*Adv. Appl. Prob. (SGSA)* **47,** 1–26 (2015)
*Printed in Northern Ireland*
© *Applied Probability Trust* 2015

# CAPACITY AND ERROR EXPONENTS OF STATIONARY POINT PROCESSES UNDER RANDOM ADDITIVE DISPLACEMENTS

VENKAT ANANTHARAM,* *University of California, Berkeley*
FRANÇOIS BACCELLI,** *University of Texas at Austin and INRIA*

**Abstract**

Consider a real-valued discrete-time stationary and ergodic stochastic process, called the noise process. For each dimension $n$, we can choose a stationary point process in $\mathbb{R}^n$ and a translation invariant tessellation of $\mathbb{R}^n$. Each point is randomly displaced, with a displacement vector being a section of length $n$ of the noise process, independent from point to point. The aim is to find a point process and a tessellation that minimizes the probability of decoding error, defined as the probability that the displaced version of the typical point does not belong to the cell of this point. We consider the Shannon regime, in which the dimension $n$ tends to $\infty$, while the logarithm of the intensity of the point processes, normalized by dimension, tends to a constant. We first show that this problem exhibits a sharp threshold: if the sum of the asymptotic normalized logarithmic intensity and of the differential entropy rate of the noise process is positive, then the probability of error tends to 1 with $n$ for all point processes and all tessellations. If it is negative then there exist point processes and tessellations for which this probability tends to 0. The error exponent function, which denotes how quickly the probability of error goes to 0 in $n$, is then derived using large deviations theory. If the entropy spectrum of the noise satisfies a large deviations principle, then, below the threshold, the error probability goes exponentially fast to 0 with an exponent that is given in closed form in terms of the rate function of the noise entropy spectrum. This is obtained for two classes of point processes: the Poisson process and a Matérn hard-core point process. New lower bounds on error exponents are derived from this for Shannon's additive noise channel in the high signal-to-noise ratio limit that hold for all stationary and ergodic noises with the above properties and that match the best known bounds in the white Gaussian noise case.

*Keywords:* Point process; random tessellation; high-dimensional stochastic geometry; information theory; entropy spectrum; large deviations theory

2010 Mathematics Subject Classification: Primary 60G55; 94A15
Secondary 60D05; 60F10

## 1. Introduction

To study communication over an additive noise channel, information theorists consider transmission via, and decoding from, the noise-corrupted reception. For the purposes of this paper, think of a *codeword* as a sequence of real numbers (called *symbols*) of a fixed length (called the *block length*). A *codebook* is a set of codewords. The allowed codewords in

Received 26 January 2012; revision received 25 March 2014.
* Postal address: Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, 94720, USA. Email address: ananth@eecs.berkeley.edu
** Postal address: Department of Mathematics, The University of Texas at Austin, Austin, TX 78712-1202, USA. Email address: baccelli@math.utexas.edu



https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

the codebook are generally subject to constraints such as power or magnitude constraints, or more complicated constraints such as run length constraints (which are constraints on the allowed patterns of symbols). In this paper we focus on just the power constraint, which is the most important case. The transmitter chooses a codeword to transmit its data over the noisy communication medium. For instance, if there are $2^k$ codewords in the codebook, the transmitter can convey $k$ bits by the choice of the codeword if the communication medium is noise free. Only *additive noise channels* are considered; this is the case where the receiver sees the sum of the transmitted codeword and a noise vector. The receiver does not know the transmitted codeword. The aim is to design the codebook so that the receiver's probability of error is small, assuming that the transmitter was *a priori* equally likely to have transmitted any one of the codewords. One of the main preoccupations of the subject of information theory, initiated by Shannon [17], is to study how to design codes for various communication channels in the asymptotic limit as the block length goes to $\infty$. While situations involving multiple transmitters and receivers are also of great interest, only the single transmitter and receiver case (this is called the *point-to-point* case) is considered in this paper.

In the asymptotic analysis, we require the error probability to be asymptotically vanishing in the block length. Typically, we can do this while having codebooks whose cardinality grows exponentially with the block length. Communication channels are thus characterized first of all by their Shannon capacity, which is the largest possible such exponent. The next question of interest is how quickly the error probability can be made to go to 0 when using codebooks with a *rate* (i.e. exponent of the exponentially growing size of the codebook) that is less than the Shannon capacity. The best possible exponent, as a function of rates below the Shannon capacity, is called the *error exponent function* or *reliability function* of the channel. Characterizing this is largely an open problem and is one of the most challenging mathematical problems in information theory. There are two major classes of lower bounds that can be proved for the error exponent. One is the *random coding bound*, which follows, in the power-constrained case, by considering codewords drawn uniformly at random from the sphere of points that satisfy the power constraint [18]. The second is the *expurgated bound*, which follows from refining this random coding ensemble by specifically eliminating codeword pairs that are too close to each other, while only slightly changing the rate of the codebook, in a way that is asymptotically negligible [9].

Our main contribution is to bring the techniques of point process theory, and more specifically Palm theory [6], [12], to bear on this problem. Our approach is closely related to the earlier work of Poltyrev [16]. However, the Palm theory viewpoint which is brought into play here is not apparent in [16]; this allows us to go well beyond the contribution of that work, which deals only with independent and identically distributed (i.i.d.) Gaussian noise. In this framework, at block length $n$, we need to think about a stationary marked point process on $\mathbb{R}^n$. Each realization of the points is now thought of as a codebook. The power constraint has now vanished, so we can think of being in the infinite signal-to-noise ratio (SNR) limit. A mark is associated to each point. This mark is thought of as the realization of the noise vector when that codeword (synonymous with the point of the process) is 'transmitted'. The noise vectors are independent from point to point, and have the law of a section of length $n$ of a given underlying stationary and ergodic centered real-valued discrete-time stochastic process, which characterizes the communication channel. The 'received noise-corrupted codeword' is thus represented by the sum of the point (codeword) and its mark (noise vector). The decoding problem is to figure out the mother point by knowing just the law of the noise process, the realization of the entire point process, and the sum of the mother point and its mark (without knowing what the mother point is).

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

For instance, in the case of i.i.d. Gaussian noise a natural way to do this would be to consider the Voronoi tessellation [15] of $\mathbb{R}^n$ associated to the point process. Note that the Voronoi cells can also be thought of as marks of the point process. A decoding rule is characterized by its error probability, defined as the limit over large cubes of the error probability per point for the points in that cube (i.e. in order to compute the error probability, we assume that the transmitter chooses one of the points within the cube uniformly at random *a priori*). It is not hard to see that it suffices to consider decoding rules that are jointly stationary with the underlying point process. This means that the error probability can be computed using Palm theory.

For the connection with information theory, the intensity of the underlying point process is itself thought of as scaling exponentially in $n$. The logarithm of the point process intensity on a per unit dimension basis will be called the *normalized logarithmic intensity*. The first question that arises then is: for a given noise process, how large can the asymptotic normalized logarithmic intensity be while still allowing for a choice of the point process (this corresponds to the codebook) and choice of decoding rule for that codebook such that the Palm error probability asymptotically vanishes? Proving the existence of and identifying this threshold would give a point process analog of Shannon's capacity formula. This is the first problem we treat in this paper. There are no surprises here, since it boils down to volume counting. The threshold turns out to be the negative of the differential entropy rate of the noise process. In honor of the pioneering work in [16], we propose to call this threshold the *Poltyrev capacity* of the associated noise process.

Much more interesting is the following question: for a given noise process and a given asymptotic normalized logarithmic intensity that is less than the Poltyrev capacity of the noise process, how large can we make the exponent of the rate at which the Palm error probability can be made to go to 0? Here point process analogs of the random coding and expurgated exponents are found. The random coding exponent comes from considering the Poisson process, while the expurgated exponent comes from considering a Matérn point process. Furthermore, just as the capacity is determined by the differential entropy rate of the noise process, the associated lower bounds on the infinite SNR error exponent are derived from a large deviations principle (LDP) on its *entropy spectrum* (the entropy spectrum for each dimension $n$ is the asymptotic law of the *information density*, which, in turn, is the random variable defined by the logarithm of the density whose expectation, on a per symbol basis, asymptotically yields the differential entropy rate). Identifying this connection is one of the main contributions of the point process formulation investigated in this paper.

Finally, all these results obtained in the infinite SNR setting can be translated back to give lower bounds on the error exponents in the original power-constrained additive noise channel which are new in information theory.

This Palm theory approach was first introduced in [1], where the i.i.d. Gaussian case (called the additive white Gaussian noise (AWGN) case in the information theory literature) was investigated, where it was shown that the infinite SNR random coding and expurgated exponents of [16] could be recovered with this viewpoint. The main contribution of the present paper is to go beyond the i.i.d. Gaussian case to the general stationary and ergodic noise processes, subject to a mild technical condition needed to have an LDP for the entropy spectrum.

