

# Capacity of Channels With Frequency-Selective and Time-Selective Fading

Antonia M. Tulino, *Senior Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, Shlomo Shamai, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

**Abstract**—This paper finds the capacity of single-user discrete-time channels subject to both frequency-selective and time-selective fading, where the channel output is observed in additive Gaussian noise. A coherent model is assumed where the fading coefficients are known at the receiver. Capacity depends on the first-order distributions of the fading processes in frequency and in time, which are assumed to be independent of each other, and a simple formula is given when one of the processes is independent identically distributed (i.i.d.) and the other one is sufficiently mixing. When the frequency-selective fading coefficients are known also to the transmitter, we show that the optimum normalized power spectral density is the waterfilling power allocation for a reduced signal-to-noise ratio (SNR), where the gap to the actual SNR depends on the fading distributions. Asymptotic expressions for high/low SNR and easily computable bounds on capacity are also provided.

**Index Terms**—Additive Gaussian noise, channel capacity, coherent communications, frequency-flat fading, frequency-selective fading, orthogonal frequency-division multiplexing (OFDM), random matrices, waterfilling.

## I. INTRODUCTION

THE simplest discrete-time additive-noise channel subject to fading is the time-selective coherent model

$$y_i = \sqrt{\gamma} A_i x_i + n_i, \quad i = 1, \dots, n \quad (1)$$

where the complex-valued input codeword  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{C}^n$  is subject to a unit average power constraint,  $\{n_i \in \mathbb{C}\}$  is a unit variance independent identically distributed (i.i.d.) complex Gaussian random process,  $\{A_i \in \mathbb{C}\}$  is a stationary ergodic

Manuscript received August 02, 2008; revised September 11, 2009. Current version published March 10, 2010. This work was supported by the U.S.–Israel Binational Science Foundation. The work of G. Caire was supported by the National Science Foundation (NSF) under Grant TF 0729162. The work of S. Verdú was supported by NSF under Grant TF 0728445. The material in this paper was presented in part at the 2010 Information Theory Workshop, Cairo, Egypt, January 6–8, 2010, and at the UCSD Workshop on Information Theory and Applications, San Diego, CA, January 31–February 5, 2010.

A. M. Tulino is with the Department of Wireless Communications, Bell Laboratories, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: a.tulino@alcatel-lucent.com).

G. Caire is with the University of Southern California, Los Angeles, CA 90089 USA (e-mail: caire@usc.edu).

S. Shamai is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

S. Verdú is with Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by L. Zheng, Associate Editor for Communications.

Color versions of Figures 2 and 7 in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2009.2039041

fading process known at the receiver and  $\gamma$  stands for signal-to-noise ratio (SNR). In vector form, (1) becomes

$$\mathbf{y} = \sqrt{\gamma} \mathbf{A} \mathbf{x} + \mathbf{n} \quad (2)$$

where  $\mathbf{A} = \text{diag}\{A_1, \dots, A_n\}$ . If the decoder (but not the encoder) knows the actual fading realization, the capacity of (1) is equal to [1]

$$C(\gamma) = \mathbb{E} [\log(1 + \gamma |A|^2)] \quad (3)$$

where the expectation is with respect to the random variable  $A$  distributed according to the first-order marginal distribution of the fading process  $\{A_i\}$ .

Another important model is the discrete-time frequency-selective fading channel, given by

$$\mathbf{y} = \sqrt{\gamma} \mathbf{F} \mathbf{G} \mathbf{F}^\dagger \mathbf{x} + \mathbf{n} \quad (4)$$

where  $\mathbf{F}$  is an  $n \times n$  unitary Fourier matrix with coefficients

$$\mathbf{F}_{i,k} = \frac{1}{\sqrt{n}} e^{-j\frac{2\pi}{n}(i-1)(k-1)}. \quad (5)$$

The columns of  $\mathbf{F}$  form an  $n$ -dimensional unitary discrete-time Fourier basis, and the fading coefficients affecting the transmitted signal frequency components are denoted by  $\mathbf{G} = \text{diag}\{G_1, \dots, G_n\}$ . Note that the random channel matrix  $\mathbf{F} \mathbf{G} \mathbf{F}^\dagger$  is circulant.

The model in (4) encompasses the random linear time-invariant channel

$$y_i = \sqrt{\gamma} \sum_{\ell=0}^L h_\ell x_{i-\ell} + z_i, \quad i = 1, \dots, n \quad (6)$$

where  $\{h_\ell\}$  denotes the (random) channel impulse response, under the assumption of cyclic prefix precoding and  $L \ll n$  [2].

In most physically meaningful frequency-selective models (see [1] and references therein) the diagonal coefficients of  $\mathbf{G}$  are identically distributed. If, moreover, they are cyclically stationary (the joint distribution is invariant to cyclic shifts), then the impulse response coefficients are uncorrelated, which is a common assumption. Using the fact that  $\mathbf{F}$  is unitary and under ergodicity and stationarity assumptions on the fading coefficients, the capacity of (4) is given by (again, assuming knowledge of  $\mathbf{G}$  at the decoder but not at the encoder)

$$C(\gamma) = \mathbb{E} [\log(1 + \gamma |G|^2)] \quad (7)$$

Both (3) and (7) are achieved by Gaussian i.i.d. input vectors  $\mathbf{x}$ . When the encoder knows  $\mathbf{G}$ , then it allocates power according

to the waterfilling formula [3]. In fact, in the familiar case of a deterministic linear time-invariant system with transfer function  $H(f)$ ,  $-1/2 \leq f \leq 1/2$ , the mutual information achieved by a stationary Gaussian input process with power spectral density  $S_x(f)$  is equal to the right side of (7) with  $G = S_x(U)|H(U)|^2$  and  $U$  uniformly distributed on  $[-\frac{1}{2}, \frac{1}{2}]$ .

A general discrete-time coherent fading model is given by the noisy version of the output of a linear time-varying system with random impulse response  $\{h_{i,\ell}\}$  known at the receiver

$$y_i = \sqrt{\gamma} \sum_{\ell=0}^L h_{i,\ell} x_{i-\ell} + z_i, \quad i = 1, \dots, n \quad (8)$$

or, equivalently, in vector form

$$\mathbf{y} = \sqrt{\gamma} \mathbf{H} \mathbf{x} + \mathbf{n} \quad (9)$$

where  $\mathbf{H}$  is the matrix representation of the convolution operator in (8). Subject to suitable stationarity and ergodicity assumptions on  $\{h_{i,\ell}\}$  the capacity is given by [1]

$$C(\gamma) = \lim_{n \rightarrow \infty} \sup_{\Sigma_x \geq 0: \text{tr}(\Sigma_x) \leq n} \frac{1}{n} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{H} \Sigma_x \mathbf{H}^\dagger)]. \quad (10)$$

A general closed-form formula for (10) in terms of the statistics of  $\{h_{i,\ell}\}$  has not been found yet either with or without knowledge of  $\mathbf{H}$  at the transmitter.

Since most mobile wireless systems are subject to both frequency-selective fading (e.g., due to multipath) and to time-selective fading (e.g., due to shadowing), it is of interest to consider a channel model that incorporates both effects. In this paper, we consider the following model (Fig. 1):

$$\mathbf{y} = \sqrt{\gamma} \mathbf{A} \mathbf{F} \mathbf{G} \mathbf{F}^\dagger \mathbf{x} + \mathbf{n} \quad (11)$$

obtained by concatenating a *random circulant* matrix  $\mathbf{F} \mathbf{G} \mathbf{F}^\dagger$ , with a time-domain diagonal fading matrix, where, as defined before,  $\mathbf{A}$  and  $\mathbf{G}$  are random diagonal matrices modeling the time-selective and frequency-selective fading coefficients, respectively. Note that (11) is a special case of (8), which captures some interesting features of time and frequency selectivity. For example, we may consider a case where signaling takes place over a set of orthogonal carriers [as in orthogonal frequency-division multiplexing (OFDM)], each attenuated by a random coefficient, with the whole signal then subject to a form of time-selective fading. Examples of time-selective (frequency-flat) fading include shadowing, impulsive noise/jamming that saturates the receiver input thereby erasing some of the received values [4], and satellite communication with the presence of a line-of-sight path modeled as a Markov chain [5].

Throughout this paper, we assume that the fading random processes  $\{A_i : i \in \mathbb{Z}\}$  and  $\{G_i : i \in \mathbb{Z}\}$  are mutually independent, stationary, and ergodic. Furthermore, either the time-domain fading or the frequency-domain fading is assumed to be i.i.d., while the other is *strong mixing* (Definition 12 in Appendix IV). We denote by  $A$  and  $G$  two independent random variables with the same first-order marginal distributions of  $\{A_i\}$  and  $\{G_i\}$ , respectively. Notice that  $A$  and  $G$  may have different distributions.

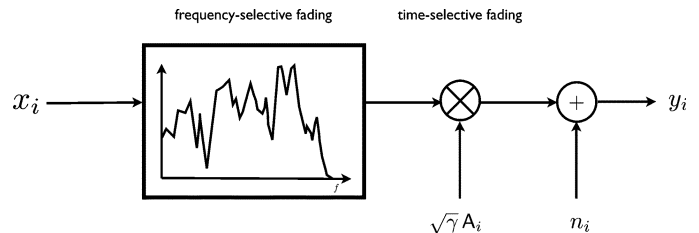


Fig. 1. Frequency-selective time-selective fading channel.

These are assumed to be sufficiently well behaved, such that all moments exist.

The main technical advance required to solve the capacity of the channel model (11) is the asymptotic spectral distribution of the matrix  $\mathbf{A} \Sigma \mathbf{A}^\dagger$ , when  $\Sigma$  is a random symmetric non-negative definite circulant matrix independent of  $\mathbf{A}$ . When the fading is known to the receiver only, the capacity is given by Theorem 1, which represents the main result of this paper. In Theorem 2, we show that when the frequency-domain fading is known also to the transmitter, the capacity achieving power allocation on the channel frequency components takes on the form of the well-known “waterfilling” solution for a scaled channel SNR, where the scaling coefficient can be characterized as the solution of a fixed-point equation. We also provide a number of easily computable upper and lower bounds to capacity, and simple formulas for the asymptotic behavior of capacity in the limits of small and large SNR are also presented.

The rest of this paper is organized as follows. Section II states the main results on capacity with fading coefficients known at the receiver only; on capacity when the frequency-selective fading is known also to the transmitter; on the bounds to capacity and on the low/high SNR asymptotic regimes. Section III presents some auxiliary results and the proofs of our main results, except those particularly technical, which are relegated to Appendixes I–IX. Finally, Section IV summarizes our conclusions.

## II. CHANNEL CAPACITY RESULTS

### A. Main Results

*Theorem 1:* The capacity of the channel model (11) with fading unknown to the transmitter is given by

$$C(\gamma) = \mathbb{E} [\log (1 + \alpha \gamma |G|^2)] + \mathbb{E} [\log (1 + \nu \gamma |A|^2)] - \log(1 + \alpha \nu \gamma) \quad (12)$$

where

$$0 \leq \alpha \leq \mathbb{E} [|A|^2] \quad (13)$$

$$0 \leq \nu \leq \mathbb{E} [|G|^2] \quad (14)$$

are coefficients that depend on  $\gamma$  and on the fading distributions, and are defined by the solution to

$$\mathbb{E} \left[ \frac{1}{1 + \alpha \gamma |G|^2} \right] = \frac{1}{1 + \alpha \nu \gamma} = \mathbb{E} \left[ \frac{1}{1 + \nu \gamma |A|^2} \right]. \quad (15)$$

*Proof:* See Section III-F. □

Notice the interesting duality between the frequency and time domains: the distributions of the random variables  $|A|^2$  and  $|G|^2$  play exactly the same role in the evaluation of capacity. In that respect, note that if  $\Sigma_{\mathbf{x}}$  in (10) is a multiple of the identity, the determinant is the same whether  $\mathbf{H} = \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{F}^\dagger$  or  $\mathbf{H} = \mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger\mathbf{F}$ . In the absence of time-domain fading ( $|A|^2$  deterministic) the solution to (15) satisfies  $\alpha = |A|^2$ . In the absence of frequency-domain fading ( $|G|^2$  deterministic) the solution to (15) satisfies  $\nu = |G|^2$ . Intuitively,  $\alpha$  captures the effect of time-domain fading variations, and  $\nu$  captures the effect of frequency-domain fading variations. Furthermore, (15) can be written as the pair of equations

$$\mathbb{E} \left[ \frac{|G|^2}{1 + \alpha\gamma|G|^2} \right] = \frac{\nu}{1 + \alpha\nu\gamma} \quad (16)$$

and

$$\mathbb{E} \left[ \frac{|A|^2}{1 + \nu\gamma|A|^2} \right] = \frac{\alpha}{1 + \alpha\nu\gamma}. \quad (17)$$

The left-hand side of (16) is the minimum mean square error (MMSE) for estimating a nonstationary independent Gaussian process  $\{w_i\}$  with variance  $|G_i|^2$  from the observation of  $\{G_i; y_i = \sqrt{\alpha\gamma}w_i + n_i : i \in \mathbb{Z}\}$ . Hence,  $\nu$  can be interpreted as the variance of a white Gaussian process  $\{w'_i\}$  that, if observed through the same additive white Gaussian noise (AWGN) channel, yields the same MMSE as  $\{w_i\}$ . The same observation holds for (17), exchanging  $\nu$  with  $\alpha$  and  $G_i$  with  $A_i$ .

Consider the following special cases of the setup in Section I.

- **Frequency-selective fading.** In the absence of time-domain fading ( $|A|^2 = 1$ ), the solution to (15) satisfies  $\alpha = 1$  and the second and third terms in (12) cancel, recovering (7).
- **Time-selective fading.** For a deterministic frequency-flat channel,  $|G|^2 = 1$ . Thus, (15) is solved by  $\nu = |G|^2 = 1$ , in which case the first and third terms in (12) cancel and we obtain (3).
- **Frequency-selective fading with on-off time-selective fading.** In the special case where  $|A|^2$  takes on the values 0 or 1 with probability  $e$  and  $1 - e$ , we obtain

$$C_e(\gamma) = \mathbb{E} \left[ \log \left( 1 + (1 - \hat{e})\gamma|G|^2 \right) \right] + d(e||\hat{e}) \quad (18)$$

where the binary divergence is defined as

$$d(a||b) = a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b} \quad (19)$$

and  $\hat{e} \geq e$  is the ( $\gamma$ -dependent) solution to

$$\frac{e}{\hat{e}} = \mathbb{E} \left[ \frac{1}{1 + (1 - \hat{e})\gamma|G|^2} \right]. \quad (20)$$

Note that in the special case in which the frequency selective fading is also ON-OFF, i.e.,  $|G|$  takes values 0 and 1, (20) becomes a quadratic equation and  $\hat{e}$  admits a closed-form solution [4].

- **Independent Rayleigh fading and Markov-correlated shadowing.** Here we have i.i.d. Rayleigh fading in the frequency domain and a two-state Markov shadowing process

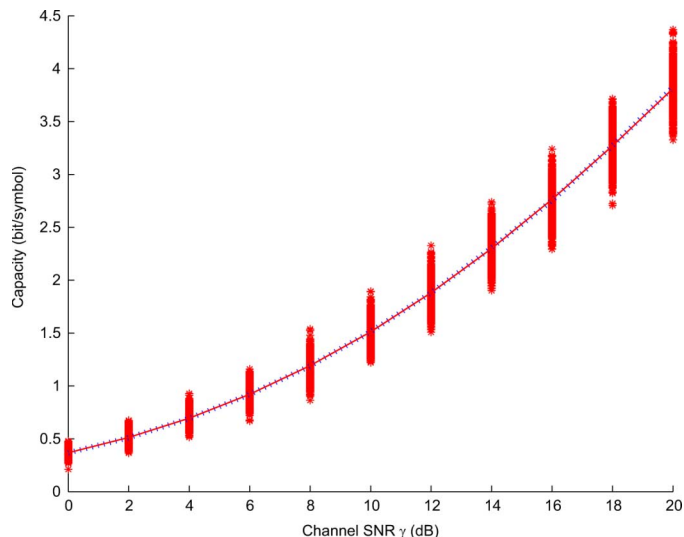


Fig. 2. Rayleigh frequency-selective fading and two-state Markov shadowing with  $a_0 = 0.1$ ,  $a_1 = 1$  and with transition probabilities  $\mathbb{P}[a_0 \rightarrow a_1] = 0.3$  and  $\mathbb{P}[a_1 \rightarrow a_0] = 0.7$ . Solid line: solution to Theorem 1. Dotted line: Monte Carlo evaluation of (24) for  $n = 64$ . The clouds of points correspond to realizations of the random variable inside the expectation in (24). (1000 points per cluster).

in the time domain. In particular,  $|G_i|^2$  is a sequence of independent exponential random variables with mean 1, and  $|A_i|^2$  is a Markov chain with two states  $a_0$  and  $a_1$  with stationary distribution  $(\pi_0, \pi_1)$ . In order to solve (15) for  $\alpha$  and  $\nu$  as a function of  $\gamma$ , we proceed as follows. For any given  $\eta \in [0, 1]$ , let  $\alpha$  be the solution of the equation

$$\mathbb{E} \left[ \frac{1}{1 + \gamma\alpha|G|^2} \right] = \frac{e^{\frac{1}{\gamma\alpha}}}{\gamma\alpha} \int_{1/\alpha\gamma}^{\infty} \frac{e^{-t}}{t} dt = \eta \quad (21)$$

and let  $\nu$  be the solution of the equation

$$\mathbb{E} \left[ \frac{1}{1 + \gamma\nu|A|^2} \right] = \frac{\pi_0}{1 + \gamma\nu a_0} + \frac{\pi_1}{1 + \gamma\nu a_1} = \eta \quad (22)$$

Then, using the second equality in (15), we find (e.g., using the bisection method) the value of  $\eta$  that satisfies

$$\alpha\nu\gamma\eta = 1 - \eta. \quad (23)$$

Finally, using the values of  $\alpha$  and  $\nu$  and  $\eta$  so obtained, calculate  $C(\gamma)$  using (12). Fig. 2 shows the comparison between  $C(\gamma)$  and Monte Carlo simulation of the finite-dimensional mutual information formula

$$\frac{1}{n} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger \right) \right] \quad (24)$$

for  $n = 64$ . We also show the realization of the normalized log-det without the expectation, in order to give an idea of the spread of the finite-dimensional mutual information for given (random) realization of the fading processes. We notice that the agreement between simulation and the result of Theorem 1 is remarkable for even a relatively small value of  $n$ .

### B. Optimality of Waterfilling With Power Penalty

If the transmitter knows the frequency-domain fading coefficients, then it can choose the input covariance matrix  $\Sigma_x$  as a function of  $\mathbf{G}$  in order to maximize the mutual information. It is sufficient to consider a circulant input covariance in the form  $\Sigma_x = \mathbf{F}\mathbf{Y}\mathbf{F}^\dagger$ . In the absence of time-domain fading, maximizing the mutual information of the frequency-selective fading channel given in (6) and (4) with respect to the input power spectral density yields (e.g., [6]) the well-known waterfilling formula

$$C(\gamma) = \mathbb{E} [\log (1 + \gamma \bar{S}_x(\gamma, |G|^2) |G|^2)] \quad (25)$$

$$= \mathbb{E} \left[ \left[ \log (\zeta_\gamma \gamma |G|^2) \right]^+ \right] \quad (26)$$

where  $\bar{S}_x(\gamma, \cdot)$  is the waterfilling power allocation function

$$\bar{S}_x(\gamma, z) = \left[ \zeta_\gamma - \frac{1}{\gamma z} \right]^+, \quad z \in \mathbb{R}_+ \quad (27)$$

and the water level  $1 < \zeta_\gamma < \infty$  is chosen in order to satisfy the transmit power constraint, i.e., such that

$$\mathbb{E} \left[ \left[ \zeta_\gamma - \frac{1}{\gamma |G|^2} \right]^+ \right] = 1. \quad (28)$$

The input power spectral density is implicitly given by the function  $\bar{S}_x(\gamma, \cdot)$  and by the realization of the frequency-selective fading ( $|G_1|^2, \dots, |G_n|^2$ ). In particular, for any given blocklength  $n$  and discrete Fourier transform (DFT) frequencies  $\{f_i = (i-1)/n : i = 1, \dots, n\}$ , the input energy associated with the  $i$ th frequency component is given by  $S_x(f_i) = \bar{S}_x(\gamma, |G_i|^2)$  and satisfies

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[S_x(f_i)] = 1. \quad (29)$$

Letting  $n \rightarrow \infty$  yields the power spectral density defined on the discrete-time frequency domain, that without loss of generality can be taken to be the interval  $[0, 1]$ .<sup>1</sup>

In the presence of time-domain fading, the capacity-achieving input power spectral density is defined by the optimal power allocation function  $S_x^*(\gamma, \cdot)$ , given implicitly by the following result.

**Theorem 2:** For all  $\gamma > 0$  and  $z \geq 0$ , the capacity-achieving input power spectral density is given by

$$S_x^*(\gamma, z) = \bar{S}_x(\gamma', z) \quad (30)$$

where  $\bar{S}_x(\cdot, \cdot)$  is the waterfilling power allocation in (27) and  $\gamma = \mu(\gamma')\gamma'$  with  $\mu(\gamma') \geq 1$  being the solution of the equation

$$\mathbb{E} \left[ \frac{\zeta_{\gamma'}}{\zeta_{\gamma'} - 1 + \mu |A|^2} \right] = 1 \quad (31)$$

<sup>1</sup>Recall that the Fourier transform of discrete-time signals is periodic of period 1.

where  $\zeta_{\gamma'}$  is the fading-free water level in (27) for the reduced SNR  $\gamma'$ .

*Proof:* See Section III-H.  $\square$

We notice that the power allocation function  $S_x^*(\cdot, \cdot)$  coincides with the waterfilling power allocation function  $\bar{S}_x(\cdot, \cdot)$  for the case without time-domain fading, calculated for a lower value of the SNR parameter: namely,  $\gamma'$  instead of  $\gamma$ . In order to evaluate the capacity  $C(\gamma)$  when the frequency-domain fading is known to the transmitter, we search for the value  $\gamma' \in (0, \gamma]$  such that  $\gamma = \mu(\gamma')\gamma'$ . Then,  $C(\gamma)$  is equal to (12) with the modified fading random variable given by  $|G|^2 \left[ \zeta_{\gamma'} - \frac{1}{\gamma' |G|^2} \right]^+$ .

### C. Bounds

**Theorem 3:** The capacity (12) is lower bounded by

$$C(\gamma) \geq \mathbb{E} [\log (1 + \gamma |A|^2 |G|^2)]. \quad (32)$$

*Proof:* See Section III-I.  $\square$

**Theorem 4:** The capacity in (12) is lower bounded by

$$C(\gamma) \geq \mathbb{E} [\log (1 + \alpha \gamma |G|^2)] \quad (33)$$

$$C(\gamma) \geq \mathbb{E} [\log (1 + \nu \gamma |A|^2)] \quad (34)$$

$$C(\gamma) \geq \log(1 + \nu \alpha \gamma) \quad (35)$$

where  $(\alpha, \nu)$  are given in Theorem 1.

*Proof:* See Section III-J.  $\square$

The following result yields upper bounds to capacity and shows that in the presence of one type of fading, the fading in the other domain is deleterious.

**Theorem 5:** The capacity in (12) is upper bounded by

$$C(\gamma) \leq \mathbb{E} [\log (1 + \gamma \mathbb{E}[|G|^2] |A|^2)] \quad (36)$$

$$C(\gamma) \leq \mathbb{E} [\log (1 + \gamma |G|^2 \mathbb{E}[|A|^2])]. \quad (37)$$

*Proof:* See Section III-K.  $\square$

### D. Asymptotics

**1) Low SNR Asymptotics:** In this section, we characterize the behavior of capacity for vanishing  $\gamma$ . We define the *kurtosis* of a real random variable  $Z$  as

$$\kappa(Z) = \frac{\mathbb{E}[Z^4]}{\mathbb{E}^2[Z^2]}. \quad (38)$$

**Theorem 6:** When the frequency-selective fading is known to the transmitter, the minimum energy per bit and the wideband slope  $\mathcal{S}_0$  [7] of the spectral efficiency of channel (11) are given as follows. In the case of no channel state information at the transmitter

$$\left( \frac{E_b}{N_0} \right)_{\min} = \frac{\ln 2}{\mathbb{E}[|A|^2] \mathbb{E}[|G|^2]} \quad (39)$$

$$\mathcal{S}_0 = \frac{2}{\kappa(|G|) + \kappa(|A|) - 1}. \quad (40)$$

When the transmitter knows the frequency-domain fading coefficients, then

$$\left(\frac{E_b}{N_0}\right)_{\min} = \frac{\ln 2}{\mathbb{E}[|A|^2]G_{\max}} \quad (41)$$

$$S_0 = \frac{2}{\kappa(|A|) + \frac{1}{B_{\max}} - 1} \quad (42)$$

where  $G_{\max}$  is the essential supremum of the frequency-selective fading, defined as

$$G_{\max} = \sup\{z : \mathbb{P}(|G|^2 \leq z) < 1\} \quad (43)$$

and  $B_{\max} = \mathbb{P}(|G|^2 = G_{\max})$  is the probability mass at  $G_{\max}$ .<sup>2</sup>

*Proof:* See Section III-L.  $\square$

We notice that  $G_{\max}$  takes on the meaning of the ‘‘peak’’ of the frequency-domain channel fading transfer function and  $B_{\max}$  corresponds to the ‘‘bandwidth’’ (i.e., the probability measure of the set of frequencies) over which the fading takes on its maximum value. When the transmitter has knowledge of the frequency-domain fading channel, the optimal power allocation of Theorem 2 puts constant power  $1/B_{\max}$  over the frequency components  $f_i = (i - 1)/n$  for which  $|G_i|^2 = G_{\max}$  and zero power elsewhere. This explains the quite different behavior of  $(E_b/N_0)_{\min}$  and  $S_0$  in the cases of unknown or known frequency-domain fading at the transmitter.

