Parameter Estimation Based on Noisy Chaotic Signals in the Weak-Noise Regime

Neri Merhav
The Viterbi Faculty of ECE
Technion - Israel Institute of Technology
Technion City, Haifa 3200003, Israel
email: merhav@technion.ac.il

***Abstract*—We consider the problem of parameter estimation, based on noisy chaotic signals, from the viewpoint of twisted modulation for waveform communication. In particular, we study communication systems where the parameter to be estimated is conveyed as the initial condition of a chaotic dynamical system of a certain class and we examine its estimation performance in terms of the expectation of a given convex function of the estimation error at high SNR, under the demand that the probability of anomaly is kept small. We derive a lower bound on the weak-noise estimation error for this class of chaotic modulators, and argue that it can be outperformed by using the itinerary signal associated with the chaotic system instead of the main chaotic output signal.**

Index Terms: modulation, estimation, chaos, dynamical system, channel capacity.

I. Introduction

We revisit the well known problem of twisted modulation and estimation, that is, conveying the value of a parameter θ by n uses of an additive white Gaussian noise (AWGN) channel,

$$y_i = x_i + z_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where x_i is the i–th component of a channel input vector, $x^n = (x_1, x_2, \ldots, x_n) = f_n(\theta)$, that depends on the parameter θ, and that is subjected to a power constraint, $\|\boldsymbol{x}\|^2 \leq nQ$, $\{z_i\}$ are independent, zero–mean, Gaussian random variables with variance σ^2, and y_i is the i–th coordinate of the channel output vector, $y^n = (y_1, y_2, \ldots, y_n)$. In general, the main interest, in twisted modulation and estimation, is to quantify how well can one estimate θ based on y^n when it is allowed to optimize both the modulator, f_n, and the estimator. In particular, how fast does the estimation error decay as a function of n when the best modulator and estimator are used? In this work, these questions are addressed in the context of modulators that are based on a certain class of chaotic dynamical systems, but before we confine ourselves to those chaotic modulators, we first discuss the twisted modulation problem in general.

A central well known problem, that is inherent to all non–linear (twisted) modulators and receivers, is the *threshold effect* (see, e.g., [18, Chap. 8]). The threshold effect amounts to a sudden transition between two modes of behavior when the signal-to-noise ratio (SNR) crosses a certain critical value. For large SNR, namely, in the *weak–noise* regime, the estimation error of the ML estimator behaves similarly as that of a (locally) linear modulator, and so, it roughly achieves the Cramér–Rao lower bound (CRLB). But beyond a certain noise level, the estimation performance breaks down completely and rather abruptly. As explained in [18, Chap. 8], for a given non–linear modulator, one can identify a certain *anomaly event* (or outage event), whose probability becomes considerably large as the threshold SNR is crossed downwards.

In [12], the approach that was adopted was to separate the weak-noise performance from the probability of the anomaly event. Both in the lower bound and in the upper bound, an anomaly event was allowed, with the freedom to define it, depending on the communication system itself. Under the constraint that the probability of anomaly would not exceed the function $e^{-\alpha n}$ for some given constant $\alpha > 0$, the fastest possible decay rate of the weak-noise estimation error was studied in [12]. As the performance metric in the weak-noise regime, we chose in [12] the expectation of an arbitrary symmetric, convex function of the estimation error. In particular, we derived in [12] an upper bound and a lower bound, which agree in the limit of high SNR for a certain range of values of α.

In this paper, we study twisted modulators that are based on a certain class of chaotic dynamical systems, where the parameter θ to be conveyed is encoded in the initial condition of the chaotic sequence that is transmitted into the channel. More specifically, we consider modulators based on chaotic systems whose transmitted output sequence is (a scaled and shifted version of) $s^n = (s_1, s_2, \ldots, s_n)$, that is generated recursively according to

$$s_{i+1} = q(s_i), \qquad i = 0, 1, 2, \ldots \tag{2}$$

where the initial condition is $s_0 = \theta$ and $q : [0,1] \to [0,1]$ is a piece-wise linear function whose (absolute) slope is larger than unity at all points of continuity of q (i.e., almost everywhere). The use of chaotic dynamical systems for modulation and estimation has become very popular in the last three decades, see, e.g., [1], [3], [2], [5], [8], [10], [11], [13], [14], [16], [17], [19], [20] for a non-exhaustive sample of references, which altogether cover the topic from a variety of aspects, including upper and lower bounds on Bayesian/non-Bayesian estimation, numerical aspects, algorithmic efficiency, system optimization, extensions and applications in Turbo coding, in hybrid coding, in spread spectrum systems, and in MIMO systems, just to name a few. There are several reasons for the popularity of chaotic systems in the context of signal modulation and parameter estimation:

1) The output of a dynamical chaotic system is extremely senstive to the initial condition. This is a very desirable property for accurate estimation of θ at the receiver which observes a noisy version of s^n, or some memoryless tranformation of s^n. At least at high SNR, the estimation error can be made exponentially small as a function of n.
2) It is computationally easy to generate the modulated signal, simply by recursive application of the non-linear mapping of the dynamical system.
3) There exist computationally efficient estimation algorithms, such as the halving method [9] and its extensions [17], which are especially suitable to the kind of systems considered here. As reported in [9] and [17], the performance of these methods is close to the CRLB at least at high SNR.
4) As discussed in [9], once we have a good estimate of the initial condition, $s_0 = \theta$, we actually have an estimate of the entire clean signal, and so, a solution to the parameter estimation problem is also useful for filtering.

The contribution of the present work is that it studies chaotic modulators systematically within the framework of waveform communication established in [18, Chap. 8]. In particular, we examine the above described class of chaotic systems from the perspective of the general results of [12] and thereby derive a lower bound on the weak-noise estimation error subject to the requirement of small probability of anomaly. It is believed that this framework can be useful for examining other classes of modulators (chaotic and others) as well. In contrast to many of the above cited papers on chaotic modulation, we avoid the use of the Cramér-Rao lower bound, which is problematic due to the fact that, for our class of chaotic modulators, the signal vector is a not continuous function (and a-fortiori, not differentiable) of the parameter. We use instead the lower bound of [12] with the appropriate adjustments required. An additional benefit of avoiding the Cramér-Rao lower bound is that our figure of merit is more general than the mean square error - it is defined as the asymptotic exponential decay rate of the expectation of $\rho(\epsilon)$, where $\epsilon = \hat{\theta} - \theta$ is estimation error and $\rho(\cdot)$ is an arbitrary convex even cost function with $\rho(0) = 0$.

The derivation of this lower bound yields a limitation on a certain design parameter of the system, depending on the SNR (details will follow in the sequel). Then, we show that by feeding the channel with a secondary output sequence from the chaotic system, called the itinerary sequence (which is obtained by a certain quantization of $\boldsymbol{s}$), instead of $\boldsymbol{s}$, yields better estimation performance than that of any modulator from the aforementioned class. In fact, it asymptotically achieves the lower bound of [12], which applies to any non-linear modulator operating at a given high SNR. At first glance, this may seem counter-intuitive because the itinerary sequence seems to convey "less information" about the parameter, but a little thought easily settles this conflict.

Finally, we discuss the usefulness of the same class of chaotic systems for the purpose of simulation of random processes, and well as its extension to systems with long range memory, where eq. (2) is replaced by

$$s_{i+1} = q_i(s_0, s_1, \ldots, s_i), \qquad i = 0, 1, 2, \ldots. \tag{3}$$

It turns out that for the memoryless Gaussian channel considered here, our main conclusions continue to apply, in other words, there is no benefit in using memory of the remote past.

The outline of the remaining part of the paper is as follows. In Section II, we define the class of chaotic dynamical systems and their corresponding modulators. In Section III, we explore some of the properties of these dynamical systems. In Section IV, we derive a lower bound to the weak-noise estimation performance in the non-outage event. In Section V, describe an alternative method of using the itinerary signal and show that its weak-noise performance can approach the the lower bound arbitrarily closely. Finally, in Section VI, we outline the extension to systems with long-range memory and also discuss how such systems can serve also as optimal process simulators.

II. The Communication System Setting

In this section, we present the class of chaotic systems that we study in this work, explain how they are used for parameter modulation, and formalize the objectives of this work.

We begin with the description of the class of dynamical systems that the modulators are based upon. Each system in this class is defined by a piece-wise linear map, parametrized by a positive integer $r \geq 2$ and a probability vector $P = \{p(0), p(1), \ldots, p(r-1)\}$ with strictly positive entries, i.e., $p(x) > 0$ for all $x \in \{0, 1, \ldots, r-1\}$ and $\sum_{x=0}^{r-1} p(x) = 1$. Given r and P, we define

$$F(x) = \sum_{x'=0}^{x-1} p(x'), \qquad x \in \{0, 1, \ldots, r-1\}, \tag{4}$$

with the convention that the summation over the empty set is defined as zero, namely, $F(0) = 0$. Clearly, $F(r) = 1$. Next, define

$$G(x) = \frac{F(x)}{p(x)}, \qquad x \in \{0, 1, \ldots, r-1\}. \tag{5}$$

Given a real $s \in [0, 1]$, let $\phi(s)$ be defined as the unique value of $x \in \{0, 1, \ldots, r-1\}$ such that $F(x) \leq s < F(x+1)$, where for $x = r-1$, the strong inequality is allowed to be weak, namely, $\phi(1) = r - 1$. This class of dynamical systems was also considered in earlier works, such as [17].

