Channels that Heat Up

Tobias Koch Amos Lapidoth
ETH Zurich
Zurich, Switzerland
Email: {tkoch, lapidoth}@isi.ee.ethz.ch

Paul P. Sotiriadis
Johns Hopkins University
Baltimore, MD, USA
Email: pps@jhu.edu

Abstract

This work considers an additive noise channel where the time-k noise variance is a weighted sum of the channel input powers prior to time k. This channel is motivated by point-to-point communication between two terminals that are embedded in the same chip. Transmission heats up the entire chip and hence increases the thermal noise at the receiver. 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-vs-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.

1 Introduction

Thermal heating in electronic systems is strongly related to performance limitation, aging, reliability and safety issues. High performance-density and small physical size (area or volume) make thermal heating important and challenging to address. This is enhanced 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; more accurately, the temperature depends on the circuit activity. The temperature influences the power of the intrinsic noise in the circuit which in turn reduces the effective communication or computation capacity of the circuit. This "negative" performance feedback is expected to become a bottleneck of future technology [1], [2].

This work aims to add this dimension to our understanding of the coupling mechanism between communication and computation performance and thermal heating. To this end a class of communication channels is introduced, where the channel's noise power depends dynamically on the channel's activity, and its channel capacity is studied.

To support the previous statements and 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 thermal heating. Thermal heating is unavoidable in electronic circuits. Every circuit block converts part of the power it draws from the power supply network (and to certain extent from its interconnections with other blocks) into heat which raises the local temperature.

A circuit block in a microchip occupies certain physical space within which heat is distributively generated and diffused according to the *heat diffusion equation* (ignoring other heat sources)

$$\mathsf{C}_{\mathrm{hv}} \frac{\partial T}{\partial t} = \nabla \cdot \left(\frac{1}{\rho_{\mathrm{thd}}} \nabla T \right) + \mathsf{E}' \tag{1}$$

The material in this paper was presented in part at the 2007 IEEE International Symposium on Information Theory (ISIT), Nice, France, and at the 2007 IEEE Information Theory Workshop (ITW), Lake Tahoe, CA, USA.

where C_{hv} is the volumetric heat capacity of the material, $\partial T/\partial t$ is the change in temperature over time, $\nabla\cdot$ is the divergence, ρ_{thd} is the distributed thermal resistance, ∇T is the temperature gradient, and E' is the power density of the added heat, [3], [4].

In many cases the diffusion equation can be replaced by the corresponding *ordinary differential equation* (ODE) that provides a lumped model of the thermal dynamics. Consider for example a microchip (die), made out of material of lower thermal resistance, which is internally heated by the activity of circuits and transfers the heat to the environment (e.g., air) which has much higher resistance. In this case we can write

$$\mathsf{C}_{\mathrm{h}}\frac{dT}{dt} = \frac{T_{\mathrm{e}} - T}{\rho_{\mathrm{th}}} + \mathsf{E} \tag{2}$$

where C_{h} is the heat capacity of the microchip (die), ρ_{th} is the thermal resistance between the die and the environment (e.g., air), T_{e} is the temperature of the environment, and E is the instantaneous heat generated, i.e., the electrical power converted into heat by the circuit.

Solving (2) with the assumption that at time $t = 0$ we have $T = T_e$ with T_{e} being fixed, we obtain

$$T(t) = T_{\mathrm{e}} + \frac{1}{\mathsf{C}_{\mathrm{h}}}\int_0^t e^{\frac{\xi - t}{\rho_{\mathrm{th}}\mathsf{C}_{\mathrm{h}}}}\mathsf{E}(\xi)d\xi, \qquad t \in \mathbb{R}. \tag{3}$$

If the circuit operates based on a reference clock of period τ, (3) can be approximated by its discrete version

$$T_k = T_{\mathrm{e}} + \sum_{\ell=1}^{k-1}\frac{\tau}{\mathsf{C}_{\mathrm{h}}}e^{-\frac{\tau}{\rho_{\mathrm{th}}\mathsf{C}_{\mathrm{h}}}(k-\ell)}\mathsf{E}_\ell, \qquad k \in \mathbb{Z}^+, \tag{4}$$

where $\mathbb{Z}^+$ denotes the set of positive integers, and where the sequences $\{T_k\}$ and $\{\mathsf{E}_k\}$ are the samples at integer multiples of τ of $T(\cdot)$ and $\mathsf{E}(\cdot)$, respectively. Equation (4) shows the fading memory effect of temperature. Note that (4) also captures discrete versions of distributed or higher order lumped approximations of the diffusion equation (1).

Every electronic circuit has some intrinsically generated noise. This noise is added to the received signal degrading its quality. Especially in the popular class of circuits based on MOS transistors [5], this noise is dominated by a thermal noise component that is stationary Gaussian, and in most applications it can be considered white. The variance of the thermal noise N follows the Johnson-Nyquist formula

$$\mathsf{N} = \lambda T W \tag{5}$$

where W is the considered bandwidth, T is the temperature of the receiver circuit block, and λ is a proportionality constant [5], [6], [7].

The transmission of information is typically associated with dissipation of energy into heat. Thus, in view of (4) and (5), this motivates 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.,

$$\theta^2(x_1, \ldots, x_{k-1}) = \sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}x_\ell^2, \qquad k \in \mathbb{Z}^+, \tag{6}$$

where x_ℓ is the transmitted symbol at time $\ell \in \mathbb{Z}^+$, and where σ^2 and $\{\alpha_\ell\}$ will be defined in Section 2.

The rest of this paper is organized as follows. Section 2 describes the channel model in more detail. Section 3 discusses channel capacity and lists some important properties thereof. The main results are presented in Section 4. The proofs of the results are given in Sections 5 and 6. Section 7 concludes with a summary.

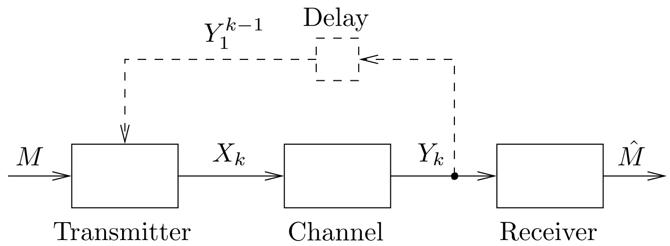

Figure 1: A schema of the communication system.

2 Channel Model

We consider the communication system depicted in Figure 1. The message M to be transmitted over the channel is assumed to be 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 the *block-length*. 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 $A_m, \ldots, A_n$, and $\mathbb{R}$ denotes the set of real numbers. 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 $X_1 = x_1, \ldots, X_k = x_k \in \mathbb{R}$, the time-$k$ channel output $Y_k \in \mathbb{R}$ is given by

$$Y_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=1}^{k-1} \alpha_{k-\ell} x_\ell^2\right)} \cdot U_k, \qquad k \in \mathbb{Z}^+, \tag{7}$$

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.,

$$\mathsf{E}\big[U_k^4\big] < \infty \qquad \text{and} \qquad h\big(U_k \big| U_{-\infty}^{k-1}\big) > -\infty. \tag{8}$$

See [8] for a definition of weak mixing. For example, $\{U_k\}$ could be a stationary & ergodic Gaussian process [9]. In particular, the case of most interest is when $\{U_k\}$ are independent and identically distributed (IID), 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, $\ell \in \mathbb{Z}^+$ are nonnegative and bounded, i.e.,

$$\alpha_\ell \geq 0, \quad \ell \in \mathbb{Z}^+ \qquad \text{and} \qquad \sup_{\ell \in \mathbb{Z}^+} \alpha_\ell < \infty. \tag{9}$$

They characterize the dissipation of the heat produced by the transmission of the message M.[1]

An example for a heat dissipation profile that satisfies (9) is the *geometric* heat dissipation profile where $\{\alpha_\ell\}$ is a geometric sequence, i.e.,

$$\alpha_\ell = \rho^\ell, \qquad \ell \in \mathbb{Z}^+ \tag{10}$$

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 (10), 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.

[1] It seems reasonable to assume that the sequence $\{\alpha_\ell\}$ is monotonically nonincreasing, i.e., $\alpha_\ell \leq \alpha_{\ell'}$ for $\ell \geq \ell'$. This assumption is, however, not required for the results stated in this paper.

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 such that—averaged over the message M and channel outputs Y_1^n—the sequence X_1^n satisfies

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[X_k^2\right] \leq \mathsf{P}, \tag{11}$$

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

$$\text{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}. \tag{12}$$

Remark 1. *The results presented in this paper do not change when* (11) *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 such 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*

$$\frac{1}{n}\sum_{k=1}^{n} x_k^2 \leq \mathsf{P}. \tag{13}$$

Indeed, all achievability results (which are based on schemes that ignore the feedback) are derived under (13)*, whereas all converse results are derived under* (11)*. Since all mappings ϕ_n and $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ that satisfy* (13) *also fulfill* (11)*, this implies that the achievability results as well as the converse results derived in this paper hold irrespective of whether constraint* (11) *or* (13) *is imposed.*

3 Channel Capacity

Let the *rate* R (in nats per channel use) be defined as

$$R \triangleq \frac{\log|\mathcal{M}|}{n}, \tag{14}$$

where $\log(\cdot)$ denotes the natural logarithm function. A rate is said to be *achievable* if there exists a sequence of mappings $\{\phi_n\}$ (without feedback) or $\{(\varphi_n^{(1)}, \ldots, \varphi_n^{(n)})\}$ (with feedback) and $\{\psi_n\}$ such that the error probability $\Pr(\hat{M} \neq M)$ tends to zero as n goes to infinity. The *capacity* C is the supremum of all achievable rates. We denote by $C(\text{SNR})$ the capacity under the input constraint (11) when there is no feedback, and we add the subscript "FB" to indicate that there is a feedback link. Clearly

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

as we can always ignore the feedback link.

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

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

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (11). When there is a feedback link, then we define the information capacity as

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

where the supremum is over all mappings $\varphi_n^{(1)}, \ldots, \varphi_n^{(n)}$ satisfying (11). By Fano's inequality [10, Thm. 2.11.1] no rate above $C_{\text{Info}}(\text{SNR})$ and $C_{\text{Info,FB}}(\text{SNR})$ is achievable, i.e.,

$$C(\text{SNR}) \leq C_{\text{Info}}(\text{SNR}) \qquad \text{and} \qquad C_{\text{FB}}(\text{SNR}) \leq C_{\text{Info,FB}}(\text{SNR}). \tag{18}$$

See [11] for conditions that guarantee that $C_{\text{Info}}(\text{SNR})$ is achievable. Note that the channel (7) is not stationary[2] 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) hold with equality.

In this paper, we shall investigate the capacities $C(\text{SNR})$ and $C_{\text{FB}}(\text{SNR})$ at low SNR and at high SNR. To study capacity at low SNR, we compute the *capacities per unit cost* defined as [12]

$$\dot{C}(0) \triangleq \sup_{\text{SNR}>0} \frac{C(\text{SNR})}{\text{SNR}} \qquad \text{and} \qquad \dot{C}_{\text{FB}}(0) \triangleq \sup_{\text{SNR}>0} \frac{C_{\text{FB}}(\text{SNR})}{\text{SNR}}. \tag{19}$$

It will become apparent later that the suprema in (19) are attained when SNR tends to zero. Note that (15) implies

$$\dot{C}(0) \leq \dot{C}_{\text{FB}}(0). \tag{20}$$

At high SNR, we study conditions under which capacity is unbounded in the SNR. Notice that when the allowed transmit power is large, then there is a trade-off 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. *Prima facie* it is not clear that, as we increase the allowed transmit power, the capacity tends to infinity. We shall see that this is not necessarily the case.

