}^h)_{l,l} &= H \left(-f + l\frac{f_q}{b} \right). \end{cases}$$

Proof: See Appendix F-D.   ∎

Define $\mathbf{G}^\eta$ such that its (α, β) subblock is $\mathbf{G}_\alpha^\eta \mathbf{G}_\beta^{\eta*}$, and $\mathbf{G}^h$ such that its (α, β) subblock is $\mathbf{G}_\alpha^h \mathbf{G}_\beta^{h*}$. Proceeding similarly as in the proof of Theorem 4, we obtain

$$\mathcal{F}\left(\mathbf{G}^\eta\right) = ak\mathbf{F}^\eta \mathbf{F}^{\eta*} \quad \text{and} \quad \mathcal{F}\left(\mathbf{G}^h\right) = ak\mathbf{F}^\eta \mathbf{F}^h \mathbf{F}^{h*} \mathbf{F}^{\eta*}$$

where $\mathbf{F}^\eta$ contain $M \times 1$ submatrices. The $(\alpha, 1)$ submatrix of $\mathbf{F}^\eta$ is given by $\mathbf{F}_\alpha^\eta \mathbf{F}_\alpha^p$.

Denote $\mathbf{F}^{\eta\ddagger} \triangleq \left(\mathbf{F}^\eta \mathbf{F}^{\eta*}\right)^{-\frac{1}{2}} \mathbf{F}^\eta$. For any continuous function $g(x)$, [29, Th. 5.4] implies that

$$\lim_{n \to \infty} \frac{1}{naM} \sum_{i=1}^{naM} g\left(\lambda_i \left\{ (\mathbf{G}^\eta)^{-\frac{1}{2}} \mathbf{G}^h (\mathbf{G}^\eta)^{-\frac{1}{2}} \right\} \right) \\ = \int_{-\frac{\tilde{f}_s}{2a}}^{\frac{\tilde{f}_s}{2a}} \sum_{i=1}^{aM} g\left(\lambda_i \left(\mathbf{F}^{\eta\ddagger} \mathbf{F}^h \mathbf{F}^{h*} \mathbf{F}^{\eta\ddagger*} \right) \right) df.$$

Then, capacity of parallel channels, achieved via water-filling power allocation, yields

$$C(f_s) = \lim_{n \to \infty} \sum_{i=1}^{naM} \frac{\log^+ \left(\nu \lambda_i \left\{ (\mathbf{G}^\eta)^{-\frac{1}{2}} \mathbf{G}^h (\mathbf{G}^\eta)^{-\frac{1}{2}} \right\} \right)}{naM} \\ = \int_{-\frac{\tilde{f}_s}{2a}}^{\frac{\tilde{f}_s}{2a}} \frac{1}{2} \sum_{i=1}^{aM} \log^+ \left(\nu \lambda_i \left(\mathbf{F}^{\eta\ddagger} \mathbf{F}^h \mathbf{F}^{h*} \mathbf{F}^{\eta\ddagger*} \right) \right) df,$$

where the water level ν can be computed through the following parametric equation:

$$P = \lim_{n \to \infty} \frac{1}{naM} \sum_{i=1}^{naM} \left[\nu - \frac{1}{\lambda_i \left\{ (\mathbf{G}^\eta)^{-\frac{1}{2}} \mathbf{G}^h (\mathbf{G}^\eta)^{-\frac{1}{2}} \right\}} \right]^+ \\ = \int_{-\frac{\tilde{f}_s}{2a}}^{\frac{\tilde{f}_s}{2a}} \sum_{i=1}^{aM} \left[\nu - \frac{1}{\lambda_i \left\{ (\mathbf{G}^\eta)^{-\frac{1}{2}} \mathbf{G}^h (\mathbf{G}^\eta)^{-\frac{1}{2}} \right\}} \right]^+ df.$$

APPENDIX E
PROOF OF PROPOSITION 3

Denote by $y^k(t)$ the analog signal after passing through the kth prefilter prior to ideal sampling. When both the input signal $x(t)$ and the noise $\eta(t)$ are Gaussian, the MMSE estimator of $x(t)$ from samples $\{y^k[n] \mid 1 \le k \le M\}$ is linear. Recall that $\tilde{T}_s = MT_s$ and $\tilde{f}_s = f_s/M$. A linear estimator of $x(t)$ from $\mathbf{y}[n]$ can be given as

$$\hat{x}(t) = \sum_{k \in \mathbb{Z}} \mathbf{g}^T(t - k\tilde{T}_s) \cdot \mathbf{y}(k\tilde{T}_s) \tag{58}$$

where we use the vector form $\mathbf{g}(t) = [g^1(t), \ldots, g^M(t)]^T$ and $\mathbf{y}(t) = [y^1(t), \ldots, y^M(t)]^T$ for notational simplicity. Here, $g^l(t)$ denotes the interpolation function operating upon the samples in the lth branch. We propose to find the optimal estimator $\mathbf{g}(t)$ that minimizes the mean square estimation error $\mathbb{E}\left(|x(t) - \hat{x}(t)|^2\right)$ for some t.

