Joint Data Compression and Time-Delay Estimation for Distributed Systems via Extremum Encoding

Amir Weiss, *Senior Member, IEEE*, Yuval Kochman, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—Motivated by the ubiquity of mobile devices and their potential capabilities as distributed systems (e.g., for localization), we consider a time-delay estimation (TDE) problem in which there are two non-colocated sensors and communication constraints between them. When the communication bandwidth is particularly limited, there is a need for compression techniques that are specifically tailored for the TDE application. For the discrete-time version of this problem, we propose such a joint compression-estimation strategy based on what we term "extremum encoding", whereby the (time-) index of the maximum of the observed signal in a finite observation window is sent from one sensor to another. Subsequent joint processing of the encoded message with the locally observed, time-delayed, noisy signal gives rise to our proposed time-delay "maximum-index"-based estimator. We analyze the performance of the proposed scheme in the asymptotic regime of large message size and delay spread, but with their ratio fixed. We derive the error probability exponent for this estimator, and its consistency. We validate the analysis via simulations, and further demonstrate the performance gains over traditional alternatives.**

***Index Terms*—Time-delay estimation, compression, distributed estimation, compression for estimation, max-index estimator.**

I. Introduction

TIME-DELAY estimation (TDE) is a fundamental problem that is found at the core of numerous important applications in various scientific fields and physical domains (i.e., acoustic, optics, radiofrequency). Examples, among others, include localization, tracking, communication, sensor calibration, medical imaging and more [1], [2], [3], [4]. In this respect, it is perhaps one of the most important problems in signal processing, and as such, it has been extensively studied in past decades. For a collection of important results, which nevertheless by no means serves as an exhaustive survey, see [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17].

Notwithstanding the foregoing, the aspect of *data compression for TDE*, namely compression whose sole objective is accurate TDE (in contrast to the traditional objective of signal reconstruction), which is highly motivated by recent technological developments [18], has only been comparatively sparsely addressed. A brief, but relatively comprehensive summary of past efforts on this front is given in what follows.

A. Data Compression for TDE: Motivation and Prior Art

As mentioned above, TDE is a necessary (algorithmic) building block in many systems for various applications. A classical example is that of passive radar/sonar, which served as a driver for some of the most important theoretical findings on this problem (e.g., the best attainable performance [8], [9]). However, in these classical settings, which involve two data-acquiring sensors, it is typically assumed that the central computing unit has access to both of the received signals. This is a fairly solid assumption when the two sensors are co-located or when there are (essentially) no constraints on the relevant communication links for sending the received signals.

In contrast, for some of the modern emerging applications, this is no longer the case. Consider, for example, the problem of passive acoustic indoor localization [19], building on power- and communication-limited "smart" devices. We envision that for such applications, these devices would opportunistically be used as *ad-hoc* sensors that could measure an acoustic signal and—with limited resources—convey a corresponding message for the purpose of TDE (e.g., for range estimation).

In such scenarios, it is not only desirable, but is necessary to reduce as much as possible the resources requirements on the spatially distributed devices that serve as the sensors. Note that, clearly, the sensors in this case are not co-located. Moreover, the assumption of an essentially unlimited communication link is weak in some cases, and unrealistic in others. Similar constraints arise in sensor networks [20], [21], which by design consist of a large number of small, low-power, untethered devices that measure a common signal for the collective purpose of some inference task (e.g., [22]).

This motivation, that is relevant for other systems as well (e.g., mobile platforms), has led to several works focused on compression techniques that are specifically designed for TDE.

In the early 1980s, Matthiesen and Miller [23] considered the fundamental two-receiver passive TDE problem, and proposed

Received 7 October 2024; revised 27 February 2025 and 19 April 2025; accepted 23 April 2025. Date of publication 7 May 2025; date of current version 27 June 2025. The associate editor coordinating the review of this article and approving it for publication was Cihan Tepedelenlioglu. *(Corresponding author: Amir Weiss.)*

Amir Weiss is with the Alexander Kofkin Faculty of Engineering, Bar-Ilan University, Ramat Gan 5290002, Israel (e-mail: amir.weiss@biu.ac.il).

Yuval Kochman is with the Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel.

Gregory W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

This article has supplementary downloadable material available at https://doi.org/10.1109/TSP.2025.3567902, provided by the authors.

Digital Object Identifier 10.1109/TSP.2025.3567902

two "simple, robust data transfer reduction techniques" for the approximated computation of the standard cross-correlation-based estimator (CCE). They considered a rate-distortion (RD) benchmark, wherein the signal to be sent is RD-optimally compressed for the square error loss with respect to the signal itself. The authors' approach in [23] therefore differs from a potentially more TDE-oriented framework, where the bits sent aim solely to maximize TDE fidelity, and may thus be oblivious to most of the features of the received signal that carry less information about the desired time-delay.

Two consistent lines of work on this more involved variant of the TDE problem are due to Fowler et al. [24], [25], [26], [27], [28], [29], [30] and Vasudevan et al. [31], [32], [33], [34]. In the first one, Fowler's general approach for compression is to maximize the Fisher information (FI) for TDE, which is derived for *uncompressed* data, for a given number of bits. In this case, the objective to be maximized (namely, the FI) is not necessarily informative for compressed observations. In the second line of work, a so-called maximum mutual information (MaxMI) quantizer is developed, where an approximation of the maximum *a posteriori* objective function for the quantized signals is maximized. In this approach, the MaxMI quantizer relies on available training data in order to estimate the signals' distributions, which are then used for numerical computation of the quantizer. In particular, the MaxMI is designed (only) for temporally white signals, and requires the noise level to be known or estimated. In both of these approaches, the quantizers do not admit closed-form expressions, and specifically, it is not clear how to optimally choose the number of quantization levels.

B. Contributions

Motivated by the potential applications and the limitations of previously proposed methods mentioned above, and inspired by information-theoretic ideas of using the asymptotic statistical properties of a maximum random variable (RV) presented in recent work by Hadar and Shayevitz [35] and Kochman and Wang [36], we propose a new method for distributed TDE with communication constraints. While the present work is focused on a specific signal model, the conceptual insight stemming from it is of both theoretical and practical value for the design of compression methods for TDE in other different settings.

Specifically, our main contributions are the following:

- *A method for joint compression-TDE for distributed systems*: We propose a compression-estimation method that is superior to existing alternatives in terms of the inherent trade-off between compression rate and estimation fidelity. Our proposed method is based on sending only the index of the maximal observed value in some prescribed range, a notion of what we term "extremum encoding". Consequently, both our compressor and estimator admit closed-form expressions, they are computationally simple, and can be interpreted intuitively. Moreover, our method is universal, in the sense that it is agnostic to the noise level, hence it does not require prior knowledge, unlike other previously proposed methods (e.g., [32]).
- *Performance Analysis*: We analyze the error probability of our proposed estimator, and derive its error exponent (Theorem 1) for white Gaussian signals. As a simple implication, it follows that it is consistent in the communication sense (i.e., with respect to the number of bits sent). Our analysis can be used to guide the design of systems that involve distributed TDE (e.g., networks of smart devices). We further analyze our proposed method for signals that are arbitrarily correlated, and derive the asymptotic error exponent for this case. We also provide exponentially tight lower and upper bounds on the estimation error absolute moments (e.g., mean-absolute error and mean-square error).

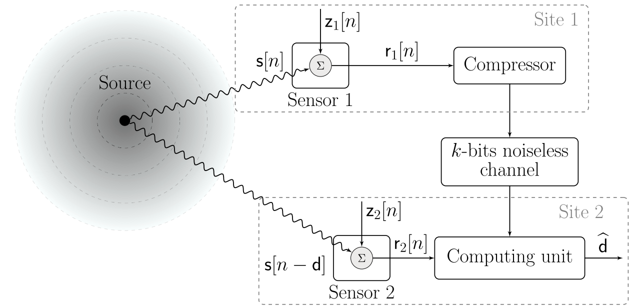

Fig. 1. A simplified illustration of the distributed discrete-TDE problem.

The rest of the paper is organized as follows. The remainder of this section is devoted to an outline of our notation. In Section II we present the problem of TDE for distributed systems depicted in Fig. 1, and formulate the joint compression-TDE problem considered in this work, where one of the sensors is co-located with the computing unit. In Section III we present our proposed compression-estimation strategy, followed by interpretations of its operation, performance analysis, and a discussion on computational complexity. We provide simulation results in Section IV that corroborate our analysis and demonstrate that our method is superior to relevant existing alternatives. Concluding remarks are given in Section V.

C. Notation

We use lowercase letters with standard font and sans-serif font, e.g., x and x, to denote deterministic and RVs,[1] respectively. Similarly, we use $\mathbf{x}$ and $\mathbf{\mathsf{x}}$ for deterministic and random vectors, respectively. The uniform distribution over a set $\mathcal{D}$ is denoted as $\mathcal{U}(\mathcal{D})$, the standard normal distribution as $\mathcal{N}(0,1)$, and $\overset{\text{iid}}{\sim}$ for an independent, identically distributed (iid) random process. We use $\mathbb{E}[\cdot]$ and $\mathbb{V}\text{ar}(\cdot)$ to denote the expectation and variance operators, respectively, and $\mathbb{P}(\mathcal{A})$ for the probability of the event $\mathcal{A}$. The entropy of a discrete RV x is denoted by $H(\mathsf{x})$. The indicator function $\mathbb{1}_{\mathcal{A}}$ is equal to one if $\mathcal{A}$ is true and zero otherwise. The natural logarithm is denoted by $\log(\cdot)$, and $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathrm{d}t$ is the standard Q-function. We use the standard order notation; let f and g be positive functions with discrete or continuous domain. We write $f = o(g)$ to indicate that $\lim f/g = 0$, and $f = \mathcal{O}(g)$ to indicate that $\limsup f/g < \infty$, where the arguments and implied limits should be clear from the context.

[1]In some cases, Greek letters would be used as RVs as well.

II. Joint Compression and Time-Delay Estimation

We consider a simplified signal model (as in [31], [32], [33], [34]) for the classical TDE problem, where the delay is in discrete-time. In particular, consider the observed discrete-time signals[2]

$$\begin{aligned} \mathsf{r}_1[n] &= \mathsf{s}[n] + \mathsf{z}_1[n], && \text{(sensor 1)} \\ \mathsf{r}_2[n] &= \mathsf{s}[n-\mathsf{d}] + \mathsf{z}_2[n], && \text{(sensor 2)} \end{aligned} \tag{1}$$

defined for all $n \in \mathbb{Z}$, where

- $\mathsf{r}_1[n], \mathsf{r}_2[n]$ are the signals observed by the first and second sensors, respectively, which are assumed to be time-synchronized;
- $\mathsf{s}[n] \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ is the common (unobservable) signal to both sensors with a relative time-delay $\mathsf{d} \in \mathcal{D}$, where $\mathcal{D} \triangleq \{-d_m, \ldots, d_m\}$ is the "uncertainty interval" (or the "delay spread") with cardinality $D \triangleq |\mathcal{D}| = 2d_m + 1$, and $d_m \in \mathbb{N}$ is the maximum (absolute) delay; and
- $\mathsf{z}_1[n] \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_1^2)$ and $\mathsf{z}_2[n] \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_2^2)$ are statistically independent white Gaussian noise processes, that are also statistically independent of $\mathsf{s}[n]$.

We assume for simplicity that the parameters σ_1^2 and σ_2^2 are known. However, it will become clear that our proposed scheme is agnostic to them.

For ease of notation, we begin with the following proposition regarding the simplification of the signal model, whose proof appears in Appendix A.

Proposition 1: Model (1) is statistically equivalent (up to scaling coefficients) to the observation model

$$\begin{aligned} &\mathsf{x}[n], && \text{(sensor 1, "encoder")} \\ &\mathsf{y}[n] = \rho \mathsf{x}[n-\mathsf{d}] + \bar{\rho}\mathsf{z}[n], && \text{(sensor 2, "decoder")} \end{aligned} \tag{2}$$

where $\mathsf{x}[n] \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $\mathsf{z}[n] \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ are statistically independent, $\rho \in (0,1]$ is the (Pearson) correlation coefficient between $\mathsf{x}[n]$ and $\mathsf{y}[n+\mathsf{d}]$ that is related to the signal-to-noise ratios (SNRs) of (1), i.e., to $1/\sigma_1^2$ and $1/\sigma_2^2$, and $\bar{\rho} \triangleq \sqrt{1-\rho^2}$.

Notice that equivalence up to scaling is sufficient, since any scheme may apply these factors, which are a function of the known parameters. Therefore, we will henceforth work with model (2), and accordingly, we shall refer to the quantity $\texttt{SNR} \triangleq \frac{\rho^2}{\bar{\rho}^2} = \frac{\rho^2}{1-\rho^2}$ as the SNR.[3] For the relation between $\texttt{SNR}$ and $1/\sigma_1^2, 1/\sigma_2^2$ (SNRs in model (1)), see Appendix A.

In (2), one sensor (the "encoder") observes[4] $\mathsf{x}[n]$ and needs to produce a message $\mathbf{m} \in \{0,1\}^{k \times 1}$ of length $k \in \mathbb{N}$ bits to be sent to the central computing unit, where the second sensor (the "decoder") is located. The latter observes $\mathsf{y}[n]$, a noisy version of (the ρ-scaled) $\mathsf{x}[n]$, delayed by d samples. The goal of the central computing unit is to estimate[5] d based on the signal $\mathsf{y}[n]$ and the received message $\mathbf{m}$. The problem setup is illustrated in Fig. 2.

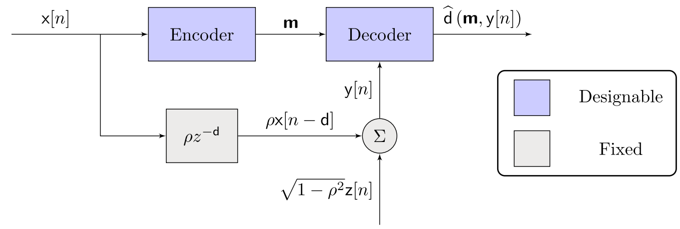

Fig. 2. Equivalent representation of the problem considered in this work. The encoder observes the signal $\mathsf{x}[n]$ and generates a message of length k bits $\mathbf{m} \in \{0,1\}^{k\times 1}$. The decoder observes the signal $\mathsf{y}[n]$ and receives the message $\mathbf{m}$, from which it constructs the estimator $\widehat{\mathsf{d}}(\mathbf{m}, \mathbf{y})$ of d, where $\mathbf{y} \triangleq [\mathsf{y}[1] \,\ldots\, \mathsf{y}[N]]^{\mathsf{T}}$.

