Capacity of the Gaussian Erasure Channel

A. M. Tulino, S. Verdú, G. Caire, S. Shamai

July 11, 2007

Abstract

This paper finds the capacity of linear time-invariant additive Gaussian noise channels observed through a memoryless erasure channel. This problem requires obtaining the asymptotic spectral distribution of a submatrix of a nonnegative definite Toeplitz matrix obtained by retaining each column/row independently and with identical probability. We show that the optimum normalized power spectral density is the waterfilling solution for reduced signal-to-noise ratio.

Index Terms: Channel capacity, Gaussian channels, intersymbol interference, erasure channels, fading, random matrices, Toeplitz matrices.

1 Introduction

The erasure channel plays an important role in information theory and coding theory. It is a very useful idealization of situations where the symbols observed by the receiver have either very high or very low reliability. Applications of erasure channels range from communication subject to jamming to packet-switched store-and-forward networks, from magnetic recording to wireless communications subject to fading, from powerline communications subject to impulsive noise to frequency-hopped multiaccess channels.

The capacity of the memoryless binary erasure channel is equal to $1 - \mathsf{e}$ bits, where e is the erasure probability. In fact, regardless of the statistics of the erasures, the capacity of the erasure channel is equal to the liminf of the proportion of non-erased symbols [23] times the logarithm of the cardinality of the alphabet.

For many applications, discrete erasure channels where symbols are either received without error or erased are rather coarse idealizations. Within the paradigm of discrete memoryless noisy channels, it is straightforward to find the capacity of channels that incorporate errors as well as erasures. Even in the presence of memory in the erasures it has been the capacity of the concatenation of a discrete memoryless channel with capacity C and an erasure channel (possibly with memory) with erasure rate e is equal to $(1 - \mathsf{e})C$ [5, 24]. Also straightforward is to deal with power-constrained memoryless Gaussian channels observed through memoryless erasure channels: the capacity is also equal to the capacity of the memoryless Gaussian channel

*This work was partially supported by NSF Grants NCR-0074277 and CCR-0312879.

†Antonia M. Tulino is with the Department of Electrical Eng., University of Napoli, "Federico II", Napoli, Italy, 80125. Email: atulino@princeton.EDU. Sergio Verdú is with the Department of Electrical Eng., Princeton University, Princeton, NJ 08544, USA. Email: verdu@Princeton.EDU. Giuseppe Caire is with the Department of Electrical Eng., Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA. E-mail: caire@usc.edu. Shlomo Shamai (Shitz) is with the Department of Electrical Eng., Technion Israel Institute of Technology, 32000 Haifa, Israel. E-mail: sshlomo@ee. technion.ac.il.

times the proportion of non-erased symbols [10, 11]. In that case the receiver obtains either the noisy output symbol or an erasure. Equivalently, we can view this channel as one with on-off fading where the receiver (but not the transmitter) knows the $\{0, 1\}$-valued fading coefficients. If the Gaussian channel has memory we are led to the setup depicted in Figure 1.

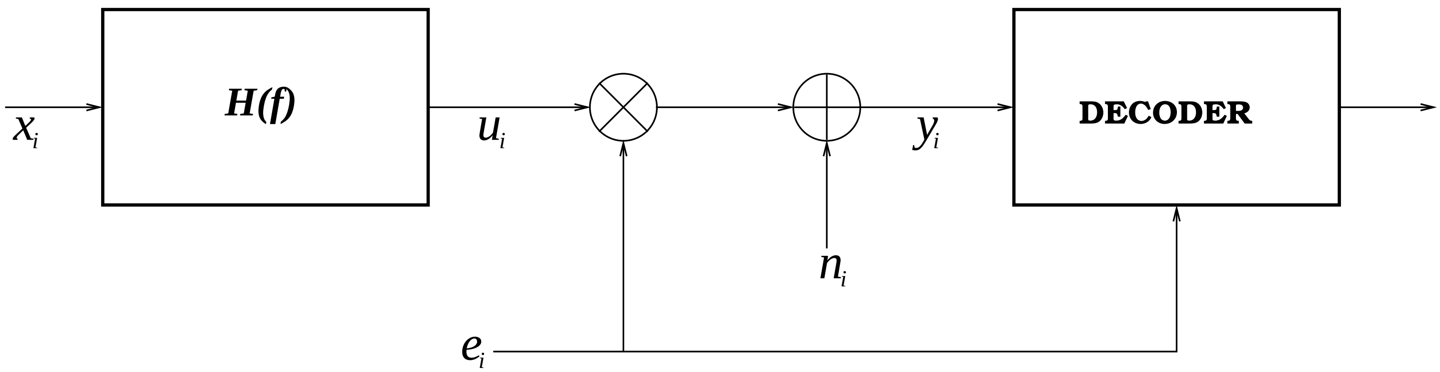

Figure 1: Linear Gaussian Erasure Channel.

This model is relevant for example in the case of a Gaussian channel with impulsive noise, typical of powerline communications [14]. The receiver automatic gain control applies a renormalization of the incoming signal, such that in the presence of a noise impulse (perfectly detected by the receiver) the corresponding symbol is completely erased. In this way, the on-off fading is a simplified but realistic model for impulse noise where the power of the impulses is much larger than the average received power. The Gaussian erasure model can also be used to assess the throughput in uplink cellular systems subject to topological randomness where intercell interference follows the classical Wyner model [26] and cell sites are either "on" or "off" in an independent and identically distributed manner. The setup is also relevant to multiaccess channels whose outputs are switched to one of the users at a time. In the fountain capacity setup introduced in [16], reliability has to be maintained for any individual sequence of erasures not just memoryless erasures. In that case, not much is known when the channel has memory.

The rest of the article has the following structure. Section 2 sets up the channel and the key random matrix problem associated with the analysis of capacity. Section 3 provides, in addition to the necessary background on random matrix theory, a new general random matrix result which is central in our development and is of independent interest. Section 4 solves the central problem of finding the asymptotic eigenvalue distribution of a submatrix of a nonnegative definite Toeplitz (or circulant) matrix obtained by retaining each column/row independently and with identical probability. The mutual information achieved in the Gaussian-erasure channel is obtained as the integral on $[0, 1-\mathsf{e}]$ of a certain function that does not depend on e, thus providing a pleasing interpretation of the effect of erasures along with a generalization of the elementary memoryless capacity $(1-\mathsf{e})C$. Section 5 finds an explicit form for the optimal (capacity achieving) input spectrum as a function of the transfer function of the channel and the erasure probability. Section 6 presents some upper and lower bounds on the Gaussian-erasure input/output mutual information that are considerably easier to compute. Section 7 deals with various asymptotic regimes for low/high signal-to-noise ratio and low/high erasure rate.

In order to maintain the presentation flow, some particularly technical proofs and auxiliary results have been relegated to the appendices. Those proofs that yield particular insight into the main results have been included in the body of the paper.

2 Problem Setup

In this paper we analyze the discrete-time channel with memory where the input codeword $\mathbf{x} = (x_1, \dots x_n) \in \mathbb{R}^n$ is subject to an average power constraint and goes through a linear time-invariant discrete-time linear system with transfer function

$$H(f), -\frac{1}{2} \leq f \leq \frac{1}{2};$$

the output of the linear system $(u_1, \dots u_n) \in \mathbb{R}^n$ is contaminated by independent identically distributed Gaussian noise $(n_1, \dots n_n) \in \mathbb{R}^n$; finally, a process of erasures $\mathbf{e} = (e_1, \dots e_n) \in \{0,1\}^n$, known to the receiver, controls which noisy outputs are available to the receiver. An equivalent version of the channel, depicted in Figure 1, is

$$y_i = \sqrt{\gamma}\, e_i u_i + n_i, \quad i = 1, \dots, n \tag{1}$$

$$u_i = \sum_{\ell=0}^{i-1} h[\ell] x_{i-\ell} \tag{2}$$

$$h[i] = \int_{-1/2}^{1/2} H(f) e^{\mathrm{j}2\pi f i}\, \mathrm{d}f \tag{3}$$

where $\{n_i\}$ are independent Gaussian with unit variance; $\{e_i\}$ are independent binary with

$$P[e_i = 0] = \mathsf{e} \tag{4}$$

where e is the erasure rate, and the codewords are restricted to satisfy

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq 1 \tag{5}$$

Because the noise is memoryless, (1) is equivalent to

$$y_i = e_i \left[\sqrt{\gamma}\, u_i + n_i \right], \quad i = 1, \dots, n \tag{6}$$

Since the receiver knows the location of the erasures, the capacity is equal to

$$C_{\mathsf{e}}(\gamma) = \lim_{n \to \infty} \frac{1}{n} \max_{\mathbf{x}} I(\mathbf{x}; \mathbf{y} | \mathbf{e}) \tag{7}$$

$$= \lim_{n \to \infty} \frac{1}{2n} \max_{\mathbf{\Sigma}} \mathbb{E}\left[\log\det\left(\mathbf{I} + \gamma \mathbf{E} \mathbf{\Sigma} \mathbf{E} \right) \right] \tag{8}$$

where we have denoted the random matrix

$$\mathbf{E} = \mathrm{diag}\{e_1, \dots e_n\} \tag{9}$$

and (8) follows from the optimality of Gaussian inputs since conditioned on $\mathbf{e}$, the channel is Gaussian; the maximum in (8) is over all matrices that can be written as

$$\mathbf{\Sigma} = \mathbf{H} \mathbf{\Sigma}_x \mathbf{H}^\dagger \tag{10}$$

with $\mathbf{H}$ denoting the Toeplitz channel matrix whose (i, j) entry is $h[i-j]$ and

$$\text{tr}\{\mathbf{\Sigma}_x\} = n. \tag{11}$$

Since there is no loss of optimality in restricting the input to be stationary (with power spectral density $S_x(f)$), we can restrict attention to circulant matrices $\mathbf{\Sigma}_x$ (Appendix B). The power spectral density of the signal at the output of the linear system is denoted by

$$S(f) = S_x(f)|H(f)|^2. \tag{12}$$

For brevity we denote the mutual information achieved with input power spectral density $S_x(f)$ by [1]

$$I_{\mathsf{e}}(\gamma) = \lim_{n\to\infty} \frac{1}{2n} \mathbb{E}\left[\log\det\left(\mathbf{I} + \gamma \mathbf{E}\mathbf{\Sigma}\mathbf{E}\right)\right] \tag{13}$$

In the absence of erasures ($\mathsf{e} = 0$), we obtain the familiar

$$I_0(\gamma) = \frac{1}{2}\int_{-1/2}^{1/2} \log(1 + \gamma S(f))\, df \tag{14}$$

which when maximized with respect to the input power spectral density yields (e.g. [3])

$$C_0(\gamma) = \frac{1}{2}\int_{-1/2}^{1/2} \log\left(1 + \gamma S_x^*(f)|H(f)|^2\right)\, \mathrm{d}f \tag{15}$$

where $S_x^*(f)$ is the waterfilling input power spectral density:

$$S_x^*(f) = \left[\zeta - \frac{1}{\gamma|H(f)|^2}\right]^+ \tag{16}$$

and the water level $1 < \zeta < \infty$ is chosen so that

$$\int_{-1/2}^{1/2} S_x^*(f)\,\mathrm{d}f = \int_{-1/2}^{1/2} \left[\zeta - \frac{1}{\gamma|H(f)|^2}\right]^+ \mathrm{d}f \tag{17}$$

$$= 1 \tag{18}$$

It is useful to realize that the normalization of the input power implies the following behavior for the water level as a function of γ:

$$\lim_{\gamma\to 0} \zeta_\gamma = \infty \tag{19}$$

$$\lim_{\gamma\to\infty} \zeta_\gamma = \frac{1}{1 - \mu\left(\{f : |H(f)| = 0\}\right)} \tag{20}$$

where μ stands for Lebesgue measure on $[-1/2, 1/2]$.

The solution (15), known since 1949 [18] can be rigorously justified by means of the Grenander-Szëgo theorem on the distribution of the eigenvalues of large deterministic Toeplitz matrices [6].

In view of (8), obtaining the capacity of the linear Gaussian erasure channel involves analyzing the asymptotic distribution of the eigenvalues of the random matrix $\mathbf{E}\mathbf{\Sigma}\mathbf{E}$ where $\mathbf{\Sigma} = \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^\dagger$ is asymptotically circulant, and $\mathbf{E}$ is a random 0-1 diagonal matrix. We will not be able to invoke an existing result to find the required asymptotic distribution. Such a result, at the intersection of the asymptotic eigenvalue distribution of Toeplitz matrices (e.g. [6]) and of random matrices (e.g. [20]), is the main contribution of this paper.

The case of colored Gaussian noise falls outside the scope of this paper. We note that (1) and (6) are no longer equivalent channel models in that more general setup.

[1] Because of the stationarity and ergodicity of the iid erasures, the expectation in (13) is unnecessary.

3 Random Matrix Theory

In addition to a new result (Theorem 1), in this section we collect a number of definitions and results on random matrix theory from [20] that will be useful in the sequel.

3.1 Eigenvalue Distributions

Given an $n \times n$ Hermitian matrix $\mathbf{A}$, the empirical cdf of the eigenvalues (also referred to as the empirical spectral distribution) of $\mathbf{A}$ is defined as

$$\mathsf{F}^n_{\mathbf{A}}(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{\lambda_i(\mathbf{A}) \le x\} \tag{21}$$

where $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$ and $1\{\cdot\}$ is the indicator function. If $\mathsf{F}^n_{\mathbf{A}}(\cdot)$ converges almost surely as $n \to \infty$, then the corresponding limit (asymptotic empirical spectral distribution) is denoted by $\mathsf{F}_{\mathbf{A}}(\cdot)$.[2]

Example 1 *Let $\mathbf{E}$ be an $n \times n$ diagonal matrix whose diagonal entries $e_1, \ldots e_n$ are independent and*

$$P[e_i = 0] = \mathsf{e} \tag{22}$$

$$P[e_i = 1] = 1 - \mathsf{e} \tag{23}$$

The asymptotic empirical spectral distribution of $\mathbf{E}$ is equal to:

$$\mathsf{F}_{\mathbf{E}}(x) = \begin{cases} \mathsf{e} & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases} \tag{24}$$

Example 2 *Let $\mathbf{\Sigma}$ be the $n \times n$ nonnegative definite Toeplitz matrix*

$$\Sigma_{i,j} = \sigma_{|i-j|} \tag{25}$$

for an absolutely summable sequence $\sigma_0, \sigma_1, \ldots$. Using the Grenander-Szëgo theorem, it can be shown [6] that the limiting eigenvalue distribution of $\mathbf{\Sigma}$ is given by

$$\mathsf{F}_{\mathbf{\Sigma}}(x) = \int_{-1/2}^{1/2} 1\{S(f) \le x\} \, \mathrm{d}f \tag{26}$$

where

$$S(f) = \sum_{k=0}^{\infty} \sigma_k e^{-\mathrm{j}2\pi f k} \tag{27}$$

3.2 Transforms

For our purposes, it is advantageous to make use of the η-transform and the Shannon transform, whose introduction was motivated by the application of random matrix theory to various problems in the information theory of noisy communication channels [20]. These transforms, intimately related to each other and to the Stieltjes transform traditionally used in random matrix theory [20], characterize the spectrum of a random matrix while carrying certain engineering intuition, as explained in [20].

[2] A slight abuse of notation, customary in asymptotic random matrix theory, is to avoid a dimension subscript in $\mathbf{A}$, which depending on the context stands for a matrix of a given dimension or for a sequence of matrices.

Definition 1 *Given a nonnegative definite random matrix* $\mathbf{A}$*, its* η*-transform is*

$$\eta_{\mathbf{A}}(\gamma) = \mathbb{E}\left[\frac{1}{1+\gamma X}\right] \tag{28}$$

where X *is a nonnegative random variable whose distribution is the asymptotic empirical spectral distribution of* $\mathbf{A}$ *while* γ *is a nonnegative real number.*

Note that

$$P[X=0] < \eta_X(\gamma) \leq 1 \tag{29}$$

The lower bound is asymptotically approached as $\gamma \to \infty$ and is equal to the fraction of zero eigenvalues of $\mathbf{A}$.

Then, $\mathbb{E}[X^k]$ is the kth asymptotic moment of $\mathbf{A}$, i.e., $\lim_{n\to\infty} \frac{1}{n}\mathrm{tr}\{\mathbf{A}^k\}$. Expressing the rational function in (24) in a series expansion, $\eta_{\mathbf{A}}(\gamma)$ can be regarded as a generating function for the asymptotic moments of $\mathbf{A}$.

Example 3 *Let* $\mathbf{E}$ *be the sequence of random matrices defined as in Example 1. The* η*-transform of* $\mathbf{E}$ *is given by:*

$$\eta_{\mathbf{E}}(\gamma) = \frac{1+\mathsf{e}\gamma}{1+\gamma} \tag{30}$$

Example 4 *Let* $\mathbf{\Sigma}$ *be the sequence of deterministic matrices defined as in Example 2. The* η*-transform of* $\mathbf{\Sigma}$ *is:*

$$\eta_{\mathbf{\Sigma}}(\gamma) = \int_{-1/2}^{1/2} \frac{1}{1+\gamma S(f)}\,\mathrm{d}f \tag{31}$$

Note that

$$\lim_{\gamma\to\infty} \eta_{\mathbf{\Sigma}}(\gamma) = \mu\left(\{f : S(f) = 0\}\right) \tag{32}$$

Analogously,

$$\eta_{\mathbf{H\Sigma_x H}}(\gamma) = \int_{-1/2}^{1/2} \frac{1}{1+\gamma S_x(f)|H(f)|^2}\,\mathrm{d}f \tag{33}$$

For future reference, it is easy to check that for the (maximal mutual information) water-filling power spectral density in (16) the η-transform is directly related to the water level:

$$1 - \eta_{\mathbf{H\Sigma^*_x H^\dagger}}(\gamma) = \frac{1}{\zeta} \tag{34}$$

Definition 2 *Given a nonnegative definite random matrix* $\mathbf{A}$*, its Shannon transform is defined as*

$$\mathcal{V}_{\mathbf{A}}(\gamma) = \mathbb{E}[\log(1+\gamma X)] \tag{35}$$

where X *is a nonnegative random variable whose distribution is the asymptotic empirical spectral distribution of* $\mathbf{A}$ *while* γ *is a nonnegative real number.*

Note that according to (13),

$$I_{\mathsf{e}}(\gamma) = \frac{1}{2}\mathcal{V}_{\mathbf{E\Sigma E}}(\gamma) \tag{36}$$

Assuming that the logarithm in (35) is natural, the η and Shannon transforms are related through

$$\frac{d}{d\gamma}\mathcal{V}_{\mathbf{A}}(\gamma) = \frac{1-\eta_{\mathbf{A}}(\gamma)}{\gamma} \tag{37}$$

We proceed to give a new connection between the η-transform and the Shannon transform which will be useful in our derivation of the capacity of the Gaussian-erasure channel.

Theorem 1 *Let* $\mathbf{A}$ *be a nonnegative definite random matrix. Let* $\rho = \lim_{n\to\infty} rank(\mathbf{A})/n$. *The Shannon transform and* η *transforms are related through*

$$\mathcal{V}_{\mathbf{A}}(\gamma) \;=\; \rho\int_0^1 \log\left(1+\daleth(y,\gamma)\right)\, dy \tag{38}$$

where $\daleth$ *is defined by the fixed-point equation*

$$\rho y\frac{\daleth(y,\gamma)}{1+\daleth(y,\gamma)} = 1-\eta_{\mathbf{A}}\left(\frac{\gamma y}{1+(1-y)\daleth(y,\gamma)}\right) \tag{39}$$

Proof: First, we show the special case where $\rho = 1$. For an $n\times n$ matrix $\mathbf{A}$, choose a decomposition $\mathbf{A} = \mathbf{C}\mathbf{C}^\dagger$, and an arbitrary unitary $n\times n$ matrix $\mathbf{U}$. Denoting the ith column of $\mathbf{CU}$ by $\mathbf{b}_i$, we can write

$$\begin{aligned}\log\det\left(\mathbf{I}+\gamma\mathbf{A}\right) &= \log\det\left(\mathbf{I}+\gamma\mathbf{C}\mathbf{C}^\dagger\right)\\ &= \log\det\left(\mathbf{I}+\gamma\mathbf{C}\mathbf{U}\mathbf{U}^\dagger\mathbf{C}^\dagger\right)\\ &= \sum_{i=1}^{n}\log\left(1+\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i\right) \qquad (40)\end{aligned}$$

where (40) follows from the definition of the determinant

$$\det(\mathbf{L})\prod_{i=1}^{n}((\mathbf{L}^{(i)})^{-1})_{ii} = 1 \tag{41}$$

with $\mathbf{L}^{(i)}$ denoting the ith principal minor of $\mathbf{L}$, in the case $\mathbf{L} = \mathbf{I}+\gamma\mathbf{C}\mathbf{U}\mathbf{U}^\dagger\mathbf{C}^\dagger$.

Assume now $\mathbf{U}$ is uniformly distributed on the set of $n\times n$ unitary matrices (i.e. it is a Haar matrix). Lemma 2 (Appendix C), implies that as $n\to\infty$ and $i = \lceil ny\rceil$ for some $y\in[0,1]$, the quadratic form in (40) converges almost surely to a constant which we denote by:

$$\lim_{n\to\infty}\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i = \daleth(y,\gamma) \tag{42}$$

In fact, Lemma 2 does not directly deal with the quadratic form in the left side of (42) since in that quadratic form $\mathbf{b}_i\mathbf{b}_i^\dagger$ is excluded from the sum. But the matrix inversion lemma gives

$$\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i = \frac{\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i}{1+\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i} \tag{43}$$

which, according to Lemma 2 and (42), allows us to conclude that

$$1 - \frac{y \daleth(y,\gamma)}{1+\daleth(y,\gamma)} = \eta \tag{44}$$

$$= \eta_{\mathbf{A}}\left(\gamma \frac{y-1+\eta}{\eta}\right) \tag{45}$$

$$= \eta_{\mathbf{A}}\left(\frac{\gamma y}{1+(1-y)\daleth(y,\gamma)}\right) \tag{46}$$

where (46) follows from (44).

Applying (42) to (40), we obtain that

$$\mathcal{V}_{\mathbf{A}}(\gamma) = \lim_{n\to\infty} \frac{1}{n} \log\det\left(\mathbf{I}+\gamma\mathbf{A}\right) \tag{47}$$

$$= \int_0^1 \log(1+\daleth(y,\gamma))\, \mathrm{d}y \tag{48}$$

$$\tag{49}$$

Together with (44-46), this establishes the desired result for $\rho = 1$. To deal with the general case in which ρ need not be equal to 1, we can express

$$\mathcal{V}_{\mathbf{A}}(\gamma) = \rho\mathcal{V}_{\hat{\mathbf{A}}}(\gamma) \tag{50}$$

$$\eta_{\mathbf{A}}(\gamma) = \rho\eta_{\hat{\mathbf{A}}}(\gamma) + 1 - \rho \tag{51}$$

where $\mathcal{V}_{\hat{\mathbf{A}}}$ and $\eta_{\hat{\mathbf{A}}}$ are the Shannon and η transforms of the asymptotic distribution of the positive eigenvalues of $\mathbf{A}$. Applying the special case of (38) and (39) with $\rho = 1$ to $\mathcal{V}_{\hat{\mathbf{A}}}$ and $\eta_{\hat{\mathbf{A}}}$, we obtain the desired general relationship between $\mathcal{V}_{\mathbf{A}}(\gamma)$ and $\eta_{\mathbf{A}}(\gamma)$ upon substitution of (50) and (51). ■

An operational and intuitive characterization of $\daleth(y,\gamma)$ is obtained by using the chain rule of mutual information and the related successive decoding interpretation. In particular, (40) follows by noticing that $\log\det(\mathbf{I}+\gamma\sum_i \mathbf{b}_i\mathbf{b}_i^\dagger)$ is the mutual information $I(\mathbf{x};\mathbf{y})$ corresponding to the Gaussian linear channel $\mathbf{y} = \sqrt{\gamma}\mathbf{B}\mathbf{x}+\mathbf{n}$, with fixed $\mathbf{B}$ and $\mathbf{x} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$. Applying the mutual information chain rule [3]

$$I(\mathbf{x};\mathbf{y}) = \sum_{i=1}^{n} I(x_i;\mathbf{y}|x_{i+1},\ldots,x_n)$$

$$= \sum_{i=1}^{n} I\left(x_i; \sqrt{\gamma}\mathbf{b}_i x_i + \sqrt{\gamma}\sum_{j=1}^{i-1}\mathbf{b}_j x_j + \mathbf{n}\right)$$

$$= \sum_{i=1}^{n} \log\left(1+\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i\right) \tag{52}$$

For Haar distributed $\mathbf{U}$, independent of $\mathbf{C}$, we have that $\mathbf{B} = \mathbf{CU}$ and $\mathbf{B\Pi}$ are identically distributed for any arbitrary permutation matrix $\mathbf{\Pi}$. Then,

$$\gamma\mathbf{b}_i^\dagger\left(\mathbf{I}+\sum_{j=1}^{i-1}\mathbf{b}_j\mathbf{b}_j^\dagger\right)^{-1}\mathbf{b}_i$$

is the Signal-to-Interference plus Noise Ratio (SINR) at the i-th output of a MMSE decision-feedback receiver that successively detects the symbols $x_n, x_{n-1}, \ldots, x_1$ and perfectly subtracts them from $\mathbf{y}$. In the limit of large n, noticing that the detection order is irrelevant because of the invariance by right-multiplication by permutations, it follows that $\daleth(y,\gamma)$ is the asymptotic SINR of the MMSE decision-feedback receiver when a fraction y of symbols is yet to be removed.

Some properties of the solution to (39) are:

1. $\daleth(y,\gamma)$ is monotonically increasing with γ.

2. $$\daleth(y,0) = 0 \tag{53}$$

3. $$\lim_{\gamma\to\infty} \frac{\daleth(y,\gamma)}{\gamma} = \digamma(y) \tag{54}$$

   which is the solution to the equation

   $$\eta_{\mathbf{A}}\left(\frac{y}{(1-y)\digamma}\right) = 1 - \rho y \tag{55}$$

4. $\daleth(y,\gamma)$ is monotonically decreasing with y (Appendix D).

4 Asymptotic Eigenvalue Distribution of $\mathbf{E\Sigma E}$

In this section we find the η-transform and the Shannon transform of a submatrix of a large nonnegative definite Toeplitz or circulant matrix obtained by retaining each column/row independently and with identical probability. Throughout this section, $\mathbf{E}$ stands for a diagonal matrix of 0-1 coefficients chosen independently with probability of 0 equal to $\mathtt{e}$.

The first step is to show that, as in the conventional case without erasures, the asymptotic eigenvalue distribution is the same as if $\mathbf{\Sigma}$ in (10) were replaced by a circulant matrix. The sufficient condition in the following lemma is satisfied because of the conventional asymptotic equivalence of products of Toeplitz matrices to circulant matrices (see [6, Thm. 5.3] and Appendix A).

Lemma 1 *Denote the diagonal matrix of the eigenvalues of $\mathbf{\Sigma}$ given in (10) by*

$$\mathbf{\Lambda} = diag\{\lambda_1, \ldots \lambda_n\}. \tag{56}$$

Further, denote the unitary discrete Fourier transform (DFT) matrix

$$\mathbf{F} = \frac{1}{\sqrt{n}}\left[e^{-\mathrm{j}\frac{2\pi}{n}(i-1)(p-1)} \middle| \begin{array}{l} i = 1,\ldots,n \\ p = 1,\ldots,n \end{array} \right] \tag{57}$$

and the circulant matrix

$$\mathbf{\Psi} = \mathbf{F\Lambda F}^\dagger \tag{58}$$

For all $\gamma > 0$,

$$\begin{aligned} \eta_{\mathbf{E\Sigma E}}(\gamma) &= \eta_{\mathbf{E\Psi E}}(\gamma) \qquad (59)\\ \mathcal{V}_{\mathbf{E\Sigma E}}(\gamma) &= \mathcal{V}_{\mathbf{E\Psi E}}(\gamma). \qquad (60) \end{aligned}$$

Proof: *Appendix A*

The next theorem yields the desired characterization of the asymptotic eigenvalue distribution of $\mathbf{E\Sigma E}$:

Theorem 2 *The η-transform of $\mathbf{E\Sigma E}$ is*

$$\eta_{\mathbf{E\Sigma E}}(\gamma) = \eta \tag{61}$$

where η is the solution of the fixed-point equation:

$$\eta = \eta_{\mathbf{\Sigma}}\left(\gamma - \gamma\frac{\mathsf{e}}{\eta}\right) \tag{62}$$

Proof: According to Lemma 1, it is equivalent to show that the η-transform of $\mathbf{E\Psi E}$ is equal to

$$\eta_{\mathbf{E\Psi E}}(\gamma) = \eta \tag{63}$$

where η is the solution of the fixed-point equation:

$$\eta = \eta_{\mathbf{\Psi}}\left(\gamma - \gamma\frac{\mathsf{e}}{\eta}\right) \tag{64}$$

Let

$$\mathbf{Q} = \mathbf{EF} \tag{65}$$

and denote by $\mathbf{q}_i$ the ith column of $\mathbf{Q}$, and let

$$\begin{aligned}\mathbf{A}_i &= \mathbf{I} + \gamma\sum_{j\neq i}\lambda_j\mathbf{q}_j\mathbf{q}_j^\dagger &(66)\\ &= \mathbf{I} + \gamma\mathbf{Q\Lambda Q}^\dagger - \gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger &(67)\end{aligned}$$

The matrix inversion lemma [12] states that:

$$\left(\mathbf{I} + \gamma\mathbf{Q\Lambda Q}^\dagger\right)^{-1} = \mathbf{A}_i^{-1} - \frac{\gamma\lambda_i}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}\mathbf{A}_i^{-1}\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1} \tag{68}$$

Multiplying both sides of (68) by $\gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger$ we obtain

$$\begin{aligned}\gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger(\mathbf{I}+\gamma\mathbf{Q\Lambda Q}^\dagger)^{-1} &= \gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1} - \frac{\gamma^2\lambda_i^2}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1} &(69)\\ &= \gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\left(1 - \frac{\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}\right) &(70)\\ &= \frac{\gamma\lambda_i}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1} &(71)\end{aligned}$$

Summing over i we have

$$\begin{aligned}\sum_{i=1}^{n}\gamma\lambda_i\mathbf{q}_i\mathbf{q}_i^\dagger(\mathbf{I}+\gamma\mathbf{Q\Lambda Q}^\dagger)^{-1} &= \gamma\mathbf{Q\Lambda Q}^\dagger(\mathbf{I}+\gamma\mathbf{Q\Lambda Q}^\dagger)^{-1} &(72)\\ &= \sum_{i=1}^{n}\frac{\gamma\lambda_i}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}\mathbf{q}_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1} &(73)\end{aligned}$$

Taking the trace on both sides of (73) and dividing by n we get

$$1-\frac{1}{n}\mathrm{tr}\{(\mathbf{I}+\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger)^{-1}\} = \frac{1}{n}\mathrm{tr}\{\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger(\mathbf{I}+\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger)^{-1}\} \tag{74}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i} \tag{75}$$

$$= 1-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i} \tag{76}$$

Therefore,

$$\eta_{\mathbf{E}\mathbf{\Psi}\mathbf{E}}(\gamma)=\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(\mathbf{I}+\gamma\mathbf{E}\mathbf{\Psi}\mathbf{E})^{-1} = \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(\mathbf{I}+\gamma\mathbf{E}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger\mathbf{E})^{-1} \tag{77}$$

$$= \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(\mathbf{I}+\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger)^{-1} \tag{78}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\gamma\lambda_i\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i} \tag{79}$$

$$= \eta_{\mathbf{\Psi}}\left(\alpha\gamma\right) \tag{80}$$

where (80) follows from Lemma 13 (Appendix H) (which is the most technically challenging result in this paper):

$$\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i \overset{a.s}{\to} \alpha. \tag{81}$$

