Channels That Heat Up

Tobias Koch, *Member, IEEE*, Amos Lapidoth, *Fellow, IEEE*, and Paul P. Sotiriadis, *Senior Member, IEEE*

***Abstract*—This paper considers an additive noise channel where the time-k noise variance is a weighted sum of the squared magnitudes of the previous channel inputs plus a constant. This channel model accounts for the dependence of the intrinsic thermal noise on the data due to the heat dissipation associated with the transmission of data in electronic circuits: the data determine the transmitted signal, which in turn heats up the circuit and thus influences the power of the thermal noise.**

The capacity of this channel (both with and without feedback) is studied at low transmit powers and at high transmit powers. At low transmit powers, the slope of the capacity-versus-power curve at zero is computed and it is shown that the heating-up effect is beneficial. At high transmit powers, conditions are determined under which the capacity is bounded, i.e., under which the capacity does not grow to infinity as the allowed average power tends to infinity.

***Index Terms*—Capacity per unit cost, channel capacity, channels with memory, high signal-to-noise ratio (SNR), on-chip communication.**

I. Introduction

HEATING in electronics is strongly related to performance limitation, aging, and reliability issues. High performance-density and small physical size make heating important and challenging to address. This is reinforced by the trend of modern (micro-)electronics technology to pack more and faster operations within the smallest possible physical area in order to increase performance, reduce cost and size, and therefore expand the potential applications of the product and make it more profitable.

Electrical power dissipation into heat raises the local temperature of the circuit, so the temperature depends on the circuit activity. The raised temperature results in higher intrinsic noise in the circuit which in turn reduces its effective communication and computation capacity. This "negative" performance feedback is expected to become an important issue in the years to come [1]–[3].

This work aims to add this dimension to our understanding of the coupling mechanism between communication and computation performance and heating. To this end, a class of communication channels, whose noise power depends dynamically on their activity, is introduced and studied.

Manuscript received May 29, 2009; revised April 09, 2009. Current version published July 15, 2009. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Nice, France, June 2007 and at the 2007 IEEE Information Theory Workshop (ITW), Lake Tahoe, CA, September 2007.

T. Koch and A. Lapidoth are with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland.

P. P. Sotiriadis was with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD USA. He is now with Sotekco Electronics LLS, Baltimore, MD 21201 USA.

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2009.2023753

To motivate the mathematical development of this new class of channels we first discuss the underlying physical mechanism that connects circuit activity with power consumption and heating. Heating is unavoidable in electronic circuits since they convert part of the power they draw from the power supply network (and other circuits they are connected to) into heat.

Every circuit is a three-dimensional object embedded inside the substrate and the surrounding packaging material. It generates heat, in a distributed manner, that is diffused according to the *heat diffusion equation*

$$\mathsf{C}_{\mathrm{hv}} \frac{\partial T}{\partial t} = \nabla \cdot (\mathsf{k} \, \nabla T) + \dot{q}. \quad (1)$$

Here C_{hv} is the volumetric heat capacity of the material, T is the point temperature, k is the thermal conductivity, and q is the heat flux generated by the distributed conversion of electrical power into heat [4], [5]. (If other heat sources exist in the volume of the circuit, they should be included in the heat diffusion equation as well.)

In many cases, (1) can be simplified to the corresponding *ordinary differential equation* (2) providing a lumped model of the thermal dynamics

$$\mathsf{C}_{\mathrm{h}} \frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_{\mathrm{e}} - T}{\mathsf{R}_{\mathrm{th}}} + \mathsf{P}_{\mathrm{th}}. \quad (2)$$

Here C_{h} is the lumped heat capacity of the circuit (partially including the substrate and packaging), R_{th} is the thermal resistance between the circuit and the external heat-sinking environment (e.g., the air) whose temperature is T_{e}, and P_{th} is the instantaneous electrical power in the circuit that is converted into heat.

Assuming that the environmental temperature T_{e} is fixed and that $T(0) = T_e$, the solution of (2) is given by

$$T(t) = T_{\mathrm{e}} + \frac{1}{\mathsf{C}_{\mathrm{h}}} \int_0^t e^{\frac{\xi - t}{\mathsf{R}_{\mathrm{th}} \mathsf{C}_{\mathrm{h}}}} \, \mathsf{P}_{\mathrm{th}}(\xi) \mathrm{d}\xi, \quad t > 0. \quad (3)$$

Now suppose that our circuit operates according to a reference clock of period τ, i.e., it transmits an output value $x_k \in \mathbb{R}$ at the beginning of every clock period $t_k = k\tau$, $k \in \mathbb{N}$. (Here $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{N}$ denotes the set of positive integers.) Further assume that the part of the electrical energy converted into heat due to the transmission of x_k is (proportional to) $x_k^2 \tau$—a typical case in circuits when x_k is voltage or current. Then (3) can be approximated by its discrete version

$$T_k = T_{\mathrm{e}} + \frac{1}{\mathsf{C}_{\mathrm{h}}} \sum_{\ell=1}^{k-1} e^{-\frac{\tau}{\mathsf{R}_{\mathrm{th}} \mathsf{C}_{\mathrm{h}}} (k-\ell)} \, \tau x_\ell^2, \quad k \in \mathbb{N}. \quad (4)$$

By defining

$$a_\ell \triangleq \frac{\tau}{\mathsf{C}_{\mathrm{h}}} e^{-\frac{\tau}{\mathsf{R}_{\mathrm{th}} \mathsf{C}_{\mathrm{h}}} \ell}, \quad \ell \in \mathbb{N}$$

(4) becomes

$$T_k = T_{\mathrm{e}} + \sum_{\ell=1}^{k-1} a_{k-\ell}\, x_\ell^2, \quad k \in \mathbb{N}. \tag{5}$$

Equation (5) describes the relation between the local temperature of the electronic circuit and the circuit activity. Note that (5), being a general discrete-time convolution, also captures discretized versions of higher order lumped approximations of the diffusion equation (1). It therefore represents a general model of the circuit-heating process, despite the simplifying assumptions used in its derivation.

Every electronic circuit has some intrinsically generated noise, which is added to the signal and degrades its quality. In wideband circuits, the dominant type of noise is typically thermal noise [6]–[8]. Thermal noise is stationary Gaussian, and in most applications it can be considered white within the bandwidth of interest. The variance of the thermal noise θ^2 follows the Johnson–Nyquist formula

$$\theta^2 = \eta T \mathsf{W} \tag{6}$$

where W is the circuit's bandwidth, T is the absolute temperature of the circuit, and η is a proportionality constant.

Applying (5) to (6), and assuming that the intrinsic noise is only additive, yields a channel model where the variance θ^2 of the additive noise is determined by the history of the power of the transmitted signal, i.e.,

$$\theta^2(x_1, \ldots, x_{k-1}) = \sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} x_\ell^2, \quad k \in \mathbb{N} \tag{7}$$

where σ^2 and $\{\alpha_\ell\}$ are discussed in more detail in Section II (proportionality constants like η are incorporated into the parameters σ and $\{\alpha_\ell\}$).

While in today's microelectronics technology the increase in thermal noise due to data transmission is often marginal compared to the signal power and can therefore be neglected, there are scenarios where the thermal coupling of data and noise becomes significant. For example, consider a communication system where the transmission of data is assisted by a repeater, which receives the transmitted signal, amplifies it, and retransmits it. The signal at the repeater's input is typically corrupted by thermal noise, which is then amplified together with the signal. When the repeater is a monolithic circuit, the temperature of the repeater's receiving end (input)—and hence also the variance of the thermal noise—depends on the power of the signal sent out by the repeater's transmitting end (output), which in turn depends on the power of the signal sent out by the transmitter. Since the signal power at the repeater's output is much larger than that at the repeater's input, the increase in thermal noise due to retransmission of data can be significant compared to the repeater's input-signal power.

We also expect that the above channel model will be relevant to the next generation of nanoscale electronic technologies based on silicon or biological substrates [3], [9], as well as to the interface between nanocircuits and conventional microelectronics [10].

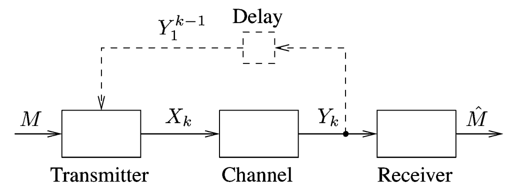

Fig. 1. A schema of the communication system.

The rest of this paper is organized as follows. Section II describes the channel model in more detail. Section III discusses channel capacity and lists some important properties thereof. Section IV presents our main results. Sections V and VI provide the proofs of these results. Section VII concludes with a summary and a discussion of our results.

II. Channel Model

We consider the communication system depicted in Fig. 1. We assume that the message M is uniformly distributed over the set $\mathcal{M} = \{1, \ldots, |\mathcal{M}|\}$ for some positive integer $|\mathcal{M}|$. The encoder maps the message to the length-n sequence $X_1, \ldots, X_n$, where n is called the *blocklength*. In the absence of feedback, the sequence X_1^n is a function of the message M, i.e., $X_1^n = \phi_n(M)$ for some mapping $\phi_n : \mathcal{M} \to \mathbb{R}^n$. Here A_m^n stands for $A_m, \ldots, A_n$. If there is a feedback link, then X_k, $k = 1, \ldots, n$ is not only a function of the message M but also of the past channel output symbols Y_1^{k-1}, i.e., $X_k = \varphi_n^{(k)}(M, Y_1^{k-1})$ for some mapping $\varphi_n^{(k)} : \mathcal{M} \times \mathbb{R}^{k-1} \to \mathbb{R}$. The receiver guesses the transmitted message M based on the n-channel output symbols Y_1^n, i.e., $\hat{M} = \psi_n(Y_1^n)$ for some mapping $\psi_n : \mathbb{R}^n \to \mathcal{M}$.

Conditional on $(X_1, \ldots, X_k) = (x_1, \ldots, x_k) \in \mathbb{R}^k$, the time-$k$ channel output $Y_k \in \mathbb{R}$ is given by

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} x_\ell^2\right)} \cdot U_k, \quad k \in \mathbb{N} \tag{8}$$

where $\{U_k\}$ is a zero-mean, unit-variance, stationary, weakly mixing random process, drawn independently of M, and being of finite fourth moment and of finite differential entropy rate, i.e.,

$$\mathsf{E}\left[U_k^4\right] < \infty \quad \text{and} \quad h\left(U_k \,\middle|\, U_{-\infty}^{k-1}\right) > -\infty. \tag{9}$$

See [11] for a definition of weak mixing. For example, $\{U_k\}$ could be a stationary ergodic Gaussian process [12] (see also [13, Sec. II]). In particular, the case of most interest is when $\{U_k\}$ are independent and identically distributed (i.i.d.), zero-mean, unit-variance, Gaussian random variables, and the reader is encouraged to focus on this case.

The parameter σ^2 is assumed to be positive. It accounts for the temperature of the device when the transmitter is silent. The coefficients α_ℓ, $\ell \in \mathbb{N}$ are nonnegative and bounded, i.e.,

$$\alpha_\ell \geq 0, \quad \ell \in \mathbb{N} \quad \text{and} \quad \sup_{\ell \in \mathbb{N}} \alpha_\ell < \infty. \tag{10}$$

They characterize the dissipation of the heat produced by transmitting message M. (It seems reasonable to assume that the sequence $\{\alpha_\ell\}$ is monotonically nonincreasing, i.e., $\alpha_\ell \geq \alpha_{\ell'}$ for

$\ell \leq \ell'$. This assumption is, however, not required for the results that are derived in this paper.)

An example of a heat dissipation profile that satisfies (10) is the *geometric* heat dissipation profile where $\{\alpha_\ell\}$ is a geometric sequence, i.e.,

$$\alpha_\ell = \rho^\ell, \quad \ell \in \mathbb{N} \tag{11}$$

for some $0 < \rho < 1$.

The heat dissipation depends *inter alia* on the efficiency of the heat sink that is employed in order to absorb the produced heat. In the above example (11), the heat sink's efficiency is described by the parameter ρ: the smaller ρ, the more efficient the heat sink. In general, an efficient heat sink is modeled by a heat dissipation profile for which the sequence $\{\alpha_\ell\}$ decays fast.

We study the above channel under an average-power constraint on the inputs, i.e., the mappings ϕ_n (without feedback) and $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ (with feedback) are chosen so that—averaged over the message M and channel outputs Y_1^n—the sequence X_1^n satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[X_k^2\right] \leq \mathsf{P}, \tag{12}$$

and we define the signal-to-noise ratio (SNR) as

$$\mathrm{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}. \tag{13}$$

Note 1: The results presented in this paper do not change when (12) is replaced by a *per-message* average-power constraint, i.e., when the mappings ϕ_n and $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ are chosen so that, for each message $m \in \mathcal{M}$ and for any given sequence of output symbols $Y_1^n = y_1^n$, the sequence x_1^n satisfies

$$\frac{1}{n}\sum_{k=1}^{n} x_k^2 \leq \mathsf{P}. \tag{14}$$

Indeed, all achievability results (which are based on schemes that ignore the feedback) are derived under (14), whereas all converse results are derived under (12). Since all mappings ϕ_n and $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ that satisfy (14) also satisfy (12), this implies that the achievability results as well as the converse results derived in this paper hold irrespective of whether constraint (12) or (14) is imposed.

The channel (8) is reminiscent of a multipath fading channel, when the transmitter and the receiver are not aware of the realization of the fading but only of its statistics (noncoherent setting). In fact, some of the techniques used this work can be extended to study the high-SNR asymptotic behavior of the capacity of such channels [14]. For more studies of noncoherent fading channels at high SNR see, e.g., [15]–[17].

