Bounds on the Capacity of MIMO Broadband Power Line Communications Channels - Full Version

Nir Shlezinger, Roee Shaked, and Ron Dabora

Abstract

Communications over power lines in the frequency range above 2 MHz, commonly referred to as broadband (BB) power line communications (PLC), is a central communications scenario for smart power grids. BB-PLC channels are characterized by a dominant colored *non-Gaussian* additive noise, as well as by periodic variations of the channel impulse response and the noise statistics, induced by the mains voltage. In this work we study the fundamental rate limits for multiple input-multiple output (MIMO) BB-PLC channels, modeled as periodic channels with additive non-Gaussian noise and finite memory. We present bounds on the capacity of these channels by exploiting a bijection with time-invariant MIMO channels of extended dimensions. We illustrate the resulting fundamental limits in a numerical analysis corresponding to practical MIMO BB-PLC channels.

I. Introduction

Power line communications (PLC) is an emerging technology which utilizes the existing power grid infrastructure for data transmission. PLC systems operating in the frequency range of $2-100$ MHz are commonly referred to as broadband (BB) PLC. In order to improve performance, BB-PLC systems may utilize all three wires of the indoor power transmission network to realize multiple transmit and receive ports, giving rise to multiple input-multiple output (MIMO) BB PLC scenarios [1]. The resulting MIMO channel exhibits periodicity in both the channel impulse response (CIR) as well as in the noise statistics. In particular, the CIR in MIMO BB-PLC channels is typically modeled as a multipath channel [2] with periodic variations [3], [4], where the channel outputs contain crosstalk from other wires [1], [5], while MIMO BB-PLC noise is generally modeled as a temporally correlated [2], [6], [7], spatially correlated [1], [8], cyclostationary

N. Shlezinger is with the faculty of EE, Technion, Haifa, Israel (e-mail: nirshlezinge@technion.ac.il). R. Shaked and R. Dabora are with the department of ECE, Ben-Gurion University, Be'er-Sheva, Israel (e-mail:shroee@post.bgu.ac.il; ron@ee.bgu.ac.il).
This work was supported by the Israel Science Foundation under Grant 1685/16.

[9] multivariate process. Furthermore, MIMO BB-PLC noise is typically non-Gaussian [2], [6], [10], where common models for the marginal probability density function (PDF) of BB-PLC noise include the Nakagami-m distribution [10] and the Gaussian mixture (GM) distribution [6]. Consequently, MIMO BB-PLC channels fall into the class of *MIMO periodic channels with additive non-Gaussian noise and finite memory*.

The unique model of MIMO BB-PLC channels introduces several major challenges when attempting to characterize their capacity. To avoid handling the technical difficulties, previous works which attempted to characterize the fundamental rate limits for BB-PLC channels used very simplified models which do not capture many of the special characteristics of these channels: The work [7] evaluated the capacity of BB-PLC channels by modeling them as having a linear time-invariant (LTI) CIR with additive *colored stationary Gaussian* noise; the work [4] modeled BB-PLC channels as linear, periodically time-varying (LPTV) channels with additive *white Gaussian* noise (AWGN), and evaluated the achievable rate by using a transmission scheme which utilizes orthogonal frequency division multiplexing (OFDM) signalling. Other related works are [11], which characterized the capacity of PLC channels in the *narrowband* frequency range ($0 - 500$ kHz), modeled as periodic channels with finite memory in which the noise is additive and *Gaussian*, and [12], which studied the capacity of periodic MIMO channels where again the noise was additive and *Gaussian*. We emphasize that [4], [7], [11], [12] derived expressions assuming *Gaussian noise*. Previous works on the capacity of channels with additive non-Gaussian noise, e.g., [13], [14], considered channels with memoryless and fixed CIR with i.i.d. non-Gaussian noise, and are thus not applicable to the characterization of the fundamental rate limits of MIMO BB-PLC channels. To the best of our knowledge, the fundamental limits for MIMO BB-PLC channels, accounting for the *the periodic variations of the CIR and of the noise statistics*, as well as for the *non-Gaussianity and the temporal correlation of the noise*, have not been characterized to date.

***Main Contributions*:** In this work we study the fundamental rate limits for discrete-time (DT) MIMO periodic channels with additive non-Gaussian noise and finite memory, which is the common model for MIMO BB-PLC channels. We note that when the noise is not a Gaussian process, obtaining a closed-form expression for the capacity is generally not a simple task, even for memoryless channels, and often times the approach is to characterize upper and lower bounds on the capacity, see, e.g., [15, Ch. 7.4]. To facilitate the derivation of such bounds, we first derive bounds on the capacity of a general LTI MIMO channel with additive *stationary non-*

Gaussian noise. Then, we prove that the capacity of finite-memory periodic MIMO channels with additive non-Gaussian noise can be obtained from the capacity of LTI MIMO channels with additive stationary non-Gaussian noise having extended dimensions compared to those of the original periodic MIMO channel, via a proper selection of the parameters of the extended LTI channel. Lastly, we apply the bounds on the capacity of the extended LTI model to obtain the corresponding bounds for the MIMO BB-PLC channel. This approach yields capacity bounds which depend on the PDF of the noise process *only through its entropy rate and autocorrelation function*. We use the derived bounds to numerically evaluate the capacity of practical MIMO BB-PLC models, and demonstrate that, in the high signal-to-noise ratio (SNR) regime, the achievable rate of *cyclostationary Gaussian signaling* is within a small gap of capacity. We also show that assuming the noise is Gaussian may result in *significantly underestimating* the capacity.

The rest of this paper is organized as follows: Section II formulates the problem; Section III derives the capacity bounds, and Section IV presents numerical examples; Lastly, Section V provides concluding remarks.

II. Problem Definition

***Notations*:** We use upper-case letters, e.g., X, to denote random variables (RVs), and lower-case letters, e.g., x, to denote deterministic values. Column vectors are denoted with boldface letters, e.g., $\mathbf{x}$ for a deterministic vector and $\mathbf{X}$ for a random vector; the i-th element of $\mathbf{x}$ ($i \geq 0$) is denoted with $(\mathbf{x})_i$. We use Sans-Sarif fonts to denote matrices, e.g., A, the element at the i-th row and the j-th column of A is denoted with $(\mathsf{A})_{i,j}$, the all-zero $k \times l$ matrix is denoted with $\mathsf{0}_{k \times l}$, and the $n \times n$ identity matrix is denoted with I_n. Complex conjugate, transpose, Hermitian transpose, Euclidean norm, determinant, stochastic expectation, covariance, differential entropy, and mutual information are denoted by $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\|\cdot\|$, $|\cdot|$, $\mathbb{E}\{\cdot\}$, $\mathrm{Cov}(\cdot)$, $h(\cdot)$, and $I(\cdot;\cdot)$, respectively, and a^+ denotes $\max\{0, a\}$. The sets of non-negative integers, integers, and of real numbers are denoted by $\mathcal{N}$, $\mathcal{Z}$, and $\mathcal{R}$, respectively. All logarithms are taken to base-2. Lastly, for any sequence $\mathbf{y}[i]$, $i \in \mathcal{Z}$, and integers $b_1 < b_2$, $\mathbf{y}_{b_1}^{b_2}$ is the column vector obtained by stacking $\left[(\mathbf{y}[b_1])^T, \ldots, (\mathbf{y}[b_2])^T \right]^T$ and $\mathbf{y}^{b_2} \equiv \mathbf{y}_0^{b_2}$.

***Definitions*:** We shall use of the following definitions:

Definition 1 (MIMO channel with finite-memory)**.** *A DT $n_{\rm r} \times n_{\rm t}$ MIMO channel with finite memory consists of an input sequence $\mathbf{X}[i] \in \mathcal{R}^{n_{\rm t}}$, $i \in \mathcal{N}$, an output sequence $\mathbf{Y}[i] \in \mathcal{R}^{n_{\rm r}}$, $i \in \mathcal{N}$, an initial state vector $\mathbf{S}_0 \in \mathcal{S}_0$ of finite dimensions, and a sequence of PDFs $\left\{ p\left(\mathbf{Y}^n | \mathbf{X}^n, \mathbf{S}_0\right) \right\}_{n=0}^{\infty}$.*

Definition 2 (Code)**.** *An* $[R,l]$ *code with rate* R *and blocklength* $l \in \mathcal{N}$ *consists of:* 1) *A message set* $\mathcal{U} \triangleq \{1,2,\ldots,2^{lR}\}$. 2) *An encoder* e_l *which maps each message* $u \in \mathcal{U}$ *into a codeword* $\mathbf{x}_{(u)}^{l-1} = \big[\mathbf{x}_{(u)}[0], \mathbf{x}_{(u)}[1], \ldots, \mathbf{x}_{(u)}[l-1]\big]$. 3) *A decoder* d_l *which maps the channel output* $\mathbf{y}^{l-1}$ *into a message* $\hat{u} \in \mathcal{U}$*. The encoder and decoder operate independently of the initial state.*

The set $\big\{\mathbf{x}_{(u)}^{l-1}\big\}_{u=1}^{2^{lR}}$ is referred to as the *codebook* of the $[R,l]$ code. Assuming U is uniformly selected from $\mathcal{U}$, the average probability of error, when the initial state is $\mathbf{s}_0$, is:

$$P_{\rm e}^{l}(\mathbf{s}_0) = \frac{1}{2^{lR}} \sum_{u=1}^{2^{lR}} \Pr\left(d_l\left(\mathbf{Y}^{l-1}\right) \neq u \middle| U=u, \mathbf{S}_0 = \mathbf{s}_0\right).$$

Definition 3 (Achievable rate)**.** *A rate* $R_{\rm c}$ *is called achievable if, for every* $\epsilon_1, \epsilon_2 > 0$*, there exists a positive integer* $l_0 > 0$ *such that for all integer* $l > l_0$*, there exists an* $[R,l]$ *code which satisfies* $\sup\limits_{\mathbf{s}_0 \in \mathcal{S}_0} P_{\rm e}^{l}(\mathbf{s}_0) < \epsilon_1$*, and* $R \geq R_{\rm c} - \epsilon_2$*.*

Definition 4 (Capacity)**.** *Capacity is defined as the supremum of all achievable rates.*

***Model and Problem Formulation*:** We consider a DT MIMO BB-PLC channel modeled as a *linear, non-Gaussian, MIMO periodic channel (LNGMPC)* with $\tilde{n}_{\rm r}$ receive ports and $\tilde{n}_{\rm t}$ transmit ports. Let $\tilde{m}$ be a non-negative integer which represents the *length of the memory of the channel*, $\tilde{p}_{\mathsf{G}}$ be a positive integer which represents the *period of the CIR*, and $\tilde{p}_{\mathbf{W}}$ be a positive integer which represents the *period of the noise statistics*. Let $\tilde{\mathbf{W}}[i] \in \mathcal{R}^{\tilde{n}_{\rm r}}$ be a real-valued, $\tilde{n}_{\rm r}$-dimensional, zero-mean, strict-sense cyclostationary, non-Gaussian additive noise. Thus, for any set of k integer indexes $\{i_l\}_{l=1}^{k}$, $k>0$, the joint PDF of $\tilde{\mathbf{W}}[i_1], \tilde{\mathbf{W}}[i_2], \ldots, \tilde{\mathbf{W}}[i_k]$ is equal to the joint PDF of $\tilde{\mathbf{W}}[i_1 + \tilde{p}_{\mathbf{W}}], \tilde{\mathbf{W}}[i_2 + \tilde{p}_{\mathbf{W}}], \ldots, \tilde{\mathbf{W}}[i_k + \tilde{p}_{\mathbf{W}}]$. Since the channel memory is $\tilde{m}$, then noise vectors which are more than $\tilde{m}$ instances apart are mutually independent, i.e., $\forall i_1, i_2, l_1, l_2 \in \mathcal{N}$ such that $i_2 > i_1 + l_1 + \tilde{m}$, the random vectors $\tilde{\mathbf{W}}_{i_1}^{i_1+l_1}$ and $\tilde{\mathbf{W}}_{i_2}^{i_2+l_2}$ are mutually independent. We further assume that there is no deterministic dependence between instances of $\tilde{\mathbf{W}}[i]$, i.e., $\nexists i_0$ for which $\tilde{\mathbf{W}}[i_0]$ can be expressed as a linear combination of $\big\{\tilde{\mathbf{W}}[i]\big\}_{i \neq i_0}$. Let $\big\{\tilde{\mathsf{G}}[i,\tau]\big\}_{\tau=0}^{\tilde{m}}$ denote the LPTV CIR of the MIMO BB-PLC channel, $\tilde{\mathsf{G}}[i,\tau] \in \mathcal{R}^{\tilde{n}_{\rm r} \times \tilde{n}_{\rm t}}$. The periodicity of the CIR implies that $\tilde{\mathsf{G}}[i,\tau] = \tilde{\mathsf{G}}[i + \tilde{p}_{\mathsf{G}}, \tau]$, $\forall i \in \mathcal{Z}, \tau \in \{0,1,\ldots,\tilde{m}\}$. The input-output relationship for the LNGMPC with input codeword length $\tilde{l}$ is given by

$$\tilde{\mathbf{Y}}[i] = \sum_{\tau=0}^{\tilde{m}} \tilde{\mathsf{G}}[i,\tau] \tilde{\mathbf{X}}[i-\tau] + \tilde{\mathbf{W}}[i], \quad i \in \{0,1,\ldots,\tilde{l}-1\}, \tag{1}$$

where the initial state of the channel (i.e., prior to the beginning of reception) is given by

$\tilde{\mathbf{S}}_0 = \left[\left(\tilde{\mathbf{X}}_{-\tilde{m}}^{-1}\right)^T, \left(\tilde{\mathbf{W}}_{-\tilde{m}}^{-1}\right)^T\right]^T$. The channel input is subject to a time-averaged power constraint $\tilde{P}$, as in [11, Eq. (7)]:

$$\frac{1}{\tilde{l}} \sum_{i=0}^{\tilde{l}-1} \mathbb{E}\left\{\left\|\tilde{\mathbf{X}}\left[i\right]\right\|^2\right\} \leq \tilde{P}. \tag{2}$$

Letting $\tilde{p}$ be the least common multiple of $\tilde{p}_{\mathsf{G}}$ and $\tilde{p}_{\mathbf{W}}$ which satisfies $\tilde{p} > \tilde{m}$, we obtain that the CIR and the statistics of the noise of the LNGMPC (1) are periodic with a period $\tilde{p}$, hence we refer to $\tilde{p}$ as the *period of the channel*. We also note that while the above model was stated for real signals, complex (baseband) channels can be accommodated by representing all complex vectors and matrices using real vectors and matrices.

