On the Capacity of Communication Channels With Memory and Sampled Additive Cyclostationary Gaussian Noise

Ron Dabora, *Senior Member, IEEE*, and Emeka Abakasanga

***Abstract*— In this work we study the capacity of interference-limited channels with memory. These channels model non-orthogonal communications scenarios, such as the non-orthogonal multiple access (NOMA) scenario and underlay cognitive communications, in which the interference from other communications signals is much stronger than the thermal noise. Interference-limited communications is expected to become a very common scenario in future wireless communications systems, such as 5G, WiFi6, and beyond. As communications signals are inherently cyclostationary in continuous time (CT), then after sampling at the receiver, the discrete-time (DT) received signal model contains the sampled desired information signal with additive sampled CT cyclostationary noise. The sampled noise can be modeled as either a DT cyclostationary process or a DT *almost*-cyclostationary process, where in the latter case the resulting channel is not information-stable. In a previous work we characterized the capacity of this model for the case in which the DT noise is *memoryless*. In the current work we come closer to practical scenarios by modelling the resulting DT noise as a *finite-memory* random process. The presence of memory requires the development of a new set of tools for analyzing the capacity of channels with additive non-stationary noise which has memory. Our results show, for the first time, the relationship between memory, sampling frequency synchronization and capacity, for interference-limited communications. The insights from our work provide a link between the analog and the digital time domains, which has been missing in most previous works on capacity analysis. Thus, our results can help improving spectral efficiency and suggest optimal transceiver designs for future communications paradigms.**

***Index Terms*— Interference-limited communications, cyclostationary processes, almost-cyclostationary processes, information spectrum, capacity.**

I. Introduction

WE CONSIDER maximizing the information rates in interference-limited communications, which is a communications scenario in which message decoding is impeded by another communications signal, instead of by the commonly-studied thermal noise. Interference-limited communications has been attracting much interest in recent years; One major reason is the emergence of non-orthogonal multiple access (NOMA) as a major paradigm for 5G communications [1]. Another important motivation is that interference-limited communications corresponds to multiple existing communications scenarios, including, for example, digital subscriber line (DSL), in which crosstalk is limiting the rate of information [2], and underlay cognitive communications, in which the secondary user is the major source of interference to the primary user [3].

Since communications signals are man-made, then they inherently possess cyclostationary statistics [4, Ch. 1.3], which follows as the signal generation process repeats at every symbol interval. Consequently, when communications is limited by interference, the corresponding continuous-time (CT) channel is modeled as a linear channel with additive wide-sense cyclostationary (WSCS) noise. In modern communications, the receiver first samples the received CT signal in order to facilitate digital processing. When the sampling interval at the receiver is commensurate with the period of the CT WSCS interference process, a situation referred to in this work as *synchronous sampling*, the resulting sampled discrete-time (DT) interference is also WSCS. This channel model was extensively analyzed in previous works: The capacity of point-to-point (PtP) DT channels with a finite memory and with additive WSCS Gaussian noise (ACGN) was derived in [5] for the case in which the channel input is subject to a time-averaged per-symbol power constraint. Capacity characterization in [5] was obtained via both a time-domain approach and a frequency-domain approach. Subsequently, the capacity of DT multiple input-multiple output (MIMO) channels with finite-memory ACGN was derived in [6], the secrecy capacity of DT finite-memory channels with ACGN was derived in [7], and bounds on the capacity of DT channels with non-Gaussian WSCS noise were presented in [8]. Algorithmic aspects of reception in the presence of ACGN have also been studied: In [9] a receiver structure which uses the periodicity of the noise correlation function for noise cancellation was presented, and in [10] optimal adaptive filtering based on least-mean-squares adaptation for DT jointly WSCS processes was studied.

We note that, practically, the frequency of an oscillator cannot be deterministically set due to its inherent physical properties, see, e.g., [11]. Thus, even though in the system design, symbol clocks are typically set to nominal values given

Manuscript received 30 January 2022; revised 29 January 2023; accepted 13 May 2023. Date of publication 30 May 2023; date of current version 15 September 2023. This work was supported by the Israel Science Foundation under Grant 584/20. *(Corresponding author: Ron Dabora.)*

Ron Dabora is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er-Sheva 8410501, Israel, and also with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: daborona@bgu.ac.il).

Emeka Abakasanga is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Be'er-Sheva 8410501, Israel (e-mail: abakasan@post.bgu.ac.il).

Communicated by A. Somekh-Baruch, Associate Editor for Shannon Theory and Information Measures.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3281519.

Digital Object Identifier 10.1109/TIT.2023.3281519

by finite-precision decimal numbers, see, e.g., [12], then, as the interference and the signal-of-interest (SOI) are clocked by physically separate oscillators, there is no reason to assume that their actual symbol intervals are related by a rational factor, due to the clocks' inherent frequency variability. This is the main motivation for the model[1] considered in the current work. When the sampling interval at the receiver is incommensurate with the period of the CT WSCS interference process, a situation referred to in this work as *asynchronous sampling*, the resulting sampled DT interference is no longer WSCS, but rather it is a wide-sense *almost* cyclostationary (WSACS) random process [13, Sec. 3.9]. As WSACS processes are *non-stationary*, the resulting DT channel is generally not information-stable, namely, the conditional distribution of the channel output given the input does not behave ergodically [14]. As a consequence, standard information-theoretic tools (e.g., based on joint typicality) cannot be applied in the capacity characterization of such channels. It is noted that, as in practice, a receiver synchronizes its sampling rate with the symbol rate of the *desired* information signal, rather than with the symbol rate of the interference, asynchronous sampling is necessarily a frequent situation in practical systems, and thus, analysis of scenarios with asynchronous sampling carries practical, as well as theoretical, importance.

While communications with synchronous sampling was extensively analyzed, communications scenarios with asynchronous sampling have not been treated until recently. In [15], we took a first step towards the capacity analysis of asynchronously-sampled interference-limited Communication Channels by considering the *memoryless* case. In this context, a DT memoryless interference process is obtained by sampling a CT finite-memory WSCS process with a sampling interval that is greater than its correlation length. Thus, the correlation function of the resulting DT process is either a periodically time-varying or an almost periodically time-varying, scaled Kornecker's delta functions, which, for Gaussian processes implies that different samples are independent. It follows that [15] *restricts the shape of the CT correlation function as well as restricts the sampling rate to be low*, which restricts the information rate carried by the SOI. For this scenario, [15] derived a limiting expression for the capacity. As the channel is not information-stable, analysis was carried out within the framework of information spectrum, leading to a capacity expressed as the limit-inferior of a sequence of capacities corresponding to synchronously-sampled CT channels with ACGN. The work [15] presented several interesting insights: First, it was shown that when sampling is *synchronous*, capacity depends on the sampling interval and on the sampling phase, even when the sampling interval is smaller than half the period of the noise correlation function. Another important insight obtained from [15] is that when sampling is *asynchronous*, capacity does not vary with the sampling rate or the sampling phase. Finally, it was observed that for some synchronous sampling rates, capacity in higher than the capacity obtained with asynchronous sampling rates arbitrarily close to the corresponding synchronous sampling rates. This means that practically, capacity of sampled CT interference-limited channels should be computed assuming *asynchronous* sampling, to avoid a false notion of a high capacity which hinges on an impractically accurate sampling frequency synchronization between the receiver and the interference. The impact of sampling frequency synchronization was subsequently studied in [16] for the dual problem of compressing a DT memoryless Gaussian random source process, obtained by asynchronously sampling a CT WSCS Gaussian source process. The rate-distortion function (RDF) for this scenario was derived for the low distortion regime, as a limit of RDFs obtained by synchronously sampling the CT source process. It was observed that asynchronous sampling can result in higher compression rates than those obtained for synchronous sampling, mirroring the conclusions of [15] on the channel capacity.

The relationship between the analog domain and the digital domain has also attracted attention from additional aspects, as part of the research effort to accurately characterize the information rates for communications over *physical* channels: The work of [17] considered sampling of a CT linear, time-invariant (LTI) additive stationary noise channel, and showed that sampling rates higher than the Nyquist rate do not facilitate increase in capacity. The work of [18], provided a quantitative analysis of the rate of convergence of the mutual information between the message and the sampled (i.e., digital) additive white Gaussian noise (AWGN) channel outputs, to the CT (i.e., analog) channel's mutual information, with and without feedback, under certain conditions.

In the current work we extend the scenario considered in the previous work of [15] by analyzing the capacity of sampled CT interference-limited channels, in which the interfering CT process is a *correlated* WSCS process, and the sampling interval is *allowed to be shorter than the correlation length* of the interference. Therefore, the scenario considered in the current work places restrictions on the shape of the CT noise correlation function, while allowing sampling intervals shorter than the correlation length. In contrast, [15] restricts both the shape of the CT noise correlation function and requires the sampling interval to be longer than the correlation length, in order to obtain an impulse-shaped DT lag profile. One important consequence of this difference is that samples of the DT interference process in the current scenario may be statistically dependent, while in [15] they must be independent. When sampling is synchronous, we arrive at the model studied in [6], thus, the focus of the current work is on asynchronous sampling. The studied setup provides a connection between the analog model and the digital model obtained after sampling at the receiver, when sampling results in a non-stationary DT channel model with memory – a situation which has a practical relevance for current and future communications setups, but has not been analyzed previously.

A. Main Contributions

In this work we analyze the fundamental rate limits for DT channels with *correlated* WSACS Gaussian noise having a

[1]It is noted that while our model is closer to practicality than the models in previous works, still it is assumed that the variations of the clocks' frequencies around their respective actual values in practice, can be ignored. A complete model would account also for the impact of such variations on the statistics of the interference, but this is outside the framework of the current analysis.

finite correlation length, arising from sampling the output of CT channels with ACGN. Since additive WSACS Gaussian noise channels are *not information-stable*, it is not possible to employ standard information-theoretic arguments in the study of their capacity, and we resort to information-spectrum characterization of the capacity [19], within which we derive a new set of tools for the capacity analysis of DT channels with sampled finite-memory cyclostationary noise. We first observe that due to non-stationarity, the distribution function of the sampled noise process depends on the sampling time offset w.r.t. the period of the CT noise correlation function, referred to in this work as the *sampling phase*. Then, we obtain a general expression for the capacity when *transmission delay is not allowed*, namely when the transmitter must start transmitting the next message immediately upon completion of the transmission of the current message. Finally, for the case in which the correlation function decreases sufficiently fast with the lag, and the transmitter is allowed to delay the transmission of the next message by a finite and bounded delay, s.t. the optimal sampling phase is attained for subsequent message transmissions, a situation we refer to as *transmission delay is allowed*, then capacity can be expressed as the limit-inferior of a sequence of capacities corresponding to DT ACGN channels with finite memory, such that the correlation function of the sequence of DT WSCS noise processes approaches the correlation function of the DT WSACS noise process as the sequence index increases. Our characterization leads to important observations on the relationship between channel memory, sampling frequency synchronization and the achievable rate: We show that for synchronous sampling, capacity varies with the sampling rate and the sampling phase. It is then numerically demonstrated that when the power of the white thermal noise at the receiver is much smaller than the power of the interference, then an increase in the sampling rate results in higher capacity values. This follows as at higher sampling rates, the power spectral density (PSD) of the DT noise varies in the frequency domain, thereby facilitating a better allocation of the transmit power across the noise spectrum.

B. Organization

The rest of this paper is organized as follows: Section II sets the mathematical notations and quantities applied in this study. Section III presents the problem formulation and discusses the initial channel state; Section IV states the capacity characterization for asynchronously-sampled ACGN channels with finite memory when transmission delay is not allowed. Section V characterizes the capacity when a finite and bounded transmission delay is allowed. Section VI presents numerical results to demonstrate the impact of the different scenario parameters on capacity. Lastly, Section VII concludes the paper. The proofs of the theorems are detailed in the appendices.

II. Preliminaries

A. Notations

We use upper-case letters, e.g., X, to denote random variables (RVs), lower-case letters, e.g., x, to denote deterministic values, and calligraphic letters, e.g., $\mathcal{X}$, to denote sets. The sets of real numbers, rational numbers, non-negative integers and integers are denoted by $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{N}$, and $\mathbb{Z}$ respectively, where $\mathbb{N}^+$ denotes the set of positive integers. Sans-Serif font is used for denoting matrices, e.g., B, where the element at the i-th row and the j-th column of B is denoted with $(\mathsf{B})_{i,j}$, $i,j \in \mathbb{N}$. For $k,l \in \mathbb{N}^+$ we denote the all-zero $k \times l$ matrix with $\mathsf{0}_{k\times l}$, all-zero $k\times k$ square matrix with $\mathsf{0}_k$ and the $k\times k$ identity matrix with I_k. For a $k \times k$ matrix C we use $\max\mathrm{Eig}\{\mathsf{C}\}$ to denote its maximal eigenvalue, $\Lambda_i^{(k)}\{\mathsf{C}\}$ to denote its i-th ordered eigenvalue in descending order, $0 \le i \le k-1$, $\mathrm{Tr}\{\mathsf{C}\}$ to denote its trace, and $\mathrm{Det}(\mathsf{C})$ to denote its determinant. We use $\mathrm{rank}(\mathsf{B})$ to denote the rank of a matrix B and $\mathrm{range}(\mathsf{B})$ to denote its column range. Column vectors are denoted with boldface letters, where lower-case letters denote deterministic vectors, e.g., $\mathbf{x}$, and upper-case letters denote random vectors, e.g., $\mathbf{X}$; the i-th element of $\mathbf{x}$, $i \in \mathbb{N}$, is denoted with $(\mathbf{x})_i$, and for $a,b \in \mathbb{N}$, $b > a$, we write $x_a^b \triangleq \big[(\mathbf{x})_a, (\mathbf{x})_{a+1}, \ldots, (\mathbf{x})_b\big]^T$, where we also denote $x^{(b)} \equiv x_0^{b-1}$. We use $X \sim F_X$ to denote that the cumulative distribution function (CDF) of the RV X is F_X. Specifically, $\mathbf{X} \sim \mathbb{N}(\mathbf{x}_0, \mathsf{C})$ denotes a real Gaussian random vector $\mathbf{X}$ with mean $\mathbf{x}_0$ and covariance matrix C. For the RVs X and Y we use $X \perp\!\!\!\perp Y$ to denote that they are statistically independent. Transpose, Euclidean norm, 1-norm, ∞-norm, stochastic expectation, differential entropy, and mutual information are denoted by $(\cdot)^T$, $\|\cdot\|$, $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\mathbb{E}\{\cdot\}$, $h(\cdot)$, and $I(\cdot;\cdot)$, respectively, and we define $a^+ \triangleq \max\{0, a\}$. A square $k \times k$ real matrix A is called positive definite, denoted $\mathsf{A} \succ 0$ (positive semidefinite, resp., denoted $\mathsf{A} \succcurlyeq 0$) if for any $x^{(k)} \in \mathbb{R}^k$, $x^{(k)} \neq \mathsf{0}_{k\times 1}$, it holds that $(x^{(k)})^T \cdot \mathsf{A} \cdot x^{(x)} > 0$ ($(x^{(k)})^T \cdot \mathsf{A} \cdot x^{(x)} \ge 0$, resp.). We use $\stackrel{(dist.)}{=}$ to denote equality in distribution, $\log(\cdot)$ to denote the base-2 logarithm, and $\ln(\cdot)$ to denote the natural logarithm. Lastly, for any sequence $\mathbf{y}[i]$, $i \in \mathbb{N}$, and $b \in \mathbb{N}^+$, we use $\mathbf{y}^{(b)}$ to denote the column vector obtained by stacking the first b sequence elements $\Big[\big(\mathbf{y}[0]\big)^T, \ldots, \big(\mathbf{y}[b-1]\big)^T\Big]^T$.

B. Wide-Sense Cyclostationary Processes

We next review some preliminaries from the theory of cyclostationarity, beginning with the definition of a wide-sense cyclostationary process:

Definition 1 (A Wide-Sense Cyclostationary Process [20, Def. 17.1], [4, Pg. 296]): A scalar stochastic process $\{X(t)\}_{t\in\mathcal{T}}$, where $\mathcal{T} = \mathbb{Z}$ or $\mathcal{T} = \mathbb{R}$, is referred to as *WSCS* if both its mean and its correlation function are periodic with respect to $t \in \mathcal{T}$ with some period $T_p \in \mathcal{T}$: $\mu_X(t) \triangleq \mathbb{E}\{X(t)\} = \mu_X(t+T_p)$ and $c_X(t,\tau) \triangleq \mathbb{E}\{X(t+\tau)X(t)\} = c_X(t+T_p,\tau)$, $\forall t,\tau \in \mathcal{T}$.

Next, we consider WSACS random processes, recalling first the definition of an almost-periodic function:

Definition 2 (An Almost-Periodic Function [21, Defs. 10, 11], [22, Ch.1.2]): A function $f(t), t \in \mathcal{T}$, where $\mathcal{T} = \mathbb{Z}$ or $\mathcal{T} = \mathbb{R}$, is referred to as *almost-periodic* if for every $\epsilon > 0$, there exists a number $l(\epsilon) \in \mathcal{T}$, $l(\epsilon) > 0$, such that for any $t_0 \in \mathcal{T}$, there exists $\tau_\epsilon \in \big(t_0, t_0 + l(\epsilon)\big)$, such that

$$\sup_{t\in\mathcal{T}} |f(t+\tau_\epsilon) - f(t)| < \epsilon.$$

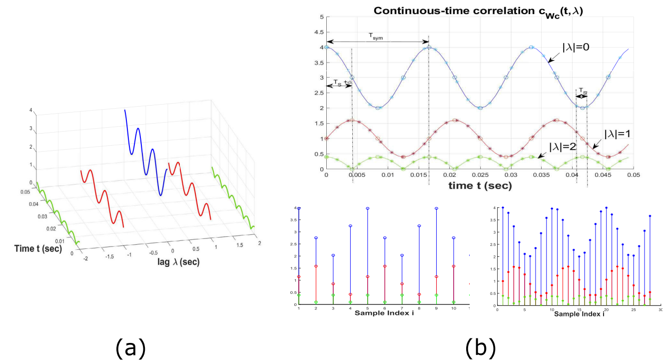

Fig. 1. (a). 3D visualization of the correlation function periodic for a CT WSCS random process at three lag values: $|\lambda| = [0, 1, 2]$; (b). Illustration of the DT correlation function obtained by sampling at different sampling rates and offsets. $'$o$'$ marks synchronous sampling, with $\frac{T_\mathrm{s}}{T_\mathrm{sym}} = \frac{1}{4} \in \mathbb{Q}$, and $'*'$ marks asynchronous sampling, with $\frac{T_\mathrm{s}}{T_\mathrm{sym}} = \frac{1}{3\pi} \notin \mathbb{Q}$.

Definition 3 (A Wide-Sense Almost-Cyclostationary Process [20, Def. 17.2], [4, Pg. 296], [22, Ch. 1.3]): A scalar stochastic process $\{X(t)\}_{t\in\mathcal{T}}$, where $\mathcal{T} = \mathbb{Z}$ or $\mathcal{T} = \mathbb{R}$, is called *WSACS* if its mean and its autocorrelation function are almost-periodic functions with respect to $t \in \mathcal{T}$.

C. Sampling of CT WSCS Random Processes

To facilitate the application of digital processing, the received signal is sampled at the receiver. Consider a DT random process $X_{T_\mathrm{s},\tau_0}[i]$, $i \in \mathbb{Z}$, obtained by sampling the CT WSCS random process $X(t)$, which has a period of T_p, with a sampling interval of T_s and at a sampling phase of $\tau_0 \in [0, T_p)$, i.e., $X_{T_\mathrm{s},\tau_0}[i] \triangleq X(i \cdot T_\mathrm{s} + \tau_0)$. In the following, we demonstrate that contrary to sampled stationary processes, for CT cyclostationary processes, the values of T_s and τ_0 have a significant impact on the statistics of the resulting sampled process $X_{T_\mathrm{s},\tau_0}[i]$. Consequently, the common practice of applying stationary theory to such scenarios can lead to erroneous results, e.g., [23]. As an example, we illustrate in Fig 1(a) the correlation function of a CT WSCS random process at three CT lag values: $|\lambda| = 0, 1, 2$ [sec], where the period in t is $T_\mathrm{sym} = 1/60$ [sec]. The bottom plots in Fig 1(b) depict two sampling scenarios: In the first case, shown in the bottom-left plot, the CT signal is sampled at a sampling interval of $T_\mathrm{s} = \frac{T_\mathrm{sym}}{4}$ and the sampling phase is $\tau_0 = \frac{T_\mathrm{s}}{2\pi}$. This is depicted by the $'$o$'$ markers on the CT correlation function at the top figure in (b). Observe that in this case, the correlation function at the three lag values, $|\lambda| = 0, 1, 2$, depicted by the blue, red and green plots, respectively, is periodic in DT. As stated in Section I, such sampling, which maintains the periodicity of the statistics in DT, is referred to as *synchronous sampling*, and the resulting DT process is WSCS. The bottom-right plot of Fig 1(b) depicts the DT correlation function obtained for $\tau_0 = 0$ and $T_\mathrm{s} = \frac{T_\mathrm{sym}}{3\pi}$, represented by the $'*'$ markers on the CT correlation function at the top figure in (b). Hence, $\frac{T_\mathrm{s}}{T_\mathrm{sym}}$ is an irrational number, and the DT correlation function is not periodic but is almost-periodic at all lags, contrary to the first case. Therefore, the resulting DT random process is not WSCS but WSACS [13, Sec. 3.9]. As stated in Section I such a sampling scenario is referred to as *asynchronous sampling*.

This example clearly demonstrates that when sampling CT WSCS processes, slight variations in the sampling interval and the sampling phase may result in significantly different statistics for the sampled processes. As explained in Section I, such sampling scenarios exist in many Communication Channels, e.g., in interference-limited communications, in which the noise component corresponds to a sampled CT WSCS process. The consequence of the variability of the statistics of sampled CT WSCS processes is that channel capacity of a DT channel obtained by sampling the output of a CT additive WSCS noise channel strongly depends on the actual sampling rate. In a recent study, [15], we characterized the capacity of such DT channels assuming the sampled additive noise is Gaussian and *memoryless*. This study aims to generalize the capacity characterization to DT channels with additive Gaussian noise with a finite memory, where the noise processes arise from the sampling of CT finite-memory WSCS processes representing communications signals, as developed in the subsequent sections.

III. Model and Problem Formulation

In this section we derive the considered channel model and detail the different assumptions on the communications scenario. We begin with the mathematical definition of the setup.

Definition 4: A finite-memory DT real-valued random process $W[i]$, $i \in \mathbb{Z}$, with memory $\tau_\mathrm{m} \in \mathbb{N}^+$, satisfies that $\forall i \in \mathbb{Z}$,

$$\mathbb{E}\big\{W[i] \cdot W[i-\lambda]\big\} = \mathbb{E}\big\{W[i]\big\} \cdot \mathbb{E}\big\{W[i-\lambda]\big\}, \ \forall |\lambda| > \tau_\mathrm{m}. \tag{1}$$

Such a model is appropriate for processes representing single-carrier digitally modulated signals with inter-symbol interference (ISI), where the memory of the process is determined by the finite length of the overall channel impulse response (CIR) (i.e., the CIR which accounts for the pulse shape, propagation through the physical medium and analog filters in the signal paths at the transmitter and at the receiver). This model is also appropriate for processes representing orthogonal frequency division multiplexing (OFDM) modulated signals, as in such processes the lack of correlation follows from the finite duration of the OFDM symbol, where the cyclic prefix (CP) interval and receiver processing induce statistical independence between the received samples corresponding to different OFDM symbols.

Definition 5: An $[R, l]$ *code* with rate $R \in \mathbb{R}^{++}$ and blocklength $l \in \mathbb{N}^+$ consists of: *1)* A message set $\mathcal{U} \triangleq \{1, 2, \ldots, 2^{lR}\}$; *2)* An encoder e_l which maps a message $u \in \mathcal{U}$ into a codeword $x_u^{(l)} = \Big[x_u[0], x_u[1], \ldots, x_u[l-1]\Big]$; and *3)* A decoder d_l which maps the sequence of channel outputs, denoted $y^{(l)}$, into a message $\hat{u} \in \mathcal{U}$.

The set $\{x_u^{(l)}\}_{u=1}^{2^{lR}}$ is referred to as the *codebook* and the message u is selected uniformly and independently from $\mathcal{U}$. Note that as the noise process is generally non-stationary, then the probability distribution of the channel output sequence, denoted $Y^{(l)}$, depends on an initial channel state $s_0 \in \mathcal{S}_0$, where $\mathcal{S}_0$ is the set of all possible initial channel states. The set $\mathcal{S}_0$ will be explicitly stated in the context of this work in

Sec. III-C. The average probability of error when the initial channel state is s_0 is defined as:

$$P_e^l(s_0) = \frac{1}{2^{lR}} \sum_{u=1}^{2^{lR}} \Pr\left(d_l\left(Y^{(l)}\right) \neq u \middle| U=u, s_0\right).$$

Definition 6: A rate $R_c \in \mathbb{R}^{++}$ is *achievable* if for every $\eta_1, \eta_2 > 0$, $\exists l_0 \in \mathbb{N}^+$ such that $\forall l > l_0$ there exists an $[R, l]$ code which satisfies

$$\sup_{s_0 \in \mathcal{S}_0} P_e^l(s_0) < \eta_1, \tag{2a}$$

and

$$R \geq R_c - \eta_2. \tag{2b}$$

Capacity is defined as the supremum over all achievable rates.

Lastly, we recall the definition of the limit-inferior in probability [19, Def. 1.3.2]:

Definition 7: The *limit-inferior in probability* of a sequence of real RVs $\{Z_k\}_{k\in\mathbb{N}^+}$ is defined as

$$\text{p}-\liminf_{k\to\infty} Z_k \triangleq \sup\left\{\alpha \in \mathbb{R} \middle| \lim_{k\to\infty} \Pr\left(Z_k < \alpha\right) = 0\right\} \triangleq \alpha_0. \tag{3}$$

Hence, $\forall \tilde{\alpha} > \alpha_0$, $\exists \epsilon > 0$, such that there exist countably many $k \in \mathbb{N}^+$ for which $\Pr(Z_k < \tilde{\alpha}) > \epsilon$.

As was stated in [15], see also [19, Pg. VIII], the quantity $\text{p}-\liminf_{k\to\infty} Z_k$ is well-defined even when the sequence of RVs $\{Z_k\}_{k\in\mathbb{N}^+}$ does not converge in distribution. This makes the limit-inferior in probability applicable to the analysis of scenarios in which methods based on the law of large numbers cannot be applied, e.g., when non-stationary and non-ergodic processes are considered [24]. We note, however, that the application of Def. 7 in information-theoretic analysis typically results in expressions which are very difficult to compute.

A. Problem Formulation

Consider a real-valued zero-mean CT WSCS Gaussian random process $W_c(t) \in \mathbb{R}$, whose autocorrelation function, $c_{W_c}(t, \lambda) \triangleq \mathbb{E}\{W_c(t+\lambda)W_c(t)\}$, is continuous in t and in λ, periodic in t with a period $T_{\rm pw} \in \mathbb{R}^{++}$ and has a finite correlation length $\lambda_m \in \mathbb{R}^{++}$, i.e., $c_{W_c}(t, \lambda) = c_{W_c}(t + T_{\rm pw}, \lambda)$, $\forall t, \lambda \in \mathbb{R}$, and $c_{W_c}(t, \lambda) = 0$ for any $|\lambda| \geq \lambda_m > 0$.

Since the autocorrelation function $c_{W_c}(t, \lambda)$ is continuous and periodic, it is sufficient to characterize its properties only over a compact interval $\mathcal{T}_0 \in \mathbb{R}$ where $\mathcal{T}_0 = [t_0, t_0 + T_{\rm pw}]$, for some arbitrary $t_0 \in \mathbb{R}$; it follows that $c_{W_c}(t, \lambda)$ is bounded and uniformly continuous with respect to time $t \in \mathcal{T}_0$ and lag $\lambda \in \mathbb{R}$ [25, Ch. III, Thm. 3.13].