III. Channel Capacity

A *rate* R (in nats per channel use) is said to be *achievable* if for every $\delta > 0$ there exist sequences of mappings $\{\phi_n\}$ (without feedback) or $\{\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}, n \in \mathbb{N}\}$ (with feedback) and $\{\psi_n\}$ such that for each $n \in \mathbb{N}$

$$\frac{\log|\mathcal{M}|}{n} > R - \delta$$

(where $\log(\cdot)$ denotes the natural logarithm function), and such that the error probability $\Pr(\hat{M} \neq M)$ tends to zero as n goes to infinity. The *capacity* is the supremum of all achievable rates. We denote by $C(\mathrm{SNR})$ the capacity under the input constraint (12) when there is no feedback, and we add the subscript "FB" to indicate that there is a feedback link. Clearly

$$C(\mathrm{SNR}) \leq C_{\mathrm{FB}}(\mathrm{SNR}) \tag{15}$$

as we can always ignore the feedback.

In the absence of feedback, the *information capacity* is defined as

$$C_{\mathrm{Info}}(\mathrm{SNR}) \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I\left(X_1^n; Y_1^n\right) \tag{16}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (12). Here we denote by $\varliminf$ the *limit inferior*; likewise, we shall denote the *limit superior* by $\varlimsup$. When there is a feedback link, we define the information capacity as

$$C_{\mathrm{Info,FB}}(\mathrm{SNR}) \triangleq \varliminf_{n\to\infty} \frac{1}{n} \sup I\left(M; Y_1^n\right) \tag{17}$$

where the supremum is over all mappings $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ satisfying (12). By Fano's inequality [18, Theorem 2.11.1] we have

$$C(\mathrm{SNR}) \leq C_{\mathrm{Info}}(\mathrm{SNR}) \tag{18}$$

and

$$C_{\mathrm{FB}}(\mathrm{SNR}) \leq C_{\mathrm{Info,FB}}(\mathrm{SNR}). \tag{19}$$

See [19] for conditions that guarantee that $C_{\mathrm{Info}}(\mathrm{SNR})$ is achievable. Note that the channel (8) is not stationary[1] since the variance of the additive noise depends on the time-index k. It is therefore *prima facie* not clear whether the inequalities in (18) and (19) hold with equality.

In this paper, we shall investigate the capacities $C(\mathrm{SNR})$ and $C_{\mathrm{FB}}(\mathrm{SNR})$ at low and high SNR. To study capacity at low SNR, we compute the *capacities per unit cost* defined as [20]

$$\dot{C}(0) \triangleq \sup_{\mathrm{SNR}>0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \tag{20}$$

and

$$\dot{C}_{\mathrm{FB}}(0) \triangleq \sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}}. \tag{21}$$

It will become apparent later that the suprema in (20) and (21) are attained when SNR tends to zero. Note that (15) implies

$$\dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0). \tag{22}$$

At high SNR, we study conditions under which the capacity is unbounded in the SNR. Notice that when the allowed transmit power is large, there is a tradeoff between optimizing the present transmission and minimizing the interference to future transmissions. Indeed, increasing the transmission power may help to overcome the present ambient noise, but it also heats up the chip and thus increases the noise variance in future receptions.

[1]By a *stationary channel* we mean a channel where for any stationary sequence of channel inputs $\{X_k\}$ and corresponding channel outputs $\{Y_k\}$, the pair $\{(X_k, Y_k)\}$ is jointly stationary.

We shall see that, as we increase the allowed transmit power, the capacity does not necessarily tend to infinity.

IV. Main Results

Our main results are presented in the following two sections. Section IV-A focuses on capacity at low SNR and presents our results on the capacity per unit cost. Section IV-B provides a sufficient condition and a necessary condition on $\{\alpha_\ell\}$ under which the capacity is bounded in the SNR.

A. Capacity Per Unit Cost

The results presented in this subsection hold under the additional assumptions that $\{U_k\}$ is i.i.d. and that

$$\sum_{\ell=1}^{\infty} \alpha_\ell < \infty. \tag{23}$$

To shorten notation, we denote this sum by α, i.e.,

$$\alpha \triangleq \sum_{\ell=1}^{\infty} \alpha_\ell. \tag{24}$$

Proposition 1: Consider the above channel model, and assume additionally that the sequence $\{\alpha_\ell\}$ satisfies (23) and that $\{U_k\}$ is i.i.d. Then

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\mathrm{SNR}>0} \frac{C_{\alpha=0}(\mathrm{SNR})}{\mathrm{SNR}} \tag{25}$$

where $C_{\alpha=0}(\mathrm{SNR})$ denotes the capacity of the channel

$$Y_k = x_k + \sigma\, U_k$$

which is a special case of (8) for $\alpha = 0$.

Proof: See Appendix A. □

For $\alpha = 0$, (8) describes a channel with an ideal heat sink or, equivalently, a channel that does not heat up. Proposition 1 thus demonstrates that the heating up can only increase the information capacity per unit cost. In other words, at low SNR, the heating-up effect is not harmful.

For *Gaussian* noise, i.e., when $\{U_k\}$ is a sequence of i.i.d., zero-mean, unit-variance, *Gaussian* random variables, the heating-up effect is beneficial.

Theorem 2: Consider the above channel model. Assume additionally that the sequence $\{\alpha_\ell\}$ satisfies (23) and that $\{U_k\}$ is a sequence of i.i.d., zero-mean, unit-variance, Gaussian random variables. Then, irrespective of whether feedback is available or not

$$\dot{C}_{\mathrm{FB}}(0) = \dot{C}(0) = \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} = \frac{1}{2}\left(1 + \sum_{\ell=1}^{\infty} \alpha_\ell\right). \tag{26}$$

Proof: See Section V. □

For example, for the geometric heat dissipation profile (11) we obtain from Theorem 2

$$\dot{C}_{\mathrm{FB}}(0) = \dot{C}(0) = \frac{1}{2}\frac{1}{1-\rho}, \quad 0 < \rho < 1. \tag{27}$$

Thus the capacity per unit cost is monotonically *decreasing* in ρ.

The above result might be counterintuitive, because it suggests not to use heat sinks at low SNR. Nevertheless, it can be heuristically explained by noting that the heating-up effect increases the *channel gain*.[2] Indeed, if we split up the channel output

$$Y_k = X_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} X_\ell^2\right)} \cdot U_k$$

into a data-dependent part

$$\tilde{X}_k = X_k + \sqrt{\left(\sum_{\ell=1}^{k-1} \alpha_{k-\ell} X_\ell^2\right)} \cdot U_k$$

and a data-independent part Z_k (with $\{Z_k\}$ being a sequence of i.i.d., zero-mean, variance-σ^2, Gaussian random variables drawn independently of $\{(U_k, X_k)\}$), then the channel gain G for (8) is given by

$$\mathsf{G} \triangleq \lim_{n\to\infty} \sup \frac{\sum_{k=1}^{n} \mathsf{E}\left[\tilde{X}_k^2\right]}{\sum_{k=1}^{n} \mathsf{E}\left[X_k^2\right]} = 1 + \sum_{\ell=1}^{\infty} \alpha_\ell \tag{28}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (12). Thus, in view of (28), Theorem 2 demonstrates that the capacity per unit cost is determined by the channel gain G. This result is not specific to (8) but has also been observed for other channel models. For example, the same is true for fading channels whenever the additive noise is Gaussian [21], [22].

B. Conditions for Bounded Capacity

While at low SNR the heating-up effect is beneficial, at high SNR it is detrimental. In fact, it turns out that the capacity can be even bounded in the SNR, i.e., the capacity does not tend to infinity as the SNR tends to infinity. The following theorem provides a sufficient condition and a necessary condition on $\{\alpha_\ell\}$ for the capacity to be bounded. Note that the results presented in this section do not require the additional assumptions made in Section IV-A: we neither assume that the sequence $\{\alpha_\ell\}$ satisfies (23) nor that $\{U_k\}$ is i.i.d.

Theorem 3: Consider the channel model described in Section II. Then

$$\text{(i)} \quad \left(\underline{\lim}_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0\right) \Longrightarrow \left(\sup_{\mathrm{SNR}>0} C_{\mathrm{FB}}(\mathrm{SNR}) < \infty\right) \tag{29}$$

$$\text{(ii)} \quad \left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0\right) \Longrightarrow \left(\sup_{\mathrm{SNR}>0} C(\mathrm{SNR}) = \infty\right) \tag{30}$$

where we define $a/0 \triangleq \infty$ for every $a > 0$ and $0/0 \triangleq 0$.

Proof: See Section VI. □

For example, for the geometric heat dissipation (11) we have

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \rho, \quad 0 < \rho < 1$$

[2] The channel gain is given by the ratio of the "desired" power at the channel output to the "desired" power at the channel input.

and it follows from Theorem 3 that the capacity is bounded. On the other hand, for a supergeometric heat dissipation, i.e., when

$$\alpha_\ell = \rho^{\ell^\kappa}, \quad \ell \in \mathbb{N}$$

for some $0 < \rho < 1$ and $\kappa > 1$, we obtain

$$\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = \lim_{\ell\to\infty} \rho^{(\ell+1)^\kappa - \ell^\kappa} = 0$$

and Theorem 3 implies that the capacity is unbounded. Roughly speaking, we can say that when the sequence of coefficients $\{\alpha_\ell\}$ decays *not faster than geometrically*, the capacity is *bounded* in the SNR, and when the sequence of coefficients $\{\alpha_\ell\}$ decays *faster than geometrically*, the capacity is *unbounded* in the SNR.

Note 2: For Part (i) of Theorem 3, the assumptions that the process $\{U_k\}$ is weakly mixing and that it has a finite fourth moment are not needed. These assumptions are only needed for Lemma 5 in the proof of Part (ii). In Part (ii) of Theorem 3, the condition on the left-hand side (LHS) of (30) can be replaced by

$$\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = \infty. \tag{31}$$

This condition (31) is weaker than the original condition (30) because

$$\left(\lim_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0\right) \Longrightarrow \left(\lim_{\ell\to\infty} \frac{1}{\ell} \log \frac{1}{\alpha_\ell} = \infty\right).$$

If neither the LHS of (29) nor the LHS of (30) holds, i.e.,

$$\varlimsup_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \quad \text{and} \quad \varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \tag{32}$$

then the capacity can be bounded or unbounded. Example 1 gives a sequence $\{\alpha_\ell\}$ satisfying (32) for which the capacity is bounded, and Example 2 provides a sequence $\{\alpha_\ell\}$ satisfying (32) for which the capacity is unbounded. (These sequences $\{\alpha_\ell\}$ are not monotonically decreasing in ℓ. Consequently, Examples 1 and 2 are rather of mathematical than of practical interest. Nevertheless, they show that when neither condition of Theorem 3 is satisfied, one can construct simple examples yielding a bounded capacity or an unbounded capacity, thus demonstrating the difficulty of finding conditions that are necessary *and* sufficient for the capacity to be bounded.)

Example 1: Consider the sequence $\{\alpha_\ell\}$ where all coefficients with an even index are equal to 1, and where all coefficients with an odd index are 0. It satisfies (32) because $\varlimsup_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell = \infty$ and $\varliminf_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell = 0$. Then the time-$k$ channel output Y_k corresponding to the channel inputs $(x_1, \ldots, x_k)$ is given by

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{\lfloor (k-1)/2 \rfloor} x_{k-2\ell}^2\right)} \cdot U_k, \quad k \in \mathbb{N}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Thus, at even times the output Y_{2k}, $k \in \mathbb{N}$ only depends on the "even" inputs $(X_2, X_4, \ldots, X_{2k})$, while at odd times the output Y_{2k+1}, $k \in \mathbb{N}_0$ only depends on the "odd" inputs $(X_1, X_3, \ldots, X_{2k+1})$. (Here $\mathbb{N}_0$ denotes the set of nonnegative integers.) By proceeding along the lines of the proof of Part (i) of Theorem 3 while choosing in (62) $\beta = 1/y_{k-2}^2$, it can be shown that the capacity of this channel is bounded. (Intuitively, the channel can be viewed as consisting of two parallel channels, one connecting the inputs and outputs at even times, and the other connecting the inputs and outputs at odd times. By Theorem 3, the capacity of both parallel channels is bounded, and it is therefore plausible that the capacity of the original channel is bounded as well.)

Example 2: Consider the sequence $\{\alpha_\ell\}$ where all coefficients with an even positive index are 0, and where all other coefficients are 1. (Again, we have $\varlimsup_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell = \infty$ and $\varliminf_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell = 0$.) In this case the time-k channel output Y_k corresponding to $(x_1, \ldots, x_k)$ is given by

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{\lfloor k/2 \rfloor} x_{k-2\ell+1}^2\right)} \cdot U_k, \quad k \in \mathbb{N}.$$

Using Gaussian inputs of power 2P at even times while setting the inputs to be zero at odd times, and measuring the channel outputs only at even times, reduces the channel to a memoryless additive noise channel and demonstrates (using the result of [23]) the achievability of

$$R = \frac{1}{4} \log(1 + 2\,\mathrm{SNR})$$

which is unbounded in the SNR.

The two seemingly similar examples thus lead to completely different capacity results. The crucial difference between Example 1 and Example 2 is that in the former example, at even times the interference is caused by the past channel inputs at *even* times, whereas in the latter example, at even times the interference is caused by the past channel inputs at *odd* times. Thus, in Example 2, setting all "odd" inputs to zero cancels (at even times) the interference from past channel inputs and hence transforms the channel into an additive noise channel whose capacity is unbounded. Evidently, this approach does not work for Example 1.

V. Proof of Theorem 2

In Section V-A, we derive an upper bound on the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$, and in Section V-B we derive a lower bound on the capacity $C(\mathrm{SNR})$ in the absence of feedback. These bounds are used in Section V-C to derive an upper bound on $\dot{C}_{\mathrm{FB}}(0)$ and a lower bound on $\dot{C}(0)$, which are then both shown to be equal to $1/2(1+\alpha)$. Together with (22) this proves Theorem 2.