In the following section we study the capacity of LNGMPCs, defined in (1)–(2), denoted $C_{\rm P}$.

III. The Capacity of LNGMPCs

Our main result is the characterization of upper and lower bounds on the capacity of LNGMPCs, defined in (1)–(2). This result is obtained via three steps: First, in Subsection III-A, we define a general LTI $n_{\rm r} \times n_{\rm t}$ MIMO channel with stationary non-Gaussian noise, to which we refer as the *linear non-Gaussian MIMO channel (LNGMC)*. We express the capacity of the LNGMC as a limit of the mutual information between its input and its output when the blocklength increases to infinity. Next, in Subsection III-B, We derive computable upper and lower bounds on the capacity of the LNGMC, which are stated in terms of the CIR, and of the entropy rate and autocorrelation function of the noise. Lastly, in Subsection III-C, we prove that the capacity of the LNGMPC can be obtained as the capacity of an equivalent $\tilde{p} \times \tilde{p}$ LNGMC, and use the bounds derived in Subsection III-B to state the corresponding capacity bounds for the LNGMPC.

A. The Capacity of the LNGMC

We begin with the definition of the LNGMC: Let m be a non-negative integer which represents the *length of the memory of the channel*, and let $\{\mathsf{G}[\tau]\}_{\tau=0}^{m}$ denote a set of $m+1$ real-valued $n_{\rm r}\times n_{\rm t}$ CIR matrices. Additionally, let $\mathbf{W}[i] \in \mathcal{R}^{n_{\rm r}}$ be a multivariate, real-valued, strict-sense stationary non-Gaussian additive noise process, whose mean is zero and whose temporal dependence spans a finite interval of length m, i.e., $\forall i_1, i_2, l_1, l_2 \in \mathcal{N}$ such that $i_2 > i_1 + l_1 + m$, the random vectors $\mathbf{W}_{i_1}^{i_1+l_1}$ and $\mathbf{W}_{i_2}^{i_2+l_2}$ are mutually independent. For the transmission of a block of l symbols, $\{\mathbf{X}[i]\}_{i=0}^{l-1}$, the input-output relationship is defined as

$$\mathbf{Y}[i] = \sum_{\tau=0}^{m} \mathsf{G}[\tau]\mathbf{X}[i-\tau] + \mathbf{W}[i], \quad i \in \{0, 1, \ldots, l-1\}, \tag{3}$$

where the initial state of the channel is given by $\mathbf{S}_0 = \left[\left(\mathbf{X}_{-m}^{-1}\right)^T, \left(\mathbf{W}_{-m}^{-1}\right)^T\right]^T$. The channel input is subject to a time-averaged power constraint P, i.e.,

$$\frac{1}{l}\sum_{i=0}^{l-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P. \tag{4}$$

The capacity of the LNGMC defined above is stated in the following proposition:

Proposition 1. *The capacity of the LNGMC defined in* (3)*, subject to* (4)*, is given by*

$$C_{\mathrm{L}} = \lim_{n\to\infty}\frac{1}{n}\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P} I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right).$$

[A proof is given in Appendix A]

Comment 1. Previous works on the capacity of finite-memory channels with Gaussian noise, e.g., [20], obtained a capacity result in the frequency domain, by transforming the channel into a set of parallel independent channels, for which capacity is expressed as an explicit integral. When the noise is non-Gaussian, switching to the frequency domain results in the noise components at different frequency bins having statistical dependence (even if the noise samples are independent in the time domain). For this reason, our analysis is carried out in the time domain, and the capacity is stated in terms of an asymptotic limit. Nonetheless, the *bounds* on the capacity of LNGMCs, derived in Prop. 2, are stated in closed-form (not as limiting expressions) in the frequency domain.

Prop. 1 implies that the capacity of the LNGMC can be computed by setting $\mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}$. We note that setting the signal component in the initial state to zero was stated as a model assumption in [20] and [21], which studied the capacity of point-to-point channels with memory and Gaussian noise.

B. Bounds on the Capacity of the LNGMC

Next, based on the capacity expression in Prop. 1, we derive upper and lower bounds on C_{L}, which depend on the PDF of the non-Gaussian noise $\mathbf{W}[i]$ only through its autocorrelation function, $\mathsf{C}_{\mathbf{W}}[\tau] \triangleq \mathbb{E}\left\{\mathbf{W}[i+\tau]\left(\mathbf{W}[i]\right)^T\right\}$, and its entropy rate, $\bar{H}_{\mathbf{W}} \triangleq \lim\limits_{l\to\infty}\frac{1}{l}h\left(\mathbf{W}^{l-1}\right)$. Note that the strict-sense stationarity and finite memory of $\mathbf{W}[i]$ imply that $\bar{H}_{\mathbf{W}} = h\left(\mathbf{W}[m]|\,\mathbf{W}^{m-1}\right)$ [17, Ch. 12.5].

In the statement of the bounds we make use of the following additional definitions: For any $\omega \in [-\pi, \pi)$, we define the $n_{\rm r} \times n_{\rm t}$ matrix $\mathsf{G}'(\omega) \triangleq \sum\limits_{\tau=0}^{m} \mathsf{G}[\tau] e^{-j\omega\tau}$, and the $n_{\rm r} \times n_{\rm r}$ matrix $\mathsf{C}'_{\mathbf{W}}(\omega) \triangleq \sum\limits_{\tau=-m}^{m} \mathsf{C}_{\mathbf{W}}[\tau] e^{-j\omega\tau}$, and we let $\left\{\alpha'_k(\omega)\right\}_{k=0}^{n_{\rm r}-1}$ and $\left\{\lambda'_k(\omega)\right\}_{k=0}^{n_{\rm t}-1}$ denote the eigenvalues of $\mathsf{G}'(\omega)\big(\mathsf{G}'(\omega)\big)^H$ and of $\big(\mathsf{G}'(\omega)\big)^H\big(\mathsf{C}'_{\mathbf{W}}(\omega)\big)^{-1}\mathsf{G}'(\omega)$, respectively. Next, let $\bar{H}_{G,\mathbf{W}}$ denote the entropy rate of a zero-mean $n_{\rm r} \times 1$ multivariate *Gaussian* process whose autocorrelation function is equal to $\mathsf{C}_{\mathbf{W}}[\tau]$. From [22, Sec. III] the entropy rate $\bar{H}_{G,\mathbf{W}}$ can be expressed as

$$\bar{H}_{G,\mathbf{W}} = \frac{1}{4\pi} \int\limits_{\omega=-\pi}^{\pi} \log \left|2\pi e \mathsf{C}'_{\mathbf{W}}(\omega)\right| d\omega. \tag{5a}$$

Let $C_{\rm G}$ denote the capacity of the channel defined in (3) when the noise $\mathbf{W}[i]$ is *Gaussian*, subject to the constraint (4) and to setting $\mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_{\rm t}\cdot m}$. In [21, Eqn. (9)] the capacity of LTI MIMO channels with additive stationary Gaussian noise was characterized[1], assuming $\mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_{\rm t}\cdot m}$. Using [21, Eqn. (9)] we can write

$$C_{\rm G} = \frac{1}{4\pi} \sum_{k=0}^{n_{\rm t}-1} \int\limits_{\omega=-\pi}^{\pi} \Big(\log\left(\Delta' \cdot \lambda'_k(\omega)\right)\Big)^{+} d\omega, \tag{5b}$$

where Δ' is set s.t. $\frac{1}{2\pi}\sum\limits_{k=0}^{n_{\rm t}-1}\int\limits_{\omega=-\pi}^{\pi}\Big(\Delta' - \left(\lambda'_k(\omega)\right)^{-1}\Big)^{+} d\omega = P$.

We next state an upper bound and two lower bounds on the capacity of the LNGMC using $\bar{H}_{\mathbf{W}}$, $\bar{H}_{G,\mathbf{W}}$, and $C_{\rm G}$. These bounds are stated in the following proposition:

Proposition 2. *The capacity of the LNGMC defined in* (3)*, subject to the constraint* (4)*, satisfies*

$$C_{\rm G} \leq C_{\rm L} \leq C_{\rm G} + \bar{H}_{G,\mathbf{W}} - \bar{H}_{\mathbf{W}}. \tag{6a}$$

Moreover, if $n_{\rm r} = n_{\rm t}$ *and* $\mathsf{G}[0]$ *is invertible, then* $C_{\rm L}$ *satisfies*

$$C_{\rm L} \geq \frac{n_{\rm t}}{2} \log\left(\frac{2\pi e P}{n_{\rm t}} \cdot 2^{\frac{1}{2\pi\cdot n_{\rm t}}\sum\limits_{k=0}^{n_{\rm t}-1}\int\limits_{\omega=-\pi}^{\pi}\log\left(\alpha'_k(\omega)\right)d\omega} + 2^{\frac{2}{n_{\rm r}}\bar{H}_{\mathbf{W}}}\right) - \bar{H}_{\mathbf{W}}. \tag{6b}$$

[A proof is given in Appendix B]

[1]We note that [21, Thm. 1] is stated for a per-codeword power constraint. However, it follows from [21, Sec. 3.1] and from [15, Ch. 7.3] that the proof of [21, Thm. 1] also holds subject to the time-averaged power constraint (4).

C. Capacity Analysis for LNGMPCs

In order to obtain bounds on the capacity of LNGMPCs, we first prove that any LNGMPC can be equivalently represented as an LNGMC, and then apply the capacity bounds derived for LNGMCs in Prop. 2 to bound the capacity of the equivalent representation. To that aim, define the $\tilde{p}\cdot\tilde{n}_{\rm r}\times 1$ random vector $\mathbf{W}_{\rm DCD}\left[\tilde{i}\right]\triangleq\tilde{\mathbf{W}}_{\tilde{i}\cdot\tilde{p}}^{(\tilde{i}+1)\cdot\tilde{p}-1}$, and additionally, define two $\tilde{p}\cdot\tilde{n}_{\rm r}\times\tilde{p}\cdot\tilde{n}_{\rm t}$ matrices, $\mathsf{G}_{\rm DCD}[0]$ and $\mathsf{G}_{\rm DCD}[1]$, as follows:

$$\mathsf{G}_{\rm DCD}[0]\triangleq\begin{bmatrix}\tilde{\mathsf{G}}[0,0] & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & & \ddots & \vdots\\ \tilde{\mathsf{G}}[\tilde{m},\tilde{m}] & \cdots & \tilde{\mathsf{G}}[\tilde{m},0] & \cdots & 0\\ \vdots & \ddots & & \ddots & \vdots\\ 0 & \cdots & \tilde{\mathsf{G}}[\tilde{p}-1,\tilde{m}] & \cdots & \tilde{\mathsf{G}}[\tilde{p}-1,0]\end{bmatrix},$$

$$\mathsf{G}_{\rm DCD}[1]\triangleq\begin{bmatrix}0\cdots 0 & \tilde{\mathsf{G}}[0,\tilde{m}] & \cdots & \tilde{\mathsf{G}}[0,1]\\ \vdots\quad\vdots & & \ddots & \vdots\\ 0\cdots 0 & 0 & & \tilde{\mathsf{G}}[\tilde{m}-1,\tilde{m}]\\ \vdots\quad\vdots & \vdots & & \vdots\\ 0\cdots 0 & 0 & \cdots & 0\end{bmatrix}.$$