The problem is formally set up in Section 2. In Section 3 we prove that the Poltyrev capacity is the threshold for the asymptotic normalized logarithmic intensity in the sense described above. In Section 4 we state representations of the error probability which will be instrumental in analyzing the logarithmic asymptotics of the error probability. In Section 5 we develop the infinite SNR random coding exponent, based on the Poisson process and maximum

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

likelihood (ML) decoding, while in Section 6 we develop the expurgated exponent, based on a Matérn process and ML decoding. Section 7 is devoted to the connections between the results from the point process framework and the motivating problem of information theory, namely, how to translate the lower bounds on the infinite SNR error exponent to those for the problem with power constraints. Several examples of noise processes of practical interest are studied in Section 8. In particular, the AWGN case is studied in depth. In Section 9 we briefly describe how the results of this paper generalize to the case of *mismatched decoding*, which is of significant practical interest.

Throughout the paper, all logarithms are to the natural base. When discussing a family of random variables indexed by the points of a point process, notation such as $\{Z_k\}_k$ is used (this would mean that $Z_k$ is associated to the $k$th point of the process). For all basic definitions pertaining to point process theory, see [6]; information theory, see [5] and [8]; and large deviations theory, see [7] and [19].

## 2. Statement of the problem

### 2.1. Encoding; normalized logarithmic intensity

Fix an integer $n$, and let $(K = K_n, \mathcal{K} = \mathcal{K}_n)$ be a measurable space. Let $\mathbb{M}(K)$ and $\mathbb{M}$ respectively denote the sets of simple marked counting measures $\nu$ on $\mathbb{R}^n \times K$ and simple counting measures $\nu$ on $\mathbb{R}^n$. They are endowed with the $\sigma$-algebras $\mathcal{M}(K)$ and $\mathcal{M}$, respectively, which are generated by the events $\nu(B \times L) = k$ and $\nu(B) = k$, respectively, where $B$ ranges over the Borel sets of $\mathbb{R}^n$, $L$ over the measurable sets of $K_n$, and $k$ over the nonnegative integers (see, e.g. [12]).

Each $\nu \in \mathbb{M}$ has a representation of the form

$$\nu = \sum_k \varepsilon_{t_k},$$

with $\varepsilon_x$ the Dirac measure at $x$ and $\{t_k\}_k$ the atoms of the counting measure $\nu$. Similarly, each $\nu \in \mathbb{M}(K)$ has a representation of the form

$$\nu = \sum_k \varepsilon_{t_k, m_k},$$

with $\{(t_k, m_k)\}_k$ the atoms of $\nu$, where $t_k \in \mathbb{R}^n$ and $m_k \in K_n$. The set $\{t_k\}_k$ is the set of *points* of $\nu^n$ and the set $\{m_k\}_k$ is its set of *marks*. Let $\mathbb{M}_0(K)$ and $\mathbb{M}_0$ respectively denote the sets of all simple marked counting measures and simple counting measures with an atom whose first coordinate is 0.

Below, only stationary and ergodic marked point processes are considered. Thus, it is assumed that, for each $n \geq 1$, there exists a probability space $(\Omega, \mathcal{G}, \mathbb{P}, \theta_t)$, endowed with an ergodic and measure preserving shift $\theta_t$ indexed by $t \in \mathbb{R}^n$. A *stationary marked point process* $\mu$ on $\mathbb{R}^n \times K_n$ is a measurable map from $(\Omega, \mathcal{G})$ to $(\mathbb{M}(K), \mathcal{M}(K))$ such that, for all $t \in \mathbb{R}^n$,

$$\mu(\theta_t(\omega)) = \tau_t(\mu(\omega)),$$

where $\tau_t(\mu)$ is the translation of $\mu$ by $-t \in \mathbb{R}^n$: if $\mu(\omega) = \sum_k \varepsilon_{T_k(\omega), M_k(\omega)}$ then

$$\mu(\theta_t(\omega)) = \sum_k \varepsilon_{-t+T_k(\omega), M_k(\omega)}.$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

Let $\lambda_n$ denote the intensity of $\mu = \mu^n$. The scaling where the normalized logarithmic intensity approaches a limit as the dimension $n$ goes to $\infty$ is of particular interest. Here $R_n$, the *normalized logarithmic intensity* of $\mu^n$, is defined via $\lambda_n = \mathrm{e}^{nR_n}$. Denote by $\mathbb{P}_0$ the Palm probability [6] of $\mu$ (by convention, under $\mathbb{P}_0$, $T_0 = 0$), and by $\mathcal{V}_k$ the Voronoi cell of point $T_k$ with respect to the point process $\mu$ (see, e.g. [15]), which is taken here to be an open set.

As described informally in Section 1, the points of this point process are thought of as representing the codewords used by a transmitter in a communication system in the *infinite SNR limit*. This connection with information theory provides the motivation for the scaling considered here. Realizations of the noise vector, as well as decoding regions (see below), are typical examples of the kinds of marks considered.

## 2.2. Decoding

To each point $T_k$ of the point process $\mu = \mu^n$ (the superscript $n$ is omitted in this section), we associate the independent mark $D_k$, a random vector taking values in $\mathbb{R}^n$, called the *displacement vector*. When the point $T_k$ of $\mu$ is thought of as a codeword, the transmission over an additive noise channel adds to it the displacement vector $D_k$, so that the received point is $Y_k = T_k + D_k$.

Decoding is discussed in terms of a sequence of marks of $\mu$ which are measurable sets of $\mathbb{R}^n$. The mark of point $T_k$ will be denoted by $\mathcal{C}_k$. The set $\mathcal{C}_k$ is the decoding region of $T_k$. The sets $\{\mathcal{C}_k\}_k$ are required to form a *tessellation of* $\mathbb{R}^n$, namely, they are all disjoint and the union of their closures is $\mathbb{R}^n$.

The displacement sequence $\{D_k\}_k$ is assumed to be i.i.d. and independent of the marked point process $\{T_k, \mathcal{C}_k\}_k$. This makes $\{T_k, (D_k, \mathcal{C}_k)\}_k$ a marked point process.

The canonical example to keep in mind, which is motivated by the AWGN channel of information theory, is when the vectors associated to the individual points of the process are i.i.d. zero-mean Gaussian random vectors each with i.i.d. coordinates and independent of the points. Then the natural choice of the decoding region of a point is its Voronoi cell in the realization of the point process.

The most general setting concerning the noise (or displacement vectors) considered in this paper will feature a real-valued centered stationary and ergodic stochastic process $\Delta = \{\Delta_l\}$, and displacement vectors $\{D_k\}_k$ independent of the point process, i.i.d. in $k$, and with a law defined by $D = D^n = (\Delta_1, \ldots, \Delta_n)$ for all $n$. As will be seen, more elaborate though natural decoding tessellations then show up, determined by the law of $\Delta$.

The decoding strategy associated with the sequence of marks $\{\mathcal{C}_k\}_k$ expects that, when $T_k$ is transmitted, then the received point $Y_k$ lands in $\mathcal{C}_k$. An error happens if this is not the case. The error probability is now formally defined in a Palm theory setting. Our eventual goal, as informally described in Section 1, is to study the exponent of decay in $n$ of the error probability.

## 2.3. Probability of error

Within the above setting, for all $n$, when $(\mu^n, \mathcal{C}^n)$ and the law of $D^n$ are given, define the associated *probability of error* as

$$p_{\mathrm{e}}(n) = \lim_{W\to\infty} \frac{\sum_k \mathbf{1}_{\{T_n^k \in B^n(0,W)\}} \mathbf{1}_{\{Y_k^n \notin \mathcal{C}_k^n\}}}{\sum_k \mathbf{1}_{\{T_n^k \in B^n(0,W)\}}} . \tag{1}$$

The limit in (1) exists almost surely and is nonrandom. This follows from the assumption that the marked point process $\mu^n$ with marks $(D_k^n, \mathcal{C}_k^n)$ is stationary and ergodic. The pointwise ergodic theorem implies that

$$p_{\mathrm{e}}(n) = \mathbb{P}_0^n(Y_0^n \notin \mathcal{C}_0^n) = \mathbb{P}_0^n(D_0^n \notin \mathcal{C}_0^n). \tag{2}$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

## 3. Poltyrev capacity of an additive noise channel

The infinite SNR additive noise channel for dimension $n$ is characterized by the law of the typical displacement vector $D^n$, with $D^n = (\Delta_1, \ldots, \Delta_n)$, with $\Delta = \{\Delta_l\}$ as defined above. It will also be assumed that these displacement vectors $D^n$ have a density $f^n$ admitting a differential entropy rate

$$h(\Delta) = -\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln f^n(D^n)]. \tag{3}$$

We define $-h(\Delta)$ to be the *Poltyrev capacity* of the additive noise channel with displacement vectors defined in terms of the process $\Delta$.