The dynamical system is defined as follows. Given an initial state, $s_0 \in [0, 1]$, it generates recursively two sequences, $\boldsymbol{s} = (s_1, s_2, \ldots)$ and $\boldsymbol{x} = (x_1, x_2, \ldots)$, as follows: For $t = 1, 2, \ldots$, let

$$x_t = \phi(s_{t-1}); \quad s_t = \frac{s_{t-1} - F(x_t)}{p(x_t)}. \tag{6}$$

We henceforth refer to $\boldsymbol{s}$ as the *main output sequence*, or as the *state sequence* of the dynamical system, and to $\boldsymbol{x}$ – as the *itinerary sequence* (see also [17]). Note that for $r = 2$ and

$p(0) = p(1) = \frac{1}{2}$, this is nothing but the well known dyadic map, where $s_t = (2s_{t-1}) \bmod 1$ where $u \bmod 1$ designates the fractional part of u, namely, $u \bmod 1 = u - \lfloor u \rfloor$. Accordingly, $x_1, x_2, \ldots$ are the bits of the binary representation of s_0 as $0.x_1x_2,\ldots$. Likewise, for a general r and $p(0) = p(1) = \cdots = p(r-1) = \frac{1}{r}$ this is the more general saw-tooth map $s_t = (rs_{t-1}) \bmod 1$, and then $s_0 = 0.x_1x_2\ldots$ is the radix-r representation, that is, $s_0 = \sum_{i=1}^{\infty} x_i r^{-i}$. Accordingly, the system (6) can be thought of as an extension of these maps that forms a mixed basis representation of s_0. More details will follow in Section III.

The modulators that we consider in this work, are induced by the above defined dynamical systems as follows. Given a parameter value, $\theta \in [0,1]$, we set the initial state to be $s_0 = \theta$ and then generate the vector $(s_1, s_2, \ldots, s_n)$ according to (6).[1] Given a prescribed power budget, Q, we then define the channel input signal according to

$$u_t = \sqrt{12Q}\left(s_t - \frac{1}{2}\right), \qquad t = 1, 2, \ldots, n, \tag{7}$$

so that if each s_t is uniformly distributed across the interval $[0,1]$ (as will be established below), then u_t is zero-mean and its variance (which is also its power) is exactly Q. The vector $u^n = (u_1, u_2, \ldots, u_n)$ is fed into an additive white Gaussian noise (AWGN) channel with noise variance σ^2. The signal-to-noise ratio (SNR) is then defined as

$$\gamma = \frac{Q}{\sigma^2}. \tag{8}$$

Let the channel output vector be denoted by $y^n = (y_1, y_2, \ldots, y_n)$, where for $i = 1, 2, \ldots, n$, $y_i = u_i + z_i$, $\{z_i\}$ being realizations of $\{Z_i\}$, which are Gaussian, zero-mean, independently and identically distributed (i.i.d.) random variables with variance σ^2. Correspondingly, $\{y_i\}$ are realizations of $\{Y_i\}$ with $Y_i = u_i + Z_i$, $i = 1, 2, \ldots, n$. The receiver implements an estimator of θ based on Y^n, namely,

$$\hat{\theta} = \varphi(y^n). \tag{9}$$

For every $\theta \in [0,1]$, let $\mathcal{O}_n(\theta) \subset \mathbb{R}^n$ designate an event defined in the space of noise vectors, $\{\boldsymbol{z}\}$, which is henceforth referred to as the *outage event* (or the *anomalous error event*) given θ. Following the framework of [12], for a given convex, even error cost function $\rho(\cdot)$, with $\rho(0) = 0$, our first objective is to derive a lower bound to

$$\sup_{\theta \in [0,1]} \boldsymbol{E}\left\{\rho(\varphi(Y^n) - \theta)\middle|\mathcal{O}_n^{\rm c}(\theta)\right\} \tag{10}$$

for any given family of outage events (to be optimized as part of the design of the communication system), subject to the constraint that

$$\sup_{\theta \in [0,1]} \Pr\{\mathcal{O}_n(\theta)\} \le e^{-\alpha n}, \tag{11}$$

where $\alpha > 0$ is prescribed constant, henceforth referred to as the *outage exponent*. Here, the expectation $\boldsymbol{E}\{\cdot\}$ in (10) and the probability $\Pr\{\cdot\}$ in (11) are defined with respect to (w.r.t.) the randomness of the noise vector, $Z_1, Z_2, \ldots, Z_n$. The achievability of this lower bound will amount to the specification of a modulator, a receiver that estimates θ, and a family of outage events $\{\mathcal{O}_n(\theta),\ 0 \le \theta \le 1\}$, similarly as in [12], except that here we confine ourselves to chaotic modulators. We will be interested, first and foremost, in the most relaxed version of the constraint, where $\sup_{\theta \in [0,1]} \Pr\{\mathcal{O}_n(\theta)\}$ is merely required to tend to zero, without commitment to a particular exponential rate. This amounts to the choice $\alpha = \alpha_n \to 0$ in the limit of large n, but not too fast. In particular, α_n should be $\omega(1/n)$ so that $e^{-n\alpha_n}$ still tends to zero as n grows without bound.

The reason for this relaxation of the constraint (11) is as follows: while the lower bound to be presented in the sequel can be derived with a general value of $\alpha > 0$ in mind (and indeed, α will be taken to zero only at the final step of the derivation therein), the achievability of such a bound remains ellusive unless α vanishes, as we are not aware of any method to achieve the lower bound using chaotic systems. We are only aware of the fact that lattice codes with nearly spherical Voronoi cells, as described in [12], are optimal for general values of α in a certain range. The special property possessed by such lattice codes is that their decoding error event (considered as the outage event) is asymptotically equivalent to the event that the norm of the noise vector would exceed a certain threshold (which is the effective radius of the Voronoi cell).

III. Properties of the Dynamical Systems

In this section, we explore some properties of the non-linear dynamical systems defined in Section II. Some of these properties may be of interest on their own right, and most of them are moreover crucial for our derivation of the lower bound on the weak-noise estimation performance in Section IV.

1. *Reconstructing the initial state from the itinerary sequence.* The first property that is important to mention is the following: given the infinite itinerary sequence, $\boldsymbol{x} = (x_1, x_2, \ldots)$, one can reconstruct the initial state, s_0, using the following expression:

$$\begin{aligned} s_0 &= w(x_1, x_2, \ldots) \\ &\stackrel{\triangle}{=} \sum_{t=1}^{\infty} F(x_t) \prod_{i=1}^{t-1} p(x_i) \\ &\equiv \sum_{t=1}^{\infty} G(x_t) \prod_{i=1}^{t} p(x_i), \end{aligned} \tag{12}$$

where a product over an empty set of indices $(\prod_{i=1}^{0} p(x_i))$ is defined as unity. It is therefore implied that although the itinerary sequence, $\boldsymbol{x} = (x_1, x_2, \ldots)$, is a certain quantized version of the state sequence, $\boldsymbol{s} = (s_1, s_2, \ldots)$, they both bear exactly the same information, as there is one-to-one correspondence between them via s_0: $\boldsymbol{x}$ determines s_0 via (12), which in turn determines $\boldsymbol{s}$. But it should be kept in mind that this one-to-one correspondence holds only for the infinite

[1] More generally, we may set s_0 to be some given function of θ.

sequences. For finite n, this is no longer true. As mentioned earlier, the last line of (12) can be thought of as a mixed basis representation of s_0 in the sense that in the special case where P is the uniform distribution, i.e., $p(x) = 1/r$ for all $x \in \{0, 1, \ldots, r\}$ and then $G(x) = x$, it boils down to the radix r representation associated with the saw-tooth map, that is,

$$s_0 = \sum_{t=1}^{\infty} x_t \cdot r^{-t}, \tag{13}$$

with $\{x_t\}$ in the role of the digits, i.e., $s_0 = 0.x_1x_2x_3\ldots$.

The relation (12) applies, of course, also to the time-shifted versions of the sequences, i.e.,

$$\begin{aligned} s_\tau &= w(x_{\tau+1}, x_{\tau+2}, \ldots) \\ &= \sum_{t=\tau+1}^{\infty} F(x_t) \prod_{i=\tau+1}^{t-1} p(x_i) \end{aligned} \tag{14}$$

for every positive integer τ.

To see why eq. (12) holds true, first observe that if $s_0 \in [0, 1]$, then so is s_1, and then by the same token, the same is true for s_2, etc., which means that $s_t \in [0, 1]$ for all $t \geq 0$. Inverting the recursion (6) in $\{s_t\}$, we have

$$s_{t-1} = F(x_t) + p(x_t)s_t, \tag{15}$$

and so,

$$\begin{aligned} s_0 &= F(x_1) + s_1 p(x_1) \\ &= F(x_1) + p(x_1)[F(x_2) + s_2 p(x_2)] \\ &= F(x_1) + p(x_1)\{F(x_2) + p(x_2)[F(x_3) + s_3 p(x_3)]\} \\ &= F(x_1) + F(x_2)p(x_1) + F(x_3)p(x_1)p(x_2) + \\ &\quad s_3 p(x_1)p(x_2)p(x_3). \end{aligned} \tag{16}$$

Likewise, more generally, for every positive integer τ,

$$s_0 = \sum_{t=1}^{\tau} F(x_t) \prod_{i=1}^{\tau-1} p(x_i) + s_\tau \prod_{i=1}^{\tau} p(x_i), \tag{17}$$

which implies that

$$\begin{aligned} &\sum_{t=1}^{\tau} F(x_t) \prod_{i=1}^{\tau-1} p(x_i) \leq s_0 \\ \leq &\sum_{t=1}^{\tau} F(x_t) \prod_{i=1}^{\tau-1} p(x_i) + [\max_x p(x)]^\tau, \end{aligned} \tag{18}$$

where we have used the fact that $0 \leq s_\tau \leq 1$. It now follows that when $\tau \to \infty$, s_0 is sandwiched between two sequences (indexed by τ) that both tend to $\sum_{t=1}^{\infty} F(x_t) \prod_{i=1}^{t-1} p(x_i)$.