As $\mathbf{W}_{\rm DCD}\left[\tilde{i}\right]$ is given by the decimated components decomposition (DCD) [23] of $\tilde{\mathbf{W}}[i]$, the strict-sense cyclostationarity of $\tilde{\mathbf{W}}[i]$ induces a strict-sense stationarity for $\mathbf{W}_{\rm DCD}\left[\tilde{i}\right]$. Using these definitions, we construct an LNGMC with a $\tilde{p}\cdot\tilde{n}_{\rm t}\times 1$ input $\mathbf{X}_{\rm DCD}\left[\tilde{i}\right]$ and a $\tilde{p}\cdot\tilde{n}_{\rm r}\times 1$ output $\mathbf{Y}_{\rm DCD}\left[\tilde{i}\right]$, which satisfies the following input-output relationship for a sequence of l channel inputs:

$$\mathbf{Y}_{\rm DCD}\left[\tilde{i}\right]=\sum_{\tilde{\tau}=0}^{1}\mathsf{G}_{\rm DCD}\left[\tilde{\tau}\right]\mathbf{X}_{\rm DCD}\left[\tilde{i}-\tilde{\tau}\right]+\mathbf{W}_{\rm DCD}\left[\tilde{i}\right], \tag{7}$$

$\tilde{i}\in\{0,1,\ldots,l-1\}$, where the channel input to the LNGMC (7) has to satisfy an average power constraint

$$\frac{1}{l}\sum_{\tilde{i}=0}^{l-1}\mathbb{E}\left\{\left\|\mathbf{X}_{\rm DCD}\left[\tilde{i}\right]\right\|^2\right\}\leq P_{\rm DCD}=\tilde{p}\cdot\tilde{P}. \tag{8}$$

Since $\tilde{p}>\tilde{m}$, the initial state of the LNGMC is $\mathbf{S}_{0,{\rm DCD}}=\left[\mathbf{X}_{\rm DCD}^T[-1],\mathbf{W}_{\rm DCD}^T[-1]\right]^T$. The relationship between the capacity of the LNGMPC in (1)–(2), denoted $C_{\rm P}$, and the capacity of the LNGMC in (7)–(8), denoted $C_{\rm DCD}$, is stated in the following theorem:

Theorem 1. *The capacity of the LNGMPC defined in* (1)*, subject to* (2)*, satisfies*

$$C_{\rm P} = \frac{1}{\tilde{p}} C_{\rm DCD}. \tag{9}$$

[A proof is given in Appendix C]

Next, using Thm. 1 and Prop. 2, we derive lower and upper bounds on the capacity of the LNGMPC. To that aim, define the $\tilde{p} \cdot \tilde{n}_{\rm r} \times \tilde{p} \cdot \tilde{n}_{\rm r}$ autocorrelation function $\mathsf{C}_{\mathbf{W}_{\rm DCD}}[\tilde{\tau}] \triangleq \mathbb{E}\Big\{ \mathbf{W}_{\rm DCD}\left[\tilde{i}+\tilde{\tau}\right] \left(\mathbf{W}_{\rm DCD}\left[\tilde{i}\right]\right)^T \Big\}$, the entropy rate $\bar{H}_{\mathbf{W}_{\rm DCD}} \triangleq \lim\limits_{n\to\infty} \frac{1}{n} h\left(\mathbf{W}_{\rm DCD}^{n-1}\right)$, the $\tilde{p}\cdot\tilde{n}_{\rm r}\times\tilde{p}\cdot\tilde{n}_{\rm t}$ matrix $\mathsf{G}'_{\rm DCD}(\omega) \triangleq \sum\limits_{\tilde{\tau}=0}^{1} \mathsf{G}_{\rm DCD}[\tilde{\tau}] e^{-j\omega\tilde{\tau}}$, and the $\tilde{p}\cdot\tilde{n}_{\rm r}\times\tilde{p}\cdot\tilde{n}_{\rm r}$ matrix $\mathsf{C}'_{\mathbf{W}_{\rm DCD}}(\omega) \triangleq \sum\limits_{\tilde{\tau}=-1}^{1} \mathsf{C}_{\mathbf{W}_{\rm DCD}}[\tilde{\tau}] e^{-j\omega\tilde{\tau}}$. Additionally, let $\left\{\alpha'_{{\rm DCD},k}(\omega)\right\}_{k=0}^{\tilde{p}\cdot\tilde{n}_{\rm r}-1}$ and $\left\{\lambda'_{{\rm DCD},k}(\omega)\right\}_{k=0}^{\tilde{p}\cdot\tilde{n}_{\rm t}-1}$ be the eigenvalues of $\mathsf{G}'_{\rm DCD}(\omega)\left(\mathsf{G}'_{\rm DCD}(\omega)\right)^H$ and of $\left(\mathsf{G}'_{\rm DCD}(\omega)\right)^H \left(\mathsf{C}'_{\mathbf{W}_{\rm DCD}}(\omega)\right)^{-1} \mathsf{G}'_{\rm DCD}(\omega)$, respectively, and also let $\bar{H}_{G,\mathbf{W}_{\rm DCD}}$ denote the entropy rate of a zero mean $\tilde{p}\cdot\tilde{n}_{\rm r}\times 1$ *Gaussian* process with autocorrelation function $\mathsf{C}_{\mathbf{W}_{\rm DCD}}[\tilde{\tau}]$. $\bar{H}_{G,\mathbf{W}_{\rm DCD}}$ can be computed via (5a) with $\mathsf{C}'_{\mathbf{W}_{\rm DCD}}(\omega)$ instead of $\mathsf{C}'_{\mathbf{W}}(\omega)$. Lastly, let $C_{{\rm DCD},G}$ be the capacity of the LNGMC (7) when the noise $\mathbf{W}_{\rm DCD}\left[\tilde{i}\right]$ is assumed to be *Gaussian* with autocorrelation function $\mathsf{C}_{\mathbf{W}_{\rm DCD}}[\tilde{\tau}]$. Thus, $C_{{\rm DCD},G}$ is obtained using (5b) with $\tilde{p}\cdot\tilde{n}_{\rm t}$, $\lambda'_{{\rm DCD},k}(\omega)$, and $P_{\rm DCD}$ replacing $n_{\rm t}$, $\lambda'_k(\omega)$, and P, respectively. Noting that $\mathsf{G}_{\rm DCD}[0]$ has a full rank if and only if $\tilde{\mathsf{G}}\left[\tilde{i},0\right]$ has a full rank for every $\tilde{i}\in\{0,1,\ldots,\tilde{p}-1\}\triangleq\tilde{\mathcal{P}}$, then, by combining Thm. 1 with Prop. 2, the following bounds on $C_{\rm P}$ are obtained:

Corollary 1. *The capacity of the LNGMPC defined in* (1)*, subject to* (2)*, satisfies*

$$\frac{1}{\tilde{p}} C_{{\rm DCD},G} \le C_{\rm P} \le \frac{1}{\tilde{p}}\left(C_{{\rm DCD},G} + \bar{H}_{G,\mathbf{W}_{\rm DCD}} - \bar{H}_{\mathbf{W}_{\rm DCD}}\right). \tag{10a}$$

Moreover, if $\tilde{n}_{\rm r}=\tilde{n}_{\rm t}$ *and* $\tilde{\mathsf{G}}\left[\tilde{i},0\right]$ *is non-singular for every* $\tilde{i}\in\tilde{\mathcal{P}}$*, then* $C_{\rm P}$ *also satisfies*

$$C_{\rm P} \ge \frac{\tilde{n}_{\rm t}}{2}\log\left(\frac{2\pi e\tilde{P}}{\tilde{n}_{\rm t}}\cdot 2^{\frac{1}{2\pi\cdot\tilde{p}\cdot\tilde{n}_{\rm t}}\sum\limits_{k=0}^{\tilde{p}\cdot\tilde{n}_{\rm t}-1}\int\limits_{\omega=-\pi}^{\pi}\log\left(\alpha'_{{\rm DCD},k}(\omega)\right)d\omega} + 2^{\frac{2}{\tilde{p}\cdot\tilde{n}_{\rm r}}\bar{H}_{\mathbf{W}_{\rm DCD}}}\right) - \frac{1}{\tilde{p}}\bar{H}_{\mathbf{W}_{\rm DCD}}. \tag{10b}$$

Comment 2. Note that (6b) also lower bounds the achievable rate of the LNGMC with *stationary Gaussian input*. This implies that (10b) constitutes a lower bound on the achievable rate of LNGMPCs with cyclostationary Gaussian input. Consequently, when (10b) coincides with the upper bound in (10a), then cyclostationary Gaussian inputs are optimal.

IV. Numerical Examples

In this section we numerically evaluate the capacity bounds derived in Section III for MIMO BB-PLC channels. As BB-PLC channels exhibit a broad range of frequency responses and noise power values, depending on the topology of the power line network and on the appliances connected to the network, [2], [7], [10], we consider a wide range of SNRs.

We consider a passband 2×2 MIMO BB-PLC channel. The multivariate LPTV CIR $\tilde{\mathsf{G}}[i,\tau]$ is generated using the method proposed in [5] for generating MIMO BB-PLC channels: Specifically, we first generate four real LPTV CIRs with period $\tilde{p}_{\mathsf{G}} = 240$ and memory length $\tilde{m} = 4$ using the channel generator proposed in [3], and denote the generated channels as $\left\{\tilde{g}_k[i,\tau]\right\}_{k=1}^{4}$. Then, setting $\rho = 0.9$, the multivariate LPTV CIR is obtained via

$$\tilde{\mathsf{G}}[i,\tau] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{1/2} \begin{bmatrix} \tilde{g}_1[i,\tau] & \tilde{g}_2[i,\tau] \\ \tilde{g}_3[i,\tau] & \tilde{g}_4[i,\tau] \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{1/2}.$$

The additive multivariate noise is a temporally and spatially correlated cyclostationary GM process generated as follows: First, we generate a real i.i.d. 2×1 GM process $\tilde{\mathbf{U}}[i]$ whose PDF is a weighted sum of three Gaussian PDFs $\mathcal{N}\big([5,4]^T, 5\cdot \mathsf{I}_2\big)$, $\mathcal{N}\big([-8,-16]^T, 2\cdot \mathsf{I}_2\big)$, and $\mathcal{N}\big([-19,4]^T, \mathsf{I}_2\big)$, with weights $\{0.7, 0.2, 0.1\}$, respectively. Then, we generate a multivariate LPTV spectral shaping filter, $\tilde{\mathsf{F}}[i,\tau]$, with period $\tilde{p}_{\mathbf{W}} = 120$ (i.e., $\tilde{p} = 240$) and memory length $\tilde{m} = 4$, based on the method described in [8] for constructing a spectral correlation profile for MIMO BB-PLC channels: Let $\rho_{\mathbf{W}}(\omega)$ be a 2π-periodic function representing the spectral variations in the spatial correlation. Following [8, Fig. 5], we set $\rho_{\mathbf{W}}(\omega) = 0.7 - \frac{|\omega|}{2\pi}$ for $|\omega| < \pi$. Let $s[i,\omega]$ be the instantaneous power spectral densitys (PSDs), corresponding to the *'heavily disturbed'* profile based on [9]. Finally, we set

$$\tilde{\mathsf{F}}'[i,\omega] = \begin{bmatrix} 1 & \rho_{\mathbf{w}}(\omega) \\ \rho_{\mathbf{w}}(\omega) & 1 \end{bmatrix}^{1/2} \begin{bmatrix} s[i,\omega] & 0 \\ 0 & s[i,\omega] \end{bmatrix}^{1/2},$$

and $\tilde{\mathsf{F}}[i,\tau] = \frac{1}{2\pi}\int\limits_{\omega=-\pi}^{\pi} \tilde{\mathsf{F}}'[i,\omega] e^{j\omega\tau} d\omega$. The noise signal $\tilde{\mathbf{W}}[i]$ is obtained as the output of the LPTV filter $\tilde{\mathsf{F}}[i,\tau]$ whose input is $\tilde{\mathbf{U}}[i]$.