The terminology is chosen in honor of Poltyrev's work [16]. The justification for this terminology comes from the following two simple theorems, which together give an analog of Shannon's capacity theorem for additive noise channels in information theory. Before stating and proving these theorems, recall that, for $\delta > 0$, if we let

$$\mathcal{A}_\delta^n = \left\{x^n \in \mathbb{R}^n : \left| -\frac{1}{n}\ln(f^n(x^n)) - h(\Delta) \right| < \delta \right\},$$

then we have

$$\mathbb{P}(D^n \in \mathcal{A}_\delta^n) \to 1 \quad \text{as } n \to \infty. \tag{4}$$

This can be seen as a consequence of either one of the generalized Shannon–McMillan–Breiman theorems in [4] or [13].

**Theorem 1.** *For all point processes $\mu^n$ such that $\liminf_n R_n > -h(\Delta)$, and all choices of decoding regions $\mathcal{C}_k^n$ which are subsets of $\mathbb{R}^n$ jointly stationary with the points and forming a tessellation of $\mathbb{R}^n$, we have $\lim_{n\to\infty} p_{\mathrm{e}}(n) = 1$.*

*Proof.* For all stationary tessellations $\{\mathcal{C}_k^n\}_k$, we have

$$\mathbb{P}_0^n(D_0^n \in \mathcal{C}_0^n) \le \mathbb{E}_0^n\big(\mathbf{1}_{\{D_0^n\in\mathcal{C}_0^n\cap\mathcal{A}_\delta^n\}}\big) + \mathbb{E}_0^n\big(\mathbf{1}_{\{D_0^n\notin\mathcal{A}_\delta^n\}}\big).$$

The second term tends to 0 as $n$ tends to $\infty$ because of (4). The first term is

$$\mathbb{E}_0^n\bigg(\int_{\mathcal{C}_0^n\cap\mathcal{A}_\delta^n} f^n(x^n)\,\mathrm{d}x^n\bigg).$$

It is bounded from above by $\mathrm{e}^{-n(h(D)-\delta)}\mathbb{E}_0^n(\mathrm{vol}(\mathcal{C}_0^n))$, and, for all translation invariant tessellations of the Euclidean space, $\mathbb{E}_0^n(\mathrm{vol}(\mathcal{C}_0^n)) = \mathrm{e}^{-nR_n}$, which allows us to complete the proof.

**Theorem 2.** *Let $\mu^n$ be a Poisson point process of intensity $\lambda_n = \mathrm{e}^{nR_n}$. If $\limsup_n R_n < -h(\Delta)$ then it is possible to choose decoding regions $\mathcal{C}_k^n$ that are subsets of $\mathbb{R}^n$ jointly stationary with the points and the displacements, forming a stationary tessellation of $\mathbb{R}^n$, such that $\lim_{n\to\infty} p_{\mathrm{e}}(n) = 0$.*

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

*Proof.* Let $\{\mathcal{C}_k^n\}_k$ be the following tessellation of $\mathbb{R}^n$:

$$\mathcal{C}_k^n = \left\{ (T_k^n + \mathcal{A}_\delta^n) \cap \left\{ \bigcup_{l \neq k} (T_l^n + \mathcal{A}_\delta^n) \right\}^c \right\}$$
$$\cup \left\{ \mathcal{V}_k^n \cap \left\{ \bigcup_{l \neq l'} [(T_l^n + \mathcal{A}_\delta^n) \cap (T_{l'}^n + \mathcal{A}_\delta^n)] \right\} \right\}$$
$$\cup \left\{ \mathcal{V}_k^n \cap \left\{ \bigcup_{l} (T_l^n + \mathcal{A}_\delta^n)^c \right\} \right\},$$

where $\mathcal{V}_k^n$ denotes the Voronoi cell of $T_k^n$. In words, $\mathcal{C}_k^n$ contains all the locations $x$ which belong to the set $T_k^n + \mathcal{A}_\delta^n$ and to no other set of the form $T_l^n + \mathcal{A}_\delta^n$, all the locations $x$ that are ambiguous (i.e. belong to two or more such sets) and which are closer to $T_k^n$ than to any other point, and all the locations which are uncovered (i.e. belong to no such set) and which are closer to $T_k^n$ than to any other point. This scheme will be referred to as *typicality decoding* in what follows.

Let $\mu_!^n = \mu^n - \varepsilon_0$. Consider the bound

$$\mathbb{P}_0^n(D_0^n \notin \mathcal{C}_0^n) \leq \mathbb{P}_0^n(D_0^n \notin \mathcal{A}_\delta^n) + \mathbb{P}_0^n(D_0^n \in \mathcal{A}_\delta^n, \mu_!^n(D_0^n - \mathcal{A}_\delta^n) > 0).$$

The first term tends to 0 due to (4). For the second term, Slivnyak's theorem [6] is used to bound it from above by

$$\mathbb{P}^n(\mu^n(D_0^n - \mathcal{A}_\delta^n) > 0) \leq \mathbb{E}^n(\mu^n(D_0^n - \mathcal{A}_\delta^n)) = \mathbb{E}^n(\mu^n(-\mathcal{A}_\delta^n)) = \mathrm{e}^{nR_n} |\mathcal{A}_\delta^n| \,.$$

But

$$\begin{aligned} 1 &\geq \mathbb{P}^n(D_0^n \in \mathcal{A}_\delta^n) \\ &= \int_{\mathcal{A}_\delta^n} f^n(x^n)\,\mathrm{d}x^n \\ &= \int_{\mathcal{A}_\delta^n} \mathrm{e}^{n(1/n)\ln(f^n(x^n))}\,\mathrm{d}x^n \\ &\geq \int_{\mathcal{A}_\delta^n} \mathrm{e}^{n(-h(D)-\delta)}\,\mathrm{d}x \\ &= \mathrm{e}^{-n(h(D)+\delta)} |\mathcal{A}_\delta^n|, \end{aligned}$$

so that $|\mathcal{A}_\delta^n| \leq \mathrm{e}^{n(h(D)+\delta)}$, which allows us to complete the proof.

Examples of stationary and ergodic noise processes are considered in Section 8. The reader may wish to consult some of the examples at this stage for concrete instances of the result above.

## 4. Maximum likelihood decoding

In this section we present representations of the ML decoding error probability that will be instrumental for the evaluation of error exponents in the forthcoming sections.

As in Section 3, $f^n$ denotes the density of the displacement vector $D^n = (\Delta_1, \ldots, \Delta_n)$ which is a section of $\Delta = \{\Delta_l\}$, a real-valued centered stationary and ergodic stochastic process. The function

$$y^n \in \mathbb{R}^n \to \ell_{f^n}(y^n) = \frac{1}{n} \ln(f^n(y^n)) \in \mathbb{R}$$

is the (rescaled) log-likelihood of $f^n$ at $y^n$. Note that $\ell_{f^n}(y^n) \in [-\infty, +\infty]$ in general.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

Below, $-\ell_{f^n}$ is often used rather than $\ell_{f^n}$. The reason is that the real-valued random variable $-\ell_{f^n}(D^n)$ is well known and referred to as the *normalized entropy density* of $D^n$ [11]. Its law, denoted by $\rho^n_\Delta(\mathrm{d}u)$, is referred to as the *entropy spectrum* of $D^n$ [11]. Note that the existence of a density for $D^n$ does not imply that $\rho^n_\Delta(\cdot)$ admits a density. Furthermore, the support of $\rho^n_\Delta(\cdot)$ is not necessarily the whole real line.