Thus, the result will follow if we can show that

$$\alpha = 1-\frac{\mathsf{e}}{\eta_{\mathbf{E}\mathbf{\Psi}\mathbf{E}}(\gamma)} \tag{82}$$

To that end, note the following identities

$$\alpha\,\eta_{\mathbf{E}\mathbf{\Psi}\mathbf{E}}(\gamma) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i}{1+\lambda_i\,\gamma\mathbf{q}_i^\dagger\mathbf{A}_i^{-1}\mathbf{q}_i} \tag{83}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbf{q}_i^\dagger(\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger+\mathbf{I})^{-1}\mathbf{q}_i \tag{84}$$

$$= \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}\left(\mathbf{Q}^\dagger(\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger+\mathbf{I})^{-1}\mathbf{Q}\right) \tag{85}$$

$$= \eta_{\mathbf{E}\mathbf{\Psi}\mathbf{E}}(\gamma)-\mathsf{e} \tag{86}$$

where (83) follows from Lemma 13 and (79); (84) follows by taking the quadratic form of the matrix in (68) with the vector $\mathbf{q}_i$. Finally, (86) follows by writing

$$\begin{aligned}
\mathrm{tr}\left(\mathbf{Q}^\dagger(\mathbf{I}+\gamma\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger)^{-1}\mathbf{Q}\right) &= \mathrm{tr}\left(\mathbf{F}^\dagger\mathbf{E}(\mathbf{I}+\gamma\mathbf{E}\mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger\mathbf{E})^{-1}\mathbf{E}\mathbf{F}\right)\\
&= \mathrm{tr}\left(\mathbf{E}(\mathbf{I}+\gamma\mathbf{E}\mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger\mathbf{E})^{-1}\mathbf{E}\right)\\
&= \mathrm{tr}\left(\mathbf{I}+\gamma\mathbf{E}\mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger\mathbf{E}\right)^{-1}-\sum_{i=1}^{n}1\{e_i=0\}\left[(\mathbf{I}+\gamma\mathbf{E}\mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger\mathbf{E})^{-1}\right]_{i,i}\\
&= \mathrm{tr}\left(\mathbf{I}+\gamma\mathbf{E}\mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger\mathbf{E}\right)^{-1}-\sum_{i=1}^{n}1\{e_i=0\} \qquad (87)
\end{aligned}$$

and by taking $\lim_{n\to\infty} \frac{1}{n}$ of both sides. ■

Using Lemma 1, Theorem 2 and (31), we can write (62) as

$$\eta_{\mathbf{E\Sigma E}}(\gamma) = \int_{-1/2}^{1/2} \frac{1}{1+\gamma S(f)(1-\mathsf{e}/\eta_{\mathbf{E\Sigma E}}(\gamma))} \, \mathrm{d}f \tag{88}$$

Since the horizontal asymptote of the η-transform is equal to the asymptotic fraction of zero eigenvalues, we have:

Corollary 1 *The fraction of zero eigenvalues of the product* $\mathbf{E\Sigma E}$ *satisfies*

$$\lim_{\gamma\to\infty} \eta_{\mathbf{E\Sigma E}}(\gamma) = \max\left\{\mathsf{e}, \mu\left(\{f : S(f) = 0\}\right)\right\} \tag{89}$$

Proof: First note that (88) requires that its fixed-point solution satisfy $\mathsf{e} \leq \eta_{\mathbf{E\Sigma E}}(\gamma) \leq 1$. Denote $\mathcal{I} = \{f : S(f) > 0\}$, and let $\mu(\mathcal{I}) = B \in [0, 1)$ be the fraction of nonzero eigenvalues of $\mathbf{\Sigma}$. We can express (88) as

$$\eta_{\mathbf{E\Sigma E}}(\gamma) = 1 - B + \int_{\mathcal{I}} \frac{1}{1+\gamma S(f)(1-\mathsf{e}/\eta_{\mathbf{E\Sigma E}}(\gamma))} \, \mathrm{d}f \tag{90}$$

If $\mathsf{e} < 1 - B$, then we see that $\eta_{\mathbf{E\Sigma E}}(\infty) = 1 - B$ solves (90) as $\gamma \to \infty$. On the other hand, suppose $\mathsf{e} \geq 1 - B$. As $\gamma \to \infty$, if $\eta_{\mathbf{E\Sigma E}}(\gamma)$ is bounded away from e, then the integral in (90) vanishes as $\gamma \to \infty$, and thus $\lim_{\gamma\to\infty} \eta_{\mathbf{E\Sigma E}}(\gamma) = 1 - B$, which contradicts $\mathsf{e} \leq \eta_{\mathbf{E\Sigma E}}(\gamma)$ unless $\mathsf{e} = 1 - B$ or $\lim_{\gamma\to\infty} \eta_{\mathbf{E\Sigma E}}(\gamma) = \mathsf{e}$. Consequently, the asymptotic fixed point is $\max\{\mathsf{e}, 1 - B\}$. ■

It is intriguing to speculate on a possible shortcut to (88) by application of the framework of free probability. For a.s. asymptotically free matrices $\mathbf{A}, \mathbf{B}$, we have that [20, eq. 2.209]

$$\eta_{\mathbf{AB}}(\gamma) = \eta_{\mathbf{A}}\left(\frac{\gamma}{\Sigma_{\mathbf{B}}\left(\eta_{\mathbf{AB}}(\gamma) - 1\right)}\right) \tag{91}$$

where $\Sigma_{\mathbf{A}}(z)$ denotes the S-transform of $\mathbf{A}$ (see [20, Section 2.2.6] and references therein), that yields the asymptotic eigenvalue distribution of the product of two a.s. asymptotically free matrices. Identifying $\mathbf{B}$ with $\mathbf{E}$ and $\mathbf{A}$ with $\mathbf{\Sigma}$, using Example 3 and [20, eq. 2.86], we have

$$\Sigma_{\mathbf{E}}(z) = -\frac{z+1}{z}\eta_{\mathbf{E}}^{-1}(1+z) = \frac{1+z}{1-\mathsf{e}+z} \tag{92}$$

Then, applying (31) and (92) in (91) we obtain that $\eta_{\mathbf{E\Sigma}}(\gamma)$ satisfies the fixed-point equation (88). Noticing that $\eta_{\mathbf{E\Sigma}}(\gamma) = \eta_{\mathbf{E\Sigma E}}(\gamma)$, the result of Theorem 2 would follow by direct application of known formulas. Unfortunately, $\mathbf{E}$ and $\mathbf{\Sigma}$ are not a.s. asymptotically free in general, and therefore the direct application of (91) is not possible. It is however interesting to notice that the final rigorous result is *as if* $\mathbf{E}$ and $\mathbf{\Sigma}$ were a.s. asymptotically free.

Using (37) with $\mathbf{A} = \mathbf{E\Sigma E}$ the mutual information rate as a function of the signal-to-noise ratio γ can be characterized in terms of the η-transform of $\mathbf{E\Sigma E}$:

$$I_{\mathsf{e}}(\gamma) \;=\; \frac{1}{2}\int_0^{\gamma} \frac{1-\eta_{\mathbf{E\Sigma E}}(x)}{x} \, \mathrm{d}x \tag{93}$$

An alternative characterization of the capacity is given by the following theorem obtained by using Theorem 1.

Theorem 3 *The mutual information rate achieved with output power spectral density $S(f)$ and erasure rate e is equal to*

$$I_{\mathsf{e}}(\gamma) \quad = \quad \frac{1}{2}\int_0^{1-\mathsf{e}} \log\left(1+\daleth_0(y,\gamma)\right)\, dy \tag{94}$$

where $\daleth_0$ is the solution to

$$\frac{\daleth_0(y,\gamma)}{1+\daleth_0(y,\gamma)} = \int_{-1/2}^{1/2} \frac{\gamma S(f)}{1+y\,\gamma S(f)+(1-y)\daleth_0(y,\gamma)}\,\mathrm{d}f \tag{95}$$

Proof: Instead of showing (94) directly we will show that

$$I_{\mathsf{e}}(\gamma) \quad = \quad \frac{1-\mathsf{e}}{2}\int_0^{1} \log\left(1+\daleth_{\mathsf{e}}(y,\gamma)\right) dy \tag{96}$$

with $\daleth_e$ the solution of the fixed-point equation

$$\frac{\daleth_{\mathsf{e}}(y,\gamma)}{1+\daleth_{\mathsf{e}}(y,\gamma)} = \int_{-1/2}^{1/2} \frac{\gamma S(f)}{(1-\mathsf{e})\,y\,\gamma S(f)+1+(1-(1-\mathsf{e})\,y)\daleth_{\mathsf{e}}(y,\gamma)}\,\mathrm{d}f \tag{97}$$

Once this alternative form of the result is established, note that the special case of (97) when $\mathsf{e}=0$ yields (95), which is equivalent to (39) in the special case of Toeplitz $\mathbf{A}$. The desired result (94) will then follow from (96) since the solutions to (97) and (95) are related through

$$\daleth_{\mathsf{e}}(y,\gamma) = \daleth_0(y-\mathsf{e}y,\gamma). \tag{98}$$

To show (96)-(97), we first assume that $S(f)>0$ for $-1/2<f<1/2$, in which case the asymptotic normalized rank of $\mathbf{E\Sigma E}$ is $1-\mathsf{e}$. From Theorem 1 applied to $\mathbf{A}=\mathbf{E\Sigma E}$ we obtain that the Shannon transform is

$$\mathcal{V}_{\mathbf{E\Sigma E}}(\gamma) = (1-\mathsf{e})\int_0^1 \log\left(1+\daleth_{\mathsf{e}}(y,\gamma)\right) dy \tag{99}$$

with $\daleth_{\mathsf{e}}$ satisfying

$$(1-\mathsf{e})y\frac{\daleth_{\mathsf{e}}(y,\gamma)}{1+\daleth_{\mathsf{e}}(y,\gamma)} = 1-\eta_{\mathbf{E\Sigma E}}\left(\frac{\gamma y}{1+(1-y)\daleth_{\mathsf{e}}(y,\gamma)}\right) \tag{100}$$

and $\eta_{\mathbf{E\Sigma E}}$ given as the solution in Theorem 2:

$$\eta_{\mathbf{E\Sigma E}}(t) = \eta_{\mathbf{\Sigma}}\left(t-t\frac{\mathsf{e}}{\eta_{\mathbf{E\Sigma E}}(t)}\right). \tag{101}$$

Letting

$$t = \frac{\gamma y}{1+(1-y)\daleth_{\mathsf{e}}(y,\gamma)} \tag{102}$$

the argument in the right side of (101) becomes

$$\begin{aligned}
\frac{\gamma y}{1+(1-y)\daleth_{\mathsf{e}}(y,\gamma)}\left(1-\frac{\mathsf{e}}{\eta_{\mathbf{E\Sigma E}}(t)}\right) &= \frac{\gamma y}{1+(1-y)\daleth_{\mathsf{e}}(y,\gamma)}\left(1-\frac{\mathsf{e}}{1-(1-\mathsf{e})y\frac{\daleth_{\mathsf{e}}(y,\gamma)}{1+\daleth_{\mathsf{e}}(y,\gamma)}}\right)\\
&= \frac{\gamma y(1-\mathsf{e})}{1+(1-y)\daleth_{\mathsf{e}}(y,\gamma)}\left(\frac{1-y\frac{\daleth_{\mathsf{e}}(y,\gamma)}{1+\daleth_{\mathsf{e}}(y,\gamma)}}{1-(1-\mathsf{e})y\frac{\daleth_{\mathsf{e}}(y,\gamma)}{1+\daleth_{\mathsf{e}}(y,\gamma)}}\right)\\
&= \frac{(1-\mathsf{e})y\gamma}{1+(1-(1-\mathsf{e})\,y)\daleth_{\mathsf{e}}(y,\gamma)}
\end{aligned} \tag{103}$$

Thus we obtain the fixed-point equation

$$(1-\mathsf{e})y\frac{\beth_{\mathsf{e}}(y,\gamma)}{1+\beth_{\mathsf{e}}(y,\gamma)} = 1-\eta_{\mathbf{\Sigma}}\left(\frac{(1-\mathsf{e})y\gamma}{1+(1-(1-\mathsf{e})\,y)\beth_{\mathsf{e}}(y,\gamma)}\right) \tag{104}$$

which straightforward algebra reveals to be identical to (97).

To remove the restriction that $S(f)$ has no zeros, it is enough to realize that the solution to (97) is continuous in $S(f)$. Thus, an infinitesimal perturbation of a nonnegative power spectral density leads to a strictly positive $S(f)$ which achieves essentially the same mutual information. ■

Despite the fact that both the evaluation of (93) and of (96) require numerical integration of a function obtained as the solution of a fixed point equation, the expression of Theorem 3 is in general much easier to compute since the integral with respect to y is on the fixed interval $[0,1]$, independent on the SNR γ.

An immediate consequence of Theorem 3 is:

Corollary 2 *The mutual information $I_{\mathsf{e}}(\gamma)$ is a concave decreasing function of* e.

Proof: Using (94), the first and second derivatives of $I_{\mathsf{e}}(\gamma)$ with respect to e are

$$\frac{\partial I_{\mathsf{e}}(\gamma)}{\partial \mathsf{e}} = -\frac{1}{2}\log(1+\beth_0(1-\mathsf{e},\gamma)) \tag{105}$$

$$\frac{\partial^2 I_{\mathsf{e}}(\gamma)}{\partial \mathsf{e}^2} = \frac{1}{2}\frac{\left.\frac{\partial \beth_0(y,\gamma)}{\partial y}\right|_{y=1-\mathsf{e}}}{1+\beth_0(1-\mathsf{e},\gamma)} \tag{106}$$

Both first and second derivative are negative, since $\beth_0(y,\gamma)$ is non-negative and decreasing with y (see Appendix D). ■

Example 5 *Figures 2 and 3 show the limiting mutual information together with the realizations of $\frac{1}{2n}\log det(\mathbf{I}+\gamma\mathbf{E\Sigma E})$ for $n=100$ and $n=1000$, respectively, for the case $\mathsf{e}=0.3$ and $\mathbf{\Sigma}$ defined by*

$$\sigma_{|i-j|} = e^{-0.2(i-j)^2}$$

As expected, as n increases the mutual information for n-dimensional blocks for given erasure realization concentrates around the deterministic value given by Theorem 2.

Example 6 *Let*

$$S(f)=1 \qquad -1/2\le f\le 1/2 \tag{107}$$

Then it is easy to check that

$$\beth_0(y,\gamma)=\gamma \qquad 0\le y\le 1 \tag{108}$$

and

$$I_{\mathsf{e}}(\gamma)=C_{\mathsf{e}}(\gamma)=(1-\mathsf{e})C_0(\gamma) \tag{109}$$

as is to be expected since the effect of erasures on the capacity of the memoryless channel is a factor of $(1-\mathsf{e})$.

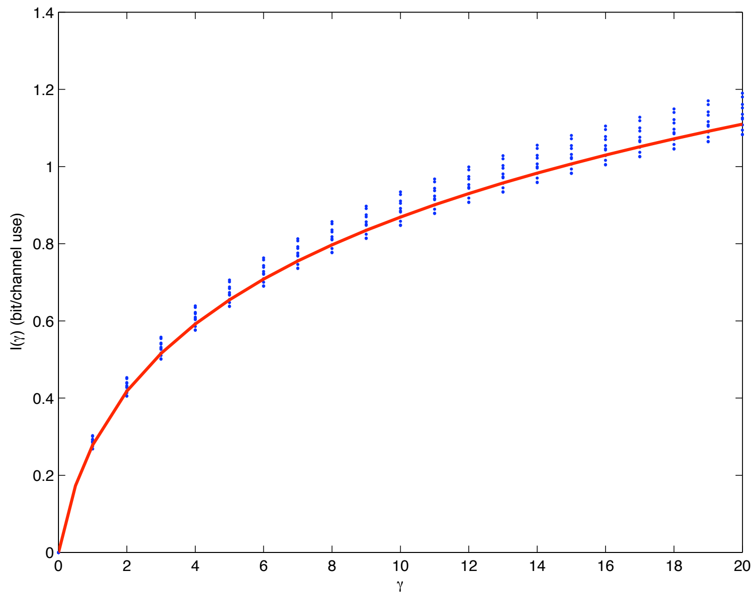

Figure 2: Mutual information $I_{\mathsf{e}}(\gamma)$ and its (random) finite length realizations for $n = 100$, $\mathsf{e} = 0.3$ and $\sigma_{|i-j|} = e^{-0.2(i-j)^2}$.

Example 7 *Suppose the Gaussian stationary random process $\{u_i\}$ in (1) has an ideal low-pass power spectral density*

$$S^{\mathsf{u}}(f) = \begin{cases} \frac{1}{B} & |f| \le B/2 \\ 0 & B/2 < |f| \le 1/2 \end{cases} \tag{110}$$

for some $B \in (0, 1]$. Plugging (110) in (95) we find the quadratic equation

$$\frac{\beth_0^{\mathsf{u}}(y,\gamma)}{1 + \beth_0^{\mathsf{u}}(y,\gamma)} = \frac{\gamma}{1 + y\,\gamma/B + (1-y)\beth_0^{\mathsf{u}}(y,\gamma)} \tag{111}$$

that yields

$$\beth_0^{\mathsf{u}}(y,\gamma) = \frac{1}{2(1-y)}\left[\gamma - 1 - \gamma y/B + \sqrt{(\gamma - 1 - \gamma y/B)^2 + 4\gamma(1-y)}\right] \tag{112}$$

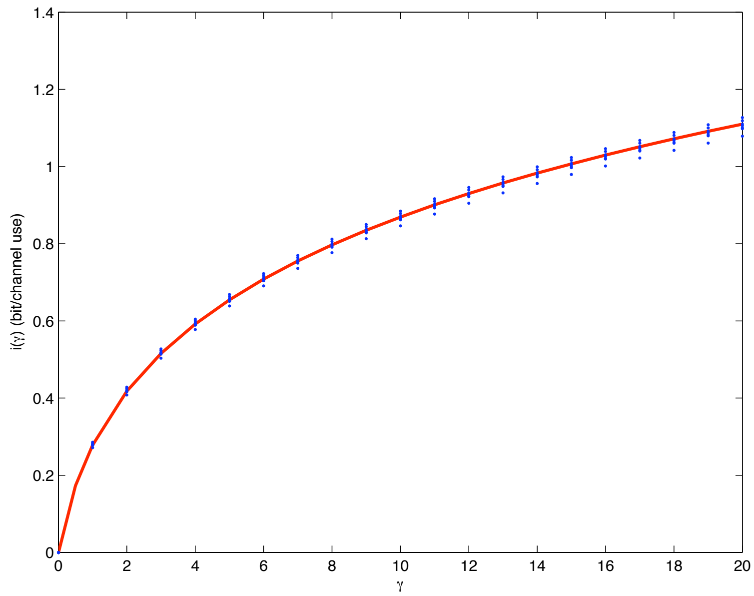

Figure 3: Mutual information $I_{\mathsf{e}}(\gamma)$ and its (random) finite length realizations for $n = 1000$, $\mathsf{e} = 0.3$ and $\sigma_{|i-j|} = e^{-0.2(i-j)^2}$.

Using (94) and (112), it can be checked that

$$I^{\mathsf{u}}_{\mathsf{e}}(\gamma, 1) = \frac{1-\mathsf{e}}{2}\log(1+\gamma) \tag{113}$$

$$I^{\mathsf{u}}_{0}(\gamma, B) = \frac{B}{2}\log\left(1+\frac{\gamma}{B}\right) \tag{114}$$

$$\begin{aligned} I^{\mathsf{u}}_{\mathsf{e}}(\gamma, \tfrac{1}{2}) &= \frac{1-\mathsf{e}}{2}\log\left(\frac{2\mathsf{e}(1+\gamma)-1-\gamma+\sqrt{1+2\gamma+(1-2\mathsf{e})^2\gamma^2}}{2\mathsf{e}}\right) \\ &\quad +\frac{1}{4}\log\left(\frac{1+2\gamma+\gamma^2(1-2\mathsf{e})+(1+\gamma)\sqrt{1+2\gamma+(1-2\mathsf{e})^2\gamma^2}}{2(1+2\gamma)}\right) \end{aligned} \tag{115}$$

$$I^{\mathsf{u}}_{\mathsf{e}}(\gamma, B) = \frac{1-\mathsf{e}}{2}\gamma\log e - \frac{1-\mathsf{e}}{4B}(1-\mathsf{e}+B\mathsf{e})\gamma^2\log e + o(\gamma^2) \tag{116}$$

$$\begin{aligned} I^{\mathsf{u}}_{\mathsf{e}}(\gamma, B) &= \frac{1-2B}{4}\log\frac{2(1-B)B}{B-2B^2-B\gamma+(1-\mathsf{e})\gamma+\Delta} + \frac{1-\mathsf{e}}{2}\log\frac{B(2\mathsf{e}+\gamma-1)-(1-\mathsf{e})\gamma+\Delta}{2B\mathsf{e}} \\ &\quad +\frac{1}{4}\log\frac{2(1-B)B\gamma^2\mathsf{e}^2}{-B^2(1-2(1-\mathsf{e})\gamma+\gamma^2)+B(\gamma-1)\gamma(2-\mathsf{e})-(1-\mathsf{e})\gamma^2+(B+\gamma-B\gamma)\Delta} \end{aligned} \tag{117}$$

where

$$\Delta = \sqrt{B^2((1+\gamma)^2-4(1-\mathsf{e})\gamma)+2B(1-\mathsf{e})(1-\gamma)\gamma+(1-\mathsf{e})^2\gamma^2} \tag{118}$$

Example 8 *Consider a wireless LAN where the access points (cell sites or hot-spots) are connected to a centralized processor via some Internet infrastructure, subject to congestion and iid packet losses. This system is described by the classical Wyner [26] model, where each received signal is affected by independent erasures. Notice that in this case erasures are not independent "in time", but are independent in space, i.e., the links between the access points and the central* *processors are either on or off, with probability* $1 - \mathsf{e}$ *and* e*, respectively, and it is assumed that their state changes in time at a very slow rate with respect to the duration of a codeword.*

In particular, consider the linear Wyner model with interference coefficient $\alpha \in [0, 1]$*: A system with* N *cells,* K *users per cell, and a central processor where the signal vector received at the central processor from the* i*-th cell is given by*

$$\mathbf{y}_i = e_i \left(\sum_{k=1}^{K} \mathbf{x}_{i,k} + \alpha \left(\sum_{k=1}^{K} \mathbf{x}_{i-1,k} + \sum_{k=1}^{K} \mathbf{x}_{i+1,k} \right) \right) + \mathbf{z}_i \tag{119}$$

where $e_i \in \{0, 1\}$ *is the link erasure random variable. It is easy to see that an intra-cell orthogonal access (TDMA/FDMA) strategy is optimal, as in the original Wyner model [26]. Restricting (119) to this case, the centralized processor at any given time receives the signal vector*

$$\mathbf{Y} = \mathbf{E}\mathbf{H}(\alpha)\mathbf{X} + \mathbf{Z} \tag{120}$$

where $\mathbf{E} = diag(e_1, \ldots, e_N)$*,* $\mathbf{X} = (x_1, \ldots, x_N)^T$ *is the vector of transmitted signals from all terminals scheduled in the same time slot, one for each cell, and*

$$\mathbf{H}(\alpha) = \begin{bmatrix} 1 & \alpha & 0 & \cdots & & 0 \\ \alpha & 1 & \alpha & 0 & & \vdots \\ 0 & \alpha & 1 & \alpha & & \\ \vdots & & \ddots & \ddots & & \alpha \\ 0 & & & \cdots & \alpha & 1 \end{bmatrix} \tag{121}$$

Conditioned on $\mathbf{E}$*, the* average *per-cell capacity is*

$$C^{(N)}(\gamma) = \frac{1}{N} \log det \left(\mathbf{I} + \gamma \mathbf{E}\mathbf{H}(\alpha)\mathbf{H}^{\dagger}(\alpha)\mathbf{E} \right) \tag{122}$$

where $\gamma = KP/N_0$ *denotes the total cell SNR [26] (aggregate transmit power of all users in the cell over the Gaussian noise power spectral density). For finite* N*, (122) is a function of the random erasure matrix* $\mathbf{E}$*. For* $N \to \infty$*, the per-cell capacity converges to a deterministic quantity, that can be computed via the results for the Gaussian-erasure channel developed in this paper. In particular, the spectrum* $S(f)$ *corresponding to the asymptotic eigenvalue distribution of* $\mathbf{\Sigma} = \mathbf{H}(\alpha)\mathbf{H}^{\dagger}(\alpha)$ *is given by [26]*

$$S(f) = (1 + 2\alpha \cos(2\pi f))^2 \tag{123}$$

and in this setup no correlation between the users is allowed so the input power spectral density (spatial frequency) is flat. Figures 4 and 5 show the per-cell capacity for $\gamma = 10$ *dB versus the interference parameter* α*, for* $\mathsf{e} = 0.1$ *and* $\mathsf{e} = 0.9$*, respectively. We notice that when the erasure probability is large, the per-cell capacity is increasing with* α*: the "macro-diversity" effect due to the inter-cell cross-talk is very helpful for large* e*. Intuitively, when a cell-site link is down, the terminal in that cell has still a chance to be successfully decoded from the signal received at the neighboring cells. On the contrary, for small* e*, the cell capacity is non-monotone with* α *as in the conventional non-erasure Wyner model [26].*

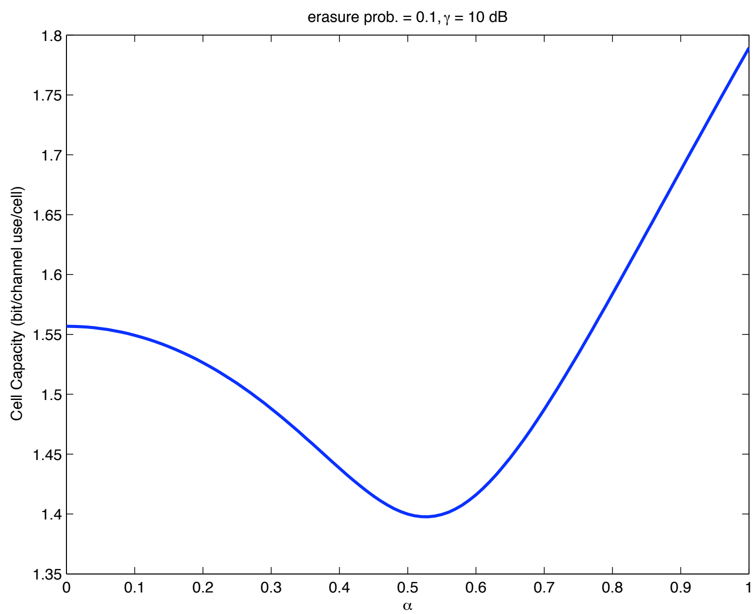

Figure 4: Cell capacity versus interference parameter α for the Wyner model with erasures, $\mathsf{e} = 0.1$ and cell SNR $\gamma = 10$ dB.

5 Input Optimization

The goal of this section is to find the optimum power spectral density as a function of the signal-to-noise ratio γ and the erasure rate e. For that purpose, we will use the following general finite-dimensional result which is of independent interest.

Theorem 4 *Let $\mathbf{\Phi}$ be an $m \times n$ complex valued random matrix whose ith column is denoted by $\boldsymbol{\phi}_i$. Consider the optimization problem*

$$\max_{\mathbf{D}} \mathbb{E}\left[\log\det\left(\mathbf{I} + \gamma \mathbf{\Phi} \mathbf{D} \mathbf{\Phi}^\dagger\right)\right] \tag{124}$$

where the maximum is over all diagonal matrices whose trace is equal to a constant ξ. Then, for $i = 1, \ldots, n$, d_i^*, the ith diagonal element of the diagonal matrix $\mathbf{D}^*$ that achieves the maximum in (124) is the positive solution to

$$\mathbb{E}\left[\frac{Z_i}{1 + \gamma d_i^* Z_i}\right] = \frac{1}{\nu\gamma} \tag{125}$$

$$Z_i = \boldsymbol{\phi}_i^\dagger \left(\mathbf{I} + \gamma \sum_{j \neq i} d_j^* \boldsymbol{\phi}_j \boldsymbol{\phi}_j^\dagger\right)^{-1} \boldsymbol{\phi}_i \tag{126}$$

if it exists (i.e. if $\nu\gamma\mathbb{E}[Z_i] > 1$); otherwise, $d_i^* = 0$. The parameter ν is chosen so that $\sum_{i=1}^n d_i^* = \xi$.

Proof: Fix $\gamma > 0$ and define the strictly concave function

$$I(\mathbf{D}) = \mathbb{E}\left[\log_e \det\left(\mathbf{I} + \gamma \mathbf{\Phi} \mathbf{D} \mathbf{\Phi}^\dagger\right)\right] \tag{127}$$

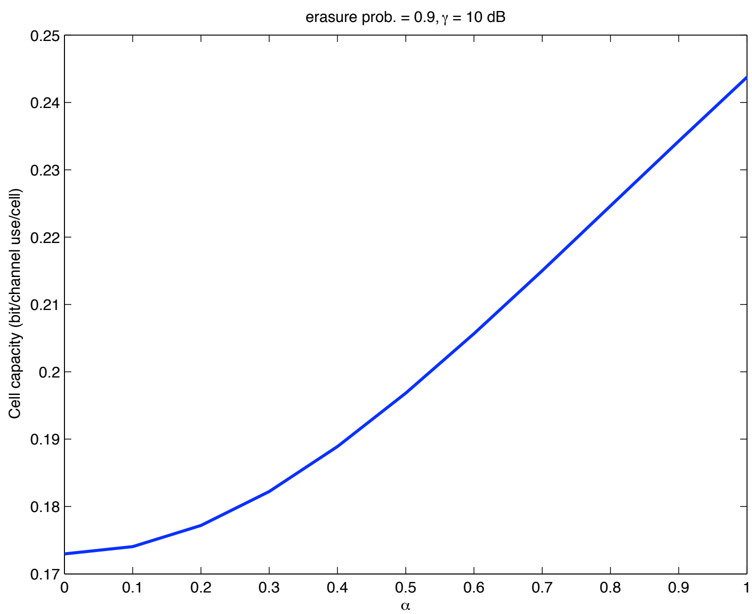

Figure 5: Cell capacity versus interference parameter α for the Wyner model with erasures, $\mathsf{e} = 0.9$ and cell SNR $\gamma = 10$ dB.