Defining $\mathrm{SNR} = \frac{\tilde{P}}{\frac{1}{\tilde{p}}\sum\limits_{i=1}^{\tilde{p}} \mathbb{E}\left\{\|\tilde{\mathbf{W}}[i]\|^2\right\}}$, we depict in Fig. 1 the capacity bounds for the MIMO BB-PLC channel vs. SNR. Note that the lower bound in (10b) is much tighter than the lower bound in (10a) for the entire SNR region, indicating the significant mismatch induced by assuming

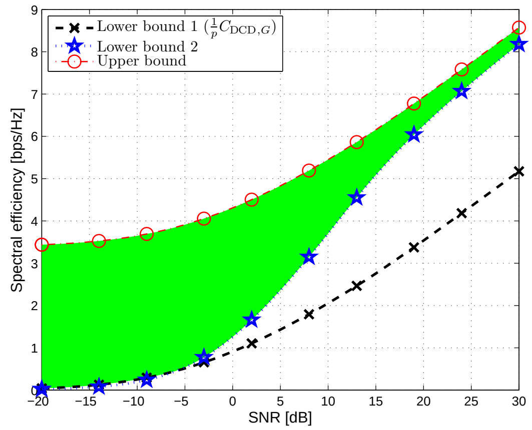

Fig. 1. Capacity comparison. MIMO BB-PLC channel.

that the noise is Gaussian. We also note that the gap between the maximal lower bound and the upper bound in Fig. 1 varies from 3.05 bps/Hz at SNR of 0 dB to 0.45 bps/Hz at high SNRs. We conclude that by using cyclostationary Gaussian inputs, it is possible to obtain an achievable rate which is very close to capacity at high SNRs. Finally, we note that for the considered channel, a 2×2 MIMO BB-PLC system utilizing a frequency band of 100 MHz, as in the ITU-T G.9963 standard [24], can achieve data rates approaching one Gbps at high SNRs.

V. Conclusions

In this paper we derived upper and lower bounds on the capacity of MIMO BB-PLC channels, modeled as finite-memory periodic MIMO channels with additive non-Gaussian noise. The capacity bounds derived depend on the noise distribution only through its entropy rate and autocorrelation function. Our numerical evaluations demonstrate the tightness of the proposed bounds at high SNRs, and illustrate the significant loss resulting from assuming that the noise is Gaussian in the computation of the capacity. We conclude that the Gaussian noise assumption may lead to inherently suboptimal schemes.

Appendix

A. *Proof of Proposition 1*

The information stability of stationary channels with finite memory, in which the input and the output are taken from *discrete and finite alphabets*, was shown in [25], see also [18, Sec. 1.5]. This results also holds for arbitrary alphabets, see, e.g. [19, Thm. 6]. In the following we provide a detailed proof of Proposition 1, namely, we prove the information stability of LNGMCs, without using information-spectrum arguments.

The following definitions are used in the sequel:

Definition A.1 (A memoryless channel)**.** *A channel with input* $\mathbf{X}[i]$*, output* $\mathbf{Y}[i]$*, and initial state* $\mathbf{S}_0$*, is called memoryless*[2] *if for every* $l \in \mathcal{N}$*,*

$$p\left(\mathbf{Y}^{l-1}|\mathbf{X}^{l-1}, \mathbf{S}_0\right) = \prod_{i=0}^{l-1} p\left(\mathbf{Y}[i]\,|\,\mathbf{X}[i]\right). \tag{A.1}$$

Definition A.2 (An n-block memoryless channel)**.** *A channel with input* $\mathbf{X}[i]$*, output* $\mathbf{Y}[i]$*, and initial state* $\mathbf{S}_0$*, is called* n*-block memoryless if for every positive integer* b

$$p\left(\mathbf{Y}^{n\cdot b-1}|\mathbf{X}^{n\cdot b-1}, \mathbf{S}_0\right) = \prod_{b'=1}^{b} p\left(\mathbf{Y}_{n\cdot(b'-1)}^{n\cdot b'-1}\,\middle|\,\mathbf{X}_{n\cdot(b'-1)}^{n\cdot b'-1}\right).$$

Def. A.2 is a specialization of the definition of n-block memoryless broadcast channels, introduced in [26, Eq. (8)], to point-to-point channels. Note that codewords of any length can be transmitted over n-block memoryless channels, however, when the length of the codeword is an integer multiple of the channel block memory n, then *the average probability of error is independent of the initial state* $\mathbf{S}_0$ [26, Sec. II]. This follows since the output of the channel corresponding to the transmitted codeword is independent of the initial channel state, by the definition of the channel.

The outline of the proof is as follows:

- First, for a given $n > m$ we construct an n-block memoryless non-Gaussian MIMO channel (n-MNGMC) based on the LNGMC by considering the last $n - m$ vector channel outputs of each n-block: The output of the n-MNGMC at time $i \in \mathcal{N}$ is defined as the output of the LNGMC if the remainder of the division of i by n, denoted $\left((i)\right)_n$, satisfies $\left((i)\right)_n \geq m$, while for $\left((i)\right)_n < m$ the output of the n-MNGMC is set to "not defined". The n-MNGMC is subject to the power constraint (4) on the channel input, similarly to the LNGMC.
- Next, in Lemma A.1 we characterize the capacity of the n-MNGMC.
- Then, we show that the capacity of the LNGMC can be obtained as the capacity of the n-MNGMC by taking $n \to \infty$. This step is carried out in Lemmas A.2–A.5.

[2]The general notion of memoryless channels as in, e.g., [34, Sec. II-A], requires that $p\left(\mathbf{Y}[i]|\,\mathbf{Y}^{i-1}, \mathbf{X}^{i}, \mathbf{S}_0\right) = p\left(\mathbf{Y}[i]|\,\mathbf{X}[i]\right)$ for all $i \in \mathcal{N}$. When no feedback to the transmitter is present, the general definition specializes Def. A.1. Since we assume that no feedback is present, we use (A.1) as the definition for memorylessness, as in [15, Ch. 7.1].

- Lastly, in Lemma A.6 we show that in the asymptotic regime of $n \to \infty$, the capacity of the n-MNGMC coincides with Prop. 1.

The capacity of the n-MNGMC is stated in the following lemma:

Lemma A.1. *The capacity of the n-MNGMC is given by*

$$C_n^{\mathrm{M}} = \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \frac{1}{n} I\left(\mathbf{X}^{n-1}; \mathbf{Y}_m^{n-1}\right). \tag{A.2}$$

Proof: In order to obtain the capacity of the n-MNGMC, we first show that (A.2) denotes the maximum achievable rate for the n-MNGMC when considering *only codes whose blocklength is an integer multiple of* n, i.e, $[R, b \cdot n]$ codes, where $b \in \mathcal{N}$. Then, we show that any rate achievable for the n-MNGMC can be achieved by considering only codes whose blocklength is an integer multiple of n.

Let us consider the n-MNGMC *constrained to using only codes whose blocklength is an integer multiple of* n. In this case, we can represent the n-MNGMC as an equivalent $n \cdot n_{\mathrm{t}} \times (n-m) \cdot n_{\mathrm{r}}$ MIMO channel (i.e., without loss of information), via the following assignments: Define the input of the transformed channel to be the $n \cdot n_{\mathrm{t}} \times 1$ vector $\mathbf{X}_{\mathrm{eq}}[i'] \triangleq \mathbf{X}_{i' \cdot n}^{(i'+1) \cdot n-1}$, $i' \geq 0$, and the output of the transformed channel to be the $(n-m) \cdot n_{\mathrm{r}} \times 1$ vector $\mathbf{Y}_{\mathrm{eq}}[i'] \triangleq \mathbf{Y}_{i' \cdot n+m}^{(i'+1) \cdot n-1}$. The transformation between the inputs and the outputs of the original n-MNGMC and of the transformed MIMO channel is clearly bijective, and thus, the capacity of the transformed MIMO channel is equal to the capacity of the original n-MNGMC, as long as the power constraint of the equivalent channel is obtained from that of the original channel as

$$\frac{1}{l'} \sum_{i'=0}^{l'-1} \mathbb{E}\left\{\|\mathbf{X}_{\mathrm{eq}}[i']\|^2\right\} \leq n \cdot P, \tag{A.3}$$

which follows since the n-MNGMC is constrained to using codes whose blocklength is $n \cdot l'$, $l' \in \mathcal{N}$. Note that the power constraint (A.3) is an average power constraint, the transmitter in the transformed channel has n times more antennas than the transmitter in the n-MNGMC, and the receiver in the transformed channel has $(n-m)$ times more antennas than the receiver in the n-MNGMC. Since the n-MNGMC is n-block memoryless, it follows that the *transformed MIMO channel is memoryless*, hence, from [15, Ch. 7.3] we obtain that its achievable rate, for a given distribution on the input $\mathbf{X}_{\mathrm{eq}}$, is given by $I(\mathbf{X}_{\mathrm{eq}}; \mathbf{Y}_{\mathrm{eq}}) = I(\mathbf{X}^{n-1}; \mathbf{Y}_m^{n-1})$. As the original power constraint (4) becomes (A.3), and as each channel use in the transformed MIMO channel corresponds to n channel uses in the n-MNGMC, it follows that the maximal achievable rate of the n-MNGMC, measured in bits per channel use, subject to the restriction that only codes whose blocklength is an integer multiple of n are allowed, is given by (A.2).

It remains to show that any rate achievable for the n-MNGMC can be achieved by considering only codes whose blocklength is an integer multiple of n: Consider a rate $R_{\rm s}$ that is achievable for the n-MNGMC and fix $\epsilon_1 > 0$ and $\epsilon_2 > 0$. From Def. 3 it follows that $\exists l_0 > 0$ such that $\forall l > l_0$ there exists an $[R, l]$ code which satisfies Def. 3. Thus, by setting b_0 as the smallest integer for which $b_0 \cdot n \geq l_0$, it follows that for all integer $b > b_0$ there exists an $[R, b \cdot n]$ code which satisfies Def. 3, which implies that the rate $R_{\rm s}$ is also achievable when considering only codes whose blocklength is an integer multiple of n. It thus directly follows that (A.2) is the maximum achievable rate for the n-MNGMC. ■

Next, we define

$$C_n^{\rm L}\left(\mathbf{s}_0\right) \triangleq \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \frac{1}{n} I\left(\mathbf{X}^{n-1}; \mathbf{Y}^{n-1} \middle| \mathbf{S}_0 = \mathbf{s}_0\right). \tag{A.4}$$

Lemma A.2 upper bounds $C_{\rm L}$ using $C_n^{\rm L}\left(\mathbf{s}_0\right)$:

Lemma A.2. *For every $\epsilon_1 \in (0,1)$, $\epsilon_2 > 0$, the capacity of the LNGMC satisfies*

$$C_{\rm L} \leq \frac{1}{1-\epsilon_1} \inf_{\mathbf{s}_0 \in \mathcal{S}_0} C_n^{\rm L}\left(\mathbf{s}_0\right) + \frac{1}{(1-\epsilon_1)n} + \epsilon_2. \tag{A.5}$$

Proof: We prove the lemma by showing that any rate $R_{\rm c}$ achievable for the LNGMC satisfies $R_{\rm c} \leq \liminf\limits_{n\to\infty} C_n^{\rm L}\left(\mathbf{s}_0\right)$ for any initial state $\mathbf{s}_0$. By definition, if R is achievable for the LNGMC then, for every $\epsilon_1, \epsilon_2 > 0$ and for all sufficiently large n, there exists an $[R_1, n]$ code, i.e., a code with rate R_1, blocklength n, and a message set $\mathcal{U}$ containing 2^{nR_1} messages, where each transmitted message U is uniformly selected from $\mathcal{U}$, such that Def. 3 is satisfied. Let $\hat{U}$ denote the estimate of U at the receiver. Fix an initial state $\mathbf{s}_0 \in \mathcal{S}_0$, and recall that from Fano's inequality [17, Sec. 2.10] it follows that

$$\begin{aligned} H\left(U \middle| \mathbf{Y}^{n-1}, \mathbf{S}_0 = \mathbf{s}_0\right) &\leq 1 + \Pr\left(U \neq \hat{U} \middle| \mathbf{S}_0 = \mathbf{s}_0\right) nR_1 \\ &\stackrel{(a)}{\leq} 1 + \epsilon_1 \cdot nR_1, \end{aligned} \tag{A.6}$$

where (a) follows from Def. 3 since $\Pr\left(U \neq \hat{U} \middle| \mathbf{S}_0 = \mathbf{s}_0\right) \leq \sup\limits_{\mathbf{s}_0 \in \mathcal{S}_0} \Pr\left(U \neq \hat{U} | \mathbf{S}_0 = \mathbf{s}_0\right) \leq \epsilon_1$.