4 Main Results

Our main results are presented in the following two sections. Section 4.1 focuses on capacity at low SNR and presents our results on the capacity per unit cost. Section 4.2 provides a sufficient condition and a necessary condition on $\{\alpha_\ell\}$ under which capacity is bounded in the SNR.

4.1 Capacity per Unit Cost

The results presented in this section hold under the additional assumptions that

$$\sum_{\ell=1}^{\infty} \alpha_\ell \triangleq \alpha < \infty \tag{21}$$

and that $\{U_k\}$ is IID.

Proposition 1. *Consider the above channel model, and assume additionally that the sequence $\{\alpha_\ell\}$ satisfies* (21) *and that $\{U_k\}$ is IID. Then*

$$\sup_{\text{SNR}>0} \frac{C_{\text{Info}}(\text{SNR})}{\text{SNR}} \geq \sup_{\text{SNR}>0} \frac{C_{\alpha=0}(\text{SNR})}{\text{SNR}}, \tag{22}$$

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

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

which is a special case of (7) *for $\alpha = 0$.*

Proof. See Appendix A. □

This proposition demonstrates that the heating up can only increase the information capacity per unit cost. Thus at low SNR the heating effect is unharmful.

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

[2]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.

Theorem 2. *Consider the above channel model, and assume additionally that the sequence $\{\alpha_\ell\}$ satisfies* (21) *and that $\{U_k\}$ is a sequence of IID, zero-mean, unit-variance Gaussian random variables. Then, irrespective of whether feedback is available or not, the corresponding capacity per unit cost is given by*

$$\dot{C}_{\mathrm{FB}}(0) = \dot{C}(0) = \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} = \frac{1}{2}\left(1 + \sum_{\ell=1}^{\infty} \alpha_\ell\right). \tag{23}$$

Proof. See Section 5. □

For example, for the geometric heat dissipation profile (10) we obtain from Theorem 2

$$\dot{C}_{\mathrm{FB}}(0) = \dot{C}(0) = \frac{1}{2}\frac{1}{1-\rho}, \qquad 0 < \rho < 1. \tag{24}$$

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 effect increases the *channel gain*[3]. 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 IID, zero-mean, variance-σ^2, Gaussian random variables drawn independently of $\{(U_k, X_k)\}$), then the channel gain G for (7) is given by

$$\mathsf{G} \triangleq \lim_{n\to\infty} \sup \frac{\sum_{k=1}^{n} \mathsf{E}\left[\tilde{X}_k^2\right]}{\sum_{k=1}^{n} \mathsf{E}[X_k^2]} = 1 + \sum_{\ell=1}^{\infty} \alpha_\ell, \tag{25}$$

where the supremum is over all joint distributions on $X_1, \ldots, X_n$ satisfying (11). Thus, in view of (25), Theorem 2 demonstrates that the capacity per unit cost is determined by the channel gain G. This result is not specific to (7) but has also been observed for other channel models. For example, the same is true for fading channels whenever the additive noise is Gaussian [13], [14].

4.2 Conditions for Bounded Capacity

While at low SNR the heating effect is beneficial, at high SNR it is detrimental. In fact, it turns out that 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 4.1: we neither assume that the sequence $\{\alpha_\ell\}$ satisfies (21) nor that $\{U_k\}$ is IID.

Theorem 3. *Consider the channel model described in Section 2. Then*

$$\text{i)} \quad \left(\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0\right) \quad \Longrightarrow \quad \left(\sup_{\mathrm{SNR}>0} C_{\mathrm{FB}}(\mathrm{SNR}) < \infty\right) \tag{26}$$

$$\text{ii)} \quad \left(\varlimsup_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} = 0\right) \quad \Longrightarrow \quad \left(\sup_{\mathrm{SNR}>0} C(\mathrm{SNR}) = \infty\right), \tag{27}$$

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

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

Proof. See Section 6. □

For example, for a geometric heat dissipation (10) we have

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

and it follows from Theorem 3 that the corresponding capacity is bounded. On the other hand, for a sub-geometric heat dissipation, i.e.,

$$\alpha_\ell=\rho^{\ell^\kappa},\qquad \ell\in\mathbb{Z}^+$$

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

Remark 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 in the proof of Part ii).*[4] *In Part ii) of Theorem 3, the condition on the left-hand side (LHS) of* (27) *can be replaced by*

$$\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell}=\infty. \tag{28}$$

This condition (28) *is weaker than the original condition* (27) *because*

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

When neither the LHS of (26) nor the LHS of (27) hold, i.e.,

$$\varlimsup_{\ell\to\infty}\frac{\alpha_{\ell+1}}{\alpha_\ell}>0\qquad\text{and}\qquad\varliminf_{\ell\to\infty}\frac{\alpha_{\ell+1}}{\alpha_\ell}=0, \tag{29}$$

then capacity can be bounded or unbounded. Example 1 exhibits a sequence $\{\alpha_\ell\}$ satisfying (29) for which the capacity is bounded, and Example 2 provides a sequence $\{\alpha_\ell\}$ satisfying (29) for which the capacity is unbounded.[5]

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* (29) *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*

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

where $\lfloor\cdot\rfloor$ denotes the floor function. Thus at even times the output Y_{2k}, $k\in\mathbb{Z}^+$ only depends on the "even" inputs $(X_2,X_4,\ldots,X_{2k})$, while at odd times the output Y_{2k+1}, $k\in\mathbb{Z}_0^+$ only depends on the "odd" inputs $(X_1,X_3,\ldots,X_{2k+1})$. By proceeding along the lines of the proof of Part i) of Theorem 3 while choosing in (60) *$\beta=1/y_{k-2}^2$, it can be shown that the capacity of this channel is bounded.*[6]

[4] They are needed to prove Lemma 5.

[5] The provided sequences $\{\alpha_\ell\}$ are not monotonically decreasing in ℓ. Consequently, Examples 1 & 2 are rather of mathematical than of practical interest. Nevertheless they show that when neither condition of Theorem 3 is satisfied, then 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.

[6] Intuitively, with this choice of $\{\alpha_\ell\}$ the channel can be divided into two parallel channels, one connecting the inputs and outputs at even times, and the other connecting the inputs and outputs at odd times. As both channels have the coefficients $\tilde{\alpha}_0=\tilde{\alpha}_1=\ldots=1$, it follows from Theorem 3 that the capacity of each parallel channel is bounded and therefore also the capacity of the original channel.

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 $\overline{\lim}_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = \infty$ and $\underline{\lim}_{\ell\to\infty} \alpha_{\ell+1}/\alpha_\ell = 0$.) In this case the time-k channel output Y_k corresponding to $(x_1, \ldots, x_k)$ is given by*

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

Using Gaussian inputs of power $2\mathsf{P}$ *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 [15]) 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.

5 Proof of Theorem 2

In Section 5.1 we derive an upper bound on the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$, and in Section 5.2 we derive a lower bound on the capacity $C(\mathrm{SNR})$ in the absence of feedback. These bounds are used in Section 5.3 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 (20) this proves Theorem 2.

5.1 Converse

The upper bound on $C_{\mathrm{FB}}(\mathrm{SNR})$ is based on (18) and on an upper bound on $\frac{1}{n}I(M;Y_1^n)$, which for our channel can be expressed, using the chain rule for mutual information, as

$$\begin{aligned}
\frac{1}{n}I(M;Y_1^n) &= \frac{1}{n}\sum_{k=1}^{n}\Big(h\big(Y_k\big|Y_1^{k-1}\big) - h\big(Y_k\big|Y_1^{k-1},M\big)\Big) \\
&= \frac{1}{n}\sum_{k=1}^{n}\Big(h\big(Y_k\big|Y_1^{k-1}\big) - h\big(Y_k\big|Y_1^{k-1},M,X_1^k\big)\Big) \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(h\big(Y_k\big|Y_1^{k-1}\big) - h(U_k) - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]\right), \qquad (30)
\end{aligned}$$

where the second equality follows because X_1^k is a function of M and Y_1^{k-1}; and the last equality follows from the behavior of differential entropy under translation and scaling [10, Thms. 9.6.3 & 9.6.4], and because U_k is independent of $\big(Y_1^{k-1}, M, X_1^k\big)$.

Evaluating the differential entropy $h(U_k)$ of a Gaussian random variable, and using the trivial lower bound $\mathsf{E}\Big[\log\Big(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\Big)\Big] \geq \log\sigma^2$, we obtain the final upper bound

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

$$
\begin{aligned}
&\le \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{k=1}^{n}\sum_{\ell=1}^{k}\alpha_{k-\ell}\mathsf{E}\left[X_\ell^2\right]/\sigma^2\right)\\
&= \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\sum_{\ell=0}^{n-k}\alpha_\ell\right)\\
&\le \frac{1}{2}\log\left(1+(1+\alpha)\,\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\left[X_k^2\right]/\sigma^2\right)\\
&\le \frac{1}{2}\log\left(1+(1+\alpha)\,\mathrm{SNR}\right),
\end{aligned}
\tag{31}
$$

where we define $\alpha_0 \triangleq 1$. Here the second inequality follows because conditioning cannot increase entropy and from the entropy maximizing property of Gaussian random variables [10, Thm. 9.6.5]; the next inequality follows by Jensen's inequality; the following equality by rewriting the double sum; the subsequent inequality follows because the coefficients are nonnegative which implies that $\sum_{\ell=0}^{n-k}\alpha_\ell \le \sum_{\ell=0}^{\infty}\alpha_\ell = 1+\alpha$; and the last inequality follows from the power constraint (11).

5.2 Direct Part

As aforementioned, the above channel (7) 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

$$
\tilde{Y}_k = x_k + \sqrt{\left(\sigma^2 + \sum_{\ell=-\infty}^{k-1}\alpha_{k-\ell}x_\ell^2\right)}\cdot U_k, \qquad k \in \mathbb{Z}^+,
\tag{32}
$$

where $\{U_k\}$ and $\{\alpha_\ell\}$ are defined in Section 2. This channel has the advantage that it is stationary & ergodic in the sense that when $\{X_k\}$ is a stationary & ergodic process then 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 (7) can be converted into (32) by adding

$$
S_k = \sqrt{\left(\sum_{\ell=-\infty}^{0}\alpha_{k-\ell}X_\ell^2\right)}\cdot U_{-k}
$$

to the channel output Y_k,[7] and, since the independence of $\{X_k\,,\ k = 0,-1,\ldots\}$ and $\{X_k\,,\ k = 1,2,\ldots\}$ ensures that the sequence $\{S_k\,,\ k \in \mathbb{Z}^+\}$ is independent of the message M, it follows that any rate achievable over (32) can be achieved over (7) by using a receiver that generates $\{S_k\,,\ k \in \mathbb{Z}^+\}$ and guesses then M based on $(Y_1 + S_1, \ldots, Y_n + S_n)$.[8]

We shall consider channel inputs $\{X_k\}$ that are blockwise IID in blocks of L symbols (for some $L \in \mathbb{Z}^+$). Thus denoting $\mathbf{X}_b = (X_{bL+1}, \ldots, X_{(b+1)L})^{\mathsf{T}}$ (where $(\cdot)^{\mathsf{T}}$ denotes the transpose), $\{\mathbf{X}_b\}$ is a sequence of IID random length-L vectors with $\mathbf{X}_b$ 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 (11) we shall choose ξ and δ so that

$$
\frac{\xi^2}{\sigma^2}\delta = L\,\mathrm{SNR}.
\tag{33}
$$

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

[8] Note that this approach is specific to the case where $\{U_k\}$ is a sequence of Gaussian random variables. Indeed, it relies heavily on the fact that given $\{X_k\} = \{x_k\}$ the additive noise term on the right-hand side of (32) can be written as the sum of two independent random variables, of which one only depends on $\{X_k\,,\ k = 0,-1,\ldots\}$ and the other only on $\{X_k\,,\ k = 1,2,\ldots\}$. This surely holds for Gaussian random variables, but it does not necessarily hold for other distributions on $\{U_k\}$.