Without communication constraints between the sensors, one can let $k \to \infty$, and $\mathbf{m}$[6] would simply be the observed signal $\mathsf{x}[n]$. In that case, assuming for example that d is deterministic unknown, the (well-known) maximum likelihood estimator (MLE) would correspond to a CCE [37]. However, when k is finite and limited, it is no longer trivial (or even clear) how to develop a good[7] joint compression-estimation strategy. After formulating the problem, we shall propose such a strategy, that is inherently different from any of the other previously proposed methods mentioned above in Section I-A. For clarity, we elaborate on this difference in Section III-A only after presenting our method at the outset of Section III.

We note in passing that the Gaussian signal model, while only approximating the underlying statistical models of signals encountered in practice, is still a well-adopted one in the literature (e.g., [38], [39]). Beyond lending itself to amenable analysis (e.g., [40]), which frequently leads to insightful observations, the Gaussian model oftentimes yields robust, functioning algorithmic solutions even for non-Gaussian signals.[8]

A. Problem Formulation

Formally, we consider the problem above when described in a Bayesian setting, where we assume a simple non-informative prior $\mathsf{d} \sim \mathcal{U}(\mathcal{D})$, as follows. For a given number of bits k and an observation interval of length N samples (possibly letting $N \to \infty$), consider the minimum attainable *risk* over all messages of length k and corresponding estimators:

$$\epsilon_k^* \triangleq \inf_{\substack{\mathbf{m}: \{\mathsf{x}[n] | n \in \{1,\ldots,N\}\} \to \{0,1\}^{k\times 1} \\ \widehat{\mathsf{d}}: \{\mathsf{y}[n] | n\in\mathbb{Z}\} \times \mathbf{m} \to \mathcal{D}}} \epsilon\left(\widehat{\mathsf{d}}\left(\mathsf{y}[n], \mathbf{m}\right)\right), \tag{3}$$

where the estimator-dependent risk $\epsilon\left(\widehat{\mathsf{d}}\right) \triangleq \mathbb{E}\left[\ell(\widehat{\mathsf{d}}, \mathsf{d})\right]$ is defined for some loss function $\ell : \mathcal{D} \times \mathcal{D} \to \mathbb{R}_+$, and the expectation is taken with respect to all sources of randomness, i.e., $\mathsf{x}[n], \mathsf{z}[n]$ and d. A pair $(\mathbf{m}_{\text{o}}, \widehat{\mathsf{d}}_{\text{o}})$ that attains (3) is called an "optimal strategy". The problem that naturally arises from this formulation can be concisely stated as follows:

Problem: *Given a message length of k bits and a loss function $\ell(\cdot,\cdot)$, find an optimal strategy for TDE.*

[2]Notice that possible attenuation to the common signal is implicitly taken into account via the variances of the processes.

[3]Observe that $\texttt{SNR} \xrightarrow{\rho\to 1} \infty$ and $\texttt{SNR} \xrightarrow{\rho \to 0} 0$, as desired, and also that $\rho^2 = \frac{1}{1+\texttt{SNR}^{-1}}$ is in fact what is (usually) known as the "Wiener coefficient".

[4]With a slight abuse of notation, we write that "one observes $\mathsf{x}[n]$" when we mean that one observes the entire process $\{\mathsf{x}[n]\}_{n\in\mathbb{Z}}$ or a snippet of it.

[5]Despite d being from a discrete set in this formulation, we will persist with "estimate" rather than "detect" to conform with TDE classical terminology.

[6]In the limit, an infinite sequence rather than a finite-dimensional vector.

[7]The sense of "goodness" in this work is to be defined in Section II-A.

[8]This can be shown analytically for certain cases, see [41, Thm. 1].

We are interested in the trade-off between the number of transmitted bits k and its associated optimal risk ϵ_k^*. Since in this work we assume a discrete time-delay, we focus on the error loss $\ell(a,b)=\mathbb{1}_{a\neq b}$ that yields the error probability risk.

To the best of our knowledge, a solution to this simplified, though fundamental problem is currently unknown. However, and while we have not been able to find such a solution, this work proposes a new type of a solution strategy that constitutes an achievability result, which is (asymptotically) superior to existing TDE methods in terms of the trade-off (3). Moreover, and importantly, our proposed strategy warrants a rethinking of the communication efficiency in the context of this specific, but ubiquitous task. We will later discuss how our proposed solution to this basic problem, formulated in discrete-time, leads to new ideas for localization methods that are attractive in scenarios with limited communication resources.

III. The "Maximum-Index"-Based Estimator

Inspired by [35], [36], our proposed strategy is as follows.
Extremum Encoding: The encoder observes a sequence of length $N=2^k$,[9] namely $\mathcal{X}_N\triangleq\{\mathsf{x}[n]\}_{n=0}^{N-1}$, and sends as the message $\mathbf{m}$ the *index* of the maximum sample among $\mathcal{X}_N$,

$$\mathsf{j}\triangleq\arg\max_{0\leq n\leq N-1}\mathsf{x}[n], \tag{4}$$

where $\mathbf{m}\in\{0,1\}^{k\times 1}$ is the binary representation of j.
Maximum-Index-Based Estimation: The decoder, which in particular observes $\mathcal{Y}_N^{\mathcal{D}}\triangleq\{\mathsf{y}[n]\}_{n=-d_m}^{N-1+d_m}$, upon receiving $\mathbf{m}$, constructs the "maximum-index"-based estimator (MIE),

$$\widehat{\mathsf{d}}_{\text{MIE}}\triangleq\arg\max_{\ell\in\mathcal{D}}\mathsf{y}[\mathsf{j}+\ell]. \tag{5}$$

Put simply, the message to the decoder dictates the center of its search (discrete) interval, whose size is the delay spread, i.e., $D=2d_m+1$. The estimated time-delay is then chosen to be the shift (in the opposite direction) relative to that center, for which the observed signal at the decoder is maximized.

Notice that both the encoder and decoder depend on the observed sequences via maximizers of subsequences. Since such statistics are invariant to positive multiplicative constant factors, the performance is not affected by the scaling factors of Proposition 1. Thus, the scheme is *universal* for any positive correlation parameter ρ. Of course, the correlation still needs to be known in order to choose an adequate message size k.

A. Interpretation of the MIE

The underlying logic of (5) is in fact quite clear and intuitive. Let us first consider the maximum *a posteriori* estimator of d, which, in our case, coincides with the MLE since $\mathsf{d}\sim\mathcal{U}(\mathcal{D})$, and can be easily shown to be given by

$$\widehat{\mathsf{d}}_{\text{MIE}}=\arg\max_{\ell\in\mathcal{D}}\frac{1}{N}\sum_{n=0}^{N-1}\mathsf{x}[n]\mathsf{y}[n+\ell]\triangleq\arg\max_{\ell\in\mathcal{D}}\widehat{\rho}_{\text{MIE}}(\ell). \tag{6}$$

Although by different means, the MIE (5) is doing exactly what the MLE (6) is doing without communication constraints, which is simply trying to identify the time-lag at which the cross-correlation between $\mathsf{x}[n]$ and $\mathsf{y}[n]$ is maximized. To see this more clearly, we recall the following useful result, by Hadar and Shayevitz [35, Theorem. 1]: Consider model (2) with $\mathsf{d}\equiv 0$, i.e., $\{\mathsf{x}[n]\}$ and $\{\mathsf{y}[n]\}$ are zero-mean unit-variance white Gaussian processes with a correlation coefficient ρ. Let $\widehat{\rho}_{\text{MIE}}\triangleq\frac{\mathsf{y}[\mathsf{j}]}{\mathbb{E}[\mathsf{x}[\mathsf{j}]]}$. Then $\widehat{\rho}_{\text{MIE}}$ is an unbiased estimator of ρ with

$$\mathbb{V}\text{ar}\left(\widehat{\rho}_{\text{MIE}}\right)=\frac{1}{k}\left(\frac{1-\rho^2}{2\log(2)}+o(1)\right)=\frac{1-\rho^2}{2\log(N)}+o(1). \tag{7}$$

Moreover, $\widehat{\rho}_{\text{MIE}}$ is asymptotically efficient[10] given $(\mathsf{x}[\mathsf{j}],\mathsf{y}[\mathsf{j}])$.

A natural extension (/application) of the above is to define

$$\widehat{\rho}_{\text{MIE}}(\ell)\triangleq\frac{\mathsf{y}[\mathsf{j}+\ell]}{\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right]},\quad\forall\ell\in\mathbb{Z}, \tag{8}$$

which is of course an unbiased, asymptotically efficient estimator of ρ when $\ell=\mathsf{d}$, and of 0 when $\ell\neq\mathsf{d}$. This is simply because, for any shift ℓ, we end up with exactly the same formulation considered in [35], but for a different correlation coefficient, since $\mathbb{E}[\mathsf{y}[n]\mathsf{x}[n-\ell]\mid\mathsf{d}]=\rho\cdot\mathbb{1}_{\ell=\mathsf{d}}$. We will return to this point later when considering a generalization of model (2) that we currently focus on.

With (8), we can revisit (5), and using the fact that $\mathbb{E}[\mathsf{x}[\mathsf{j}]]$ is constant with respect to the optimization index ℓ, we have

$$\widehat{\mathsf{d}}_{\text{MIE}}=\arg\max_{\ell\in\mathcal{D}}\frac{\mathsf{y}[\mathsf{j}+\ell]}{\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right]}=\arg\max_{\ell\in\mathcal{D}}\widehat{\rho}_{\text{MIE}}(\ell). \tag{9}$$

Indeed, it is now evident that the MIE (9) (for limited communication) and the MLE (6) (for unlimited communication) are similar in nature—both are choosing the hypothesized delay at which the empirical cross-correlation is maximized.

A further insightful interpretation of (5) is the following. Resorting again to the MLE so as to focus on the conceptual nature of the cross-correlation operation, the time-delay estimate is chosen as the one for which the two (relatively-)shifted versions of the two received signals are the "most similar" to each other. Therefore, if we apply some signal compression method that would preserve[11] this similarity under the correct time-delay, we would be able (and most likely need) to use an estimation procedure that implements the same concept.

With this in mind, consider the (coarse) signal compression method, that zeros all the samples below the highest threshold possible, such that *only* the largest sample, namely the maximum, passes and is left uncompressed. In the ideal noiseless case, when the two received signals are merely two differently time-shifted versions of the same waveform, the specific form of similarity discussed above is indeed preserved. Consequently, a sensible time-delay estimator would look for the hypothesized time-delay for which the (relatively-)shifted versions of the two received signals are, again, the "most similar" to each other. An illustration of this concept is given in Fig. 3. Loosely speaking, since the maximum of an N-sample long realization of a white Gaussian process has a vanishing variance (see, e.g., [42]), it becomes asymptotically (almost) deterministic, and this is exactly what happens when we send its position on the time axis; its magnitude and sign are already known with high

[9] This assumption is merely for notational convenience, and can of course be relaxed, in the sense that N can be any natural number.

[10] An unbiased estimator that attains the Cramér-Rao lower bound.

[11] Possibly (only) with high probability.

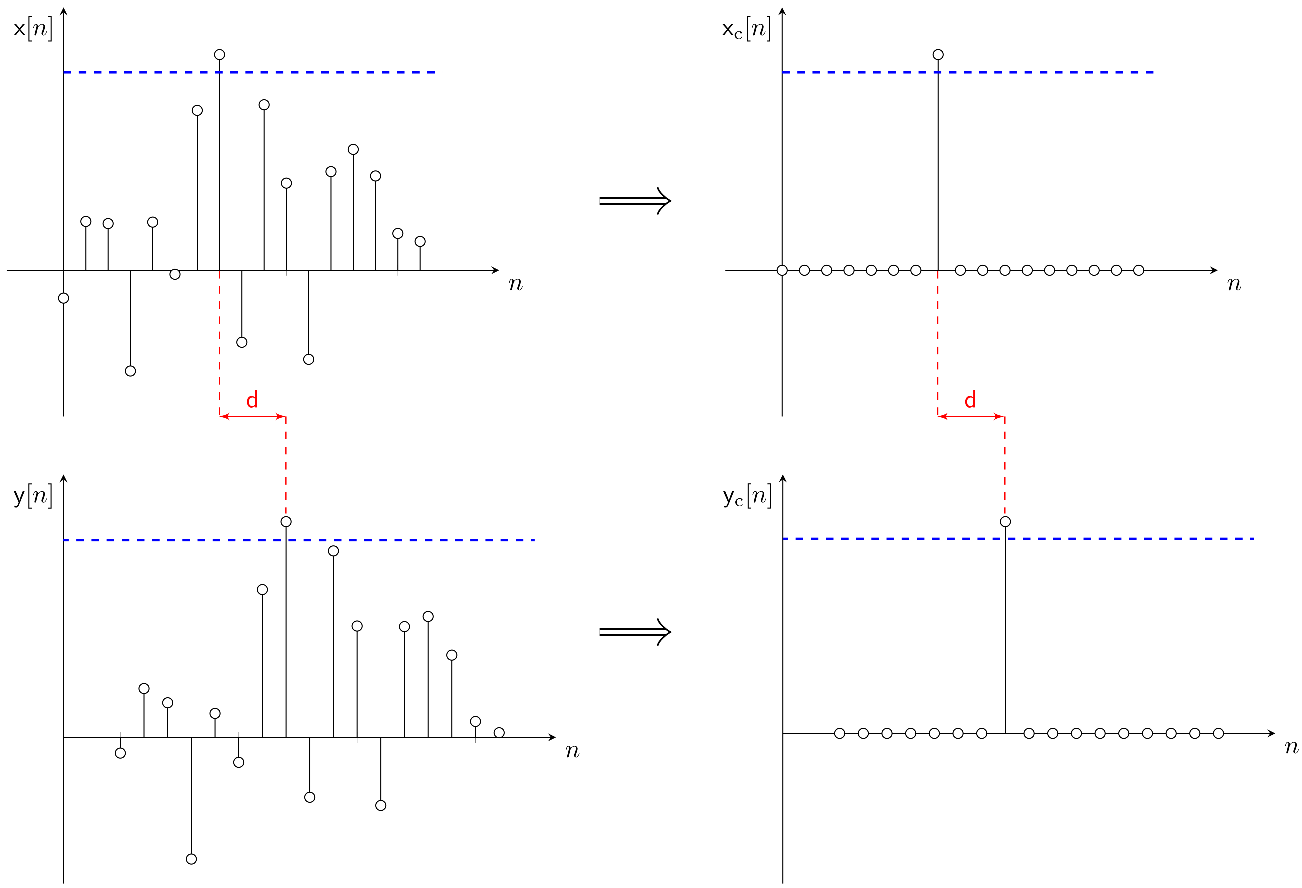

Fig. 3. Illustration of the compression-estimation strategy that is indirectly performed by the proposed extremum encoding and MIE. Here, $\mathsf{x}_{\mathrm{c}}[n]$ and $\mathsf{y}_{\mathrm{c}}[n]$ denote the compressed versions of $\mathsf{x}[n]$ and $\mathsf{y}[n]$, respectively. When the noise is sufficiently small, the (relative) location in time of the maximum is unchanged, and the delay can be identified even after this crude compression. Since the maximum of a finite-length Gaussian sequence is asymptotically deterministic (see, e.g., [42]), only its location in time needs to be sent.

probability. This intuition generalizes to the noisy case, since the instantaneous SNR of the sample $\mathsf{y}[\mathsf{j}+\mathsf{d}]$ is ever growing as $k \to \infty$ ($\Rightarrow N \to \infty$).[12]

Due to the asymptotic concentration of $\mathsf{x}[\mathsf{j}]$—the maximum of a finite-length realization of a white Gaussian process—around its mean (see, e.g., [42]), the intuitive interpretation of (9) can also be rigorously justified, as we show next.