2) *High-SNR Asymptotics:* The following result finds the high-SNR slope  $S_\infty$  and the high-SNR decibel offset  $\mathcal{L}_\infty$  (see [8]). For the sake of brevity, we give the results in the case of no fading knowledge at the transmitter, for which the optimal input is i.i.d.

*Theorem 7:* Let  $u_0 = \mathbb{P}(|G|^2 = 0)$  and  $v_0 = \mathbb{P}(|A|^2 = 0)$  denote the masses at 0 of the two fading distributions. If  $u_0 > v_0$ , define  $\varsigma$  by

$$u_0 = \mathbb{E}\left[\frac{1}{1 + \varsigma|A|^2}\right]. \quad (44)$$

If  $u_0 < v_0$ , define  $\psi$  by

$$v_0 = \mathbb{E}\left[\frac{1}{1 + \psi|G|^2}\right]. \quad (45)$$

For large SNR, the capacity (in bits/complex dimension), when the transmitter has no knowledge of the fading realization, behaves like

$$C(\gamma) = S_\infty (\log_2 \gamma - \mathcal{L}_\infty) + o(1) \quad (46)$$

<sup>2</sup>Notice that, depending on the distribution of  $|G|^2$ ,  $G_{\max}$  may be equal to  $+\infty$ . Also,  $B_{\max} = 0$  if the cumulative density function (cdf) of  $|G|^2$  has no probability mass at  $G_{\max}$ .

where

$$S_\infty = 1 - \max\{u_0, v_0\} \quad (47)$$

and where (48), shown at the bottom of the page, holds, where  $h(\cdot)$  is the binary entropy function (in bits) and the high-SNR offsets in the absence of time-domain and frequency-domain fading are given by, respectively

$$\mathcal{L}_\infty^{\text{no-tdf}} = -\mathbb{E}[\log_2 |G|^2 | |G|^2 > 0] \quad (49)$$

$$\mathcal{L}_\infty^{\text{no-fdf}} = -\mathbb{E}[\log_2 |A|^2 | |A|^2 > 0]. \quad (50)$$

*Proof:* See Section III-L.  $\square$

Looking at the channel model (11), it is expected that the high-SNR slope  $S_\infty$ , also referred to as ‘‘multiplexing gain,’’ or ‘‘pre-log’’ factor of capacity, is given by the asymptotic normalized rank of the matrix  $\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{F}^\dagger$ . In fact, the asymptotic normalized rank is given by  $\frac{1}{n} \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{G})\}$ , which converges almost surely to (47).

### III. PROOFS AND AUXILIARY RESULTS

In this section, we recall some useful definitions in random matrix theory and we give several analytical properties of the solution in Theorem 1, as well as an alternative representation. Then, we proceed to the proofs of the main results. In particular, the main technical result is given in Theorem 11 and Lemma 1 of Section III-G.

#### A. Transforms in Random Matrix Theory [9]

*Definition 1:* The  $\eta$ -transform of a nonnegative random variable  $X$  is

$$\eta_X(\gamma) = \mathbb{E}\left[\frac{1}{1 + \gamma X}\right] \quad (51)$$

with  $\gamma \geq 0$ .

Note that

$$\mathbb{P}(X = 0) < \eta_X(\gamma) \leq 1 \quad (52)$$

with the lower bound asymptotically tight as  $\gamma \rightarrow \infty$ .

*Definition 2:* The Shannon transform of a nonnegative random variable  $X$  is defined as

$$\mathcal{V}_X(\gamma) = \mathbb{E}[\log(1 + \gamma X)] \quad (53)$$

with  $\gamma \geq 0$ .

$$\mathcal{L}_\infty = \begin{cases} \mathcal{L}_\infty^{\text{no-tdf}} + \frac{h(u_0)}{1 - u_0} + \log_2 \varsigma - \frac{1}{1 - u_0} \mathbb{E}[\log_2(1 + \varsigma|A|^2)], & u_0 > v_0 \\ \mathcal{L}_\infty^{\text{no-tdf}} + \mathcal{L}_\infty^{\text{no-fdf}} + \frac{h(u_0)}{1 - u_0} & u_0 = v_0 \\ \mathcal{L}_\infty^{\text{no-fdf}} + \frac{h(v_0)}{1 - v_0} + \log_2 \psi - \frac{1}{1 - v_0} \mathbb{E}[\log_2(1 + \psi|G|^2)] & u_0 < v_0 \end{cases} \quad (48)$$

Assuming that the logarithm in (53) is natural, the  $\eta$  and Shannon transforms are related through

$$\frac{d}{d\gamma} \mathcal{V}_X(\gamma) = \frac{1 - \eta_X(\gamma)}{\gamma}. \quad (54)$$

Also, it is useful to recall here the definition of the  $S$ -transform of free probability (see [9] and references therein), which is used in some of the proofs that follow.

*Definition 3:* The  $S$ -transform of a nonnegative random variable  $X$  is defined as

$$\Sigma_X(z) = -\frac{z+1}{z} \eta_X^{-1}(z+1) \quad (55)$$

where  $\eta_X^{-1}(\cdot)$  denotes the inverse function of the  $\eta$ -transform.

It is common to denote the  $\eta$ -transform, the Shannon transform and the  $S$ -transform of the spectral distribution of a sequence of nonnegative-definite  $n \times n$  random matrices  $\mathbf{B}$ , for  $n \rightarrow \infty$ , by  $\eta_{\mathbf{B}}(\cdot)$ ,  $\mathcal{V}_{\mathbf{B}}(\cdot)$ , and  $\Sigma_{\mathbf{B}}(\cdot)$ , respectively. In this case, the lower bound in (52) corresponds to the limiting fraction of zero eigenvalues of  $\mathbf{B}$ .

### B. Properties of the Solution in Theorem 1

- Fix  $\gamma$ , and denote the right-hand side of (12) by  $I(\alpha, \nu)$ . Then, the solution to (15) is a stationary point of  $I(\alpha, \nu)$ .
- Let  $u_0, v_0, \psi$ , and  $\varsigma$  be defined as in Theorem 7. As  $\gamma \rightarrow \infty$ , the solution to (15) becomes

$$\lim_{\gamma \rightarrow \infty} \frac{1}{1 + \gamma\alpha\nu} = \max\{u_0, v_0\} \quad (56)$$

Case  $u_0 > v_0$

$$\lim_{\gamma \rightarrow \infty} \gamma\nu = \varsigma \quad (57)$$

$$\lim_{\gamma \rightarrow \infty} \alpha = \frac{1 - u_0}{\varsigma u_0}. \quad (58)$$

Case  $u_0 < v_0$

$$\lim_{\gamma \rightarrow \infty} \gamma\alpha = \psi \quad (59)$$

$$\lim_{\gamma \rightarrow \infty} \nu = \frac{1 - v_0}{\psi v_0}. \quad (60)$$

Case  $u_0 = v_0$

$$\lim_{\gamma \rightarrow \infty} \gamma\nu = \lim_{\gamma \rightarrow \infty} \gamma\alpha = \infty. \quad (61)$$

- As  $\gamma \rightarrow 0$ , the solution to (15) converges to

$$\lim_{\gamma \rightarrow 0} \alpha(\gamma) = \mathbb{E}[|A|^2] \quad (62)$$

$$\lim_{\gamma \rightarrow 0} \nu(\gamma) = \mathbb{E}[|G|^2]. \quad (63)$$

- Applying Jensen's inequality to both identities in (15), we obtain the right inequalities in (13) and (14).

### C. Alternative Characterization

We give an alternative characterization of capacity that hinges on the positive function  $\mathfrak{J}_0(y, \gamma)$ ,  $0 \leq y \leq 1$ , defined as the solution of the fixed-point equation

$$\frac{1}{1 + \mathfrak{J}_0(y, \gamma)} = \mathbb{E} \left[ \frac{1}{1 + y\gamma|G|^2 + (1-y)\mathfrak{J}_0(y, \gamma)} \right]. \quad (64)$$

*Theorem 8:* The capacity of (12) can be written in the alternative form

$$C(\gamma) = \int_0^1 \log(1 + \mathfrak{J}_0(y, \bar{\gamma}(y))) dy \quad (65)$$

where, for given  $\gamma$  and  $y$ ,  $\bar{\gamma}(y)$  is the value of  $\bar{\gamma}$  that solves the equation

$$\frac{1}{1 + \mathfrak{J}_0(y, \bar{\gamma})} = \mathbb{E} \left[ \frac{\bar{\gamma}}{\bar{\gamma} + \mathfrak{J}_0(y, \bar{\gamma})(\bar{\gamma}(1-y) + \gamma y|A|^2)} \right]. \quad (66)$$

*Proof:* See Section III-D.  $\square$

While in Theorem 1 the time-domain and frequency-domain fading play symmetric roles, in Theorem 8 their role is asymmetric. Of course, a completely equivalent formulation of Theorem 8 can be obtained by exchanging the roles of  $|G|^2$  and  $|A|^2$ . The two alternative forms of Theorem 8 may facilitate computation depending on the distributions of  $|G|^2$  and  $|A|^2$ .

The following general auxiliary result is quite useful in the proof of Theorem 8.

*Theorem 9:* Let  $X$  be a nonnegative random variable and let  $U$  be uniformly distributed on  $[0, 1]$ . For each  $u \in (0, 1]$ , let  $g(u) \in [0, \infty)$  be the solution to

$$\frac{1}{1 + g(u)} = \mathbb{E} \left[ \frac{1}{1 + (1-u)g(u) + uX} \right]. \quad (67)$$

Then

$$\mathbb{E}[\log(1 + g(U))] = \mathbb{E}[\log(1 + X)]. \quad (68)$$

*Proof:* See Section III-D.  $\square$

- It is not difficult to show that  $\mathfrak{J}_0(y, \gamma)$  is monotonically decreasing with  $y$  for fixed  $\gamma$  and monotonically increasing with  $\gamma$  for fixed  $y$ .
- Applying Jensen's inequality to (66) yields that

$$\bar{\gamma}(y, \gamma) \leq \gamma \mathbb{E}[|A|^2]. \quad (69)$$

- Theorem 9 can be stated alternatively as the following result of independent interest tying the Shannon transform and the  $\eta$ -transform.

*Theorem 10:* The Shannon transform of a nonnegative random variable  $X$ , defined in (2), is given by

$$\mathcal{V}_X(\gamma) = \int_0^1 \log(1 + \mathfrak{J}(y, \gamma)) dy \quad (70)$$

where  $\mathfrak{J}$  is defined by the fixed-point equation

$$\frac{1}{1 + \mathfrak{J}(y, \gamma)} = \mathbb{E} \left[ \frac{1}{1 + (1 - y)\mathfrak{J}(y, \gamma) + \gamma y X} \right] \quad (71)$$

$$= \frac{1}{1 + (1 - y)\mathfrak{J}(y, \gamma)} \eta_X \left( \frac{\gamma y}{1 + (1 - y)\mathfrak{J}(y, \gamma)} \right). \quad (72)$$

*Proof:* Using the definition of Shannon transform (see Definition 2) and simply replacing in Theorem 9  $g(u)$  with  $\mathfrak{J}(y, \gamma)$  and  $X$  with  $\gamma X$ , we obtain (71). Finally, (72) is obtained via straightforward manipulation of the definition of the  $\eta$ -transform (see Definition 1).  $\square$

- Straightforward algebra reveals that (64) is equivalent to the fixed-point equation

$$\frac{\mathfrak{J}_0(y, \gamma)}{1 + \mathfrak{J}_0(y, \gamma)} = \mathbb{E} \left[ \frac{\gamma |G|^2}{1 + y\gamma |G|^2 + (1 - y)\mathfrak{J}_0(y, \gamma)} \right]. \quad (73)$$

Using (73), accounting for the monotonicity of  $\mathfrak{J}_0(\cdot, \gamma)$  and taking the limit as  $y \rightarrow 0$ , we obtain that for any  $\gamma > 0$

$$\mathfrak{J}_0(y, \gamma) \leq \mathfrak{J}_0(0, \gamma) = \gamma \mathbb{E}[|G|^2]. \quad (74)$$

#### D. Proof of Theorem 9

Let  $\mathbf{D}_X = \text{diag}(X_1, \dots, X_n)$  denote an  $n \times n$  diagonal matrix with random i.i.d. diagonal entries distributed as the non-negative random variable  $X$ , and let  $\mathbf{U}$  be an arbitrary unitary matrix (i.e.,  $\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$ ). Let  $\mathbf{B} = \mathbf{D}_X^{\frac{1}{2}}\mathbf{U}$  and let  $\mathbf{b}_i$  denote the  $i$ th column of  $\mathbf{B}$ . For an arbitrary  $n$ , we have

$$\begin{aligned} & \mathbb{E} [\log(1 + X)] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\log(1 + X_i)] \end{aligned} \quad (75)$$

$$= \frac{1}{n} \mathbb{E} [\log(\det(\mathbf{I} + \mathbf{D}_X))] \quad (76)$$

$$= \frac{1}{n} \mathbb{E} \left[ \log \left( \det \left( \mathbf{I} + \mathbf{D}_X^{\frac{1}{2}} \mathbf{U} \mathbf{U}^\dagger \mathbf{D}_X^{\frac{1}{2}} \right) \right) \right] \quad (77)$$

$$= \frac{1}{n} \mathbb{E} \left[ \log \left( \det \left( \mathbf{I} + \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\dagger \right) \right) \right] \quad (78)$$

$$= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \log \left( 1 + \mathbf{b}_i^\dagger \left( \mathbf{I} + \sum_{j=1}^{i-1} \mathbf{b}_j \mathbf{b}_j^\dagger \right)^{-1} \mathbf{b}_i \right) \right] \quad (79)$$

where (79) follows from the chain-rule of mutual information of an appropriate Gaussian vector model.<sup>3</sup>

Assume now that  $\mathbf{U}$  is uniformly distributed on the set of  $n \times n$  unitary matrices (i.e., it is a Haar matrix); then, as  $n \rightarrow \infty$

<sup>3</sup>In fact, (79) can be shown from purely matrix-theoretic arguments. However, it is nice to see it in terms of the chain-rule decomposition of the following mutual information. Consider the vector Gaussian model  $\mathbf{y} = \mathbf{B}\mathbf{s} + \mathbf{z}$ , where  $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ ,  $\mathbf{s} \sim \mathcal{CN}(0, 1)$ , and  $\mathbf{B}$  is fixed. Then,  $I(\mathbf{s}; \mathbf{y}) = \log \det(\mathbf{I} + \sum_{i=1}^n \mathbf{b}_i \mathbf{b}_i^\dagger)$ ,  $I(s_i; \mathbf{y} | s_{i+1}, \dots, s_n) = \log(1 + \mathbf{b}_i^\dagger (\mathbf{I} + \sum_{j=1}^{i-1} \mathbf{b}_j \mathbf{b}_j^\dagger)^{-1} \mathbf{b}_i)$  and the chain rule yields  $I(\mathbf{s}; \mathbf{y}) = \sum_{i=1}^n I(s_i; \mathbf{y} | s_{i+1}, \dots, s_n)$ , from which the identity follows.

the quadratic form in (79) converges almost surely to a deterministic quantity that can be computed via a fixed-point equation. Specifically, using the result in [9, eq. 3.112], we have that

$$\lim_{n \rightarrow \infty} \mathbf{b}_i^\dagger \left( \mathbf{I} + \sum_{j=1}^i \mathbf{b}_j \mathbf{b}_j^\dagger \right)^{-1} \mathbf{b}_i = \frac{1 - \eta}{u} \quad (80)$$

where  $\eta$  is the solution of

$$\eta = \eta_X \left( \frac{u - 1 + \eta}{\eta} \right) \quad (81)$$

and where  $u \in [0, 1]$  is such that  $i = \lceil nu \rceil$ . Using the matrix inversion lemma [10], we can write

$$\mathbf{b}_i^\dagger \left( \mathbf{I} + \sum_{j=1}^i \mathbf{b}_j \mathbf{b}_j^\dagger \right)^{-1} \mathbf{b}_i = \frac{\mathbf{b}_i^\dagger \left( \mathbf{I} + \sum_{j=1}^{i-1} \mathbf{b}_j \mathbf{b}_j^\dagger \right)^{-1} \mathbf{b}_i}{1 + \mathbf{b}_i^\dagger \left( \mathbf{I} + \sum_{j=1}^{i-1} \mathbf{b}_j \mathbf{b}_j^\dagger \right)^{-1} \mathbf{b}_i}. \quad (82)$$

Hence, as  $n \rightarrow \infty$ , we have that the quadratic form  $\mathbf{b}_i^\dagger (\mathbf{I} + \sum_{j=1}^{i-1} \mathbf{b}_j \mathbf{b}_j^\dagger)^{-1} \mathbf{b}_i$  converges to a deterministic limit  $g(u)$ , that satisfies

$$\frac{1 - \eta}{u} = \frac{g(u)}{1 + g(u)}. \quad (83)$$

Eliminating  $\eta$  from (81) and (83), we obtain

$$1 - \frac{ug(u)}{1 + g(u)} = \eta_X \left( \frac{u}{1 + (1 - u)g(u)} \right) \quad (84)$$

which is equivalent to (67). Using the limit  $g(u)$  in (79), we obtain (68) as desired.

#### E. Proof of Theorem 8

When the transmitter has no knowledge of the fading realization, the optimal input covariance is  $\Sigma_x = \mathbf{I}$  (see Appendix I, Theorem 13). It follows that

$$C(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{x}; \mathbf{y} | \mathbf{A}, \mathbf{G}) \quad (85)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log \det(\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (86)$$

$$= \mathcal{V}_{\mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger}(\gamma) \quad (87)$$

$$= \int_0^1 \log(1 + \mathfrak{J}(y, \gamma)) dy \quad (88)$$

where according to Theorem 10,  $\mathfrak{J}(y, \gamma)$  is defined by the fixed-point equation

$$y \frac{\mathfrak{J}(y, \gamma)}{1 + \mathfrak{J}(y, \gamma)} = 1 - \eta_{\mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger} \left( \frac{\gamma y}{1 + (1 - y)\mathfrak{J}(y, \gamma)} \right). \quad (89)$$

The  $\eta$ -transform  $\eta_{\mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger}(\gamma)$  is given in Theorem 11 in Section III-G. Fix  $y$  and  $\gamma$  and consider the result of Theorem 11 for

$$\tau = \frac{\gamma y}{1 + (1 - y)\mathfrak{J}(y, \gamma)} \quad (90)$$

and denote the corresponding  $\alpha$  therein by  $\alpha(y, \gamma)$ . Furthermore, define

$$\bar{\gamma}(y, \gamma) = \gamma \alpha(y, \gamma) \quad (91)$$

$$\mathfrak{J}_0(y, \bar{\gamma}(y, \gamma)) = \mathfrak{J}_0(y, \gamma). \quad (92)$$

The remaining task is to show that both fixed-point equations (64) and (66) are satisfied by (91) and (92). To that end, putting together (89) and (105), we obtain (dropping the arguments of  $\bar{\gamma}$ )

$$\frac{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})}{1 + \mathfrak{J}_0(y, \bar{\gamma})} = 1 - y \frac{\mathfrak{J}_0(y, \bar{\gamma})}{1 + \mathfrak{J}_0(y, \bar{\gamma})} \quad (93)$$

$$\begin{aligned} &= \eta_{\mathbf{G}\mathbf{G}^\dagger} \left( \frac{\bar{\gamma}y}{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})} \right) \quad (94) \\ &= \mathbb{E} \left[ \frac{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})}{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma}) + \bar{\gamma}y|G|^2} \right] \quad (95) \end{aligned}$$

which is equivalent to (64) since  $\bar{\gamma}(y, \cdot)$  is a one-to-one mapping of the positive real line.

Using (89) and (108) and (107), we can write the product  $\tau\nu$ , argument of the  $\eta$ -transform in (106) as

$$\frac{1 - \eta}{\alpha(y, \gamma)\eta} = \frac{1}{\alpha(y, \gamma)} \frac{y\mathfrak{J}_0(y, \bar{\gamma})}{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})}. \quad (96)$$

Thus, (89) and (106) lead to

$$\begin{aligned} \frac{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})}{1 + \mathfrak{J}_0(y, \bar{\gamma})} &= \eta_{\mathbf{A}\mathbf{A}^\dagger} \left( \frac{1}{\alpha(y, \gamma)} \frac{y\mathfrak{J}_0(y, \bar{\gamma})}{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})} \right) \quad (97) \\ &= \mathbb{E} \left[ \frac{1}{1 + |A|^2 \frac{1}{\alpha(y, \gamma)} \frac{y\mathfrak{J}_0(y, \bar{\gamma})}{1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})}} \right] \quad (98) \end{aligned}$$

which upon dividing both sides by  $1 + (1-y)\mathfrak{J}_0(y, \bar{\gamma})$  is readily seen to be equivalent to (66) in view of (91).

#### F. Proof of Theorem 1

Theorem 11 yields  $\eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}$ . In order to find the corresponding Shannon transform in terms of the solution of a fixed-point equation, we follow an idea originated in [8]: for any differentiable function  $f$ , the definition of the Shannon transform of an arbitrary nonnegative random variable  $X$  leads to

$$\frac{d}{dx} \mathcal{V}_X(xf(x)) = \mathbb{E} \left[ \frac{x\dot{f}(x)X + f(x)X}{1 + xf(x)X} \right]. \quad (99)$$

Since both sides of (12) are equal to zero at  $\gamma = 0$ , it is sufficient to show that the derivatives with respect to  $\gamma$  of both sides of (12) coincide. The derivative of the right side minus the left-hand side of (12) is equal to

$$\begin{aligned} &\mathbb{E} \left[ \frac{(\dot{\alpha}\gamma + \alpha)|G|^2}{1 + \alpha\gamma|G|^2} \right] + \mathbb{E} \left[ \frac{(\dot{\nu}\gamma + \nu)|A|^2}{1 + \nu\gamma|A|^2} \right] \\ &- \frac{\alpha\nu + \alpha\dot{\nu}\gamma + \dot{\alpha}\nu\gamma}{1 + \alpha\nu\gamma} - \frac{1 - \eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\gamma)}{\gamma} \end{aligned}$$

$$\begin{aligned} &= \frac{\dot{\alpha}\gamma + \alpha}{\alpha\gamma} (1 - \eta_{\mathbf{G}\mathbf{G}^\dagger}(\alpha\gamma)) + \frac{\dot{\nu}\gamma + \nu}{\nu\gamma} (1 - \eta_{\mathbf{A}\mathbf{A}^\dagger}(\nu\gamma)) \\ &- \frac{\alpha\nu + \alpha\dot{\nu}\gamma + \dot{\alpha}\nu\gamma}{1 + \alpha\nu\gamma} - \frac{1 - \eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\gamma)}{\gamma} \quad (100) \end{aligned}$$

$$\begin{aligned} &= \frac{\dot{\alpha}\gamma + \alpha}{\alpha\gamma} (1 - \eta) + \frac{\dot{\nu}\gamma + \nu}{\nu\gamma} (1 - \eta) \\ &- \frac{\alpha\nu + \alpha\dot{\nu}\gamma + \dot{\alpha}\nu\gamma}{1 + \alpha\nu\gamma} - \frac{1 - \eta}{\gamma} \quad (101) \end{aligned}$$

$$= (\alpha\nu + \alpha\dot{\nu}\gamma + \dot{\alpha}\nu\gamma) \left( \frac{1 - \eta}{\alpha\nu\gamma} - \eta \right) \quad (102)$$

$$= 0 \quad (103)$$

where in addition to (99), we have used the fact that  $C(\gamma)$  is given by (87) to write the left-hand side of (100); the right-hand side of (100) follows from the definition of the  $\eta$ -transform; (101) follows from Theorem 11 applied at  $\tau = \gamma$  and the fact that (15) is equivalent to (105) and (106) if

$$\eta + \eta\nu\alpha\gamma = 1 \quad (104)$$

which is also responsible for (103).