The sets

$$\mathcal{S}^n_\Delta(u) = \{y^n \in \mathbb{R}^n \colon -\ell_{f^n}(y^n) \le u\}, \qquad u \in \mathbb{R},$$

will be referred to as the *log-likelihood level sets* of $D^n$. The volume $W^n_\Delta(u)$ of $\mathcal{S}^n_\Delta(u)$ will be referred to as the *log-likelihood level volume* for $u$. The measure $w^n_\Delta$ on $\mathbb{R}$ defined by

$$w^n_\Delta(B) = \mathrm{vol}\{y^n \in \mathbb{R}^n \colon -\ell_{f^n}(y^n) \in B\}$$

for all Borel sets $B$ of the real line, will be called the *log-likelihood level measure*. It turns out that the measures $w^n_\Delta$ and $\rho^n_\Delta$ are mutually absolutely continuous. Indeed, we have

$$\rho^n_\Delta(B) = \int \mathbf{1}_{\{-\ell_{f^n}(x^n)\in B\}}\, f^n(x^n)\,\mathrm{d}x^n,$$

which implies that, for all bounded Borel sets $B$ of the real line,

$$\mathrm{e}^{-n\sup(B)} w^n_\Delta(B) \le \rho^n_\Delta(B) \le \mathrm{e}^{-n\inf(B)} w^n_\Delta(B). \tag{5}$$

From (5), it immediately follows that the measure $w^n_\Delta$ is $\sigma$-finite. Also, for all $u$,

$$W^n_\Delta(u) = \int_{(-\infty,u]} w^n_\Delta(\mathrm{d}s) = \int_{(-\infty,u]} \mathrm{e}^{ns}\rho^n_\Delta(\mathrm{d}s). \tag{6}$$

Since $\mu^n$ is a point process, for all $x^n$, $\mathbb{P}^n_0$-almost surely ($\mathbb{P}$-a.s.), the $\mathbb{R}^n$–valued sequence $\{x^n - T^n_k\}_k$ has no accumulation point. Hence, $\mathbb{P}^n_0$-a.s., the set

$$\mathrm{argmax}_k \ell_{f^n}(x^n - T^n_k)$$

is nonempty, i.e. the supremum is achieved by at least one $k$. By definition, under ML decoding, when $x^n$ is received, we return the codeword $\mathrm{argmax}_k \ell_{f^n}(x^n - T^n_k)$ if the latter is uniquely defined. If there is *ambiguity*, i.e. if there are several solutions to the above maximization problem, then we return any one of them.

Given that $0 = T^n_0$ is 'transmitted' and that the realization of the additive noise is $x^n$, a sufficient condition for ML decoding to be successful is that $\mu^n$ has no point $T^n_k$ other than $T^n_0 = 0$ such that $\ell_{f^n}(x^n - T^n_k) \ge \ell_{f^n}(x^n)$. But, for all $x^n$,

$$\ell_{f^n}(x^n - T^n_k) < \ell_{f^n}(x^n) \quad \text{for all } k \ne 0 \quad \text{if and only if} \quad (\mu^n - \varepsilon_0)(\mathcal{F}(x^n)) = 0,$$

with

$$\mathcal{F}(x^n) = \{y^n \in \mathbb{R}^n \colon \ell_{f^n}(x^n - y^n) \ge \ell_{f^n}(x^n)\}.$$

Hence,

$$p_{\mathrm{e}}(n) \le \mathbb{P}^n_0((\mu^n - \varepsilon_0)(\mathcal{F}(D^n)) > 0).$$

Also, note that the volume of the set $\mathcal{F}(x^n)$ depends only on $\ell_{f^n}(x^n)$. If this last quantity is equal to $-u$, the associated volume is

$$\mathrm{vol}\{y^n \in \mathbb{R}^n \colon -\ell_{f^n}(x^n - y^n) \le -u\} = \mathrm{vol}\{y^n \in \mathbb{R}^n \colon -\ell_{f^n}(y^n) \le -u\},$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

i.e.

$$\mathrm{vol}(\mathcal{F}(x^n)) = W_\Delta^n(-\ell_{f^n}(x^n)). \tag{7}$$

The main result of this section is stated in the following theorem.

**Theorem 3.** *For all stationary and ergodic point processes $\mu^n$ and all i.i.d. displacement vectors, under ML decoding,*

$$p_{\mathrm{e}}(n) \leq 1 - \int_{x^n \in \mathbb{R}^n} \mathbb{P}_0^n((\mu^n - \varepsilon_0)(\mathcal{F}(x^n)) = 0) f^n(x^n)\, \mathrm{d}x^n. \tag{8}$$

*If $\mu^n$ is such that, under $\mathbb{P}_0^n$, the point process $\mu^n - \varepsilon_0$ admits an intensity bounded from above by the function $g^n(\cdot)$ on $\mathbb{R}^d$, then*

$$p_{\mathrm{e}}(n) \leq \int_{x^n \in \mathbb{R}^n} \min\bigg(1, \int_{\mathcal{F}(x^n)} g^n(y^n)\, \mathrm{d}y^n\bigg) f^n(x^n)\, \mathrm{d}x^n. \tag{9}$$

*If $\mu^n$ is a Poisson of intensity $\lambda_n$ then*

$$p_{\mathrm{e}}(n) \leq 1 - \int_{u \in \mathbb{R}} \exp(-\lambda_n W_\Delta^n(u)) \rho_\Delta^n(\mathrm{d}u), \tag{10}$$

*where $\rho_\Delta^n(\mathrm{d}u)$ is the entropy spectrum of $f^n$.*

*Proof.* The probability of success (given that 0 is sent and that the additive noise is $x^n$) is the probability that $\mu^n$ has no point other than 0 in $\mathcal{F}(x^n)$, which proves (8). Equation (9) is immediate from (8) and the definition of $\mathcal{F}(x^n)$. Equation (10) follows from (7), (8), and Slivnyak's theorem.

With the preceding discussion of ML decoding in view, it is convenient to define the *(log-)likelihood cell* $\mathcal{L}_k^n(\Delta)$ of point $T_k^n$ as follows:

$$\mathcal{L}_k^n(\Delta) = \{x^n : \ell_{f^n}(x^n - T_k^n) > \inf_{l \neq k} \ell_{f^n}(x^n - T_l^n)\}$$
$$\cup \{x^n : \ell_{f^n}(x^n - T_k^n) = \ell_{f^n}(x^n - T_l^n) \text{ for some } l \neq k\} \cap \mathcal{V}_k^n.$$

It is comprised of the locations $x^n$ with a likelihood (with respect to $f^n$) to $T_k^n$ larger than that to any other point; as well as the locations $x^n$ with an ambiguous log-likelihood but which are closer to $T_k^n$ for Euclidean distance than to all other points of $\mu^n$. These cells form a stationary tessellation of the Euclidean space which we refer to as the *likelihood tessellation* with respect to the point process $\mu^n$ for the noise $\Delta$ (more precisely, $D^n$ or $f^n$). The likelihood tessellation with respect to additive white Gaussian noise with positive variance is the Voronoi tessellation for all dimensions $n$, all point processes $\mu^n$ on $\mathbb{R}^n$, and all $k$.

The resolution of ambiguity in this definition is somewhat Gaussian centric. Any other tessellation whose cells satisfy the conditions of Section 2 could be used in place of the Voronoi tessellation.

## 5. Random coding exponent: the Poisson process

Let $\Delta$ be a stationary, ergodic, centered, discrete-time, real-valued stochastic process. For all stationary and ergodic point processes $\mu^n$ of normalized logarithmic intensity $-h(\Delta) - \ln(\alpha)$, with $\alpha > 1$, and decoding regions $\mathcal{C}^n = \{\mathcal{C}_k^n\}_k$ jointly stationary and ergodic with $\mu^n$, let

$$p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha, \Delta)$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

denote the probability of error associated with these data, as defined in (2). The '*pp*' superscript is used to recall that the setting is the point process setting described in Section 2.

For a fixed family $(\mu, \mathcal{C}) = (\mu^n, \mathcal{C}^n)$ of a jointly stationary and ergodic point process and decoding region for each dimension $n$, with normalized logarithmic intensity $-h(\Delta) - \ln(\alpha_n)$ for all dimensions $n \geq 1$ and with $\alpha_n \to \alpha$ as $n \to \infty$, let

$$\overline{\pi}(\mu, \mathcal{C}, \alpha, \Delta) = \limsup_n -\frac{1}{n} \ln(p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha_n, \Delta)),$$
$$\underline{\pi}(\mu, \mathcal{C}, \alpha, \Delta) = \liminf_n -\frac{1}{n} \ln(p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha_n, \Delta)). \tag{11}$$

The assumptions on the density $f^n$ on $\mathbb{R}^n$ of $D^n = (\Delta_1, \dots, \Delta_n)$ under which error exponents will be analyzed in the point process formulation are summarized below (where H-SEN stands for hypothesis on stationary ergodic noise).