A. Converse

The upper bound on $C_{\mathrm{FB}}(\mathrm{SNR})$ is based on (19) and on an upper bound on $\frac{1}{n} I(M; Y_1^n)$. We have

$$\frac{1}{n} I\left(M; Y_1^n\right) = \frac{1}{n} \sum_{k=1}^{n} \left(h\left(Y_k \middle| Y_1^{k-1}\right) - h\left(Y_k \middle| Y_1^{k-1}, M\right) \right)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\Big(h\left(Y_k \left| Y_1^{k-1}\right.\right) - h\left(Y_k \left| Y_1^{k-1}, M, X_1^k\right.\right)\Big)$$
$$= \frac{1}{n}\sum_{k=1}^{n}\left(h\left(Y_k \left| Y_1^{k-1}\right.\right) - h(U_k) - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]\right) \quad (33)$$

where the first step follows from the chain rule for mutual information [18, Theorem 2.5.2]; the second step follows because X_1^k is a function of M and Y_1^{k-1}; and the last step follows from the behavior of differential entropy under translation and scaling [18, Theorems 9.6.3 and 9.6.4], and because U_k is independent of (Y_1^{k-1}, M, X_1^k).

Evaluating the differential entropy $h(U_k)$ of a Gaussian random variable, and using the trivial lower bound

$$\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2)\right)\right] \geq \log\sigma^2$$

we obtain the final upper bound

$$\begin{aligned}\frac{1}{n}I(M;Y_1^n) &\leq \frac{1}{n}\sum_{k=1}^{n}\left(h\left(Y_k \left| Y_1^{k-1}\right.\right) - \frac{1}{2}\log\left(2\pi e\sigma^2\right)\right)\\
&\leq \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\mathsf{E}\left[X_\ell^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log\left(1 + \frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\alpha_{k-\ell}\mathsf{E}\left[X_\ell^2\right]/\sigma^2\right)\\
&= \frac{1}{2}\log\left(1 + \frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\sum_{\ell=0}^{n-k}\alpha_\ell\right)\\
&\leq \frac{1}{2}\log\left(1 + (1+\alpha)\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\right)\\
&\leq \frac{1}{2}\log(1 + (1+\alpha)\,\mathrm{SNR}),\end{aligned} \quad (34)$$

where we define $\alpha_0 \triangleq 1$. Here, the second step follows because conditioning cannot increase entropy and from the entropy maximizing property of Gaussian random variables [18, Theorem 9.6.5]; the third step follows from Jensen's inequality; the fourth step by rewriting the double sum; the fifth step follows because the coefficients are nonnegative which implies that

$$\sum_{\ell=0}^{n-k}\alpha_\ell \leq \sum_{\ell=0}^{\infty}\alpha_\ell = 1 + \alpha;$$

and the last step follows from the power constraint (12).

B. Direct Part

As mentioned earlier, the above channel (8) is not stationary, and it is therefore *prima facie* not clear whether $C_{\text{Info}}(\mathrm{SNR})$ is achievable. We shall sidestep this problem by studying the capacity of a different channel whose time-k channel output $\tilde{Y}_k \in \mathbb{R}$ is, conditional on the sequence $\{X_k\} = \{x_k\}$, given by

$$\tilde{Y}_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=-\infty}^{k-1}\alpha_{k-\ell}x_\ell^2\right)}\cdot U_k, \quad k \in \mathbb{N} \quad (35)$$

where $\{U_k\}$ and $\{\alpha_\ell\}$ are defined in Section II. This channel has the advantage that it is stationary and ergodic in the sense that when $\{X_k\}$ is a stationary ergodic process, the pair $\{(X_k, \tilde{Y}_k)\}$ is jointly stationary ergodic. It follows that if the sequences $\{X_k, k = 0, -1, \ldots\}$ and $\{X_k, k = 1, 2, \ldots\}$ are independent of each other, and if the random variables X_k, $k = 0, -1, \ldots$ are bounded, then any rate that can be achieved over this new channel is also achievable over the original channel. Indeed, the original channel (8) can be converted into (35) by adding

$$S_k = \sqrt{\left(\sum_{\ell=-\infty}^{0}\alpha_{k-\ell}X_\ell^2\right)}\cdot V_k, \quad k \in \mathbb{N}$$

to the channel output Y_k (where $\{V_k\}$ is a sequence of i.i.d., zero-mean, unit-variance, Gaussian random variables drawn independently of $\{(U_k, X_k)\}$),[3] and since the independence of $\{X_k, k = 0, -1, \ldots\}$ and $\{X_k, k = 1, 2, \ldots\}$ ensures that the sequence $\{S_k, k \in \mathbb{N}\}$ is independent of the message M, it follows that any rate achievable over (35) can be achieved over (8) by using a receiver that generates $\{S_k, k \in \mathbb{N}\}$ and then guesses M based on $(Y_1 + S_1, \ldots, Y_n + S_n)$.[4]

We shall consider channel inputs $\{X_k\}$ that are blockwise i.i.d. in blocks of L symbols (for some $L \in \mathbb{N}$). Thus, denoting $\boldsymbol{X}_b = (X_{bL+1}, \ldots, X_{(b+1)L})^{\mathsf{T}}$ (where $(\cdot)^{\mathsf{T}}$ denotes the transpose), $\{\boldsymbol{X}_b\}$ is a sequence of i.i.d. random length-L vectors with $\boldsymbol{X}_b$ taking on the value $(\xi, 0, \ldots, 0)^{\mathsf{T}}$ with probability δ and $(0, \ldots, 0)^{\mathsf{T}}$ with probability $1-\delta$, for some $\xi \in \mathbb{R}$. Note that to satisfy the average-power constraint (12) we shall choose ξ and δ so that

$$\frac{\xi^2}{\sigma^2}\delta = L\,\mathrm{SNR}. \quad (36)$$

Let $\tilde{\boldsymbol{Y}}_b = (\tilde{Y}_{bL+1}, \ldots, \tilde{Y}_{(b+1)L})^{\mathsf{T}}$. Noting that the pair $\{(\boldsymbol{X}_b, \tilde{\boldsymbol{Y}}_b)\}$ is jointly stationary ergodic, it follows from [19] that the rate

$$\lim_{n\to\infty}\frac{1}{n}I\left(\boldsymbol{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\right)$$

is achievable over the new channel (35) and thus yields a lower bound on the capacity $C(\mathrm{SNR})$ of the original channel (8). We have

$$\frac{1}{n}I\left(\boldsymbol{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\right) = \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1}I\left(\boldsymbol{X}_b; \tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\left|\boldsymbol{X}_0^{b-1}\right.\right)$$

[3]The boundedness of the random variables X_k, $k = 0, -1, \ldots$ guarantees that $\sum_{\ell=-\infty}^{0}\alpha_{k-\ell}x_\ell^2$ is finite for any realization of $\{X_k, k = 0, -1, \ldots\}$.

[4]This approach is specific to the case where $\{U_k\}$ is a Gaussian process. Indeed, it relies heavily on the fact that, given $\{X_k\} = \{x_k\}$, the additive noise term on the right-hand side (RHS) of (35) can be written as the sum of two independent random variables, where one random variable depends only on $\{X_k, k = 0, -1, \ldots\}$, and where the other random variable depends only on $\{X_k, k = 1, 2, \ldots\}$. This certainly holds for Gaussian random variables, but it does not necessarily hold for other distributions on $\{U_k\}$.

$$\begin{aligned}&\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(\boldsymbol{X}_b;\tilde{\boldsymbol{Y}}_b \middle| \boldsymbol{X}_0^{b-1}\right)\\ &\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(\boldsymbol{X}_b;\tilde{\boldsymbol{Y}}_b \middle| \boldsymbol{X}_{-\infty}^{b-1}\right)\\ &\quad -\frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b \middle| \boldsymbol{X}_0^{b}\right) \end{aligned}\tag{37}$$

where we use the chain rule and the nonnegativity of mutual information. It is shown in Appendix B that

$$\lim_{b\to\infty} I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b \middle| \boldsymbol{X}_0^{b}\right)=0.\tag{38}$$

This together with a Cesáro-type theorem [18, Theorem 4.2.3] yields

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n} I\left(\boldsymbol{X}_0^{\lfloor n/L\rfloor-1};\tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\right)\\ &\quad\geq \frac{1}{L} I\left(\boldsymbol{X}_0;\tilde{\boldsymbol{Y}}_0 \middle| \boldsymbol{X}_{-\infty}^{-1}\right)\\ &\quad\quad -\frac{1}{L}\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b \middle| \boldsymbol{X}_0^{b}\right)\\ &\quad=\frac{1}{L} I\left(\boldsymbol{X}_0;\tilde{\boldsymbol{Y}}_0 \middle| \boldsymbol{X}_{-\infty}^{-1}\right) \end{aligned}\tag{39}$$

where the first step follows by the stationarity of $\{(\boldsymbol{X}_b,\tilde{\boldsymbol{Y}}_b)\}$, which implies that $I(\boldsymbol{X}_b;\tilde{\boldsymbol{Y}}_b \,|\, \boldsymbol{X}_{-\infty}^{b-1})$ does not depend on b, and by noting that $\lim_{n\to\infty}\frac{\lfloor n/L\rfloor}{n}=1/L$.

We proceed to analyze $I(\boldsymbol{X}_0;\tilde{\boldsymbol{Y}}_0 \,|\, \boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1})$ for a given sequence $\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}$. Making use of the canonical decomposition of mutual information (e.g., [20, eq. (10)]), we have

$$\begin{aligned}&I\left(\boldsymbol{X}_0;\tilde{\boldsymbol{Y}}_0 \middle| \boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}\right)\\ &\quad= I\left(X_1;\tilde{\boldsymbol{Y}}_0 \middle| \boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}\right)\\ &\quad=\int D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=x,\boldsymbol{x}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right)\mathrm{d}P_{X_1}(x)\\ &\quad\quad - D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right)\\ &\quad=\delta D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=\xi,\boldsymbol{x}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right)\\ &\quad\quad - D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right) \end{aligned}\tag{40}$$

where the first step follows because, for our choice of input distribution, $X_2=\cdots=X_L=0$ and hence X_1 conveys as much information about $\tilde{\boldsymbol{Y}}_0$ as $\boldsymbol{X}_0$. Here $D(\cdot\|\cdot)$ denotes relative entropy, i.e.,

$$D(P_1\,\|\,P_0)=\begin{cases}\displaystyle\int \log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}\mathrm{d}P_1, & \text{if } P_1\ll P_0\\ +\infty, & \text{otherwise}\end{cases}$$

P_{X_1} denotes the distribution of X_1; and $P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=\xi,\boldsymbol{x}_{-\infty}^{-1}}$, $P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}$, and $P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}$ denote the distributions of $\tilde{\boldsymbol{Y}}_0$ conditional on the inputs $(X_1=\xi,\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1})$, $(X_1=0,\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1})$, and $\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}$, respectively. Thus, $P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=\xi,\boldsymbol{x}_{-\infty}^{-1}}$ is the law of an L-variate Gaussian random vector of mean $(\xi,0,\ldots,0)^{\mathsf{T}}$ and of diagonal covariance matrix $\mathsf{K}^{(\xi)}_{\boldsymbol{x}_{-\infty}^{-1}}$ with diagonal entries

$$\mathsf{K}^{(\xi)}_{\boldsymbol{x}_{-\infty}^{-1}}(1,1)=\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L}x^2_{\ell L+1}$$

$$\mathsf{K}^{(\xi)}_{\boldsymbol{x}_{-\infty}^{-1}}(i,i)=\sigma^2+\alpha_{i-1}\xi^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}x^2_{\ell L+1},\qquad i=2,\ldots,L;$$

$P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}$ is the law of an L-variate, zero-mean, Gaussian random vector of diagonal covariance matrix $\mathsf{K}^{(0)}_{\boldsymbol{x}_{-\infty}^{-1}}$ with diagonal entries

$$\mathsf{K}^{(0)}_{\boldsymbol{x}_{-\infty}^{-1}}(i,i)=\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}x^2_{\ell L+1},\quad i=1,\ldots,L;$$

and $P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}$ is given by

$$P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}=\delta P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=\xi,\boldsymbol{x}_{-\infty}^{-1}}+(1-\delta)P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}.$$

In order to evaluate the first term on the RHS of (40), we note that the relative entropy of two real, L-variate Gaussian random vectors of means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and of covariance matrices K_1 and K_2 is given by

$$\begin{aligned}&D\Big(\mathcal{N}(\boldsymbol{\mu}_1\mathsf{K}_1)\,\|\,\mathcal{N}(\boldsymbol{\mu}_2\mathsf{K}_2)\Big)\\ &\quad=\frac{1}{2}\log\det\mathsf{K}_2-\frac{1}{2}\log\det\mathsf{K}_1+\frac{1}{2}\mathrm{tr}\left(\mathsf{K}_1\mathsf{K}_2^{-1}-\mathsf{I}_L\right)\\ &\quad\quad+\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\mathsf{K}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\end{aligned}\tag{41}$$

where $\det(\cdot)$ and $\mathrm{tr}(\cdot)$ denote the determinant and the trace, and where I_L denotes the $L\times L$ identity matrix.