B. Performance Analysis

We now analyze the performance of our proposed method. In particular, we derive an explicit upper bound on the error probability of (5) that decays exponentially as a function of the number of bits sent. As a result, this also shows that (5) is consistent in the communication sense. We further show that the exponential behavior of the bound is tight for our scheme.

In contrast to our recent work [43], where the parameter d_m controlling the delay spread was considered to be fixed, we now focus on the regime $k \to \infty$, where the *rate* of this problem, defined as,

$$R \triangleq \frac{k}{\log_2(D)} = \frac{k}{\log_2(2d_m+1)}, \tag{10}$$

namely the message length in bits over the (shortest) binary description length of the delay spread, is kept fixed.[13] In other words, R is the information conveyed relative to the minimal information required for description of d. This implies that for a given (non-zero) rate R, we have $d_m \to \infty$ at an appropriate rate. Thus, [43] addresses a special case of the more general framework considered in this work. We note that our definition (10) to the rate is not the standard one (i.e., in bits per sample) that is used in the context of classical compression. Rather, (10) is in message length in bits per the delay spread ("uncertainty") in bits. This way, the rate of an "oracle" encoder, that somehow knows the (unknown) delay and would have simply sent it to the decoder, is $R = 1$. Since we focus on asymptotics, we ignore rounding issues in (10).

Before we proceed to the analysis of our proposed encoding-estimation strategy, we present a general result that pertains to the fundamental limitations of *any* encoding-estimation strategy. In particular, it states that the number of bits to be conveyed (/transmitted) from the encoder to the decoder has to be larger than the (minimal) number of bits that is required

[12] This will be shown rigorously, and in detail, throughout the proofs of our analysis in Section III-B.

[13] Generally, per the definition (10), only the cardinality of $\mathcal{D}$, D, matters for our results (rather than the maximum delay d_m). However, since d_m represents a meaningful quantity—the maximum delay spread in samples—in the TDE problem, we will occasionally use $2d_m$ rather than $D-1$.

for the description of any possible delay from the delay spread. Formally, we have the following proposition.

Proposition 2 (Lower bound on rate): Let k be the number of bits that is allocated in order to send the message $\mathbf{m}$, and $\widehat{\mathsf{d}}$ an estimator of d based on $\mathsf{y}[n]$ and $\mathbf{m}$. Then, for any encoding-estimation strategy, namely any pair $\left(\widehat{\mathsf{d}}, \mathbf{m}\right)$,

$$\lim_{k\to\infty} \mathbb{P}\left(\widehat{\mathsf{d}} \neq \mathsf{d}\right) = 0 \implies R \geq 1. \tag{11}$$

Proof: Say that you give the encoder as genie information the sequence $\mathsf{y}[n]$ and the delay d, and now it needs to convey the delay to the decoder. First, the encoder can throw away $\mathsf{x}[n]$, since, having the delay d, it can always create locally a signal that is statistically equivalent (i.e., that has the same joint distribution with $\mathsf{y}[n]$). Now, in order to convey the delay to the decoder, the number of bits it must use is

$$k \geq H\left(\mathsf{d} \mid \mathsf{y}[n]\right) = H(\mathsf{d}) = \log_2(D), \tag{12}$$

where we have used the fact that $\mathsf{y}[n]$ is independent of d, and that $\mathsf{d} \sim \mathcal{U}(\mathcal{D})$. Thus, $R = k/\log_2(D) \geq 1$. ■

C. Main Results

In order to establish our main result—the exact error exponent of the MIE (5)—we first prove two propositions, from which our main theorem is readily obtained. Specifically, we will derive exponentially tight upper and lower bounds that will facilitate the main theorem.

We start by upper bounding the error probability of (5).

Proposition 3 (Threshold rate and upper bound): Consider the loss $\ell(a,b) = \mathbb{1}_{a\neq b}$, which yields the error probability risk $\epsilon \triangleq \mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right)$, and let R be as defined in (10). Then, a sufficient condition for a vanishing error probability, i.e., that $\epsilon \to 0$ as $k \to \infty$ while R is kept fix, is

$$R > R_0 \triangleq \frac{1}{\rho^2} = 1 + \frac{1}{\mathtt{SNR}}, \tag{13}$$

where R_0 is the threshold rate. Furthermore, for any rate above R_0 and any $\rho \in (0,1]$,[14] namely any $\mathtt{SNR} > 0$, we have

$$\epsilon \leq \bar{\epsilon}(k,\rho,d_m)\left(1 + o(1)\right), \tag{14}$$

where

$$\bar{\epsilon}(k,\rho,d_m) \triangleq Q\left(\sqrt{\tfrac{2k\log(2)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right) + 2d_m Q\left(\rho\sqrt{\tfrac{2k\log(2)}{2-\rho^2}}\right). \tag{15}$$

Proof: See Appendix B. ■

Remark 1: When $\rho \to 1$ (the "high SNR regime"),

$$\bar{\epsilon}(k,\rho,d_m) \xrightarrow[\text{Noiseless case}]{\rho\to 1} \frac{2d_m}{2^k}. \tag{16}$$

Therefore, if the bound is tight, then even for $\rho = 1$, the error probability is *not* zero for a finite observation interval.[15] This is because for finite intervals, the maximum has a nonvanishing variance, and one of the samples (or more) that are observed by the decoder at the "edges" (due to the inherent time-delay uncertainty), but not by the encoder, can be greater than the one reported by the encoder. This behavior is what one would expect from the true error probability, and thus it is reassuring (as a "sanity check") that the upper bound reflects it.

Remark 2: Further to Remark 1, in the infinite SNR (or, equivalently, the noiseless) regime, viz $\rho = 1$, in the limit $k \to \infty$ when R is held fixed, we have $R_0 = 1$. Thus, in this case the lower bound on the rate R for our (realizable) method coincides with the genie lower bound (11) of Proposition 2.

Remark 3: Although in this work we consider a discrete time-delay, hence focusing on the error loss $\ell(a,b) = \mathbb{1}_{a\neq b}$, with Proposition (3) we can also obtain an upper bound on the q-th absolute moments of the estimation error, $\mathbb{E}\left[\left|\widehat{\mathsf{d}}_{\text{MIE}} - \mathsf{d}\right|^q\right]$ (for any $q \in \mathbb{Z}$), which is given in (S58), Appendix SI, in the supplementary materials.

Remark 4: While we have chosen to model the delay d as random, and specifically as a uniform RV, it follows from our derivation (specifically, from (58)) that our upper bound holds even when d is not uniformly distributed.

An immediate corollary of Proposition 3 is the following.

Corollary 1 (Communication consistency): Under the same conditions of Proposition 3, the MIE (5) is consistent, namely,

$$\lim_{k\to\infty} \mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right) = 0. \tag{17}$$

This also implies that $\lim_{k\to\infty} \mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right) = 0$ for a fixed d_m (as in [43]). Furthermore, note that $\widehat{\mathsf{d}}_{\text{MIE}}$ is also consistent (in the "standard" sense) with respect to the number of samples that need to be used, N, i.e., $\lim_{N\to\infty} \mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right) = 0$.

Proof: We trivially obtain (17) by taking the limit $k \to \infty$ of $\bar{\epsilon}$ in (15), with which the upper bound in (14) vanishes. ■

Remark 5: For Corollary 1, R need not be fixed, and it is only required that $d_m = o\left(2^{k\rho^2/(2-\rho^2)}\right)$, namely the uncertainty interval grows "sufficiently slow" with k (and N). This case, wherein R and d_m are not fixed, can be thought of as the intermediate regime between the one we focus on here (R fixed), and the one previously considered in [43] ($R \to \infty$).

Next, we present the second key ingredient for our main result—a lower bound on the error probability of the MIE (5).

Proposition 4 (Threshold rate and lower bound): Under the same conditions of Proposition 3, in the limit $k \to \infty$ with R fixed, we have

$$\epsilon \geq \underline{\epsilon}(k,\rho,d_m)(1 + o(1)), \tag{18}$$

where

$$\underline{\epsilon}(k,\rho,d_m) \triangleq \max\left\{Q\left(\sqrt{\tfrac{2k\log(2)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right), 2d_m Q\left(\rho\sqrt{\tfrac{2\log(2)k}{2-\rho^2}}\right)\right\}, \tag{19}$$

and a necessary condition on the rate R for the lower bound (19) to approach zero is (13).

[14] The case $\rho = 0$ is less interesting, and trivial to analyze, since both sensors observe (purely) statistically independent white Gaussian noise.

[15] And this is due to the finite sample size available for estimation of ρ.

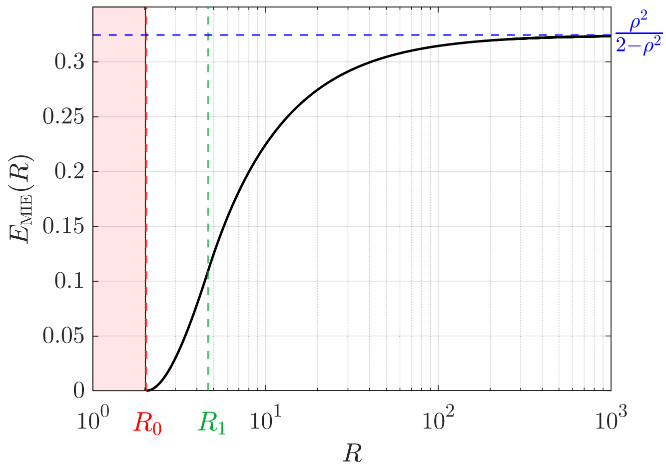

Fig. 4. Error exponent $E_{\text{MIE}}(R)$ (21) of the MIE vs. R for a fixed $\rho = 0.7$. The vertical red dashed line is located at the threshold rate R_0, and the vertical green dashed line is located at the critical rate R_1. The horizontal dashed blue line is the asymptotic value $\lim_{R\to\infty} E_{\text{MIE}}(R)$. The red shaded region corresponds to rates at which $\mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right)$ cannot vanish for $\rho = 0.7$.

Proof: See Appendix C. ∎

Our main result can now be readily obtained.

Theorem 1 (Exact error exponent of the MIE): For $R > R_0$, in the limit $k \to \infty$, we have

$$-\frac{1}{k}\log_2(\epsilon) = E_{\text{MIE}}(R) + o(1), \tag{20}$$

where

$$E_{\text{MIE}}(R) \triangleq \begin{cases} \frac{\left(\sqrt{R}-\sqrt{R_0}\right)^2}{R(R_0-1)}, & R_0 < R < R_1 \\ \frac{R-2R_0+1}{(2R_0-1)R}, & R_1 < R \end{cases}, \tag{21}$$

and where $R_1 \triangleq \left(\frac{2-\rho^2}{\rho}\right)^2$ is the critical rate. Furthermore, it follows that (13) is necessary and sufficient for ϵ to approach zero in the limit $k \to \infty$ for a fixed R.

Proof: See Appendix D. ∎

Theorem 1 provides an exact characterization of the asymptotic rate of decay of the error probability of the MIE. Specifically, it determines the exact threshold rate R_0 above which the MIE has a vanishing error probability. Moreover, it shows that there exists a critical rate R_1, below and above which the error exponent changes its functional form. It further shows that as R grows, the error exponent approaches the limit

$$\lim_{R\to\infty} E_{\text{MIE}}(R) = \frac{\rho^2}{2-\rho^2} = \frac{\texttt{SNR}}{2+\texttt{SNR}}. \tag{22}$$

Thus, the SNR conditions—encapsulated here by ρ—effect the performance of the MIE even when $R \to \infty$. Fig. 4 shows the error exponent (21) vs. the rate R for a fixed correlation (SNR-encapsulating) coefficient $\rho = 0.7$. The red shaded area corresponds to rates at which the error probability of the MIE ϵ cannot vanish for this value of ρ (i.e., this SNR level).

In addition, Theorem 1 has an important practical implication, which pertains to the relation between the SNR and the observation length, stated concisely as follows:

Given the SNR conditions in both sensors, (13) *quantifies the relative minimal observation (discrete-) time interval that is required for TDE with* $\widehat{\mathsf{d}}_{\text{MIE}}$ *at a desired estimation fidelity.*

D. Extension to the Generally Correlated Signal Case

Thus far, we have considered the model (2) in which the signals are white, namely, the cross-correlation between $\mathsf{x}[n]$ and $\mathsf{y}[n]$ is a scaled (Kronecker) delta function. If we think of these discrete-time signals as sampled versions of some bandlimited, continuous-time signals (say, $\mathsf{x}(t)$ and $\mathsf{y}(t)$), then model (2) is accurate when their cutoff frequency is identical, the time-delay is an integer multiplication of the sampling period,[16] and the sampling rate is exactly their Nyquist rate.

However, when this is not the case, e.g., when the sampling rate is above (the maximum between) their Nyquist rate, the sampled signals $\mathsf{x}[n], \mathsf{y}[n]$ will generally be *temporally* correlated. In fact, for such systems, the physical justification for the assumption that the true time-delay can be modeled sufficiently well as an integer multiplication of the sampling period becomes even stronger. Therefore, the extension to the case of general correlation structures, beyond white signals, is an important and relevant one from a practical standpoint.

Motivated by the above, we have the following result.

