Shannon Meets Nyquist:  
Capacity Limits of Sampled Analog Channels

Yuxin Chen, Yonina C. Eldar, and Andrea J. Goldsmith

Abstract

We explore two fundamental questions at the intersection of sampling theory and information theory: how is channel capacity affected by sampling below the channel’s Nyquist rate, and what sub-Nyquist sampling strategy should be employed to maximize capacity. In particular, we first derive the capacity of sampled analog channels for two prevalent sampling mechanisms: filtering followed by sampling and sampling following filter banks. Connections between sampling and MIMO Gaussian channels are illuminated based on this analysis. Optimal prefilters that maximize capacity are identified for both cases, as well as several kinds of channels for which these sampling mechanisms are optimal to maximize capacity at sub-Nyquist rates. We also highlight connections between sampled analog channel capacity and minimum mean squared error estimation from sampled data. In particular, it is shown that for both filtering and filter-band sampling strategies, the filters maximizing capacity and minimizing mean squared error are equivalent. We also investigate a more general sampling strategy by adding modulation banks to filter-bank sampling. This general sampling method subsumes most nonuniform sampling techniques applied in both theory and practice. We also show a connection between this general sampling method and MIMO Gaussian channels. We then identify the optimal sampling strategy for piece-wise flat sampled channels to be a single branch of modulation and filtering. These results demonstrate the tradeoffs between channel capacity and sampling rate, illustrate the interplay between sampling techniques and capacity of...
sampled analog channels, and identify a simple optimal sampling strategy to maximize capacity for a large class of channels.

Index Terms

sampling rate, channel capacity, sampled analog channels, sub-Nyquist sampling

I. INTRODUCTION

The capacity of continuous-time waveform 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, channels 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 of bandlimited channels, especially for wideband communication systems. This gives rise to a natural question at the intersection of sampling theory and information theory: how much information, in the Shannon sense, can be conveyed through undersampled analog channels. There are two types of fundamental capacity metrics that are of interests: (1) the capacity for a given sampling mechanism, which incorporates the sampling mechanism into the channel model; (2) the capacity for a fixed sampling rate when optimizing over a general class of sampling mechanisms at that rate. The second metric requires that the sampling mechanism be optimized relative to capacity. In this paper, we study several sampling mechanisms of increasing complexity, and investigate the interplay between capacity and these sampling strategies. In particular, for each of these sampling strategies, we determine the capacity as a function of the sampling rate, and identify classes of channels for which the sampling strategy is optimal at any given sampling rate below the Nyquist rate.

A. Related Work

The derivation of the capacity of linear time-invariant (LTI) waveform channels was pioneered by Shannon et. al. [2], [3]. Making use of the asymptotic spectral properties of Toeplitz operators [4] or, alternatively, Fourier analysis [5], this capacity result established the optimality of a water-filling power allocation based on signal-to-noise ratio (SNR) across the frequency domain [1], which has motivated practical power and bit loading in multicarrier communications [6]. The Shannon framework has also been extended to wideband fading channels [7]–[9], multiple-input-multiple-output (MIMO) channels [10]–[12], and non-coherent channels [13]–[15]. 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 since no information is lost, has been used since the early days of information theory to transform analog channels into their discrete counterparts, e.g. [16]. Forney et. al. [17] surveys minimum-bandwidth orthogonal pulse amplitude modulation (PAM) techniques for serial transmission over linear Gaussian channels, which allows the lossless conversion between analog and digital channels through Nyquist-rate sampling. This paradigm of discretization has also been employed by Medard et. al. to bound the maximum mutual information in time-varying channels [8], [13]. 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 [18], [19]. Various sampling methods at this rate for bandlimited functions were reviewed by Jerri [20], including both uniform and nonuniform pointwise sampling techniques. Examples include recurrent non-uniform sampling proposed by Yen [21], 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 filter-bank sampling first analyzed by Papoulis [22], in which the input signal is sampled through \( M \) linear systems. For perfect reconstruction, these methods require sampling at an aggregate rate equal to or above the Nyquist rate.

In practice, however, the Nyquist rate may be excessive for perfect reconstruction of signals that possess certain structure. 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 [23], [24]. If the spectral support is known \textit{a priori}, then the sampling rate requirement for perfect recovery is the sum of the subband bandwidths (including both positive and negative frequencies), termed the \textit{Landau rate} [25]. One type of sampling mechanism that can reconstruct multiband signals sampled at the Landau rate is a filter bank followed by sampling, or “generalized” sampling, studied in [26], [27]. The basic paradigm is to apply a bank of prefilters to the analog signal, each followed by a uniform sampler. A bank of filters followed by sampling is an effective class of non-uniform sampling methods that is widely applied in theory and practice [21].

When the channel or signal structure is unknown, other sampling methods have been investigated to see under what conditions a signal can be recovered from sub-Nyquist samples. Inspired by recent “compressive sensing” ideas [28], [29], sub-Nyquist sampling approaches have been developed to exploit the structure of various classes of input signals, such as multiband signals [30], [31] and signals with finite rate of innovation [32], [33]. Modulation and filter banks followed by sampling, where the signal is passed
through modulation banks and filter banks before sampling, has proven to be very effective for signal reconstruction at sub-Nyquist sampling rates. By scrambling spectral contents from different subbands through the modulation operation, this method performs well in subsampling sparse multiband signals with unknown spectral support. One example of this sampling mechanism is the modulated wideband converter (MWC) proposed by Mishali et al. \cite{30}, \cite{34}, where all post-modulation filters are chosen to be low-pass filters. In fact, modulation and filter banks followed by sampling represents the most general class of realizable nonuniform sampling techniques, although it does not include certain techniques such as sampling at random sample times.

Most of the above sampling theoretic work aims at finding optimal sampling and reconstruction mechanisms that achieve perfect reconstruction 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. MSE \cite{35}, \cite{36}). Another line of work pioneered by Berger et. al. \cite{37}–\cite{41} investigated joint optimization of the transmitted pulse shape and receiver prefiltering in PAM over an analog communication channel under sub-Nyquist sampling. In this work the optimal receiver prefilter that minimizes the MSE between the original signal and the signal reconstructed from the samples is shown to prevent aliasing. However, this work does not consider optimal sampling techniques based on the information-theoretic metric of channel capacity achievable through noisy samples of the channel output. In addition, these optimal filters derived in \cite{37}, \cite{39} are used to determine an SNR metric which in turn is used to compute an approximation to sampled channel capacity based on the formula for capacity of the bandlimited AWGN channels. This approximation does not correspond to the precise capacity of undersampled bandlimited AWGN channels we derive herein, nor is the capacity of more general undersampled analog channels considered. Guo et. al. explored the connection between mutual information and MMSE for Gaussian channels \cite{42}, but focused on the discrete domain and therefore did not account for undersampling of analog signals.

The tradeoff between capacity and hardware complexity has been studied in another line of work focused on sampling precision \cite{43}–\cite{45}. These works demonstrate that, due to quantization of samples, sampling above the Nyquist rate can be beneficial in increasing achievable data rates. The focus of this quantization analysis is on the effect of increasing the sampling rate beyond the Nyquist rate to combat quantization error, whereas this paper is concerned with determining capacity and optimal sub-Nyquist sampling strategies for channels based on the channel structure, without considering quantization errors.
B. Contribution

Sampling structures typically rely on general prefiltering prior to sampling [18], which can suppress aliasing and post-sampling noise, minimize the recovery error for certain classes of input signals, and account for non-ideal linear distortion features of practical acquisition devices [46], [47]. Here, we explore sampled analog channels with the following three classes of sampling mechanisms: (1) a filter followed by sampling: the analog channel output is prefiltered by a single linear filter followed by an ideal uniform sampler (see Fig. 2); (2) sampling following filter banks: the analog channel output is passed through a bank of LTI filters, each followed by an ideal uniform sampler (see Fig. 3); (3) modulation and filter banks followed by sampling: the channel output is split into $M$ branches, where each branch is prefiltered by an LTI filter, modulated by a different modulation sequence, passed through another LTI filter and then sampled uniformly. Our main contributions are summarized as follows.

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

- **Filter banks followed by sampling.** A closed-form expression for sampled channel capacity is derived, along with analysis that relates it to a MIMO Gaussian channel. The input should be chosen to decouple the dimensions of the equivalent MIMO channel. We also derive optimal filter banks that maximize capacity. The $M$ filters select the $M$ frequencies with the $M$ highest SNRs and zero out signals from all other frequencies. This 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 non-monotonic, and it achieves the analog capacity of multiband channels sampled at the Landau rate if the number of branches is appropriately chosen.

- **Modulation and filter banks followed by sampling.** For modulation sequences that are periodic with period $T_q = 1/f_q$, we derive the sampled channel capacity and show its connection to a more general MIMO Gaussian channel in the frequency domain than in the case without modulation banks. For sampling following a single branch of modulation and filtering, we provide an algorithm
to identify the optimal modulation sequence for piece-wise flat channels when $T_q$ is an integer multiple of the sampling period. This single-branch mechanism achieves the same performance as employing an optimal filter bank with each branch sampled at a period $T_q$.

One interesting fact we discover for all these techniques is the non-monotonicity of capacity with sampling rate, which indicates that more sophisticated sampling techniques are needed to maximize achievable data rates under sub-Nyquist sampling.

C. Organization

The remainder of this paper is organized as follows. In Section II we describe the problem formulation of sampled analog channels for the sampling mechanisms described above. We then state our formal capacity results for each of these sampling mechanisms: sampling with a filter, with a filter bank, and with modulation and filter banks, along with their implications in Sections III–V. In particular, in each section the main theorems are analyzed and interpreted based on Fourier analysis and classical MIMO channel results, with numerical examples provided to illustrate the loss in capacity due to reduced-rate sampling. Optimal sampling structures that maximize capacity under reduced-rate sampling are derived under both a filter followed by sampling and sampling following filter banks. These optimal structures are shown to minimize the MSE between the input and a linearly reconstructed signal in Section VI. Proofs of the main theorems (Theorems 2–4) are provided in the appendices. Our notation is summarized in Table I.

II. PRELIMINARIES: CAPACITY OF UNDERSAMPLED CHANNELS

A. Capacity Definition

We consider the same waveform channel model of Gallager [1, Chapter 8]. The transmit signal $x(t)$ is time constrained to the interval $(0, T]$. 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 analog channel output is given as

$$r(t) = h(t) \ast x(t) + \eta(t),$$

and is observed over $(0, T]$, where $\eta(t)$ is stationary zero-mean Gaussian noise, as illustrated in Fig. 1(a). We assume throughout the remainder of the paper that perfect channel state information, i.e.

1We 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, Chapter 8], we first calculate the capacity for finite time intervals and then take the limit of the interval to infinity.
Table I

SUMMARY OF NOTATION AND PARAMETERS

| Symbol | Description |
|--------|-------------|
| $\mathcal{L}_1$ | set of measurable functions $f$ such that $\int |f| \, d\mu < \infty$ |
| $\mathbb{S}_+$ | 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 $i$th prefilter |
| $S_\eta(f), S_x(f)$ | 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 ($T_s = 1/f_s$) |
| $p_i(t)$ | modulating sequence in the $i$th channel |
| $T_p$ | period of the modulating sequence $p_i(t)$ |
| $\|\cdot\|_F, \|\cdot\|_2$ | Frobenius norm, $\ell_2$ norm |
| $\mathbf{v}^T$ | transpose of vector $\mathbf{v}$ |

Boldface used for vectors and matrices.

Perfect knowledge of $h(t)$, is known at both the transmitter and the receiver. The traditional Shannon framework investigates the maximum mutual information between $x(t)$ and $r(t)$ under a power constraint. Specifically, the analog channel capacity is defined as [1, Section 8.1]

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

where the supremum is over all input distributions subject to an average power constraint $\mathbb{E}(\frac{1}{T} \int_0^T |x(\tau)|^2 \, d\tau) \leq P$, and we explicitly indicate the interval $(0,T]$ over which the signal is transmitted and observed. For completeness, we repeat the classical analog capacity result from Gallager as follows.

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

$$C = \frac{1}{2} \int_{f \in \mathcal{F}(\nu)} \log \left( \frac{\nu |H(f)|^2}{S_\eta(f)} \right) \, df,$$

where $\mathcal{F}(\nu)$ and $\nu$ satisfy

$$\mathcal{F}(\nu) = \left\{ f : \frac{S_\eta(f)}{|H(f)|^2} \leq \nu \right\};$$

$$\int_{f \in \mathcal{F}(\nu)} \left[ \nu - \frac{S_\eta(f)}{|H(f)|^2} \right] \, df = P.$$
The channel is frequency-selective where the SNR at each frequency $f$ in the spectral domain is captured by $|H(f)|^2/S_n(f)$. The optimal transmission strategy is to perform water-filling power allocation over frequency. For a channel of bandwidth $B$ over positive frequencies, if we remove the noise outside the channel bandwidth via prefiltering and sample the output at a rate $f \geq 2B$, then we can perfectly recover all spectral contents of the received signal (including transmitted signal and noise) within the channel bandwidth, which allows capacity (2) to be achieved without loss of data rate due to sampling. 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 typically 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 \( \{y_i(t) : 1 \leq i \leq M\} \). 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 \( \{y_i[n] : n \in \mathbb{Z}, 1 \leq i \leq M\} \), as illustrated in Fig. 1(b). Here, each branch is uniformly sampled at a sampling rate of $f_s/M$ samples per second.

Defining the sampled sequence as $y[n] = [y_1[n], \cdots, y_M[n]]$, the problem of finding the capacity $C(f_s)$ of sampled analog channels can be posed as quantifying the maximum mutual information between the input signal $x(t)$ on the interval \((0, T]\) and the output sequence sampled at an aggregate rate $f_s$ on the interval \((0, T]\) in the limit as $T \to \infty$. The sampled channel capacity can thus be expressed as

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

where the supremum is taken over all possible input distributions subject to an average power constraint $E\left( \frac{1}{T} \int_0^T |x(\tau)|^2 \, d\tau \right) \leq P$, and we explicitly indicate the interval \((0, T]\) over which the samples are taken.

The sequence of discrete-time samples depends on the sampling rate and the sampling mechanism we employ, which impacts the information conveyed through this sampled sequence. In our analysis we focus on developing a general analytic framework for sampled channel capacity that accommodates a large class of sampling mechanisms, going beyond ideal uniform Nyquist-rate sampling. For ease of exposition, we organize this paper in incremental steps associated with increasingly complex sampling strategies, where the first steps lay the analytical foundation for the later steps. In particular, starting from sampling following a single filter, we extend our results to incorporate filter banks and modulation
banks, which is the most general class of realizable non-uniform sampling strategies.

B. Sampling Mechanisms

In this subsection, we formally describe the three classes of sampling strategies we investigate.

1) Filtering followed by samplings: Ideal uniform sampling is performed by sampling the analog signal uniformly at a rate \( f_s = T_s^{-1} \). In order to avoid aliasing, suppress out-of-band noise and compensate for distortion, a prefilter is often added prior to the ideal uniform sampler [18]. Adding a prefilter can also be used to model the linear distortion features of practical sampling devices. 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

\[
y(t) = s(t) * (h(t) * x(t) + \eta(t)) \quad t \in (0, T).
\]
We then sample \( y(t) \) using an ideal uniform sampler, leading to the sampled sequence
\[
y[n] = y(nT_s),
\]
where \( T_s \) denotes the sampling interval. The metric of interest is then the maximum mutual information between \( x(0, T) \) and \( \{y[n]\}_{0}^{T} \).

![Figure 2. Filtering followed by sampling: the analog channel output \( r(t) \) is linearly filtered prior to ideal uniform sampling.](image)

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 a rate approaching the Landau rate (i.e. the total widths of spectral support). That is because uniform sampling at sub-Nyquist rate may suppress information by collapsing subbands, resulting in fewer degrees of freedom. This motivates us to investigate certain nonuniform sampling mechanisms. In particular, we now consider the class of non-uniform sampling mechanisms that is most widely used in practice, where the received signal is preprocessed by a bank of filters. Most nonuniform sampling techniques that have been studied in theory and applied in practice \([22], [26], [27]\) fall under filter-bank sampling and modulation-bank sampling (as described in II-B3). Note that the filters may introduce any given delay, so 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 followed by ideal sampling at rate \( f_s/M \) for each branch, as illustrated in Fig. 3. We denote by \( s_i(t) \) and \( S_i(f) \) the impulse response and frequency response of the \( i \)th linear filter, respectively. The filtered analog output in the \( i \)th branch prior to sampling is then given as
\[
y_i(t) = (h(t) \ast s_i(t)) \ast x(t) + s_i(t) \ast \eta(t), \quad t \in (0, T].
\]
These filtered signals are then passed through \( M \) ideal samplers to yield
\[
y_i[n] \overset{\Delta}{=} y_i(nMT_s) \quad \text{and} \quad y[n] \overset{\Delta}{=} \begin{bmatrix} y_1[n], y_2[n], \ldots, y_M[n] \end{bmatrix},
\]
where \( T_s = f_s^{-1} \). The capacity now is the maximum mutual information between \( x(0, T) \) and \( y[n] \).
3) **Modulation and Filter Banks Followed by Sampling:** We generalize a filter bank followed by sampling 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 $M$ different branches. In the $i$th branch, the received analog 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 structures. By taking advantage of random modulation sequences to achieve incoherence among different branches, this sampling mechanism has proven useful for sub-sampling analog multiband signals by exploiting spectral sparsity [30]. Note that as with previous methods, the filters can introduce arbitrary delay, so that the branches may be sampled at different times.

In the $i$th branch, the received prefiltered analog signal in the time interval $(0, T]$ prior to sampling can be written as

$$y_i(t) = s_i(t) \ast (q_i(t) \cdot (p_i(t) \ast h(t) \ast x(t) + p_i(t) \ast \eta(t))),$$

resulting in the digital sequence of samples

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

---

**Figure 3.** A filter bank followed by sampling: the received analog signal $r(t)$ is passed through $M$ branches. In the $i$th branch, the signal $r(t)$ is passed through an LTI prefilter with frequency response $S_i(f)$, and then sampled uniformly by an ideal uniform sampler.
Figure 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 $f_s/M$.

C. Sampling Preliminaries

Before proceeding, we recall some preliminaries from sampling theory. Suppose that a sampled sequence $x[n]$ of a continuous-time signal $x(t)$ is obtained by sampling $x(t)$ at a rate $f_s$, i.e., $x[n] = x(nT_s)$. The discrete time Fourier transform of $x[n]$ is given by

$$\frac{1}{T_s} \sum_{n=\infty}^{\infty} X(f - nf_s),$$

(10)

where $X(f) \overset{\Delta}{=} \int_{-\infty}^{\infty} x(t) \exp(-j2\pi ft) dt$ is the continuous-time Fourier transform of $x(t)$. In other words, the uniformly sampled signal depends on the periodic extension of the original signal in the frequency domain.