**Assumption H-SEN.** (i) *For all n, the differential entropy of $f^n$, $h(D^n)$, is well defined.*

(ii) *The differential entropy rate of $\Delta = \{\Delta_l\}$, i.e. $h(\Delta)$, as defined in (3), exists and is finite.*

(iii) *The entropy spectrum $\rho_\Delta^n(\mathrm{d}u)$, i.e. the law of the random variables $\{-(1/n)\ln(f^n(D^n))\}$, satisfies an LDP (on the real line endowed with its Borel $\sigma$-field), with good (in particular lower semicontinuous) and convex rate function $I(x)$ [7], [19].*

A simple sufficient condition for Assumption H-SEN(iii) to hold is that the conditions of the Gärtner-Ellis theorem hold, namely that the limit

$$\lim_{n\to\infty} \frac{1}{n} \ln(\mathbb{E}((f^n(D^n))^{-\theta})) =: G(\theta)$$

exists as an extended real number, is finite in some neighborhood of the origin, and is essentially smooth (see [7, Definition 2.3.5]). From the Gärtner-Ellis theorem, the family of measures $\rho_\Delta^n(\mathrm{d}x)$ then satisfies an LDP with good and convex rate function

$$I(x) = \sup_\theta(\theta x - G(\theta)). \tag{12}$$

The following lemma gives the log-scale asymptotics of the log-likelihood level volumes. (Some of the results derived below do not require this convexity assumption.)

**Lemma 1.** *Suppose that Assumption H-SEN holds. Then*

$$\sup_{s<u}(s - I(s)) \leq \liminf_{n\to\infty} \frac{1}{n} \ln(W_\Delta^n(u)) \leq \limsup_{n\to\infty} \frac{1}{n} \ln(W_\Delta^n(u)) \leq \sup_{s\leq u}(s - I(s)).$$

*Furthermore, the function*

$$J(u) = \sup_{s\leq u}(s - I(s)), \tag{13}$$

*which will be referred to as the* volume exponent, *is upper semicontinuous.*

*Proof.* From (6),

$$W_\Delta^n(u) \geq \int \mathrm{e}^{n\phi(s)} \rho_\Delta^n(\mathrm{d}s),$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

where

$$\phi(s) = \begin{cases} 1 & \text{if } s < u, \\ -\infty & \text{if } s \geq u. \end{cases}$$

Since $\rho_\Delta^n(\mathrm{d}x)$ satisfies an LDP and since the function $\phi$ is lower semicontinuous, the lower bound is proved as in Lemma 4.3.4 of [7]. Similarly,

$$\widetilde{\phi}(s) = \begin{cases} s & \text{if } s \leq u, \\ -\infty & \text{if } s > u, \end{cases}$$

is upper semicontinuous and the upper bound is proved as in Lemma 4.3.6 of [7]. In both cases, it should be noted that the proofs in [7] actually allow for functions $\phi$ with values in $\{-\infty\} \cup \mathbb{R}$.

We now show that the upper semicontinuity of the function $g(s) = s - I(s)$ implies the upper semicontinuity of the function $J(u) = \sup_{s \leq u} g(s)$. We have to show that

$$J(u) \geq \lim_{\varepsilon \to 0} \sup_{s \in [u-\varepsilon, u+\varepsilon]} J(s) = \lim_{\varepsilon \to 0} J(u+\varepsilon), \tag{14}$$

where the rightmost equality follows from the fact that $J$ is nondecreasing. Hence, using monotonicity again, we have to show that $J$ is right continuous.

We have

$$J(u+\varepsilon) = J(u) + \sup_{s \in [u, u+\varepsilon]} (g(s) - J(u))^+,$$

with $a^+ = \max(a, 0)$. So, either $g(s) \leq J(u)$ for all $s \in [u, u+\varepsilon]$, in which case $J(u+\varepsilon) = J(u)$ and the right continuity is trivially satisfied, or $g(s) > J(u)$ for some $s \in [u, u+\varepsilon]$, in which case

$$J(u+\varepsilon) = \sup_{[u, u+\varepsilon]} g(s).$$

It then follows from the upper semicontinuity of the function $g(s)$ that

$$J(u) \geq g(u) \geq \lim_{\varepsilon \to 0} \sup_{[u, u+\varepsilon]} g(s) = \lim_{\varepsilon \to 0} J(u+\varepsilon),$$

so that (14), and hence right continuity, holds in this case too.

Since $I(h(\Delta)) = 0$, it follows from (13) that $J(h(\Delta)) \geq h(\Delta)$. The concavity of the function $x \to x - I(x)$ implies that this function is nondecreasing on the interval $(-\infty, h(\Delta)]$. Hence, from (13), we have

$$J(h(\Delta)) = h(\Delta).$$

Furthermore, we may conclude that at all points $u$ of continuity of $J$ we have

$$\lim_{n \to \infty} \frac{1}{n} \ln(W_\Delta^n(u)) = J(u).$$

The following theorem, which follows from considering the family of Poisson point processes with ML decoding, yields the random coding exponent for the problem formulation adopted here.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

**Theorem 4.** *Assume that $\mu^n$ is Poisson with normalized logarithmic intensity $-h(\Delta)-\ln(\alpha)$ with $\alpha>1$ and that the decoder uses ML decoding. Suppose that Assumption H-SEN holds. Then the associated error exponent is such that*

$$\underline{\pi}(\text{Poi}, \mathcal{L}(\Delta), \alpha, \Delta) \geq \inf_u \{F(u)+I(u)\}, \tag{15}$$

*where $I(u)$ is the rate function of $\rho_\Delta^n$ (defined in (12)) and*

$$F(u) = (\ln(\alpha)+h(\Delta)-J(u))^+,$$

*where $J(u)=\sup_{s\leq u}(s-I(s))$ is the volume exponent defined in Lemma 1.*

*Proof.* From (10),

$$p_{\mathrm{e}}(n) \leq \int_{u\in\mathbb{R}} (1-\exp(-\lambda_n W_\Delta^n(u)))\rho_\Delta^n(\mathrm{d}u). \tag{16}$$

Using (16) and the bound

$$1-\mathrm{e}^{-\lambda_n W_\Delta^n(u)} \leq \min(1, \lambda_n W_\Delta^n(u)),$$

we can write

$$p_{\mathrm{e}}(n) \leq \int_u \mathrm{e}^{-n\phi_n(u)}\rho_\Delta^n(\mathrm{d}u),$$

with

$$\phi_n(u) = \left(\ln(\alpha)+h(\Delta)-\frac{1}{n}\ln(W_\Delta^n(u))\right)^+.$$

In order to conclude, we use Theorem 2.3 of [19]. Since the law $\rho_\Delta^n(\mathrm{d}u)$ satisfies an LDP with good rate function $I(u)$, it is enough to prove that, for all $\delta>0$, there exists $\varepsilon>0$ such that

$$\liminf_{n\to\infty} \inf_{(u-\varepsilon,u+\varepsilon)} \left(\ln(\alpha)+h(\Delta)-\frac{1}{n}\ln(W_\Delta^n(u))\right)^+ \geq (\ln(\alpha)+h(\Delta)-J(u))^+ - \delta.$$

Since the function $u\to W_\Delta^n(u)$ is nondecreasing, it is enough to show that, for all $\delta>0$, there exists $\varepsilon>0$ such that

$$\lim_{n\to\infty}\left(\ln(\alpha)+h(\Delta)-\sup_{m\geq n}\frac{1}{m}\ln(W_\Delta^m(u+\varepsilon))\right)^+ \geq (\ln(\alpha)+h(\Delta)-J(u))^+ - \delta.$$

There are two cases to consider: if $\ln(\alpha)+h(\Delta)-J(u)\leq 0$, the result is obvious, and if $\ln(\alpha)+h(\Delta)-J(u)>0$ then we have to prove that, for all $\delta$, there exists an $\varepsilon$ such that

$$\lim_{n\to\infty}\sup_{m\geq n}\frac{1}{m}\ln(W_\Delta^m(u+\varepsilon)) \leq \sup_{s\leq u}(s-I(s))+\delta.$$

But, from Lemma 1,

$$\lim_{n\to\infty}\sup_{m\geq n}\frac{1}{m}\ln(W_\Delta^m(u+\varepsilon)) \leq \sup_{s\leq u+\varepsilon}(s-I(s)).$$

Hence, it is enough to show that, for all $\delta$, there exists an $\varepsilon$ such that

$$\sup_{s\leq u}(s-I(s)) \geq \sup_{s\leq u+\varepsilon}(s-I(s))-\delta.$$

This follows from the fact that the function $J(u)$ is upper semicontinuous.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

Note that all terms in the final expression to be minimized, namely,

$$(\ln(\alpha) + h(\Delta) - J(u))^+ + I(u),$$

have a simple conceptual meaning. Here $e^{-(\ln(\alpha)+h(\Delta))}$ is the intensity, i.e. $\lambda_n$; $e^{nJ(u)}$ is the volume of the log-likelihood level set for level $u$; $e^{-nI(u)}$ is the value of the density of the entropy spectrum at $u$; and, finally, the positive part stems from the minimum of the mean number of points in the above set and the number 1.