Proposition 5 (Upper bound, general autocorrelation): Let $\mathsf{x}[n]$ be a zero-mean unit-variance stationary Gaussian process with an autocorrelation $C_{\mathsf{xx}}[\ell] \triangleq \mathbb{E}\left[\mathsf{x}[n]\mathsf{x}[n-\ell]\right]$, such that $\mathsf{y}[n]$ (as in (2)) and $\mathsf{x}[n]$ are (conditionally) jointly Gaussian with a (conditional) cross-correlation $\mathbb{E}\left[\mathsf{y}[n]\mathsf{x}[n-\ell]|\mathsf{d}\right] = \rho \cdot C_{\mathsf{xx}}[\ell - \mathsf{d}] \triangleq \rho(\ell)$. Then, a sufficient condition for a vanishing error probability of the MIE (5), i.e., that $\epsilon \to 0$ as $k \to \infty$ with a fixed rate R, is

$$R > R_0(\ell^*) \triangleq R_0 \cdot \frac{1}{(1-\varepsilon_0^*)^2\delta(\ell^*)}, \tag{23}$$

where

$$\ell^* \triangleq \arg\max_{\ell\in\mathbb{Z}\setminus\{0\}} C_{\mathsf{xx}}[\ell], \quad \delta(\ell^*) \triangleq 1 - C_{\mathsf{xx}}[\ell^*], \tag{24}$$

$$\varepsilon_0^* \triangleq \inf\left\{\varepsilon_0 \middle| \varepsilon_0 \in (0,1), \frac{(1-\delta(\ell^*))(\rho-\rho(\ell^*))^2}{2-\rho^2-\rho^2(\ell^*)} < \frac{\varepsilon_0^2}{(1-\varepsilon_0)^2}\right\}. \tag{25}$$

Furthermore, for any rate above $R_0(\ell^*)$ and any $\rho \in (0,1]$,

$$\epsilon \leq \left[Q\left(\sqrt{\frac{2\log(N)}{1-\rho^2}}\left(\rho(1-\varepsilon_0^*)\sqrt{\delta(\ell^*)} - \frac{1}{\sqrt{R}}\right)\right)\right. \tag{26}$$

$$\left. + 2d_m Q\left(\frac{(1-\varepsilon_0^*)\sqrt{2\log(N)\delta(\ell^*)(\rho-\rho(\ell^*))^2}}{\sqrt{2-\rho^2-\rho^2(\ell^*)}}\right)\right](1+o(1)). \tag{27}$$

Proof: See Appendix SII (supplementary materials). ∎

Similarly to the iid case (Section III-C), a direct consequence of Proposition 5 is the following corollary, whose proof is identical to that of Corollary 1, and is therefore omitted.

Corollary 2 (Communication consistency, general autocorrelation): Under the same conditions of Proposition 5, the MIE (5) is consistent in the communication sense, namely,

$$\lim_{k\to\infty} \mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}} \neq \mathsf{d}\right) = 0, \tag{28}$$

[16] If this is not the case, then the discrete-time-delay is an approximation.

This implies $\lim_{k\to\infty}\mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}}\neq\mathsf{d}\right)=0$ for a fixed d_m, which, in turn, implies $\lim_{N\to\infty}\mathbb{P}\left(\widehat{\mathsf{d}}_{\text{MIE}}\neq\mathsf{d}\right)=0$, i.e., consistency (in the classical sense) with respect to the sample size.

Remark 6: Similarly to Corollary 1, for Corollary 2, R need not be fixed, and it is only required that

$$d_m = o\left(2^{k\frac{(1-\varepsilon_0^*)^2\delta(\ell^*)(\rho-\rho(\ell^*))^2}{2-\rho^2-\rho^2(\ell^*)}}\right), \tag{29}$$

i.e., the uncertainty interval can grow, but "sufficiently slow" with k (and consequently with N).

While the upper bound already establishes the performance guarantees for the general autocorrelation case, we further provide the following lower bound, which is informative for the high-rate asymptotic regime.

Proposition 6 (Lower bound, general autocorrelation): Under the conditions of Proposition 5, but in the regime where $k,R\to\infty$, we have

$$\epsilon \geq Q\left(\frac{(1-\varepsilon_0^*)\sqrt{2\log(N)\delta(\ell^*)(\rho-\rho(\ell^*))^2}}{\sqrt{2-\rho^2-\rho^2(\ell^*)}}\right)(1+o(1)). \tag{30}$$

Proof: See Appendix SIII (supplementary materials). ■

The lower bound (30) is exponentially tight in the asymptotic regime $k,R\to\infty$, namely when d_m is either fixed or growing sufficiently slow relative to k. Indeed, from (30) and (27), we have the following theorem.

Theorem 2 (Asymptotic error exponent of the MIE, general autocorrelation): Under the same conditions of Proposition 6,

$$-\frac{1}{k}\log_2(\epsilon) = \frac{(1-\varepsilon_0^*)^2\delta(\ell^*)(\rho-\rho(\ell^*))^2}{2-\rho^2-\rho^2(\ell^*)} + o(1), \tag{31}$$

where in (31) the little-o is with respect to k and R.

The (simple) proof of Theorem 2, based on Propositions 5 and 6, follows exactly the same reasoning as the one for Theorem 1, and is therefore omitted. As a "sanity check", if $C_{\mathsf{xx}}[\ell^*]=0$, it follows that $C_{\mathsf{xx}}[\ell]=0$ for all $\ell\in\mathbb{Z}\backslash\{0\}$ and that $\delta(\ell^*)=1$, $\rho(\ell^*)=0$ and $\varepsilon_0^*\to 0$. In this case, observe that the right-hand side of (31) coincides with (21) for $R\to\infty$, i.e., (22).

Remark 7: Although we assume throughout this work that $\mathsf{d}\sim\mathcal{U}(\mathcal{D})$, our specific analysis technique is such that in large parts we condition on a particular value $\mathsf{d}=d$. Hence, it can be relevant for other prior distributions, and it is rather simple to derive similar results in a non-Bayesian formulation, where the delay is assumed to be deterministic and unknown.

While other extensions are deferred to future work, we end this section by emphasizing that our compression-estimation strategy indeed generalizes beyond the (seemingly) simplistic iid model to a less trivial setting. We consider this aspect a key element in the contribution of our proposed method.

E. Computational Complexity

Beyond its performance in terms of the trade-off between accuracy and communication efficiency, another appealing property of the MIE (5) is its computational complexity, in particular relative to standard CCEs, such as the MLE (6). For such standard CCEs, computing the empirical cross-correlation at $D=2d_m+1$ time-lags based on N samples amounts to $\mathcal{O}(Nd_m)$ operations. In contrast, the MIE simply requires two searches for the maximum of two arrays of sizes N (encoder) and D (decoder), hence its computational complexity is $\mathcal{O}(N+d_m)$. It is therefore evident that our proposed method is not only more efficient in terms of communication, but is also attractive in terms of the required computational resources.

IV. Simulation Results

In our simulation experiments presented in this section, we generate the signals according to the model (2) and compare our proposed method (MIE) with the following benchmarks:

- The CCE (6) where $\mathsf{x}[n]$ is replaced by $\widehat{\mathsf{x}}_{\text{RD}}[n]$,[17] a RD-optimally compressed version thereof, where the distortion measure is the squared error $(\mathsf{x}[n]-\widehat{\mathsf{x}}_{\text{RD}}[n])^2$. This is the lower bound considered in [23, Figure 10];
- The CCE (6) where $\mathsf{x}[n]$ is replaced by $\widehat{\mathsf{x}}_{1-\text{bit}}[n]\triangleq\text{sign}(\mathsf{x}[n])$. This compression method is used, e.g., in [22, Section 4.1], and has recently gained renewed interest for various related tasks [45], [46], and specifically also in the context of time-delay-based localization [47], [48], [49];

In this numerically evaluated part of the paper, for fairness, we compare our method to benchmarks that are also universal, i.e., that do not require knowledge of the SNR. All empirical results were obtained by averaging 10^6 independent trials.

A. White Signal

In each independent trial of this simulation experiment, we generate 2^k-samples long realizations of standard iid Gaussian processes for $\mathsf{x}[n]$ and $\mathsf{z}[n]$, draw d uniformaly from $\mathcal{D}$, and compute $\mathsf{y}[n]$, as depicted in Fig. 2, for a fixed ρ. For $\widehat{\mathsf{x}}_{\text{RD}}[n]$, we use a compression rate $k/N=k/2^k$ bits per sample, such that $\{x[n]\}_{n=0}^{N-1}$ is compressed into a message of k bits. For $\widehat{\mathsf{x}}_{1-\text{bit}}[n]$, we take the signs of the first k samples as the message.

Fig. 5 presents the error probability vs. k, the size of the message $\mathbf{m}$ in bits, for fixed $\texttt{SNR}=15$dB and $d_m=250$. Evidently, our proposed method is able to achieve superior performance relative to the alternatives, where the gain is most significant at high rates (higher k for a fixed d_m). Furthermore, a good empirical fit is reflected between the error probability of the MIE and our (logarithmically shifted[18]) error exponent curve (21), which corroborates our analytical derivations.

Next, we fix $k=12$ bits (as well as the sample size $N=2^{12}$), and we vary d_m for a fixed $\texttt{SNR}=20$dB. Fig. 6 presents the error probability vs. the rate R. Here, rather than presenting the error exponent (21), we present the upper and lower bounds, (15) and (19), respectively. It is observed that, even in this non-asymptotic regime, the upper and lower bounds closely track the true error probability curve. In addition, it is observed that, except for low rates near R_0, our method outperforms the two benchmark alternatives.

[17]In this case, this is the still MLE (after the said RD compression), since $\widehat{\mathsf{x}}_{\text{RD}}[n]$ is Gaussian [44, Ch. 10.3.2], and $\widehat{\mathsf{x}}_{\text{RD}}[n],\mathsf{y}[n]$ are jointly Gaussian.

[18]For details on how it was vertically shifted, see the caption of Fig. 5.

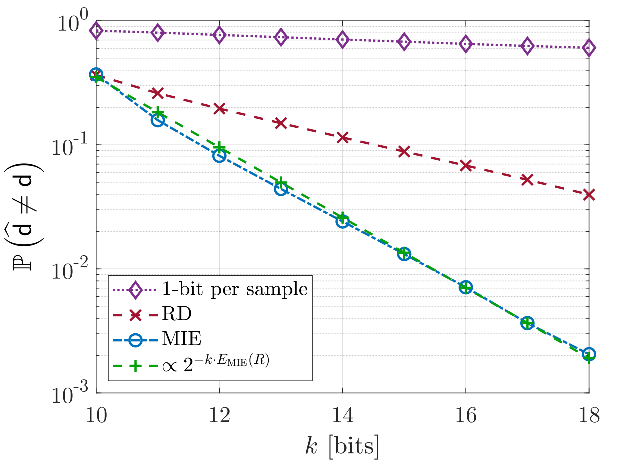

Fig. 5. Error probability vs. number of transmitted bits, for $\texttt{SNR} = 15$dB and $d_m = 250$. The green dashed line is $\widehat{c} \cdot 2^{-k \cdot E_{\text{MIE}}(R)}$, where $\widehat{c}$ is the best least-squares-fitted constant for the empirical curve of the MIE. The performance of the MIE, whose exponential behavior is accurately predicted by our analysis, becomes increasingly favorable relative to the alternatives as k grows.

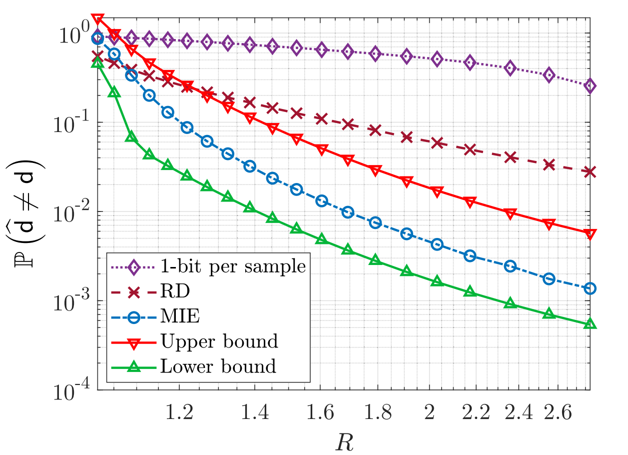

Fig. 6. Error probability vs. the rate, as defined in Proposition 3, for $\texttt{SNR} = 20$dB and $k = 12$ bits. The error probability of the MIE rapidly drops below the two benchmarks when departing from R_0 (on the left edge), and is contained within the margin between the upper and lower bounds.

In Fig. 7 we consider the same setting, with $k = 12$ fixed, but we now vary the SNR, and present four curves corresponding to different values of d_m (alternatively, to different rates). From condition (13), we see that given a fixed rate R, we have a condition for the minimal SNR above which the error probability asymptotically vanishes,

$$\texttt{SNR} > \frac{1}{R-1} = \frac{\log_2(2d_m+1)}{k - \log_2(2d_m+1)} \triangleq \texttt{SNR}_{\text{thr}}(d_m), \quad (32)$$

where $\texttt{SNR}_{\text{thr}}(d_m)$ is the threshold SNR value for a given k.

As can be observed from Fig. 7, all methods exhibit unacceptable performance at low SNRs. However, the SNR from which the MIE outperforms the benchmarks methods decrease as the rate increases. Additionally, while the benchmark methods exhibit a "saturation"-like trend already above 10^{-1}, our

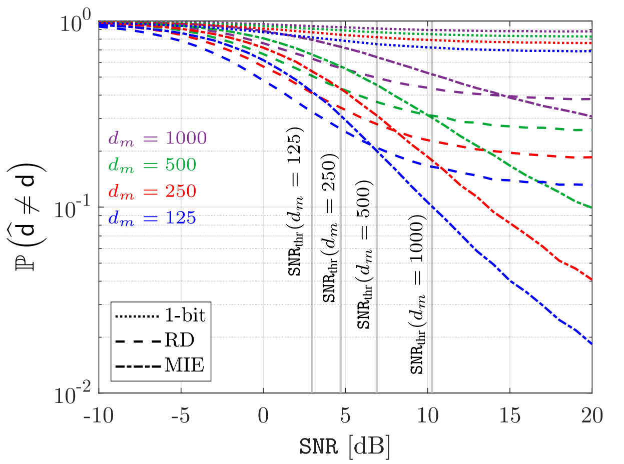

Fig. 7. Error probability vs. the SNR, for $k = 12$ bits and four different delay spreads d_m, namely four different rates. For a given rate R, (13) provides the minimal SNR above which the error probability asymptotically vanishes, where $\texttt{SNR}_{\text{thr}}(d_m) = \frac{1}{R-1} = \frac{\log_2(2d_m+1)}{k-\log_2(2d_m+1)}$ is the threshold SNR.

method continues to improve, and it will saturate only at the asymptotic level (16).[19]

B. Correlated Signal

We repeat the first simulation experiment described in Section IV-A, but this time with the signal

$$\mathsf{x}[n] = \sum_{m=0}^{M-1} h_m \tilde{\mathsf{x}}[n-m], \quad (33)$$

where $\tilde{\mathsf{x}}[n] \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $\{h_m\}_{m=0}^{M-1}$ is a finite impulse response filter, such that $\sum_{m=0}^{M-1} h_m^2 = 1$. It follows that $C_{\mathsf{xx}}[\ell] = \sum_{m=0}^{M-1} h_{\ell+m} h_m$,[20] and $\mathsf{x}[n]$ is (still) unit-variance. Since we assume that knowledge regarding the correlation structure $C_{\mathsf{xx}}[\ell]$ is unknown, $\widehat{\mathsf{x}}_{\text{RD}}[n]$ is still compressed as if it is white, namely at a rate $R = \frac{1}{2}\log_2\left(C_{\mathsf{xx}}[0]/\xi^2\right) = -\frac{1}{2}\log_2\left(\xi^2\right)$, where ξ^2 denotes the mean-square error distortion of $\widehat{\mathsf{x}}_{\text{RD}}[n]$.[21]

Fig. 8 shows the error probability vs. k for $\texttt{SNR} = 5$dB and the correlated signal (33) with the filter (of order $M = 4$)

$$\left\{\tilde{h}_0, \tilde{h}_1, \tilde{h}_2, \tilde{h}_3\right\} = \{1, -0.2, 0.3, 0.1\}, \; h_m = \frac{\tilde{h}_m}{\|\tilde{h}_m\|_2}, \quad (34)$$

in two different operational regimes, where d_m is computed as detailed in Table I. In the first regime, the rate R is below R_0 (13), which is obviously lower than the threshold rate for the non-iid case, whereas in the second R is above $R_0(\ell^*)$ (23). Fig. 8 corroborates our result in Proposition 5, and demonstrates that it is sufficient to operate at a (fixed) rate R above $R_0(\ell^*)$ to have $\epsilon \to 0$ when $k \to \infty$. The curves of the first regime (at a rate below R_0), however, suggest that there indeed exists a

[19]The upper and lower bounds are not presented in Fig. 7 to enhance clarity, and since the asymptotic analysis is less relevant for $k = 12$, rendering the bounds less informative relative to an extensive empirical comparison.