2. *The itinerary sequence as a random process.* Let S_0 be a random variable, uniformly distributed across the unit interval, $[0, 1]$, and let $\boldsymbol{S} = (S_1, S_2, \ldots)$ and $\boldsymbol{X} = (X_1, X_2, \ldots)$ the corresponding sequences of random variables generated from S_0 according to the recursion (6). Observe that the numbers, $F(x)$, $x \in \{0, 1, \ldots, r-1\}$ can be viewed as endpoints of successive, non-overlapping sub-intervals of lengths $p(x)$, $x \in \{0, 1, \ldots, r-1\}$, forming a partition of the unit interval. It is therefore clear that X_1 is distributed according to P. Now, given $X_1 = x_1$, S_0 is clearly uniformly distributed across the interval $[F(x_1), F(x_1 + 1)]$. Thus, the equation $S_1 = [S_0 - F(X_1)]/p(X_1)$ shifts and stretches this uniform distribution, such that S_1 is again uniform across $[0, 1]$, which means that the uniform distribution over $[0, 1]$ is invariant under the recursion. Since S_1 is uniform over $[0, 1]$ independently of X_1, then X_2 is distributed according to P for the same reason that X_1 was distributed according to P for $S_0 \sim \text{Unif}[0, 1]$. Moreover, since S_1 is independent of X_1, then $X_2 = \phi(S_1)$ is also independent of X_1. Likewise, S_2 is uniformly distributed over $[0, 1]$, independently of (X_1, X_2), and so, $X_3 = \phi(S_2)$ is distributed according to P, independently of (X_1, X_2), and so on. It follows then that $\boldsymbol{X}$ is a sequence of i.i.d. random variables, all with distribution P. In the sequel, it will be useful to use the equivalence between the uniform distribution over the initial state, S_0, and the fact that the induced process, $X_1, X_2, \ldots$, is i.i.d. with a marginal distribution P, via the one-to-one relationship between them, as was established in item no. 1 above.

3. *The Lyapunov exponent.* Since the class of dynamical systems considered here are chaotic in general, one of the important characteristics is the Lyapunov exponent, which measures the local exponential rate of divergence between two state sequences that begin from two very close initial states (see, e.g., [6]). Assuming that S_0 is uniformly distributed over $[0, 1]$, so that $X_1, X_2, \ldots$ are i.i.d. with distribution P, the Lyapunov exponent is given by

$$\begin{aligned} \lambda &= \lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\left\{\log\left|\frac{\partial s_n}{\partial s_0}\right|\right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\left\{\log\left|\prod_{i=1}^{n} \frac{\partial s_i}{\partial s_{i-1}}\right|\right\} \\ &= \lim_{n\to\infty} \frac{1}{n} \boldsymbol{E}\left\{\log\prod_{i=1}^{n} \frac{1}{p(X_i)}\right\} \\ &= H, \end{aligned} \tag{19}$$

where H is the entropy the source P, i.e.,

$$H = -\sum_{x=0}^{r-1} p(x) \log p(x). \tag{20}$$

In words, the Lyapunov exponent coincides with the entropy of the itinerary process, which is maximized when P is the uniform distribution.

4. *The length of the signal locus.* An important factor in the weak-noise performance of a twisted modulation system is the length of the signal locus [18, Chap. 8] (see also [12]): Consider the channel input vector (7), with the (temporary) notation $u^n(\theta) = (u_1(\theta), \ldots, u_n(\theta))$ that emphasizes the dependence on the parameter θ. While θ exhausts the interval

$[0,1]$, the vector $u^n(\theta)$ draws a curve in the vector space of dimension n. The length of this curve, in general, is given by

$$L_n = \int_0^1 \|\dot{u}^n(\theta)\|\mathrm{d}\theta, \tag{21}$$

where $\dot{u}^n(\theta) = (\dot{u}_1(\theta), \ldots, \dot{u}_n(\theta))$, $\dot{u}_i(\theta)$ being the derivative of $u_i(\theta)$ w.r.t. θ (provided that it exists), $i = 1, 2, \ldots, n$. In our case, each $u_i(\theta)$ is a piece-wise linear function of $\theta = s_0$ with discontinuities. Therefore, the vector $u^n(\theta)$, as a whole, is also piece-wise linear and piece-wise continuous in θ and there are exactly $r^n - 1$ points of discontinuity along the interval $0 \leq \theta \leq 1$, because as s_0 varies from 0 to 1, the itinerary vector x^n exhausts all r^n possible values, beginning from $(0, 0, \ldots, 0)$ and ending at $(r-1, r-1, \ldots, r-1)$. Each change in one of the components of x^n corresponds to one discontinuity point of s^n. The integral in (21) expresses the sum of lengths of the continuous pieces of the signal locus. To assess L_n in our case, consider the following. Along each sub-interval of continuity, we have

$$\begin{aligned}
\dot{u}_i(\theta) &= \frac{\mathrm{d}u_i(\theta)}{\mathrm{d}\theta} \\
&= \frac{\mathrm{d}u_i(s_0)}{\mathrm{d}s_0} \\
&= \frac{\mathrm{d}u_i}{\mathrm{d}s_i} \prod_{j=1}^{i} \frac{\mathrm{d}s_j}{\mathrm{d}s_{j-1}} && (22) \\
&= \sqrt{12Q} \cdot \prod_{j=1}^{i} \frac{1}{p(x_j)} && (23)
\end{aligned}$$

Thus, L_n, is upper bounded as follows:

$$\begin{aligned}
L_n &= \sqrt{12Q} \int_0^1 \mathrm{d}s_0 \cdot \sqrt{\sum_{i=1}^{n} \prod_{j=1}^{i} \frac{1}{p^2(x_j)}} \\
&= \sqrt{12Q} \boldsymbol{E} \left\{ \sqrt{\sum_{i=1}^{n} \prod_{j=1}^{i} \frac{1}{p^2(X_j)}} \right\} \\
&\leq \sqrt{12Q} \boldsymbol{E} \left\{ \sqrt{n \cdot \prod_{j=1}^{n} \frac{1}{p^2(X_j)}} \right\} \\
&= \sqrt{12Qn} \cdot \sum_{\boldsymbol{x}} \prod_{i=1}^{n} p(x_i) \sqrt{\prod_{j=1}^{n} \frac{1}{p^2(x_j)}} \\
&= \sqrt{12Qn} \cdot r^n, && (24)
\end{aligned}$$

where we have use the fact integration along the interval $s_0 \in [0,1]$ is equivalent to expectation w.r.t. the randomness of the random variable $S_0 \sim \text{unif}[0,1]$, which in turn is equivalent to expectation w.r.t. the randomness of $\{X_i\}$ being an i.i.d. process with probability distribution P. The inequality in the above chain amounts to the fact that the sum of n positive terms cannot exceed n times the largest term, which is the last one. Likewise,

$$\begin{aligned}
L_n &= \sqrt{12Q} \boldsymbol{E} \left\{ \sqrt{\sum_{i=1}^{n} \prod_{j=1}^{i} \frac{1}{p^2(X_j)}} \right\} \\
&\geq \sqrt{12Q} \boldsymbol{E} \left\{ \sqrt{\prod_{j=1}^{n} \frac{1}{p^2(X_j)}} \right\} \\
&= \sqrt{12Q} \sum_{\boldsymbol{x}} \prod_{i=1}^{n} p(x_i) \sqrt{\prod_{j=1}^{n} \frac{1}{p^2(x_j)}} \\
&= \sqrt{12Q} r^n, && (25)
\end{aligned}$$

which means that $L_n \doteq r^n$, independent of P. Here, the notation $a_n \doteq b_n$, for two positive sequences, $\{a_n\}$ and $\{b_n\}$, means that $\frac{1}{n} \log \frac{a_n}{b_n} \to 0$, as $n \to \infty$.