For convenience of exposition, we introduce the following notation of an infinite-dimensional vector, which is the sampled vector associated with $X(f)$:

$$V_X(f, f_s) = \frac{1}{\sqrt{T_s}} \left[ \cdots, X(f - f_s), X(f), X(f + f_s), \cdots \right]^T$$

(11)

such that the $l$th $(l \in \mathbb{Z})$ coordinate of $V_X(f, f_s)$ is $X(f + lf_s)$. If $x(t)$ is uniformly sampled with sampling rate $f_s$, then the discrete time Fourier transform of $x[n]$ can be written as

$$\frac{1}{\sqrt{T_s}} V_X^T(f, f_s) \mathbf{1},$$

(12)

where $\mathbf{1}$ is an infinite vector whose entries are all equal to 1. We will see later that the $\ell_2$ norm of the sampled vector plays a key role in the capacity of the sampled analog channel. This norm is given by

$$\|V_X(f, f_s)\|_2 \overset{\Delta}{=} \frac{1}{\sqrt{T_s}} \left( \sum_{l=\infty}^{\infty} |X(f - lf_s)|^2 \right)^{\frac{1}{2}},$$

(13)
which characterizes the folded signal magnitude at frequency $f$. For notational convenience, we use

$$V_{XY + Z}(f, f_s) \triangleq \frac{1}{\sqrt{T_s}} [\cdots, X(f - f_s)Y(f - f_s) + Z(f - f_s), X(f)Y(f) + Z(f), \cdots]^T$$  \hspace{1cm} (14)

to denote the sampled vector associated with $X(f)Y(f) + Z(f)$. In particular, if $Z(f) = 0$ for all $f$, then we have the notion

$$V_{XY}(f, f_s) \triangleq \frac{1}{\sqrt{T_s}} [\cdots, X(f - f_s)Y(f - f_s), X(f)Y(f), \cdots]^T.$$ \hspace{1cm} (15)

D. Exposition Outline

The main results for the three sampling strategies introduced in Section II-B are formally stated and analyzed in the following sections. For each scenario, we first provide an approximate treatment based on Fourier analysis by relating the sampled channel to a traditional MISO or MIMO Gaussian channel. This approach, while not strictly rigorous, allows for a more intuitive and informative understanding of our results from a communication theoretic perspective. We also provide complementary interpretations of these results along with several numerical examples. In particular, we analyze how to optimize capacity for both filtering followed by sampling and sampling following filter banks, and interpret these optimization results from both information theoretic and sampling theoretic viewpoints. The rigorous analyses of the main theorems are deferred to the appendices, which make heavy use of asymptotic spectral properties of Toeplitz and block-Toeplitz matrices. This method of exposition, with an approximate analysis based on Fourier analysis followed by a rigorous treatment using Toeplitz properties, is similar to the exposition of waveform channel capacity used by Gallager in [1, Chapter 8].

III. A Filter Followed by Sampling

A. Main Results

Applying a prefilter generates a new equivalent channel with channel gain $H(f)S(f)$. The noise is also prefiltered and is therefore non-white in general. The ideal uniform sampler that follows the prefilter creates an aliased version of the prefiltered signal in the frequency domain, as reflected in the following capacity expression.

**Theorem 2.** Consider the system shown Fig. 2, where $\eta(t)$ is Gaussian noise with power spectral density $S_\eta(f)$. Assume that $h(t), s(t)$ are both continuous, bounded and absolutely Riemann integrable, and that there exists some constant $\epsilon_s$ such that

$$\left\|V_S(\sqrt{S_\eta}(f, f_s)) \right\|_2^2 = \sum_{l=-\infty}^{\infty} |S(f - lf_s)|^2 S_\eta(f - lf_s) \geq \epsilon_s > 0$$ \hspace{1cm} (16)
holds. 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 \( \epsilon > 1 \). The capacity \( C(f_s) \) of the sampled channel with a power constraint \( P \) is then given parametrically as

\[
C(f_s) = \frac{1}{2} \int_{f \in \mathcal{F}(\nu)} \log \left( \frac{\sum_{l=\infty}^{\infty} |H(f - lf_s)S(f - lf_s)|^2}{\nu} \right) df = \frac{1}{2} \int_{f \in \mathcal{F}(\nu)} \log \left( \frac{\nu \|VHS(f, f_s)\|_2^2}{\nu \|V S \sqrt{S_\eta}(f, f_s)\|_2^2} \right) df
\]

where

\[
\mathcal{F}(\nu) = \left\{ f : \frac{\|V S \sqrt{S_\eta}(f, f_s)\|_2^2}{\nu \|VHS(f, f_s)\|_2^2} \leq \nu \text{ and } f \in \left[ -\frac{f_s}{2}, \frac{f_s}{2} \right] \right\}
\]

and \( \nu \) satisfies

\[
\int_{f \in \mathcal{F}(\nu)} \left[ \nu - \frac{\|V S \sqrt{S_\eta}(f, f_s)\|_2^2}{\nu \|VHS(f, f_s)\|_2^2} \right] df = P.
\]

**Remark 1.** The assumption (16) ensures that \( \|V S \sqrt{S_\eta}\|_2 \) is bounded away from zero ensures that the filter response satisfies \( s(t) \neq 0 \) for all \( t \).

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 (17) at each \( f \in \left[ -\frac{f_s}{2}, \frac{f_s}{2} \right] \) within the sampling bandwidth. The linear time invariance of both the channel and the prefilter response leads to an equivalent frequency-selective channel, and the ideal uniform sampling that follows generates a folded version of the non-sampled channel capacity. Specifically, this capacity expression differs from the analog capacity given in Theorem 1 in that the SNR in the sampled scenario is \( \gamma_s(f) := \frac{\|VHS(f, f_s)\|_2^2}{\|V S \sqrt{S_\eta}(f, f_s)\|_2^2} \) in contrast to \( \gamma_0(f) := \frac{|H(f)|^2}{S_\eta(f)} \) for the non-sampled scenario. Water filling over the inverse sampled SNR \( \gamma_s^{-1}(f) \) is the optimal power allocations.

\[\text{This condition is used in Appendix A as a sufficient condition to guarantee asymptotic properties of Toeplitz matrices. A similar condition will be used in Theorems 3 and 4.}\]
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 interpretation as in [1, Chapter 8.3]. The rigorous analysis, which is deferred to Appendix A, makes use of a discretization argument and asymptotic spectral properties of Toeplitz matrices.

Consider first the equivalence between the sampled channel and a MISO channel at a single frequency \( f \in [-f_s/2, f_s/2] \). Suppose the transmitted signal has a frequency response \( X(f) \). The Fourier transform of the sampled signal is given by

\[
\frac{1}{T_s} \sum_{k=-\infty}^{+\infty} H(f - kf_s) S(f - kf_s) X(f - kf_s), \quad \forall f \in \left[ -\frac{f_s}{2}, \frac{f_s}{2} \right]
\]

(21)
due to aliasing. The summing operation allows us to treat the aliased channel at each frequency \( f \) within the sampling bandwidth as a separate MISO channel with countably many input branches and a single output branch, as illustrated in Fig. 5(a).

By assumption, the noise is of spectral density \( S_\eta(f) \), and hence the prefiltered noise has power spectral density \( S_\eta(f) |S(f)|^2 \). The power spectral density of the sampled noise sequence at \( f \in [-f_s/2, f_s/2] \) is then given by

\[
\| V_S \sqrt{S_\eta}(f, f_s) \|^2 = \sum_{l=-\infty}^{\infty} S_\eta(f - lf_s) |S(f - lf_s)|^2.
\]

If we term \( \{ f - lf_s : l \in \mathbb{Z} \} \) 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. It follows from known results [48] that the MISO channel effectively has only one degree of freedom, and that the capacity-achieving strategy for a MISO Gaussian channel, which is often referred to as transmit maximum ratio combining (MRC) or beamforming, exploits the transmit diversity to maximize the received SNR. Specifically, denote by \( G(f) \) the transmitted signal for each \( f \in [-f_s/2, f_s/2] \). This signal is multiplied by a constant gain \( c\alpha_l(l \in \mathbb{Z}) \), and sent through the \( l \)th input branch, i.e.

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

(22)

where \( \alpha_l = \frac{H^*(f-lf_s)S^*(f-lf_s)}{\| V_{HS}(f, f_s) \|^2} \) and \( c \) is a normalizing constant determined by the power constraint. The resulting SNR can be expressed as the sum of SNRs (as shown in [48]) at each branch

\[
c^2 \frac{\| V_{HS}(f, f_s) \|^2}{\| V_S \sqrt{S_\eta}(f, f_s) \|^2}.
\]

This is an approximate analysis since the Fourier transform of the input signal may not even exist. The proof we provide later does not use Fourier analysis but rather the convergence properties of Toeplitz operators.
Since the sampling operation combines signal components at frequencies from each aliased set \(\{f - lf_s : l \in \mathbb{Z}\}\), it is equivalent to having a set of parallel MISO channels, each indexed by some \(f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]\). Since each MISO channel has one degree of freedom, it can be converted to a set of parallel SISO channels, where the channel at \(f\) has an equivalent channel gain \(\tilde{H}(f) = \|V_{HS}(f, f_s)\|_2\), as illustrated in Fig. 5(b). The water-filling strategy is optimal in allocating power among the set of parallel channels, which yields the parametric equations (19) and (20) and completes our approximate analysis.

Figure 5. Equivalent representations for filtering followed by sampling: (a) Equivalent MISO Gaussian channel for a given \(f \in [-f_s/2, f_s/2]\); (b) The equivalent set of parallel SISO channels representing all \(f \in [-f_s/2, f_s/2]\), where the SISO channel at a given frequency is equivalent to the MISO channel in Fig. 5(a).

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] 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 Gallager’s approach [1 Chapter 8.4] does not accommodate 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 process in the time domain by point-wise sampling via an interval $\Delta$, 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. Thanks to the continuity assumption, an exact capacity expression can be obtained by letting $\Delta$ go to zero.

2) Spectral properties of block-Toeplitz matrices: After discretization, the input-output relation is similar to a MIMO discrete system. Applying MIMO channel capacity results leads to the capacity for a given $T$ and $\Delta$. The channel capacity is then obtained by taking $T$ to infinity and $\Delta$ to zero, which can be related to the channel matrix’s spectrum using Toeplitz theory. Since the prefiltered noise is non-white 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 introduce the concept of asymptotic equivalence 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: As we can see from Theorem 2 different prefilters lead to different channel capacities. A natural question then is how to choose $S(f)$ to maximize capacity under filtering followed by sampling. The optimizing prefilter is given in the following corollary.

**Corollary 1.** Consider the system shown in Fig. 2. Suppose that in each aliased set $\{f - lf_s : l \in \mathbb{Z}\}$, there exists $k$ such that $\frac{|H(f-kf_s)|^2}{S_n(f-kf_s)} = \sup_{l \in \mathbb{Z}} \frac{|H(f-lf_s)|^2}{S_n(f-lf_s)}$. Then the capacity in (17) is maximized by the filter with frequency response

$$S(f-kf_s) = \begin{cases} 
1, & \text{if } \frac{|H(f-kf_s)|^2}{S_n(f-kf_s)} = \sup_{l \in \mathbb{Z}} \frac{|H(f-lf_s)|^2}{S_n(f-lf_s)}, \\
0, & \text{otherwise},
\end{cases}$$

for any $f \in [-f_s/2, f_s/2]$ and any $k \in \mathbb{Z}$.

**Proof:** It can be observed from (17) 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 $f_2$ ($f_1 \neq f_2; f_1, f_2 \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$)
separately. Specifically, the SNR at each \( f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right] \) in the aliased channel is given by

\[
\text{SNR}(f) := \frac{\| V_{HS}(f,f_s) \|_2^2}{\| V_{S\sqrt{\mathcal{S}_\eta}}(f,f_s) \|_2^2} = \sum_{l=-\infty}^{+\infty} \frac{|H(f-lf_s)|^2}{\mathcal{S}_\eta(f-lf_s)} \lambda_l,
\]

(25)

where

\[
\lambda_l = \frac{\mathcal{S}_\eta(f-lf_s) |S(f-lf_s)|^2}{T_s \| V_{S\sqrt{\mathcal{S}_\eta}}(f,f_s) \|_2^2}.
\]

Note that \( 0 \leq \lambda_l \leq 1 \) and \( \sum_l \lambda_l = 1 \). Thus, \( \text{SNR}(f) \) is a convex combination of \( \left\{ \frac{|H(f-lf_s)|^2}{\mathcal{S}_\eta(f-lf_s)}, l \in \mathbb{Z} \right\} \), which is upper bounded by

\[
\text{SNR}_{\max}(f) = \sup_{l \in \mathbb{Z}} \frac{|H(f-lf_s)|^2}{\mathcal{S}_\eta(f-lf_s)}.
\]

(26)

This bound can be attained by the filter given in (24). In other words, the optimal prefilter puts all its mass in those frequency components with the highest SNR within each aliased frequency set \( \{f-lf_s : l \in \mathbb{Z}\} \).

2) Interpretations: Recall that \( S(f) \) is assumed to be right-invertible and is applied after the noise is added. In the analog channel, the prefilter would become useless in terms of capacity benefits since we are always able to recover the non-filtered signal by applying an inverse filter on \( y(t) \), i.e. by the data processing inequality no addition information can be obtained by filtering the received signal. However, in the aliased channel, the prefiltering operation is non-invertible. As we show 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 SNR should be given higher weight, while those that suffer from poor channel gain should be suppressed.

The problem of finding optimal prefilters can indeed be posed as a joint optimization over all input and filter responses. Looking at the equivalent aliased channel for a given frequency \( f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right] \) as illustrated in Fig. 5(a), we have full control over both \( X(f) \) and \( S(f) \). The channel associated with the frequency \( f \) differs from a standard MISO channel model in that the prefiltering \( S(f) \) allows us to weight the branch input gain to the combiner. Thus it is equivalent to applying transmit beamforming (i.e. transmit branch weighting) and receiver shaping (i.e. multiplication by \( S(f) \)) followed by combining in a MIMO Gaussian channel. A related joint optimization problem over both prefiltering and postfiltering has been investigated in [40], [41], but their approach is based on the MSE measure instead of an information theoretic metric. We will discuss the relation with MSE optimization later in Section VI.

Although MRC at the transmitter side maximizes the combiner SNR for a MISO channel [48], it turns out to be suboptimal for our joint optimization problem. The optimal solution is to perform
selection combining [48] by setting $S(f - lf)$ to one for some $l = l_0$, as well as noise suppression by setting $S(f - lf)$ to zero for all other $l$s. The prefilter outputs the signal on the branch with the highest SNR, and suppresses noise from all remaining branches. The resulting combiner SNR becomes $\max_l \text{SNR}_l$ times the number of branches, which exceeds the SNR achieved by MRC ($\sum_l \text{SNR}_l$). Here, $\text{SNR}_l$ denotes the channel gain $|H(f - lf)|^2$ divided by the noise. 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 aliased-free channel. After passing through an optimal prefilter, all frequencies modulo $f_s$ except the one with the highest SNR are removed. Unless there exist multiple branches that possess the highest SNR, the optimal filtering followed by sampling indeed suppresses aliasing as well as noise. This alias-suppressing phenomena, while different from many sub-Nyquist works that advocate mixing instead of alias suppressing [28], [30], arises from the fact that we have control over the input shape. When the input is given, the prefilter that maximizes the mutual information is the MRC type filter which indeed mixes different frequency components from each aliased set. But a joint optimization over both the input and the prefilter yields an input whose frequency support is equal to the sampling bandwidth, thus resulting in an alias-suppressing filter in order to remove noise.

E. Numerical examples

1) Additive Gaussian Noise Channel without Prefiltering: The first numerical example we consider is an additive Gaussian noise channel. The channel gain is flat within the channel bandwidth $B$ (here, we set $B = 0.5$), i.e. $H(f) = 1$ if $f \in [-B,B]$ and $H(f) = 0$ otherwise. The noise process is modeled as a measurable and stationary Gaussian process with the power spectral density plotted in Fig. 6. In fact, this is the noise model adopted by Lapidoth in [49] to approximate white noise, which avoids the infinite variance of the standard model for unfiltered white noise. In this example, we employ ideal point-wise 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
Fig. 7 illustrates the capacity-sampling tradeoff curve for the raised-cosine channel, for different roll-off factors. Prefilter for a monotone channel reduces to a low-pass filter with cutoff frequency whose channel response obeys a monotone channel. For certain classes of channel models, this channel class is a discontinuous, which may be hard to realize in practice. However, for certain classes of channel models, information can be harvested. 

Increasing the sampling rate above twice the channel bandwidth already preserves all contents of the channel output – no further increase further when the sampling rate exceeds twice the noise bandwidth, since oversampling at any rate above twice the noise bandwidth already preserves all contents of the channel output – no further information can be harvested.

2) Optimally Prefiltered 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 \( \frac{|H(f_1)|^2}{S_n(f_1)} \geq \frac{|H(f_2)|^2}{S_n(f_2)} \) for any \( f_1 > f_2 \). Corollary \( \square \) implies that the optimizing prefilter for a monotone channel reduces to a low-pass filter with cutoff frequency \( f_s/2 \). As an example, Fig. \( \square \) illustrates the capacity-sampling tradeoff curve for the raised-cosine channel, for different roll-off factors.

Figure 6. Additive Gaussian noise channel. The channel gain and the power spectral density of the noise is plotted in the left plot. The sampling mechanism employed here is ideal uniform sampling without filtering. The power constraint is \( P = 5 \). The sampled capacity, as illustrated in the right plot, does not achieve analog capacity when sampling at a rate equal to twice the channel bandwidth, but does achieve it when sampling at a rate equal to twice the noise bandwidth.
The frequency response of the channel is given by

\[
H(f) = \begin{cases} 
  T, & |f| \leq \frac{1-\beta}{2T}, \\
  \frac{T}{2} \left[1 + \cos \left( \frac{\pi T}{2} \left[ |f| - \frac{1-\beta}{2T} \right] \right) \right], & \frac{1-\beta}{2T} \leq |f| \leq \frac{1}{2},
\end{cases}
\]

(27)

where \( \beta \) denotes the roll-off factor and \( T \) is a given period. It can be observed that below the Nyquist rate, capacity increases with \( f_s \) since the effective sampling bandwidth increases, while oversampling beyond the Nyquist rate does not increase capacity. As expected, sampling at or above the Nyquist rate creates an alias-free capacity expression that can be simplified as

\[
C(f_s) = \frac{1}{2} \int_{f \in F(\nu)} \log \left( \nu \frac{|H(f)|^2}{\mathcal{S}_\eta(f)} \right) \, df,
\]

(28)

which equals the classical Nyquist-rate (i.e. the analog) channel capacity derived in [1]. For non-monotone channels, the optimal prefilter may not be a low-pass filter, as illustrated in Fig. [8] We plot in Fig. [8] b) the optimal filter for the channel given in Fig. [8] a) when the sampling rate \( f_s = 0.4f_{NYQ} \). It can be seen that this filter is no longer a low-pass filter but is of support size \( 0.4f_{NYQ} \).

