

On Finite-Time Mutual Information

Jieao Zhu, Zijian Zhang, Zhongzhichao Wan, and Linglong Dai
Beijing National Research Center for Information Science and Technology (BNRist)
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
E-mails: zja21, zhangzj20, wzzc20@mails.tsinghua.edu.cn; dail@tsinghua.edu.cn

Abstract—Shannon-Hartley theorem can accurately calculate the channel capacity when the signal observation time is infinite. However, the calculation of finite-time mutual information, which remains unknown, is essential for guiding the design of practical communication systems. In this paper, we investigate the mutual information between two correlated Gaussian processes within a finite-time observation window. We first derive the finite-time mutual information by providing a limit expression. Then we numerically compute the mutual information within a single finite-time window. We reveal that the number of bits transmitted per second within the finite-time window can exceed the mutual information averaged over the entire time axis, which is called the exceed-average phenomenon. Furthermore, we derive a finite-time mutual information formula under a typical signal autocorrelation case by utilizing the Mercer expansion of trace class operators, and reveal the connection between the finite-time mutual information problem and the operator theory. Finally, we analytically prove the existence of the exceed-average phenomenon in this typical case, and demonstrate its compatibility with the Shannon capacity.

Index Terms—Finite-time mutual information, exceed-average, Mercer expansion, trace class, operator theory.

I. INTRODUCTION

The Shannon-Hartley theorem [1] has accurately revealed the fundamental theoretical limit of information transmission rate C , which is also called as the Shannon capacity, over a Gaussian waveform channel of a limited bandwidth W . The expression for Shannon capacity is $C = W \log(1 + S/N)$, where S and N denote the signal power and the noise power, respectively. The derivation of Shannon-Hartley Theorem heavily depends on the Nyquist sampling principle [2]. The Nyquist sampling principle, which is also named as the $2WT$ theorem [3], claims that one can only obtain $2WT + o(2WT)$ independent samples within an observation time window T in a channel band-limited to W [4].

Based on the Nyquist sampling principle, the Shannon capacity is derived by multiplying the capacity $1/2 \log(1 + P/N)$ of a Gaussian symbol channel [5, p.249] with $2WT + o(2WT)$ at first, and then dividing the result by T , finally letting $T \rightarrow \infty$. Note that this approximation only holds when $T \rightarrow \infty$. Therefore, the Shannon capacity only asymptotically holds as T becomes sufficiently large. When T is of finite value, the approximation fails to work. Thus, when the observation time T is finite, the Shannon-Hartley Theorem cannot be directly applied to calculate the capacity in a finite-time window. To the best of our knowledge, the evaluation of the finite-time mutual information has not yet been investigated in the literature. It is worth noting that real-

world communication systems transmit signals in a finite-time window, thus evaluating the finite-time mutual information is of practical significance.

In this paper, to fill in this gap, we analyze the finite-time mutual information instead of the traditional infinite-time counterpart, and prove the existence of exceed-average phenomenon within a finite-time observation window¹. Specifically, our contributions are summarized as follows:

- We derive the mutual information expressions within a finite-time observation window by using dense sampling and limiting methods. In this way, we can overcome the continuous difficulties that appear when analyzing the information contained in a continuous time interval. These finite-time mutual information expressions make the analysis of finite-time problems possible.
- We conduct numerical experiments based on the discretized formulas. In the numerical results under a special setting, we reveal the exceed-average phenomenon, i.e., the mutual information within a finite-time observation window exceeds the mutual information averaged over the whole time axis.
- In order to analytically prove the exceed-average phenomenon, we first derive an analytical finite-time mutual information formula based on Mercer expansion [6], where we can find the connection between the mutual information problem and the operator theory [7]. To make the problem tractable, we construct a typical case in which the transmitted signal has certain statistical properties. Utilizing this construction, we obtain a closed-form mutual information solution in this typical case, which leads to a rigorous proof of the exceed-average phenomenon.