## 6. Expurgated exponent: a Matérn process

A Matérn I point process is created by deleting points from a Poisson process as follows. Choose some positive radius called the exclusion radius. Any point in the initial Poisson process that has another point within this fixed exclusion radius is deleted (note that both points will be deleted since the first point will also be within the same fixed radius of the second point). This is the simplest type of *hard-sphere exclusion*. For an information theorist, this is reminiscent of expurgation [9] and this term will also be used below to describe the transformation of the Poisson into a Matérn point process. This process, and a related process called the Matérn II process, were introduced in [14]. The Matérn II process will not be considered in this paper.

Mimicking this idea, a new class of Matérn point processes is introduced in order to cope with the general stationary and ergodic noise in the present problem formulation. Assume for simplicity that $f^n(x^n) = f^n(-x^n)$. If two points $S$ and $T$ of the Poisson point process $\mu^n$ are such that $-\ell_{f^n}(T - S) < \xi$, with $\xi \in \mathbb{R}$ some threshold, then both $T$ and $S$ are deleted ($-\ell_{f^n}$ may be thought of as a surrogate distance; two points which are 'too close' are discarded). The surviving points form the Matérn-$\Delta$-$\xi$ point process $\widehat{\mu}^n$.

**Theorem 5.** *Under the assumptions of Theorem 3, the probability of error for the Matérn-$\Delta$-$\xi$ point process satisfies the bound*

$$p_e(n) \le \int_{x^n \in \mathbb{R}^n} \min\left(1, \lambda_n \int_{y^n \in \mathbb{R}^n} \mathbf{1}_{\{-\ell_{f^n}(y^n) \ge \xi\}} \mathbf{1}_{\{\ell_{f^n}(x^n - y^n) \le \ell_{f^n}(x^n)\}} \, dy^n\right) f^n(x^n) \, dx^n. \quad (17)$$

*Proof.* Let $\widehat{\mathbb{P}}_0^n$ denote the Palm probability of $\widehat{\mu}^n$. Under $\widehat{\mathbb{P}}_0^n$, the point process $\widehat{\mu}^n - \varepsilon_0$ has an intensity bounded from above by $\lambda_n \mathbf{1}_{\{-\ell_{f^n}(y^n) \ge \xi\}}$ at $y^n$. The result then follows from (9).

Note that the Matérn-AWGN-$\xi$ reduces to the Matérn I model for the exclusion radius

$$r_n(\xi) = \sqrt{2n\sigma^2}\sqrt{\xi - \tfrac{1}{2}\ln(2\pi\sigma^2)} \quad (18)$$

for $\xi > \frac{1}{2}\ln(2\pi\sigma^2)$. Hence, the following special case holds.

**Theorem 6.** *In the AWGN case,*

$$p_e(n) \le \int_{r>0} \min(1, \lambda_n \mathrm{vol}(B^n(0, r_n(\xi))^c \cap B^n(x^n(r), r))) g_\sigma^n(r) \, dr \quad (19)$$

*with $x^n(r) = (r, 0, \ldots, 0) \in \mathbb{R}^n$ and $r_n(\cdot)$ defined in (18).*

*Proof.* The result immediately follows from (17) and (18).

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

In the general case, the unfortunate fact that the volume of the vulnerability set (the set which ought to be empty of points for no error to occur) now depends on the point $x^n$, and not only on the value of $\ell_{f^n}(x^n)$, can be taken care of by introducing the upper bound

$$M^n_\Delta(u,\xi) = \sup_{\{x^n\colon -\ell_{f^n}(x^n)=u\}} \int_{y^n\in\mathbb{R}^n} \mathbf{1}_{\{-\ell_{f^n}(y^n)\geq\xi\}}\,\mathbf{1}_{\{-\ell_{f^n}(x^n-y^n)\leq u\}}\,\mathrm{d}y^n,$$

which depends only on $\ell_{f^n}(x^n)$. This quantity will be referred to as the *expurgated log-likelihood level volume*. By the same arguments as above, we obtain the following result.

**Corollary 1.** *The probability of error for the Matérn-$\Delta$-$\xi$ point process satisfies the bound*

$$p_{\mathrm{e}}(n) \leq \int_{u\in\mathbb{R}} \min(1, \lambda_n M^n_\Delta(u,\xi))\rho^n_\Delta(\mathrm{d}u).$$

In Section 8 we calculate the expurgated exponent based on the Matérn-$\Delta$-$\xi$ process in some examples. Particular attention is paid to the AWGN case, where it is shown that the expurgated exponent of [16] can be recovered.

## 7. The channel with power constraints

In the traditional model for point-to-point communication over an additive noise channel with power-constrained inputs, the codewords, of block length $n$, are subject to the power constraint $P$. A codebook is thus a finite, nonempty subset, call it $\mathcal{T}$, of points in $B^n(0,\sqrt{nP})$ (the closed ball of radius $\sqrt{nP}$ around the origin), whose elements are the codewords. Then $R(\mathcal{T}) = (1/n)\ln\mid\mathcal{T}\mid\;\geq 0$ is the rate of the code. The noise vector for block length $n$, $D^n = (\Delta_1,\dots,\Delta_n)$, is assumed to have the law of the first $n$ values of the centered, real-valued stationary, and ergodic stochastic process $\Delta = \{\Delta_l\}$. Suppose that Assumption H-SEN holds, and that the marginals of $\Delta$ have finite variance.

The transmitter is assumed to pick a codeword to transmit uniformly at random from the codebook. The receiver sees the sum of the codeword and the noise vector, and, without knowing which codeword was picked, is required to determine it from the received noise-corrupted codeword. The optimum decision rule is maximum likelihood decoding, i.e. to choose as the decision for the transmitted codeword one of those for which the conditional probability of seeing the given observation is largest among all codewords. The probability of error of the codebook, $p_{\mathrm{e}}(\mathcal{T})$, is defined to be the average probability of error over all codewords, where the probability of error of a codeword is the probability of error of the maximum likelihood decision rule, conditioned on this codeword having been transmitted. Shannon [17], [18] proved that, asymptotically in the block length, there is a threshold on the rate such that, for rates below this threshold, it is possible to choose codebooks for which the probability of error goes asymptotically to 0, while, for rates above this threshold, this is not possible. This threshold is given by the *Shannon capacity*, defined by

$$C_P(\Delta) = \lim_{n\to\infty}\frac{1}{n} \sup_{T^n,\mathbb{E}(\sum_{i=1}^n (T_i^n)^2)<nP} I(T^n, T^n + D^n),$$

where the supremum is over all distribution functions for $T^n = (T_1^n,\dots,T_n^n)\in\mathbb{R}^n$ such that $\mathbb{E}(\sum_{i=1}^n (T_i^n)^2) < nP$. This limit is known to exist. Here, for jointly distributed vector-valued random variables $(X,Y)$, the expression $I(X;Y)$ denotes their *mutual information* [5], [8].

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

Note that the Shannon capacity is a characteristic of both the noise process $\Delta$ and the power constraint $P$. Let $\sigma^2$ denote the variance of $\Delta_0$. The relation between the Shannon capacity and the Poltyrev capacity is given by the following lemma, due to Shannon [17]. We give a proof, since it is illuminating.

**Lemma 2.** *Under the foregoing assumptions,*

$$\tfrac{1}{2}\ln(2\pi \mathrm{e} P) - h(\Delta) \le C_P(\Delta) \le \tfrac{1}{2}\ln(2\pi \mathrm{e}(P+\sigma^2)) - h(\Delta). \tag{20}$$

*Proof.* We have

$$I(T^n + \Delta^n; T^n) = h(T^n + \Delta^n) - h(T^n + \Delta^n \mid T^n) = h(T^n + \Delta^n) - h(\Delta^n).$$

It is well known that, for all stationary sequences $\{A_k\}$, we have

$$h(A_1, A_2, \ldots, A_n) \le \frac{n}{2}\ln(2\pi \mathrm{e}\,\mathrm{var}(A_1)).$$

Hence,

$$\frac{1}{n}I(T^n + \Delta^n; T^n) \le \frac{1}{2}\ln(2\pi \mathrm{e}(P+\sigma^2)) - \frac{1}{n}h(\Delta^n).$$

For the lower bound, the inequality $h(T^n + \Delta^n) \ge h(T^n)$ is used to deduce that

$$I(T^n + \Delta^n; T^n) = h(T^n + \Delta^n) - h(T^n + \Delta^n \mid T^n) \ge h(T^n) - h(\Delta^n).$$

Taking now $T^n$ Gaussian with i.i.d. $\mathcal{N}(0, P)$ coordinates, we obtain

$$C_P(\Delta) \ge \tfrac{1}{2}\ln(2\pi \mathrm{e} P) - h(\Delta).$$

In the power-constrained scenario, we define

$$\mathcal{E}(n, R, P, \Delta) = -\frac{1}{n}\ln p_{\mathrm{e,opt}}(n, R, P, \Delta)\,,$$

with $p_{\mathrm{e,opt}}(n, R, P, \Delta)$ the infimum of $p_{\mathrm{e}}(\mathcal{T})$ over all codes in $\mathbb{R}^n$ of rate at least $R \ge 0$ and all decoding rules, when the signal power is $P$ and the noise is $\Delta$. We then define

$$\bar{\mathcal{E}}(R, P, \Delta) = \limsup_n \mathcal{E}(n, R, P, \Delta) \quad \text{and} \quad \underline{\mathcal{E}}(R, P, \Delta) = \liminf_n \mathcal{E}(n, R, P, \Delta).$$

Assuming that these are identical, we denote this common limit by $\mathcal{E}(R, P, \Delta)$. For fixed $P$ and $\Delta$, the function $R \mapsto \mathcal{E}(R, P, \Delta)$, defined for rates less than the Shannon capacity, is known as the error exponent function or the reliability function in information theory.