[20]Here, $h_m = 0$ for all $m \notin \{0, \ldots, M-1\}$.

[21]When the autocorrelation $C_{\mathsf{xx}}[\ell]$ is known, $\widehat{\mathsf{x}}_{\text{RD}}[n]$ can be better compressed so as to obtain the optimal RD trade-off of the (memoryless) optimal prediction error of $\mathsf{x}[n]$ (aka "innovation process") from its past samples [50].

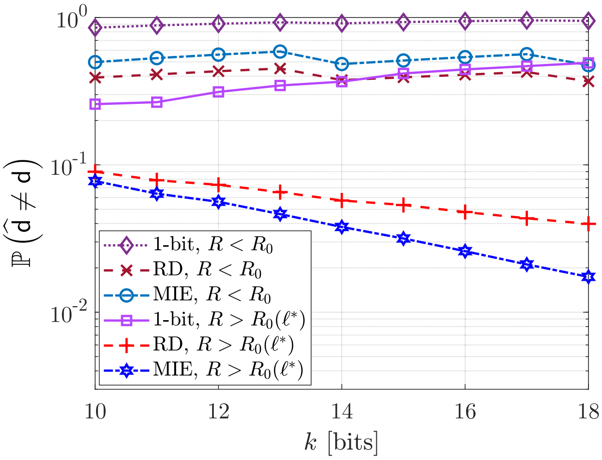

Fig. 8. Error probability vs. number of transmitted bits for the temporally correlated signal (33) at $\mathtt{SNR} = 5$dB. Two regimes are considered, as detailed in Table I. For a rate above $R_0(\ell^*)$, the error probability vanishes as $k \to \infty$.

TABLE I
DIFFERENT RATE-OPERATION REGIMES AT $\mathtt{SNR} = 5$dB

Regime 1: $R < R_0 = 1.3162$	**Regime 2:** $R > R_0(\ell^*) = 2.9404$
$d_m = 2^{\lfloor \frac{k}{1.3} \rfloor} \implies R \leq 1.3$	$d_m = \left\lfloor 2^{\frac{k}{3}-1} \right\rceil \implies R \geq 3$

threshold rate below which the error probability cannot vanish asymptotically. As in (13) for the iid case, we expect this threshold rate to be SNR-dependent. However, we also expect it to depend on the autocorrelation $C_{\mathsf{xx}}[\ell]$.[22]

We note in passing that the results for the first regime also show that there exist a rate below R_0 for which extremum encoding with the MIE is not an optimal strategy (with respect to the criterion (3)), at least for correlated signals, and at least in the non-asymptotic regime. Hence, theoretically, it is possible to devise an improved "hybrid" strategy that uses the "standard" RD signal compression below the rate at which it is superior to our scheme (if such a rate exists). Nevertheless, it should be emphasized that such a strategy would not be realizable, since it would be based on the impractical implementation of optimal RD signal compression.

V. Concluding Remarks

In the broad context of distributed systems, we present a realizable joint compression-TDE method that, to the best of our knowledge, currently constitutes the state-of-the-art asymptotic performance in terms of the trade-off between communication resources and TDE accuracy. Our method is universal in the sense that it does not require knowledge about the SNR conditions at either of the sensors. Furthermore, it reduces the overall computational complexity, and specifically the computational load from the central computing unit. For white Gaussian processes, we derive the exact error exponent of our method, thus providing an accurate characterization of its asymptotic performance. We further show that our method extends to correlated processes, and demonstrate by simulations its superiority over two benchmark alternatives.

While our method was developed for a simplified signal model, its underlying conceptual framework holds promise for the development of techniques applicable to more complex signal models in practical systems. This opens the door to several exciting research directions associated with our work that remain to be explored. One key avenue is extending (or adapting) the method to account for continuous time-delays, which is the more realistic, and at the same time the technically more challenging one. Additionally, our work naturally paves the way for innovative localization approaches in distributed systems. Since some of the most popular localization methods rely on estimated time-differences-of-arrival (TDOAs), our proposed method could serve as the fundamental step of TDOA estimation for such localization solutions in contemporary distributed networks, such as IoT-based sensor networks composed of interconnected "smart" devices, operating in an *ad-hoc* manner [51]. Further research on these important extensions, alongside additional ones, is underway [52] and will be addressed in future work.

Appendix A
Proof of Proposition 1

We shall show that the joint distribution of $\mathsf{r}_1[n]$ and $\mathsf{r}_2[n]$ is identical to that of $\mathsf{x}[n]$ and $\mathsf{y}[n]$. For this, we first condition on $\mathsf{d} = d$ (and henceforth omit the notation of this condition for brevity). Now, since $\mathbb{V}\mathrm{ar}(\mathsf{r}_1[n])$ and $\mathbb{V}\mathrm{ar}(\mathsf{r}_2[n])$ are known, and since $\mathsf{s}[n], \mathsf{z}_1[n]$ and $\mathsf{z}_2[n]$ are all mutually statistically independent, zero-mean white Gaussian processes, $\mathsf{r}_1[n]/\sqrt{\mathbb{V}\mathrm{ar}(\mathsf{r}_1[n])}$ and $\mathsf{r}_2[n]/\sqrt{\mathbb{V}\mathrm{ar}(\mathsf{r}_2[n])}$ are zero-mean, unit-variance jointly Gaussian with a cross-correlation function $\frac{\mathbb{E}[\mathsf{r}_2[n]\mathsf{r}_1[n-\ell]]}{\sqrt{\mathbb{V}\mathrm{ar}(\mathsf{r}_1[n])\mathbb{V}\mathrm{ar}(\mathsf{r}_2[n])}} = \rho \cdot \mathbb{1}_{\ell=d}$, with $\rho \triangleq \frac{1}{\sqrt{(1+\sigma_1^2)(1+\sigma_2^2)}}$.

Now, for (2), again condition on $\mathsf{d} = d$. Clearly, since $\mathsf{x}[n]$ and $\mathsf{z}[n]$ are mutually statistically independent, zero-mean white Gaussian processes, by construction of $\mathsf{y}[n]$, $\mathsf{x}[n]$ and $\mathsf{y}[n]$ are zero-mean, unit-variance jointly Gaussian with a cross-correlation function $\mathbb{E}\left[\mathsf{y}[n]\mathsf{x}[n-\ell]\right] = \rho \cdot \mathbb{1}_{\ell=d}$. Since the conditional joint distributions of $\mathsf{r}_1[n]$ and $\mathsf{r}_2[n]$, and $\mathsf{x}[n]$ and $\mathsf{y}[n]$ are identical, and since the dependence on d (via the cross-correlation function) is identical, re-introducing the (uniform) randomness of d does not change the statistical equivalence of the two models.

We note that the correlation parameter ρ encapsulates the SNR conditions in *both* sensors. That is, although (2) seemingly suggests that only the decoder observes a noisy signal, we recall the mapping $\rho = \frac{1}{\sqrt{(1+\sigma_1^2)(1+\sigma_2^2)}}$, which reflects the translation between model (1) and model (2). Therefore, it suffices, for example, that only $\sigma_1 \to \infty$ (or $\sigma_2 \to \infty$) in order to approach the zero correlation regime. Conversely, for $\rho \to 1$, we must have that $\sigma_1, \sigma_2 \to 0$, as expected.

Appendix B
Proof of Proposition 3

In order to prove the proposition, we shall use the following key lemma, whose proof is given below.

[22]Since (31) is obtained for the regime where $k, R \to \infty$, which in particular renders it inaccurate for parameter values amenable for simulation experiments, we exclude such corresponding empirical results from Fig. 8.

Lemma 1 (Lower tail upper bound of max): For any $\tau \in \mathbb{R}_+$,

$$\mathbb{P}(\mathsf{x}[\mathsf{j}] < \tau) \le e^{-2^k \left(\frac{\tau}{1+\tau^2}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{\tau^2}{2}}}. \tag{35}$$

Furthermore, if we choose $\tau_*(k) \triangleq \sqrt{2\log(2)k(1-\varepsilon(k))} = \sqrt{2\log(N)}(1+o(1))$, where $\varepsilon(k) \triangleq \frac{1}{\sqrt{k}} = o(1)$, we obtain

$$\mathbb{P}(\mathsf{x}[\mathsf{j}] < \tau_*(k)) = o\left(2^{-k}\right). \tag{36}$$

Proof of Lemma 1: For $\tau > 0$, we have

$$\begin{aligned}\mathbb{P}(\mathsf{x}[\mathsf{j}] < \tau) &= \mathbb{P}\left(\max_{0\le n\le N-1} \mathsf{x}[n] < \tau\right) && (37)\\ &= \mathbb{P}(\mathsf{x}[0] < \tau, \ldots, \mathsf{x}[N-1] < \tau) && (38)\\ &= \mathbb{P}(\mathsf{x}[0] < \tau)^N && (39)\\ &= (1 - Q(\tau))^N && (40)\\ &\le \left(1 - \left(\frac{\tau}{1+\tau^2}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{\tau^2}{2}}\right)^N && (41)\\ &\le e^{-N\left(\frac{\tau}{1+\tau^2}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{\tau^2}{2}}}, && (42)\end{aligned}$$

where we have used:

- $Q(x) \ge \frac{x}{(1+x^2)}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ for $x > 0$ in (41) [53, Eq. (10)]; and
- $1 - x \le e^{-x} \Rightarrow (1-x)^N \le e^{-xN}$ in (42).

Choosing $\tau = \tau_*(k)$ gives, after simplifying,

$$\mathbb{P}(\mathsf{x}[\mathsf{j}] < \tau_*(k)) \le e^{-2^{\sqrt{k}}\cdot\frac{1}{\sqrt{2\pi}}\left(\frac{\sqrt{2\log(2)k(1-\varepsilon(k))}}{1+2\log(2)k(1-\varepsilon(k))}\right)}, \tag{43}$$

which establishes (36). ■

Proof of Proposition 3: Our proof strategy is as follows. We first obtain an upper bound on the conditional error probability, when conditioning on the delay d and the maximal value observed by the encoder, namely $\mathsf{x}[\mathsf{j}]$. We then show that:

1) The conditional upper bound is independent of d. Therefore, since d is uniformly distributed, averaging with respect to it does not affect the bound; and
2) The randomness in $\mathsf{x}[\mathsf{j}]$ is asymptotically negligible with respect to the randomness in $\mathsf{z}[\mathsf{j}+\mathsf{d}]$ (see Lemma 1), so that we can replace $\mathsf{x}[\mathsf{j}]$ with $\mathbb{E}[\mathsf{x}[\mathsf{j}]]$ to obtain an upper bound on the unconditional error probability.

We begin by upper bounding the aforementioned conditional error probability. For this, let $\mathsf{v} \sim \mathcal{N}(0,1)$ and denote $a \wedge b \triangleq \min(a,b)$ for brevity. With these notations, we have,

$$\begin{aligned}&\mathbb{P}\left(\widehat{\mathsf{d}} \ne \mathsf{d} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) && (44)\\ &= \mathbb{P}\left(\bigcup_{\substack{\ell\in\mathcal{D}\\ \ell\ne\mathsf{d}}} \widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) && (45)\\ &\le \mathbb{E}\left[1 \wedge \left(2d_m \cdot \mathbb{P}\left(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}]\right)\right) \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right] && (46)\\ &\le \mathbb{E}\left[1 \wedge \left(2d_m \cdot \mathbb{P}\left(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{v} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}]\right)\right) \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right] && (47)\\ &\le \mathbb{P}\left(\mathsf{z}[\mathsf{j}+\mathsf{d}] < z_m \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) && (48)\\ &+ 2d_m \mathbb{P}\left(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{v} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}] \ge z_m\right) && (49)\\ &\cdot \mathbb{P}\left(\mathsf{z}[\mathsf{j}+\mathsf{d}] \ge z_m \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right), && (50)\end{aligned}$$

where:

- In (46), we have used the union bound;
- In (47), replacing $\mathsf{y}[\mathsf{j}+\ell]$ by v can only increase the probability. To see this more clearly, we first recall that $\widehat{\rho}(\ell)$ is merely a scaled version of $\mathsf{y}[\mathsf{j}+\ell]$. Then, we observe that $\mathsf{y}[\mathsf{j}+\ell] = \rho\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell] + \bar{\rho}\mathsf{z}[\mathsf{j}+\ell]$ is a convex combination of a (possibly) one-sided (upper bounded) truncated standard Gaussian RV ($\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell]$)[23] and a standard Gaussian RV ($\mathsf{z}[\mathsf{j}+\ell]$), which are independent. Since v can be thought of as a convex combination with the same coefficients of two independent standard Gaussian RVs, it is interpreted as replacing the truncated Gaussian $\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell]$ with a standard Gaussian, which can only increase the probability that $\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell]$;
- The inequality (48) holds for any $z_m \in \mathbb{R}$, since below (respectively above) z_m we bound the minimum by its first (respectively second) argument.

We are now left with the task of judicially choosing z_m, and evaluating or bounding the probabilities in (48)–(50). To this end, let

$$z_m = \frac{\sqrt{2\log(D)} - \rho\mathsf{x}[\mathsf{j}]}{\sqrt{1-\rho^2}}, \tag{51}$$

where we recall that (48)–(50) is conditioned on $\mathsf{x}[\mathsf{j}]$ (and d).