From the orthogonality principle, the MMSE estimate $\hat{x}(t)$ obeys

$$\mathbb{E}\left(x(t)\mathbf{y}^*(l\tilde{T}_s) \right) = \mathbb{E}\left(\hat{x}(t)\mathbf{y}^*(l\tilde{T}_s) \right), \quad \forall l \in \mathbb{Z}. \tag{59}$$

Since $x(t)$ and $\eta(t)$ are both stationary Gaussian processes, we can define $\mathbf{R}_{XY}(\tau) := \mathbb{E}(x(t)\mathbf{y}^*(t - \tau))$ to be the cross correlation function between $x(t)$ and $\mathbf{y}(t)$, and $\mathbf{R}_Y(\tau) := \mathbb{E}(\mathbf{y}(t)\mathbf{y}^*(t - \tau))$ the autocorrelation function of $\mathbf{y}(t)$. Plugging (58) into (59) leads to the following relation:

$$\mathbf{R}_{XY}\left(t - l\tilde{T}_s \right) = \sum_{k \in \mathbb{Z}} \mathbf{g}^T\left(t - k\tilde{T}_s \right) \mathbf{R}_Y\left(k\tilde{T}_s - l\tilde{T}_s \right).$$

Replacing t by $t + l\tilde{T}_s$, we can equivalently express it as

$$
\begin{aligned}
\mathbf{R}_{XY}(t) &= \sum_{k \in \mathbb{Z}} \mathbf{g}^T \left(t + l\tilde{T}_s - k\tilde{T}_s \right) \mathbf{R}_Y \left(k\tilde{T}_s - l\tilde{T}_s \right) \\
&= \sum_{l \in \mathbb{Z}} \mathbf{g}^T \left(t - l\tilde{T}_s \right) \mathbf{R}_Y \left(l\tilde{T}_s \right),
\end{aligned}
\tag{60}
$$

which is equivalent to the convolution of $\mathbf{g}(t)$ and $\mathbf{R}_Y(t) \cdot \sum_{l \in \mathbb{Z}} \delta \left(t - l\tilde{T}_s \right)$.

Let $\mathcal{F}(\cdot)$ denote Fourier transform operator. Define the cross spectral density $\mathbf{S}_{XY}(f) := \mathcal{F}(\mathbf{R}_{XY}(t))$ and $\mathbf{S}_Y(f) = \mathcal{F}(\mathbf{R}_Y(t))$. By taking the Fourier transform on both sides of (60), we have

$$
\mathbf{S}_{XY}(f) = \mathbf{G}(f)\mathcal{F}\left(\mathbf{R}_Y(\tau) \sum_{l \in \mathbb{Z}} \delta(\tau - l\tilde{T}_s) \right)
$$

which immediately yields that $\forall f \in \left[-\tilde{f}_s/2, \tilde{f}_s/2 \right]$

$$
\begin{aligned}
\mathbf{G}(f) &= \mathbf{S}_{XY}(f) \left[\mathcal{F}\left(\mathbf{R}_Y(\tau) \sum_{l \in \mathbb{Z}} \delta(\tau - l\tilde{T}_s) \right) \right]^{-1} \\
&= \mathbf{S}_{XY}(f) \left(\sum_{l \in \mathbb{Z}} \mathbf{S}_Y \left(f - l\tilde{f}_s \right) \right)^{-1}.
\end{aligned}
$$

Since the noise $\eta(t)$ is independent of $x(t)$, the cross correlation function $\mathbf{R}_{XY}(t)$ is

$$
\begin{aligned}
\mathbf{R}_{XY}(\tau) = \mathbb{E}\,(x(t+\tau)\cdot \\
\left[(s_1 * h * x)^*(t), \ldots, (s_M * h * x)^*(t) \right])
\end{aligned}
$$

which allows the cross spectral density to be derived as

$$
\mathbf{S}_{XY}(f) = H^*(f)\mathcal{S}_X(f)\left[S_1^*(f), \ldots, S_M^*(f) \right].
\tag{61}
$$

Additionally, the spectral density of $\mathbf{y}(t)$ can be given as the following $M \times M$ matrix:

$$
\mathbf{S}_Y(f) = \left(|H(f)|^2 \mathcal{S}_X(f) + \mathcal{S}_\eta(f) \right) \mathbf{S}(f)\mathbf{S}^*(f)
\tag{62}
$$

with $\mathcal{S}_\eta(f)$ denoting the spectral density of the noise $\eta(t)$, and $\mathbf{S}(f) = [S_1(f), \ldots, S_m(f)]^T$.

Define

$$
\begin{aligned}
\mathbf{K}(f) := \sum_{l \in \mathbb{Z}} \left(\left| H(f - l\tilde{f}_s) \right|^2 \mathcal{S}_X(f - l\tilde{f}_s) + \mathcal{N}(f - l\tilde{f}_s) \right) \\
\mathbf{S}(f - l\tilde{f}_s)\mathbf{S}^*(f - l\tilde{f}_s).
\end{aligned}
$$

The Wiener–Hopf linear reconstruction filter can now be written as

$$
\mathbf{G}(f) = H^*(f)\mathcal{S}_X(f)\mathbf{S}^*(f)\mathbf{K}^{-1}(f).
$$

Define $R_X(\tau) = \mathbb{E}\,(x(t)x^*(t - \tau))$. Since $\int_{-\infty}^{\infty} \mathcal{S}_X(f)df = R_X(0)$, the resulting MSE at any sample time is

$$
\begin{aligned}
\xi(t) &= \mathbb{E}\left(|x(t)|^2 \right) - \mathbb{E}\left(|\hat{x}(t)|^2 \right) \\
&= \mathbb{E}\left(|x(t)|^2 \right) - \mathbb{E}\left(x(t)\hat{x}^*(t) \right) \\
&= R_X(0) - \mathbb{E}\left(x(t) \left(\sum_{l \in \mathbb{Z}} \mathbf{g}^T(t - l\tilde{T}_s)\mathbf{y}(l\tilde{T}_s) \right)^* \right) \\
&= R_X(0) - \sum_{l \in \mathbb{Z}} \mathbf{R}_{XY}(t - l\tilde{T}_s)\mathbf{g}(t - l\tilde{T}_s).
\end{aligned}
$$