Organization: In the rest of this paper, the finite-time mutual information is formulated and evaluated numerically in Section II, where the exceed-average phenomenon is first discovered. Then, in Section III, we derive a closed-form finite-time mutual information formula under a typical case. Based on this formula, in Section IV, the exceed-average phenomenon is rigorously proved. Finally, conclusions are drawn in Section V.

Notations: $X(t)$ denotes a Gaussian Process; $R_X(t_1, t_2)$ denotes the autocorrelation function; $S_X(f), S_X(\omega)$ are the

¹Simulation codes will be provided to reproduce the results presented in this paper: <http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html>.

Power Spectral Density (PSD) of the corresponding process $X(t)$; Boldface italic symbols $\mathbf{X}(t_1^n)$ denotes the column vector generated by taking samples of $X(t)$ on instants $t_i, 1 \leq i \leq n$; Upper-case boldface letters such as Φ denote matrices; $\mathbb{E}[\cdot]$ denotes the expectation; $\mathbb{1}_A(\cdot)$ denotes the indicator function of the set A ; $\mathcal{L}^2([0, T])$ denotes the collection of all the square-integrable functions on window $[0, T]$; i denotes the imaginary unit.

II. NUMERICAL ANALYSIS OF THE FINITE-TIME MUTUAL INFORMATION

In this section, we numerically evaluate the finite-time mutual information. In Subsection II-A, we model the transmission problem by Gaussian processes, and derive the mutual information expressions within a finite-time observation window; In Subsection II-B, we approximate the finite-time mutual information by discretized matrix-based formulas; In Subsection II-C, we reveal the exceed-average phenomenon by numerically evaluating the finite-time mutual information.

A. The Expressions for finite-time mutual information

Inspired by [8], we model the transmitted signal by a zero-mean stationary Gaussian stochastic process, denoted as $X(t)$, and the received signal by $Y(t) := X(t) + N(t)$. The noise process $N(t)$ is also a stationary Gaussian process independent of $X(t)$. The receiver can only observe the signal within a finite-time window $[0, T]$, where $T > 0$ is the observation window span. We aim to find the maximum mutual information that can be acquired within this time window.

To analytically express the amount of acquired information, we first introduce n sampling points inside the time window, denoted by $(t_1, t_2, \dots, t_n) := t_1^n$, and then let $n \rightarrow \infty$ to approximate the finite-time mutual information. By defining $\mathbf{X}(t_1^n) \equiv (X(t_1), X(t_2), \dots, X(t_n))$ and $\mathbf{Y}(t_1^n) \equiv (Y(t_1), Y(t_2), \dots, Y(t_n))$, the mutual information on these n samples can be expressed as

$$I(t_1^n) = I(\mathbf{X}(t_1^n); \mathbf{Y}(t_1^n)), \quad (1)$$

and the finite-time mutual information is defined as

$$I(T) = \lim_{n \rightarrow \infty} \sup_{\{t_1^n\} \subset [0, T]} I(t_1^n). \quad (2)$$

Then, the transmission rate $C(T)$ can be defined as $C(T) = I(T)/T$. After these definitions, we can then define the limit mutual information as $C(\infty) = \lim_{T \rightarrow \infty} C(T)$ by letting $T \rightarrow \infty$.

B. Discretization

Without loss of generality, we fix the sampling instants uniformly onto fractions of T : $t_i = (i-1)T/n, 1 \leq i \leq n$. Since the random vectors $\mathbf{X}(t_1^n)$ and $\mathbf{Y}(t_1^n)$ are samples of a Gaussian process, they are both Gaussian random vectors with mean zero and covariance matrices \mathbf{K}_X and \mathbf{K}_Y , where

$\mathbf{K}_X, \mathbf{K}_Y \in \mathbb{R}^{n \times n}$ are symmetric positive-definite matrices defined as

$$\begin{aligned} (\mathbf{K}_X)_{i,j} &= R_X(t_i, t_j) := \mathbb{E}[X(t_i)X(t_j)], \\ (\mathbf{K}_Y)_{i,j} &= R_Y(t_i, t_j) := \mathbb{E}[Y(t_i)Y(t_j)]. \end{aligned} \quad (3)$$

Note that $Y(t)$ is the independent sum of $X(t)$ and $N(t)$, thus the autocorrelation functions satisfy $R_Y(t_1, t_2) = R_X(t_1, t_2) + R_N(t_1, t_2)$, and similarly the covariance matrices satisfy $\mathbf{K}_Y = \mathbf{K}_X + \mathbf{K}_N$.