The following result shows how to obtain lower bounds on the error exponent function for power-constrained additive noise channels from error exponents coming out of the point process formulation (such as the random coding exponent and the expurgated exponent developed in Sections 5 and 6, respectively).

The next theorem features a sequence of stationary point processes $\mu^n$ in $\mathbb{R}^n$ with normalized logarithmic intensities converging to the finite limit $-h(\Delta) - \ln(\alpha')$, where $\alpha > \alpha' > 1$. The following condition will be required on this collection: for all $\gamma > 0$ and all $P > 0$,

$$\ln(\mathbb{P}^n(\mu^n(B^n(0, \sqrt{nP})) \ge (2\pi \mathrm{e} P)^{n/2}\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha'+\gamma)})) = o(n). \tag{21}$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

This condition is satisfied, e.g. by homogeneous Poisson and Matérn point processes as both are such that, for all Borel sets $B$ of $\mathbb{R}^n$,

$$\mathbb{E}(\mu^n(B)^2) \le \mathbb{E}(\widehat{\mu}^n(B)^2), \tag{22}$$

where $\widehat{\mu}^n$ denotes the homogeneous Poisson point process with the same intensity as $\mu^n$. For Matérn point processes, (22) follows from the evaluation of the reduced second moment measure, which is classical. For all collections $\{\mu_n\}$ satisfying (22) for all $n$, we obtain (21) from Chebyshev's inequality.

**Theorem 7.** *Let $\Delta$ be a centered, real-valued, stationary, and ergodic stochastic process, and let $\alpha > 1$. Let $(\mu, \mathcal{C}) := (\mu^n, \mathcal{C}^n)$ be a sequence where, for each $n \ge 1$, $\mu^n$ is a stationary and ergodic point process in $\mathbb{R}^n$ with normalized logarithmic intensity $-h(\Delta) - \ln(\alpha_n)$, with $\alpha_n \to \alpha'$ as $n \to \infty$, where $\alpha > \alpha' > 1$, and the sequence $\{\mu^n\}$ satisfies (21), and where, for each $n \ge 1$, the tessellation $\mathcal{C}^n$ is jointly stationary with $\mu^n$. Then, for all $P > 0$ such that $\frac{1}{2}\ln(2\pi \mathrm{e} P) > h(\Delta) + \ln(\alpha)$, we have*

$$\underline{\mathcal{E}}\big(\tfrac{1}{2}\ln(2\pi \mathrm{e} P) - h(\Delta) - \ln(\alpha), P, \Delta\big) \ge \underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta) \tag{23}$$

*and*

$$\underline{\mathcal{E}}\bigg(C_P(\Delta) - \ln(\alpha) - \frac{1}{2}\ln\bigg(1 + \frac{\sigma^2}{P}\bigg), P, \Delta\bigg) \ge \underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta). \tag{24}$$

*Here $\underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta)$ is the error exponent without restriction for the family $(\mu, \mathcal{C})$, as defined in (11). In addition,*

$$\liminf_{P\to\infty} \underline{\mathcal{E}}(C_P(\Delta) - \ln(\alpha), P, \Delta) \ge \underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta). \tag{25}$$

*Proof.* From the very definition of Palm probabilities, for all $n$,

$$p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha_n, \Delta) = \frac{\mathbb{E}^n\big(\sum_{k\colon T_k^n \in B^n(0,\sqrt{nP})} p_{\mathrm{e,k}}\big)}{\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha_n)} V_B^n(\sqrt{nP})},$$

where $p_{\mathrm{e,k}}$ denotes the probability that $T_k^n + D_k^n$ does not belong to $\mathcal{C}_k^n$ given $\{T_l^n, \mathcal{C}_l^n\}_l$. Hence, for all $\gamma > 0$, we can write

$$\begin{aligned}
&p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha_n, \Delta)\\
&\ge \frac{\mathbb{E}^n \sum_{\{k\colon T_k^n \in B^n(0,\sqrt{nP})\}} p_{\mathrm{e,k}} \mathbf{1}_{\{\mu^n(B^n(0,\sqrt{nP}))\ge(2\pi \mathrm{e} P)^{n/2}\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha'+\gamma)}\}}}{\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha_n)} V_B^n(\sqrt{nP})}\\
&\ge \mathbb{P}^n(\mu^n(B^n(0,\sqrt{nP})) \ge (2\pi eP)^{n/2}\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha'+\gamma)})\\
&\quad\times p_{\mathrm{e,opt}}\bigg(n, \frac{1}{2}\ln(2\pi eP) - h(\Delta) - \ln(\alpha'+\gamma), P, \Delta\bigg)\mathrm{e}^{-n\ln(\alpha'+\gamma)}\mathrm{e}^{n\ln(\alpha_n)}\frac{(2\pi \mathrm{e} P)^{n/2}}{V_B^n(\sqrt{nP})},
\end{aligned}$$

where we have used the fact that $p_{\mathrm{e,opt}}(n, R, P, \Delta)$ is nondecreasing in $R$ and $\mathrm{e}^{nR}$ is non-

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

decreasing in $R$. If $\gamma > 0$ is sufficiently small, we can then write:

$$
\begin{aligned}
&-\frac{1}{n}\ln(p_{\mathrm{e}}^{\mathrm{pp}}(n, \mu^n, \mathcal{C}^n, \alpha_n, \Delta)) \\
&\quad \le -\frac{1}{n}\ln\big(p_{\mathrm{e,opt}}\big(n, \tfrac{1}{2}\ln(2\pi \mathrm{e} P) - h(\Delta) - \ln\alpha, P, \Delta\big)\big) \\
&\qquad - \frac{1}{n}\ln(\mathbb{P}^n(\mu^n(B^n(0, \sqrt{nP})) \ge (2\pi\mathrm{e}P)^{n/2}\mathrm{e}^{-nh(\Delta)}\mathrm{e}^{-n\ln(\alpha'+\gamma)})) \\
&\qquad - \ln(\alpha_n) + \ln(\alpha' + \gamma) - \frac{1}{n}\ln\bigg(\frac{(2\pi\mathrm{e}P)^{n/2}}{V_B^n(\sqrt{nP})}\bigg).
\end{aligned}
$$

When taking the limit in $n$, the second term on the right-hand side tends to 0 (from (21)), and the last term of the right-hand side tends to 0 as well (from classical asymptotics on the volume of the $d$-ball). Hence, first taking the limit as $n \to \infty$ and then letting $\gamma \to 0$, (23) follows.

We obtain (24) from (23) when using the second inequality of (20) and the fact that the function $x \to \underline{\mathcal{E}}(x, P, \Delta)$ is nonincreasing.

To prove (25), pick $\tilde{\alpha}$ such that $\alpha > \tilde{\alpha} > \alpha' > 1$. It suffices to observe that from the preceding proof we have

$$
\underline{\mathcal{E}}\bigg(C_P(\Delta) - \ln(\tilde{\alpha}) - \frac{1}{2}\ln\bigg(1 + \frac{\sigma^2}{P}\bigg), P, \Delta\bigg) \ge \underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta).
$$

The preceding theorem can, in particular, be used with the family $(\mu, \mathcal{C})$ taken to be either (Poi, $\mathcal{L}(\Delta)$) or (Mat, $\mathcal{L}(\Delta)$), for which $\underline{\pi}(\mu, \mathcal{C}, \alpha', \Delta)$ has been studied in detail in this paper. An excellent survey of the known upper and lower bounds for the error exponent function in the power-constrained AWGN case is given in [3].