5. *The autocorrelation function and the spectrum of the state process.* We now calculate the autocorrelation individual function of the stationary state process $\{S_t\}$, induced by a uniformly distributed initial state, S_0, across the interval $[0,1]$. In Appendix A, we derive the following expression for $R_S(k) \stackrel{\triangle}{=} \boldsymbol{E}\{S_0 S_k\}$ (k – integer):

$$R_S(k) = \frac{1}{4} + \frac{1}{12} \cdot \left(\sum_{x=0}^{r-1} p^2(x) \right)^{|k|}, \tag{26}$$

where the first term comes from the DC component of $\{S_t\}$, which is $\frac{1}{2}$, as each S_t is distributed uniformly over $[0,1]$. The exponential term is the more interesting term. Using the relation (see (7),

$$U_t = \sqrt{12Q} \left(S_t - \frac{1}{2} \right), \tag{27}$$

it is clear that the autocorrelation function of $\{U_t\}$, is given by

$$R_U(k) = \boldsymbol{E}\{U_0 U_k\} = Q \cdot \left(\sum_{x=0}^{r-1} p^2(x) \right)^{|k|}. \tag{28}$$

For the sake of brevity, let us denote

$$\bar{p} = \boldsymbol{E}\{p(X)\} = \sum_{x=0}^{r-1} p^2(x), \tag{29}$$

and then the power spectrum is

$$S_U(e^{j\theta}) = \frac{Q(1 - \bar{p}^2)}{(1 - \bar{p}e^{-j\theta})(1 - \bar{p}e^{j\theta})}, \qquad j \stackrel{\triangle}{=} \sqrt{-1}. \tag{30}$$

The channel output spectrum is then

$$\begin{aligned}
S_Y(e^{j\theta}) &= \frac{Q(1 - \bar{p}^2)}{(1 - \bar{p}e^{-j\theta})(1 - \bar{p}e^{j\theta})} + \sigma^2 \\
&= \frac{A(1 - \tau e^{-j\theta})(1 - \tau e^{j\theta})}{(1 - \bar{p}e^{-j\theta})(1 - \bar{p}e^{j\theta})} && (31)
\end{aligned}$$

where A and τ are readily found to be given by:

$$\begin{aligned}\tau &= 2\sigma^2\bar{p}\cdot\Bigg[\sigma^2(1+\bar{p}^2)+Q(1-\bar{p}^2)+\\ &\quad\sqrt{[\sigma^2(1+\bar{p}^2)+Q(1-\bar{p}^2)]^2-4\sigma^2\bar{p}^2}\Bigg]^{-1}\\ &< 1 \end{aligned}\tag{32}$$

and

$$\begin{aligned}A &= \frac{1}{2}\Bigg[\sigma^2(1+\bar{p}^2)+Q(1-\bar{p}^2)+\\ &\quad\sqrt{[\sigma^2(1+\bar{p}^2)+Q(1-\bar{p}^2)]^2-4\sigma^2\bar{p}^2}\Bigg].\end{aligned}\tag{33}$$

6. Ergodic properties. Since $R_U(k)$ tends to zero as $k\to\infty$, then so does $\frac{1}{2k+1}\sum_{j=-k}^{k}R_U(j)$, and then it follows by Slutsky's theorem [15, p. 432] that the process $\{U_t\}$ is mean-ergodic, that is,

$$\begin{aligned}&\lim_{n\to\infty}\boldsymbol{E}\left\{\left(\frac{1}{n}\sum_{t=1}^{n}U_t-\boldsymbol{E}\{U_1\}\right)^2\right\}\\ &= \lim_{n\to\infty}\boldsymbol{E}\left\{\left(\frac{1}{n}\sum_{t=1}^{n}U_t\right)^2\right\}\\ &= 0.\end{aligned}\tag{34}$$

Using the inequalities (18), it is straightforward to check (see Appendix B for the details) that $\{U_t\}$ is also autocovariance-ergodic [15, p. 437], namely,

$$\lim_{n\to\infty}\boldsymbol{E}\left\{\left[\frac{1}{n}\sum_{t=1}^{n}U_tU_{t+k}-R_U(k)\right]^2\right\}=0\tag{35}$$

for every fixed non-negative integer k. This is done by applying Slutsky's theorem to establish mean-ergodicity of the process $V_t=U_tU_{t+k}$. It then follows by Chebychev's inequality that for every $\epsilon>0$,

$$\lim_{n\to\infty}\Pr\left\{\left|\frac{1}{n}\sum_{t=1}^{n}U_tU_{t+k}-R_U(k)\right|>\epsilon\right\}=0.\tag{36}$$

Since $\{u_t\}$ are completely determined by s_0, then the randomness of $\{U_t\}$ fully stems from the randomness of the uniformly distributed initial state, S_0, and then it follows that the Lebesgue measure of the set

$$\left\{s_0:\left|\frac{1}{n}\sum_{t=1}^{n}u_tu_{t+k}-R_U(k)\right|\le\epsilon,\ k=0,1\right\},\tag{37}$$

tends to unity as $n\to\infty$. Since $\{u_t\}$ are piece-wise continuous functions of s_0, then so is $\frac{1}{n}\sum_{t=1}^{n}u_tu_{t+k}$, which implies that the above set is the union of disjoint sub-intervals of $[0,1]$ whose total Lebesgue measure (length) tends to unity.

7. An upper bound on the mutual information between the channel input and output. The induced normalized mutual information associated with the channel is then:

$$\begin{aligned}C_0 &= \lim_{n\to\infty}\frac{h(Y^n)-h(Y^n|U^n)}{n}\\ &\le \frac{1}{4\pi}\int_{-\pi}^{\pi}\ln[2\pi eS_Y(e^{j\theta})]\mathrm{d}\theta-\frac{1}{2}\ln(2\pi e\sigma^2)\\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\ln[S_Y(e^{j\theta})]\mathrm{d}\theta-\frac{1}{2}\ln\sigma^2\\ &= \frac{1}{2}\ln\left(\frac{A}{\sigma^2}\right)\\ &\stackrel{\triangle}{=} C_1.\end{aligned}\tag{38}$$

Note that A can also be written as

$$\begin{aligned}A &= \frac{1}{2}\Bigg[\sigma^2+Q+\bar{p}^2(\sigma^2-Q)+\\ &\quad\sqrt{(\sigma^4+Q^2)(1-\bar{p}^2)^2+2\sigma^2Q(1-\bar{p}^4)}\Bigg],\end{aligned}\tag{39}$$

and so,

$$\begin{aligned}\frac{A}{\sigma^2} &= \frac{1}{2}\Bigg[1+\gamma+\bar{p}^2(1-\gamma)+\\ &\quad\sqrt{(1+\gamma^2)(1-\bar{p}^2)^2+2\gamma(1-\bar{p}^4)}\Bigg],\end{aligned}\tag{40}$$

which makes it clear that, at least when $\gamma>1$, A/σ^2 and hence also C_1, is a decreasing function of $\bar{p}$.

IV. A Lower Bound on the Weak-Noise Estimation Performance

In this section, we derive a lower bound on the weak-noise estimation performance of twisted modulators that are based on the class of chaotic dynamical systems that were defined in Section II. We do this on the basis of the generic lower bound of [12], but with the appropriate adjustments that are particular to this class of modulators, using the results of Section III. The resulting bound is then tighter than the generic bound of [12]. For the sake of completeness, we begin by presenting the generic lower bound of [12] as background, and the first step to this end, is to provide a few definitions and to list the assumptions therein.

Given a positive real α, which designates the outage exponent (11), let $w(\alpha)$ be the unique solution w to the equation[2]

$$w-\ln(1+w)=2\alpha.\tag{41}$$

Let us also define

$$R(\alpha,\gamma)=\frac{1}{2}\ln\frac{\gamma}{1+w(\alpha)}.\tag{42}$$

The following assumptions are imposed in [12].

[2] As explained in [12], $w(\alpha)$ as defined in eq. (41), is the value of $w>0$ for which $\Pr\{\sum_{i=1}^{n}Z_i^2\ge n\sigma^2(1+w)\}$ decays exponentially with the exponential order of $e^{-\alpha n}$.

A.1 For any given constant $c > 0$, the error cost function ρ obeys $\rho(e^{-nc}) \doteq e^{-n\zeta(c)}$, where $\zeta(\cdot)$ is some continuous function with the property that $c > 0$ implies $\zeta(c) > 0$.

A.2 Denoting $M_n = e^{nR(\alpha,\gamma)}$, consider the partition of the unit interval into M_n non–overlapping sub–intervals, each of size $1/M_n$. Then, for all sufficiently large n, the number, M_n^{c}, of sub–intervals in which $s^n(\theta)$ is continuous, obeys $M_n^{\mathrm{c}} \doteq e^{nR(\alpha,\gamma)}$.

A.3 If, for a certain vector $\boldsymbol{z} \in \mathcal{O}_n(\theta)$, one of the components vanishes, then upon replacing this component by any non–zero number, the resulting vector remains in $\mathcal{O}_n(\theta)$. In addition, it is assumed that as noise variance tends to zero, the covering radius of $\mathcal{O}_n^{\mathrm{c}}(\theta)$ tends to zero as well.

Denoting

$$E(\alpha,\gamma) = \zeta[R(\alpha,\gamma)], \tag{43}$$

the following lower bound is derived in [12, eq. (21)] under assumptions A1-A3:

$$\begin{aligned} &\sup_{0\leq\theta\leq 1} \boldsymbol{E}\left\{\rho(\varphi(Y^n)-\theta)\middle|\mathcal{O}_n^{\mathrm{c}}(\theta)\right\} \\ \geq\;& \exp\{-n[E(\alpha,\gamma)+o(\gamma)]\}, \end{aligned} \tag{44}$$

where $o(\gamma)$ designates a function that tends to zero as $\gamma \to \infty$. In [12], this lower bound is derived in two steps. In the first step, it is shown that for every positive integer M,

$$\begin{aligned} &\sup_{0\leq\theta\leq 1} \boldsymbol{E}\left\{\rho(\varphi(Y^n)-\theta)\middle|\mathcal{O}_n^{\mathrm{c}}(\theta)\right\} \\ \geq\;& 2\rho\left(\frac{1}{2M}\right)\cdot\left[Q\left(\frac{L_n}{2\sigma M}\right)-e^{-\alpha n}\right], \end{aligned} \tag{45}$$

where L_n is the length of the locus of the signal fed into the channel (for our modulators, $\{u^n(\theta),\ 0 \leq \theta \leq 1\}$), as described in Section III. In the second step, it is proved that to meet the outage constraint, L_n must be upper bounded according to