The mutual information $I(t_1^n)$ is defined as $I(t_1^n) = h(\mathbf{Y}(t_1^n)) - h(\mathbf{Y}(t_1^n) | \mathbf{X}(t_1^n)) = h(\mathbf{Y}(t_1^n)) - h(\mathbf{N}(t_1^n))$, where $h(\cdot)$ denotes the differential entropy. Utilizing the entropy formula for n -dimensional Gaussian vectors [5], we obtain

$$I(t_1^n) = \frac{1}{2} \log \left(\frac{\det(\mathbf{K}_X + \mathbf{K}_N)}{\det(\mathbf{K}_N)} \right). \quad (4)$$

C. Numerical Analysis

In order to study the properties of mutual information $I(t_1^n)$ as a function of n , we set the autocorrelation function of the signal process $X(t)$ and noise process $N(t)$ to the following special case

$$\begin{aligned} R_X(t_1, t_2) &= \text{sinc}(10(t_1 - t_2)), \\ R_N(t_1, t_2) &= \exp(-|t_1 - t_2|), \end{aligned} \quad (5)$$

where $\text{sinc}(x) := \sin(\pi x)/(\pi x)$, and the corresponding PSDs are $S_X(f) = 0.1 \times \mathbb{1}_{\{-5 \leq f \leq 5\}}$ and $S_N(f) = \frac{2}{1+(2\pi f)^2}$. In order to compare the finite-time mutual information with the averaged mutual information, the Shannon limit with colored noise spectrum $S_N(f)$ is utilized, which is a generalized version of the well-known formula $C = W \log(1 + S/N)$. The averaged mutual information of colored noise PSD [5] is expressed as

$$C_{\text{av}} := \frac{1}{2} \int_{-\infty}^{+\infty} \log \left(1 + \frac{S_X(f)}{S_N(f)} \right) df. \quad (6)$$

Then, plugging (5) into (6) yields the numerical result for C_{av} .

We calculate the finite-time transmission rate $C(T)$ and the average mutual information C_{av} against the number of samples n within the observation window $[0, T]$. The numerical results are collected in Fig. 1. It is shown that $I(t_1^n)$ is an increasing function of n , and for fixed values of T , the approximated finite-time mutual information $I(t_1^n)$ tends to a finite limit under the correlation assumptions given by (5). The most amazing observation is that, it is possible to obtain more information within finite-time window $[0, T]$ than the prediction TC_{av} given by averaging the mutual information along the entire time axis (6). We call this phenomenon the exceed-average phenomenon.

III. A CLOSED-FORM FINITE-TIME MUTUAL INFORMATION FORMULA

In this section, we first introduce the Mercer expansion in Subsection III-A as a basic tool for our analysis. Then we derive the series representation of the finite-time mutual information, and the corresponding power constraint in Subsection III-B, under the assumption of AWGN noise.

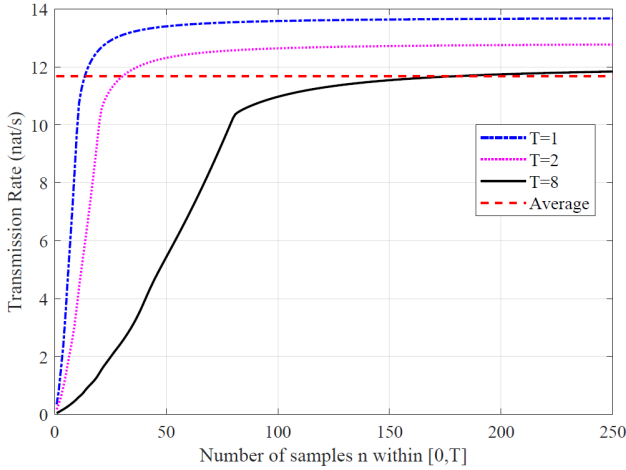


Fig. 1. A first glance to the exceed-average phenomenon. The red dashed horizontal line is the Shannon limit, and the $T = 1, 2, 8$ curves illustrate the dependence of $C(t_1^n)$ on n .