Denote

$$\mathbf{D}_\mu = \mathbf{D}^* + \mu \mathbf{T} \tag{128}$$

for $0 \le \mu \le 1$, with $\mathbf{D}^*$ the unique maximizer of (127) and $\mathbf{T}$ a diagonal matrix with zero trace and with ith diagonal coefficient $\tau_i \ge 0$ if $d_i^* = 0$. Then, for all sufficiently small μ, $\mathbf{D}_\mu$ belongs to the feasible set of matrices over which we are optimizing. Denote

$$\mathbf{A}_\ell = \left(\mathbf{I} + \gamma \sum_{j \neq \ell} d_j^* \boldsymbol{\phi}_j \boldsymbol{\phi}_j^\dagger \right)^{-1} \tag{129}$$

It is easy to check that

$$\begin{aligned}
\frac{d}{d\mu} I(\mathbf{D}_\mu)|_{\mu=0^+} &= \mathbb{E}\left[\mathrm{tr}\left\{ \left(\gamma \mathbf{\Phi} \mathbf{T} \mathbf{\Phi}^\dagger\right) \left(\mathbf{I} + \gamma \mathbf{\Phi} \mathbf{D}^* \mathbf{\Phi}^\dagger\right)^{-1} \right\} \right]. && (130)\\
&= \gamma \sum_{\ell=1}^n \tau_\ell \, \mathbb{E}\left[\boldsymbol{\phi}_\ell^\dagger \left(\mathbf{I} + \gamma \mathbf{\Phi} \mathbf{D}^* \mathbf{\Phi}^\dagger\right)^{-1} \boldsymbol{\phi}_\ell \right] && (131)\\
&= \gamma \sum_{\ell=1}^n \tau_\ell \, \mathbb{E}\left[\boldsymbol{\phi}_\ell^\dagger \left(\mathbf{A}_\ell + \gamma d_\ell^* \boldsymbol{\phi}_\ell \boldsymbol{\phi}_\ell^\dagger\right)^{-1} \boldsymbol{\phi}_\ell \right] && (132)\\
&= \gamma \sum_{\ell=1}^n \tau_\ell \, \mathbb{E}\left[\boldsymbol{\phi}_\ell^\dagger \mathbf{A}_\ell^{-1} \boldsymbol{\phi}_\ell - \frac{\gamma d_\ell^*}{1 + \gamma d_\ell^* \boldsymbol{\phi}_\ell^\dagger \mathbf{A}_\ell^{-1} \boldsymbol{\phi}_\ell} (\boldsymbol{\phi}_\ell^\dagger \mathbf{A}_\ell^{-1} \boldsymbol{\phi}_\ell)^2 \right] && (133)\\
&= \gamma \sum_{\ell=1}^n \tau_\ell \, \mathbb{E}\left[Z_\ell \left(1 - \frac{\gamma d_\ell^* Z_\ell}{1 + \gamma d_\ell^* Z_\ell}\right) \right] && (134)\\
&= \gamma \sum_{\ell=1}^n \tau_\ell \, \mathbb{E}\left[\frac{Z_\ell}{1 + \gamma d_\ell^* Z_\ell} \right] && (135)
\end{aligned}$$

where (133) follows from the matrix inversion lemma, and (134) follows from (126).

To conclude the proof we argue by contradiction. Suppose that for some (i, j), $d_i^* > 0$, $d_j^* > 0$ and

$$\mathbb{E}\left[\frac{Z_i}{1+\gamma d_i^* Z_i}\right] > \mathbb{E}\left[\frac{Z_j}{1+\gamma d_j^* Z_j}\right] \tag{136}$$

Then let $\mathbf{T}$ be zero everywhere except $\tau_i = 1 = -\tau_j$. Then (135) is strictly positive contradicting the optimality of $\mathbf{D}^*$. Thus, there must exist a ν such that (125) holds for all those i such that $d_i^* > 0$. This implies that if $\nu\gamma\mathbb{E}[Z_j] \leq 1$, then $d_j^* = 0$. Now, let us suppose that for some j, $d_j^* = 0$ and $\nu\gamma\mathbb{E}[Z_j] > 1$. Then let $\mathbf{T}$ be zero everywhere except $\tau_j = \nu = -\tau_i$ where i is such that $d_i^* > 0$. The resulting value of (135) is $\nu\gamma\mathbb{E}[Z_j] - 1 > 0$, contradicting the optimality of $\mathbf{D}^*$. ■

A result related to Theorem 4 is obtained in [21] by using an MMSE representation. Note that in the special case where $\mathbf{\Phi}$ is a deterministic diagonal matrix, we recover the conventional waterfilling solution with

$$Z_i = |\phi_{ii}|^2. \tag{137}$$

The waterfilling solution is relevant when the channel matrix is known at transmitter and receiver in which case parallel orthogonal channels can be created. But, in general, the solution given by Theorem 4 does not correspond to waterfilling on any statistical measure of the channel. However, as in the special case of waterfilling, the high-SNR solution is

$$\lim_{\gamma\to\infty} \mathbf{D}^* = \frac{\xi}{n}\mathbf{I}. \tag{138}$$

We now return to the Gaussian erasure channel and show, using Theorem 4, an appealing solution for the capacity-achieving power spectral density.

Theorem 5 *The capacity-achieving input power spectral density is*

$$S_x^*(f, \gamma, \mathsf{e}) = \frac{1}{\theta(\mathsf{e}, \zeta)}\left[\zeta - \frac{1}{\gamma|H(f)|^2}\right]^+ \tag{139}$$

where

$$\theta(\mathsf{e}, \zeta) = \frac{1}{2}\left[\zeta + 1 - \sqrt{(\zeta-1)^2 + 4\zeta\mathsf{e}}\right] \tag{140}$$

and ζ is chosen so that the integral of (139) is equal to 1.

Proof: Using Theorem 12 in Appendix B and the asymptotic equivalence of the product of Toeplitz matrices to the product of the corresponding circulant matrices we can write the objective function as

$$C_{\mathsf{e}}(\gamma) = \lim_{n\to\infty} \frac{1}{2n} \max_{\Lambda_x} \mathbb{E}\left[\log\det\left(\mathbf{I} + \gamma\mathbf{Q}\mathbf{\Lambda}_H\mathbf{\Lambda}_x\mathbf{\Lambda}_H\mathbf{Q}^\dagger\right)\right] \tag{141}$$

where recall from (65) that $\mathbf{Q} = \mathbf{EF}$, $\mathbf{\Lambda}_H = \mathrm{diag}\{H_1, \ldots, H_n\}$ is the diagonal matrix of the singular values of $\mathbf{H}$ and the maximization is over the set of nonnegative diagonal matrices with trace equal to n. To solve (141) we make use of Theorem 4 with $\mathbf{\Phi} = \mathbf{Q}\mathbf{\Lambda}_H$ and $\mathbf{D} = \mathbf{\Lambda}_x$. In this case, (126) takes the form

$$Z_i = H_i^2 \mathbf{q}_i^\dagger \left(\mathbf{I} + \gamma \sum_{j \neq i} H_j^2 d_j^* \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i \tag{142}$$

Taking the limit of (142) as $n \to \infty$, Lemma 1, Lemma 13, and (82) imply that almost surely

$$\lim_{n\to\infty} \mathbf{q}_i^\dagger \left(\mathbf{I} + \gamma \sum_{j \neq i} H_j^2 d_j^* \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i = \alpha \tag{143}$$

$$= 1 - \frac{\mathsf{e}}{\eta_{\mathbf{E\Sigma E}}(\gamma)} \tag{144}$$

Thus,

$$\mathbb{E}\left[\frac{Z_i}{1 + \gamma d_i^* Z_i} \right] \to \frac{H_i^2 \alpha}{1 + \gamma d_i^* H_i^2 \alpha} \tag{145}$$

Using (145) in Theorem 4, the sought-after power spectral density satisfies

$$\frac{|H(f)|^2 \alpha \nu \gamma}{1 + \gamma \alpha S_x^*(f) |H(f)|^2} = 1 \tag{146}$$

if $\alpha\nu\gamma|H(f)|^2 > 1$, and $S_x^*(f) = 0$ otherwise. Thus, using (144), we get

$$S_x^*(f) = \left[\nu - \frac{1}{\gamma |H(f)|^2 (1 - \frac{\mathsf{e}}{\eta_{\mathbf{E\Sigma E}}(\gamma)})} \right]^+ \tag{147}$$

Choosing the water level so that the integral of (147) is equal to 1, leads, according to (34), to

$$\nu = \frac{1}{1 - \eta_{\mathbf{\Sigma}}(\alpha\gamma)} \tag{148}$$

$$= \frac{1}{1 - \eta_{\mathbf{E\Sigma E}}(\gamma)} \tag{149}$$

where we have used Theorem 2.

To obtain the final result (139), we change variables and let $\zeta = \nu\alpha$, thereby expressing (147) as

$$S_x^*(f) = \frac{1}{\alpha} \left[\zeta - \frac{1}{\gamma |H(f)|^2} \right]^+ \tag{150}$$

where in view of (144) and (149), α satisfies the quadratic equation

$$\alpha = \left(1 - \frac{\mathsf{e}}{1 - \alpha} \right) \zeta \tag{151}$$

whose solution in the interval $[0, 1 - \mathsf{e}]$ is denoted by $\theta(\mathsf{e}, \zeta)$ and given in (140). ∎

The function $\theta(\mathsf{e}, \zeta)$ is monotonically decreasing with e, and increasing with ζ. It satisfies

$$\begin{aligned}
\theta(0,\zeta) &= 1 \\
\theta(1,\zeta) &= 0 \\
\theta(\mathsf{e},0) &= 0 \\
\lim_{\zeta\to\infty} \theta(\mathsf{e},\zeta) &= 1-\mathsf{e}
\end{aligned}$$

As the following corollary to Theorem 5 shows, the effect of erasures on the capacity-achieving input power spectral density is tantamount to a reduction in the signal to noise ratio.

Corollary 3 *For all* $0 \le \mathsf{e} \le 1$, $\gamma > 0$,

$$S_x^*(f, \gamma/\kappa, \mathsf{e}) = S_x^*(f, \gamma, 0) \tag{152}$$

where

$$\kappa = 1 - \frac{\mathsf{e}\,\zeta_\gamma}{\zeta_\gamma - 1} \tag{153}$$

and ζ_γ *is the erasure-free water level for* γ, *i.e. the solution to (18).*

Proof: Let

$$\begin{aligned}
\bar{\zeta} &= \kappa\,\zeta_\gamma &(154)\\
&= \frac{\kappa}{1-\frac{\mathsf{e}}{1-\kappa}} &(155)
\end{aligned}$$

where (155) follows by solving for ζ_γ in (153). Comparing (155) to (151) we conclude that $\kappa = \theta(\mathsf{e}, \bar{\zeta})$, and consequently:

$$\begin{aligned}
\frac{\bar{\zeta}}{\theta(\mathsf{e},\bar{\zeta})} &= \zeta_\gamma &(156)\\
\frac{\gamma\theta(\mathsf{e},\bar{\zeta})}{\kappa} &= \gamma. &(157)
\end{aligned}$$

Thus, (152) follows in view of (139) and its particularization to $\mathsf{e} = 0$:

$$S_x^*(f,\gamma,0) = \left[\zeta_\gamma - \frac{1}{\gamma|H(f)|^2}\right]^+ \tag{158}$$

■

To illustrate the application of Corollary 3, we find an explicit expression for the optimum power spectral density for arbitrary erasure rate e, as a function of the optimum solution for $\mathsf{e} = 0$ under the assumption that the signal-to-noise ratio is high enough that the support of $S_x^*(f,\gamma,0)$ is the whole interval $[-1/2, 1/2]$. This implies that $|H(f)|$ is bounded away from zero and we can define

$$\begin{aligned}
Z(f) &= \mathsf{h} - \frac{1}{|H(f)|^2} &(159)\\
\mathsf{h} &= \int_{-1/2}^{1/2} \frac{1}{|H(f)|^2} df &(160)
\end{aligned}$$

and therefore

$$\int_{-1/2}^{1/2} Z(f)\, df = 0 \tag{161}$$

Because of the assumption on γ, the capacity-achieving power spectral density in the absence of erasures is equal to (cf. (158))

$$S_x^*(f,\gamma,0) = \zeta_\gamma - \frac{1}{\gamma|H(f)|^2} \tag{162}$$

$$= 1 + \frac{Z(f)}{\gamma} \tag{163}$$

where in (163) we have used the fact that the area of $S_x^*(f,\gamma,0)$ is normalized to 1, and thus

$$\zeta_\gamma = 1 + \frac{\mathsf{h}}{\gamma} \tag{164}$$

According to Corollary 3

$$S_x^*(f,\gamma,\mathsf{e}) = S_x^*(f,\gamma\kappa,0) \tag{165}$$

with

$$\kappa = 1 - \frac{e\,\zeta_{\gamma\kappa}}{\zeta_{\gamma\kappa}-1} \tag{166}$$

$$= 1 - \mathsf{e}\left(\frac{\gamma\kappa}{\mathsf{h}} + 1\right) \tag{167}$$

$$= \frac{1-\mathsf{e}}{1+\frac{\gamma\mathsf{e}}{\mathsf{h}}} \tag{168}$$

where (168) follows from solving the first-order equation (167).

Thus,

$$S_x^*(f,\gamma,\mathsf{e}) = 1 + \frac{Z(f)}{\gamma}\frac{1+\frac{\gamma\mathsf{e}}{\mathsf{h}}}{1-\mathsf{e}} \tag{169}$$

$$= S_x^*(f,\gamma,0) + \mathsf{e}\,\frac{\frac{\gamma}{\mathsf{h}}+1}{1-\mathsf{e}} Z(f) \tag{170}$$

6 Bounds on the Mutual Information

The purpose of this section is to develop bounds on the mutual information $I_{\mathsf{e}}(\gamma)$ for given output spectrum $S(f)$ that are significantly easier to compute. While none of the bounds developed in this section captures the behavior of mutual information for both high and low SNR and any arbitrary $S(f)$, it turns out that by taking the minimum of the upper bounds and the maximum of the lower bounds we obtain a generally accurate and asymptotically tight approximation.

The proposed lower and upper bounds are given in Theorems 6 and 7 respectively.

Theorem 6 *Let $y_i = \sqrt{\gamma}\, e_i u_i + n_i$, where $\{u_i\}$ is stationary Gaussian with power spectral density $S(f)$, and $\{n_i\}$ is stationary Gaussian with unit power spectral density. Furthermore, denote*

$$G = \int_{-1/2}^{1/2} S(f)\, \mathrm{d}f. \tag{171}$$

and let B the Lebesgue measure of $\mathcal{I} \in [-1/2, 1/2]$, the support of $S(f)$ (or "generalized bandwidth" of $\{u_i\}$). Then, the mutual information rate is lower bounded by

$$
\begin{aligned}
I_{\mathsf{e}}(\gamma) &\geq I_{\mathsf{e}}^{\mathsf{u}}(G\gamma, B) + \frac{1}{2}\int_{\mathcal{I}} \log\left(\frac{1+\gamma S(f)}{1+\gamma G/B}\right) \mathrm{d}f & (172)\\
I_{\mathsf{e}}(\gamma) &\geq (1-\mathsf{e})I_0(\gamma) & (173)\\
I_{\mathsf{e}}(\gamma) &\geq I_0(\gamma) - \frac{\mathsf{e}}{2}\log(1+\daleth_0(1-\mathsf{e},\gamma)) & (174)
\end{aligned}
$$

where an explicit expression for $I_{\mathsf{e}}^{\mathsf{u}}(\gamma, B)$ is given in (117).

Proof:

(172) For the purposes of proving this bound, we will only show the special case $G = 1$, as then, the general formula follows because the mutual information achieved with $(\gamma, S(f))$ is the same as that achieved by $(a\gamma, S(f)/a)$. For a given arbitrary unit-power $S(f)$, we define the associated unit-power on-off spectrum

$$
S^{\mathsf{u}}(f) = \frac{1}{B} 1\{S(f) > 0\} \tag{175}
$$

Let $\daleth_0(y,\gamma)$ and $\daleth_0^{\mathsf{u}}(y,\gamma)$ be the solutions of (95) for $S(f)$ and $S^{\mathsf{u}}(f)$ respectively. Applying Jensen's inequality to (95) we have

$$
\begin{aligned}
\frac{\daleth_0(y,\gamma)}{1+\daleth_0(y,\gamma)} &= \mathbb{E}\left[\frac{\gamma B\, S(F)}{1+y\gamma S(F)+(1-y)\daleth_0(y,\gamma)}\right]\\
&\leq \frac{\gamma}{1+y\gamma/B+(1-y)\daleth_0(y,\gamma)} & (176)
\end{aligned}
$$

where the expectation is with respect to the random variable F uniformly distributed on $\mathcal{I} = \{f : S(f) > 0\}$, and $\mathbb{E}[S(F)] = 1/B$ since we are assuming $G = 1$.

Since it is apparent that the solution of (95) is invariant to swapping frequency bands, $\daleth_0^{\mathsf{u}}(y,\gamma)$ is given by (112) that was obtained assuming a brickwall low-pass spectrum. Thus, we see from (111) that $\daleth_0^{\mathsf{u}}(y,\gamma) = x$ with

$$
\frac{x(1+y\gamma/B+(1-y)x)}{1+x} = \gamma \tag{177}
$$

Moreover, since the left side of (177) is monotonically increasing in x, and in view of (176)

$$
\daleth_0(y,\gamma) \leq \daleth_0^{\mathsf{u}}(y,\gamma) \quad \text{for any } y \in [0,1], \gamma \geq 0, \tag{178}
$$

Using (178), the mutual informations in (172) satisfy ($G = 1$)

$$
\begin{aligned}
I_{\mathsf{e}}(\gamma) - I_{\mathsf{e}}^{\mathsf{u}}(\gamma, B) &= \frac{1}{2}\int_0^{1-\mathsf{e}} \log\frac{1+\daleth_0(y,\gamma)}{1+\daleth_0^{\mathsf{u}}(y,\gamma)}\,\mathrm{d}y & (179)\\
&= \frac{1}{2}\int_0^{1} \log\frac{1+\daleth_0(y,\gamma)}{1+\daleth_0^{\mathsf{u}}(y,\gamma)}\,\mathrm{d}y - \frac{1}{2}\int_{1-\mathsf{e}}^{1} \log\frac{1+\daleth_0(y,\gamma)}{1+\daleth_0^{\mathsf{u}}(y,\gamma)}\,\mathrm{d}y & (180)\\
&\geq \frac{1}{2}\int_0^{1} \log\frac{1+\daleth_0(y,\gamma)}{1+\daleth_0^{\mathsf{u}}(y,\gamma)}\,\mathrm{d}y & (181)\\
&= I_0(\gamma) - I_0^{\mathsf{u}}(\gamma, B) & (182)\\
&= \frac{1}{2}\int_{\mathcal{I}} \log\left(\frac{1+\gamma S(f)}{1+\gamma/B}\right)\mathrm{d}f & (183)
\end{aligned}
$$

(173) follows from the concavity of $I_{\mathsf{e}}(\gamma)$ in e (Corollary 2) and $I_1(\gamma) = 0$.

(174) is obtained using the monotonicity of $\beth_0(y, \gamma)$ with respect to y (see Appendix D), which yields

$$\begin{aligned} I_{\mathsf{e}}(\gamma) &= \frac{1}{2}\int_0^{1-\mathsf{e}} \log\left(1 + \beth_0(y,\gamma)\right) \mathrm{d}y \\ &= I_0(\gamma) - \frac{1}{2}\int_{1-\mathsf{e}}^{1} \log\left(1 + \beth_0(y,\gamma)\right) \mathrm{d}y \\ &\geq I_0(\gamma) - \frac{\mathsf{e}}{2}\log(1 + \beth_0(1-\mathsf{e},\gamma)) \end{aligned} \tag{184}$$

■

Interestingly, the lower bound (172) implies that

$$I_0(\gamma) - I_{\mathsf{e}}(\gamma) \leq I_0^{\mathsf{u}}(\gamma, B) - I_{\mathsf{e}}^{\mathsf{u}}(\gamma, B) \tag{185}$$

This shows that the largest mutual information penalty caused by the presence of erasures for spectra with fixed generalized bandwidth B occurs when $S(f)$ is on/off. This conforms to the intuition that correlation in the process $\{u_i\}$ helps estimating the erased components, and therefore if $\{u_i\}$ is as "white" as possible given its bandwidth B, then erasures are maximally harmful in terms of mutual information.

The lower bound (173) demonstrates that contiguous erasures, giving rise to the mutual information at the RHS of (173) are in fact worst case, and in that sense are associated with the fountain capacity [16] of a filtered Gaussian channel.

Theorem 7 *With the same notation as in Theorem 6, the mutual information is upper bounded by*

$$I_{\mathsf{e}}(\gamma) \leq I_{\mathsf{e}}^{\mathsf{u}}(G\gamma, B) \tag{186}$$

$$I_{\mathsf{e}}(\gamma) \leq I_0((1-\mathsf{e})\gamma) \tag{187}$$

$$I_{\mathsf{e}}(\gamma) \leq \frac{1-\mathsf{e}}{2}\log(1 + G\gamma) \tag{188}$$

$$I_{\mathsf{e}}(\gamma) \leq I_0(\gamma) + \frac{\mathsf{e}}{2}\log\left(\int_{-1/2}^{1/2} \frac{1}{1+\gamma S(f)}\,\mathrm{d}f\right) \tag{189}$$

Proof:

(186) follows from (178).

(187) follows from the concavity of log det and Jensen's inequality:

$$\mathbb{E}\left[\log\det(\mathbf{I} + \gamma\mathbf{E}\boldsymbol{\Sigma}\mathbf{E})\right] = \mathbb{E}\left[\log\det(\mathbf{I} + \gamma\mathbf{E}\boldsymbol{\Sigma})\right] \leq \log\det(\mathbf{I} + \gamma(1-\mathsf{e})\boldsymbol{\Sigma}) \tag{190}$$

(188) follows from the Hadamard inequality:

$$\mathbb{E}\left[\log\det(\mathbf{I} + \gamma\mathbf{E}\boldsymbol{\Sigma}\mathbf{E})\right] \leq \frac{(1-\mathsf{e})n}{2}\log(1 + G\gamma) \tag{191}$$

(189) follows from the monotonicity of $\beth_0(y, \gamma)$ with respect to y (Appendix D):

$$\begin{aligned} I_{\mathsf{e}}(\gamma) &= I_0(\gamma) - \frac{1}{2}\int_{1-\mathsf{e}}^{1} \log(1 + \beth_0(y, \gamma))\, \mathrm{d}y \\ &\leq I_0(\gamma) - \frac{\mathsf{e}}{2}\log(1 + \beth_0(1, \gamma)) \\ &= I_0(\gamma) + \frac{\mathsf{e}}{2}\log\left(\int_{-1/2}^{1/2} \frac{1}{1 + \gamma S(f)}\, \mathrm{d}f\right) \end{aligned} \tag{192}$$

where (192) follows from (95) particularized for $y = 1$.

■

It is interesting to notice that (188) can also be derived as an application for the MMSE-I relation of [7]. Starting from

$$\frac{d}{d\gamma}\frac{1}{n}\log\det(\mathbf{I} + \gamma\mathbf{E\Sigma E}) = \frac{1}{n}\sum_{i:e_i=1} \mathbb{E}[|u_i - \widehat{u}_i|^2|\mathbf{E}] \tag{193}$$

for any i such that $e_i = 1$, we can upper bound the corresponding MMSE $\mathbb{E}\left[|u_i - \widehat{u}_i|^2|\mathbf{E}\right]$ by the MMSE obtained by estimating u_i from the observation y_i alone. We obtain

$$\frac{d}{d\gamma}I_{\mathsf{e}}(\gamma) \leq \frac{1-\mathsf{e}}{2}\frac{G}{1 + G\gamma} \tag{194}$$

Integrating both sides of (194) we obtain (188). Furthermore, the lower bound (173) can also be shown as an application of the MMSE-I relation of [7].

Figures 6 and 7 compare the bounds for the case of an ideal low-pass transfer function with bandwidth $B = 0.5$, and erasure probability $\mathsf{e} = 0.1$ and $\mathsf{e} = 0.7$, respectively. In those figures, LB1, LB2 and LB3 refer to (172), (173) and (174), respectively, and UB1, UB2, UB3 and UB4 refer to (186), (187), (188) and (189), respectively. Obviously, for an ideal low-pass transfer function we have that (172) and (186) coincide with the exact mutual information and are therefore tight. Figure 8 shows the mutual information and the bounds for a Gauss-Markov process defined as

$$u_i = \rho\, u_{i-1} + \sqrt{1 - \rho^2}\, v_i \tag{195}$$

with $\{v_i\}$ i.i.d. standard Gaussian, with $\rho = 0.9$, and for $\mathsf{e} = 0.5$. Here, all bounds with the exception of UB2 yield the correct high-SNR slope, given by $1 - \mathsf{e} = 0.5$ (see Theorem 10). However, LB1 and UB1 that were tight in the ideal low-pass spectrum case are now quite loose, while LB2, LB3 and UB4 are quite accurate.

No bound uniformly outperforms the others over all power spectral densities $S(f)$. Often, a simple and accurate estimate of the mutual information for a given $S(f)$ is obtained by taking the minimum of UB1, UB2, UB3, UB4 and the maximum of LB1, LB2, LB3.

7 Asymptotics

In this section we focus on the asymptotic behavior of $I_{\mathsf{e}}(\gamma)$ for given $S(f)$, and of the capacity $C_{\mathsf{e}}(\gamma)$ with the optimal input spectrum $S_x^*(f, \gamma, \mathsf{e})$ given in Theorem 5, in four limiting cases: $\mathsf{e} \to 0$, $\mathsf{e} \to 1$, $\gamma \to 0$, and $\gamma \to \infty$. Often, the limiting behaviors capture interesting qualitative properties and offer easily computable and asymptotically tight approximations.

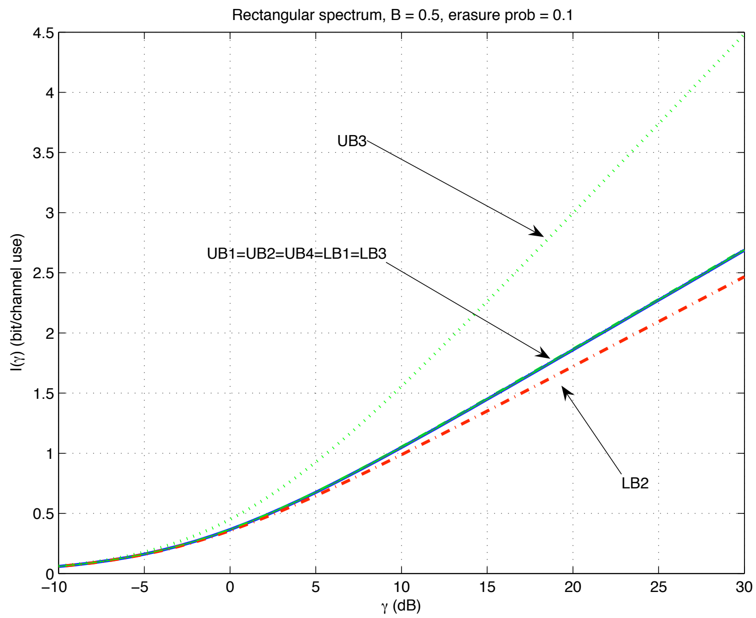

Figure 6: Mutual information and the bounds for an ideal low-pass transfer function with $B = 0.5$ and $\mathsf{e} = 0.1$.

7.1 Sporadic Erasures

The regime of sporadic erasures $\mathsf{e} \to 0$ is of particular interest. The following result shows that the upper bound (189) is asymptotically tight.

Theorem 8 *For any output power spectral density $S(f)$ and $\mathsf{e} \to 0$,*

$$I_{\mathsf{e}}(\gamma) = I_0(\gamma) - \frac{\mathsf{e}}{2} \log \frac{1}{\eta_{\mathbf{\Sigma}}(\gamma)} + o(\mathsf{e}) \tag{196}$$

Proof: Since $I_{\mathsf{e}}(\gamma)$ is a concave decreasing function of $\mathsf{e} \in [0, 1]$ (Corollary 2), for sufficiently small e it satisfies

$$I_{\mathsf{e}}(\gamma) = I_0(\gamma) + \left.\frac{\partial I_{\mathsf{e}}(\gamma)}{\partial \mathsf{e}}\right|_{\mathsf{e}=0} \mathsf{e} + o(\mathsf{e}) \tag{197}$$

Using Theorem 3, we have

$$\left.\frac{\partial I_{\mathsf{e}}(\gamma)}{\partial \mathsf{e}}\right|_{\mathsf{e}=0} = -\frac{1}{2} \log\left(1 + \beth_0(1, \gamma)\right) \tag{198}$$

Particularizing (95) we have

$$\frac{\beth_0(1, \gamma)}{1 + \beth_0(1, \gamma)} = \int_{-1/2}^{1/2} \frac{\gamma S(f)}{1 + \gamma S(f)} \, \mathrm{d}f \tag{199}$$

Solving for $\beth_0(1, \gamma)$ we obtain

$$1 + \beth_0(1, \gamma) = \left[\int_{-1/2}^{1/2} \frac{1}{1 + \gamma S(f)} \, \mathrm{d}f \right]^{-1} = \frac{1}{\eta_{\mathbf{\Sigma}}(\gamma)} \tag{200}$$

which, with (197), yields (196). ■

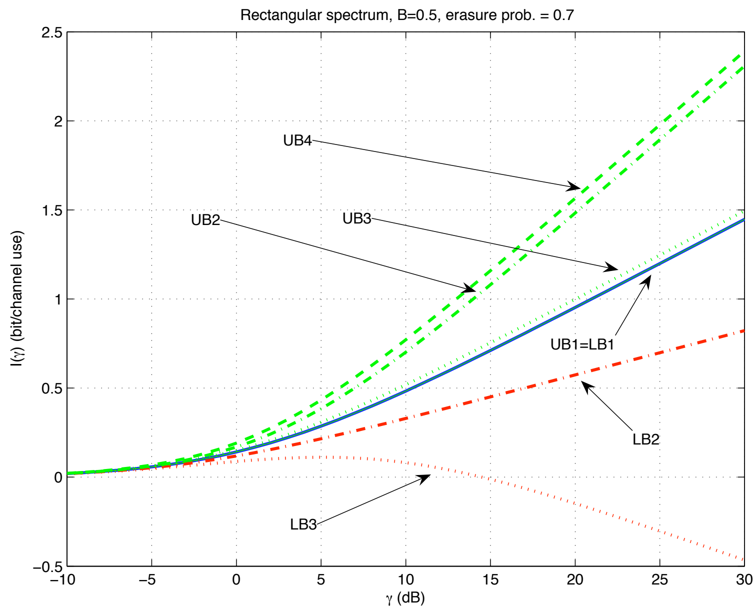

Figure 7: Mutual information and the bounds for an ideal low-pass transfer function with $B = 0.5$ and $\mathsf{e} = 0.7$.

The next result deals specifically with the capacity of the Gaussian erasure channel for sporadic erasures:

Theorem 9 *In the regime of sporadic erasures the capacity of the Gaussian erasure channel satisfies:*

$$C_{\mathsf{e}}(\gamma) = C_0(\gamma) - \frac{\mathsf{e}}{2} \log \frac{\zeta}{\zeta - 1} + o(\mathsf{e}) \tag{201}$$

where ζ is the water level of the power spectral density that achieves $C_0(\gamma)$.

Proof: In the proof, we specifically denote the dependence on γ of the water level of the power spectral density that achieves $C_0(\gamma)$, i.e. ζ_γ.

Let $S_x^*(f, \gamma, \mathsf{e})$ denote the capacity-achieving input power spectral density for erasure rate e, and denote the η-transform of the corresponding output spectrum by

$$\eta_{\mathbf{\Sigma}_{\mathsf{e}}}(\gamma) = \int_{-1/2}^{1/2} \frac{1}{1 + \gamma S_x^*(f, \gamma, \mathsf{e}) |H(f)|^2} \mathrm{d}f \tag{202}$$

Using (196) we get,

$$\begin{aligned} I_{\mathsf{e}}(S_x^*(\cdot, \gamma, \mathsf{e})) - I_0(S_x^*(\cdot, \gamma, 0)) &\geq I_{\mathsf{e}}(S_x^*(\cdot, \gamma, 0)) - I_0(S_x^*(\cdot, \gamma, 0)) && (203) \\ &= -\frac{\mathsf{e}}{2} \log \frac{1}{\eta_{\mathbf{\Sigma_0}}(\gamma)} + o(\mathsf{e}) && (204) \\ &= -\frac{\mathsf{e}}{2} \log \frac{1}{1 - \frac{1}{\zeta_\gamma}} + o(\mathsf{e}) && (205) \end{aligned}$$

where (205) follows from (34).