![Channel Capacity vs Sampling Rate](image)

Figure 7. The sampled channel capacity vs sampling rate for a raised-cosine channel with an optimal prefilter. The channel bandwidth is assumed to be \([-\frac{1}{2}, \frac{1}{2}]\), the power constraint \( P = 10 \), and the noise is white with flat spectral density \( \sigma^2_\eta = 1 \). The frequency response \( H(f) \) of the channel is assumed to be a raised cosine function with \( \beta = 0.9 \) and \( T = 1.6 \). The tradeoff curves for two types of prefilters are illustrated: (1) the optimal prefilter (which is a low-pass filter); (2) the matched filter whose frequency response obeys \( S(f) = H^*(f) \). In the sub-Nyquist sampling rate regime, the optimal prefilter outperforms the matched filter, while the two curves coincide when sampling is performed above the Nyquist rate.

3) Capacity Non-monotonicity: When the channel is not monotone, a somewhat counter-intuitive fact arises: the channel capacity \( C(f_s) \) is not necessarily a non-decreasing function of the sampling rate \( f_s \). Examples include specific multiband channels as illustrated in Fig. [9] Here, the Fourier transform of the channel response is concentrated in two sub-intervals. Specifically, assuming that the entire bandwidth is
If the channel is sampled at a rate $f_s = \frac{3}{5} f_{\text{NYQ}}$, 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 $f_s = \frac{2}{5} f_{\text{NYQ}}$, it can be easily verified that the two subbands remain non-overlapping 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. 9. 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.

Figure 8. Capacity of optimally prefiltered channel: (a) frequency response of the original channel; (b) optimal prefilter associated with this channel for sampling rate 0.4; (c) optimally prefiltered channel response with sampling rate 0.4; (d) capacity vs sampling rate for the optimal prefilter and for the matched filter. The optimal prefilter has support size $f_s$ in the frequency domain, hence its output is alias-free. In the sub-Nyquist regime, this alias-suppressing filter outperforms the matched filter in the resulting capacity.
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 multiple branches. Since the sampled output at these branches are all functions of the same input and noise and hence mutually dependent, the optimal transmission scheme must account for their correlation. Specifically, the transmit signal should be chosen to decouple mutual interference across different branches. This is reflected in the capacity expression given in Theorem 3.

In order to state our theorem formally, we introduce two Fourier symbol matrices $F_s$ and $F_h$. Here, $F_s$ is an infinite matrix of $m$ rows and infinite columns and $F_h$ is a diagonal infinite matrix such that

$$\forall 1 \leq i \leq k, \forall l \in \mathbb{Z}: \begin{cases} (F_s(f))_{i,l} = S_i \left( f - \frac{lf}{M} \right) \sqrt{S_\eta \left( f - \frac{lf}{M} \right)}, \\ (F_h(f))_{l,l} = \frac{H(f - \frac{lf}{M})}{\sqrt{S_\eta(f - \frac{lf}{M})}}. \end{cases}$$ (30)

**Theorem 3.** 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 $h(\eta) := \mathcal{F}^{-1} \left( \frac{H(f)}{\sqrt{S_\eta(f)}} \right)$.
satisfies $h_\eta(t) = o(t^{-\epsilon})$ for some constant $\epsilon > 1$, and that $F_s$ is right-invertible for every $f$. The capacity $C(f_s)$ of the sampled channel with a power constraint $P$ can be given as

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

where

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

**Remark 2.** Using the same argument as used by Telatar in [10], we can express this capacity alternatively as

$$C(f_s) = \max_{\{Q(f)\} \in Q} \int_{-f_s/2M}^{f_s/2M} \frac{1}{2} \log \det \left( I_M + \tilde{F}_s F_h Q F_h^* \tilde{F}_s^* \right) \, df,$$

(31)

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

$$Q = \left\{ \left\{ Q(f) : |f| \leq \frac{f_s}{2M}, Q(f) \in \mathbb{S}_+ \right\} : \int_{-f_s/2M}^{f_s/2M} \text{Tr}(Q(f)) \, df = P \right\}.$$

(32)

The optimal $\{Q(f)\}$ corresponds to a water-filling power allocation strategy based on the singular values of the equivalent channel matrix $\tilde{F}_s F_h$, where $F_h$ is associated with the original channel and $\tilde{F}_s$ arises from prefiltering and noise whitening. For each $f \in [-f_s/2M, f_s/2M]$, the integrand in (31) can be interpreted as a MIMO Gaussian channel capacity formula with degrees of freedom associated with the frequency domain, as illustrated in Fig. 10(a). We still have full control over a countable number of input branches $\left\{ X \left( f - \frac{lf_s}{2M} \right) \mid l \in \mathbb{Z} \right\}$, but this time we have $M$ receive branches instead of a single branch (as in the MISO case for sampling following a single filter). The channel capacity can be achieved when the transmit signals are designed to decouple this MIMO channel into $M$ parallel channels (and hence $M$ degrees of freedom), each corresponding to one of its singular directions. Unlike traditional MIMO Gaussian channels, the noise samples in each output sample set $\{y_i[n] : 1 \leq i \leq M\}$ result from the same process $\eta(t)$ (as shown in Fig. 3(a)) and hence noise samples are correlated.

**B. Approximate Analysis**

The sampled analog channel under filter banks followed by sampling can be studied through its connection with MIMO Gaussian channels (see Fig. 10). Consider first a single frequency $f \in [-f_s/2M, f_s/2M]$. Since we employ a bank of filters with each filter followed by an ideal uniform
MIMO Gaussian channel capacity results [10] immediately imply that the channel capacity at a given frequency \( f \) has the following channel matrix

\[
(F_s(f)F_s^*(f))^{-1} F_s(f)F_h(f) = \tilde{F}_s(f)F_h(f).
\]

(34)

MIMO Gaussian channel capacity results [10] immediately imply that the channel capacity at a given frequency \( f \) has the following channel matrix

\[
\sum_{l=-\infty}^{\infty} \left| S_l \left( f - \frac{lf_s}{M} \right) \right|^2 S_n \left( f - \frac{lf_s}{M} \right), \quad \left( -\frac{f_s}{2M} \leq f \leq \frac{f_s}{2M} \right)
\]

indicating the mutual correlation of noise at different branches. The received noise vector can be whitened by multiplying \( Y(f) = [\cdots, Y(f), Y(f - f_s), \cdots]^T \) by an \( M \times M \) whitening matrix \((F_s(f)F_s^*(f))^{-1}\). Since whitening here is an invertible operation, it preserves capacity. After whitening, the channel of Fig. [10(a)] associated with frequency \( f \) can be expressed as

\[
\max \frac{1}{Q} \log \det \left[ I + \tilde{F}_s(f)F_h(f)Q(f)F_h^*(f)\tilde{F}_s^*(f) \right]
\]

subject to the constraints \( \text{trace}(Q(f)) \leq P(f) \) and \( Q(f) \in S_+ \), where \( Q(f) \) denotes the power allocation matrix. Ranging over all \( f \in [-f_s/2M, f_s/2M] \), we have the set of parallel MIMO channels for each frequency \( f \) illustrated in Fig. [10(b)], where each MIMO channel in this figure is equivalent to the set of parallel channels in Fig. [10(a)] for the given frequency. Performing the water-filling power allocation strategy across all parallel channels leads to our capacity expression.

Figure 10. Equivalent representations for a bank of \( M \) filters followed by sampling: (a) Equivalent MIMO Gaussian channel for a single \( f \in [-f_s/2M, f_s/2M] \); (b) An equivalent set of parallel channels representing all \( f \in [-f_s/2M, f_s/2M] \), where the MIMO channel at a given frequency \( f \) is equivalent to the MIMO channel of Fig. [10(a)].
C. Optimal Filter Bank

1) Derivation of optimal banks of filters: In general, \( \log \det \left[ I_M + \tilde{F}_s \tilde{F}_h \tilde{F}_s^* \right] \) is not perfectly determined by \( \tilde{F}_s(f) \) and \( F_h(f) \) at a single frequency \( f \), but also depends on the water-level due to the fact that the optimal power allocation strategy relies on the power constraint \( P/\sigma^2 \eta \) as well as \( F_s \) and \( F_h \) across all \( f \). In other words, \( \log \det \left[ I_M + \tilde{F}_s \tilde{F}_h \tilde{F}_s^* \right] \) is a function of all singular values of \( \tilde{F}_s \tilde{F}_h \) 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 element-wise 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 Corollary 2.

Since \( F_h(f) \) is a diagonal matrix, \( \lambda_k (F_h F_h^*) \) denotes the \( k \)th largest entry of \( F_h F_h^* \) or, equivalently, the \( k \)th largest element in the set \( \{ |H(f - lf_s M)|^2 : l \in \mathbb{Z} \} \). The optimal filter bank can then be given as follows.

**Corollary 2.** Consider the system shown in Fig. 3. Suppose that for each aliased set \( \{ f - if_s M : i \in \mathbb{Z} \} \) and each \( k \) \((1 \leq k \leq M)\), there exists an integer \( l \) such that \( |H(f - lf_s M)|^2 \) is equal to the \( k \)th largest element in \( \{ |H(f - if_s M)|^2 : i \in \mathbb{Z} \} \). The capacity (31) under filter-bank sampling can be maximized by a bank of filters for which the frequency response of the \( k \)th filter is given by

\[
S_k \left( f - \frac{lf_s}{M} \right) = \begin{cases} 
1, & \text{if } |H \left( f - \frac{lf_s}{M} \right)|^2 = \lambda_k \left( F_h(f) F_h^*(f) \right); \\
0, & \text{otherwise},
\end{cases}
\]  

(36)

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

\[
C(f_s) = \frac{1}{2} \int_{-f_s/2M}^{f_s/2M} \sum_{k=1}^{M} (\log (\nu - \lambda_k (F_h F_h^*)))_+ \, df,
\]  

(37)

where \( \nu \) is chosen such that

\[
\int_{-f_s/2M}^{f_s/2M} \sum_{k=1}^{M} [\nu - \lambda_k (F_h F_h^*)]_+ \, df = P.
\]  

(38)

Here, we use the notation \( (x)_+ \) to denote \( \max \{ x, 0 \} \).

**Proof:** See Appendix D.

The choice of prefilters in (36) achieves the upper bounds on all singular values, and is hence universally optimal regardless of the water level. Since \( \tilde{F}_s \) has orthonormal rows, it acts as an orthogonal projection.
and outputs the $M$-dimensional subspace closest to the channel space spanned by $F_h$. By observing that the rows of the diagonal matrix $F_h$ are orthogonal to each other, the subspace with the strongest signal strength corresponds to the $M$ rows of $F_h$ that contain the highest channel gain out of the entire aliased frequency set $\{ f - \frac{l f_s}{M} : l \in \mathbb{Z} \}$. Hence, the maximum data rate is 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. Define the aliased frequency set $B_k(f) = \{ f + k \frac{f_s}{M} + l f_s : l \in \mathbb{Z} \}$. Then sampling following a single optimal filter selects the strongest frequency from each set $B_k(f)$ $(0 \leq k < M)$, while sampling following an optimal filter bank selects the $M$ largest points from $B_0(f) \cup B_1(f) \cup \cdots \cup B_{M-1}(f)$, and hence outperforms a single-branch with a filter followed by sampling.

To illustrate this, consider the example given in Fig. 11 where we compare sampling following a single filter and sampling following two filters, with optimal filters designed in each case. Consider two aliased sets $B_1 = \{ f - l f_s : l \in \mathbb{Z} \}$ and $B_2 = \{ f + \frac{f_s}{M} - l f_s : l \in \mathbb{Z} \}$. Single-branch sampling extracts out the frequency with the best SNR from $B_1$ and another one from $B_2$, while two-branch sampling can select the two frequencies with the best SNRs from $B_1 \cup B_2$. In the example shown in Fig. 11 the latter corresponds to selecting two frequencies from $B_1$, which strictly outperforms sampling following a single branch.

D. Discussion and Numerical Examples

We note that 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 cut-off frequency $f_s/2$, which is equivalent to single-branch sampling with an optimal filter. In other words, using filter banks harvests no gain in capacity compared to a single branch with a filter followed by sampling. In this case, the sampled capacity with the optimal filter bank increases monotonically with the sampling rate up to the Nyquist-rate capacity.

For more general channels, however, the capacity is not a monotone function of $f_s$. Consider again the multiband sparse channel where the channel response is concentrated in two sub-intervals, as illustrated in Fig. 9(a). As discussed above, sampling following a single filter only allows us to select the best $f$ out of the set $\{ f - l f_s : l \in \mathbb{Z} \}$, while sampling following filter banks allows us to select the best $f$ out of the set $\{ f - l \frac{f_s}{M} : l \in \mathbb{Z} \}$. For example, when many frequencies in the set $\{ f - l f_s : l \in \mathbb{Z} \}$ have higher channel gain than all points in the set $\{ f + \frac{f_s}{M} - l f_s : l \in \mathbb{Z} \}$, filter-bank sampling allows these desirable frequencies to be used for multiple branches. In the single-filter sampling, however, at most one from each set can be effectively used. Thus, the sampled channel capacity with a filter bank exceeds
Figure 11. Sampling following a single filter v.s. sampling following two filters. The blue solid lines represent the SNRs at frequencies in the aliased frequency set \( \{ f - lf_s : l \in \mathbb{Z} \} \), while the red dotted lines represent the SNRs at frequencies in the aliased set \( \{ f + \frac{f_s}{2} - lf_s : l \in \mathbb{Z} \} \). Sampling following an optimal prefilter must select the frequency with the best SNR from each aliased set separately, while sampling following 2 filters can simultaneously select two frequencies from the same aliased set, thus strictly outperforming sampling following a single filter.

that of with a single filter, but neither capacity is monotonically increasing in \( f_s \). This is shown in Fig. 9b). Specifically, we see in this figure that when we apply a bank of two filters prior to sampling, the capacity curve is still non-monotonic but outperforms a single filter followed by sampling.

Another consequence of our results is that when the number of prefilters is appropriately 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 \( f \), \( \{ \tilde{H}(f - lf_s) : l \in \mathbb{Z} \} = \{ H(f - lf_s) : l \in \mathbb{Z} \} \). In other words, \( \cdots, \tilde{H}(f - f_s), \tilde{H}(f), \tilde{H}(f + f_s), \cdots \) is a permutation of \( \cdots, H(f - f_s), H(f), H(f + f_s), \cdots \) for every \( f \). 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. This is an immediate consequence of the data processing inequality which implies that permutation of the channel response at rate \( f_s/M \) preserves capacity.
Proposition 1. If there exists a permutation $\tilde{H}(f)$ of $H(f)$ at rate $f_s/m$ such that the support of $\tilde{H}(f)$ is $\left[ -\frac{f_L}{2}, \frac{f_L}{2} \right]$, 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 non-zero channel gain. For any $f_s \geq f_L = \frac{K}{N} f_{NYQ}$, we are always able to permutate the channel at rate $f_s/K$ to generate a band-limited channel of bandwidth $f_L$. Hence, sampling above the Landau rate following $K$ filters achieves the Nyquist-rate channel capacity. This is illustrated in Fig. 9 where a four-branch filter bank followed by sampling has a higher capacity than that with a single prefilter followed by sampling, and achieves the Nyquist-rate capacity whenever $f_s \geq \frac{2}{5} f_{NYQ}$.

V. Modulation and Filter Banks Followed by Sampling

A. Main Results

We now treat modulation and filter banks followed by sampling. The Fourier transform of each of the periodic modulation sequences $q_i(t)$ is a delta train with spacing equal to the inverse period $1/T_q$. Since multiplication in the time domain corresponds to convolution in the spectral domain, the modulation bank mixes frequency components among different aliased sets. This is reflected in Theorem 4 that characterizes the sampled analog channel capacity.

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 $F^\eta$ and $F^h$. The $aM \times \infty$-dimensional matrix $F^\eta$ contains $M$ submatrices with the $\alpha$th submatrix given by an $a \times \infty$-dimensional matrix $F^\eta_{\alpha}$.

$$F^\eta_{\alpha,l,v} = \sum_{u} c^{u}_{\alpha} S_{\alpha} \left( -f + uf_q + v \frac{f_q}{b} \right) \exp \left( -j2\pi lMT_s \left( f - uf_q - v \frac{f_q}{b} \right) \right).$$

Also, $F^p_{\alpha,l,l} = P_{\alpha} \left( -f + l \frac{f_q}{b} \right) \sqrt{S_{\eta} \left( -f + l \frac{f_q}{b} \right)}$ and $F^h_{\alpha,l,l} = \frac{H \left( -f + l \frac{f_q}{b} \right)}{\sqrt{S_{\eta} \left( -f + l \frac{f_q}{b} \right)}}$.

Theorem 4. 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, $F^\eta$ 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 $h_\eta(t) := F^{-1} \left( H(f) / \sqrt{S_{\eta}(f)} \right)$
satisfies \( h_\eta(t) = a(t^{-\epsilon}) \) for some constant \( \epsilon > 1 \). We also 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

\[
C(f_s) = \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \frac{1}{2} \sum_{i=1}^{aM} \left[ \log \left( \nu \lambda_i \left( (F^\eta F^\eta^*) - \frac{1}{2} F^\eta F^h F^h^* F^\eta F^\eta^* (F^\eta F^\eta^*)^{-\frac{1}{2}} \right)^+ \right) \right] \, df,
\]

where \( \nu \) is chosen such that

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

**Remark 3.** The right invertibility of \( F^\eta \) ensures that the sampling method is non-degenerate, e.g. the modulation sequence cannot be zero.