#### G. Asymptotic Spectrum of $\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger$

Theorems 1 and 8 hinge on the following characterization of the asymptotic distribution of the singular values of a random circulant matrix (as defined in this paper) premultiplied by an independent random diagonal matrix.

*Theorem 11:* Let  $\mathbf{G} = \text{diag}(G_1, \dots, G_n)$  and  $\mathbf{A} = \text{diag}(A_1, \dots, A_n)$  be mutually independent random diagonal matrices according to the assumptions of Section I. For  $\tau \geq 0$ , let  $(\eta, \alpha, \nu)$  be the solution of the system of equations

$$\eta = \eta_{|G|^2}(\tau\alpha) \quad (105)$$

$$\eta = \eta_{|A|^2}(\tau\nu) \quad (106)$$

$$\eta = \frac{1}{1 + \alpha\nu\tau}. \quad (107)$$

Then, the  $\eta$ -transform of  $\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger$  is given by

$$\eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\tau) = \eta. \quad (108)$$

*Proof:* The key technical result from which the proof of Theorem 11 follows as a corollary is Lemma 1.

*Lemma 1:* Let  $\mathbf{G} = \text{diag}(G_1, \dots, G_n)$  and  $\mathbf{A} = \text{diag}(A_1, \dots, A_n)$  be mutually independent random diagonal matrices according to the assumptions of Section I, and let  $\mathbf{F}$  denote a unitary Fourier  $n \times n$  matrix. Then,  $\mathbf{A}^\dagger\mathbf{A}$  and  $\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger$  are asymptotically free, for  $n \rightarrow \infty$ .

*Proof:* The proof is given in Appendix VII where the definition of freeness (see [9] and references therein) is also recalled.  $\square$

For two asymptotically free sequences of  $n \times n$  nonnegative definite matrices  $\mathbf{B}$  and  $\mathbf{C}$ , the  $\eta$ -transform of their product satisfies [9, eq. 2.209]

$$\eta_{\mathbf{C}\mathbf{B}}(\tau) = \eta_{\mathbf{C}} \left( \frac{\tau}{\Sigma_{\mathbf{B}}(\eta_{\mathbf{C}\mathbf{B}}(\tau) - 1)} \right) \quad (109)$$

where  $\Sigma_{\mathbf{B}}(z)$  denotes the  $S$ -transform of  $\mathbf{B}$  (see [9, Sec. 2.2.6], and references therein), already introduced in (55). By exchanging the role of  $\mathbf{C}$  and  $\mathbf{B}$  in (109), we obtain the symmetric expression

$$\eta_{\mathbf{CB}}(\tau) = \eta_{\mathbf{B}} \left( \frac{\tau}{\Sigma_{\mathbf{C}}(\eta_{\mathbf{CB}}(\tau) - 1)} \right). \quad (110)$$

Letting  $\eta = \eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\tau)$  denote the desired  $\eta$ -transform computed at the argument  $\tau$ , and applying (109) and (110) to  $\mathbf{B} = \mathbf{A}^\dagger\mathbf{A}$  and  $\mathbf{C} = \mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger$ , we obtain

$$\eta = \eta_{|\mathbf{G}|^2}(\tau\alpha) = \eta_{|\mathbf{A}|^2}(\tau\nu) \quad (111)$$

where we defined

$$\alpha = \frac{1}{\Sigma_{|\mathbf{A}|^2}(\eta - 1)} \quad (112)$$

and

$$\nu = \frac{1}{\Sigma_{|\mathbf{G}|^2}(\eta - 1)}. \quad (113)$$

Using the expression of the  $S$ -transform in terms of the corresponding  $\eta$ -transform and the equalities in (111), we have

$$\Sigma_{|\mathbf{G}|^2}(\eta - 1) = -\frac{\eta}{\eta - 1} \eta_{|\mathbf{G}|^2}^{-1}(\eta) = -\frac{\eta}{\eta - 1} \tau\alpha \quad (114)$$

$$\Sigma_{|\mathbf{A}|^2}(\eta - 1) = -\frac{\eta}{\eta - 1} \eta_{|\mathbf{A}|^2}^{-1}(\eta) = -\frac{\eta}{\eta - 1} \tau\nu. \quad (115)$$

Eliminating  $\Sigma_{|\mathbf{G}|^2}(\eta - 1)$  and  $\Sigma_{|\mathbf{A}|^2}(\eta - 1)$  from the system of equations given by (112)–(114), we obtain (107).  $\square$

#### H. Proof of Theorem 2

Consider the case without time-domain fading, given in (25) with waterfilling power allocation  $\bar{S}_x(\cdot, \cdot)$  defined by (27) and (28). The following lemma shows an interesting relation between the “water level” in the waterfilling solution and the  $\eta$ -transform of the modified fading distribution obtained by concatenating the actual fading with the optimal frequency-domain “power controller.”

*Lemma 2:* Consider the waterfilling power allocation  $\bar{S}_x(\gamma, \cdot)$  defined by

$$\bar{S}_x(\gamma, z) = \left[ \zeta_\gamma - \frac{1}{\gamma z} \right]^+, \quad z \geq 0 \quad (116)$$

with “water level”  $1 < \zeta_\gamma < \infty$  solution of

$$\mathbb{E} \left[ \left[ \zeta_\gamma - \frac{1}{\gamma |\mathbf{G}|^2} \right]^+ \right] = 1. \quad (117)$$

Define the modified fading coefficients  $P_i$ , identically distributed as  $P = \bar{S}_x(\gamma, |\mathbf{G}|^2)|\mathbf{G}|^2$ . Then

$$\zeta_\gamma = \frac{1}{1 - \eta_P(\gamma)}. \quad (118)$$

*Proof:* Noticing that  $P = 0$  if  $|\mathbf{G}|^2 \leq \frac{1}{\gamma\zeta_\gamma}$ , we can write

$$\eta_P(\gamma) = \mathbb{E} \left[ \frac{1}{1 + \gamma P} \right] \quad (119)$$

$$\begin{aligned} &= \mathbb{P} \left( |\mathbf{G}|^2 \leq \frac{1}{\gamma\zeta_\gamma} \right) + \mathbb{P} \left( |\mathbf{G}|^2 > \frac{1}{\gamma\zeta_\gamma} \right) \\ &\quad \times \mathbb{E} \left[ \frac{1}{\gamma\zeta_\gamma |\mathbf{G}|^2} \middle| |\mathbf{G}|^2 > \frac{1}{\gamma\zeta_\gamma} \right]. \end{aligned} \quad (120)$$

From (117), we can write

$$\begin{aligned} 1 &= \left( 1 - \mathbb{P} \left( |\mathbf{G}|^2 \leq \frac{1}{\gamma\zeta_\gamma} \right) \right) \zeta_\gamma \\ &\quad - \mathbb{P} \left( |\mathbf{G}|^2 > \frac{1}{\gamma\zeta_\gamma} \right) \mathbb{E} \left[ \frac{1}{\gamma |\mathbf{G}|^2} \middle| |\mathbf{G}|^2 > \frac{1}{\gamma\zeta_\gamma} \right]. \end{aligned} \quad (121)$$

Dividing both sides of (121) by  $\zeta_\gamma$  and comparing with (119), we obtain the result.  $\square$

Lemma 2 is instrumental in proving the form of the optimal power allocation function given by Theorem 2. Consider the case where the transmitter knows the realization of the frequency-domain fading, and multiplies each  $i$ th signal frequency component by the factor  $\sqrt{S_x^*(\gamma, |\mathbf{G}_i|^2)}$ , where the function  $S_x^*(\gamma, \cdot)$  is defined in Theorem 2. It is clear that the mutual information achieved in this case is equal to the mutual information of an equivalent channel where the frequency-domain power controller is seen as a part of the channel, and the transmitter has no channel state information and, because of symmetry, transmits a white Gaussian input. Let  $P_i = |\mathbf{G}_i|^2 S_x^*(\gamma, |\mathbf{G}_i|^2)$  denote the resulting modified fading process. Since the function  $S_x^*(\gamma, \cdot)$  is a memoryless stationary deterministic mapping and, by construction,  $\mathbb{E}[S_x^*(\gamma, |\mathbf{G}_i|^2)] = 1$ , it follows that  $\{P_i\}$  inherits the same stationarity and ergodicity properties of the original fading process  $\{G_i\}$ . Therefore, Theorem 11 and Lemma 1 apply verbatim also for the modified fading distribution. Letting  $P$  denote a random variable with the same first-order marginal distribution of  $\{P_i\}$ , we notice also that the modified fading distribution satisfies the *compatibility condition*  $\mathbb{E}[P/|\mathbf{G}|^2] \leq 1$ ,<sup>4</sup> that reflects the original channel input power constraint.

The proof of Theorem 2 is obtained in two steps. First, we find an optimality condition for the best possible modified fading distribution, subject to the compatibility condition given above, even allowing the new fading to be dependent of the whole  $\mathbf{A}$  and  $\mathbf{G}$  and without any requirement of stationarity and ergodicity. Then, we will show that in fact this condition is met, asymptotically for  $n \rightarrow \infty$ , by  $P_i = |\mathbf{G}_i|^2 S_x^*(\gamma, |\mathbf{G}_i|^2)$ .

Let  $\mathbf{Q} = \mathbf{A}\mathbf{F} = [\mathbf{q}_1, \dots, \mathbf{q}_n]$  and consider the optimization problem

$$\begin{aligned} &\max_{D_i \geq 0} \frac{1}{n} \log \det (\mathbf{I} + \gamma \mathbf{Q} \text{diag}(D_1, \dots, D_n) \mathbf{Q}^\dagger) \\ &\text{subject to} \quad \frac{1}{n} \sum_{i=1}^n \frac{D_i}{|\mathbf{G}_i|^2} \leq 1. \end{aligned} \quad (122)$$

<sup>4</sup>Here we let  $\frac{0}{0} = 0$  by continuity.

The above problem is solved for each realization of the fading matrices  $\mathbf{A}$  and  $\mathbf{G}$ , and therefore, the solution  $\{D_i\}$  is a new fading process possibly dependent on the whole  $\mathbf{A}$  and  $\mathbf{G}$ .

The convexity of (122) enables us to appeal to the Karush–Kuhn–Tucker (KKT) conditions [11] to characterize necessary and sufficient conditions for the solution. Letting  $\mathbf{D} = \text{diag}(D_1, \dots, D_n)$ , the Lagrangian function is given by

$$\mathcal{L}(\mathbf{D}, \lambda) = \frac{1}{n} \log \det (\mathbf{I} + \gamma \mathbf{Q} \mathbf{D} \mathbf{Q}^\dagger) - \lambda \left( \frac{1}{n} \sum_{i=1}^n \frac{D_i}{|\mathbf{G}_i|^2} - 1 \right). \quad (123)$$

After straightforward algebra, obtain the KKT conditions in the form

$$D_i = |\mathbf{G}_i|^2 \left[ \frac{1}{\lambda} - \frac{1}{\gamma Z_i |\mathbf{G}_i|^2} \right]^+ \quad (124)$$

where  $Z_i = \mathbf{q}_i^\dagger \left( \mathbf{I} + \sum_{j \neq i} \mathbf{q}_j \mathbf{q}_j^\dagger D_j \right)^{-1} \mathbf{q}_i$  and where  $\lambda$  is chosen such that

$$\frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{\lambda} - \frac{1}{\gamma Z_i |\mathbf{G}_i|^2} \right]^+ = 1. \quad (125)$$

Then, letting  $n \rightarrow \infty$ , the solution (124) and (125) defines the modified the frequency-domain fading process  $\{D_i\}$  that maximizes the mutual information per symbol subject to the compatibility condition  $\frac{1}{n} \sum_{i=1}^n D_i / |\mathbf{G}_i|^2 = 1$  for all  $n$ , which obviously implies  $\frac{1}{n} \sum_{i=1}^n D_i / |\mathbf{G}_i|^2 \rightarrow 1$  with probability 1.

Now, consider a fading distribution  $P_i = |\mathbf{G}_i|^2 g(|\mathbf{G}_i|^2)$ , for some fixed function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ . In this case, following a rather involved moment calculation sketched in Appendix VIII, the following convergence result can be proved:

$$Z_i = \mathbf{q}_i^\dagger \left( \mathbf{I} + \sum_{j \neq i} \mathbf{q}_j \mathbf{q}_j^\dagger P_j \right)^{-1} \mathbf{q}_i \xrightarrow{\text{a.s.}} \alpha \quad (126)$$

where  $\alpha$  is the positive solution of the fixed-point equation in Theorem 11.

Using (105)–(108), we have

$$\eta \triangleq \eta_{\mathbf{A} \mathbf{F} \mathbf{P} \mathbf{F}^\dagger \mathbf{A}^\dagger}(\gamma) = \eta_{\mathbf{P}}(\alpha \gamma) \quad (127)$$

where we let  $\mathbf{P} = \text{diag}(P_1, \dots, P_n)$ .

For such fading distribution, we can rewrite (125) in the limit of large  $n$  as

$$\mathbb{E} \left[ \left[ \frac{1}{\lambda} - \frac{1}{\gamma \alpha |\mathbf{G}_i|^2} \right]^+ \right] = 1. \quad (128)$$

This is formally identical to the solution for the “water level” in the case of no time-domain fading, for a modified SNR  $\gamma' = \alpha \gamma$ . For consistency with the notation introduced in (27) and (28) and in Lemma 2, we let the solution of (128) be denoted by  $1/\lambda = \zeta_{\gamma'}$ . Using the notation introduced in Theorem 2, we define  $\mu(\gamma') = 1/\alpha$ , so that the modified SNR  $\gamma'$  satisfies  $\gamma = \mu(\gamma') \gamma'$ .

Choosing the function  $g(\cdot)$  in the definition of  $\{P_i\}$  to be equal to the waterfilling solution  $\bar{S}_x(\gamma', \cdot)$ , Lemma 2 combined with (127) yields

$$\zeta_{\gamma'} = \frac{1}{1 - \eta_{\mathbf{P}}(\gamma')} = \frac{1}{1 - \eta}. \quad (129)$$

Finally, we can eliminate  $\eta$  from the system of resulting equations and state the power allocation directly in terms of the modified SNR  $\gamma'$ . Using (129) and the expression of  $\alpha$  in terms of  $\eta$  given in (112), we can write

$$\mu(\gamma') = \Sigma_{|\mathbf{A}|^2}(\eta - 1) \quad (130)$$

$$= \Sigma_{|\mathbf{A}|^2} \left( -\frac{1}{\zeta_{\gamma'}} \right). \quad (131)$$

Finally, using the definition of the  $S$ -transform and after straightforward algebra it is easy to see that (130) is equivalent to (31).

Notice that the statistics of the modified fading process  $\{P_i\}$  was defined using the function  $g(\cdot) = \bar{S}_x(\gamma', \cdot) = S_x^*(\gamma, \cdot)$ . In view of the above derivation, the modified fading process  $\{P_i\}$  satisfies both the KKT conditions (124) and the compatibility condition (125) in the limit of  $n \rightarrow \infty$ . Therefore, this is the asymptotically optimal frequency-domain fading distribution for the equivalent channel with no state information at the transmitter, with fading subject to the compatibility condition in (122). For the argument said at the beginning, this implies that  $S_x^*(\gamma, \cdot)$  is the optimal power allocation function for the original channel, when the transmitter knows the frequency-domain fading. At this point, Theorem 2 is proved.

We conclude this section by showing that  $Z_i$  in (126) converges to the factor  $\alpha$  in the equality  $\eta_{\mathbf{A} \mathbf{F} \mathbf{P} \mathbf{F}^\dagger \mathbf{A}^\dagger}(\gamma) = \eta_{\mathbf{P}}(\alpha \gamma)$ . Let again  $\mathbf{A} \mathbf{F} = \mathbf{Q}$  with  $i$ th column denoted by  $\mathbf{q}_i$ , define  $\mathbf{B}_i = \mathbf{I} + \gamma \sum_{j \neq i} P_j \mathbf{q}_j \mathbf{q}_j^\dagger$ , so that  $Z_i = \mathbf{q}_i^\dagger \mathbf{B}_i^{-1} \mathbf{q}_i$ . Recalling the definition of the  $\eta$ -transform, we can write

$$1 - \eta_{\mathbf{A} \mathbf{F} \mathbf{P} \mathbf{F}^\dagger \mathbf{A}^\dagger}(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left\{ \mathbf{I} - (\mathbf{I} + \gamma \mathbf{Q} \mathbf{P} \mathbf{Q}^\dagger)^{-1} \right\} \right] \quad (132)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left\{ \gamma \mathbf{Q} \mathbf{P} \mathbf{Q}^\dagger (\mathbf{I} + \gamma \mathbf{Q} \mathbf{P} \mathbf{Q}^\dagger)^{-1} \right\} \right] \quad (133)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left\{ \left( \sum_{i=1}^n \gamma P_i \mathbf{q}_i \mathbf{q}_i^\dagger \right) (\mathbf{I} + \gamma \mathbf{Q} \mathbf{P} \mathbf{Q}^\dagger)^{-1} \right\} \right] \quad (134)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n \gamma P_i \text{tr} \left\{ \mathbf{q}_i \mathbf{q}_i^\dagger \left( \mathbf{B}_i^{-1} - \frac{\gamma P_i \mathbf{B}_i^{-1} \mathbf{q}_i \mathbf{q}_i^\dagger \mathbf{B}_i^{-1}}{1 + \gamma P_i Z_i} \right) \right\} \right] \quad (135)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n \gamma P_i \text{tr} \left\{ \mathbf{q}_i \mathbf{q}_i^\dagger \mathbf{B}_i^{-1} \right\} \left( 1 - \frac{\gamma P_i Z_i}{1 + \gamma P_i Z_i} \right) \right] \quad (136)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \sum_{i=1}^n \frac{\gamma P_i Z_i}{1 + \gamma P_i Z_i} \right] \quad (137)$$

$$= \mathbb{E} \left[ \frac{\gamma P \alpha}{1 + \gamma P \alpha} \right] \quad (138)$$

$$= 1 - \mathbb{E} \left[ \frac{1}{1 + \alpha\gamma\mathbf{P}} \right] \quad (139)$$

$$= 1 - \eta_{\mathbf{P}}(\alpha\gamma) \quad (140)$$

where in (135) we applied the matrix inversion lemma and in (138) we have used Lemma 12 (see Appendix VIII) which proves the a.s. convergence of  $Z_i$  to some constant  $\alpha$  independent of the normalized index  $i/n$ .

### I. Proof of Theorem 3

Fix  $\gamma > 0$ . For the purposes of this proof it is convenient to temporarily switch notation and let  $(\alpha^*, \nu^*)$  denote the solution of (15). In addition, we denote the right-hand side of (12) evaluated at  $\alpha = \alpha^*$  by  $I(\alpha^*, \nu)$ , i.e., for fixed  $\gamma$  and  $\alpha = \alpha^*$ , we consider this as a function of a dummy variable  $\nu \in \mathbb{R}_+$ . This function is continuous and differentiable for all  $\nu \in \mathbb{R}_+$ , with derivative that satisfies

$$\frac{1}{\gamma} \frac{\partial I(\alpha^*, \nu)}{\partial \nu} = \mathbb{E} \left[ \frac{|A|^2}{1 + \nu\gamma|A|^2} \right] - \frac{\alpha^*}{1 + \nu\gamma\alpha^*}. \quad (141)$$

Because of the first property in Section III-B, or simply from the second equality in (15), we have that

$$\left. \frac{\partial I(\alpha^*, \nu)}{\partial \nu} \right|_{\nu=\nu^*} = 0. \quad (142)$$

We now show that not only is  $\nu^*$  a stationary point of  $I(\alpha^*, \nu)$  but that it is in fact a global maximum. To that end, it is enough to show that

$$\frac{\partial I(\alpha^*, \nu)}{\partial \nu} \geq 0, \quad 0 \leq \nu \leq \nu^* \quad (143)$$

$$\frac{\partial I(\alpha^*, \nu)}{\partial \nu} < 0, \quad \nu^* < \nu. \quad (144)$$

We recognize that each of terms in the right-hand side of (141) can be interpreted as a MMSE. The first term is the MMSE for estimating  $X$  from the observation of  $(\sqrt{\nu\gamma}X + Z, |A|^2)$  with  $Z \sim \mathcal{CN}(0, 1)$ , where  $X \sim \mathcal{CN}(0, a^2)$  when conditioned on  $|A|^2 = a^2$ . The second term in the right-hand side of (141) is the MMSE for estimating  $X'$  from  $\sqrt{\nu\gamma}X' + Z$ , when  $X' \sim \mathcal{CN}(0, \alpha^*)$ . The single crossing-point property [12, Proposition 12], of MMSEs in additive white Gaussian noise dictates that since the MMSE terms in (141) coincide at  $\gamma\nu^*$ , one of them must be strictly higher for lower SNRs, and strictly lower for higher SNRs. Then, to see that (143) and (144) is indeed satisfied it is enough to verify the behavior of  $\frac{\partial I(\alpha^*, \nu)}{\partial \nu}$  at  $\nu = 0$ . From (141), we get

$$\left. \frac{\partial I(\alpha^*, \nu)}{\partial \nu} \right|_{\nu=0} = \gamma (\mathbb{E} [|A|^2] - \alpha^*) > 0 \quad (145)$$

where (145) holds [see(13)] whenever  $|A|^2$  is not deterministic (if  $|A|^2$  is deterministic, Theorem 3 trivially holds with equality).

Since  $\nu^*$  attains the global maximum, it follows that for any nonnegative random variable  $\mathbf{P}$ , we have

$$C(\gamma) = I(\alpha^*, \nu^*) \geq \mathbb{E}[I(\alpha^*, \mathbf{P})]. \quad (146)$$

In particular, by choosing  $\mathbf{P} = |\mathbf{G}|^2$ , we find

$$C(\gamma) \geq \mathbb{E}[I(\alpha^*, |\mathbf{G}|^2)] = \mathbb{E} [\log (1 + \gamma|A|^2|\mathbf{G}|^2)] \quad (147)$$

as we wanted to show.

Notice that in this proof we could have taken a dual approach considering the function  $I(\alpha, \nu^*)$  as a function of  $\alpha$  for fixed  $\nu^*$ .

### J. Proof of Theorem 4

Following identical steps as in Section III-I and choosing  $\mathbf{P} = 0$ , we find

$$C(\gamma) \geq I(\alpha, 0) = \mathbb{E} [\log (1 + \gamma\alpha|\mathbf{G}|^2)] \quad (148)$$

which coincides with (33). Repeating the same steps while exchanging the role of  $\nu$  and  $\alpha$ , we find (34).

To show (35), we add and subtract  $\log(1 + \nu\alpha\gamma)$  from (12), yielding

$$C(\gamma) = \log(1 + \nu\alpha\gamma) + \mathbb{E} \left[ \log \left( \frac{1 + \nu\gamma|A|^2}{1 + \alpha\nu\gamma} \right) \right] + \mathbb{E} \left[ \log \left( \frac{1 + \alpha\gamma|\mathbf{G}|^2}{1 + \alpha\nu\gamma} \right) \right]. \quad (149)$$

Then, notice both the second and thirds term in the right side of (149) are nonnegative. For the second term, by using convexity of  $-\log(\cdot)$  and applying Jensen's inequality, we have

$$\mathbb{E} \left[ -\log \left( \frac{1 + \alpha\nu\gamma}{1 + \nu\gamma|A|^2} \right) \right] \geq -\log \left( \mathbb{E} \left[ \frac{1 + \alpha\nu\gamma}{1 + \nu\gamma|A|^2} \right] \right) = 0 \quad (150)$$

where the identity follows from the second equality in (15). A similar argument holds for the third term in (149).

### K. Proof of Theorem 5

In order to show (36) and (37), we write

$$C(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (151)$$

and use Jensen's inequality with respect to  $\mathbf{G}$  and  $\mathbf{A}$ , respectively.

### L. Proof of Theorem 6

In order to obtain the low-SNR behavior of channel capacity, we use the general results in [7, eq. (35)] and [7, Th. 9] and write

$$\left( \frac{E_b}{N_0} \right)_{\min} = \frac{1}{\dot{C}(0)} \quad (152)$$

$$S_0 = -2 \frac{\dot{C}'(0)^2}{\ddot{C}(0)} \quad (153)$$

where we define  $\dot{C}(0) = \left. \frac{\partial C(\gamma)}{\partial \gamma} \right|_{\gamma=0}$  and  $\ddot{C}(0) = \left. \frac{\partial^2 C(\gamma)}{\partial \gamma^2} \right|_{\gamma=0}$ , where  $C(\gamma)$  is the capacity as a function of the SNR  $\gamma$  expressed in nats per symbol.