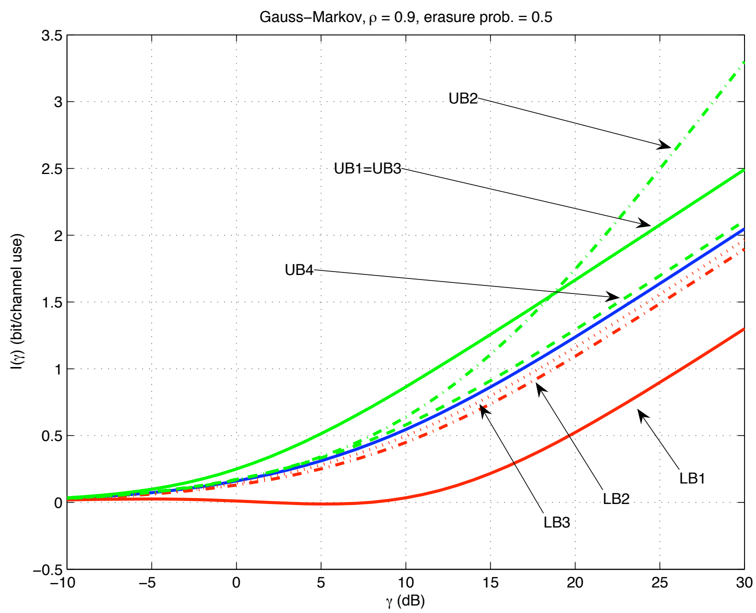

Figure 8: Mutual information and the bounds for a Gauss-Markov spectrum with $\rho = 0.9$ and $\mathsf{e} = 0.5$.

To show the reverse inequality we use the upper bound (189), that yields

$$\begin{aligned} I_{\mathsf{e}}(S_x^*(\cdot,\gamma,\mathsf{e})) - I_0(S_x^*(\cdot,\gamma,0)) &\le I_0(S_x^*(\cdot,\gamma,\mathsf{e})) - I_0(S_x^*(\cdot,\gamma,0)) \\ &\quad + \frac{\mathsf{e}}{2}\log\left(\int_{-1/2}^{1/2} \frac{1}{1+\gamma S_x^*(f,\gamma,\mathsf{e})|H(f)|^2}\, df\right) \quad (206) \\ &= I_0(S_x^*(\cdot,\gamma,\mathsf{e})) - I_0(S_x^*(\cdot,\gamma,0)) \\ &\quad - \frac{\mathsf{e}}{2}\log\frac{1}{\eta_{\mathbf{\Sigma}_{\mathsf{e}}}(\gamma)} + o(\mathsf{e}) \quad (207) \\ &= I_0(S_x^*(\cdot,\gamma,\mathsf{e})) - I_0(S_x^*(\cdot,\gamma,0)) \\ &\quad - \frac{\mathsf{e}}{2}\log\frac{1}{1-\frac{1}{\zeta_{\gamma\kappa}}} + o(\mathsf{e}) \quad (208) \end{aligned}$$

where κ is defined such that (recall Corollary 3)

$$S_x^*(f,\gamma,\mathsf{e}) = S_x^*(f,\gamma\kappa,0) \quad (209)$$

Now let us consider the following expansion of $S_x^*(f,\gamma,\mathsf{e})$

$$S_x^*(f,\gamma,\mathsf{e}) = S_x^*(f,\gamma,0) + \mathsf{e}\, a_1(f) + o(\mathsf{e}) \quad (210)$$

(For sufficiently large γ, an explicit expression is given in (170).) Since $\kappa < 1$, the support of $S_x^*(f,\gamma,\mathsf{e})$ is a subset of that of $S_x^*(f,\gamma,0)$. Thus, $a_1(f) = 0$ for those frequencies such that $S_x^*(f,\gamma,0) = 0$. Moreover, since the integrals of the spectral densities in (210) are equal to 1, $\int a_1(f)\,\mathrm{d}f = 0$. Using (210), the difference of mutual informations in the right side of (208) becomes

$$I_0(S_x^*(\cdot,\gamma,\mathsf{e})) - I_0(S_x^*(\cdot,\gamma,0)) = \frac{1}{2}\int_{-1/2}^{1/2} \log\left(\frac{1+\gamma S_x^*(f,\gamma,\mathsf{e})|H(f)|^2}{1+\gamma S_x^*(f,\gamma,0)|H(f)|^2}\right) df \tag{211}$$

$$= \frac{1}{2}\int_{-1/2}^{1/2} \log\left(1+\gamma\frac{\mathsf{e}\, a_1(f)|H(f)|^2}{1+\gamma S_x^*(f,\gamma,0)|H(f)|^2}\right) df + o(\mathsf{e}) \tag{212}$$

$$\leq \frac{\mathsf{e}\log e}{2}\int_{-1/2}^{1/2} \frac{\gamma a_1(f)|H(f)|^2}{1+\gamma S_x^*(f,\gamma,0)|H(f)|^2}\, df + o(\mathsf{e}) \tag{213}$$

$$= \frac{\mathsf{e}\log e}{2\zeta}\int_{-1/2}^{1/2} a_1(f)\, df + o(\mathsf{e}) \tag{214}$$

$$= o(\mathsf{e}) \tag{215}$$

where (214) follows from the fact that $a_1(f) = 0$ if $S_x^*(f,\gamma,0) = 0$, and according to (16), if $S_x^*(f,\gamma,0) > 0$ it satisfies

$$1+\gamma S_x^*(f,\gamma,0)|H(f)|^2 = \zeta\gamma|H(f)|^2 \tag{216}$$

Since $\kappa \to 1$ as $\mathsf{e} \to 0$, the result follows in view of (205), (208) and (215). ∎

The first-order decrease of the capacity of finite-alphabet channels with memory observed through an erasure channel is shown in [24] to be equal to the so-called erasure mutual information rate evaluated at the capacity-achieving distribution. Although the channel in this paper falls outside the scope of that finite-alphabet result in [24], (201) is consistent with it. Using (201), it is interesting to compare the impact of erasures on the capacity of the channel with and without memory. In the high signal-to-noise ratio regime the behavior depends on whether the channel has full bandwidth or not. If $B < 1$, then the linear decrease with e converges to

$$\frac{1}{2}\log\frac{1}{1-B}$$

as $\gamma \to \infty$, in contrast to an unbounded term

$$\frac{B}{2}\log(1+\frac{\gamma}{B})$$

when the channel is flat on its passband. If $B = 1$, then the linear decrease with e is also unbounded but can be shown to be smaller than $\frac{1}{2}\log(1+\gamma)$. In the low signal-to-noise ratio regime, the linear decrease with e can also be shown to be smaller than the channel capacity.

7.2 Sporadic non-erasures

As $\mathsf{e} \to 1$, the solution in Theorem 3 takes the limiting expression:

$$\lim_{\mathsf{e}\to 1} \frac{I_{\mathsf{e}}(\gamma)}{1-\mathsf{e}} = \frac{1}{2}\log\left(1+\gamma\int_{-1/2}^{1/2} S(f)\,\mathrm{d}f\right) \tag{217}$$

Optimizing (217) over unit input power spectrum $S_x(f)$ with $S(f) = |H(f)|^2 S_x(f)$ results in $S_x(f)$ that places all its power at the most favorable frequency (or frequencies), yielding

$$\lim_{\mathsf{e}\to 1} \frac{C_{\mathsf{e}}(\gamma)}{1-\mathsf{e}} = \frac{1}{2}\log\left(1+\gamma G_{\max}\right) \tag{218}$$

where the maximum channel gain is denoted by

$$G_{\max} = \max_f |H(f)|^2. \tag{219}$$

The solution (218) obtained with an input that has very long range memory (if $G_{\max}$ is achieved at only one frequency) may be surprising until realizing that in the sporadic non-erasure regime, the channel effectively breaks any long-term memory in the input process and achieving maximum efficiency in power transfer becomes paramount.

Example 9 *Figure 9 shows the capacity $C_{\mathsf{e}}(\gamma)$ for fixed $\gamma = 10$ dB, as a function of the erasure probability e, for a channel with transfer function $|H(f)|^2$ defined by its η-transform*[3]

$$\eta_{|H|^2}(z) = 0.2 + \frac{0.2}{1+z} + \frac{0.2}{1+2z} + \frac{0.2}{1+5z} + \frac{0.1}{1+7z} + \frac{0.1}{1+10z}$$

Figure 9 shows also the affine approximations of capacity for sporadic erasures ($\mathsf{e} \downarrow 0$) and sporadic non-erasures ($\mathsf{e} \to 1$) obtained before.

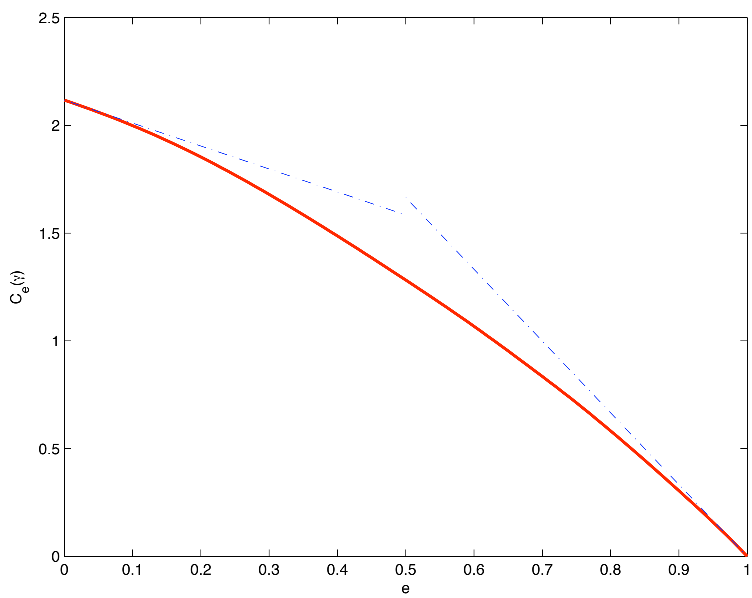

Figure 9: Capacity versus erasure rate at $\gamma = 10dB$ (Example 9), with its affine asymptotic approximations for small and large erasure probability.

7.3 Low SNR Asymptotics

We characterize the behavior of capacity for fixed e and vanishing γ. In order to be consistent with the notation and definitions introduced in [22] to characterize the low-SNR (or *wideband*) regime, we shall consider the complex circularly symmetric version of our channel model, where $\gamma = E_s/N_0$ is the *transmitter* SNR, and N_0 denotes the complex noise variance per component.

[3]Since capacity depends on $|H(f)|^2$ only via its η-transform, defined by $\eta_{|H|^2}(z) = \int_{-1/2}^{1/2} \frac{1}{1+z|H(f)|^2}\, \mathrm{d}f$, it is convenient to give examples in terms of the η-transform itself.

The system spectral efficiency C (measured in bit/s/Hz) as a function of E_b/N_0, where E_b denotes the transmitted energy per information bit, is given implicitly by

$$\begin{aligned} \mathsf{C} &= C_{\mathsf{e}}(\gamma) \\ \frac{E_b}{N_0} &= \frac{\gamma}{C_{\mathsf{e}}(\gamma)} \end{aligned} \tag{220}$$

where $C_{\mathsf{e}}(\gamma)$ is the channel capacity expressed in bits per (complex) channel use.

At low SNR or, equivalently, for small C, E_b/N_0 in dB as a function of C takes on the convenient form

$$\left.\frac{E_b}{N_0}\right|_{\mathrm{dB}} = \left.\left(\frac{E_b}{N_0}\right)_{\min}\right|_{\mathrm{dB}} + \frac{\mathsf{C}}{\mathcal{S}_0} 10 \log_{10} 2 + o(\mathsf{C}) \tag{221}$$

We have the following result which states that the presence of erasures requires boosting the energy per bit by a factor of $\frac{1}{1-\mathsf{e}}$ at the transmitter.

Theorem 10 *The minimum energy per bit and wideband slope $\mathcal{S}_0$ of the spectral efficiency of the Gaussian erasure channel are equal to*

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{1}{(1-\mathsf{e})G_{\max}\log_2 e} \tag{222}$$

$$\mathcal{S}_0 = \frac{2(1-\mathsf{e})B_{\max}}{\mathsf{e}B_{\max}+1-\mathsf{e}} \tag{223}$$

where $B_{\max} = \mu(\{f : |H(f)|^2 = G_{\max}\})$.

Proof: Let $\mathcal{I}_{\max}$ denote the set of frequencies $f \in [-1/2, 1/2]$ such that $|H(f)|^2 = G_{\max}$. For the time being, we assume that $B_{\max} = \mu(\mathcal{I}_{\max}) > 0$. For sufficiently small γ, the capacity-achieving input spectrum is given by

$$S_x^*(f,\gamma,\mathsf{e}) = \begin{cases} \frac{1}{B_{\max}} & f \in \mathcal{I}_{\max} \\ 0 & f \neq \mathcal{I}_{\max} \end{cases} \tag{224}$$

It follows that for sufficiently low γ, we can use the Taylor series expansion in (116) to obtain

$$\begin{aligned} \left(\frac{E_b}{N_0}\right)_{\min} &= \frac{1}{\left.\frac{\partial C_{\mathsf{e}}(\gamma)}{\partial\gamma}\right|_{\gamma=0}} \\ &= \frac{1}{(1-\mathsf{e})G_{\max}\log_2 e} \end{aligned} \tag{225}$$

and

$$\begin{aligned} \mathcal{S}_0 &= -2\frac{\left[\left.\frac{\partial C_{\mathsf{e}}(\gamma)}{\partial\gamma}\right|_{\gamma=0}\right]^2}{\left.\frac{\partial^2 C_{\mathsf{e}}(\gamma)}{\partial\gamma^2}\right|_{\gamma=0}} \\ &= \frac{2(1-\mathsf{e})^2 G_{\max}^2}{(1-\mathsf{e})G_{\max}^2 + \frac{1-B_{\max}}{B_{\max}}(1-\mathsf{e})^2 G_{\max}^2} \\ &= \frac{2(1-e)B_{\max}}{\mathsf{e}B_{\max}+1-\mathsf{e}} \end{aligned} \tag{226}$$

where we used the formulas in [22, eq. (35)] and [22, Theorem 9]. The case $B_{\max} = 0$, as for example when $|H(f)|^2$ has a unique maximum $G_{\max}$ on $f \in [-1/2, 1/2]$, can be approximated by a sequence of channel transfer functions that are equal to $G_{\max}$ on intervals of positive but vanishing measure. Since $(E_b/N_0)_{\min}$ does not depends on $B_{\max}$, each of these approximations yields the same minimum E_b/N_0, which is therefore equal to the limit. The slope $\mathcal{S}_0$ can be obtained as the limit for $B_{\max} \to 0$ of these approximations, yielding $\mathcal{S}_0 = 0$. It follows that (222) is valid also for the case $B_{\max} = 0$. ∎

Example 10 *Figure 10 shows the spectral efficiency* C *vs* E_b/N_0 *in dB for two Gaussian erasure channels with the same erasure probability* $\mathsf{e} = 0.5$*. The first, denoted as "Channel 1", is given in Example 9, and the second, denoted as "Channel 2", has* η *transform given by*

$$\eta_{|H|^2}(z) = 0.9 + \frac{0.1}{1 + 10z}$$

Both channels have $B_{\max} = 0.1$ *and* $G_{\max} = 10$*, hence they have the same* $(E_b/N_0)_{\min}$ *and* $\mathcal{S}_0$*, although for non-vanishing* γ *their capacity is very different. Also shown in Figure 10 is the affine approximation (221) of* C*, that becomes exact as* $(E_b/N_0) \downarrow (E_b/N_0)_{\min}$*.*

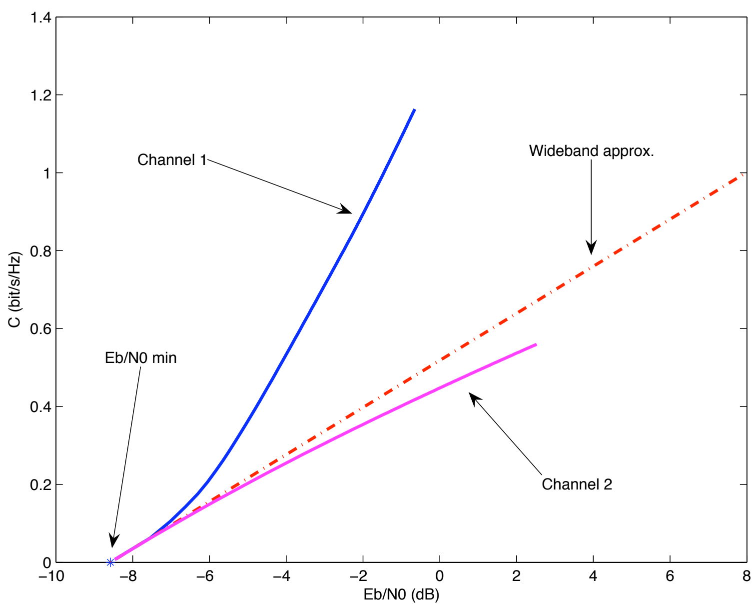

Figure 10: Spectral efficiency vs E_b/N_0 in dB, for two channels with the same minimum E_b/N_0 and slope $\mathcal{S}_0$. The dash-dotted straight line shows the wideband affine approximation.

7.4 High-SNR Asymptotics

For large SNR, the spectral efficiency C introduced in the previous section behaves like [17]

$$\left.\frac{E_b}{N_0}\right|_{\text{dB}} = \frac{\mathsf{C}}{\mathcal{S}_\infty} 10 \log_{10} 2 - 10 \log_{10}(\mathsf{C}) + \mathcal{L}_\infty 10 \log_{10} 2 + o(1) \tag{227}$$

where $\mathcal{S}_\infty$ and $\mathcal{L}_\infty$ are known as the high-SNR slope and the high-SNR dB offset respectively [17]. Using (220), it follows that for large γ,

$$C_{\mathsf{e}}(\gamma) = \mathcal{S}_\infty \left(\log_2 \gamma - \mathcal{L}_\infty\right) + o(1) \tag{228}$$

(expressed in bits per complex dimension). The high-SNR slope $\mathcal{S}_\infty$ captures the channel degrees of freedom, also referred to as "multiplexing gain", i.e.,

$$\mathcal{S}_\infty = \lim_{\gamma\to\infty} \frac{C_{\mathsf{e}}(\gamma)}{\log\gamma} \tag{229}$$

Theorem 11 *Consider a Gaussian erasure channel with given channel transfer function $H(f)$ and erasure probability e. Let $\mathcal{I} = \{f : |H(f)|^2 > 0\}$ and $B = \mu(\mathcal{I})$ denote its generalized bandwidth. The high-SNR slope is given by*

$$\mathcal{S}_\infty = \min\{1-\mathsf{e}, B\} \tag{230}$$

and the high-SNR dB offset is

$$\mathcal{L}_\infty = -\int_0^1 \log_2 \digamma(y)\ \mathrm{d}y \tag{231}$$

where $\digamma(y)$ is the solution of the fixed point equation

$$B - y\mathcal{S}_\infty = \int_{\mathcal{I}} \frac{1}{1 + \frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\digamma}\frac{|H(f)|^2}{B}}\ \mathrm{d}f \tag{232}$$

Proof: It is well-known that a white input ($\mathbf{\Sigma}_x = \mathbf{I}$) achieves the same $\mathcal{S}_\infty$ as the optimal input covariance $\mathbf{\Sigma}_x^*$. This can be shown by the following upper and lower bounds that hold for any finite n. Clearly, we have

$$\mathbb{E}[\log\det(\mathbf{I} + \gamma\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})] \le \max_{\mathbf{\Sigma}_x : \mathrm{tr}(\mathbf{\Sigma}_x)\le n} \mathbb{E}[\log\det(\mathbf{I} + \gamma\mathbf{E}\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^\dagger\mathbf{E})] \tag{233}$$

Moreover, since $\log\det(\cdot)$ is increasing on the cone of positive definite matrices and for any input covariance that satisfies the trace constraint $\mathrm{tr}(\mathbf{\Sigma}_x) \le n$ we have that $n\mathbf{I} - \mathbf{\Sigma}_x$ is non-negative definite, it follows that

$$\max_{\mathbf{\Sigma}_x : \mathrm{tr}(\mathbf{\Sigma}_x)\le n} \mathbb{E}[\log\det(\mathbf{I} + \gamma\mathbf{E}\mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^\dagger\mathbf{E})] \le \mathbb{E}[\log\det(\mathbf{I} + n\gamma\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})] \tag{234}$$

Hence, using a sandwich argument, it is immediate to see that for any finite n,

$$\begin{aligned}
\lim_{\gamma\to\infty} \frac{C_{\mathsf{e}}(\gamma)}{\log\gamma} &= \lim_{\gamma\to\infty} \frac{\mathbb{E}[\log\det(\mathbf{I} + \gamma\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})]}{\log\gamma} \\
&= \lim_{\gamma\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^n \left[\log\gamma\lambda_i(\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})\right]^+}{\log\gamma}\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n 1\{\lambda_i(\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E}) > 0\}\right] \\
&= \mathbb{E}[\mathrm{rank}(\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})]
\end{aligned} \tag{235}$$

Corollary 1 yields

$$\frac{\mathrm{rank}(\mathbf{E}\mathbf{H}\mathbf{H}^\dagger\mathbf{E})}{n} \to \min\{1-\mathsf{e}, B\}, \quad \text{a.s.}$$

and (230) is proved.

We turn our attention now to the high-SNR offset $\mathcal{L}_\infty$. From Theorem 1 we have that

$$C_{\mathsf{e}}(\gamma) = \mathcal{S}_\infty \int_0^1 \log(1 + \beth(y,\gamma))\,\mathrm{d}y \tag{236}$$

where $\beth(y,\gamma)$ is the solution of the fixed-point equation (39), with $\mathbf{A} = \mathbf{E}\mathbf{\Sigma}^*\mathbf{E}$ and $\rho = \mathcal{S}_\infty$. Furthermore, in the high-SNR regime $\beth(y,\gamma) = \digamma(y)\gamma + O(1)$ as per (54) where $\digamma(y)$ is the solution to (55).

From (228) and (236), we have

$$\begin{aligned}\mathcal{L}_\infty &= \lim_{\gamma\to\infty}\left(\log\gamma - \frac{1}{\mathcal{S}_\infty}C_{\mathsf{e}}(\gamma)\right)\\ &= \lim_{\gamma\to\infty}\left(\log\gamma - \int_0^1 \log_2\left(1+\gamma\digamma(y)\right)\,\mathrm{d}y\right)\\ &= -\int_0^1 \log_2\digamma(y)\ \mathrm{d}y \end{aligned}\tag{237}$$

Hence, it remains only to be shown that $\digamma(y)$ is given by the solution of the fixed-point equation (232). The fixed-point equation (55) takes the form

$$\mathcal{S}_\infty y = 1 - \eta_{\mathbf{E}\mathbf{\Sigma}^*_\infty\mathbf{E}}\left(\frac{y}{(1-y)\digamma}\right) \tag{238}$$

where $\mathbf{\Sigma}^*_\infty$ denotes the linear-system output covariance matrix driven by the capacity achieving input in the limit of large γ. Notice that in contrast to the case of γ-independent $\mathbf{A}$ in (55), here $\mathbf{\Sigma}^*$ depends on γ via the capacity-achieving input spectrum.

Using Theorem 2, we have that

$$\eta_{\mathbf{E}\mathbf{\Sigma}^*_\infty\mathbf{E}}(t) = \eta_{\mathbf{\Sigma}^*_\infty}\left(t - t\frac{\mathsf{e}}{\eta_{\mathbf{E}\mathbf{\Sigma}^*_\infty\mathbf{E}}(t)}\right) \tag{239}$$

Letting $t = \frac{y}{(1-y)\digamma}$ and using (238), we obtain

$$y\mathcal{S}_\infty = 1 - \eta_{\mathbf{\Sigma}^*_\infty}\left(\frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\digamma}\right) \tag{240}$$

Now, we notice that as $\gamma\to\infty$ the optimal input spectrum satisfies:

$$S_x^*(f,\gamma,\mathsf{e}) = \begin{cases} \frac{1}{B} - O(1/\gamma) & f\in\mathcal{I}\\ 0 & f\notin\mathcal{I}\end{cases} \tag{241}$$

It follows that

$$\begin{aligned}\eta_{\mathbf{\Sigma}^*_\infty}(t) &= \lim_{\gamma\to\infty}\int_{-1/2}^{1/2}\frac{1}{1+tS_x^*(f,\gamma,\mathsf{e})|H(f)|^2}\,\mathrm{d}f\\ &= 1 - B + \int_{\mathcal{I}}\frac{1}{1+t\frac{|H(f)|^2}{B}}\,\mathrm{d}f\end{aligned}\tag{242}$$

Using (242) into (240) we arrive at

$$B - y\mathcal{S}_\infty = \int_{\mathcal{I}}\frac{1}{1+\frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\digamma}\frac{|H(f)|^2}{B}}\,\mathrm{d}f \tag{243}$$

and the theorem is proved. ∎

As an example of the application of Theorem 11, we obtain $\mathcal{L}_\infty$ in closed form for a channel with ideal low-pass transfer function, by explicitly solving the fixed point equation for $\mathsf{F}(y)$. We also provide upper and lower bounds on $\mathcal{L}_\infty$ for a general channel transfer function that require no fixed-point equation solution and numerical integration.

Considering the case where $H(f)$ is an ideal low-pass transfer function with bandwidth B and power gain G. In this case, the fixed-point equation (232) becomes

$$B - y\mathcal{S}_\infty = \frac{B}{1 + \frac{y(1-\mathsf{e}-y\mathcal{S}_\infty}{(1-y)(1-y\mathcal{S}_\infty)\mathsf{F}}\frac{G}{B}} \tag{244}$$

We consider the case $\mathcal{S}_\infty = B$ and $\mathcal{S}_\infty = 1 - \mathsf{e}$ separately. If $\mathcal{S}_\infty = B \leq 1 - \mathsf{e}$, then (244) becomes

$$1 - y = \frac{1}{1 + \frac{y(1-\mathsf{e}-yB}{(1-y)(1-yB)\mathsf{F}}\frac{G}{B}} \tag{245}$$

Solving for F we find

$$\mathsf{F}(y) = G\frac{\frac{1-\mathsf{e}}{B} - y}{1 - By} \tag{246}$$

which yields

$$\begin{aligned}
\mathcal{L}^{\mathsf{u}}_\infty &= -\int_0^1 \log_2\left(G\frac{\frac{1-\mathsf{e}}{B} - y}{1 - By}\right) \mathrm{d}y \\
&= -\log_2 G + \left(\frac{1-\mathsf{e}}{B} - 1\right)\log_2\left(1 - \frac{B}{1-\mathsf{e}}\right) - \frac{1-B}{B}\log_2(1-B) + \log_2\frac{B}{1-\mathsf{e}}
\end{aligned} \tag{247}$$

If $\mathcal{S}_\infty = 1 - \mathsf{e} \leq B$, then (244) becomes

$$1 - y\frac{1-\mathsf{e}}{B} = \frac{1}{1 + \frac{y(1-\mathsf{e})}{(1-(1-\mathsf{e})y)\mathsf{F}}\frac{G}{B}} \tag{248}$$

Solving for F we find

$$\mathsf{F}(y) = G\frac{1 - y\frac{1-\mathsf{e}}{B}}{1 - (1-\mathsf{e})y} \tag{249}$$

which yields

$$\begin{aligned}
\mathcal{L}^{\mathsf{u}}_\infty &= -\int_0^1 \log_2\left(G\frac{1 - y\frac{1-\mathsf{e}}{B}}{1 - (1-\mathsf{e})y}\right) \mathrm{d}y \\
&= -\log_2 G + \left(\frac{B}{1-\mathsf{e}} - 1\right)\log_2\left(1 - \frac{1-\mathsf{e}}{B}\right) - \frac{\mathsf{e}}{1-\mathsf{e}}\log_2 \mathsf{e}
\end{aligned} \tag{250}$$

Next, consider the case of a general channel transfer function with generalized bandwidth B and average gain $\overline{G}$. Using Jensen's inequality in the fixed-point equation (232) we have

$$\int_{\mathcal{I}} \frac{1}{1 + \frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\mathsf{F}}\frac{|H(f)|^2}{B}} \mathrm{d}f \geq \frac{B}{1 + \frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\mathsf{F}}\frac{\overline{G}}{B}} \tag{251}$$

Let $\mathsf{F}(y)$ denote the solution of (232) for the given $H(f)$ and $\mathsf{F}^{\mathsf{u}}(y)$ denote the solution of the new fixed-point equation

$$B - y\mathcal{S}_\infty = \frac{B}{1 + \frac{y(1-\mathsf{e}-y\mathcal{S}_\infty)}{(1-y)(1-y\mathcal{S}_\infty)\mathsf{F}}\frac{\overline{G}}{B}} \tag{252}$$

Since the right-hand side of (232) is an increasing function of $\digamma$, it follows that $\digamma^{\mathsf{u}}(y) \geq \digamma(y)$ for all $y \in [0,1]$. This yields the lower bound

$$\begin{aligned}\mathcal{L}_\infty &= -\int_0^1 \log_2 \digamma(y)\,\mathrm{d}y \\ &\geq -\int_0^1 \log_2 \digamma^{\mathsf{u}}(y)\,\mathrm{d}y \\ &= \mathcal{L}_\infty^{\mathsf{u}} \end{aligned} \tag{253}$$

where $\mathcal{L}_\infty^{\mathsf{u}}$ is given by (247) or by (250), depending on B and e, replacing G by $\overline{G}$.

Finally, by operating similarly to what done for the mutual information lower bound (172) we can show the upper bound

$$\mathcal{L}_\infty \leq \mathcal{L}_\infty^{\mathsf{u}} + \beta \tag{254}$$

where $\beta \geq 0$ is the channel Jensen's penalty factor, defined by

$$\beta = \frac{B}{\mathcal{S}_\infty}\left[\log_2(\overline{G}) - \frac{1}{B}\int_{\mathcal{I}} \log_2(|H(f)|^2)\,\mathrm{d}f\right] \tag{255}$$

In fact, we can write

$$\begin{aligned} C_{\mathsf{e}}(\gamma) &\geq \int_{\mathcal{I}} \log(1+\gamma|H(f)|^2 S_x^*(f,\gamma,\mathsf{e}))\,\mathrm{d}f - \int_{1-\mathsf{e}}^1 \log\left(1+\daleth_0^{\mathsf{u}}(y,\overline{G}\gamma)\right)\,\mathrm{d}y \\ &= \int_{\mathcal{I}} \log(1+\gamma|H(f)|^2 S_x^*(f,\gamma,\mathsf{e}))\,\mathrm{d}f - \left(\int_0^1 \log\left(1+\daleth_0^{\mathsf{u}}(y,\overline{G}\gamma)\right)\,\mathrm{d}y - \int_0^{1-\mathsf{e}} \log\left(1+\daleth_0^{\mathsf{u}}(y,\overline{G}\gamma)\right)\,\mathrm{d}y\right) \\ &= \int_0^{1-\mathsf{e}} \log\left(1+\daleth_0^{\mathsf{u}}(y,\overline{G}\gamma)\right)\,\mathrm{d}y - \left(B\log(1+\overline{G}\gamma/B) - \int_{\mathcal{I}} \log(1+\gamma|H(f)|^2 S_x^*(f,\gamma,\mathsf{e}))\,\mathrm{d}f\right) \\ &= \int_0^{1-\mathsf{e}} \log\left(1+\daleth_0^{\mathsf{u}}(y,\overline{G}\gamma)\right)\,\mathrm{d}y - \int_{\mathcal{I}} \log\left(\frac{1+\overline{G}\gamma/B}{1+\gamma|H(f)|^2 S_x^*(f,\gamma,\mathsf{e})}\right)\,\mathrm{d}f \end{aligned} \tag{256}$$

where first term in (256) coincides with the capacity of a Gaussian erasure channel with ideal low-pass transfer function, bandwidth B and gain $\overline{G}$, while the second term in (256), in the limit of large γ is easily recognized to yield $\int_{\mathcal{I}} \log \frac{\overline{G}}{|H(f)|^2}\,\mathrm{d}f$. The upper bound (254) follows by noticing that a lower bound on capacity yields an upper bound to the high-SNR dB offset.