Since $\mathcal{F}(\mathbf{g}(-t)) = (\mathbf{G}^*(f))^T$ and $\mathbf{S}_{XY} = H^*(f)\mathcal{S}_X(f)\mathbf{S}^*(f)$, Parseval's identity implies that

$$
\begin{aligned}
\xi(t) &= \int_{-\infty}^{\infty} \left[\mathcal{S}_X(f) - \mathbf{G}^*(f)\mathbf{S}_{XY}^T \right] df \\
&= \int_{-\infty}^{\infty} \left[\mathcal{S}_X(f) - |H(f)|^2 \mathcal{S}_X(f)|^2 \, \mathbf{S}^*(f)\mathbf{K}^{-1}(f)\mathbf{S}(f) \right] df \\
&= \int_{-\tilde{f}_s/2}^{\tilde{f}_s/2} \left[\sum_{l=-\infty}^{\infty} \mathcal{S}_X(f - l\tilde{f}_s) - \tilde{T}_s\mathbf{V}_\zeta^T(f, \tilde{f}_s) \cdot \mathbf{1} \right] df.
\end{aligned}
$$

Suppose that we impose power constraints $\sum_{l \in \mathbb{Z}} \mathcal{S}_X(f - l\tilde{f}_s) = P(f)$, and define $\zeta(f) := |H(f)\mathcal{S}_X(f)|^2 \, \mathbf{S}^*(f)\mathbf{K}^{-1}(f)\mathbf{S}(f)$. For a given input process $x(t)$, the problem of finding the optimal prefilter $\mathbf{S}(f)$ that minimizes MSE then becomes

$$
\underset{\{S(f - lf_s), l \in \mathbb{Z}\}}{\text{maximize}} \quad \mathbf{V}_\zeta^T(f, \tilde{f}_s) \cdot \mathbf{1}
$$

where the objective function can be alternatively rewritten in matrix form

$$
\text{trace}\left\{ \mathbf{F}_X^{\frac{1}{2}} \mathbf{F}_h^* \mathbf{F}_s^* \left(\mathbf{F}_s \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right) \mathbf{F}_s^* \right)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_X^{\frac{1}{2}} \right\}.
\tag{63}
$$

Here $\mathbf{F}_X$ and $\mathbf{F}_\eta$ are diagonal matrices such that $(\mathbf{F}_X)_{l,l} = \mathcal{S}_X(f - lf_s)$ and $(\mathbf{F}_\eta)_{l,l} = \mathcal{S}_\eta(f + kf_s)$. We observe that

$$
\begin{aligned}
&\text{trace}\left\{ \mathbf{F}_X^{\frac{1}{2}} \mathbf{F}_h^* \mathbf{F}_s^* \left(\mathbf{F}_s \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right) \mathbf{F}_s^* \right)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_X^{\frac{1}{2}} \right\} \\
&= \text{trace}\left\{ \left(\mathbf{F}_s \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right) \mathbf{F}_s^* \right)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{F}_s^* \right\} \\
&\overset{(a)}{=} \text{trace}\left\{ \left(\mathbf{Y}\mathbf{Y}^* \right)^{-1} \mathbf{Y} \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{-\frac{1}{2}} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \right. \\
&\qquad\qquad\qquad \left. \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{-\frac{1}{2}} \mathbf{Y}^* \right\} \\
&\overset{(b)}{=} \text{trace}\left\{ \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{Y}^* \left(\mathbf{Y}\mathbf{Y}^* \right)^{-1} \mathbf{Y} \right\} \\
&\overset{(c)}{=} \text{trace}\left\{ \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \tilde{\mathbf{Y}}^* \tilde{\mathbf{Y}} \right\} \\
&\overset{(d)}{\leq} \underset{\mathbf{Z} \cdot \mathbf{Z}^* = \mathbf{I}_M}{\sup} \text{trace}\left\{ \mathbf{Z} \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{Z}^* \right\} \\
&= \sum_{i=1}^{M} \lambda_i(\mathbf{D}),
\end{aligned}
\tag{64}
$$

where (a) follows by introducing $\mathbf{Y} := \mathbf{F}_s \left(\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta \right)^{\frac{1}{2}}$, (b) follows from the fact that $\mathbf{F}_h, \mathbf{F}_X, \mathbf{F}_\eta$ are all diagonal matrices, (c) follows by introducing $\tilde{\mathbf{Y}} = (\mathbf{YY}^*)^{-\frac{1}{2}} \mathbf{Y}$, and (d) follows by observing that $\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^* = (\mathbf{YY}^*)^{-\frac{1}{2}} \mathbf{YY}^* (\mathbf{YY}^*)^{-\frac{1}{2}} = \mathbf{I}$. Here, $\mathbf{D}$ is an infinite diagonal matrix such that

$$\mathbf{D}_{l,l} = \frac{|H(f - l f_s)|^2 \, \mathcal{S}_X(f - l f_s)}{|H(f - l f_s)|^2 \, \mathcal{S}_X(f - l f_s) + \mathcal{S}_\eta(f - l f_s)}$$

. In other words, the upper bound is the sum of the M largest $\mathbf{D}_{i,i}$ which are associated with M frequency points of highest SNR $\frac{|H(f + l f_s)|^2 \mathcal{S}_X(f + l f_s)}{\mathcal{S}_\eta(f + l f_s)}$.

Therefore, when restricted to the set of all permutations of $\{\mathcal{S}_X(f), \mathcal{S}_X(f \pm f_s), \ldots\}$, the minimum MSE is achieved when assigning the M largest $\mathcal{S}_X(f + l f_s)$ to M branches with the largest SNR. In this case, the corresponding optimal filter can be chosen such that

$$S_k(f - l f_s) = \begin{cases} 1, & \text{if } l = \hat{k} \\ 0, & \text{otherwise} \end{cases} \tag{65}$$

where $\hat{k}$ is the index of the kth largest element in $\left\{ |H(f - l f_s)|^2 / \mathcal{S}_\eta(f - l f_s) : l \in \mathbb{Z} \right\}$.