The process $W_c(t)$ is sampled at a finite, positive sampling interval $T_s(\epsilon)$ such that $T_{\rm pw} = (p + \epsilon) \cdot T_s(\epsilon)$ where $p \in \mathbb{N}^+$ and $\epsilon \in [0, 1)$, resulting in the DT random process $W_\epsilon[i] \triangleq W_c(\tau_0 + i \cdot T_s(\epsilon))$, $i \in \mathbb{Z}$. In this study we focus on the case in which the sampling interval is smaller than the correlation span of the CT process $W_c(t)$, hence, letting $\tau_0 \in \mathbb{R}$ denote the sampling phase corresponding to index $i = 0$, then $W_\epsilon[i]$ is a zero-mean Gaussian random process whose autocorrelation function is given by:

$$\begin{aligned} c_{W_\epsilon}^{\{\tau_0\}}[i, \Delta] &\equiv \mathbb{E}\Big\{W_\epsilon[i+\Delta] \cdot W_\epsilon[i] \Big| \tau_0\Big\} \\ &\triangleq \mathbb{E}\left\{W_c\left(\frac{(i+\Delta) \cdot T_{\rm pw}}{p+\epsilon} + \tau_0\right) \cdot W_c\left(\frac{i \cdot T_{\rm pw}}{p+\epsilon} + \tau_0\right)\right\} \\ &= c_{W_c}\left(i \cdot \frac{T_{\rm pw}}{p+\epsilon} + \tau_0, \Delta \cdot \frac{T_{\rm pw}}{p+\epsilon}\right). \end{aligned} \tag{4}$$

It follows that $c_{W_\epsilon}^{\{\tau_0\}}[i, \Delta] = 0$ for all $|\Delta| > \left\lceil \frac{(p+1)\cdot\lambda_m}{T_{\rm pw}} \right\rceil \triangleq \tau_{\rm m} < \infty$, hence, the correlation length of $W_\epsilon[i]$ is finite. In the following we say that $W_\epsilon[i]$ has a *finite memory* $\tau_{\rm m} < \infty$, referring to the finite correlation length of the sampled process $W_\epsilon[i]$. Due to the fact that $W_\epsilon[i]$ is a sampled physical noise process, it can be assumed that the correlation matrix corresponding to any sequence length is positive definite, see elaboration in Comment A.1.

Next, observe that from (4), see also [15], it follows that when $\epsilon \in \mathbb{Q}^{++}$, i.e., $\exists u, v \in \mathbb{N}^+$ s.t. $\epsilon = \frac{u}{v}$, then the process $W_\epsilon[i]$ is WSCS with a period which is equal to $p_{u,v} \triangleq p \cdot v + u$. Such a sampling scenario corresponds to *synchronous sampling*, for which capacity was characterized in [5]. On the other hand, when ϵ is irrational ($\epsilon \notin \mathbb{Q}$), the resulting DT process $W_\epsilon[i]$ is WSACS [13, Sec. 3.9], which corresponds to *asynchronous sampling*. In this work, we consider DT channels with real-valued additive, finite-memory WSACS Gaussian noise. Letting $X[i]$ and $Y_\epsilon[i]$ denote the real-valued channel input and output, respectively, at time $i \in \mathbb{Z}$, the input-output relationship for the transmission of a sequence of $l \in \mathbb{N}^+$ channel inputs is given by:

$$Y_\epsilon[i] = X[i] + W_\epsilon[i], \qquad i \in \{0, 1, \ldots, l-1\}, \tag{5}$$

where the subscript ϵ is retained to indicate the synchronization mismatch between the period of the CT noise and the sampling interval at the transmitter. The noise process $W_\epsilon[i]$ has a temporal correlation length of $\tau_{\rm m}$ samples. The channel input, $X[i]$, is subject to a per-codeword power constraint P,

$$\frac{1}{l} \sum_{i=0}^{l-1} \left(x_u[i]\right)^2 \leq P, \qquad u \in \mathcal{U}. \tag{6}$$

Lastly, it is assumed that the input and the noise in (5) satisfy the independence assumption:

Assumption 1: $\big\{X[i]\big\}_{i\in\mathbb{Z}}$ is independent of $\big\{W_\epsilon[i]\big\}_{i\in\mathbb{Z}}$.

The channel model in (5) is particularly relevant for modelling channels in which $W_c(t)$ is a digitally modulated interfering communications signal, and is thus a CT WSCS process with a period which is equal to its symbol duration [25, Sec. 5]. For example, when $W_c(t)$ is an OFDM modulated signal with a sufficiently large number of subcarriers, then it can be modeled as a Gaussian process [26], and the correlation function is periodic with a period which is equal to the duration of an OFDM symbol. As another example, when $W_c(t)$ is a linearly modulated quadrature amplitude modulation (QAM) signal with a partial response pulse shaping, then, when the ISI spans a sufficiently long interval, it follows that $W_c(t)$ is modeled as a Gaussian process, see [27, Sec. III-A]. Here, again the correlation function is periodic, where the period is equal to the duration of an information symbol. In both examples, the process $W_{\rm c}(t)$ has a finite correlation length. In such interference-limited setups, as the sampling rate at the receiver is generally not synchronous with the symbol rate of the interferer, then the resulting DT interfering signal can be modeled as a finite-memory WSACS process, giving rise to the channel input-output relationship in (5). *Our goal is to characterize the capacity of the channel (5) subject to the power constraint (6).* We also note that the model of Eqn. (5) has previously been used in the analysis of communications systems, e.g., in [28], which studied feedback capacity for stationary, finite-dimensional Gaussian channels, hence, the current model adds the non-stationary noise characteristics to the model considered in [28] (without feedback).

The analysis in this work relies also on the following two assumptions:

Assumption 2: The transmitter (Tx) and the receiver (Rx), are both assumed to know the CT noise correlation function $c_{W_{\rm c}}(t,\lambda)$.

Note that, as explained in detail in the next subsection, non-stationarity of the sampled noise statistics implies that knowledge of $c_{W_{\rm c}}(t,\lambda)$ is not sufficient for maximizing the rate, which is in contrast to the situation for DT channels with additive stationary noise. Therefore, it is also assumed that

Assumption 3: The propagation delay between the transmitter and receiver is negligible compared to the period and to the maximal slope of the (uniformly continuous) noise correlation function.

This assumption implies that the receiver can attain perfect sampling time synchronization with the transmitter, as well as that both the transmitter and the receiver can identify the temporal phase within the period of the noise correlation function at any time instant. This is referred to in the following as *perfect Tx-Rx timing synchronization*. Note that the sampling interval used by the transmitter and receiver is not synchronized with the period of the noise correlation function in the sense that their ratio is an irrational number.

B. An Example Scenario: Lowpass Channel With ACGN

As another motivating example for the DT model of Eqn. (5), consider a CT baseband channel model with memory, in which the received signal is given by $Y(t)=h(t)*X(t)+W(t)$, where $'*'$ denotes the linear convolution, and $W(t)$ is a WSCS Gaussian random process with a finite memory (i.e., an interfering communications signal). For simplicity of the discussion assume that the channel can be approximated as a first-order stable lowpass filter with a transfer function (TF)

$$H(s)=\frac{1}{s+a}, \qquad a\in\mathbb{R}^{++},\ \Re\{s\}>-a.$$

Letting $X[m]$ denote the DT information sequence at a rate of $\frac{1}{T_s}$, the overall received CT signal component can be modeled as

$$\sum_{m=-\infty}^{\infty}X[m]\delta(t-mT_s)*h(t)=\sum_{m=-\infty}^{\infty}X[m]\cdot h(t-mT_s).$$

Thus, sampling at intervals of $T_s\in\mathbb{R}^{++}$, we obtain the following DT relationship between $X[m]$, $Y[n]\triangleq Y(n\cdot T_{\rm s})$, $W[n]\triangleq W(n\cdot T_{\rm s})$ and $h[n]\triangleq h(n\cdot T_s)$:

$$Y[n]=\sum_{m=-\infty}^{\infty}X[m]\cdot h[n-m]+W[n],$$

which is a real-valued linear, time-invariant DT channel with memory, where $W[n]$ is a DT Gaussian random process with a finite memory. Observe that the equivalent DT CIR is obtained from the CIR of the CT channel by sampling (i.e., via the so-called impulse-invariance method, see, e.g., [20, Sec. 11.3.2.2]), which results in a DT TF of the form:

$$H(z)=\frac{1}{1-e^{-a\cdot T_s}\cdot z^{-1}}, \quad |z|>e^{-a\cdot T_s}.$$

Since $\alpha\triangleq e^{-a\cdot T_s}<1$ for all considered T_s, then $H(z)$ corresponds to a stable and causal channel with a stable and causal inverse. Such a channel model is very popular in communications, see, e.g., [29]. The zero-forcing equalizer for this model is the highpass filter whose TF is obtained by taking the inverse of $H(z)$: $H_{ZF}(z)=1-\alpha\cdot z^{-1}$, and its impulse response is $h_{ZF}[n]=\delta[n]-\alpha\cdot\delta[n-1]$. Observe that after filtering $Y[n]$ with $h_{ZF}[n]$, the resulting interference process, namely, $W[n]*h_{ZF}[n]$, is a random process with a finite memory, hence, after zero-forcing equalization, the resulting overall DT channel is appropriately modeled via Eqn. (5). In particular, when the interference is a single-carrier QAM signal, filtering effectively increases the duration of the CIR, hence, filtering increases the interference's memory. Note that when the interference is an OFDM signal, then, if the length of the overall CIR is shorter than the length of the CP (which should hold by design), then subsequent symbols remain independent.

C. The Initial Channel State

Note that the channel model in Eqn. (5) with the input power constraint of Eqn. (6) corresponds to an additive Gaussian noise channel, where the Gaussian noise is non-stationary. Due to Gaussianity, the distribution of the noise is completely characterized by the first two moments. As the noise has a zero-mean, then non-stationarity of the noise manifests itself via the fact that its correlation matrix is not Toeplitz. It is noted that such a general model was discussed in [30], subject to an *average sum-power* constraint. In this context we make the following observation: In order to transmit a message, the transmitter has to select the respective codeword. As the sampled noise correlation function is generally non-periodic, then the noise correlation matrices corresponding to *different message intervals* may be *completely different*. Accordingly, to facilitate analysis of such channels, it is necessary to introduce an initial state variable which identifies the noise correlation function present during the transmission of the message. For the case of sampled cyclostationary noise, the initial state corresponds to the relative location of the first sample of the DT noise correlation function within the period of the CT noise correlation function, which, for the subsequent discussion is denoted with τ_0. Due to the periodicity of $c_{W_{\rm c}}(t,\lambda)$ in $t\in\mathbb{R}$, it is enough to consider $\tau_0\in[0,T_{\rm pw})$, hence, the initial state space $\mathcal{S}_0$ is the interval $[0,T_{\rm pw})$. In this context, it is noted that the models [30] and of [31] do not consider an initial channel state in the model while pertaining to be relevant to non-stationary channels. This implies that the models in [30] and [31] make a *hidden assumption* that while the channel is non-stationary, *there exists a synchronization mechanism that sets the channel statistics to be identical for subsequent message intervals*.

Moreover, as the correlation function of the optimal input is a function of the sampled noise correlation function, it follows that knowledge of the noise correlation function $c_{W_c}(t,\lambda)$ alone at the transmitter is *not sufficient* for obtaining the optimal performance, and knowledge of the initial state τ_0 for each message is required. Due to *Assumption 3*, this knowledge is available in our setup. This knowledge requirement is not necessary when the additive noise is stationary.

IV. Capacity Characterization When Transmission Delay Is Not Allowed

When the initial state τ_0 is known at the transmitter, it is able to adapt the statistics of the codebook such that the achievable rate is maximized. When *transmission delay is not allowed*, then once a message transmission has been completed, the subsequent message is transmitted immediately, without further delay, at the sampling phase τ_0 present at the time the transmission of the current message has finished. Due to *Assumption 3*, of perfect Tx-Rx timing synchronization, it follows that both the receiver and the transmitter know the DT noise correlation function at every message transmission. Letting $X_{\text{opt}}^{(k)}$ denote the input process which maximizes the mutual information between $X^{(k)}$ and $Y_{\epsilon}^{(k)}$ when the sampling phase is τ_0, i.e., maximizes $I(X^{(k)};Y_{\epsilon}^{(k)}|\tau_0)$, we obtain the following capacity characterization:

Theorem 1: Consider the channel (5) with power constraint (6), when the transmitter can identify the sampling phase within a period of the CT noise correlation function, $\tau_0 \in [0, T_{\text{pw}})$, and is allowed to adapt its information rate and codebook accordingly. If no transmission delay is allowed, then capacity is given by

$$C_{\epsilon} = \frac{1}{T_{\text{pw}}} \int_{\tau_0=0}^{T_{\text{pw}}} C_{\epsilon}(\tau_0) \mathrm{d}\tau_0,$$

where $C_{\epsilon}(\tau_0) \triangleq \liminf_{k\to\infty} \frac{1}{k} I(X_{\text{opt}}^{(k)};Y_{\epsilon}^{(k)}|\tau_0)$, as long as the maximizing input $X_{\text{opt}}^{(k)}$ is Gaussian, with a distribution which depends on τ_0 and satisfies $\frac{1}{k}\text{Tr}\Big\{\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big\} \leq P$ and $\frac{1}{k^2}\text{Tr}\Big\{\big(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\big)^2\Big\} \underset{k\to\infty}{\longrightarrow} 0$.

Proof: The proof is provided in Appendix A. ∎

Comment 1 (The Requirement on the Trace of the Squared Input Correlation Matrix): We note that the condition $\frac{1}{k^2}\text{Tr}\Big\{\big(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\big)^2\Big\} \underset{k\to\infty}{\longrightarrow} 0$ is introduced in order to satisfy the per-codeword power constraint (6). Such a condition was also considered in [30, Sec. VIII].

Comment 2 (Capacity With an Average Power Constraint): Instead of the per-codeword power constraint (6) one may consider a more-relaxed average sum-power constraint, as in, e.g., [32, Eqn. (7)], [30, Eqn. (7)]:

$$\frac{1}{l}\sum_{i=0}^{l-1} \mathbb{E}_U |x_U[i]|^2 \leq P. \tag{7}$$

With such a constraint then Thm. 1 holds *without* requiring the consideration of $\text{Tr}\Big\{\big(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\big)^2\Big\}$. This follows as any codebook of length k, generated randomly according to a Gaussian distribution $X_{\text{opt}}^{(k)} \sim \mathbb{N}\Big(0_{k\times 1}, \mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big)$ such that $\frac{1}{k}\text{Tr}\Big\{\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big\} = P - \delta$, where $\delta > 0$ is arbitrarily small, will satisfy

$$\begin{aligned}\frac{1}{k}\sum_{i=0}^{k-1}\mathbb{E}_U\left\{|x_{U,\text{opt}}[i]|^2|\right\} &\overset{(a)}{=} \frac{1}{2^{kR}}\sum_{u=0}^{2^{kR}-1}\frac{1}{k}\sum_{i=0}^{k-1}|x_{u,\text{opt}}[i]|^2\\ &\overset{(b)}{=} \frac{1}{2^{kR}}\sum_{u=0}^{2^{kR}-1}\omega_u\\ &\underset{k\to\infty}{\overset{(c)}{\longrightarrow}} \mathbb{E}\{\Omega_1\} \qquad \text{(in probability)}\\ &\leq P,\end{aligned}$$

where (a) follow by the uniform selection of codewords for transmission; in (b) we consider the realizations $\omega_u \triangleq \frac{1}{k}\sum_{i=0}^{k-1}|x_{u,\text{opt}}[i]|^2$: Note that for different indexes $u \in \mathcal{U}$, the realizations ω_u are generated independently using the same multivariate distribution for all messages $u \in \mathcal{U}$. The expectation of the generating RV, $\mathbb{E}\{\Omega_1\}$, is equal to

$$\begin{aligned}\mathbb{E}\{\Omega_1\} &= \mathbb{E}\Big\{\frac{1}{k}\sum_{i=0}^{k-1}|X_{1,\text{opt}}[i]|^2|\Big\}\\ &= \frac{1}{k}\text{Tr}\Big\{\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big\} \leq P, \qquad u\in\mathcal{U}.\end{aligned}$$

Step (c) follows by the weak law of large numbers [33, Sec. 7.4], as the mean of 2^{kR} independent realizations of the independent and identically distributed (i.i.d) RVs $\{\Omega_u\}_{u\in\mathcal{U}}$ converges in probability to its expectation.

Then, we can conclude that the corresponding $\text{p-}\liminf_{k\to\infty} Z_{k,\epsilon}\Big(F_{X_{\text{opt}}^{(k)}|\tau_0}|\tau_0\Big)$, defined in (A.11), is achievable by considering the proof of the direct part of [19, Thm. 3.2.1], as there is no need to restrict the selected codewords when generating them according to the distribution $\mathbb{N}\Big(0_{k\times 1}, \mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big)$. This follows as for sufficiently large k the codebooks generated according to this Gaussian distribution satisfy the average constraint (7) with a probability arbitrarily close to 1, as k increases. Thus, subject to (7), the optimal input for Thm. 1 is $X_{\text{opt}}^{(k)} \sim \mathbb{N}\Big(0_{k\times 1}, \mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big)$, with $\frac{1}{k}\text{Tr}\Big\{\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0)\Big\} \leq P$.

When codebook adaptation is allowed but the rate has to be fixed, the following corollary is immediate:

Corollary 1: Consider the channel (5) with power constraint (6), when the transmitter can identify the sampling phase within a period of the CT noise correlation function, $\tau_0 \in [0, T_{\text{pw}})$, and is allowed to adapt its codebook accordingly. If the message rate has to be fixed, and no transmission delay is allowed, then capacity is given by

$$C_\epsilon = \min_{\tau_0 \in [0, T_{\rm pw})} \liminf_{k\to\infty} \frac{1}{k} I(X_{\rm opt}^{(k)}; Y_\epsilon^{(k)} | \tau_0),$$

as long as the maximizing input $X_{\rm opt}^{(k)}$ is Gaussian, with a distribution which depends on τ_0 and satisfies $\frac{1}{k}{\rm Tr}\left\{ \mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) \right\} \le P$ and $\frac{1}{k^2}{\rm Tr}\left\{ \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^2 \right\} \underset{k\to\infty}{\longrightarrow} 0$.

V. Capacity Characterization When Transmission Delay Is Allowed

In this section we consider a transmission scenario in which the transmitter is allowed to delay the transmission of the next message such that it would begin at the optimal sampling phase within the period of the noise correlation function. In such a scenario, capacity can be expressed via a sequence of capacities of DT additive WSCS Gaussian noise channels.

A. Approaching the Relationship (5) via a Sequence of DT ACGN Channels

To characterize the capacity of the channel (5), we define for each $n \in \mathbb{N}^+$ a rational number $\epsilon_n \triangleq \frac{\lfloor n\cdot\epsilon \rfloor}{n}$ and a corresponding DT process $W_n[i] \triangleq W_{\rm c}\left(\frac{i\cdot T_{\rm pw}}{p+\epsilon_n} + \tau_0\right)$, $i \in \mathbb{Z}$, $\tau_0 \in [0, T_{\rm pw})$. As follows from the discussion in Sec. III-A, the DT process $W_n[i]$ is a *zero-mean WSCS Gaussian random process* with period $p_n = p \cdot n + \lfloor n \cdot \epsilon \rfloor$. Note that as

$$\tau_{\rm m} \triangleq \left\lceil \frac{(p+1)\cdot\lambda_m}{T_{\rm pw}} \right\rceil \ge \left\lceil \frac{(p+\epsilon_n)\cdot\lambda_m}{T_{\rm pw}} \right\rceil,$$

then the correlation length of the noise $W_n[i]$ can be set to $\tau_{\rm m}$ for all $n \in \mathbb{N}^+$, hence, the noise process $W_n[i]$ has a finite memory of $\tau_{\rm m}$.

Next, we define a channel with input $X[i]$ and output $Y_n[i]$ via the input-output relationship:

$$Y_n[i] = X[i] + W_n[i], \tag{8}$$

where the channel input is subject to the per-codeword power constraint (6). The channel (8) is an additive noise channel with correlated, finite-memory WSCS Gaussian noise $W_n[i]$, whose period is p_n. The capacity of the channel (8) was explicitly derived in [6, Thm. 1], by transforming the DT channel (8) into a MIMO channel via the decimated component decomposition (DCD) [20, Sec. 17.2]. For blocklengths which are integer multiples of p_n, the DCD transforms the process $W_n[i]$ into an equivalent p_n-dimensional stationary process $\tilde{\boldsymbol{W}}_n^{\{p_n\}}[i]$, such that $\left(\tilde{\boldsymbol{W}}_n^{\{p_n\}}[i]\right)_b = W_n[i\cdot p_n + b]$, $0 \le b \le p_n - 1$. We define the correlation matrix for sampling phase τ_0 as $\mathsf{C}_{\tilde{\boldsymbol{W}}_n^{\{p_n\}}}[\tau;\tau_0] \triangleq \mathbb{E}\left\{ \tilde{\boldsymbol{W}}_n^{\{p_n\}}[i+\tau] \cdot \left(\tilde{\boldsymbol{W}}_n^{\{p_n\}}[i]\right)^T \middle| \tau_0 \right\}$. From the finite correlation length of the process $W_n[i]$, it follows that for all n such that $p_n > \tau_{\rm m}$, $\left(\mathsf{C}_{\tilde{\boldsymbol{W}}_n^{\{p_n\}}}[\tau;\tau_0]\right)_{k_1,k_2} = 0$, $\forall |\tau| > 1$, $\forall k_1, k_2 \in \{0, 1, \ldots, p_n - 1\}$, see [5, Sec. IV]. Next, for all $\theta \in [-\pi, \pi)$, define the $p_n \times p_n$ matrix $\mathsf{C}'_{\tilde{\boldsymbol{W}}_n^{\{p_n\}}}(\theta;\tau_0) \triangleq \sum_{\tau=-1}^{1} \mathsf{C}_{\tilde{\boldsymbol{W}}_n^{\{p_n\}}}[\tau;\tau_0] e^{-j\theta\tau}$, let $\{\Lambda'_{k,n}(\theta;\tau_0)\}_{k=0}^{p_n-1}$ be the eigenvalues of $\left(\mathsf{C}'_{\tilde{\boldsymbol{W}}_n^{\{p_n\}}}(\theta;\tau_0)\right)^{-1}$, and let $\bar{\Delta}^{\{p_n;\tau_0\}}$ be the unique solution to

$$\frac{1}{2\pi\cdot p_n} \sum_{k=0}^{p_n-1} \int_{\theta=-\pi}^{\pi} \left(\bar{\Delta}^{\{p_n;\tau_0\}} - \left(\Lambda'_{k,n}(\theta;\tau_0)\right)^{-1} \right)^+ {\rm d}\theta = P. \tag{9}$$

Then, the capacity of the channel (8), denoted $C_n(\tau_0)$, is given as [6, Thm. 1]:

$$C_n(\tau_0) = \frac{1}{4\pi\cdot p_n} \sum_{k=0}^{p_n-1} \int_{\theta=-\pi}^{\pi} \left(\log\left(\bar{\Delta}^{\{p_n;\tau_0\}} \cdot \Lambda'_{k,n}(\theta;\tau_0) \right) \right)^+ {\rm d}\theta$$
$$\text{[bits per channel use]}. \tag{10}$$

Note that the capacity $C_n(\tau_0)$ generally depends on the initial sampling phase τ_0. Then, maximizing over the initial sampling phase we define

$$C_n = \max_{\tau_0 \in [0, T_{\rm pw}]} C_n(\tau_0). \tag{11}$$

With the aid of (9)-(11), we subsequently obtain a characterization for the capacity of the asynchronously-sampled channel (5), denoted C_ϵ, when transmission delay of up to $\tau_{\rm m} \cdot T_{\rm s}(\epsilon) + T_{\rm pw}$ between subsequent messages is allowed. This is stated in the following theorem:

Theorem 2: Consider the channel (5) with power constraint (6), when the transmitter can identify the sampling phase within a period of the noise correlation function, $\tau_0 \in [0, T_{\rm pw})$, and may delay its message transmission time by up to $\tau_{\rm m} \cdot T_{\rm s}(\epsilon) + T_{\rm pw}$ time units. If the noise correlation function $c_{W_{\rm c}}(t,\tau)$, characterized in Section III-A, satisfies

$$\min_{0 \le t \le T_{\rm pw}} \left\{ c_{W_{\rm c}}(t,0) - 2\tau_{\rm m} \cdot \max_{|\lambda| > \frac{T_{\rm pw}}{p+1}} \left\{ |c_{W_{\rm c}}(t,\lambda)| \right\} \right\} \ge \gamma_1 > 0, \tag{12}$$

and the power constraint P satisfies

$$P > \max_{t \in [0, T_{\rm pw}]} \left(c_{W_{\rm c}}(t,0) + 2\tau_{\rm m} \cdot \max_{|\lambda| > \frac{T_{\rm pw}}{p+1}} \left\{ |c_{W_{\rm c}}(t,\lambda)| \right\} \right), \tag{13}$$

then, for any fixed value of $\epsilon \in (0,1)$, $\epsilon \notin \mathbb{Q}$, capacity is given by

$$C_\epsilon = \liminf_{n\to\infty} C_n, \tag{14}$$

where C_n is obtained via (9)-(11). Furthermore, Gaussian inputs are optimal.

Proof: The proof is detailed in Appendix B. ■

Comment 3 (On the Capacity for Synchronous Sampling): When $\epsilon \in \mathbb{Q}^{++}$, i.e., $\epsilon = \frac{u}{v}$ for some $u, v \in \mathbb{N}^+$, then the DT noise process $W_\epsilon[i]$ is WSCS with a period which is equal to $p_{u,v} = p \cdot v + u$. As noted in Sections I and II-C, such a sampling scenario corresponds to *synchronous sampling*, whose capacity, when τ_0 is given, was characterized in [5], [6] and is given by (9)−(10), where p_n is replaced by $p_{u,v}$, $\tilde{\boldsymbol{W}}_n^{\{p_n\}}[i]$ is replaced by $\tilde{\boldsymbol{W}}_{u,v}^{\{p_{u,v}\}}[i]$, and the quantities appearing in the statement of (9)−(10) are replaced by appropriate corresponding quantities. Note that for such an ϵ, then when $n = b \cdot v$, $b \in \mathbb{N}^+$, we have that $\epsilon_n = \frac{u}{v}$, consequently, for $\epsilon \in \mathbb{Q}^{++}$ it follows that $\liminf_{n\to\infty} C_n \leq C_{\frac{u}{v}}$. Yet, from the upper bound in (B.17) it holds that $C_\epsilon \leq \liminf_{n\to\infty} C_n$, hence for $\epsilon \in \mathbb{Q}^{++}$ we immediately obtain $C_\epsilon = \liminf_{n\to\infty} C_n$.