Example 11 *Figure 11 shows the spectral efficiency* C *vs* E_b/N_0 *in dB for large SNR for two Gaussian erasure channels with the same erasure probability* $\mathsf{e} = 0.1$ *and the following transfer function* η*-transform,*
Channel 3:

$$\eta_{|H_3|^2}(z) = 0.3 + \frac{0.7}{1+10z} \tag{257}$$

Channel 4:

$$\eta_{|H_4|^2}(z) = 0.3 + \frac{0.2}{1+z} + \frac{0.2}{1+5z} + \frac{0.1}{1+8z} + \frac{0.1}{1+20z} + \frac{0.1}{1+30z} \tag{258}$$

Channel 3 is a low-pass with gain 10 and $B = 0.7$*. Channel 4 has five different nonzero gains, the same bandwidth and average gain* $\overline{G} = 10$*. Also shown in Figure 11 is the approximation (227) of* C*, that becomes exact as* $(E_b/N_0) \to \infty$*. In this case, since* $1-\mathsf{e} > B$*, we have* $\mathcal{S}_\infty = B = 0.7$*.*

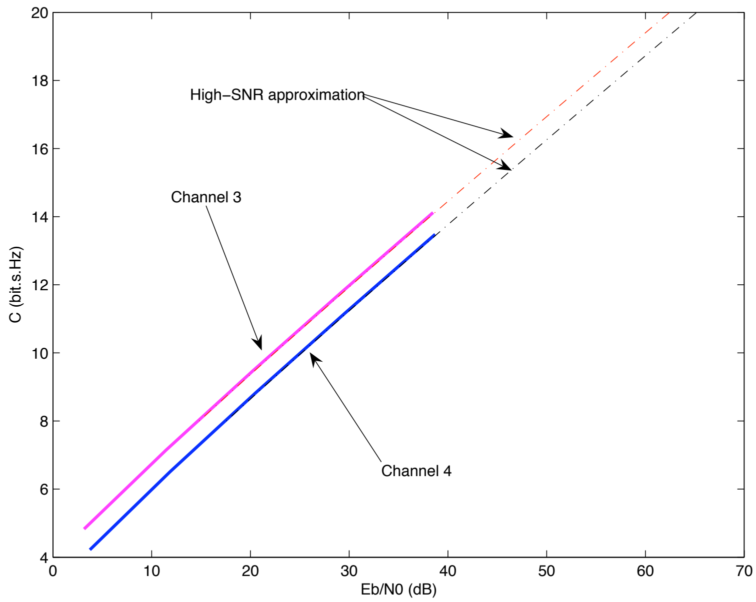

Figure 11: Spectral efficiency vs E_b/N_0 in dB, for two channels with the same generalized bandwidth $B = 0.7$ and average channel gain $\overline{G} = 10$ and $\mathsf{e} = 0.1$. The dash-dotted line shows the high-SNR approximations.

8 Conclusion

We have found the capacity of the discrete-time linear filter additive Gaussian channel whose observations are randomly erased by a memoryless process known to the receiver only. This model has a variety of applications in wireline and wireless communications scenarios ranging from classical impulse jammed Gaussian channels to macro-diversity cellular systems with a backbone network infrastructure affected by packet losses.

Stationary Gaussian inputs maximize the mutual information subject to an input power constraint. Because of the presence of erasures the classical Toeplitz asymptotic spectral theory for deterministic matrices is insufficient in order to characterize the mutual information. Our solution is based on a novel result in random matrix theory, which is of independent interest: the asymptotic spectral distribution of random sub-matrices of a nonnegative definite Toeplitz matrix, obtained by retaining columns/rows independently and with equal probability.

The input-output mutual information takes the form of a fixed-point equation given by Theorem 3, and which demonstrates explicitly the role played by erasures in reducing the number of effective degrees of freedom. This representation generalizes the well-known fact that for memoryless channels observed through erasure channels the capacity is reduced by a multiplicative factor equal to the non-erasure rate. Furthermore, we have been able to find a remarkable closed-form solution of the fixed-point equation of Theorem 3 in the case of a rectangular input spectrum (ideal low-pass filter).

We have given a general result for optimum power allocation for linear vector channels unknown at the transmitter. The application of this result to the optimization of the input power spectral density leads to the pleasing conclusion that the classical waterfilling solution remains optimal in the presence of erasures: their only effect is to introduce a penalty in the signal-to-noise ratio that should be assumed for the waterfilling solution.

We have provided closed form expressions for asymptotic low and high signal-to-noise ratio and for sporadic erasures and sporadic nonerasures. In addition, we have derived a number of simple bounds, which are useful in the asymptotic analysis and shed further insight into the impact of erasures on Gaussian channels with memory.

Our development has been limited to the important special case of memoryless erasures. Are these results either upper or lower bounds to the general case of erasures with memory? Neither. Fixing an arbitrary channel transfer function, we can always find a sufficiently slow binary Markov chain governing the erasures that the capacity will be arbitrarily close to $(1-\mathsf{e})C_0(\gamma)$, which is a lower bound to the memoryless-erasure capacity according to Theorem 6. Conversely, consider an ideal low-pass channel with bandwidth $B = \frac{1}{2}$. According to Example 7, the memoryless-erasure capacity with $\mathsf{e} = \frac{1}{2}$ is equal to $\frac{1}{4}\log(\sqrt{1+2\gamma}+1+\gamma) - \frac{1}{4}\log 2$, which is strictly smaller than $\frac{1}{2}\log(1+\gamma)$ which is the capacity when every other symbol is erased.

The technical development of this work advances the application of random matrix theory to information theoretic problems. We are currently applying these technical developments to problems related to estimation and prediction subject to missing random observations. Furthermore, we are working on the nontrivial extension of the setup to the basic problem of frequency-flat fading dispersive channels with receiver side information.

APPENDIX

A Proof of Lemma 1

Recall the standard definition of asymptotic equivalence of matrices [6, Thm. 5.3]

Definition 3 *The $n \times n$ matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ are asymptotically equivalent if*

i) $\|\mathbf{A}_n\| \leq M < \infty$ *and* $\|\mathbf{B}_n\| \leq M' < \infty$

ii) $\lim_{n\to\infty} |\mathbf{A}_n - \mathbf{B}_n| = 0$

where the strong norm and the weak norm of an $n \times n$ matrix $\mathbf{A}$ are defined as

$$\|\mathbf{A}\|^2 = \max_i \lambda_i(\mathbf{A}\mathbf{A}^\dagger) \tag{259}$$

$$|\mathbf{A}|^2 = \frac{1}{n}\mathrm{tr}\{\mathbf{A}\mathbf{A}^\dagger\} \tag{260}$$

respectively.

We first show that the matrices $\mathbf{E\Sigma E}$ and $\mathbf{E\Psi E}$ are asymptotically equivalent for any realization of $\mathbf{E}$. To see this note first that both matrices are uniformly bounded in strong norm: From the assumption that the sequence in (25) is absolutely summable we get

$$\|\mathbf{E\Sigma E}\|^2 \leq \|\mathbf{E}\|^4\|\mathbf{\Sigma}\|^2 \tag{261}$$

$$\leq \max_i \lambda_i^2(\mathbf{\Sigma}) \tag{262}$$

$$< M \tag{263}$$

and analogously for $\mathbf{E\Psi E}$. Furthermore,

$$|\mathbf{E\Sigma E} - \mathbf{E\Psi E}|^2 = |\mathbf{E}\,(\mathbf{\Sigma} - \mathbf{\Psi})\,\mathbf{E}|^2 \tag{264}$$

$$\leq |\mathbf{E}|^4|\,|\mathbf{\Sigma} - \mathbf{\Psi}|^2 \tag{265}$$

$$\to 0 \tag{266}$$

where (266) follows from the fact that $\mathbf{\Sigma}$ and the circulant matrix $\mathbf{\Psi}$ are asymptotically equivalent ($\mathbf{\Sigma}$ is the product of Toeplitz matrices) and $|\mathbf{E}| \leq 1$ for every realization of $\mathbf{E}$.

Because $\mathbf{E\Sigma E}$ and $\mathbf{E\Psi E}$ are asymptotically equivalent, the averages of any continuous function $f(\cdot)$ evaluated at their respective eigenvalues converge to the same quantity. In particular taking

$$f(x) = \frac{1}{1+\gamma x} \tag{267}$$

we obtain that for each realization of $\mathbf{E}$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\gamma\lambda_i(\mathbf{E\Sigma E})} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\gamma\lambda_i(\mathbf{E\Psi E})} \tag{268}$$

Averaging both sides of (268) with respect to $\mathbf{E}$ we obtain the desired equality (59). To show (60) we simply use $f(x) = \log(1 + \gamma x)$.

B Optimality of Stationary Inputs

Theorem 12 *The capacity*

$$C(\gamma) \;=\; \lim_{n\to\infty}\frac{1}{2n}\max_{\boldsymbol{\Sigma}_x}\mathbb{E}\left[\log det\left(\mathbf{I}+\gamma\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}\right)\right] \tag{269}$$

is achieved by circulant input covariance $\boldsymbol{\Sigma}_x$.

Proof: Denote the diagonal matrix of the eigenvalues of $\mathbf{H}$ by $\boldsymbol{\Lambda}_H$. Further, recall that $\mathbf{F}$ denotes the unitary discrete Fourier transform (DFT) matrix defined in (57). Define the circulant matrix

$$\tilde{\mathbf{H}} = \mathbf{F}\boldsymbol{\Lambda}_H\mathbf{F}^\dagger. \tag{270}$$

For any feasible choice of $\mathbf{E}$ and $\boldsymbol{\Sigma}_x$, the matrices $\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}$ and $\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\tilde{\mathbf{H}}^\dagger\mathbf{E}$ are asymptotically equivalent. Similarly to the proof of Lemma 1:

$$\begin{aligned}
& |\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}-\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\tilde{\mathbf{H}}^\dagger\mathbf{E}|^2 \\
= \;& |\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}-\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}+\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}-\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\tilde{\mathbf{H}}^\dagger\mathbf{E}|^2 \\
\leq \;& |\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}-\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}|^2+|\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\mathbf{H}^\dagger\mathbf{E}-\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\tilde{\mathbf{H}}^\dagger\mathbf{E}|^2 \\
\leq \;& 2|\mathbf{E}|^4|\boldsymbol{\Sigma}_x|^2|\mathbf{H}|^2|\tilde{\mathbf{H}}-\mathbf{H}|^2 \\
\to \;& 0
\end{aligned} \tag{271}$$

Thus, the result will follow if we can show the statement of the Lemma replacing $\mathbf{H}$ by $\tilde{\mathbf{H}}$ in (269).

Let $\mathbf{X}$ the elementary circulant permutation matrix, defined as

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & & 0 & 1 \\ 1 & \cdots & & 0 & 0 \end{bmatrix} \tag{272}$$

and let

$$\boldsymbol{\Pi}_\ell = \mathbf{X}^\ell \tag{273}$$

$$\boldsymbol{\Sigma}_x^{(\ell)} = \boldsymbol{\Pi}_\ell\boldsymbol{\Sigma}_x\boldsymbol{\Pi}_\ell^\dagger \tag{274}$$

Invoking the Jensen's inequality,

$$\mathbb{E}\left[\log\det\left(\mathbf{I}+\gamma\mathbf{E}\mathbf{H}\left(\frac{1}{n}\sum_{\ell=0}^{n}\boldsymbol{\Sigma}_x^{(\ell)}\right)\mathbf{H}^\dagger\mathbf{E}\right)\right] \;\geq\; \frac{1}{n}\sum_{\ell=0}^{n-1}\mathbb{E}\left[\log\det\left(\mathbf{I}+\gamma\mathbf{E}\mathbf{H}\boldsymbol{\Sigma}_x^{(\ell)}\mathbf{H}^\dagger\mathbf{E}\right)\right] \tag{275}$$

$$\begin{aligned}
&= \frac{1}{n}\sum_{\ell=0}^{n-1}\mathbb{E}\left[\log\det\left(\mathbf{I}+\gamma\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Pi}_\ell\boldsymbol{\Sigma}_x\boldsymbol{\Pi}_\ell^\dagger\tilde{\mathbf{H}}^\dagger\mathbf{E}\right)\right] \\
&= \frac{1}{n}\sum_{\ell=0}^{n-1}\mathbb{E}\left[\log\det\left(\mathbf{I}+\gamma\mathbf{E}\boldsymbol{\Pi}_\ell^\dagger\tilde{\mathbf{H}}\boldsymbol{\Pi}_\ell\boldsymbol{\Sigma}_x\boldsymbol{\Pi}_\ell^\dagger\tilde{\mathbf{H}}^\dagger\boldsymbol{\Pi}_\ell\mathbf{E}\right)\right]
\end{aligned} \tag{276}$$

$$= \mathbb{E}\left[\log\det\left(\mathbf{I}+\gamma\mathbf{E}\tilde{\mathbf{H}}\boldsymbol{\Sigma}_x\tilde{\mathbf{H}}^\dagger\mathbf{E}\right)\right] \tag{277}$$

where in (276) we have used the fact $\mathbf{E}$ is a diagonal matrix whose diagonal elements are i.i.d., while in (277) we have used the fact that $\tilde{\mathbf{H}}$ is circulant thus $\mathbf{\Pi}_\ell^\dagger \tilde{\mathbf{H}} \mathbf{\Pi}_\ell = \tilde{\mathbf{H}}$. Finally, since $\frac{1}{n}\sum_{\ell=0}^{n} \mathbf{\Sigma}_x^{(\ell)}$ is a circulant matrix with the same trace as $\mathbf{\Sigma}_x$ that achieves at least as good objective function the desired result is obtained. ■

C Auxiliary Lemma

Lemma 2 *[20, Eq. 3.112], [4] Let* $\mathbf{C}$ *be a* $n \times n$ *random matrix uniformly bounded in operator norm and having a limiting empirical spectral distribution. Let* $\mathbf{V}$ *be an* $n \times p$ *(with* $n \geq p$*) random matrix uniformly distributed over the Stiefel manifold of complex* $n \times p$ *matrices such that* $\mathbf{V}^\dagger \mathbf{V} = \mathbf{I}$*. Denote the* i*th column of* $\mathbf{CV}$ *by* $\mathbf{b}_i$*, and let* $\frac{p}{n} \to y$*. Then for all* $i = 1, \ldots, p$

$$\lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{C}\mathbf{V}\mathbf{V}^\dagger \mathbf{C}\right)^{-1} \mathbf{b}_i = \frac{1}{y}\left(1 - \eta_{\mathbf{C}\mathbf{V}\mathbf{V}^\dagger\mathbf{C}}(\gamma)\right) \tag{278}$$

almost surely, where $\eta_{\mathbf{C}\mathbf{V}\mathbf{V}^\dagger\mathbf{C}}(\gamma) = \eta$ *is the solution to:*

$$\eta = \eta_{\mathbf{C}\mathbf{C}^\dagger}\left(\gamma \frac{y - 1 + \eta}{\eta}\right) \tag{279}$$

D Proof of the monotonicity of $\daleth(\cdot, \gamma)$.

Let denote by y and $y + \delta$ the following ratios:

$$y = \lim_{n\to\infty} \frac{i}{n}, \qquad y + \delta = \lim_{n\to\infty} \frac{i+d}{n}.$$

where $d > 0$.

Denoting the j-th column of $\mathbf{CU}$ by $\mathbf{b}_j$ with $\mathbf{C}$ and $\mathbf{U}$ defined as in the proof of Theorem 1, let

$$\mathbf{B}_i \mathbf{B}_i^\dagger = \sum_{j=1}^{i} \mathbf{b}_j \mathbf{b}_j^\dagger \tag{280}$$

and

$$\begin{aligned} \mathbf{B}_{i+d}\mathbf{B}_{i+d}^\dagger &= \sum_{j=1}^{i+d} \mathbf{b}_j \mathbf{b}_j^\dagger \\ &= \mathbf{B}_i \mathbf{B}_i^\dagger + \sum_{j=i+1}^{i+d} \mathbf{b}_j \mathbf{b}_j^\dagger \qquad (281) \\ &= \mathbf{B}_i \mathbf{B}_i^\dagger + \mathbf{B}_d \mathbf{B}_d^\dagger \qquad (282) \end{aligned}$$

Using Lemma 2 and the inversion lemma [12], we have that

$$\begin{aligned} \lim_{n\to\infty} \gamma \mathbf{b}_{i+d}^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_{i+d}\mathbf{B}_{i+d}^\dagger\right)^{-1} \mathbf{b}_{i+d} &= \lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_{i+d}\mathbf{B}_{i+d}^\dagger\right)^{-1} \mathbf{b}_i \qquad (283) \\ &= \lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_i\mathbf{B}_i^\dagger + \gamma \mathbf{B}_d \mathbf{B}_d\right)^{-1} \mathbf{b}_i \qquad (284) \\ &= \lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_i\mathbf{B}_i^\dagger\right)^{-1} \mathbf{b}_i - \lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \mathbf{R} \mathbf{b}_i \qquad (285) \end{aligned}$$

where (283) follows from Lemma 2, (285) follows from the inversion lemma, and finally $\mathbf{R}$ is a definite positive matrix given by:

$$\mathbf{R} = \left(\mathbf{I} + \gamma \mathbf{B}_i \mathbf{B}_i^\dagger\right)^{-1} \mathbf{B}_d \left(\mathbf{I} + \gamma \mathbf{B}_d^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_i \mathbf{B}_i^\dagger\right)^{-1} \mathbf{B}_d\right)^{-1} \mathbf{B}_d^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_i \mathbf{B}_i^\dagger\right)^{-1}$$

From (285), it follows that

$$\begin{aligned} \frac{\beth(y+\delta,\gamma)}{1+\beth(y+\delta,\gamma)} &= \lim_{n\to\infty} \gamma \mathbf{b}_{i+d}^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_{i+d} \mathbf{B}_{i+d}^\dagger\right)^{-1} \mathbf{b}_{i+d} && (286) \\ &\le \lim_{n\to\infty} \gamma \mathbf{b}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{B}_i \mathbf{B}_i^\dagger\right)^{-1} \mathbf{b}_i && (287) \\ &= \frac{\beth(y,\gamma)}{1+\beth(y,\gamma)} && (288) \end{aligned}$$

from which it follows that

$$\beth(y+\delta,\gamma) \le \beth(y,\gamma) \tag{289}$$

for any $\delta \ge 0$.

E Combinatorial Definitions and Facts

Definition 4 *[19, 15] Let $\mathcal{X}$ be a set. An ℓ-partition of $\mathcal{X}$ is a set $\rho[\ell] = \{\mathcal{V}_1, \dots, \mathcal{V}_\ell\}$ of subsets of $\mathcal{X}$ such that*

$$\begin{aligned} \mathcal{V}_i \neq \emptyset \ \ \forall\, i = 1, \dots, \ell \\ \mathcal{V}_i \cap \mathcal{V}_j = \emptyset \ \ \forall\, i \neq j \\ \textstyle\bigcup_{i=1}^{\ell} \mathcal{V}_i = \mathcal{X} \end{aligned} \tag{290}$$

The elements $\mathcal{V}_i$ of $\rho[\ell]$ are called the blocks of the partition.

Definition 5 *Denote by $|\mathcal{X}|$ the cardinality of the set $\mathcal{X}$. Let $\mathbf{m} = [m_1, \dots, m_\ell]$ be an ℓ-dimensional vector whose entries are positive integers such that*

$$m_1 + m_2 + \ldots + m_\ell = |\mathcal{X}|$$

with $0 < m_1 \le m_2 \le \ldots m_\ell \le |\mathcal{X}|$.

An $(\mathbf{m}, \ell)$-partition of $\mathcal{X}$, denoted by $\rho[\mathbf{m}, \ell]$, is an ℓ-partition where the blocks have cardinalities $\mathbf{m} = [m_1, m_2, \dots m_\ell]$. The i-th block of $\rho[\mathbf{m}, \ell]$ shall be denoted by $\mathcal{V}(m_i)$.

In this paper, the set $\mathcal{X}$ with respect to which partitions are defined will always be the natural index set $\{1, 2, \dots, |\mathcal{X}|\}$. Furthermore, we define a partition of a vector $\mathbf{x}$ in the following way. Let $\mathbf{x}$ be a vector of $|\mathcal{X}|$ not necessarily distinct elements over some finite alphabet (to be defined later). For $\mathcal{V} \subseteq \mathcal{X}$, we denote by $\mathbf{x}(\mathcal{V})$ the *multi-set* (i.e., set with possibly repeated elements) given by

$$\mathbf{x}(\mathcal{V}) = \{x_j : j \in \mathcal{V}\}$$

Let $\rho[\ell]$ be an ℓ-partition of $\mathcal{X}$. The corresponding partition of $\mathbf{x}$ is briefly denoted by $\rho[\ell, \mathbf{x}]$. This is indeed the partition of the multiset $\{x_1, \dots, x_{|\mathcal{X}|}\}$ with blocks $\mathbf{x}(\mathcal{V}_i)$ for $i = 1, \dots, \ell\}$. Similarly, we shall use the notation $\rho[\mathbf{m}, \ell, \mathbf{x}]$ when we want to point out the cardinality of the blocks. We also adopt the convention that the vector $\mathbf{m}$ defining a $(\mathbf{m}, \ell)$-partition has non-decreasing non-zero entries, as in Definition 5.

We denote the set of all ℓ-partitions of $\mathcal{X}$ by $\mathfrak{P}_\ell(|\mathcal{X}|)$. Notice that $\mathfrak{P}_\ell(|\mathcal{X}|)$ is uniquely identified by the integers ℓ and $|\mathcal{X}|$. The set of all partitions of $\mathcal{X}$ will be denoted by $\mathfrak{P}(|\mathcal{X}|)$.

Lemma 3 *The number of $(\mathbf{m}, \ell)$-partitions of $\mathcal{X}$ is given by the Faá di Bruno coefficient, $\mathcal{P}(\mathbf{m}, \ell)$, [2]:*

$$\mathcal{P}(\mathbf{m}, \ell) = \frac{|\mathcal{X}|!}{m_1! \dots m_\ell! f(m_1, \dots, m_\ell)}$$

where we define

$$f(m_1, \dots, m_\ell) = f_1! \dots f_{|\mathcal{X}|}!$$

with f_i equal to the number of entries of the vector $\mathbf{m} = [m_1, \dots, m_\ell]$ equal to i.

Lemma 4 *The number of ℓ-partitions of $\mathcal{X}$ is the Stirling number of the second kind, $S(|\mathcal{X}|, \ell)$, [19] while the number of partitions of $\mathcal{X}$ is the $|\mathcal{X}|$-th Bell number,*

$$B_{|\mathcal{X}|} = \sum_{\ell=1}^{|\mathcal{X}|} S(|\mathcal{X}|, \ell)$$

.

E.1 Lattice of partitions and the degree of inclusion

The natural partial order relation for partitions is the *refinement order* $\rho \leq \sigma$ defined as follows:

Definition 6 *Given two partitions $\rho[\ell] = \{\mathcal{V}_1, \dots, \mathcal{V}_\ell\}$ and $\sigma[w] = \{\mathcal{U}_1, \dots, \mathcal{U}_w\}$ of $\mathcal{X}$, we say that $\rho[\ell]$ is finer than $\sigma[w]$, or, equivalently, that $\sigma[w]$ is coarser than $\rho[\ell]$, if for every $i = 1, \dots, \ell$ there exists $j = 1, \dots, w$ such that $\mathcal{V}_i \subset \mathcal{U}_j$. In other words, every element of $\rho[\ell]$ is a subset of some element of $\sigma[w]$. In this case, we write $\rho[\ell] \leq \sigma[w]$.*

When $\rho[\ell] \leq \sigma[w]$, but $\rho[\ell] \neq \sigma[w]$ (this condition is equivalent to $\ell > w$), then we write $\rho[\ell] < \sigma[w]$. If $\rho[\ell] < \sigma[w]$, but there does not exist any partition $\pi \in \mathfrak{P}(|\mathcal{X}|)$ such that $\rho[\ell] < \pi < \sigma[w]$, then we say that $\rho[\ell]$ covers $\sigma[w]$, and write $\rho[\ell] \prec \sigma[w]$. In this case, $\sigma[w]$ is an immediate successor to $\rho[\ell]$ in the hierarchy imposed by the ordering relation.

The coarsest element of $\mathfrak{P}(|\mathcal{X}|)$ corresponds to the unique 1-partition $\{\{1, \dots, |\mathcal{X}|\}\}$ while the finest element of $\mathfrak{P}(|\mathcal{X}|)$ is the unique $|\mathcal{X}|$-partition $\{\{1\}, \dots, \{|\mathcal{X}|\}\}$.

The set of partitions $\mathfrak{P}(|\mathcal{X}|)$ is a partially ordered set under the refinement ordering defined above. Furthermore, we can define two operations $\vee$ and $\wedge$ such that $\rho \vee \sigma$ is the finest partition π such that $\pi \geq \rho$ and $\pi \geq \sigma$ (least upper bound), and $\rho \wedge \sigma$ is the coarsest partition π such that $\pi \leq \rho$ and $\pi \leq \sigma$ (largest lower bound). $\mathfrak{P}(|\mathcal{X}|)$ is clearly closed under $\vee$ and $\wedge$. The refinement ordering relation $\leq$ is reflexive ($\rho \leq \rho$), antisymmetric (if $\rho \leq \sigma$ and $\sigma \leq \rho$, then $\rho = \sigma$) and transitive (if $\rho \leq \sigma$ and $\sigma \leq \tau$ then $\rho \leq \tau$). It can also be checked that for any $\rho, \sigma \in \mathfrak{P}(|\mathcal{X}|)$, $\rho \vee \sigma$ and $\rho \wedge \sigma$ are uniquely determined (that is, $\vee$ and $\wedge$ are properly defined operators $\mathfrak{P}(|\mathcal{X}|) \times \mathfrak{P}(|\mathcal{X}|) \to \mathfrak{P}(|\mathcal{X}|)$). Under these conditions, $\mathfrak{P}(|\mathcal{X}|)$ is a lattice (or algebra) with respect to the operations $\vee$ and $\wedge$.

The lattice of $\mathfrak{P}(|\mathcal{X}|)$ admits a graphical representation given by a graph called *Hasse diagram*, obtained as follows: for $\ell = 1, 2, \dots, |\mathcal{X}|$, draw layers of nodes such that each layer ℓ has $S(|\mathcal{X}|, \ell)$ nodes. Each node in layer ℓ represents a distinct partition with ℓ blocks. Then, an edge (ρ, σ) in the graph exists if and only if $\rho \prec \sigma$. Figure 12 shows an example of Hasse diagram for the set of partitions of $\mathcal{X} = \{1, 2, 3, 4\}$, which we use as a running example to illustrate various definitions and facts in the sequel.

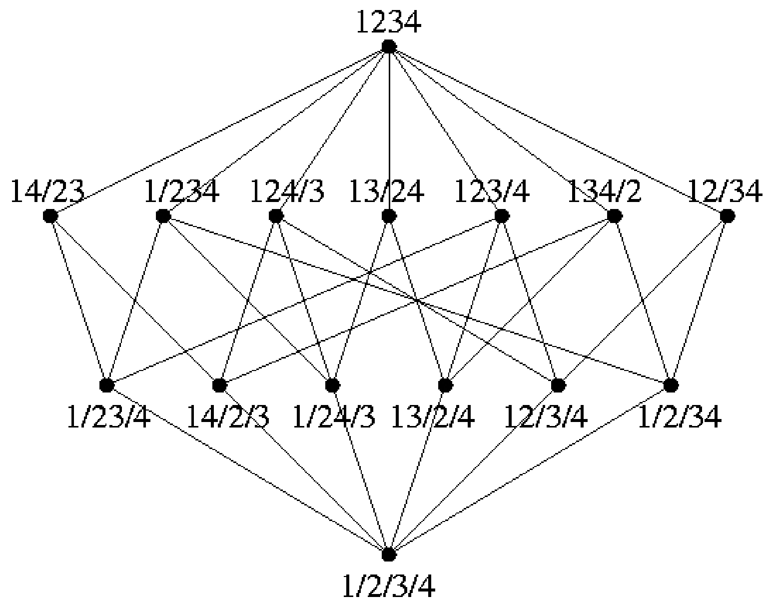

Figure 12: Hasse diagram of the partially ordered set $\mathfrak{P}(4)$ of partitions of $\mathcal{X} = \{1, 2, 3, 4\}$.

Example 12 *Referring to the set* $\{1, 2, 3, 4\}$ *of Figure 12, the number of* $([1, 3], 2)$*-partitions is*

$$\frac{4!}{1!3!f(1,3)} = 4$$

Example 13 *The number of* 2*-partitions of* $\mathcal{X} = \{1, 2, 3, 4\}$ *is* $S(4, 2) = 7$*, while the number of partitions of* $\mathcal{X}$ *is* $B_4 = \sum_{\ell=1}^{4} S(4, \ell) = 15$

Next, we introduce a function $\zeta : \mathfrak{P}(|\mathcal{X}|) \times \mathfrak{P}(|\mathcal{X}|) \to \mathbb{Z}$ that turns out to play a fundamental role in the proofs of the main results of the following appendices. This function, referred to as degree of inclusion, it is useful to carry out counting operations on the set of partitions.

Definition 7 *Consider two partitions* $\rho[\ell] < \sigma[w]$ *in* $(|\mathcal{X}|)$*. For any integer* $w \leq q \leq \ell$*, let*

$$[\rho, \sigma]_q = \{\pi \in \mathfrak{P}_q(|\mathcal{X}|) : \rho \leq \pi \leq \sigma\} \tag{291}$$

denote the set of q*-partitions between* $\rho[\ell]$ *and* $\sigma[\ell]$ *(with respect to the refinement ordering).*

The degree of inclusion ζ *maps pairs of partitions into the integers, and it is defined as:*

$$\zeta\left(\rho \to \sigma\right) = \begin{cases} 0 & \rho \not\leq \sigma \\ 1 & \rho = \sigma \\ -1 & w = \ell - 1, \text{ and } \rho < \sigma \end{cases} \tag{292}$$

and for $w < \ell - 1$ *with* $\rho < \sigma$

$$\zeta(\rho \to \sigma) = \sum_{a=w+1}^{\ell-1} \left(|[\rho, \sigma]_a| + \sum_{b=w+1}^{a-1} (-1)^{a-b} \sum_{\pi \in [\rho,\sigma]_{b+1}} |[\pi, \sigma]_b| \right) - 1 \tag{293}$$

The degree of inclusion can be easily computed from the Hasse diagram. In fact, interpreting the diagram as a directed graph where edges point upward, we notice that $\sum_{a=w+1}^{\ell-1} |[\rho,\sigma]_a|$ is equal to the total number of nodes in the sub-graph formed by all (directed) paths joining ρ with σ. Furthermore, for any $a \in \{w+2,\ldots,\ell-1\}$, and $b \in \{w+1,\ldots,a-1\}$, $\sum_{\pi\in[\rho,\sigma]_{b+1}} |[\pi,\sigma]_b|$ is given by the total number of edges pointing upward of the $(b+1)$-th layer in the sub-graph of the paths joining ρ with σ.