APPENDIX F
PROOFS OF AUXILIARY LEMMAS

A) Proof of Lemma 3: For any $i \leq j$, we have

$$\left| \left(\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} - \hat{\mathbf{H}}^n \right)_{ij} \right|$$
$$\leq \left| \sum_{t=-\infty}^{-j} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^* \right| + \left| \sum_{t=n-j+1}^{\infty} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^* \right|. \tag{66}$$

Since $h(t)$ is absolutely summable and Riemann integrable, for sufficiently small Δ, there exists a constant c such that $\sum_{i=-\infty}^{\infty} \left\| \tilde{\mathbf{h}}_i \right\|_1 \leq c$. In the following analysis, we define $\mathbf{R}^1$ and $\mathbf{R}^2$ to capture the two residual terms respectively, i.e.,

$$\mathbf{R}_{ij}^1 = \sum_{t=-\infty}^{-j} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^*, \quad \text{and} \quad \mathbf{R}_{ij}^2 = \sum_{t=n-j+1}^{\infty} \tilde{\mathbf{h}}_{j-i+t} \tilde{\mathbf{h}}_t^*.$$

In order to prove that $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \sim \hat{\mathbf{H}}^n$, we need to prove
1) $\lim_{n \to \infty} \frac{1}{n} \left\| \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} - \hat{\mathbf{H}}^n \right\|_F^2 = 0$, or equivalently, $\lim_{n \to \infty} \frac{1}{n} \left\| \mathbf{R}^2 \right\|_F^2 = 0$ and $\lim_{n \to \infty} \frac{1}{n} \left\| \mathbf{R}^1 \right\|_F^2 = 0$; and
2) the ℓ_2 norms of both $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ and $\hat{\mathbf{H}}^n$ are uniformly bounded, i.e., $\exists M_u$ such that $\left\| \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \right\|_2 \leq M_u < \infty$ and $\left\| \hat{\mathbf{H}}^n \right\|_2 \leq M_u < \infty$ for all n.

  1) We first prove that $\lim_{n \to \infty} \frac{1}{n} \left\| \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} - \hat{\mathbf{H}}^n \right\|_F^2 = 0$. By our assumptions, we have $h(t) = o(t^{-\epsilon})$ for some $\epsilon > 1$.

Since $s(t)$ is absolutely integrable, $\tilde{h}(t) = o(t^{-\epsilon})$ also holds. Without loss of generality, we suppose that $j \geq i$.
  a) if $i \geq n^{\frac{1}{2\epsilon}}$, by the assumption $\tilde{h}(t) = o\left(\frac{1}{t^\epsilon}\right)$ for some $\epsilon > 1$, one has

$$\left| \mathbf{R}_{ij}^1 \right| \leq \sum_{t=-\infty}^{-j} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \left\| \tilde{\mathbf{h}}_t \right\|_\infty$$
$$\leq \left(\max_{\tau \geq n^{\frac{1}{2\epsilon}}} \left\| \tilde{\mathbf{h}}_{-\tau} \right\|_1 \right) \sum_{t=-\infty}^{-j} \left\| \tilde{\mathbf{h}}_t \right\|_\infty$$
$$\leq \left(\max_{\tau \geq n^{\frac{1}{2\epsilon}}} \left\| \tilde{\mathbf{h}}_{-\tau} \right\|_1 \right) \sum_{t=-\infty}^{-j} \left\| \tilde{\mathbf{h}}_t \right\|_1 \leq c \max_{\tau \geq n^{\frac{1}{2\epsilon}}} \left\| \tilde{\mathbf{h}}_{-\tau} \right\|_1$$
$$= kc \cdot o\left(\frac{1}{\sqrt{n}} \right) = o\left(\frac{1}{\sqrt{n}} \right) \tag{67}$$

  b) if $j \geq n^{\frac{1}{2\epsilon}}$,

$$\left| \mathbf{R}_{ij}^1 \right| \leq \sum_{t=-\infty}^{-j} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \left\| \tilde{\mathbf{h}}_t \right\|_\infty$$
$$\leq \left(\sum_{t=-\infty}^{-j} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \right) \max_{\tau \leq -j} \left\| \tilde{\mathbf{h}}_\tau \right\|_\infty$$
$$\leq c \max_{\tau \geq n^{\frac{1}{2\epsilon}}} \left\| \tilde{\mathbf{h}}_{-\tau} \right\|_\infty$$
$$= c \cdot o\left(\frac{1}{\sqrt{n}} \right) = o\left(\frac{1}{\sqrt{n}} \right) \tag{68}$$

  c) if $j < n^{\frac{1}{2\epsilon}}$ and $i < n^{\frac{1}{2\epsilon}}$, we have

$$\left| \mathbf{R}_{ij}^1 \right|^2 \leq \left(\sum_{t=-\infty}^{\infty} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \left\| \tilde{\mathbf{h}}_t \right\|_\infty \right)^2$$
$$\leq \left(\sum_{t=-\infty}^{\infty} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \right)^2 \left(\max_t \left\| \tilde{\mathbf{h}}_t \right\|_\infty \right)^2$$
$$\leq \left(\sum_{t=-\infty}^{\infty} \left\| \tilde{\mathbf{h}}_{j-i+t} \right\|_1 \right)^2 \left(\sum_{t=-\infty}^{\infty} \left\| \tilde{\mathbf{h}}_t \right\|_1 \right)^2$$
$$\leq c^4. \tag{69}$$