Let $\tilde{\mathbf{Y}}_b = (\tilde{Y}_{bL+1}, \ldots, \tilde{Y}_{(b+1)L})^{\mathsf{T}}$. Noting that the pair $\{(\mathbf{X}_b, \tilde{\mathbf{Y}}_b)\}$ is jointly stationary & ergodic, it follows from [11] that the rate

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

is achievable over the new channel (32) and thus yields a lower bound on the capacity $C(\mathrm{SNR})$ of the original channel (7). We lower bound $\frac{1}{n} I\big(\mathbf{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\big)$ as

$$\begin{aligned} \frac{1}{n} I\big(\mathbf{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\big) &= \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1} \big| \mathbf{X}_0^{b-1}\big) \\ &\geq \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_b \big| \mathbf{X}_0^{b-1}\big) \\ &\geq \frac{1}{n} \sum_{b=0}^{\lfloor n/L\rfloor-1} \Big(I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_b \big| \mathbf{X}_{-\infty}^{b-1}\big) - I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \big| \mathbf{X}_0^{b}\big)\Big), \end{aligned} \tag{34}$$

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

$$\lim_{b\to\infty} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \big| \mathbf{X}_0^{b}\big) = 0. \tag{35}$$

This together with a Cesáro type theorem [10, Thm. 4.2.3] yields

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} I\big(\mathbf{X}_0^{\lfloor n/L\rfloor-1}; \tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\big) &\geq \frac{1}{L} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \big| \mathbf{X}_{-\infty}^{-1}\big) - \frac{1}{L} \lim_{n\to\infty} \frac{1}{\lfloor n/L\rfloor} \sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(\mathbf{X}_{-\infty}^{-1}; \tilde{\mathbf{Y}}_b \big| \mathbf{X}_0^{b}\big) \\ &= \frac{1}{L} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \big| \mathbf{X}_{-\infty}^{-1}\big), \end{aligned} \tag{36}$$

where the first inequality follows by the stationarity of $\{(\mathbf{X}_b, \tilde{\mathbf{Y}}_b)\}$ which implies that $I\big(\mathbf{X}_b; \tilde{\mathbf{Y}}_b | \mathbf{X}_{-\infty}^{b-1}\big)$ 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 $I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 | \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\big)$ for a given sequence $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$. Making use of the canonical decomposition of mutual information (e.g., [12, Eq. (10)]), we have

$$\begin{aligned} I\big(\mathbf{X}_0; \tilde{\mathbf{Y}}_0 \big| \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\big) &= I\big(X_1; \tilde{\mathbf{Y}}_0 \big| \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}\big) \\ &= \int D\Big(P_{\tilde{\mathbf{Y}}_0|X_1=x,\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \mathrm{d}P_{X_1}(x) \\ &\quad - D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\ &= \delta D\Big(P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) \\ &\quad - D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}} \Big\| P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big), \end{aligned} \tag{37}$$

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

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

and

$$P_{\tilde{\mathbf{Y}}_0|X_1=\xi,\mathbf{x}_{-\infty}^{-1}}, \quad P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}, \quad \text{and} \quad P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}$$

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

with diagonal entries

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

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

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

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

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

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

In order to evaluate the first term on the right-hand side (RHS) of (37) 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

$$\begin{aligned} D\big(\mathcal{N}(\boldsymbol{\mu}_1,\mathsf{K}_1) \,\|\, \mathcal{N}(\boldsymbol{\mu}_2,\mathsf{K}_2)\big) = {} & \frac{1}{2}\log\det \mathsf{K}_2 - \frac{1}{2}\log\det\mathsf{K}_1 + \frac{1}{2}\operatorname{tr}\left(\mathsf{K}_1\mathsf{K}_2^{-1} - \mathsf{I}_L\right) \\ & + \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^{\mathsf{T}}\mathsf{K}_2^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2), \end{aligned} \tag{38}$$

with $\det \mathsf{A}$ and $\operatorname{tr}(\mathsf{A})$ denoting the determinant and the trace of the matrix A, and where I_L denotes the $L\times L$ identity matrix. The second term on the RHS of (37) is analyzed in the next subsection.

Let $\mathsf{E}\Big[D\big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\big)\Big]$ denote the second term on the RHS of (37) averaged over $\mathbf{X}_{-\infty}^{-1}$, i.e.,

$$\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] = \mathsf{E}_{\mathbf{X}_{-\infty}^{-1}}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big].$$

Then using (38) & (37) and taking expectations over $\mathbf{X}_{-\infty}^{-1}$, we obtain, again defining $\alpha_0 \triangleq 1$,

$$\begin{aligned} \frac{1}{L} I\big(\mathbf{X}_0;\tilde{\mathbf{Y}}_0\big|\mathbf{X}_{-\infty}^{-1}\big) = {} & \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2/\sigma^2}\right] \\ & - \frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\mathsf{E}\left[\log\left(1+\frac{\alpha_{i-1}\xi^2}{\sigma^2+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}X_{\ell L+1}^2}\right)\right] \\ & - \frac{1}{L}\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \\ \geq {} & \frac{\delta}{L}\frac{\xi^2}{\sigma^2}\frac{1}{2}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1}\mathsf{E}\left[X_{\ell L+1}^2\right]/\sigma^2} \\ & - \frac{\delta}{L}\frac{1}{2}\sum_{i=2}^{L}\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right) \\ & - \frac{1}{L}\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \\ \geq {} & \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}} \\ & - \frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2} \\ & - \frac{1}{L}\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big], \end{aligned} \tag{39}$$

where the first inequality follows by the lower bound $\mathsf{E}[1/(1+X)] \geq 1/(1+\mathsf{E}[X])$, which is a consequence of Jensen's inequality applied to the convex function $1/(1+x)$, $x>0$, and by the upper bound

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

and the second inequality follows by (33) and by upper bounding

$$\sum_{\ell=-\infty}^{-1}\alpha_{-\ell L+i-1} \leq \sum_{\ell=1}^{\infty}\alpha_\ell = \alpha, \qquad i=1,\ldots,L.$$

The final lower bound follows now by (39) and (36)

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}I\big(\mathbf{X}_0^{\lfloor n/L\rfloor-1};\tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\big) \geq \frac{1}{2}\mathrm{SNR}\sum_{i=1}^{L}&\frac{\alpha_{i-1}}{1+\alpha\,L\,\mathrm{SNR}}\\ -\frac{1}{2}\mathrm{SNR}\sum_{i=2}^{L}&\frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}\\ -\frac{1}{L}\mathsf{E}&\left[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\right]\end{aligned} \tag{40}$$

and by recalling that

$$C(\mathrm{SNR}) \geq \lim_{n\to\infty}\frac{1}{n}I\big(\mathbf{X}_0^{\lfloor n/L\rfloor-1};\tilde{\mathbf{Y}}_0^{\lfloor n/L\rfloor-1}\big). \tag{41}$$

5.3 Asymptotic Analysis

We start with analyzing the upper bound (31). Using that $\log(1+x)\leq x$, $x>-1$ we have

$$\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{\frac{1}{2}\log(1+(1+\alpha)\,\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha), \tag{42}$$

and we thus obtain

$$\dot{C}_{\mathrm{FB}}(0) = \sup_{\mathrm{SNR}>0}\frac{C_{\mathrm{FB}}(\mathrm{SNR})}{\mathrm{SNR}} \leq \frac{1}{2}(1+\alpha). \tag{43}$$

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

$$\dot{C}(0) = \sup_{\mathrm{SNR}>0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \geq \lim_{\mathrm{SNR}\downarrow 0}\frac{C(\mathrm{SNR})}{\mathrm{SNR}} \tag{44}$$

and proceed by analyzing the limiting ratio of the lower bound (40) to SNR as SNR tends to zero. To this end we first shall show that

$$\lim_{\mathrm{SNR}\downarrow 0}\frac{\mathsf{E}\left[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\right]}{\mathrm{SNR}} = 0. \tag{45}$$

We recall that for any pair of distributions P_0 and P_1 satisfying $P_1 \ll P_0$ [12, p. 1023]

$$\lim_{\beta\downarrow 0}\frac{D\left(\beta P_1+(1-\beta)P_0\|\,P_0\right)}{\beta} = 0. \tag{46}$$

Thus, for any given $\mathbf{X}_{-\infty}^{-1}=\mathbf{x}_{-\infty}^{-1}$, (46) together with $\delta=\mathrm{SNR}\,L\,\sigma^2/\xi^2$ implies that

$$\lim_{\mathrm{SNR}\downarrow 0}\frac{D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)}{\mathrm{SNR}} = 0. \tag{47}$$

In order to show that this also holds when $D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)$ is averaged over $\mathbf{X}_{-\infty}^{-1}$, we derive in the following the uniform upper bound

$$\sup_{\mathbf{x}_{-\infty}^{-1}} D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) = D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0}. \tag{48}$$

The claim (45) follows then by upper bounding

$$\mathsf{E}\Big[D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{X}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{X}_{-\infty}^{-1}}\Big)\Big] \leq D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0} \tag{49}$$

and by (47).