Now the probability in (48) is readily given by

$$\mathbb{P}\left(\mathsf{z}[\mathsf{j}+\mathsf{d}] < z_m \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) = Q\left(\frac{\rho\mathsf{x}[\mathsf{j}] - \sqrt{2\log(D)}}{\sqrt{1-\rho^2}}\right), \tag{52}$$

which is independent of d. As for the probability in (49), we have,

$$\begin{aligned}&\mathbb{P}\left(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{v} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}] \ge z_m\right) && (53)\\ &= \mathbb{P}\left(\rho\mathsf{x}[\mathsf{j}] < \mathsf{v} - \sqrt{1-\rho^2}\mathsf{z}[\mathsf{j}+\mathsf{d}] \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}] \ge z_m\right) && (54)\\ &\le \mathbb{P}\left(\rho\mathsf{x}[\mathsf{j}] < \mathsf{v} - \sqrt{1-\rho^2}\mathsf{z}[\mathsf{j}+\mathsf{d}] \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) && (55)\\ &= Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]}{\sqrt{2-\rho^2}}\right) && (56)\end{aligned}$$

where in (55), by removing the lower bound of $\mathsf{z}[\mathsf{j}+\mathsf{d}]$, we have only increased the probability that $\rho\mathsf{x}[\mathsf{j}] < \mathsf{v} - \sqrt{1-\rho^2}\mathsf{z}[\mathsf{j}+\mathsf{d}]$, and in (56), we have used the independence of (the standard Gaussian RVs) v and $\mathsf{z}[\mathsf{j}+\mathsf{d}]$. We further note that, similarly to (52), (56) is independent of d. At this point, upper bounding the probability (50) by 1, and further using (52) and the upper bound (56), we arrive at,

$$\begin{aligned}&\mathbb{P}\left(\widehat{\mathsf{d}} \ne \mathsf{d} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right) && (57)\\ &\le Q\left(\frac{\rho\mathsf{x}[\mathsf{j}] - \sqrt{2\log(D)}}{\sqrt{1-\rho^2}}\right) + 2d_m Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]}{\sqrt{2-\rho^2}}\right). && (58)\end{aligned}$$

Since the error probability is nothing but the expectation of the conditional error probability, we obtain,

$$\mathbb{P}\left(\widehat{\mathsf{d}} \ne \mathsf{d}\right) = \mathbb{E}\left[\mathbb{P}\left(\widehat{\mathsf{d}} \ne \mathsf{d} \,\middle|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right)\right] \tag{59}$$

[23]Recall that, if $\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell] \in \mathcal{X}_N$ then $\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell] < \mathsf{x}[\mathsf{j}]$ and is therefore truncated from above, and if $\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell] \notin \mathcal{X}_N$ then $\mathsf{x}[\mathsf{j}-\mathsf{d}+\ell] \sim \mathcal{N}(0,1)$.

$$\leq \mathbb{P}(\mathsf{x}[\mathsf{j}] < \tau_*(k)) + \mathbb{P}\Big(\widehat{\mathsf{d}} \neq \mathsf{d} \;\Big|\; \mathsf{d}, \mathsf{x}[\mathsf{j}] = \tau_*(k)\Big) \tag{60}$$

$$= \Bigg[Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2} \cdot \left(\rho\sqrt{1-\varepsilon(k)} - \tfrac{1}{\sqrt{R}}\right)^2}\right) + 2d_m Q\left(\sqrt{\tfrac{2\log(N)}{2-\rho^2} \cdot \rho^2(1-\varepsilon(k))}\right)\Bigg] (1+o(1)) \tag{61}$$

$$= \Bigg[Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2} \cdot \left(\rho - \tfrac{1}{\sqrt{R}}\right)^2}\right) + 2d_m Q\left(\sqrt{\tfrac{2\log(N)}{2-\rho^2} \cdot \rho^2}\right)\Bigg] (1+o(1)), \tag{62}$$

which gives (15) with $N = 2^k$, where we have used

$$\mathbb{P}(\mathsf{x}[\mathsf{j}] \geq \tau_*(k))\, \mathbb{E}\left[\mathbb{P}\Big(\widehat{\mathsf{d}} \neq \mathsf{d} \;\Big|\; \mathsf{d}, \mathsf{x}[\mathsf{j}]\Big) \;\big|\; \mathsf{x}[\mathsf{j}] \geq \tau_*(k)\right] \leq \mathbb{P}\Big(\widehat{\mathsf{d}} \neq \mathsf{d} \;\Big|\; \mathsf{d}, \mathsf{x}[\mathsf{j}] = \tau_*(k)\Big) \tag{63}$$

and (58) with $\mathsf{x}[\mathsf{j}] = \tau_*(k)$ in (60) and Lemma 1, particularly (36), in (62).

Finally, we observe that in order to have a vanishing upper bound, the Q-function's argument in (62) must be positive. Writing this inequality explicitly gives

$$\rho > \frac{1}{\sqrt{R}} \implies R > \frac{1}{\rho^2}, \tag{64}$$

thus establishing (13). ■

Appendix C
Proof of Proposition 4

To prove the proposition, we shall use the following lemmas, whose proofs are given in Appendix E.

Lemma 2: For any $R > R_0$,

$$\mathbb{P}\left(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \bigcap \Big\{ \bigcap_{\substack{k\in\mathcal{D} \\ k\neq \mathsf{d},\ell}} \widehat{\rho}(k) < \widehat{\rho}(\ell) \Big\}\right) = \mathbb{P}(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell))\,(1+o(1)). \tag{65}$$

Lemma 3: For any $\ell \in \mathcal{D}\backslash\{\mathsf{d}\}$,

$$\mathbb{P}(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)) \geq Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)). \tag{66}$$

Lemma 4: Let $\mathsf{q} = \max_{\ell\in\mathcal{D}\backslash\{\mathsf{d}\}} \mathsf{y}[\mathsf{j}+\ell]$, as defined in (75) (below). Then, for any fixed $\rho \in (0,1)$,

$$\mathbb{P}(\rho\mathsf{x}[\mathsf{j}] > \mathsf{q}) \geq \left(1 - \frac{1}{N^{\frac{1-\varepsilon(k)}{R_0} - \frac{1}{R}}}\right)(1+o(1)). \tag{67}$$

Lemma 5: For brevity, let $\mathsf{q} = \max_{\ell\in\mathcal{D}\backslash\{\mathsf{d}\}} \mathsf{y}[\mathsf{j}+\ell]$, as defined in (75) (below). Then,

$$\mathbb{E}\left[Q\left(\tfrac{\mathsf{q}-\rho\mathsf{x}[\mathsf{j}]}{\sqrt{1-\rho^2}}\right) \middle| \rho\mathsf{x}[\mathsf{j}] > \mathsf{q}\right] \tag{68}$$

$$\geq Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right)(1+o(1)). \tag{69}$$

Proof of Proposition 4: We will bound ϵ from below in two different ways, and then take the maximum between the two resulting lower bounds.

For the first lower bound, we have

$$\mathbb{P}\Big(\widehat{\mathsf{d}} \neq \mathsf{d}\Big) = \mathbb{P}\left(\bigcup_{\substack{\ell\in\mathcal{D} \\ \ell\neq\mathsf{d}}} \widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)\right) \tag{70}$$

$$\geq \mathbb{P}\left(\bigcup_{\substack{\ell\in\mathcal{D} \\ \ell\neq\mathsf{d}}} \Big\{\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \bigcap \Big\{ \bigcap_{\substack{k\in\mathcal{D} \\ k\neq \mathsf{d},\ell}} \widehat{\rho}(k) < \widehat{\rho}(\ell) \Big\}\Big\}\right) \tag{71}$$

$$= \sum_{\substack{\ell\in\mathcal{D} \\ \ell\neq\mathsf{d}}} \mathbb{P}\left(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \bigcap \Big\{ \bigcap_{\substack{k\in\mathcal{D} \\ k\neq \mathsf{d},\ell}} \widehat{\rho}(k) < \widehat{\rho}(\ell) \Big\}\right). \tag{72}$$

Using Lemma 2 in (72) for each summand, we obtain

$$\mathbb{P}\left(\widehat{\mathsf{d}} \neq \mathsf{d}\right) \geq 2d_m \cdot \mathbb{P}\left(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)\right)(1+o(1)) \tag{73}$$

$$\geq 2d_m \cdot Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)), \tag{74}$$

where we have used Lemma (3) in (74), thus establishing the second argument of the max operator in (19).

For the second lower bound, for brevity in the following derivation, let

$$\mathsf{q} \triangleq \max_{\ell\in\mathcal{D}\backslash\{\mathsf{d}\}} \mathsf{y}[\mathsf{j}+\ell], \tag{75}$$

with which we have,

$$\mathbb{P}\Big(\widehat{\mathsf{d}} \neq \mathsf{d}\Big) \geq \mathbb{P}\left(\widehat{\rho}(\mathsf{d}) < \max_{\ell\in\mathcal{D}\backslash\{\mathsf{d}\}} \widehat{\rho}(\ell)\right) \tag{76}$$

$$= \mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{q}) \tag{77}$$

$$= \mathbb{E}\left[\mathbb{P}\left(\mathsf{z}[\mathsf{j}+\mathsf{d}] < \tfrac{\mathsf{q}-\rho\mathsf{x}[\mathsf{j}]}{\sqrt{1-\rho^2}} \middle| \mathsf{q}, \mathsf{x}[\mathsf{j}]\right)\right] \tag{78}$$

$$= \mathbb{E}\left[Q\left(\tfrac{\mathsf{q}-\rho\mathsf{x}[\mathsf{j}]}{\sqrt{1-\rho^2}}\right)\right] \tag{79}$$

$$\geq \mathbb{P}(\rho\mathsf{x}[\mathsf{j}] > \mathsf{q})\, \mathbb{E}\left[Q\left(\tfrac{\mathsf{q}-\rho\mathsf{x}[\mathsf{j}]}{\sqrt{1-\rho^2}}\right) \middle| \rho\mathsf{x}[\mathsf{j}] > \mathsf{q}\right], \tag{80}$$

where, in fact, all the transitions above are trivial.

Now, first, for any fixed $\rho \in (0,1)$,[24] from Lemma 4,

$$\mathbb{P}(\rho\mathsf{x}[\mathsf{j}] > \mathsf{q}) \geq \left(1 - \frac{1}{N^{\frac{1-\varepsilon(k)}{R_0} - \frac{1}{R}}}\right)(1+o(1)) \tag{81}$$

$$\Rightarrow \mathbb{P}(\rho\mathsf{x}[\mathsf{j}] < \mathsf{q}) < \frac{1+o(1)}{N^{\frac{1-\varepsilon(k)}{R_0} - \frac{1}{R}}} = o(1), \tag{82}$$

where (82) holds for $R > R_0$.

As for the expectation in (80), using Lemma 5,

$$\mathbb{E}\left[Q\left(\tfrac{\rho\mathsf{x}[\mathsf{j}]-\mathsf{q}}{\sqrt{1-\rho^2}}\right) \middle| \rho\mathsf{x}[\mathsf{j}] > \mathsf{q}\right] \tag{83}$$

$$\geq Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right)(1+o(1)). \tag{84}$$

[24]The cases $\rho = 0$ and $\rho = 1$ can be addressed and analyzed separately and easily.

Thus, using (81) and (84) in (80) readily gives us

$$\mathbb{P}\left(\widehat{\mathsf{d}} \neq \mathsf{d}\right) \geq Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right)(1+o(1)). \quad (85)$$

Taking the maximum between (85) and (74) gives the lower bound $\underset{\sim}{\epsilon}(k,\rho,d_m)$ in (19), which completes the proof. ■

Appendix D
Proof of Theorem 1

We first obtain the upper bound

$$\widetilde{\epsilon}(k,\rho,d_m) \triangleq \left[2^{-k\frac{(\sqrt{R}-\sqrt{R_0})^2}{R(R_0-1)}} + 2^{-k\frac{R-2R_0+1}{(2R_0-1)R}}\right] \geq \bar{\epsilon}(k,\rho,d_m), \quad (86)$$

using the notation of the threshold rate R_0. For this, with $R > 1/\rho^2 = R_0$, the first term in (15) can simply be further upper bounded using $Q(x) \leq e^{-\frac{x^2}{2}}, \forall x > 0$, as

$$Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2}}\left(\rho - \tfrac{1}{\sqrt{R}}\right)\right) \leq \exp\left\{-\log(N)\frac{\left(\rho-\frac{1}{\sqrt{R}}\right)^2}{1-\rho^2}\right\} \quad (87)$$

$$= 2^{-k\frac{\left(\rho-\frac{1}{\sqrt{R}}\right)^2}{1-\rho^2}} = 2^{-k\frac{(\sqrt{R}-\sqrt{R_0})^2}{R(R_0-1)}}, \quad (88)$$

where we have used $N = 2^k$. Similarly, for the second term,

$$2d_m Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right) \leq 2d_m 2^{-k\frac{\rho^2}{2-\rho^2}} = 2^{-k\frac{R-2R_0+1}{(2R_0-1)R}}, \quad (89)$$

which readily gives (86).

Next, we obtain the lower bound

$$\underset{\sim}{\epsilon}(k,\rho,d_m) \triangleq \max\left\{\frac{c_1(k,R,R_0)}{2^{k\frac{(\sqrt{R}-\sqrt{R_0})^2}{R(R_0-1)}}}, \frac{c_2(k,R_0)}{2^{k\frac{R-2R_0+1}{(2R_0-1)R}}}\right\} \quad (90)$$

$$\leq \underline{\epsilon}(k,\rho,d_m), \quad (91)$$

with

$$c_1(k,R,R_0) \triangleq \sqrt{\frac{R(R_0-1)}{4\pi\log(2)k\left(\sqrt{R}-R_0\right)^2}}, \quad (92)$$

$$c_2(k,R_0) \triangleq \sqrt{\frac{2R_0-1}{4\pi\log(2)k}}, \quad (93)$$

which is similar in form to (86), and is obtained as follows. Using a lower bound on the Q-function [54, Eq. (13)],

$$2d_m Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right) \geq 2d_m\sqrt{\frac{2-\rho^2}{4\pi\rho^2\log(N)}}e^{-\frac{\rho^2}{2-\rho^2}\log(N)} \quad (94)$$

$$= 2d_m\sqrt{\frac{2-\rho^2}{4\pi\rho^2\log(N)}}\cdot 2^{-k\frac{\rho^2}{2-\rho^2}} \quad (95)$$

$$= \sqrt{\frac{2R_0-1}{4\pi\log(2)k}}\cdot 2^{-k\frac{R-2R_0+1}{(2R_0-1)R}}, \quad (96)$$

and similarly,

$$Q\left(\sqrt{\tfrac{2\log(N)}{1-\rho^2}}\left(\rho-\tfrac{1}{\sqrt{R}}\right)\right) \quad (97)$$

$$\geq \sqrt{\frac{(1-\rho^2)R}{4\pi\log(N)\left(\rho\sqrt{R}-1\right)^2}}\cdot e^{-\log(N)\frac{\left(\frac{1}{\sqrt{R_0}}-\frac{1}{\sqrt{R}}\right)^2}{1-\frac{1}{R_0}}} \quad (98)$$

$$= \sqrt{\frac{R(R_0-1)}{4\pi\log(2)k\left(\sqrt{R}-R_0\right)^2}}\cdot 2^{-k\frac{(\sqrt{R}-\sqrt{R_0})^2}{R(R_0-1)}}, \quad (99)$$

which gives (91).