Therefore,

$$\begin{aligned}I\left(U;\mathbf{Y}^{n-1}\middle|\,\mathbf{S}_0=\mathbf{s}_0\right)&=H\left(U|\,\mathbf{S}_0=\mathbf{s}_0\right)-H\left(U|\,\mathbf{Y}^{n-1},\mathbf{S}_0=\mathbf{s}_0\right)\\&\stackrel{(a)}{\geq}H\left(U|\,\mathbf{S}_0=\mathbf{s}_0\right)-1-\epsilon_1\cdot nR_1\\&\stackrel{(b)}{=}nR_1-1-\epsilon_1\cdot nR_1,\end{aligned}\tag{A.7}$$

where (a) follows from (A.6), and (b) follows since U is uniformly distributed over $\mathcal{U}$ and independent of $\mathbf{S}_0$, thus $H\left(U|\,\mathbf{S}_0=\mathbf{s}_0\right)=H\left(U\right)=nR_1$. Combining Def. 3 and (A.7) leads to

$$R_{\rm c}-\epsilon_2\leq\frac{I\left(U;\mathbf{Y}^{n-1}|\,\mathbf{S}_0=\mathbf{s}_0\right)+1}{n\left(1-\epsilon_1\right)},$$

thus,

$$\begin{aligned}\left(1-\epsilon_1\right)\left(R_{\rm c}-\epsilon_2\right)-\frac{1}{n}&\leq\frac{1}{n}I\left(U;\mathbf{Y}^{n-1}\middle|\,\mathbf{S}_0=\mathbf{s}_0\right)\\&\stackrel{(a)}{\leq}\sup_{p\left(\mathbf{X}^{n-1}\right):\frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}\frac{1}{n}I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}\middle|\,\mathbf{S}_0=\mathbf{s}_0\right)\\&=C_n^{\rm L}\left(\mathbf{s}_0\right),\end{aligned}\tag{A.8}$$

where (a) follows from the data processing lemma [17, Ch. 2.8] as $U|\mathbf{S}_0=\mathbf{s}_0\leftrightarrow\mathbf{X}^{n-1}|\mathbf{S}_0=\mathbf{s}_0\leftrightarrow\mathbf{Y}^{n-1}|\mathbf{S}_0=\mathbf{s}_0$ form a Markov chain. Dividing both sides of (A.8) by $1-\epsilon_1$ yields

$$R_{\rm c}\leq\frac{C_n^{\rm L}\left(\mathbf{s}_0\right)}{1-\epsilon_1}+\frac{1}{n(1-\epsilon_1)}+\epsilon_2,\qquad\forall\mathbf{s}_0\in\mathcal{S}_0,$$

hence,

$$\begin{aligned}C_{\rm L}&=\sup R_{\rm c}\\&\stackrel{(a)}{\leq}\inf_{\mathbf{s}_0\in\mathcal{S}_0}\left\{\frac{C_n^{\rm L}\left(\mathbf{s}_0\right)}{1-\epsilon_1}+\frac{1}{n(1-\epsilon_1)}+\epsilon_2\right\}\\&=\frac{1}{1-\epsilon_1}\inf_{\mathbf{s}_0\in\mathcal{S}_0}C_n^{\rm L}\left(\mathbf{s}_0\right)+\frac{1}{(1-\epsilon_1)n}+\epsilon_2,\end{aligned}\tag{A.9}$$

where (a) follows from the following proposition: Suppose that $\mathcal{A}$ and $\mathcal{B}$ are nonempty sets of real numbers such that $a\leq b\quad\forall a\in\mathcal{A}$ and $b\in\mathcal{B}$. Then, $\sup\mathcal{A}\leq\inf\mathcal{B}$.[3] Note that (A.9) coincides with (A.5). ■

[3]Proof: Fix $b\in\mathcal{B}$. Since $a\leq b$ for all $a\in\mathcal{A}$, then b is an upper bound on $\mathcal{A}$, thus, $b\geq\sup\mathcal{A}$. This inequality implies that $\sup\mathcal{A}$ is a lower bound on $\mathcal{B}$, hence $\sup\mathcal{A}\leq\inf\mathcal{B}$.

Lemma A.3. *For all $n > 2m$, the capacity of the n-MNGMC satisfies* $\inf_{\mathbf{s}_0 \in \mathcal{S}_0} C_n^{\rm L}(\mathbf{s}_0) \leq \frac{n+2m}{n} C_{n+2m}^{\rm M}$.

Proof: Note that for every $n > 2m$ and for every distribution on $\mathbf{X}^{n+2m-1}$, it follows from the mutual information chain rule [17, Ch. 2.4] that the input and output of the LNGMC satisfy

$$\begin{aligned}
I\left(\mathbf{X}^{n+2m-1}; \mathbf{Y}_m^{n+2m-1}\right) &= I\left(\mathbf{X}^{n+2m-1}; \mathbf{Y}_m^{2m-1}\right) + I\left(\mathbf{X}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{Y}_m^{2m-1}\right) \\
&= I\left(\mathbf{X}^{n+2m-1}; \mathbf{Y}_m^{2m-1}\right) + I\left(\mathbf{X}^{2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{Y}_m^{2m-1}\right) \\
&\quad + I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}^{2m-1}, \mathbf{Y}_m^{2m-1}\right) \\
&\geq I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}^{2m-1}, \mathbf{Y}_m^{2m-1}\right) \\
&= I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}_m^{2m-1}, \mathbf{Y}_m^{2m-1}, \mathbf{X}^{m-1}\right) \\
&\stackrel{(a)}{=} I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}, \mathbf{X}^{m-1}\right), \qquad \text{(A.10)}
\end{aligned}$$

where (a) follows from the definition of the LNGMC in (3). Applying the definition of conditional mutual information [15, Ch. 2.4] to the expression in (A.10), we can write:

$$\begin{aligned}
&I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}, \mathbf{X}^{m-1}\right) \\
&= \int\limits_{\substack{\tilde{\mathbf{s}}_0 \in \mathcal{S}_0 \\ \tilde{\mathbf{x}}^{m-1} \in \mathcal{R}^m}} I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \tilde{\mathbf{s}}_0, \mathbf{X}^{m-1} = \tilde{\mathbf{x}}^{m-1}\right) p_{\mathbf{S}_0, \mathbf{X}^{m-1}}\left(\tilde{\mathbf{s}}_0, \tilde{\mathbf{x}}^{m-1}\right) d\tilde{\mathbf{s}}_0 d\tilde{\mathbf{x}}^{m-1}.
\end{aligned}$$

Next, we note that from the generalized mean value inequality for integration [33, Ch. 9.10.2], it follows that for any input distribution on $\mathbf{X}^{n+2m-1}$, there exist $\hat{\mathbf{s}}_0 \in \mathcal{S}_0$ and $\hat{\mathbf{x}}^{m-1} \in \mathcal{R}^m$ such that[4]

$$\begin{aligned}
&\int\limits_{\substack{\tilde{\mathbf{s}}_0 \in \mathcal{S}_0 \\ \tilde{\mathbf{x}}^{m-1} \in \mathcal{R}^m}} I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \tilde{\mathbf{s}}_0, \mathbf{X}^{m-1} = \tilde{\mathbf{x}}^{m-1}\right) p_{\mathbf{S}_0, \mathbf{X}^{m-1}}\left(\tilde{\mathbf{s}}_0, \tilde{\mathbf{x}}^{m-1}\right) d\tilde{\mathbf{s}}_0 d\tilde{\mathbf{x}}^{m-1} \\
&= I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \hat{\mathbf{s}}_0, \mathbf{X}^{m-1} = \hat{\mathbf{x}}^{m-1}\right) \int\limits_{\substack{\tilde{\mathbf{s}}_0 \in \mathcal{S}_0 \\ \tilde{\mathbf{x}}^{m-1} \in \mathcal{R}^m}} p_{\mathbf{S}_0, \mathbf{X}^{m-1}}\left(\tilde{\mathbf{s}}_0, \tilde{\mathbf{x}}^{m-1}\right) d\tilde{\mathbf{s}}_0 d\tilde{\mathbf{x}}^{m-1} \\
&= I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \hat{\mathbf{s}}_0, \mathbf{X}^{m-1} = \hat{\mathbf{x}}^{m-1}\right) \qquad \text{(A.11)}
\end{aligned}$$

By plugging (A.11) into (A.10) we conclude that for the input and output of the LNGMC it

[4]Note that the mean value theorem for integration requires the integral to be defined over a finite interval. However, as $\int_{\tilde{\mathbf{s}}_0 \in \mathcal{S}_0} p_{\mathbf{S}_0}(\tilde{\mathbf{s}}_0) d\tilde{\mathbf{s}}_0 = 1$, it follows that the PDF approaches 0 for $\|\tilde{\mathbf{s}}_0\| \to \infty$, therefore, the integral can be approached arbitrarily close by considering a finite $2m$-dimensional volume, i.e., integrating over $\|\tilde{\mathbf{s}}_0\| \leq \Omega$ for sufficiently large Ω, instead of over $\mathcal{S}_0$, see also [35, Discussion after Eq. (B.15)].

follows that for every $n > 2m$ and for every distribution on $\mathbf{X}^{n+2m-1}$, $\exists\hat{\mathbf{s}}_0 \in \mathcal{S}_0$, $\hat{\mathbf{x}}^{m-1} \in \mathcal{R}^m$, such that

$$I\left(\mathbf{X}^{n+2m-1}; \mathbf{Y}_m^{n+2m-1}\right) \geq I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \hat{\mathbf{s}}_0, \mathbf{X}^{m-1} = \hat{\mathbf{x}}^{m-1}\right). \quad \text{(A.12)}$$

Next, for $C_n^{\mathrm{L}}(\mathbf{s}_0)$, defined in (A.4), we write

$$\begin{aligned} C_n^{\mathrm{L}}(\mathbf{s}_0) &= \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \frac{1}{n} I\left(\mathbf{X}^{n-1}; \mathbf{Y}^{n-1} \middle| \mathbf{S}_0 = \mathbf{s}_0\right) \\ &\stackrel{(a)}{=} \sup_{p\left(\mathbf{X}_{2m}^{n+2m-1}\right): \frac{1}{n}\sum\limits_{i=2m}^{n+2m-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \frac{1}{n} I\left(\mathbf{X}_{2m}^{n+2m-1}; \mathbf{Y}_{2m}^{n+2m-1} \middle| \left\{\mathbf{X}_m^{2m-1}, \mathbf{W}_m^{2m-1}\right\} = \mathbf{s}_0\right), \end{aligned} \quad \text{(A.13)}$$

where (a) follows from the fact that the LNGMC has an LTI CIR with a memory length of m (3), combined with the fact that $\mathbf{W}[i]$ is strict-sense stationary, as well as the definition of the initial state $\mathbf{S}_0 = \left[\left(\mathbf{X}_{-m}^{-1}\right)^T, \left(\mathbf{W}_{-m}^{-1}\right)^T\right]^T$: These characteristics imply that the conditional probability function of the channel output given the channel input is invariant to index shifting, i.e., $p\left(\mathbf{Y}^{n-1} \middle| \mathbf{X}_{-m}^{n-1} = \mathbf{x}, \mathbf{W}_{-m}^{-1} = \mathbf{w}\right) = p\left(\mathbf{Y}_{2m}^{n+2m-1} \middle| \mathbf{X}_m^{n+2m-1} = \mathbf{x}, \mathbf{W}_m^{2m-1} = \mathbf{w}\right)$ for all $\mathbf{x} \in \mathcal{R}^{n_{\mathrm{t}}\cdot(n+m)}$, $\mathbf{w} \in \mathcal{R}^{n_{\mathrm{r}}\cdot m}$. By combining (A.13) and (A.12), and recalling that (A.12) holds for any distribution of $\mathbf{X}^{n+2m-1}$, it follows that for any distribution of $\mathbf{X}^{2m-1}$

$$\begin{aligned}
&\inf_{\mathbf{s}_0\in\mathcal{S}_0} C_n^{\mathrm{L}}(\mathbf{s}_0)\\
&= \inf_{\mathbf{s}_0\in\mathcal{S}_0}\left\{\sup_{p\left(\mathbf{X}_{2m}^{n+2m-1}\right):\frac{1}{n}\sum\limits_{i=2m}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}\frac{1}{n}I\left(\mathbf{X}_{2m}^{n+2m-1};\mathbf{Y}_{2m}^{n+2m-1}\middle|\left\{\mathbf{X}_m^{2m-1},\mathbf{W}_m^{2m-1}\right\}=\mathbf{s}_0\right)\right\}\\
&\stackrel{(a)}{=} \inf_{\substack{\mathbf{s}_0\in\mathcal{S}_0\\ \tilde{\mathbf{x}}^{m-1}\in\mathcal{R}^m}}\left\{\sup_{\substack{p\left(\mathbf{X}_{2m}^{n+2m-1}|\mathbf{X}^{2m-1}=\tilde{\mathbf{x}}^{2m-1}\right):\\ \frac{1}{n}\sum\limits_{i=2m}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}}\frac{1}{n}I\left(\mathbf{X}_{2m}^{n+2m-1};\mathbf{Y}_{2m}^{n+2m-1}\middle|\left\{\mathbf{X}_m^{2m-1},\mathbf{W}_m^{2m-1}\right\}=\mathbf{s}_0,\mathbf{X}^{m-1}=\tilde{\mathbf{x}}^{m-1}\right)\right\}\\
&\stackrel{(b)}{\leq} \sup_{p(\mathbf{X}^{n+2m-1}):\frac{1}{n}\sum\limits_{i=2m}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P,\frac{1}{2m}\sum\limits_{i=0}^{2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}\frac{1}{n}I\left(\mathbf{X}^{n+2m-1};\mathbf{Y}_m^{n+2m-1}\right), \qquad \text{(A.14)}\\
&\stackrel{(c)}{\leq} \sup_{p(\mathbf{X}^{n+2m-1}):\frac{1}{n+2m}\sum\limits_{i=0}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}\frac{1}{n}I\left(\mathbf{X}^{n+2m-1};\mathbf{Y}_m^{n+2m-1}\right)\\
&= \frac{n+2m}{n}C_{n+2m}^{\mathrm{M}}, \qquad \text{(A.15)}
\end{aligned}$$

where (a) follows since introducing the conditioning into the PDF in the supremum (note that $\tilde{\mathbf{x}}_m^{2m-1}$ in the conditioning in the PDF $p\left(\mathbf{X}_{2m}^{n+2m-1}|\mathbf{X}^{2m-1}=\tilde{\mathbf{x}}^{2m-1}\right)$ is taken from $\mathbf{s}_0$) as well as adding the conditioning in the mutual information does not change the search space of the input distributions and consequently the value of the supremum does not change, (b) follows from (A.12), and (c) follows since the set over which the supremum is evaluated contains the set over which the supremum is evaluated in (A.14), $\left\{p\left(\mathbf{X}^{n+2m-1}\right):\frac{1}{n+2m}\sum\limits_{i=0}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P\right\}\supseteq\left\{p\left(\mathbf{X}^{n+2m-1}\right):\frac{1}{2m}\sum\limits_{i=0}^{2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P,\ \frac{1}{n}\sum\limits_{i=2m}^{n+2m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P\right\}$, hence the supremum can only be increased. It follows from (A.15) that $\inf\limits_{\mathbf{s}_0\in\mathcal{S}_0} C_n^{\mathrm{L}}(\mathbf{s}_0)\leq\frac{n+2m}{n}C_{n+2m}^{\mathrm{M}}$. ∎

