Forty-Ninth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 28 - 30, 2011

Shannon Meets Nyquist: The Interplay between Capacity and Sampling

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

***Abstract*—We explore several 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 is the optimal input and sampling strategy at a given sub-Nyquist sampling rate. In particular, we derive the capacity of sampled analog channels for several sampling mechanisms, including a filter followed by sampling and a filter bank followed by sampling. Connections between sampling and MIMO Gaussian channels are illuminated based on this analysis. These results demonstrate the tradeoff between channel capacity and sampling rate, and also illustrate the interplay between sampling techniques and capacity of sampled analog channels.**

***Index Terms*—sampled analog channels, sub-Nyquist sampling, channel capacity**

I. Introduction

The capacity of continuous-time waveform channels and the corresponding capacity-achieving water-filling power allocation strategy over frequency is well known [1] and underlies practical protocol design in OFDM systems. Although receiver analysis and design in modern communication systems is based on digital sequences (obtained by sampling the received analog signals), the information content of these signals can be preserved if noise outside the channel bandwidth is filtered out and the filtered output is sampled above the Nyquist rate associated with the signal bandwidth. The majority of information theoretic work thus implicitly assumes Nyquist rate sampling without accounting for hardware limitations, which may preclude sampling at this rate, especially for wideband communications. In this paper we explore how channel capacity is affected by reduced-rate sampling, namely, how much information, in the Shannon sense, can be conveyed through a sampled analog channel which is sampled at a rate below the Nyquist rate. Bridging sampling theory and information theory, our work attempts to characterize the tradeoff between fundamental data rate limits and sampling rate constraints.

A. Related Work and Motivation

The derivation of the capacity of linear time-invariant (LTI) waveform channels was pioneered by Shannon. Making use of the asymptotic spectral properties of Toeplitz operators [1] or, alternatively, Fourier analysis [2], this capacity result established the optimality of water-filling power allocation based on signal-to-noise power across the frequency domain. Information theoretic work on communication channels often implicitly assumes that sampling is performed above the Nyquist rate, and therefore preserves information. Forney *et. al.* [3] surveys minimum-bandwidth orthogonal pulse amplitude modulation (PAM) techniques for transmission over 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 [4], [5]. 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 a fundamental sampling rate requirement for perfect reconstruction of band-limited signals. This sampling rate requirement may be excessive for more general classes of signals like multiband signals, whose spectral contents reside continuously within several subbands over a wide spectrum. Landau [6] characterized the minimum sampling rate required in this setting – the sum of the bandwidths of the spectral support (termed the *Landau rate*). However, these works aim at finding optimal sampling and reconstruction mechanisms that achieve perfect reconstruction of a class of analog signals from *noiseless* samples. Another line of work pioneered by Berger *et. al.* [7]–[11] investigated joint optimization of the transmitted pulse shape and receiver prefiltering in PAM over an analog communication channel under sub-Nyquist sampling. 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, the optimal filters derived in [7], [9] are used to determine an SNR metric which in turn is used to compute an approximation to sampled channel 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.

The tradeoff between capacity and hardware complexity has been studied in another line of work focused on sampling precision [12]–[14]. 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.

Y. Chen and A. Goldsmith are with Stanford University (email: {yxchen@, andrea@ee.}stanford.edu), and Y. C. Eldar is with Technion, Israel Institute of Technology (email: yonina@ee.technion.ac.il). This work is supported in part by the NSF Center for Science of Information, the Interconnect Focus Center of the Semiconductor Research Corporation, and BSF Transformative Science Grant 2010505.

978-1-4577-1818-2/11/$26.00 ©2011 IEEE

B. 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 $H(f) = \int_{-\infty}^{\infty} h(t) \exp(-j2\pi ft)\mathrm{d}t$. The analog channel output is

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

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

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

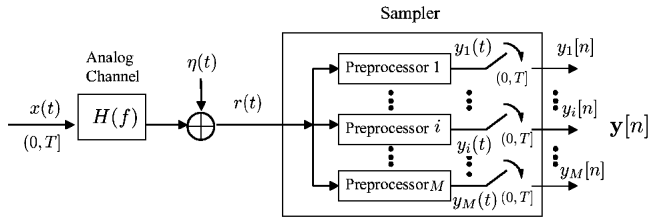

Figure 1. Sampled analog channel: The input $x(t)$, constrained to $(0, T]$, is passed through M branches of the receiver analog front end to yield analog outputs $\{y_i(t) : 1 \le i \le M\}$; each analog output $y_i(t)$ is observed over $(0, T]$ and uniformly sampled by a sampler at a rate $f_s/M = (MT_s)^{-1}$ samples per second to yield the sampled sequence $y_i[n]$. The preprocessor can be a filter, or a filter and a modulator followed by another filter.