The degree of inclusion ζ satisfies the following additive decomposition:

$$\zeta(\rho \to \sigma) = -\sum_{a=w}^{\ell-1} \sum_{\pi\in[\rho,\sigma]_a} \zeta(\pi \to \sigma) \tag{294}$$

Example 14 *Referring to the diagram of Figure 12, we have that:*

$$\begin{aligned}
\zeta(\underbrace{\{\{1\};\{2\};\{3\};\{4\}\}}_{\rho} \to \underbrace{\{\{1,2,3,4\}\}}_{\sigma}) &= \sum_{a=2}^{3}\left(|[\rho,\sigma]_a| + \sum_{b=2}^{a-1}(-1)^{a-b}\sum_{\pi\in[\rho,\sigma]_{b+1}}|[\pi,\sigma]_b|\right) - 1 \\
&= \sum_{a=2}^{3}|[\rho,\sigma]_a| - \sum_{\pi\in[\rho,\sigma]_3}|[\pi,\sigma]_2| - 1 \\
&= 6+7-6\cdot 3-1 = -6
\end{aligned} \tag{295}$$

Next, we wish to check the validity of (294). We have one partition at layer 1, namely $\{\{1,2,3,4\}\}$. At layer 2 we have 7 partitions $\pi[2]$, with degree of inclusion $\zeta(\pi[2] \to \{\{1,2,3,4\}\}) = -1$. Then, we have 6 partitions $\pi[3]$ at layer 3. Their degree of inclusion is $\zeta(\pi[3] \to \{\{1,2,3,4\}\}) = 2$. In order to see this, notice that the sub-graph of partitions $\pi[3] \leq \pi \leq \{\{1,2,3,4\}\}$ consists has three intermediate nodes $\pi[2]$ and one top node . Hence, $\zeta(\pi[3] \to \{\{1,2,3,4\}\}) = -(1-3) = 2$. Eventually, using (294) we have

$$\zeta(\{\{1\},\{2\},\{3\},\{4\}\} \to \{\{1,2,3,4\}\}) = -(1+7\cdot(-1)+6\cdot 2) = -6 \tag{296}$$

which coincides with the previous direct calculation.

At this point it should be clear that if the refinement of σ to $\rho \leq \sigma$ involves the partition of a single block of σ into d blocks of ρ, then $\zeta(\rho \to \sigma)$ is uniquely determined by d. For example, any 2-way partition ($d = 2$) has $\zeta = \zeta(2) = -1$ (this corresponds to a single block of σ split into two blocks of ρ). Any 3-way partition has $\zeta = \zeta(3) = 2$ (this corresponds to a single block of σ split into three blocks of ρ). Any 4-way partition has $\zeta = \zeta(4) = -6$ (this corresponds to a single block of σ split into four blocks of ρ). It should be remarked that the graph corresponding to a d-way partition of a single block depends only on d (e.g., a 4-way partition has always the graph given in Figure 12), no matter how many other blocks (that do not split) ρ and σ have, and what the cardinality of the blocks is.

Clearly, there are refinements that involve the splitting of more than one block of the top partition. For example, consider the sub-graph of Figure 12 of all paths joining $\{\{1\},\{2\},\{3\},\{4\}\}$ (bottom) with $\{\{1,2\},\{3,4\}\}$ (top). In this case, the two blocks $\{1,2\}$ and $\{3,4\}$ of the top partition are split into two subblocks. The corresponding graph is obtained as the Cartesian product graph of two 2-way partitions, as shown in Figure 13.

In general, consider two nested partitions $\rho[\mathbf{m}, \ell] \leq \sigma[\mathbf{v}, w]$ such that each block $\mathcal{V}(v_i), i = 1, \ldots, w$, of $\sigma[\mathbf{v}, w]$, is partitioned into d_i blocks $\mathcal{U}(m_j)$ of $\rho[\mathbf{m}, \ell]$, with $j \in \left\{\sum_{h=1}^{i-1} d_h + 1, \ldots, \sum_{h=1}^{i} d_h\right\}$, with the consistency conditions:

$$\ell = \sum_{i=1}^{w} d_i,$$

and, for all $i = 1, \ldots, w$,

$$v_i = \sum_{j=\sum_{h=1}^{i-1} d_h+1}^{\sum_{h=1}^{i} d_h} m_j$$

It can be shown that ζ satisfies the following multiplicative decomposition:

$$\zeta(\rho[\mathbf{m}, \ell] \rightarrow \sigma[\mathbf{v}, w]) = \prod_{i=1}^{w} \zeta(d_i) \tag{297}$$

where $\zeta(d_i)$ is the degree of inclusion for a d_i-way partition.

The sum and product rules (294) and (297) allow very simple recursive computation of the inclusion index.

Example 15 *Referring to the diagram of Figure 13, direct calculation shows that*

$$\zeta(\{\{1\}, \{2\}, \{3\}, \{4\}\} \rightarrow \{\{1,2\}, \{3,4\}\}) = 2 - 1 = 1$$

Using the product rule we have

$$\begin{aligned} \zeta(\{\{1\}, \{2\}, \{3\}, \{4\}\} \rightarrow \{\{1,2\}, \{3,4\}\}) &= \zeta(\{\{1\}, \{2\}\} \rightarrow \{\{1,2\}\})\zeta(\{\{3\}, \{4\}\} \rightarrow \{\{3,4\}\}) \\ &= (-1)(-1) = 1 \end{aligned}$$

A more involved example is given in Figure 14: consider partitions

$$\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\} < \{\{1,2,3\}, \{4,5\}\}.$$

The first is obtained by a 3-way partition of the block $\{1,2,3\}$ *and a 2-way partition of the block* $\{1,2\}$ *of the second. Hence, the inclusion index is readily given by* $\zeta(3)\zeta(2) = 2(-1) = -2$. *The corresponding Hasse diagram of Figure 14 is obtained as the Cartesian product of a 3-way and a 2-way partition. One can check by direct calculation that, indeed,*

$$\zeta(\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\} \rightarrow \{\{1,2,3\}, \{4,5\}\}) = -2$$

E.2 Good partitions

From now on we will consider vectors $\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|}$, where $\mathbb{Z}_n$ denotes the ring of integer residues modulo n. This implies that all operations on the elements of $\mathbf{x}$ are defined in the $\mathbb{Z}_n$ arithmetic.

Next, we introduce a special type of partitions that we refer to as *good* ℓ-partition.

Definition 8 *Fix* $\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|}$. *We say that a* $(\mathbf{m}, \ell)$*-partition* $\rho[\mathbf{m}, \ell] = \{\mathcal{V}(m_1), \ldots, \mathcal{V}(m_\ell)\}$ *of* $\mathcal{X}$ *is a good partition of* $\mathbf{x}$ *if*

$$\mathsf{sum}[\mathbf{x}(\mathcal{V}(m_r))] \triangleq \sum_{j \in \mathcal{V}(m_r)} x_j = 0 \qquad \forall r = 1, \ldots, \ell \tag{298}$$

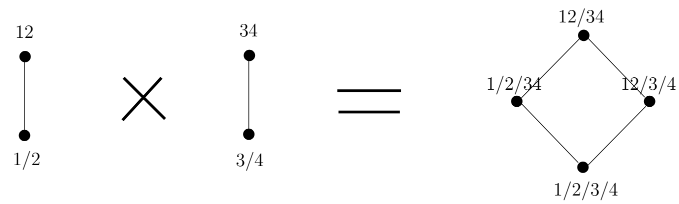

Figure 13: Hasse diagram of the 2×2-way partition refinement from $\{\{1,2\},\{3,4\}\}$ to $\{\{1\},\{2\},\{3\},\{4\}\}$.

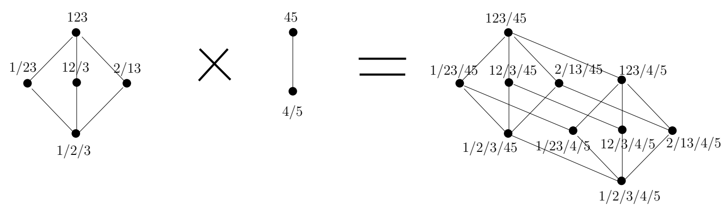

Figure 14: Hasse diagram of the 3×2-way partition refinement from $\{\{1,2,3\},\{4,5\}\}$ to $\{\{1\},\{2\},\{3\},\{4\},\{5\}\}$.

Example 16 *The vector* $\mathbf{x} = (1, n-1, 0, n-2, 2)$ *with elements in* $\mathbb{Z}_n$ *has four* good *partitions:*

- *one* good $([5], 1)$*-partition:* $\{1,2,3,4,5\}$*;*
- *one* $([1,2,2], 3)$*-partition:* $\{\{3\},\{1,2\},\{4,5\}\}$;
- *two* $([2,3], 2)$*-partitions:*

$$\{\{1,2,3\},\{4,5\}\}; \quad \{\{1,2\},\{3,4,5\}\}.$$

Example 17 *The vector* $\mathbf{x} = (1,1,0,2,4)$ *with elements in* $\mathbb{Z}_7$ *has no* good *partitions.*

The condition that $\rho[\mathbf{m}, \ell]$ is a good partition of $\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|}$ is equivalent to say that $\mathbf{x}$ lies in the solution space of a linear equation over $\mathbb{Z}_n$. In particular, the partition $\rho[\mathbf{m}, \ell]$ is associated to the incidence matrix $\mathbf{A}_\rho$ with $|\mathcal{X}|$ rows and ℓ columns, such that the i-th column of $\mathbf{A}_\rho$ contains 1s for all positions $j \in \mathcal{V}(m_i)$ and 0s elsewhere. $\rho[\mathbf{m}, \ell]$ is a good partition of $\mathbf{x}$ if and only if $\mathbf{x}$ is a solution of the linear equation $\mathbf{x}\mathbf{A}_\rho = \mathbf{0}$. In other words, $\mathbf{x}$ is an element of the kernel of $\mathbf{A}_\rho$. The kernel of a linear transformation over the ring $\mathbb{Z}_n$ is a $\mathbb{Z}_n$-module. For later use, we wish to find its cardinality. The coefficients in each linear equation of the type (298) are either 0 or 1. Hence, the greatest common divisor of each equation is 1, which is obviously relatively prime with n. Hence, the solution space of each r-th equation in (298) is isomorphic[4] to $\mathbb{Z}_n^{m_r - 1}$. Furthermore, by definition of partition it follows that the columns of $\mathbf{A}_\rho$

[4]Notice that while this is a completely trivial conclusion if $\mathbb{Z}_n$ was a field, the condition that the coefficients of the equation are relatively prime with n is important in a ring that has zero divisors, as in the case where n is not a prime. For example, if $n = 8$ the equation $x + y = 0$ has 8 solutions, but the equation $4x + 2y = 0$ 16 solutions.

are mutually orthogonal (in fact, they have disjoint support corresponding to the disjoint blocks $\{\mathcal{V}_r : r = 1, \dots, \ell\}$ of the partition $\rho[\mathbf{m}, \ell]$). This implies that the kernel of $\mathbf{A}_\rho$ is isomorphic to the Cartesian product

$$\mathbb{Z}_n^{m_1-1} \times \mathbb{Z}_n^{m_2-1} \times \cdots \times \mathbb{Z}_n^{m_\ell-1} \tag{299}$$

It follows that $|\mathrm{Ker}(\mathbf{A}_\rho)| = n^{|\mathcal{X}|-\ell}$ depends on the partition only through the number of blocks ℓ. We have the following result:

Lemma 5 *For two partitions, $\rho[\mathbf{m}, \ell]$ and $\sigma[\mathbf{v}, w]$ of $\mathcal{X}$, we have that $\mathrm{Ker}(\mathbf{A}_\rho) \subseteq \mathrm{Ker}(\mathbf{A}_\sigma)$ if an only if $\rho[\mathbf{m}, \ell] \le \sigma[\mathbf{v}, w]$.*

Proof: Suppose that $\rho[\mathbf{m}, \ell] \le \sigma[\mathbf{v}, w]$. Hence, each block σ is partitioned into blocks of ρ. Consider a block $\mathcal{U}_j$ of σ and, without loss of generality, let $\mathcal{V}_{i_1}, \dots, \mathcal{V}_{i_{r_j}}$ denote the blocks of ρ that partition $\mathcal{U}_j$. For any $\mathbf{x} \in \mathrm{Ker}(\mathbf{A}_\rho)$ it follows that

$$\mathsf{sum}[\mathbf{x}(\mathcal{U}_j)] = \sum_{h=1}^{r_j} \mathsf{sum}[\mathbf{x}(\mathcal{V}_{i_h})] = 0$$

Hence, $\mathbf{x} \in \mathrm{Ker}(\mathbf{A}_\sigma)$. This shows sufficiency. In order to show necessity, without loss of generality suppose that $\ell \ge w$, $\rho[m, \ell] \not\le \sigma[\mathbf{v}, w]$. There must exist a block $\mathcal{V}$ of ρ with non empty intersection with at least two blocks of σ (otherwise, ρ would be a refinement of σ). Denote these blocks as $\mathcal{U}$ and $\mathcal{U}'$. We choose a vector $\mathbf{x} \in \mathrm{Ker}(\mathbf{A}_\rho)$ such that all components are equal to zero but two non-zero components, $x_i = 1$ for $i \in \mathcal{V} \cap \mathcal{U}$ and $x_{i'} = -1$ for $i' \in \mathcal{V} \cap \mathcal{U}'$. Clearly, $\mathbf{x} \notin \mathrm{Ker}(\mathbf{A}_\sigma)$. This shows that $\mathrm{Ker}(\mathbf{A}_\rho) \not\subseteq \mathrm{Ker}(\mathbf{A}_\sigma)$. ■

In the proofs of the main results of the following appendices, we shall deal with vectors $\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|}$ such that $\mathsf{sum}[\mathbf{x}] = 0$. The partition $\rho[1] = \{\{1, \dots, |\mathcal{X}|\}\}$ is a good partition of such vectors. Since any partition $\rho[\mathbf{m}, \ell]$ is a refinement of $\rho[1]$, Lemma 5 yields that $\mathrm{Ker}(\mathbf{A}_{\rho[\mathbf{m},\ell]}) \subseteq \mathrm{Ker}(\mathbf{A}_{\rho[1]})$. It follows that

$$\left|\mathrm{Ker}(\mathbf{A}_{\rho[\mathbf{m},\ell]}) \cap \mathrm{Ker}(\mathbf{A}_{\rho[1]})\right| = \left|\mathrm{Ker}(\mathbf{A}_{\rho[\mathbf{m},\ell]})\right| = n^{|\mathcal{X}|-\ell} \tag{300}$$

E.3 Partitions of concatenated vectors

Given s vectors $\mathbf{x}_1, \dots, \mathbf{x}_s$ of possibly different dimension $|\mathcal{X}_1|, \dots, |\mathcal{X}_s|$, respectively, we define the concatenation $(\mathbf{x}_1|\dots|\mathbf{x}_s)$. The vectors $\mathbf{x}_1, \dots, \mathbf{x}_s$ are referred to as the *components* of the concatenation. The corresponding index set of $\mathbf{x}$ shall be denoted by $\mathcal{X}_1^s$ and has cardinality $|\mathcal{X}_1^s| = \sum_{i=1}^s |\mathcal{X}_i|$. For a partition $\varpi[\boldsymbol{\nu}, \iota]$ of the index set $\{1, \dots, s\}$ with blocks $\varpi(\nu_1), \dots, \varpi(\nu_\iota)$, we define the induced ι-partition of the concatenation $(\mathbf{x}_1|\dots|\mathbf{x}_s)$ as the partition $\rho[\mathbf{m}, \iota, (\mathbf{x}_1|\dots|\mathbf{x}_s)]$ with blocks

$$\mathbf{x}^{\varpi(\nu_j)} \triangleq \{\mathbf{x}_i : i \in \mathcal{V}(\nu_j)\}. \tag{301}$$

With some abuse of notation, the induced partition shall be denoted by $\rho[\varpi[\boldsymbol{\nu}, \iota], (\mathbf{x}_1|\dots|\mathbf{x}_s)]$. The blocks (301) of the induced partition $\rho[\varpi[\boldsymbol{\nu}, \iota], (\mathbf{x}_1|\dots|\mathbf{x}_s)]$ are called "groups". Notice that these blocks are unions (in the sense of multi-sets, as there might be repeated elements) of components of $(\mathbf{x}_1|\dots|\mathbf{x}_s)$. The index set of the components of the group $\mathbf{x}^{\varpi(\nu_j)}$, a subset of $\mathcal{X}_1^s$, is also referred to as a "group", and it is briefly indicated by $\mathcal{X}^{\varpi(\nu_j)}$.

Example 18 *Let* $\mathbf{x}_1 = (x_1)$, $\mathbf{x}_2 = (x_2)$ *and* $\mathbf{x}_3 = (x_3, x_4)$ *and their concatenation* $(\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3) = (x_1|x_2|(x_3, x_4))$. *Consider the partition*

$$\varpi[[1,2],2] = \{\{2\},\{1,3\}\},$$

of $\{1,2,3\}$. *The induced* 2*-partition of* $(x_1|x_2|(x_3,x_4))$ *is:*

$$\rho[\{\{2\},\{1,3\}\},2,(x_1|x_2|(x_3,x_4))] = \{\{x_2\},\{x_1,x_3,x_4\}\}$$

Definition 9 *Let* $\mathbf{x} = (\mathbf{x}_1|\ldots|\mathbf{x}_s)$ *be a concatenation of* s *vectors as defined above. For a given* $(\mathbf{m},\ell)$*-partition* $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$, *let*

$$\iota^* = arg\max_{\iota}\{1 \le \iota \le s : \rho[\varpi[\boldsymbol{\nu},\iota],(\mathbf{x}_1|\ldots|\mathbf{x}_s)] \ge \sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]\} \tag{302}$$

Then, there is a unique partition $\varpi[\boldsymbol{\nu}^*,\iota^*]$ *of* $\{1,\ldots,s\}$ *such that* $\rho[\varpi[\boldsymbol{\nu}^*,\iota^*],(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ *includes* $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$. *We refer to this unique partition* $\varpi[\boldsymbol{\nu}^*,\iota^*]$ *and to* ι^* *as the* canonical decomposition *and the* order of the canonical decomposition *of* $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$.

Example 19 *Consider the same concatenation of Example 18, and consider the* $([1,1,2],3)$*-partition:*

$$\sigma[[1,1,2],3,(x_1|x_2|(x_3,x_4))] = \{\{x_2\},\{x_4\},\{x_1,x_3\}\}$$

The partitions $\varpi[\boldsymbol{\nu},\iota]$, *of* $\{1,2,3\}$ *such that* $\rho[\varpi[\boldsymbol{\nu},\iota],(x_1|x_2|(x_3,x_4))] \ge \sigma[[1,1,2],3,(x_1|x_2|(x_3,x_4))]$ *are:*

$$\{1,2,3\} \quad : \quad \textit{for which} \quad \rho[\{1,2,3\},(x_1|x_2|(x_3,x_4))] = \{\{x_1,x_2,x_3,x_4\}\} \tag{303}$$

$$\{\{2\},\{1,3\}\} \quad : \quad \textit{for which} \quad \rho[\{\{2\},\{1,3\}\},(x_1|x_2|(x_3,x_4))] = \{\{x_2\},\{x_1,x_3,x_4\}\}. \tag{304}$$

Thus, the canonical decomposition of $\{\{x_2\},\{x_4\},\{x_1,x_3\}\}$ *is:*

$$\varpi[\boldsymbol{\nu}^*,\iota^*] = \{\{2\},\{1,3\}\},$$

and its order is $\iota^* = 2$.

Definition 10 *Consider the concatenation* $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$. *Let* $\varpi[\boldsymbol{\nu}^*,\iota^*]$ *be the* canonical decomposition *of a partition* $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$, *and let* $\rho[\varpi[\boldsymbol{\nu}^*,\iota^*],(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ *denote the induced partition of* $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$, *with groups* $\mathbf{x}^{\varpi(\nu_1^*)},\ldots,\mathbf{x}^{\varpi(\nu_{\iota^*}^*)}$. *The canonical representation of* $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ *is:*

$$\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)] = \{\sigma[\mathbf{m}_1,\ell_1,\mathbf{x}^{\varpi(\nu_1^*)}],\ldots,\sigma[\mathbf{m}_{\iota^*},\ell_{\iota^*},\mathbf{x}^{\varpi(\nu_{\iota^*}^*)}]\} \tag{305}$$

where $\sum_{i=1}^{\iota^*}\ell_i = \ell$ *and the vectors* $\mathbf{m}_i$, *for* $i = 1,\ldots,\iota^*$, *have dimensions* ℓ_i *and components that sum to the* i*-th group cardinality* $|\mathbf{x}^{\varpi(\nu_i^*)}|$, *respectively.*

This implies that, for a given concatenation $\mathbf{x} = (\mathbf{x}_1|\ldots|\mathbf{x}_s)$ of s vectors, $\sigma[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ is isomorphic to some Cartesian-product partition in

$$\mathfrak{P}(|\mathcal{X}^{\varpi(\nu_1^*)}|) \times \ldots \times \mathfrak{P}(|\mathcal{X}^{\varpi(\nu_{\iota^*}^*)}|) \tag{306}$$

It should be clear that both concepts of *canonical decomposition* and *canonical representation* are relative to a given concatenation of vectors. In other words, a given partition τ of the index set of size k may admit different canonical decompositions and representations with respect to two different concatenations $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$ and $(\mathbf{x}'_1|\ldots|\mathbf{x}'_r)$ of the same total length k.

Example 20 *Referring to Example 19, we have that the canonical decomposition of* $\{\{x_2\},\{x_4\},\{x_1,x_3\}\}$ *w.r.t. the concatenation* $(x_1|x_2|(x_3,x_4))$ *is:*

$$\varpi[\boldsymbol{\nu}^*,\iota^*] = \{\{2\},\{1,3\}\}.$$

It follows that the corresponding canonical representation of $\{\{x_2\},\{x_4\},\{x_1,x_3\}\}$ *is given by:*

$$\{\{x_2\},\{x_4\},\{x_1,x_3\}\} = \{\sigma[[1],1,(x_2)],\sigma[[1,2],2,(x_1,x_3,x_4)]\}$$

with $\sigma[[1],1,(x_2)],=\{x_2\}$ *and* $\sigma[[1,2],2,(x_1,x_3,x_4)] = \{\{x_4\},\{x_1,x_3\}\}$.

Definition 11 *Let the disjoint sets* $\mathfrak{D}^{(i)}(|\mathcal{X}_1^s|)$ *denote the set of partitions* $\sigma[\mathbf{m},\ell]$ *of* $\mathcal{X}_1^s$ *whose order of the canonical decomposition relative to the concatenation* $\mathbf{x} = (\mathbf{x}_1|\ldots|\mathbf{x}_s)$ *is* $\iota^* = i$.

Since the canonical decomposition index is uniquely defined, it is clear that $\mathfrak{D}^{(i)}(|\mathcal{X}_1^s|) \cap \mathfrak{D}^{(j)}(|\mathcal{X}_1^s|) = \emptyset$ for $i \neq j$, and that $\cup_{i=1}^s \mathfrak{D}^{(i)}(|\mathcal{X}_1^s|) = \mathfrak{P}(|\mathcal{X}_1^s|)$. In words, the collection of sets of partitions $\mathfrak{D}^{(i)}(|\mathcal{X}_1^s|)$ for $i = 1,\ldots,s$ forms a partition of the set of all partitions of the concatenated index set $\mathcal{X}_1^s$.

Canonical decomposition and canonical representation allow us to establish a generalization of (300) for concatenated vectors such that each component has zero sum. We have the following:

Lemma 6 *Consider concatenated vectors* $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$ *with equal-length components* $\mathbf{x}_i \in \mathbb{Z}_n^{|\mathcal{X}|}$, *for some integer* $|\mathcal{X}|$. *Consider a partition* $\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ *and let* $\varpi(\boldsymbol{\nu}^*,\iota^*)$ *denote its canonical decomposition with respect to* $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$. *Denote by* $\mathbb{G}(\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)])$ *the set of such concatenated vectors satisfying the conditions* $\mathsf{sum}[\mathbf{x}_i] = 0$ *for all* $i = 1,\ldots,s$ *and for which* $\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ *is a good partition. Then,*

$$|\mathbb{G}(\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)])| = n^{s|\mathcal{X}|-\ell-s+\iota^*} \tag{307}$$

Proof: Consider the partition $\rho[\varpi[\boldsymbol{\nu}^*,\iota^*],(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ induced by $\varpi(\boldsymbol{\nu}^*,\iota^*)$. By definition of canonical decomposition, $\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ is a refinement of $\rho[\varpi[\boldsymbol{\nu}^*,\iota^*],(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$. Hence, the blocks of $\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ are sub-blocks of the groups $\mathbf{x}^{\mathcal{V}(\nu_i^*)}$, for $i = 1,\ldots,\iota^*$. Let $\mathbf{A}_\rho$ denote the incidence matrix (as previously defined) of ρ. The canonical representation of $\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ implies that we can partition the columns of $\mathbf{A}_\rho$ into mutually orthogonal sets of ℓ_i columns, for $i = 1,\ldots,\iota^*$, where each set corresponds to the ℓ_i subblocks of $\mathbf{x}^{\mathcal{V}(\nu_i^*)}$.

Next, consider the partition

$$\sigma[[|\mathcal{X}|,\ldots,|\mathcal{X}|],s,(\mathbf{x}_1|\ldots|\mathbf{x}_s)] = \{\{\mathbf{x}_1\},\ldots,\{\mathbf{x}_s\}\}$$

with blocks that coincide with the components of the concatenation. It is clear that the zero-sum condition $\mathsf{sum}[\mathbf{x}_i] = 0$ for all $i = 1,\ldots,s$ defines the kernel of $\mathbf{A}_\sigma$. Therefore, we have that

$$\mathbb{G}(\rho[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)]) = \mathrm{Ker}(\mathbf{A}_\rho) \cap \mathrm{Ker}(\mathbf{A}_\sigma) \tag{308}$$

Notice also that $\mathbf{A}_\sigma$ admits a corresponding partition into ι^* mutually orthogonal set of columns of size ν_i^*. Let $\mathbf{A}_\rho^{\ell_i}$ and $\mathbf{A}_\sigma^{\nu_i^*}$ denote the submatrices obtained from these mutually orthogonal sets of columns, respectively. Because of the mutual orthogonality, it follows that

$$|\mathrm{Ker}(\mathbf{A}_\rho) \cap \mathrm{Ker}(\mathbf{A}_\sigma)| = \prod_{i=1}^{\iota^*} \left|\mathrm{Ker}(\mathbf{A}_\rho^{\ell_i}) \cap \mathrm{Ker}(\mathbf{A}_\sigma^{\nu_i^*})\right| \tag{309}$$

Next, we evaluate each term in the above product. Consider the i-th term. By definition of canonical decomposition, it follows that any sum of $h < \nu_i^*$ columns of $\mathbf{A}_\sigma^{\nu_i^*}$ is linearly independent of the columns of $\mathbf{A}_\rho^{\ell_i}$. In fact, since the columns of $\mathbf{A}_\rho^{\ell_i}$ are mutually orthogonal (recall that the blocks of a partition are disjoint sets), we can write a column of $\mathbf{A}_\sigma^{\nu_i^*}$ as a linear combination of columns of $\mathbf{A}_\rho^{\ell_i}$ if and only if there exists a set of blocks of $\rho[\mathbf{m}, \ell, (\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ that forms a partition of the union of $h < \nu_i^*$ components of the concatenation $(\mathbf{x}_1|\ldots|\mathbf{x}_s)$. But if this was true, then the canonical decomposition of $\rho[\mathbf{m}, \ell, (\mathbf{x}_1|\ldots|\mathbf{x}_s)]$ would contain an additional group and the degree would be larger than ι^*, contradicting the definition of ι^*. Hence, we conclude that the rank of the concatenated matrix $[\mathbf{A}_\rho^{\ell_i}|\mathbf{A}_\sigma^{\nu_i^*}]$ is at least $\ell_i + \nu_i^* - 1$. We can show that the rank is indeed exactly equal to $\ell_i + \nu_i^* - 1$ by noticing that the sum of all columns of $\mathbf{A}_\rho^{\ell_i}$ is equal to the sum of all columns of $\mathbf{A}_\sigma^{\nu_i^*}$ and it coincides with the all-one vector, by construction. Hence, the columns of $[\mathbf{A}_\rho^{\ell_i}|\mathbf{A}_\sigma^{\nu_i^*}]$ are linearly dependent over $\mathbb{Z}_n$. It follows that

$$\left|\mathrm{Ker}(\mathbf{A}_\rho^{\ell_i}) \cap \mathrm{Ker}(\mathbf{A}_\sigma^{\nu_i^*})\right| = n^{\nu_i^*|\mathcal{X}| - \ell_i - \nu_i^* + 1} \tag{310}$$

Using (310) into (309) and recalling that $\sum_{i=1}^{\iota^*} \ell_i = \ell$ and $\sum_{i=1}^{\iota^*} \nu_i^* = s$, we obtain (307). ∎

As said before, notice that for $s = 1$ we have $\mathbb{G}(\rho[\mathbf{m}, \ell, \mathbf{x}]) = \mathrm{Ker}(\mathbf{A}_{\rho[\mathbf{m},\ell]}) \cap \mathrm{Ker}(\mathbf{A}_{\rho[1]})$ and we obtain (300) as a special case of Lemma 6, since in this case $\iota^* = 1$.

F Partitions and equivalence relations

A partition $\rho[\ell]$ of the index set $\mathcal{X}$ induces an equivalence relation on the elements of $\mathcal{X}$. In particular, we say that two indices $i, j \in \mathcal{X}$ are equivalent (and write $i \sim j$) if they belongs to the same block of $\pi[\ell]$. Similarly, for a vector $\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|}$ we have the equivalence relation $x_i \sim x_j$ if elements x_i and x_j belongs to the same block of $\rho[\ell, \mathbf{x}]$.

Definition 12 *We define $\mathbb{S}(\rho[\ell])$ to be the subset of $\mathbb{Z}_n^{|\mathcal{X}|}$ of all vectors that are constant over the blocks of $\rho[\ell]$, i.e.,*

$$\mathbb{S}(\rho[\ell]) = \left\{\mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|} : x_i = x_j \;\; \textit{if} \;\; i \sim j\right\} \tag{311}$$

Lemma 7 *Consider two partitions $\rho[\ell]$ and $\sigma[w]$ of $\mathcal{X}$ with $w \leq \ell$. Then, $\mathbb{S}(\sigma[w]) \subseteq \mathbb{S}(\rho[\ell])$ if an only if $\sigma[w] \geq \rho[\ell]$.*

Consequently, we can define a partial order on the set:

$$\mathfrak{S}(|\mathcal{X}|) = \{\mathbb{S}(\rho[\ell]) : \rho[\ell] \in \mathfrak{P}(|\mathcal{X}|)\},$$

with respect to the inclusion relation.

The (unique) smallest and biggest elements of $\mathfrak{S}(|\mathcal{X}|)$ correspond to the biggest and smallest elements of $\mathfrak{P}(|\mathcal{X}|)$, i.e., the 1-partition, $\{\{1, \ldots, |\mathcal{X}|\}\}$ and the $|\mathcal{X}|$-partition, $\{\{1\}, \ldots, \{|\mathcal{X}|\}\}$. For these sets we adopt the short-hand notation $\mathbb{S}_1$ and $\mathbb{S}_{|\mathcal{X}|}$, respectively. Notice that $\mathbb{S}_1$ is the set of constant vectors, and has cardinality n, and $\mathbb{S}_{|\mathcal{X}|}$ is the set of all vectors (i.e., it coincides with $\mathbb{Z}_n^{|\mathcal{X}|}$, and has cardinality $n^{|\mathcal{X}|}$.