By combining inequalities (67), (68), and (69), we can obtain

$$\lim_{n \to \infty} \frac{1}{n} \left\| \mathbf{R}^1 \right\|_F^2$$
$$= \lim_{n \to \infty} \frac{1}{n} \left(\sum_{i,j < n^{\frac{1}{2\epsilon}}} \left| \mathbf{R}_{ij}^1 \right|^2 + \sum_{i \geq n^{\frac{1}{2\epsilon}} \text{ or } j \geq n^{\frac{1}{2\epsilon}}} \left| \mathbf{R}_{ij}^1 \right|^2 \right)$$
$$\leq \lim_{n \to \infty} \frac{1}{n} \left[n^{\frac{1}{\epsilon}} \max_{i,j < n^{\frac{1}{2\epsilon}}} \left| \mathbf{R}_{ij}^1 \right|^2 + 2 n^{1 + \frac{1}{2\epsilon}} \max_{i \text{ or } j \geq n^{\frac{1}{2\epsilon}}} \left| \mathbf{R}_{ij}^1 \right|^2 \right]$$
$$= \lim_{n \to \infty} \frac{1}{n} \left[n^{\frac{1}{\epsilon}} c^4 + 2 n^{1 + \frac{1}{2\epsilon}} o\left(\frac{1}{n} \right) \right] = 0.$$

Similarly, we can show that

$$\lim_{n \to \infty} \frac{1}{n} \left\| \mathbf{R}^2 \right\|_F^2 = 0$$

which immediately implies that

$$\lim_{n \to \infty} \frac{1}{n} \left\| \hat{\mathbf{H}}^n - \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \right\|_{\mathrm{F}}^2 = 0.$$

2) We now proceed to show that $\left\| \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \right\|_2$ and $\left\| \hat{\mathbf{H}}^n \right\|_2$ are uniformly bounded. Since $\hat{\mathbf{H}}^n$ is a Toeplitz matrix, applying [24, Lemma 6] and [24, Sec. 4.1] yields

$$\left\| \hat{\mathbf{H}}^n \right\|_2 \le 2 \sum_{i=0}^{\infty} \sum_{t=-\infty}^{\infty} \left| \tilde{\mathbf{h}}_{i+t} \tilde{\mathbf{h}}_t^* \right|$$

$$\le 2 \sum_{t=-\infty}^{+\infty} \left\| \tilde{\mathbf{h}}_t \right\|_{\infty} \sum_{i=0}^{\infty} \left\| \tilde{\mathbf{h}}_{i+t} \right\|_1 \le 2c^2.$$

Additionally, since $\tilde{\mathbf{H}}^n$ is a block Toeplitz matrix, [29, Corollary 4.2] allows us to bound the norm as

$$\left\| \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} \right\|_2 = \left\| \hat{\mathbf{H}}^n \right\|_2^2 \le \left\| \mathbf{F}_{\tilde{h}}(\omega) \right\|_{\infty}^2 = \sup_{\omega} \sum_{i=0}^{k-1} \left| \mathbf{F}_{\tilde{h},i}(\omega) \right|^2$$

$$\le \sum_{j=0}^{\infty} \left(\sum_{i=0}^{k-1} \left| \left(\tilde{\mathbf{h}}_j \right)_i \right| \right)^2 \le \left(\sum_{j=0}^{\infty} \left\| \tilde{\mathbf{h}}_j \right\|_1 \right)^2 \le c^2.$$

Hence, by definition of asymptotic equivalence, we have $\hat{\mathbf{H}}^n \sim \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$.

B) Proof of Lemma 4: We know that $\mathbf{S}^n \mathbf{S}^{n*} = \hat{\mathbf{S}}^n$; hence, $\mathbf{C}^n \sim \hat{\mathbf{S}}^n = \mathbf{S}^n \mathbf{S}^{n*}$. Recall that $\left(\hat{\mathbf{S}}^n \right)_{1i} = \sum_{t=-\infty}^{\infty} \mathbf{s}_{i-1+t} \mathbf{s}_t^*$. For a given k, the Fourier series related to $\{ \mathbf{C}^n \}$ can be given as

$$F_c^k(\omega) = \sum_{i=-\infty}^{\infty} \left(\sum_{t=-\infty}^{\infty} \mathbf{s}_{i+t} \mathbf{s}_t^* \right) \exp(ji\omega). \tag{70}$$

By Lemma 2, in order to show $(\mathbf{C}^n)^{-1} \sim (\mathbf{S}^n \mathbf{S}^{n*})^{-1}$, we will need to show that $F_c^k(\omega)$ is uniformly bounded away from 0.

When k is sufficiently large, the Riemann integrability of $s(t)$ implies that

$$F_c^k(\omega) \doteq \Delta \sum_{i=-\infty}^{\infty} \left(\int_{-\infty}^{+\infty} s(t + iT_s) \mathbf{s}(t)^* dt \right) \exp(ji\omega)$$

$$= \Delta \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} s(t + \tau) s(t)^* dt \right)$$

$$\cdot \left(\sum_{i=-\infty}^{\infty} \delta(\tau - iT_s) \right) \exp\left(j \frac{\omega}{T_s} \tau \right) d\tau.$$

We observe that

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{\infty} s(t + \tau) s(t)^* dt \right) \exp\left(j \frac{\omega}{T_s} \tau \right) d\tau$$

$$= \left(\int_{-\infty}^{\infty} s(t + \tau) \exp\left(j \frac{\omega}{T_s} (t + \tau) \right) d\tau \right)$$

$$\left(\int_{-\infty}^{+\infty} s(t) \exp\left(j \frac{\omega}{T_s} t \right) dt \right)^*$$

$$= \left| S\left(-j \frac{\omega}{T_s} \right) \right|^2.$$

Since $F_c^k(\omega)$ corresponds to the Fourier transform of the signals obtained by uniformly sampling $\int_{-\infty}^{\infty} s(t + \tau) s(t)^* dt$, we can immediately see that

$$\lim_{\Delta \to 0} F_c^k(\omega) = \frac{\Delta}{T_s} \sum_{i=-\infty}^{\infty} \left| S\left(-j \left(\frac{\omega}{T_s} - \frac{i2\pi}{T_s} \right) \right) \right|^2. \tag{71}$$

If for all $\omega \in [-\pi, \pi]$, we have

$$\sum_{i=-\infty}^{\infty} \left| S\left(-j \left(\frac{\omega}{T_s} - \frac{i2\pi}{T_s} \right) \right) \right|^2 \ge \epsilon_s > 0 \tag{72}$$

for some constant ϵ_s, then $\sigma_{\min}(\mathbf{C}^n) = \inf_{\omega} F_c^k(\omega) \ge \frac{\Delta \epsilon_s}{T_s}$, which leads to $\left\| (\mathbf{C}^n)^{-1} \right\|_2 \le \frac{T_s}{\Delta \epsilon_s}$.