Let $\mathsf{E}[D(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{X}_{-\infty}^{-1}}\,\|\,P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{X}_{-\infty}^{-1}})]$ denote the second term on the RHS of (40) averaged over $\boldsymbol{X}_{-\infty}^{-1}$, i.e.,

$$\begin{aligned}&\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{X}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{X}_{-\infty}^{-1}}\right)\right]\\ &\quad=\mathsf{E}_{\boldsymbol{X}_{-\infty}^{-1}}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right)\right].\end{aligned}$$

Then using (41) and (40) and taking expectations over $\boldsymbol{X}_{-\infty}^{-1}$, we obtain, again defining $\alpha_0\triangleq 1$,

$$\begin{aligned}&\frac{1}{L}I\left(\boldsymbol{X}_0;\tilde{\boldsymbol{Y}}_0\middle|\boldsymbol{X}_{-\infty}^{-1}\right)\\ &\quad=\frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X^2_{\ell L+1}/\sigma^2}\right]\\ &\quad\quad-\frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\mathsf{E}\left[\log\left(1+\frac{\alpha_{i-1}\xi^2}{\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X^2_{\ell L+1}}\right)\right]\\ &\quad\quad-\frac{1}{L}\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{X}_{-\infty}^{-1}}\middle\| P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{X}_{-\infty}^{-1}}\right)\right]\\ &\quad\geq\frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}\mathsf{E}\left[X^2_{\ell L+1}\right]/\sigma^2}\\ &\quad\quad-\frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{L}\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{X}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{X}_{-\infty}^{-1}}\right)\right] \\
&\geq \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}} \\
&\quad - \frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\
&\quad - \frac{1}{L}\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{X}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{X}_{-\infty}^{-1}}\right)\right] \qquad (42)
\end{aligned}
$$

where the second step follows from the lower bound $\mathsf{E}[1/(1+X)] \geq 1/(1+\mathsf{E}[X])$ (which follows by applying Jensen's inequality to the convex function $f(x)=1/(1+x)$, $x>0$), and from the upper bound

$$
\mathsf{E}\left[\log\left(1+\frac{\alpha_{i-1}\xi^2}{\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2}\right)\right] \leq \log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right), \quad i=2,\ldots,L;
$$

and the third step follows from (36) and by upper-bounding $\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1} \leq \alpha$. The final lower bound follows now by (42) and (39)

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}I\left(\boldsymbol{X}_0^{\lfloor n/L\rfloor-1};\tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\right) \\
&\quad\geq \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}} \\
&\qquad - \frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\
&\qquad - \frac{1}{L}\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{X}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{X}_{-\infty}^{-1}}\right)\right] \qquad (43)
\end{aligned}
$$

and by recalling that

$$
C(\mathrm{SNR}) \geq \lim_{n\to\infty}\frac{1}{n}I\left(\boldsymbol{X}_0^{\lfloor n/L\rfloor-1};\tilde{\boldsymbol{Y}}_0^{\lfloor n/L\rfloor-1}\right). \qquad (44)
$$

C. Asymptotic Analysis

We start with analyzing the upper bound (34). Using that $\log(1+x) \leq x$ for $x>-1$ we have

$$
\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{\frac{1}{2}\log\left(1+(1+\alpha)\,\mathrm{SNR}\right)}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha) \qquad (45)
$$

and we thus obtain

$$
\dot{C}_{\mathrm{FB}}(0) \triangleq \sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha). \qquad (46)
$$

In order to derive a lower bound on $\dot{C}(0)$ we note that

$$
\dot{C}(0) \triangleq \sup_{\mathrm{SNR}>0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \geq \lim_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \qquad (47)
$$

and proceed by analyzing the limiting ratio of the lower bound (43) to SNR as SNR tends to zero. To this end, we first show that

$$
\lim_{\mathrm{SNR}\downarrow 0}\frac{\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{X}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{X}_{-\infty}^{-1}}\right)\right]}{\mathrm{SNR}} = 0. \qquad (48)
$$

We recall that for any pair of distributions P_0 and P_1 satisfying $P_1 \ll P_0$ [20, p. 1023]

$$
\lim_{\beta\downarrow 0}\frac{D\left(\beta P_1+(1-\beta)P_0 \| P_0\right)}{\beta} = 0. \qquad (49)
$$

Thus, for any $\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}$, (49) together with $\delta=\mathrm{SNR}L\sigma^2/\xi^2$ implies that

$$
\lim_{\mathrm{SNR}\downarrow 0}\frac{D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{x}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{x}_{-\infty}^{-1}}\right)}{\mathrm{SNR}} = 0. \qquad (50)
$$

In order to show that this also holds when

$$
D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{x}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{x}_{-\infty}^{-1}}\right)
$$

is averaged over $\boldsymbol{X}_{-\infty}^{-1}$, we derive in the following the uniform upper bound

$$
\begin{aligned}
&\sup_{\boldsymbol{x}_{-\infty}^{-1}} D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{x}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{x}_{-\infty}^{-1}}\right) \\
&\qquad = D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{x}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{x}_{-\infty}^{-1}}\right)\Big|_{\boldsymbol{x}_{-\infty}^{-1}=0}. \qquad (51)
\end{aligned}
$$

The claim (48) follows then by upper-bounding

$$
\mathsf{E}\left[D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{X}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{X}_{-\infty}^{-1}}\right)\right] \leq D\left(P_{\tilde{\boldsymbol{Y}}_0 \mid \boldsymbol{x}_{-\infty}^{-1}} \middle\| P_{\tilde{\boldsymbol{Y}}_0 \mid X_1=0, \boldsymbol{x}_{-\infty}^{-1}}\right)\Big|_{\boldsymbol{x}_{-\infty}^{-1}=0}
$$

and by (50).

To prove (51) we use that every Gaussian random vector can be expressed as the sum of two independent Gaussian random vectors to write the channel output $\tilde{\boldsymbol{Y}}_0$ as

$$
\tilde{\boldsymbol{Y}}_0 = \boldsymbol{X}_0 + \boldsymbol{V} + \boldsymbol{W} \qquad (52)
$$

where, conditional on $\boldsymbol{X}_{-\infty}^{0}=\boldsymbol{x}_{-\infty}^{0}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ are L-variate, zero-mean Gaussian random vectors, drawn independently of each other and having the respective diagonal covariance matrices $\mathsf{K}_{\boldsymbol{V}\mid\boldsymbol{x}_0}$ and $\mathsf{K}_{\boldsymbol{W}\mid\boldsymbol{x}_{-\infty}^{-1}}$ whose diagonal entries are given by

$$
\begin{aligned}
\mathsf{K}_{\boldsymbol{V}\mid\boldsymbol{x}_0}(1,1) &= \sigma^2 \\
\mathsf{K}_{\boldsymbol{V}\mid\boldsymbol{x}_0}(i,i) &= \sigma^2+\alpha_{i-1}x_1^2, \quad i=2,\ldots,L
\end{aligned}
$$

and

$$
\mathsf{K}_{\boldsymbol{W}\mid\boldsymbol{x}_{-\infty}^{-1}}(i,i) = \sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}x_{\ell L+1}^2, \quad i=1,\ldots,L.
$$

Thus, $\boldsymbol{W}$ is the portion of the noise due to $\boldsymbol{X}_{-\infty}^{-1}$, and $\boldsymbol{V}$ is the portion of the noise that remains after subtracting $\boldsymbol{W}$. Note that $\boldsymbol{X}_0+\boldsymbol{V}$ and $\boldsymbol{W}$ are independent of each other because $\boldsymbol{X}_0$ is, by construction, independent of $\boldsymbol{X}_{-\infty}^{-1}$. The upper bound (51) follows now by

$$\begin{aligned} D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}\,\middle\|\,P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right) &= D\left(P_{\boldsymbol{X}_0+\boldsymbol{V}+\boldsymbol{W}\,|\,\boldsymbol{x}_{-\infty}^{-1}}\,\middle\|\,P_{\boldsymbol{X}_0+\boldsymbol{V}+\boldsymbol{W}\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right) \\ &\le D\left(P_{\boldsymbol{X}_0+\boldsymbol{V}}\,\middle\|\,P_{\boldsymbol{X}_0+\boldsymbol{V}\,|\,X_1=0}\right) \\ &= D\left(P_{\tilde{\boldsymbol{Y}}_0\,|\,\boldsymbol{x}_{-\infty}^{-1}}\,\middle\|\,P_{\tilde{\boldsymbol{Y}}_0\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}\right)\Big|_{\boldsymbol{x}_{-\infty}^{-1}=0} \end{aligned} \tag{53}$$

where $P_{\boldsymbol{X}_0+\boldsymbol{V}+\boldsymbol{W}\,|\,\boldsymbol{x}_{-\infty}^{-1}}$ and $P_{\boldsymbol{X}_0+\boldsymbol{V}+\boldsymbol{W}\,|\,X_1=0,\boldsymbol{x}_{-\infty}^{-1}}$ denote the distributions of $\boldsymbol{X}_0+\boldsymbol{V}+\boldsymbol{W}$ conditional on the inputs $\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1}$ and on $(X_1=0,\boldsymbol{X}_{-\infty}^{-1}=\boldsymbol{x}_{-\infty}^{-1})$, respectively; $P_{\boldsymbol{X}_0+\boldsymbol{V}}$ denotes the unconditional distribution of $\boldsymbol{X}_0+\boldsymbol{V}$; and $P_{\boldsymbol{X}_0+\boldsymbol{V}\,|\,X_1=0}$ denotes the distribution of $\boldsymbol{X}_0+\boldsymbol{V}$ conditional on $X_1=0$. Here, the inequality follows by the data processing inequality for relative entropy [18, Section 2.9] and by noting that $\boldsymbol{X}_0+\boldsymbol{V}$ is independent of $\boldsymbol{X}_{-\infty}^{-1}$.

Returning to the analysis of (43), we obtain from (43) and (47)

$$\begin{aligned} \dot{C}(0) &\ge \varliminf_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \\ &\ge \varliminf_{\mathrm{SNR}\downarrow 0}\frac{1}{2}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha L\mathrm{SNR}} \\ &\quad -\frac{1}{2}\sum_{i=2}^{L}\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\ &= \frac{1}{2}\sum_{i=1}^{L}\alpha_{i-1}-\frac{1}{2}\sum_{i=2}^{L}\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}. \end{aligned} \tag{54}$$

By letting first ξ^2 go to infinity while holding L fixed, and by letting then L go to infinity, we obtain the desired lower bound

$$\dot{C}(0)\ge\varliminf_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}}\ge\frac{1}{2}(1+\alpha). \tag{55}$$

Thus, (55), (22), and (46) yield

$$\frac{1}{2}(1+\alpha)\le\varliminf_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}}\le\dot{C}(0)\le\dot{C}_{\mathrm{FB}}(0)\le\frac{1}{2}(1+\alpha) \tag{56}$$

which proves Theorem 2.

VI. Proof of Theorem 3

A. Part (i)

In order to show that

$$\varliminf_{\ell\to\infty}\frac{\alpha_{\ell+1}}{\alpha_\ell}>0 \tag{57}$$

implies that the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$ is bounded, we derive an upper bound on the capacity that is based on (19) and on an upper bound on $\frac{1}{n}I(M;Y_1^n)$. Again, we define $\alpha_0\triangleq 1$.

We first note that, according to (57), we can find an $\ell_0\in\mathbb{N}$ and a $0<\rho<1$ such that

$$\alpha_{\ell_0}>0\quad\text{and}\quad\frac{\alpha_{\ell+1}}{\alpha_\ell}\ge\rho,\quad\ell\ge\ell_0. \tag{58}$$

We continue with the chain rule for mutual information

$$\begin{aligned} \frac{1}{n}I(M;Y_1^n) &= \frac{1}{n}\sum_{k=1}^{\ell_0}I\left(M;Y_k\,\middle|\,Y_1^{k-1}\right) \\ &\quad+\frac{1}{n}\sum_{k=\ell_0+1}^{n}I\left(M;Y_k\,\middle|\,Y_1^{k-1}\right). \end{aligned} \tag{59}$$

Each summand in the first sum on the RHS of (59) is upper-bounded by

$$\begin{aligned} I\left(M;Y_k\,\middle|\,Y_1^{k-1}\right) &\le h(Y_k)-h\left(Y_k\,\middle|\,Y_1^{k-1},M\right) \\ &= h(Y_k)-\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]-h\left(U_k\,\middle|\,U_1^{k-1}\right) \\ &\le \frac{1}{2}\log\left(2\pi e\left(1+\sum_{\ell=1}^{k}\alpha_{k-\ell}\frac{\mathsf{E}\left[X_\ell^2\right]}{\sigma^2}\right)\right)-h\left(U_k\,\middle|\,U_1^{k-1}\right) \\ &\le \frac{1}{2}\log\left(2\pi e\left(1+\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\sum_{\ell=1}^{k}\frac{\mathsf{E}\left[X_\ell^2\right]}{\sigma^2}\right)\right) \\ &\qquad -h\left(U_k\,\middle|\,U_1^{k-1}\right) \\ &\le \frac{1}{2}\log\left(2\pi e\left(1+\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\,n\,\mathrm{SNR}\right)\right)-h\left(U_k\,\middle|\,U_1^{k-1}\right) \\ &\le \frac{1}{2}\log\left(2\pi e\left(1+\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}\,n\,\mathrm{SNR}\right)\right)-h\left(U_k\,\middle|\,U_{-\infty}^{k-1}\right). \end{aligned} \tag{60}$$

Recall that $\sup_{\ell'\in\mathbb{N}_0}\alpha_{\ell'}$ is finite (10). Here, the first step follows because conditioning cannot increase entropy; the second step follows because (X_1^k,U_1^{k-1}) is a function of (M,Y_1^{k-1}), from the behavior of entropy under translation and scaling [18, Theorems 9.6.3 and 9.6.4], and from the fact that, conditional on U_1^{k-1}, the random variable U_k is independent of (X_1^k,M,Y_1^{k-1}); the third step follows from the entropy maximizing property of Gaussian random variables and by lower-bounding $\mathsf{E}[\log(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2)]\ \ge\ \log\sigma^2$; the fourth step follows by upper-bounding each coefficient α_ℓ by the supremum of α_ℓ, $\ell\in\mathbb{N}_0$; the fifth step follows from the power constraint (12); and the last step follows because conditioning cannot increase entropy.