Defining the sampled sequence as $\mathbf{y}[n] = [y_1[n], \cdots, y_M[n]]$, 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

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

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

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

C. Contribution

Sampling structures typically rely on general prefiltering prior to sampling [15], 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 [16], [17]. Here, we explore sampled analog channels with the following two 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) filter banks followed by sampling: the analog channel output is passed through a bank of LTI filters, each followed by an ideal uniform sampler (see Fig. 7). 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.
- **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. 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.

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

II. A Filer Followed by Sampling

A. Problem Formulation

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 [15]. 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

$$y(t) = s(t) * (h(t) * x(t) + \eta(t)), \quad t \in (0, T]. \tag{2}$$

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 $\{y[n]\}_{(0,T]}$.

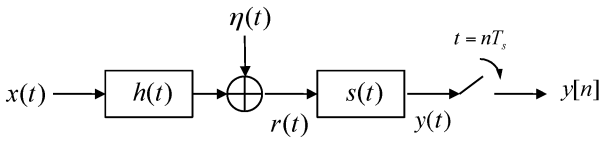

Figure 2. Filtering followed by sampling: the analog channel output $r(t)$ is linearly filtered prior to ideal uniform sampling.

B. Capacity Results

Since both the channel and the prefilter are assumed to be LTI, applying the prefilter equivalently generates a new channel with channel gain $H(f)S(f)$. Meanwhile, the noise is prefiltered and hence non-white in general. The ideal uniform sampler that follows creates an aliased version of the prefiltered signal in the frequency domain, as reflected in the following capacity expression.

Theorem 1. *Consider the system shown Fig. 2, where $\eta(t)$ is Gaussian noise with power spectral density $\mathcal{S}_\eta(f)$. Assume that $h(t)$, $s(t)$ are both continuous, bounded and absolutely Riemann integrable. Define*

$$\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2 \triangleq \sum_{l=-\infty}^{\infty} |S(f-lf_s)|^2 \mathcal{S}_\eta(f-lf_s),$$

$$\|\mathbf{V}_{HS}(f,f_s)\|_2^2 \triangleq \sum_{l=-\infty}^{\infty} |H(f-lf_s)S(f-lf_s)|^2.$$

Suppose that there exists some constant ϵ_s such that

$$\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2 \geq \epsilon_s > 0 \tag{3}$$

holds. Additionally, suppose that $h_\eta(t) := \mathcal{F}^{-1}\left(\frac{H(f)}{\sqrt{\mathcal{S}_\eta(f)}}\right)$ satisfies $h_\eta(t) = o\left(t^{-\epsilon}\right)$ for some constant $\epsilon > 1$ [2]. The capacity $C(f_s)$ of the sampled channel with a power constraint P is then given parametrically as

$$C(f_s) = \frac{1}{2}\int_{f\in\mathcal{F}(\nu)} \log\left(\nu \frac{\|\mathbf{V}_{HS}(f,f_s)\|_2^2}{\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2}\right) \mathrm{d}f \tag{4}$$

where

$$\mathcal{F}(\nu) = \left\{ f : \frac{\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2}{\|\mathbf{V}_{HS}(f,f_s)\|_2^2} \leq \nu \text{ and } |f| \leq \frac{f_s}{2} \right\} \tag{5}$$

and ν satisfies

$$\int_{f\in\mathcal{F}(\nu)} \left[\nu - \frac{\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2}{\|\mathbf{V}_{HS}(f,f_s)\|_2^2}\right] \mathrm{d}f = P. \tag{6}$$

[2]This condition is used in the theorem proof as a sufficient condition to guarantee asymptotic properties of Toeplitz matrices. A similar condition will be used in Theorem 2.