Comment 4 (Elaboration on Condition (12)): Condition (12) guarantees that for any sequence of k samples of the noise process, denoted $W_n^{(k)} \equiv \{W_n[i]\}_{i=0}^{k-1}$, and for any sampling phase $\tau_0 \in [0, T_{\rm pw})$, the correlation matrix, denoted $\mathsf{C}_{W_n^{(k)}}(\tau_0)$, is strictly diagonally dominant (SDD) [34, Eqn. (4)]. To see this, recall that by definition, $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,v} \triangleq \mathbb{E}\{W_n[u] \cdot W_n[v] | \tau_0\} \equiv c_{W_n}^{\{\tau_0\}}[v, u-v]$, $0 \leq u, v \leq k-1$. The diagonal dominance can be verified by noting that for any $0 \leq u \leq k-1$ it holds that

$$\begin{aligned}
&\left|\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,u}\right| - \sum_{v=0, v\neq u}^{k-1} \left|\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,v}\right| \\
&= \left|\mathbb{E}\{(W_n[u])^2 | \tau_0\}\right| - \sum_{v=0, v\neq u}^{k-1} \left|\mathbb{E}\{W_n[u] \cdot W_n[v] | \tau_0\}\right| \\
&\overset{(a)}{=} c_{W_{\rm c}}\left(u \cdot \frac{T_{\rm pw}}{p+\epsilon_n} + \tau_0, 0\right) \\
&\quad - \sum_{v=0, v\neq u}^{k-1} \left| c_{W_{\rm c}}\left(u \cdot \frac{T_{\rm pw}}{p+\epsilon_n} + \tau_0, (v-u) \cdot \frac{T_{\rm pw}}{p+\epsilon_n}\right)\right| \\
&\geq \min_{0 \leq t \leq T_{\rm pw}} \left\{ c_{W_{\rm c}}(t, 0) - 2\tau_{\rm m} \cdot \max_{|\lambda| > \frac{T_{\rm pw}}{p+1}} \{|c_{W_{\rm c}}(t, \lambda)|\}\right\} \\
&\geq \gamma_1 > 0, \qquad (15)
\end{aligned}$$

where (a) follows from the definition of the autocorrelation function (4), since for the real-valued random process $W_n[i]$, $i \in \mathbb{N}$ we can write $\mathbb{E}\{W_n[u] \cdot W_n[v] \; |\tau_0\} = \mathbb{E}\{W_n[v] \cdot W_n[u] \; |\tau_0\}$, $0 \leq u, v \leq k-1$.

Thus, condition (12) guarantees that the correlation decreases sufficiently fast as the lag increases, such that a strictly diagonally dominant noise correlation matrix is obtained for any $n, k \in \mathbb{N}^+$. This facilitates upper bounding the eigenvalues of the inverse noise correlation matrix, see Appendix B-A. It is noted, however, that condition (12) is stricter than the actual requirement, which is more involved to state analytically: In fact, from step (c) in the derivation of Eqn. (B.6), it is only required that for every $n \in \mathbb{N}^+$ sufficiently large, as well as for ϵ, the correlation matrices $\mathsf{C}_{W_n^{(k)}}(\tau_0)$ and $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ are SDD for all $\tau_0 \in [0, T_{\rm pw}]$. In the simulations in Sec. VI we directly verify the SDD condition.

Comment 5 (Elaboration on Condition (13)): The lower bound on the power P guarantees that for any sequence of k noise samples, the eigenvalues of the corresponding noise correlation matrix are smaller than P. Then, when in Step (b) in the derivation of (B.29), waterfilling is applied over the eigenvalues of the noise correlation matrix, see, e.g., [30, Eqns. (15)-(16)], it follows that power is allocated to all eigenvalues. This facilitates the bounding of the WSCS channel capacity by the mutual information of any segment of length k, where k is sufficiently large, up to an arbitrarily small error.

Comment 6 (Intuition From Stationary Analysis Does Not Apply Here): We emphasize that while it seems intuitive that the limit in (14) holds, our result shows that for this limit to hold, additional conditions on the noise statistics are required. This highlights the fact that when considering non-stationary channels, then intuition based on stationary processes may lead to incorrect perceptions. In the current work, we obtain capacity characterization when the noise correlation decays sufficiently fast. If this is not the case, it is not possible to uniformly bound the difference between the mutual information expressions corresponding to the channels (5) and (8), subject to (6), and consequently, showing the interchangeability of the limits in n (the approximation index) and in k (the sequence length) becomes an involved task. We also require P to be sufficiently large to allow relating the mutual information of a finite segment of length k and capacity. Let F_X denote the CDF of the RV X. and consider, for example, the limit $\liminf_{k\to\infty} \frac{1}{k} I(X^{(k)}; Y_\epsilon^{(k)} | \tau_0)$ in Corollary 1. In Lemma B.1 we show that $\lim_{n\to\infty} \frac{1}{k} I(X_{n,\rm opt}^{(k)}; Y_n^{(k)} | \tau_{n,k}^{\rm opt}) = \frac{1}{k} I(X_{\rm opt}^{(k)}; Y_\epsilon^{(k)} | \tau_{\epsilon,k}^{\rm opt})$, where $(F_{X_{n,\rm opt}^{(k)}}, \tau_{n,k}^{\rm opt})$ and $(F_{X_{\rm opt}^{(k)}}, \tau_{\epsilon,k}^{\rm opt})$ maximize the left-hand side (LHS) and right-hand side (RHS) respectively. Following the proof of Lemma B.1 it is straightforward to conclude that

$$\begin{aligned}
\liminf_{k\to\infty} \frac{1}{k} & I(X_{\rm opt}^{(k)}; Y_\epsilon^{(k)} | \tau_{\epsilon,k}^{\rm opt}) \\
&= \liminf_{k\to\infty} \lim_{n\to\infty} \frac{1}{k} I(X_{n,\rm opt}^{(k)}; Y_n^{(k)} | \tau_{n,k}^{\rm opt}).
\end{aligned}$$

However, since the convergence of the sequence $\left\{\frac{1}{k} I(X_{n,\rm opt}^{(k)}; Y_n^{(k)} | \tau_{n,k}^{\rm opt})\right\}_{n\in\mathbb{N}^+}$ is generally not uniform in $k \in \mathbb{N}^+$, it is not possible to switch the order of the limits on the RHS, and we cannot relate $\liminf_{k\to\infty} \frac{1}{k} I(X_{\rm opt}^{(k)}; Y_\epsilon^{(k)} | \tau_{\epsilon,k}^{\rm opt})$ and C_n. This lack of uniform convergence follows as we show in the proof of Lemma B.1 that the distance $\left|\frac{1}{k} I(X_{\rm opt}^{(k)}; Y_\epsilon^{(k)} | \tau_{\epsilon,k}^{\rm opt}) - \frac{1}{k} I(X_{n,\rm opt}^{(k)}; Y_n^{(k)} | \tau_{n,k}^{\rm opt})\right|$ is proportional to the distance $\zeta[i] \triangleq \left| c_{W_{\rm c}}\left(\frac{i \cdot T_{\rm pw}}{p+\epsilon} + \tau_{\epsilon,k}^{\rm opt}, \frac{\Delta \cdot T_{\rm pw}}{p+\epsilon}\right) - c_{W_{\rm c}}\left(\frac{i \cdot T_{\rm pw}}{p+\epsilon_n} + \tau_{n,k}^{\rm opt}, \frac{\Delta \cdot T_{\rm pw}}{p+\epsilon_n}\right)\right|$, $0 \leq i \leq k-1$. For a fixed $n \in \mathbb{N}^+$, we obtain that $\zeta[i]$ periodically increases and decreases over the range $0 \leq i \leq k-1$. Then, as k increases, this distance may increase up to the maximal magnitude of the correlation function, and as consequence the mutual information expressions do not converge as k increases. Thus, to keep this distance bounded as k increases, n has to increase as well, which implies that convergence in n is not uniform in k.

Comment 7 (Relationship With the Work of Cover and Pombra): In [30], the capacity of additive Gaussian noise channels with and without feedback was considered. By analyzing the distribution of a quadratic form in Gaussian RVs, it is shown in [30] that the asymptotic equipartition property applies to nonergodic Gaussian processes. While in Appendix A we also analyze a quadratic form in Gaussian RVs, it is emphasized that the analysis for our situation is considerably more involved than for the situation in [30, Sec. V], as in our scenario the weighting matrix is not the inverse of the correlation matrix of the Gaussian vector, and moreover, it is an indefinite matrix. The resulting RV is thus *a weighted sum* of chi-square RVs, which is not distributed as a chi-square RV with a higher degree-of-freedom, differently from [30]. In fact, there is no explicit expression for the probability density function (PDF) of the above resulting RV, which necessities the use of a completely different set of arguments in the analysis in Appendix A. It is also noted that the information-spectrum framework, which was introduced several years after the work of [30], has not been applied, as far as we know, to the capacity analysis of channels with additive non-stationary Gaussian noise. Lastly, note that in the achievability proof in [30], the exponent of the probability of decoding error depends on the blocklength, see [30, Eqns. (66)-(67)]. Thus, it is not clear how it is possible to conclude a vanishing probability of error for a given rate $C_{n,\mathrm{FB}}$ in the asymptotic as the blocklength increases to infinity, using the arguments in [30, Sec, VII] without additional conditions.

Comment 8 (Evaluating Capacity in the Presence of Multiple Interferers): The setup in Sec. III-A considered the case of a single interferer. We note that when multiple interferers are present at fixed locations and when the channels between the interferers and the receiver are invariant, then the aggregate interference is a CT WSCS Gaussian process. In addition to Gaussianity of the aggregate interference, we note that, in order to apply the scheme derived in the proof of Thm. 2, the transmitter should acquire and synchronize with the noise correlation function. In this context we may consider two possible scenarios: In the first scenario, referred to as *partial coordination*, the receiver and transmitter can obtain (e.g., through a control channel) the signal parameters of each interferer (e.g., modulation type, symbol duration, pulse shape for single carrier or subcarrier frequencies for OFDM). With these parameters, the receiver and transmitter can obtain the CT correlation function of each interferer. Then, to obtain the aggregate CT correlation function, the receiver needs to inform the transmitter the delays at which each interferer is received. In multi-interferers scenarios in which this is feasible, then the approach of Thm. 2 can be applied. In the second scenario, referred to as *uncoordinated interferers*, the transmitter and receiver each need to independently obtain the correlation function of the aggregate CT interference. In such a case, as the relative delays from each interferer to the receiver and to the transmitter are different, then the estimated aggregate correlation function will likely be different between the transmitter and the receiver. Therefore, in such a scenario, multiple uncoordinated interferers cannot be handled via the scheme of Thm. 2.

VI. Numerical Examples and Discussion

In this section we use numerical evaluations to derive insights from the analytic capacity characterization of Thm. 2. First, in Subsection VI-A, we consider the evolution of $C_n(\tau_0)$ w.r.t the index n and the impact of the sampling phase $\phi = \tau_0/T_{\mathrm{pw}} \in [0, 1)$ on capacity. Next, in Subsection VI-B, we study the variations of the capacity of the sampled DT channel for different sampling rates and different sampling phases. We also compare the capacity results with the capacity obtained for additive *memoryless* WSCS Gaussian noise channels having the same noise power and signal power.

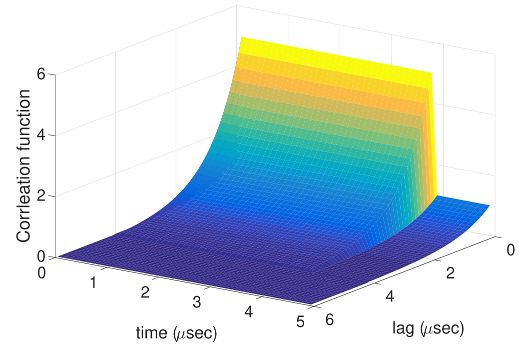

Fig. 2. The correlation function $c_{W_{\mathrm{c}}}(t, \lambda)$ of Eqn. (17) at positive lags $\lambda \geq 0$, with normalized sampling time offset $\phi = 0$ and $t_{\mathrm{dc}} = 0.75$.

To model the correlation function of the CT WSCS noise we define a periodic pulse function, $\Pi_{t_{\mathrm{dc}},t_{\mathrm{rf}}}(t)$, having a rise/fall time of $t_{\mathrm{rf}} = 0.01$, a period of 1, and a duty cycle (DC) of t_{dc}, which is varied in the range $0 \leq t_{\mathrm{dc}} \leq 0.75$; hence, $\Pi_{t_{\mathrm{dc}},t_{\mathrm{rf}}}(t) = \Pi_{t_{\mathrm{dc}},t_{\mathrm{rf}}}(t+1)\ \forall t \in \mathbb{R}$, and for $t \in [0, 1)$ the pulse function is expressed mathematically as:

$$\Pi_{t_{\mathrm{dc}},t_{\mathrm{rf}}}(t) = \begin{cases} \dfrac{t}{t_{\mathrm{rf}}} & t \in [0, t_{\mathrm{rf}}] \\ 1 & t \in (t_{\mathrm{rf}}, t_{\mathrm{dc}} + t_{\mathrm{rf}}) \\ 1 - \dfrac{t - t_{\mathrm{dc}} - t_{\mathrm{rf}}}{t_{\mathrm{rf}}} & t \in [t_{\mathrm{dc}} + t_{\mathrm{rf}}, t_{\mathrm{dc}} + 2 \cdot t_{\mathrm{rf}}] \\ 0 & t \in (t_{\mathrm{dc}} + 2 \cdot t_{\mathrm{rf}}, 1). \end{cases} \tag{16}$$

Let the period of the CT correlation function $c_{W_{\mathrm{c}}}(t, \lambda)$ be $T_{\mathrm{pw}} = 5$ [μsec]. Then, given a normalized sampling time offset $\phi \in [0, 1)$, we express the time-varying variance, $c_{W_{\mathrm{c}}}(t, 0)$, as

$$c_{W_{\mathrm{c}}}(t, 0) = 1 + 4 \cdot \Pi_{t_{\mathrm{dc}},t_{\mathrm{rf}}}\left(\frac{t}{T_{\mathrm{pw}}} - \phi\right).$$

For our setup, the correlation length of the noise process in CT is set to $\lambda_m = 4$ [μsec] and the temporal correlation is modeled as a decaying exponential function for all lags $|\lambda| \leq \lambda_m$, i.e., the correlation at any lag $\lambda > 0$ is given by

$$c_{W_{\mathrm{c}}}(t, \lambda) = \begin{cases} e^{-\lambda \cdot 10^6} \cdot c_{W_{\mathrm{c}}}(t, 0) & , 0 \leq \lambda \leq \lambda_m \\ 0 & , \lambda > \lambda_m \end{cases}, \tag{17}$$

and for $\lambda < 0$ we use $c_{W_{\mathrm{c}}}(t, \lambda) = c_{W_{\mathrm{c}}}(t + \lambda, -\lambda)$. This correlation function is depicted in Fig. 2 for a single period, $0 \leq t \leq T_{\mathrm{pw}}$, $t_{\mathrm{dc}} = 0.75$, $\phi = 0$, and $0 \leq \lambda \leq 6$.

A. Convergence of $\{C_n(\tau_0)\}_{n\in\mathbb{N}^+}$

As stated in Theorem 2, if the correlation function of the DT noise satisfies the condition (12) and the power satisfies condition (13), then the capacity with asynchronous sampling, C_ϵ, is equal to the limit-inferior of a sequence of capacities corresponding to synchronous sampling, $\{C_n(\tau_0)\}_{n\in\mathbb{N}^+}$. For evaluating this sequence, we set the following parameter

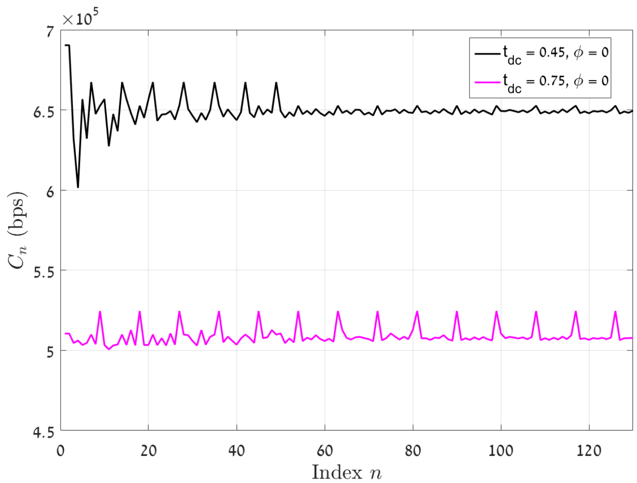

Fig. 3. $C_n(\tau_0)$ versus n, for $\tau_0 = 0$.

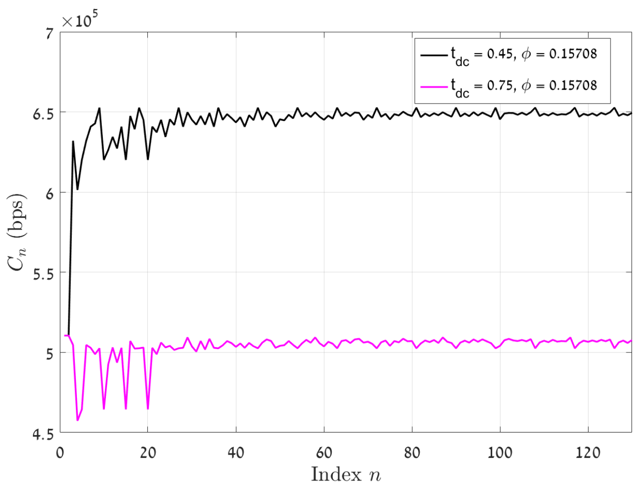

Fig. 4. $C_n(\tau_0)$ versus n, $\tau_0 = \frac{\pi}{20}T_{\rm pw}$.

values: $\epsilon = \frac{\pi}{7}$, $p = 2$, $t_{\rm dc} \in \{0.45, 0.75\}$, $\phi \in \{0, \frac{\pi}{20}\}$, and the input power constraint $P = 10$. First, we evaluate $C_n(\tau_0)$ using (9)–(10) for each n and then normalize it by its respective sampling interval $T_{\rm s}(\epsilon_n) \triangleq \frac{T_{\rm pw}}{p+\epsilon_n}$ to obtain $C_n(\tau_0)$ in bits per second (bps). We note that $C_n(\tau_0)$ can be evaluated irrespective of condition (12), yet to conclude about C_ϵ, either (12) or the SDD condition have to be verified as discussed in Comment 4, in addition to condition (13). Recall the definition of ϵ_n: $\epsilon_n = \frac{\lfloor n\cdot\epsilon\rfloor}{n} \to \epsilon$ as $n \to \infty$; then, it follows that the sampling interval $T_{\rm s}(\epsilon_n)$ converges to $T_{\rm s}(\epsilon) \triangleq \frac{T_{\rm pw}}{p+\epsilon}$ as n increases. We recall that since ϵ_n is rational, then the resulting DT sampled noise is WSCS with a fundamental period of $p_n = p \cdot n + \lfloor n \cdot \epsilon \rfloor$.

Figs. 3 and 4 depict $C_n(\tau_0)$ for normalized sampling time offsets of 0 and $\frac{\pi}{20}$ respectively, for both considered $t_{\rm dc}$ values, where $n = \{1, 2, \ldots, 130\}$. We observe from the figures that capacity is lower when $t_{\rm dc}$ is higher. This can be explained by the fact that the time-averaged noise power increases as $t_{\rm dc}$ increases. We also observe that the variations in the capacity $C_n(\tau_0)$ are more pronounced at smaller n. This is because at smaller n, the resulting fundamental period of the DT noise correlation function, p_n, consists of only a few samples, which are sparsely spaced across the period of the CT noise correlation function. Then, for the smaller values of n, as n varies, the sampling interval varies significantly, and consequently, the values of the sampled noise correlation function may significantly vary as well. At higher n, (i.e., higher p_n), it is observed that, as expected, the sequence $\{C_n(\tau_0)\}_{n\in\mathbb{N}^+}$ does not converge to a limiting value, since the limiting noise process $W_\epsilon[i]$ is non-stationary. However, for sufficiently large n, the variations of $C_n(\tau_0)$ as n increases, seem to follow a regular pattern. This can be explained by noting that for higher n, as n increases, the variations of the sampling instances of the CT noise correlation function become smaller, and accordingly, the values of the sampled correlation function do not vary significantly with $n \in \mathbb{N}^+$.

It is also observed from both Figs. 3 and 4 that at the smaller values of n, the nature of the variations in $C_n(\tau_0)$ is highly dependent on $\phi = \tau_0/T_{\rm pw}$. For example, with $t_{\rm dc} = 0.45$ and at $n \in [2, 25]$ the value of $C_n(0)$ is within the range $C_n(0) \in [0.601, 0.690]$ megabits per second (Mbps) and $C_n(\frac{\pi}{20} \cdot T_{\rm pw}) \in [0.510, 0.652]$ Mbps; with $t_{\rm dc} = 0.75$ and $n \in [2, 25]$ then $C_n(0) \in [0.501, 0.525]$ Mbps and $C_n(\frac{\pi}{20} \cdot T_{\rm pw}) \in [0.457, 0.510]$ Mbps. At higher n, these capacity variations become periodic within a constant range.

Figs. 3 and 4 also clearly demonstrate that the capacity with *synchronous* sampling may depend on the *sampling phase* ϕ (i.e., the values of $C_n(\phi \cdot T_{\rm pw})$ as n increases may depend on ϕ). Note that the *capacity* with *asynchronous* sampling, which is the limit-inferior of $\{C_n\}_{n\in\mathbb{N}^+}$, is *independent* of the *sampling phase*. This is in agreement with engineering intuition: Since with asynchronous sampling the resulting DT process is WSACS, it is reasonable that capacity should be affected mainly by the DC and not by the sampling phase. In the setup of Thm. 2 this follows as the transmitter may delay the transmission of a message to start at the optimal phase, which is also known at the receiver (via knowledge of the autocorrelation function), thereby facilitating Tx-Rx coordination. Numerically, the limit-inferior of $C_n(\phi\cdot T_{\rm pw})$ for $t_{\rm dc} = 0.45$ was evaluated at 0.648 Mbps for both $\phi = 0$ and $\phi = \frac{\pi}{20}$, and for $t_{\rm dc} = 0.75$ it was evaluated at 0.503 Mbps for both values of ϕ.

To further illustrate this behaviour, Figs. 5 and 6 depict the capacity values for sampling time offsets $\phi \in [0, 2]$ for two values of the approximation indices n: At $n = 1$ ($p_n = p = 2$) we observe significant variations of $C_1(\phi \cdot T_{\rm pw})$ for both DC values, 45% and 75%. On the other hand, at a sufficiently high n, e.g., $n = 40$ ($p_n = 97$), we observe that capacity $C_{40}(\phi \cdot T_{\rm pw})$ varies very little with ϕ for both DC values. This follows as at longer periods the correlation function of the process $W_n[i]$ more closely resembles $c_{W_{\rm c}}(t, \lambda)$. We also observe that the variations are periodic in all setups, which is expected, as for a fixed and finite $n \in \mathbb{N}^+$, the DT noise process $W_n[i]$ is WSCS, thus, its DT correlation function will repeat identically after a single period shift (i.e., an integer value of ϕ).

B. Variations of Capacity With the Sampling Rate

Next, we examine how variation of the sampling rate affects the capacity of DT channels obtained by sampling CT channels with additive WSCS Gaussian noise having a finite

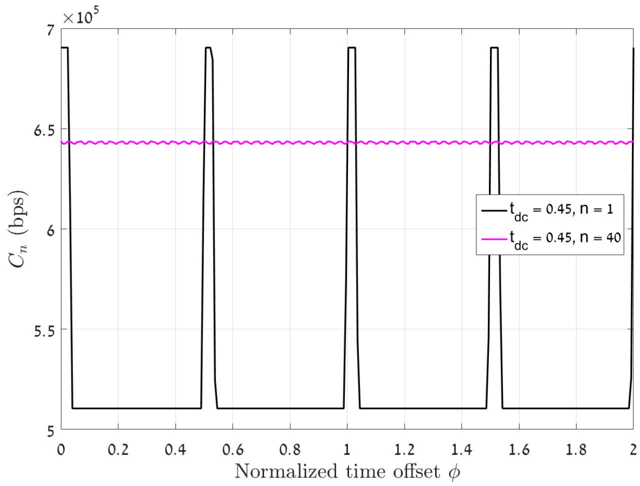

Fig. 5. $C_n(\tau_0)$ versus $\phi = \tau_0/T_{\rm pw}$; $t_{\rm dc} = 0.45$.

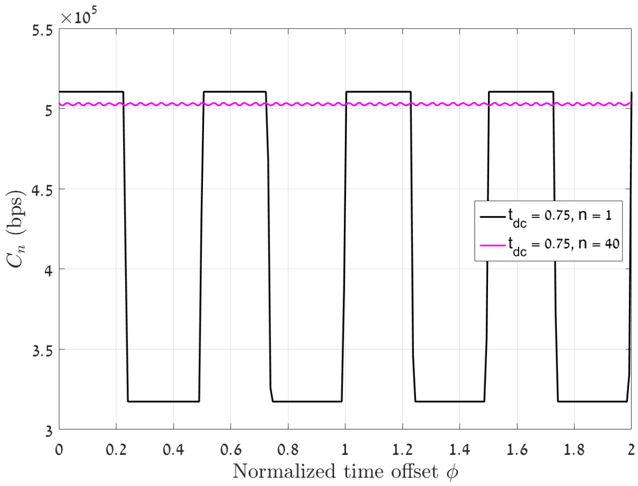

Fig. 6. $C_n(\tau_0)$ versus $\phi = \tau_0/T_{\rm pw}$; $t_{\rm dc} = 0.75$.

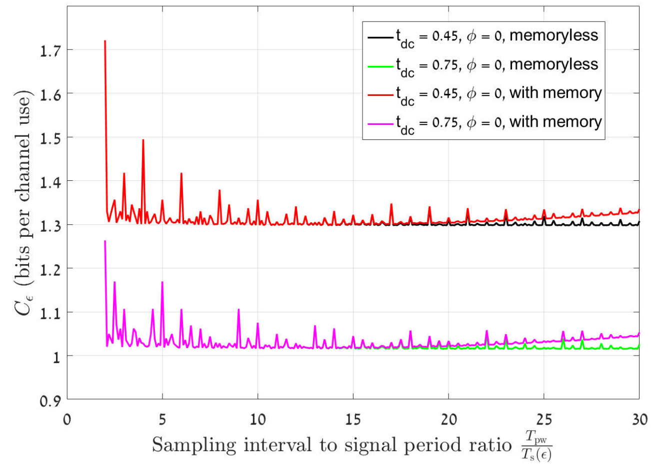

Fig. 7. $C_{\frac{u}{v}}(\tau_0)$ versus $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)}$ for offset $\tau_0 = 0$.

memory, and compare their capacity with that of DT channels with memoryless sampled noise having the same variance as the noise with finite memory. The results are depicted in Figs. 7 and 8, which present the evaluated capacity (in bits per channel use) for $\phi = 0$ and for $\phi = \frac{\pi}{20}$, respectively, for sampling intervals in the range $2 \leq \frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} \leq 30$. Recall from the problem formulation in Section III-A that $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = p + \epsilon$ where $p \in \mathbb{N}^+$ and $\epsilon \in [0, 1)$. In Figs. 7 and 8 we plot the capacity values C_ϵ for synchronous sampling, i.e., when ϵ can be written as $\epsilon = \frac{u}{v}$, $u, v \in \mathbb{N}^+$, and hence, the fundamental period of the noise statistics is given by $p_{u,v} = p \cdot v + u$. Recall that in this case, capacity depends on τ_0, thus we denote $C_\epsilon \equiv C_{\frac{u}{v}}(\tau_0)$, yet, this dependence becomes weaker as the period $p_{u,v}$ increases. To highlight the transition from memoryless channels to channels with memory, we use 10^7 instead of 10^6 in the power of the exponential function in (17).