We start by considering the case where the transmitter has no knowledge of the frequency-domain fading, and therefore, the optimal input is white and stationary. The relevant expression for  $C(\gamma)$  is given in (12) where  $\alpha = \alpha(\gamma)$  and  $\nu = \nu(\gamma)$  are the solution of (15). It is straightforward to check that the functions satisfy

$$\alpha(\gamma) = \mathbb{E}[|A|^2] - \mathbb{E}[|G|^2] \text{Var}(|A|^2)\gamma + o(\gamma) \quad (154)$$

$$\nu(\gamma) = \mathbb{E}[|G|^2] - \mathbb{E}[|A|^2] \text{Var}(|G|^2)\gamma + o(\gamma). \quad (155)$$

Furthermore, taking a Taylor series expansion of the expressions (12) in nats, we have

$$C(\gamma) = \mathbb{E}[\log(1 + \alpha\gamma|G|^2)] + \mathbb{E}[\log(1 + \nu\gamma|A|^2)] - \log(1 + \alpha\nu\gamma) \quad (156)$$

$$= \gamma(\alpha(\gamma)\mathbb{E}[|G|^2] + \nu(\gamma)\mathbb{E}[|A|^2] - \alpha(\gamma)\nu(\gamma)) - \frac{\gamma^2}{2}(\alpha^2(\gamma)\mathbb{E}[|G|^4] + \nu^2(\gamma)\mathbb{E}[|A|^4] - \alpha^2(\gamma)\nu^2(\gamma)) + o(\gamma^2) \quad (157)$$

$$= \mathbb{E}[|A|^2]\mathbb{E}[|G|^2]\gamma - \frac{\gamma^2}{2}\mathbb{E}^2[|A|^2]\mathbb{E}^2[|G|^2](\kappa(|G|) + \kappa(|A|) - 1) + o(\gamma^2). \quad (158)$$

Therefore

$$\dot{C}(0) = \mathbb{E}[|A|^2]\mathbb{E}[|G|^2] \quad (159)$$

$$\ddot{C}(0) = \mathbb{E}^2[|A|^2]\mathbb{E}^2[|G|^2](\kappa(|G|) + \kappa(|A|) - 1) \quad (160)$$

where the kurtosis  $\kappa(Z)$  is defined in (38). Thus, (39) follows by using (159) and (160) in (152) and in (153).

When  $\mathbf{G}$  is known at the transmitter, from Theorem 2 and the proof in Section III-H, we know that the capacity formula (12) holds after replacing  $|G|^2$  with the modified fading  $\mathbb{P} = |G|^2 S_x^*(\gamma, |G|^2)$ , where the modified ‘‘waterfilling’’ power allocation function  $S_x^*(\gamma, \cdot)$  is provided by Theorem 2.

Suppose for the time being that there exists some value  $G_{\max} < \infty$  such that  $\mathbb{P}(|G|^2 \leq G_{\max}) = 1$ , and such that  $B_{\max} = \mathbb{P}(|G|^2 = G_{\max}) > 0$ . In this case, it is simple to show

that in the limit of very small  $\gamma$  the power allocation function becomes

$$S_x^*(\gamma, z) = \frac{1}{B_{\max}} 1\{z = G_{\max}\}. \quad (161)$$

In this case, it is immediate to see that  $\mathbb{E}[\mathbb{P}] = G_{\max}$  and  $\kappa(\sqrt{\mathbb{P}}) = \frac{\mathbb{E}[\mathbb{P}^2]}{\mathbb{E}^2[\mathbb{P}]} = \frac{1}{B_{\max}}$ . Hence, (41) follows in the same way as before.

In order to handle the general case where the distribution of  $|G|$  may have unbounded support (i.e.,  $G_{\max} = +\infty$ ) or no mass point at its essential supremum (i.e.,  $B_{\max} = 0$ ), we operate as follows. For some arbitrary value  $\bar{G} < 0$ , the actual fading distribution  $|G|^2$  can be approximated by a truncated distribution

$$|\bar{G}|^2 = \begin{cases} |G|^2, & \text{for } |G|^2 < \bar{G} \\ \bar{G}, & \text{for } |G|^2 \geq \bar{G}. \end{cases} \quad (162)$$

We can choose  $\bar{G}$  such that, for the truncated fading distribution, we have  $\bar{G}_{\max} = \bar{G} < \infty$  and  $\bar{B}_{\max} = \mathbb{P}(|G|^2 \geq \bar{G}) > 0$ . Then, from what said before Theorem 6 holds for the truncated fading distribution. Finally, the result is seen to hold in general by a continuity argument, letting  $\bar{G} \rightarrow \infty$ .

#### M. Proof of Theorem 7

Throughout this section, we assume a white stationary input with covariance matrix  $\Sigma_x = \mathbf{I}$ . From property (56) of the solution of (15), and using the fact that the limit for  $\gamma \rightarrow \infty$  of the  $\eta$ -transform yields the fraction of zero eigenvalues, we obtain that the asymptotic normalized rank of the channel matrix  $\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger$  is equal to  $\mathcal{S}_\infty = 1 - \max\{u_0, v_0\}$ .

Then, we consider the high-SNR offset  $\mathcal{L}_\infty$ . In the absence of time-domain fading, it is easy to see (cf. [8, (33)]) that  $\mathcal{L}_\infty = \mathcal{L}_\infty^{\text{no-tdf}}$  given in (49). Similarly, in the absence of frequency-domain fading, we obtain  $\mathcal{L}_\infty = \mathcal{L}_\infty^{\text{no-fdf}}$  given in (49).

Consider now the case where both fadings are present, and  $u_0 > v_0$ . We have (163)–(166), shown at the bottom of the page, where (164) follows from (12), and (165) follows from the limits in (57) and (58). In a completely symmetric way, using the limits in (59) and (60) we obtain the expression for the case  $u_0 < v_0$ . Finally, for the case  $u_0 = v_0$ , we use the fact that [cf. (61)]  $\gamma\alpha$  and  $\gamma\nu$  both diverge to infinity as  $\gamma \rightarrow \infty$ , while  $\gamma\alpha\nu \rightarrow 1 - u_0/v_0$ . Therefore, we have (167)–(170), shown at the bottom of the next page.

$$\mathcal{L}_\infty = \lim_{\gamma \rightarrow \infty} \left( \log_2 \gamma - \frac{C(\gamma)}{\mathcal{S}_\infty} \right) \quad (163)$$

$$= \lim_{\gamma \rightarrow \infty} \left( \log_2 \gamma - \frac{1}{1 - u_0} \left[ \mathbb{E}[\log_2(1 + \alpha\gamma|G|^2)] + \mathbb{E}[\log_2(1 + \nu\gamma|A|^2)] - \log_2(1 + \alpha\nu\gamma) \right] \right) \quad (164)$$

$$= \frac{1}{1 - u_0} \log_2 \frac{1}{u_0} - \log_2 \frac{1 - u_0}{u_0} + \log_2 \varsigma - \frac{1}{1 - u_0} \mathbb{E}[\log_2(1 + \varsigma|A|^2)] - \mathbb{E}[\log_2 |G|^2 | |G|^2 > 0] \quad (165)$$

$$= \mathcal{L}_\infty^{\text{no-tdf}} + \frac{h(u_0)}{1 - u_0} + \log_2 \varsigma - \frac{1}{1 - u_0} \mathbb{E}[\log_2(1 + \varsigma|A|^2)] \quad (166)$$

## IV. CONCLUDING REMARKS

We obtained the channel capacity of a channel model that captures the effect of fading in both the time domain and the frequency domain. The central technical result of this paper is the asymptotic freeness of the random diagonal matrix  $\mathbf{A}\mathbf{A}^\dagger$  and the random circulant matrix  $\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger$ , when the coefficients in  $\mathbf{G}$  are i.i.d. independent of those of  $\mathbf{A}$  which satisfy relatively mild assumptions (or *vice versa*). This allows us to obtain the asymptotic eigenvalue distribution of  $\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger$  in terms of its  $\eta$ -transform, which yields the channel capacity in the case where the transmitter has no information about the realization of the fading, but only knows its statistics and the channel SNR. Along the way, we obtained new and relevant auxiliary results that have interest on their own. For example, Theorems 9 and 10 offer a novel general characterization of the Shannon transform of a nonnegative random variable.

For the case when the frequency-domain fading is known to the transmitter, we found the optimal frequency-domain power allocation function that takes on the form of a modified water-filling power allocation for an SNR value lower than the actual channel SNR. This means that in the presence of time-selective fading it is preferable to focus the signal energy on a subset of favorable frequency bands, thereby extending the correlation in the time domain to cope with the time selectivity of the channel more effectively. Appendix IX deals with the case where the frequency selectivity originates from a deterministic linear time-invariant filter; there is considerable evidence that such a case can also be encompassed by the main result of this paper.

The capacity formulas of Theorems 1 and 8 are given in terms of the solution of coupled fixed-point equations. Although the numerical computation of such formulas is quite straightforward, we have also provided simple upper and lower bounds that can be computed from their closed-form expressions. Finally, we have provided simple and closed-form expressions for the low-SNR and the high-SNR capacity approximation in terms of the fundamental asymptotic parameters  $(E_b/N_0)_{\min}$ ,  $\mathcal{S}_0$ ,  $\mathcal{S}_\infty$ , and  $\mathcal{L}_\infty$ .

As illustrated numerically, and typical in random matrix theory, the convergence of the average mutual information rate

$$\frac{1}{n} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger)]$$

is very fast. Just like with multiantenna systems where large-size asymptotic formulas are useful proxies for even small arrays, in the present case, the main result is an accurate approximation to the capacity of standardized OFDM [number of carriers ranging from 52 (IEEE802.11a) to 6817 (DVB)].

## APPENDIX I

## OPTIMALITY OF STATIONARY INPUTS

*Theorem 12:* Suppose that both  $\{A_i, i \in \mathbb{Z}\}$  and  $\{G_i, i \in \mathbb{Z}\}$  are stationary processes; the receiver knows both  $\mathbf{A}$  and  $\mathbf{G}$ , while the transmitter has no knowledge of the  $\mathbf{A}$  other than its probability distribution. Then, the maximization in

$$C(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \times \max_{\Sigma_x} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger)] \quad (171)$$

can be restricted to circulant input covariance  $\Sigma_x$ , regardless of whether  $\mathbf{G}$  is known at the transmitter.

*Proof:* Let  $\mathbf{\Pi}$  denote the elementary circulant permutation matrix, defined as

$$\mathbf{\Pi} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & & & 0 & 1 \\ 1 & \cdots & & & 0 & 0 \end{bmatrix} \quad (172)$$

and denote for an arbitrary  $\Sigma_x$

$$\Sigma_x^{(\ell)} = \mathbf{\Pi}^\ell \Sigma_x \mathbf{\Pi}^{\ell\dagger}. \quad (173)$$

Invoking Jensen's inequality

$$\begin{aligned} & \mathbb{E} \left[ \log \det \left( \mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger \left( \frac{1}{n} \sum_{\ell=0}^{n-1} \Sigma_x^{(\ell)} \right) \mathbf{F}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger \right) \right] \\ & \geq \frac{1}{n} \sum_{\ell=0}^{n-1} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger \mathbf{\Pi}^\ell \Sigma_x \mathbf{\Pi}^{\ell\dagger} \mathbf{F}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger \right) \right] \end{aligned} \quad (174)$$

$$= \frac{1}{n} \sum_{\ell=0}^{n-1} \mathbb{E} \left[ \log \det \left( \mathbf{I} + \gamma \mathbf{A}\mathbf{\Pi}^{\ell\dagger} \mathbf{F}\mathbf{G}\mathbf{G}^\dagger \mathbf{\Pi}^\ell \Sigma_x \mathbf{\Pi}^{\ell\dagger} \mathbf{F}\mathbf{G}^\dagger\mathbf{F}^\dagger \mathbf{A}^\dagger \right) \right] \quad (175)$$

$$= \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger \Sigma_x \mathbf{F}\mathbf{G}^\dagger\mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (176)$$

where (175) holds under the assumption that the diagonal elements of  $\mathbf{A}$  are circularly stationary, while in (176) we have used the fact that  $\mathbf{F}\mathbf{G}\mathbf{G}^\dagger$  is circulant thus  $\mathbf{\Pi}^{\ell\dagger} \mathbf{F}\mathbf{G}\mathbf{G}^\dagger \mathbf{\Pi}^\ell = \mathbf{F}\mathbf{G}\mathbf{G}^\dagger$ . Therefore, the objective function achieved by an arbitrary  $\Sigma_x$  can only be improved by substituting with the circulant  $\frac{1}{n} \sum_{\ell=0}^{n-1} \Sigma_x^{(\ell)}$  with identical trace. To drop the assumption

$$\mathcal{L}_\infty = \lim_{\gamma \rightarrow \infty} \left( \log_2 \gamma - \frac{C(\gamma)}{\mathcal{S}_\infty} \right) \quad (167)$$

$$= \lim_{\gamma \rightarrow \infty} \left( \log_2 \gamma - \frac{1}{1-u_0} \left[ \mathbb{E} [\log_2(1 + \alpha\gamma|G|^2)] + \mathbb{E} [\log_2(1 + \nu\gamma|A|^2)] - \log_2(1 + \alpha\nu\gamma) \right] \right) \quad (168)$$

$$= \frac{1}{1-u_0} \log_2 \frac{1}{u_0} - \log_2 \frac{1-u_0}{u_0} - \lim_{\gamma \rightarrow \infty} \mathbb{E} \left[ \log_2 \frac{(1 + \gamma\alpha|G|^2)(1 + \gamma\nu|A|^2)}{\gamma^2\alpha\nu} \middle| |G|^2 > 0, |A|^2 > 0 \right] \quad (169)$$

$$= \mathcal{L}_\infty^{\text{no-tdf}} + \mathcal{L}_\infty^{\text{no-fdf}} + \frac{h(u_0)}{1-u_0}. \quad (170)$$

of circularly stationary  $\mathbf{A}$ , it is necessary to go to the limit: because of stationarity, the matrix inversion lemma leads to

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^\dagger)] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \boldsymbol{\Pi} \boldsymbol{\Pi}^\dagger \boldsymbol{\Sigma} \boldsymbol{\Pi} \mathbf{A}^\dagger)] \end{aligned} \quad (177)$$

for positive definite  $\boldsymbol{\Sigma}$  for which the limit exists.  $\square$

We proceed to the case in which neither  $\mathbf{G}$  nor  $\mathbf{A}$  are known at the transmitter but only their statistics are available.

*Theorem 13:* In addition to the setup and assumptions of Theorem 12 suppose that  $\{\mathbf{A}_i, i \in \mathbb{Z}\}$  is i.i.d.,  $\{\mathbf{G}_i, i \in \mathbb{Z}\}$  is strongly mixing and are both unknown to the transmitter. Then, the maximization in (171) is achieved by  $\boldsymbol{\Sigma}_x = \mathbf{I}$ .

*Proof:* Theorem 12 shows that the capacity of the channel defined in (11) is achieved by complex circularly symmetric Gaussian stationary inputs with covariance  $\boldsymbol{\Sigma}_x = \mathbf{F} \boldsymbol{\Upsilon} \mathbf{F}^\dagger$ , where  $\boldsymbol{\Upsilon}$  is a diagonal matrix. Then, (171) can be rewritten as

$$\begin{aligned} C(\gamma) &= \lim_{n \rightarrow \infty} \frac{1}{n} \max_{\boldsymbol{\Upsilon}} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \boldsymbol{\Upsilon} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (178) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \max_{\boldsymbol{\Upsilon}} \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G}^{(k)} \boldsymbol{\Upsilon}^{(k)} \mathbf{G}^{\dagger(k)} \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (179) \end{aligned}$$

where  $\boldsymbol{\Upsilon}^{(k)}$  is the  $k$ -circularly shifted version of the diagonal matrix  $\boldsymbol{\Upsilon}$ . Note that (179) follows from the fact that  $\mathbf{A} \mathbf{F} \mathbf{D}^{(k)} \mathbf{F}^\dagger \mathbf{A}^\dagger$  and  $\mathbf{A} \mathbf{F} \mathbf{D} \mathbf{F}^\dagger \mathbf{A}^\dagger$  have the same eigenvalues.

If we were to assume that  $\{\mathbf{G}_i : i \in \mathbb{Z}\}$  is an i.i.d. process, the result would easily follow, since for an arbitrary diagonal  $\boldsymbol{\Upsilon}$  such that  $\frac{1}{n} \text{tr}(\boldsymbol{\Upsilon}) = 1$ , we can write the identity as the average of all circular shifts of  $\boldsymbol{\Upsilon}$  and

$$\begin{aligned} & \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger)] \\ & \geq \sum_{k=1}^n \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \boldsymbol{\Upsilon}^{(k)} \mathbf{G}^{\dagger(k)} \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (180) \\ & = \mathbb{E} [\log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{F} \mathbf{G} \boldsymbol{\Upsilon} \mathbf{G}^\dagger \mathbf{F}^\dagger \mathbf{A}^\dagger)] \quad (181) \end{aligned}$$

where (180) follows due to concavity of the log determinant and (181) follows from the fact that  $\{\mathbf{G}_i : i \in \mathbb{Z}\}$  is an i.i.d. process.

For the case of  $\{\mathbf{G}_i : i \in \mathbb{Z}\}$  with memory, we will use the following general finite-dimensional result which is of independent interest.

*Theorem 14 [13]:* Let  $\boldsymbol{\Phi}$  be an  $m \times n$  complex valued random matrix whose  $i$ th column is denoted by  $\boldsymbol{\phi}_i$ . Consider the optimization problem

$$\max_{\mathbf{D}} \mathbb{E} [\log \det (\mathbf{I} + \gamma \boldsymbol{\Phi} \mathbf{D} \boldsymbol{\Phi}^\dagger)] \quad (182)$$

where the maximum is over all diagonal matrices whose trace is equal to a constant  $\xi$ . Then, for  $i = 1, \dots, n$ ,  $d_i^*$ , the  $i$ th diagonal

element of the diagonal matrix  $\mathbf{D}^*$  that achieves the maximum in (182) is the positive solution to

$$\mathbb{E} \left[ \frac{Z_i}{1 + \gamma d_i^* Z_i} \right] = \frac{1}{\nu \gamma} \quad (183)$$

$$Z_i = \boldsymbol{\phi}_i^\dagger \left( \mathbf{I} + \gamma \sum_{j \neq i} d_j^* \boldsymbol{\phi}_j \boldsymbol{\phi}_j^\dagger \right)^{-1} \boldsymbol{\phi}_i \quad (184)$$

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

We make use of Theorem 14 with

$$\xi = n \quad (185)$$

$$\boldsymbol{\Phi} = \mathbf{Q} \mathbf{G} \quad (186)$$

$$\mathbf{Q} = \mathbf{A} \mathbf{F} = [\mathbf{q}_1, \dots, \mathbf{q}_n] \quad (187)$$

$$\mathbf{D} = \boldsymbol{\Upsilon} \quad (188)$$

and (184) takes the form

$$Z_i = |\mathbf{G}_i|^2 \mathbf{q}_i^\dagger \left( \mathbf{I} + \gamma \sum_{j \neq i} |\mathbf{G}_j|^2 d_j^* \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i. \quad (189)$$

Taking the limit of (189) as  $n \rightarrow \infty$ , Lemma 12 (see Appendix VIII) implies that almost surely

$$\lim_{n \rightarrow \infty} \mathbf{q}_i^\dagger \left( \mathbf{I} + \gamma \sum_{j \neq i} |\mathbf{G}_j|^2 \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i = \alpha. \quad (190)$$

Thus, the KKT condition in (183) becomes in the limit

$$\mathbb{E} \left[ \frac{\alpha \gamma |\mathbf{G}|^2}{1 + \alpha \gamma d_i^* |\mathbf{G}|^2} \right] = \frac{1}{\nu} \quad (191)$$

implying that the optimal  $d_i^*$  must be a constant for all  $i$ , which in turns yields that  $\mathbf{D}^* = \mathbf{I}$  is the unique maximizer in (182), in the limit of large  $n$ .  $\square$

## APPENDIX II

### COMBINATORIAL DEFINITIONS AND FACTS

*Definition 4 [14], [15]:* Let  $\mathcal{X}$  denote the index set  $\{1, \dots, |\mathcal{X}|\}$ . An  $\ell$ -partition of  $\mathcal{X}$  is a set  $\rho[\ell] = \{\mathcal{V}_1, \dots, \mathcal{V}_\ell\}$  of subsets  $\mathcal{V}_i \subseteq \mathcal{X}$  such that

$$\begin{aligned} & \mathcal{V}_i \neq \emptyset \quad \forall i = 1, \dots, \ell \\ & \mathcal{V}_i \cap \mathcal{V}_j = \emptyset \quad \forall i \neq j \\ & \bigcup_{i=1}^{\ell} \mathcal{V}_i = \mathcal{X}. \end{aligned} \quad (192)$$

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

*Definition 5:* Let  $\mathbf{m} = [m_1, \dots, m_\ell]$  be an  $\ell$ -dimensional vector whose entries are positive integers such that

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

and  $0 < m_1 \leq \dots \leq m_\ell$ . An  $(\mathbf{m}, \ell)$ -partition of  $\mathcal{X}$ , denoted by  $\rho[\mathbf{m}, \ell]$ , is an  $\ell$ -partition with blocks of cardinality  $m_1, \dots, m_\ell$ .<sup>5</sup>

Let  $\mathbf{z} = (z_1, \dots, z_{|\mathcal{X}|})$  with components indexed by  $\mathcal{X}$ . A partition  $\rho[\mathbf{m}, \ell]$  of  $\mathcal{X}$  induces a corresponding partition of  $\mathbf{z}$  into subvectors, or “multisets,”  $\{z_j : j \in \mathcal{V}_i\}$ . Adopting a Matlab-like notation, we will indicate these subvectors as  $\mathbf{z}(\mathcal{V}_i)$ . We denote the set of all partitions of  $\mathcal{X}$  as  $\mathfrak{P}(\mathcal{X})$ , and the set of all  $\ell$ -partitions as  $\mathfrak{P}_\ell(\mathcal{X})$ . Obviously,  $\mathfrak{P}(\mathcal{X}) = \bigcup_{\ell=1}^{|\mathcal{X}|} \mathfrak{P}_\ell(\mathcal{X})$ .

### A. Lattice of Partitions and the Degree of Inclusion

The natural partial order relation on  $(\mathcal{X})$  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 a refinement of  $\sigma[w]$ , 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 block of  $\rho[\ell]$  is a subset of some block 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  $\mathfrak{P}(\mathcal{X})$  is the 1-partition  $\{\mathcal{X}\}$  and the finest element of  $\mathfrak{P}(\mathcal{X})$  is the  $|\mathcal{X}|$ -partition  $\{\{1\}, \dots, \{|\mathcal{X}|\}\}$ .

The set  $\mathfrak{P}(\mathcal{X})$  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  $\pi$  such that  $\pi \geq \rho$  and  $\pi \geq \sigma$  (least upper bound), and  $\rho \wedge \sigma$  is the coarsest partition  $\pi$  such that  $\pi \leq \rho$  and  $\pi \leq \sigma$  (largest lower bound).  $\mathfrak{P}(\mathcal{X})$  is 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  $\rho \leq \tau$ ). Also, 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  $\mathfrak{P}(\mathcal{X}) \times \mathfrak{P}(\mathcal{X}) \rightarrow \mathfrak{P}(\mathcal{X})$ ). 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  $\ell$  has a node for each partition in  $\mathfrak{P}_\ell(\mathcal{X})$ . Then, an edge  $(\rho, \sigma)$  in the graph exists if and only if  $\rho \prec \sigma$ . Fig. 3 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.

Next, we introduce a function  $\zeta : \mathfrak{P}(\mathcal{X}) \times \mathfrak{P}(\mathcal{X}) \rightarrow \mathbb{Z}$ , referred to as *degree of inclusion*, that plays an important role in some computations needed in the proofs of our main results in the following.

<sup>5</sup>It is customary to indicate the “type” of the partition by specifying  $\mathbf{m}$  as an ordered set. For example, the partitions  $\{\{1, 2, 3\}, \{4\}\}$  and  $\{\{2\}, \{1, 3, 4\}\}$  of the set  $\mathcal{X} = \{1, 2, 3, 4\}$  are both of type  $[1, 3]$ : they are both  $([1, 3], 2)$  partitions.