The optimal \( \nu \) corresponds to a water-filling power allocation strategy based on the singular values of the equivalent channel matrix \( (F^\eta F^\eta^*) - \frac{1}{2} F^\eta F^h F^h^* F^\eta F^\eta^* \), where \( (F^\eta F^\eta^*)^{-\frac{1}{2}} \) is due to noise prewhitening and \( F^\eta F^h \) is the equivalent channel matrix after modulation and filtering. This result can again be interpreted by viewing (39) as the MIMO Gaussian channel capacity of the equivalent channel matrix. We note that a closed-form capacity expression may be hard to obtain for general modulated sequences \( q_i(t) \). That 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. However, when \( q_i(t) \) is periodic, it can be mapped to a spike train in the frequency domain, which still exhibits block Toeplitz properties, as we describe in more detail in Appendix C.

**B. Approximate Analysis**

The Fourier transform of the signal prior to modulation in the \( i \)th branch at a given frequency \( f \) can be expressed as \( P_i(f) (H(f)X(f) + N(f)) \). Multiplication of this prefiltered signal with the modulation sequence \( q_i(t) \) corresponds to convolution in the frequency domain. The modulation sequence \( q_i(t) \) is assumed to be periodic with frequency response \( \sum_l c_l \delta(f - lf_q) \). Define

\[
R(f) = H(f)X(f) + N(f).
\]

The channel output is sampled at a rate \( \tilde{f}_s = f_s/M \) in the \( i \)th branch. We observe that \( T_q \) does not coincide with \( \tilde{T}_s := MT_s \), and that the sampling system is periodic with period \( bT_q = a\tilde{T}_s \). Specifically, if we denote by \( h(t, \tau) \) the output of the sampling system at time \( t \) due to an input at time \( \tau \), then \( h(t - bT_q, \tau - bT_q) = h(t, \tau) \). We therefore divide all samples in the \( i \)th branch into \( a \) groups, where the \( l \)th \((0 \leq l < a)\) group contains \( \{y_i[l + ka] \mid k \in \mathbb{Z} \} \), as illustrated in Fig. 12(a). Hence, each group is sampled uniformly at rate \( f_q/b \). 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. For the $i$th branch, denote by
\[ g_i^\eta(t, \tau) := \int s_i(t - \tau_1)q_i(\tau_1)p(\tau_1 - \tau)d\tau_1 \]
the output response of the preprocessing system at time $t$ due to an input impulse at time $\tau$. Then, the equivalent impulse response of the sampling system for the $l$th group is given by $g_i^l := g_i^\eta(l\tilde{T}_s, l\tilde{T}_s - t)$. Thus, the equivalent Fourier transform of the system output before ideal sampling in the $l$th group of the $i$th branch can be written as
\[ \tilde{Y}_l^i(f) = \frac{P_i(f)R(f)}{\sum_u c_i^u S_i(f - uf_q)} \sum_u c_i^u S_i(f - uf_q) \exp(j2\pi f l\tilde{T}_s) \],
which further leads to the sampled sequence in the $l$th group of the $i$th branch as
\[ Y_l^i(f) = \sum_v \tilde{Y}_l^i \left( f - \frac{vf_q}{b} \right) \]
\[ = \sum_v P_i \left( f - \frac{vf_q}{b} \right) R \left( f - \frac{vf_q}{b} \right) \sum_u c_i^u S_i \left( f - uf_q - \frac{vf_q}{b} \right) \exp(j2\pi l\tilde{T}_s \left( f - uf_q - \frac{vf_q}{b} \right)) \]
\[ = \sum_v A_{l,v} P_i \left( f - \frac{vf_q}{b} \right) H \left( f - \frac{vf_q}{b} \right) X \left( f - \frac{vf_q}{b} \right) + A_{l,v} P_i \left( f - \frac{vf_q}{b} \right) N \left( f - \frac{vf_q}{b} \right), \]
where
\[ A_{l,v} := \sum_u c_i^u S_i \left( f - uf_q - \frac{vf_q}{b} \right) \exp(j2\pi l\tilde{T}_s \left( f - uf_q - \frac{vf_q}{b} \right)). \]

Fig. 12 illustrates this representation for sampling with a single branch of modulation and filtering when $f_s = 2f_q$. All the information of the entire sampled data is contained in \( \{Y_l^i(f) \mid 0 \leq l < a, 1 \leq i \leq M\} \), and hence the sampling system can be equivalently represented as a MIMO channel with an infinite number of input branches and $aM$ output branch.

Due to the convolution operation in the spectral domain, the frequency response of the sampled output at frequency $f$ becomes a linear combination of frequency components \( \{X(f)\} \) and \( \{N(f)\} \) from several different aliased frequency sets. Here, we introduce the definition of a modulated aliased frequency set as a generalization of the aliased set. Specifically, for each $f$, its modulated aliased set is the set

\[ \{ \frac{vf_q}{b} \mid 0 \leq v < q, 1 \leq i \leq M\} \]
\( \left\{ f - l f_q - k \tilde{f}_s \mid l, k \in \mathbb{Z} \right\} \). 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
\[
\left\{ f_0 - l f_q - k \tilde{f}_s \mid l, k \in \mathbb{Z} \right\} = \left\{ f_0 - l \frac{f_q}{b} \mid l \in \mathbb{Z} \right\} = \left\{ f_0 - l \frac{\tilde{f}_s a}{b} \mid l \in \mathbb{Z} \right\}.
\]
(41)

In other words, for a given \( f_0 \in \left[ -\frac{f_q}{2b}, \frac{f_q}{2b} \right] \), 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 \( \left\{ f_0 - l \frac{\tilde{f}_s a}{b} \mid l \in \mathbb{Z} \right\} \) 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 \( \left\{ f_0 - l \frac{\tilde{f}_s a}{b} \mid l \in \mathbb{Z} \right\} \) and \( aM \) groups of output branches, each associated with one group of sample sequences in one branch. As an example, we illustrate in Fig. 12 the equivalent MIMO Gaussian channel under sampling following a single branch of modulation and filtering, when \( S(f) = 0 \) for all \( f \notin \left[ -\frac{f_q}{2b}, \frac{f_q}{2b} \right] \).

The effective frequencies of this frequency-selective MIMO Gaussian channel range from \( -\frac{f_q}{2b} \) to \( \frac{f_q}{2b} \), 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.

The rigorous analysis of Theorem 4 based on Toeplitz properties is deferred to Appendix C.

C. An Upper Bound on Sampled Capacity

Following the same analysis of optimal filter-bank sampling as in Section IV-C, we can derive an upper bound on the sampled channel capacity, as characterized in Corollary 3.

**Corollary 3.** Consider the system shown in Fig. 4. Suppose that for each aliased set \( \left\{ f - i f_q/b \mid i \in \mathbb{Z} \right\} \) and each \( k \) \((1 \leq k \leq aM)\), there exists an integer \( l \) such that \( |H (f - i f_q/b)|^2 \) is equal to the \( k \)th largest element in \( \left\{ |H (f - i f_q/b)|^2 \mid i \in \mathbb{Z} \right\} \). The capacity (39) under sampling following modulation and filter banks can be upper bounded by
\[
C(f_s) = \frac{1}{2} \int_{-f_s/2b}^{f_s/2b} \sum_{k=1}^{aM} \left( \log \left( \nu \cdot \lambda_k (F_h F_h^*) \right) \right)_+ \ df,
\]
(42)

where \( \nu \) is chosen such that
\[
\int_{-f_s/2b}^{f_s/2b} \sum_{k=1}^{aM} \left[ \nu - \lambda_k (F_h F_h^*) \right]_+ \ df = P.
\]
(43)

**Proof:** By observing that \( (F^0 F^{\eta^*})^{-\frac{1}{2}} F^0 \) has orthonormal rows, we can immediately derive the result using Proposition 5.
Figure 12. (a) Grouping of the sampling set when $f_s = 3f_q$. The sampling grid is divided into 3 groups, where each group forms a uniform set with rate $f_s/3$; (b) Equivalent MIMO Gaussian channel for a given $f \in [0, \frac{f_s}{2}]$ under sampling following a single branch of modulation and filtering, where $f_q = \frac{1}{2} f_s$; (c) An equivalent set of parallel MIMO channels representing all $f \in [-f_q/2b, f_q/2b]$, where the MIMO channel at a given frequency is equivalent to the MIMO channel of Fig. 12(a).
It can be observed that this upper bound 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 do 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 the filter-bank sampling.

We caution that this upper bound may not be tight for many scenarios, since we restrict our analysis framework to periodic modulation sequences. A general optimal modulation has not been identified in this work, which is left for future investigation.

D. Single-branch Sampling with Modulation and Filtering v.s. 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 \( T_q \). We consider here two special cases of single-branch modulation sampling, and investigate whether any hardware benefit can be harvested.

1) \( \frac{1}{M} f_s = \frac{1}{a} f_q \) for some integer \( a \): In this case, the modulated aliased set is 
\[
\left\{ f - k \frac{f_s}{M} - l f_q \mid k, l \in \mathbb{Z} \right\} = \left\{ f - k \frac{f_s}{M} \mid k \in \mathbb{Z} \right\},
\]
which is equivalent to the original aliased frequency set. That said, the sampled output \( Y(f) \) is still a linear combination of 
\[
\left\{ H \left( f - k \frac{f_s}{M} \right) X \left( f - k \frac{f_s}{M} \right) + N \left( f - k \frac{f_s}{M} \right) \mid k \in \mathbb{Z} \right\}.
\]
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 more design degrees of freedom. Therefore, the maximum sampled channel capacity achievable by adding an additional modulation bank is no larger than the one achievable without the modulation sequences.

2) \( \frac{1}{M} f_s = b f_q \) for some integer \( b \): In this case, the modulated aliased set is enlarged to 
\[
\left\{ f - k \frac{f_s}{M} - l f_q \mid k, l \in \mathbb{Z} \right\} = \left\{ f - l f_q \mid k \in \mathbb{Z} \right\},
\]
which may potentially provide implementation gain compared with filter-bank sampling with the same number of branches. We consider the following example. Suppose that the channel contains 3 flat subbands with channel gains as plotted in Fig. 13 and that the noise is of unit spectral density within these 3 subbands and zero otherwise. Here, single-branch filtering followed by sampling is employed, where the sampling rate is \( f_s = 2 \) and the period of the modulation sequence \( T_q = 2 T_s \). Due to aliasing, Subband 1 and Subband 3 (as illustrated in Fig. 13) are mixed together. According to Section III-D, the optimal prefilter without modulation would be a band-pass filter with passband \(-1.5 \leq f \leq 0.5\), resulting in a channel consisting of two subbands with respective channel gains 2 and 1.
SNR

modulated sampling following optimal filtering. As illustrated in Fig. 13, equivalent parallel channels experiencing respective channel gains for all $a$ two-subband channel with respective channel gains both arbitrarily close to 2. That said, sampling $f$ filter is exploited. Specifically, suppose that the modulation sequence has a period of $2T_s$ and that the noise is of unit spectral density within these 3 subbands and zero otherwise. Here, single-

components of the sampled response under modulation followed by sampling.

Figure 13. The left plot illustrates the channel gain, where the sampling rate is $f_s = 2$. The right plot illustrates the signal components of the sampled response under modulation followed by sampling.

Now if we employ modulation sampling with a lowpass filter, the channel structure can be better exploited. Specifically, suppose that the modulation sequence has a period of $2T_s$ and obeys $c^0 = 1$, $c^1 = 100$, $c^2 = 1$, $c^{-2} = 1000$ and $c^i = 0$ for all other $i$'s, and that the cutoff frequency of the low-pass filter is $f_{\text{cutoff}} = 1$. By simple manipulation,

$$
\begin{bmatrix}
\exp\left(j\pi T_q \left(f - \frac{f_s}{2}\right)\right) Y\left(f - \frac{f_s}{2}\right) \\
\exp\left(j\pi T_q f\right) Y\left(f\right)
\end{bmatrix}
= \begin{bmatrix}
2 & 100 & 1001 \\
100 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
2X\left(f - \frac{f_s}{2}\right) + N\left(f - \frac{f_s}{2}\right) \\
X\left(f\right) + N\left(f\right) \\
2X\left(f + \frac{f_s}{2}\right) + N\left(f + \frac{f_s}{2}\right)
\end{bmatrix}
$$

for all $f \in \left[0, \frac{f_s}{2}\right]$. Through noise whitening and eigenvalue decomposition, we can derive a pair of equivalent parallel channels experiencing respective channel gains 2 and 1.99, which outperforms non-modulated sampling following optimal filtering. As illustrated in Fig. 13, $Y\left(f - \frac{f_s}{2}\right)$ primarily depends on the frequency component at $f + \frac{f_s}{2}$, while $Y\left(f\right)$ primarily depends on the frequency component at $f - \frac{f_s}{2}$: both frequencies have SNR 4. In fact, by increasing $c^{-2}$ and $c^1$ correspondingly, we can obtain a two-subband channel with respective channel gains both arbitrarily close to 2. That said, sampling following a single branch of modulation can achieve the same capacity as applying the optimal filter bank.

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 $f_q = \frac{1}{K} f_s$ for some integers $K$ and $L$. The SNR $\frac{\left|H(f)\right|^2}{S_n(f)}$ 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 [48], the channel gain (and hence the SNR) is roughly
equal across the subband. Take any \( f \in \left[-\frac{f_q}{2}, \frac{f_q}{2}\right] \), and run the following simple algorithm to determine the modulation sequence.

**Algorithm 1**

1. **Initialize.** Find the \( K \) largest elements in \( \left\{ \frac{\left| H(f-lf_q) \right|^2}{S_n(f-lf_q)} : l \in \mathbb{Z}, -L \leq l \leq L-1 \right\} \). Denote by \( \{l_i : 1 \leq i \leq K\} \) the index set of these \( K \) elements such that \( l_1 > l_2 > \cdots > l_K \). Set \( i = 1 \), \( I_{\text{max}} = -L \), \( J = \emptyset \), and \( c^i = 0 \) for all \( i \in \mathbb{Z} \). Let \( A \) be a large given number.

2. For \( i = 1 : K \)

   For \( m = I_{\text{max}} : I_{\text{max}} + K - 1 \)

   if \( (m \mod K) \notin J \), do

   \[ J = J \cup \{m \mod K\}, I_k = m, I_{\text{max}} = m + L - 1 - l_i, c^{m-l_i} = A^{K+1-i} \]

   and break;

3. For \( i = -L : L - 1 \)

   if \( i \in \{I_1, \cdots, I_K\} \), then \( S(f + if_q) = 1 \);

   else \( S(f + if_q) = 0 \).

The goal of this algorithm is to generate \( K \) subbands with high SNR. Due to convolution, the signal in each subband is a linear combination of the frequency components in all frequencies in the modulated aliased set. Adjusting the values of \( \{c^i : i \in \mathbb{Z}\} \) results in different weights for each component. Here, the signal in each subband being selected through Step 2 will contain one primary component accounting for most of the power of the entire signal. Filtering is further used in Step 3 in order to suppress aliasing. The performance of this algorithm is characterized in the following proposition.

**Proposition 2.** Consider the piecewise flat channel with \( 2L \) subbands as described above. For a given \( f_q \), the modulation sequence found by Algorithm 1 maximizes capacity when \( A \rightarrow \infty \).

In fact, the performance of this algorithm is asymptotically 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 multi-branch filter-bank sampling. This is in general the preferred approach since building multiple analog filters is more expensive (in terms of power consumption, size, or cost) than a single modulator. We note, however, that for a given overall sampling rate, modulation-bank sampling does not outperform filter-bank sampling in terms of sampled capacity, which is formally stated as follows.
Proposition 3. Consider the setup in Theorem 4. 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 Section III-D and Section 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 channel input and the signal reconstructed from sampling the channel output, as detailed below.

The sampling problem we consider can be formally stated as follows. For a given analog channel, suppose that the channel input \( x(t) \) is any zero-mean WSS stochastic signal whose power spectral density (PSD) \( S_X(f) \) satisfies a power constraint \( \int_{-\infty}^{\infty} S_X(f)df = P \). This input is passed through the channel where it is filtered by the channel frequency response and contaminated by stationary Gaussian noise. We sample the channel output using a filter bank at a fixed rate \( f_s/M \) in each branch, and recover an 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 \( t \in \mathbb{R} \). Since the samples \( \{y[n]\} \) are Gaussian random variables, the MMSE estimate \( \hat{x}(t) \) for a given input process \( x(t) \) is linear in \( \{y[n]\} \). 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 \( t \in \mathbb{R} \).

In this joint optimization, 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 4. Suppose the channel input \( x(t) \) is any WSS signal. For a given sampling system, let \( \hat{x}(t) \) denote the corresponding optimal linear estimate of \( x(t) \) from the digital sequence \( \{y[n]\} \). Then the optimal filter bank given in (36) and its corresponding optimal input \( x(t) \) minimizes the MSE reconstruction error \( \mathbb{E} (|x(t) - \hat{x}(t)|^2) \) over all possible LTI filter banks.

\(^6\)Instead of giving an analysis that accommodates a general class of signals, we restrict our attention to wide-sense stationary (WSS) Gaussian input signals. This restriction, while falling short of generality, allows us to derive informative sampling results in a simple way.
Proof: See Appendix F.

Proposition 4 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. Thus, under sampling with filter-banks, information theory reconciles with sampling theory through the SNR metric when determining optimal systems. Intuitively, high SNR typically leads to large capacity and low MSE.

Proposition 4 includes the optimal prefilter under single-prefilter sampling as a special case. That said, the optimal prefilter puts all its mass in the frequency with the highest SNR in each aliased frequency set, and thus suppresses aliasing, which coincides with the capacity-optimizing filter derived in Section III-D. We note that a similar MSE minimization problem was investigated decades ago with applications in pulse amplitude modulation (PAM) [40], [41]: 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, as we do in Proposition 4.

VII. Conclusions and Future Work

We characterize sampled channel capacity as a function of sampling rate for different sampling mechanisms, thereby forging a new connection between sampling theory and information theory. We show that the capacity of a sampled analog channel degrades with reduced sampling rate and identify optimal sampling structures for several classes of sampling methods, which exploits structure in the sampling design. These results also indicate that capacity was not always monotonic in sampling rate, and illuminate an intriguing connection between MIMO channel capacity and capacity of undersampled analog channels. Our work establishes a framework for using the information-theoretic metric of capacity to optimize sampling structure, offering a different angle from traditional design of sampling methods based on other statistical measures such as MSE.

Our work also opens more 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 achieve or at least close the gap to capacity. Investigation of capacity under more general nonuniform sampling techniques is a topic of future work. In addition, the optimal sampling structure for time-varying channels will require different analysis than used in the time-invariant case, and it remains to be seen what sampling mechanisms are optimal for channels when the channel state is partially or fully unknown. A deeper understanding of how to exploit the 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 provide first 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 a couple of auxiliary lemmas are deferred to Appendix H.