For both figures, we observe an increase in the capacity (in bits per channel use) as the sampling rate increases. In addition, we note that when the values of p and $\epsilon = \frac{u}{v}$ result in a smaller value of the period $p_{u,v}$, the capacity varies significantly, as can be seen by the peaks and dips in both the memoryless Gaussian noise plot and the plot for Gaussian noise with a finite memory; it is also observed that the variations are different for different sampling time offsets. On the other hand, when the period $p_{u,v}$ is large, the capacity approaches the asynchronous-sampling capacity and the peaks/dips notably reduce. Moreover, at the longer periods, it is observed that capacity values are very similar for both sampling time offsets, which is reasonable when approaching the asynchronous sampling situation. Finally, it is evident from the figures that a slight change in the sampling rate can result in a significant change in the capacity. As an example, consider the plots for the finite-memory noise in Figs. 7 and 8, at $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 5$ (i.e., $p_{u,v} = 5$, which is a relatively small period) and $\phi = 0$: The capacity values for the noise with finite memory are 1.356 and 1.170 bits per channel use, for $t_{\rm dc} = 45\%$ and $t_{\rm dc} = 75\%$, respectively. However, when the sampling rate changes to $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 5.2$, these values change to 1.302 and 1.039, respectively. The impact of the sampling phase is more pronounced at smaller $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)}$: For $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 5$, the capacities at $\phi = \frac{\pi}{20}$ for $t_{\rm dc} = 45\%$ and $t_{\rm dc} = 75\%$ are 1.170 and 0.980 bits per channel use, respectively, which are very different from the respective values at $\phi = 0$ noted above. Lastly, consider $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 23.2$, i.e. $p_{u,v} = 116$, which is a relatively long period for the DT correlation function. For this sampling rate, the capacities (with memory) for $t_{\rm dc} = 45\%$ and $t_{\rm dc} = 75\%$ are 1.298 and 1.015 bits per channel use, respectively, for both $\phi = \frac{\pi}{20}$ and for $\phi = 0$.

Another property observed from Figs. 7 and 8 is that at relatively low sampling rates (i.e., when $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)}$ is smaller, e.g., $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} < 12$), the sampled channels for memoryless noise and for noise with a finite memory have approximately the same capacity. As the sampling rate is increased, it is observed that the sampled channel with a finite-memory noise has a higher capacity than the sampled memoryless channel, whose capacity does not vary much with the sampling rate variation. This is explained by observing that as the sampling rate increases, the sampled noise for the case of finite-memory CT noise begins to exhibit noise correlation, which can be utilized to increase capacity via waterfilling. As an Example, at $t_{\rm dc} = 75\%$, $\phi = \frac{\pi}{20}$ and $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 8.5$ the capacity is 1.014 bits per

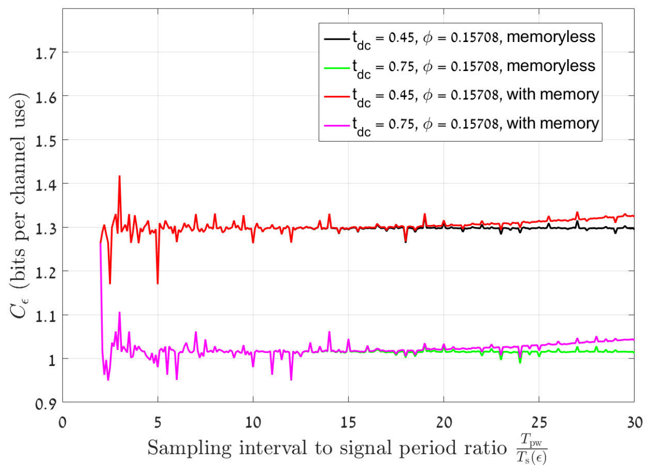

Fig. 8. $C_{\frac{u}{v}}(\tau_0)$ versus $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)}$ for offset $\tau_0 = \frac{\pi}{20}T_{\rm pw}$.

channel use for both the finite-memory and the memoryless cases. However, as $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)}$ increases, e.g., at $\frac{T_{\rm pw}}{T_{\rm s}(\epsilon)} = 29$, and $\phi = \frac{\pi}{20}$, the capacity is 1.310 bits per channel use for the channel with a finite-memory Gaussian noise, whereas it is 1.285 bits per channel use for the channel with memoryless Gaussian noise. Observe that this gap, in favor of the channel with sampled finite-memory noise widens as the sampling rate is further increased. That said, we note that while the model considered (5) does not account for additive thermal noise (since $c_{W_{\rm c}}(t,\lambda)$ has finite values), then at higher sampling rates, the impact of the thermal noise should also be accounted for in addition to the interference, as higher sampling rates are associated with higher receiver bandwidths. Accounting for the thermal noise will limit the capacity increase observed in the figures.

Our numerical evaluations reveal a very interesting phenomenon that should be considered when designing communications systems: It is observed that capacity is greatly dependent upon the precise value of the sampling rate. It is thus recommended to take the asynchronous capacity as the practical capacity value, even if the analytical capacity value due to the nominal sampling rate used in the system design is higher. The results also imply that increasing the sampling rate can increase the capacity even when the sampling rate is higher than the Nyquist rate.[2] This observation stands in contrast to the observation in [17], which studied linear, time-invariant channels with stationary Gaussian noise. Intuitively, this follows as in the current scenario, sampling is applied to a two-dimensional periodic function, hence it is not enough to be able to identify the temporal correlation profile, but also the periodicity of the correlation function, which may require higher sampling rates.

VII. Conclusion

In this work we analyzed the capacity of additive Gaussian noise channels obtained by sampling CT channels with additive WSCS Gaussian noise, focusing on the scenario in which the sampled noise is non-stationary. We first explained that in this case, maximizing the information rate requires Tx-Rx time synchronization w.r.t the correlation function of the CT noise, and it is not sufficient to have both the transmitter and the receiver know the noise correlation function without such synchronization. Subsequently, we derived a general capacity characterization when transmission delay is not allowed. Finally, we considered the scenario in which transmission delay of up to sum of the noise memory the noise period is allowed, for which we obtained a limiting capacity expression derived using original bounds on the optimal mutual information density rate of the channel. We then used the limiting expression to examine the impact of the combination of channel memory and sampling on the information rates of the resulting DT channel, and presented novel insights arising from this examination. This work is another step in the study of the relationship between sampling and capacity, which provides a much needed missing link between the analog domain models and the respective digital models obtained after sampling.

Appendix A
Proof of Thm. 1

We consider the mutual information density rate for the channel (5): Let $\tau_0 \in [0, T_{\rm pw})$ denote the sampling phase within a period of the correlation function of the CT noise process $W_{\rm c}(t)$. For a given $k \in \mathbb{N}^+$ and a given $\tau_0 \in [0, T_{\rm pw})$, let $F_{X^{(k)}|\tau_0} \equiv F_{X^{(k)}|\tau_0}\left(x^{(k)}|\tau_0\right)$ denote the CDF of the random vector $\left\{X[i]\right\}_{i=0}^{k-1}$, which is the channel input process when transmission begins at the sampling phase τ_0. Recall that the transmitter is aware of τ_0, hence, it can choose its codebook accordingly. Furthermore, the transmission scheme appends each codeword with $\tau_{\rm m}$ zeros, and the receiver discards the last $\tau_{\rm m}$ received channel outputs for each message reception. Thus, the received channel output sequences for different messages are statistically independent. In a similar manner as in [5, Appendix A], it follows that for sufficiently large k, this assumption does not affect the capacity. Lastly, define the random variable corresponding to the mutual information density rate for this transmission as (see [36, Lemma 7.16]):

$$Z_{k,\epsilon}\big(F_{X^{(k)}|\tau_0}|\tau_0\big) \triangleq \frac{1}{k}\log\frac{p_{Y_\epsilon^{(k)}|X^{(k)},\tau_0}\big(Y_\epsilon^{(k)}\big|X^{(k)},\tau_0\big)}{p_{Y_\epsilon^{(k)}|\tau_0}\big(Y_\epsilon^{(k)}|\tau_0\big)}.$$

Note that by [30], when τ_0 is given, then the mutual information for each $k \in \mathbb{N}^+$, for the additive Gaussian noise channel (5), $\frac{1}{k}I\big(X^{(k)};Y_\epsilon^{(k)}|\tau_0\big)$, is maximized, subject to an average sum-power constraint, by a Gaussian random input vector. We now analyze $Z_{k,\epsilon}\big(F_{X^{(k)}|\tau_0}|\tau_0\big)$ when $\left\{X[i]\right\}_{i=0}^{k-1}$ is distributed according to the Gaussian distribution which maximizes $\frac{1}{k}I\big(X^{(k)};Y_\epsilon^{(k)}|\tau_0\big)$ subject to the constraint $\frac{1}{k}\sum_{i=0}^{k-1}\mathbb{E}\big\{\big(X[i]\big)^2\big\} \le P$. In Lemma A.1, at the end of the proof, we will show that such a random codebook generation process results in the constraint (6) satisfied with a probability which is arbitrarily close to 1, as long as $\left\{X[i]\right\}_{i=0}^{k-1}$ satisfies the trace constraint which appears in the statement of the theorem. Let $X_{\rm opt}^{(k)}$ denote a random process

[2]See [35, Ch. 12.4.3] for elaboration on the Nyquist rate for bandlimited WSCS processes.

generated according to this maximizing input distribution, let $\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)$, $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$, and $\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)$ denote the correlation matrices of $Y_\epsilon^{(k)}$, $W_\epsilon^{(k)}$ and of $X_{\rm opt}^{(k)}$, respectively, *when* τ_0 *is given*, and recall the definition of the correlation matrix $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$:

$$\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)_{u,v} \triangleq \mathbb{E}\left\{W_\epsilon[u]\cdot W_\epsilon[v]\middle|\tau_0\right\} \equiv c_{W_\epsilon}^{\{\tau_0\}}[v,u-v], \tag{A.1}$$

for $(u,v)\in\mathcal{K}\times\mathcal{K}$, see Eqn. (4).

Comment A.1: Note that as $W_\epsilon[i]$ is a sampled *physical* noise process then *the matrix* $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ *has a full rank*. This follows as if $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ does not have a full rank, then by the definition of a multivariate Normal RVs, see [37, Def. 16.1], we obtain that at least one element in the vector $W_\epsilon^{(k)}$ is *identically equal* to a linear combination of the other elements. Such a linear relationship can be used to design a linear transformation at the receiver which completely eliminates the noise at one or more time indexes of the received sequence, leading to an infinite capacity value, which naturally does not correspond to physical scenarios.

Consider the scalar RV $V_{k,\epsilon}(\tau_0)\triangleq k\cdot Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)$:

$$\begin{aligned}
&V_{k,\epsilon}(\tau_0)\\
&\triangleq \log\frac{p_{Y_\epsilon^{(k)}|X_{\rm opt}^{(k)},\tau_0}\left(Y_\epsilon^{(k)}\middle|X_{\rm opt}^{(k)},\tau_0\right)}{p_{Y_\epsilon^{(k)}|\tau_0}\left(Y_\epsilon^{(k)}|\tau_0\right)}\\
&=\log\left(p_{W_\epsilon^{(k)}|\tau_0}\left(Y_\epsilon^{(k)}-X_{\rm opt}^{(k)}|\tau_0\right)\right)\\
&\qquad -\log\left(p_{Y_\epsilon^{(k)}|\tau_0}\left(Y_\epsilon^{(k)}|\tau_0\right)\right)\\
&\stackrel{(a)}{=}\frac{1}{2}\log\left(\frac{\mathrm{Det}\left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)}{\mathrm{Det}\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)}\right)\\
&\quad+\frac{\log(e)}{2}\left(Y_\epsilon^{(k)}\right)^T\left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)^{-1}Y_\epsilon^{(k)}\\
&\quad-\frac{\log(e)}{2}\left(Y_\epsilon^{(k)}-X_{\rm opt}^{(k)}\right)^T\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\left(Y_\epsilon^{(k)}-X_{\rm opt}^{(k)}\right),
\end{aligned}$$

where (a) follows since $p_{Y_\epsilon^{(k)}|\tau_0}\left(y^{(k)}|\tau_0\right)$ and $p_{W_\epsilon^{(k)}|\tau_0}\left(y^{(k)}-x^{(k)}|\tau_0\right)$ are Gaussian PDFs (note that since $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\succ 0$ then also $\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\succ 0$). Next, consider the scalar RV $\tilde{V}_{k,\epsilon}(\tau_0)$ defined in Eqn. (A.2), as shown at the bottom of the next page. Using $\tilde{V}_{k,\epsilon}(\tau_0)$ we can write $V_{k,\epsilon}(\tau_0)\stackrel{(dist.)}{=}\frac{1}{2}\log\left(\frac{\mathrm{Det}\left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)}{\mathrm{Det}\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)}\right)+\frac{\log(e)}{2}\tilde{V}_{k,\epsilon}(\tau_0)$. Define next the matrix $\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)$ via Eqn. (A.3), as shown at the bottom of the next page. Combining this definition with (A.2), we can express $\tilde{V}_{k,\epsilon}$ as $\tilde{V}_{k,\epsilon}=-\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}^T\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}$. Note that the matrix $\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)$ is a real, symmetric, full-rank, *indefinite* matrix, which is *different from the inverse correlation matrix of the Gaussian vector* $\left[\left(X_{\rm opt}^{(k)}\right)^T,\left(W_\epsilon^{(k)}\right)^T\right]^T$, hence, it is not possible to apply a simple decomposition as was done in, e.g., [30, Sec. V], to express the distribution of $Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)$.

As generally $\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\triangleq\begin{bmatrix}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) & 0_k\\ 0_k & \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\end{bmatrix}$ may not be a full-rank matrix, then let $\mathrm{rank}\left(\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\right)=2k-\tilde{\tilde{k}}_{\epsilon,k}$, where $\tilde{\tilde{k}}_{\epsilon,k}\in\mathbb{N}$ denotes the number of degenerate elements of $X_{\rm opt}^{(k)}$. We can now write the distribution of the Gaussian random vector $\left[\left(X_{\rm opt}^{(k)}\right)^T,\left(W_\epsilon^{(k)}\right)^T\right]^T$, separating the degenerate and the non-degenerate components, as follows [38, Sec. III-A–III-B]: First, decompose $\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)$ as

$$\begin{aligned}
\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)&=\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]\\
&\quad\cdot\begin{bmatrix}\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0) & 0_{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times\tilde{\tilde{k}}_{\epsilon,k}}\\ 0_{\tilde{\tilde{k}}_{\epsilon,k}\times(2k-\tilde{\tilde{k}}_{\epsilon,k})} & 0_{\tilde{\tilde{k}}_{\epsilon,k}\times\tilde{\tilde{k}}_{\epsilon,k}}\end{bmatrix}\\
&\quad\cdot\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]^T,
\end{aligned}\tag{A.4}$$

where $\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]$ is an orthogonal $2k\times 2k$ matrix, $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\in\mathbb{R}^{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$ is a symmetric positive-definite matrix, $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\succ 0$, the columns of $\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\in\mathbb{R}^{2k\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$ form an orthonormal basis for $\mathrm{range}\left(\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\right)$ and the columns of $\mathsf{P}_{\epsilon,k}^{0}(\tau_0)\in\mathbb{R}^{2k\times\tilde{\tilde{k}}_{\epsilon,k}}$ form an orthonormal basis for the null space of $\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)$. Then,

$$\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]^T\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}\stackrel{(dist.)}{=}\begin{bmatrix}\mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\\ 0_{\tilde{\tilde{k}}_{\epsilon,k}\times 1}\end{bmatrix},$$

$$\mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\sim\mathbb{N}\left(0_{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times 1},\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\right),$$

and we obtain that

$$\begin{aligned}
&\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}^T\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}\\
&=\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}^T\cdot\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]\\
&\quad\cdot\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]^T\\
&\quad\cdot\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]\\
&\quad\cdot\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]^T\cdot\begin{bmatrix}X_{\rm opt}^{(k)}\\W_\epsilon^{(k)}\end{bmatrix}\\
&\stackrel{(dist.)}{=}\left(\mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\right)^T\cdot\left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T\cdot\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\\
&\quad\cdot\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\cdot\mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}.
\end{aligned}$$

Observe that $\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)$ is a full-rank matrix, and since $\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)$ is also a full-rank matrix, then $\left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T\cdot\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)$ is full-rank. Since $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\succ 0$ and symmetric, it can be expressed as [39, Thm. 13.11] $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)=\left(\mathsf{R}_{\epsilon,k}(\tau_0)\right)^2$; where $\mathsf{R}_{\epsilon,k}(\tau_0)\in\mathbb{R}^{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$ is a positive-definite symmetric matrix. Then, letting $\mathsf{R}_{\epsilon,k}^{-1}(\tau_0) \in \mathbb{R}^{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$ denote the inverse of $\mathsf{R}_{\epsilon,k}(\tau_0)$, we can write

$$\begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix}^T \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix} \overset{(dist.)}{=} \left(\mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\right)^T \cdot \mathsf{R}_{\epsilon,k}^{-1}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0)\right)^T \cdot \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}^{-1}(\tau_0) \cdot \mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}.$$

Next, observe that

$$\Gamma_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \triangleq \mathsf{R}_{\epsilon,k}^{-1}(\tau_0) \cdot \mathrm{B}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \sim \mathbb{N}\left(0_{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times 1}, \mathsf{I}_{2k-\tilde{\tilde{k}}_{\epsilon,k}}\right),$$

and note that since $\mathsf{R}_{\epsilon,k}(\tau_0)$ and $\left(\mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0)\right)^T \cdot \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0)$ are full-rank matrices, then also $\mathsf{R}_{\epsilon,k}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0)\right)^T \cdot \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0)$ is a full-rank, square, symmetric, real, indefinite matrix, whose rank is $2k - \tilde{\tilde{k}}_{\epsilon,k}$. Thus, we can write [39, Thm. 11.27]:

$$\begin{aligned} \tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0) &\triangleq \mathsf{R}_{\epsilon,k}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0)\right)^T \cdot \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\text{CW}}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0) \\ &= \left(\mathsf{P}_{\epsilon,k}(\tau_0)\right)^T \cdot \mathsf{D}_{\epsilon,k}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}(\tau_0), \end{aligned}$$

where $\left(\mathsf{P}_{\epsilon,k}(\tau_0)\right)^T \cdot \mathsf{P}_{\epsilon,k}(\tau_0) = \mathsf{I}_{2k-\tilde{\tilde{k}}_{\epsilon,k}}$ and $\mathsf{D}_{\epsilon,k}(\tau_0) \in \mathbb{R}^{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0) \in \mathbb{R}^{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times(2k-\tilde{\tilde{k}}_{\epsilon,k})}$. Let $d_{\epsilon,ii,k}^{\{\tau_0\}}$ denote the i-th eigenvalue of the matrix $\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)$. Since $\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)$ is full-rank it follows that $d_{\epsilon,ii,k}^{\{\tau_0\}} \neq 0$ for $0 \le i \le 2k - \tilde{\tilde{k}}_{\epsilon,k} - 1$. Using this representation, we write

$$\begin{aligned} \begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix}^T \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) \begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix} &\overset{(dist.)}{=} \left(\Gamma_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\right)^T \cdot \left(\mathsf{P}_{\epsilon,k}(\tau_0)\right)^T \cdot \mathsf{D}_{\epsilon,k}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}(\tau_0) \cdot \Gamma_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \\ &\overset{(dist.)}{=} \left(\tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\right)^T \cdot \mathsf{D}_{\epsilon,k}(\tau_0) \cdot \tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}, \end{aligned} \tag{A.5}$$

where we define

$$\tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \triangleq \mathsf{P}_{\epsilon,k}(\tau_0) \cdot \Gamma_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \sim \mathbb{N}\left(0_{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times 1}, \mathsf{I}_{2k-\tilde{\tilde{k}}_{\epsilon,k}}\right). \tag{A.6}$$

Eventually, we obtain

$$\begin{aligned} \tilde{V}_{k,\epsilon}(\tau_0) &\overset{(dist.)}{=} -\left(\tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}\right)^T \cdot \mathsf{D}_{\epsilon,k}(\tau_0) \cdot \tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})} \\ &= \sum_{i=0}^{2k-\tilde{\tilde{k}}_{\epsilon,k}-1} \left(-d_{\epsilon,ii,k}^{\{\tau_0\}}\right) \cdot \left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2, \quad d_{\epsilon,ii,k}^{\{\tau_0\}} \in \mathbb{R}, \end{aligned}$$

where $\tilde{\Gamma}_{\epsilon,i,k}$, $0 \le i \le 2k - \tilde{\tilde{k}}_{\epsilon,k} - 1$ denotes the i-th element of the vector $\tilde{\Gamma}_\epsilon^{(2k-\tilde{\tilde{k}}_{\epsilon,k})}$. Observe from (A.6) that the elements $\tilde{\Gamma}_{\epsilon,i,k}$ are i.i.d Gaussian RVs, hence, $\left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2$ is a central chi-square random variable with a single degree of freedom [40, Example 5.2], which is denoted as $\left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2 \sim \mathcal{X}^2(1)$, and $\left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2$, $0 \le i \le 2k - \tilde{\tilde{k}}_{\epsilon,k} - 1$, are mutually independent

$$\begin{aligned} \tilde{V}_{k,\epsilon}(\tau_0) &\triangleq \left(Y_\epsilon^{(k)}\right)^T \left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)^{-1} Y_\epsilon^{(k)} - \left(Y_\epsilon^{(k)} - X_{\text{opt}}^{(k)}\right)^T \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \left(Y_\epsilon^{(k)} - X_{\text{opt}}^{(k)}\right) \\ &\overset{(dist.)}{=} \left(X_{\text{opt}}^{(k)} + W_\epsilon^{(k)}\right)^T \left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \left(X_{\text{opt}}^{(k)} + W_\epsilon^{(k)}\right) \\ &\qquad - \left(W_\epsilon^{(k)}\right)^T \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \left(W_\epsilon^{(k)}\right) \\ &= -\begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix}^T \begin{bmatrix} \mathsf{I}_k & \mathsf{I}_k \\ 0_k & \mathsf{I}_k \end{bmatrix}^T \begin{bmatrix} -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} & 0_k \\ 0_k & \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathsf{I}_k & \mathsf{I}_k \\ 0_k & \mathsf{I}_k \end{bmatrix} \begin{bmatrix} X_{\text{opt}}^{(k)} \\ W_\epsilon^{(k)} \end{bmatrix} \end{aligned} \tag{A.2}$$

$$\begin{aligned} \tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0) &\triangleq \begin{bmatrix} \mathsf{I}_k & 0_k \\ \mathsf{I}_k & \mathsf{I}_k \end{bmatrix} \begin{bmatrix} -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} & 0_k \\ 0_k & \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathsf{I}_k & \mathsf{I}_k \\ 0_k & \mathsf{I}_k \end{bmatrix} \\ &= \begin{bmatrix} -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} & -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \\ -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} & -\left(\mathsf{C}_{X_{\text{opt}}^{(k)}}(\tau_0) + \mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} + \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \end{bmatrix} \end{aligned} \tag{A.3}$$

$\mathcal{X}^2(1)$ RVs. We can now express $Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)$ as:

$$Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right) \stackrel{(dist.)}{=} \frac{1}{2k}\cdot\log\left(\frac{{\rm Det}\left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)}{{\rm Det}\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)}\right) + \frac{\log(e)}{2k}\sum_{i=0}^{2k-\tilde{\tilde{k}}_{\epsilon,k}-1}\left(-d_{\epsilon,ii,k}^{\{\tau_0\}}\right)\cdot\left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2.$$

Examining $Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)$, we note that since $\mathbb{E}\left\{Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)\right\} = \frac{1}{2k}\cdot\log\left(\frac{{\rm Det}\left(\mathsf{C}_{Y_\epsilon^{(k)}}(\tau_0)\right)}{{\rm Det}\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)}\right)$ [36, Eqn. (7.31)], then it necessarily should hold that $\sum_{i=0}^{2k-\tilde{\tilde{k}}_{\epsilon,k}-1}\left(-d_{\epsilon,ii,k}^{\{\tau_0\}}\right) = 0$. This can be verified via a direct derivation detailed in Eqns. (A.7), as shown at the bottom of the next page. In the steps leading to (A.7c), (a) follows since

$$\begin{bmatrix}\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) & \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\end{bmatrix}\cdot\begin{bmatrix}\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0) & \mathsf{0}_{(2k-\tilde{\tilde{k}}_{\epsilon,k})\times\tilde{\tilde{k}}_{\epsilon,k}}\\ \mathsf{0}_{\tilde{\tilde{k}}_{\epsilon,k}\times(2k-\tilde{\tilde{k}}_{\epsilon,k})} & \mathsf{0}_{\tilde{\tilde{k}}_{\epsilon,k}\times\tilde{\tilde{k}}_{\epsilon,k}}\end{bmatrix}\cdot\begin{bmatrix}\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) & \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\end{bmatrix}^T = \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\cdot\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\cdot\left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T;$$

and (b) follows from (A.4).

We now compute ${\rm Tr}\left\{\left(\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)\right)^2\right\}$: Begin by using Eqn. (A.7a) and write

$${\rm Tr}\left\{\left(\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)\right)^2\right\} = {\rm Tr}\left\{\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\cdot\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\right\}.$$

By computing explicitly the matrix product we obtain that

$$\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\cdot\tilde{\mathsf{C}}_\epsilon^{(k)}(\tau_0)\cdot\mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0) = \begin{bmatrix}\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) & \mathsf{0}_{k\times k}\\ \mathsf{0}_{k\times k} & \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\end{bmatrix}, \quad \text{(A.8)}$$

hence,

$${\rm Tr}\left\{\left(\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)\right)^2\right\} = 2\cdot{\rm Tr}\left\{\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right\}.$$

With this result we can compute the variance of $\tilde{V}_{k,\epsilon}(\tau_0)$, denoted by $\tilde{\sigma}_{\epsilon,k}^2(\tau_0)$, as follows:

$$\begin{aligned}\tilde{\sigma}_{\epsilon,k}^2(\tau_0) &= \sum_{i=0}^{2k-\tilde{\tilde{k}}_{\epsilon,k}-1}{\rm var}\left(d_{\epsilon,ii,k}^{\{\tau_0\}}\cdot\left(\tilde{\Gamma}_{\epsilon,i,k}\right)^2\right)\\ &= 2\cdot\sum_{i=0}^{2k-\tilde{\tilde{k}}_{\epsilon,k}-1}\left(d_{\epsilon,ii,k}^{\{\tau_0\}}\right)^2\\ &= 2\cdot{\rm Tr}\left\{\left(\tilde{\tilde{\mathsf{C}}}_\epsilon^{(k)}(\tau_0)\right)^2\right\}\\ &= 4\cdot\left(k-{\rm Tr}\left\{\left(\mathsf{I}_k+\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^{-1}\right\}\right).\end{aligned} \quad \text{(A.9)}$$

From the results of (A.7c) and (A.9) it follows that $\mathbb{E}\left\{Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)\right\} = \frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)$, and ${\rm var}\left(Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)\right) \leq \frac{3}{k}$. Since the variance of $Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)$ decreases as k increases, then, by Chebyshev's inequality [40, Eqn. (5-88)], we obtain $\Pr\left(\left|Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)-\frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)\right| > \frac{1}{k^{1/3}}\right) < \frac{3}{k^{1/3}}$, and we conclude that $\forall\delta>0$, $\exists k_0(\delta)\in\mathbb{N}^+$ s.t. $\forall k>k_0(\delta)$, it follows that

$$\Pr\left(Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right) < \frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)-\delta\right) < 3\delta. \quad \text{(A.10)}$$

Note that by definition of the limit-inferior, $\forall\delta>0$, $\exists k_1(\delta)\in\mathbb{N}^+$ s.t. $\forall k>k_1(\delta)$ (recall that $X_{\rm opt}^{(k)}$ maximizes $\frac{1}{k}I\left(X^{(k)};Y_\epsilon^{(k)}|\tau_0\right)$)

$$\frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right) > \liminf_{k\to\infty}\frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)-\delta,$$

hence, for all $k>\max\left\{k_0(\delta),k_1(\delta)\right\}$.

$$\Pr\left(Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right) < \liminf_{k\to\infty}\frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)-2\delta\right) < 3\delta,$$

and we conclude that

$$\begin{aligned}&{\rm p-}\liminf_{k\to\infty}Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)\\ &\triangleq \sup\left\{\alpha\in\mathbb{R}\,\Big|\,\lim_{k\to\infty}\Pr\left(Z_{k,\epsilon}\left(F_{X_{\rm opt}^{(k)}|\tau_0}|\tau_0\right)<\alpha\right)=0\right\}\\ &\geq \liminf_{k\to\infty}\frac{1}{k}I\left(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_0\right).\end{aligned} \quad \text{(A.11)}$$

Next, we consider the power constraint. The following lemma asserts that a Gaussian codebook generated according to a distribution which satisfies the trace constraints $\frac{1}{k}{\rm Tr}\left\{\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right\}\leq P$ and $\frac{1}{k^2}{\rm Tr}\left\{\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^2\right\}\xrightarrow[k\to\infty]{}0$, satisfies the per-codeword power constraint (6) asymptotically as $k\to\infty$ with a probability which is arbitrarily close to 1:

Lemma A.1: Let $X_{\rm opt}^{(k)}\sim\mathbb{N}\left(\mathsf{0}_k,\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)$, with $\frac{1}{k}{\rm Tr}\left\{\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right\}\leq P$, and assume that $\frac{1}{k^2}{\rm Tr}\left\{\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^2\right\}\xrightarrow[k\to\infty]{}0$. Then, $\forall\delta>0$, there exists

$k_\delta \in \mathbb{N}^+$ such that $\forall k \in \mathbb{N}^+$, $k > k_\delta$ it holds that $\Pr\left(\frac{1}{k}\sum_{i=0}^{k-1}\left(X_{\rm opt}[i]\right)^2 \leq P\right) > 1-\delta$.