Now, by sandwiching the left-hand side of (20) with $-\frac{1}{k}\log_2\left(\widetilde{\epsilon}(k,\rho,d_m)\right)$ and $-\frac{1}{k}\log_2(\underset{\sim}{\epsilon}(k,\rho,d_m))$ from below and above, respectively, and taking the limit $k\to\infty$ for any fixed $R > R_0$, we obtain

$$E_{\text{MIE}}(R) = \begin{cases} \frac{\left(\rho-1/\sqrt{R}\right)^2}{1-\rho^2}, & \frac{1}{\rho^2} < R < \left(\frac{2-\rho^2}{\rho}\right)^2 \\ \frac{\rho^2}{2-\rho^2} - \frac{1}{R}, & \left(\frac{2-\rho^2}{\rho}\right)^2 < R \end{cases}. \quad (100)$$

The form (21) is obtained by straightforward algebra using the definitions of the rate R, the threshold rate R_0 and the critical rate R_1.

Appendix E
Proofs of Lemmas 3, 2, 4 and 5

To prove Lemma 2, we shall use the following auxiliary lemma, whose proof is given below.

Lemma 6: For any $k \neq \ell \neq \mathsf{d}$, where $k,\ell,\mathsf{d} \in \mathcal{D}$,

$$\mathbb{P}(\widehat{\rho}(k) > \widehat{\rho}(\mathsf{d})|\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell))$$
$$= Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)) = o(1). \quad (101)$$

Proof of Lemma 6: Let $\mathsf{v} \sim \mathcal{N}(0,1)$ be independent of the process $\mathsf{y}[n]$. Then,

$$\mathbb{P}(\widehat{\rho}(k) > \widehat{\rho}(\mathsf{d})|\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)) \quad (102)$$
$$= \mathbb{P}(\mathsf{y}[\mathsf{j}+k] > \mathsf{y}[\mathsf{j}+\mathsf{d}]|\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell]) \quad (103)$$
$$= \frac{\mathbb{P}(\mathsf{y}[\mathsf{j}+k] > \mathsf{y}[\mathsf{j}+\mathsf{d}], \mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell])}{\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell])} \quad (104)$$
$$\leq \frac{\mathbb{P}(\mathsf{v} > \mathsf{y}[\mathsf{j}+\mathsf{d}], \mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell])}{\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell])} \quad (105)$$
$$= \frac{\mathbb{E}\left[\mathbb{P}(\mathsf{v} > \mathsf{y}[\mathsf{j}+\mathsf{d}], \mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell]|\,\mathsf{y}[\mathsf{j}+\mathsf{d}])\right]}{\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell])} \quad (106)$$
$$= \frac{1}{\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}]<\mathsf{y}[\mathsf{j}+\ell])}\Big[\mathbb{E}\left[\mathbb{P}(\mathsf{v} > \mathsf{y}[\mathsf{j}+\mathsf{d}]|\,\mathsf{y}[\mathsf{j}+\mathsf{d}])\right]$$
$$\times \underbrace{\mathbb{E}\left[\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{y}[\mathsf{j}+\ell]|\,\mathsf{y}[\mathsf{j}+\mathsf{d}])\right]}_{=\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}]<\mathsf{y}[\mathsf{j}+\ell])}\Big] \quad (107)$$
$$= \mathbb{E}\left[\mathbb{P}(\mathsf{v} > \mathsf{y}[\mathsf{j}+\mathsf{d}]|\,\mathsf{y}[\mathsf{j}+\mathsf{d}])\right] = \mathbb{P}(\mathsf{v} > \mathsf{y}[\mathsf{j}+\mathsf{d}]) \quad (108)$$
$$= \mathbb{E}\left[\mathbb{P}(\mathsf{v} - \bar{\rho}\mathsf{z}[\mathsf{j}+\mathsf{d}] > \rho\mathsf{x}[\mathsf{j}]|\,\mathsf{x}[\mathsf{j}])\right] \quad (109)$$
$$= \mathbb{E}\left[Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]}{\sqrt{2-\rho^2}}\right)\right] \leq Q\left(\rho\sqrt{\tfrac{2\log(N)(1-\varepsilon(k))}{2-\rho^2}}\right)(1+o(1)), \quad (110)$$
$$= Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)), \quad (111)$$

where we have used:

- The definition of $\widehat{\rho}(\ell)$ in (103);
- Bayes' rule in (104);

- The fact that $\mathsf{y}[\mathsf{j}+k]$ is a convex combination of a one-sided truncated normal RV ($\mathsf{x}[\mathsf{j}+k-\mathsf{d}]$, which is upper bound by $\mathsf{x}[\mathsf{j}]$) and a normal RV ($\mathsf{z}[\mathsf{j}+k]$), whereas v can be viewed as a convex combination with the same coefficients but of two independent normal RVs. Thus, when replacing $\mathsf{y}[\mathsf{j}+k]$ by v (105), the integration interval increases, and the probability cannot be decreased;
- The independence of v of the process $\mathsf{y}[n]$ in (107); and
- In (110), Lemma 1, particularly (36). ■

Proof of Lemma 2: For brevity, define $\mathcal{A} \triangleq \{\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)\}$ and $\mathcal{B} \triangleq \left\{\bigcap_{\substack{k\in\mathcal{D}\\ k\neq\mathsf{d},\ell}} \widehat{\rho}(k) < \widehat{\rho}(\ell)\right\}$. Since $\mathbb{P}(\mathcal{A}\cap\mathcal{B}) = \mathbb{P}(\mathcal{A})\,\mathbb{P}(\mathcal{B}|\mathcal{A})$, it suffices to show that $\mathbb{P}(\mathcal{B}|\mathcal{A}) = 1-o(1)$ or, equivalently, $\mathbb{P}(\overline{\mathcal{B}}|\mathcal{A}) = o(1)$, which is what we show next.

Focusing on this conditional probability, we have,

$$\mathbb{P}(\overline{\mathcal{B}}|\mathcal{A}) \tag{112}$$
$$= \mathbb{P}\left(\bigcup_{\substack{k\in\mathcal{D}\\ k\neq\ell\neq\mathsf{d}}} \widehat{\rho}(k) > \widehat{\rho}(\ell) \,\middle|\, \widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)\right) \tag{113}$$
$$\le (2d_m - 1)\mathbb{P}(\widehat{\rho}(k) > \widehat{\rho}(\ell) \,|\, \widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)) \tag{114}$$
$$\le (2d_m - 1)\mathbb{P}(\widehat{\rho}(k) > \widehat{\rho}(\mathsf{d}) \,|\, \widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)) \tag{115}$$
$$\le (2d_m - 1)Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)) \tag{116}$$
$$\Rightarrow\ \mathbb{P}(\overline{\mathcal{B}}|\mathcal{A}) = o(1) \ \Rightarrow\ \mathbb{P}(\mathcal{B}|\mathcal{A}) = 1-o(1), \tag{117}$$

where we have used the union bound in (114), the fact that replacing $\widehat{\rho}(\ell)$ by $\widehat{\rho}(\mathsf{d})$ can only increase the probability in (115), Lemma 6 in (116), and (117) is for any $R > R_0$. ■

To prove Lemma 3, we shall use the following lemma, whose proof is given below.

Lemma 7: Let $\mathsf{v},\mathsf{z} \sim \mathcal{N}(0,1)$ be independent, and $\mathsf{u} \triangleq \min(\mathsf{v}, V)$, for some $V\in\mathbb{R}$. Then, for any $a\in\mathbb{R}$,

$$\mathbb{P}(a < \rho\mathsf{u} + \bar{\rho}\mathsf{z}) \ge \mathbb{P}(a < \mathsf{v}) - Q(V). \tag{118}$$

Proof of Lemma 7: We have,

$$\mathbb{P}(a < \rho\mathsf{u} + \bar{\rho}\mathsf{z}) \tag{119}$$
$$= \mathbb{E}\left[\mathbb{P}(a < \rho\mathsf{u} + \bar{\rho}\mathsf{z} \,|\, \mathsf{z})\right] \tag{120}$$
$$= \mathbb{E}\left[\mathbb{P}\left(\tfrac{a-\bar{\rho}\mathsf{z}}{\rho} < \mathsf{u} \,|\, \mathsf{z}\right)\right] \tag{121}$$
$$= \mathbb{E}\left[\frac{1}{1-Q(V)}\int_{\frac{a-\bar{\rho}\mathsf{z}}{\rho}}^{V} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\mathrm{d}x\right] \tag{122}$$
$$\ge \mathbb{E}\left[\int_{\frac{a-\bar{\rho}\mathsf{z}}{\rho}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\mathrm{d}x - \int_{V}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\mathrm{d}x\right] \tag{123}$$
$$= \mathbb{E}\left[\mathbb{P}\left(\tfrac{a-\bar{\rho}\mathsf{z}}{\rho} < \mathsf{v} \,|\, \mathsf{z}\right)\right] - Q(V) \tag{124}$$
$$= \mathbb{P}(a < \mathsf{v}) - Q(V), \tag{125}$$

where (123) is from $[1-Q(V)]^{-1} > 1$, and (125) follows from the fact that $\rho^2 + \bar{\rho}^2 = 1$ and that v and z are independent. ■

Proof of Lemma 3: For brevity, in the following derivation let $\mathsf{v}_1 \sim \mathcal{N}(0,1)$ be independent and denote $\mathsf{t} \triangleq \mathsf{x}[\mathsf{j}+\ell-\mathsf{d}]$. With these notations, we have,

$$\mathbb{P}(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]) \tag{126}$$
$$= \mathbb{E}\left[\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \rho\mathsf{t} + \bar{\rho}\mathsf{v}_1 \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}], \mathsf{v}_1) \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right] \tag{127}$$
$$\ge \mathbb{E}\left[\mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{v}_1 \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}], \mathsf{z}[\mathsf{j}+\mathsf{d}], \mathsf{v}_1) \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]\right] - Q(\mathsf{x}[\mathsf{j}]) \tag{128}$$
$$= \mathbb{P}(\mathsf{y}[\mathsf{j}+\mathsf{d}] < \mathsf{v}_1 \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}]) - Q(\mathsf{x}[\mathsf{j}]) \tag{129}$$
$$= \mathbb{P}(\rho\mathsf{x}[\mathsf{j}] < \mathsf{v}_1 - \bar{\rho}\mathsf{z}[\mathsf{j}+\mathsf{d}] \,|\, \mathsf{x}[\mathsf{j}]) - Q(\mathsf{x}[\mathsf{j}]) \tag{130}$$
$$= Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]}{\sqrt{2-\rho^2}}\right) - Q(\mathsf{x}[\mathsf{j}]), \tag{131}$$

where (128) is from Lemma 7 and that fact that t can either be standard Gaussian or (a one-sided) truncated standard Gaussian with an upper bound $\mathsf{x}[\mathsf{j}]$, and in (130) we removed the condition on d since the noise $\mathsf{z}[n]$ is white.

Now, using Lemma 1, we have,

$$\mathbb{P}(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell)) = \mathbb{E}\left[\mathbb{P}(\widehat{\rho}(\mathsf{d}) < \widehat{\rho}(\ell) \,|\, \mathsf{d}, \mathsf{x}[\mathsf{j}])\right] \tag{132}$$
$$\ge Q\left(\rho\sqrt{\tfrac{2\log(N)}{2-\rho^2}}\right)(1+o(1)). \tag{133}$$

■

Proof of Lemma 4:

$$\mathbb{P}(\rho\mathsf{x}[\mathsf{j}] > \mathsf{q}) \ge \mathbb{P}\left(\rho\mathsf{x}[\mathsf{j}] > \max_{1\le\ell\le 2d_m} \mathsf{v}_\ell\right) \tag{134}$$
$$= \mathbb{E}\left[\mathbb{P}\left(\rho\mathsf{x}[\mathsf{j}] > \max_{1\le\ell\le 2d_m} \mathsf{v}_\ell \,\middle|\, \mathsf{x}[\mathsf{j}]\right)\right] \tag{135}$$
$$= \mathbb{E}\left[(1-Q(\rho\mathsf{x}[\mathsf{j}]))^{2d_m}\right] \tag{136}$$
$$\ge \mathbb{E}\left[(1-Q(\rho\mathsf{x}[\mathsf{j}]))^{D}\right] \tag{137}$$
$$\ge \mathbb{E}\left[\left(1-e^{-\frac{\rho^2\mathsf{x}^2[\mathsf{j}]}{2}}\right)^{N^{\frac{1}{R}}}\right](1+o(1)) \tag{138}$$
$$\ge \left(1-e^{-\rho^2\log(N)(1-\varepsilon(k))}\right)^{N^{\frac{1}{R}}}(1+o(1)) \tag{139}$$
$$\ge \left(1-\frac{N^{\frac{1}{R}}}{N^{\rho^2(1-\varepsilon(k))}}\right)(1+o(1)) \tag{140}$$
$$= \left(1-\frac{1}{N^{\frac{1-\varepsilon(k)}{R_0}-\frac{1}{R}}}\right)(1+o(1)), \tag{141}$$

where we have used:

- $\{\mathsf{v}_\ell\}$ are iid standard normal RVs independent from $\mathsf{x}[\mathsf{j}]$, and the inequality (134) follows from the fact that we replaced one-sided (upper bounded) truncated normal RVs with (unbounded) standard normal RVs, thus increasing their support, so the probability can only be decreased;
- $D > 2d_m \Rightarrow \gamma^{2d_m} \ge \gamma^D$ for any $\gamma\in[0,1]$ in (137);
- $\forall x \ge 0 : Q(x) \le e^{-\frac{x^2}{2}}$, $D = N^{\frac{1}{R}}$ (by (10)) in (138);[25]
- Lemma 1 in (139); and
- $(1-\varepsilon)^n \ge 1 - n\varepsilon$ [55] in (140), which holds for any $0\le\varepsilon\le 1$ and $n\in\mathbb{R}_+$ (recall $N^{1/R} = D \in \mathbb{R}_+$). ■

[25]The $1+o(1)$ factor is due to the fact that $\mathbb{P}(\mathsf{x}[\mathsf{j}] > 0) = 1 - \left(\frac{1}{2}\right)^{2^k}$.