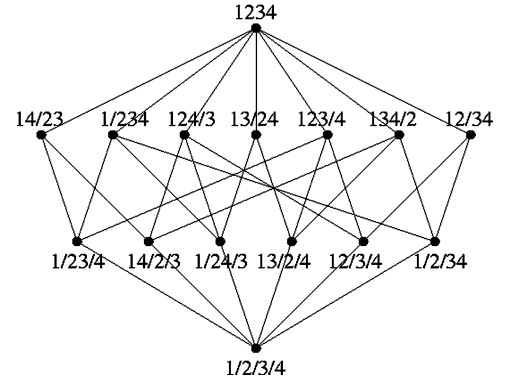


Fig. 3. Hasse diagram of the partially ordered set  $\mathfrak{P}(\{1, 2, 3, 4\})$ .

*Definition 7:* Consider two partitions  $\rho[\ell] < \sigma[w]$  in  $[\mathcal{X}]$ . For any integer  $w \leq q \leq \ell$ , define the set of  $q$ -partitions “in between”  $\rho[\ell]$  and  $\sigma[w]$ , i.e.,

$$[\rho, \sigma]_q = \{\pi \in \mathfrak{P}_q(\mathcal{X}) : \rho \leq \pi \leq \sigma\}. \quad (193)$$

The degree of inclusion  $\zeta$  of the pairs  $\rho[\ell], \sigma[w]$  is defined as

$$\zeta(\rho[\ell] \rightarrow \sigma[w]) = \begin{cases} 0, & \rho[\ell] \not\leq \sigma[w] \\ 1, & \rho[\ell] = \sigma[w] \\ -1, & w = \ell - 1 \text{ and } \rho[\ell] < \sigma[w] \end{cases} \quad (194)$$

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

$$\zeta(\rho[\ell] \rightarrow \sigma[w]) = \sum_{a=w+1}^{\ell-1} \left( |[\rho[\ell], \sigma[w]]_a| + \sum_{b=w+1}^{a-1} (-1)^{a-b} \times \sum_{\pi \in [\rho[\ell], \sigma[w]]_{b+1}} |[\pi, \sigma[w]]_b| \right) - 1. \quad (195)$$

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  $\sum_{a=w+1}^{\ell-1} |[\rho[\ell], \sigma[w]]_a|$  is equal to the total number of nodes in the subgraph formed by all (directed) paths joining  $\rho[\ell]$  with  $\sigma[w]$ . Furthermore, for any  $a \in \{w+2, \dots, \ell-1\}$ , and  $b \in \{w+1, \dots, a-1\}$ ,  $\sum_{\pi \in [\rho[\ell], \sigma[w]]_{b+1}} |[\pi, \sigma[w]]_b|$  is given by the total number of edges pointing upward of the  $(b+1)$ th layer in the subgraph of the paths joining  $\rho$  with  $\sigma$ .

The degree of inclusion  $\zeta$  satisfies the following additive decomposition:

$$\zeta(\rho[\ell] \rightarrow \sigma[w]) = - \sum_{a=w}^{\ell-1} \sum_{\pi \in [\rho[\ell], \sigma[w]]_a} \zeta(\pi \rightarrow \sigma[w]). \quad (196)$$

*Example 1:* Referring to the diagram of Fig. 3

$$\zeta(\underbrace{\{\{1\}; \{2\}; \{3\}; \{4\}\}}_{\rho} \rightarrow \underbrace{\{\{1, 2, 3, 4\}\}}_{\sigma})$$

$$= \sum_{a=2}^3 \left( \left| [\rho, \sigma]_a \right| + \sum_{b=2}^{a-1} (-1)^{a-b} \sum_{\pi \in [\rho, \sigma]_{b+1}} \left| [\pi, \sigma]_b \right| \right) - 1 \quad (197)$$

$$= \sum_{a=2}^3 \left| [\rho, \sigma]_a \right| - \sum_{\pi \in [\rho, \sigma]_3} \left| [\pi, \sigma]_2 \right| - 1 \quad (198)$$

$$= 6 + 7 - 6 \cdot 3 - 1 = -6. \quad (199)$$

Next, we wish to check the validity of (196). We have one partition at layer 1, namely  $\{\{1, 2, 3, 4\}\}$ . At layer 2, we have seven partitions  $\pi[2]$ , with degree of inclusion  $\zeta(\pi[2] \rightarrow \{\{1, 2, 3, 4\}\}) = -1$ . Then, we have six partitions  $\pi[3]$  at layer 3. Their degree of inclusion is  $\zeta(\pi[3] \rightarrow \{\{1, 2, 3, 4\}\}) = 2$ . In order to see this, notice that the subgraph of partitions  $\pi[3] \leq \pi \leq \{\{1, 2, 3, 4\}\}$  consists has three intermediate nodes  $\pi[2]$  and one top node. Hence,  $\zeta(\pi[3] \rightarrow \{\{1, 2, 3, 4\}\}) = -(1 - 3) = 2$ . Eventually, using (196), we have

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

$$= -(1 + 7 \cdot (-1) + 6 \cdot 2) = -6 \quad (200)$$

which coincides with the previous direct calculation.

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

For refinements that involve the splitting of more than one block of the top partition, the corresponding graph is obtained as the Cartesian product graph of single-block partitions. For example, consider the subgraph of Fig. 3 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, and the corresponding graph is given by the Cartesian product of the graphs of the two two-way partitions, as shown in Fig. 4.

In general, consider two nested partitions  $\rho[\ell] \leq \sigma[w]$  such that each  $i$ th block of  $\sigma[w]$  is partitioned into  $d_i$  blocks of  $\rho[\ell]$ . It can be shown that  $\zeta$  satisfies the following multiplicative decomposition:

$$\zeta(\rho[\ell] \rightarrow \sigma[w]) = \prod_{i=1}^w \zeta(d_i) \quad (201)$$

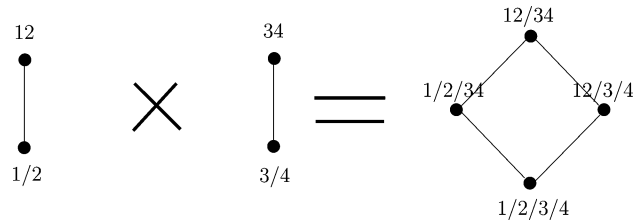


Fig. 4. Hasse diagram of the  $2 \times 2$ -way partition refinement from  $\{\{1, 2\}, \{3, 4\}\}$  to  $\{\{1\}, \{2\}, \{3\}, \{4\}\}$ .

where, with some notational abuse, we denote by  $\zeta(d)$  the value of the degree of inclusion for a  $d$ -way partition that depends only on  $d$  as noticed before.

The sum and product rules (196) and (201) allow very simple recursive computation of the inclusion index.

*Example 2:* Referring to the diagram of Fig. 4, direct calculation shows that

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

Using the product rule, we have

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

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

$$= (-1)(-1) = 1.$$

A more involved example is given in Fig. 5. Consider partitions

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

The first is obtained by a three-way partition of the block  $\{1, 2, 3\}$  and a two-way partition of the block  $\{4, 5\}$  of the second. Hence, the inclusion index is readily given by  $\zeta(3)\zeta(2) = 2(-1) = -2$ . The corresponding Hasse diagram of Fig. 5 is obtained as the Cartesian product of a three-way and a two-way partition. One can check by direct calculation that, indeed

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

## B. Good Partitions

In some calculations in Appendixes I–IX, we will work with vectors  $\mathbf{z}$  of  $k$  components defined on the ring  $\mathbb{Z}_n$  of integer residues modulo- $n$ . In this section, we consider the index set  $\mathcal{X} = \{1, \dots, k\}$  and the corresponding partitions in  $\mathfrak{P}(\mathcal{X})$ , inducing the partition of a vector  $\mathbf{z}$  into subvectors as said before.

*Definition 8:* Fix  $\mathbf{z} \in \mathbb{Z}_n^k$ . We say that  $\rho[\mathbf{m}, \ell] = \{\mathcal{V}_1, \dots, \mathcal{V}_\ell\}$  of  $\{1, \dots, k\}$  is a good partition for  $\mathbf{z}$  if

$$\text{sum}[\mathbf{z}(\mathcal{V}_i)] = 0 \quad \forall i = 1, \dots, \ell \quad (202)$$

where  $\text{sum}[\cdot]$  denotes the sum modulo- $n$  (i.e., in the ring  $\mathbb{Z}_n$ ) of the components of the argument vector.

The condition that  $\rho[\mathbf{m}, \ell]$  is a good partition of  $\mathbf{z} \in \mathbb{Z}_n^k$  is equivalently expressed by saying that  $\mathbf{z}$  belongs to the solution space of a linear equation over  $\mathbb{Z}_n$ . In particular, a partition  $\rho[\mathbf{m}, \ell]$  is associated to the incidence matrix  $\mathbf{A}_\rho$  with  $k$  rows and

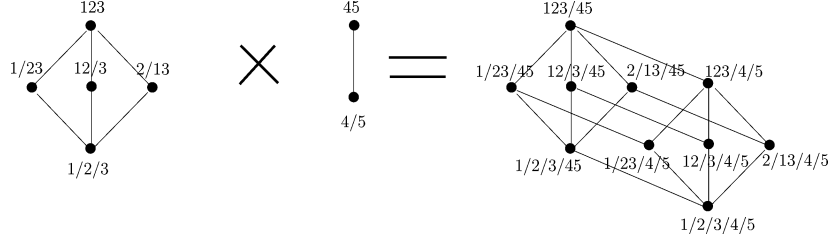


Fig. 5. Hasse diagram of the  $3 \times 2$ -way partition refinement from  $\{\{1, 2, 3\}, \{4, 5\}\}$  to  $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$ .

$\ell$  columns, such that the  $i$ th column of  $\mathbf{A}_\rho$  contains 1s for all positions  $j \in \mathcal{V}_i$  and 0s elsewhere.  $\rho[\mathbf{m}, \ell]$  is a good partition for  $\mathbf{z}$  if and only if  $\mathbf{z}$  is a solution of the linear equation  $\mathbf{z}\mathbf{A}_\rho = \mathbf{0}$ . Therefore, by definition, the set of the vectors  $\mathbf{z} \in \mathbb{Z}_n^k$  for which  $\rho[\mathbf{m}, \ell]$  is good is given by  $\text{Ker}(\mathbf{A}_\rho)$  (the kernel of the linear map  $\mathbb{Z}_n^k \rightarrow \mathbb{Z}_n^\ell$  defined by  $\mathbf{A}_\rho$ ). The kernel of a linear transformation over the ring  $\mathbb{Z}_n$  is a  $\mathbb{Z}_n$ -module. Rather than using this standard algebraic term, we will use a more familiar coding theoretic terminology:  $\text{Ker}(\mathbf{A}_\rho)$  is a *linear code* of length  $k$  over  $\mathbb{Z}_n$ , with *parity-check equation* given by  $\mathbf{z}\mathbf{A}_\rho = \mathbf{0}$ .

Notice that the columns of  $\mathbf{A}_\rho$  contain only elements 0 and 1. Therefore, there is no column which greatest common divisor that is a divisor of zero in  $\mathbb{Z}_n$ . It follows that the solution space of each  $r$ th parity-check equation defining  $\text{Ker}(\mathbf{A}_\rho)$  is isomorphic to  $\mathbb{Z}_n^{|\mathcal{V}_r|-1}$ . Furthermore, by definition of partition it follows that the columns of  $\mathbf{A}_\rho$  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  $\text{Ker}(\mathbf{A}_\rho)$  is isomorphic to the Cartesian product

$$\mathbb{Z}_n^{m_1-1} \times \mathbb{Z}_n^{m_2-1} \times \dots \times \mathbb{Z}_n^{m_\ell-1}. \quad (203)$$

It follows that  $|\text{Ker}(\mathbf{A}_\rho)| = n^{k-\ell}$  depends on the partition only through the number of blocks  $\ell$ .

*Lemma 3:* Consider the partitions,  $\rho[\mathbf{m}, \ell]$  and  $\sigma[\mathbf{v}, w]$  of  $\mathcal{X} = \{1, \dots, k\}$ , then  $\text{Ker}(\mathbf{A}_\rho) \subseteq \text{Ker}(\mathbf{A}_\sigma)$  if and only if  $\rho[\mathbf{m}, \ell] \leq \sigma[\mathbf{v}, w]$ .

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

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

Hence,  $\mathbf{z} \in \text{Ker}(\mathbf{A}_\sigma)$ . This shows sufficiency. In order to show necessity, without loss of generality suppose that  $\ell \geq w$ ,  $\rho[\mathbf{m}, \ell] \not\leq \sigma[\mathbf{v}, w]$ . There must exist a block  $\mathcal{V}$  of  $\rho$  with nonempty intersection with at least two blocks of  $\sigma$  (otherwise,  $\rho$  would be a refinement of  $\sigma$ ). Denote these blocks as  $\mathcal{U}$  and  $\mathcal{U}'$ . We choose a vector  $\mathbf{z} \in \text{Ker}(\mathbf{A}_\rho)$  such that all components

<sup>6</sup>Notice that while this would be a trivial conclusion if  $\mathbb{Z}_n$  were a field, the condition that the coefficients of the equation are relatively prime with  $n$  is important in a ring that has divisors of zero, as in the case where  $n$  is not a prime. For example, if  $n = 8$ , the equation  $x + y = 0$  has eight solutions, but the equation  $4x + 2y = 0$  has 16 solutions.

are equal to zero but two nonzero components,  $z_i = 1$  for  $i \in \mathcal{V} \cap \mathcal{U}$  and  $z_{i'} = -1$  for  $i' \in \mathcal{V} \cap \mathcal{U}'$ . Clearly,  $\mathbf{z} \notin \text{Ker}(\mathbf{A}_\sigma)$ . This shows that  $\text{Ker}(\mathbf{A}_\rho) \not\subseteq \text{Ker}(\mathbf{A}_\sigma)$ .  $\square$

### APPENDIX III

#### SUMMING FUNCTIONS $f : \mathbb{Z}_n^k \rightarrow \mathbb{C}$

A partition  $\rho$  of the index set  $\mathcal{X} = \{1, \dots, k\}$  induces an equivalence relation on the elements of  $\mathcal{X}$ . In particular, we say that two indices  $i, j \in \mathcal{X}$  are equivalent with respect to the partition  $\rho$  (and write  $i \sim j$ ) if they belong to the same block of  $\rho$ .

*Definition 9:* We define  $\mathcal{S}(\rho)$  to be the subset of  $\mathbb{Z}_n^k$  of all vectors that are constant over the blocks of  $\rho[\ell]$ , i.e.,

$$\mathcal{S}(\rho) = \left\{ \mathbf{z} \in \mathbb{Z}_n^k : z_i = z_j \text{ if } i \sim j \right\}. \quad (204)$$

*Lemma 4:* Consider two partitions  $\rho[\ell]$  and  $\sigma[w]$  of  $\mathcal{X}$  with  $w \leq \ell$ . Then,  $\mathcal{S}(\sigma[w]) \subseteq \mathcal{S}(\rho[\ell])$  if and only if  $\sigma[w] \geq \rho[\ell]$ .

*Definition 10:* We define  $\mathcal{S}^-(\rho) \subset \mathbb{Z}_n^k$  of all vectors that are constant over the blocks of  $\rho$  and take on distinct values in different blocks, i.e.,

$$\mathcal{S}^-(\rho) = \left\{ \mathbf{z} \in \mathbb{Z}_n^k : z_i = z_j \text{ if } i \sim j, \text{ otherwise } z_i \neq z_j \right\}. \quad (205)$$

It is easy to see that

$$\mathcal{S}^-(\rho[\ell]) = \mathcal{S}(\rho[\ell]) - \bigcup_{\sigma[\ell-1] > \rho[\ell]} \mathcal{S}(\sigma[\ell-1]) \quad (206)$$

$$= \mathcal{S}(\rho[\ell]) - \bigcup_{w=1}^{\ell-1} \bigcup_{\sigma[w] > \rho[\ell]} \mathcal{S}^-(\sigma[w]). \quad (207)$$

For all  $1 \leq \ell, w \leq k$  and corresponding distinct partitions  $\rho[\ell] \neq \sigma[w]$

$$\mathcal{S}^-(\rho[\ell]) \cap \mathcal{S}^-(\sigma[w]) = \emptyset. \quad (208)$$

Furthermore, the union of such sets over all the partitions exhausts the whole  $\mathbb{Z}_n^k$ , i.e.,

$$\mathbb{Z}_n^k = \bigcup_{\rho \in \mathfrak{P}(\mathcal{X})} \mathcal{S}^-(\rho). \quad (209)$$

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

*Lemma 5:* Consider a function  $f : \mathbb{Z}_n^k \rightarrow \mathbb{C}$ , and a partition  $\rho[\ell] \in \mathfrak{P}(\mathcal{X})$ . Then

$$\sum_{\mathbf{z} \in \mathbb{S}^-(\rho[\ell])} f(\mathbf{z}) = \sum_{w=1}^{\ell} \sum_{\sigma[w] \geq \rho[\ell]} \zeta(\rho[\ell] \rightarrow \sigma[w]) \sum_{\mathbf{z} \in \mathbb{S}(\sigma[w])} f(\mathbf{z}) \quad (210)$$

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

*Proof:* The proof is by induction. For  $\ell = 1$ , (210) follows from the facts that  $\mathbb{S}_1^-(\rho[1])$  is the set of constant vectors  $\{\mathbf{z} = (z, z, \dots, z) : z \in \mathbb{Z}_n\}$ ,  $\zeta(\rho \rightarrow \rho) = 1$ , and the sum over  $\sigma$  contains only the term  $\sigma[1] = \rho[1]$ .

Now let us assume that (210) holds for all  $1 \leq h \leq \ell - 1$ . We wish to show that it holds also for  $\ell$ . Using (207) and (208), we have (211)–(217), shown at the bottom of the page, where (214) follows by changing the summation order, (216) follows from (196), and (217) follows from the definition of  $\zeta$ .  $\square$

#### APPENDIX IV STRONG MIXING PROCESSES

*Definition 11:* Let  $(\Omega, \mathcal{F}, \mathcal{P})$  be a probability space. For any  $\sigma$ -fields  $\mathcal{A} \subset \mathcal{F}$  and  $\mathcal{B} \subset \mathcal{F}$ , define the following measure of dependence, which we refer to as the strong mixing coefficient:

$$\rho(\mathcal{A}, \mathcal{B}) = \sup\{|\mathcal{P}(B|A) - \mathcal{P}(B)|\} \quad (218)$$

where  $B \in \mathcal{B}$ ,  $A \in \mathcal{A}$  and  $\mathcal{P}(A) \geq 0$ .

For  $\infty \leq J \leq L \leq \infty$ , the  $\sigma$ -field generated by  $X_p$ ,  $p = J, \dots, L$  is denoted by

$$\mathcal{F}_J^L = \sigma(X_p, p = J, \dots, L). \quad (219)$$

*Definition 12:* Let  $\{X_p; p \in \mathbb{Z}\}$  be a stationary random real process with  $\mathbb{E}[X_p] = \mu$  and  $\mathbb{E}[X_p^2] = \sigma^2 + \mu^2$ .  $\{X_p; p \in \mathbb{Z}\}$  is a strong mixing process if

$$\lim_{n \rightarrow \infty} \sup_{j \in \mathbb{Z}} \rho(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty}) = 0. \quad (220)$$

Furthermore, we say that  $\{X_p; p \in \mathbb{Z}\}$  is a strong mixing process with polynomial convergence rate if

$$\left| \sup_{j \in \mathbb{Z}} \rho(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^{\infty}) \right| \leq r(n) \quad (221)$$

with  $r(n) = \frac{1}{(n-1)^\beta}$  with  $\beta > 0$ .

*Example 3:* Irreducible and aperiodic chains, either with a countable state space or a finite state space are strong mixing processes with polynomial convergence rate.

*Proposition 1:* Let  $\{X_p; p \in \mathbb{Z}\}$  be a strong mixing process with polynomial convergence rate as in Definition 12. For any polynomials  $p_0 \dots p_k$  and indices  $r_1 \dots r_k$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[p_1(X_{r_1}) \dots p_k(X_{r_k}) p_0(X_n)] \\ = \mathbb{E}[p_1(X_{r_1}) \dots p_k(X_{r_k})] \lim_{n \rightarrow \infty} \mathbb{E}[p_0(X_n)] \end{aligned} \quad (222)$$

where the convergence rate is polynomial in the sense specified in Definition 12.

A special case of strong mixing processes is the stationary process where  $\{X_p; p \in \mathbb{Z}_n\}$  are independent identically distributed.

#### APPENDIX V FOURIER COEFFICIENTS OF STATIONARY PROCESSES

We summarize some of the statistical properties of the Fourier coefficients  $c_\ell$  of a stationary process. Let  $\{X_p; p \in \mathbb{Z}\}$  be a

$$\sum_{\mathbf{z} \in \mathbb{S}^-(\rho[\ell])} f(\mathbf{z}) = \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) - \sum_{h=1}^{\ell-1} \sum_{\sigma[h] > \rho[\ell]} \sum_{\mathbf{z} \in \mathbb{S}^-(\sigma[h])} f(\mathbf{z}) \quad (211)$$

$$= \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) - \sum_{h=1}^{\ell-1} \sum_{\sigma[h] > \rho[\ell]} \sum_{w=1}^h \sum_{\tau[w] > \sigma[h]} \zeta(\sigma[h] \rightarrow \tau[w]) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (212)$$

$$= \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) - \sum_{h=1}^{\ell-1} \sum_{w=1}^{\ell-1} \sum_{\sigma[h] > \rho[\ell]} \sum_{\tau[w] > \sigma[h]} \zeta(\sigma[h] \rightarrow \tau[w]) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (213)$$

$$= \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) - \sum_{h=1}^{\ell-1} \sum_{w=1}^{\ell-1} \sum_{\tau[w] > \rho[\ell]} \sum_{\sigma \in [\rho[\ell], \tau[w]]_h} \zeta(\sigma \rightarrow \tau[w]) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (214)$$

$$= \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) - \sum_{w=1}^{\ell-1} \sum_{\tau[w] > \rho[\ell]} \left( \sum_{h=1}^{\ell-1} \sum_{\sigma \in [\rho[\ell], \tau[w]]_h} \zeta(\sigma \rightarrow \tau[w]) \right) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (215)$$

$$= \sum_{\mathbf{z} \in \mathbb{S}(\rho[\ell])} f(\mathbf{z}) + \sum_{w=1}^{\ell-1} \sum_{\tau[w] > \rho[\ell]} \zeta(\rho[\ell] \rightarrow \tau[w]) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (216)$$

$$= \sum_{w=1}^{\ell} \sum_{\tau[w] > \rho[\ell]} \zeta(\rho[\ell] \rightarrow \tau[w]) \sum_{\mathbf{z} \in \mathbb{S}(\tau[w])} f(\mathbf{z}) \quad (217)$$

stationary real-valued random process with  $\mathbb{E}[X_p] = \mu$  and  $\mathbb{E}[X_p^2] = \sigma^2 + \mu^2$ .

Denote the DFT coefficients of  $\{X_p; p = 0, \dots, n-1\}$  by

$$c_\ell = \frac{1}{n} \sum_{p=0}^{n-1} X_p e^{-j\frac{2\pi}{n}p\ell}. \quad (223)$$

As usual, the index set  $\ell = 0, 1, \dots, n-1$  is identified with the ring  $\mathbb{Z}_n$ , with the corresponding modulo- $n$  ring operations. Since  $\{X_p\}$  is real-valued, it follows that

$$c_\ell^* = c_{-\ell} = c_{n-\ell}. \quad (224)$$

Furthermore, both  $c_0$  and (if  $n$  is even)  $c_{n/2}$  are real.

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

$$\mathbb{E}[c_0] = \frac{1}{n} \sum_{p=0}^{n-1} \mathbb{E}[X_p] = \mu \quad (225)$$

$$\mathbb{E}[c_\ell] = \frac{\mu}{n} \sum_{p=0}^{n-1} e^{-j\frac{2\pi}{n}p\ell} = 0, \quad \ell \neq 0. \quad (226)$$

*Lemma 6 [16]:* For any integer  $k = 1, 2, \dots$  and  $\{0 < \alpha_i < \frac{1}{2}\}_{i=1}^k$  such that  $\alpha_i \neq \alpha_j$  if  $i \neq j$ , let  $\mathcal{L} = \{\lfloor \alpha_1 n \rfloor, \dots, \lfloor \alpha_k n \rfloor\}$ . As  $n \rightarrow \infty$ , the joint distribution of the Fourier coefficients (223) of a stationary process with variance  $\sigma^2$

$$\{\sqrt{nc_\ell}\}, \quad \ell \in \mathcal{L} \quad (227)$$

converges to a proper-complex Gaussian product distribution with zero mean and variances  $\sigma^2$ . Furthermore,  $\sqrt{n}(c_0 - \mu)$  and (if  $n$  is even)  $\sqrt{nc_{n/2}}$  are real valued with variance  $\sigma^2$ , asymptotically jointly Gaussian with (227).