$$L_n \leq \sigma \cdot \exp\{n[R(\alpha,\gamma)+o(\gamma)]\}. \tag{46}$$

By substituting this upper bound instead of L_n in the right-hand side of (45), we further lower bound $\sup_{0\leq\theta\leq 1} \boldsymbol{E}\left\{\rho(\varphi(Y^n)-\theta)\middle|\mathcal{O}_n^{\mathrm{c}}(\theta)\right\}$ (as $Q(\cdot)$ is a decreasing function), and then by selecting M to be of the exponential order of $\exp\{n[R(\alpha,\gamma)\}$, the lower bound (44) is obtained. The upper bound on L_n is obtained by a 'tube-packing' argument, in the spirit of the one in [18, pp. 669–674], which is a version of the sphere-packing argument of coding for the Gaussian channel (see, e.g., [4, p. 265]), but adapted to parameter modulation and estimation. According to this argument, the tube generated by the union of spheres of radius $\sqrt{n\sigma^2}$ around all vectors along the curve $\{u^n(\theta),\ 0 \leq \theta \leq 1\}$, must be packed within the sphere of vectors $\{y^n\}$ that are obtained as $y^n = u^n + z^n$, where u^n is within a sphere of radius $\sqrt{nQ}$ and z^n is within a sphere of radius $\sqrt{n\sigma^2}$ and is orthogonal to u^n, namely, a sphere of radius $\sqrt{n(Q+\sigma^2)}$. Another way to look at this sphere is as the set of y^n-vectors defined as follows: Let $Y = U + Z$, where $U \sim \mathcal{N}(0,Q)$ and $Z \sim \mathcal{N}(0,\sigma^2)$ be independent random variables, and let $f_Y(y)$ be the (Gaussian) probability density function (pdf) obtained by the convolution between $f_U(u)$ - the pdf of U, and $f_Z(z)$ - the pdf of Z. Now, let $h(Y)$ be the differential entropy of Y. Then, the sphere of radius $\sqrt{n(Q+\sigma^2)}$ can be thought of as the typical set

$$\mathcal{T}_n(Y) = \left\{y^n : \frac{1}{n}\sum_{i=1}^{n}\ln f_Y(y_i) \geq -h(Y)-\epsilon\right\}, \tag{47}$$

and then let ϵ tends to zero (after letting n grow without bound). Indeed,

$$\begin{aligned} 1 &\geq \int_{\mathcal{T}_n(Y)} \mathrm{d}y^n \cdot \prod_{i=1}^{n} f_Y(y_i) \\ &\geq \int_{\mathcal{T}_n(Y)} \mathrm{d}y^n \cdot e^{-n[h(Y)+\epsilon]} \\ &= \mathrm{Vol}\{\mathcal{T}_n(Y)\}\cdot e^{-n[h(Y)+\epsilon]}, \end{aligned} \tag{48}$$

and so,

$$\begin{aligned} \mathrm{Vol}\{\mathcal{T}_n(Y)\} &\leq e^{n[h(Y)+\epsilon]} \\ &= \exp\left\{n\left[\frac{1}{2}\ln(2\pi e\sigma^2)+\epsilon\right]\right\}. \end{aligned} \tag{49}$$

We now modify this lower bound, along with its derivation in [12], by particularizing it to the class of chaotic modulators considered here. The difference is in the upper bound to L_n. Note that the vector $u^n(\theta)$ must lie within the n-dimensional hyper-cube $[-\sqrt{12Q}, +\sqrt{12Q}]^n$. A typical noise vector, z^n, is essentially orthogonal (uncorrelated) to $u^n(\theta)$, and so, a typical channel output vector, $y^n = u^n(\theta) + z^n$, is the sum of a vector in $[-\sqrt{12Q}, +\sqrt{12Q}]^n$ and an orthogonal vector in a sphere of radius $\sqrt{n\sigma^2}$. To assess the volume of typical channel output vectors, we proceed along the same line as before: Consider a random variable $Y = U + Z$, where U is uniformly distributed across the interval $[-\sqrt{12Q}, +\sqrt{12Q}]$ and $Z \sim \mathcal{N}(0,\sigma^2)$ independent of U. Accordingly, let $f_Y(y)$ be the pdf of Y, which is now the convolution between the uniform pdf of U over $[-\sqrt{12Q}, +\sqrt{12Q}]$ and the Gaussian pdf of Z. Let $h(Y)$ denote the differential entropy associated with f_Y, and consider again the set $\mathcal{T}_n(Y)$ but with the new definitions of f_Y and $h(Y)$. Then, here too, $\mathrm{Vol}\{\mathcal{T}_n(Y)\}$ is essentially upper bounded by $e^{nh(Y)} = e^{nh(U+Z)}$. According to a tube-packing bound similar to that of [18] and [12], we now have (ignoring the ϵ-term):

$$\begin{aligned} L_n &\overset{\cdot}{\leq} \frac{e^{nh(Y)}}{(2\pi e\sigma^2[1+w(\alpha)])^{(n-1)/2}} \\ &\doteq \sqrt{2\pi e}\cdot\sigma\cdot\exp\left\{n\left[h(Y)-\right.\right. \\ &\qquad \left.\left.\frac{1}{2}\ln(2\pi e\sigma^2[1+w(\alpha)])\right]\right\}. \end{aligned} \tag{50}$$

For large γ, the convolution of the rectangular f_U with the Gaussian f_Z is very close to the original f_U, and so, $h(Y) = h(U) + o(\sigma^2) = \frac{1}{2}\ln[12Q(1+\Delta(\gamma))]$, where $\Delta(\gamma) \to 0$ as $\gamma \to \infty$. Thus,

$$\begin{aligned} L_n &\leq \sqrt{2\pi e}\cdot\sigma\cdot\exp\left\{\frac{n}{2}\ln\left(\frac{12Q(1+\Delta(\gamma))}{2\pi e\sigma^2[1+w(\alpha)]}\right)\right\} \\ &= \sqrt{2\pi e}\cdot\sigma\cdot\exp\left\{n\left(\frac{1}{2}\ln\gamma - \frac{1}{2}\ln\left(\frac{2\pi e}{12}\right) + \right.\right. \\ &\qquad \left.\left.\frac{1}{2}\ln[1+\Delta(\gamma)] - \frac{1}{2}\ln[1+w(\alpha)]\right)\right\} \\ &\triangleq L_n^*. \end{aligned} \tag{51}$$

Consider now the case of very high SNR ($\gamma \gg 1$) and auppose that $\alpha = \alpha_n$ tend to zero, yet $n\alpha_n \to \infty$, which means that although we wish to keep the probability of anomaly arbitrarily small for large n, we do not insist on an exponential decay of this probability. In this case, we can neglect the last two terms at the exponent and only the first two terms remain. The term $\frac{1}{2}\ln\gamma$ approximates the channel capacity, $C = \frac{1}{2}\ln(1+\gamma)$ for large γ. The constant term $\frac{1}{2}\ln(2\pi e/12) \approx 0.1765$, which we henceforth denote by μ, is the loss due to channel input shaping mismatch.

Next, recall that in Section III, we have seen that for the chaotic modulators considered here, $L_n \geq \sqrt{12Q}r^n$, independent of P. Together with the upper bound $L_n \leq L_n^* \doteq \exp\{n[C-\mu+o(\gamma)]\}$, it follows that the anomaly probability constraint imposes a limitation on the alphabet size, r. Specifically, neglecting the $o(\gamma)$ term at the exponent of L_n^*, we see that r must not exceed $e^{C-\mu}$. Upon selecting $r < e^{C-\mu}$, we have $r^n \ll e^{n(C-\mu)}$ for large n, which means, among other things, that the number of discontinuity points of $u^n(\theta)$, as function of θ, which is $r^n - 1$, is exponentially smaller than $e^{n(C-\mu)}$, a fact that in turn allows us to select, similarly as in [12], $M \doteq L_n^*/(2\pi s)$ (with s being an arbitrary positive constant) in (45) and be assured that Assumption A.2 is satisfied. The resulting lower bound would then be

$$\begin{aligned} &\sup_{0\leq\theta\leq 1}\boldsymbol{E}\left\{\rho(\phi(Y^n)-\theta)\middle|\mathcal{O}_n^c(\theta)\right\} \\ \geq\ & \rho\left(\frac{\sigma s}{L_n^*}\right)\left[Q(s) - e^{-n\alpha_n}\right] \doteq e^{-n\zeta(C-\mu)}, \end{aligned} \tag{52}$$

and so, the loss compared to the generic lower bound of [12] is at least in the shaping loss reduction by μ.

Discussion. It should be pointed out the shaping mismatch is not the only factor that causes loss in performance to our class of chaotic modulators considered here. The reason is that there is an additional inherent limitation on the channel input signals generated by these chaotic modulators, and this is the fact that they possess memory. More precisely, owing to the ergodic properties of the dynamical system (as discussed in Section III), for most values of $\theta \in [0,1]$ (in the Lebesgue measure sense), $u^n(\theta)$ possesses an empirical autocorrelation function that is close to $R_U(k) = Q\cdot\bar{p}^{|k|}$. This further reduces the volume of the object in which the tube of noisy signals must be packed. A detailed analysis of this volume reveals that it is upper bounded by the exponential order of $A^{n/2}$, where A is defined as in eq. (33), and so, another upper bound to L_n^* is of the exponential order of e^{nC_1}, where C_1 is defined as in (38). This alternative upper bound, however, ignores the fact that $u^n(\theta)$ is limited to lie in the hyper-cube $[-\sqrt{12Q}, +\sqrt{12Q}]^n$. Combining the two limitations on $u^n(\theta)$ in a joint manner is not a trivial task, but of course, one can always take the better between the two individual bounds. Anyway, we will not delve here into this analysis because the first lower bound above suffices on its own right to make the point that the performance of the chaotic modulators cannot approach the generic lower bound, not even in the exponential order. Nonetheless, there is an alternative way to use these modulators so that the lower bound of [12] would be approached arbitrarily closely in the exponential order. This is the subject of the next section.