Lemma A.4. *The capacity of the LNGMC, C_{L}, satisfies $\sup\limits_{n>m} C_n^{\mathrm{M}}\leq C_{\mathrm{L}}$.*

Proof: We will show that for any $n>m$, any rate R_{c} achievable for the n-MNGMC, is also achievable for the LNGMC, i.e., for any $\epsilon_1,\epsilon_2>0$, then, taking l to be sufficiently large, it is possible to construct an $[R_1,l]$ code for the LNGMC such that Def. 3 is satisfied. The proof follows the same reasoning as that of [26, Lemma 2]: We first show that the statement of the lemma holds for l which are integer multiples of n, following which we prove that this statement is true for all sufficiently large integer l.

Fix $n > m$ and consider a rate $R_{\rm c}$, achievable for the n-MNGMC. Then, by definition, for any $\epsilon_1, \epsilon_2 > 0$, $\exists b_0$ sufficiently large, such that for all integer $b > b_0$, there exists an $[R_1, b \cdot n]$ code for the n-MNGMC with an average error probability which satisfies Def. 3, and code rate which satisfies

$$R_1 \geq R_{\rm c} - \frac{\epsilon_2}{2}. \tag{A.16}$$

We denote this code by $\mathcal{C}^{\rm M}_{b\cdot n}$. Note that, by the definition of the n-MNGMC, the code $\mathcal{C}^{\rm M}_{b\cdot n}$ considers only the last $b \cdot n - m$ channel outputs out of each $b \cdot n$ channel outputs when decoding the message.

Next, apply the code $\mathcal{C}^{\rm M}_{b\cdot n}$ to the LNGMC. Note that the code rate is unchanged. Since the decoder considers the last $b \cdot n - m$ channel outputs out of each $b \cdot n$ channel outputs, the error probability is also the same as that of the n-MNGMC. It thus follows that for $R_{\rm c}$ achievable for the n-MNGMC, then for sufficiently large codelengths $l = b \cdot n$, there exists a code for the LNGMC with a rate which is arbitrarily close to $R_{\rm c}$, and with an arbitrarily small error probability for all $\mathbf{s}_0$.

We next use the coding scheme of the n-MNGMC to construct a coding scheme for the LNGMC, for any sufficiently large integer value of $l \in \mathcal{N}$: Let $l = b \cdot n + a$, where b is a positive integer and $a \in \{0, 1, \ldots, n-1\}$. We construct an $[R_2, l]$ code for the LNGMC by appending a arbitrary symbols to the codewords of $\mathcal{C}^{\rm M}_{b\cdot n}$. The decoder discards the last a channel outputs. Clearly, for any $\mathbf{s}_0$, the error probability is the same as that of the code $\mathcal{C}^{\rm M}_{b\cdot n}$ since the decoder operates on the same received symbols. The code rate R_2 is obtained by

$$R_2 = R_1 \frac{b \cdot n}{b \cdot n + a} \stackrel{(a)}{\geq} \left(R_{\rm c} - \frac{\epsilon_2}{2}\right) \frac{b \cdot n}{b \cdot n + a},$$

where (a) follows from (A.16). Thus, for sufficiently large b, i.e., $b > \max\left\{\frac{2a(R_{\rm c} - \epsilon_2)}{n\epsilon_2}, b_0\right\}$, it follows that $R_2 \geq R_{\rm c} - \epsilon_2$. Consequently, if $R_{\rm c}$ is achievable for the n-MNGMC, then for any $\epsilon_1, \epsilon_2 > 0$ we can find a sufficiently large value of $l_0 \in \mathcal{N}$, such that for any $l > l_0$ we can construct an $[R_2, l]$ code for the LNGMC, which satisfies $R_2 > R_{\rm c} - \epsilon_2$ and the probability of error is less than ϵ_1. This implies that $R_{\rm c}$ is an achievable rate for the LNGMC and hence, $C^{\rm M}_n \leq C_{\rm L}$ for all $n > m$, from which it follows that $\sup\limits_{n>m} C^{\rm M}_n \leq C_{\rm L}$. ■

Lemma A.5. *The capacity of the LNGMC satisfies* $C_{\rm L} = \lim\limits_{n\to\infty} C^{\rm M}_n$.

Proof: By combining Lemmas A.2-A.4 we conclude that for every $\epsilon_1 \in (0,1)$, $\epsilon_2 > 0$, and $n > 2m$ it holds that:

$$\begin{aligned}
\sup_{n>m} C_n^{\mathrm{M}} &\overset{(a)}{\leq} C_{\mathrm{L}} \\
&\overset{(b)}{\leq} \frac{1}{1-\epsilon_1} \inf_{\mathbf{s}_0 \in \mathcal{S}_0} C_n^{\mathrm{L}}(\mathbf{s}_0) + \frac{1}{(1-\epsilon_1)n} + \epsilon_2 \\
&\overset{(c)}{\leq} \frac{1}{1-\epsilon_1} \frac{n+2m}{n} C_{n+2m}^{\mathrm{M}} + \frac{1}{(1-\epsilon_1)n} + \epsilon_2,
\end{aligned} \tag{A.17}$$

where (a) follows from Lemma A.4; (b) follows from Lemma A.2; and (c) follows from Lemma A.3. As (A.17) is satisfied for all $n > 2m$, it follows from [36, Thm. 3.19] that

$$\begin{aligned}
C_{\mathrm{L}} &\leq \liminf_{n\to\infty} \left(\frac{1}{1-\epsilon_1} \frac{n+2m}{n} C_{n+2m}^{\mathrm{M}} + \frac{1}{(1-\epsilon_1)n} \right) + \epsilon_2 \\
&= \frac{1}{1-\epsilon_1} \liminf_{n\to\infty} C_n^{\mathrm{M}} + \epsilon_2.
\end{aligned} \tag{A.18}$$

Since ϵ_1, ϵ_2 can be made arbitrarily small, (A.18) implies that

$$C_{\mathrm{L}} \leq \liminf_{n\to\infty} C_n^{\mathrm{M}}. \tag{A.19}$$

Lastly, it follows from the definition of $\limsup$ [28, Def. 5.4] that $\limsup\limits_{n\to\infty} C_n^{\mathrm{M}} \leq \sup\limits_{n>m} C_n^{\mathrm{M}}$. Since $\liminf\limits_{n\to\infty} C_n^{\mathrm{M}} \leq \limsup\limits_{n\to\infty} C_n^{\mathrm{M}}$, it follows from (A.17) that $C_{\mathrm{L}} = \lim\limits_{n\to\infty} C_n^{\mathrm{M}}$, and that the limit exists. ■

Lemma A.6. *For $n \to \infty$, the capacity of the n-MNGMC satisfies*

$$\lim_{n\to\infty} C_n^{\mathrm{M}} = \lim_{n\to\infty} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \frac{1}{n} I\left(\mathbf{X}^{n-1}; \mathbf{Y}^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right). \tag{A.20}$$

Proof: Note that for any distribution of $\mathbf{X}^{n-1}$, $n > m$, it follows from the mutual information chain rule that

$$I\left(\mathbf{X}^{n-1}; \mathbf{Y}_m^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) \leq I\left(\mathbf{X}^{n-1}; \mathbf{Y}^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right). \tag{A.21}$$

It thus follows from Lemma A.1 that

$$\begin{aligned}
C_n^{\mathrm{M}} &= \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n} I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}\right) \\
&\overset{(a)}{=} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n} I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right) \\
&\le \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n} I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right), \qquad \text{(A.22)}
\end{aligned}$$

where (a) follows since given $\mathbf{X}^{n-1}$, $\mathbf{Y}_m^{n-1}$ is independent of $\mathbf{X}_{-m}^{-1}$, and fixing $\mathbf{X}_{-m}^{-1}$ does not change the supremum over the PDF $p\left(\mathbf{X}^{n-1}\right)$. Next, we note that the definition of the LNGMC (3) and the strict-sense stationarity of $\mathbf{W}[i]$ imply that the conditional probability function of the channel output given the channel input is invariant to index shifting, i.e., $p\Big(\mathbf{Y}^{n-m-1}\Big|\mathbf{X}_{-m}^{n-m-1} = \mathbf{x}\Big) = p\Big(\mathbf{Y}_m^{n-1}\Big|\mathbf{X}^{n-1} = \mathbf{x}\Big)$ for all $\mathbf{x} \in \mathcal{R}^{n_{\mathrm{t}}\cdot n}$. Therefore,

$$\begin{aligned}
&\sup_{p(\mathbf{X}^{n-m-1}): \frac{1}{n-m}\sum\limits_{i=0}^{n-m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n-m} I\left(\mathbf{X}^{n-m-1};\mathbf{Y}^{n-m-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right) \\
&= \sup_{p\left(\mathbf{X}_m^{n-1}\right): \frac{1}{n-m}\sum\limits_{i=m}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n-m} I\left(\mathbf{X}_m^{n-1};\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right) \\
&\overset{(a)}{=} \sup_{p\left(\mathbf{X}_m^{n-1}\right): \frac{1}{n-m}\sum\limits_{i=m}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n-m} I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right) \\
&\overset{(b)}{\le} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{m}\sum\limits_{i=0}^{m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P, \frac{1}{n-m}\sum\limits_{i=m}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n-m} I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}\right) \\
&\overset{(c)}{\le} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\le P} \frac{1}{n-m} I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}\right) \\
&= \frac{n}{n-m} C_n^{\mathrm{M}}, \qquad \text{(A.23)}
\end{aligned}$$

where (a) follows since $h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}_m^{n-1}, \mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big) = h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}^{n-1}, \mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big)$, thus $I\left(\mathbf{X}_m^{n-1};\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right) = h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big) - h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}_m^{n-1}, \mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big) = h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big) - h\Big(\mathbf{Y}_m^{n-1}|\mathbf{X}^{n-1}, \mathbf{X}^{m-1} = \mathbf{0}_{n_{\mathrm{t}}\cdot m}\Big) = I\left(\mathbf{X}^{n-1};\mathbf{Y}_m^{n-1}|\mathbf{X}^{m-1}=\mathbf{0}_{n_{\mathrm{t}}\cdot m}\right)$; (b) follows since the set of distributions of $\mathbf{X}^{m-1}$ which satisfy $\frac{1}{m}\sum\limits_{i=0}^{m-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \le P$ includes the deterministic assignment $\mathbf{X}^{m-1}=\mathbf{0}_{n_{\rm t}\cdot m}$; and (c) follows since $\Big\{p\left(\mathbf{X}^{n-1}\right):\frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}\left[i\right]\|^2\right\}\leq P\Big\}\supseteq\Big\{p\left(\mathbf{X}^{n-1}\right):\frac{1}{m}\sum\limits_{i=0}^{m-1}\mathbb{E}\left\{\|\mathbf{X}\left[i\right]\|^2\right\}\leq P,\frac{1}{n-m}\sum\limits_{i=m}^{n-1}\mathbb{E}\left\{\|\mathbf{X}\left[i\right]\|^2\right\}\leq P\Big\}$, hence the supremum can only be increased. Letting $n\to\infty$ in (A.22) and (A.23) proves the lemma. ■