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

\[
C(f_s) = \frac{1}{2} \int_{-f_s/2}^{f_s/2} \left[ \log \left( \nu - \frac{\sum_{l=-\infty}^{\infty} |H(f-lf_s)S(f-lf_s)|^2}{\sum_{l=-\infty}^{\infty} |S(f-lf_s)|^2} \right) \right]^{+} \, df
\]

where the water level \( \nu \) can be calculated through the following equation:

\[
\int_{-f_s/2}^{f_s/2} \left( \nu - \frac{\|V_{HS}(f,f_s)\|_2^2}{\|V_S(f,f_s)\|_2^2} \right)^{+} \, df = P.
\]

This capacity result under white noise can then be immediately extended to accommodate for colored noise. Suppose the additive noise is of power spectral density \( S_\eta(f) \). We can then split the channel filter \( H(f) \) into two parts with respective frequency response \( H(f) / \sqrt{S_\eta(f)} \) and \( \sqrt{S_\eta(f)} \). Since the colored
noise is equivalent to a white Gaussian noise passed through a filter with transfer function $\sqrt{S_\eta(f)}$, the original system can be redrawn as in Fig. 14. This equivalent representation immediately leads to the capacity in the presence of colored noise by plugging in corresponding terms in (44).

\[ \eta(t) \]

\[ \begin{array}{c}
x(t) \xrightarrow{H(f)} \frac{\eta(t)}{\sqrt{S_\eta(f)}} \xrightarrow{\sqrt{S_\eta(f)}S(f)} y[n]
\end{array} \]

Figure 14. Equivalent representation of filtering followed by sampling in the presence of colored noise.

A. Channel Discretization and Diagonalization

Similar to the analysis for non-filtered ideal sampling, we set $T = nT_s$ and $T_s = k\Delta$. For notational simplicity, we define

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

for any function $g(t)$. If $g(t)$ is a continuous function, then we have $\lim_{\Delta \to 0} g_{u,v} = g(uT_s - v\Delta)$.

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 let $\tilde{h}_i = \Delta \cdot [\tilde{h}_{i,0}, \tilde{h}_{i,1}, \cdots, \tilde{h}_{i,k-1}]$ and $s_i = \Delta \cdot [s_{i,0}, s_{i,1}, \cdots, s_{i,k-1}]$ be $1 \times k$ vectors. We define the transmit signal vector $x^n$ and noise vector $\eta$ as

\[ (x^n)_i = \frac{1}{\Delta} \int_0^\Delta x(i\Delta + \tau) \, d\tau \quad (0 \leq i < nk) \]

and

\[ (\eta)_i = \frac{1}{\Delta} \int_0^\Delta \eta(i\Delta + \tau) \, d\tau \quad (i \in \mathbb{Z}). \]

We also introduce

\[ \tilde{H}^n := \begin{bmatrix}
\tilde{h}_0 & \tilde{h}_{-1} & \cdots & \tilde{h}_{-n+1} \\
\tilde{h}_1 & \tilde{h}_0 & \cdots & \tilde{h}_{-n+2} \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{h}_{n-1} & \tilde{h}_{n-2} & \cdots & \tilde{h}_0
\end{bmatrix}, \quad \text{and} \quad S^n := \begin{bmatrix}
\cdots & s_0 & s_{-1} & \cdots \\
\cdots & s_1 & s_0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\cdots & s_{n-1} & s_{n-2} & \cdots
\end{bmatrix}. \]

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

\[ y^n = \tilde{H}^n x^n + S^n \eta. \]

As can be seen, $\tilde{H}^n$ is a fat block Toeplitz matrix. Moreover, $S^n S^{n*}$ is asymptotically equivalent to a Toeplitz matrix, as will be shown later. We note that each element $\eta_i$ is a zero-mean Gaussian variable.
with variance
\[
\mathbb{E} \left( |\eta|^2 \right) = \frac{1}{\Delta^2} \int_0^\Delta \int_0^\Delta \mathbb{E} (\eta(i\Delta + \tau)\eta^*(i + t)) \, d\tau \, dt
\]
\[
= \frac{1}{\Delta^2} \int_0^\Delta \int_0^\Delta \delta(\tau - t) \, d\tau \, dt
\]
\[
= \frac{1}{\Delta},
\]
and that \( \mathbb{E} (\eta_i \eta_l^*) = 0 \) for any \( i \neq l \), thus implying that \( \eta \) is an i.i.d. Gaussian vector with each entry having variance \( 1/\Delta \). The filtered noise \( S^\eta \) is no longer i.i.d. Gaussian, and hence we first attempt to whiten the noise.

Setting \( \tilde{S}^n = (S^nS^{*n})^{-\frac{1}{2}} S^n \), we see that
\[
\mathbb{E} \tilde{S}^n \eta \left( \tilde{S}^n \eta \right)^* = \tilde{S}^n \mathbb{E} (\eta \eta^*) \tilde{S}^{*n} = \tilde{S}^n \tilde{S}^{*n} = (S^nS^{*n})^{-\frac{1}{2}} S^n S^{*n} (S^nS^{*n})^{-\frac{1}{2}} = 1/\Delta I^n.
\]

This basically implies that \( \tilde{S}^n \) projects the i.i.d. Gaussian noise \( \eta \) onto an orthogonal \( n \) dimensional subspace, i.e. \( (S^nS^{*n})^{-\frac{1}{2}} S^n \eta^n \) is still i.i.d. Gaussian noise. Applying this whitening operation, we get
\[
\tilde{y}^n := (S^nS^{*n})^{-\frac{1}{2}} y^n
\]
\[
= (S^nS^{*n})^{-\frac{1}{2}} \tilde{H}^n x^n + (S^nS^{*n})^{-\frac{1}{2}} S^n \eta^n
\]
\[
= (S^nS^{*n})^{-\frac{1}{2}} \tilde{H}^n x^n + \tilde{\eta}^n.
\]
Here, \( \tilde{\eta}^n \) consists of independent zero-mean Gaussian elements with variance \( 1/\Delta \). Since \( S^n \) is of full row rank, the data processing inequality immediately yields that
\[
I (x^n, \tilde{y}^n) = I (x^n, y^n).
\]

Consequently, we can express the capacity of the sampled analog channel as the following limit:
\[
C(f_s) = \lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{nT_s} \sup_{p(x): \frac{1}{n} \mathbb{E}(\|x^n\|^2) \leq P} I (x^n, y^n)
\]
\[
= \lim_{k \to \infty} \lim_{n \to \infty} \frac{f_s}{n} \sup_{p(x): \frac{1}{n} \mathbb{E}(\|x^n\|^2) \leq P} (x^n, \tilde{y}^n).
\]

We note that even when there exists no integer \( n \) such that \( T = nT_s \), the capacity can be bounded through the following fact. Since the proof of this lemma is straightforward, we defer it to Appendix H-A.

**Lemma 1.** Suppose the following limit
\[
\lim_{n \to \infty} \frac{1}{nT_s} \sup I (x(0, nT_s]; \{y[n]\})
\]
exists. Then we have

\[
\lim_{T \to \infty} \frac{1}{T} \sup_{I} \left( x(0, T); \{ y[n] \} \right) = \lim_{n \to \infty} \frac{1}{nT_s} \sup_{I} \left( x(0, nT_s); \{ y[n] \} \right).
\]

(52)

Hence, it suffices to investigate the case when \( T \) is integer multiples of \( T_s \).

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 [4], [51] for more details.

A Toeplitz matrix is an \( n \times n \) matrix \( T_n \) where \( (T_n)_{k,l} = t_{k-l} \), which implies that a Toeplitz matrix \( T_n \) is uniquely defined by the sequence \( \{ t_k \} \). A special case of Toeplitz matrices is circulant matrices where every row of the matrix \( C_n \) is a right cyclic shift of the row above it. The Fourier series (or symbol) with respect to the sequence of Toeplitz matrices \( \{ T_n := [t_{k-l}; k, l = 0, 1, \ldots, n-1] : n \in \mathbb{Z} \} \) is given by

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

(53)

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

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

(54)

One key concept in our proof is asymptotic equivalence, which is formally defined as follows [51].

Definition 1 (Asymptotic Equivalence). Two sequences of \( n \times n \) matrices \( \{ A^n \} \) and \( \{ B^n \} \) are said to be asymptotically equivalent if

1. \( A^n \) and \( B^n \) are uniformly bounded, i.e. there exists a constant \( c \) independent of \( n \) such that

\[
\|A^n\|_2, \|B^n\|_2 \leq c < \infty, \quad n = 1, 2, \ldots
\]

(55)

2. \( \lim_{n \to \infty} \frac{1}{\sqrt{n}} \|A^n - B^n\|_F = 0. \)

We will abbreviate asymptotic equivalence of \( \{ A^n \} \) and \( \{ B^n \} \) by \( A^n \sim B^n \). Two important results regarding asymptotic equivalence are given in the following lemmas [51].
Lemma 2. Suppose $A^n \sim B^n$ 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
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\alpha_{n,k}) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(\beta_{n,k}). \tag{56}
\]

Lemma 3. (a) Suppose a sequence of Toeplitz matrices $T^n$ where $(T^n)_{ij} = t_{i-j}$ satisfies that $\{t_i\}$ is absolutely summable. Suppose the Fourier series $F(\omega)$ related to $T^n$ is positive and $T^n$ is Hermitian. Then we have $T^n(F) \sim C^n(F)$. \tag{57}

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 $T^n(F)^{-1} \sim C^n(F)^{-1} = C^n(1/F) \sim T^n(1/F). \tag{58}$

(b) Suppose $A^n \sim B^n$ and $C^n \sim D^n$, then $A^nC^n \sim B^nD^n$.

Toeplitz or block Toeplitz matrices have well-known asymptotic spectral properties [4], [52]. The notion of asymptotic equivalence allows us to approximate non-Toeplitz matrices by Toeplitz matrices, which we will use in the next subsection to analyze the spectral properties of the channel matrix.

C. Capacity via Convergence of the Discrete Model

To prove the capacity theorem, we first construct an asymptotically equivalent channel matrix and obtain its capacity. This requires us to first exploit the asymptotic spectral properties of the Hermitian matrices $S^nS^{n*}$ and $\tilde{H}^n\tilde{H}^{n*}$. In particular, for any $1 \leq i \leq j \leq n$, we have
\[
(S^nS^{n*})_{ij} = (S^nS^{n*})_{ji}^* = \sum_{t=-\infty}^{\infty} s_{j-i+t}s_t^*. \tag{59}
\]
Hence, the Hermitian matrix $\tilde{S}^n := S^nS^{n*}$ is still Toeplitz. On the other hand,
\[
(\tilde{H}^n\tilde{H}^{n*})_{ij} = (\tilde{H}^n\tilde{H}^{n*})_{ji}^* = \sum_{t=-j+1}^{n-j} \tilde{h}_{j-i+t}\tilde{h}_t^*. \tag{60}
\]
Obviously, $\tilde{H}^n\tilde{H}^{n*}$ is not a Toeplitz matrix. Instead of investigating the eigenvalue distribution of $\tilde{H}^n\tilde{H}^{n*}$ directly, we look at a new Hermitian Toeplitz matrix $\hat{H}^n$ associated with $\tilde{H}^n\tilde{H}^{n*}$ such that for any $i \leq j$:
\[
(\hat{H}^n)_{ij} = (\hat{H}^n)_{ji}^* = \sum_{t=-\infty}^{\infty} \hat{h}_{j-i+t}\hat{h}_t^*. \tag{61}
\]
Lemma 4. The above definition of $\hat{H}_n$ implies that

$$\hat{H}_n \sim \hat{H}^n \hat{H}^{n*}. \quad (62)$$

Proof: See Appendix [H-B] for the detailed derivation.

Next, we construct the circulant matrix $C_n$ as defined in (54). The following lemma relates $(C_n)^{-1}$ with $(S^n S^{n*})^{-1}$.

Lemma 5. If there exists some constant $\epsilon_s > 0$ such that for all $f \in \left[ -\frac{f_s}{2}, \frac{f_s}{2} \right]$,

$$\sum_{l=-\infty}^{\infty} |S(f - lf_s)|^2 \geq \epsilon_s > 0$$

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

Proof: See Appendix [H-C].

One of the most useful properties of a circulant matrix $C_n$ is that its eigenvectors $\{u_c^{(m)}\}$ are

$$u_c^{(m)} = \frac{1}{\sqrt{n}} \left(1, e^{-2\pi jm/n}, \ldots, e^{-2\pi j(n-1)m/n}\right). \quad (64)$$

Suppose the eigenvalue decomposition of $C_n$ is given as

$$C_n = U_c \Lambda_c U_c^*, \quad (65)$$

where $U_c$ is a Fourier coefficient matrix, and $\Lambda_c$ 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 6. For any continuous function $g(x)$, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g(\lambda_i) = T_s \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} g \left( \frac{\|V_{HS}(f, f_s)\|_2^2}{\|V_S(f, f_s)\|_2^2} \right) df,$$

where $\lambda_i$ denotes the $i$th eigenvalue of $(S^n S^{n*})^{-\frac{1}{2}} \hat{H}^n \hat{H}^{n*} (S^n S^{n*})^{-\frac{1}{2}}$.

Proof: See Appendix [H-D].

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

\[ C_T = \frac{1}{2T} \sum_{i : \lambda_i \geq \nu^{-1}} \log (\nu \lambda_i), \]  \hspace{1cm} (66)

\[ Pnk_k/\Delta = \sum_{i : \lambda_i \geq \nu^{-1}} \left[ \nu - \frac{1}{\lambda_i} \right], \]  \hspace{1cm} (67)

where \(\{\lambda_i\}\) denote the eigenvalues of \((S^\dagger S^{\ast})^{-\frac{1}{2}} \tilde{H}^n \tilde{H}^{n\ast} (S^\dagger S^{\ast})^{-\frac{1}{2}}\), and \(\nu\) is the water level of the optimal power allocation over this discrete model, as can be calculated through (67).

By passing to the limit \(T \to \infty\), we can exploit the asymptotic spectral properties of Toeplitz matrices as

\[
\lim_{T \to \infty} C_T(\nu) = \lim_{T \to \infty} \frac{1}{T} \sum_{i : \lambda_i \geq 1/\nu} \frac{1}{2} \log [\nu \lambda_i]
\]

\[
= \frac{1}{2} \int_{f \in \mathcal{F}(\nu)} \log \nu \left( \frac{\|V_{HS}(f,f_s)\|_2^2}{\|V_s(f,f_s)\|_2^2} \right) df,
\]

where \(\mathcal{F}(\nu) = \{ f : \frac{\|V_{HS}(f,f_s)\|_2^2}{\|V_s(f,f_s)\|_2^2} \geq \frac{1}{\nu} \}\). Similarly, (67) can be transformed into

\[
PT_s = \frac{Pk}{1/\Delta} = \frac{1}{n} \sum_{i : \lambda_i \geq \nu^{-1}} \left[ \nu - \frac{1}{\lambda_i} \right], \]  \hspace{1cm} (68)

\[
= T_s \int_{f \in \mathcal{F}(\nu)} \left[ \nu - \frac{\|V_s(f,f_s)\|_2^2}{\|V_{HS}(f,f_s)\|_2^2} \right] df, \]  \hspace{1cm} (69)

which completes the proof.

**APPENDIX B**

**PROOF OF THEOREM 3**

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

\[ \tilde{h}_i(t) := h(t) * s_i(t), \quad \text{and} \quad \tilde{h}_i^l = \left[ \tilde{h}_i \left( l\tilde{T}_s \right), \tilde{h}_i \left( l\tilde{T}_s - \Delta \right), \cdots, \tilde{h}_i \left( l\tilde{T}_s - (k-1)\Delta \right) \right]. \]
We introduce the following two matrices as
\[
\tilde{H}_n = \begin{bmatrix}
\tilde{h}_0 & \tilde{h}_1^{-1} & \cdots & \tilde{h}_{-n+1}
\tilde{h}_1 & \tilde{h}_0 & \cdots & \tilde{h}_{-n+2}
\vdots & \vdots & \ddots & \vdots
\tilde{h}_{-n+1} & \tilde{h}_{-n+2} & \cdots & \tilde{h}_0
\end{bmatrix}
\] and
\[
S_n = \begin{bmatrix}
s_0 & s_1^{-1} & \cdots & s_{-n+1}
s_1 & s_0 & \cdots & s_{-n+2}
\vdots & \vdots & \ddots & \vdots
s_{-n+1} & s_{-n+2} & \cdots & s_0
\end{bmatrix}.
\]

We also set \((x_n)_i = \frac{1}{\Delta} \int_0^{\Delta} x(i\Delta + \tau) \, d\tau \text{ (0 \leq i < nk)}, \) and \((\eta)_i = \frac{1}{\Delta} \int_0^{\Delta} \eta(i\Delta + \tau) \, d\tau \text{ (i \in \mathbb{Z})} \). Defining \(y^n = [y_1[0], \ldots, y_1[n-1], y_2[0], \ldots, y_2[n-1], \ldots, y_M[n-1]]^T \) leads to the discretized channel model
\[
y^n = \begin{bmatrix}
\tilde{H}_1^n & \tilde{H}_2^n & \cdots & \tilde{H}_M^n
\end{bmatrix} x^n + \begin{bmatrix}
S_1^n & S_2^n & \cdots & S_M^n
\end{bmatrix} \eta.
\] (70)

Whitening the noise gives us
\[
\tilde{y}^n = \left( \begin{bmatrix}
S_1^n \\
S_2^n \\
\vdots \\
S_M^n
\end{bmatrix}
\begin{bmatrix}
S_1^n & \cdots & S_M^n
\end{bmatrix} \right)^{-\frac{1}{2}} \begin{bmatrix}
\tilde{H}_1^n \\
\tilde{H}_2^n \\
\vdots \\
\tilde{H}_M^n
\end{bmatrix} x^n + \tilde{\eta},
\] (71)
where \(\tilde{\eta}\) 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
\[
C(f_s) = \lim_{k \to \infty} \lim_{n \to \infty} \frac{f_s}{Mn} \sup_{x^n} I(x^n; \tilde{y}^n),
\]

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

B. Capacity via Convergence of the Discrete Model

We can see that for any \(1 \leq u, v \leq m\),
\[
S_u^n S_v^{n*} = \tilde{S}_{u,v}^n,
\] (72)
where the Toeplitz matrix \(\tilde{S}_{u,v}^n\) is defined such that for any \(1 \leq i \leq j \leq n\)
\[
(\tilde{S}_{u,v}^n)_{i,j} = \sum_{t=-\infty}^{\infty} s_u^{i-t} (s_v^t)^*.
\] (73)
Let $S^n = [S^n_1, S^n_2, \ldots, S^n_M]^*$. Then the Hermitian block Toeplitz matrix