Proof: We consider the distribution of $\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)}$. First note that $\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)$ is in general positive semidefinite. Let $\tilde{k}_{X,k}$ denote the number of zero eigenvalue of $\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)$. Then, as in (A.4), $\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)$ can be decomposed as

$$\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) = \left[\ \mathsf{P}_{X,k}^{\rm CW}(\tau_0) \quad \mathsf{P}_{X,k}^{0}(\tau_0)\ \right] \cdot \begin{bmatrix} \mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0) & \mathsf{0}_{(k-\tilde{k}_{X,k})\times\tilde{k}_{X,k}} \\ \mathsf{0}_{\tilde{k}_{X,k}\times(k-\tilde{k}_{X,k})} & \mathsf{0}_{\tilde{k}_{X,k}\times\tilde{k}_{X,k}} \end{bmatrix} \cdot \left[\ \mathsf{P}_{X,k}^{\rm CW}(\tau_0) \quad \mathsf{P}_{X,k}^{0}(\tau_0)\ \right]^T, \quad \text{(A.12)}$$

where $\left[\ \mathsf{P}_{X,k}^{\rm CW}(\tau_0) \quad \mathsf{P}_{X,k}^{0}(\tau_0)\ \right]$ is an orthogonal $k\times k$ matrix, $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0) \in \mathbb{R}^{(k-\tilde{k}_{X,k})\times(k-\tilde{k}_{X,k})}$ is a symmetric positive-definite matrix. Repeating the steps leading to the derivation of (A.6) we obtain

$$\Gamma_X^{(k)} \triangleq \left(\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0)\right)^{-\frac{1}{2}} \cdot \left(\mathsf{P}_{X,k}^{\rm CW}(\tau_0)\right)^T \cdot X_{\rm opt}^{(k)} \sim \mathbb{N}\left(\mathsf{0}_k, \mathsf{I}_{k-\tilde{k}_{X,k}}\right).$$

It follows that the RV $\Gamma_X^{(k)}$ is a vector of $k-\tilde{k}_{X,k}$ i.i.d. Gaussian RVs, $\left\{\Gamma_{X,i}\right\}_{i=0}^{k-\tilde{k}_{X,k}-1}$, each has a zero mean and unit variance. Therefore, as derived in (A.13), shown at the bottom of the next page, we can write $\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)} \overset{(dist.)}{=} \left(\Gamma_X^{(k)}\right)^T \cdot \mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0) \cdot \Gamma_X^{(k)}$. As $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0) \succ 0$, symmetric, we can write its eigenvalue decomposition as $\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0) = \left(\mathsf{P}_X(\tau_0)\right)^T \cdot \mathsf{D}_X(\tau_0) \cdot \mathsf{P}_X(\tau_0)$, where $\mathsf{P}_X(\tau_0)$ is an orthogonal matrix and $\mathsf{D}_X(\tau_0)$ a diagonal matrix with $k-\tilde{k}_{X,k}$ positive elements $\left\{d_{X,i}^{\{\tau_0\}}\right\}_{i=0}^{k-\tilde{k}_{X,k}}$. Finally we conclude that

$$\frac{1}{k}\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)} \overset{(dist.)}{=} \frac{1}{k}\sum_{i=0}^{k-\tilde{k}_{X,k}-1} d_{X,i}^{\{\tau_0\}} \cdot \Gamma_{X,i}^2,$$

where $\Gamma_{X,i}^2 \sim \chi^2(1)$, chi-square RVs, mutually independent over the index i. Then

$$\begin{aligned}
\mathbb{E}\left\{\frac{1}{k}\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)}\right\} &= \frac{1}{k}\sum_{i=0}^{k-\tilde{k}_{X,k}-1} d_{X,i}^{\{\tau_0\}} \\
&= \frac{1}{k}{\rm Tr}\left\{\mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}}(\tau_0)\right\} \\
&= \frac{1}{k}{\rm Tr}\left\{\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right\} \\
&\overset{(a)}{\leq} P \\
{\rm var}\left(\frac{1}{k}\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)}\right) &= \sum_{i=0}^{k-\tilde{k}_{X,k}-1} \frac{1}{k^2}\left(d_{X,i}^{\{\tau_0\}}\right)^2 \cdot {\rm var}\left(\Gamma_{X,i}^2\right) \\
&= \frac{2}{k^2} \cdot {\rm Tr}\left(\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^2\right),
\end{aligned}$$

where (a) follows by choice of the statistics used for generating the channel input $X_{\rm opt}^{(k)}$. As by assumption $\frac{1}{k^2}{\rm Tr}\left\{\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)\right)^2\right\} \underset{k\to\infty}{\longrightarrow} 0$, then repeating the argument in the discussion after (A.9), we can apply Chebyshev's inequality and conclude that for any arbitrary δ, taking k sufficiently large we obtain $\Pr\left(\left|\frac{1}{k}\left(X_{\rm opt}^{(k)}\right)^T \cdot X_{\rm opt}^{(k)} - (P-\delta)\right| > \delta\right) < \delta$. ■

$$\begin{aligned}
\sum_{i=0}^{2k-\approx{k}_{\epsilon,k}-1} d_{\epsilon,ii,k}^{\{\tau_0\}} &= {\rm Tr}\left\{\tilde{\tilde{\mathsf{C}}}_{\epsilon}^{(k)}(\tau_0)\right\} \\
&= {\rm Tr}\left\{\mathsf{R}_{\epsilon,k}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T \cdot \tilde{\mathsf{C}}_{\epsilon}^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0)\right\} \\
&= {\rm Tr}\left\{\tilde{\mathsf{C}}_{\epsilon}^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0) \cdot \mathsf{R}_{\epsilon,k}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T\right\} \\
&= {\rm Tr}\left\{\tilde{\mathsf{C}}_{\epsilon}^{(k)}(\tau_0) \cdot \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) \cdot \mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0) \cdot \left(\mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0)\right)^T\right\} \\
&\overset{(a)}{=} {\rm Tr}\left\{\tilde{\mathsf{C}}_{\epsilon}^{(k)}(\tau_0) \cdot \left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) \quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]\begin{bmatrix} \mathsf{C}_{\tilde{X}_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0) & \mathsf{0}_{(2k-\approx{k}_{\epsilon,k})\times\approx{k}_{\epsilon,k}} \\ \mathsf{0}_{\approx{k}_{\epsilon,k}\times(2k-\approx{k}_{\epsilon,k})} & \mathsf{0}_{\approx{k}_{\epsilon,k}\times\approx{k}_{\epsilon,k}} \end{bmatrix}\left[\ \mathsf{P}_{\epsilon,k}^{\rm CW}(\tau_0) \quad \mathsf{P}_{\epsilon,k}^{0}(\tau_0)\ \right]^T\right\} \\
&\overset{(b)}{=} {\rm Tr}\left\{\tilde{\mathsf{C}}_{\epsilon}^{(k)}(\tau_0) \cdot \mathsf{C}_{X_{\rm opt}^{(k)}W_\epsilon^{(k)}}(\tau_0)\right\} \qquad \text{(A.7a)} \\
&= -{\rm Tr}\left\{\begin{bmatrix} \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) & \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0) \\ \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) & \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)-\mathsf{I}_k \end{bmatrix}\right\} \\
&= {\rm Tr}\left\{-\left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0) - \left(\mathsf{C}_{X_{\rm opt}^{(k)}}(\tau_0)+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)+\mathsf{I}_k\right\} \qquad \text{(A.7b)} \\
&= 0 \qquad \text{(A.7c)}
\end{aligned}$$

As $X_{\text{opt}}^{(k)}$ has a specific distribution, and it satisfies the per-codeword power constraint (6) with a probability arbitrarily close to 1, as k increases, then, from the general capacity formula [19, Thm. 3.6.1] it follows that $\text{p-}\liminf_{k\to\infty} Z_{k,\epsilon}\left(F_{X_{\text{opt}}^{(k)}|\tau_0}|\tau_0\right)$ is a lower bound on capacity $C_\epsilon(\tau_0)$. Hence, we obtain the following lower bound on capacity:

$$\begin{aligned}C_\epsilon(\tau_0) &\geq \text{p-}\liminf_{k\to\infty} Z_{k,\epsilon}\left(F_{X_{\text{opt}}^{(k)}|\tau_0}|\tau_0\right)\\ &\geq \liminf_{k\to\infty}\frac{1}{k}I\big(X_{\text{opt}}^{(k)};Y_\epsilon^{(k)}|\tau_0\big).\end{aligned}\tag{A.14}$$

Next, recall that from Fano's inequality, for any $[R,k]$ code, designed for delay τ_0, having an average probability of error $P_e^k(\tau_0)\leq\rho$, $\rho\in[0,1)$, we obtain (see, e.g., [31, Thm. 3]):

$$\begin{aligned}R &\overset{(a)}{\leq} \frac{1}{1-\rho}\cdot\frac{1}{k}I\left(\bar{X}^{(k)};\bar{Y}_\epsilon^{(k)}|\tau_0\right)+\frac{h(\rho)}{k}\\ &\overset{(b)}{\leq} \frac{1}{1-\rho}\cdot\frac{1}{k}\sup_{\left\{\mathbb{E}_U\left\{\frac{1}{k}\sum_{i=0}^{k-1}\left(x_U[i]\right)^2\right\}\leq P\right\}_{k\in\mathbb{N}^+}} I\left(\bar{X}^{(k)};\bar{Y}_\epsilon^{(k)}|\tau_0\right)\\ &\quad+\frac{h(\rho)}{k}\\ &\overset{(c)}{\leq} \frac{1}{1-\rho}\cdot\frac{1}{k}\sup_{\left\{\substack{F_{\bar{X}^{(k)}|\tau_0}:\\ \mathbb{E}\left\{\frac{1}{k}\sum_{i=0}^{k-1}\left(\bar{X}[i]\right)^2\right\}\leq P}\right\}_{k\in\mathbb{N}^+}} I\left(\bar{X}^{(k)};Y_\epsilon^{(k)}|\tau_0\right)\\ &\quad+\frac{h(\rho)}{k}\end{aligned}$$

where in (a) $\bar{X}^{(k)}$ is an RV which places probability mass of $\frac{1}{|\mathcal{U}|}$ on each codeword $x_u^{(l)}$, $u\in\mathcal{U}$, and $\bar{Y}_\epsilon^{(k)}=\bar{X}^{(k)}+W_\epsilon^{(k)}$; (b) follows as when each codeword $u\in\mathcal{U}$ satisfies $\frac{1}{k}\sum_{i=0}^{k-1}\left(x_u[i]\right)^2\leq P$, then the average over all codewords in the codebook satisfies the same constraint; (c) follows as we define $\bar{X}[i]=x_U[i]$ and then relax the restrictions on the input codebook by directly maximizing over the RV $\bar{X}^{(k)}$. Hence, we obtain the upper bound on capacity as

$$C_\epsilon(\tau_0)\leq\liminf_{k\to\infty}\sup_{\left\{\substack{F_{\bar{X}^{(k)}|\tau_0}:\\ \mathbb{E}\left\{\frac{1}{k}\sum_{i=0}^{k-1}\left(\bar{X}[i]\right)^2\right\}\leq P}\right\}_{k\in\mathbb{N}^+}}\frac{1}{k}I\left(\bar{X}^{(k)};Y_\epsilon^{(k)}|\tau_0\right).\tag{A.15}$$

By [30], for every given $k\in\mathbb{N}^+$, the supremum in Eqn. (A.15) is achieved by Gaussian random vector. If the optimal distribution in (A.15) satisfies the conditions of Lemma A.1, then (A.15) is maximized by the distribution $X_{\text{opt}}^{(k)}\sim F_{X_{\text{opt}}^{(k)}|\tau_0}$, which is used in the derivation of the lower bound (A.14), i.e.,

$$C_\epsilon(\tau_0)\leq\liminf_{k\to\infty}\frac{1}{k}I\left(X_{\text{opt}}^{(k)};Y_\epsilon^{(k)}|\tau_0\right),\tag{A.16}$$

which, combined with the lower bound of (A.14), results in $C_\epsilon(\tau_0)=\liminf_{k\to\infty}\frac{1}{k}I\big(X^{(k)};Y_\epsilon^{(k)}|\tau_0\big)$.

Finally, as the sampling interval is incommensurate with T_{pw}, then for the transmission of asymptotically long sequence of *messages*, the sequence of sampling phases τ_0 is a uniformly distributed sequence over the interval $[0,T_{\text{pw}})$, [41, Example 2.1]. As transmitter's and receiver's knowledge of $\tau_0\in[0,T_{\text{pw}})$ allows both units to select the appropriate codebook with a rate of $\liminf_{k\to\infty}\frac{1}{k}I\big(X_{\text{opt}}^{(k)};Y_\epsilon^{(k)}|\tau_0\big)$, it follows that when rate adaptation is allowed, capacity can be expressed as the average rate

$$C_\epsilon=\frac{1}{T_{\text{pw}}}\int_{\tau_0=0}^{T_{\text{pw}}}C_\epsilon(\tau_0)\mathrm{d}\tau_0.$$

Appendix B
Proof of Thm. 2

A. Convergence of the Noise Correlation Matrices and Their Inverses

Define the set $\mathcal{K}\triangleq\{0,1,2,\ldots,k-1\}$, and consider the k-dimensional, zero-mean, real random vectors $W_n^{(k)}$ and $W_\epsilon^{(k)}$. Recall the definition of the correlation matrix $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ in Eqn. (A.1) and define the correlation matrix $\mathsf{C}_{W_n^{(k)}}(\tau_0)$ in a similar manner:

$$\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,v}\triangleq\mathbb{E}\big\{W_n[u]\cdot W_n[v]\big|\tau_0\big\}\equiv c_{W_n}^{\{\tau_0\}}[v,u-v],\tag{B.1}$$

for $(u,v)\in\mathcal{K}\times\mathcal{K}$. Note that since $\mathbb{E}\big\{W_n[u]\cdot W_n[v]\big|\tau_0\big\}=\mathbb{E}\big\{W_n[v]\cdot W_n[u]\big|\tau_0\big\}$, then $c_{W_n}^{\{\tau_0\}}[v,u-v]=c_{W_n}^{\{\tau_0\}}[u,v-u]$, and $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,v}=\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{v,u}$. Similarly, $c_{W_\epsilon}^{\{\tau_0\}}[v,u-v]=c_{W_\epsilon}^{\{\tau_0\}}[u,v-u]$, and $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)_{u,v}=\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)_{v,u}$.

Next, we note that by the definition of $\epsilon_n\triangleq\frac{\lfloor n\cdot\epsilon\rfloor}{n}$ it directly follows that $\frac{n\epsilon-1}{n}\leq\epsilon_n\leq\frac{n\epsilon}{n}$, hence,

$$\lim_{n\to\infty}\epsilon_n=\epsilon.\tag{B.2}$$

Define $c_{W_n}^{\{\tau_0\}}[i,\Delta]\triangleq c_{W_c}\left(i\cdot\frac{T_{\text{pw}}}{p+\epsilon_n}+\tau_0,\Delta\cdot\frac{T_{\text{pw}}}{p+\epsilon_n}\right)$. Then, by the definition of a continuous function, we obtain that

$$\begin{aligned}\left(X_{\text{opt}}^{(k)}\right)^T\cdot X_{\text{opt}}^{(k)} &\overset{(dist.)}{=}\left(X_{\text{opt}}^{(k)}\right)^T\cdot\left[\begin{array}{cc}\mathsf{P}_{X,k}^{\text{CW}}(\tau_0) & \mathsf{P}_{X,k}^{0}(\tau_0)\end{array}\right]\cdot\left[\begin{array}{cc}\mathsf{P}_{X,k}^{\text{CW}}(\tau_0) & \mathsf{P}_{X,k}^{0}(\tau_0)\end{array}\right]^T\cdot X_{\text{opt}}^{(k)}\\ &\overset{(dist.)}{=}\left(X_{\text{opt}}^{(k)}\right)^T\cdot\left(\mathsf{P}_{X,k}^{\text{CW}}(\tau_0)\right)\cdot\left(\mathsf{P}_{X,k}^{\text{CW}}(\tau_0)\right)^T\cdot X_{\text{opt}}^{(k)}\\ &\overset{(dist.)}{=}\left(X_{\text{opt}}^{(k)}\right)^T\cdot\left(\mathsf{P}_{X,k}^{\text{CW}}(\tau_0)\right)\cdot\left(\mathsf{C}_{\tilde{X}_{\text{opt}}^{(k)}}(\tau_0)\right)^{-\frac{1}{2}}\cdot\mathsf{C}_{\tilde{X}_{\text{opt}}^{(k)}}(\tau_0)\cdot\left(\mathsf{C}_{\tilde{X}_{\text{opt}}^{(k)}}(\tau_0)\right)^{-\frac{1}{2}}\cdot\left(\mathsf{P}_{X,k}^{\text{CW}}(\tau_0)\right)^T\cdot X_{\text{opt}}^{(k)}\\ &\overset{(dist.)}{=}\left(\Gamma_X^{(k)}\right)^T\cdot\mathsf{C}_{\tilde{X}_{\text{opt}}^{(k)}}(\tau_0)\cdot\Gamma_X^{(k)}\end{aligned}\tag{A.13}$$

continuity of $c_{W_c}(t,\lambda)$ in t and in λ, combined with (B.2), implies that $\forall i,\Delta\in\mathbb{Z}$,

$$\lim_{n\to\infty} c_{W_n}^{\{\tau_0\}}[i,\Delta] = \lim_{n\to\infty} c_{W_c}\left(i\cdot\frac{T_{\rm pw}}{p+\epsilon_n}+\tau_0,\Delta\cdot\frac{T_{\rm pw}}{p+\epsilon_n}\right) = c_{W_c}\left(i\cdot\frac{T_{\rm pw}}{p+\epsilon}+\tau_0,\Delta\cdot\frac{T_{\rm pw}}{p+\epsilon}\right)\equiv c_{W_\epsilon}^{\{\tau_0\}}[i,\Delta]. \quad (B.3)$$

Recall that as the autocorrelation function $c_{W_c}(t,\lambda)$ is bounded and continuous in $(t,\lambda)\in\mathbb{R}^2$, periodic in $t\in\mathbb{R}$, and is zero $\forall|\lambda|\geq\lambda_m$, then it is bounded and uniformly continuous with respect to time $t\in[0,T_{\rm pw}]$ and lag $\lambda\in\mathbb{R}$ [23, Ch. III, Thm. 3.13]. Also recall that the correlation matrices $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ and $\mathsf{C}_{W_n^{(k)}}(\tau_0)$ are non-singular (the rationale for this assumption is given in Comment A.1). Combining these properties with the definitions of the correlation matrices $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ and $\mathsf{C}_{W_n^{(k)}}(\tau_0)$ in (A.1) and (B.1), respectively, and with the limit in (B.3) we obtain that

$$\lim_{n\to\infty}\max_{\substack{\tau_0\in[0,T_{\rm pw}],\\(u,v)\in\mathcal{K}\times\mathcal{K}}}\left\{\left|\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)_{u,v}-\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)_{u,v}\right|\right\}=0. \quad (B.4)$$

Next, consider the mapping $m_k:\mathbb{R}^{k^2}\mapsto\mathbb{R}^{k^2}$, defined via

$$m_k\big(\mathsf{C}^{(k)}\big)=\big(\mathsf{C}^{(k)}\big)^{-1},\qquad \mathsf{C}^{(k)}\in\mathbb{R}^{k^2}.$$

This is a continuous mapping over the set of positive-definite $k\times k$ matrices $\mathsf{C}^{(k)}\succ 0$, see, e.g., [42, Eqns. (1.5)-(1.6)]. Consider now the positive-definite matrix $\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}$. By the strict diagonal dominance of $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ (see condition (12)) we obtain that

$$\begin{aligned}
&\max\mathrm{Eig}\left\{\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\} \qquad (B.5)\\
&\overset{(a)}{\leq}\left\|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\|_1\\
&\overset{(b)}{=}\left\|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\|_\infty\\
&\overset{(c)}{\leq}\left(\min_{0\leq u\leq k-1}\left\{\left|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)_{u,u}\right|-\sum_{v=0,v\neq u}^{k-1}\left|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)_{u,v}\right|\right\}\right)^{-1}\\
&\overset{(d)}{\leq}\left(\min_{0\leq t\leq T_{\rm pw}}\left\{c_{W_c}(t,0)-2\tau_{\rm m}\cdot\max_{|\lambda|>\frac{T_{\rm pw}}{p+1}}\left\{|c_{W_c}(t,\lambda)|\right\}\right\}\right)^{-1}, \qquad (B.6)
\end{aligned}$$

where (a) follows from the upper bound of [43, Thm. 5]; (b) follows from the symmetry of $\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}$, due to which we can obtain explicitly

$$\begin{aligned}
\left\|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\|_1 &\triangleq \max_{0\leq v\leq k-1}\left\{\sum_{u=0}^{k-1}\left|\left(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right)_{u,v}\right|\right\}\\
&=\max_{0\leq v\leq k-1}\left\{\sum_{u=0}^{k-1}\left|\left(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right)_{v,u}\right|\right\}\\
&=\max_{0\leq u\leq k-1}\left\{\sum_{v=0}^{k-1}\left|\left(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right)_{u,v}\right|\right\}\\
&=\left\|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\|_\infty;
\end{aligned}$$

(c) follows from the bound in [34, Eqn. (4)] (see also [44, Eqn. (3)]), as, by condition (12), the matrix $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ is SDD, see Comment 4; and lastly, (d) follows from Eqn. (15). Evidently, the same bound applies to $\max\mathrm{Eig}\left\{\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}\right\}$ and we note that it is independent of τ_0. Then, since the magnitudes of the elements of a positive-definite real, symmetric matrix are upper-bounded by its largest eigenvalue[3] we conclude that the elements of the matrices $\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}$ and $\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}$ belong to a finite interval whose end points are independent of k, τ_0 and n.

Next, note that boundedness of $c_{W_c}(t,\lambda)$ (see Section III-A) implies that the elements of $\mathsf{C}_{W_n^{(k)}}(\tau_0)$ and of $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$ are all bounded. It therefore follows from (B.4), the boundedness of the elements of $\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)$, $\mathsf{C}_{W_n^{(k)}}(\tau_0)$, $\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}$ and $\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}$, boundedness and uniform continuity of the CT correlation function $c_{W_c}(t,\lambda)$ in t and in λ, continuity of the mapping $m_k:\mathbb{R}^{k^2}\mapsto\mathbb{R}^{k^2}$,[4] and from [46, Thm. 11.2.3] that

$$\lim_{n\to\infty}\max_{\substack{\tau_0\in[0,T_{\rm pw}],\\(u,v)\in\mathcal{K}\times\mathcal{K}}}\left\{\left|\left(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}\right)_{u,v}-\left(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right)_{u,v}\right|\right\}=0. \quad (B.7)$$

B. Showing That $\lim_{n\to\infty}\frac{1}{k}I(X_{n,\rm opt}^{(k)};Y_n^{(k)}|\tau_{n,k}^{\rm opt})=\frac{1}{k}I(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\rm opt})$ for the Optimal Sampling Phases and Inputs Distributions $\big(\tau_{n,k}^{\rm opt},F_{X_{n,\rm opt}^{(k)}|\tau_{n,k}^{\rm opt}}\big)$ and $\big(\tau_{\epsilon,k}^{\rm opt},F_{X_{\rm opt}^{(k)}|\tau_{\epsilon,k}^{\rm opt}}\big)$

Consider a fixed $k\in\mathbb{N}^+$, let

$$\mathcal{C}_{X^{(k)}}\triangleq\left\{\tau\in[0,T_{\rm pw}),\mathsf{C}_{X^{(k)}}\in\mathbb{R}^{k\times k}\,\middle|\,\sum_{i=0}^{k-1}\big(\mathsf{C}_{X^{(k)}}\big)_{ii}\leq k\cdot P,\ \mathsf{C}_{X^{(k)}}=\big(\mathsf{C}_{X^{(k)}}\big)^T,\mathsf{C}_{X^{(k)}}\succcurlyeq 0\right\}, \quad (B.8)$$

[3]for a positive-definite real, symmetric matrix C, we have

$$\begin{aligned}
0&<(\mathbf{e}_i-\mathbf{e}_j)^T\cdot\mathsf{C}\cdot(\mathbf{e}_i-\mathbf{e}_j)\\
&=\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j-\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j-\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_i\\
\Rightarrow&\frac{1}{2}\cdot(\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j)>\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j\\
0&<(\mathbf{e}_i+\mathbf{e}_j)^T\cdot\mathsf{C}\cdot(\mathbf{e}_i+\mathbf{e}_j)\\
&=\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j+\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_i\\
\Rightarrow&-\frac{1}{2}\cdot(\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j)<\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j
\end{aligned}$$

Hence, $0\leq\left|\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j\right|<\frac{1}{2}\cdot\big(\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j\big)$. Now, letting $\mathbf{e}_i$ denote the all-zero vector except for 1 at the i-th element ($i\in\mathbb{N}$), we note that the (i,j)-th element of C is given by

$$\left|(\mathsf{C})_{i,j}\right|=\left|\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_j\right|<\frac{1}{2}\cdot\big(\mathbf{e}_i^T\cdot\mathsf{C}\cdot\mathbf{e}_i+\mathbf{e}_j^T\cdot\mathsf{C}\cdot\mathbf{e}_j\big)\leq\max\mathrm{Eig}\left\{\mathsf{C}\right\},$$

where the last inequality follows from [45, Thm. 4.2.2].