Combining Lemmas A.5 and A.6 completes the proof of Proposition 1. □

B. Proof of Proposition 2

To prove the proposition, we first prove the upper bound in (6a), and then we prove the lower bounds in (6a) and (6b). To that aim, we define

$$S_n \triangleq \frac{1}{n}\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P} I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right).$$

We begin by stating an identity which will be used in the sequel: Prop. 1 implies that the capacity of the LNGMC can be computed by setting $\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}$. Hence, by defining the $l\cdot n_{\rm r}\times l\cdot n_{\rm t}$ matrix $\tilde{\mathsf{G}}_l$ such that

$$\tilde{\mathsf{G}}_l \triangleq \begin{bmatrix} \mathsf{G}[0] & \cdots & 0 & \cdots & 0\\ \vdots & \ddots & & \ddots & \vdots\\ \mathsf{G}[m] & \cdots & \mathsf{G}[0] & \cdots & 0\\ \vdots & \ddots & & \ddots & \vdots\\ 0 & \cdots & \mathsf{G}[m] & \cdots & \mathsf{G}[0]\end{bmatrix}, \tag{B.1}$$

and setting $\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}$, the output of the LNGMC for blocklength l can be expressed as

$$\mathbf{Y}^{l-1}=\tilde{\mathsf{G}}_l\mathbf{X}^{l-1}+\mathbf{W}^{l-1}. \tag{B.2}$$

In order to prove the upper bound in (6a), let $\mathbf{W}_G[i]$ be a zero-mean Gaussian process with an autocorrelation function $\mathsf{C}_{\mathbf{W}}[\tau]$, defined in Subsection III-A, s.t. $\mathbf{W}_G[i]$ is mutually independent of the channel input. Note that

$$\begin{aligned}\frac{1}{n}I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)&=\frac{1}{n}h\left(\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right)\\&=\frac{1}{n}\Big(h\left(\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)-h\left(\mathbf{W}_G^{n-1}\right)\Big)\\&\quad+\frac{1}{n}h\left(\mathbf{W}_G^{n-1}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right).\end{aligned} \tag{B.3}$$

Consequently

$$S_n = \frac{1}{n} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} \Big(h\left(\mathbf{Y}^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) - h\left(\mathbf{W}_G^{n-1}\right) \Big) + \frac{1}{n} h\left(\mathbf{W}_G^{n-1}\right) - \frac{1}{n} h\left(\mathbf{W}^{n-1}\right).$$

Since, for a given correlation function, Gaussian distribution maximizes the differential entropy [17, Thm. 8.6.5], $h\left(\mathbf{Y}^{n-1}\right)$ is maximized for a Gaussian distribution of $\mathbf{Y}^{n-1}$ with the same first and second-order moments as the original vector $\mathbf{Y}^{n-1}$. By letting $\left\{\mathbf{Y}_G[i]\right\}_{i=0}^{n-1}$ be a Gaussian process with the same first and second-order statistical moments as $\left\{\mathbf{Y}[i]\right\}_{i=0}^{n-1}$, we have that

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \sup_{p(\mathbf{X}^{n-1}): \frac{1}{n}\sum\limits_{i=0}^{n-1} \mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\} \leq P} h\left(\mathbf{Y}^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) - h\left(\mathbf{W}_G^{n-1}\right) \\ &\stackrel{(a)}{\leq} \lim_{n\to\infty} \frac{1}{n} \sup_{\mathrm{Cov}(\mathbf{X}^{n-1}): \mathrm{Tr}\left(\mathrm{Cov}(\mathbf{X}^{n-1})\right) \leq nP} h\left(\mathbf{Y}_G^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) - h\left(\mathbf{W}_G^{n-1}\right) \stackrel{(b)}{=} C_{\mathrm{G}}, \end{aligned} \tag{B.4}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, (a) follows from [17, Thm. 8.6.5], and since the differential entropy of a Gaussian random vector depends only on its covariance matrix [17, Thm. 8.4.1], hence the supremum is carried out over the covariance of the input; and (b) follows from [21, Lemma 3], noting that $h\left(\mathbf{Y}_G^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) - h\left(\mathbf{W}_G^{n-1}\right)$ denotes the mutual information between the input and the output of an LTI MIMO channel with additive Gaussian noise $\mathbf{W}_G^{n-1}$ and Gaussian output $\mathbf{Y}_G^{n-1} = \tilde{\mathsf{G}}_n \mathbf{X}^{n-1} + \mathbf{W}^{n-1}$, as in (B.2). Plugging (B.3)–(B.4) into Prop. 1 yields

$$\begin{aligned} C_{\mathrm{L}} = \lim_{n\to\infty} S_n &< C_{\mathrm{G}} + \lim_{n\to\infty} \left(\frac{1}{n} h\left(\mathbf{W}_G^{n-1}\right) - \frac{1}{n} h\left(\mathbf{W}^{n-1}\right) \right) \\ &= C_{\mathrm{G}} + \bar{H}_{G,\mathbf{W}} - \bar{H}_{\mathbf{W}}, \end{aligned} \tag{B.5}$$

which proves the upper bound in (6a).

The lower bound in (6a) follows since it can be concluded from [32], [15, Thm. 7.4.3][5], that for a given noise covariance matrix, then Gaussian noise is the worst-case noise distribution in terms of capacity, i.e., it results in the smallest capacity. Specifically, the supremum of $I\left(\mathbf{X}^{n-1}; \mathbf{Y}^{n-1} | \mathbf{X}_{-m}^{-1} = \mathbf{0}_{n_t \cdot m}\right) = I\left(\mathbf{X}^{n-1}; \tilde{\mathsf{G}}_n \mathbf{X}^{n-1} + \mathbf{W}^{n-1}\right)$ over all input distributions is lower bounded by the mutual information between the channel inputs and the channel outputs in which the additive non-Gaussian noise is replaced with an additive Gaussian noise with the

[5]While [15, Thm. 7.4.3] is stated for scalar channels, the same proof also applies to MIMO channels.

same second-order moments as that of the non-Gaussian noise. Consequently, in the limit of $n \to \infty$, the lower bound in Eq. (6a) directly follows from Prop. 1.

Next, from (B.2) we note that since both $\mathbf{X}^{n-1}$ and $\mathbf{W}^{n-1}$ are independent of $\mathbf{X}_{-m}^{-1}$, then

$$h\left(\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)=h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}+\mathbf{W}^{n-1}\right)\overset{(a)}{\geq}\frac{n\cdot n_{\rm r}}{2}\log\left(2^{\frac{2h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)}{n\cdot n_{\rm r}}}+2^{\frac{2h\left(\mathbf{W}^{n-1}\right)}{n\cdot n_{\rm r}}}\right), \quad \text{(B.6)}$$

where (a) follows from the entropy power inequality [17, Thm. 17.7.3]. Thus, we have that

$$\begin{aligned}
S_n &= \sup_{p(\mathbf{X}^{n-1}):\ \frac{1}{n}\sum\limits_{i=0}^{n-1}\mathbb{E}\left\{\|\mathbf{X}[i]\|^2\right\}\leq P}\frac{1}{n}I\left(\mathbf{X}^{n-1};\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)\\
&= \sup_{p(\mathbf{X}^{n-1}):\ \frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{1}{n}h\left(\mathbf{Y}^{n-1}|\mathbf{X}_{-m}^{-1}=\mathbf{0}_{n_{\rm t}\cdot m}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right)\\
&\overset{(a)}{\geq} \sup_{p(\mathbf{X}^{n-1}):\ \frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{n_{\rm r}}{2}\log\left(2^{\frac{2h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)}{n\cdot n_{\rm r}}}+2^{\frac{2h\left(\mathbf{W}^{n-1}\right)}{n\cdot n_{\rm r}}}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right), \quad \text{(B.7)}
\end{aligned}$$

where (a) follows from (B.6). Note that for any positive constants a_1, a_2, a_3 and a real constant t, the function $\log\left(a_1 2^{a_2 t}+a_3\right)$ is monotonically increasing w.r.t. t, therefore

$$\begin{aligned}
&\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{n_{\rm r}}{2}\log\left(2^{\frac{2}{n\cdot n_{\rm r}}h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)}+2^{\frac{2}{n\cdot n_{\rm r}}h\left(\mathbf{W}^{n-1}\right)}\right)\\
&\qquad=\frac{n_{\rm r}}{2}\log\left(2^{\sup\limits_{p(\mathbf{X}^{n-1}):\frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{2}{n\cdot n_{\rm r}}h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)}+2^{\frac{2}{n\cdot n_{\rm r}}h\left(\mathbf{W}^{n-1}\right)}\right). \quad \text{(B.8)}
\end{aligned}$$

Next, consider Eq. (B.8): Note that when $n_{\rm t}=n_{\rm r}$ and $\mathsf{G}[0]$ is invertible, it follows from (B.1) that $\tilde{\mathsf{G}}_n$ is also invertible, hence, by letting $\mathcal{M}_{n\cdot P}$ be the set of $n_{\rm t}\times n_{\rm t}$ positive semi-definite real symmetric matrices $\mathsf{C}_{\mathbf{X}}$ such that $\mathrm{Tr}\left(\mathsf{C}_{\mathbf{X}}\right)\leq n\cdot P$, we have that

$$\begin{aligned}
&\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{2}{n\cdot n_{\rm r}}h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)\overset{(a)}{=}\frac{2}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n|+\frac{2}{n\cdot n_{\rm r}}\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}h\left(\mathbf{X}^{n-1}\right)\\
&\overset{(b)}{=}\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n|^2+\frac{1}{n\cdot n_{\rm r}}\sup_{\mathrm{Cov}(\mathbf{X}^{n-1})\in\mathcal{M}_{n\cdot P}}\log\left(2\pi e\right)^{n\cdot n_{\rm r}}\left|\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)\right|\\
&=\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T|+\log\left(2\pi e\right)+\frac{1}{n\cdot n_{\rm r}}\sup_{\mathrm{Cov}(\mathbf{X}^{n-1})\in\mathcal{M}_{n\cdot P}}\log\left|\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)\right|, \quad \text{(B.9)}
\end{aligned}$$

where (a) follows from [17, Eq. (8.71)], and (b) follows from [17, Thm. 8.6.5]. Since $\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)$ is positive semi-definite, it follows from the inequality of the arithmetic and geometric means [28, Pg. 326] that $|\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)|\leq\left(\frac{1}{n\cdot n_{\rm t}}\mathrm{Tr}\Big(\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)\Big)\right)^{n\cdot n_{\rm t}}$, and thus $\frac{1}{n\cdot n_{\rm t}}\log|\mathrm{Cov}\left(\mathbf{X}^{n-1}\right)|\leq$

$\log\left(\frac{1}{n\cdot n_{\rm t}}{\rm Tr}\Big({\rm Cov}\left(\mathbf{X}^{n-1}\right)\Big)\right)$. Consequently,

$$\frac{1}{n\cdot n_{\rm t}}\sup_{{\rm Cov}(\mathbf{X}^{n-1})\,\in\mathcal{M}_{n\cdot P}}\log\left|{\rm Cov}\left(\mathbf{X}^{n-1}\right)\right|\leq\sup_{{\rm Cov}(\mathbf{X}^{n-1})\in\mathcal{M}_{n\cdot P}}\log\left(\frac{1}{n\cdot n_{\rm t}}{\rm Tr}\Big({\rm Cov}\left(\mathbf{X}^{n-1}\right)\Big)\right)$$
$$\stackrel{(a)}{\leq}\log\left(\frac{P}{n_{\rm t}}\right), \tag{B.10}$$

where (a) follows since $\log(\cdot)$ is monotonically increasing over $\mathcal{R}^+$. Note that for ${\rm Cov}\left(\mathbf{X}^{n-1}\right)=\frac{P}{n_{\rm t}}\cdot\mathsf{I}_{n\cdot n_{\rm t}}$ the right hand side of (B.10) is obtained with equality. Plugging this assignment into (B.9), and recalling that $n_{\rm t}=n_{\rm r}$, yields

$$\sup_{p(\mathbf{X}^{n-1}):\frac{1}{n}\mathbb{E}\left\{\|\mathbf{X}^{n-1}\|^2\right\}\leq P}\frac{2}{n\cdot n_{\rm r}}h\left(\tilde{\mathsf{G}}_n\mathbf{X}^{n-1}\right)=\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T|+\log\left(2\pi e\right)+\log\left(\frac{P}{n_{\rm t}}\right)$$
$$=\log\left(2\pi e\frac{P}{n_{\rm t}}\right)+\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T|. \tag{B.11}$$