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

*Lemma 7:* For any integer  $k > 0$ , consider an index vector  $\mathbf{j} = [j_1, j_2, \dots, j_k] \in \mathbb{Z}_n^k$ , such that  $\text{sum}[\mathbf{j}] \neq 0$ . Then

$$\mathbb{E}[c_{j_1} c_{j_2} \cdots c_{j_k}] = 0. \quad (228)$$

*Lemma 8:* For any integer  $k > 0$ , let  $\mathbf{j} = [j_1, j_2, \dots, j_k] \in \mathbb{Z}_n^k$  be such that  $\text{sum}[\mathbf{j}] = 0$ . Let  $\{X_p; p \in \mathbb{Z}_n\}$  be independent identically distributed and let  $c_\ell$  with  $\ell \in \mathbb{Z}_n$  be the Fourier coefficient defined in (223). Then

$$\mathbb{E}[c_{j_1} c_{j_2} \cdots c_{j_k}] = \frac{1}{n^k} \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \mathcal{Q}(\kappa, \tau[\mathbf{m}, \ell]) \times \sum_{\mathbf{p} \in \mathcal{S}(\tau[\mathbf{m}, \ell])} e^{-j\frac{2\pi}{n}\mathbf{j} \cdot \mathbf{p}} \quad (229)$$

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

$$\mathcal{Q}(k, \tau[\mathbf{m}, \ell]) = \sum_{w=\ell}^k \sum_{\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]} \left( \prod_{i=1}^w \mu_{v_i} \right) \times \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]). \quad (230)$$

with  $\mu_j = \mathbb{E}[X_j^2]$ , with  $\tau[\mathbf{m}, \ell]$  and  $\sigma[\mathbf{v}, w]$  denoting partitions of the index set  $\{1, \dots, k\}$ , and where  $\mathcal{S}(\tau[\mathbf{m}, \ell])$  is given in Definition 9 (see Appendix III) and denotes the set of all vectors  $\mathbf{z} \in \mathbb{Z}_n^k$  with constant values on the blocks of the partition  $\tau[\mathbf{m}, \ell]$ . Finally,  $v_i$  with  $i = 1, \dots, w$  denotes the cardinalities of the blocks  $\mathcal{V}_1, \dots, \mathcal{V}_w$  of the partition  $\sigma[\mathbf{v}, w]$ .

*Proof:* We can write

$$\mathbb{E}[c_{j_1} c_{j_2} \cdots c_{j_k}] = \frac{1}{n^k} \sum_{\mathbf{p} \in \mathbb{Z}_n^k} \mathbb{E}[X_{p_1} \cdots X_{p_k}] e^{-j\frac{2\pi}{n}\mathbf{j} \cdot \mathbf{p}}. \quad (231)$$

Recalling the decomposition (209) of  $\mathbb{Z}_n^k$ , we have (232)–(236), shown at the bottom of the next page, where (236) coincides with the desired result, (232) follows from the definition of the set  $\mathcal{S}^-(\sigma[\mathbf{v}, w])$ , and from the fact that the process is stationary, (233) follows by denoting  $\mathbb{E}[X_{p_1}^2] = \mu_{v_1}$ , (234) is an application of Lemma 5, and (235) and (236) follow by rearranging the terms in the summations.  $\square$

Given a polynomial  $q_r(\cdot)$  and a stationary random process  $\{X_p; p = 0, \dots, n-1\}$ , let us now denote by  $c_{\ell, q_r}$  the Fourier coefficients of the new stationary random process  $\{q_r(X_p); p = 0, \dots, n-1\}$  by

$$c_{\ell, q_r} = \frac{1}{n} \sum_{p=0}^{n-1} q_r(X_p) e^{-j\frac{2\pi}{n}p\ell}. \quad (237)$$

Following in the footsteps of the proof of Lemma 8, we can show the following.

*Lemma 9:* For any integer  $k > 0$ , let  $\mathbf{j} = [j_1, j_2, \dots, j_k] \in \mathbb{Z}_n^k$  be such that  $\text{sum}[\mathbf{j}] = 0$  and let  $q_1(\cdot), \dots, q_k(\cdot)$  denote  $k$  polynomials. Let  $\{X_p; p \in \mathbb{Z}_n\}$  be i.i.d. and let  $c_{\ell, q_r}$  with  $\ell \in \mathbb{Z}_n$  and  $r = 1, \dots, k$  be the Fourier coefficient defined in (237). Then

$$\mathbb{E}[c_{j_1, q_1} c_{j_2, q_2} \cdots c_{j_k, q_k}] = \frac{1}{n^k} \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \mathcal{P}_{q_1, \dots, q_k}(k, \tau[\mathbf{m}, \ell]) \sum_{\mathbf{p} \in \mathcal{S}(\tau[\mathbf{m}, \ell])} e^{-j\frac{2\pi}{n}\mathbf{j} \cdot \mathbf{p}} \quad (238)$$

where

$$\mathcal{P}_{q_1, \dots, q_k}(k, \tau[\mathbf{m}, \ell]) = \sum_{w=\ell}^k \sum_{\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]} \left( \prod_{i=1}^w \mu_{\mathcal{V}_i}(q_1, \dots, q_k) \right) \times \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]) \quad (239)$$

and where

$$\mu_{\mathcal{V}_i}(q_1, \dots, q_k) = \mathbb{E} \left[ \prod_{r \in \mathcal{V}_i} q_r(X_1) \right] \quad (240)$$

where  $\mathcal{V}_1, \dots, \mathcal{V}_w$  denoting the blocks of the partition  $\sigma[\mathbf{v}, w]$ .

Notice that for  $q_1(x) = \dots = q_k(x) = x$ , we get back the result of Lemma 8.

#### APPENDIX VI

##### ON THE MOMENTS OF THE PRODUCTS OF RANDOM DIAGONAL MATRICES WITH CIRCULANT MATRICES

In this Appendix, we prove some auxiliary results on the moments of products of diagonal random matrices with random circulant matrices that we are going to use in the proof of our main result. Since it is useful to exploit the structure of the ring  $\mathbb{Z}_n$  for the indices of the DFT coefficients, we will index the elements of  $n \times n$  matrices from 0 to  $n-1$  instead of from 1 to  $n$ .

Let  $\mathbf{X} = \text{diag}\{X_0, \dots, X_{n-1}\}$  be a semidefinite random diagonal matrix whose diagonal elements  $X_i$  are i.i.d. random variables. As introduced before, we will use the notation  $\mu_j = \mathbb{E}[X_i^j]$  in order to indicate the  $j$ th moment of the diagonal elements of  $\mathbf{X}$ .

Let  $\mathbf{F}$  be the unitary DFT matrix as defined in (5) and let  $\mathbf{\Lambda}$  be a real diagonal matrix with diagonal elements  $[\mathbf{\Lambda}]_{ii} = \lambda_i$ . Then

$$\mathbf{\Psi} = \mathbf{F}\mathbf{\Lambda}\mathbf{F}^\dagger = \sum_{i=0}^{n-1} \lambda_i \mathbf{f}_i \mathbf{f}_i^\dagger \quad (241)$$

is a circulant matrix.

For an  $n \times n$  Hermitian random matrix  $\mathbf{V}$ , we define the normalized expected trace operator  $\phi(\mathbf{V})$  as

$$\phi(\mathbf{V}) = \frac{1}{n} \mathbb{E}[\text{tr} \mathbf{V}]. \quad (242)$$

We have the following results.

*Theorem 15:* For  $\mathbf{X}$  and  $\mathbf{\Psi}$  defined as above and for any positive integer  $k$

$$\phi((\mathbf{X}\mathbf{\Psi})^k) = \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \frac{\mathcal{Q}(k, \tau[\mathbf{m}, \ell])}{n^{k-\ell+1}} \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \left( \prod_{r=1}^k \lambda_{z_r} \right) \quad (243)$$

where  $\mathcal{Q}(k, \tau[\mathbf{m}, \ell])$  is defined in (230) of Lemma 8 with  $\tau[\mathbf{m}, \ell]$  denoting a the partition of the index set  $\{1, \dots, k\}$  and where  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  is the linear code over  $\mathbb{Z}_n$  defined by

$$\mathbb{G}(\tau[\mathbf{m}, \ell]) = \{\mathbf{z} \in \mathbb{Z}_n^k : \mathbf{z}\mathbf{B}\mathbf{A}_\tau = \mathbf{0}\} \quad (244)$$

with  $\mathbf{B}$  denoting the matrix of dimension  $k \times k$

$$\mathbf{B} = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \quad (245)$$

and  $\mathbf{A}_\tau$  denoting the incidence matrix of  $\tau[\mathbf{m}, \ell]$ , of dimension  $k \times \ell$ .

*Proof:* Note that

$$\begin{aligned} \phi((\mathbf{X}\mathbf{\Psi})^k) &= \phi \left( \left( \mathbf{X} \sum_{z=0}^{n-1} \mathbf{f}_z \lambda_z \mathbf{f}_z^\dagger \right)^k \right) \\ &= \phi \left( \sum_{z_1, \dots, z_k} \mathbf{X} \mathbf{f}_{z_1} \lambda_{z_1} \mathbf{f}_{z_1}^\dagger \mathbf{X} \dots \mathbf{X} \mathbf{f}_{z_k} \lambda_{z_k} \mathbf{f}_{z_k}^\dagger \right) \\ &= \phi \left( \sum_{z_1, \dots, z_k} \lambda_{z_1} \mathbf{f}_{z_1}^\dagger \mathbf{X} \mathbf{f}_{z_2} \lambda_{z_2} \mathbf{f}_{z_2}^\dagger \mathbf{X} \mathbf{f}_{z_3} \dots \lambda_{z_k} \mathbf{f}_{z_k}^\dagger \mathbf{X} \mathbf{f}_{z_1} \right) \\ &= \phi \left( \sum_{z_1, \dots, z_k} \left( \prod_{r=1}^k \lambda_{z_r} \right) \mathbf{f}_{z_1}^\dagger \mathbf{X} \mathbf{f}_{z_2} \mathbf{f}_{z_2}^\dagger \mathbf{X} \mathbf{f}_{z_3} \dots \mathbf{f}_{z_k}^\dagger \mathbf{X} \mathbf{f}_{z_1} \right) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[c_{j_1} c_{j_2} \cdots c_{j_k}] &= \frac{1}{n^k} \sum_{w=1}^k \sum_{\sigma[\mathbf{v}, w]} \sum_{\mathbf{p} \in \mathbb{S}^-(\sigma[\mathbf{v}, w])} \mathbb{E}[X_{p_1} \cdots X_{p_k}] e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \\ &= \frac{1}{n^k} \sum_{w=1}^k \sum_{\sigma[\mathbf{v}, w]} \left( \prod_{i=1}^w \mathbb{E}[X_1^{v_i}] \right) \sum_{\mathbf{p} \in \mathbb{S}^-(\sigma[\mathbf{v}, w])} e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \quad (232) \end{aligned}$$

$$= \frac{1}{n^k} \sum_{w=1}^k \sum_{\sigma[\mathbf{v}, w]} \left( \prod_{i=1}^w \mu_{v_i} \right) \sum_{\mathbf{p} \in \mathbb{S}^-(\sigma[\mathbf{v}, w])} e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \quad (233)$$

$$= \frac{1}{n^k} \sum_{w=1}^k \sum_{\sigma[\mathbf{v}, w]} \left( \prod_{i=1}^w \mu_{v_i} \right) \sum_{\ell=1}^w \sum_{\tau[\mathbf{m}, \ell] \geq \sigma[\mathbf{v}, w]} \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]) \sum_{\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])} e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \quad (234)$$

$$= \frac{1}{n^k} \sum_{\ell=1}^k \sum_{w=1}^k \sum_{\sigma[\mathbf{v}, w]} \sum_{\tau[\mathbf{m}, \ell] \geq \sigma[\mathbf{v}, w]} \left( \prod_{i=1}^w \mu_{v_i} \right) \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]) \sum_{\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])} e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \quad (235)$$

$$= \frac{1}{n^k} \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \sum_{w=\ell}^k \sum_{\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]} \left( \prod_{i=1}^w \mu_{v_i} \right) \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]) \sum_{\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])} e^{-j \frac{2\pi}{n} \mathbf{j} \cdot \mathbf{p}} \quad (236)$$

$$\begin{aligned}
&= \phi \left( \sum_{z_1, \dots, z_k} \left( \prod_{r=1}^k \lambda_{z_r} \right) c_{z_2 - z_1} c_{z_3 - z_2} \cdots c_{z_1 - z_k} \right) \\
&= \frac{1}{n} \sum_{z_1, \dots, z_k} \left( \prod_{r=1}^k \lambda_{z_r} \right) \mathbb{E} [c_{z_2 - z_1} c_{z_3 - z_2} \cdots c_{z_1 - z_k}]. \quad (246)
\end{aligned}$$

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

$$\begin{aligned}
v_1 &= z_2 - z_1 \\
v_s &= z_{s+1} - z_s, \quad s = 2, \dots, k-1 \\
v_k &= z_1 - z_k. \quad (247)
\end{aligned}$$

In vector form, we have that  $\mathbf{v} = \mathbf{zB}$  where  $\mathbf{B}$  is defined in (245). Furthermore, by construction,  $\sum_{r=1}^k v_r = 0$  over  $\mathbb{Z}$ , and therefore also over  $\mathbb{Z}_n$  (i.e., with respect to the modulo- $n$  sum). Using a notation already introduced, we have that  $\text{sum}[\mathbf{v}] = 0$ . By Lemma 8

$$\begin{aligned}
\mathbb{E} [c_{v_1} c_{v_2} \cdots c_{v_k}] &= \frac{1}{n^k} \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \mathcal{Q}(k, \tau[\mathbf{m}, \ell]) \\
&\quad \times \sum_{\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])} e^{-j \frac{2\pi}{n} \mathbf{v} \cdot \mathbf{p}} \quad (248)
\end{aligned}$$

$\mathcal{Q}(k, \tau[\mathbf{m}, \ell])$  is defined in (230) and where  $\mathbb{S}(\tau[\mathbf{m}, \ell])$  is defined in Appendix III and indicates the set of  $k$ -vectors over  $\mathbb{Z}_n$  with constant values over the blocks of the partition  $\tau[\mathbf{m}, \ell]$ . Denoting these blocks by  $\mathcal{V}_1, \dots, \mathcal{V}_\ell$ , for any  $\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])$ , the subvector  $\mathbf{p}(\mathcal{V}_r)$  has the form

$$\mathbf{p}(\mathcal{V}_r) = \underbrace{(h_r, h_r, \dots, h_r)}_{|\mathcal{V}_r| \text{ times}}$$

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

$$\begin{aligned}
&\sum_{\mathbf{p} \in \mathbb{S}(\tau[\mathbf{m}, \ell])} e^{-j \frac{2\pi}{n} \mathbf{v} \cdot \mathbf{p}} \\
&= \sum_{h_1, \dots, h_\ell=0}^{n-1} e^{-j \frac{2\pi}{n} \sum_{r=1}^{\ell} h_r \text{sum}[\mathbf{v}(\mathcal{V}_r)]} \\
&= \prod_{r=1}^{\ell} \left( \sum_{h=0}^{n-1} e^{-j \frac{2\pi}{n} h \text{sum}[\mathbf{v}(\mathcal{V}_r)]} \right) \\
&= \begin{cases} n^\ell, & \text{if } \text{sum}[\mathbf{v}(\mathcal{V}_r)] = 0 \forall r = 1, \dots, \ell \\ 0, & \text{otherwise.} \end{cases} \quad (249)
\end{aligned}$$

It follows that this term is not identically zero only if  $\tau[\mathbf{m}, \ell]$  is a good partition for the vector of indices  $\mathbf{v}$  (see Definition 8), i.e., if  $\mathbf{vA}_\tau = \mathbf{0}$ .

Now, we examine the set of index vectors  $\mathbf{z} \in \mathbb{Z}_n^k$  in the sum (246) that correspond to nonzero terms. Noticing that  $\mathbf{v} = \mathbf{zB}$  and that the nonzero terms correspond to index vectors  $\mathbf{v}$  such that  $\mathbf{vA}_\tau = \mathbf{0}$ , it follows that  $\mathbf{z}$  must satisfy the linear equation  $\mathbf{zBA}_\tau = \mathbf{0}$ . It follows that the sum over all  $\mathbf{z} \in \mathbb{Z}_n^k$  as in (246) is equivalent to summing over  $\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])$  as defined in (244). This concludes the proof.  $\square$

Following in the footsteps of the proof of Theorem 15, the following result can be also proved.

*Theorem 16:* Let  $\mathbf{X}$  and  $\Psi$  be defined as in Theorem 15. For any integer  $k > 0$ , and polynomials  $\{p_i, q_i : i = 1, \dots, k\}$

$$\begin{aligned}
&\phi(p_1(\Psi) q_1(\mathbf{X}) \cdots p_k(\Psi) q_k(\mathbf{X})) \\
&= \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \frac{\mathcal{P}_{q_1, \dots, q_k}(k, \tau[\mathbf{m}, \ell])}{n^{k-\ell+1}} \\
&\quad \times \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \left( \prod_{r=1}^k p_r(\lambda_{z_r}) \right) \quad (250)
\end{aligned}$$

where  $\mathcal{P}_{q_1, \dots, q_k}(k, \tau[\mathbf{m}, \ell])$  is defined as in (239) of Lemma 9 with  $\tau[\mathbf{m}, \ell]$  denoting a the partition of the index set  $\{1, \dots, k\}$  and where  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  is defined in (244).

Next, we examine the structure of the linear code  $\mathbb{G}(\tau[\ell])$  defined in (244), for some arbitrary partition  $\tau[\ell]$  of  $\{1, \dots, k\}$ . All ‘‘codewords’’  $\mathbf{z} \in \mathbb{G}(\tau[\ell])$  satisfy the parity-check equation  $\mathbf{zBA}_\tau = \mathbf{0}$ . This is a set of  $\ell$  linear equations in  $k$  variables. However, the rank of  $\mathbf{BA}_\tau$  is only  $\ell-1$ , in fact, by construction, the sum of the columns of  $\mathbf{BA}_\tau$  is equal to  $\mathbf{0}$ . It follows that the linear code  $\mathbb{G}(\tau[\ell])$  over  $\mathbb{Z}_n$  defined by the (redundant) parity-check equation  $\mathbf{zBA}_\tau$  is isomorphic to  $\mathbb{Z}_n^{k-\ell+1}$  and has size  $n^{k-\ell+1}$ .

Given the above algebraic structure,  $\mathbb{G}(\tau[\ell])$  after a suitable permutation of its components can be given in *systematic form*. In particular, there exists a matrix,  $\mathbf{K}_\tau \in \mathbb{Z}_n^{(k-\ell+1) \times (\ell-1)}$ , such that

$$\mathbb{G}(\tau[\ell]) = \{ [\mathbf{z} | \mathbf{zK}_\tau] : \mathbf{z} \in \mathbb{Z}_n^{k-\ell+1} \}. \quad (251)$$

As a consequence, the sum with respect to  $\mathbf{z} \in \mathbb{G}(\tau[\ell])$  in (250) can be more conveniently written as a sum with respect to the independent variables  $z_1, \dots, z_{k-\ell+1}$ , referred to in the following as the ‘‘information symbols,’’ with reference to the systematic form of  $\mathbb{G}(\tau[\ell])$ . The ‘‘parity symbols’’ corresponding to the last  $\ell-1$  components  $\mathbf{zK}_\tau$  in (251) are obtained as linear combination of the information symbols. For future use, we define the elements  $\kappa_{r,s} = [\mathbf{K}_\tau]_{r,s}$  of the matrix  $\mathbf{K}_\tau$  in (251). Therefore, the  $s$ th parity symbol of  $\mathbb{G}(\tau[\ell])$  is given by  $\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}$  (where operations are in the ring  $\mathbb{Z}_n$ ).

## APPENDIX VII FREEDOM OF $\mathbf{A}^\dagger \mathbf{A}$ AND $\mathbf{F} \mathbf{G} \mathbf{G}^\dagger \mathbf{F}^\dagger$

In this section, we recall the definition of freeness and we provide the proof of Lemma 1, which is at the heart of our main results.

*Definition 13:* The Hermitian random matrices  $\mathbf{C}$  and  $\mathbf{B}$  are asymptotically free if for all  $k$  and for all polynomials  $p_i(\cdot)$  and  $q_i(\cdot)$  with  $1 \leq i \leq k$  such that<sup>7</sup>

$$\lim_{n \rightarrow \infty} \phi(p_i(\mathbf{C})) = \lim_{n \rightarrow \infty} \phi(q_i(\mathbf{B})) = 0 \quad (252)$$

we have

$$\lim_{n \rightarrow \infty} \phi(p_1(\mathbf{C}) q_1(\mathbf{B}) \cdots p_k(\mathbf{C}) q_k(\mathbf{B})) = 0 \quad (253)$$

with  $\phi(\cdot)$  defined in (242).

<sup>7</sup>This includes polynomials with constant (zero-order) terms.

For the sake of the proof of Lemma 1 it is convenient to make the following assumptions:  $\mathbf{A}$  is a random diagonal matrix

$$\mathbf{A} = \text{diag}\{A_1, \dots, A_n\} \quad (254)$$

with i.i.d. diagonal elements  $[\mathbf{A}]_{ii} = A_i$ ;  $\mathbf{F}$  is the unitary DFT matrix defined as in (5); and  $\mathbf{G}$  is a random diagonal matrix

$$\mathbf{G} = \text{diag}\{G_1, \dots, G_n\}, \quad (255)$$

whose diagonal elements  $[\mathbf{G}]_{ii} = G_i$  are either i.i.d., or distributed according to a strong mixing process with polynomial convergence rate (see Definition 12). It is immediate to see that the role of  $\mathbf{A}$  and  $\mathbf{G}$  can be exchanged, so that the statement of Lemma 1 follows.

Then, considering the random circulant matrix

$$\Psi = \mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger = \sum_{i=1}^n |G_i|^2 \mathbf{f}_i \mathbf{f}_i^\dagger, \quad (256)$$

we wish to show that  $\mathbf{A}^\dagger \mathbf{A}$  and  $\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger$  are asymptotically free.

We define the polynomials

$$\bar{q}_j(\mathbf{A}^\dagger \mathbf{A}) = q_j(\mathbf{A}^\dagger \mathbf{A}) - \mathbb{E}[q_j(|A|^2)]\mathbf{I} \quad (257)$$

and

$$\bar{p}_j(\Psi) = p_j(\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger) - \mathbb{E}[p_j(|G|^2)]\mathbf{I}. \quad (258)$$

It follows that

$$\phi(\bar{q}_j(\mathbf{A}^\dagger \mathbf{A})) = \phi(\bar{p}_j(\Psi)) = 0. \quad (259)$$

Thus, from Definition 13, in order to prove freeness, it is sufficient to prove that, for all  $k$  and for all polynomials  $p_j(\cdot)$  and  $q_j(\cdot)$  with  $1 \leq j \leq k$ , we have

$$\lim_{n \rightarrow \infty} \phi(\bar{p}_1(\Psi) \bar{q}_1(\mathbf{A}^\dagger \mathbf{A}) \dots \bar{p}_k(\Psi) \bar{q}_k(\mathbf{A}^\dagger \mathbf{A})) = 0. \quad (260)$$

For notational convenience, from now on, we let  $\lambda_i = |G_i|^2$  and  $X_i = |A_i|^2$ . Using Theorem 16, we have

$$\begin{aligned} & \phi(\bar{p}_1(\Psi) \bar{q}_1(\mathbf{A}^\dagger \mathbf{A}) \dots \bar{p}_k(\Psi) \bar{q}_k(\mathbf{A}^\dagger \mathbf{A})) \\ &= \sum_{\ell=1}^k \sum_{\tau[\mathbf{m}, \ell]} \frac{\mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(k, \tau[\mathbf{m}, \ell])}{n^{k-\ell+1}} \\ & \times \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \mathbb{E} \left[ \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) \right]. \end{aligned} \quad (261)$$

In order to prove (260), we will show that for all  $\ell = 1, \dots, k$  and for all  $\tau[\mathbf{m}, \ell]$  either  $\mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(k, \tau[\mathbf{m}, \ell]) = 0$  or

$$\lim_{n \rightarrow \infty} \frac{1}{n^{k-\ell+1}} \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \mathbb{E} \left[ \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) \right] = 0. \quad (262)$$

In order to prove this result, we examine the structure of the linear code  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  in the systematic form (251) and consider separately the following cases.