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 (4) 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 in that the SNR in the sampled scenario is $\gamma_s(f) := \|\mathbf{V}_{HS}(f,f_s)\|_2^2 / \left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2$ in contrast to $\gamma_0(f) := |H(f)|^2/\mathcal{S}_\eta(f)$ for the non-sampled scenario. Water filling over the inverse sampled SNR $\gamma_s^{-1}(f)$ determines the optimal power allocation.

C. Approximate Analysis

Rather than providing here a rigorous proof of Theorem 1, 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 **??**, 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[3] $X(f)$. The Fourier transform of the sampled signal is given as follows: for all $f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$

$$\frac{1}{T_s}\sum_{k=-\infty}^{+\infty} H(f-kf_s)S(f-kf_s)X(f-kf_s) \tag{7}$$

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. 3(a).

By assumption, the noise is of spectral density $\mathcal{S}_\eta(f)$, and hence the prefiltered noise has power spectral density $\mathcal{S}_\eta(f)|S(f)|^2$. The power spectral density of the sampled noise sequence at $f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$ is then given by $\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2 = \sum_{l=-\infty}^{\infty} \mathcal{S}_\eta(f-lf_s)\,|S(f-lf_s)|^2$. 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 [18] 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 $f \in \left[-\frac{f_s}{2}, \frac{f_s}{2}\right]$. This

[3]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.

signal is multiplied by a constant gain $c\alpha_l (l \in \mathbb{Z})$, and sent through the lth input branch, i.e.

$$X(f - lf_s) = c\alpha_l G(f), \quad \forall l \in \mathbb{Z}, \tag{8}$$

where $\alpha_l = \frac{H^*(f-lf_s)S^*(f-lf_s)}{\|\mathbf{V}_{HS}(f,f_s)\|_2}$ 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 [18]) at each branch

$$\frac{c^2 \|\mathbf{V}_{HS}(f,f_s)\|_2^2}{\left\|\mathbf{V}_{S\sqrt{\mathcal{S}_\eta}}(f,f_s)\right\|_2^2}. \tag{9}$$

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 $\tilde{H}(f) = \|\mathbf{V}_{HS}(f,f_s)\|_2$, as illustrated in Fig. 3(b). The water-filling strategy is optimal in allocating power among the set of parallel channels, which yields the parametric equations (5) and (6) and completes our approximate analysis.

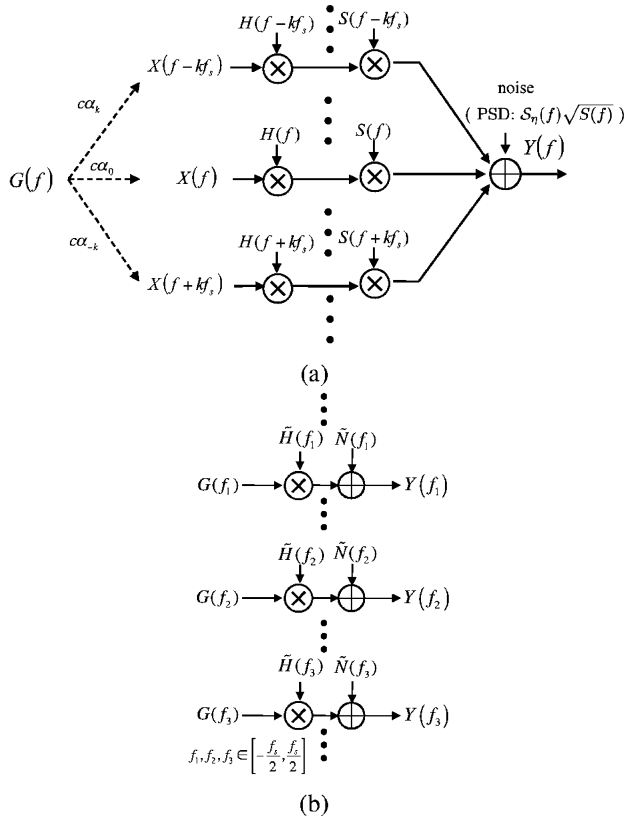

Figure 3. Equivalent representations for filtering followed by sampling: (a) Equivalent MISO Gaussian channel for a given $f \in [-f_s/2, f_s/2]$; (b) The equivalent set of parallel SISO channels representing all $f \in [-f_s/2, f_s/2]$, where the SISO channel at a given frequency is equivalent to the MISO channel in Fig. 3(a).