The summands in the second sum on the RHS of (59) are upper-bounded using the general upper bound for mutual information [15, Theorem 5.1]

$$I(X;Y)\le\int D\Big(W(\cdot\,|\,x)\,\Big\|\,R(\cdot)\Big)\mathrm{d}Q(x) \tag{61}$$

where $W(\cdot\,|\,\cdot)$ is the channel law, $Q(\cdot)$ is the distribution on the channel input X, and $R(\cdot)$ is any distribution on the output alphabet. Thus, any choice of output distribution $R(\cdot)$ yields an upper bound on the mutual information.

For each $k = \ell_0 + 1, \dots, n$, we upper-bound the conditional mutual information $I(M; Y_k \mid Y_1^{k-1} = y_1^{k-1})$ for a given $Y_1^{k-1} = y_1^{k-1}$ by choosing $R(\cdot)$ to be a Cauchy distribution whose density is given by

$$\frac{\sqrt{\beta}}{\pi}\frac{1}{1+\beta y_k^2}, \quad y_k \in \mathbb{R} \tag{62}$$

where we choose $\beta = 1/(\tilde{\beta}\, y_{k-\ell_0}^2)$ with

$$\tilde{\beta} = \min\left\{\rho^{\ell_0-1}\,\frac{\alpha_{\ell_0}}{\max_{0\le\ell'<\ell_0}\alpha_{\ell'}}, \rho^{\ell_0}\right\}$$

where $0 < \rho < 1$ and $\ell_0 \in \mathbb{N}$ are given by (58).[5] Note that (58) together with (10) implies that

$$0 < \tilde{\beta} < 1 \quad \text{and} \quad \tilde{\beta}\alpha_\ell \le \alpha_{\ell+\ell_0}, \quad \ell \in \mathbb{N}_0. \tag{63}$$

Applying (62) to (61) yields

$$\begin{aligned} I\left(M; Y_k \left| Y_1^{k-1} = y_1^{k-1}\right.\right) &\le \mathsf{E}\left[\log\left(1 + \frac{Y_k^2}{\tilde{\beta}\,Y_{k-\ell_0}^2}\right)\middle| Y_1^{k-1} = y_1^{k-1}\right] \\ &\quad + \frac{1}{2}\log\left(\tilde{\beta}\,y_{k-\ell_0}^2\right) + \log\pi \\ &\quad - h\left(Y_k \left| M, Y_1^{k-1} = y_1^{k-1}\right.\right), \end{aligned} \tag{64}$$

and we thus obtain, averaging over Y_1^{k-1}

$$\begin{aligned} I\left(M; Y_k \left| Y_1^{k-1}\right.\right) \le \log\pi - h\left(Y_k \left| Y_1^{k-1}, M\right.\right) &+ \frac{1}{2}\mathsf{E}\left[\log\left(\tilde{\beta}\,Y_{k-\ell_0}^2\right)\right] \\ &+ \mathsf{E}\left[\log\left(\tilde{\beta}\,Y_{k-\ell_0}^2 + Y_k^2\right)\right] \\ &- \mathsf{E}\left[\log\left(Y_{k-\ell_0}^2\right)\right] - \log\tilde{\beta}. \end{aligned} \tag{65}$$

We evaluate the terms on the RHS of (65) individually. We begin with

$$\begin{aligned} &h\left(Y_k \left| Y_1^{k-1}, M\right.\right) \\ &\quad \ge \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] + h\left(U_k \left| U_{-\infty}^{k-1}\right.\right) \end{aligned} \tag{66}$$

where we use the same arguments as in the second step in (60). The next term is upper-bounded by

$$\begin{aligned} &\mathsf{E}\left[\log\left(\tilde{\beta}Y_{k-\ell_0}^2\right)\right] \\ &= \mathsf{E}\left[\mathsf{E}\left[\log\left(\tilde{\beta}\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2\right)\middle| X_1^{k-\ell_0}\right]\right] \\ &\le \mathsf{E}\left[\log\left(\tilde{\beta}\mathsf{E}\left[\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2\middle| X_1^{k-\ell_0}\right]\right)\right] \\ &= \mathsf{E}\left[\log\left(\tilde{\beta}\,X_{k-\ell_0}^2 + \tilde{\beta}\,\sigma^2 + \tilde{\beta}\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] \\ &\le \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2\right)\right] \end{aligned} \tag{67}$$

where we define, for a given $X_1^{k-1} = x_1^{k-1}$,

$$\theta\left(x_1^{k-1}\right) \triangleq \sqrt{\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}\,x_\ell^2}. \tag{68}$$

Here, the second step in (67) follows from Jensen's inequality, and the last step follows from (63). Similarly, we use Jensen's inequality along with (63) to upper-bound

$$\begin{aligned} &\mathsf{E}\left[\log\left(\tilde{\beta}Y_{k-\ell_0}^2 + Y_k^2\right)\right] \\ &\quad \le \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}\,X_\ell^2 + \sigma^2 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2\right)\right] \\ &\quad \le \log 2 + \mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k}\alpha_{k-\ell}\,X_\ell^2\right)\right]. \end{aligned} \tag{69}$$

In order to lower-bound $\mathsf{E}[\log(Y_{k-\ell_0}^2)]$ we need the following lemma.

Lemma 4: Let X be a random variable of density $f_X(x)$, $x \in \mathbb{R}$. Then for any $0 < \delta \le 1$ and $0 < \eta < 1$

$$\sup_{c\in\mathbb{R}}\mathsf{E}\left[\log|X+c|^{-1}\,\mathrm{I}\{|X+c| \le \delta\}\right] \le \epsilon(\delta,\eta) + \frac{1}{\eta}h^-(X) \tag{70}$$

where $\mathrm{I}\{\text{statement}\}$ is 1 if the statement is true and 0 otherwise; $h^-(X)$ is defined as

$$h^-(X) \triangleq \int_{\{x\in\mathbb{R}: f_X(x)>1\}} f_X(x)\log f_X(x)\mathrm{d}x; \tag{71}$$

and where $\epsilon(\delta,\eta) > 0$ tends to zero as $\delta \downarrow 0$.

Proof: See [15, Lemma 6.7]. □

We write the expectation as

$$\begin{aligned} &\mathsf{E}\left[\log\left(Y_{k-\ell_0}^2\right)\right] \\ &\quad = \mathsf{E}\left[\mathsf{E}\left[\log\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2\middle| X_1^{k-\ell_0}\right]\right] \end{aligned}$$

and lower-bound the conditional expectation for a given $X_1^{k-\ell_0} = x_1^{k-\ell_0}$

$$\begin{aligned} &\mathsf{E}\left[\log\left(X_{k-\ell_0} + \theta\left(X_1^{k-\ell_0-1}\right)U_{k-\ell_0}\right)^2\middle| X_1^{k-\ell_0} = x_1^{k-\ell_0}\right] \\ &= \log\theta^2\left(x_1^{k-\ell_0-1}\right) \\ &\quad - 2\,\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\left(X_1^{k-\ell_0-1}\right)} + U_{k-\ell_0}\right|^{-1}\middle| X_1^{k-\ell_0} = x_1^{k-\ell_0}\right] \\ &\ge \log\theta^2\left(x_1^{k-\ell_0-1}\right) - 2\epsilon(\delta,\eta) - \frac{2}{\eta}h^-(U_{k-\ell_0}) + \log\delta^2 \end{aligned} \tag{72}$$

[5]When $y_{k-\ell_0} = 0$, with this choice of β the density of the Cauchy distribution (62) is undefined. However, this event is of zero probability and has therefore no impact on the mutual information $I(M; Y_k \mid Y_1^{k-1})$.

$$\mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)}+U_{k-\ell_0}\right|^{-1}\right]=\mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)}+U_{k-\ell_0}\right|^{-1}\mathrm{I}\left\{\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)}+U_{k-\ell_0}\right|\leq\delta\right\}\right]$$
$$+\mathsf{E}\left[\log\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)}+U_{k-\ell_0}\right|^{-1}\mathrm{I}\left\{\left|\frac{x_{k-\ell_0}}{\theta\left(x_1^{k-\ell_0-1}\right)}+U_{k-\ell_0}\right|>\delta\right\}\right] \quad (73)$$

for some $0<\delta\leq 1$ and $0<\eta<1$. Here, the inequality follows by splitting the conditional expectation into the two expectations given in (73) at the top of the page, and by upper-bounding then the first expectation on the RHS of (73) using Lemma 4 and the second expectation by $-\log\delta$. Averaging (72) over $X_1^{k-\ell_0}$ yields

$$\mathsf{E}\left[\log\left(Y_{k-\ell_0}^2\right)\right]\geq\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right]-2\epsilon(\delta,\eta)-\frac{2}{\eta}h^-(U_{k-\ell_0})+\log\delta^2. \quad (74)$$

Note that the fact that $U_{k-\ell_0}$ is of unit variance together with [15, Lemma 6.4] implies that $h^-(U_{k-\ell_0})$ is finite.

Turning back to the upper bound (65), we obtain from (66), (67), (69), and (74)

$$\begin{aligned}I\left(M;Y_k\left|Y_1^{k-1}\right.\right)&\leq\log\pi-\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]-h\left(U_k\left|U_{-\infty}^{k-1}\right.\right)\\&\quad+\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2\right)\right]\\&\quad+\log 2+\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right]\\&\quad-\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right]\\&\quad+2\,\epsilon(\delta,\eta)+\frac{2}{\eta}h^-(U_{k-\ell_0})-\log\delta^2-\log\tilde\beta\\&\leq\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right]\\&\quad-\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right]+\mathsf{K}\end{aligned} \quad (75)$$

where

$$\mathsf{K}\triangleq\log\frac{2\pi}{\tilde\beta\,\delta^2}-h\left(U_k\left|U_{-\infty}^{k-1}\right.\right)+\frac{2}{\eta}h^-(U_{k-\ell_0})+2\,\epsilon(\delta,\eta)$$

is a finite constant. The last step in (75) follows because we have with probability one $\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2\leq\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2$. Note that K does not depend on k since the process $\{U_k\}$ is stationary.

Turning back to the evaluation of the second sum on the RHS of (59), we use that, for any sequences $\{a_k\}$ and $\{b_k\}$

$$\sum_{k=\ell_0+1}^{n}(a_k-b_k)=\sum_{k=n-2\ell_0+1}^{n}(a_k-b_{k-n+3\ell_0})+\sum_{k=\ell_0+1}^{n-2\ell_0}(a_k-b_{k+2\ell_0}). \quad (76)$$

Defining

$$a_k\triangleq\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right] \quad (77)$$

and

$$b_k\triangleq\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] \quad (78)$$

we have for the first sum on the RHS of (76)

$$\begin{aligned}&\sum_{k=n-2\ell_0+1}^{n}(a_k-b_{k-n+3\ell_0})\\&=\sum_{k=n-2\ell_0+1}^{n}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k-n+2\ell_0-1}\alpha_{k-n+2\ell_0-\ell}X_\ell^2}\right)\right]\\&\leq 2\,\ell_0\,\log\left(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\,n\,\mathrm{SNR}\right)\end{aligned} \quad (79)$$

which follows by lower-bounding the denominator by σ^2, and by using then Jensen's inequality together with the last two steps in (60). For the second sum on the RHS of (76) we have

$$\begin{aligned}&\sum_{k=\ell_0+1}^{n-2\ell_0}(a_k-b_{k+2\ell_0})\\&\quad=\sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}X_\ell^2}\right)\right]\\&\quad\leq\sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k+\ell_0-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}X_\ell^2}\right)\right]\\&\qquad-(n-3\ell_0)\,\log\tilde\beta\\&\quad\leq-(n-3\ell_0)\log\tilde\beta\end{aligned} \quad (80)$$

where the second step follows by adding $\log\tilde\beta$ to the expectation and by upper-bounding then $\tilde\beta\,\sigma^2\leq\sigma^2$ and $\tilde\beta\,\alpha_\ell\leq\alpha_{\ell+\ell_0}$ (63); and the third step follows because we have with probability one

$$\sum_{\ell=1}^{k}\alpha_{k+\ell_0-\ell}X_\ell^2\leq\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}X_\ell^2.$$

We combine now (75), (76), (79), and (80) to upper-bound

$$\begin{aligned}&\frac{1}{n}\sum_{\ell=\ell_0+1}^{n}I\left(M;Y_k\left|Y_1^{k-1}\right.\right)\\&\quad\leq\frac{n-\ell_0}{n}\mathsf{K}+\frac{2\ell_0}{n}\log\left(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\,n\,\mathrm{SNR}\right)\end{aligned}$$

$$- \frac{n-3\ell_0}{n} \log \tilde{\beta} \quad (81)$$

which together with (59) and (60) yields

$$\begin{aligned} &\frac{1}{n} I\left(M; Y_1^n\right) \\ &\le \frac{n-\ell_0}{n}\mathsf{K} - \frac{n-3\ell_0}{n}\log\tilde{\beta} + \frac{\ell_0}{2n}\log(2\pi e) \\ &\quad + \frac{\ell_0}{n}\frac{5}{2}\log\left(1+\sup_{\ell\in\mathbb{N}_0}\alpha_\ell\, n\,\mathrm{SNR}\right) - \frac{\ell_0}{n} h\left(U_k \,\middle|\, U_{-\infty}^{k-1}\right). \end{aligned} \quad (82)$$

This converges to $\mathsf{K} - \log\tilde{\beta} < \infty$ as we let n tend to infinity, thus proving that $\underline{\lim}_{\ell\to\infty}\alpha_{\ell+1}/\alpha_\ell > 0$ implies that the capacity $C_{\mathrm{FB}}(\mathrm{SNR})$ is bounded in the SNR.