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

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

$$
(H^n_{u,v})_{i,j} = \sum_{t=-\infty}^{\infty} \tilde{h}^{j-i+t} \left( \tilde{h}^t_v \right)^*, \quad (74)
$$

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

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

satisfies

$$
\lim_{n \to \infty} \frac{1}{\sqrt{nM}} \left\| \hat{H}^n - \tilde{H}^n \tilde{H}^{n*} \right\|_F \leq \lim_{n \to \infty} \frac{1}{\sqrt{nM}} \sum_{1 \leq u,v \leq M} \frac{1}{\sqrt{n}} \left\| H^n_{u,v} - \tilde{H}^{n}_u \tilde{H}^{n*}_v \right\| = 0,
$$

$$
\Rightarrow \hat{H}^n \sim \tilde{H}^n \tilde{H}^{n*}. \quad (75)
$$

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

$$
[F_\delta(f)]_{u,v} = \frac{\Delta^2}{T^2} \sum_{k=0}^{k-1} \left( \sum_{l_1} S_u \left( -f + l_1 \tilde{f}_s \right) \exp \left( -j2\pi \left( f - l_1 \tilde{f}_s \right) i\Delta \right) \right) \left( \sum_{l_2} S_v \left( -f + l_2 \tilde{f}_s \right) \exp \left( -j2\pi \left( f - l_2 \tilde{f}_s \right) i\Delta \right) \right)^* \quad (76)
$$

$$
= \frac{\Delta^2}{T^2} \sum_{k=0}^{k-1} \left( \sum_{l_1, l_2} S_u \left( -f + l_1 \tilde{f}_s \right) S^*_v \left( -f + l_2 \tilde{f}_s \right) \exp \left( -j2\pi \left( l_2 - l_1 \right) \tilde{f}_s i\Delta \right) \right) \quad (77)
$$

$$
= \frac{\Delta}{T} \sum_{l_1, l_2} \left( \sum_{l_1, l_2} S_u \left( -f + l_1 \tilde{f}_s \right) S^*_v \left( -f + l_2 \tilde{f}_s \right) \exp \left( -j2\pi \left( l_2 - l_1 \right) \tilde{f}_s i\Delta \right) \right) \quad (78)
$$

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

$$
\sum_{l_1=1}^{M} T^n_{l_1, l_2} \left( F^{-1}_\delta \right) \cdot \tilde{S}^n_{l_1, l_2} \sim T^n \left( \sum_{l_1=1}^{M} \left[ F^{-1}_\delta \right]_{l_1, l_2} \tilde{S}^n_{l_1, l_2} \right) = T^n (\delta \{ l_1 - l_1 \}), \quad (80)
$$
which immediately yields
\[ T^n \left( F^{-1}_s \right) \tilde{S}^n \sim I \implies T^n \left( F^{-1}_s \right) \sim \left( \tilde{S}^n \right)^{-1} \sim (S^n S^n*)^{-1}. \] (81)

Therefore, for any continuous function \( g(x) \), [52, Theorem 5.4] implies that
\[ \lim_{n \to \infty} \frac{1}{nM} \sum_{i=1}^{nM} g \left( \lambda_i \left( (S^n S^n*)^{-1/2} \tilde{H}^n \tilde{H}^n* (S^n S^n*)^{-1/2} \right) \right) = \int_{-\frac{T}{2M}}^{\frac{T}{2M}} \sum_{i=1}^{M} g \left( \lambda_i \left( F^{-1/2}_s F_h F^{-1/2}_s \right) \right) df. \]

Finally, the capacity of parallel channels [1], which is achieved via water filling power allocation, yields
\[ C(f_s) = \lim_{n \to \infty} \frac{1}{2nMT_s} \sum_{i=1}^{nM} \left[ \log \left( \nu \lambda_i \left( (S^n S^n*)^{-1/2} \tilde{H}^n \tilde{H}^n* (S^n S^n*)^{-1/2} \right) \right) \right] + \] (82)
\[ = \int_{-\frac{T}{2M}}^{\frac{T}{2M}} \frac{1}{2} \sum_{i=1}^{M} \left[ \log \left( \nu \lambda_i \left( F^{-1/2}_s F_h F^{-1/2}_s \right) \right) \right]^{+} df \] (83)
\[ = \int_{-\frac{T}{2M}}^{\frac{T}{2M}} \frac{1}{2} \sum_{i=1}^{M} \left[ \log \left( \nu \lambda_i \left( (F_s F_s^*)^{-1/2} F_s F_h F_s^* (F_s F_s^*)^{-1/2} \right) \right) \right]^{+} df, \] (84)

where
\[ P = \int_{-\frac{T}{2M}}^{\frac{T}{2M}} \sum_{i=1}^{M} \left[ \nu - \lambda_i \left( F^{-1/2}_s F_h F^{-1/2}_s \right) \right]^{+} df \] (85)
\[ = \int_{-\frac{T}{2M}}^{\frac{T}{2M}} \sum_{i=1}^{M} \left[ \nu - \lambda_i \left( (F_s F_s^*)^{-1/2} F_s F_h F_s^* (F_s F_s^*)^{-1/2} \right) \right]^{+} df. \] (86)

This completes the proof.

**APPENDIX C**

**PROOF OF THEOREM 4**

Following similar steps as in the proof of Theorem 3, 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 \( P_i(f) \sqrt{S_\eta(f)} \) and \( H(f) \) with \( H(f)/\sqrt{S_\eta(f)} \).
In the $i$th branch, the noise component at time $t$ is given by
\[
s_i(t) * (q_i(t) \cdot (p_i(t) * \eta(t))) = \int_{\tau_1} d\tau_1 s_i(t - \tau_1) \int_{\tau_2} q_i(\tau_1) p_i(\tau_1 - \tau_2) \eta(\tau_2) \, d\tau_2
\]
\[
= \int_{\tau_2} \left( \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau_2) \, d\tau_1 \right) \eta(\tau_2) \, d\tau_2
\]
\[
= \int_{\tau_2} g_i^n(t, \tau_2) \eta(\tau_2) \, d\tau_2,
\]
where
\[
g_i^n(t, \tau_2) \triangleq \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau_2) \, d\tau_1.
\]
Let $\tilde{T}_s = MT_s$. Our assumption $bT_q = a\tilde{T}_s$ immediately leads to
\[
g_i^n(t + a\tilde{T}_s, \tau + bT_q) = \int_{\tau_1} s_i(t + a\tilde{T}_s - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau - a\tilde{T}_s) \, d\tau_1
\]
\[
= \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1 + bT_q) p_i(\tau_1 - \tau) \, d\tau_1
\]
\[
= \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) p_i(\tau_1 - \tau) \, d\tau_1 = g_i^n(t, \tau),
\]
implying that $g_i^n(t, \tau)$ 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
\[
g_i^h(t, \tau_2) \triangleq \int_{\tau_1} s_i(t - \tau_1) q_i(\tau_1) \int_{\tau_3} p_i(\tau_1 - \tau_2 - \tau_3) h(\tau_3) \, d\tau_3 \, d\tau_1,
\]
which also satisfies the block-Toeplitz property $g_i^h(t + a\tilde{T}_s, \tau + bT_q) = g_i^h(t, \tau)$.

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 $G_i^n$ and $G_i^h$ such that $\forall m \in \mathbb{Z}, 0 \leq l < n$
\[
\begin{cases}
(G_i^n)_{l,m} = g_i^n(l\tilde{T}_s, m\Delta), \\
(G_i^h)_{l,m} = g_i^h(l\tilde{T}_s, m\Delta).
\end{cases}
\]
Setting $y_i^n = [y_i[0], y_i[1], \cdots, y_i[n - 1]]^T$ leads to similar discretized approximation as in the proof of Theorem 2 as follows
\[
y_i^n = G_i^h x^n + G_i^n \eta.
\]
Here, $\eta$ is a i.i.d. zero-mean Gaussian vector where each entry is of variance $1/\Delta$.

Hence, $G^h_i$ and $G^\eta_i$ are block Toeplitz matrices satisfying $(G^h_i)_{l+a,m+ak} = (G^h_i)_{l,m}$ and $(G^\eta_i)_{l+a,m+ak} = (G^\eta_i)_{l,m}$. Using the same definition of $x^n$ and $\eta$ as in Appendix B, we can express the system equation as

$$y^n = \begin{bmatrix} G^h_1 \\ G^h_2 \\ \vdots \\ G^h_M \end{bmatrix} x^n + \begin{bmatrix} G^\eta_1 \\ G^\eta_2 \\ \vdots \\ G^\eta_M \end{bmatrix} \eta.$$  \hspace{1cm} (88)

Whitening the noise component yields

$$\tilde{y}_n = \left( \begin{bmatrix} G^\eta_1 \\ G^\eta_2 \\ \vdots \\ G^\eta_M \end{bmatrix} \right)^{-\frac{1}{2}} \begin{bmatrix} G^h_1 \\ G^h_2 \\ \vdots \\ G^h_M \end{bmatrix} x_n + \tilde{\eta}, \hspace{1cm} (89)$$

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 7.** At a given frequency $f$, the Fourier symbol with respect to $G^\eta_\alpha G^\eta_\beta$ is given by $ak F^\eta_\alpha F^\eta_\alpha^\ast$, and the Fourier symbol with respect to $G^h_\alpha G^h_\beta$ is given by $ak F^\eta_\alpha F^h_\alpha F^h_\beta F^h_\beta^\ast F^\eta_\beta^\ast$. Here for any $(l,v)$ such that $v \in \mathbb{Z}$ and $1 \leq l \leq a$, we have

$$(F^\eta_\alpha)_{l,v} = \sum_u c^u \delta\left( -f + uf_q + v \frac{f_q}{b} \right) \exp \left( -j 2 \pi l \frac{T_a}{M} \left( f - uf_q - v \frac{f_q}{b} \right) \right).$$

Also, $F^\eta_\alpha$ and $F^h_\alpha$ are infinite diagonal matrices such that for all $l \in \mathbb{Z}$

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

**Proof:** See Appendix [H-E].

Define $G^\eta$ such that its $(\alpha, \beta)$ subblock is $G^\eta_\alpha G^\eta_\beta^\ast$, and $G^h$ such that its $(\alpha, \beta)$ subblock is $G^h_\alpha G^h_\beta^\ast$. Proceeding similarly as in the proof of Theorem 5, we obtain

$$\mathcal{F}(G^\eta) = ak F^\eta F^\eta^\ast \quad \text{and} \quad \mathcal{F}(G^h) = ak F^h F^h^\ast F^\eta^\ast,$$

where $F^\eta$ contain $M \times 1$ submatrices. The $(\alpha,1)$ submatrix of $F^\eta$ is given by $F^\eta_\alpha F^\eta_\alpha^\ast$. 

For any continuous function \(g(x)\), Theorem 5.4 implies that

\[
\lim_{n \to \infty} \frac{1}{naM} \sum_{i=1}^{naM} g \left( \lambda_i \left\{ (G^\eta)^{-\frac{1}{2}} G^h (G^\eta)^{-\frac{1}{2}} \right\} \right)
\]

\[
= \int_{-\frac{\nu}{2}}^{\frac{\nu}{2}} \sum_{i=1}^{aM} g \left( \lambda_i \left( (F^\eta F^\eta^*)^{-\frac{1}{2}} F^\eta F^h F^h^* F^\eta^* (F^\eta F^\eta^*)^{-\frac{1}{2}} \right) \right) \, df.
\]

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

\[
C(f_s) = \lim_{n \to \infty} \frac{1}{naM} \sum_{i=1}^{naM} \left[ \log \left( \nu \lambda_i \left\{ (G^\eta)^{-\frac{1}{2}} G^h (G^\eta)^{-\frac{1}{2}} \right\} \right) \right]^+
\]

\[
= \int_{-\frac{\nu}{2}}^{\frac{\nu}{2}} \sum_{i=1}^{aM} \left[ \log \left( \nu \lambda_i \left( (F^\eta F^\eta^*)^{-\frac{1}{2}} F^\eta F^h F^h^* F^\eta^* (F^\eta F^\eta^*)^{-\frac{1}{2}} \right) \right) \right]^+ \, df,
\]

where the water level \(\nu\) can be computed through the following parametric equation

\[
P = \lim_{n \to \infty} \frac{1}{naM} \sum_{i=1}^{naM} \left[ \nu - \lambda_i \left\{ (G^\eta)^{-\frac{1}{2}} G^h (G^\eta)^{-\frac{1}{2}} \right\} \right]^+
\]

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

**APPENDIX D**

**PROOF OF COROLLARY 2**

Corollary 2 immediately follows from the following proposition.

**Proposition 5.** The \(k\)th largest eigenvalue \(\lambda_k\) of the positive semidefinite matrix \(\tilde{F}_s F_h F^*_h \tilde{F}_s^*\) is bounded by

\[
0 \leq \lambda_k \leq \lambda_k (F_h F^*_h), \quad 1 \leq k \leq M.
\]

These upper bounds can be attained simultaneously by the filter (36).

**Proof:** See Appendix C

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

**APPENDIX E**

**PROOF OF PROPOSITION 2**

It can be easily observed that Algorithm 1 keeps \(K\) subbands in total while zeroing out all others through filtering. Define \(R(f) = H(f)X(f) + N(f)\). In step 2 of the algorithm, the frequency response of the \(i\)th subband being chosen is a linear combination of
\{H \left( f - lf_q \right) X \left( f - lf_q \right) + N \left( f - lf_q \right) \mid l \in \{I_i, I_{i+1}, \cdots, I_K\}\}. More specifically,

\[ Y \left( f - I_i f_q \right) = A^{K+1-i} R \left( f - I_k f_q \right) + \sum_{k=i+1}^{K} B_k R \left( f - I_k f_q \right), \]

where \(|B_k|\) is either \(|A^{K+1-k}|\) or 0. Treating the residual term as noise, the SNR at the \(i\)th branch is

\[
\lim_{A \to \infty} \text{SNR}_i = \frac{|H \left( f - I_i f_p \right)|^2}{S_\eta \left( f - I_i f_p \right)}.
\]

Thus, for large \(A\), this sampling method extracts out \(K\) subbands of highest SNR, and suppresses all other subbands.

Now we need to prove that this is optimal. For any given \(f\), modulation and filtering act as two right-invertible operators on both the signal and the noise. After noise whitening, the equivalent channel matrix is given by

\[
\left( F^\eta F^{\eta^*} \right)^{-\frac{1}{2}} F^\eta F^h,
\]

where \(\left( F^\eta F^{\eta^*} \right)^{-\frac{1}{2}} F^\eta\) is of \(K\) rows orthonormal to each other and \(F^h\) is a diagonal matrix. This implies that modulation along with filtering also plays the role of projecting the frequency components onto a \(K\) dimensional subspace, albeit with respect to the modulated aliased set which is larger than the original aliased set. Applying the same proof as for Corollary\(^2\) the optimal method is to extract out \(K\) subbands with the highest SNR, which coincides with the asymptotic performance of the sampling method derived by Algorithm 1.

**APPENDIX F**

**PROOF OF PROPOSITION\(^4\)**

Denote by \(y^k(t)\) the analog signal after passing through the \(k\)th 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 \(\{y^k[n] \mid 1 \leq k \leq M\}\) is linear. Recall that \(\hat{T}_s = MT_s\) and \(\hat{f}_s = f_s/M\). A linear estimator of \(x(t)\) from \(y[n]\) can be given as

\[
\hat{x}(t) = \sum_{k \in \mathbb{Z}} g^T(t-k\hat{T}_s) \cdot y(k\hat{T}_s),
\]

(95)

where we use the vector form \(g(t) = [g^1(t), \cdots, g^M(t)]^T\) and \(y(t) = [y^1(t), \cdots, y^M(t)]^T\) for notational simplicity. Here, \(g^l(t)\) denotes the interpolation function operating upon the samples in the \(l\)th branch. We propose to find the optimal estimator \(g(t)\) that minimizes the mean square estimation error

\[ \mathbb{E} \left( |x(t) - \hat{x}(t)|^2 \right) \] for some \(t\).
From the orthogonality principle [50], the MMSE estimate \( \hat{x}(t) \) obeys
\[
\mathbb{E} \left( x(t)y^*(lT_s) \right) = \mathbb{E} \left( \hat{x}(t)y^*(lT_s) \right), \quad \forall l \in \mathbb{Z}.
\]
(96)

Since \( x(t) \) and \( \eta(t) \) are both stationary Gaussian processes, we can define \( R_{XY}(\tau) := \mathbb{E} (x(t)y^*(t-\tau)) \) to be the cross correlation function between \( x(t) \) and \( y(t) \), and \( R_Y(\tau) := \mathbb{E} (y(t)y^*(t-\tau)) \) the autocorrelation function of \( y(t) \). Plugging (95) into (96) leads to the following relation
\[
R_{XY}(t-lT_s) = \sum_{k \in \mathbb{Z}} g^T(kT_s - kT_s) R_Y(kT_s - lT_s).
\]
(97)

Replacing \( t \) by \( t + lT_s \), we can equivalently express it as
\[
R_{XY}(t) = \sum_{k \in \mathbb{Z}} g^T(t + lT_s - kT_s) R_Y(kT_s - lT_s) = \sum_{l \in \mathbb{Z}} g^T(t - lT_s) R_Y(lT_s),
\]
(98)

which is equivalent to the convolution of \( g(t) \) and \( R_Y(t) \cdot \sum_{l \in \mathbb{Z}} \delta(t - lT_s) \).