[4]Since the inverse is a continuous mapping from a compact and finite set to a compact and finite set.

and denote

$$(\tau_{n,k}^{\rm opt}, \mathsf{C}_{X_n^{(k)}}^{\rm opt}) = \mathop{\rm argmax}_{(\tau_0, \mathsf{C}_{X^{(k)}})\in\mathcal{C}_{X^{(k)}}} \frac{1}{2k}\log\left(\frac{{\rm Det}\Big(\mathsf{C}_{X^{(k)}}+\mathsf{C}_{W_n^{(k)}}(\tau_0)\Big)}{{\rm Det}\Big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\Big)}\right) \tag{B.9a}$$

$$(\tau_{\epsilon,k}^{\rm opt}, \mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt}) = \mathop{\rm argmax}_{(\tau_0, \mathsf{C}_{X^{(k)}})\in\mathcal{C}_{X^{(k)}}} \frac{1}{2k}\log\left(\frac{{\rm Det}\Big(\mathsf{C}_{X^{(k)}}+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\Big)}{{\rm Det}\Big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\Big)}\right). \tag{B.9b}$$

Then, the zero-mean Gaussian random vectors $X_{n,\rm opt}^{(k)}$ and $X_{\rm opt}^{(k)}$, with covariance matrices $\mathsf{C}_{X_n^{(k)}}^{\rm opt}$ and $\mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt}$, respectively, at the respective sampling phases $\tau_{n,k}^{\rm opt}$ and $\tau_{\epsilon,k}^{\rm opt}$, maximize the mutual information expressions $\frac{1}{k}I(X_n^{(k)};Y_n^{(k)}|\tau_0)$ and $\frac{1}{k}I(X^{(k)};Y_\epsilon^{(k)}|\tau_0)$, respectively, when maximization is over all sampling phases and associated input distributions which satisfy the respective trace constraint, $\frac{1}{k}{\rm Tr}\big\{\mathsf{C}_{X_n^{(k)}}\big\}\le P$, $\frac{1}{k}{\rm Tr}\big\{\mathsf{C}_{X^{(k)}}\big\}\le P$, see, e.g., [30, Eqn. (6)]. We now have the following lemma:

Lemma B.1: When $k\in\mathbb{N}^+$ is fixed, and $X_{n,\rm opt}^{(k)}$ and $X_{\rm opt}^{(k)}$ are zero-mean Gaussian random vectors with covariance matrices $\mathsf{C}_{X_n^{(k)}}^{\rm opt}$ and $\mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt}$, respectively, where $(\tau_{n,k}^{\rm opt}, \mathsf{C}_{X_n^{(k)}}^{\rm opt})$ and $(\tau_{\epsilon,k}^{\rm opt}, \mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt})$ satisfy (B.9), then

$$\lim_{n\to\infty}\frac{1}{k}I(X_{n,\rm opt}^{(k)};Y_n^{(k)}|\tau_{n,k}^{\rm opt}) = \frac{1}{k}I(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\rm opt}). \tag{B.10}$$

Proof: First, recall that by [30], when the covariance matrix and sampling phase pairs are given as $(\tau_{n,k}^{\rm opt}, \mathsf{C}_{X_n^{(k)}}^{\rm opt})$ and $(\tau_{\epsilon,k}^{\rm opt}, \mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt})$ satisfying Eqns. (B.9), then the mutual information expressions in (B.10), evaluated for $X_{n,\rm opt}^{(k)}$ and $X_{\rm opt}^{(k)}$ Gaussian inputs processes corresponding to the sampling phase-correlation matrix pairs of Eqns. (B.9), are equal to the maximal values of the objective functions in (B.9):

$$\frac{1}{k}I(X_{n,\rm opt}^{(k)};Y_n^{(k)}|\tau_{n,k}^{\rm opt}) = \frac{1}{2k}\log\left(\frac{{\rm Det}\Big(\mathsf{C}_{X_n^{(k)}}^{\rm opt}+\mathsf{C}_{W_n^{(k)}}(\tau_{n,k}^{\rm opt})\Big)}{{\rm Det}\Big(\mathsf{C}_{W_n^{(k)}}(\tau_{n,k}^{\rm opt})\Big)}\right)$$

$$\frac{1}{k}I(X_{\rm opt}^{(k)};Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\rm opt}) = \frac{1}{2k}\log\left(\frac{{\rm Det}\Big(\mathsf{C}_{X_\epsilon^{(k)}}^{\rm opt}+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_{\epsilon,k}^{\rm opt})\Big)}{{\rm Det}\Big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_{\epsilon,k}^{\rm opt})\Big)}\right).$$

Therefore, convergence of the limit in (B.10) corresponds to having that the optimal values of the objective function in the optimization problem (B.9a) converge, as $n\to\infty$, to the optimal value of the objective function in (B.9b). To prove this convergence we employ [47, Thm. 2.1][5]. The main requirement for the application of [47, Thm. 2.1] is that $\lim_{n\to\infty}\frac{1}{2k}\log\Big({\rm Det}\Big(\mathsf{C}_{X^{(k)}}+\mathsf{C}_{W_n^{(k)}}(\tau_0)\Big)/{\rm Det}\Big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\Big)\Big) = \frac{1}{2k}\log\Big({\rm Det}\Big(\mathsf{C}_{X^{(k)}}+\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\Big)/{\rm Det}\Big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\Big)\Big)$ *uniformly* over $\mathcal{C}_{X^{(k)}}$. In the following we show that such a uniform convergence holds.

It follows from the limit in (B.7) that for any $\delta_1>0$ there exists $n_0(\delta_1)\in\mathbb{N}^+$ sufficiently large such that for all $0\le l,q\le k-1$, $\tau_0\in[0,T_{\rm pw})$, and for all $n>n_0(\delta_1)$, it holds that $\left|\Big(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}-\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\Big)_{l,q}\right|\le\delta_1$. Now, we note that since $\mathsf{C}_{X^{(k)}}$ is a positive semi-definite matrix which satisfies a constraint on the sum of its diagonal elements (B.8), then from the Cauchy-Schwartz inequality [40, Eqn. (9-176)], [33, Sec. 3.6] we have that

$$\begin{aligned}\left|\big(\mathsf{C}_{X^{(k)}}\big)_{l,q}\right| &= \left|\mathbb{E}\{X[l]X[q]\}\right|\\ &\le\sqrt{\mathbb{E}\left\{(X[l])^2\right\}\mathbb{E}\left\{(X[q])^2\right\}}\\ &\le\sqrt{(k\cdot P)^2}=k\cdot P,\qquad\forall\, 0\le l,q\le k-1.\end{aligned} \tag{B.11}$$

It thus follows that all the matrices $\mathsf{C}_{X^{(k)}}\in\mathcal{C}_{X^{(k)}}$ have bounded elements. We now bound $\left|\Big(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}-\Big(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}\right|$ for all $n>n_0(\delta_1)$ as follows:

$$\begin{aligned}&\left|\Big(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}-\Big(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}\right|\\ &=\left|\sum_{m=0}^{k-1}\Big(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}-\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\Big)_{l,m}\big(\mathsf{C}_{X^{(k)}}\big)_{m,q}\right|\\ &\le\delta_1\cdot\sum_{m=0}^{k-1}\left|\big(\mathsf{C}_{X^{(k)}}\big)_{m,q}\right|\le\delta_1\cdot P\cdot k^2,\end{aligned}$$

where we recall that P and k are finite and given. Thus, for any $\delta>0,\exists n_0(\delta)\in\mathbb{N}^+$ sufficiently large such that for all $0\le l,q\le k-1$, $n>n_0(\delta)$, and for all $(\tau_0,\mathsf{C}_{X^{(k)}})\in\mathcal{C}_{X^{(k)}}$

$$\left|\Big(\big(\mathsf{C}_{W_n^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}-\Big(\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big)_{l,q}\right|\le\delta. \tag{B.12}$$

Observe that the product $\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}$ has non-negative eigenvalues, [48, Thm. 7.5], and the maximal eigenvalue, denoted ${\rm maxEig}\Big\{\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\Big\}$, can be upper bounded as follows:

$$\begin{aligned}&{\rm maxEig}\left\{\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\mathsf{C}_{X^{(k)}}\right\}\\ &\stackrel{(a)}{\le}{\rm maxEig}\left\{\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\}\cdot{\rm maxEig}\left\{\mathsf{C}_{X^{(k)}}\right\}\\ &\stackrel{(b)}{\le}\left\|\big(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\big)^{-1}\right\|_1\cdot\left\|\mathsf{C}_{X^{(k)}}\right\|_1\\ &\stackrel{(c)}{\le}\underbrace{\frac{k^2\cdot P}{\min\limits_{0\le t\le T_{\rm pw}}\left\{c_{W_{\rm c}}(t,0)-2\tau_{\rm m}\cdot\max\limits_{|\lambda|>\frac{T_{\rm pw}}{p+1}}\left\{|c_{W_{\rm c}}(t,\lambda)|\right\}\right\}}}_{\triangleq\beta_0(k)},\end{aligned} \tag{B.13}$$

[5] [47, Thm. 2.1]: Let $f_n\to f$ uniformly as $n\to\infty$. Then, the sequence of problems $P_n: a_n=\inf\limits_{x\in\mathcal{X}}f_n(x)$ converges to the problem $P: a=\inf\limits_{x\in\mathcal{X}}f(x)$ as $n\to\infty$.
See additional conditions regarding the application of [47, Thm. 2.1] in Footnotes 7 and 8.

where (a) follows from [49, Eqn. (9)], see also [48, Thm. 8.12]; in (b) we use the upper bound from [43, Thm. 5]; and step (c) follows from the bounds in (B.6) and (B.11).

Since the magnitudes of the elements of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}$ are upper bounded by (B.6) (see Footnote 3) and the magnitudes of the elements of $\mathsf{C}_{X^{(k)}}$ are upper bounded by $k \cdot P$, we obtain that the magnitudes of the elements of the matrix product $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ are upper bounded by $k^2 \cdot \left(\min_{0\le t\le T_{\rm pw}}\left\{c_{W_{\rm c}}(t,0)-2\tau_{\rm m}\cdot\max_{|\lambda|>\frac{T_{\rm pw}}{p+1}}\left\{|c_{W_{\rm c}}(t,\lambda)|\right\}\right\}\right)^{-1}\cdot P \equiv \beta_0(k)$. As the upper bound $\beta_0(k)$ is finite and independent of ϵ, then the same upper bound applies also to the magnitudes of the elements of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$, and we conclude that the magnitudes of the elements of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ and of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ are all finite and upper bounded by a bound which increases as k^2, and is independent of n and τ_0.

Consider next the ordered sets of eigenvalues of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ and of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ arranged in descending order: Let $\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) \equiv \Lambda_i^{(k)}\left\{\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}\right\}$ and $\Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) \equiv \Lambda_i^{(k)}\left\{\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}\right\}$, $0 \le i \le k-1$. Then, continuity of the eigenvalues of square real matrices [45, Sec. 2.4.9, Thm. 2.4.9.2], combined with the boundedness of the elements of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ and of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$, boundedness of their corresponding eigenvalues, and the fact that convergence of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ to $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ is uniform over $\mathcal{C}_{X^{(k)}}$, see Eqn. (B.12), imply that the ordered sets of eigenvalues of $\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ converge to the ordered set of eigenvalues of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$ uniformly in $(\tau_0,\mathsf{C}_{X^{(k)}}) \in \mathcal{C}_{X^{(k)}}$, namely, $\forall\delta > 0$, $\exists\tilde{n}_0(\delta) \in \mathbb{N}^+$ sufficiently large such that for all $n > \tilde{n}_0(\delta)$, and for all $(\tau_0,\mathsf{C}_{X^{(k)}}) \in \mathcal{C}_{X^{(k)}}$

$$\left|\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) - \Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right| \le \delta, \quad 0 \le i \le k-1. \tag{B.14}$$

Lastly, consider the distance between the objective functions in (B.9): For any $(\tau_0,\mathsf{C}_{X^{(k)}}) \in \mathcal{C}_{X^{(k)}}$, the distance between the objective functions can now be expressed as[6]:

$$\begin{aligned}
&\left|\frac{1}{2k}\,\mathrm{logdet}\left(\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right) - \frac{1}{2k}\,\mathrm{logdet}\left(\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right)\right|\\
&= \frac{1}{2k}\left|\sum_{i=0}^{k-1}\log\left(1+\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right) - \sum_{i=0}^{k-1}\log\left(1+\Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\right|.
\end{aligned}$$

[6]Note that since $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1} \succ 0$, symmetric, then $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\cdot \mathsf{C}_{X^{(k)}} + \mathsf{I}_k = \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-\frac{1}{2}} \cdot \left(\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-\frac{1}{2}}\mathsf{C}_{X^{(k)}} \cdot \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-\frac{1}{2}} + \mathsf{I}_k\right) \cdot \left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{\frac{1}{2}}$.

Using the first order Taylor expansion [50, Pgs. 415-418] we can write

$$\begin{aligned}
&\log\left(1+\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\\
&= \log\left(1+\Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\\
&\quad + \frac{1}{\ln 2}\frac{1}{1+\Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)}\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) - \Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\\
&\quad + \xi_i\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right),
\end{aligned}$$

where $\xi_i\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)$ is a reminder term. Thus, $\forall n > \tilde{n}_0(\delta)$ such that (B.14) is satisfied, we obtain $\forall(\tau_0,\mathsf{C}_{X^{(k)}}) \in \mathcal{C}_{X^{(k)}}$ that

$$\begin{aligned}
&\frac{1}{2k}\left|\log\det\left(\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right) - \mathrm{logdet}\left(\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right)\right|\\
&= \frac{1}{2k}\left|\sum_{i=0}^{k-1}\left(\frac{1}{\ln 2}\frac{1}{1+\Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)}\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) - \Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right) + \xi_i\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\right)\right|\\
&\le 2\frac{1}{2k}\sum_{i=0}^{k-1}\left|\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) - \Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right| + \frac{1}{2k}\sum_{i=0}^{k-1}\left|\xi_i\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\right|\\
&\le \delta + \frac{1}{2}\max_{0\le i\le k-1}\left|\xi_i\left(\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right)\right|\\
&\overset{(a)}{\le} \delta + \frac{1}{2}\delta^2.
\end{aligned}$$

Note that uniformity in $\mathcal{C}_{X^{(k)}}$ of the inequality in step (a) follows from the uniform boundedness of $\left|\Lambda_{i,n}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0) - \Lambda_{i,\epsilon}^{(k)}(\mathsf{C}_{X^{(k)}};\tau_0)\right|$ over $\mathcal{C}_{X^{(k)}}$, as established in (B.14), combined with the bound in (B.15), as shown at the bottom of the next page, derived using [50, Pg. 418], which holds uniformly over $\mathcal{C}_{X^{(k)}}$. Note that step (a′) in the derivation of (B.15) follows since $\zeta \ge 0$, by the non-negativity of the eigenvalues of $\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}$.

It follows that the distance between the logdet functions is uniformly upper bounded for all $(\tau_0,\mathsf{C}_{X^{(k)}}) \in \mathcal{C}_{X^{(k)}}$, hence, convergence of $\frac{1}{2k}\,\mathrm{logdet}\left(\left(\mathsf{C}_{W_n^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right)$ to $\frac{1}{2k}\,\mathrm{logdet}\left(\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_0)\right)^{-1}\mathsf{C}_{X^{(k)}}+\mathsf{I}_k\right)$ as $n \to \infty$ is uniform over the feasible set $\mathcal{C}_{X^{(k)}}$. We conclude that for a sequence of optimization problems (B.9a), the objective functions converge uniformly to a limiting objective function

(B.9b). Thus, it follows from the proof of [47, Thm. 2.1][7,8] that the sequence of optimal objective values of the sequence of problems (B.9a) converges to the optimal objective value of the limiting problem (B.9b):

$$\lim_{n\to\infty} \frac{1}{2k} \operatorname{logdet}\left(\left(\mathsf{C}_{W_n^{(k)}}(\tau_{n,k}^{\text{opt}})\right)^{-1} \mathsf{C}^{\text{opt}}_{X_n^{(k)}} + \mathsf{I}_k\right) = \frac{1}{2k} \operatorname{logdet}\left(\left(\mathsf{C}_{W_\epsilon^{(k)}}(\tau_{\epsilon,k}^{\text{opt}})\right)^{-1} \mathsf{C}^{\text{opt}}_{X^{(k)}} + \mathsf{I}_k\right).$$

■

C. Equivalence Between and C_ϵ and $\liminf_{n\to\infty} C_n$ for the Setup of Theorem 2

Let $F_{X_{\text{opt}}^{(k)}|\tau_{\epsilon,k}^{\text{opt}}}$ denote a k-dimensional Gaussian CDF with a correlation matrix denoted by $\mathsf{C}^{\text{opt}}_{X^{(k)}}$, such that $\left(\tau_{\epsilon,k}^{\text{opt}}, \mathsf{C}^{\text{opt}}_{X^{(k)}}\right)$ maximizes $\frac{1}{k} I(X^{(k)}; Y_\epsilon^{(k)}|\tau_{\epsilon,k})$. Let $Z_{k,\epsilon}\left(F_{X_{\text{opt}}^{(k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\right)$ denote the corresponding mutual information density rate. We now have the following Lemma:

Lemma B.2: For the setup of Theorem 2 it holds that

$$C_\epsilon = \text{p-}\liminf_{k\to\infty} Z_{k,\epsilon}\left(F_{X_{\text{opt}}^{(k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\right) = \liminf_{n\to\infty} C_n, \tag{B.16}$$

where C_n is defined in Eqn. (11).

Proof: First, consider the upper bound on capacity: Recall that by Lemma B.1, for every finite $k \in \mathbb{N}^+$, letting $\tau_{n,k}^{\text{opt}}$ and $\tau_{\epsilon,k}^{\text{opt}}$ denote the optimal sampling phases within the noise period, and letting $X_{n,\text{opt}}^{(k)}$ and $X_{\text{opt}}^{(k)}$ denote the corresponding Gaussian inputs with the optimal correlation matrices, it holds that $\lim_{n\to\infty} \frac{1}{k} I(X_{n,\text{opt}}^{(k)}; Y_n^{(k)}|\tau_{n,k}^{\text{opt}}) = \frac{1}{k} I(X_{\text{opt}}^{(k)}; Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\text{opt}})$. Then, we can write

$$\begin{aligned}
C_\epsilon &\overset{(a)}{\leq} \liminf_{k\to\infty} \frac{1}{k} I\left(X_{\text{opt}}^{(k)}; Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\text{opt}}\right) \\
&\overset{(b)}{=} \liminf_{k\to\infty} \lim_{n\to\infty} \frac{1}{k} I\left(X_{n,\text{opt}}^{(k)}; Y_n^{(k)}|\tau_{n,k}^{\text{opt}}\right) \\
&\overset{(c)}{=} \liminf_{k\to\infty} \liminf_{n\to\infty} \frac{1}{k} I\left(X_{n,\text{opt}}^{(k)}; Y_n^{(k)}|\tau_{n,k}^{\text{opt}}\right) \\
&\overset{(d)}{\leq} \liminf_{k\to\infty} \liminf_{n\to\infty} C_n(\tau_{n,k}^{\text{opt}}) \\
&\overset{(e)}{\leq} \liminf_{k\to\infty} \liminf_{n\to\infty} C_n \\
&= \liminf_{n\to\infty} C_n,
\end{aligned} \tag{B.17}$$

where (a) follows from similar arguments as in the derivation of (A.15):

$$\begin{aligned}
C_\epsilon &\leq \liminf_{k\to\infty} \left\{ \sup_{\substack{\left(F_{\bar{X}^{(k)}|\bar{\tau}_{\epsilon,k}}, \bar{\tau}_{\epsilon,k}\right): \\ \frac{1}{k}\sum_{i=0}^{k-1}(\bar{x}_u[i])^2 \leq P,\ u\in\mathcal{U}, \\ \bar{\tau}_{\epsilon,k}\in[0,T_{\text{pw}}]}} \frac{1}{k} I\left(\bar{X}^{(k)}; \bar{Y}_\epsilon^{(k)}|\bar{\tau}_{\epsilon,k}\right)\right\}_{k\in\mathbb{N}^+} \\
&\leq \liminf_{k\to\infty} \left\{ \sup_{\substack{\left(F_{\bar{X}^{(k)}|\bar{\tau}_{\epsilon,k}}, \bar{\tau}_{\epsilon,k}\right): \\ \mathbb{E}\left\{\frac{1}{k}\sum_{i=0}^{k-1}(\bar{X}[i])^2\right\} \leq P, \\ \bar{\tau}_{\epsilon,k}\in[0,T_{\text{pw}}]}} \frac{1}{k} I\left(\bar{X}^{(k)}; \bar{Y}_\epsilon^{(k)}|\bar{\tau}_{\epsilon,k}\right)\right\}_{k\in\mathbb{N}^+},
\end{aligned}$$

where $\bar{Y}_\epsilon^{(k)}$ is the channel output for input $\bar{X}^{(k)}$, and $\bar{\tau}_{\epsilon,k}$ is the sampling phase, and in the first inequality we maximize over all input distributions which facilitate selection of a codebook which satisfies the per-codeword power constraint. As the channel (5) is an additive Gaussian noise channel, then, subject to a trace constraint on the input correlation matrix, it follows that for every given $k \in \mathbb{N}^+$, the mutual information expression between the channel input and its output is maximized by Gaussian inputs [30, Eqns. (4), (30)], which satisfy the trace constraint, and we use the maximizing sampling phase according to (B.9b). Step (b) follows from Lemma B.1; and step (c) follows since the limit in n exists and is finite [46, Thm. 33.1.1]. Step (d) follows as the channel (8) is a DT ACGN channel, then, by the discussion in [6, Eqns. (14)-(16)] it is equivalent to a finite-memory stationary multivariate Gaussian channel. Hence, Step (d) follows directly from the converse in [52, Eqn. (21)]: For every finite $k \in \mathbb{N}^+$ and a given $\tau_{n,k}^{\text{opt}}$, the achievable rate is not greater than the capacity: $C_n\left(\tau_{n,k}^{\text{opt}}\right) \geq \frac{1}{k} I\left(X_n^{(k)}; Y_n^{(k)}|\tau_{n,k}\right)$. Lastly, step (e) follows as by (11), $C_n \geq C_n\left(\tau_{n,k}^{\text{opt}}\right)$.

To show achievability of $\liminf_{n\to\infty} C_n$, we consider a transmission scheme which partitions the transmitted sequence into finite-length blocks, all of length $k \in \mathbb{N}^+$, and appends each

[7]For the application of [47, Thm. 2.1]: X in the theorem corresponds to the union of the interval $[0, T_{\text{pw}})$ and the space of real symmetric, positive semidefinite matrices, subject to a constraint on their trace: $\left\{\mathsf{C}_{X^{(k)}} \in \mathbb{R}^{k\times k} \middle| \frac{1}{k}\sum_{i=0}^{k-1}\left(\mathsf{C}_{X^{(k)}}\right)_{ii} \leq P,\ \mathsf{C}_{X^{(k)}} = \left(\mathsf{C}_{X^{(k)}}\right)^T, \mathsf{C}_{X^{(k)}} \succcurlyeq 0\right\}$, which is a convex space. Y in the theorem corresponds to the set of real numbers and the positive cone corresponds to the set of non-negative real numbers, thus, this cone is clearly normal [51, Example 6.3.5].

[8]Note that while [47, Thm. 2.1] is stated for convex objectives, convexity of the objective is not required for the convergence of the optimal objective values, only for the convergence of the optimal solutions. As we are not interested in the convergence of the optimal solutions (i.e., not interested in the convergence of $\left(\tau_{n,k}^{\text{opt}}, \mathsf{C}^{\text{opt}}_{X_n^{(k)}}\right)$), then we can apply the steps in proof of [47, Thm. 2.1] to conclude that the optimal objective *values* converge also for non-convex objectives, without considering convergence of the solutions.

$$\begin{aligned}
&\left|\xi_i\left(\Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right)\right| = \left|\frac{1}{2}\cdot\frac{1}{\ln 2}\cdot\frac{1}{(1+\zeta)^2}\left(\Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right) - \Lambda_{i,\epsilon}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right)^2\right|, \\
&\min\left\{\Lambda_{i,\epsilon}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right), \Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right\} \leq \zeta \leq \max\left\{\Lambda_{i,\epsilon}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right), \Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right\} \\
&\overset{(a')}{\Rightarrow} \left|\xi_i\left(\Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right)\right| \leq \frac{1}{2}\cdot\frac{1}{\ln 2}\left(\Lambda_{i,n}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right) - \Lambda_{i,\epsilon}^{(k)}\left(\mathsf{C}_{X^{(k)}};\tau_0\right)\right)^2 \leq \delta^2
\end{aligned} \tag{B.15}$$

k-block with a guard interval, sufficient to facilitate statistical independence between the noise process samples belonging to the different k-blocks as well as to facilitate synchronization of the transmission start times of all k-blocks to begin at the initial sampling phase $\tau_{\epsilon,k}^{\text{opt}}$, which is selected to maximize the mutual information of the k-block subject to a trace constraint, as in (B.9b). Note that such synchronization is permissible as the setup of Thm. 2 allows for transmission delay. We now detail the operations at the transmitter and at the receiver, recalling that $W_\epsilon[i]$ has a *finite memory* $\tau_{\rm m} < \infty$. In the following description we use the tilde symbol to denote the channel input and output with the additional guard intervals, as well as the appropriate noise sequence. The channel inputs which carry information and the processed outputs used for decoding are denoted without the tilde symbol, and so does the corresponding noise sequence.

1) Transmitter's Operations: Let $k \in \mathbb{N}^+$ denote the blocklength, and set the duration of the guard interval between subsequent blocks to $\tau_{\rm m} \cdot T_{\rm s}(\epsilon) + \Delta_g$ time units (in CT), $0 \leq \Delta_g \leq T_{\rm pw}$ will be explicitly defined later in the proof. Due to the finite memory and Gaussianity of the noise process it follows that noise sequences belonging to different k-blocks are statistically independent. Moreover, the additional guard time of Δ_g beyond $\tau_{\rm m}$ allows the transmitter to synchronize the sampling phase of the i-th k-block, denoted $\tau_{\epsilon,k}^{(i)}$, to the sampling phase which maximizes the k-block mutual information, i.e., $\tau_{\epsilon,k}^{(i)} = \tau_{\epsilon,k}^{\text{opt}}$. The codewords are generated according to a Gaussian distribution $F_{X_{\text{opt}}^{(k)}|\tau_{\epsilon,k}^{\text{opt}}}$ with a correlation matrix $\mathsf{C}_{X^{(k)}}^{\text{opt}}$, s.t. $\left(\tau_{\epsilon,k}^{\text{opt}}, \mathsf{C}_{X^{(k)}}^{\text{opt}}\right)$ are selected according to (B.9b). Hence, a codeword consisting of $l \cdot k$ code symbols is transmitted over a time interval corresponding to $l \cdot (k + \tau_{\rm m} + \Delta_g/T_{\rm s}(\epsilon))$ channel symbols. Note that as the transmitter knows the correlation function of the noise, and, naturally, knows its own symbol interval $T_{\rm s}(\epsilon)$, it can deterministically compute the delay Δ_g needed to arrive again at $\tau_{\epsilon,k}^{\text{opt}}$, which is given by (B.18), as shown at the bottom of the next page. A codebook $\mathcal{CB}_k^{(i)}(\tau_{\epsilon,k}^{\text{opt}})$ of rate R for the i-th k-block is generated by selecting 2^{kR} codewords randomly and independently according to $F_{X_{\text{opt}}^{(k)}|\tau_{\epsilon,k}^{\text{opt}}}$. The codeword for the i-th k-block, $1 \leq i \leq l$, denoted $X_{(i-1)\cdot k}^{i\cdot k-1}$, is independent of the codewords selected for the other k-blocks.

Note that with this construction, the codebook satisfies that for each message $u \in \mathcal{U}$,

$$\begin{aligned}\frac{1}{l\cdot k}\sum_{i=0}^{l\cdot k-1}\left(x_u[i]\right)^2 &= \frac{1}{l}\sum_{l'=0}^{l-1}\left(\frac{1}{k}\sum_{k'=0}^{k-1}\left(x_u[l'\cdot k+k']\right)^2\right)\\ &= \frac{1}{k}\sum_{k'=0}^{k-1}\left(\frac{1}{l}\sum_{l'=0}^{l-1}\left(x_u[l'\cdot k+k']\right)^2\right)\\ &\underset{l\to\infty}{\overset{\text{(in prob.)}}{\longrightarrow}} \frac{1}{k}\sum_{k'=0}^{k-1}\mathbb{E}\left\{\left(X[k']\right)^2\right\}\leq P, \quad \text{(B.19)}\end{aligned}$$

where the limit as $l \to \infty$ follows from the weak law of large numbers, see e.g., [53, Sec. 9.1], and the inequality follows from the trace constraint in the definition of $\mathcal{C}_{X^{(k)}}$. It follows that for every $\delta > 0$, we can select $l \in \mathbb{N}^+$ sufficiently large s.t. $\forall u \in \mathcal{U}$,

$$\Pr\left(\frac{1}{l\cdot k}\sum_{i=0}^{l\cdot k-1}\left(X_u[i]\right)^2 > P\right) \leq \delta.$$

A message u of rate R is transmitted via a codeword $X_u^{(l\cdot k)} \equiv \left\{X_{u,(i-1)\cdot k}^{i\cdot k-1}\right\}_{i=1}^{l}$, by splitting the information bit sequence of length $l \cdot k \cdot R$ into l blocks, each contains $k \cdot R$ bits, where each block of $k \cdot R$ bits is mapped into a codeword of length k, e.g., the i-th block of the message u is mapped into $X_{u,(i-1)\cdot k}^{i\cdot k-1} \in \mathcal{CB}_k^{(i)}(\tau_{\epsilon,k}^{\text{opt}})$. Lastly, the transmitted $X_u^{(l\cdot k)}$ is transmitted as a sequence $\tilde{X}^{(l\cdot(k+\tau_{\rm m}))}$ of $l \cdot (k + \tau_{\rm m})$ samples, which is sent over $l \cdot (k + \tau_{\rm m}) \cdot T_{\rm s}(\epsilon) + l \cdot \Delta_g$ time units. The rate of this scheme is then $R \cdot \frac{k}{k+\tau_{\rm m}+\Delta_g/T_{\rm s}(\epsilon)} = R \cdot \left(1 - \frac{\tau_{\rm m}+\Delta_g/T_{\rm s}(\epsilon)}{k+\tau_{\rm m}+\Delta_g/T_{\rm s}(\epsilon)}\right)$.