V. Feeding the Channel by the Itinerary Signal

So far, we have considered our chaotic modulators as devices that generate and feed the channel with u^n, which is a scaled and shifted version of the state sequence, s^n, as was also done in most of the earlier works on twisted modulation using chaotic dynamical systems. However, our dynamical system generates also an additional output, that is, the itinerary sequence $x^n = (x_1,\ldots,x_n)$. In this section, we consider the option of feeding the channel with x^n, for the purpose estimating θ. More precisely, we allow some mapping $c:\{0,1,\ldots,r-1\}\to\mathcal{V}$, where $\mathcal{V}$ is a set of r real numbers $\{c(x),\ x\in\{0,1,\ldots,r-1\}\}$ that designate legitimate channel inputs that meet the power constraint, $\sum_{x=0}^{r-1}p(x)[c(x)]^2 \leq Q$, and we transmit $v^n = c(x^n) = (c(x_1),\ldots,c(x_n))$.

At first glance, it may not seem like a good idea to use x^n or any function of it, such as $c(x^n)$, as an alternative to s^n, because each x_i is a quantized version of s_{i-1}, and so, it seems to convey less information about θ. Somewhat surprisingly, it turns out that with the correct choices, it works better than with s^n, or u^n. Note that while it is true that the itinerary vector is a quantized version of the corresponding state vector, it turns out that in the limit of large n, the difference between them vanish in terms of the information they convey on θ. Indeed, recall that the infinite sequence $\boldsymbol{x} = (x_1, x_2, \ldots)$ fully determines θ, which in turn fully determines $\boldsymbol{s} = (s_1, s_2, \ldots)$. Also, in contrast to the transmission of u^n, if one uses $v^n = c(x^n)$ as the channel input vector, there is no longer a limitation on r as we had before, and there are therefore many degrees of freedom. In fact, r can even be countably infinite. As it turns out, by using x^n with the correct choices of r, P, $c(\cdot)$, and a certain map $s_0 = \psi(\theta)$, where $\psi:[0,1]\to[0,1]$ (see footnote no. 1), we can approach the generic lower bound of [12] arbitrarily closely at least for the case $\alpha\to 0$.

Consider the following construction: Let $\epsilon > 0$ be arbitrarily small and select $R = C - \epsilon$, where $C = \frac{1}{2}\ln(1+\gamma)$ is the capacity of the Gaussian channel. Let $M = e^{nR}$ and consider the grid of M points, $\mathcal{G} = \{1/(2M), 3/(2M), 5/(2M), \ldots, (2M-1)/(2M)\}$. For each grid point $\hat{\theta}_i = (2i-1)/2M$, $i = 1,2,\ldots,M$, select independently at random a number, denoted $\psi(\hat{\theta}_i)$, under the uniform distribution over the interval $[0,1]$. Reveal these independent random selections to both transmitter and receiver. With probability one, these M random numbers are all distinct, and then the inverse map ψ^{-1} is well defined. Let r be an arbitrarily large odd number and let

$$c(x) = \delta \cdot \left(x - \frac{r-1}{2}\right), \qquad x = 0, 1, \ldots, r-1, \tag{53}$$

where $\delta > 0$ is an arbitrarily small number such that $r\delta \gg \sqrt{Q}$. Finally, let P be defined as follows:

$$p(x) = \begin{cases} \int_{-\infty}^{-(r/2-1)\delta} g(a)\mathrm{d}a & x = 0 \\ \int_{\delta(x-r/2)}^{\delta(x-r/2+1)} g(a)\mathrm{d}a & x \in \{1, 2, \ldots, r-2\} \\ \int_{(r/2-1)\delta}^{\infty} g(a)\mathrm{d}a & x = r-1 \end{cases} \tag{54}$$

where $g(a) = (2\pi Q)^{-1/2} e^{-a^2/(2Q)}$.

Our modulation scheme works as follows. Given a parameter value $\theta \in [0,1]$, consider its quantization to the nearest grid point $\hat{\theta}_i \in \mathcal{G}$ and then define $s[i] = \psi(\hat{\theta}_i)$ as the initial state. Let $x^n[i]$ be the corresponding itinerary sequence generated by the system, starting at $s[i]$. Transmit $c(x^n[i])$ over the channel, and at the receiver side apply a ML channel decoder for the code $\mathcal{C} = \{c(x^n[i]),\ i = 1, 2, \ldots, M\}$. Let $\hat{i}$ be the index of the decoded message. Then, the estimated parameter is $\hat{\theta} = \psi^{-1}[s(\hat{i})]$.

To see why this scheme nearly achieves the lower bound, is conceptually simple. The quantization error in θ cannot exceed $\frac{1}{2M} = \frac{1}{2}e^{-n(C-\epsilon)}$ whose cost is $\rho\left(\frac{1}{2}e^{-n(C-\epsilon)}\right) \doteq e^{-n\zeta(C-\epsilon)}$. Since $\{\psi(\hat{\theta}_i),\ i = 1, \ldots, M\}$ are independent uniformly distributed random variables, their corresponding itinerary vectors $\{c(x^n[i]),\ i = 1, \ldots, M\}$ are independent i.i.d. random vectors that form a codebook for the Gaussian channel, where the random coding distribution is a finely quantized version of the capacity-achieving pdf, $g(a)$, and hence it is nearly capacity-achieving for small δ and large r. Since $R < C$, the error probability of the decoding, which is viewed as the probability of anomaly, is arbitrarily small. In the event of correct decoding, $\hat{i} = i$, the correct $\hat{\theta}_i$ is reconstructed and error cost remains $e^{-n\zeta(C-\epsilon)}$, namely, the quantization error cost only.

VI. Extension to Dynamical Systems with Long-Range Memory

The class of chaotic dynamical systems that we considered so far can be naturally extended to possess longer range memory of the past and the lower bound derived in Section IV will continue to apply. Specifically, given an integer $r \geq 2$ and a sequence of conditional probability distributions, $\{p_t(x_t|x_1, \ldots, x_{t-1}),\ x_1, x_2, \ldots, x_t \in \{0, 1, \ldots, r-1\}\}$, $t = 1, 2, \ldots$, defining a probability law of a process P, let us define

$$F_t(x|x_1, \ldots, x_{t-1}) = \sum_{x'=0}^{x-1} p_t(x'|x_1, \ldots, x_{t-1}), \tag{55}$$

and let $\phi_t(s|x_1, \ldots, x_{t-1})$ be defined as the unique value of $x \in \{0, 1, \ldots, r-1\}$ such that

$$F_t(x|x_1, \ldots, x_{t-1}) \leq s < F_t(x+1|x_1, \ldots, x_{t-1}), \tag{56}$$

where for $x = r-1$, the strong inequality is allowed to be weak, namely, $\phi(1|x_1, \ldots, x_{t-1}) = r-1$. The dynamical system is now defined by the following recursion:

$$\begin{aligned} x_t &= \phi_t(s_{t-1}|x_1, \ldots, x_{t-1}), \\ s_t &= \frac{s_{t-1} - F_t(x_t|x_1, \ldots, x_{t-1})}{p_t(x_t|x_1, \ldots, x_{t-1})}; \quad t = 1, 2, \ldots \end{aligned} \tag{57}$$

Eq. (12) extends to this case reads as follows:

$$\begin{aligned} s_0 &= \sum_{t=1}^{\infty} F_t(x_t|x_1, \ldots, x_{t-1}) \prod_{j=1}^{t-1} p_j(x_j|x_1, \ldots, x_{j-1}) \\ &= \sum_{t=1}^{\infty} F_t(x_t|x_1, \ldots, x_{t-1}) p_{t-1}(x_1, \ldots, x_{t-1}). \end{aligned} \tag{58}$$

If the initial state is a random variable, S_0, uniformly distributed over the interval $[0,1]$, then the induced itinerary sequence, $X_1, X_2, \ldots$, is a random process whose probability law is governed by the given sequence of conditional probability distributions, $\{p_t(x_t|x_1, \ldots, x_{t-1}),\ x_1, x_2, \ldots, x_t \in \{0, 1, \ldots, r-1\}\}$, $t = 1, 2, \ldots$. Here, the Lyapunov exponent is given by the entropy rate,

$$\lambda = \bar{H} = \lim_{n\to\infty} \frac{H(X_1, \ldots, X_n)}{n}, \tag{59}$$

provided that the limit exists. The length of the signal locus continues to be upper bounded by $\sqrt{12Qn} \cdot r^n$ and lower bounded by $\sqrt{12Q}r^n$. If the itinerary signal is used for transmission over a general channel (that may not be necessarily memoryless), it naturally suggests to select P to be capacity-achieving channel input process.