In order to prove (48) we use that any Gaussian random vector can be expressed as the sum of two independent Gaussian random vectors to write the channel output $\tilde{\mathbf{Y}}_0$ as

$$\tilde{\mathbf{Y}}_0 = \mathbf{X}_0 + \mathbf{V} + \mathbf{W}, \tag{50}$$

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

$$\begin{aligned}
\mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(1,1) &= \sigma^2\\
\mathsf{K}_{\mathbf{V}|\mathbf{x}_0}(i,i) &= \sigma^2 + \alpha_{i-1}x_1^2, \qquad i = 2,\dots,L,
\end{aligned}$$

and

$$\mathsf{K}_{\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}(i,i) = \sum_{\ell=-\infty}^{-1} \alpha_{-\ell L+i-1}x_{\ell L+1}^2, \qquad i = 1,\dots,L.$$

Thus $\mathbf{W}$ is the portion of the noise due to $\mathbf{X}_{-\infty}^{-1}$, and $\mathbf{V}$ is the portion of the noise that remains after subtracting $\mathbf{W}$. Note that $\mathbf{X}_0 + \mathbf{V}$ and $\mathbf{W}$ are independent of each other because $\mathbf{X}_0$ is, by construction, independent of $\mathbf{X}_{-\infty}^{-1}$. The upper bound (48) follows now by

$$\begin{aligned}
D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big) &= D\Big(P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\\
&\leq D\big(P_{\mathbf{X}_0+\mathbf{V}}\big\|P_{\mathbf{X}_0+\mathbf{V}|X_1=0}\big)\\
&= D\Big(P_{\tilde{\mathbf{Y}}_0|\mathbf{x}_{-\infty}^{-1}}\Big\|P_{\tilde{\mathbf{Y}}_0|X_1=0,\mathbf{x}_{-\infty}^{-1}}\Big)\Big|_{\mathbf{x}_{-\infty}^{-1}=0},
\end{aligned} \tag{51}$$

where

$$P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|\mathbf{x}_{-\infty}^{-1}} \quad \text{and} \quad P_{\mathbf{X}_0+\mathbf{V}+\mathbf{W}|X_1=0,\mathbf{x}_{-\infty}^{-1}}$$

denote the distributions of $\mathbf{X}_0 + \mathbf{V} + \mathbf{W}$ conditional on the inputs $\mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1}$ and on $(X_1 = 0, \mathbf{X}_{-\infty}^{-1} = \mathbf{x}_{-\infty}^{-1})$, respectively; $P_{\mathbf{X}_0+\mathbf{V}}$ denotes the unconditional distribution of $\mathbf{X}_0 + \mathbf{V}$; and $P_{\mathbf{X}_0+\mathbf{V}|X_1=0}$ denotes the distribution of $\mathbf{X}_0 + \mathbf{V}$ conditional on $X_1 = 0$. Here the inequality follows by the data processing inequality for relative entropy (see [10, Sec. 2.9]) and by noting that $\mathbf{X}_0 + \mathbf{V}$ is independent of $\mathbf{X}_{-\infty}^{-1}$.

Returning to the analysis of (40), we obtain from (44) and (45)

$$\begin{aligned}
\dot{C}(0) &\geq \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}}\\
&\geq \lim_{\mathrm{SNR}\downarrow 0} \frac{1}{2}\sum_{i=1}^{L} \frac{\alpha_{i-1}}{1+\alpha\, L\,\mathrm{SNR}} - \frac{1}{2}\sum_{i=2}^{L} \frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}\\
&= \frac{1}{2}\sum_{i=1}^{L}\alpha_{i-1} - \frac{1}{2}\sum_{i=2}^{L} \frac{\log\left(1+\alpha_{i-1}\xi^2/\sigma^2\right)}{\xi^2/\sigma^2}.
\end{aligned} \tag{52}$$

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 on the capacity per unit cost

$$\dot{C}(0) \geq \lim_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{1}{2}(1+\alpha). \tag{53}$$

Thus (53), (20), and (43) yield

$$\frac{1}{2}(1+\alpha) \leq \varliminf_{\mathrm{SNR}\downarrow 0} \frac{C(\mathrm{SNR})}{\mathrm{SNR}} \leq \dot{C}(0) \leq \dot{C}_{\mathrm{FB}}(0) \leq \frac{1}{2}(1+\alpha) \tag{54}$$

which proves Theorem 2.

6 Proof of Theorem 3

6.1 Part i)

In order to show that

$$\varliminf_{\ell\to\infty} \frac{\alpha_{\ell+1}}{\alpha_\ell} > 0 \tag{55}$$

implies that the feedback capacity $C_{\mathrm{FB}}(\mathrm{SNR})$ is bounded, we derive a capacity upper bound which is based on (18) 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 (55), we can find an $\ell_0 \in \mathbb{Z}^+$ and a $0<\rho<1$ so that

$$\alpha_{\ell_0} > 0 \qquad \text{and} \qquad \frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell \geq \ell_0. \tag{56}$$

We continue with the chain rule for mutual information

$$\frac{1}{n}I(M;Y_1^n) = \frac{1}{n}\sum_{k=1}^{\ell_0} I\big(M;Y_k\big|Y_1^{k-1}\big) + \frac{1}{n}\sum_{k=\ell_0+1}^{n} I\big(M;Y_k\big|Y_1^{k-1}\big). \tag{57}$$

Each summand in the first sum on the RHS of (57) is upper bounded by

$$\begin{aligned}
I\big(M;Y_k\big|Y_1^{k-1}\big) &\leq h(Y_k) - h\big(Y_k\big|Y_1^{k-1},M\big) \\
&= h(Y_k) - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2 + \sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] - h\big(U_k\big|U_1^{k-1}\big) \\
&\leq \frac{1}{2}\log\left(2\pi e\left(1+\sum_{\ell=1}^{k}\alpha_{k-\ell}\frac{\mathsf{E}\big[X_\ell^2\big]}{\sigma^2}\right)\right) - h\big(U_k\big|U_1^{k-1}\big) \\
&\leq \frac{1}{2}\log\left(2\pi e\left(1+\big(\sup_{\ell'\in\mathbb{Z}_0^+}\alpha_{\ell'}\big)\sum_{\ell=1}^{k}\frac{\mathsf{E}\big[X_\ell^2\big]}{\sigma^2}\right)\right) - h\big(U_k\big|U_1^{k-1}\big) \\
&\leq \frac{1}{2}\log\left(2\pi e\left(1+\big(\sup_{\ell'\in\mathbb{Z}_0^+}\alpha_{\ell'}\big)\,n\,\mathrm{SNR}\right)\right) - h\big(U_k\big|U_1^{k-1}\big) \\
&\leq \frac{1}{2}\log\left(2\pi e\left(1+\big(\sup_{\ell'\in\mathbb{Z}_0^+}\alpha_{\ell'}\big)\,n\,\mathrm{SNR}\right)\right) - h\big(U_k\big|U_{-\infty}^{k-1}\big).
\end{aligned} \tag{58}$$

Recall that $\sup_{\ell'\in\mathbb{Z}_0^+}\alpha_{\ell'}$ is finite (9). Here the first inequality follows because conditioning cannot increase entropy; the following equality follows because $\big(X_1^k, U_1^{k-1}\big)$ is a function of $\big(M, Y_1^{k-1}\big)$, from the behavior of entropy under translation and scaling [10, Thms. 9.6.3 & 9.6.4], and from the fact that, conditional on U_1^{k-1}, U_k is independent of $\big(X_1^k, M, Y_1^{k-1}\big)$; the subsequent inequality follows from the entropy maximizing property of Gaussian random variables and by lower bounding $\mathsf{E}\Big[\log\Big(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\Big)\Big] \geq \log\sigma^2$; the next inequality by upper bounding each coefficient $\alpha_\ell \leq \sup_{\ell'\in\mathbb{Z}_0^+}\alpha_{\ell'}$, $\ell=1,\ldots,k$; the subsequent inequality follows from the power constraint (11); and the last inequality follows because conditioning cannot increase entropy.

The summands in the second sum on the RHS of (57) are upper bounded using the general upper bound for mutual information [16, Thm. 5.1]

$$I(X;Y) \leq \int D\big(W(\cdot|x)\big\|R(\cdot)\big)\,\mathrm{d}Q(x), \tag{59}$$

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.

We upper bound $I\big(M;Y_k\big|Y_1^{k-1}=y_1^{k-1}\big)$, $k=\ell_0+1,\ldots,n$ 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

$$\frac{\sqrt{\beta}}{\pi}\frac{1}{1+\beta y_k^2}, \qquad y_k\in\mathbb{R}, \tag{60}$$

where we choose the scale parameter β to be[9]

$$\beta=\frac{1}{\tilde{\beta}y_{k-\ell_0}^2} \qquad \text{and} \qquad \tilde{\beta}=\min\left\{\rho^{\ell_0-1}\frac{\alpha_{\ell_0}}{\max\limits_{\ell'=0,\ldots,\ell_0-1}\alpha_{\ell'}},\alpha_{\ell_0},\rho^{\ell_0}\right\}, \tag{61}$$

with $0<\rho<1$ and $\ell_0\in\mathbb{Z}^+$ given by (56). Note that (56) together with (9) implies that

$$0<\tilde{\beta}<1 \qquad \text{and} \qquad \tilde{\beta}\alpha_\ell\le\alpha_{\ell+\ell_0}, \quad \ell\in\mathbb{Z}_0^+. \tag{62}$$

Applying (60) to (59) yields

$$\begin{aligned} I\big(M;Y_k\big|Y_1^{k-1}=y_1^{k-1}\big) &\le \mathsf{E}\left[\log\left(1+\frac{Y_k^2}{\tilde{\beta}Y_{k-\ell_0}^2}\right)\,\middle|\,Y_1^{k-1}=y_1^{k-1}\right]+\frac{1}{2}\log\big(\tilde{\beta}y_{k-\ell_0}^2\big) \\ &\quad +\log\pi-h\big(Y_k\big|M,Y_1^{k-1}=y_1^{k-1}\big), \end{aligned} \tag{63}$$

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

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

We evaluate the terms on the RHS of (64) individually. We begin with

$$h\big(Y_k\big|Y_1^{k-1},M\big)\ge\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right]+h\big(U_k\big|U_{-\infty}^{k-1}\big), \tag{65}$$

where we use the same steps as in the equality in (58) and that conditioning cannot increase entropy. The next term is upper bounded by

$$\begin{aligned} \mathsf{E}\Big[\log\big(\tilde{\beta}Y_{k-\ell_0}^2\big)\Big] &= \mathsf{E}\left[\mathsf{E}\left[\log\left(\tilde{\beta}\big(X_{k-\ell_0}+\theta\big(X_1^{k-\ell_0-1}\big)\cdot U_{k-\ell_0}\big)^2\right)\,\middle|\,X_1^{k-\ell_0}\right]\right] \\ &\le \mathsf{E}\left[\log\left(\tilde{\beta}\mathsf{E}\left[\big(X_{k-\ell_0}+\theta\big(X_1^{k-\ell_0-1}\big)\cdot U_{k-\ell_0}\big)^2\,\middle|\,X_1^{k-\ell_0}\right]\right)\right] \\ &= \mathsf{E}\left[\log\left(\tilde{\beta}X_{k-\ell_0}^2+\tilde{\beta}\sigma^2+\tilde{\beta}\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] \\ &\le \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2\right)\right], \end{aligned} \tag{66}$$

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

$$\theta\big(x_1^{k-1}\big)\triangleq\sqrt{\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}x_\ell^2}. \tag{67}$$

[9]When $y_{k-\ell_0}=0$ then with this choice of β the density of the Cauchy distribution (60) is undefined. However, this event is of zero probability and has therefore no impact on the mutual information $I\big(M;Y_k\big|Y_1^{k-1}\big)$.