A. The Mercer Expansion

By analyzing the structure of the autocorrelation functions, it is possible to obtain $I(t_1^n)$ and $I(T)$ directly. In fact, if we know the Mercer expansion [6] of the autocorrelation function $R_X(t_1, t_2)$ on interval $[0, T]$, then we can calculate $h(\mathbf{X}(t_1^n))$ more easily [9]. In the following discussion, we assume the Mercer expansion of the source autocorrelation function $R_X(t_1, t_2)$ to be in the following form

$$\lambda_k \phi_k(t_1) = \int_0^T R_X(t_1, t_2) \phi_k(t_2) dt_2; k > 0, k \in \mathbb{N}, \quad (7)$$

where the eigenvalues are positive: $\lambda_k > 0$, and the eigenfunctions form an orthonormal set:

$$\int_0^T \phi_i(t) \phi_j(t) dt = \delta_{ij}. \quad (8)$$

The Mercer's theorem [6] ensures the existence and uniqueness of the eigenpairs $(\lambda_k, \phi_k(t))_{k=1}^{\infty}$, and furthermore, the kernel itself can be expanded under the eigenfunctions:

$$R_X(t_1, t_2) = \sum_{k=1}^{+\infty} \lambda_k \phi_k(t_1) \phi_k(t_2). \quad (9)$$

B. Finite-Time Mutual Information Formula

Based on Mercer expansion, we obtain a closed-form formula in the following **Theorem 1**.

Theorem 1 (Source expansion, AWGN noise): Suppose the information source, modeled by the transmitted process $X(t)$, has autocorrelation function $R_X(t_1, t_2)$. An AWGN noise of PSD $n_0/2$ is imposed onto $X(t)$, resulting in the received process $Y(t)$. The Mercer expansion of $R_X(t_1, t_2)$ on $[0, T]$ is given by (7), satisfying (8). Then the finite-time mutual information $I(T)$ within the observation window $[0, T]$ between the processes $X(t)$ and $Y(t)$ can be expressed as

$$I(T) = \frac{1}{2} \sum_{k=1}^{+\infty} \log \left(1 + \frac{\lambda_k}{n_0/2} \right). \quad (10)$$

Proof: The Mercer expansion decomposes the AWGN channel into an infinite number of independent parallel sub-channels, each with signal power λ_k and noise variance $n_0/2$. Thus, accumulating all the mutual information values of these subchannels yields the finite-time mutual information $I(T)$. ■ From **Theorem 1** we can conclude that the finite-time mutual information of AWGN channel is uniquely determined by the Mercer spectra λ_k of $R_X(t_1, t_2)$ within $[0, T]$. However, it remains unknown whether the series representation (10) converges. In fact, the convergence is closely related to the signal power, which is calculated in the following **Lemma 1**.

Lemma 1 (Operator Trace Coincide with Power Constraint): Given stationary Gaussian process $X(t)$ with mean zero and autocorrelation $R_X(t_1, t_2)$. The Mercer expansion of $R_X(t_1, t_2)$ on $[0, T]$ is given by (7), satisfying (8). The Mercer operator $M(\cdot) : \mathcal{L}^2([0, T]) \rightarrow \mathcal{L}^2([0, T])$ is defined by the integral $(M\phi)(s) = \int_0^T R_X(s, \tau) \phi(\tau) d\tau$. Then the sum of all the eigenvalues λ_k of operator M is equal to the signal energy PT within $[0, T]$:

$$\text{tr}(M) := \sum_{k=1}^{+\infty} \lambda_k = PT, \quad (11)$$

where $P = R_X(0, 0)$.