- **Category 1.** There exists at least one information symbol that does not appear in any parity check equation, i.e., there exist a position  $1 \leq r \leq k - \ell + 1$  such that the  $r$ th row of  $\mathbf{K}_\tau$  is formed by all 0s.
- **Category 2.** There exists at least one parity symbol whose parity-check equation contains more than one information symbol, and this linear equation is unique, i.e., there exist a position  $1 \leq s \leq \ell - 1$  such that the  $s$ th column of  $\mathbf{K}_\tau$  has at least two nonzero elements, and there are no other columns of  $\mathbf{K}_\tau$  equal to the  $s$ th column.
- **Category 3.** There exists at least one information symbol that does not uniquely determine any parity symbol, i.e., there exist a position  $1 \leq r \leq k - \ell + 1$  such that for any  $1 \leq s \leq \ell - 1$  such that  $\kappa_{r,s} \neq 0$ , then there exists some  $r'$  such that also  $\kappa_{r',s} \neq 0$ .
- **Category 4.** This case includes all the cases not in Categories 1, 2, or 3. In particular, this includes the case where all the information variables uniquely determine a parity variable, i.e., for each  $1 \leq r \leq k - \ell + 1$  there exist some  $1 \leq s \leq \ell - 1$  such that  $\kappa_{r,s}$  is the only nonzero element in the  $s$ th column of  $\mathbf{K}_\tau$ .

Notice that Categories 1, 2, and 3 are not necessarily mutually exclusive. The proof of Lemma 1 follows from the following lemmas, the proof of which is given in the following subsections.

*Lemma 10:* If  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  belongs to Category 1, 2, or 3, then (262) holds.

*Lemma 11:* If  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  belongs to Category 4, then  $\mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(\kappa, \tau[\mathbf{m}, \ell]) = 0$ .

#### A. Proof of Lemma 10

From the expression of  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  in systematic form (251), we can write

$$\begin{aligned} \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) &= \sum_{\mathbf{z} \in \mathbb{Z}_n^{k-\ell+1}} \prod_{r=1}^{k-\ell+1} \bar{p}_r(\lambda_{z_r}) \\ & \times \prod_{s=1}^{\ell-1} p_{k-\ell+1+s} \left( \lambda_{\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}} \right). \end{aligned} \quad (263)$$

We examine the various cases separately.

**Category 1 (i.i.d. Case).** Without loss of generality, assume that the first row of  $\mathbf{K}_\tau$  is all zero. Then, the sum in (263) can be separated into two sum, over the domains  $\mathcal{S}$  and  $\mathcal{S}^c$  such that

$$\mathcal{S} = \left\{ (z_1, \dots, z_{k-\ell+1}) \in \mathbb{Z}_n^{k-\ell+1} : \bigcap_{r=2}^{k-\ell+1} \{z_1 \neq z_r\} \bigcap_{s=1}^{\ell-1} \left\{ z_1 \neq \sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s} \right\} \right\} \quad (264)$$

and  $\mathcal{S}^c$  is the complement of  $\mathcal{S}$  in  $\mathbb{Z}_n^{k-\ell+1}$ . We notice that  $\mathcal{S}^c$  is given as the union of linear subcodes of  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  that satisfy one additional linear equation of the type  $z_1 = z_r$  for some  $r = 2, \dots, k - \ell + 1$  or  $z_1 = \sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}$  for some  $s = 1, \dots, \ell - 1$ . There are at most  $k$  such linear subcodes of size  $n^{k-\ell}$ . Using (263) and taking expectation, we have

$$\begin{aligned} & \sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \mathbb{E} \left[ \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) \right] \\ &= \sum_{\mathbf{z} \in \mathcal{S}^c} \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} \bar{p}_r(\lambda_{z_r}) \prod_{s=1}^{\ell-1} p_{k-\ell+1+s} \left( \lambda_{\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}} \right) \right] \end{aligned} \quad (265)$$

since

$$\begin{aligned} & \sum_{\mathbf{z} \in \mathcal{S}} \mathbb{E} [\bar{p}_1(\lambda_{z_1})] \\ & \times \mathbb{E} \left[ \prod_{r=2}^{k-\ell+1} \bar{p}_r(\lambda_{z_r}) \prod_{s=1}^{\ell-1} p_{k-\ell+1+s} \left( \lambda_{\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}} \right) \right] \\ &= 0 \end{aligned}$$

due to the fact that  $\mathbb{E}[\bar{p}_r(\lambda_z)] = 0$  for any  $z \in \mathbb{Z}_n$  and  $r = 1, \dots, k$  by assumption. Dividing by  $n^{k-\ell+1}$  and noticing that the term in the right-hand side of (265) is  $O(n^{k-\ell})$  as  $n \rightarrow \infty$ , we have that (262) holds.

**Category 1 (Strong Mixing Case).** Again, without loss of generality, assume that the first row of  $\mathbf{K}_\tau$  is all zero. In the strong mixing case with polynomial rate, since independence does not hold any longer, we have to replace the notion of components “different from” with the notion of components “sufficiently far apart.” For some fixed  $n_0$ , we define the set

$$\begin{aligned} \mathcal{S}_{n_0} = & \left\{ (z_1, \dots, z_{k-\ell+1}) \in \mathbb{Z}_n^{k-\ell+1} : \right. \\ & \left. \bigcap_{r=2}^{k-\ell+1} \{|z_1 - z_r| > n_0\} \bigcap_{s=1}^{\ell-1} \left| \left| z_1 - \sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s} \right| > n_0 \right| \right\} \end{aligned} \quad (266)$$

and  $\mathcal{S}^c$  is the complement of  $\mathcal{S}$  in  $\mathbb{Z}_n^{k-\ell+1}$ . We notice that  $\mathcal{S}^c$  is given as the union of linear subcodes of  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  and a finite number  $O(n_0)$  of cosets of these. It follows that the size of  $\mathcal{S}^c$  is  $O(n_0 n^{k-\ell})$ .

For any  $\epsilon > 0$ , fix  $n_0 = n_0(\epsilon)$  such that

$$\mathbb{E} \left[ \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) \right] \leq \epsilon \mathbb{E} \left[ \prod_{r=2}^k \bar{p}_r(\lambda_{z_r}) \right] = \epsilon \mathcal{K}$$

for some finite constant  $\mathcal{K}$  and all  $\mathbf{z} \in \mathcal{S}$ , where the existence of such  $n_0(\epsilon)$  is guaranteed by Proposition 1 in Appendix IV and by the fact that  $\mathbb{E}[\bar{p}(\lambda_z)] = 0$  for all  $z \in \mathbb{Z}_n$ .

Then, using (263), separating the summation into the contribution of all  $\mathbf{z} \in \mathcal{S}_{n_0(\epsilon)}$  and the contribution of all  $\mathbf{z} \in \mathcal{S}_{n_0(\epsilon)}^c$  and taking expectation, we have (for sufficiently large  $n$ )

$$\sum_{\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \mathbb{E} \left[ \prod_{r=1}^k \bar{p}_r(\lambda_{z_r}) \right] = \epsilon O(n^{k-\ell+1}) + O(n_0(\epsilon) n^{k-\ell}). \quad (267)$$

Dividing by  $n^{k-\ell+1}$  and letting  $n \rightarrow \infty$ , we have that the term in (262) is (in absolute value) dominated by some quantity  $\delta(\epsilon)$  that is vanishing as  $\epsilon \downarrow 0$ , so that the limit in (262) holds.

**Category 2.** In this case, there exists a parity symbol that is not identically equal to an information symbol (its parity equation contains at least two nonzero coefficients) and it is not identically equal to another parity symbol (its parity equation is unique). Without loss of generality, we can assume that this symbol is  $z_k$ . Hence, for the i.i.d. case, we define the set

$$\begin{aligned} \mathcal{S} = & \left\{ (z_1, \dots, z_{k-\ell+1}) \in \mathbb{Z}_n^{k-\ell+1} : \right. \\ & \times \bigcap_{r=1}^{k-\ell+1} \left\{ z_r \neq \sum_{r=1}^{k-\ell+1} z_r \kappa_{r,k} \right\} \\ & \left. \times \bigcap_{s=1}^{\ell-2} \left\{ \sum_{r=1}^{k-\ell+1} z_r (\kappa_{r,s} - \kappa_{r,k}) \neq 0 \right\} \right\} \end{aligned}$$

and  $\mathcal{S}^c$  is the complement of  $\mathcal{S}$  in  $\mathbb{Z}_n^{k-\ell+1}$ . Again, we notice that  $\mathcal{S}^c$  is the union of linear subcodes of  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  defined by one additional linear equation, and therefore, it has size  $O(n^{k-\ell})$ . At this point, the proof for the i.i.d. case follows from the same argument used before for Category 1. The proof for the strong mixing case follows along the same lines, by replacing “different from” with the “sufficiently separated” condition in the definition of the summation sets. Details are omitted for the sake of brevity.

**Category 3.** By now, the argument that leads to the proof of Lemma 10 should be clear: we separate the sum over all  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  in (262) into two terms. One term contains codewords  $\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])$  that have one symbol distinct from all other symbols, and the other term is the complement. The first term is identically zero when taking expectation, since the term corresponding to the distinct symbol factors out of the product and it is, by construction, equal to zero. It turns out that for Categories 1 and 2 we can show that the complement set of codewords  $\mathcal{S}^c$  is formed by the union of a small (i.e., constant in  $n$ ) number of linear subcodes of size at most  $n^{k-\ell}$  that is vanishing with respect to  $n^{k-\ell+1}$  when we take the limit for  $n \rightarrow \infty$ . For the strong mixing case, term by term independence cannot be invoked. However, we can identify a subset of codewords for which one symbol is sufficiently separated by the others by more than  $n_0(\epsilon)$  modulo  $n$ , such that the expectation of the product of polynomials “almost factors out” in the sense of Proposition 1 in Appendix IV.

This proof pattern applies also for Category 3. In this case, there exists an information symbol (say,  $z_1$ ) that is not replicated into any parity symbol, i.e., there is no parity equation of the type  $z_{k-\ell+1+s} = z_1 \kappa_{1,s}$ . Therefore, in this case, the sets  $\mathcal{S}$  and  $\mathcal{S}_{n_0(\epsilon)}$  take on the same form of (264) and (266), respectively.

## B. Proof of Lemma 11

This lemma follows immediately by noticing that if  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  belongs to Category 4, then all the information symbols are replicated into some parity symbols (up to a multiplicative coefficient), that is, for every  $1 \leq r \leq k - \ell + 1$ ,

there exists some  $1 \leq s \leq \ell - 1$  such that  $z_{k-\ell+1+s} = z_{r\kappa_{r,s}}$  with  $\kappa_{r,s} \neq 0$ . This is possible only for partitions  $\tau[\mathbf{m}, \ell]$  with  $\ell - 1 \geq k - \ell + 1$ , implying  $\ell \geq \lceil \frac{k}{2} \rceil + 1$ . However, this condition implies that  $\tau[\mathbf{m}, \ell]$  must have a block of size 1. In fact, if all blocks of  $\tau[\mathbf{m}, \ell]$  were of size at least 2, we would have  $2\ell \leq k$ , which is a contradiction.

Lemma 11 follows by showing that if  $\tau[\mathbf{m}, \ell]$  has a block of size 1, then the coefficient  $\mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(k, \tau[\mathbf{m}, \ell])$  defined in (239) is identically zero. Recalling the expression

$$\begin{aligned} & \mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(k, \tau[\mathbf{m}, \ell]) \\ &= \sum_{w=\ell}^k \sum_{\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]} \left( \prod_{i=1}^w \mu_{\mathcal{V}_i}(\bar{q}_1, \dots, \bar{q}_k) \right) \\ & \quad \times \zeta(\sigma[\mathbf{v}, w] \rightarrow \tau[\mathbf{m}, \ell]) \end{aligned}$$

where

$$\mu_{\mathcal{V}_i}(\bar{q}_1, \dots, \bar{q}_k) = \mathbb{E} \left[ \prod_{r \in \mathcal{V}_i} \bar{q}_r(X_1) \right]$$

and where  $\mathcal{V}_1, \dots, \mathcal{V}_w$  are the blocks of the partition  $\sigma[\mathbf{v}, w] \leq \tau[\ell]$ , we notice that all such partitions  $\sigma[\mathbf{v}, w]$  in the above formula are refinements of  $\tau[\mathbf{m}, \ell]$ . If  $\tau[\mathbf{m}, \ell]$  has a block of size 1, then all such  $\sigma[\mathbf{v}, w]$  have also a block of size 1. Without loss of generality, let this block be  $\mathcal{V}_1 = \{r_1\}$  for some  $1 \leq r_1 \leq k$ . Therefore, for all  $\sigma[\mathbf{v}, w] \leq \tau[\mathbf{m}, \ell]$ , we have

$$\mu_{\mathcal{V}_1}(\bar{q}_1, \dots, \bar{q}_k) = \mathbb{E}[\bar{q}_{r_1}(X_1)] = 0$$

by definition of the polynomials  $\{\bar{q}_r\}$ . Since  $\mathcal{P}_{\bar{q}_1, \dots, \bar{q}_k}(k, \tau[\mathbf{m}, \ell])$  is a weighted sum of products each of which contains a zero term, it must be equal to zero.

#### APPENDIX VIII PROOF OF THE LIMIT (126)

The a.s. limit (126) is formally stated by the following.

*Lemma 12:* Let  $\mathbf{P} = \text{diag}(P_1, \dots, P_n)$  with diagonal elements either i.i.d. or distributed according to a strong mixing process with polynomial convergence rate (Definition 12). Let  $\mathbf{A} = \text{diag}(A_1, \dots, A_n)$  with i.i.d. diagonal elements, and let  $\mathbf{F}$  denote the unitary DFT matrix as defined in (5). Let  $\mathbf{Q} = \mathbf{A}\mathbf{F}$  and let  $\mathbf{q}_i$  denote the  $i$ th column of  $\mathbf{Q}$ . For any fixed  $\nu \in [0, 1]$ , as  $n \rightarrow \infty$ , we have

$$\mathbf{q}_{\lfloor \nu n \rfloor}^\dagger \left( \mathbf{I} + \gamma \sum_{j \neq \lfloor \nu n \rfloor} \mathbf{q}_j \mathbf{q}_j^\dagger P_j \right)^{-1} \mathbf{q}_{\lfloor \nu n \rfloor} \xrightarrow{\text{a.s.}} \alpha \quad (268)$$

where  $\alpha$  depends on the asymptotic distribution of  $\mathbf{A}$  and  $\mathbf{P}$  but it does not depend on  $\nu$ .

The proof of this result is based on calculations that are similar (in the spirit) but more involved than those presented before

for the proof of freeness. In the following, we outline the main ideas of the proof.

*Proof (Sketch):* Again, it is convenient to use indices in  $\mathbb{Z}_n$ . Therefore, the matrix and vector components will be numbered from 0 to  $n - 1$  instead of from 1 to  $n$ . Fix  $i = \lfloor \nu n \rfloor$ , and define

$$\begin{aligned} \mathbf{\Lambda}^{(i)} &= \text{diag}\{P_0, P_1, \dots, P_{i-1}, 1, P_{i+1}, \dots, P_{n-1}\} \\ \alpha_i &= \mathbf{q}_i^\dagger \left( \mathbf{I} + \gamma \sum_{j \neq i} \lambda_j^{(i)} \mathbf{q}_j \mathbf{q}_j^\dagger \right)^{-1} \mathbf{q}_i \end{aligned} \quad (269)$$

where  $\lambda_j^{(i)}$  denotes the  $j$ th diagonal element of  $\mathbf{\Lambda}^{(i)}$ , proving Lemma 12 is equivalent to proving that  $\beta_i \xrightarrow{\text{a.s.}} \beta$ , where

$$\begin{aligned} \beta_i &= \frac{\alpha_i}{1 + \gamma \alpha_i} \\ &= \mathbf{q}_i^\dagger \left( \mathbf{I} + \gamma \mathbf{q}_i \mathbf{q}_i^\dagger + \gamma \sum_{j \neq i} \lambda_j^{(i)} \mathbf{q}_i \mathbf{q}_i^\dagger \right)^{-1} \mathbf{q}_i \end{aligned} \quad (270)$$

where (270) follows from the matrix inversion lemma.

Using the series expansion of the matrix inverse and writing  $\mathbf{q}_i = \mathbf{A}\mathbf{f}_i$ , we obtain

$$\begin{aligned} \beta_i &= \mathbf{f}_i^\dagger \mathbf{A}^\dagger \left( \mathbf{I} + \gamma \mathbf{A} \left( \sum_{j \in \mathbb{Z}_n} \lambda_j^{(i)} \mathbf{f}_j \mathbf{f}_j^\dagger \right) \mathbf{A}^\dagger \right)^{-1} \mathbf{A}\mathbf{f}_i \\ &= \mathbf{f}_i^\dagger \left( \sum_{k=0}^{\infty} (-\gamma)^k (\mathbf{X}\Phi_i)^k \mathbf{X} \right) \mathbf{f}_i \\ &= \sum_{k=0}^{\infty} (-\gamma)^k \theta_{k,i} \end{aligned} \quad (271)$$

where we defined the  $n \times n$  diagonal matrix  $\mathbf{X} = \mathbf{A}\mathbf{A}^\dagger$  with i.i.d. diagonal elements  $X_j = |A_j|^2$ , and the  $n \times n$  circulant matrix  $\Phi_i = \mathbf{F}\mathbf{\Lambda}^{(i)}\mathbf{F}^\dagger$  and in (271), we defined

$$\theta_{k,i} = \mathbf{f}_i^\dagger (\mathbf{X}\Phi_i)^k \mathbf{X}\mathbf{f}_i. \quad (272)$$

It follows that it is sufficient for our purposes to show that the random variables  $\theta_{k,i}$ , where  $i = \lfloor \nu n \rfloor$  and  $\nu \in [0, 1]$  converges almost surely to some limit  $\varrho_k$  independent of  $\nu$ . The proof proceeds through a sequence of lemmas.

*Lemma 13:* The limit  $\varrho_k = \lim_{n \rightarrow \infty} \mathbb{E}[\theta_{k, \lfloor \nu n \rfloor}]$  exists and does not depend on  $\nu$  but only on the asymptotic distribution of  $\mathbf{A}$  and  $\mathbf{P}$ .

*Lemma 14:* The central moments of  $\theta_{k,i}$  of order 2 and 4 satisfy

$$\text{Var}\{\theta_{k,i}\} = O\left(\frac{1}{n}\right) \quad (273)$$

$$\mathbb{E}\left[|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]|^4\right] = O\left(\frac{1}{n^2}\right). \quad (274)$$

The last step of the proof of Lemma 12 follows as an application of Markov's inequality and of the Borel–Cantelli lemma. For  $\epsilon > 0$ , we have

$$\mathbb{P}(|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]| > \epsilon) \leq \frac{\mathbb{E}[|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]|^4]}{\epsilon^4} \quad (275)$$

$$= O\left(\frac{1}{n^2}\right) \quad (276)$$

where we used (274). This, combined with Lemma 13, shows that  $\theta_{k,i} \rightarrow \varrho_k$  in probability. Furthermore, since the sequence of probabilities  $\{\mathbb{P}(|\theta_{k,i} - \mathbb{E}[\theta_{k,i}]| > \epsilon) \mid n = 1, 2, \dots\}$  is summable for all  $\epsilon > 0$ , we have that  $\theta_{k,i} \rightarrow \varrho_k$  almost surely.

We conclude this section by proving Lemma 13 in details. The proof of Lemma 14 follows along the same lines but it is considerably longer. For the sake of space limitation, we omit this rather technical and tedious proof here.

We wish to compute  $\mathbb{E}[\theta_{k,i}]$  where  $\theta_{k,i}$  is defined in (272). Following similar steps as in the proof of Theorem 15, we arrive at the expression

$$\mathbb{E}[\theta_{k,i}] = \sum_{\mathbf{z} \in \mathbb{Z}_n^k} \mathbb{E}\left[\prod_{r=1}^k \lambda_{z_r}^{(i)}\right] \times \mathbb{E}\left[c_{z_1-i} c_{z_2-z_1} \cdots c_{z_k-z_{k-1}} c_{i-z_k}\right] \quad (277)$$

where, as usual,  $c_\ell$  denotes the  $\ell$ th DFT coefficient of  $(X_0, \dots, X_{n-1})$ .

Noticing that the indices of the DFT coefficients,  $z_1 - i, z_2 - z_1, \dots, i - z_k$  have zero sum, we can use Lemma 8 and algebraic manipulation similar to what done in the proof of Theorem 15 in order to arrive at the expression

$$\mathbb{E}[\theta_{k,i}] = \frac{1}{n^{k-\ell+1}} \sum_{\ell=1}^{k+1} \sum_{\tau[\mathbf{m}, \ell]} \mathcal{Q}(k+1, \tau[\mathbf{m}, \ell]) \times \sum_{(i, z_1, \dots, z_k) \in \mathbb{G}(\tau[\mathbf{m}, \ell])} \mathbb{E}\left[\prod_{r=1}^k \lambda_{z_r}^{(i)}\right]. \quad (278)$$

where  $\tau[\mathbf{m}, \ell]$  denotes an  $(\mathbf{m}, \ell)$ -partition of the index set  $\{1, \dots, k+1\}$ , where  $\mathcal{Q}(k+1, \tau[\mathbf{m}, \ell])$  is defined in (230) of Lemma 8 (see Appendix V), and where  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  is a linear code of length  $k+1$  over  $\mathbb{Z}_n$  defined by

$$\mathbb{G}(\tau[\mathbf{m}, \ell]) = \{\mathbf{z} \in \mathbb{Z}_n^{k+1} : \mathbf{z}\mathbf{B}\mathbf{A}_\tau = \mathbf{0}\} \quad (279)$$

with  $\mathbf{B}$  denoting the matrix of dimension  $(k+1) \times (k+1)$

$$\mathbf{B} = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & -1 & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}. \quad (280)$$

and  $\mathbf{A}_\tau$  denoting the incidence matrix of  $\tau[\mathbf{m}, \ell]$ , of dimension  $(k+1) \times \ell$ .

Notice that the  $(k+1)$ -tuple  $(i, z_1, \dots, z_k) \in \mathbb{G}(\tau[\mathbf{m}, \ell])$  belongs to a coset of the linear subcode  $\mathbb{G}_0(\tau[\mathbf{m}, \ell]) \subseteq \mathbb{G}(\tau[\mathbf{m}, \ell])$

formed by all codewords  $\mathbf{z} \in \mathbb{G}(\tau[\mathbf{m}, \ell])$  with first component equal to 0. In particular, since the constant vector  $(i, i, \dots, i)$  is a codeword of  $\mathbb{G}(\tau[\mathbf{m}, \ell])$  but is not a codeword of  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$ , this coset is given by

$$\mathbb{G}_0(\tau[\mathbf{m}, \ell]) + (i, i, \dots, i). \quad (281)$$

Consider the code of length  $k$  obtained by eliminating the identically zero first component of  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$ . This operation is known as “shortening” in coding theory and, with some abuse of notation, the shortened code is also denoted by  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$ . We will adopt this notation here. Notice that  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  is defined by the parity-check equation

$$\tilde{\mathbf{B}}\mathbf{A}_\tau = \mathbf{0} \quad (282)$$

where

$$\tilde{\mathbf{B}} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix} \quad (283)$$

has dimension  $k \times (k+1)$ . Furthermore, the parity-check (282) is redundant (the sum of the columns of the matrix  $\tilde{\mathbf{B}}\mathbf{A}_\tau$  is equal to zero), so that  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  has dimension  $k - \ell + 1$  over  $\mathbb{Z}_n$  (i.e., size  $n^{k-\ell+1}$ ).

We can write (278) by summing over  $\mathbf{z} \in \mathbb{G}_0(\tau[\mathbf{m}, \ell])$  and, using (281), we arrive at

$$\mathbb{E}[\theta_{k,i}] = \frac{1}{n^{k-\ell+1}} \sum_{\ell=1}^{k+1} \sum_{\tau[\mathbf{m}, \ell]} \mathcal{Q}(k+1, \tau[\mathbf{m}, \ell]) \times \sum_{\mathbf{z} \in \mathbb{G}_0(\tau[\mathbf{m}, \ell])} \mathbb{E}\left[\prod_{r=1}^k \lambda_{z_r+i}^{(i)}\right]. \quad (284)$$

Up to irrelevant component permutation, we can write  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  in systematic form as

$$\mathbb{G}_0(\tau[\mathbf{m}, \ell]) = \{\mathbf{z} | \mathbf{z}\mathbf{K}_\tau : \mathbf{z} \in \mathbb{Z}_n^{k-\ell+1}\}. \quad (285)$$

As a consequence, the sum with respect to  $\mathbf{z} \in \mathbb{G}_0(\tau[\mathbf{m}, \ell])$  in (284) can be more conveniently written as a sum with respect to the information symbols (i.e., independent variables)  $z_1, \dots, z_{k-\ell+1}$ . Without loss of generality, we choose  $\mathbf{K}_\tau$  such that its identically zero columns (if any) are placed in the last  $p$  positions, for some  $p$  (which may be equal to zero) that generally depends on the specific code  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$ . In general, a codeword of  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  in systematic form is given by

$$\left( z_1, \dots, z_{k-\ell+1}, \sum_{r=1}^{k-\ell+1} z_r k_{r,1}, \dots, \sum_{r=1}^{k-\ell+1} z_r k_{r,m}, \underbrace{0, \dots, 0}_p \right)$$

where  $m+p = \ell-1$  and where, by construction, we assume that the first  $m$  columns of  $\mathbf{K}_\tau$  have at least one nonzero element.