Here the first inequality in (66) follows from Jensen's inequality, and the second inequality follows from (62). Similarly we use Jensen's inequality along with (62) to upper bound

$$\begin{aligned}\mathsf{E}\Big[\log\big(\tilde{\beta}Y_{k-\ell_0}^2+Y_k^2\big)\Big] &\leq \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2+\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right]\\ &\leq \log 2+\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right]. \end{aligned} \tag{68}$$

In order to lower bound $\mathsf{E}\big[\log\big(Y_{k-\ell_0}^2\big)\big]$ 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 $0<\delta\leq 1$ and $0<\eta<1$ we have*

$$\sup_{c\in\mathbb{R}}\mathsf{E}\big[\log|X+c|^{-1}\cdot \mathrm{I}\{|X+c|\leq\delta\}\big]\leq \epsilon(\delta,\eta)+\frac{1}{\eta}h^-(X) \tag{69}$$

where $\mathrm{I}\{\cdot\}$ denotes the indicator function[10]*; $h^-(X)$ is defined as*

$$h^-(X)\triangleq\int_{\{x\in\mathbb{R}:f_X(x)>1\}} f_X(x)\log f_X(x)\,dx; \tag{70}$$

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

Proof. See [16, Lemma 6.7]. □

We write the expectation as

$$\mathsf{E}\big[\log\big(Y_{k-\ell_0}^2\big)\big]=\mathsf{E}\Big[\mathsf{E}\Big[\log\Big(X_{k-\ell_0}+\theta\big(X_1^{k-\ell_0-1}\big)\cdot U_{k-\ell_0}\Big)^2\,\Big|\,X_1^{k-\ell_0}\Big]\Big]$$

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

$$\begin{aligned}&\mathsf{E}\Big[\log\Big(X_{k-\ell_0}+\theta\big(X_1^{k-\ell_0-1}\big)\cdot U_{k-\ell_0}\Big)^2\,\Big|\,X_1^{k-\ell_0}=x_1^{k-\ell_0}\Big]\\ &\quad=\log\theta^2\big(x_1^{k-\ell_0-1}\big)-2\,\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|^{-1}\,\middle|\,X_1^{k-\ell_0}=x_1^{k-\ell_0}\right]\\ &\quad\geq\log\theta^2\big(x_1^{k-\ell_0-1}\big)-2\epsilon(\delta,\eta)-\frac{2}{\eta}h^-(U_{k-\ell_0})+\log\delta^2 \end{aligned}\tag{71}$$

for some $0<\delta\leq 1$ and $0<\eta<1$. Here the inequality follows by splitting the conditional expectation into the two expectations

$$\begin{aligned}&\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|^{-1}\,\middle|\,X_1^{k-\ell_0}=x_1^{k-\ell_0}\right]\\ &\quad=\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|^{-1}\cdot\mathrm{I}\left\{\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|\leq\delta\right\}\,\middle|\,X_1^{k-\ell_0}=x_1^{k-\ell_0}\right]\\ &\qquad+\mathsf{E}\left[\log\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|^{-1}\cdot\mathrm{I}\left\{\left|\frac{X_{k-\ell_0}}{\theta\big(X_1^{k-\ell_0-1}\big)}+U_{k-\ell_0}\right|>\delta\right\}\,\middle|\,X_1^{k-\ell_0}=x_1^{k-\ell_0}\right]\end{aligned}$$

and by upper bounding then the first term on the RHS using Lemma 4 and the second term by $-\log\delta$. Averaging (71) over $X_1^{k-\ell_0}$ yields

$$\mathsf{E}\big[\log\big(Y_{k-\ell_0}^2\big)\big]\geq\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right]-2\epsilon(\delta,\eta)-\frac{2}{\eta}h^-(U_{k-\ell_0})+\log\delta^2. \tag{72}$$

[10]The indicator function $\mathrm{I}\{\text{statement}\}$ takes on the value 1 if the statement is true and 0 otherwise.

Note that, since $U_{k-\ell_0}$ is of unit variance, (8) together with [16, Lemma 6.4] implies that $h^{-}(U_{k-\ell_0})$ is finite.

Turning back to the upper bound (64) we obtain from (65), (66), (68), and (72)

$$\begin{aligned}
I\big(M;Y_k\big|Y_1^{k-1}\big) &\\
\leq \log\pi - \frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-1}\alpha_{k-\ell}X_\ell^2\right)\right] - h\big(U_k\big|U_{-\infty}^{k-1}\big) &\\
+\frac{1}{2}\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}X_\ell^2\right)\right] + \log 2 + \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right] &\\
-\mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] + 2\epsilon(\delta,\eta) + \frac{2}{\eta}h^{-}(U_{k-\ell_0}) - \log\delta^2 - \log\tilde{\beta} &\\
\leq \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right] - \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] + \mathsf{K}, & \qquad (73)
\end{aligned}$$

where

$$\mathsf{K} \triangleq \log\frac{2\pi}{\tilde{\beta}\delta^2} - h\big(U_k\big|U_{-\infty}^{k-1}\big) + \frac{2}{\eta}h^{-}(U_{k-\ell_0}) + 2\epsilon(\delta,\eta) \tag{74}$$

is a finite constant, and where the last inequality in (73) follows because for any $X_{k-\ell_0+1}^{k-1} = x_{k-\ell_0+1}^{k-1}$ we have $\sum_{\ell=1}^{k-\ell_0}\alpha_{k-\ell}\,x_\ell^2 \leq \sum_{\ell=1}^{k-1}\alpha_{k-\ell}\,x_\ell^2$. Note that K does not depend on k as the process $\{U_k\}$ is stationary.

Turning back to the evaluation of the second sum on the RHS of (57), we use that for any sequences $\{a_k\}$ and $\{b_k\}$

$$\sum_{k=\ell_0+1}^{n}(a_k-b_k) = \sum_{k=n-2\ell_0+1}^{n}(a_k-b_{k-n+3\ell_0}) + \sum_{k=\ell_0+1}^{n-2\ell_0}(a_k-b_{k+2\ell_0}). \tag{75}$$

Defining

$$a_k \triangleq \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2\right)\right] \tag{76}$$

and

$$b_k \triangleq \mathsf{E}\left[\log\left(\sigma^2+\sum_{\ell=1}^{k-\ell_0-1}\alpha_{k-\ell_0-\ell}X_\ell^2\right)\right] \tag{77}$$

we have for the first sum on the RHS of (75)

$$\begin{aligned}
\sum_{k=n-2\ell_0+1}^{n}(a_k-b_{k-n+3\ell_0}) &= \sum_{k=n-2\ell_0+1}^{n}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k-n+2\ell_0-1}\alpha_{k-n+2\ell_0-\ell}X_\ell^2}\right)\right]\\
&\leq 2\ell_0\log\left(1+\big(\sup_{\ell\in\mathbb{Z}_0^+}\alpha_\ell\big)\,n\,\mathrm{SNR}\right) \qquad (78)
\end{aligned}$$

which follows by lower bounding the denominator by σ^2, and by using then Jensen's inequality together with the third and fourth inequality in (58). For the second sum on the RHS of (75) we have

$$\begin{aligned}
\sum_{k=\ell_0+1}^{n-2\ell_0}(a_k-b_{k+2\ell_0}) &= \sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}X_\ell^2}\right)\right]\\
&\leq \sum_{k=\ell_0+1}^{n-2\ell_0}\mathsf{E}\left[\log\left(\frac{\sigma^2+\sum_{\ell=1}^{k}\alpha_{k+\ell_0-\ell}X_\ell^2}{\sigma^2+\sum_{\ell=1}^{k+\ell_0-1}\alpha_{k+\ell_0-\ell}X_\ell^2}\right)\right] - (n-3\ell_0)\log\tilde{\beta}\\
&\leq -(n-3\ell_0)\log\tilde{\beta}, \qquad (79)
\end{aligned}$$

where the first inequality follows by adding $\log\tilde{\beta}$ to the expectation and by upper bounding then $\tilde{\beta}\alpha_\ell < \alpha_{\ell+\ell_0}$, $\ell \in \mathbb{Z}_0^+$ (62); and the last inequality follows because for any given $X_{k+1}^{k+\ell_0-1} = x_{k+1}^{k+\ell_0-1}$ we have $\sum_{\ell=1}^{k} \alpha_{k+\ell_0-\ell}\, x_\ell^2 \leq \sum_{\ell=1}^{k+\ell_0-1} \alpha_{k+\ell_0-\ell}\, x_\ell^2$.

We apply now (73), (75), (78), and (79) to upper bound

$$\frac{1}{n}\sum_{\ell=\ell_0+1}^{n} I\big(M;Y_k\big|Y_1^{k-1}\big) \leq \frac{n-\ell_0}{n}\mathsf{K} + \frac{2\ell_0}{n}\log\left(1+\big(\sup_{\ell\in\mathbb{Z}_0^+}\alpha_\ell\big)\,n\,\mathrm{SNR}\right) - \frac{n-3\ell_0}{n}\log\tilde{\beta} \tag{80}$$

which together with (57) and (58) yields

$$\begin{aligned}\frac{1}{n}I(M;Y_1^n) \leq{}& \frac{n-\ell_0}{n}\mathsf{K} - \frac{n-3\ell_0}{n}\log\tilde{\beta} + \frac{\ell_0}{2n}\log(2\pi e) - \frac{\ell_0}{n}h\big(U_k\big|U_{-\infty}^{k-1}\big)\\ &+ \frac{\ell_0}{n}\frac{5}{2}\log\left(1+\big(\sup_{\ell\in\mathbb{Z}_0^+}\alpha_\ell\big)\,n\,\mathrm{SNR}\right).\end{aligned} \tag{81}$$

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

6.2 Part ii)

We shall show that

$$\lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty \tag{82}$$

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

$$\varlimsup_{\ell\to\infty}\frac{\alpha_{\ell+1}}{\alpha_\ell} = 0 \quad\Longrightarrow\quad \lim_{\ell\to\infty}\frac{1}{\ell}\log\frac{1}{\alpha_\ell} = \infty. \tag{83}$$

We prove the claim by proposing a coding scheme that achieves an unbounded rate. We first note that (82) implies that for any $0<\varrho<1$ we can find an $\ell_0\in\mathbb{Z}^+$ so that

$$\alpha_\ell < \varrho^\ell, \quad \ell=\ell_0, \ell\geq\ell_0. \tag{84}$$

If there exists an $\ell_0\in\mathbb{Z}^+$ so that $\alpha_\ell=0$, $\ell\geq\ell_0$, then we can achieve the (unbounded) rate

$$R = \frac{1}{2L}\log(1+L\,\mathrm{SNR}), \qquad L\geq\ell_0 \tag{85}$$

by a coding scheme where the channel inputs $\{X_{kL+1}\,,\, k\in\mathbb{Z}_0^+\}$ are IID, 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 that 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 [15].

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 $L\geq\ell_0$ 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.[11]

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 $\|\mathbf{y}-\mathbf{x}(m')\|^2$ for each $m'\in\mathcal{M}$ and decides on the message that satisfies

$$\hat{M} = \arg\min_{m'\in\mathcal{M}}\|\mathbf{y}-\mathbf{x}(m')\|^2, \tag{86}$$

[11] It follows from the weak law of large numbers that, for any $m\in\mathcal{M}$, $\frac{1}{n}\sum_{k=1}^{n}x_k^2(m)$ 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 (13)—and hence also the average-power constraint (11)—vanishes as n tends to infinity.

where ties are resolved with a fair coin flip. Here, $\|\cdot\|$ denotes the Euclidean norm, and $\mathbf{y}$ and $\mathbf{x}(m')$ denote the respective 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 $\Pr\bigl(\hat{M} \neq M\bigr)$, 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 m-th message $\Pr\bigl(\hat{M}\neq M \bigm| M=m\bigr)$ does not depend on m, and we thus conclude that $\Pr\bigl(\hat{M}\neq M\bigr) = \Pr\bigl(\hat{M}\neq M \bigm| M=1\bigr)$. We further note that

$$\Pr\bigl(\hat{M}\neq M \bigm| M=1\bigr) \leq \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{Y}-\mathbf{X}(m')\|^2 \leq \|\mathbf{Z}\|^2 \,\middle|\, M=1\right), \tag{87}$$

where

$$\mathbf{Z} = \Bigl(\theta\bigl(X_1(1)\bigr)\cdot U_1, \theta\bigl(X_1^L(1)\bigr)\cdot U_{L+1}, \ldots, \theta\bigl(X_1^{(\lfloor n/L\rfloor-1)L+1}(1)\bigr)\cdot U_{(\lfloor n/L\rfloor -1)L+1}\Bigr)^{\mathsf{T}}$$

which is, conditional on $M=1$, equal to $\|\mathbf{Y}-\mathbf{X}(1)\|^2$. In order to analyze (87) we need the following lemma.