Proof of Lemma 5: As for the expectation in (80), using Jensen's inequality,

$$\mathbb{E}\left[Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]-\mathsf{q}}{\sqrt{1-\rho^2}}\right)\middle|\,\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}\right]\geq Q\left(\frac{\rho\mathbb{E}[\mathsf{x}[\mathsf{j}]|\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}]-\mathbb{E}[\mathsf{q}|\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}]}{\sqrt{1-\rho^2}}\right), \tag{142}$$

since the expectation is conditioned on $\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}$, which implies that the Q-function's argument is positive, and the Q-function is convex on $\mathbb{R}_+$. Proceeding, we have,

$$\mathbb{E}\left[\mathsf{x}[\mathsf{j}]|\,\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}\right]=\frac{\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right]-\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathsf{x}[\mathsf{j}]|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q}\right]}{\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]>\mathsf{q})} \tag{143}$$

$$\leq\frac{\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right]}{\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]>\mathsf{q})}=\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right](1+o(1)), \tag{144}$$

where we have used (141), and that (143) is positive, in (144). Furthermore, we have,

$$\mathbb{E}\left[\mathsf{q}|\,\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}\right]=\frac{\mathbb{E}\left[\mathsf{q}\right]-\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathsf{q}|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q}\right]}{\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]>\mathsf{q})} \tag{145}$$

$$\geq\sqrt{2\log(D)}(1+o(1)), \tag{146}$$

since $\mathbb{E}\left[\mathsf{q}\right]=\sqrt{2\log(2d_m)}(1+o(1))=\sqrt{2\log(D)}(1+o(1))$ and

$$\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathsf{q}|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q}\right] \tag{147}$$

$$=\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathbb{E}\left[\mathsf{q}|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q},\mathsf{q}<\mathsf{x}[\mathsf{j}],\mathsf{x}[\mathsf{j}]\right]|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q}\right] \tag{148}$$

$$\leq\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathsf{x}[\mathsf{j}]|\,\rho\mathsf{x}[\mathsf{j}]<\mathsf{q}\right] \tag{149}$$

$$\leq\mathbb{P}(\rho\mathsf{x}[\mathsf{j}]<\mathsf{q})\,\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right] \tag{150}$$

$$\leq\frac{\sqrt{2\log(N)}}{N^{\frac{1-\varepsilon(k)}{R_0}-\frac{1}{R}}}(1+o(1))=o(1), \tag{151}$$

where we have used (82) and $\mathbb{E}\left[\mathsf{x}[\mathsf{j}]\right]=\sqrt{2\log(N)}(1+o(1))$ in (151).

Therefore, with (141) and (146), we obtain,

$$\mathbb{E}\left[Q\left(\frac{\rho\mathsf{x}[\mathsf{j}]-\mathsf{q}}{\sqrt{1-\rho^2}}\right)\middle|\,\rho\mathsf{x}[\mathsf{j}]>\mathsf{q}\right] \tag{152}$$

$$\geq Q\left(\frac{\rho\mathbb{E}[\mathsf{x}[\mathsf{j}]]-\sqrt{2\log(D)}}{\sqrt{1-\rho^2}}(1+o(1))\right) \tag{153}$$

$$=Q\left(\sqrt{\frac{2\log(N)}{1-\rho^2}}\left(\rho-\frac{1}{\sqrt{R}}\right)\right)(1+o(1)). \tag{154}$$

■

Acknowledgment

The first author wishes to thank Omri Lev and Alejandro Lancho for helpful discussions on the analysis of the proposed method.

References

[1] F. Viola and W. F. Walker, "A comparison of the performance of time-delay estimators in medical ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 4, pp. 392–401, Apr. 2003.

[2] Y. Li, "Position and time-delay calibration of transducer elements in a sparse array for underwater ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 53, no. 8, pp. 1458–1467, Aug. 2006.

[3] D. Musicki, R. Kaune, and W. Koch, "Mobile emitter geolocation and tracking using TDOA and FDOA measurements," *IEEE Trans. Signal Process*, vol. 58, no. 3, pp. 1863–1874, Mar. 2009.

[4] A. Weiss, T. Arikan, H. Vishnu, G. B. Deane, A. C. Singer, and G. W. Wornell, "A semi-blind method for localization of underwater acoustic sources," *IEEE Trans. Signal Process*, vol. 70, pp. 3090–3106, 2022.

[5] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 15, no. 3, pp. 386–391, May 1969.

[6] A. Quazi, "An overview on the time delay estimate in active and passive systems for target localization," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 29, no. 3, pp. 527–533, Jun. 1981.

[7] J. P. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 30, no. 6, pp. 998–1003, Dec. 1982.

[8] A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation–part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 31, no. 2, pp. 472–486, Apr. 1983.

[9] E. Weinstein and A. Weiss, "Fundamental limitations in passive time-delay estimation–part II: Wide-band systems," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 32, no. 5, pp. 1064–1078, Oct. 1984.

[10] M. Azaria and D. Hertz, "Time delay estimation by generalized cross correlation methods," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 32, no. 2, pp. 280–285, Apr. 1984.

[11] A. Fertner and A. Sjolund, "Comparison of various time delay estimation methods by computer simulation," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 34, no. 5, pp. 1329–1330, Jan. 1986.

[12] G. C. Carter, "Coherence and time delay estimation," *Proc. IEEE*, vol. 75, no. 2, pp. 236–255, Feb. 1987.

[13] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process*, vol. 41, no. 2, pp. 525–533, Feb. 1993.

[14] M. S. Brandstein and H. F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. Int. Conf. Acoust., Speech, Signal Process., vol.* 1, Apr. 1997, pp. 375–378.

[15] S. Björklund, "A survey and comparison of time-delay estimation methods in linear systems," M.S. thesis, 2003. Dept. Elect. Eng., Linkoping Univ., Linköping, Sweden.

[16] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 509–519, Sep. 2004.

[17] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process*, vol. 2006, pp. 1–19, Dec. 2006.

[18] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey," *IEEE Trans. Ind. Informat*, vol. 10, no. 4, pp. 2233–2243, Jan. 2014.

[19] M. Liu, L. Cheng, K. Qian, J. Wang, J. Wang, and Y. Liu, "Indoor acoustic localization: A survey," *Human-Centric Comput. Inf. Sci.*, vol. 10, pp. 1–24, Jan. 2020.

[20] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[21] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, Mar. 2002.

[22] H. Wang, D. Estrin, and L. Girod, "Preprocessing in a tiered sensor network for habitat monitoring," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 4, pp. 1–10, 2003.

[23] D. J. Matthiesen and G. Miller, "Data transfer minimization for coherent passive location systems," *Report No. esDTRY81Y129, Air Force Project*, no. 4110, 1981.

[24] M. L. Fowler, "Coarse quantization for data compression in coherent location systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1269–1278, Oct. 2000.

[25] M. L. Fowler, "Data compression for emitter location," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA: Princeton Univ., Mar. 2000.

[26] M. L. Fowler, "New distortion measures for data compression for emitter location," in *Proc. Conf. Rec. 35th Asilomar Conf. Signals Syst. Comput. (Cat. No. 01CH37256)*, vol. 1, pp. 658–662, Nov. 2001.

[27] M. L. Fowler, "Non-MSE wavelet-based data compression for emitter location," *Math. Data/Image Coding, Compression, Encryption IV, Appl.*, vol. 4475, pp. 13–22, Dec. 2001.

[28] M. Chen and M. L. Fowler, "Optimizing non-MSE distortion for data compression in emitter location systems," in *Proc. Conf. Inf. Sci. Syst.*, Mar. 2003, pp. 243–254.

[29] M. L. Fowler and M. Chen, "Fisher-information-based data compression for estimation using two sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 1131–1137, Jul. 2005.

[30] M. Chen and M. L. Fowler, "Data compression for multi-parameter estimation for emitter location," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 308–322, Jan. 2010.

[31] L. Vasudevan and A. Ortega, "Processing-aware compression for sensor networks," in *Proc. Sensor Array and Multichannel Signal Process. Workshop Proc.,* Aug. 2002, pp. 239–243.
[32] L. Vasudevan, A. Ortega, and U. Mitra, "Application-specific compression for time delay estimation in sensor networks," in *Proc. 1st Int. Conf. Embedded Netw. Sensor Syst.*, Nov. 2003, pp. 243–254.
[33] L. Vasudevan, A. Ortega, and U. Mitra, "Jointly optimized quantization and time delay estimation for sensor networks," in *Proc. 1st Int. Symp. Control, Commun. Signal Process.,* Mar. 2004, pp. 203–208.
[34] L. Vasudevan, A. Ortega, and U. Mitra, "Application-optimized quantizers for time delay estimation in sensor networks," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Sep. 2003, pp. 9–12.
[35] U. Hadar and O. Shayevitz, "Distributed estimation of Gaussian correlations," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5323–5338, Sep. 2019.
[36] Y. Kochman and L. Wang, "On the communication exponent of distributed testing for Gaussian correlations," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Apr. 2021, pp. 1–5.
[37] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
[38] M. L. Fowler and X. Hu, "Signal models for TDOA/FDOA estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1543–1550, Oct. 2008.
[39] X. Zhang, H. Li, J. Liu, and B. Himed, "Joint delay and doppler estimation for passive sensing with direct-path interference," *IEEE Trans. Signal Process*, vol. 64, no. 3, pp. 630–640, Feb. 2015.
[40] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.
[41] A. Weiss, A. Lancho, Y. Bu, and G. W. Wornell, "A bilateral bound on the mean-square error for estimation in model mismatch," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2023, pp. 2655–2660.
[42] H. A. David and H. N. Nagaraja, *Order Statistics.* Hoboken, NJ, USA: Wiley, 2004.
[43] A. Weiss, Y. Kochman, and G. W. Wornell, "A joint data compression and time-delay estimation distributed systems via extremum encoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2024, pp. 9366–9370.
[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Hoboken, NJ, USA: Wiley, 1999.
[45] X. Huang and B. Liao, "One-bit MUSIC," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 961–965, Apr. 2019.
[46] A. Bhandari, M. H. Conde, and O. Loffeld, "One-bit time-resolved imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1630–1641, Apr. 2020.
[47] S. Sedighi, K. V. Mishra, M. B. Shankar, and B. Ottersten, "Localization with one-bit passive radars in narrowband internet-of-things using multivariate polynomial optimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 2525–2540, 2021.
[48] A. Weiss and G. W. Wornell, "One-bit direct position determination of narrowband Gaussian signals," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Jul. 2021, pp. 466–470.
[49] L. Ni, D. Zhang, Y. Sun, N. Liu, J. Liang, and Q. Wan, "Detection and localization of one-bit signal in multiple distributed subarray systems," *IEEE Trans. Signal Process*, vol. 71, pp. 2776–2791, 2023.
[50] R. Zamir, Y. Kochman, and U. Erez, "Achieving the Gaussian rate–distortion function by prediction," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3354–3364, Jul. 2008.
[51] R. C. Shit, S. Sharma, D. Puthal, and A. Y. Zomaya, "Location of things (LoT): A review and taxonomy of sensors localization in IoT infrastructure," *IEEE Commun. Surv. Tuts.*, vol. 20, no. 3, pp. 2028–2061, Jan. 2018.
[52] A. Weiss, Y. Kochman, and G. W. Wornell, "Extremum encoding for joint baseband signal compression and time-delay estimation for distributed systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2025., pp. 1–5.
[53] R. D. Gordon, "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *Ann. Math. Statist.*, vol. 12, no. 3, pp. 364–366, 1941.
[54] G. Abreu, "Very simple tight bounds on the Q-function," *IEEE Trans. Commun.*, vol. 60, no. 9, pp. 2415–2420, Sep. 2012.
[55] M. Zakai and J. Ziv, "On the threshold effect in radar range estimation (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 15, no. 1, pp. 167–170, Jan. 1969.

Amir Weiss (Senior Member, IEEE) received the B.Sc. (*magna cum laude*), M.Sc. and Ph.D. degrees in electrical engineering from Tel Aviv University (TAU), Tel-Aviv, Israel, in 2013, 2015, and 2020, respectively. From 2019 to 2020, he was a Postdoctoral Fellow with the Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel, and from 2020 to 2023, he was a Postdoctoral Associate with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. His main research areas are in statistical and digital signal processing, estimation theory and machine learning, with applications to localization, communication, and compression. He has held Researcher position with Elbit Systems, EW and SIGINT Elisra Ltd., Holon, Israel, from 2013 to 2020, specializing in detection and estimation of RADAR and SONAR signals. Since 2024, he has been on the faculty with Bar-Ilan University, Ramat Gan, Israel, where he is a Senior Lecturer (Associate Professor equivalent) with the Faculty of Engineering. He is also the Head of AI and Technological Innovation with Medibyt Ltd., Israel. He received the scholarship for excellent M.Sc. students from the Faculty of Engineering, TAU, in 2015, and The David and Paulina Trotsky Foundation Award for outstanding Ph.D. students in 2019. He is a recipient of the 2021 ICASSP Outstanding Paper Award. He also received the Nadav Levanon Studies Prize for graduate students in 2016, the Scientific Publication Prize (three times) in 2017, 2018, and 2020, and the David Burshtein Scientific Publication Prize in 2019, all from The Yitzhak and Chaya Weinstein Research Institute for Signal Processing. He is a recipient of the 2024 Alon Scholarship for the Integration of Outstanding Faculty.

Yuval Kochman (Member, IEEE) received the B.Sc. (*cum laude*), M.Sc. (*cum laude*), and Ph.D. degrees from Tel Aviv University, in 1993, 2003, and 2010, respectively, all in electrical engineering. From 2009 to 2011, he was a Postdoctoral Associate with the Signals, Information and Algorithms Laboratory, MIT. Since 2012, he has been with the School of Computer Science and Engineering, Hebrew University of Jerusalem. Outside academia, he has worked in the areas of radar and digital communications. His research interests include information theory, communications, and signal processing.

Gregory W. Wornell (Fellow, IEEE) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from Massachusetts Institute of Technology, in 1985, 1987, and 1991, respectively, all in electrical engineering and computer science. Since 1991, he has been on the faculty with MIT, where he is the Sumitomo Professor of Engineering in the Department of Electrical Engineering and Computer Science (EECS) and the Schwarzman College of Computing. At MIT he leads the Signals, Information, and Algorithms Laboratory, and is affiliated with the Research Laboratory of Electronics (RLE), the Computer Science and Artificial Intelligence Laboratory (CSAIL), and the Institute for Data, Systems and Society (IDSS). He has held visiting appointments with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA, in 1999–2000, with Hewlett-Packard Laboratories, Palo Alto, CA, in 1999, and with AT&T Bell Laboratories, Murray Hill, NJ, in 1992–1993. His research interests and publications span the areas of signal processing, information theory, statistical inference, artificial intelligence, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the information theory and signal processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching, including the 2019 IEEE Leon K. Kirchmayer Graduate Teaching Award.