Combining (B.11), (B.8), and (B.7) results in $S_n\geq\frac{n_{\rm r}}{2}\log\left(\frac{2\pi eP}{n_{\rm t}}\cdot 2^{\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T|}+2^{\frac{2}{n\cdot n_{\rm r}}h\left(\mathbf{W}^{n-1}\right)}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right)$, for any n. Lastly, we note that in the limit as $n\rightarrow\infty$, it follows from the extension of Szego's theorem to block-Toeplitz matrices [21, Appendix A.2], [29, Thm. 5] that $\lim\limits_{n\rightarrow\infty}\frac{1}{n}\log\left|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T\right|=\frac{1}{2\pi}\sum\limits_{k=0}^{n_{\rm t}-1}\int\limits_{\omega=-\pi}^{\pi}\log\left(\alpha'_k(\omega)\right)d\omega$, therefore, since 2^t is continuous w.r.t. $t\in\mathcal{R}$, letting n tend to infinity in (B.7), it follows from Prop. 1 and [28, Pg. 224] that

$$C_{\rm L}=\lim_{n\rightarrow\infty}S_n$$
$$\geq\lim_{n\rightarrow\infty}\frac{n_{\rm r}}{2}\log\left(\frac{2\pi eP}{n_{\rm t}}\cdot 2^{\frac{1}{n\cdot n_{\rm r}}\log|\tilde{\mathsf{G}}_n\tilde{\mathsf{G}}_n^T|}+2^{\frac{2}{n\cdot n_{\rm r}}h\left(\mathbf{W}^{n-1}\right)}\right)-\frac{1}{n}h\left(\mathbf{W}^{n-1}\right)$$
$$=\frac{n_{\rm r}}{2}\log\left(\frac{2\pi eP}{n_{\rm t}}\cdot 2^{\frac{1}{2\pi\cdot n_{\rm r}}\sum\limits_{k=0}^{n_{\rm t}-1}\int\limits_{\omega=-\pi}^{\pi}\log\left(\alpha'_k(\omega)\right)d\omega}+2^{\frac{2}{n_{\rm r}}\bar{H}_{\mathbf{W}}}\right)-\bar{H}_{\mathbf{W}}, \tag{B.12}$$

which completes the proof of (6b). □

C. Proof of Theorem 1

The outline of the proof is as follows: First, in Lemma C.1 we show that the capacity of the MIMO BB-PLC channel (1), can be characterized by considering only codes whose blocklength is an integer multiple of $\tilde{p}$. Then, we show that the capacity of MIMO BB-PLC channels constrained to using only codes whose blocklength is an integer multiple of $\tilde{p}$ satisfies (9).

Lemma C.1. *The capacity of the MIMO BB-PLC channel is identical to the maximum achievable rate obtained by considering only codes whose blocklength is an integer multiple of $\tilde{p}$.*

Proof: The proof follows by first showing that any rate achievable for the MIMO BB-PLC channel can be achieved by considering only codes whose blocklength is an integer multiple of $\tilde{p}$, and then showing any rate achievable for the MIMO BB-PLC channel when considering such codes, is an achievable rate for the MIMO BB-PLC channel. As these steps are essentially the same as in the proof of [12, Lemma 1], they are not repeated here. ■

Next, we note that the MIMO BB-PLC channel (1) subject to the constraint that only codes whose blocklength is an integer multiple of $\tilde{p}$ are used, i.e., $\tilde{l} = l \cdot \tilde{p}$ where $l \in \mathcal{N}$, can be represented as an equivalent $\tilde{p} \times \tilde{p}$ LNGMC with code blocklength l via the following assignments: Let the $\tilde{p} \cdot \tilde{n}_{\rm t} \times 1$ vector $\mathbf{X}_{\rm DCD}[i] \triangleq \tilde{\mathbf{X}}_{i \cdot \tilde{p}}^{(i+1)\cdot\tilde{p}-1}$ be the input to the transformed channel and the $\tilde{p} \cdot \tilde{n}_{\rm t} \times 1$ vector $\mathbf{Y}_{\rm DCD}[i] \triangleq \tilde{\mathbf{Y}}_{i \cdot \tilde{p}}^{(i+1)\cdot\tilde{p}-1}$ be the output of the channel. The transformation is clearly bijective as for the BB-PLC channel we consider only codes whose blocklength is an integer multiple of $\tilde{p}$. For each blocklength l, the input to the equivalent LNGMC satisfies

$$\frac{1}{l}\sum_{i=0}^{l-1}\mathbb{E}\left\{\|\mathbf{X}_{\rm DCD}[i]\|^2\right\} = \frac{1}{l}\sum_{i=0}^{l-1}\sum_{k=0}^{\tilde{p}-1}\mathbb{E}\left\{\left\|\tilde{\mathbf{X}}[i\cdot\tilde{p}+k]\right\|^2\right\} = \frac{\tilde{p}}{\tilde{l}}\sum_{\tilde{i}=0}^{\tilde{l}-1}\mathbb{E}\left\{\left\|\tilde{\mathbf{X}}\left[\,\tilde{i}\,\right]\right\|^2\right\} \stackrel{(a)}{\leq} \tilde{p}\cdot\tilde{P},$$

where (a) follows from (2). Consequently, the equivalent LNGMC input is subject to a maximal power constraint $P_{\rm DCD} = \tilde{p} \cdot \tilde{P}$. Next, we note that the input-output relationship of the BB-PLC channel (1) implies that the input-output relationship of the transformed channel is given by (7), and that the equivalent LNGMC noise $\mathbf{W}_{\rm DCD}[i]$ appearing in (7), is a zero-mean strict-sense stationary process. Moreover, as $\tilde{p} > \tilde{m}$, it follows that the temporal dependence of $\mathbf{W}_{\rm DCD}[i]$ spans an interval of length $m = 1$. Recall that $C_{\rm DCD}$ denotes the capacity of the channel (7)–(8).

As each channel use in the equivalent LNGMC (7)–(8) corresponds to $\tilde{p}$ channel uses in the BB-PLC channel (1)–(2), it follows that the maximal achievable rate of the BB-PLC channel, measured in bits per channel use, subject to the restriction that only codes whose blocklength is an integer multiple of $\tilde{p}$ are allowed, can be obtained from the maximal achievable rate of the equivalent LNGMC as $C_{\rm P} = \frac{1}{\tilde{p}} C_{\rm DCD}$. Finally, from Lemma C.1, we conclude that $C_{\rm P}$ is the maximum achievable rate for the BB-PLC channel, thus proving the theorem. □

References

[1] C. Cano, A. Pittolo, D. Malone, L. Lampe, A. M. Tonello, and A. G. Dabak. "State of the art in power line communications: From the applications to the medium," *IEEE J. Sel. A. Commun.*, vol. 34, no. 7, Jul. 2016, pp. 1935–1952.

[2] M. Gotz, M. Rapp, and K. Dostert. “Power line channel characteristics and their effect on communication system design,” *IEEE Commun. Mag.*, vol. 42, no. 4, Apr. 2004, pp. 78–86.

[3] F. J. Cañete, J. A. Cortés, L. Díez, and J. T. Entrambasaguas. “A channel model proposal for indoor power line communications,” *IEEE Commun. Mag.*, vol. 49, no. 12, Dec. 2011, pp. 166–174.

[4] M. A. Tunc, E. Perrins, and L. Lampe. “Optimal LPTV-aware bit loading in broadband PLC,” *IEEE Trans. Commun.*, vol. 61, no. 12, Dec. 2013, pp. 5152–5162.

[5] D. Veronesi, R. Riva, P. Bisaglia, F. Osnato, K. Afkhamie, A. Nayagam, D. Rende, and L. Yonge. “Characterization of in-home MIMO power line channels,” *Proc. IEEE ISPLC*, Udine, Italy, Apr. 2011, pp. 42–47.

[6] F. Gianaroli, F.. Pancaldi, and G. M. Vitetta. “The impact of statistical noise modeling on the error-rate performance of OFDM power-line communications,” *IEEE Trans. Power Del.*, vol. 29, no. 6, Apr. 2014, pp. 2622–2630.

[7] A. M. Tonello, F. Versolatto, and A. Pittolo. “In-home power line communication channel: Statistical characterization,” *IEEE Trans. Commun.*, vol. 62, no. 6, Jun. 2014, pp. 2096–2106.

[8] D. Rende, A. Nayagam, K. Afkhamie, L. Yonge, R. Riva, D. Veronesi, F. Osnato, and P. Bisaglia. “Noise correlation and its effect on capacity of inhome MIMO power line channels,” *Proc. IEEE ISPLC*, Udine, Italy, Apr. 2011, pp. 60 - 65.

[9] J. A. Cortés, L. Díez, F. J. Cañete, and J. J. Sánchez-Martińez,. “Analysis of the indoor broadband power-line noise scenario,” *IEEE Trans. Electromagn. Compat.*, vol. 52, no. 4, Nov. 2010, pp. 849–858.

[10] H. Meng, Y. L. Guan, and S. Chen. “Modeling and analysis of noise effects on broadband power-line communications,” *IEEE Trans. Power Del.*, vol. 20, no. 2, Apr. 2005, pp. 630–637.

[11] N. Shlezinger and R. Dabora. “On the capacity of narrowband PLC channels,” *IEEE Trans. Commun.*, vol. 63, no. 4, Apr. 2015, pp. 1191 - 1201.

[12] N. Shlezinger and R. Dabora. “The capacity of discrete-time Gaussian MIMO channels with periodic characteristics,” *Proc. IEEE ISIT*, Barcelona, Spain, Jun. 2016, pp. 1058 - 1062.

[13] S. Ihara. “On the capacity of channels with additive non-Gaussian noise,” *Information and Control*, vol. 37, no. 1, Apr. 1978, pp. 34–39.

[14] M. S. Pinsker, V. V. Prelov, and S. Verdu. “Sensitivity of channel capacity,” *IEEE Trans. Inform. Theory*, vol. 41, no. 6, Nov. 1995, pp. 1877–1888.

[15] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley Press, 1968.

[16] N. Shlezinger, R. Shaked, and R. Dabora. “Bounds on the capacity of MIMO broadband power line communications channels - Full version,” Available online at http://www.ee.bgu.ac.il/˜daborona, 2018.

[17] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Press, 2006.

[18] R. L. Dobrushin. “General formulation of Shannons main theorem in information theory,” *Amer. Math. Soc. Translations: Series 2*, vol. 33, 1963, pp. 323–438.

[19] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer, 2003.

[20] W. Hirt and J. L. Massey. “Capacity of discrete-time Gaussian channel with intersymbol interference,” *IEEE Trans. Inform. Theory*, vol. 34, no. 3, May 1988, pp. 380–388.

[21] L. H. Brandenburg and A. D. Wyner. “Capacity of the Gaussian channel with memory: The multivariate case,” *Bell System Technical Journal*, vol. 53, no. 5, May. 1974, pp. 745–778.

[22] S. Verdu. “The capacity region of the symbol-asynchronous Gaussian multiple-access channel,” *IEEE Trans. Inform. Theory*, vol. 35, no. 4, Aug. 1989, pp. 733–751.

[23] G. B. Giannakis. “Cyclostationary signal analysis,” *Digital Signal Processing Handbook*, CRC Press, 1998, pp. 17.1–17.31.

[24] International Telecommunications Union (ITU). “ITU-T Recommendation G.9963, Unified high-speed wire-line based home networking transceivers Multiple Input/Multiple Output (MIMO),” Sep. 2011.

[25] I. P. Tsaregradskii. “A note on the capacity of a stationary channel with finite memory,” *Theory of Probability and its Applications*, vol. 3, no. 1, 1958, pp. 79–91.

[26] A. Goldsmith and M. Effros. “The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise,” *IEEE Trans. Inform. Theory*, vol. 47, no. 1, Jan. 2001, pp. 219–240.

[27] A. Goldsmith. *Wireless Communications*. Cambridge, 2005.

[28] H. Amann and J. Escher. *Analysis I*. Birkhauser Verlag, Basel, 2005.

[29] J. Guttiérez-Guttiérez and P. M. Crespo. “Asymptotically equivalent sequences of matrices and hermitian block toeplitz matrices with continuous symbols: Applications to MIMO systems,” *IEEE Trans. Inform. Theory*, vol. 54, no. 12, Dec. 2008, pp. 5671–5680.

[30] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck. “On entropy approximation for Gaussian mixture random vectors,” *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Seoul, South Korea, Aug. 2008, pp. 181-188.

[31] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.

[32] S. N. Diggavi and T. M. Cover. “The worst additive noise under a covariance constraint,” *IEEE Trans. Inform. Theory*, vol. 47, no. 7, Dec. 2001, pp. 3072–3081.

[33] S. C. Malik and S. Arora. *Mathematical Analysis*. New Age International, 1992.

[34] Y. Murin, R. Dabora, and D. Gunduz. “Source-channel coding theorems for the multiple-access relay channel,” *IEEE Trans. Inform. Theory*, vol. 59, no. 9, Sep. 2013, pp. 5446–5465.

[35] N. Shlezinger and R. Dabora. “On the capacity of narrowband PLC channels: Complete version with detailed proofs,” *ArXiv preprint*, arXiv:1505.2934, 2015.

[36] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill New York, 1976.