Proof: Integrating both sides of (9) on the interval $[0, T]$ gives the conclusion immediately. ■

Remark 1: From the above **Lemma 1**, we can conclude that the sum of λ_k is finite when T is finite. It can be immediately derived that $I(T) < \infty$, since $\log(1+x) \leq x$.

Remark 2: The finite-time mutual information formula (10) is closely related to the operator theory [7] in functional analysis. The sum of all the eigenvalues λ_k is called the operator trace in linear operator theory. As is mentioned in **Lemma 1**, the autocorrelation function $R_X(t_1, t_2)$ can be treated as a linear operator M on $\mathcal{L}^2([0, T])$.

IV. PROOF OF THE EXISTENCE OF EXCEED-AVERAGE PHENOMENON

In this section, we first give a proof of the existence of the exceed-average phenomenon in a typical case, then we discuss the compatibility with the Shannon-Hartley theorem.

A. Closed-Form Mutual Information in a Typical Case

To prove the exceed-average phenomenon, we only need to show that the finite-time mutual information is greater than the averaged mutual information in a typical case. Let us consider a finite-time communication scheme with a finitely-powered stationary transmitted signal autocorrelation², which is specified as

$$R_X(t_1, t_2) = R_X(\tau) = P \exp(-\alpha|\tau|), \quad (12)$$

where $\tau = t_1 - t_2$, under AWGN channel with noise PSD being $n_0/2$. The power of signal $X(t)$ is $P = R_X(0)$. According to **Lemma 1**, the trace of the corresponding

²The signal autocorrelation $R_X(\tau)$ is often observed in many scenarios, such as passing a signal with white spectrum through an RC lowpass filter.

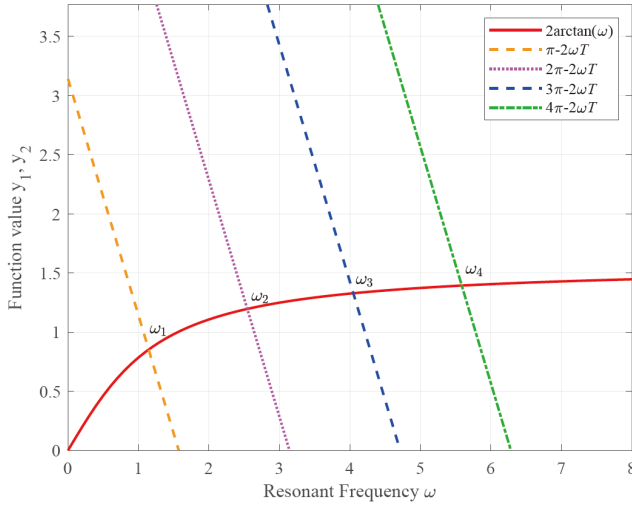


Fig. 2. Images of $2\arctan(\omega/\alpha)$ and $k\pi - \omega T$ with $T = 2, \alpha = 1$ and $k = 1, 2, 3, 4$. The desired resonant frequencies ω_k can be read from the horizontal coordinates of the intersection points of the curve $2\arctan(\omega/\alpha)$ and the parallel lines.

Mercer operator $M(\cdot)$ is finite. Then the finite-time mutual information given by **Theorem 1** is also finite, as is shown in *Remark 1*. Finding the Mercer expansion is equivalent to finding the eigenpairs $(\lambda_k, \phi_k(t))_{k=1}^{\infty}$. The eigenpairs are determined by the following characteristic integral equation [10]:

$$\lambda_k \phi_k(s) = \int_0^s P e^{-\alpha(s-t)} \phi_k(t) dt + \int_s^T P e^{-\alpha(t-s)} \phi_k(t) dt. \quad (13)$$

Differentiating both sides of (13) twice with respect to s yields the boundary conditions and the differential equation that λ_k, ϕ_k must satisfy:

$$\begin{aligned} \lambda_k \phi_k''(s) &= (\alpha^2 \lambda_k - 2\alpha P) \phi_k(s), 0 < s < T, \\ \phi_k'(0) &= \alpha \phi_k(0), \\ \phi_k'(T) &= -\alpha \phi_k(T). \end{aligned} \quad (14)$$