B. Part (ii)

We show that

$$\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty \quad (83)$$

implies that the capacity $C(\mathrm{SNR})$ is unbounded in the SNR. Part (ii) of Theorem 3 follows then by noting that

$$\lim_{\ell\to\infty}\frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \Longrightarrow \lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty.$$

We prove the claim by proposing a coding scheme that achieves an unbounded rate. We first note that (83) implies that for any $0 < \varrho < 1$ we can find an $\ell_0 \in \mathbb{N}$ such that

$$\alpha_\ell < \varrho^\ell, \quad \ell \ge \ell_0. \quad (84)$$

If there exists an $\ell_0 \in \mathbb{N}$ such that $\alpha_\ell = 0$, $\ell \ge \ell_0$, then we can achieve the (unbounded) rate

$$R = \frac{1}{2L}\log(1 + L\mathrm{SNR}), \quad L \ge \ell_0$$

by a coding scheme where $\{X_{kL+1}, k \in \mathbb{N}_0\}$ is a sequence of i.i.d., zero-mean, Gaussian random variables of variance $L\mathsf{P}$, and where the other inputs are deterministically zero. Indeed, by waiting L time-steps, the chip's temperature cools down to the ambient one, so the noise variance is independent of the previous channel inputs and we can achieve—after appropriate normalization—the capacity of the additive white Gaussian noise (AWGN) channel [23].

For the more general case (84) we propose the following encoding and decoding scheme. Let $x_1^n(m)$, $m \in \mathcal{M}$ denote the codeword sent out by the transmitter that corresponds to the message $M = m$. We choose some $L \ge \ell_0$ and generate the components $x_{kL+1}(m)$, $m \in \mathcal{M}$, $k = 0, \ldots, \lfloor n/L \rfloor - 1$ independently of each other according to a zero-mean Gaussian law of variance P. The other components are set to zero. (It follows from the weak law of large numbers that $\frac{1}{n}\sum_{k=1}^n x_k^2(m)$ converges to P/L in probability as n tends to infinity. This guarantees that the probability that a codeword does not satisfy the per-message power constraint (14)—and hence also the average-power constraint (12)—vanishes as n tends to infinity.)

The receiver uses a *nearest neighbor decoder* in order to guess M based on the received sequence of channel outputs y_1^n. Thus, it computes $\|\boldsymbol{y} - \boldsymbol{x}(m')\|^2$ for each $m' \in \mathcal{M}$ and decides on the message that satisfies

$$\hat{M} = \arg\min_{m'\in\mathcal{M}} \|\boldsymbol{y} - \boldsymbol{x}(m')\|^2 \quad (85)$$

where ties are resolved with a fair coin flip. Here $\|\cdot\|$ denotes the Euclidean norm, and $\boldsymbol{y}$ and $\boldsymbol{x}(m')$ denote the vectors

$$(y_1, y_{L+1}, \ldots, y_{(\lfloor n/L\rfloor-1)L+1})^{\mathsf{T}}$$

and

$$(x_1(m'), x_{L+1}(m'), \ldots, x_{(\lfloor n/L\rfloor-1)L+1}(m'))^{\mathsf{T}}.$$

We are interested in the average probability of error $\Pr(\hat{M} \ne M)$, averaged over all codewords in the codebook, and averaged over all codebooks. By the symmetry of the codebook construction, the probability of error corresponding to the mth message $\Pr(\hat{M} \ne M \,|\, M = m)$ does not depend on m, and we thus conclude that

$$\Pr(\hat{M} \ne M) = \Pr(\hat{M} \ne M \,|\, M = 1).$$

We further note that

$$\begin{aligned} &\Pr(\hat{M} \ne M \,|\, M = 1) \\ &\le \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\boldsymbol{Y} - \boldsymbol{X}(m')\|^2 \le \|\boldsymbol{Z}\|^2 \,\middle|\, M = 1\right) \end{aligned} \quad (86)$$

where

$$\boldsymbol{Z} = \begin{pmatrix} \theta(X_1(1))U_1 \\ \theta\left(X_1^L(1)\right)U_{L+1} \\ \vdots \\ \theta\left(X_1^{(\lfloor n/L\rfloor-1)L}(1)\right)U_{(\lfloor n/L\rfloor-1)L+1} \end{pmatrix}$$

which is, conditional on $M = 1$, equal to $\boldsymbol{Y} - \boldsymbol{X}(1)$. In order to analyze (86) we need the following lemma.

Lemma 5: Consider the channel described in Section II, and assume that $\{\alpha_\ell\}$ satisfies (83). Further assume that $\{X_{kL+1}, k \in \mathbb{N}_0\}$ is a sequence of i.i.d., zero-mean, Gaussian random variables of variance P, and that $X_k = 0$ for $k \bmod L \ne 1$. (Here $k \bmod L$ denotes the remainder upon dividing k by L). Let the set $\mathcal{D}_\epsilon$ be defined as

$$\begin{aligned} \mathcal{D}_\epsilon \triangleq \Bigg\{ &(\boldsymbol{y}, \boldsymbol{z}) \in \mathbb{R}^{\lfloor n/L\rfloor} \times \mathbb{R}^{\lfloor n/L\rfloor} : \\ &\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{y}\|^2 - \left(\sigma^2 + \mathsf{P} + \alpha^{(L)}\mathsf{P}\right)\right| \le \epsilon, \\ &\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{z}\|^2 - \left(\sigma^2 + \alpha^{(L)}\mathsf{P}\right)\right| \le \epsilon \Bigg\} \end{aligned} \quad (87)$$

where we define

$$\alpha^{(L)} \triangleq \sum_{\ell=1}^{\infty} \alpha_{\ell L}. \quad (88)$$

Then

$$\lim_{n\to\infty} \Pr\left((\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{D}_\epsilon\right) = 1 \quad (89)$$

for any $\epsilon > 0$.

Proof: See Appendix C. □

In order to upper-bound the RHS of (86) we proceed along the lines of [23], [22]. Using that, by the symmetry of the codebook construction, the law of $(\boldsymbol{Y}, \boldsymbol{Z})$ does not depend on m, and using that the codewords are independent of each other so, conditional on $M = 1$, the distribution of $(\boldsymbol{X}(2), \ldots, \boldsymbol{X}(|\mathcal{M}|))$ does not depend on $(\boldsymbol{y}, \boldsymbol{z})$, we obtain

$$\begin{aligned}\Pr&\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\boldsymbol{Y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{Z}\|^2 \,\middle|\, M=1\right)\\ &\leq \Pr\Big((\boldsymbol{Y},\boldsymbol{Z}) \notin \mathcal{D}_\epsilon\Big)\\ &\quad + \int_{\mathcal{D}_\epsilon} \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\boldsymbol{y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{z}\|^2\right) \mathrm{d}P(\boldsymbol{y},\boldsymbol{z}).\end{aligned} \tag{90}$$

It follows from Lemma 5 that the first term on the RHS of (90) vanishes as n tends to infinity. To evaluate the second term on the RHS of (90), we note that by the union of events bound

$$\begin{aligned}\Pr&\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\boldsymbol{y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{z}\|^2\right)\\ &\leq \sum_{m'=2}^{|\mathcal{M}|} \Pr\left(\|\boldsymbol{y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{z}\|^2\right).\end{aligned} \tag{91}$$

By upper-bounding

$$\|\boldsymbol{z}\|^2 \leq \lfloor n/L \rfloor \left(\sigma^2 + \alpha^{(L)}\mathsf{P} + \epsilon\right), \quad (\boldsymbol{y},\boldsymbol{z}) \in \mathcal{D}_\epsilon,$$

by lower-bounding

$$\|\boldsymbol{y}\|^2 \geq \lfloor n/L \rfloor \left(\sigma^2 + \mathsf{P} + \alpha^{(L)}\mathsf{P} - \epsilon\right), \quad (\boldsymbol{y},\boldsymbol{z}) \in \mathcal{D}_\epsilon,$$

and by applying Chernoff's bound [24, Section 5.4], we obtain for each $m' = 2, \ldots, |\mathcal{M}|$ and for any $s < 0$

$$\begin{aligned}\Pr&\left(\|\boldsymbol{y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{z}\|^2\right)\\ &\leq \exp\Big(-s\lfloor n/L\rfloor(\sigma^2+\alpha^{(L)}\mathsf{P}+\epsilon)\Big)\\ &\quad\times \exp\left(s\frac{\lfloor n/L\rfloor\left(\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}-\epsilon\right)}{1-2s\mathsf{P}}\right)\\ &\quad\times \exp\left(-\frac{1}{2}\lfloor n/L\rfloor \log(1-2\,s\,\mathsf{P})\right), \quad (\boldsymbol{y},\boldsymbol{z}) \in \mathcal{D}_\epsilon.\end{aligned} \tag{92}$$

Applying (91) and (92) to (90), it follows that

$$\Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\boldsymbol{Y}-\boldsymbol{X}(m')\|^2 \leq \|\boldsymbol{Z}\|^2 \,\middle|\, M=1\right)$$

tends to zero as n tends to infinity if for some $s < 0$ the rate R satisfies

$$\begin{aligned}R < \frac{s}{L}\left(\sigma^2+\alpha^{(L)}\mathsf{P}+\epsilon\right) &+ \frac{1}{2L}\log(1-2s\mathsf{P})\\ &- \frac{s}{L}\frac{\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}-\epsilon}{1-2s\mathsf{P}}.\end{aligned} \tag{93}$$

Hence, by choosing $s = -\frac{1}{2}\frac{1}{1+\alpha^{(L)}\mathsf{P}}$, it follows that any rate below

$$\begin{aligned}-\frac{1}{2L}\frac{\sigma^2+\alpha^{(L)}\mathsf{P}+\epsilon}{1+\alpha^{(L)}\mathsf{P}} &+ \frac{1}{2L}\log\left(1+\frac{\mathsf{P}}{1+\alpha^{(L)}\mathsf{P}}\right)\\ &+\frac{1}{2L}\frac{\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}-\epsilon}{1+\alpha^{(L)}\mathsf{P}}\frac{1}{1+\mathsf{P}/\left(1+\alpha^{(L)}\mathsf{P}\right)}\end{aligned}$$

is achievable. As P tends to infinity this converges to

$$\frac{1}{2L}\log\left(1+\frac{1}{\alpha^{(L)}}\right) > \frac{1}{2L}\log\frac{1}{\alpha^{(L)}}. \tag{94}$$

It remains to show that given (84) we can make $-\frac{1}{2L}\log\alpha^{(L)}$ arbitrarily large. Indeed, (84) implies that

$$\alpha^{(L)} = \sum_{\ell=1}^{\infty}\alpha_{\ell L} < \sum_{\ell=1}^{\infty}\varrho^{\ell L} = \frac{\varrho^L}{1-\varrho^L}$$

and the RHS of (94) can therefore be further lower-bounded by

$$\frac{1}{2L}\log(1-\varrho^L) + \frac{1}{2}\log\frac{1}{\varrho}.$$

Letting L tend to infinity yields that we can achieve any rate below

$$\frac{1}{2}\log\frac{1}{\varrho}.$$

Since this can be made arbitrarily large by choosing ϱ sufficiently small, we conclude that $\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty$ implies that the capacity is unbounded.

VII. Conclusion

We studied a model for on-chip communication with nonideal heat sinks. To account for the heating-up effect, we proposed a channel model where the variance of the additive noise depends on a weighted sum of the past channel input powers. The weights are related to the efficiency of the heat sink.

To study the capacity of this channel at low SNR, we computed the capacity per unit cost. We showed that, irrespective of the distribution on the (i.i.d.) noise sequence $\{U_k\}$, the heating-up effect is not harmful in the sense that the capacity per unit cost cannot be smaller than the capacity per unit cost of the channel with an ideal sink (i.e., for $\alpha = 0$). We further showed that if the noise $\{U_k\}$ is i.i.d. Gaussian, then the heating-up effect is even beneficial in the sense that the capacity per unit cost is *larger* than the capacity per unit cost of the channel with an ideal heat sink. This suggests that at low SNR no heat sinks should be used. (Of course, there may be other reasons to use heat sinks.) Studying capacity at high SNR, we derived a sufficient condition and a necessary condition for the capacity to be bounded in the SNR. We showed that if $\{\alpha_\ell\}$ decays not faster than geometrically, then the capacity is bounded in the SNR. On the other hand, if $\{\alpha_\ell\}$ decays faster than geometrically, then the capacity is unbounded in the SNR. This result demonstrates the importance of an efficient heat sink at high SNR.

APPENDIX A
PROOF OF PROPOSITION 1

We first note that by the expression of the capacity per unit cost of a memoryless channel [20] we have

$$\sup_{\mathrm{SNR}>0} \frac{C_{\alpha=0}(\mathrm{SNR})}{\mathrm{SNR}} = \sup_{\zeta>0} \frac{D\Big(W_{\alpha=0}(\cdot\,|\,\zeta)\big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\zeta^2/\sigma^2} \tag{95}$$

where $W_{\alpha=0}(\cdot\,|\,\cdot)$ denotes the channel law of the channel

$$Y_k = x_k + \sigma U_k. \tag{96}$$

Thus, to prove Proposition 1 it suffices to show that

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\zeta>0} \frac{D\Big(W_{\alpha=0}(\cdot\,|\,\zeta)\big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\zeta^2/\sigma^2}.$$

We shall obtain this result by deriving a lower bound on $C_{\mathrm{Info}}(\mathrm{SNR})$ and by computing then its limiting ratio to SNR as SNR tends to zero.

In order to lower-bound $C_{\mathrm{Info}}(\mathrm{SNR})$, which we defined in (16) as

$$C_{\mathrm{Info}}(\mathrm{SNR}) = \varliminf_{n\to\infty} \frac{1}{n} \sup I(X_1^n; Y_1^n),$$

we evaluate $\frac{1}{n}I(X_1^n;Y_1^n)$ for inputs $\{X_k\}$ that are blockwise i.i.d. in blocks of L symbols (for some $L \in \mathbb{N}$). Thus, $\{(X_{bL+1},\dots,X_{(b+1)L}), b\in\mathbb{N}_0\}$ is a sequence of i.i.d. random length-L vectors with $(X_{bL+1},\dots,X_{(b+1)L})$ taking on the value $(\xi,0,\dots,0)$ with probability δ and $(0,\dots,0)$ with probability $1-\delta$, for some $\xi\in\mathbb{R}$. To satisfy the power constraint (12) we shall choose ξ and δ such that

$$\frac{\xi^2}{\sigma^2}\delta = L\,\mathrm{SNR}. \tag{97}$$

We use the chain rule for mutual information to write

$$\begin{aligned} \frac{1}{n} I\left(X_1^n;Y_1^n\right) &= \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(X_{bL+1};Y_1^n\,\middle|\,X_1^{bL}\right) \\ &\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\left(X_{bL+1};Y_{bL+1}\,\middle|\,X_1^{bL}\right) \end{aligned} \tag{98}$$

where the inequality follows because reducing observations cannot increase mutual information.