Definition 13 *We define* $\mathbb{S}^-(\rho[\ell])$ *to be the subset of* $\mathbb{Z}_n^{|\mathcal{X}|}$ *of all vectors that are constant over the blocks of* $\rho[\ell]$ *but take on distinct values in different blocks, i.e.,*

$$\mathbb{S}^-(\rho[\ell]) = \left\{ \mathbf{x} \in \mathbb{Z}_n^{|\mathcal{X}|} : x_i = x_j \;\; \textit{if} \;\; i \sim j, \;\; \textit{otherwise} \;\; x_i \neq x_j \right\} \tag{312}$$

It is easy to see that

$$\mathbb{S}^-(\rho[\ell]) = \mathbb{S}(\rho[\ell]) - \bigcup_{\sigma[\ell-1] > \rho[\ell]} \mathbb{S}(\sigma[\ell-1]) \tag{313}$$

$$= \mathbb{S}(\rho[\ell]) - \bigcup_{w=1}^{\ell-1} \bigcup_{\sigma[w] > \rho[\ell]} \mathbb{S}^-(\sigma[w]) \tag{314}$$

The sets $\mathbb{S}^-(\rho[\ell])$ are disjoint, in fact, for all $\ell, w \in \{1, \ldots, |\mathcal{X}|\}$ and corresponding distinct partitions $\rho[\ell] \neq \sigma[w]$ we have

$$\mathbb{S}^-(\rho[\ell]) \cap \mathbb{S}^-(\sigma[w]) = \emptyset \tag{315}$$

Furthermore, we have that their union exhausts the whole $\mathbb{Z}_n^{|\mathcal{X}|}$, i.e.,

$$\mathbb{S}_{|\mathcal{X}|} = \bigcup_{\rho \in \mathfrak{P}(|\mathcal{X}|)} \mathbb{S}^-(\rho) \tag{316}$$

Therefore, the set $\{\mathbb{S}^-(\rho) : \rho \in \mathfrak{P}(|\mathcal{X}|)\}$ is a partition of the set of all vectors $\mathbb{Z}_n^{|\mathcal{X}|}$.

Lemma 8 *Consider a function* $f : \mathbb{Z}_n^{|\mathcal{X}|} \to \mathbb{C}$*, and a fixed partition* $\rho[\ell]$ *of* $\mathcal{X}$*. Then,*

$$\sum_{\mathbf{x} \in \mathbb{S}^-(\rho[\ell])} f(\mathbf{x}) = \sum_{w=1}^{\ell} \sum_{\sigma[w] \geq \rho[\ell]} \zeta(\rho[\ell] \to \sigma[w]) \sum_{\mathbf{x} \in \mathbb{S}(\sigma[w])} f(\mathbf{x}) \tag{317}$$

where $\zeta(\rho[\ell] \to \sigma[w])$ *is the degree of inclusion, defined in (7).*

Proof: The proof is by induction. For $\ell = 1$, (317) follows immediately from the facts that $\mathbb{S}_1^-(\rho[1]) = \mathbb{S}_1$, $\zeta(\rho \to \rho) = 1$, and the sum over σ contains only the term $\sigma[1] = \rho[1]$.

Now let us assume that (317) holds for all $1 \leq h \leq \ell - 1$. We wish to show that it holds also for ℓ. Using (314) and (315) we have:

$$\begin{aligned}
\sum_{\mathbf{x}\in\mathbb{S}^-(\rho[\ell])} f(\mathbf{x}) &= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) - \sum_{h=1}^{\ell-1}\sum_{\sigma[h]\geq\rho[\ell]}\sum_{\mathbf{x}\in\mathbb{S}^-(\sigma[h])} f(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) - \sum_{h=1}^{\ell-1}\sum_{\sigma[h]\geq\rho[\ell]}\sum_{w=1}^{h}\sum_{\tau[w]\geq\sigma[h]} \zeta(\sigma[h]\to\tau[w]) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) - \sum_{h=1}^{\ell-1}\sum_{w=1}^{\ell-1}\sum_{\sigma[h]\geq\rho[\ell]}\sum_{\tau[w]\geq\sigma[h]} \zeta(\sigma[h]\to\tau[w]) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) - \sum_{h=1}^{\ell-1}\sum_{w=1}^{\ell-1}\sum_{\tau[w]\geq\rho[\ell]}\sum_{\sigma\in[\rho[\ell],\tau[w]]_h} \zeta(\sigma\to\tau[w]) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \qquad (318) \\
&= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) - \sum_{w=1}^{\ell-1}\sum_{\tau[w]\geq\rho[\ell]} \left(\sum_{h=1}^{\ell-1}\sum_{\sigma\in[\rho[\ell],\tau[w]]_h} \zeta(\sigma\to\tau[w])\right) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \\
&= \sum_{\mathbf{x}\in\mathbb{S}(\rho[\ell])} f(\mathbf{x}) + \sum_{w=1}^{\ell-1}\sum_{\tau[w]\geq\rho[\ell]} \zeta(\rho[\ell]\to\tau[w]) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \qquad (319) \\
&= \sum_{w=1}^{\ell}\sum_{\tau[w]\geq\rho[\ell]} \zeta(\rho[\ell]\to\tau[w]) \sum_{\mathbf{x}\in\mathbb{S}(\tau[w])} f(\mathbf{x}) \qquad (320)
\end{aligned}$$

where (318) follows by changing the summation order, (319) follows from (294) and (320) follows from the definition of ζ. ■

G Fourier coefficients of stationary processes

We summarize some of the statistical properties of the Fourier coefficients c_ℓ of a stationary process. Let $\{\mathsf{X}_p; p \in \mathbb{Z}\}$ be a stationary random real process with $\mathbb{E}[\mathsf{X}_p] = \mu$ and $\mathbb{E}[\mathsf{X}_p^2] = \sigma^2 + \mu^2$.

Denote the Fourier coefficients of $\{\mathsf{X}_p; p = 0, \ldots, n-1\}$ by

$$c_\ell = \frac{1}{n}\sum_{p=0}^{n-1} \mathsf{X}_p e^{-\mathrm{j}\frac{2\pi}{n}p\ell} \qquad (321)$$

Of course $c_\ell^* = c_{-\ell} = c_{n-\ell}$. For this reason, it is convenient to view the index set $\ell = 0, 1, \ldots, n-1$ as the ring $\mathbb{Z}_n$.

The expectations of the Fourier coefficients of a stationary process depend on whether $\ell = 0$

or $\ell \neq 0$:

$$\mathbb{E}[c_0] = \frac{1}{n}\sum_{p=0}^{n-1}\mathbb{E}[\mathsf{X}_p] \tag{322}$$

$$= \mu \tag{323}$$

$$\mathbb{E}[c_\ell] = \frac{1}{n}\sum_{p=0}^{n-1}\mathbb{E}[\mathsf{X}_p]e^{-\mathrm{j}\frac{2\pi}{n}p\ell} \tag{324}$$

$$= \mu\frac{1}{n}\sum_{p=0}^{n-1}e^{-\mathrm{j}\frac{2\pi}{n}p\ell} \tag{325}$$

$$= 0 \tag{326}$$

Lemma 9 *The Fourier coefficients $\{c_\ell; \ell \in \mathbb{Z}_n : \ell \neq 0\}$ of a stationary process, are (marginally) identically distributed.*

Lemma 10 *[13] Let $\mathcal{L}$ denote a fixed set of indices (that does not depend on n). As $n \to \infty$, the joint distribution of the coefficients*

$$\sqrt{n}\{c_\ell - \mathbb{E}[c_\ell]\}, \ell \in \mathcal{L}$$

converges to an independent complex Gaussian product distribution with zero mean and variances σ^2.

The mixed moments play an important role in our analysis. The following result easily follows from the definition of the Fourier coefficients.

Lemma 11 *Consider an index vector $\mathbf{j} = [j_1, j_2, \ldots, j_\kappa] \in \mathbb{Z}_n^\kappa$, such that $\mathsf{sum}[\mathbf{j}] \neq 0$. Then* [5]

$$\mathbb{E}[c_{j_1}c_{j_2}\cdots c_{i_\kappa}] = 0. \tag{327}$$

For the case where the sum of the indices in $\mathbf{j}$ is zero, we have the following result

Lemma 12 *Let $\mathbf{j} = [j_1, j_2 \ldots, j_\kappa] \in \mathbb{Z}_n^\kappa$ be such that $\mathsf{sum}[\mathbf{j}] = 0$. Let $\{\mathsf{X}_p; p \in \mathbb{Z}_n\}$ be independent identically distributed with $P[\mathsf{X}_1 = 0] = 1 - P[\mathsf{X}_1 = 1] = \mathsf{e}$. Then*

$$\mathbb{E}[c_{j_1}c_{j_2}\cdots c_{j_\kappa}] = \frac{1}{n^\kappa}\sum_{\ell=1}^{\kappa}\sum_{\tau[\ell]}\mathcal{B}(\kappa, \mathsf{e}, \tau[\ell])\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} \tag{328}$$

where $\mathbf{j}\cdot\mathbf{p} = \sum_{r=1}^{\kappa} j_r p_r$ and where

$$\mathcal{B}(\kappa, \mathsf{e}, \tau[\ell]) = (1-\mathsf{e})^\ell + \sum_{w=\ell+1}^{\kappa}(1-\mathsf{e})^w\sum_{\sigma[w]\leq\tau[\ell]}\zeta(\sigma[w] \to \tau[\ell]) \tag{329}$$

with $\tau[\ell]$ and $\sigma[w]$ denoting the partitions of the index set $\mathcal{X} = \{1, \ldots, \kappa\}$ and $\mathbb{S}(\tau[\ell])$ given in Definition 12 (see Appendix F).

[5]The squares (as opposed to the magnitude squares) of the complex coefficients satisfy $\mathbb{E}[c_\ell^2] = 0\ \forall\ell \neq 0$

Proof: We can write

$$\mathbb{E}[c_{j_1}c_{j_2}\cdots c_{j_\kappa}] = \frac{1}{n^\kappa}\sum_{\mathbf{p}\in\mathbb{Z}_n^\kappa}\mathbb{E}\left[\mathsf{X}_{p_1}\cdots\mathsf{X}_{p_\kappa}\right]e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} \tag{330}$$

Recalling that $\mathbb{Z}_n^\kappa = \mathbb{S}_\kappa$ and using the decomposition (316), we have

$$\begin{aligned}
\mathbb{E}[c_{j_1}c_{j_2}\cdots c_{j_\kappa}] &= \frac{1}{n^\kappa}\sum_{w=1}^{\kappa}\sum_{\sigma[\mathbf{v},w]}\sum_{\mathbf{p}\in\mathbb{S}^-(\sigma[\mathbf{v},w])}\mathbb{E}\left[\mathsf{X}_{p_1}\cdots\mathsf{X}_{p_\kappa}\right]e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} \\
&= \frac{1}{n^\kappa}\sum_{w=1}^{\kappa}\sum_{\sigma[\mathbf{v},w]}\left(\prod_{i=1}^{w}\mathbb{E}\left[\mathsf{X}_1^{v_i}\right]\right)\sum_{\mathbf{p}\in\mathbb{S}^-(\sigma[\mathbf{v},w])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (331)\\
&= \frac{1}{n^\kappa}\sum_{w=1}^{\kappa}\sum_{\sigma[w]}(1-\mathsf{e})^w\sum_{\mathbf{p}\in\mathbb{S}^-(\sigma[w])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (332)\\
&= \frac{1}{n^\kappa}\sum_{w=1}^{\kappa}\sum_{\sigma[w]}(1-\mathsf{e})^w\sum_{\ell=1}^{w}\sum_{\tau[\ell]\geq\sigma[w]}\zeta(\sigma[w]\to\tau[\ell])\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (333)\\
&= \frac{1}{n^\kappa}\sum_{\ell=1}^{\kappa}\sum_{w=1}^{\kappa}\sum_{\sigma[w]}\sum_{\tau[\ell]\geq\sigma[w]}(1-\mathsf{e})^w\zeta(\sigma[w]\to\tau[\ell])\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (334)\\
&= \frac{1}{n^\kappa}\sum_{\ell=1}^{\kappa}\sum_{\tau[\ell]}\sum_{w=\ell}^{\kappa}\sum_{\sigma[w]\leq\tau[\ell]}(1-\mathsf{e})^w\zeta(\sigma[w]\to\tau[\ell])\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (335)\\
&= \frac{1}{n^\kappa}\sum_{\ell=1}^{\kappa}\sum_{\tau[\ell]}\left[(1-\mathsf{e})^\ell+\sum_{w=\ell+1}^{\kappa}(1-\mathsf{e})^w\sum_{\sigma[w]\leq\tau[\ell]}\zeta(\sigma[w]\to\tau[\ell])\right]\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])}e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{j}\cdot\mathbf{p}} && (336)
\end{aligned}$$

where (336) coincides with the desired result, (331) follows from the definition of the set $\mathbb{S}^-(\sigma[\mathbf{v},w])$, and from the fact that the process is stationary, (332) follows by noticing that for any non-zero integer v, $\mathbb{E}[\mathsf{X}_1^v] = \mathbb{E}[\mathsf{X}_1] = 1-\mathsf{e}$, (333) is an application of Lemma 8, and (334), (335) follow by rearranging the terms in the summations. ■

H Lemma 13

Lemma 13 *Let* $\mathbf{E}$ *be the erasure matrix defined in Example 1. Let* $\mathbf{\Psi}$ *be the circulant matrix and* $\mathbf{F}$ *denote the* $n\times n$ *unitary DFT matrix defined in Lemma 1, and let* $\mathbf{\Lambda} = \mathit{diag}(\lambda_1,\ldots,\lambda_n)$ *denote the eigenvalues of* $\mathbf{\Psi}$*. Let*

$$\mathbf{Q} = \mathbf{EF} \tag{337}$$

and denote by $\mathbf{q}_i$ *the ith column of* $\mathbf{Q}$*. Then as the dimension of the matrices grows,*

$$\mathbf{q}_i^\dagger\left(\mathbf{I}+\gamma\sum_{j\neq i}\lambda_j\mathbf{q}_j\mathbf{q}_j^\dagger\right)^{-1}\mathbf{q}_i \stackrel{a.s.}{\to} \alpha \tag{338}$$

where α *depends on* e *and on the asymptotic empirical distribution of* $\mathbf{\Lambda}$ *but it does not depend on* i*.*

Proof: Denote[6]

$$\mathbf{\Lambda}_i = \text{diag}\{\lambda_0, \lambda_1, \ldots, \lambda_{i-1}, 1, \lambda_{i+1}, \ldots, \lambda_{n-1}\} \tag{339}$$

$$\alpha_i = \mathbf{q}_i^\dagger \left(\mathbf{I} + \gamma \sum_{j \neq i} \lambda_j \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i \tag{340}$$

The lemma is equivalent to proving that $\beta_i \overset{a.s.}{\to} \beta$, where

$$\begin{aligned} \beta_i &= \frac{\alpha_i}{1+\gamma\alpha_i} \\ &= \mathbf{q}_i^\dagger \left(\mathbf{I} + \gamma \mathbf{q}_i \mathbf{q}_i^\dagger + \gamma \sum_{j \neq i} \lambda_j \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i & (341) \\ &= \left(\mathbf{Q}^\dagger (\mathbf{I} + \gamma \mathbf{Q} \mathbf{\Lambda}_i \mathbf{Q}^\dagger)^{-1} \mathbf{Q} \right)_{i,i} & (342) \end{aligned}$$

where (341) follows from the matrix inversion lemma. Using the series expansion of the matrix inverse, we can write the matrix in the right side of (342) as

$$\begin{aligned} \mathbf{Q}^\dagger (\mathbf{I} + \gamma \mathbf{Q} \mathbf{\Lambda}_i \mathbf{Q}^\dagger)^{-1} \mathbf{Q} &= \mathbf{Q}^\dagger \sum_{k=0}^{\infty} (-\gamma)^k (\mathbf{Q} \mathbf{\Lambda}_i \mathbf{Q}^\dagger)^k \mathbf{Q} & (343) \\ &= \sum_{k=0}^{\infty} (-\gamma)^k \mathbf{Q}^\dagger (\mathbf{Q} \mathbf{\Lambda}_i \mathbf{Q}^\dagger)^k \mathbf{Q} & (344) \\ &= \sum_{k=0}^{\infty} (-\gamma)^k (\mathbf{C} \mathbf{\Lambda}_i)^k \mathbf{C} & (345) \end{aligned}$$

where we have denoted the $n \times n$ circulant matrix

$$\begin{aligned} \mathbf{C} &= \mathbf{Q}^\dagger \mathbf{Q} & (346) \\ &= \mathbf{F}^\dagger \mathbf{E}^2 \mathbf{F} & (347) \\ &= \mathbf{F}^\dagger \mathbf{E} \mathbf{F} & (348) \end{aligned}$$

Thus,

$$\mathbf{C}_{i,j} = \frac{1}{n} \sum_{p=0}^{n-1} e_p e^{\mathrm{j} \frac{2\pi}{n} p(i-j)} = c_{i-j} \tag{349}$$

Notice that the coefficients of the first row of $\mathbf{C}$ are obtained as the DFT of the sequence $\{e_p : p = 0, \ldots, n-1\}$.

According to (342) and (345), it will be sufficient for our purposes to show that the random variables

$$\theta_{k,i} = \left((\mathbf{C} \mathbf{\Lambda}_i)^k \, \mathbf{C} \right)_{i,i}, \tag{350}$$

converge almost surely

$$\lim_{n \to \infty} \theta_{k,i} \overset{a.s.}{\to} \varrho_k \tag{351}$$

[6]For the sake of notation simplicity, it is convenient to use indices in $\mathbb{Z}_n$. Therefore, the matrix and vector components shall be numbered from 0 to $n-1$ rather than from 1 to n.

where ϱ_k does not depend on i. At this point, the proof proceeds through a sequence of lemmas, proved in the following sections.

Lemma 14 *The limit $\varrho_k = \lim_{n\to\infty} \mathbb{E}[\theta_{k,i}]$ exists and does not depend on the index i but only on the erasure probability e and on the asymptotic empirical distribution of $\mathbf{\Lambda}$.*

Proof: Appendix I ∎

Lemma 15 *The central moments of $\theta_{k,i}$ of order 2,3 and 4 satisfy*

$$\mathrm{Var}\{\theta_{k,i}\} = O\left(\frac{1}{n}\right) \quad (352)$$

$$\mathbb{E}\left[|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]|^4\right] = O\left(\frac{1}{n^2}\right) \quad (353)$$

Proof: Appendix J ∎

The last step of the proof of Lemma 13 follows as an application of Markov's inequality and of the Borel-Cantelli lemma. For $\epsilon > 0$ we have

$$P\left(|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]| > \epsilon\right) \leq \frac{\mathbb{E}[|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]|^4]}{\epsilon^4} \quad (354)$$

$$= O\left(\frac{1}{n^2}\right) \quad (355)$$

where (355) follows from (353). This, combined with Lemma 14, shows that $\theta_{k,i} \to \varrho_k$ in probability. Furthermore, since the sequence of probabilities $\{P(|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]| > \epsilon) \; : \; n = 1, 2, \ldots\}$ is summable for all $\epsilon > 0$, we have that $\theta_{k,i} \to \varrho_k$ almost surely. ∎

I Appendix: Proof of Lemma 14

We wish to compute $\mathbb{E}[\theta_{k,i}]$ where $\theta_{k,i}$ is defined in (350). For notation simplicity, we shall prove the expression for finite n and k for a general diagonal non-negative matrix $\mathbf{B} = \mathrm{diag}(b_0, \ldots, b_{n-1})$ and substitute $\mathbf{B} = \mathbf{\Lambda}_i$ at the end. We organize the proof in steps, for the sake of readability.

Step 1.

$$\begin{aligned}\theta_{k,i} &= \left((\mathbf{CB})^k \mathbf{C}\right)_{i,i} \\ &= \sum_{\mathbf{j}\in\mathbb{Z}_n^k} \left(\prod_{r=1}^{k} b_{j_r}\right) c_{j_1-i} c_{i-j_k} \left(\prod_{s=1}^{k-1} c_{j_{s+1}-j_s}\right) \end{aligned} \quad (356)$$

To prove (356), we notice that since $\mathbf{C}$ is circulant, we can write

$$\mathbf{C} = \sum_{\ell=0}^{n-1} c_\ell \mathbf{X}^\ell \quad (357)$$

where $\mathbf{X}$ is the elementary cyclic permutation matrix defined in (272). For any $j \in \mathbb{Z}_n$, denote by $\mathbf{1}_j$ the n-dimensional vector of all zeros except a "1" in position j. Then the following rules hold:

1. $\mathbf{X}^{\ell}\mathbf{1}_j = \mathbf{1}_{j-\ell}$.
2. $\mathbf{1}_j^T\mathbf{X}^{\ell} = \mathbf{1}_{\ell+j}^T$.
3. $\mathbf{1}_j^T\mathbf{1}_{\ell} = \mathbf{0}$ unless $\ell = j$.

Notice that the above indices relations holds modulo n, i.e., index operations are in the ring $\mathbb{Z}_n$.

Furthermore, since $\mathbf{B}$ is diagonal, we can write

$$\mathbf{B} = \sum_{j=0}^{n-1} b_j \mathbf{1}_j \mathbf{1}_j^T \tag{358}$$

As far as notation is concerned, we shall use extensively the notation $\sum_{\mathbf{v}\in\mathbb{Z}_n^k}$ to indicate a sum with respect to k indices $v_1, \ldots, v_k$ that range over $\mathbb{Z}_n = \{0, \ldots, n-1\}$. Using the above definitions, we can write

$$\begin{aligned}
(\mathbf{CB})^k\,\mathbf{C} &= \left(\sum_{\ell=0}^{n-1} c_\ell \mathbf{X}^\ell \sum_{j=0}^{n-1} b_j \mathbf{1}_j\mathbf{1}_j^T\right)^k \sum_{\ell=0}^{n-1} c_\ell \mathbf{X}^\ell \\
&= \sum_{\mathbf{j}\in\mathbb{Z}_n^k}\left(\prod_{r=1}^k b_{j_r}\right)\sum_{\ell=0}^{n-1}\sum_{\boldsymbol{\ell}\in\mathbb{Z}_n^k} c_\ell \left(\prod_{s=1}^k c_{\ell_s}\right)\mathbf{X}^{\ell_1}\mathbf{1}_{j_1}\mathbf{1}_{j_1}^T\mathbf{X}^{\ell_2}\mathbf{1}_{j_2}\mathbf{1}_{j_2}^T\mathbf{X}^{\ell_3}\cdots\mathbf{1}_{j_k}\mathbf{1}_{j_k}^T\mathbf{X}^{\ell} \\
&= \sum_{\mathbf{j}\in\mathbb{Z}_n^k}\left(\prod_{r=1}^k b_{j_r}\right)\sum_{\ell=0}^{n-1}\sum_{\boldsymbol{\ell}\in\mathbb{Z}_n^k} c_\ell \left(\prod_{s=1}^k c_{\ell_s}\right)\mathbf{X}^{\ell_1}\mathbf{1}_{j_1}\left(\prod_{g=1}^{k-1}\mathbf{1}_{j_g}^T\mathbf{X}^{\ell_{g+1}}\mathbf{1}_{j_{g+1}}\right)\mathbf{1}_{j_k}^T\mathbf{X}^{\ell} \\
&= \sum_{\mathbf{j}\in\mathbb{Z}_n^k}\left(\prod_{r=1}^k b_{j_r}\right)\sum_{\ell=0}^{n-1}\sum_{\boldsymbol{\ell}\in\mathbb{Z}_n^k} c_\ell \left(\prod_{s=1}^k c_{\ell_s}\right)\mathbf{1}_{j_1-\ell_1}\left(\prod_{g=1}^{k-1}\mathbf{1}_{j_g+\ell_{g+1}}^T\mathbf{1}_{j_{g+1}}\right)\mathbf{1}_{j_k+\ell}^T \\
&= \sum_{\mathbf{j}\in\mathbb{Z}_n^k}\left(\prod_{r=1}^k b_{j_r}\right)\sum_{\ell,\ell_1=0}^{n-1} c_\ell c_{\ell_1}\left(\prod_{s=1}^{k-1} c_{j_{s+1}-j_s}\right)\mathbf{1}_{j_1-\ell_1}\mathbf{1}_{j_k+\ell}^T
\end{aligned} \tag{359}$$

where (359) holds because only the terms for $j_g + \ell_{g+1} = j_{g+1}$ modulo n are not identically zero. Now, we take the (i,i)-th element of the matrix in (359) and get (356).

Step 2. We have

$$\mathbb{E}\left[\theta_{k,i}\right] \;=\; \sum_{\ell=1}^{k+1}\sum_{\tau[\ell]}\mathcal{B}(k+1,\mathsf{e},\tau[\ell])\frac{1}{n^{k+1-\ell}}\sum_{\mathbf{j}\in\mathbb{G}(i,\tau[\ell])}\left(\prod_{r=1}^k b_{j_r}\right) \tag{360}$$

where $\mathbb{G}(i,\tau[\ell])$ is a subset of $\mathbb{Z}_n^k$ defined in (368), and where $\mathcal{B}(\kappa,\mathsf{e},\tau[\ell])$ is the coefficient defined in (329) of Lemma 12.

In order to show (360), we use (356) and write

$$\mathbb{E}[\theta_{k,i}] = \sum_{\mathbf{j}\in\mathbb{Z}_n^k}\left(\prod_{r=1}^k b_{j_r}\right)\mathbb{E}\left[c_{j_1-i}c_{i-j_k}\left(\prod_{s=1}^{k-1} c_{j_{s+1}-j_s}\right)\right] \tag{361}$$

We notice that the above expectation reduces to computing the joint moment of order $k+1$ of the Fourier coefficients

$$\{c_\ell : \ell = j_1 - i, j_2 - j_1, j_3 - j_2, \ldots, j_k - j_{k-1}, i - j_k\}$$

where i is fixed, and $\mathbf{j} = (j_1, \ldots, j_k)$ ranges over $\mathbb{Z}_n^k$.

Define the indices $\mathbf{v} \in \mathbb{Z}_n^{k+1}$ such that

$$\begin{aligned} v_1 &= j_1 - i \\ v_s &= j_s - j_{s-1}, \quad s = 2, \ldots, k \\ v_{k+1} &= i - j_k \end{aligned} \tag{362}$$

Notice that, by construction, the index vector $\mathbf{v}$ satisfies $\mathsf{sum}[\mathbf{v}] = 0$. Hence, the expectation in (361) is obtained by applying Lemma 12. We have

$$\mathbb{E}\left[c_{v_1} c_{v_2} \cdots c_{v_{k+1}}\right] = \frac{1}{n^{k+1}} \sum_{\ell=1}^{k+1} \sum_{\tau[\ell]} \mathcal{B}(k+1, \mathsf{e}, \tau[\ell]) \sum_{\mathbf{p} \in \mathbb{S}(\tau[\ell])} e^{-\mathrm{j}\frac{2\pi}{n} \mathbf{v} \cdot \mathbf{p}} \tag{363}$$

Next, we wish to evaluate the sum of complex exponentials in (363). Recall that $\mathbb{S}(\tau[\ell]) \subset \mathbb{Z}_n^{k+1}$ is the set of vectors with constant values over the blocks of the partition $\tau[\ell]$. Denoting these blocks by $\mathcal{V}(m_1), \ldots, \mathcal{V}(m_\ell)$, for any $\mathbf{p} \in \mathbb{S}(\tau[\ell])$ we have

$$\mathbf{p}(\mathcal{V}(m_r)) = (\underbrace{h_r, h_r, \ldots, h_r}_{m_r \text{ times}})$$

for some value $h_r \in \mathbb{Z}_n$, for all $r = 1, \ldots, \ell$. Then, we have

$$\begin{aligned} \sum_{\mathbf{p} \in \mathbb{S}(\tau[\ell])} e^{-\mathrm{j}\frac{2\pi}{n} \mathbf{v} \cdot \mathbf{p}} &= \sum_{h_1, \ldots, h_\ell = 0}^{n-1} e^{-\mathrm{j}\frac{2\pi}{n} \sum_{r=1}^{\ell} h_r \mathsf{sum}[\mathbf{v}(\mathcal{V}(m_r))]} \\ &= \prod_{r=1}^{\ell} \left(\sum_{h=0}^{n-1} e^{-\mathrm{j}\frac{2\pi}{n} h \, \mathsf{sum}[\mathbf{v}(\mathcal{V}(m_r))]} \right) \\ &= \begin{cases} n^\ell & \text{if } \mathsf{sum}[\mathbf{v}(\mathcal{V}(m_r))] = 0 \ \ \forall\, r = 1, \ldots, \ell \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{364}$$

Expressed in words, the condition on $\tau[\ell]$ for which the term in (364) is non-zero is that $\tau[\ell]$ is a good partition for the vector of indices $\mathbf{v}$ (see Definition 8).

Now, we examine the set of index vectors $\mathbf{j} \in \mathbb{Z}_n^k$ such that a given partition $\tau[\ell]$ of size $k+1$ is a good partition for $\mathbf{v}$, where $\mathbf{v}$ and $\mathbf{j}$ are related by (362) and i is given and fixed. We write (362) in matrix form as

$$\mathbf{v} = \mathbf{j}\boldsymbol{\Phi} + (-i, 0, \ldots, 0, i) \tag{365}$$

where $\boldsymbol{\Phi}$ is the double-diagonal matrix of dimension $k \times (k+1)$ given by

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \tag{366}$$

Letting $\mathbf{A}_\tau$ denote the incidence matrix of $\tau[\ell]$, of dimension $(k+1)\times\ell$, we have that all $\mathbf{j}$ such that $\tau[\ell]$ is a good partition of $\mathbf{v}$ satisfy the linear equation

$$\mathbf{j}\mathbf{\Phi}\mathbf{A}_\tau = (i,0,0,\ldots,-i)\mathbf{A}_\tau \tag{367}$$

The solution of the above (undertermined) system of linear equations over $\mathbb{Z}_n$ is a translate of the Kernel (null-space)of the linear transformation $\mathbb{Z}_n^k \to \mathbb{Z}_n^\ell$ defined by the matrix $\mathbf{\Phi}\mathbf{A}_\tau$. In particular, notice that

$$(i,i,i,\ldots,i)\mathbf{\Phi} = (i,0,0,\ldots,-i)$$

Hence, the set of solutions of (367) is given by

$$\mathbb{G}(i,\tau[\ell]) = \mathrm{Ker}(\mathbf{\Phi}\mathbf{A}_\tau) + (i,i,i,\ldots,i) \tag{368}$$

For later use, notice that $\mathbf{\Phi}$ maps $\mathbb{Z}_n^k$ into the subset of $\mathbb{Z}_n^{k+1}$ of vectors with zero sum of their components. Hence, using (307) in Lemma 6 we have that, irrespectively of i and $\tau[\ell]$,

$$|\mathbb{G}(i,\tau[\ell])| = n^{k+1-\ell} \tag{369}$$

By using (364) into (363), and the result thereof into (361) we finally arrive at (360).

Step 3. Substituting $b_j = \lambda_j$ for $j \neq i$ and $b_i = 1$ in (360), we notice that $\mathbb{E}[\theta_{k,i}]$ depends on i only through the term

$$\widetilde{\mathcal{T}}_{k,i}^{(n)}(\mathbf{\Lambda},\tau[\ell]) = \frac{1}{n^{k+1-\ell}} \sum_{\mathbf{j}\in\mathbb{G}(i,\tau[\ell])} \left(\prod_{r=1}^{k} b_{j_r} \right) \tag{370}$$

Lemma 14 is proved if we can show that, as $n\to\infty$, $\widetilde{\mathcal{T}}_i^{(n)}(\mathbf{\Lambda},\tau[\ell]) \to \mathcal{T}_k(\mathbf{\Lambda},\tau[\ell])$, where the latter is some limit independent of i.

Since $|\mathbb{G}(i,\tau[\ell])| = n^{k+1-\ell}$ and the eigenvalues in $\mathbf{\Lambda}$ are bounded[7], it follows that $\lim_{n\to\infty} \widetilde{\mathcal{T}}_{k,i}^{(n)}(\mathbf{\Lambda},\tau[\ell])$ exists (finite).

In order to see that the limit of $\widetilde{\mathcal{T}}_{k,i}^{(n)}(\mathbf{\Lambda},\tau[\ell])$ is indeed independent of i, we need some observations on the structure of the set $\mathbb{G}(i,\tau[\ell])$ defined in (368).