Let $\omega_k > 0$ denote the resonant frequency of the above harmonic oscillator equation, and let $\phi_k(t) = A_k \cos(\omega_k t) + B_k \sin(\omega_k t)$ be the sinusoidal form of the eigenfunction. Using the boundary conditions we obtain

$$\begin{aligned} B_k \omega_k &= \alpha A_k, \\ B_k \omega_k \cos(\omega_k T) - A_k \omega_k \sin(\omega_k T) &= -\alpha (A_k \cos(\omega_k T) + B_k \sin(\omega_k T)). \end{aligned} \quad (15)$$

To ensure the existence of solution to the homogeneous linear equations (15) with unknowns A_k, B_k , the determinant must be zero. Exploiting this condition, we find that ω_k naturally satisfy the transcendental equation $\tan(\omega_k T) = (2\omega_k \alpha) / (\omega_k^2 - \alpha^2)$. By introducing an auxiliary variable $\theta_k = \arctan(\omega_k / \alpha) \in [0, \pi/2]$, this transcendental equation can be simplified as $\tan(\omega_k T) = -\tan(2\theta_k)$, i.e., there exists positive integer m such that $2\arctan(\omega_k / \alpha) = m\pi - \omega_k T$. The integer m can be chosen to be equal to k . From the

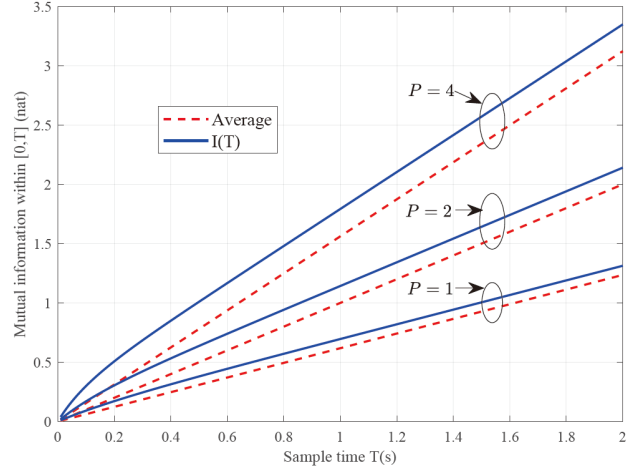


Fig. 3. Theoretical verification of the exceed-average effect. The blue lines represent the finite-time mutual information $I(T)$. The red lines are the average mutual information TC_{av} , calculated from (6). All the curves are evaluated under hypothesis (12), where $P = 1, 2, 4, n_0 = 1$ and $\alpha = 1$.

function images of $2\arctan(\omega/\alpha)$ and $k\pi - \omega T$ (Fig. 2), we can determine ω_k , and then λ_k . To sum up, the solution to the characteristic equation (13) are collected into (16) as follows.

$$\begin{aligned} 2\arctan(\omega_k/\alpha) &= k\pi - \omega_k T, \\ \lambda_k &= \frac{2\alpha P}{\alpha^2 + \omega_k^2}, \\ \phi_k(t) &= \frac{1}{Z_k} (\omega_k \cos(\omega_k t) + \alpha \sin(\omega_k t)), \end{aligned} \quad (16)$$

where Z_k denotes the normalization constants of $\phi_k(t)$ on $[0, T]$ to ensure orthonormality.

Equation (16) gives all the eigenpairs $(\lambda_k, \phi_k)_{k=1}^{\infty}$, from which we can calculate $I(T)$ by applying **Theorem 1**. As for the Shannon limit C_{sh} , by applying (6) and evaluating the integral with [11] we can obtain

$$C_{av} = C_{sh} = \frac{1}{2} \left(\sqrt{\alpha^2 + \frac{4P\alpha}{n_0}} - \alpha \right). \quad (17)$$

After all the preparation works above, we can rigorously prove that $C(T) > C_{av}$ under the typical case of (12), as long as the transmission power P is smaller than a constant δ . The following **Theorem 2** proves this result.