Let $R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ denote the maximum rate achievable on (96) using on–off keying with on-symbol ξ and with its corresponding probability $\wp$ chosen in order to satisfy the power constraint snr, i.e.,

$$R_{\text{on-off}}^{(\xi)}(\mathsf{snr}) \triangleq \sup_{\substack{P_X(\xi)=1-P_X(0)=\wp,\\ \xi^2/\sigma^2\wp\leq\mathsf{snr}}} I(X;X+\sigma U_k). \tag{99}$$

Notice that $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is a nonnegative, monotonically nondecreasing function with $R_{\text{on-off}}^{(\xi)}(0)=0$. From the strict concavity of mutual information it follows that

$$R_{\text{on-off}}^{(\xi)}(\mathsf{snr}) > 0, \quad \mathsf{snr}>0.$$

Also, for a fixed ξ, the function $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is concave in snr. Consequently, for some $\mathsf{snr}_0>0$, we have that $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is strictly monotonic in $\mathsf{snr}\in[0,\mathsf{snr}_0]$ and hence the supremum on the RHS of (99) is attained for

$$\wp = \mathsf{snr}\,\sigma^2/\xi^2, \quad 0\leq\mathsf{snr}\leq\mathsf{snr}_0.$$

By expressing $I(X_{bL+1};Y_{bL+1}\,|\,X_1^{bL}=x_1^{bL})$ for a given $X_1^{bL}=x_1^{bL}$ as

$$\begin{aligned} &I\left(X_{bL+1};Y_{bL+1}\,\middle|\,X_1^{bL}=x_1^{bL}\right) \\ &\quad= I\left(X_{bL+1};X_{bL+1}+\theta\left(x_1^{bL}\right)U_{bL+1}\right) \\ &\quad= I\left(X_{bL+1};\frac{\sigma}{\theta\left(x_1^{bL}\right)}X_{bL+1}+\sigma\,U_{bL+1}\right) \end{aligned}$$

(where $\theta(x_1^{bL})$ is defined in (68)), and by using that for $\mathsf{snr}\in[0,\mathsf{snr}_0]$ the supremum on the RHS of (99) is attained for $\wp=\mathsf{snr}\,\sigma^2/\xi^2$, we obtain

$$\begin{aligned} &I\left(X_{bL+1};Y_{bL+1}\,\middle|\,X_1^{bL}=x_1^{bL}\right) \\ &\qquad= R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,x_{\ell L+1}^2/\sigma^2}\right) \end{aligned} \tag{100}$$

for $0\leq\mathrm{SNR}\leq\mathsf{snr}_0/L$. Averaging (100) over X_1^{bL}, and applying the result to (98), yields

$$\begin{aligned} &\frac{1}{n}I\left(X_1^n;Y_1^n\right) \\ &\quad\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1}\mathsf{E}\left[R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,X_{\ell L+1}^2/\sigma^2}\right)\right] \\ &\quad\geq \frac{\lfloor n/L\rfloor}{n}R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right), \end{aligned} \tag{101}$$

where the second inequality follows by the monotonicity of $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ and because we have with probability one $\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}\,X_{\ell L+1}^2/\sigma^2\leq\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2$. The lower bound on $C_{\mathrm{Info}}(\mathrm{SNR})$ follows by letting n tend to infinity

$$\begin{aligned} C_{\mathrm{Info}}(\mathrm{SNR}) &\geq \varliminf_{n\to\infty}\frac{1}{n}I\left(X_1^n;Y_1^n\right) \\ &\geq \frac{1}{L}R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right) \end{aligned} \tag{102}$$

for $0\leq\mathrm{SNR}\leq\mathsf{snr}_0/L$.

We continue by lower-bounding the information capacity per unit cost as

$$\begin{aligned} &\sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \\ &\quad\geq \varliminf_{\mathrm{SNR}\downarrow0}\frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \\ &\quad\geq \varliminf_{\mathrm{SNR}\downarrow0}\frac{1}{L}\frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right)}{\mathrm{SNR}} \\ &\quad= \varliminf_{\mathrm{SNR}\downarrow0}\frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}\right)}{\frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}}\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2} \\ &\quad= \varliminf_{\mathrm{SNR}'\downarrow0}\frac{R_{\text{on-off}}^{(\xi)}(\mathrm{SNR}')}{\mathrm{SNR}'}\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2} \end{aligned} \tag{103}$$

where in the last step we substitute

$$\mathrm{SNR}' = \frac{L\,\mathrm{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}.$$

Proceeding along the lines of the proof of [20, Theorem 3], it can be shown that

$$\lim_{\mathrm{SNR}'\downarrow 0}\frac{R_{\text{on-off}}^{(\xi)}(\mathrm{SNR}')}{\mathrm{SNR}'} = \frac{D\Big(W_{\alpha=0}(\cdot\,|\,\xi)\Big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\xi^2/\sigma^2} \tag{104}$$

and therefore

$$\sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{D\Big(W_{\alpha=0}(\cdot\,|\,\xi)\Big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\xi^2/\sigma^2}\,\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\,\xi^2/\sigma^2}. \tag{105}$$

Noting that (10) and (23) imply

$$0 \leq \lim_{L\to\infty}\sum_{\ell=1}^{\infty}\alpha_{\ell L} \leq \lim_{L\to\infty}\sum_{\ell=L}^{\infty}\alpha_\ell = 0 \tag{106}$$

we obtain, upon letting L tend to infinity

$$\sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{D\Big(W_{\alpha=0}(\cdot\,|\,\xi)\Big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\xi^2/\sigma^2}. \tag{107}$$

Maximizing (107) over ξ^2 yields then

$$\sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\xi>0}\frac{D\Big(W_{\alpha=0}(\cdot\,|\,\xi)\Big\|W_{\alpha=0}(\cdot\,|\,0)\Big)}{\xi^2/\sigma^2} \tag{108}$$

which, in view of (95), proves Proposition 1.

Appendix B

We prove that

$$\lim_{b\to\infty} I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b\,\middle|\,\boldsymbol{X}_0^b\right) = 0. \tag{109}$$

Let $\alpha_b^{(i)}$ be defined as

$$\alpha_0^{(1)} \triangleq 0 \tag{110}$$

$$\alpha_b^{(i)} \triangleq \alpha_{bL+i-1}, \quad (b,i)\in\mathbb{N}_0\times\mathbb{N}\setminus\{(0,1)\}. \tag{111}$$

We have

$$\begin{aligned}
I&\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b\,\middle|\,\boldsymbol{X}_0^b\right)\\
&= \sum_{i=1}^{L} I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{Y}_{bL+i}\,\middle|\,\boldsymbol{X}_0^b,\tilde{Y}_{bL+1}^{bL+i-1}\right)\\
&\leq \sum_{i=1}^{L}\left(h\left(\tilde{Y}_{bL+i}\,\middle|\,\boldsymbol{X}_0^b\right) - h\left(\tilde{Y}_{bL+i}\,\middle|\,\boldsymbol{X}_{-\infty}^b\right)\right)\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2+\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right)\right]\\
&\quad - \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=-\infty}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2\right)\right]\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2.+\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right)\right]\\
&\quad - \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2\right)\right]\\
&= \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(1+\frac{\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}}{\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2}\right)\right]\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\log\left(1+L\,\mathrm{SNR}\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right)
\end{aligned} \tag{112}$$

where the second step follows because conditioning cannot increase entropy and because, conditional on $\boldsymbol{X}_{-\infty}^b$, the random variable $\tilde{Y}_{bL+i}$ is independent of $\tilde{Y}_{bL+1}^{bL+i-1}$; the third step follows from the entropy maximizing property of Gaussian random variables and because, conditional on $\boldsymbol{X}_{-\infty}^b$, the random variable $\tilde{Y}_{bL+1}^{bL+i-1}$ is Gaussian; the fourth step follows because, with probability one, $\sum_{\ell=-\infty}^{-1}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2 \geq 0$; and the last step follows because, with probability one, $\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2 \geq 0$.

By upper-bounding $\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)} \leq \sum_{\ell=b+1}^{\infty}\alpha_\ell$, we obtain

$$I\left(\boldsymbol{X}_{-\infty}^{-1};\tilde{\boldsymbol{Y}}_b\,\middle|\,\boldsymbol{X}_0^b\right) \leq \frac{L}{2}\log\left(1+L\,\mathrm{SNR}\sum_{\ell=b+1}^{\infty}\alpha_\ell\right) \tag{113}$$

and (109) follows thus by noting that (23) implies

$$\lim_{b\to\infty}\sum_{\ell=b+1}^{\infty}\alpha_i = 0.$$

Appendix C
Proof of Lemma 5

We show that for any $\epsilon > 0$

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2-\left(\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}\right)\right|>\epsilon\right)=0 \tag{114}$$

and

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2-\left(\sigma^2+\alpha^{(L)}\mathsf{P}\right)\right|>\epsilon\right)=0. \tag{115}$$

Lemma 5 follows then by the union of events bound.

In order to prove (114) and (115), we first note that

$$\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Y}\|^2] = \sigma^2+\mathsf{P}+\frac{\mathsf{P}}{\lfloor n/L\rfloor}\sum_{k=1}^{\lfloor n/L\rfloor-1}\sum_{\ell=1}^{k}\alpha_{\ell L} \tag{116}$$

$$\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Z}\|^2] = \sigma^2+\frac{\mathsf{P}}{\lfloor n/L\rfloor}\sum_{k=1}^{\lfloor n/L\rfloor-1}\sum_{\ell=1}^{k}\alpha_{\ell L} \tag{117}$$

and hence, by Cesáro's mean [18, Theorem 4.2.3]

$$\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Y}\|^2] = \sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P} \tag{118}$$

$$\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Z}\|^2]=\sigma^2+\alpha^{(L)}\mathsf{P} \tag{119}$$

where $\alpha^{(L)}$ was defined in (88) as

$$\alpha^{(L)}=\sum_{\ell=1}^{\infty}\alpha_{\ell L}.$$

Thus, for any $\epsilon>0$ and $0<\varepsilon<\epsilon$ there exists an n_0 such that for all $n\geq n_0$

$$\left|\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Y}\|^2]-\left(\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}\right)\right|\leq\varepsilon \tag{120}$$

$$\left|\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Z}\|^2]-\left(\sigma^2+\alpha^{(L)}\mathsf{P}\right)\right|\leq\varepsilon \tag{121}$$

and it follows from the triangle inequality that

$$\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2-\left(\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}\right)\right| \leq\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Y}\|^2]\right|+\varepsilon \tag{122}$$

and

$$\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2-\left(\sigma^2+\alpha^{(L)}\mathsf{P}\right)\right| \leq\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Z}\|^2]\right|+\varepsilon. \tag{123}$$

This yields

$$\begin{aligned}\Pr&\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2-\left(\sigma^2+\mathsf{P}+\alpha^{(L)}\mathsf{P}\right)\right|>\epsilon\right)\\ &\leq\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Y}\|^2]\right|>\epsilon-\varepsilon\right)\\ &\leq\frac{\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2\right)}{(\epsilon-\varepsilon)^2}\end{aligned} \tag{124}$$

and

$$\begin{aligned}\Pr&\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2-\left(\sigma^2+\alpha^{(L)}\mathsf{P}\right)\right|>\epsilon\right)\\ &\leq\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}[\|\boldsymbol{Z}\|^2]\right|>\epsilon-\varepsilon\right)\\ &\leq\frac{\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2\right)}{(\epsilon-\varepsilon)^2}\end{aligned} \tag{125}$$

where $\mathsf{Var}(A)=\mathsf{E}[(A-\mathsf{E}[A])^2]$. Here, the last inequalities in (124) and (125) follow from Chebyshev's inequality [24, Section 5.4].

It remains to show that

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2\right)=0 \tag{126}$$

and

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Z}\|^2\right)=0. \tag{127}$$

We shall prove (126). The proof of (127) follows along the same lines. We begin by writing the variance as

$$\begin{aligned}\mathsf{Var}&\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2\right)\\ &=\frac{1}{\lfloor n/L\rfloor^2}\mathsf{Var}\left(\sum_{k=0}^{\lfloor n/L\rfloor-1}Y_{kL+1}^2\right)\\ &=\frac{1}{\lfloor n/L\rfloor^2}\sum_{k=0}^{\lfloor n/L\rfloor-1}\mathsf{Var}\left(Y_{kL+1}^2\right)\\ &\quad+\frac{2}{\lfloor n/L\rfloor^2}\sum_{k>j}\mathsf{Cov}\left(Y_{kL+1}^2,Y_{jL+1}^2\right)\end{aligned} \tag{128}$$

where $\mathsf{Cov}(A,B)=\mathsf{E}[(A-\mathsf{E}[A])(B-\mathsf{E}[B])]$. We evaluate both terms on the RHS of (128) individually. For the sake of clarity, we shall omit the details and show only the main steps. Unless otherwise stated, these steps can be derived in a straightforward way using that

i) $\{X_{kL+1},k\in\mathbb{N}_0\}$ is a sequence of i.i.d., zero-mean, variance-P, Gaussian random variables;
ii) the fourth moment of a zero-mean, variance-P, Gaussian random variable is given by $3\mathsf{P}$, and all odd moments are zero;
iii) $X_k=0$ for $k \bmod L\neq 1$;
iv) $\{U_k\}$ (and hence also $\{U_{kL+1},k\in\mathbb{N}_0\}$) is a zero-mean, unit-variance, stationary, weakly mixing random process;
v) $\{X_k\}$ and $\{U_k\}$ are independent of each other.