Lemma 5. *Consider the channel described in Section 2, and assume that $\{\alpha_\ell\}$ satisfies (82). Further assume that $\{X_{kL+1}, k\in\mathbb{Z}_0^+\}$ is a sequence of IID, zero-mean Gaussian random variables of variance P, and that $X_k = 0$ if $k \mod L \neq 1$ (where $k \mod L$ stands for the remainder upon diving k by L). Let the set $\mathcal{D}_\epsilon$ be defined as*

$$\mathcal{D}_\epsilon \triangleq \left\{ (\mathbf{y},\mathbf{z}) \in \mathbb{R}^{\lfloor n/L\rfloor}\times\mathbb{R}^{\lfloor n/L\rfloor} : \begin{array}{l} \left|\dfrac{1}{\lfloor n/L\rfloor}\|\mathbf{y}\|^2 - (\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P})\right| < \epsilon, \\ \left|\dfrac{1}{\lfloor n/L\rfloor}\|\mathbf{z}\|^2 - (\sigma^2+\alpha^{(L)}\,\mathsf{P})\right| < \epsilon \end{array} \right\}, \tag{88}$$

with $\alpha^{(L)}$ being defined as

$$\alpha^{(L)} \triangleq \sum_{\ell=1}^{\infty} \alpha_{\ell L}. \tag{89}$$

Then

$$\lim_{n\to\infty} \Pr\bigl((\mathbf{Y},\mathbf{Z})\in\mathcal{D}_\epsilon\bigr) = 1 \tag{90}$$

for any $\epsilon > 0$.

Proof. See Appendix C. ☐

In order to upper bound the RHS of (87) we proceed along the lines of [15], [14]. We have

$$\begin{aligned} &\Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{Y}-\mathbf{X}(m')\|^2 \leq \|\mathbf{Z}\|^2 \,\middle|\, M=1\right) \\ &\quad \leq \Pr\bigl((\mathbf{Y},\mathbf{Z})\notin\mathcal{D}_\epsilon\bigr) + \int_{\mathcal{D}_\epsilon} \Pr\left(\bigcup_{m'=2}^{|\mathcal{M}|} \|\mathbf{y}-\mathbf{X}(m')\|^2 \leq \|\mathbf{z}\|^2 \,\middle|\, (\mathbf{y},\mathbf{z}), M=1\right) \mathrm{d}\mathrm{P}(\mathbf{y},\mathbf{z}), \end{aligned} \tag{91}$$

where we use that, by the symmetry of the codebook construction, the law of $(\mathbf{Y},\mathbf{Z})$ does not depend on M. It follows from Lemma 5 that the first term on the RHS of (91) vanishes as n tends to infinity. Since the codewords are independent of each other, conditional on $M=1$, the distribution of $\mathbf{X}(m')$, $m'=2,\ldots,|\mathcal{M}|$ does not depend on $(\mathbf{y},\mathbf{z})$. We upper bound the second term on the RHS of (91) by analyzing $\Pr\bigl(\|\mathbf{y}-\mathbf{X}(m')\|^2 \leq \|\mathbf{z}\|^2 \bigm| (\mathbf{y},\mathbf{z}), M=1\bigr)$, $m'=2,\ldots,|\mathcal{M}|$ and by applying then the union of events bound.

For $m' = 2, \ldots, |\mathcal{M}|$, we have

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

for any $s < 0$. This follows by upper bounding $\|\mathbf{z}\|^2$ by $\lfloor n/L \rfloor (\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon)$ and from Chernoff's bound [17, Sec. 5.4]. Using that, for $(\mathbf{y}, \mathbf{z}) \in \mathcal{D}_\epsilon$,

$$\|\mathbf{y}\|^2 > \lfloor n/L \rfloor (\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon)$$

it follows from the union of events bound and from (92) that (91) goes to zero as n tends to infinity if for some $s < 0$ the rate R satisfies

$$R < \frac{s}{L}(\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon) + \frac{1}{2L}\log(1-2s\mathsf{P}) - \frac{s}{L}\frac{\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon}{1-2s\mathsf{P}}. \qquad (93)$$

Thus choosing $s = -1/2 \cdot 1/(1 + \alpha^{(L)}\,\mathsf{P})$ yields that any rate below

$$\begin{aligned}
&-\frac{1}{2L}\frac{\sigma^2 + \alpha^{(L)}\,\mathsf{P} + \epsilon}{1 + \alpha^{(L)}\,\mathsf{P}} + \frac{1}{2L}\log\left(1 + \frac{\mathsf{P}}{1 + \alpha^{(L)}\,\mathsf{P}}\right) \\
&+ \frac{1}{2L}\frac{\sigma^2 + \mathsf{P} + \alpha^{(L)}\,\mathsf{P} - \epsilon}{1 + \alpha^{(L)}\,\mathsf{P}}\frac{1}{1 + \frac{\mathsf{P}}{1+\alpha^{(L)}\,\mathsf{P}}} \qquad (94)
\end{aligned}$$

is achievable. As P tends to infinity this converges to

$$\frac{1}{2L}\log\left(1 + \frac{1}{\alpha^{(L)}}\right) > \frac{1}{2L}\log\frac{1}{\alpha^{(L)}}. \qquad (95)$$

It remains to show that given (84) we can make $-\frac{1}{L}\log\alpha^{(L)}$ 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 (95) can therefore be further lower bounded by

$$\frac{1}{2L}\log\left(1 - \varrho^L\right) + \frac{1}{2}\log\frac{1}{\varrho}. \qquad (96)$$

Letting L tend to infinity yields then that we can achieve any rate below $\frac{1}{2}\log\frac{1}{\varrho}$. As 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.

7 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 characterize 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 the heating effect is not just unharmful but can be even beneficial in the sense that the capacity per unit cost can be larger than the capacity per unit cost of a corresponding channel with ideal heat sink, i.e., where the weights describing the dependency of the noise variance on the channel input powers are zero. This suggests that at low SNR no heat sinks should be used.

Studying capacity at high SNR, we derived a sufficient condition and a necessary condition on the weights for the capacity to be bounded in the SNR. We showed that when the sequence of weights decays not faster than geometrically, then capacity is bounded in the SNR. On the other hand, if the sequence of weights decays faster than geometrically, then capacity is unbounded in the SNR. This result demonstrates the importance of an efficient heat sink at high SNR.

Acknowledgment

Fruitful discussions with Ashish Khisti and Michèle Wigger are gratefully acknowledged. Sergio Verdú's comments at the ISIT 2007 on our low SNR results are also much appreciated.

A Proof of Proposition 1

We first note that by the expression of the capacity per unit cost of a memoryless channel [12] we have

$$\sup_{\mathrm{SNR}>0} \frac{C_{\alpha=0}(\mathrm{SNR})}{\mathrm{SNR}} = \sup_{\zeta^2>0} \frac{D\big(W_{\alpha=0}(\cdot|\zeta)\big\|W_{\alpha=0}(\cdot|0)\big)}{\zeta^2/\sigma^2}, \tag{97}$$

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

$$Y_k = x_k + \sigma \cdot U_k. \tag{98}$$

Thus to prove Proposition 1 it suffices to show that

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

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 was 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 IID in blocks of L symbols (for some $L \in \mathbb{Z}^+$). Thus $\{(X_{bL+1},\ldots,X_{(b+1)L}), b\in\mathbb{Z}_0^+\}$ is a sequence of IID random length-L vectors with $(X_{bL+1},\ldots,X_{(b+1)L})$ taking on the value $(\xi,0,\ldots,0)$ with probability δ and $(0,\ldots,0)$ with probability $1-\delta$, for some $\xi\in\mathbb{R}$. To satisfy the power constraint (11) we shall choose ξ and δ such that

$$\frac{\xi^2}{\sigma^2}\delta = L\,\mathrm{SNR}. \tag{99}$$

We use the chain rule for mutual information to write

$$\begin{aligned}\frac{1}{n}I(X_1^n;Y_1^n) &= \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(X_{bL+1};Y_1^n\big|X_1^{bL}\big)\\ &\geq \frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1} I\big(X_{bL+1};Y_{bL+1}\big|X_1^{bL}\big),\end{aligned} \tag{100}$$

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 (98) 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.,

$$R_{\text{on-off}}^{(\xi)}(\mathsf{snr}) \triangleq \sup_{\substack{P_X(\xi)=1-P_X(0)=\wp,\\ \xi^2/\sigma^2\wp\leq\mathsf{snr}}} I(X;X+\sigma\cdot U_k), \qquad \mathsf{snr}\geq 0. \tag{101}$$

Notice that $R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$, $\mathsf{snr}\geq 0$ is a nonnegative, monotonically nondecreasing function of snr with $R_{\text{on-off}}^{(\xi)}(0)=0$. From the strict concavity of mutual information it follows that $R_{\text{on-off}}^{(\xi)}(\mathsf{snr}) > 0$ whenever $\mathsf{snr}>0$. Also, for a fixed ξ, $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is concave in snr. Consequently, for some $\mathsf{snr}_0>0$, the function $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is strictly monotonic in the interval $\mathsf{snr}\in[0,\mathsf{snr}_0]$, and hence the supremum on the RHS of (101) is attained for $\wp=\mathsf{snr}\,\sigma^2/\xi^2$, $\mathsf{snr}\in[0,\mathsf{snr}_0]$.

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

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

(with $\theta\big(x_1^{bL}\big)$ defined in (67)), and by using that for $\mathsf{snr}\in[0,\mathsf{snr}_0]$ the supremum on the RHS of (101) is attained for $\wp=\mathsf{snr}\,\sigma^2/\xi^2$ we obtain

$$I\big(X_{bL+1};Y_{bL+1}\big|X_1^{bL}=x_1^{bL}\big)=R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}x_{\ell L+1}^2/\sigma^2}\right),\quad \text{SNR}\in[0,\text{SNR}_0],\tag{102}$$

where $\text{SNR}_0\triangleq\mathsf{snr}_0/L$. Averaging over X_1^{bL} and combining with (100) yields

$$\begin{aligned}\frac{1}{n}I(X_1^n;Y_1^n)&\geq\frac{1}{n}\sum_{b=0}^{\lfloor n/L\rfloor-1}\mathsf{E}\left[R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}X_{\ell L+1}^2/\sigma^2}\right)\right]\\&\geq\frac{\lfloor n/L\rfloor}{n}R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}\right),\quad \text{SNR}\in[0,\text{SNR}_0],\end{aligned}\tag{103}$$

where the second inequality follows by upper bounding $\sum_{\ell=0}^{b-1}\alpha_{(b-\ell)L}X_{\ell L+1}^2/\sigma^2\leq\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2$, and by using that $\mathsf{snr}\mapsto R_{\text{on-off}}^{(\xi)}(\mathsf{snr})$ is monotonically increasing in snr. The lower bound on $C_{\text{Info}}(\text{SNR})$ follows then by letting n tend to infinity

$$C_{\text{Info}}(\text{SNR})=\varliminf_{n\to\infty}\frac{1}{n}I(X_1^n;Y_1^n)\geq\frac{1}{L}R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}\right).\tag{104}$$

With this we can lower bound the information capacity per unit cost as

$$\begin{aligned}\sup_{\text{SNR}>0}\frac{C_{\text{Info}}(\text{SNR})}{\text{SNR}}&\geq\varliminf_{\text{SNR}\downarrow0}\frac{C_{\text{Info}}(\text{SNR})}{\text{SNR}}\\&\geq\varliminf_{\text{SNR}\downarrow0}\frac{1}{L}\frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}\right)}{\text{SNR}}\\&=\varliminf_{\text{SNR}\downarrow0}\frac{R_{\text{on-off}}^{(\xi)}\left(\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}\right)}{\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}}\,\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}\\&=\varliminf_{\text{SNR}'\downarrow0}\frac{R_{\text{on-off}}^{(\xi)}(\text{SNR}')}{\text{SNR}'}\,\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2},\end{aligned}\tag{105}$$

where the first inequality follows by lower bounding the supremum by the limit; and where the last equality follows by substituting $\text{SNR}'=\frac{L\,\text{SNR}}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}$.