Let \( \mathcal{F}(\cdot) \) denote Fourier transform operator. Define the cross spectral density \( S_{XY}(f) := \mathcal{F}(R_{XY}(t)) \) and \( S_Y(f) = \mathcal{F}(R_Y(t)) \). By taking the Fourier transform on both sides of (98), we have
\[
S_{XY}(f) = G(f) \mathcal{F} \left( R_Y(\tau) \sum_{l \in \mathbb{Z}} \delta(\tau - lT_s) \right),
\]

which immediately yields
\[
G(f) = S_{XY}(f) \left[ \mathcal{F} \left( R_Y(\tau) \sum_{l \in \mathbb{Z}} \delta(\tau - lT_s) \right) \right]^{-1} = S_{XY}(f) \left( \sum_{l \in \mathbb{Z}} S_Y(f - l\tilde{f}_s) \right)^{-1}, \quad \forall f \in \left[-\frac{\tilde{f}_s}{2}, \frac{\tilde{f}_s}{2}\right].
\]
(99)

Since the noise \( \eta(t) \) is independent of \( x(t) \), the cross correlation function \( R_{XY}(t) \) is
\[
R_{XY}(\tau) = \mathbb{E} (x(t + \tau) [(s_1 * h * x)^*(t), \cdots, (s_M * h * x)^*(t)])
\]

which allows the cross spectral density to be derived as
\[
S_{XY}(f) = H^*(f)S_X(f) [S_1^*(f), \cdots, S_M^*(f)].
\]
(100)

Additionally, the spectral density of \( y(t) \) can be given as the following \( M \times M \) matrix
\[
S_Y(f) = \left( |H(f)|^2 S_X(f) + S_\eta(f) \right) S(f)S^*(f),
\]
(101)

with \( S_\eta(f) \) denoting the spectral density of the noise \( \eta(t) \), and \( S(f) = [S_1(f), \cdots, S_m(f)]^T \).

Define \( K(f) := \sum_{l \in \mathbb{Z}} \left( |H(f - l\tilde{f}_s)|^2 S_X(f - l\tilde{f}_s) + N(f - l\tilde{f}_s) \right) S(f - l\tilde{f}_s)S^*(f - l\tilde{f}_s) \). The Wiener-Hopf linear reconstruction filter can now be written as
\[
G(f) = H^*(f)S_X(f)S^*(f)K^{-1}(f),
\]
Define $R_X(\tau) = \mathbb{E}(x(t)x^*(t - \tau))$. Since $\int_{-\infty}^{\infty} S_X(f) \, df = R_X(0)$, the resulting MSE is

$$\xi(t) = \mathbb{E}\left(\left|x(t)\right|^2\right) - \mathbb{E}\left(\left|\hat{x}(t)\right|^2\right) = \mathbb{E}\left(\left|x(t)\right|^2\right) - \mathbb{E}(x(t)\hat{x}^*(t))$$

(102)

$$= R_X(0) - \mathbb{E}\left(x(t) \left(\sum_{l \in \mathbb{Z}} g^T(t - lT_s)y(lT_s)\right)^*\right)$$

(103)

$$= R_X(0) - \sum_{l \in \mathbb{Z}} R_{XY}(t - lT_s)g(t - lT_s).$$

(104)

Define $\zeta(f) := |H(f)S_X(f)|^2 S^*(f)K^{-1}(f)S(f)$. Since $\mathcal{F}(g(-t)) = (G^*(f))^T$ and $S_{XY} = H^*(f)S_X(f)S^*(f)$, Parseval’s identity implies that

$$\xi = \int_{-\infty}^{\infty} \left[S_X(f) - G^*(f)S^T_X\right] \, df$$

$$\xi = \int_{-\infty}^{\infty} \left[S_X(f) - |H(f)S_X(f)|^2 S^*(f)K^{-1}(f)S(f)\right] \, df$$

$$\xi = \int_{-\infty}^{\infty} \left[\sum_{l \in \mathbb{Z}} S_X(f - l\hat{f}_s) - \hat{T}_s V_{\zeta}^T(f, \hat{f}_s) \cdot 1\right] \, df.$$

(105)

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

$$\maximize_{\{S(f - l\hat{f}_s), l \in \mathbb{Z}\}} V_{\zeta}^T(f, \hat{f}_s) \cdot 1,$$

where the objective function can be alternatively rewritten in matrix form

$$\text{trace}\left\{F_X^{\frac{1}{2}}F^*_hF^*_s (F_s (F_h F^*_h + F_\eta) F^*_s)^{-1} F_s F_h F^\frac{1}{2}_X\right\}$$

(105)

Here $F_X$ and $F_\eta$ are diagonal matrices such that $(F_X)_{l,l} = S_X(f - l\hat{f}_s)$ and $(F_\eta)_{l,l} = S_\eta(f + k\hat{f}_s)$. We
observe that
\[
\text{trace}\left\{ \mathbf{F}_s^\dagger \mathbf{F}_h^* \mathbf{F}_s^* (\mathbf{F}_s (\mathbf{F}_h^* + \mathbf{F}_\eta) \mathbf{F}_s^*)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_s^\dagger \right\} = \text{trace}\left\{ (\mathbf{F}_s (\mathbf{F}_h^* + \mathbf{F}_\eta) \mathbf{F}_s^*)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{F}_s \right\}
\]
\[
= \text{trace}\left\{ (\mathbf{F}_s (\mathbf{F}_h^* + \mathbf{F}_\eta) \mathbf{F}_s^*)^{-1} \mathbf{F}_s \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{F}_s \right\}
\]
\[
(a) \text{trace}\left\{ (\mathbf{Y} \mathbf{Y}^*)^{-1/2} \mathbf{Y} (\mathbf{F}_h^* + \mathbf{F}_\eta)^{-1/2} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* (\mathbf{F}_h^* + \mathbf{F}_\eta)^{-1/2} \mathbf{Y}^* \right\}
\]
\[
(b) \text{trace}\left\{ (\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{Y}^* (\mathbf{Y} \mathbf{Y}^*)^{-1} \mathbf{Y} \right\}
\]
\[
(c) \text{trace}\left\{ (\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{Y}^* \right\}
\]
\[
(d) \sup_{\mathbf{Z} \cdot \mathbf{Z} = \mathbf{I}_M} \text{trace}\left\{ \mathbf{Z} (\mathbf{F}_h \mathbf{F}_h^* + \mathbf{F}_\eta)^{-1} \mathbf{F}_h \mathbf{F}_X \mathbf{F}_h^* \mathbf{Z} \right\}
\]
\[
= \sum_{i=1}^{M} \lambda_i(\mathbf{D}),
\]
where (a) follows by introducing $\mathbf{Y} := \mathbf{F}_s (\mathbf{F}_h^* + \mathbf{F}_\eta)^{1/2}$, (b) follows from the fact that $\mathbf{F}_h$, $\mathbf{F}_X$, $\mathbf{F}_\eta$ are all diagonal matrices, (c) follows by introducing $\tilde{\mathbf{Y}} = (\mathbf{Y} \mathbf{Y}^*)^{-1/2} \mathbf{Y}$, and (d) follows by observing that $\tilde{\mathbf{Y}} \mathbf{Y}^* = (\mathbf{Y} \mathbf{Y}^*)^{-1/2} \mathbf{Y} \mathbf{Y}^* (\mathbf{Y} \mathbf{Y}^*)^{-1/2} = \mathbf{I}$. Here, $\mathbf{D}$ is an infinite diagonal matrix such that $D_{l,l} = |H(f-lf_s)|^2 S_X(f-lf_s) / |H(f-lf_s)|^2 S_X(f-lf_s) + S_\eta(f-lf_s)$. In other words, the upper bound is the sum of the $M$ largest $\mathbf{D}_{l,i}$ which are associated with $M$ frequency points of highest SNR.

Therefore, when restricted to the set of all permutations of $\{S_X(f), S_X(f \pm fs), \cdots\}$, the minimum MSE is achieved when assigning the $M$ largest $S_X(f + lf_s)$ to $M$ branches with the largest SNR. In this case, the corresponding optimal filter can be chosen such that
\[
S_k(f - lf_s) = \begin{cases} 
1, & \text{if } l = \hat{k} \\
0, & \text{otherwise.}
\end{cases}
\]

where $\hat{k}$ is the index of the $k^{th}$ largest element in $\left\{ |H(f-lf_s)|^2 / S_\eta(f-lf_s) : l \in \mathbb{Z} \right\}$.

\section*{APPENDIX G}
\textbf{PROOF OF PROPOSITION}$^5$

Recall that at a given $f$, $\mathbf{F}_h$ is an infinite diagonal matrix satisfying $(\mathbf{F}_h)_{l,l} = H \left( f - \frac{lf}{4^l} \right)$ for all $l \in \mathbb{Z}$, and that $\tilde{\mathbf{F}}_s = (\mathbf{F}_s \mathbf{F}_s^*)^{-1/2} \mathbf{F}_s$. 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
\[
\tilde{\mathbf{F}}_s \left( \tilde{\mathbf{F}}_s^* \right)^* = (\mathbf{F}_s \mathbf{F}_s^*)^{-1/2} \mathbf{F}_s \mathbf{F}_s^* (\mathbf{F}_s \mathbf{F}_s^*)^{-1/2} = \mathbf{I},
\]
which indicates that the rows of $\tilde{F}$ are orthonormal. Hence, the operator norm of $\tilde{F}$ is no larger than 1, which leads to

$$
\lambda_1 \left( \tilde{F}_s F_h F^*_h \tilde{F}^*_s \right) = \left\| \tilde{F}_s F_h \right\|_2^2 \leq \left\| \tilde{F}_s \right\|_2^2 \left\| F_h \right\|_2^2 \leq \left\| F_h \right\|_2^2 = \lambda_1 \left( F_h F^*_h \right). \tag{115}
$$

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

Suppose that $v_k$ is the eigenvector associated with the $k$th largest eigenvalue $\lambda_k$ of $\tilde{F}_s F_h F^*_h \tilde{F}^*_s$, and denote by $\left( \tilde{F}_s \right)_j$ the $j$th column of $\tilde{F}_s$. Since $\tilde{F}_s F_h F^*_h \tilde{F}^*_s$ is Hermitian positive semidefinite, its eigendecomposition yields an orthogonal basis of eigenvectors. Observe that $\{v_1, \cdots, v_k\}$ spans a $k$-dimensional space and that $\left\{ \left( \tilde{F}_s \right)_j, 1 \leq j \leq k - 1 \right\}$ spans a subspace of dimension no more than $k - 1$. For any $k \geq 2$, there exists $k$ scalars $a_1, \cdots, a_k$ such that

$$
\left( \sum_{i=1}^k a_i v_i \right) \perp \left\{ \left( \tilde{F}_s \right)_j, 1 \leq j \leq k - 1 \right\} \quad \text{and} \quad \sum_{i=1}^k a_i v_i \neq 0. \tag{116}
$$

This allows us to define the following unit vector

$$
\tilde{v}_k \triangleq \sum_{i=1}^k \frac{a_i}{\sqrt{\sum_{j=1}^k |a_j|^2}} v_i, \tag{117}
$$

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

$$
\left\| \tilde{F}_s F_h F^*_h \tilde{F}^*_s \tilde{v}_k \right\|_2^2 = \left\| \sum_{i=1}^k \frac{a_i}{\sqrt{\sum_{j=1}^k |a_j|^2}} \tilde{F}_s F_h F^*_h \tilde{F}^*_s v_i \right\|_2^2 \tag{118}
$$

$$
= \left\| \sum_{i=1}^k \frac{a_i}{\sqrt{\sum_{j=1}^k |a_j|^2}} v_i \right\|_2^2 \geq \lambda_k^2. \tag{119}
$$

Define $u_k := \tilde{F}^*_s \tilde{v}_k$. From (116) we can see that $(u_k)_i = \left( \left( \tilde{F}_s \right)_i, \tilde{v}_i \right) = 0$ holds for all
\[ i \in \{i_1, i_2, \ldots, i_{k-1}\}. \text{ In other words, } u_k \perp \{e_{i_1}, \ldots, e_{i_{k-1}}\}. \text{ This further implies that} \]
\[
\lambda_k^2 \leq \left\| \tilde{F}_s F_h F_h^* \tilde{v}_k \right\|_2^2 \leq \left\| F_h F_h^* F_h^* \tilde{v}_k \right\|_2^2 \]  
(121)
\[
\leq \left\| F_h F_h^* u_k \right\|_2^2 \]  
(122)
\[
\leq \sup_{x \perp \text{span}\{e_{i_1}, \ldots, e_{i_{k-1}}\}} \left\| F_h F_h^* x \right\|_2^2 \]  
(123)
\[
= \lambda_k^2 (F_h F_h^*) \]  
(124)
by observing that \(F_h F_h^*\) is a diagonal matrix.

Setting
\[
S_k \left( f - \frac{l f_s}{M} \right) = \begin{cases} 
1, & \text{if } \left| H \left( f - \frac{l f_s}{M} \right) \right| \leq \lambda_k (F_h (f) F_h^*(f)), \\
0, & \text{otherwise}, 
\end{cases} \]  
(125)
yields \(\tilde{F}_s = F_s\) and hence \(\tilde{F}_s F_h F_h^* \tilde{F}_s\) is a diagonal matrix such that
\[
(F_s F_h F_h^* \tilde{F}_s)_{k,k} = \lambda_k (F_h F_h^*). \]  
(126)

Apparently, this choice of \(S_k(f)\) allows the upper bounds
\[
\lambda_k (\tilde{F}_s F_h F_h^* \tilde{F}_s) = \lambda_k (F_h F_h^*), \quad \forall 1 \leq k \leq M \]  
(127)
to be attained simultaneously.

**APPENDIX H**

**Proofs of Auxiliary Lemmas**

**A. Proof of Lemma 7**

Suppose that \(T = n T_s + T_0\) where \(0 < T_0 < T_s\). Then
\[
\frac{1}{T} \sup x(0, T] \{y[n]\} = \frac{n T_s}{T} \frac{1}{n T_s} \sup x(0, T] \{y[n]\} 
\geq \frac{n T_s}{T} \frac{1}{n T_s} \sup x(0, n T_s] \{y[n]\}. \]

Similarly, we have
\[
\frac{1}{T} \sup x(0, T] \{y[n]\} \leq \frac{(n + 1) T_s}{T} \frac{1}{(n + 1) T_s} \sup x(0, (n + 1) T_s] \{y[n]\}. \]

Combining the above inequalities we obtain
\[
\lim_{n \to \infty} \frac{n T_s}{T} \frac{1}{n T_s} \sup x(0, n T_s] \{y[n]\} \leq \lim_{n \to \infty} \frac{1}{n T_s + T_0} \sup x(0, n T_s + T_0] \{y[n]\} 
\leq \lim_{n \to \infty} \frac{(n + 1) T_s}{T} \frac{1}{(n + 1) T_s} \sup x(0, (n + 1) T_s] \{y[n]\}. \]
Since the transmission time $T$ is an integer multiple of $T_s$, the lower and upper bounds in the above inequality equal the channel capacity, which immediately leads to the limit when $T$ is not an integer multiple of $T_s$: 

$$
\lim_{n \to \infty} \frac{1}{nT_s + T_0} \sup_{x(0, nT_s + T_0)} I(x, \{y_s[n]\}) = \lim_{n \to \infty} \frac{1}{nT_s} \sup_{x(0, nT_s)} I(x, \{y_s[n]\}).
$$

(128)

Since $T_0$ can be arbitrarily chosen from the interval $(0, T_s)$, we conclude that

$$
\lim_{T \to \infty} \frac{1}{T} \sup_{x(0, T)} I(x, \{y_s[n]\}) = \lim_{n \to \infty} \frac{1}{nT_s} \sup_{x(0, nT_s)} I(x, \{y_s[n]\}).
$$

(129)

**B. Proof of Lemma 4**

For any $i \leq j$, we have

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

(130)

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

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

By our assumptions, we have $h(t) = o(t^{-\epsilon})$ for some constant $\epsilon > 1$. Since $s(t)$ is also absolutely integrable, $\tilde{h}(t) = o(t^{-\epsilon})$ holds as well. Without loss of generality, suppose that $j \geq i$, then we have

(a) if $i \geq n^{\frac{1}{\epsilon}}$, by the assumption $\tilde{h}(t) = o \left( \frac{1}{\sqrt{n}} \right)$ for some $\epsilon > 1$, we have

$$
\left| R_{ij}^1 \right| \leq \sum_{t = -\infty}^{-j} \left\| \tilde{h}_{j-i+t} \right\|_1 \left\| \tilde{h}_t \right\|_\infty 
$$

$$
\leq \left( \max_{\tau \geq n^{\frac{1}{\epsilon}}} \left\| \tilde{h}_{j-i+t} \right\|_1 \right) \sum_{t = -\infty}^{-j} \left\| \tilde{h}_t \right\|_\infty 
$$

$$
\leq \left( \max_{\tau \geq n^{\frac{1}{\epsilon}}} \left\| \tilde{h}_{j-i+t} \right\|_1 \right) \sum_{t = -\infty}^{j} \left\| \tilde{h}_t \right\|_1 
$$

$$
\leq c \max_{\tau \geq n^{\frac{1}{\epsilon}}} \left\| \tilde{h}_{j-i+t} \right\|_1 
$$

$$
= kc \cdot o \left( \frac{1}{\sqrt{n}} \right) = o \left( \frac{1}{\sqrt{n}} \right). 
$$
(b) if $j \geq n^{\frac{1}{2}}$

\[ |R_{ij}^1| \leq \sum_{t=-\infty}^{-j} \| \tilde{h}_{j-i-t} \|_1 \| \tilde{h}_t \|_\infty \]

\[ \leq \left( \sum_{t=-\infty}^{-j} \| \tilde{h}_{j-i-t} \|_1 \right) \max_{\tau \leq -j} \| \tilde{h}_\tau \|_\infty \]

\[ \leq c \max_{\tau \geq n^{\frac{1}{2}}} \| \tilde{h}_{-\tau} \|_\infty \]

\[ = c \cdot o \left( \frac{1}{\sqrt{n}} \right) = o \left( \frac{1}{\sqrt{n}} \right). \]

(c) if $j < n^{\frac{1}{2}}$ and $i < n^{\frac{1}{2}}$, Cauchy-Schwartz inequality yields

\[ |R_{ij}^1|^2 \leq \left( \sum_{t=-\infty}^{+\infty} \| \tilde{h}_{j-i-t} \|_1 \right)^2 \left( \sum_{t=-\infty}^{+\infty} \| \tilde{h}_t \|_1 \right)^2 \]

\[ \leq \left( \sum_{t=-\infty}^{+\infty} \| \tilde{h}_t \|_1 \right)^4 \leq c^4. \]

Hence, we have

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} |R_{ij}^1|^2 \leq \lim_{n \to \infty} \frac{1}{n} \left[ \left( n^{\frac{1}{2}} \right)^2 c^4 + 2n^{1+\frac{1}{2}} o \left( \frac{1}{n} \right) \right] = 0. \]

Similarly, we can show that

\[ \lim_{n \to \infty} \frac{1}{n} \| R^2 \|_F^2 = 0, \]

which immediately implies that

\[ \lim_{n \to \infty} \frac{1}{n} \left\| \hat{H}^n - \tilde{H}^n \tilde{H}^{n*} \right\|_F^2 = 0. \]

Since $\hat{H}^n$ is a Toeplitz matrix, applying [51, Lemma 6] and [51, Section 4.1] yields

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

\[ \leq 2 \sum_{t=-\infty}^{+\infty} \left\| \tilde{h}_t \right\|_\infty \sum_{i=0}^{\infty} \left\| \tilde{h}_{i+t} \right\|_1 \]

\[ \leq 2c^2. \]
Additionally, since $\tilde{H}^n$ is a block Toeplitz matrix, [52, Corollary 4.2] allows us to bound the norm as

$$
\| \tilde{H}^n \tilde{H}^n^* \|_2 = \| \tilde{H}^n \|_2^2 \leq \| F_{\tilde{H}}(\omega) \|_{\infty}^2
$$

$$
= \sup_{\omega} \sum_{i=0}^{k-1} | F_{\tilde{h}_{i+1}}(\omega) |^2
$$

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

$$
\leq c^2.
$$

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

C. Proof of Lemma [5]

We know that $S^n S^n^* = \hat{S}^n$, hence, $C^n \sim \hat{S}^n = S^n S^n^*$. Recall that $\left( \hat{S}^n \right)_{1i} = \sum_{t=-\infty}^{\infty} s_{i-1+t}s_t^*$. For a given $k$, the Fourier series related to $\{C^n\}$ can be given as

$$
F_{C}^k(\omega) = \sum_{i=-\infty}^{\infty} \left( \sum_{t=-\infty}^{\infty} s_{i+t}s_t^* \right) \exp(ji\omega).
$$

(134)

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

$$
F_{C}^k(\omega) \approx \Delta \sum_{i=-\infty}^{\infty} \left( \int_{-\infty}^{+\infty} s(t + \tau)s^*(t)\,dt \right) \exp(ji\omega)
$$

$$
= \Delta \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} s(t + \tau)s(t)\,dt \right) \left( \sum_{i=0}^{\infty} \delta(\tau - i\tau_s) \right) \exp(j\omega T_s \tau) \,d\tau.
$$

(135)

(136)

We observe that

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

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

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

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

$$
F_{C}^k(\omega) \approx \frac{\Delta}{T_s} \sum_{i=-\infty}^{\infty} \left| S\left(-j \left( \frac{\omega}{T_s} - \frac{i2\pi}{T_s} \right) \right) \right|^2.
$$

(137)
If for all \( \omega \in [-\pi, \pi] \), we have
\[
\sum_{i=-\infty}^{\infty} \left| S \left( -j \left( \frac{\omega}{T_s} - \frac{i2\pi}{T_s} \right) \right) \right|^2 \geq \epsilon_s > 0
\]  
for some constant \( \epsilon_s \), then \( \sigma_{\min}(C^n) = \inf_{\omega} F_c(\omega) \geq \frac{\Delta \epsilon_s}{T_s} \), which leads to \( \left\| (C^n)^{-1} \right\|_2 \leq \frac{T_s}{\Delta \epsilon_s} \).