To illustrate the structure of the summation set $\mathbb{G}(i,\tau[\ell])$, let's consider a simple example with $k=3$ and $\ell=2$, and $\tau[2]=\{\{1,2\},\{3,4\}\}$. We have

$$\mathbf{\Phi}\mathbf{A}_\tau = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix}$$

The kernel of $\mathbf{\Phi}\mathbf{A}_\tau$ is given by the set of vectors $\mathbf{j}=(j_1,0,j_3)$, with j_1, j_3 freely varying in $\mathbb{Z}_n$. Its cardinality is given by n^2, consistently with (369). Furthermore, we notice that the kernel is isomorphic to $\mathbb{Z}_n^2$. In general, it is easy to see that the kernel is isomorphic to $\mathbb{Z}_n^{k+1-\ell}$. It follows that $\mathbb{G}(i,\tau[\ell])$ is formed by all vectors of the type

$$(\underbrace{j_1+i, j_2+i, \ldots, j_{k+1-\ell}+i}_{k+1-\ell}, \underbrace{i,i,\ldots,i}_{\ell-1})$$

[7]Recall that $\mathbf{\Lambda}$ contains the eigenvalues of $\mathbf{\Sigma}$ defined as a Toeplitz matrix of an absolutely summable sequence, that has bounded strong norm.

Since $j_1, \ldots, j_{k+1-\ell}$ take all possible values in $\mathbb{Z}_n$, this set is identical (up to component permutations that are irrelevant since the argument of the summation in (370) is a product) to the set of vectors of length k given by

$$(z_1, \ldots, z_{k+1-\ell}, i, i, \ldots, i)$$

where $z_1, \ldots, z_{k+1-\ell}$ take on all possible values in $\mathbb{Z}_n$.

Recalling that $b_i = 1$, while $b_j = \lambda_j$ for all $j \neq i$, we have

$$\sum_{\mathbf{j} \in \mathbb{G}(i, \tau[\ell])} \left(\prod_{r=1}^{k} b_{j_r} \right) = \sum_{\mathbf{z} \in \mathbb{Z}_n^{k+1-\ell}} \left(\prod_{r=1}^{k+1-\ell} b_{z_r} \right) \tag{371}$$

Next, consider

$$\mathcal{T}_k^{(n)}(\mathbf{\Lambda}, \tau[\ell]) = \frac{1}{n^{k+1-\ell}} \sum_{\mathbf{z} \in \mathbb{Z}_n^{k+1-\ell}} \left(\prod_{r=1}^{k+1-\ell} \lambda_{z_r} \right) \tag{372}$$

Using (371) in (370) and comparing the resulting expression with (372) we notice that the two expressions differ by the fact that in some terms of the sum or products in (370) we have a factor 1 instead of λ_i. These are precisely the factors with index $z_r = i$. We can decompose the sum w.r.t. $\mathbf{z} \in \mathbb{Z}_n^{k+1-\ell}$ into "shells" of vectors with exactly s components fixed to i. Using the binomial identity

$$n^{k+1-\ell} = \sum_{s=0}^{k+1-\ell} \binom{k+1-\ell}{s} (n-1)^{k+1-\ell-s}$$

we have that there are exactly $\binom{k+1-\ell}{s}(n-1)^{k+1-\ell-s}$ vectors $\mathbf{z} \in \mathbb{Z}_n^{k+1-\ell}$ such that s components are equal to i and the others are different from i. Then, let $\mathcal{S}_n^{\kappa}(i)$ denote the set of vectors $\mathbf{z} \in \mathbb{Z}_n^{\kappa}$ with all components different from i. Clearly, the set $\mathcal{S}_n^{k+1-\ell-s}(i)$ is in one-to-one correspondence with the sth shell defined above. We have

$$\begin{aligned}
\left| \widetilde{\mathcal{T}}_{k,i}^{(n)}(\mathbf{\Lambda}, \tau[\ell]) - \mathcal{T}_k^{(n)}(\mathbf{\Lambda}, \tau[\ell]) \right| &= \left| \frac{1}{n^{k+1-\ell}} \sum_{s=0}^{k+1-\ell} \sum_{\mathbf{z} \in \mathcal{S}_n^{k+1-\ell-s}(i)} \left(\prod_{r=1}^{k+1-\ell-s} \lambda_{z_r} \right) (1 - \lambda_i^s) \right| \\
&\leq \sum_{s=0}^{k+1-\ell} |1 - \lambda_i^s| \, \frac{1}{n^{k+1-\ell}} \sum_{\mathbf{z} \in \mathcal{S}_n^{k+1-\ell-s}(i)} \left(\prod_{r=1}^{k+1-\ell-s} \lambda_{z_r} \right) \\
&\leq \mathcal{K} \sum_{s=1}^{k+1-\ell} \frac{1}{n^{k+1-\ell}} \sum_{\mathbf{z} \in \mathcal{S}_n^{k+1-\ell-s}(i)} \left(\prod_{r=1}^{k+1-\ell-s} \lambda_{z_r} \right)
\end{aligned} \tag{373}$$

for some constant $\mathcal{K}$ that does not depend on n and i that uniformly bounds from above the term $|1 - \lambda_i^s|$. The existence of such constant follows from the fact that $\mathbf{\Lambda}$ (i.e., $\mathbf{\Sigma}$) is bounded in strong norm.

The final result is obtained by observing that for all $s = 1, 2, \ldots, k+1-\ell$, we have

$$\sum_{\mathbf{z} \in \mathcal{S}_n^{k+1-\ell-s}(i)} \left(\prod_{r=1}^{k+1-\ell-s} \lambda_{z_r} \right) \leq (n-1)^{k+1-\ell-s} \mathcal{K}'$$

for some finite constant $\mathcal{K}'$ independent of n and i. Hence,

$$\lim_{n\to\infty} \frac{1}{n^{k+1-\ell}} \sum_{\mathbf{z}\in\mathcal{S}_n^{k+1-\ell-s}(i)} \left(\prod_{r=1}^{k+1-\ell-s} \lambda_{z_r} \right) \leq \lim_{n\to\infty} \frac{(n-1)^{k+1-\ell-s}\mathcal{K}'}{n^{k+1-\ell}} = 0 \tag{374}$$

It follows that $\lim_{n\to\infty} \widetilde{\mathcal{T}}_{k,i}^{(n)}(\mathbf{\Lambda}, \tau[\ell]) = \lim_{n\to\infty} \mathcal{T}_k^{(n)}(\mathbf{\Lambda}, \tau[\ell])$, where the latter limit does not depend on i, by construction (see (372)).

J Proof of Lemma 15

Again, for the sake of readability, we proceed by steps.

Step 1. We start by proving the following general expression for the s-th moment of $\theta_{k,i}$ defined in (350). [8] For any integer $s \geq 1$, we have

$$\mathbb{E}[\theta_{k,i}^s] = \sum_{\ell=1}^{s(k+1)} \sum_{\tau[\ell]} \mathcal{B}(s(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{sk-\ell+s}} \sum_{(\mathbf{j}_1|\dots|\mathbf{j}_s)\in\mathbb{G}^s(i,\tau[\ell])} \prod_{t=1}^{s}\prod_{r=1}^{k} b_{j_{t,r}} \tag{375}$$

where, as in Appendix I, we consider a general diagonal matrix $\mathbf{B}$ with elements $b_0, \dots, b_{n-1}$ and the replace $\mathbf{B} = \mathbf{\Lambda}_i$ at the end, where $\tau[\ell]$ in the above summation are partitions of the set of indices $\mathcal{X}_1^s$ of a concatenated vector of the type $\mathbf{x} = (\mathbf{x}_1|\dots|\mathbf{x}_s)$ with s blocks of length $k+1$, and where the set $\mathbb{G}^s(i, \tau[\ell])$ is the generalization of the set $\mathbb{G}(i, \tau[\ell])$ defined in (368) to the case of concatenated vectors with s components, and shall be explicitly defined in (381) given below (it can be noticed that for $s = 1$ the definitions (368) and (381) coincide).

In order to show (375), consider the explicit expression of $\theta_{k,i}$ given in (356). Then, we can write

$$\begin{aligned}
\mathbb{E}[\theta_{k,i}^s] &= \mathbb{E}\left[\sum_{\mathbf{j}_1\in\mathbb{Z}_n^k} \cdots \sum_{\mathbf{j}_s\in\mathbb{Z}_n^k} \left(\prod_{t=1}^{s}\prod_{r=1}^{k} b_{j_{t,r}} \right) \prod_{t=1}^{s} c_{j_{t,1}-i} c_{i-j_{t,k}} \left(\prod_{r=1}^{k-1} c_{j_{t,r+1}-j_{t,r}} \right) \right] \\
&= \sum_{(\mathbf{j}_1|\dots|\mathbf{j}_s)\in\mathbb{Z}_n^{sk}} \left(\prod_{t=1}^{s}\prod_{r=1}^{k} b_{j_{t,r}} \right) \mathbb{E}\left[\prod_{t=1}^{s} c_{j_{t,1}-i} c_{i-j_{t,k}} \left(\prod_{r=1}^{k-1} c_{j_{t,r+1}-j_{t,r}} \right) \right]
\end{aligned} \tag{376}$$

Consider the concatenated index vector $\mathbf{v} = (\mathbf{v}_1|\dots|\mathbf{v}_s)$ with components of length $k+1$, that depends on the summation indices $(\mathbf{j}_1|\dots|\mathbf{j}_s)$ of (376) as follows

$$\begin{aligned}
v_{t,1} &= j_{t,1} - i \\
v_{t,r} &= j_{t,r} - j_{t,r-1}, \quad r = 2, \dots, k \\
v_{t,k+1} &= i - j_{t,k}
\end{aligned} \tag{377}$$

for $t = 1, \dots, s$. Notice that for any $i \in \mathbb{Z}_n$ we have that $\mathsf{sum}[\mathbf{v}_t] = 0$ for all $t = 1, \dots, s$. This implies that $\mathsf{sum}[\mathbf{v}] = 0$. Then, we can apply Lemma 12 and obtain

$$\mathbb{E}\left[\prod_{t=1}^{s} c_{j_{t,1}-i} c_{i-j_{t,k}} \left(\prod_{r=1}^{k-1} c_{j_{t,r+1}-j_{t,r}} \right) \right] = \frac{1}{n^{s(k+1)}} \sum_{\ell=1}^{s(n+1)} \sum_{\tau[\ell]\in\mathfrak{P}_\ell(s(k+1))} \mathcal{B}(s(k+1), \eta, \tau[\ell]) \sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])} e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{v}\cdot\mathbf{p}} \tag{378}$$

[8] Notice that $\theta_{k,i}$ is real and non-negative. Hence, $\mathbb{E}[\theta_{k,i}^s] = \mathbb{E}[|\theta_{k,i}|^s]$ for all $s = 1, 2, \dots$.

Similarly to what done in (364), we have

$$\sum_{\mathbf{p}\in\mathbb{S}(\tau[\ell])} e^{-\mathrm{j}\frac{2\pi}{n}\mathbf{v}\cdot\mathbf{p}} = \left\{ \begin{array}{ll} n^\ell & \text{if } \mathsf{sum}[\mathbf{v}(\mathcal{V}(m_r))] = 0 \ \forall\, r = 1,\ldots,\ell \\ 0 & \text{otherwise} \end{array} \right. \tag{379}$$

where $\mathcal{V}(m_r)$ denotes the r-th block of $\tau[\ell]$. This shows that the sum in (379) is non-zero if and only if $\tau[\ell]$ is a good partition for the concatenated index vector $\mathbf{v} = (\mathbf{v}_1|\ldots|\mathbf{v}_s)$.

By replacing (379) in (378) and using the result in (376) we obtain (375), where the summation index set $\mathbb{G}^s(i,\tau[\ell])$ contains all index vectors $(\mathbf{j}_1|\ldots|\mathbf{j}_s)$ such that $\tau[\ell]$ is a good partition for the associated index vector $(\mathbf{v}_1|\ldots|\mathbf{v}_s)$ defined in (377).

Step 2. We wish to obtain a more explicit expression of the sth moment in (375). In order to do so, we have to take a closer look at the structure of $\mathbb{G}^s(i,\tau[\ell])$. Consider the $k\times(k+1)$ matrix $\mathbf{\Phi}$ defined in (366), and the incidence matrix $\mathbf{A}_\tau$ associated to the partition $\tau[\ell]$, of dimension $s(k+1)\times\ell$. Furthermore, it is clear from (377) that, for each $t = 1,\ldots,s$,

$$\mathbf{v}_t = \mathbf{j}_t\mathbf{\Phi} - (i,i,\ldots,i)\mathbf{\Phi} \tag{380}$$

It follows that the $\mathbb{G}^s(i,\tau[\ell])$ coincides with the solution set of the system of linear equations over $\mathbb{Z}_n$ given by $(\mathbf{v}_1|\ldots|\mathbf{v}_s)\mathbf{A}_\tau = \mathbf{0}$. Explicitly, this is given by

$$\mathbb{G}^s(i,\tau[\ell]) = \mathrm{Ker}\left([\mathbf{I}_s\otimes\mathbf{\Phi}]\,\mathbf{A}_\tau\right) + (\underbrace{i,i,\ldots,i}_{sk}) \tag{381}$$

where $\mathbf{I}_s$ is the $s\times s$ identity matrix and $\otimes$ denotes Kronecker product.

Let $\varpi[\boldsymbol{\nu}^*,\iota^*]$ denote the canonical decomposition of $\tau[\ell]$ with respect to the concatenated vector $(\mathbf{v}_1|\ldots|\mathbf{v}_s)$ and let $\{\tau_1[\ell_1],\ldots,\tau_{\iota^*}[\ell_{\iota^*}]\}$ denote its canonical representation, as in Definition 10. As argued in the proof of Lemma 6, the matrix $\mathbf{A}_\tau$ can be partitioned into ι^* blocks of mutually orthogonal columns, $\mathbf{A}_\tau = [\mathbf{A}_\tau^{\ell_1}|\ldots|\mathbf{A}_\tau^{\ell_{\iota^*}}]$ where the block $\mathbf{A}_\tau^{\ell_r}$ has dimension $s(k+1)\times\ell_r$ and corresponds to the partition $\tau_r[\ell_r]$ in the canonical decomposition.[9] By definition of canonical decomposition and canonical representation, we can partition $\mathbf{I}_s\otimes\mathbf{\Phi}$ into ι^* blocks with dimension $\nu_r^*k\times s(k+1)$, such that each such r-th block has rows orthogonal to the columns of $\mathbf{A}_\tau^{\nu_p^*}$ for all $p\neq r$. It follows that $\mathbb{G}^s(i,\tau[\ell])$ has a ι^*-fold Cartesian product structure.

Let's focus on the r-th block of the Cartesian product. It is apparent that this is the solution space of the system of linear equations

$$(\mathbf{z}_1|\ldots|\mathbf{z}_{\nu_r^*})\left[\mathbf{I}_{\nu_r^*}\otimes\mathbf{\Phi}\right]\mathbf{A}_{\tau_r} = (i,i,\ldots,i)\left[\mathbf{I}_{\nu_r^*}\otimes\mathbf{\Phi}\right]\mathbf{A}_{\tau_r} \tag{382}$$

where $\mathbf{A}_{\tau_r}$ is the incidence matrix of the r-th component partition of the canonical decomposition of $\tau[\ell]$, and $\mathbf{z}_1,\ldots,\mathbf{z}_{\nu_r^*}$ are k-vectors of variables over $\mathbb{Z}_n$. Hence, the solutions of (382) are given by the set

$$\mathrm{Ker}\left(\left[\mathbf{I}_{\nu_r^*}\otimes\mathbf{\Phi}\right]\mathbf{A}_{\tau_r}\right) + (\underbrace{i,i,\ldots,i}_{\nu_r^*k})$$

that is, by a translate of the Kernel (null-space) of $\left[\mathbf{I}_{\nu_r^*}\otimes\mathbf{\Phi}\right]\mathbf{A}_{\tau_r}$. Notice that the set of vectors

$$\left\{\mathbf{x} = (\mathbf{z}_1|\ldots|\mathbf{z}_{\nu_r^*})\left[\mathbf{I}_{\nu_r^*}\otimes\mathbf{\Phi}\right] : \mathbf{z}_t\in\mathbb{Z}_n^k, t = 1,\ldots,\nu_r^*\right\}$$

[9]The adverted reader will notice the close relationship between the set $\mathbb{G}(\tau[\mathbf{m},\ell,(\mathbf{x}_1|\ldots|\mathbf{x}_s)])$ examined in Lemma 6 and the set $\mathbb{G}^s(i,\tau[\ell])$ considered here.

contains all vectors of dimension $\nu_r^*(k+1)$ composed by ν_r^* blocks of length $(k+1)$ such that the sum of elements over each component is zero. Hence, Ker $\left(\left[\mathbf{I}_{\nu_r^*} \otimes \mathbf{\Phi}\right] \mathbf{A}_{\tau_r}\right)$ is isomorphic to the set $\mathbb{G}(\tau_r[\ell_r, (\mathbf{x}_1|\dots|\mathbf{x}_{\nu_r^*})])$ considered in Lemma 6. It follows that its cardinality is $n^{\nu_r^* k - \ell_r + 1}$. Eventually, the solution space of (382) is given by vectors of the form

$$(z_1, z_2, \dots, z_{\nu_r^* k - \ell_r + 1}, \underbrace{i, i, \dots, i}_{\ell_r - 1})$$

From the Cartesian product structure of $\mathbb{G}^s(i, \tau[\ell])$ discussed before, after reordering of the components, the vectors in $\mathbb{G}^s(i, \tau[\ell])$ have form

$$(z_1, z_2, \dots, z_{sk-\ell+\iota^*}, \underbrace{i, i, \dots, i}_{\ell - \iota^*})$$

where $z_1, \dots, z_{sk-\ell+\iota^*}$ are arbitrary variables in $\mathbb{Z}_n$.

The sought more explicit expression for (375) is finally obtained by recalling that, by construction, $b_i = 1$ and $b_j = \lambda_j$ for all $j \neq i$. Hence, all terms in the product of b's in (375) with index $j_{t,r} = i$ disappear (they are equal to 1) and using the structure of $\mathbb{G}^s(i, \tau[\ell])$ discussed above we arrive at

$$\mathbb{E}[\theta_{k,i}^s] = \sum_{\ell=1}^{s(k+1)} \sum_{\tau[\ell]} \mathcal{B}(s(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{sk-\ell+s}} \sum_{\mathbf{z} \in \mathbb{Z}_n^{sk-\ell+\iota^*}} \prod_{r=1}^{sk-\ell+\iota^*} b_{z_r} \tag{383}$$

where we notice that the only dependence of the inner sum (w.r.t. the indices $\mathbf{z}$) on the partition $\tau[\ell]$ is through the number of its blocks ℓ and the index of its canonical decomposition ι^* with respect to the vector concatenation with s equal-length components.

Step 3. In order to progress towards the proof of Lemma 15 we need the following factorization property of the coefficients $\mathcal{B}(\kappa, \mathsf{e}, \tau[\ell])$, where $\tau[\ell]$ denotes a partition with ℓ blocks of the index set $\mathcal{X} = \{1, \dots, \kappa\}$. Let the block sizes of $\tau[\ell]$ be $\mathbf{m} = (m_1, \dots, m_\ell)$. Let $\ell = \ell_1 + \ell_2$, and define $\mathbf{m}_1 = (m_1, \dots, m_{\ell_1})$ and $\mathbf{m}_2 = (m_{\ell_1+1}, \dots, m_\ell)$. Then, $\tau[\ell] = \tau[\mathbf{m}, \ell]$ can be written as the Cartesian product of two partitions $\tau_1[\mathbf{m}_1, \ell_1]$ and $\tau_2[\mathbf{m}_2, \ell_2]$, where $\tau_1[\mathbf{m}_1, \ell_1]$ is a partition of the index set $\mathcal{X}_1 = \{1, \dots, \kappa_1\}$ with $\kappa_1 = \sum_{r=1}^{\ell_1} m_r$ and where $\tau_2[\mathbf{m}_2, \ell_2]$ is a partition of the index set $\mathcal{X}_2 = \{1, \dots, \kappa_2\}$ with $\kappa_2 = \sum_{r=1}^{\ell_2} m_{\ell_1+r}$.

Notice that any $\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]$, by definition, must be formed by the Cartesian product of some $\sigma_1[\mathbf{v}_1, w_1]$ and $\sigma_2[\mathbf{v}_2, w_2]$ such that $\sigma_1[\mathbf{v}_1, w_1] \leq \tau_1[\mathbf{m}_1, \ell_1]$ and $\sigma_2[\mathbf{v}_2, w_2] \leq \tau_2[\mathbf{m}_2, \ell_2]$.

Then, recalling the product rule for ζ (see (297)), we can write

$$\begin{aligned}
\mathcal{B}(\kappa, \mathsf{e}, \tau[\ell]) &= (1-\mathsf{e})^{\ell} + \sum_{w=\ell+1}^{\kappa} (1-\mathsf{e})^{w} \sum_{\sigma[w]\leq\tau[\ell]} \zeta(\sigma[w] \to \tau[\ell]) \\
&= \sum_{w=\ell}^{\kappa} (1-\mathsf{e})^{w} \sum_{\sigma[w]\leq\tau[\ell]} \zeta(\sigma[w] \to \tau[\ell]) \\
&= \sum_{w_1=\ell_1}^{\kappa_1} \sum_{w_2=\ell_2}^{\kappa_2} (1-\mathsf{e})^{w_1+w_2} \\
&\quad \sum_{\sigma_1[w_1]\leq\tau_1[\ell_1]} \sum_{\sigma_2[w_2]\leq\tau_2[\ell_2]} \zeta(\sigma_1[w_1] \to \tau_1[\ell_1])\zeta(\sigma_2[w_2] \to \tau_2[\ell_2]) \qquad (384) \\
&= \left(\sum_{w_1=\ell_1}^{\kappa_1} (1-\mathsf{e})^{w_1} \sum_{\sigma_1[w_1]\leq\tau_1[\ell_1]} \zeta(\sigma_1[w_1] \to \tau_1[\ell_1]) \right) \cdot \\
&\quad \left(\sum_{w_2=\ell_2}^{\kappa_2} (1-\mathsf{e})^{w_2} \sum_{\sigma_2[w_2]\leq\tau_2[\ell_2]} \zeta(\sigma_2[w_2] \to \tau_2[\ell_2]) \right) \\
&= \mathcal{B}(\kappa_1, \mathsf{e}, \tau_1[\ell_1])\mathcal{B}(\kappa_2, \mathsf{e}, \tau_2[\ell_2]) \qquad (385)
\end{aligned}$$

The above factorization can be trivially generalized to a d-fold Cartesian product. In particular, consider $\kappa = s(k+1)$ and a concatenated vector $(\mathbf{x}_1|\dots|\mathbf{x}_s)$ with equal-length components $\mathbf{x}_r$ of length $k+1$. Consider $\tau[\ell, (\mathbf{x}_1|\dots|\mathbf{x}_s)]$ and its canonical decomposition $\varpi(\boldsymbol{\nu}^*, \iota^*)$. The canonical representation of $\tau[\ell, (\mathbf{x}_1|\dots|\mathbf{x}_s)]$ with respect to the concatenation $(\mathbf{x}_1|\dots|\mathbf{x}_s)$ is given by (see Definition 10)

$$\tau[\ell, (\mathbf{x}_1|\dots|\mathbf{x}_s)] = \left\{ \tau_1[\ell_1, \mathbf{x}^{\mathcal{V}(\nu_1^*)}], \dots, \tau_{\iota^*}[\ell_{\iota^*}, \mathbf{x}^{\mathcal{V}(\nu_{\iota^*}^*)}] \right\}$$

For what said above, it follows that

$$\mathcal{B}(s(k+1), \mathsf{e}, \tau[\ell]) = \prod_{r=1}^{\iota^*} \mathcal{B}(\nu_r^*(k+1), \mathsf{e}, \tau_r[\ell_r]) \qquad (386)$$

This is the factorization we need in order to conclude the proof of Lemma 15.

Step 4. We consider in details the statement (352) relative to the variance, and leave to the reader the statement (353) relative to the 4-th central moment that follows from an analogous development and it is just much more cumbersome.

The trick consists of summing the expression for the s-th order moment (383) by dividing the partitions $\tau[\ell]$ into "shells" with the same canonical decomposition order $\iota^* = 1, 2, \dots, s$. Notice that these "shells" were introduced in general in Definition 11 and, by specializing the notation introduced in Definition 11 to the case at hands, they are denoted by $\mathfrak{D}^{(\iota^*)}(s(k+1))$.

Using (383) for $s = 2$ we have

$$\begin{aligned}
\mathbb{E}[\theta_{k,i}^2] &= \sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{2k-\ell+2}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{2k-\ell+\iota^*}} \prod_{r=1}^{2k-\ell+\iota^*} b_{z_r} \\
&= \sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]\in\mathfrak{D}^{(1)}(2(k+1))} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{2k-\ell+2}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{2k-\ell+1}} \prod_{r=1}^{2k-\ell+1} b_{z_r} + \\
&\quad + \sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]\in\mathfrak{D}^{(2)}(2(k+1))} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{2k-\ell+2}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{2k-\ell+2}} \prod_{r=1}^{2k-\ell+2} b_{z_r} \qquad (387)
\end{aligned}$$

Consider the second term of (387), that involves only partitions with $\iota^* = 2$. Using the facts in Step 3 above, we can write

$$\begin{gathered}
\textstyle\sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]\in\mathfrak{D}^{(2)}(2(k+1))} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{2k-\ell+2}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{2k-\ell+2}} \prod_{r=1}^{2k-\ell+2} b_{z_r} = \\
\textstyle\sum_{\ell_1=1}^{k+1} \sum_{\ell_2=1}^{k+1} \sum_{\tau_1[\ell_1]\in\mathfrak{P}_{\ell_1}(k+1)} \sum_{\tau_2[\ell_2]\in\mathfrak{P}_{\ell_2}(k+1)} \mathcal{B}(k+1, \mathsf{e}, \tau_1[\ell_1]) \mathcal{B}(k+1, \mathsf{e}, \tau_2[\ell_2]) \cdot \\
\textstyle\cdot \frac{1}{n^{k-\ell_1+1} n^{k-\ell_1+1}} \sum_{\mathbf{z}_1\in\mathbb{Z}_n^{k-\ell_1+1}} \sum_{\mathbf{z}_2\in\mathbb{Z}_n^{k-\ell_2+1}} \prod_{t=1}^{2} \prod_{r=1}^{k-\ell_t+1} b_{z_{t,r}} = \\
\left(\textstyle\sum_{\ell=1}^{k+1} \sum_{\tau[\ell]\in\mathfrak{P}_\ell(k+1)} \mathcal{B}(k+1, \mathsf{e}, \tau[\ell]) \frac{1}{n^{k-\ell+1}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{k-\ell+1}} \prod_{r=1}^{k-\ell+1} b_{z_r}\right)^2 = (\mathbb{E}[\theta_{k,i}])^2 \qquad (388)
\end{gathered}$$

where the last line follows from (360) and from (368).

Using (388) in (387) we obtain

$$\begin{aligned}
\mathrm{Var}\{\theta_{k,i}\} &= \mathbb{E}[\theta_{k,i}^2] - (\mathbb{E}[\theta_{k,i}])^2 \\
&= \sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]\in\mathfrak{D}^{(1)}(2(k+1))} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{1}{n^{2k-\ell+2}} \sum_{\mathbf{z}\in\mathbb{Z}_n^{2k-\ell+1}} \prod_{r=1}^{2k-\ell+1} b_{z_r} \\
&\leq \mathcal{K} \sum_{\ell=1}^{2(k+1)} \sum_{\tau[\ell]\in\mathfrak{D}^{(1)}(2(k+1))} \mathcal{B}(2(k+1), \mathsf{e}, \tau[\ell]) \frac{n^{2k-\ell+1}}{n^{2k-\ell+2}} \\
&\leq \frac{\mathcal{K}'}{n} \qquad (389)
\end{aligned}$$

for some constants $\mathcal{K}$ and $\mathcal{K}'$ that do not depend on n and on i, where the upperbound follows from the fact that the eigenvalues of $\mathbf{\Lambda}$, and hence the elements $b_0, \ldots, b_{n-1}$, are bounded (recall that $\mathbf{\Lambda}$ is the eigenvalue diagonal matrix of the Toeplitz matrix $\mathbf{\Sigma}$, that is bounded in strong norm). We conclude that $\mathrm{Var}\{\theta_{k,i}\} = O(1/n)$, and (352) is proved.

As said before, the proof of (353) follows from an analogous development, that is significantly more cumbersome even though it does not contain any new fact, and therefore is omitted for the sake of conciseness.

References

[1] P. Billingsley, *Probability and Measure*, New York, Wiley, 1986.

[2] F. de Bruno, '"Note sur une nouvelle formule de calcul differentiel", *Quart. J. Math.*, vol. 1, pp. 359-360, 1856.

[3] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[4] M. Debbah, W. Hachem, P. Loubaton and M. de Courville, '"MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[5] S. N. Diggavi and M. Grossglauser, "Information transmission over a finite buffer channel, " *IEEE Trans. on Information Theory*, vol 52, no. 3, pp 1226-1237, March 2006.

[6] R. M. Gray, "Toeplitz and Circulant Matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol 2, issue 3, pp 155-239, 2006

[7] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Information Theory*, vol. 51, no. 4, pp. 1261-1283, Apr. 2005.

[8] P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, NJ, 1950.

[9] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Mathematical Society, 2000.

[10] D. Julian, "Erasure networks," *2002 IEEE International Symp. Information Theory*, p. 138, Lausanne, Switzerland, June 30-July 5, 2002.

[11] ——, "Dependent and Gaussian erasure networks," *2003 IEEE International Symp. Information Theory*, p. 375, Yokohama, Japan, June 29 -July 4, 2003.

[12] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, New York: Academic Press, 1985.

[13] R. A. Oldshen, "Asymptotic Properties of the Periodogram of a Discrete Stationary Process," *Journal of Applied Probability*, vol. 4, no. 3, pp. 508–528, Dec. 1967.

[14] N. Pavlidou, A.J. Han Vinck, J. Yazdani, and B. Honary "Power line communications: state of the art and future trends," *IEEE Communications Magazine*, vol. 41, pp. 34–40, 2003.

[15] J. Riordan, *An Introduction to Combinatorial Analysis*, New York: Wiley, 1980.

[16] S. Shamai and E. Telatar and S. Verdú, "Fountain Capacity," *2006 IEEE International Symp. Information Theory*, Seattle, WA, July 2006.

[17] S. Shamai and S. Verdú, "The Effect of Frequency-flat Fading on the Spectral Efficiency of CDMA," *IEEE Trans. Information Theory,* vol. 47, no. 4, pp. 1302-1327, May 2001.

[18] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[19] R. P. Stanley, *Enumerative Combinatorics: Volume 1*, Cambridge University Press, April, 1997.

[20] A. Tulino and S. Verdú, "Random Matrix Theory and Wireless Communications," *Foundations and Trends In Communications and Information Theory*, vol. 1, no. 1, pp. 1–184, 2004.

[21] A. M. Tulino and A. Lozano and S. Verdú, "Capacity-Achieving Input Covariance for Single-user Multi-Antenna Channels," *IEEE Trans. on Wireless Communications*, vol. 5, no. 3, pp. 662–671, Mar. 2006.

[22] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[23] S. Verdú and T. S. Han, "A General Formula for Channel Capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[24] S. Verdú and T. Weissman, "The Information Lost in Erasures," *preprint*, 2007

[25] D. Voiculescu, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.

[26] A. D. Wyner, "Shannon-Theoretic Approach to a Gaussian Cellular Multiple Access Channel", *IEEE Trans. Inform. Theory,* Vol. 40, No. 6, pp. 1713-1727, Nov. 1994.