Finally, as a side remark, we point out that the mapping between S_0 and $\boldsymbol{X} = (X_1, X_2, \ldots)$ can serve as a simulator for synthesizing a prescribed random process P from a sequence of purely random bits [7]. Consider the binary symmetric source of random bits, $\boldsymbol{B} = (B_1, B_2, \ldots)$ that define the uniformly distributed random variable,

$$S_0 = \sum_{i=1}^{\infty} B_i 2^{-i}, \tag{60}$$

which in turn is mapped to the sequence of random variables, $\boldsymbol{X} = (X_1, X_2, \ldots)$ governed by the process law P. Viewing this as a mapping from $\boldsymbol{B}$ to $\boldsymbol{X}$, it can be considered a simulator of a random process with an optimal average conversion rate [7] of $\bar{H}$ bits per symbol, where $\bar{H}$ is the entropy rate, assuming it exists. To see why this is true, assume first that P is a dyadic source in the sense that all conditional distributions, $p_t(x_t|x_1, \ldots, x_{t-1})$, are integer powers of $\frac{1}{2}$, and consider the recursion:

$$\begin{aligned} S_t &= \frac{S_{t-1} - F_t(X_t|X_1, \ldots, X_{t-1})}{p_t(X_t|X_1, \ldots, X_{t-1})} \\ &= \exp_2\{\log[1/p_t(X_t|X_1, \ldots, X_{t-1})]\} \times \\ &\quad [S_{t-1} - F_t(X_t|X_1, \ldots, X_{t-1})]. \end{aligned} \tag{61}$$

In terms of the binary representation of S_{t-1}, the subtraction of $F_t(X_t|X_1,\ldots,X_{t-1})$ amounts to a certain manipulation of the bits of S_{t-1} whereas the multiplication by $\exp_2\{\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]\}$ means a shift of $\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]$ bits to the left. After the shift, the left-most $\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]$ bits are discarded and S_t then depends on the tail, starting from $(\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]+1)$-th bit and onward. Thus, to generate X_t, $\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]$ bits were utilized, and so, on the average, we have used $\boldsymbol{E}\{\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]\} = H(X_t|X_1,\ldots,X_{t-1})$ bits per symbol. Averaging over $t=1,2,\ldots,n$ yields $H(X_1,\ldots,X_n)/n$, which tends to $\bar{H}$ is the limit exists. For a non-dyadic source, $\log[1/p_t(X_t|X_1,\ldots,X_{t-1})]$ are not all integers, but then to reduce the effect of rounding errors, one considers the effect of $n \gg 1$ successive iterations, with a total shift of $\sum_{t=1}^n \log[1/p_t(X_t|X_1,\ldots,X_{t-1})] = \log[1/p_t(X_1,\ldots,X_t)]$, and then applies the usual considerations well known from the theory of variable-length lossless source coding. The advantage of this simulator is that it is relatively easy to implement and that it generates a process with the exact distribution, as opposed to fixed-rate schemes with the approximate distribution.

Appendix A

Derivation of $R_S(k)$

To derive $R_S(k)$, we use the relationship between the state at a given time and the itineraries at all later times. For a given integer $k \geq 0$, we define

$$\begin{aligned} R_S(k) &\stackrel{\Delta}{=} \boldsymbol{E}\{S_0 S_k\} \\ &= \boldsymbol{E}\left\{\sum_{t=1}^{\infty} F(X_t)\prod_{i=1}^{t-1} p(X_t) \times \right. \\ &\qquad \left. \sum_{\ell=1}^{\infty} F(X_{k+\ell})\prod_{j=1}^{\ell-1} p(X_{k+j})\right\} \\ &= \sum_{t=1}^{\infty}\sum_{\ell=1}^{\infty} A(t,\ell), \end{aligned} \tag{A.1}$$

where

$$A(t,\ell) = \boldsymbol{E}\left\{F(X_t)F(X_{k+\ell})\prod_{i=1}^{t-1} p(X_t)\cdot\prod_{j=1}^{\ell-1} p(X_{k+j})\right\}.$$

We denote

$$\bar{p} = \boldsymbol{E}\{p(X)\} = \sum_{x=0}^{r-1} p^2(x) \tag{A.2}$$

$$\bar{F} = \boldsymbol{E}\{F(X)\} = \sum_{x=0}^{r-1} p(x)F(x) \tag{A.3}$$

$$\overline{p^2} = \boldsymbol{E}\{p^2(X)\} = \sum_{x=0}^{r-1} p^3(x) \tag{A.4}$$

$$\overline{F^2} = \boldsymbol{E}\{F^2(X)\} = \sum_{x=0}^{r-1} p(x)F^2(x) \tag{A.5}$$

$$\overline{pF} = \boldsymbol{E}\{p(X)F(X)\} = \sum_{x=0}^{r-1} p^2(x)F(x). \tag{A.6}$$

For calculating $A(t,\ell)$, we distinguish between several cases.

1. *The case $k+1>t$.* In this case, there is no overlap between the segments X_1^t and $X_{k+1}^{k+\ell}$, and so, they are independent. Thus,

$$A(t,\ell) = (\bar{F})^2(\bar{p})^{t+\ell-2}. \tag{A.7}$$

2. *The case $k+1=t$.* In this case, there is an overlap of one sample, $X_t = X_{k+1}$, and so,

$$A(t,\ell) = (\bar{p})^{t-1}\cdot\overline{pF}\cdot(\bar{p})^{\ell-2}\cdot\bar{F} = (\bar{p})^{t+\ell-3}\cdot\bar{F}\cdot\overline{pF}. \tag{A.8}$$

3. *The case $k+1<t$.* This case should be subdivided into three secondary sub-cases:

3a. *The sub-case $t>k+\ell$.* In this case, the segment X_1^t fully covers the segment $X_{k+1}^{k+\ell}$, and then

$$\begin{aligned} A(t,\ell) &= (\bar{p})^k(\overline{p^2})^{\ell-1}\cdot\overline{pF}\cdot(\bar{p})^{t-k-\ell-1}\cdot\bar{F} \\ &\quad (\bar{p})^{t-\ell-1}\cdot\overline{pF}\cdot\bar{F}\cdot(\overline{p^2})^{\ell-1}. \end{aligned} \tag{A.9}$$

3b. *The sub-case $t=k+\ell$.* Similar to 3a, except that $X_t \equiv X_{k+\ell}$.

$$A(t,\ell) = (\bar{p})^k\cdot(\overline{p^2})^{\ell-1}\cdot\overline{F^2}. \tag{A.10}$$

3c. *The sub-case $t<k+\ell$.* Here the segments X_1^t and $X_{k+1}^{k+\ell}$ have only a partial overlap and none of them is a sub-string of the other.

$$\begin{aligned} A(t,\ell) &= (\bar{p})^k\cdot(\overline{p^2})^{t-k-1}\cdot\overline{pF}\cdot(\bar{p})^{k+\ell-t-1}\cdot\bar{F} \\ &= (\bar{p})^{2k+\ell-t-1}\cdot(\overline{p^2})^{t-k-1}\cdot\bar{F}\cdot\overline{pF}. \end{aligned} \tag{A.11}$$

Putting everything together, we get:

$$\begin{aligned} R_S(k) &= \sum_{t=1}^{\infty}\sum_{\ell=1}^{\infty} A(t,\ell) \\ &= \sum_{t=1}^{k}\sum_{\ell=1}^{\infty} A(t,\ell) + \sum_{\ell=1}^{\infty} A(k+1,\ell) + \\ &\quad \sum_{t=k+2}^{\infty}\left[\sum_{\ell=1}^{t-k-1} A(t,\ell) + A(t,t-k) + \right. \\ &\quad \left. \sum_{\ell=t-k+1}^{\infty} A(t,\ell)\right] \\ &= \sum_{t=1}^{k}\sum_{\ell=1}^{\infty}(\bar{F})^2(\bar{p})^{t+\ell-2} + \sum_{\ell=1}^{\infty}(\bar{p})^{t+\ell-3}\cdot\bar{F}\cdot\overline{pF} + \\ &\quad \sum_{t=k+2}^{\infty}\left[\sum_{\ell=1}^{t-k-1}(\bar{p})^{t-\ell-1}\cdot\overline{pF}\cdot\bar{F}\cdot(\overline{p^2})^{\ell-1} + \right. \\ &\quad (\bar{p})^k\cdot(\overline{p^2})^{t-k-1}\cdot\overline{F^2} + \\ &\quad \sum_{\ell=t-k+1}^{\infty}(\bar{p})^{2k+\ell-t-1}\cdot(\overline{p^2})^{t-k-1}\cdot\bar{F}\cdot\overline{pF}. \end{aligned}$$

It is easy to verify that all five terms are proportional to the geometric series $\{\bar{p}^k\}$, except for the first term, which also has a constant component (independent of k), given by

$$\begin{aligned}
\left[\frac{\bar{F}}{1-\bar{p}}\right]^2 &= \left[\frac{\sum_{x=0}^{r-1} p(x)F(x)}{1-\sum_{x=0}^{r-1} p^2(x)}\right]^2 \\
&= \left[\frac{\sum_{x=0}^{r-1} p(x)\sum_{x'=0}^{x-1} p(x')}{\sum_{x=0}^{r-1}\sum_{x'=0}^{r-1} p(x)p(x') - \sum_{x=0}^{r-1} p^2(x)}\right]^2 \\
&= \left[\frac{\sum_{(x,x'):\ x'<x} p(x)p(x')}{\sum_{(x,x'):\ x'\neq x} p(x)p(x')}\right]^2 \\
&= \left(\frac{1}{2}\right)^2 = \frac{1}{4}, \qquad \text{(A.12)}
\end{aligned}$$

accounting for the DC component of $\{S_t\}$, which is $\frac{1}{2}$. In other words, $R_S(k)$ is of the form

$$R_S(k) = \frac{1}{4} + q\cdot(\bar{p})^k, \qquad \text{(A.13)}$$

where q is a constant. The constant q is easily found to be $q = \frac{1}{12}$ by using the simple fact that $R_S(0) = \boldsymbol{E}\{S_0^2\} = \frac{1}{4}+q$ must be equal to $\frac{1}{3}$ since S_0 is uniformly distributed over $[0,1]$. It follows then that

$$R_S(k) = \frac{1}{4} + \frac{(\bar{p})^k}{12}. \qquad \text{(A.14)}$$

Appendix B

Autocovariance-Ergodicity of $\{U_t\}$

Since $\{U_t\}$ is just a scaled and shifted version of $\{S_t\}$ (see (7), it is enough to show that the latter has a similar property. We recall the inequalities (18) and denote

$$\hat{S}_0 \stackrel{\Delta}{=} \sum_{t=1}^{\tau} F(X_t)\prod_{i=1}^{\tau-1} p(X_i) \le S_0 \le \hat{S}_0 + \pi^\tau, \qquad \text{(B.1)}$$

where $\pi = \max_x p(x)$. Likewise,

$$\hat{S}_k \stackrel{\Delta}{=} \sum_{t=k+1}^{\tau+k} F(X_t)\prod_{i=k+1}^{\tau+k-1} p(X_i) \le S_k \le \hat{S}_k + \pi^\tau. \qquad \text{(B.2)}$$