Let $\Xi^n = \mathbf{C}^n - \mathbf{S}^n \mathbf{S}^{n*}$. Since $\mathbf{S}^n \mathbf{S}^{n*} \sim \mathbf{C}^n$, we can have $\lim_{n \to \infty} \frac{1}{\sqrt{n}} \| \Xi^n \|_{\mathrm{F}} = 0$, which implies that

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \Xi^n (\mathbf{C}^n)^{-1} \right\|_{\mathrm{F}} \le \lim_{n \to \infty} \frac{1}{\sqrt{n}} \| \Xi^n \|_{\mathrm{F}} \left\| (\mathbf{C}^n)^{-1} \right\|_2$$

$$\le \lim_{n \to \infty} \frac{T_s}{\Delta \epsilon_s} \frac{1}{\sqrt{n}} \| \Xi^n \|_{\mathrm{F}} = 0.$$

The Taylor expansion of $(\mathbf{S}^n \mathbf{S}^{n*})^{-1}$ yields

$$(\mathbf{S}^n \mathbf{S}^{n*})^{-1} = (\mathbf{C}^n - \Xi^n)^{-1}$$

$$= (\mathbf{C}^n)^{-1} \left(\mathbf{I} + \Xi^n (\mathbf{C}^n)^{-1} + \left(\Xi^n (\mathbf{C}^n)^{-1} \right)^2 + \cdots \right).$$

Hence, we can bound

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| (\mathbf{S}^n \mathbf{S}^{n*})^{-1} - (\mathbf{C}^n)^{-1} \right\|_{\mathrm{F}}$$

$$\le \lim_{n \to \infty} \left\| (\mathbf{C}^n)^{-1} \right\|_2 \left(\sum_{i=1}^{\infty} \left(\frac{1}{\sqrt{n}} \left\| \Xi^n (\mathbf{C}^n)^{-1} \right\|_{\mathrm{F}} \right)^i \right) = 0.$$

C) Proof of Lemma 5: Since $(\mathbf{C}^n)^{-\frac{1}{2}}$ and $\left(\hat{\mathbf{S}}^n \right)^{-\frac{1}{2}}$ are both Hermitian and positive semidefinite, we have $(\mathbf{C}^n)^{-\frac{1}{2}} \sim \left(\hat{\mathbf{S}}^n \right)^{-\frac{1}{2}}$. The asymptotic equivalence allows us to relate $\frac{1}{n} \sum_{i=1}^{n} g(\lambda_i)$ to the function associated with the spectrum of the circulant matrix $\mathbf{C}^n$ instead of $\hat{\mathbf{S}}^n$. One nice property is that $(\mathbf{C}^n)^{-\frac{1}{2}} = \mathbf{U}_c \mathbf{\Lambda}_c^{-\frac{1}{2}} \mathbf{U}_c^*$ is still a circulant matrix. Combining the above results with Lemma 1 yields

$$(\mathbf{S}^n \mathbf{S}^n)^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \sim (\mathbf{C}^n)^{-\frac{1}{2}} \hat{\mathbf{H}}^n (\mathbf{C}^n)^{-\frac{1}{2}}.$$

Note that $(\mathbf{C}^n)^{-\frac{1}{2}} \hat{\mathbf{H}}^n (\mathbf{C}^n)^{-\frac{1}{2}}$ is simply multiplication of 3 Toeplitz matrices. This allows us to untangle $F_c(\omega)$ and $F_{\hat{h}}(\omega)$, hence separating $H(f)$ and $S(f)$.

Specifically, denote by $F_{c_{0.5}}(\omega), F_c(\omega), F_{\hat{h}}(\omega), \mathbf{F}_{\tilde{h}}(\omega)$ the Fourier series related to $(\mathbf{C}^n)^{\frac{1}{2}}, \mathbf{C}^n, \hat{\mathbf{H}}^n$, and $\hat{\mathbf{H}}^n$, respectively. We note that $F_{c_{0.5}}(\omega), F_c(\omega)$, and $F_{\hat{h}}(\omega)$ are all scalars since their related matrices are Toeplitz, while $\mathbf{F}_{\tilde{h}}(\omega)$ is a $1 \times k$ vector

since $\tilde{\mathbf{H}}$ is block Toeplitz. Then, for any continuous function $g(x)$, applying [24, Th. 12] yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \lambda_i \left((\mathbf{S}^n \mathbf{S}^n)^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \right) \right\}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \lambda_i \left((\mathbf{C}^n)^{-\frac{1}{2}} \hat{\mathbf{H}}^n (\mathbf{C}^n)^{-\frac{1}{2}} \right) \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} g\left(F_{c_{0.5}}^{-1}(\omega) F_{\hat{h}}(\omega) F_{c_{0.5}}^{-1}(\omega) \right) d\omega$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \lambda_i \left((\mathbf{C}^n)^{-1} \hat{\mathbf{H}}^n \right) \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} g\left(\frac{F_{\hat{h}}(\omega)}{F_c(\omega)} \right) d\omega.$$