Let \( \Xi^n = C^n - S^nS^{n*} \). Since \( S^nS^{n*} \sim C^n \), we can have \( \lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \Xi^n \right\|_F = 0 \), which implies that
\[
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \Xi^n (C^n)^{-1} \right\|_F \leq \lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| \Xi^n \right\|_F \left\| (C^n)^{-1} \right\|_2 \leq \lim_{n \to \infty} \frac{T_s}{\Delta \epsilon_s} \frac{1}{\sqrt{n}} \left\| \Xi^n \right\|_F = 0.
\]  
The Taylor expansion of \( (S^nS^{n*})^{-1} \) yields
\[
(S^nS^{n*})^{-1} = (C^n - \Xi^n)^{-1} = (C^n)^{-1} \left( I + \Xi^n (C^n)^{-1} + \left( \Xi^n (C^n)^{-1} \right)^2 + \cdots \right).
\]  
Hence, we can bound
\[
\lim_{n \to \infty} \frac{1}{\sqrt{n}} \left\| (S^nS^{n*})^{-1} - (C^n)^{-1} \right\|_F \leq \lim_{n \to \infty} \left\| (C^n)^{-1} \right\|_2 \left( \sum_{i=1}^{\infty} \left( \frac{1}{\sqrt{n}} \left\| \Xi^n \right\|_F \right)^i \right) = 0.
\]  

**D. Proof of Lemma 6**

Since \( (C^n)^{-\frac{1}{2}} \) and \( \left( \hat{S}^n \right)^{-\frac{1}{2}} \) are both Hermitian and positive semidefinite, we have \( (C^n)^{-\frac{1}{2}} \sim \left( \hat{S}^n \right)^{-\frac{1}{2}} \).

We observe that \( (C^n)^{-\frac{1}{2}} = U_c \Lambda_c^{-\frac{1}{2}} U_c^T \), which is also a circulant matrix. Combining the above results with Lemma 2 yields
\[
(S^nS^n)^{-\frac{1}{2}} \hat{H}^n \hat{H}^{n*} (S^nS^{n*})^{-\frac{1}{2}} \sim (C^n)^{-\frac{1}{2}} \hat{H}^n (C^n)^{-\frac{1}{2}}
\]  
where \( (C^n)^{-\frac{1}{2}} \) and \( \hat{H}^n \) are both Hermitian Toeplitz matrices. Denote by \( F_{c_{0.5}}(\omega), F_c(\omega), F_h(\omega), F_{\hat{h}}(\omega) \) the Fourier series related to \( (C^n)^{\frac{1}{2}}, C^n, \hat{H}^n \) and \( \hat{H}^n \), respectively. We note that \( F_{c_{0.5}}(\omega), F_c(\omega) \) and \( F_h(\omega) \) are all scalars since their related matrices are Toeplitz, while \( F_{\hat{h}}(\omega) \) is a \( 1 \times k \) vector since \( \hat{H} \) is block Toeplitz. Then for any continuous function \( g(x) \), applying [51] Theorem 12 yields
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g \left\{ \lambda_i \left( (S^nS^n)^{-\frac{1}{2}} \hat{H}^n \hat{H}^{n*} (S^nS^{n*})^{-\frac{1}{2}} \right) \right\}
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g \left\{ \lambda_i \left( (C^n)^{-\frac{1}{2}} \hat{H}^n (C^n)^{-\frac{1}{2}} \right) \right\}
= \frac{1}{2\pi} \int_{-\pi}^{\pi} g \left( F_{c_{0.5}}(\omega) F_h(\omega) F_{c_{0.5}}^{-1}(\omega) \right) d\omega
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g \left\{ \lambda_i \left( (C^n)^{-1} \hat{H}^n \right) \right\}
= \frac{1}{2\pi} \int_{-\pi}^{\pi} g \left( \frac{F_h(\omega)}{F_c(\omega)} \right) d\omega.
\]
We observe that $\hat{H}_n$ is asymptotically equivalent to $\hat{H}^n \hat{H}^{n^*}$, and the eigenvalues of $\hat{H}^n \hat{H}^{n^*}$ are exactly the square of the corresponding singular values of $\hat{H}^n$. Hence, we know from [52] that for any continuous function $g(x)$:

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g \left\{ \lambda_i \left( \hat{H}^n \right) \right\} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g \left\{ \sigma_i^2 \left( \hat{H}^n \right) \right\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g \left( \sigma^2 \left( F_{h}^{(n)}(\omega) \right) \right) d\omega \quad (144)
$$

where $F_{h}^{(n)}(\omega)$ can be expressed as $F_{h}^{(n)}(\omega) = [F_{h,0}^{(n)}(\omega), \ldots, F_{h,k-1}^{(n)}(\omega)]$. Here, for any $0 \leq i < k$:

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

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

Through algebraic manipulation, we have that

$$
F_{h,i}^{(n)}(\omega) = \Delta \sum_{l=-\infty}^{+\infty} H (-f + l f_s) \exp(-j2\pi (f - l f_s) i\Delta),
$$

which yields

$$
F_{h}^{(n)}(f) = \sigma^2 \left( F_{h}^{(n)}(2\pi f) \right) = \sum_{i=0}^{k-1} \left| F_{h,i}^{(n)}(2\pi f) \right|^2 \quad (145)
$$

and

$$
\frac{\Delta^2}{T_s^2} \sum_{i=0}^{k-1} \left| \sum_{l=-\infty}^{+\infty} \tilde{h}(-f + l f_s) \exp(-j2\pi (f - l f_s) i\Delta) \right|^2 \quad (146)
$$

Similarly, we have

$$
F_{c}(f) = \frac{\Delta}{T_s} \sum_{u=-\infty}^{+\infty} \left| S(-f + l f_s) \right|^2. \quad (151)
$$

Combining the above results yields

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

and

$$
\frac{T_s}{\int_{-f_s/2}^{f_s/2} g \left( \frac{\sum_{l=-\infty}^{+\infty} \left| H(-f + l f_s) \right| S(-f + l f_s) \right) d f}{\sum_{l=-\infty}^{+\infty} \left| S(-f + l f_s) \right|^2} \quad (153)
$$
This completes the proof.

E. Proof of Lemma 7

Denote by $K_\alpha^0$ the Fourier symbol associated with the block Toeplitz matrix $G_\alpha^0$. We know that the Fourier transform of $g_\alpha^0(t, \tau)$ with respect to $\tau$ is given by

$$
\int_{\tau_2} g_{\alpha}^0(t, \tau) \exp(-j2\pi f \tau) \, d\tau
$$

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

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

$$
= P_i(-f) \cdot \left[ S_i(-f) \exp(-j2\pi tf) \cdot \sum_{u} c_i^u \delta(f - u f_q) \right]
$$

$$
= P_i(-f) \cdot \left[ \sum_{u} c_i^u S_i(-f + u f_q) \exp(-j2\pi f (f - u f_q)) \right].
$$

For any $(l, m)$ such that $1 \leq l \leq a$ and $1 \leq m \leq ak$, the $(l, m)$ entry of the Fourier symbol $K_\alpha^0$ can be related to the sampling sequence of $g_\alpha^0(l T_s, \tau)$ at a rate $\frac{T_s}{\alpha}$ with a phase shift $m \Delta$, and hence it can be calculated as follows

$$(K_\alpha^0)_{l,m} = \sum_{v} P_i(-f + v \frac{f_q}{b}) \exp \left( j2\pi \left( f - v \frac{f_q}{b} \right) m \Delta \right).
$$

$$
\left[ \sum_{u} c_i^u S \left( -f + u f_q + v \frac{f_q}{b} \right) \exp \left( -j2\pi l T_s \left( f - u f_q - v \frac{f_q}{b} \right) \right) \right] .
$$

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

$$(K_\alpha^0 K_\beta^{*})_{l,d} = ak \sum_{v} P_{\alpha}(-f + v \frac{f_q}{b}) \left[ \sum_{u_1} c_{\alpha}^{u_1} S_{\alpha} \left( -f + u_1 f_q + v \frac{f_q}{b} \right) \exp \left( -j2\pi l T_s \left( f - u_1 f_q - v \frac{f_q}{b} \right) \right) \right].
$$

$$
P_{\beta}^{*} \left( -f + v \frac{f_q}{b} \right) \left[ \sum_{u_2} c_{\beta}^{u_2} S_{\beta} \left( -f + u_2 f_q + v \frac{f_q}{b} \right) \exp \left( -j2\pi l T_s \left( f - u_2 f_q - v \frac{f_q}{b} \right) \right) \right]^{*}.
$$

Define another matrix $F_\alpha^0$ such that

$$(F_\alpha^0)_{l,v} = \sum_{u} c_{\alpha}^{u} S_{\alpha} \left( -f + u f_q + v \frac{f_q}{b} \right) \exp \left( -j2\pi l T_s \left( f - u f_q - v \frac{f_q}{b} \right) \right).
$$

It can be easily seen that

$$
K_\alpha^0 K_\beta^{*} = ak F_\alpha^0 F_\alpha^{*} F_\alpha^{*} F_\beta^{*}.
$$

Replacing $P_{\alpha}$ by $P_{\alpha} H$ immediately gives us the Fourier symbol for $G_\alpha^h G_\beta^{h*}$.
REFERENCES

[1] R. G. Gallager, *Information theory and reliable communication*. New York: John Wiley & Sons, Inc, 1968.

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

[3] J. L. Holsinger, “Digital communication over fixed time-continuous channels with memory, with special application to telephone channels,” *MIT Research Lab of Electronics, Tech. Rept. 430*, 1964.

[4] U. Grenander and G. Szego, *Toeplitz forms and their applications*. New York: American Mathematical Society, 1984.

[5] W. Hirt and J. Massey, “Capacity of the discrete-time Gaussian channel with intersymbol interference,” *IEEE Transactions on Information Theory*, vol. 34, no. 3, pp. 38–38, May. 1988.

[6] A. Ruiz, J. Cioffi, and S. Kasturia, “Discrete multiple tone modulation with coset coding for the spectrally shaped channel,” *IEEE Transactions on Communications*, vol. 40, no. 6, pp. 1012–1029, Jun 1992.

[7] I. Telatar and D. Tse, “Eigenvalue estimates and mutual information for the linear time-varying channel,” *accepted to IEEE Transactions on Information Theory*, pp. 1384–1400, 2011.

[8] M. Medard and R. G. Gallager, “Bandwidth scaling for fading multipath channels,” *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 840–852, Apr. 2002.

[9] S. Borade and L. Zheng, “Wideband fading channels with feedback,” *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6058–6065, Dec. 2010.

[10] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. - Dec. 1999.

[11] T. Marzetta and B. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, Jan 1999.

[12] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, “Capacity limits of mimo channels,” *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 5, pp. 684–702, June 2003.

[13] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 933–946, May. 2000.

[14] G. Durisi, U. Schuster, H. Bolcskei, and S. Shamai, “Noncoherent capacity of underspread fading channels,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

[15] B. Farrell and T. Strohmer, “Eigenvalue estimates and mutual information for the linear time-varying channel,” *accepted to IEEE Transactions on Information Theory*, 2011.

[16] P. Bello, “Characterization of randomly time-variant linear channels,” *IEEE Transactions on Communications Systems*, vol. 11, no. 4, pp. 360–393, Dec. 1963.

[17] G. D. Forney and G. Ungerboeck, “Modulation and coding for linear Gaussian channels,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2384–2415, Oct 1998.

[18] Y. C. Eldar and T. Michaeli, “Beyond bandlimited sampling,” *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 48–68, May. 2009.

[19] A. Aldroubi and K. Gröchenig, “Nonuniform sampling and reconstruction in shift-invariant spaces,” *SIAM Review*, vol. 43, no. 4, pp. 585–620, 2001.

[20] A. Jerri, “The Shannon sampling theorem: Its various extensions and applications: A tutorial review,” *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977.

[21] J. Yen, “On nonuniform sampling of bandwidth-limited signals,” *IRE Transactions on Circuit Theory*, vol. 3, no. 4, pp. 251–257, Dec. 1956.
[22] A. Papoulis, “Generalized sampling expansion,” *IEEE Transactions on Circuits and Systems*, vol. 24, no. 11, pp. 652 – 654, Nov. 1977.

[23] W. Wu, S. Vishwanath, and A. Arapostathis, “Capacity of a class of cognitive radio channels: Interference channels with degraded message sets,” *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4391 –4399, Nov. 2007.

[24] S. Jafar and S. Srinivasa, “Capacity limits of cognitive radio with distributed and dynamic spectral activity,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 529 –537, April 2007.

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

[26] Y.-P. Lin and P. Vaidyanathan, “Periodically nonuniform sampling of bandpass signals,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 3, pp. 340 –351, Mar. 1998.

[27] M. Unser and J. Zerubia, “A generalized sampling theory without band-limiting constraints,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 959 –969, Aug. 1998.

[28] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[29] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[30] M. Mishali and Y. C. Eldar, “From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375 –391, Apr. 2010.

[31] ———, “Blind multiband signal reconstruction: Compressed sensing for analog signals,” *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993 –1009, March 2009.

[32] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31 –40, Mar. 2008.

[33] K. Gedalyahu, R. Tur, and Y. C. Eldar, “Multichannel sampling of pulse streams at the rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1491 –1504, April 2011.

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

[35] M. Matthews, “On the linear minimum-mean-squared-error estimation of an undersampled wide-sense stationary random process,” *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 272 –275, Jan. 2000.

[36] T. Michaeli and Y. C. Eldar, “High-rate interpolation of random signals from nonideal samples,” *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 977 –992, Mar. 2009.

[37] T. Berger and D. Tufts, “Optimum pulse amplitude modulation–I: Transmitter-receiver design and bounds from information theory,” *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 196 – 208, Apr. 1967.

[38] D. Tufts and T. Berger, “Optimum pulse amplitude modulation–ii: Inclusion of timing jitter,” *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 209 – 216, Apr 1967.

[39] T. Berger, *Nyquist’s Problem in Data Transmission Theory*. Cambridge: PhD dissertation, Harvard University, December 1965.

[40] D. Chan and R. Donaldson, “Optimum pre- and postfiltering of sampled signals with application to pulse modulation and data compression systems,” *IEEE Transactions on Communication Technology*, vol. 19, no. 2, pp. 141 –157, Apr. 1971.

[41] T. Ericson, “Optimum PAM filters are always band limited (corresp.),” *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 570 – 573, Jul 1973.
[42] D. Guo, S. Shamai, and S. Verdu, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. on Info. Theory*, vol. 51, no. 4, pp. 1261 –1282, Apr. 2005.

[43] S. Shamai, “Information rates by oversampling the sign of a bandlimited process,” *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1230 –1236, Jul. 1994.

[44] E. N. Gilbert, “Increased information rate by oversampling,” *IEEE Trans. on Information Theory*, vol. 39, no. 6, pp. 1973 –1976, Nov. 1993.

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

[46] M. Unser and A. Aldroubi, “A general sampling theory for nonideal acquisition devices,” *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2915 –2925, Nov. 1994.

[47] Y. C. Eldar and M. Unser, “Nonideal sampling and interpolation from noisy observations in shift-invariant spaces,” *IEEE Trans. on Signal Processing*, vol. 54, no. 7, pp. 2636 –2651, Jul. 2006.

[48] A. J. Goldsmith, *Wireless communications*. New York: Cambridge University Press, 2005.

[49] A. Lapidoth, *A Foundation in Digital Communication*. New York: Cambridge University Press, 2009.

[50] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.

[51] R. Gray, *Toeplitz and circulant matrices: A review*, ser. Foundations and Trends in Communications and Information Theory. NOW Publisher, 2006, vol. 2, no. 3.

[52] P. Tilli, “Singular values and eigenvalues of non-Hermitian block Toeplitz matrices,” *Linear Algebra and its Applications*, vol. 272, no. 1-3, pp. 59 – 89, 1998.