Proceeding along the lines of the proof of [12, Thm. 3], it can be shown that

$$\lim_{\text{SNR}'\downarrow0}\frac{R_{\text{on-off}}^{(\xi)}(\text{SNR}')}{\text{SNR}'}=\frac{D\big(W_{\alpha=0}(\cdot|\xi)\big\|W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2}\tag{106}$$

and therefore

$$\sup_{\text{SNR}>0}\frac{C_{\text{Info}}(\text{SNR})}{\text{SNR}}\geq\frac{D\big(W_{\alpha=0}(\cdot|\xi)\big\|W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2}\cdot\frac{1}{1+\sum_{\ell=1}^{\infty}\alpha_{\ell L}\xi^2/\sigma^2}.\tag{107}$$

Noting that (9) & (21) imply

$$0\leq\lim_{L\to\infty}\sum_{\ell=1}^{\infty}\alpha_{\ell L}\leq\lim_{L\to\infty}\sum_{\ell=L}^{\infty}\alpha_{\ell}=0\tag{108}$$

we obtain by letting L tend to infinity

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \frac{D\big(W_{\alpha=0}(\cdot|\xi)\big\|W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2}. \tag{109}$$

Maximizing (109) over ξ^2 yields then

$$\sup_{\mathrm{SNR}>0} \frac{C_{\mathrm{Info}}(\mathrm{SNR})}{\mathrm{SNR}} \geq \sup_{\xi^2>0} \frac{D\big(W_{\alpha=0}(\cdot|\xi)\big\|W_{\alpha=0}(\cdot|0)\big)}{\xi^2/\sigma^2} \tag{110}$$

which, in view of (97), proves Proposition 1.

B Appendix to Section 5.2

We shall prove that

$$\lim_{b\to\infty} I\big(\mathbf{X}_{-\infty}^{-1};\tilde{\mathbf{Y}}_b\big|\mathbf{X}_0^b\big) = 0. \tag{111}$$

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

$$\alpha_0^{(1)} \triangleq 0 \tag{112}$$

$$\alpha_b^{(i)} \triangleq \alpha_{bL+i-1}, \qquad (b,i)\in\mathbb{Z}_0^+\times\mathbb{Z}^+\setminus\{(0,1)\}. \tag{113}$$

We have

$$\begin{aligned}
I\big(\mathbf{X}_{-\infty}^{-1};\tilde{\mathbf{Y}}_b\big|\mathbf{X}_0^b\big) &= \sum_{i=1}^{L} I\big(\mathbf{X}_{-\infty}^{-1};\tilde{Y}_{bL+i}\big|\mathbf{X}_0^b,\tilde{Y}_{bL+1}^{bL+i-1}\big)\\
&\leq \sum_{i=1}^{L}\Big(h\big(\tilde{Y}_{bL+i}\big|\mathbf{X}_0^b\big) - h\big(\tilde{Y}_{bL+i}\big|\mathbf{X}_{-\infty}^b\big)\Big)\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left((2\pi e)\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2+\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right)\right)\right]\\
&\quad -\frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left((2\pi e)\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2+\sum_{\ell=-\infty}^{-1}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2\right)\right)\right]\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left((2\pi e)\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2+\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right)\right)\right]\\
&\quad -\frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left((2\pi e)\left(\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2\right)\right)\right]\\
&= \frac{1}{2}\sum_{i=1}^{L}\mathsf{E}\left[\log\left(1+\frac{\mathsf{P}\,L\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}}{\sigma^2+\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2}\right)\right]\\
&\leq \frac{1}{2}\sum_{i=1}^{L}\log\left(1+L\,\mathrm{SNR}\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)}\right),
\end{aligned} \tag{114}$$

where the first inequality follows because conditioning cannot increase entropy and because, conditional on $\mathbf{X}_{-\infty}^b$, $\tilde{Y}_{bL+i}$ is independent of $\tilde{Y}_{bL+1}^{bL+i-1}$; the next inequality follows from the entropy maximizing property of Gaussian random variables; the subsequent inequality follows because $\sum_{\ell=-\infty}^{-1}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2 \geq 0$, $i=1,\ldots,L$; and the last inequality follows because $\sum_{\ell=0}^{b}\alpha_{b-\ell}^{(i)}X_{\ell L+1}^2 \geq 0$, $i=1,\ldots,L$.

By upper bounding

$$\sum_{\ell=b+1}^{\infty}\alpha_\ell^{(i)} \leq \sum_{\ell=b+1}^{\infty}\alpha_\ell, \qquad i=1,\ldots,L \tag{115}$$

we obtain

$$I\big(\mathbf{X}_{-\infty}^{-1};\tilde{\mathbf{Y}}_b\big|\mathbf{X}_0^b\big)\leq\frac{L}{2}\log\left(1+L\,\mathrm{SNR}\sum_{\ell=b+1}^{\infty}\alpha_\ell\right),\tag{116}$$

and (111) follows by noting that (21) implies

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

C Proof of Lemma 5

We shall show that for any $\epsilon>0$

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-(\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P})\right|\geq\epsilon\right)=0\tag{117}$$

and

$$\lim_{n\to\infty}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-(\sigma^2+\alpha^{(L)}\,\mathsf{P})\right|\geq\epsilon\right)=0.\tag{118}$$

Lemma 5 follows then by the union of events bound.

In order to prove (117) & (118), we first note that

$$\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Y}\|^2\big]=\sigma^2+\mathsf{P}+\frac{\mathsf{P}}{\lfloor n/L\rfloor}\sum_{k=1}^{\lfloor n/L\rfloor-1}\sum_{\ell=1}^{k}\alpha_{\ell L}\tag{119}$$

$$\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Z}\|^2\big]=\sigma^2+\frac{\mathsf{P}}{\lfloor n/L\rfloor}\sum_{k=1}^{\lfloor n/L\rfloor-1}\sum_{\ell=1}^{k}\alpha_{\ell L}\tag{120}$$

and therefore, by Cesáro's mean [10, Thm. 4.2.3],

$$\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Y}\|^2\big]=\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P}\tag{121}$$

$$\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Z}\|^2\big]=\sigma^2+\alpha^{(L)}\,\mathsf{P},\tag{122}$$

where $\alpha^{(L)}$ was defined in (89) 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 $n\geq n_0$

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

$$\left|\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Z}\|^2\big]-(\sigma^2+\alpha^{(L)}\,\mathsf{P})\right|\leq\varepsilon\tag{124}$$

and it follows from the triangle inequality that

$$\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-(\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P})\right|\leq\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Y}\|^2\big]\right|+\varepsilon\tag{125}$$

$$\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-(\sigma^2+\alpha^{(L)}\,\mathsf{P})\right|\leq\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Z}\|^2\big]\right|+\varepsilon.\tag{126}$$

From this we obtain

$$\begin{aligned}\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-(\sigma^2+\mathsf{P}+\alpha^{(L)}\,\mathsf{P})\right|\geq\epsilon\right)&\leq\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\big[\|\mathbf{Y}\|^2\big]\right|\geq\epsilon-\varepsilon\right)\\&\leq\frac{\mathsf{Var}\Big(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\Big)}{(\epsilon-\varepsilon)^2}\end{aligned}\tag{127}$$

and

$$\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-(\sigma^2+\alpha^{(L)}\,\mathsf{P})\right|\geq\epsilon\right)\leq\Pr\left(\left|\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2-\frac{1}{\lfloor n/L\rfloor}\mathsf{E}\left[\|\mathbf{Z}\|^2\right]\right|\geq\epsilon-\varepsilon\right)$$
$$\leq\frac{\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2\right)}{(\epsilon-\varepsilon)^2},\qquad(128)$$

with $\mathsf{Var}(A)=\mathsf{E}\left[(A-\mathsf{E}[A])^2\right]$ denoting the variance of A. Here the last inequalities in (127) & (128) follow from Chebyshev's inequality [17, Sec. 5.4].

It remains to show that

$$\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)=\lim_{n\to\infty}\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Z}\|^2\right)=0.\qquad(129)$$

We shall prove (129) for $\mathbf{Y}$. The proof for $\mathbf{Z}$ follows along the same lines. We begin by writing $\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)$ as

$$\begin{aligned}
&\mathsf{Var}\left(\frac{1}{\lfloor n/L\rfloor}\|\mathbf{Y}\|^2\right)\\
&\quad=\frac{1}{\left(\lfloor n/L\rfloor\right)^2}\mathsf{Var}\left(\sum_{k=0}^{\lfloor n/L\rfloor-1}Y_{kL+1}^2\right)\\
&\quad=\frac{1}{\left(\lfloor n/L\rfloor\right)^2}\sum_{k=0}^{\lfloor n/L\rfloor-1}\mathsf{Var}\left(Y_{kL+1}^2\right)+\frac{2}{\left(\lfloor n/L\rfloor\right)^2}\sum_{\substack{k=1,j=0\\k>j}}^{\lfloor n/L\rfloor-1}\mathsf{Cov}\left(Y_{kL+1}^2,Y_{jL+1}^2\right),
\end{aligned}\qquad(130)$$

where $\mathsf{Cov}(A,B)=\mathsf{E}[(A-\mathsf{E}[A])(B-\mathsf{E}[B])]$ denotes the covariance between A and B. We shall evaluate both terms on the RHS of (130) separately. For the sake of clarity, we shall omit the details of the derivations and show only the main steps. Unless otherwise stated these steps can be derived in a straightforward way using that

i) $\{X_{kL+1}\,,\,k\in\mathbb{Z}_0^+\}$ is a sequence of IID, zero-mean, variance-P Gaussian random variables whose fourth moments are given by $3\mathsf{P}$, while all odd moments are zero;

ii) $X_k=0$ if $k \mod L\neq1$;

iii) $\{U_k\}$ (and hence also $\{U_{kL+1}\,,\,k\in\mathbb{Z}_0^+\}$) is a zero-mean, unit-variance, stationary & weakly-mixing random process;

iv) and that $\{X_k\}$ and $\{U_k\}$ are independent of each other.