D. 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 **??**.

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 Δ, e.g. $h(t)$ is transformed into $\{h[n]\}$ by setting

$$h[n] = h(n\Delta).$$

For any given T, this allows us to use a finite-dimensional matrix to approximate the continuous-time block-Toeplitz operator. Then, due the continuity assumption, an exact capacity expression can be obtained by letting Δ go to zero.

2) Spectral properties of block-Toeplitz matrices: After discretization, the input-output relation is similar to a MIMO discrete system. Applying MIMO channel capacity results leads to the capacity for a given T and Δ. The channel capacity is then obtained by taking T to infinity and Δ to zero, which can be related to the channel matrix's spectrum using Toeplitz theory. Since the 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.

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. 4. In fact, this is the noise model adopted by Lapidoth in [19] to approximate white noise, which avoids the infinite variance of the standard model for unfiltered white noise[4]. 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

[4]In fact, the white noise process only exists as a "generalized process" in stochastic calculus, and ideal uniform sampling operating on white noise without prefiltering brings in noise from high-frequency components, which results in a folded noise process with infinite spectral density. In order to avoid this mathematical difficulty, we consider in this example Lapidoth's noise model.

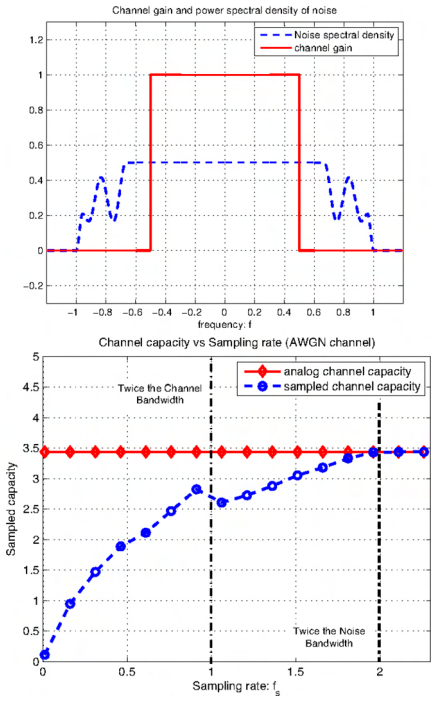

Figure 4. Additive Gaussian noise channel. The channel gain and the power spectral density of the noise is plotted in the left plot. The sampling mechanism employed here is ideal uniform sampling without filtering. The power constraint is $P = 5$. The sampled capacity, as illustrated in the right plot, 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.

allow analog capacity to be achieved when sampling at a rate equal to twice the channel bandwidth. This is because uniform sampling without prefiltering brings in noise from high-frequency components outside the channel bandwidth. Increasing the sampling rate above twice the channel bandwidth (but below the noise bandwidth) spreads the total noise power over a larger sampling bandwidth, reducing the noise density at each frequency, which allows the sampled capacity to continue increasing at sampling rates above the Nyquist rates, as illustrated in Fig. 4. It can be seen that the capacity does not increase monotonically with the sampling rate, which is a consequence of the non-monotonicity of the SNR in f_s, as described in more detail later. We also note that capacity does not 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) Optimal Prefiltering: In the full-length version of this paper [20], we have identified the optimal prefilters that maximize capacity. Although they are in general discontinuous and hence hard to realize, they reduce to low-pass filters with cutoff frequency $f_s/2$ for monotone channels (namely, channels for which the magnitude of the channel response $|H(f)|$ is non-increasing). Fig. 5 illustrates the capacity-sampling tradeoff curve for the raised-cosine channel[5] with different roll-off factors. 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, which reverts to the classical capacity results for non-sampled waveform channels.