For the first sum on the RHS of (128) it suffices to show that $\mathsf{Var}(Y_{kL+1}^2)<\infty$, $k\in\mathbb{N}_0$. Indeed, this sum contains only $\lfloor n/L\rfloor$ summands and hence, if $\mathsf{Var}(Y_{kL+1}^2)<\infty$, then its ratio to $\lfloor n/L\rfloor^2$ vanishes as n tends to infinity. We have

$$\begin{aligned}\mathsf{Var}&\left(Y_{kL+1}^2\right)\\ &=\mathsf{E}\left[Y_{kL+1}^4\right]-\left(\mathsf{E}\left[Y_{kL+1}^2\right]\right)^2\\ &\leq\mathsf{E}\left[Y_{kL+1}^4\right]\\ &=\mathsf{E}\left[\left(X_{kL+1}+\theta\left(X_1^{kL}\right)U_{kL+1}\right)^4\right]\\ &=3\mathsf{P}^2+6\mathsf{P}\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\\ &\quad+\left(\sigma^4+2\sigma^2\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right.\\ &\qquad\left.+2\mathsf{P}^2\sum_{\ell=1}^{k}\alpha_{\ell L}^2+\mathsf{P}^2\left(\sum_{\ell=1}^{k}\alpha_{kL}\right)^2\right)\mathsf{E}\left[U_{kL+1}^4\right]\\ &\leq3\mathsf{P}^2+6\mathsf{P}\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)\\ &\quad+\left(\sigma^4+2\sigma^2\mathsf{P}\alpha^{(L)}\right.\\ &\qquad\left.+2\mathsf{P}^2\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2+\mathsf{P}^2\left(\alpha^{(L)}\right)^2\right)\mathsf{E}\left[U_{kL+1}^4\right]\end{aligned} \tag{129}$$

where the last step follows by upper-bounding $\sum_{\ell=1}^{k}\alpha_{\ell L}$ by $\alpha^{(L)}$ and $\sum_{\ell=1}^{k}\alpha_{\ell L}^2$ by $\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2$. Note that (83) implies that

$$\alpha^{(L)}<\infty\quad\text{and}\quad\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2<\infty.$$

By additionally noting that U_{kL+1} has a finite fourth moment (9), it follows that (for any finite P)

$$\mathsf{Var}\left(Y_{kL+1}^2\right) < \infty, \quad k \in \mathbb{N}_0. \tag{130}$$

In order to show that the second term on the RHS of (128) vanishes as n tends to infinity, we shall evaluate

$$\begin{aligned}\mathsf{Cov}\left(Y_{kL+1}^2, Y_{jL+1}^2\right) \\ = \mathsf{E}\left[Y_{kL+1}^2 Y_{jL+1}^2\right] - \mathsf{E}\left[Y_{kL+1}^2\right]\mathsf{E}\left[Y_{jL+1}^2\right], \quad k > j.\end{aligned} \tag{131}$$

We have

$$\begin{aligned}\mathsf{E}\left[Y_{kL+1}^2 Y_{jL+1}^2\right] \\ = \mathsf{P}^2 + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j}\alpha_{\ell L}\right) \\ + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right) + 2\mathsf{P}^2\alpha_{(k-j)L} \\ + \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right) \\ \times \mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] \\ + 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right]\end{aligned} \tag{132}$$

and

$$\begin{aligned}\mathsf{E}\left[Y_{kL+1}^2\right]\mathsf{E}\left[Y_{jL+1}^2\right] \\ = \mathsf{P}^2 + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j}\alpha_{\ell L}\right) + \mathsf{P}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right) \\ + \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right).\end{aligned} \tag{133}$$

Equations (131), (132), and (133) thus yield

$$\begin{aligned}\mathsf{Cov}\left(Y_{kL+1}^2, Y_{jL+1}^2\right) \\ = 2\mathsf{P}^2\alpha_{(k-j)L} + 2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] \\ + \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right) \\ \times \left(\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] - 1\right).\end{aligned} \tag{134}$$

We continue by summing $\mathsf{Cov}(Y_{kL+1}^2, Y_{jL+1}^2)$ over (k,j)

$$\begin{aligned}\sum_{k>j}\mathsf{Cov}\left(Y_{kL+1}^2, Y_{jL+1}^2\right) \\ = \sum_{k>j}2\mathsf{P}^2\alpha_{(k-j)L} \\ + \sum_{k>j}2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] \\ + \sum_{k>j}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right) \\ \times \left(\mathsf{E}\left[U_{kL+1}^2 U_{jL+1}^2\right] - 1\right) \\ = \sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}2\mathsf{P}^2\alpha_{\nu L} \\ + \sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+\nu)L}\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] \\ + \sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right) \\ \times \left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right)\end{aligned} \tag{135}$$

where the second step follows by substituting $\nu = k - j$ and from the stationarity of $\{U_k\}$.

The first two terms on the RHS of (135) can be upper-bounded using (84), namely

$$\alpha_\ell < \varrho^\ell, \quad (0 < \varrho < 1, \quad \ell \geq \ell_0).$$

Indeed, by noting that $L \geq \ell_0$, this yields

$$\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\alpha_{\nu L} < \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\varrho^{\nu L} < \sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L} \tag{136}$$

and

$$\begin{aligned}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+\nu)L} &< \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\sum_{\ell=1}^{j}\left(\varrho^{2L}\right)^\ell\varrho^{\nu L} \\ &< \sum_{\nu=1}^{\lfloor n/L\rfloor}\sum_{\ell=1}^{\infty}(\varrho^{2L})^\ell\varrho^{\nu L} \\ &= \frac{\varrho^{2L}}{1-\varrho^{2L}}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}.\end{aligned} \tag{137}$$

Applying (136) we can upper-bound the first term on the RHS of (135) by

$$\begin{aligned}\frac{2}{\lfloor n/L\rfloor^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}2\mathsf{P}^2\alpha_{\nu L} \\ < \frac{4\mathsf{P}^2}{\lfloor n/L\rfloor^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L} \\ = 4\mathsf{P}^2\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}.\end{aligned} \tag{138}$$

Likewise, applying (137) we can upper-bound the second term on the RHS of (135) by

$$\begin{aligned}\frac{2}{\lfloor n/L\rfloor^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+\nu)L}\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] \\ \leq \frac{4\mathsf{P}^2}{\lfloor n/L\rfloor^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+\nu)L}\mathsf{E}\left[U_1^4\right] \\ < 4\mathsf{P}^2\frac{\varrho^{2L}}{1-\varrho^{2L}}\mathsf{E}\left[U_1^4\right]\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}\end{aligned} \tag{139}$$

where the first inequality follows from the Cauchy–Schwarz inequality.

As for the last term on the RHS of (135), we upper-bound each summand by

$$\begin{aligned}&\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2+\mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right)\\&\leq\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2+\mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right|\\&\leq\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)^2\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right| \qquad (140)\end{aligned}$$

where the first inequality follows by upper-bounding $\mathsf{E}[U_{\nu L+1}^2U_1^2]-1 \leq |\mathsf{E}[U_{\nu L+1}^2U_1^2]-1|$; and the second inequality follows by upper-bounding

$$\sum_{\ell=1}^{j}\alpha_{\ell L}\leq\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\leq\sum_{\ell=1}^{\infty}\alpha_{\ell L}=\alpha^{(L)}.$$

Applying (138), (139), and (140) to (135) yields

$$\begin{aligned}&\frac{2}{\lfloor n/L\rfloor^2}\sum_{k>j}\mathsf{Cov}\left(Y_{kL+1}^2,Y_{jL+1}^2\right)\\&<4\mathsf{P}^2\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}\\&\quad+4\mathsf{P}^2\frac{\varrho^{2L}}{1-\varrho^{2L}}\mathsf{E}\left[U_1^4\right]\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}\\&\quad+\frac{2}{(\lfloor n/L\rfloor)^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)^2\\&\qquad\times\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right|\\&\leq4\mathsf{P}^2\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}\\&\quad+4\mathsf{P}^2\frac{\varrho^{2L}}{1-\varrho^{2L}}\mathsf{E}\left[U_1^4\right]\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\varrho^{\nu L}\\&\quad+2\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)^2\frac{\lfloor n/L\rfloor-1}{\lfloor n/L\rfloor}\\&\qquad\times\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right|. \qquad (141)\end{aligned}$$

Here, the second step follows by upper-bounding

$$\begin{aligned}&\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j}\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)^2\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right|\\&\qquad\leq\sum_{\nu=1}^{\lfloor n/L\rfloor}\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)^2\left|\mathsf{E}\left[U_{\nu L+1}^2U_1^2\right]-1\right|.\end{aligned}$$

By Cesáro's mean [18, Theorem 4.2.3] the first two terms on the RHS of (141) tend to zero as n tends to infinity, and by the weakly mixing property of $\{U_k\}$ the third term on the RHS of (141) tends to zero as n tends to infinity [11, Theorem 6.1]. It thus follows from (128), (130), and (141) that

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\boldsymbol{Y}\|^2\right)=0.$$

Together with (124) this proves (114). The proof of (115) follows along the same lines.

Acknowledgment

The authors wish to thank Ashish Khisti and Michèle Wigger for helpful discussions and Sergio Verdú for his comments on our low-SNR results.

References

[1] R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, and J. D. Meindl, "Interconnect limits on gigascale integration (GSI) in the 21st century," *Proc. IEEE*, vol. 89, no. 3, pp. 305–324, Mar. 2001.
[2] B. E. S. Akgul, L. N. Chakrapani, P. Korkmaz, and K. V. Palem, "Probabilistic CMOS technology: A survey and future directions," in *Proc. IFIP Int. Conf. on VLSI*, Nice, France, Oct. 2006, pp. 1–6.
[3] L. B. Kish, "End of Moore's law: Thermal (noise) death of integration in micro and nano electronics," *Phys. Lett. A*, no. 3–4, pp. 144–149, Dec. 2002.
[4] K. E. Goodson, "Thermal conduction in electronic microstructures," in *CRC Handbook of Thermal Engineering*, ser. Mechanical Engineering Handbook Series, F. Keith, Ed., 1st ed. Boca Raton, FL: CRC, Dec. 1999.
[5] J. H. Lienhard IV and J. H. Lienhard V, *A Heat Transfer Textbook*, 3rd ed. Cambridge, MA: Phlogisten, 2008.
[6] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2003.
[7] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 186–201, Feb. 2000.
[8] B. Razavi, "CMOS technology characterization for analog and RF design," *IEEE J. Solid-State Circuits*, vol. 34, no. 3, pp. 268–276, Mar. 1999.
[9] S. E. Lyshevski, *Nano and Molecular Electronics Handbook*, 1st ed. Boca Raton, FL: CRC, 2007.
[10] M. M. Ziegler and M. R. Stan, "The CMOS/nano interface from a circuits perspective," in *Proc. Int. Symp. Circuits and Systems (ISCAS)*, Bangkok, Thailand, May 2003, pp. 904–907.
[11] K. Petersen, *Ergodic Theory*, ser. Cambridge Studies in Advanced Mathematics 2. Cambridge, U.K.: Cambridge Univ. Press, 1983.
[12] G. Maruyama, "The harmonic analysis of stationary stochastic processes," *Memoirs of the Faculty of Science*, ser. A, vol. 4, no. 1, pp. 45–106, 1949.
[13] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.
[14] T. Koch and A. Lapidoth, "On multipath fading channels at high SNR," *IEEE Trans. Inf. Theory* [Online]. Available: http://arxiv.org/abs/0901.0521, submitted for publication.
[15] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[16] J. Chen and V. V. Veeravalli, "Capacity results for block-stationary gaussian fading channels with a peak-power constraint," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4498–4520, Dec. 2007.
[17] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.
[19] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
[20] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.
[21] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
[22] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side-information' be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
[23] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1520–1529, Sep. 1996.
[24] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

Tobias Koch (S'02–M'09) received the M.Sc. degree in electrical engineering (with distinction) from ETH Zurich, Zurich, Switzerland, in 2004.

He is currently working toward the Ph.D. degree at ETH Zurich. His research interests include digital communication theory and information theory with particular focus on wireless communication and communication in electronic circuits.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in mathematics (*summa cum laude*, 1986), the B.Sc. degree in electrical engineering (*summa cum laude*, 1986), and the M.Sc. degree in electrical engineering (1990) all from the Technion–Israel Institute of Technology, Haifa, Israel. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1995.

In the years 1995–1999, he was an Assistant and Associate Professor at the department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, Cambridge, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. His research interests are in digital communications and information theory.

Dr. Lapidoth served as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY during 2003–2004 .

Paul P. Sotiriadis (S'99–M'02–SM'09) received the diploma in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 1994, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1996, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 2002.

In 2002 he joined The Johns Hopkins University, Baltimore, MD, as an Assistant Professor of Electrical and Computer Engineering and in 2007 he joined Apex/Eclipse INC as the Chief Technology Officer. Shortly after that, he started Sotekco Electronics LLC, an electronics research company in Maryland. His research interests include design, modeling, and optimization of analog, mixed-signal, RF, and microwave circuits, advanced frequency synthesis, biomedical instrumentation, and, interconnect networks in deep-submicrometer technologies. He has led several projects in these fields funded by US organizations and has collaborations with industry, academia, and National laboratories. He has authored and coauthored over seventy technical papers in IEEE journals and conferences, has one patent and several patents pending and has contributed chapters to technical books.

Dr. Sotiriadis serves as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II. He is and has been a member of the technical committees of several conferences. He regularly reviews for many IEEE TRANSACTIONS and conferences. He also serves on proposal review panels at the National Science Foundation.