Theorem 2 (Existence of exceed-average phenomenon in a typical case): Suppose $X(t)$ and $Y(t)$ are specified according to (12). The eigenpairs are determined by (16). Then, for any fixed positive values of T, n_0 and α , there exists $\delta > 0$ such that the exceed-average inequality

$$C(T) := \frac{1}{2T} \sum_{k=1}^{+\infty} \log \left(1 + \frac{\lambda_k}{n_0/2} \right) > C_{av} \quad (18)$$

holds strictly for arbitrary $0 < P < \delta$.

Proof: See Appendix A. ■

To support the above theoretical analysis, numerical experiments on $I(T)$ are conducted based on evaluations of (16)

and (17). As shown in Fig. 3, it seems that we can always harness more mutual information in a finite-time observation window than the averaged mutual information. This fact is somehow unsurprising because the observations $\mathbf{Y}(t_1^N)$ inside the window $[0, T]$ can always eliminate some extra uncertainty outside the window due to the autocorrelation of $X(t)$.

B. Compatibility with the Shannon-Hartley theorem

Though the exceed-average effect do imply mathematically that the mutual information within a finite-time window is higher than the average mutual information C_{av} (which coincides with the Shannon limit (6) in this case), in fact, it is still impossible to construct a long-time stable data transmission scheme above the Shannon capacity by leveraging this effect. So the exceed-average phenomenon does not contradict the Shannon-Hartley Theorem. Placing additional observation windows cannot increase the average information rate, because the posterior process $Y(t)|\mathbf{Y}(t_1^N)$ does not carry as much information as the original one, causing a rate reduction in the later windows. It is expected that, the finite-time mutual information would ultimately decrease to the average mutual information as the total observation time tends to infinity (i.e., $C(\infty) = C_{av}$), and the analytical proof, as well as the achievability of this finite-time mutual information, is still worth investigation in future works.

V. CONCLUSIONS

In this paper, we provided rigorous proofs of the existence of the exceed-average phenomenon under typical autocorrelation settings of the transmitted signal process and the noise process. Our discovery of the exceed-average phenomenon provided a generalization of Shannon's renowned formula $C = W \log(1 + S/N)$ to the practical finite-time communications. Since the finite-time mutual information is a more precise estimation of the capacity limit, the optimization target may shift from the average mutual information to the finite-time mutual information in the design of practical communication systems. Thus, it may have guiding significance for the performance improvement of modern communication systems. In future works, general proofs of $C(T) > C_{av}$, independent of the concrete autocorrelation settings, still require further investigation. Moreover, we need to answer the question of how to exploit this exceed-average phenomenon to improve the communication rate, and whether this finite-time mutual information is achievable by a certain coding scheme. In addition, although we have discovered numerically that the finite-time mutual information agrees with the Shannon capacity when $T \rightarrow \infty$, an analytical proof of this result is required in the future.

APPENDIX A PROOF OF THEOREM 2

Plugging (17) into the right-hand side of (18), and differentiate both sides w.r.t P . Notice that if $P = 0$, then both sides of (18) are equal to 0. Thus, we only need to prove that

the derivative of left-hand side is strictly larger than that of right-hand side within a small interval $P \in (0, \delta)$:

$$\frac{1}{2} \sum_{k=1}^{+\infty} \left(\frac{1}{1 + \frac{2\lambda_k}{n_0}} \frac{2\lambda_k}{n_0 P} \right) > \frac{T}{n_0} \frac{1}{\sqrt{1 + 4P/(n_0\alpha)}}. \quad (19)$$

Multiply both sides of (19) by n_0 and define $\mu_k := \lambda_k/(PT)$, and then from **Lemma 1** we obtain $\sum_k \mu_k = 1$. In this way, (19) is equivalent to

$$\sum_{k=1}^{+\infty} \frac{\mu_k}{1 + \frac{2\lambda_k}{n_0}} > \frac{1}{\sqrt{1 + 4P/(n_0\alpha)}}. \quad (20)$$