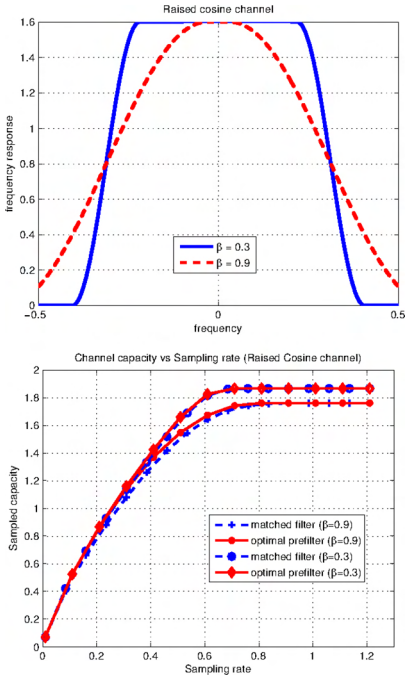

Figure 5. The sampled channel capacity vs sampling rate for a raised-cosine channel with an optimal prefilter. The channel bandwidth is assumed to be $[-\frac{1}{2}, \frac{1}{2}]$, the power constraint $P = 10$, and the noise is white with flat spectral density $\sigma_\eta^2 = 1$. The frequency response $H(f)$ 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.

However, a somewhat counter-intuitive fact arises for more general types of channels: the channel capacity $C(f_s)$ is not necessarily an increasing function of the sampling rate f_s. Consider the multiband channel in Fig. 6 with 4 subbands. If the channel is sampled at a rate $f_s = \frac{3}{5} f_{\text{NYQ}}$, aliasing occurs and collapses the two subbands down to one subband (and hence one degree of freedom). However, if sampling is done at a rate $f_s = \frac{2}{5} f_{\text{NYQ}}$, it can be easily verified that the two subbands shift in frequency but remain non-overlapping, resulting in two degrees of freedom. The tradeoff curve between capacity and sampling rate for this multiband

[5]Specifically, the frequency response of the raised cosine channel here is given as

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

where β denotes the roll-off factor.

channel with optimal prefilters is also plotted in Fig. 6. The underlying reason for the non-monotonic behavior of the capacity curve is that uniform sampling with a single filter largely constrains our ability to exploit channel and signal structures.

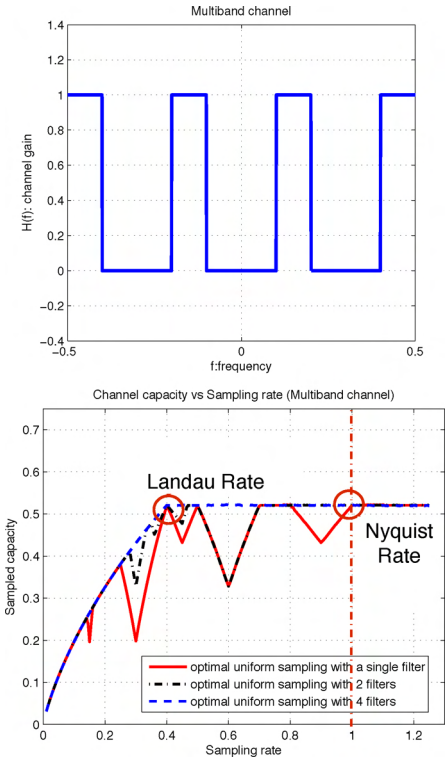

Figure 6. The sampled channel capacity vs sampling rate for (1) uniform sampling with a single filter, (2) uniform sampling with 2 filters, and (3) uniform sampling with 4 filters. The channel bandwidth is assumed to be $[-\frac{1}{2}, \frac{1}{2}]$, the power constraint $P = 10$, and the noise power $\sigma_\eta^2 = 1$. For uniform sampling with both a single filter and 2 filters, the capacity is not monotonically increasing in the sub-Nyquist regime, while sampling above the Nyquist rate does not increase data rate. For the case with 4 filters, it achieves full capacity when exceeding the Landau rate. It can be observed that applying sampling with filter banks outperforms sampling with a single filter, but whether oversampling above the Landau rate achieves full non-sampled capacity depends on the number of filters. In general, when sampling is done below the Landau rate, capacity does not in general monotonically increase with the sampling rate for either a single filter or a bank of filters.