For the first sum on the RHS of (130) it suffices to show that $\mathsf{Var}(Y_{kL+1})<\infty$, $k\in\mathbb{Z}_0^+$. Indeed, this sum contains only $\lfloor n/L\rfloor$ summands and hence, when divided by $(\lfloor n/L\rfloor)^2$, this sum vanishes as n tends to infinity, given that $\mathsf{Var}(Y_{kL+1})<\infty$, $k\in\mathbb{Z}_0^+$. We have

$$\begin{aligned}
\mathsf{Var}\left(Y_{kL+1}^2\right)&=\mathsf{E}\left[Y_{kL+1}^4\right]-\left(\mathsf{E}\left[Y_{kL+1}^2\right]\right)^2\\
&\leq\mathsf{E}\left[Y_{kL+1}^4\right]\\
&=\mathsf{E}\left[\left(X_{kL+1}+\theta\left(X_1^{kL}\right)\cdot U_{kL+1}\right)^4\right]\\
&=3\mathsf{P}^2+6\mathsf{P}\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\\
&\quad+\left(\sigma^4+2\sigma^2\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}+2\mathsf{P}^2\sum_{\ell=1}^{k}\alpha_{\ell L}^2+\mathsf{P}^2\left(\sum_{\ell=1}^{k}\alpha_{kL}\right)^2\right)\mathsf{E}\left[U_{kL+1}^4\right]\\
&\leq3\mathsf{P}^2+6\mathsf{P}\left(\sigma^2+\mathsf{P}\alpha^{(L)}\right)\\
&\quad+\left(\sigma^4+2\sigma^2\mathsf{P}\alpha^{(L)}+2\mathsf{P}^2\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2+\mathsf{P}^2\left(\alpha^{(L)}\right)^2\right)\mathsf{E}\left[U_{kL+1}^4\right]
\end{aligned}\qquad(131)$$

where the second inequality follows by upper bounding $\sum_{\ell=1}^{k}\alpha_{\ell L}\leq\alpha^{(L)}$. Note that (84) implies that $\alpha^{(L)}$ and $\sum_{\ell=1}^{\infty}\alpha_{\ell L}^2$ are bounded. It follows therefore by noting that U_{kL+1} has a finite fourth moment that (for a finite P)

$$\mathsf{Var}(Y_{kL+1})<\infty. \tag{132}$$

In order to show that the second term on the RHS of (130) vanishes as n tends to infinity, we shall evaluate

$$\mathsf{Cov}(Y_{kL+1},Y_{jL+1})=\mathsf{E}\left[Y_{kL+1}^2Y_{jL+1}^2\right]-\mathsf{E}\left[Y_{kL+1}^2\right]\mathsf{E}\left[Y_{jL+1}^2\right]$$

for $k\in\mathbb{Z}^+$, $j\in\mathbb{Z}_0^+$, $k>j$. We have

$$\begin{aligned}\mathsf{E}\left[Y_{kL+1}^2Y_{jL+1}^2\right]&=\mathsf{E}\left[\left(X_{kL+1}+\theta\big(X_1^{kL}\big)\cdot U_{kL+1}\right)^2\left(X_{jL+1}+\theta\big(X_1^{jL}\big)\cdot U_{jL+1}\right)^2\right]\\&=\mathsf{P}^2+\mathsf{P}\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{j}\alpha_{\ell L}\right)+\mathsf{P}\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)+2\mathsf{P}^2\alpha_{(k-j)L}\\&\quad+\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2+\mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\mathsf{E}\left[U_{kL+1}^2U_{jL+1}^2\right]\\&\quad+2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\,\mathsf{E}\left[U_{kL+1}^2U_{jL+1}^2\right].\end{aligned}\tag{133}$$

Evaluating

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

we obtain from (134) & (133)

$$\begin{aligned}\mathsf{Cov}(Y_{kL+1},Y_{jL+1})&=2\mathsf{P}^2\alpha_{(k-j)L}+2\mathsf{P}^2\sum_{\ell=1}^{j}\alpha_{\ell L}\alpha_{(\ell+k-j)L}\mathsf{E}\left[U_{kL+1}^2U_{jL+1}^2\right]\\&\quad+\left(\sigma^2+\mathsf{P}\sum_{\ell=1}^{k}\alpha_{\ell L}\right)\left(\sigma^2+\mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{kL+1}^2U_{jL+1}^2\right]-1\right).\end{aligned}\tag{135}$$

Summing over k and j and diving by $(\lfloor n/L\rfloor)^2$ yields

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

$$
\begin{aligned}
&= \frac{2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \alpha_{\nu L} \\
&\quad + \frac{2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} \mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] \\
&\quad + \frac{2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \left(\sigma^2 + \mathsf{P} \sum_{\ell=1}^{j+\nu} \alpha_{\ell L}\right) \left(\sigma^2 + \mathsf{P} \sum_{\ell'=1}^{j} \alpha_{\ell' L}\right) \left(\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right),
\end{aligned}
\tag{136}
$$

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

The first two terms on the RHS of (136) can be upper bounded using (84)

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

Indeed, noting that $L \geq \ell_0$, we have

$$
\sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \alpha_{\nu L} < \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \varrho^{\nu L} < \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L} \tag{137}
$$

and

$$
\begin{aligned}
\sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} &< \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \left(\varrho^{2L}\right)^\ell \varrho^{\nu L} \\
&< \sum_{\nu=1}^{\lfloor n/L \rfloor} \sum_{\ell=1}^{\infty} \left(\varrho^{2L}\right)^\ell \varrho^{\nu L} \\
&= \frac{\varrho^{2L}}{1 - \varrho^{2L}} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L}.
\end{aligned}
\tag{138}
$$

Consequently with (137) we can upper bound the first term on the RHS of (136) as

$$
\begin{aligned}
\frac{2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \alpha_{\nu L} &< \frac{4\mathsf{P}^2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L} \\
&= 4\mathsf{P}^2 \frac{\lfloor n/L \rfloor - 1}{\lfloor n/L \rfloor} \frac{1}{\lfloor n/L \rfloor} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L},
\end{aligned}
\tag{139}
$$

and it follows from Cesáro's mean that this upper bound tends to zero as n tends to infinity. Likewise with (138) we can upper bound the second term on the RHS of (136) as

$$
\begin{aligned}
&\frac{2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} 2\mathsf{P}^2 \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} \mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] \\
&\qquad \leq \frac{4\mathsf{P}^2}{(\lfloor n/L \rfloor)^2} \sum_{j=0}^{\lfloor n/L \rfloor - 2} \sum_{\nu=1}^{\lfloor n/L \rfloor - 1 - j} \sum_{\ell=1}^{j} \alpha_{\ell L} \alpha_{(\ell+\nu)L} \mathsf{E}\left[U_1^4\right] \\
&\qquad < 4\mathsf{P}^2 \frac{\varrho^{2L}}{1 - \varrho^{2L}} \mathsf{E}\left[U_1^4\right] \frac{\lfloor n/L \rfloor - 1}{\lfloor n/L \rfloor} \frac{1}{\lfloor n/L \rfloor} \sum_{\nu=1}^{\lfloor n/L \rfloor} \varrho^{\nu L},
\end{aligned}
\tag{140}
$$

where the first inequality follows from the Cauchy-Schwarz inequality. As above, it follows from Cesáro's mean that this upper bound tends to zero as n tends to infinity.

It thus remains to show that the last term on the RHS of (136) vanishes as n tends to infinity. We have for each $j = 0, \ldots, \lfloor n/L \rfloor - 2$

$$\begin{aligned}
\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} & \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right) \\
&\leq \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| \\
&\leq \sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} \left(\sigma^2 + \mathsf{P}\alpha^{(L)}\right)^2 \left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| \\
&\leq \sum_{\nu=1}^{\lfloor n/L\rfloor} \left(\sigma^2 + \mathsf{P}\alpha^{(L)}\right)^2 \left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right|, \qquad (141)
\end{aligned}$$

where the first inequality follows by upper bounding $\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1 \leq \left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right|$; and the second inequality follows by upper bounding $\sum_{\ell=1}^{j}\alpha_{\ell L} \leq \sum_{\ell=1}^{j+\nu}\alpha_{\ell L} \leq \sum_{\ell=1}^{\infty}\alpha_{\ell L} = \alpha^{(L)}$. The last term on the RHS of (136) is therefore upper bounded by

$$\begin{aligned}
\frac{2}{(\lfloor n/L\rfloor)^2} & \sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor-1-j} \left(\sigma^2 + \mathsf{P}\sum_{\ell=1}^{j+\nu}\alpha_{\ell L}\right)\left(\sigma^2 + \mathsf{P}\sum_{\ell'=1}^{j}\alpha_{\ell' L}\right)\left(\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right) \\
&\leq \frac{2}{(\lfloor n/L\rfloor)^2}\sum_{j=0}^{\lfloor n/L\rfloor-2}\sum_{\nu=1}^{\lfloor n/L\rfloor} \left(\sigma^2 + \mathsf{P}\alpha^{(L)}\right)^2 \left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| \\
&= 2\left(\sigma^2 + \mathsf{P}\alpha^{(L)}\right)^2 \frac{\lfloor n/L\rfloor - 1}{\lfloor n/L\rfloor}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right|. \qquad (142)
\end{aligned}$$

It follows now from the weakly-mixing property of $\{U_k\}$ that [8, Thm. 6.1]

$$\lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - 1\right| = \lim_{n\to\infty}\frac{1}{\lfloor n/L\rfloor}\sum_{\nu=1}^{\lfloor n/L\rfloor}\left|\mathsf{E}\left[U_{\nu L+1}^2 U_1^2\right] - \mathsf{E}\left[U_{\nu L+1}^2\right]\mathsf{E}\left[U_1^2\right]\right| = 0$$

so that the last term on the RHS of (136) vanishes as n tends to infinity.

Thus (142), (140), and (139) show that (136) vanishes as n tends to infinity which in turn shows, along with (130) and (132), that

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

Together with (127), this proves (117). The proof of (118) follows along the same lines.

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] L. B. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics," *Physics Lett. A*, no. 3–4, pp. 144–149, Dec. 2002.

[3] K. E. Goodson, "Thermal conduction in electronic microstructures," in *CRC Handbook of Thermal Engineering*, 1st ed., ser. Mechanical Engineering Handbook Series, F. Keith, Ed., Dec. 1999.

[4] J. H. Lienhard IV and J. H. Lienhard V, *A Heat Transfer Textbook*, 3rd ed. Cambridge Massachusets, USA: Phlogisten Press, 2008.

[5] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. USA: Oxford University Press, 2003.

[6] 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.

[7] B. Razavi, "CMOS technology characterization for analog and RF design," *IEEE J. Solid-State Circuits*, vol. 34, no. 3, pp. 268–276, Mar. 1999.

[8] K. Petersen, *Ergodic Theory*, ser. Cambridge Studies in Advanced Mathematics 2. Cambridge University Press, 1983.

[9] G. Maruyama, "The harmonic analysis of stationary stochastic processes," *Memoirs of the faculty of science, Series A*, vol. 4, no. 1, pp. 45–106, 1949.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley & Sons, 1991.

[11] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.

[12] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[13] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[14] A. Lapidoth and S. Shamai (Shitz), "Fading channels: how perfect need 'perfect side-information' be?" *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[15] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1520–1529, Sept. 1996.

[16] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[17] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, 1968.