Now we only need to show that both $F_{\hat{h}}(\omega)$ and $F_c(\omega)$ have simple closed-form expressions. We observe that $\hat{\mathbf{H}}_n$ is asymptotically equivalent to $\tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$, and the eigenvalues of $\hat{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*}$ are exactly the square of the corresponding singular values of $\tilde{\mathbf{H}}^n$. Hence, we know from [29] that for any continuous function $g(x)$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \lambda_i \left(\hat{\mathbf{H}}^n \right) \right\} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \sigma_i^2 \left(\tilde{\mathbf{H}}^n \right) \right\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} g\left(\sigma^2 \left(\mathbf{F}_{\tilde{h}}(\omega) \right) \right) d\omega$$

where $\mathbf{F}_{\tilde{h}}(\omega)$ can be expressed as $\mathbf{F}_{\tilde{h}}(\omega) = \left[F_{\tilde{h},0}(\omega), \cdots, F_{\tilde{h},k-1}(\omega) \right]$. Here, for any $0 \le i < k$:

$$F_{\tilde{h},i}(\omega) := \Delta \sum_{u=-\infty}^{+\infty} \tilde{h}_{u,i} \exp(ju\omega)$$

$$= \Delta \sum_{u\in\mathbb{Z}} \tilde{h}(uT_s - i\Delta) \exp(ju\omega).$$

The above analysis implies that $F_{\hat{h}}(\omega) = \sigma^2 \left(F_{\tilde{h}}(\omega) \right)$.

Through algebraic manipulation, we have that

$$F_{\tilde{h},i}(\omega) = \frac{\Delta}{T_s} \sum_{l\in\mathbb{Z}} H(-f + lf_s) \exp(-j2\pi(f - lf_s)i\Delta)$$

which yields

$$F_{\hat{h}}(f) = \sigma^2 \left(\mathbf{F}_{\tilde{h}}(2\pi f) \right) = \sum_{i=0}^{k-1} \left| F_{\tilde{h},i}(2\pi f) \right|^2$$

$$= \frac{\Delta^2}{T_s^2} \sum_{i=0}^{k-1} \left| \sum_{l=-\infty}^{+\infty} \tilde{H}(-f + lf_s) \exp(-j2\pi(f - lf_s)i\Delta) \right|^2$$

$$= \frac{\Delta^2}{T_s^2} \sum_{i=0}^{k-1} \sum_{l_1,l_2} \tilde{H}(-f + l_1 f_s) \tilde{H}^*(-f + l_2 f_s) \cdot$$

$$\exp(-j2\pi(l_2 - l_1)f_s i\Delta)$$

$$= \frac{\Delta^2}{T_s^2} \sum_{l_1,l_2} \tilde{H}(-f + l_1 f_s) \tilde{H}^*(-f + l_2 f_s)$$

$$\left[\sum_{i=0}^{k-1} \exp\left(-j2\pi(l_2 - l_1)\frac{i}{k} \right) \right]$$

$$= \frac{\Delta}{T_s} \sum_{l} |H(-f + lf_s) S(-f + lf_s)|^2.$$

Similarly, we have

$$F_c(f) = \frac{\Delta}{T_s} \sum_{l\in\mathbb{Z}} |S(-f + lf_s)|^2. \tag{73}$$

Combining the above results yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g\left\{ \lambda_i \left((\mathbf{S}^n \mathbf{S}^n)^{-\frac{1}{2}} \tilde{\mathbf{H}}^n \tilde{\mathbf{H}}^{n*} (\mathbf{S}^n \mathbf{S}^{n*})^{-\frac{1}{2}} \right) \right\}$$

$$= T_s \int_{-f_s/2}^{f_s/2} g\left(\frac{\sum_{l=-\infty}^{+\infty} |H(-f + lf_s) S(-f + lf_s)|^2}{\sum_{l=-\infty}^{+\infty} |S(-f + lf_s)|^2} \right) df.$$

This completes the proof.

D) Proof of Lemma 6: Denote by $\mathbf{K}_\alpha^\eta$ the Fourier symbol associated with the block Toeplitz matrix $\mathbf{G}_\alpha^\eta$. We know that the Fourier transform of $g_i^\eta(t,\tau)$ with respect to τ is given by

$$\int_\tau g_i^\eta(t,\tau) \exp(-j2\pi f\tau) d\tau$$

$$= \int_{\tau_2} \int_{\tau_1} s_i(t-\tau_1) q_i(\tau_1) p_i(\tau_1 - \tau_2) \exp(-j2\pi f\tau_2) d\tau_1 d\tau_2$$

$$= \int_{\tau_2} p_i(\tau_1 - \tau_2) \exp(j2\pi f(\tau_1 - \tau_2)) d\tau_2$$

$$\int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) \exp(-j2\pi f\tau_1) d\tau_1$$

$$= P_i(-f) \cdot \left[S_i(-f) \exp(-j2\pi tf) \cdot \sum_u c_i^u \delta(f - uf_q) \right]$$

$$= P_i(-f) \cdot \left[\sum_u c_i^u S_i(-f + uf_q) \exp(-j2\pi t(f - uf_q)) \right].$$

Introduce the notation $S^e(f) \triangleq S(f) \exp\left(j2\pi l\tilde{T}_s f \right)$. For any (l,m) such that $1 \le l \le a$ and $1 \le m \le ak$, the (l,m) entry of the Fourier symbol $\mathbf{K}_\alpha^\eta$ can be related to the sampling sequence of $g_\alpha^\eta\left(l\tilde{T}_s, \tau \right)$ at a rate $\frac{\tilde{f}_s}{a}$ with a phase shift $m\Delta$, and hence, it can be calculated as follows:

$$(\mathbf{K}_\alpha^\eta)_{l,m} = \sum_v P_i\left(-f + v\frac{f_q}{b} \right) \exp\left(j2\pi \left(f - v\frac{f_q}{b} \right) m\Delta \right)$$

$$\cdot \left[\sum_u c_i^u S^e\left(-f + uf_q + v\frac{f_q}{b} \right) \right].$$