Since $\varphi(x) := 1/(1 + 2x/n_0)$ is convex on $(0, +\infty)$, by applying Jensen's inequality to the left-hand side of (20), we only need to prove that

$$\frac{1}{1 + \frac{2}{n_0} \sum_k \lambda_k \mu_k} > \frac{1}{\sqrt{1 + 4P/(n_0\alpha)}}. \quad (21)$$

From the definition of μ_k we can derive that $\lambda_k \mu_k = \lambda_k^2/(PT)$. So we go on to calculate $\sum_k \lambda_k^2$. That is equivalent to calculate $\text{tr}(M^2)$, where M^2 corresponds to the integral kernel:

$$K_{M^2}(t_1, t_2) := \int_0^T P^2 \exp(-\alpha|t_1 - s|) \exp(-\alpha|t_2 - s|) ds. \quad (22)$$

Evaluating the kernel K_{M^2} on the diagonal $t = t_1 = t_2$, and integrating this kernel on the diagonal of $[0, T]^2$ gives $\sum_k \lambda_k^2$, i.e., $\text{tr}(M^2)$:

$$\begin{aligned} \sum_k \lambda_k^2 &= \frac{P^2}{2\alpha} \int_0^T \left(2 - e^{-2\alpha t} - e^{-2\alpha(T-t)} \right) dt, \\ &= \frac{P^2}{\alpha} \left(T - \frac{1}{2\alpha} (1 - e^{-2\alpha T}) \right). \end{aligned} \quad (23)$$

By substituting (23) into (21), we just need to prove that

$$\sqrt{1 + 4P/(n_0\alpha)} > 1 + \frac{2P}{n_0\alpha} \left(1 - \frac{1 - e^{-2\alpha T}}{2\alpha T} \right). \quad (24)$$

Define the dimensionless number $x = 2P/(n_0\alpha)$. Since the function $\psi(x) := (1 - \exp(-x))/x$ is strictly positive and less than 1 at $x > 0$, we can conclude that, there exists a small positive $\delta > 0$ such that (24) holds for $0 < P < \delta$. The number δ can be chosen as

$$\delta = \frac{n_0\alpha\psi(2\alpha T)}{(1 - \psi(2\alpha T))^2} > 0, \quad (25)$$

which implies that (19) holds for any $0 < P < \delta$. Thus, integrating (19) on both sides from $p = 0$ to $p = P$, $P < \delta$ gives rise to the conclusion (18), which completes the proof of **Theorem 2**.

ACKNOWLEDGMENT

This work was supported in part by the National Key Research and Development Program of China (Grant No.2020YFB1807201), in part by the National Natural Science Foundation of China (Grant No. 62031019), and in part by the European Commission through the H2020-MSCA-ITN META WIRELESS Research Project under Grant 956256.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *The Bell Syst. Techn. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [2] H. Landau, "Sampling, data transmission, and the nyquist rate," *Proc. IEEE*, vol. 55, no. 10, pp. 1701–1706, Oct. 1967.
- [3] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, Apr. 1928.
- [4] D. Slepian, "On bandwidth," *Proc. IEEE*, vol. 64, no. 3, pp. 292–300, Mar. 1976.
- [5] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
- [6] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Royal Soc.*, vol. 209, pp. 4–415, 1909.
- [7] K. Zhu, *Operator theory in function spaces*. American Mathematical Soc., 2007, no. 138.
- [8] A. Balakrishnan, "A note on the sampling principle for continuous signals," *IRE Trans. Inf. Theory*, vol. 3, no. 2, pp. 143–146, Jun. 1957.
- [9] J. Barrett and D. Lampard, "An expansion for some second-order probability distributions and its application to noise problems," *IRE Trans. Inf. Theory*, vol. 1, no. 1, pp. 10–15, Mar. 1955.
- [10] D. Cai and P. S. Vassilevski, "Eigenvalue problems for exponential-type kernels," *Comput. Methods in Applied Math.*, vol. 20, no. 1, pp. 61–78, Jan. 2020.
- [11] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic press, 2014.