III. A Bank of Filters Followed by Sampling

A. Problem Formulation

As discussed in the previous section, 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 rates 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. 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.

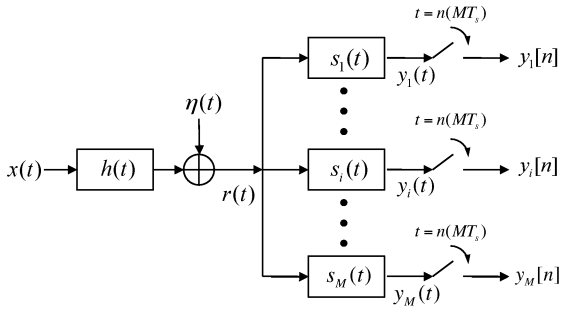

Figure 7. A filter bank followed by sampling: the received analog signal $r(t)$ is passed through M branches. In the ith branch, the signal $r(t)$ is passed through an LTI prefilter with frequency response $S_i(f)$, and then sampled uniformly by an ideal uniform sampler.

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. 7. We denote by $s_i(t)$ and $S_i(f)$ the impulse response and frequency response of the ith linear filter, respectively. The filtered analog output in the ith branch prior to sampling is then given as

$$y_i(t) = (h(t) * s_i(t)) * x(t) + s_i(t) * \eta(t), \quad t \in (0, T]. \tag{10}$$

These filtered signals are then passed through M ideal samplers to yield

$$y_i[n] \triangleq y_i(nMT_s) \quad \text{and} \quad \mathbf{y}[n] \triangleq [y_1[n], y_2[n], \cdots, y_M[n]], \tag{11}$$

where $T_s = f_s^{-1}$. The capacity now is the maximum mutual information between $x(0, T]$ and $\mathbf{y}[n]$.

B. Capacity Results

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 2.

In order to state our theorem formally, we introduce two Fourier symbol matrices $\mathbf{F}_s$ and $\mathbf{F}_h$. Here, $\mathbf{F}_s$ is an infinite matrix of m rows and infinite columns and $\mathbf{F}_h$ is a diagonal infinite matrix such that $\forall 1 \le i \le k, \forall l \in \mathbb{Z}$:

$$\begin{cases} (\mathbf{F}_s(f))_{i,l} &= S_i\left(f - \frac{lf_s}{M}\right)\sqrt{\mathcal{S}_\eta\left(f - \frac{lf_s}{M}\right)}, \\ (\mathbf{F}_h(f))_{l,l} &= \frac{H\left(f - \frac{lf_s}{M}\right)}{\sqrt{\mathcal{S}_\eta\left(f - \frac{lf_s}{M}\right)}}. \end{cases} \tag{12}$$

Theorem 2. *Consider the system shown in Fig. 7. Assume that $h(t)$ and $s_i(t)$ $(1 \le i \le M)$ are all continuous, bounded and absolutely Riemann integrable. Additionally, assume that $h_\eta(t) := \mathcal{F}^{-1}\left(\frac{H(f)}{\sqrt{\mathcal{S}_\eta(f)}}\right)$ satisfies $h_\eta(t) = o\left(t^{-\epsilon}\right)$ for some constant $\epsilon > 1$, and that $\mathbf{F}_s$ is right-invertible for every f. The*

capacity $C(f_s)$ of the sampled channel with a power constraint P can be given as

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

where

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

Remark 1. Using the same argument as used by Telatar in [21], we can express this capacity alternatively as

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

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

$$\mathcal{Q} = \left\{ \left\{ \mathbf{Q}(f) : |f| \le \frac{f_s}{2M}, \mathbf{Q}(f) \in \mathbb{S}_+ \right\} : \int_{-\frac{f_s}{2M}}^{\frac{f_s}{2M}} \mathrm{Tr}\left(\mathbf{Q}(f)\right) \mathrm{d}f = P. \right\} \tag{14}$$

The optimal $\{\mathbf{Q}(f)\}$ corresponds to a water-filling power allocation strategy based on the singular values of the equivalent channel matrix $\tilde{\mathbf{F}}_s \mathbf{F}_h$, where $\mathbf{F}_h$ is associated with the original channel and $\tilde{\mathbf{F}}_s$ arises from prefiltering and noise whitening. For each $f \in [-f_s/2M, f_s/2M]$, the integrand in (13) can be interpreted as a MIMO Gaussian channel capacity formula with degrees of freedom associated with the frequency domain, as illustrated in Fig. 8(a). We still have full control over a countable number of input branches $\left\{ X\left(f - \frac{lf_s}{M}\right) \mid l \in \mathbb{Z} \right\}$, 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 $\{y_i[n] : 1 \le i \le M\}$ result from the same process $\eta(t)$ (as shown in Fig. 7(a)) and hence noise samples are correlated.