Using the fact that $\sum_{m=0}^{ak-1} \exp\left(j2\pi \left((v_2 - v_1)\frac{f_q}{b} \right) m\Delta \right) = ak\delta[v_2 - v_1]$, we get through algebraic manipulation that

$$\left(\mathbf{K}_\alpha^\eta \mathbf{K}_\beta^{\eta*} \right)_{l,d} = ak \sum_v P_\alpha\left(-f + v\frac{f_q}{b} \right) \cdot P_\beta^*\left(-f + v\frac{f_q}{b} \right)$$

$$\left[\sum_{u_1} c_\alpha^{u_1} S_\alpha^e\left(-f + u_1 f_q + v\frac{f_q}{b} \right) \right] \cdot$$

$$\left[\sum_{u_2} c_\beta^{u_2} S_\beta^e\left(-f + u_2 f_q + v\frac{f_q}{b} \right) \right]^*.$$

Define another matrix $\mathbf{F}_\alpha^\eta$ such that

$$(\mathbf{F}_\alpha^\eta)_{l,v} = \sum_u c_\alpha^u S_\alpha \left(-f + u f_q + v \frac{f_q}{b} \right) \cdot$$
$$\exp\left(-j 2\pi l \tilde{T}_s \left(f - u f_q - v \frac{f_q}{b} \right) \right) \cdot$$

It can be easily seen that

$$\mathbf{K}_\alpha^\eta \mathbf{K}_\beta^{\eta*} = a k \mathbf{F}_\alpha^\eta \mathbf{F}_\alpha^p \mathbf{F}_\alpha^{p*} \mathbf{F}_\beta^{\eta*}.$$

Replacing P_α by $P_\alpha H$ immediately gives us the Fourier symbol for $\mathbf{G}_\alpha^h \mathbf{G}_\beta^{h*}$.

References

[1] R. G. Gallager, *Information theory and reliable communication*. New York, NY, USA: Wiley, 1968.

[2] C. E. Shannon, *A mathematical theory of communication*. Urbana, USA: Univ. Illinois Press, 1949.

[3] U. Grenander and G. Szego, *Toeplitz forms and their applications*. New York, NY, USA: Amer. Math. Soc., 1984.

[4] W. Hirt and J. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 38–38, May 1988.

[5] P. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. IT-11, no. 4, pp. 360–393, Dec. 1963.

[6] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[7] M. Medard, "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.

[8] J. Yen, "On nonuniform sampling of bandwidth-limited signals," *IRE Trans. Circuit Theory*, vol. 3, no. 4, pp. 251–257, Dec. 1956.

[9] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. IT-24, no. 11, pp. 652–654, Nov. 1977.

[10] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, pp. 37–52, 1967.

[11] Y.-P. Lin and P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II, Analog Digital Signal Process.*, vol. 45, no. 3, pp. 340–351, Mar. 1998.

[12] M. Unser and J. Zerubia, "A generalized sampling theory without band-limiting constraints," *IEEE Trans. Circuits Syst. II, Analog Digital Signal Process.*, vol. 45, no. 8, pp. 959–969, Aug. 1998.

[13] 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.

[14] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits, Devices Syst.*, vol. 5, no. 1, pp. 8–20, Jan. 2011.

[15] T. Berger and D. Tufts, "Optimum pulse amplitude modulation-I: Transmitter-receiver design and bounds from information theory," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 2, pp. 196–208, Apr. 1967.

[16] T. Berger, "Nyquist's problem in data transmission theory," Ph.D. dissertation, Harvard Univ., Cambridge, MA, USA, Dec. 1965.

[17] D. Chan and R. Donaldson, "Optimum pre-and postfiltering of sampled signals with application to pulse modulation and data compression systems," *IEEE Trans. Commun. Technol.*, vol. IT-19, no. 2, pp. 141–157, Apr. 1971.

[18] T. Ericson, "Optimum PAM filters are always band limited (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 570–573, Jul. 1973.

[19] S. Shamai, "Information rates by oversampling the sign of a bandlimited process," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1230–1236, Jul. 1994.

[20] T. Koch and A. Lapidoth, Increased capacity per unit-cost by oversampling Sep. 2010 [Online]. Available: http://arxiv.org/abs/1008.5393

[21] Y. Chen, A. J. Goldsmith, and Y. C. Eldar, Channel capacity under sub-Nyquist nonuniform sampling Apr. 2012 [Online]. Available: http://arxiv.org/abs/1204.6049

[22] Y. C. Eldar and T. Michaeli, "Beyond bandlimited sampling," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 48–68, May 2009.

[23] A. J. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.

[24] R. Gray, *Toeplitz and circulant matrices: A review*, ser. Foundations and Trends in Communications and Information Theory. Delft, The Netherlands: Now Publishers, 2006, vol. 2.

[25] A. Lapidoth, *A Foundation in Digital Communication*. New York, NY, USA: Cambridge Univ. Press, 2009.

[26] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[27] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2000.

[28] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.

[29] P. Tilli, "Singular values and eigenvalues of non-Hermitian block Toeplitz matrices," *Linear Algebra Appl.*, vol. 272, no. 1–3, pp. 59–89, 1998.

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

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 Edwards Chair in Engineering. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT and a Visiting Professor at Stanford University, Stanford, CA. Her research interests are in the broad areas of statistical signal processing, sampling theory and compressed sensing, optimization methods, and their applications to biology and optics.

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

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) is a professor of Electrical Engineering at Stanford University, and was previously an assistant professor of Electrical Engineering at Caltech. She co-founded Accelera Mobile Broadband, Inc. and Quantenna Communications Inc., and has previously 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 National Academy of Engineering Gilbreth Lecture Award, the IEEE Wireless Communications Technical Committee Recognition 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 is currently on the Steering Committee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and has previously served 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*. Dr. Goldsmith 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 been a Distinguished Lecturer for both societies, served as the President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory society, and currently chairs the Emerging Technology Committee and is a member of the Strategic Planning Committee in the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award, served as Chair of its Faculty Senate, and currently serves on its Faculty Senate and on its Budget Group.