Recall that, by definition, we have

$$\lambda_{z+i}^{(i)} = \begin{cases} P_{z+i}, & \text{for } z \neq 0 \\ 1, & \text{for } z = 0. \end{cases} \quad (286)$$

Using this and the systematic form of  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  in (284), we obtain

$$\begin{aligned} \mathbb{E}[\theta_{k,i}] &= \sum_{\ell=1}^{k+1} \sum_{\tau[\mathbf{m}, \ell]} \mathcal{Q}(k+1, \tau[\mathbf{m}, \ell]) \frac{1}{n^{k-\ell+1}} \\ &\times \sum_{\mathbf{z} \in \mathbb{Z}_n^{k-\ell+1}} \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} \lambda_{i+z_r}^{(i)} \prod_{s=1}^m \lambda_{i+\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}}^{(i)} \right]. \end{aligned} \quad (287)$$

The proof proceeds by showing that for all partitions  $\tau[\mathbf{m}, \ell]$ , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^{k-\ell+1}} &\left| \sum_{\mathbf{z} \in \mathbb{Z}_n^{k-\ell+1}} \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} \lambda_{i+z_r}^{(i)} \prod_{s=1}^m \lambda_{i+\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}}^{(i)} \right] \right. \\ &\left. - \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} P_{i+z_r} \prod_{s=1}^m P_{i+\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}} \right] \right| \\ &= 0. \end{aligned} \quad (288)$$

In order to see this, it is useful to split the above sum over  $\mathbf{z} \in \mathbb{Z}_n^{k-\ell+1}$  into two contributions defined by the sets

$$\mathcal{S} = \left\{ \mathbf{z} \in \mathbb{Z}_n^{k-\ell+1} : \bigcap_{r=1}^{k-\ell+1} \{z_r \neq 0\} \bigcap_{s=1}^m \left\{ \sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s} \neq 0 \right\} \right\}$$

and its complement  $\mathcal{S}^c$ . The contribution of all terms  $\mathbf{z} \in \mathcal{S}$  in the sum in (288) is zero because of (286). Furthermore, notice that  $|\mathcal{S}^c| \leq (k-p)n^{k-\ell}$ , in fact,  $\mathcal{S}^c$  is given by the union of all linear subcodes of  $\mathbb{G}_0(\tau[\mathbf{m}, \ell])$  with one component in positions  $1, \dots, k-p$  fixed to zero. These subcodes have size at most  $n^{k-\ell}$ , and there are at most  $k-p$  such subcodes. From the above argument, it follows that the term in the limit (288) is upper bounded by

$$\begin{aligned} \frac{1}{n^{k-\ell+1}} \sum_{\mathbf{z} \in \mathcal{S}^c} &\left| \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} \lambda_{i+z_r}^{(i)} \prod_{s=1}^m \lambda_{i+\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}}^{(i)} \right] \right. \\ &\left. - \mathbb{E} \left[ \prod_{r=1}^{k-\ell+1} P_{i+z_r} \prod_{s=1}^m P_{i+\sum_{r=1}^{k-\ell+1} z_r \kappa_{r,s}} \right] \right| \\ &\leq \frac{n^{k-\ell}}{n^{k-\ell+1}} \mathcal{K} \end{aligned} \quad (289)$$

where  $\mathcal{K}$  denotes some positive constant independent of  $n$ , so that the limit (288) follows.

Finally, we observe that for any partition  $\tau[\mathbf{m}, \ell]$  and fixed  $\nu \in [0, 1]$  it holds

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^{k-\ell+1}} \\ \times \left| \sum_{\mathbf{z} \in \mathbb{G}_0(\tau[\mathbf{m}, \ell])} \mathbb{E} \left[ \prod_{r=1}^k P_{z_r + \lfloor \nu n \rfloor} \right] - \mathbb{E} \left[ \prod_{r=1}^k P_{z_r} \right] \right| = 0 \end{aligned} \quad (290)$$

where equality follows by the fact that  $\{P_j\}$  is stationary strong-mixing with polynomial convergence rate (details are omitted).

Limits (288) and (290) imply that the limit of (284) for  $n \rightarrow \infty$  is indeed independent of  $\nu$ , and Lemma 13 is proved.

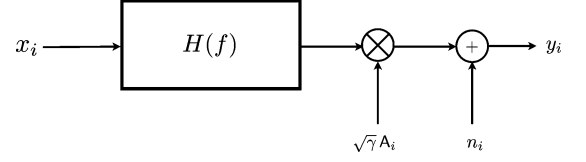


Fig. 6. Deterministic intersymbol interference with memoryless fading and Gaussian noise.

## APPENDIX IX DETERMINISTIC INTERSYMBOL INTERFERENCE WITH TIME-SELECTIVE FADING

In this Appendix, we deal with the model in Fig. 6 which incorporates both deterministic intersymbol interference and time-selective (frequency-flat) fading [17]. Through rather routine approximation of Toeplitz matrices with circulant matrices, the analysis of this model corresponds to the case where the diagonal matrix  $\mathbf{G}$  is deterministic. We conjecture that the solution is equivalent to that in Theorem 1, with  $|\mathbf{G}|^2 = S_x(U)|H(U)|^2$  where  $U$  is uniformly distributed on  $[-1/2, 1/2]$ , namely, the following.<sup>8</sup>

*Conjecture 1:* The mutual information achieved by a stationary Gaussian input with power spectral density  $S_x(f)$  is

$$\begin{aligned} I(\gamma) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(1 + \alpha\gamma S(f)) df \\ &\quad + \mathbb{E} [\log(1 + \nu\gamma|A|^2)] - \log(1 + \alpha\nu\gamma) \end{aligned} \quad (291)$$

where  $S(f) = S_x(f)|H(f)|^2$  and  $\alpha$ , and  $\nu$  are defined by the solution to

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{1 + \alpha\gamma S(f)} df = \frac{1}{1 + \alpha\nu\gamma} = \mathbb{E} \left[ \frac{1}{1 + \nu\gamma|A|^2} \right]. \quad (292)$$

The optimization of  $S_x(f)$  proceeds in the same way as in Theorem 2, i.e., it is the waterfilling solution for the same transfer function (and no time-selective fading) but computed for a reduced SNR given by Theorem 2 where  $\bar{S}_x$  is now given by

$$\bar{S}_x(\gamma, f) = \left[ \zeta_\gamma - \frac{1}{\gamma|H(f)|^2} \right]^+ \quad (293)$$

with

$$\int_{-1/2}^{1/2} \left[ \zeta_\gamma - \frac{1}{\gamma|H(f)|^2} \right]^+ df = 1. \quad (294)$$

We proceed to outline a possible path to prove Conjecture 1. The objective is to obtain an expression for  $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\log \det(\mathbf{I} + \gamma \mathbf{A} \mathbf{\Sigma} \mathbf{A}^\dagger)]$  where  $\mathbf{\Sigma} = \mathbf{H} \mathbf{\Sigma}_x \mathbf{H}^\dagger$  and  $\mathbf{H}$  is a deterministic Toeplitz channel matrix describing the linear time-invariant discrete-time linear system with transfer function  $\{H(f) : f \in [0, 1]\}$ , and  $\mathbf{\Sigma}_x$  is the Toeplitz input covariance matrix.

The first step consists of showing that the asymptotic eigenvalue distribution of  $\mathbf{A} \mathbf{\Sigma} \mathbf{A}^\dagger$  is the same as if  $\mathbf{\Sigma}$  is replaced by a circulant matrix. The sufficient condition in the following lemma is satisfied because of the conventional asymptotic

<sup>8</sup>The result in [17] is analogous to Theorem 8, but the proof (omitted in [17]) turns out to have a gap.

equivalence of products of Toeplitz matrices to circulant matrices (see [18, Th. 5.3]).

*Lemma 15:* Let  $\Sigma$  be Toeplitz with convergent eigenvalue distribution and let

$$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}. \quad (295)$$

denote the diagonal matrix of the eigenvalues of  $\Sigma$ . Further, denote the circulant matrix

$$\Psi = \mathbf{F}\Lambda\mathbf{F}^\dagger \quad (296)$$

with  $\mathbf{F}$  the unitary DFT matrix as given in (5). For all  $\gamma > 0$

$$\eta_{\mathbf{A}\Sigma\mathbf{A}}(\gamma) = \eta_{\mathbf{A}\Psi\mathbf{A}}(\gamma) \quad (297)$$

$$\mathcal{V}_{\mathbf{A}\Sigma\mathbf{A}}(\gamma) = \mathcal{V}_{\mathbf{A}\Psi\mathbf{A}}(\gamma). \quad (298)$$

The conjecture would be proved by showing that, in the limit of large  $n$

$$\eta_{\mathbf{A}\Psi\mathbf{A}^\dagger}(\gamma) = \eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\gamma) \quad (299)$$

where  $\eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}$  is given in Theorem 11 in Section III-G and where  $\{|G_i|^2\}$  is a frequency-domain fading process such that, for each  $i = 1, \dots, n$ , the term  $|G_i|^2$  is obtained by sampling independently with uniform probability a value from  $\{\lambda_1, \dots, \lambda_n\}$ .

As an intermediate step, it is not too difficult to show that, as  $n \rightarrow \infty$

$$\eta_{\mathbf{A}\tilde{\Psi}\mathbf{A}^\dagger}(\gamma) = \eta_{\mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger}(\gamma) \quad (300)$$

where  $\tilde{\Psi} = \mathbf{F}\Pi\mathbf{A}\Pi^\top\mathbf{F}^\dagger$ , and  $\Pi$  is a random permutation matrix, equiprobably distributed over the set of all permutations of  $n$  elements (i.e., over the symmetric group  $S_n$ ).

The claim (300) follows by noticing that, according to the definition of  $\eta$ -transform, it is sufficient to show that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left( \mathbf{I} + \gamma \mathbf{A}\mathbf{F}\mathbf{G}\mathbf{G}^\dagger\mathbf{F}^\dagger\mathbf{A}^\dagger \right) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n!n} \sum_{\Pi \in S_n} \mathbb{E} \left[ \text{tr} \left( \mathbf{I} + \gamma \mathbf{A}\mathbf{F}\Pi\mathbf{A}\Pi^\top\mathbf{F}^\dagger\mathbf{A}^\dagger \right) \right]. \end{aligned} \quad (301)$$

In order to show (301), we choose to follow a discrete approximation route. First, note that if we can prove the sought-after result when the empirical distribution of  $\Lambda$  only has a finite number of masses, the result will follow from continuity. Second, the possible realizations of  $\mathbf{G}$  are partitioned according to the equivalence relationship of permutation, or, in information theoretic language, into types. Because of the i.i.d. assumption, all members of each equivalence class have the same likelihood. Furthermore, the method of types, ensures that we can safely neglect all atypical types (i.e., all those that are not very similar to the empirical distribution of  $\Lambda$ .) Thus, (301) is established by including a further averaging in the right-hand side with respect to types that are close to that of  $\Lambda$ . However, that is unnecessary since the right-hand side of (301) is continuous with respect to the type of  $\Lambda$ .

At this point, (299) follows if we can show that  $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} [\text{tr}(\mathbf{I} + \gamma \mathbf{A}\mathbf{F}\Pi\mathbf{A}\Pi^\top\mathbf{F}^\dagger\mathbf{A}^\dagger)]$  (for any given

sequence of permutations  $\{\Pi\}$ ) is independent of  $\{\Pi\}$ . Let  $\mathbf{X} = \mathbf{A}\mathbf{A}^\dagger$  be a diagonal matrix whose diagonal elements  $\{X_p : p = 0, \dots, n-1\}$  are i.i.d. with a common distribution all whose moments exist. The empirical distribution of the deterministic diagonal matrix  $\Lambda$  is assumed to converge. The invariance to permutations of  $\Lambda$  of the  $\eta$ -transform follows from the invariance of all limiting  $k$ th moments, i.e.,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left( (\mathbf{X}\mathbf{F}\Lambda\mathbf{F}^\dagger)^k \right) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \text{tr} \left( (\mathbf{X}\mathbf{F}\Pi\mathbf{A}\Pi^\top\mathbf{F}^\dagger)^k \right) \right]. \end{aligned} \quad (302)$$

Fix  $k$  and let  $c_0, \dots, c_{n-1}$  denote the DFT coefficients of  $\{X_p : p = 0, \dots, n-1\}$ , as defined in (223). Denote the mean and variance of  $X_i$  by  $\mu$  and  $\sigma^2$ , respectively. A key observation is that thanks to Lemma 6 and to the expression in (246), asymptotically as  $n \rightarrow \infty$  the  $k$ th moments in (302) not depend on the distribution of  $X_i$  except through its mean and variance. Therefore, as far dealing with the limit in (302), we are free to assume that  $X_i$  is i.i.d. Gaussian. We obtain

$$\begin{aligned} & \mathbb{E} \left[ \text{tr} \left( (\mathbf{X}\mathbf{F}\Pi\mathbf{A}\Pi^\top\mathbf{F}^\dagger)^k \right) \right] \\ &= \mathbb{E} \left[ \text{tr} \left( (\mathbf{F}^\dagger\mathbf{X}\mathbf{F}\Pi\mathbf{A}\Pi^\top)^k \right) \right] \end{aligned} \quad (303)$$

$$= \mathbb{E} \left[ \text{tr} \left( ((\mathbf{C} + \mu\mathbf{I})\Pi\mathbf{A}\Pi^\top)^k \right) \right] \quad (304)$$

where we have defined the  $n \times n$  complex Gaussian circulant matrix  $\mathbf{C}$  whose first row consists of  $c_0, \dots, c_{n-1}$  where:

- $c_0 \sim \mathcal{N}(0, \sigma^2/n)$ ;
- $c_\ell \sim \mathcal{CN}(0, \sigma^2/n)$  for  $\ell = 1, \dots, \lfloor \frac{n}{2} \rfloor$ ;
- $c_{n-\ell} = c_\ell^*$ ;
- $c_{\lfloor \frac{n}{2} \rfloor + 1} \sim \mathcal{N}(0, \sigma^2/n)$  if  $n$  is odd.

Applying Newton's formula to the right-hand side of (304), all ensuing terms have the form of the expectation of the trace of a power of the product of a circulant matrix and a diagonal matrix. Then, Conjecture 1 would follow by proving that

$$\frac{1}{n} \mathbb{E} \left[ \text{tr} \left( (\mathbf{C}\Pi\mathbf{A}\Pi)^k \right) \right]$$

is invariant with respect to  $\Pi$  (at least in the limit as  $n \rightarrow \infty$ ). This is supported by extensive Monte Carlo simulation, although the proof of this invariance remains open.

Further evidence of the correctness of our conjecture is provided by the following special case, for which a direct calculation of the mutual information rate is possible using a completely different approach. Consider an ISI channel with two consecutive nonzero coefficients, denoted by  $h_0, h_1$ , and multiplicative i.i.d. time-domain "erasure" fading, i.e., such that  $A \in \{0, 1\}$  with  $\mathbb{P}[A = 0] = e \in [0, 1]$ . The corresponding time-domain channel model is given by

$$y_i = A_i \sqrt{\gamma} (h_0 x_i + h_1 x_{i-1}) + z_i, \quad i = 1, \dots, n \quad (305)$$

where we assume  $h_0, h_1 \in \mathbb{R}$  for the sake of notational simplicity. We are interested in computing the mutual information rate when the input is Gaussian i.i.d. Since the channel memory length is equal to one, every null fading components

splits the received signal into noninterfering segments. Let the two-band Toeplitz (or, equivalently, its circulant approximation, cf. Lemma 15) be denoted by  $\mathbf{H}$ . Then, after neglecting an initial transient that is irrelevant in the limit for  $n \rightarrow \infty$ , we have that

$$\frac{1}{n} \log \det (\mathbf{I} + \gamma \mathbf{A} \mathbf{H} \mathbf{H}^\dagger \mathbf{A}^\dagger) = \frac{1}{n} \sum_{k=1}^{N(n)} \log \det (\mathbf{B}_{w_k-1}) \quad (306)$$

where  $\{w_k\}$  are i.i.d. geometrically distributed intererasure times, taking values in the positive integers  $1, 2, 3, \dots$  with probability  $\mathbb{P}[w = \ell] = (1 - e)^{\ell-1} e$ ,  $N(n)$  denotes the number of runs of 1s in a sequence of length  $n$ , and we have introduced the  $\ell \times \ell$  Jacobi tridiagonal matrix

$$\mathbf{B}_\ell = \begin{bmatrix} a & b & 0 & \cdots & 0 \\ b & a & b & & \vdots \\ 0 & \ddots & \ddots & \ddots & \\ \vdots & & & & b \\ 0 & \cdots & 0 & b & a \end{bmatrix}$$

where  $a = 1 + \gamma(h_0^2 + h_1^2)$  and  $b = \gamma h_0 h_1$ . (By convention, we take  $\mathbf{B}_0 = \mathbf{1}$ .) The determinant of the block  $\mathbf{B}_\ell$  satisfies the difference equation

$$\Delta_\ell = a \Delta_{\ell-1} - b^2 \Delta_{\ell-2}$$

with initial conditions  $\Delta_0 = 1$  and  $\Delta_1 = a$ . This can be solved explicitly and yields

$$\Delta_\ell = \frac{s_1^{\ell+1} - s_2^{\ell+1}}{\sqrt{a^2 - 4b^2}} \quad (307)$$

with

$$s_{1,2} = \frac{1}{2} \left( a \pm \sqrt{a^2 - 4b^2} \right). \quad (308)$$

Eventually, we arrive at the following expression for the capacity of this channel: taking expectation with respect to the erasure fading process and the limit for  $n \rightarrow \infty$ , we find

$$C(\gamma) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \sum_{k=1}^{N(n)} \log \Delta_{w_k-1} \right] \quad (309)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ \mathbb{E} \left[ \sum_{k=1}^{N(n)} \log \Delta_{w_k-1} \middle| N(n) \right] \right] \quad (310)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[ N(n) \sum_{\ell=1}^{\infty} e(1-e)^\ell \log \left( \frac{s_1^{\ell+1} - s_2^{\ell+1}}{\sqrt{a^2 - 4b^2}} \right) \right] \quad (311)$$

$$= e^2 \sum_{\ell=1}^{\infty} (1-e)^\ell \log \left( \frac{s_1^{\ell+1} - s_2^{\ell+1}}{\sqrt{a^2 - 4b^2}} \right) \quad (312)$$

where we used the bounded convergence theorem and the fact that, by the strong law of large numbers,  $\frac{N(n)}{n} \rightarrow e$  almost surely.

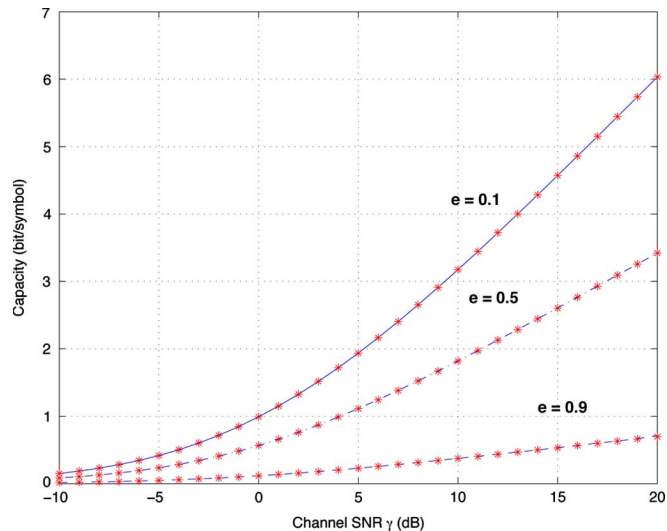


Fig. 7. Two-tap time-invariant channel ( $h_0 = 1$ ,  $h_1 = 0.5$ ), with i.i.d. erasure fading; solid lines correspond to (291) and \* correspond to (312).

Fig. 7 illustrates the comparison of (312) with the result in Conjecture 1 particularized to the case  $H(f) = h_0 + h_1 e^{-j2\pi f}$ ; perfect agreement up to any desired numerical precision is obtained.

## REFERENCES

- [1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Nov. 1998.
- [2] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, p. 38, May 1988.
- [3] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.
- [4] A. M. Tulino, S. Verdú, G. Caire, and S. Shamai, "Capacity of the Gaussian erasure channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 1721–1725.
- [5] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel—recording, statistics, and channel model," *IEEE Trans. Veh. Technol.*, vol. 40, no. 3, pp. 375–386, May 1991.
- [6] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
- [7] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
- [8] S. Shamai and S. Verdú, "The effect of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.
- [9] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–184, 2004.
- [10] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Stat.*, vol. 21, no. 1, pp. 124–127, 1950.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [12] D. Guo, S. Shamai, and S. Verdú, "Estimation of non-Gaussian random variables in Gaussian noise: Properties of the MMSE," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 6–11, 2008, pp. 1083–1087.
- [13] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input covariance for single-user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 662–671, Mar. 2006.
- [14] R. P. Stanley, *Enumerative Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 1997, vol. 1.
- [15] J. Riordan, *An Introduction to Combinatorial Analysis*. New York: Wiley, 1980.

- [16] R. A. Olshen, "Asymptotic properties of the periodogram of a discrete stationary process," *J. Appl. Probab.*, vol. 4, no. 3, pp. 508–528, Dec. 1967.
- [17] A. M. Tulino, S. Verdú, G. Caire, and S. Shamai, "Intersymbol interference with flat fading channel capacity," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1577–1581.
- [18] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.

**Antonia M. Tulino** (M'00–SM'05) received the Ph.D. degree from the Electrical Engineering Department, Seconda Università degli Studi di Napoli, Italy, in 1999.

She is currently with the Department of Wireless Communications, Bell Laboratories, Alcatel-Lucent, Holmdel, NJ. She held research positions at the Center for Wireless Communications, Oulu, Finland and at the Department of Electrical Engineering, Princeton University, Princeton, NJ. She has served on the Faculty of Engineering, Università degli Studi del Sannio, Benevento, Italy, and as Associate Professor at the Department of Electrical and Telecommunications Engineering at the Università degli Studi di Napoli "Federico II."

Dr. Tulino has received the 2009 Stephen O. Rice Prize in the Field of Communications Theory for the best paper published in the IEEE TRANSACTIONS ON COMMUNICATIONS in 2008. A frequent contributor to the IEEE TRANSACTIONS ON INFORMATION THEORY, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING, her research interests lay in the broad area of communication systems approached with the complementary tools provided by signal processing, information theory, and random matrix theory.

**Giuseppe Caire** (S'92–M'94–SM'03–F'05) was born in Torino, Italy, in 1965. He received the B.Sc. degree in electrical engineering from Politecnico di Torino, Italy, in 1990, the M.Sc. degree in electrical engineering from Princeton University, Princeton, NJ, in 1992 and the Ph.D. degree from Politecnico di Torino, in 1994.

He was a recipient of the AEI G.Someda Scholarship in 1991, has been with the European Space Agency, ESTEC, Noordwijk, The Netherlands, from May 1994 to February 1995, was a recipient of the COTRAO Scholarship in 1996 and of a CNR Scholarship in 1997. He has been visiting Princeton University in summer 1997 and Sydney University in summer 2000. He has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and he is now Professor with the Electrical Engineering Department of the Viterbi School of Engineering, University of Southern California, Los Angeles. His current interests are in the field of communications theory, information theory, and coding theory with particular focus on wireless applications.

Dr. Caire served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in 1998–2001 and as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY in 2001–2003. He received the Jack

Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004. Since November 2004, he has been member of the Board of Governors of the IEEE Information Theory Society.

**Shlomo Shamai (Shitz)** (S'82–M'85–SM'88–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1975, 1981, and 1986, respectively.

During 1975–1985, he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now the William Fondiller Professor of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI, and a corecipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 IEEE Communications Society and Information Theory Society Joint Paper Award, and the 2007 IEEE Information Theory Society Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as an Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and also has served on the Board of Governors of the Information Theory Society.

**Sergio Verdú** (S'80–M'84–SM'88–F'93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980 and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1984.

Since 1984, he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering.

Dr. Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from IEEE Communications Society. He has also received paper awards from the Japanese Telecommunications Advancement Foundation and from Eurasip. He received the 2000 Frederick E. Terman Award from the American Society for Engineering Education for his book *Multiuser Detection* (Cambridge, U.K.: Cambridge Univ. Press, 1998). He served as President of the IEEE Information Theory Society in 1997. He is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*.