2) Receiver's Operations: At the beginning of reception, the receiver identifies the start time of the received sequence. From that point, as the symbol rate at the receiver is synchronized with $T_{\rm s}(\epsilon)$, the receiver can maintain k-block synchronization as applied at the transmitter: Let $\tilde{Y}_\epsilon^{(l\cdot(k+\tau_{\rm m}))} = \tilde{X}^{(l\cdot(k+\tau_{\rm m}))} + \tilde{W}_\epsilon^{(l\cdot(k+\tau_{\rm m}))}$ denote the received samples, observed over $l\cdot(k+\tau_{\rm m})\cdot T_{\rm s}(\epsilon)+l\cdot\Delta_g$ time units, obtained by receiving a block of $k+\tau_{\rm m}$ samples, and then waiting for Δ_g times units to process the next block of $k+\tau_{\rm m}$ samples. The receiver then keeps only the first k samples of each block, discarding the last $\tau_{\rm m}$ samples. This processing results in a received sequence of l k-blocks denoted $\left\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\right\}_{i=1}^{l} \equiv Y_\epsilon^{(l\cdot k)}$, which are used by the decoder to receover the message $u \in \mathcal{U}$.

In the following we denote the mutual information density rate for the i-th k-block, with initial sampling time $\tau_{\epsilon,k}^{(i)}$ and input distribution $F_{X_{(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}}$, with $Z_{k,\epsilon}^{(i)}\left(F_{X_{(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}}|\tau_{\epsilon,k}^{(i)}\right) \triangleq \frac{1}{k}\log\left(\frac{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}^{(i)}}\left(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\middle|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}^{(i)}\right)}{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}}\left(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}\right)}\right)$. Note that the guard interval also facilitates statistical independence between channel outputs at the decoder corresponding to different transmitted *messages*. The addition of the guard interval effectively decreases the information rate, however, as the blocklength k increases, the impact of such fixed-length guard interval on the information rate becomes asymptotically negligible, and hence, it does not impact capacity. Recall that $k \in \mathbb{N}^+$ denotes the length of a block of symbols without guard interval.

Using the above scheme it is shown that when $\left\{X_{(i-1)\cdot k,\text{ opt}}^{i\cdot k-1}\right\}_{i=1}^{l}$ are Gaussian with the optimal covariance matrix, $\mathsf{C}_{X^{(k)}}^{\text{opt}}$, designed for the optimal sampling phase $\tau_{\epsilon,k}^{\text{opt}}$, and $X_{\text{opt}}^{(l\cdot k)}$ is the corresponding input sequence, then

$$C_\epsilon \geq \text{p-}\liminf_{k\to\infty} Z_{k\cdot l,\epsilon}\left(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\right) \geq \liminf_{n\to\infty} C_n, \quad \text{(B.20)}$$

where the first inequality follows from [19, Thm. 3.6.1], as by (B.19) the constructed codebook satisfies the per-codeword power constraint (6) with a probability which is arbitrarily close to 1, and the limit-inferior in probability of any given input process which satisfies the per-codeword power constraint (6), clearly does not exceed capacity. Hence, it remains to show the inequality on the RHS.

Letting $\tau_{\epsilon,k}^{\rm opt}$ denote the sampling phase at the start of the transmitted message, then as the guard interval facilitates statistical independence between the k-blocks we obtain that

$$Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big) = \frac{1}{l}\sum_{i=1}^{l} Z_{k,\epsilon}^{(i)}\big(F_{X_{(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)$$

via the derivation leading to (B.21), as shown at the bottom of the page, where in step (a) in the derivation of (B.21), $\tau_{\epsilon,k}^{(i)}$ denotes the sampling phase of the i-th k-block which is generated by the transmission scheme, and the equality follows since both the input k-blocks $\big\{X_{(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}$ and the noise k-blocks $\big\{W_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}$ are mutually independent (over $1 \leq i \leq l$): For the noise, $\big\{W_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}$, independence among the k-blocks follows as the k-blocks are separated more than $\tau_{\rm m}$ samples apart, while the noise memory is $\tau_{\rm m}$; for the channel output, $\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}$, independence follows as both the noise k-blocks are independent (as explained above) and the input k-blocks are independent by the assumption of uniformity and independence of the messages, as well as the codebook generation process; step (b) in the derivation of (B.21) follows since at each block, sampling phase synchronization is applied, which results in $\tau_{\epsilon,k}^{(i)} = \tau_{\epsilon,k}^{\rm opt}$ for all k-blocks $1 \leq i \leq l$.

Note that as the sampling phase (within a period of the noise correlation function), $\tau_{\epsilon,k}^{\rm opt}$, is identical for all k-blocks, it follows that the noise process has the *same correlation matrix* for every k-block. Therefore, the CDF $F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}$ which maximizes the mutual information for the i-th k-block, $\frac{1}{k}I(X_{(i-1)\cdot k}^{i\cdot k-1};Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt})$, is identical for all the k-blocks (i.e., for all of the l blocks, each of length k): $F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}} = F_{X_{\rm opt}^{(k)}|\tau_{\epsilon,k}^{\rm opt}}$, $i = 1,2,\ldots,l$. Letting $F_{X_{\rm opt}^{(l\cdot k)}|\tau_{\epsilon,k}^{\rm opt}} = \prod_{i=1}^{l} F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}$, we conclude that the RVs $\Big\{Z_{k,\epsilon}^{(i)}\big(F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)\Big\}_{i=1}^{l}$ all have identical distributions. Thus,

$$\begin{aligned}
&\mathbb{E}\Big\{Z_{l\cdot k,\epsilon}\big(F_{X_{\rm opt}^{(l\cdot k)}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)\Big\}\\
&\quad= \mathbb{E}\Big\{\frac{1}{l}\sum_{i=1}^{l} Z_{k,\epsilon}^{(i)}\big(F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)\Big\}\\
&\quad= \frac{1}{l}\sum_{i=1}^{l}\mathbb{E}\Big\{Z_{k,\epsilon}^{(i)}\big(F_{X_{(i-1)\cdot k,{\rm opt}}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)\Big\}\\
&\quad\overset{(a)}{=} \frac{1}{k}I\big(X_{\rm opt}^{(k)};Y_{\epsilon}^{(k)}|\tau_{\epsilon,k}^{\rm opt}\big)
\end{aligned}$$

$$\Delta_g = \begin{cases} \tau_{\epsilon,k}^{\rm opt} - \Big(\big(\tau_{\epsilon,k}^{\rm opt} + (k+\tau_{\rm m})\cdot T_{\rm s}(\epsilon)\big) \bmod T_{\rm pw}\Big) & , \Big(\big(\tau_{\epsilon,k}^{\rm opt} + (k+\tau_{\rm m})\cdot T_{\rm s}(\epsilon)\big) \bmod T_{\rm pw}\Big) \leq \tau_{\epsilon,k}^{\rm opt} \\ \tau_{\epsilon,k}^{\rm opt} + T_{\rm pw} - \Big(\big(\tau_{\epsilon,k}^{\rm opt} + (k+\tau_{\rm m})\cdot T_{\rm s}(\epsilon)\big) \bmod T_{\rm pw}\Big) & , \Big(\big(\tau_{\epsilon,k}^{\rm opt} + (k+\tau_{\rm m})\cdot T_{\rm s}(\epsilon)\big) \bmod T_{\rm pw}\Big) > \tau_{\epsilon,k}^{\rm opt} \end{cases} \tag{B.18}$$

$$\begin{aligned}
&Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)\\
&\triangleq \frac{1}{l\cdot k}\log\left(\frac{p_{Y_{\epsilon}^{(l\cdot k)}|X^{(l\cdot k)},\tau_{\epsilon,k}}\big(Y_{\epsilon}^{(l\cdot k)}\big|X^{(l\cdot k)},\tau_{\epsilon,k}^{\rm opt}\big)}{p_{Y_{\epsilon}^{(l\cdot k)}|\tau_{\epsilon,k}}\big(Y_{\epsilon}^{(l\cdot k)}|\tau_{\epsilon,k}^{\rm opt}\big)}\right)\\
&= \frac{1}{l\cdot k}\log\left(\frac{p_{\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\big\{X_{(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l},\tau_{\epsilon,k}}\Big(\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\Big|\big\{X_{(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l},\tau_{\epsilon,k}^{\rm opt}\Big)}{p_{\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\tau_{\epsilon,k}}\Big(\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\tau_{\epsilon,k}^{\rm opt}\Big)}\right)\\
&= \frac{1}{l\cdot k}\log\left(\frac{p_{\big\{W_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\tau_{\epsilon,k}}\Big(\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1} - X_{(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\Big|\tau_{\epsilon,k}^{\rm opt}\Big)}{p_{\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\tau_{\epsilon,k}}\Big(\big\{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big\}_{i=1}^{l}\big|\tau_{\epsilon,k}^{\rm opt}\Big)}\right)\\
&\overset{(a)}{=} \frac{1}{l\cdot k}\log\left(\prod_{i=1}^{l}\frac{p_{W_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1} - X_{(i-1)\cdot k}^{i\cdot k-1}\big|\tau_{\epsilon,k}^{(i)}\big)}{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{(i)}\big)}\right)\\
&\overset{(b)}{=} \frac{1}{l\cdot k}\log\left(\prod_{i=1}^{l}\frac{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}^{\rm opt}\big)}{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}\big)}\right)\\
&= \frac{1}{l}\sum_{i=1}^{l}\frac{1}{k}\log\left(\frac{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}\big|X_{(i-1)\cdot k}^{i\cdot k-1},\tau_{\epsilon,k}^{\rm opt}\big)}{p_{Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}}\big(Y_{\epsilon,(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}\big)}\right)\\
&= \frac{1}{l}\sum_{i=1}^{l} Z_{k,\epsilon}^{(i)}\big(F_{X_{(i-1)\cdot k}^{i\cdot k-1}|\tau_{\epsilon,k}^{\rm opt}}|\tau_{\epsilon,k}^{\rm opt}\big)
\end{aligned} \tag{B.21}$$

$$\begin{aligned}&\text{var}\Big(Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}_{\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}|\tau^{\text{opt}}_{\epsilon,k}\big)\Big)\\&\quad\overset{(b)}{=}\frac{1}{l^2}\sum_{i=1}^{l}\text{var}\Big(Z^{(i)}_{k,\epsilon}\big(F_{X^{i\cdot k-1}_{(i-1)\cdot k,\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}|\tau^{\text{opt}}_{\epsilon,k}\big)\Big)\\&\quad\overset{(c)}{\leq}\frac{3}{l\cdot k},\end{aligned}$$

where steps (a) and (c) follow from the derivation in Appendix A, after Eqn. (A.9): $\mathbb{E}\left\{Z_{k,\epsilon}\left(F_{X^{(k)}_{\text{opt}}|\tau_{\epsilon,k}}|\tau_{\epsilon,k}\right)\right\}=\frac{1}{k}I\big(X^{(k)}_{\text{opt}};Y^{(k)}_{\epsilon}|\tau_{\epsilon,k}\big)$ and $\text{var}\Big(Z_{k,\epsilon}\big(F_{X^{(k)}_{\text{opt}}|\tau_{\epsilon,k}}|\tau_{\epsilon,k}\big)\Big)\leq\frac{3}{k}$; step (b) follows since the mutual information density rates of the different k-blocks are mutually independent.

Since the variance of $Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}_{\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}|\tau^{\text{opt}}_{\epsilon,k}\big)$ decreases as $l\cdot k$ increases, then, by Chebyshev's inequality [40, Eqn. (5-88)], we obtain $\Pr\Big(\Big|Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}_{\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}|\tau^{\text{opt}}_{\epsilon,k}\big)-\frac{1}{k}I\big(X^{(k)}_{\text{opt}};Y^{(k)}_{\epsilon}|\tau^{\text{opt}}_{\epsilon,k}\big)\Big|>\frac{1}{(l\cdot k)^{1/3}}\Big)<\frac{3}{(l\cdot k)^{1/3}}$, and we conclude that $\forall k\in\mathbb{N}^+$ and $\forall\delta>0$, $\exists l_0(k,\delta)\in\mathbb{N}^+$ s.t. $\forall l>l_0(k,\delta)$, it follows that

$$\Pr\Big(Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}_{\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}|\tau^{\text{opt}}_{\epsilon,k}\big)<\frac{1}{k}I\big(X^{(k)}_{\text{opt}};Y^{(k)}_{\epsilon}|\tau^{\text{opt}}_{\epsilon,k}\big)-\delta\Big)<3\delta. \quad \text{(B.22)}$$

To show achievability of $\liminf_{n\to\infty}C_n$ we will show the RHS of (B.20). First, recall that

$$\begin{aligned}&\text{p-}\liminf_{k\to\infty}Z_{k,\epsilon}\big(F_{X^{(k)}|\tau_{\epsilon,k}}|\tau_{\epsilon,k}\big)\\&\quad=\sup\left\{\alpha\in\mathbb{R}\,\Big|\lim_{k\to\infty}\Pr\Big(Z_{k,\epsilon}\big(F_{X^{(k)}|\tau_{\epsilon,k}}|\tau_{\epsilon,k}\big)<\alpha\Big)=0\right\},\end{aligned}$$

hence, $\liminf_{n\to\infty}C_n$ is achievable if, for a given finite and bounded constant ξ (that will be defined later), $\forall\delta>0$, $\Pr\Big(Z_{l\cdot k,\epsilon}\big(F_{X^{(l\cdot k)}|\tau_{\epsilon,k}}|\tau_{\epsilon,k}\big)<\liminf_{n\to\infty}C_n-\delta\cdot(5+2\cdot\xi)\Big)$, can be made arbitrarily small by properly selecting $k\in\mathbb{N}^+$, $\tau_{\epsilon,k}$ and $F_{X^{(k)}|\tau_{\epsilon,k}}$, and taking l sufficiently large. Next, define $\gamma(k)$ via Eqn. (B.23), as shown at the bottom of the next page, and pick $k\in\mathbb{N}^+$ s.t. $\frac{\tau_{\text{m}}\cdot\gamma(k)}{k+\tau_{\text{m}}}<\delta$. In addition, k is selected sufficiently large such that the values of the sampled variance $c^{\{\tau_0\}}_{W_\epsilon}[i,0]$ over a k-block starting at sampling phase τ_0 satisfy

$$\Big|\frac{1}{k}\sum_{i=0}^{k-1}c^{\{\tau_0\}}_{W_\epsilon}[i,0]-\frac{1}{T_{\text{pw}}}\int_{t=0}^{T_{\text{pw}}}c_{W_{\text{c}}}(t+\tau_0,0)\text{d}t\Big|<\frac{\delta}{2},\quad\forall\tau_0\in[0,T_{\text{pw}}).\quad\text{(B.24)}$$

Such a selection is possible since the sampling interval $T_{\text{s}}(\epsilon)$ is incommensurate with the noise period T_{pw}. Then, for a sufficiently large k, the sampling points will be nearly uniformly distributed over $[0,T_{\text{pw}})$, see also discussion after Eqn. (A.16). By definition of a uniformly distributed modulo 1 sequence,[9] [41, Def. 1.1], it follows that the empirical distribution of the sampling instances approaches a uniform distribution on $[0,T_{\text{pw}}]$. Then, by [41, Thm. 1.1], as the correlation function (at any given lag, and hence also at lag $\lambda=0$) is a continuous mapping of the time, then (B.24) follows.

[9]Note that switching from modulo 1 to modulo T_{pw} amounts to scaling of the time axis, which can be incorporated into the definition in a straightforward manner.

When k is fixed, we pick $n\in\mathbb{N}^+$ s.t. $\liminf_{n_0\to\infty}C_{n_0}<C_n+\delta$, and also

$$\Big|\frac{1}{k}I(X^{(k)}_{\text{opt}};Y^{(k)}_{\epsilon}|\tau^{\text{opt}}_{\epsilon,k})-\frac{1}{k}I(X^{(k)}_{n,\text{opt}};Y^{(k)}_{n}|\tau^{\text{opt}}_{n,k})\Big|<\delta,\quad\text{(B.25)}$$

where we recall that $F_{X^{(k)}_{\text{opt}}|\tau^{\text{opt}}_{\epsilon,k}}$ and $F_{X^{(k)}_{n,\text{opt}}|\tau^{\text{opt}}_{n,k}}$ are Gaussian CDFs, $\big(\tau^{\text{opt}}_{n,k},X^{(k)}_{n,\text{opt}}\big)$ maximize $\frac{1}{k}I(X^{(k)}_n;Y^{(k)}_n|\tau_{n,k})$, and $\big(\tau^{\text{opt}}_{\epsilon,k},X^{(k)}_{\text{opt}}\big)$ maximize $\frac{1}{k}I(X^{(k)};Y^{(k)}_{\epsilon}|\tau_{\epsilon,k})$. Such $n\in\mathbb{N}^+$ exists by the convergence in Lemma B.1. Lastly, the selected n is increased to guarantee that

$$\Big|\frac{1}{k}\sum_{i=0}^{k-1}c^{\{\tau_0\}}_{W_\epsilon}[i,0]-\frac{1}{k}\sum_{i=0}^{k-1}c^{\{\tau_0\}}_{W_n}[i,0]\Big|<\frac{\delta}{2},\quad\forall\tau_0\in[0,T_{\text{pw}}),$$

which implies that

$$\Big|\frac{1}{k}\sum_{i=0}^{k-1}c^{\{\tau_0\}}_{W_n}[i,0]-\frac{1}{T_{\text{pw}}}\int_{t=0}^{T_{\text{pw}}}c_{W_{\text{c}}}(t+\tau_0,0)\text{d}t\Big|<\delta,\quad\forall\tau_0\in[0,T_{\text{pw}}).\quad\text{(B.26)}$$

After picking n, we pick $l\in\mathbb{N}^+$ such that

$$C_n\leq\frac{1}{l\cdot(k+\tau_{\text{m}})}I\big(\tilde{X}^{(l\cdot(k+\tau_{\text{m}}))}_n;\tilde{Y}^{(l\cdot(k+\tau_{\text{m}}))}_n|\tau_{n,l}\big)+\delta,\quad\tau_{n,l}\in[0,T_{\text{pw}}],$$

where $\tilde{X}_n[i]$ is the capacity-achieving input process for the channel (8), see [6, Thm. 1], and $\tilde{Y}_n[i]$ is the corresponding output process: $\tilde{Y}_n[i]=\tilde{X}_n[i]+W_n[i]$. Such $\tau_{n,l}$ exists since we can set $\tau_{n,l}=\underset{\tau_0\in[0,T_{\text{pw}}]}{\text{argmax}}\;C_n(\tau_0)\triangleq\tau^{\text{opt}}_n$. Then, for τ^{opt}_n, k and n, as $\tilde{X}_n[i]$ is the capacity-achieving input process, there is a codeword length beyond which the mutual information between the channel input and output is less than δ away from capacity. Next, define RVs the $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Y}_1$, and $\mathbf{Y}_2$ as follows:

$$\begin{aligned}\mathbf{X}_1&=\big\{\tilde{X}^{i\cdot(k+\tau_{\text{m}})-1}_{n,i\cdot\tau_{\text{m}}+(i-1)\cdot k}\big\}_{i=1}^{l},\\\mathbf{X}_2&=\big\{\tilde{X}^{(i-1)\cdot(k+\tau_{\text{m}})+\tau_{\text{m}}-1}_{n,(i-1)\cdot(k+\tau_{\text{m}})}\big\}_{i=1}^{l},\\\mathbf{Y}_1&=\big\{\tilde{Y}^{i\cdot(k+\tau_{\text{m}})-1}_{n,i\cdot\tau_{\text{m}}+(i-1)\cdot k}\big\}_{i=1}^{l},\\\mathbf{Y}_2&=\big\{\tilde{Y}^{(i-1)\cdot(k+\tau_{\text{m}})+\tau_{\text{m}}-1}_{n,(i-1)\cdot(k+\tau_{\text{m}})}\big\}_{i=1}^{l}.\end{aligned}$$

We note that $(\mathbf{X}_2^T,\mathbf{X}_1^T)^T$ is a *permutation* of the vector $\tilde{X}^{l\cdot(k+\tau_{\text{m}})}_n$. Let us denote this permutation with the matrix P, i.e., $(\mathbf{X}_2^T,\mathbf{X}_1^T)^T=\mathsf{P}\cdot\tilde{X}^{l\cdot(k+\tau_{\text{m}})}_n$. Similarly, we write $(\mathbf{W}_2^T,\mathbf{W}_1^T)^T=(\mathbf{Y}_2^T,\mathbf{Y}_1^T)^T-(\mathbf{X}_2^T,\mathbf{X}_1^T)^T=\mathsf{P}\cdot W^{l\cdot(k+\tau_{\text{m}})}_n$. Note that $\text{cov}\big((\mathbf{W}_2^T,\mathbf{W}_1^T)^T\big|\tau_{n,l}\big)=\mathbb{E}\{\mathsf{P}\cdot W^{l\cdot(k+\tau_{\text{m}})}_n\cdot\big(W^{l\cdot(k+\tau_{\text{m}})}_n\big)^T\cdot\mathsf{P}^T\big|\tau_{n,l}\}=\mathsf{P}\cdot\mathsf{C}_{W^{l\cdot(k+\tau_{\text{m}})}_n}(\tau_{n,l})\cdot\mathsf{P}^T$. With these definitions, applying the chain rules for differential entropy and for mutual information, we write

$$\begin{aligned}
&I\big(\tilde{X}_n^{l\cdot(k+\tau_{\rm m})};\tilde{Y}_n^{l\cdot(k+\tau_{\rm m})}|\tau_{n,l}\big)\\
&\equiv I(\mathbf{X}_1,\mathbf{X}_2;\mathbf{Y}_1,\mathbf{Y}_2|\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1,\mathbf{Y}_2|\tau_{n,l}) + I(\mathbf{X}_2;\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + I(\mathbf{X}_1;\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l})\\
&\qquad + I(\mathbf{X}_2;\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) - h(\mathbf{Y}_2|\mathbf{X}_1,\mathbf{Y}_1,\tau_{n,l})\\
&\qquad + h(\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\tau_{n,l}) - h(\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\mathbf{X}_2,\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) - h(\mathbf{Y}_2|\mathbf{X}_1,\mathbf{Y}_1,\tau_{n,l})\\
&\qquad + h(\mathbf{Y}_1|\mathbf{X}_1,\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{X}_1,\mathbf{Y}_1,\tau_{n,l})\\
&\qquad\quad - h(\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\mathbf{X}_2,\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) + h(\mathbf{Y}_1|\mathbf{X}_1,\tau_{n,l})\\
&\qquad - h(\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}_1,\mathbf{X}_2,\tau_{n,l})\\
&\overset{(a)}{=} I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) + h(\mathbf{W}_1|\mathbf{X}_1,\tau_{n,l})\\
&\qquad - h(\mathbf{W}_1,\mathbf{W}_2|\mathbf{X}_1,\mathbf{X}_2,\tau_{n,l})\\
&\overset{(b)}{=} I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) + h(\mathbf{W}_1|\tau_{n,l})\\
&\qquad - h(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\\
&= I(\mathbf{X}_1;\mathbf{Y}_1|\tau_{n,l}) + h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) - h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l}),
\end{aligned}$$

where (a) follows as $\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{W}_1$ and $\mathbf{Y}_2 = \mathbf{X}_2 + \mathbf{W}_2$; and (b) follows as $(\mathbf{W}_1,\mathbf{W}_2) \perp\!\!\!\perp (\mathbf{X}_1,\mathbf{X}_2)$. Next, denoting the maximal diagonal element of the matrix A with $\max\mathrm{Diag}\{\mathsf{A}\}$, we can upper bound $h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) - h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l})$ via (B.27), as shown at the bottom of the next page. In the derivation of (B.27), step (a) follows from [53, Corollary on Pg. 253]; (b) follows since $\mathbf{Y}_2|\tau_{n,l}$ and $(\mathbf{W}_2^T,\mathbf{W}_1^T)^T|\tau_{n,l}$ are Gaussian vectors, and since the conditional covariance for jointly Gaussian RVs is independent of the conditioning value, see, e.g., [54, Ch. 21.6]:

$$\begin{aligned}
&2\cdot h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l})\\
&= 2\cdot\int_{\mathbf{w}_1\in\mathbb{R}^{l\cdot k}} f_{\mathbf{W}_1|\tau_{n,l}}(\mathbf{w}_1|\tau_{n,l})h(\mathbf{W}_2|\mathbf{W}_1=\mathbf{w}_1,\tau_{n,l})\mathrm{d}\mathbf{w}_1\\
&\overset{(a')}{=}\int_{\mathbf{w}_1\in\mathbb{R}^{l\cdot k}} f_{\mathbf{W}_1|\tau_{n,l}}(\mathbf{w}_1|\tau_{n,l})\cdot\mathrm{logdet}\Big((2\pi e)\cdot\big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})\\
&\quad-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\big)\Big)\mathrm{d}\mathbf{w}_1\\
&= \mathrm{logdet}\Big((2\pi e)\cdot\big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})\\
&\quad-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\big)\Big),
\end{aligned}$$

where the expression for the conditional correlation matrix is step (a′) is given in [54, Ch. 21.6]. Step (c) follows from Hadamard's inequality [53, Eqn. (8.64)], as the determinant of a symmetric positive semidefinite matrix is upper bounded by the product of its diagonal elements, and we take $l\cdot\tau_{\rm m}$ multiples of the largest possible diagonal element to further upper bound this product. In step (d) we used [55, Lemma 11.6] and the fact that for any positive x it holds that $\log(x) = \ln(x)\cdot\log(e) \leq (x-1)\cdot\log(e)$; step (e) follows since $\Big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l}) - \mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\Big)^{-1}$ is the upper-left block of the inverse covariance matrix of the vector $(\mathbf{W}_2^T,\mathbf{W}_1^T)^T$, namely, the upper-left block of $\Big(\mathrm{cov}\big((\mathbf{W}_2^T,\mathbf{W}_1^T)^T|\tau_{n,l}\big)\Big)^{-1}$, [45, Eqn. (0.7.3.1)]. Then, having more elements can only increase the maximum. In step (f) we upper bound the largest diagonal element by the matrix 1-norm, as in step (a) in the derivation of (B.6); in step (g) we use the fact that permutation matrices are orthogonal, hence,

$$\begin{aligned}
&\big(\mathsf{P}\cdot\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\cdot\mathsf{P}^T\big)^{-1}\\
&\quad= \big(\mathsf{P}^T\big)^{-1}\cdot\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\cdot\mathsf{P}^{-1}\\
&\quad= \mathsf{P}\cdot\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\cdot\mathsf{P}^T,
\end{aligned}$$

and the fact that induced matrix norms are sub-multiplicative [45, Ch. 5.6 and Example 5.6.4], thereby $\Big\|\mathsf{P}\cdot\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\cdot\mathsf{P}^T\Big\|_1 \leq \|\mathsf{P}\|_1\cdot\Big\|\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\Big\|_1\cdot\big\|\mathsf{P}^T\big\|_1=\Big\|\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\Big\|_1$, where the last equality follows since for a permutation matrix P, $\|\mathsf{P}\|_1 = \big\|\mathsf{P}^T\big\|_1 = 1$. Lastly, step (h) in the derivation of (B.27) follows similarly to step (c) in the derivation of (B.13). Using the definition of $\gamma(k)$ in (B.23) we obtain

$$h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l}) - h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l}) \leq l\cdot\tau_{\rm m}\cdot\gamma(k).$$