Now, let $V_t = U_t U_{t+k}$. If we show that the auto-covariance function of $\{V_t\}$ tends to zero, then by Slutsky's theorem, it would be mean-ergodic, which would then imply that $\{U_t\}$ is autocovariance-ergodic. Now, let $m \ge k+\tau+1$.

$$\begin{aligned}
\boldsymbol{E}\{V_tV_{t+m}\} &= \boldsymbol{E}\{S_0S_kS_mS_{m+k}\} \\
&\le \boldsymbol{E}\{[\hat{S}_0+\pi^\tau][\hat{S}_k+\pi^\tau]S_mS_{m+k}\} \\
&\le \boldsymbol{E}\{\hat{S}_0\hat{S}_kS_mS_{m_k}\} + 2\pi^\tau + \pi^{2\tau} \\
&= \boldsymbol{E}\{\hat{S}_0\hat{S}_k\}\boldsymbol{E}\{S_mS_{m+k}\} + 2\pi^\tau + \pi^{2\tau} \\
&\le \boldsymbol{E}\{S_0S_k\}\boldsymbol{E}\{S_mS_{m+k}\} + 2\pi^\tau + \pi^{2\tau} \\
&= [R_S(k)]^2 + 2\pi^\tau + \pi^{2\tau} \\
&= \boldsymbol{E}\{V_t\}\cdot\boldsymbol{E}\{V_{t+m}\} + 2\pi^\tau + \pi^{2\tau}, \quad \text{(B.3)}
\end{aligned}$$

where the second inequality is due to the fact than $\hat{S}_0\hat{S}_k$ depends on $X_1^{\tau+k}$ whereas S_mS_{m+k} depends on $X_{m+1}, X_{m+2}, \ldots$, which are independent. By choosing $\tau = m-k-1$, we have

$$\begin{aligned}
R_V(m) &= \text{Cov}\{V_t, V_{t+m}\} \\
&= \boldsymbol{E}\{V_tV_{t+m}\} - \boldsymbol{E}\{V_t\}\cdot\boldsymbol{E}\{V_{t+m}\} \\
&\le 2\pi^{m-k-1} + \pi^{2(m-k-1)}, \qquad \text{(B.4)}
\end{aligned}$$

which tends to zero as $m\to\infty$ for every fixed k. Likewise,

$$\begin{aligned}
\boldsymbol{E}\{V_tV_{t+m}\} &\ge \boldsymbol{E}\{\hat{S}_0\hat{S}_kS_mS_{m+k}\} \\
&= \boldsymbol{E}\{\hat{S}_0\hat{S}_k\}\boldsymbol{E}\{S_mS_{m+k}\} \\
&\ge \boldsymbol{E}\{(S_0-\pi^\tau)(S_k-\pi^\tau)\}R_S(k) \\
&= [R_S(k)]^2 - 2\pi^\tau - \pi^{2\tau} \\
&= \boldsymbol{E}\{V_t\}\cdot\boldsymbol{E}\{V_{t+m}\} - 2\pi^\tau - \pi^{2\tau}, \text{(B.5)}
\end{aligned}$$

and again, we take $\tau = m-k-1$. Thus, $|\text{Cov}\{V_t, V_{t+m}\}| \le 2\pi^{m-k-1} + \pi^{2(m-k-1)} \to 0$ as $m\to\infty$.

References

[1] B. Chen, *Efficient Communication over Additive White Gaussian Noise and Intersymbol Interference Using Chaotic Sequences*, M.Sc. thesis, Department of Electrical Engineering and Computer Science, 1996.

[2] B. Chen and G. W. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 881-890, July 1998.

[3] L. Cong, W. Xiaofu, and S. Songgeng, "A general efficient method for chaotic signal estimation," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1424–1428, May 1999.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, NJ, USA, 2006.

[5] F. F. Drake, *Information's role in the estimation of chaotic signals*, Ph.D. thesis, Georgia Institute of Technology, August 1998.

[6] J.-P. Eckmann and D. Ruelle, "Ergodic theory chaos and strange attractors," *Reviews of Modern Physics*, vol. 57, no. 3, Part I, pp. 617–656, July 1985.

[7] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[8] I. Hen and N. Merhav, "On the threshold effect in the estimation of chaotic sequences," *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2894–2904, November 2004.

[9] S. Kay and V. Nagesha, "Methods for chaotic signal estimation," *IEEE Trans. Signal Processing*, vol. 43, no. 8, pp. 2013–2016, August 1995.

[10] M. P. Kennedy and G. Kolumbán, "Digital communications using chaos," *Signal Processing*, vol. 80, pp. 1307–1320, 2000.

[11] H. Leung, S. Shanmugam, N. Xie, and S. Wang, "An ergodic approach for chaotic signal estimation at low SNR with application to ultra-wide-band communication," *IEEE Trans. Signal Processing*, Vol. 54, no. 3, pp. 1091–1103, March 2006.

[12] N. Merhav, "Trade-offs between weak–noise estimation performance and outage exponents in non-linear modulation," *IEEE Trans. Inform. Theory*, vol. 65, no. 8, pp. 5189–5196, August 2019.

[13] C. Pantaleón, L. Vielva, D. Luengo and I. Santamaria, "Bayesian estimation of chaotic signals generated by piece-wise linear maps," *Signal Processing*, vol. 80, pp. 659–664, 2003.

[14] H. C. Papadopoulos and G. W. Wornell, "Maximum-likelihood estimation of a class of chaotic signals," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 312–317, January 1995.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 3rd Edition, U.S.A., 1991.

[16] G. J. Wallinger, *Information Theory and Chaotic Systems*, Ph.D. thesis, Graz University of Technology, Graz, Austria, October 2013.

[17] S. Wang, P. C. Yip, and H. Leung, "Estimating initial conditions of noisy chaotic signals generated by piece-wise linear Markov maps using itineraries," *IEEE Trans. on Signal Processing*, vol. 47, no. 12, pp. 3289–3302, December 1999.

[18] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, J. Wiley & Sons, 1965, reissued by Waveland Press, Inc., Prospect Heights, Illinois, 1990.

[19] K. Xie, P. Tant, N. B. Chongl and J. Li (Tiffany), "Analog Turbo codes: a chaotic construction," *Proc. 2009 International Symposium on Information Theory (ISIT 2009)*, pp. 894–898, Seoul, Korea, June-July, 2009.

[20] W. Yu, J. Wu, Y. Li, and B. Zhang, "A hybrid digital-analog chaotic code," *IEEE Wireless Communications Letters*, Vol. 7, no. 6, pp. 930–933, December 2018.

Neri Merhav (S'86–M'87–SM'93–F'99–LF'23) was born in Haifa, Israel, on March 16, 1957. He received the B.Sc., M.Sc., and D.Sc. degrees from the Technion, Israel Institute of Technology, in 1982, 1985, and 1988, respectively, all in electrical engineering.

From 1988 to 1990 he was with AT&T Bell Laboratories, Murray Hill, NJ, USA. Since 1990 he has been with the Electrical Engineering Department of the Technion, where he is now the Irving Shepard Professor. During 1994–2000 he was also serving as a consultant to the Hewlett–Packard Laboratories – Israel (HPL-I). His research interests include information theory, statistical communications, and statistical signal processing. He is especially interested in the areas of lossless/lossy source coding and prediction/filtering, relationships between information theory and statistics, detection, estimation, as well as in the area of Shannon Theory, including topics in joint source–channel coding, source/channel simulation, and coding with side information with applications to information hiding and watermarking systems. Another recent research interest concerns the relationships between Information Theory and statistical physics.

Dr. Merhav was a co-recipient of the 1993 Paper Award of the IEEE Information Theory Society and he is a Fellow of the IEEE since 1999. He also received the 1994 American Technion Society Award for Academic Excellence and the 2002 Technion Henry Taub Prize for Excellence in Research. More recently, he was a co-recipient of the Best Paper Award of the 2015 IEEE Workshop on Information Forensics and Security (WIFS 2015).

During 1996-1999 he served as an Associate Editor for Source Coding to the IEEE TRANSACTIONS ON INFORMATION THEORY, and during 2017-2020 – as an Associate Editor for Shannon Theory in the same journal. He also served as a co–chairperson of the Program Committee of the 2001 IEEE International Symposium on Information Theory.