C. Approximate Analysis

The sampled analog channel under filter banks followed by sampling can be studied through its connection with MIMO Gaussian channels (see Fig. 8). 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 sampler, the equivalent channel has M receive branches, each corresponding to one branch of prefiltered sampling at a rate f_s/M. The noise received in the ith branch is zero-mean Gaussian with power spectral density

$$\sum_{l=-\infty}^{\infty} \left| S_i \left(f - \frac{lf_s}{M} \right) \right|^2 \mathcal{S}_\eta \left(f - \frac{lf_s}{M} \right) \tag{15}$$

for all $|f| \le f_s/2M$, indicating the mutual correlation of noise at different branches. The received noise vector can be whitened by multiplying $\mathbf{Y}(f) = [\cdots, Y(f), Y(f - f_s), \cdots]^T$ by an $M \times M$ whitening matrix $(\mathbf{F}_s(f)\mathbf{F}_s^*(f))^{-\frac{1}{2}}$. Since whitening here is an invertible operation, it preserves capacity. After whitening, the channel of Fig. 8(a) associated with frequency f has the following channel matrix

$$(\mathbf{F}_s(f)\mathbf{F}_s^*(f))^{-\frac{1}{2}} \mathbf{F}_s(f)\mathbf{F}_h(f) = \tilde{\mathbf{F}}_s(f)\mathbf{F}_h(f). \tag{16}$$

MIMO Gaussian channel capacity results [21] immediately imply that the channel capacity at a given frequency $f \in [-f_s/2M, f_s/2M]$ corresponding to the channel in Fig. 8(a) can be expressed as

$$\max_{\mathbf{Q}} \frac{1}{2} \log \det \left[\mathbf{I} + \tilde{\mathbf{F}}_s(f)\mathbf{F}_h(f)\mathbf{Q}(f)\mathbf{F}_h^*(f)\tilde{\mathbf{F}}_s^*(f) \right] \tag{17}$$

subject to the constraints $\mathrm{trace}\,(\mathbf{Q}(f)) \le P(f)$ and $\mathbf{Q}(f) \in \mathbb{S}_+$, where $\mathbf{Q}(f)$ 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. 8(b), where each MIMO channel in this figure is equivalent to the set of M parallel channels in Fig. 8(a) for the given frequency. Performing the water-filling power allocation strategy across all parallel channels leads to our capacity expression.