With this bound, letting $E_W \triangleq \frac{1}{T_{\rm pw}}\int_{t=0}^{T_{\rm pw}} c_{W_{\rm c}}(t+\tau_0,0)\mathrm{d}t$ and setting $\xi \triangleq \frac{3}{P+E_W} < \infty$, considering $\delta \leq \frac{P+E_W}{2}$ we can bound C_n via (B.29), where in the derivation of (B.29), in step (a) τ_i denotes the sampling phase of the i-th block, and we also use the fact that conditioning decreases the differential entropy [53, Corollary on Pg. 253], and the fact that the noise process $W_n[i]$ has a memory of $\tau_{\rm m}$; in (b) $\bar{Y}_n[i] = \bar{X}_n[i] + W_n[i]$. The bound is obtained as follows: The mutual information in (B.28), corresponds to the sum of the mutual information of l blocks, each consisting of k symbols. Note that the period of the noise in this additive Gaussian WSCS noise channel is p_n, and each period of p_n samples consists of $n_k \triangleq p_n/(k+\tau_{\rm m})$ independent[10] MIMO subchannels, each of size $k\times k$. Applying the DCD, as in the proof of [5, Thm. 1] we obtain an equivalent $(n_k\cdot k)\times(n_k\cdot k)$ MIMO channel:

$$\tilde{Y}^{(n_k\cdot k)}[i] = \tilde{X}^{(n_k\cdot k)}[i] + \tilde{W}_n^{(n_k\cdot k)}[i], \qquad \text{(B.30)}$$

[10]Note that if $p_n/(k+\tau_{\rm m})$ is not an integer we can set $n_k = p_n$, since $p_n\cdot(k+\tau_{\rm m})$ consecutive samples contain an integer number of periods and is thus a period of the noise.

$$\gamma(k) \;\triangleq\; \frac{1}{2}\cdot\log\Big(P\cdot k + \max_{0\leq t\leq T_{\rm pw}}\big\{c_{W_{\rm c}}(t,0)\big\}\Big) + \frac{1}{2}\cdot\log(e)\cdot\frac{1}{\displaystyle\min_{0\leq t\leq T_{\rm pw}}\Big\{c_{W_{\rm c}}(t,0) - 2\tau_{\rm m}\cdot\max_{|\lambda|>\frac{T_{\rm pw}}{p+1}}\big\{|c_{W_{\rm c}}(t,\lambda)|\big\}\Big\}} \qquad \text{(B.23)}$$

where $\tilde{W}_n^{(n_k\cdot k)}[i]$ is a memoryless stationary noise process. Then, as argued in [32, Sec, I.D], capacity subject to the power constraint (6) is equivalent to capacity subject to a (MIMO) per-symbol average power constraint, i.e.,

$$\mathrm{Tr}\Big\{\tilde{X}^{(n_k\cdot k)}\cdot\big(\tilde{X}^{(n_k\cdot k)}\big)^T\Big\}<n_k\cdot k\cdot P.$$

Next, observe that the capacity of the channel (B.30) subject to the above power constraint is obtained by water-filling over the eigenvalues, see e.g., [30, Eqns. (15)-(16)]. As by condition (13), $P>\max\limits_{t\in[0,T_{\rm pw}]}\Big(c_{W_{\rm c}}(t,0)+2\tau_{\rm m}\cdot\max\limits_{|\lambda|>\frac{T_{\rm pw}}{p}}\big\{|c_{W_{\rm c}}(t,\lambda)|\big\}\Big)$, then, by [45, Corollary 6.1.5] it follows that that $P>\mathrm{maxEig}\big\{\mathsf{C}_{W_n^{(\tilde{k})}}(\tau_0)\big\}$ for any $\tilde{k}\in\mathbb{N}^+$. This implies that the waterfilling solution will use all the eigenvalues of the noise correlation matrix. Recall that by the selection of n, see (B.26), for each k-block the trace of the noise correlation matrix is within δ from E_W. Then, letting $\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\triangleq\mathbb{E}\Big\{\tilde{W}_n^{(n_k\cdot k)}\cdot\big(\tilde{W}_n^{(n_k\cdot k)}\big)^T\Big|\tau_0\Big\}$ we obtain from (B.26) that

$$\Big|\mathrm{Tr}\Big\{\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\Big\}-n_k\cdot k\cdot E_W\Big|<n_k\cdot k\cdot\delta.$$

$$\begin{aligned}
&2\cdot\big(h(\mathbf{Y}_2|\mathbf{Y}_1,\tau_{n,l})-h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l})\big)\\
&\stackrel{(a)}{\leq}\ 2\cdot\big(h(\mathbf{Y}_2|\tau_{n,l})-h(\mathbf{W}_2|\mathbf{W}_1,\tau_{n,l})\big)\\
&\stackrel{(b)}{=}\ \mathrm{logdet}\big((2\pi e)\cdot\mathrm{cov}(\mathbf{Y}_2|\tau_{n,l})\big)\\
&\qquad-\mathrm{logdet}\Big((2\pi e)\cdot\big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\big)\Big)\\
&=\ \mathrm{logdet}\big(\mathrm{cov}(\mathbf{Y}_2|\tau_{n,l})\big)-\mathrm{logdet}\Big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})\\
&\qquad\qquad-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\Big)\\
&\stackrel{(c)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)\\
&\qquad-\mathrm{logdet}\Big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\Big)\\
&\stackrel{(d)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)\\
&\qquad+\log(e)\cdot\bigg(\mathrm{Tr}\Big\{\Big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\Big)^{-1}\Big\}\\
&\qquad\qquad-l\cdot\tau_{\rm m}\bigg)\\
&\leq\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)\\
&\qquad+(l\cdot\tau_{\rm m})\cdot\log(e)\cdot\mathrm{maxDiag}\Big\{\Big(\mathrm{cov}(\mathbf{W}_2|\tau_{n,l})\\
&\qquad\qquad-\mathrm{cov}(\mathbf{W}_2,\mathbf{W}_1|\tau_{n,l})\big(\mathrm{cov}(\mathbf{W}_1|\tau_{n,l})\big)^{-1}\mathrm{cov}(\mathbf{W}_1,\mathbf{W}_2|\tau_{n,l})\Big)^{-1}\Big\}\\
&\stackrel{(e)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)\\
&\qquad+\log(e)\cdot(l\cdot\tau_{\rm m})\cdot\mathrm{maxDiag}\Big\{\Big(\mathrm{cov}\big((\mathbf{W}_2^T,\mathbf{W}_1^T)^T|\tau_{n,l}\big)\Big)^{-1}\Big\}\\
&\stackrel{(f)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)+(l\cdot\tau_{\rm m})\cdot\log(e)\cdot\Big\|\big(\mathsf{P}\cdot\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\cdot\mathsf{P}^T\big)^{-1}\Big\|_1\\
&\stackrel{(g)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)+(l\cdot\tau_{\rm m})\cdot\log(e)\cdot\Big\|\big(\mathsf{C}_{W_n^{(l\cdot(k+\tau_{\rm m}))}}(\tau_{n,l})\big)^{-1}\Big\|_1\\
&\stackrel{(h)}{\leq}\ (l\cdot\tau_{\rm m})\cdot\log\big(P\cdot k+\max_{0\leq t\leq T_{\rm pw}}\{c_{W_{\rm c}}(t,0)\}\big)\\
&\qquad+(l\cdot\tau_{\rm m})\cdot\log(e)\cdot\frac{1}{\min\limits_{0\leq t\leq T_{\rm pw}}\Big\{c_{W_{\rm c}}(t,0)-2\tau_{\rm m}\cdot\max\limits_{|\lambda|>\frac{T_{\rm pw}}{p+1}}\big\{|c_{W_{\rm c}}(t,\lambda)|\big\}\Big\}}
\end{aligned}\qquad\text{(B.27)}$$

The waterfilling rate for this case is then

$$2R_{tot} = \frac{1}{k\cdot n_k}\log\left(\frac{\left(P+\frac{1}{n_k\cdot k}\mathrm{Tr}\left\{\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\right\}\right)^{k\cdot n_k}}{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\right)}\right) \tag{B.31}$$

$$\leq \frac{1}{k\cdot n_k}\log\left(\frac{\left(P+E_W+\delta\right)^{k\cdot n_k}}{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\right)}\right)$$

$$= \log\left(\frac{P+E_W+\delta}{\sqrt[k\cdot n_k]{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\right)}}\right). \tag{B.32}$$

Due to the discarding of τ_{m} samples every k samples, the n_k k-blocks of the noise process $\tilde{W}_n^{(n_k\cdot k)}$ are independent, thus, (B.30) consists of n_k parallel MIMO subchannels. The waterfilling solution to channel i_c within this set results in a rate of

$$2R_{i_c} = \frac{1}{k}\log\left(\frac{\left(P+\frac{1}{k}\mathrm{Tr}\left\{\mathsf{C}_{\tilde{W}_n^{(k)}}(\tau_{i_c})\right\}\right)^{k}}{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(k)}}(\tau_{i_c})\right)}\right)$$

$$\leq \log\left(\frac{P+E_W+\delta}{\sqrt[k]{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(k)}}(\tau_{i_c})\right)}}\right)$$

Then, the sum-rate over the n_k k-blocks can be upper bounded as

$$2\frac{1}{n_k}\sum_{i_c=1}^{n_k}R_{i_c} \leq \log\left(\frac{P+E_W+\delta}{\sqrt[n_k]{\prod_{i_c=1}^{n_k}\sqrt[k]{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(k)}}(\tau_{i_c})\right)}}}\right)$$

$$= \log\left(\frac{P+E_W+\delta}{\sqrt[k\cdot n_k]{\mathrm{Det}\left(\mathsf{C}_{\tilde{W}_n^{(n_k\cdot k)}}(\tau_0)\right)}}\right),$$

which coincides with the upper bound in (B.32). Finally, recalling that $\delta \leq \frac{P+E_W}{2}$, then, as we are interested in bounding the distance between the upper bound and (B.31), we consider

$$\begin{aligned}
&\log(P+E_W+\delta)-\log(P+E_W-\delta)\\
&\quad= \log\left((P+E_W)\left(1+\frac{\delta}{P+E_W}\right)\right)\\
&\qquad\quad - \log\left((P+E_W)\left(1-\frac{\delta}{P+E_W}\right)\right)\\
&\quad= \log(e)\cdot\ln\left(1+\frac{\delta}{P+E_W}\right)\\
&\qquad\quad - \log(e)\cdot\ln\left(1-\frac{\delta}{P+E_W}\right)
\end{aligned}$$

$$\begin{aligned}
C_n &\leq \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot I\left(\tilde{X}_n^{(l\cdot(k+\tau_{\mathrm{m}}))};\tilde{Y}_n^{(l\cdot(k+\tau_{\mathrm{m}}))}\middle|\tau_{n,l}\right)+\delta\\
&\leq \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot I\left(\left\{\tilde{X}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l};\left\{\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l}\middle|\tau_{n,l}\right)+\frac{l\cdot\tau_{\mathrm{m}}\cdot\gamma(k)}{l\cdot(k+\tau_{\mathrm{m}})}+\delta\\
&= \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot\Bigg(h\left(\left\{\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l}\middle|\tau_{n,l}\right)\\
&\qquad\qquad -h\left(\left\{\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l}\middle|\left\{\tilde{X}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l},\tau_{n,l}\right)\Bigg)+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\\
&= \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot\left(h\left(\left\{\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l}\middle|\tau_{n,l}\right)-h\left(\left\{W_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\right\}_{i=1}^{l}\middle|\tau_{n,l}\right)\right)+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\\
&\overset{(a)}{\leq} \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot\sum_{i=1}^{l}\left(h\left(\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\middle|\tau_i\right)-h\left(W_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\middle|\tau_i\right)\right)+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\\
&= \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot\sum_{i=1}^{l}\left(h\left(\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\middle|\tau_i\right)-h\left(\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\middle|\tilde{X}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1},\tau_i\right)\right)\\
&\qquad\qquad +\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\\
&= \frac{1}{l\cdot(k+\tau_{\mathrm{m}})}\cdot\sum_{i=1}^{l}I\left(\tilde{X}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1};\tilde{Y}_{n,i\cdot\tau_{\mathrm{m}}+(i-1)\cdot k}^{i\cdot(k+\tau_{\mathrm{m}})-1}\middle|\tau_i\right)+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta \qquad\qquad (\mathrm{B.28})\\
&\overset{(b)}{\leq} \frac{1}{k+\tau_{\mathrm{m}}}\cdot\max_{\left\{\substack{F_{\bar{X}_n^{(k)}}:\ \frac{1}{k}\sum_{i=0}^{k-1}\mathbb{E}\{(\bar{X}_n[i])^2\}\leq P,\\ \bar{\tau}_0\in[0,T_{\mathrm{pw}}]}\right\}} I\left(\bar{X}_n^{(k)};\bar{Y}_n^{(k)}\middle|\bar{\tau}_0\right)+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\cdot(1+2\cdot\xi)\\
&\overset{(c)}{=} \frac{k}{k+\tau_{\mathrm{m}}}\cdot\frac{1}{k}I(X_{n,\mathrm{opt}}^{(k)};Y_n^{(k)}|\tau_{n,k}^{\mathrm{opt}})+\frac{\tau_{\mathrm{m}}\cdot\gamma(k)}{k+\tau_{\mathrm{m}}}+\delta\cdot(1+2\cdot\xi)\\
&\overset{(d)}{\leq} \frac{1}{k}I(X_{n,\mathrm{opt}}^{(k)};Y_n^{(k)}|\tau_{n,k}^{\mathrm{opt}})+2\delta\cdot(1+\xi) \qquad\qquad (\mathrm{B.29})
\end{aligned}$$

$$\overset{(a')}{\leq} \log(e) \cdot \left(\frac{\delta}{P + E_W} - \frac{-\frac{\delta}{P+E_W}}{1 - \frac{\delta}{P+E_W}} \right)$$
$$= \delta \cdot \log(e) \cdot \left(\frac{1}{P + E_W} + \frac{1}{P + E_W - \delta} \right)$$
$$\overset{(b')}{\leq} \delta \cdot \log(e) \cdot \frac{3}{P + E_W}$$
$$< 2 \cdot \delta \cdot \xi,$$

where in (a′) we used $\frac{x}{1+x} < \ln(1+x) < x, \forall x > -1$, which holds for $\delta \leq \frac{P+E_W}{2}$; and (b′) follows as $\delta \leq \frac{P+E_W}{2}$. Thus, step (b) in the derivation of (B.29) follows as we can upper bound the capacity of the channel in (B.28) by the capacity of the best k-block subchannel, with a maximum error of $2 \cdot \delta \cdot \xi$, and we further maximize the mutual information over the initial phase; step (c) follows as the maximization in step (b) coincides with (B.9a). As detailed in Section B-B, we use $\tau_{n,k}^{\text{opt}}$ to denote the optimal sampling phase, $X_{n,\text{opt}}^{(k)}$ to denote the corresponding optimal input vector, and $Y_n^{(k)}$ to denote the corresponding channel output, for the channel (8). Step (d) follows from our choice of $k \in \mathbb{N}^+$.

Lastly, we analyze the mutual information density rate of the scheme considered in Subsections B-C.1-B-C.2. For any $\delta \in \mathbb{R}$ s.t. $0 < \delta \leq \frac{P+E_W}{2}$, we can select k, n and l s.t.

$$\Pr\left(Z_{l\cdot k,\epsilon}\big(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\big) \leq \liminf_{n\to\infty} C_n - \delta \cdot (5 + 2 \cdot \xi) \right)$$
$$\overset{(a)}{\leq} \Pr\left(Z_{l\cdot k,\epsilon}\big(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\big) \leq C_n - \delta \cdot (4 + 2 \cdot \xi) \right)$$
$$\overset{(b)}{\leq} \Pr\left(Z_{l\cdot k,\epsilon}\big(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\big) \leq \frac{1}{k} I(X_{n,\text{opt}}^{(k)}; Y_n^{(k)}|\tau_{n,k}^{\text{opt}}) - 2\delta \right)$$
$$\overset{(c)}{\leq} \Pr\left(Z_{l\cdot k,\epsilon}\big(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\big) \leq \frac{1}{k} I\big(X_{\text{opt}}^{(k)}; Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\text{opt}}\big) - \delta \right)$$
$$\overset{(d)}{\leq} 3\delta,$$

where (a) follows as $\liminf_{n_0\to\infty} C_{n_0} < C_n + \delta$ by the selection of $n \in \mathbb{N}^+$; (b) follows from (B.29) and our selection of k, n and l. Step (c) follows as the mutual information expressions $\frac{1}{k} I(X_{n,\text{opt}}^{(k)}; Y_n^{(k)}|\tau_{n,k}^{\text{opt}})$ and $\frac{1}{k} I(X_{\text{opt}}^{(k)}; Y_\epsilon^{(k)}|\tau_{\epsilon,k}^{\text{opt}})$ are maximized by Gaussian inputs [30, Eqns. (4), (30)]. Then, due to Lemma B.1, for the fixed $k \in \mathbb{N}^+$, n can be selected such that (B.25) is satisfied. Lastly, step (d) follows from the bound in Eqn. (B.22), as $l \in \mathbb{N}^+$ can be selected arbitrarily large. Consequently, we conclude that

$$\lim_{l\to\infty} \Pr\left(Z_{l\cdot k,\epsilon}\big(F_{X_{\text{opt}}^{(l\cdot k)}|\tau_{\epsilon,k}^{\text{opt}}}|\tau_{\epsilon,k}^{\text{opt}}\big) < \liminf_{n\to\infty} C_n \right) = 0.$$

By the rate expression in Section B-C.1 we obtain that a rate of $\liminf_{n\to\infty} C_n \cdot \left(1 - \frac{\tau_\text{m}+\Delta_g/T_\text{s}(\epsilon)}{k+\tau_\text{m}+\Delta_g/T_\text{s}(\epsilon)}\right)$ is achievable. Note that fixing k sufficiently large, we can approach $\liminf_{n\to\infty} C_n$ arbitrarily close. As a final comment, we note that we considered blocklengths which are integer multiple of k. Since k is fixed, then by taking l sufficiently large, we can use zero-padding of up to $k-1$ zeros and obtain a code with any blocklength, causing only an arbitrarily small rate decrease.

It thus follows that $\liminf_{n\to\infty} C_n$ is both achievable and, by (B.17), it is the upper bound on the achievable rate, hence it is the capacity for the situation in which the transmitter can select the most appropriate sampling phase for transmission. ∎

References

[1] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[2] J. Campbell, A. Gibbs, and B. Smith, "The cyclostationary nature of crosstalk interference from digital signals in multipair cable-part II: Applications and further results," *IEEE Trans. Commun.*, vol. 31, no. 5, pp. 638–649, May 1983.

[3] R. Blasco-Serrano, J. Lv, R. Thobaben, E. Jorswieck, A. Kliks, and M. Skoglund, "Comparison of underlay and overlay spectrum sharing strategies in MISO cognitive channels," in *Proc. Int. ICST Conf. Cognit. Radio Oriented Wireless Netw. Commun. (CROWNCOM)*, Jan. 2012, pp. 224–229.

[4] W. A. Gardner, *Cyclostationarity in Communications and Signal Processing*. Piscataway, NJ, USA: IEEE Press, 1994.

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

[6] N. Shlezinger and R. Dabora, "The capacity of discrete-time Gaussian MIMO channels with periodic characteristics," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jun. 2016, pp. 1058–1062.

[7] N. Shlezinger, D. Zahavi, Y. Murin, and R. Dabora, "The secrecy capacity of Gaussian MIMO channels with finite memory," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1874–1897, Mar. 2017.

[8] N. Shlezinger, R. Shaked, and R. Dabora, "On the capacity of MIMO broadband power line communications channels," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4795–4810, Oct. 2018.

[9] N. Shlezinger and R. Dabora, "Frequency-shift filtering for OFDM signal recovery in narrowband power line communications," *IEEE Trans. Commun.*, vol. 62, no. 4, pp. 1283–1295, Apr. 2014.

[10] N. Shlezinger, K. Todros, and R. Dabora, "Adaptive filtering based on time-averaged MSE for cyclostationary signals," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1746–1761, Apr. 2017.

[11] J. R. Vig, "Quartz crystal resonators and oscillators for frequency control and timing applications: A tutorial (revision 6.1)," Army Lab. Command, Electron. Technol. Devices Lab., Fort Monmouth, NJ, USA, Tech. Rep. SLCET-TR-88-1, ADA268205, May 1993.

[12] I. Gaspar, L. Mendes, M. Matthé, N. Michailow, A. Festag, and G. Fettweis, "LTE-compatible 5G PHY based on generalized frequency division multiplexing," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Aug. 2014, pp. 209–213.

[13] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Process.*, vol. 86, no. 4, pp. 639–697, Apr. 2006.

[14] R. L. Dobrushin, "General formulation of Shannon's main theorem in information theory," *Amer. Math. Soc. Trans.*, vol. 33, no. 2, pp. 323–438, 1963.

[15] N. Shlezinger, E. Abakasanga, R. Dabora, and Y. C. Eldar, "The capacity of memoryless channels with sampled cyclostationary Gaussian noise," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 106–121, Jan. 2020.

[16] E. Abakasanga, N. Shlezinger, and R. Dabora, "On the rate-distortion function of sampled cyclostationary Gaussian processes," *Entropy*, vol. 22, no. 3, p. 345, Mar. 2020.

[17] Y. Chen, Y. C. Eldar, and A. J. Goldsmith, "Shannon meets Nyquist: Capacity of sampled Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4889–4914, Aug. 2013.

[18] G. Han and S. Shamai (Shitz), "On sampling continuous-time AWGN channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 2, pp. 782–794, Feb. 2022.

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

[20] V. K. Madisetti and D. B. Williams, *The Digital Signal Processing Handbook*. Boca Raton, FL, USA: CRC Press, 1999.

[21] Y. Guan and K. Wang, "Translation properties of time scales and almost periodic functions," *Math. Comput. Model.*, vol. 57, no. 5–6, pp. 1165–1174, Mar. 2013.

[22] A. Napolitano, *Generalizations of Cyclostationary Signal Processing: Spectral Analysis and Applications*. Hoboken, NJ, USA: Wiley, 2012.

[23] W. Gardner, "Common pitfalls in the application of stationary process theory to time-sampled and modulated signals," *IEEE Trans. Commun.*, vol. 35, no. 5, pp. 529–534, May 1987.

[24] T. S. Han and S. Verdu, “Approximation theory of output statistics,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[25] H. Amann and J. Escher, *Analysis. I*. Basel, Switzerland: Birkhäuser-Verlag, 1998.
[26] S. Wei, D. L. Goeckel, and P. A. Kelly, “Convergence of the complex envelope of bandlimited OFDM signals,” *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4893–4904, Oct. 2010.
[27] K. Metzger, “On the probability density of intersymbol interference,” *IEEE Trans. Commun.*, vol. 35, no. 4, pp. 396–402, Apr. 1987.
[28] C. Li and N. Elia, “Control approach to computing the feedback capacity for stationary finite dimensional Gaussian channels,” in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 1038–1045.
[29] G. B. Giannakis and E. Serpedin, “Blind identification of ARMA channels with periodically modulated inputs,” *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3099–3104, Nov. 1998.
[30] T. M. Cover and S. Pombra, “Gaussian feedback capacity,” *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 37–43, Jan. 1989.
[31] S. Verdu and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
[32] W. Hirt and J. L. Massey, “Capacity of the discrete-time Gaussian channel with intersymbol interference,” *IEEE Trans. Inf. Theory*, vol. 34, no. 3, p. 38, May 1988.
[33] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2001.
[34] J. H. Ahlberg and E. N. Nilson, “Convergence properties of the spline fit,” *J. Soc. Ind. Appl. Math.*, vol. 11, no. 1, pp. 95–104, Mar. 1963.
[35] W. A. Gardner, *Introduction to Random Processes: With Applications to Signals and Systems*, 2nd ed. New York, NY, USA: McGraw-Hill, 1990.
[36] R. M. Gray, *Entropy and Information Theory*, 2nd ed. New York, NY, USA: Springer, 2011.
[37] J. Jacod and P. Protter, *Probability Essentials*, 2nd ed. Berlin, Heidelberg: Springer-Verlag, 2004.
[38] E. Baktash, M. Karimi, and X. Wang, “Covariance matrix estimation under degeneracy for complex elliptically symmetric distributions,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2474–2484, Mar. 2017.
[39] S. Banerjee and A. Roy, *Linear Algebra and Matrix Analysis for Statistics*. Boca Raton, FL, USA: CRC Press, 2014.
[40] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.
[41] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*. Hoboken, NJ, USA: Wiley, 1974.
[42] G. W. Stewart, “On the continuity of the generalized inverse,” *SIAM J. Appl. Math.*, vol. 17, no. 1, pp. 33–45, Jan. 1969.
[43] A. Dembo, “Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices,” *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 352–355, Mar. 1988.
[44] N. Morača, “Bounds for norms of the matrix inverse and the smallest singular value,” *Linear Algebra Appl.*, vol. 429, no. 10, pp. 2589–2601, Nov. 2008.
[45] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[46] R. A. Rankin, *Introduction to Mathematical Analysis*. New York, NY, USA: Dover, 2007.
[47] P. Kanniappan and S. M. A. Sastry, “Uniform convergence of convex optimization problems,” *J. Math. Anal. Appl.*, vol. 96, no. 1, pp. 1–12, Oct. 1983.
[48] F. Zhang, *Matrix Theory: Basic Results and Techniques*, 2nd ed. New York, NY, USA: Springer, 2011.
[49] F. Zhang and Q. Zhang, “Eigenvalue inequalities for matrix product,” *IEEE Trans. Autom. Control*, vol. 51, no. 9, pp. 1506–1509, Sep. 2006.
[50] M. Spivak, *Calculus*, 3rd ed. Houston, TX, USA: Publish or Perish, 1994.
[51] P. Drabek and J. Milota, *Methods of Nonlinear Analysis: Applications to Differential Equations*, 2nd ed. Basel, Switzerland: Springer, 2013.
[52] L. H. Brandenburg and A. D. Wyner, “Capacity of the Gaussian channel with memory: The multivariate case,” *Bell Syst. Tech. J.*, vol. 53, no. 5, pp. 745–778, May 1974.
[53] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[54] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*. Cambridge, MA, USA: Birkhäuser Boston, 1997.
[55] G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature With Applications*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

Ron Dabora (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University in 1994 and 2000, respectively, and the Ph.D. degree in electrical engineering from Cornell University, USA, in 2007. From 1994 to 2000, he was with the Ministry of Defense of Israel, and from 2000 to 2003, he was with the Algorithms Group, Millimetrix Broadband Networks, Israel. From 2007 to 2009, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, Stanford University, USA. Since 2009, he has been with the School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel, where he is currently an Associate Professor. From 2022 to 2023, he was a Visiting Fellow with the Department of Electrical Engineering, Princeton University, USA. His research interests include network information theory, wireless communications, power line communications, and machine learning. He served as a TPC Member for several international conferences, including WCNC, PIMRC, and ICC. From 2012 to 2014, he served as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. From 2014 to 2019, he served as a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS.

Emeka Abakasanga received the B.Sc. degree in communications engineering from the University of Science and Technology, Ghana, in 2014, and the M.Sc. degree in communications engineering from Loughborough University, U.K., in 2015. He is currently pursuing the Ph.D. degree with the Network Information Laboratory, Ben-Gurion University of the Negev, Israel. His research interests include information theory and signal processing for communications.