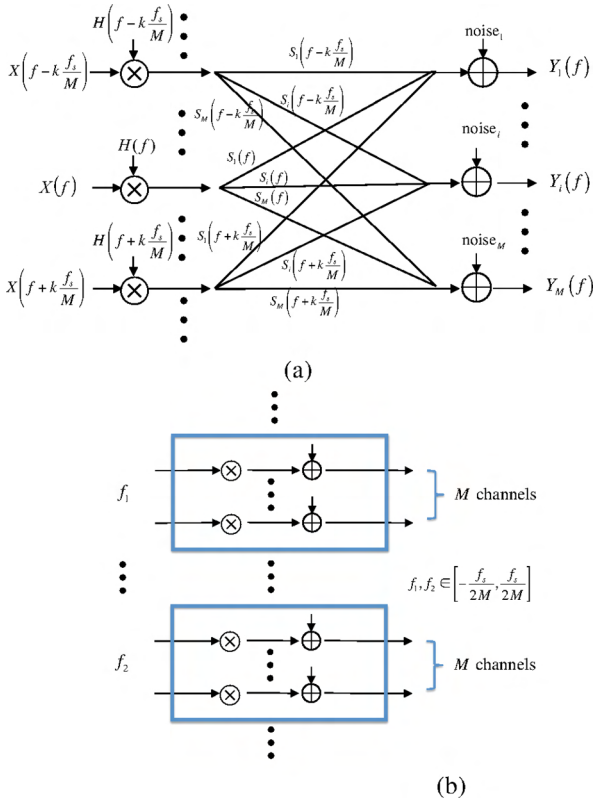

Figure 8. Equivalent representations for a bank of M filters followed by sampling: (a) Equivalent MIMO Gaussian channel for a single $f \in [-f_s/2M, f_s/2M]$; (b) An equivalent set of parallel channels representing all $f \in [-f_s/2M, f_s/2M]$, where the MIMO channel at a given frequency f is equivalent to the MIMO channel of Fig. 8(a).

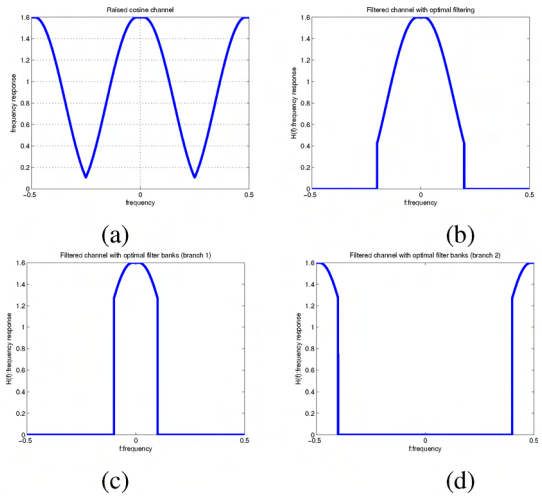

Figure 9. The frequency response of the original and the filtered channel. (a) Original channel model; (b) filtered channel using optimal filtering; (c)(d) filtered channel using optimal filter banks (2 branches).

D. Numerical Examples

We again look at the multiband sparse channel where the channel response is concentrated in two subbands, as illustrated in Fig. 6. As derived in [20], the optimal filter bank will select m frequencies with the highest SNR out of the set $\left\{f - l\frac{f_s}{m} : l \in \mathbb{Z}\right\}$. By comparison, uniform sampling with filtering only allows us to select the best single frequency with the highest SNR out of the set $\{f - lf_s \mid l \in \mathbb{Z}\}$. For example, when several frequencies in the set $\{f - lf_s\}$ enjoy higher channel gain than those in the set $\left\{f + \frac{f_s}{m} - lf_s\right\}$, sampling with filter banks allows these desirable frequencies to be used for multiple branches. For uniform sampling with filtering, however, at most one from each set can be effectively used. See Fig. 9 for an example, in which sampling with 2 filters selects better frequency components than sampling with a single filter. Because of the better exploitation of spectral contents, sampling with filter banks outperforms sampling with a single filter, as illustrated in Fig. 6. It can be observed that when the number of branches is carefully chosen based on the channel structure, sampling with filter banks can achieve full capacity when the sampling rate exceeds the Landau rate, but uniform sampling with filter banks does not necessarily result in mononicity of capacity in f_s when it is performed below the Landau rate.

IV. Concluding Remarks

This paper characterizes the interplay between sampling and channel capacity. In particular, we characterize capacity as a function of sampling rate along with the capacity-achieving input strategy for different sampling mechanisms. We show that the capacity of a sampled channel degrades with reduced sampling rate, which encourages exploitation of channel structures in the sampling mechanism. Our work also points to the need for more general sampling mechanisms based on capacity as an optimization metric. It remains to be seen how capacity of sampled analog channels can be improved by more general non-uniform sampling. A deeper understanding of sufficient statistics that favors the channel structure may also guide the design of sampling mechanisms for both single-user and multiuser channels.

References

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

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

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

[10] 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.

[11] T. Ericson, "Optimum PAM filters are always band limited (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 570 – 573, Jul 1973.

[12] 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.

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

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

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

[16] 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.

[17] 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.

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

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

[20] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: capacity limits of sampled analog channels," *submitted to IEEE Transactions on Information Theory*, September 2011. [Online]. Available: http://arxiv.org/abs/1109.5415

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