## 8. Examples

This section contains several examples of noise processes $\Delta$ of interest in applications and calculation of the concrete instantiation of the preceding results in these cases. Consider first the additive white noise (WN) case, i.e. when $\Delta = \{\Delta_l\}$ is an i.i.d. sequence, focusing on the special cases of white symmetric exponential noise and white uniform noise. Additive colored Gaussian noise (CGN) is then discussed, where $\{\Delta_l\}$ is a Gaussian sequence which is not necessarily white, and finally we discuss in detail the AWGN case, which is the case of most interest in applications. Connections to the work in [16] in the AWGN case are made. A random coding exponent is calculated in all examples, and an expurgated exponent is calculated where it was possible to give a relatively clean looking result.

### 8.1. White noise

The WN case is that where the displacement vector $\Delta$ has i.i.d. coordinates. Let $D$ be a typical coordinate random variable. The differential entropy rate of $\Delta$ is then

$$
h(\Delta) = h(D) = -\int_{\mathbb{R}} f(x)\ln(f(x))\,\mathrm{d}x,
$$

where $f(x)$ denotes the density of $D$.

From Cramér's theorem [7], [19] we have

$$
I(x) = \sup_{\theta}(\theta x - \ln(\mathbb{E}(f(D)^{-\theta}))),
$$

with $D$ a random variable with density $f$.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

Note that the rate function $I(\cdot)$ is not necessarily a good rate function. A sufficient condition is that 0 is in the interior of the set $\{\theta : \mathbb{E}((f(D))^{-\theta}) < \infty\}$ (see [7, Lemma 2.2.20]).

8.1.1. *White symmetric exponential noise.* The differential entropy of the symmetric exponential distribution of variance $\sigma^2$ is $h(D) = \ln(\sqrt{2}e\sigma)$ and

$$\mathbb{E}(f(D)^{-\theta}) = (\sqrt{2}\sigma)^{\theta}\mathbb{E}\left(\exp\left(\theta\frac{|D|\sqrt{2}}{\sigma}\right)\right) = (\sqrt{2}\sigma)^{\theta}\frac{1}{1-\theta}, \qquad \theta < 1.$$

So

$$I(u) = \sup_{\theta}(\theta u - \theta\ln(\sqrt{2}\sigma) + \ln(1-\theta)),$$

that is,

$$I(u) = \begin{cases} +\infty & \text{for } u \le \ln(\sqrt{2}\sigma), \\ u - h(D) - \ln(u - \ln(\sqrt{2}\sigma)) & \text{otherwise,} \end{cases} \tag{26}$$

which is a good and convex rate function.

From Lemma 1 we obtain

$$J(u) = \begin{cases} -\infty & \text{for } u \le \ln(\sqrt{2}\sigma), \\ \ln(\sqrt{2}e\sigma(u - \ln(\sqrt{2}\sigma))) & \text{otherwise.} \end{cases} \tag{27}$$

It follows from (26) for $I$ and (27) for $J$ that in this case the function to minimize in (15) is

$$v - 1 - \ln(v) + (\ln(\alpha) - \ln(v))^{+}$$

for $v > 0$. So in this case the random coding exponent is the right-hand side of the inequality

$$\underline{\pi}(\text{Poi}, \mathcal{L}(\Delta), \alpha, \Delta) \ge \begin{cases} \alpha - 1 - \ln\alpha & \text{if } 1 \le \alpha < 2, \\ 1 - 2\ln 2 + \ln\alpha & \text{if } \alpha \ge 2. \end{cases}$$

Consider the Matérn-$\Delta$-$\xi$ point process, where $\Delta$ is white symmetric exponential noise and where the exclusion regions are $L_1$ balls of radius

$$r^n(\xi) = \frac{n\sigma}{\sqrt{2}}(\xi - \ln(\sqrt{2}\sigma))$$

for $\xi > \ln(\sqrt{2}\sigma)$. For the target normalized logarithmic intensity $-h(\Delta) - \alpha$, we build the Matérn point process $\widetilde{\mu}^n$ from a Poisson point process $\mu^n$ of intensity $\lambda_n = e^{nR}$ with $R = -\ln(\sqrt{2}e\sigma\alpha)$, where $\alpha > 1$. The parameter $\xi$ is chosen as

$$\xi = \alpha - \varepsilon + \ln(\sqrt{2}\sigma),$$

so that the $L_1$ exclusion radius is $r_n = n\sigma(\alpha - \varepsilon)/\sqrt{2}$. The intensity of the associated Matérn point process is then $\widetilde{\lambda}_n = \lambda_n \exp(-\lambda_n V^n_{B,1}(r_n))$, with

$$V^n_{B,1}(r_n) = \frac{(2r_n)^n}{n!} = \frac{(\sqrt{2}\sigma(\alpha - \varepsilon))^n n^n}{n!}$$

the volume of the $L_1$ ball of radius $r_n$. It is easy to see that $\widetilde{\lambda}_n \le \lambda_n$ for all $n$ and that $\lim_{n\to\infty}\widetilde{\lambda}_n/\lambda_n = 1$.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

It follows from (17) that

$$p_{\mathrm{e}}(n) \le \int_{r>0} \min\Big(1, \lambda_n \sup_{\{x^n : |x^n|_1 = r\}} \mathrm{vol}(B_1^n(0, r_n)^c \cap B_1^n(x^n, r))\Big) g_\sigma^n(r)\, \mathrm{d}r, \tag{28}$$

where $|\cdot|_1$ denotes the $L_1$ norm, $B_1^n(x, r)$ the $L_1$ ball of center $x$ and radius $r$, and $g_\sigma^n(r)$ here denotes the density of the $L_1$ norm of $D^n$, given by

$$g_\sigma^n(r) = \mathrm{e}^{-\sqrt{2}r/\sigma}\left(\frac{\sqrt{2}}{\sigma}\right)^n \frac{r^{n-1}}{\Gamma(n)}, \qquad r \ge 0.$$

Making the substitution $v = (\sqrt{2}r)/n\sigma$, the right-hand side of (28) is

$$\int_{v>0} \min\Big(1, \lambda_n \sup_{\{x^n : |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v)\Big) \mathrm{e}^{-vn} \frac{(vn)^n}{v\Gamma(n)}\, \mathrm{d}v, \tag{29}$$

with

$$W(x^n, v) = \mathrm{vol}\left(B_1^n\left(0, \frac{n\sigma(\alpha-\varepsilon)}{\sqrt{2}}\right)^c \cap B_1^n\left(x^n, \frac{vn\sigma}{\sqrt{2}}\right)\right). \tag{30}$$

Let $\widetilde{\alpha} = \alpha - \varepsilon$. If $v \le \widetilde{\alpha}/2$ then $W(x^n, v) = 0$ for all $x^n$ with $|x^n|_1 = v\sigma n/\sqrt{2}$. It is proved below that if $v > \widetilde{\alpha}/2$ then

$$\lim_{n\to\infty} \frac{1}{n} \ln\Big(\sup_{\{x^n : |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v)\Big) = \ln(\sqrt{2}v\mathrm{e}\sigma). \tag{31}$$

We have

$$\sup_{\{x^n : |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v) \ge W(x^n(v), v)$$

with $x^n(v) = (v\sigma n/\sqrt{2}, 0, \dots, 0)$. The region $R(n)$ on the right-hand side of (30) includes the region

$$\left\{y^n = (y_1, \dots, y_n) \in \mathbb{R}^n : y_1 > \frac{v\sigma n}{\sqrt{2}}, \frac{(\alpha-\varepsilon)\sigma n}{\sqrt{2}} < y_1 + \sum_{i=2}^n |y_i| < \frac{2v\sigma n}{\sqrt{2}}\right\},$$

which is comprised of $2^{n-1}$ copies (one for each configuration of signs of the variables $y_2, \dots, y_n$; see Figure 1) of the following basic region:

$$\left\{y^n = (y_1, \dots, y_n) \in \mathbb{R}_+^n : y_1 > \frac{v\sigma n}{\sqrt{2}}, \frac{(\alpha-\varepsilon)\sigma n}{\sqrt{2}} < y_1 + \sum_{i=2}^n y_i < \frac{2v\sigma n}{\sqrt{2}}\right\}.$$

In Figure 1, the origin of the plane is the tagged codeword. The large ball centered at 0 and passing through point $A$ is that with radius $n\sigma(\alpha-\varepsilon)/\sqrt{2}$. Point $V$ is that with coordinate $x^n(v) = (v\sigma n/\sqrt{2}, 0, \dots, 0)$. The region $R(n)$ is depicted by the union of the dashed region and the grey region. The volume $V(n)$ is that of the grey region.

The volume $V(n)$ of this basic region is the same as that of

$$\left\{y^n = (y_1, \dots, y_n) \in \mathbb{R}_+^n : \frac{(\alpha-\varepsilon-v)\sigma n}{\sqrt{2}} < \sum_{i=1}^n y_i < \frac{v\sigma n}{\sqrt{2}}\right\},$$

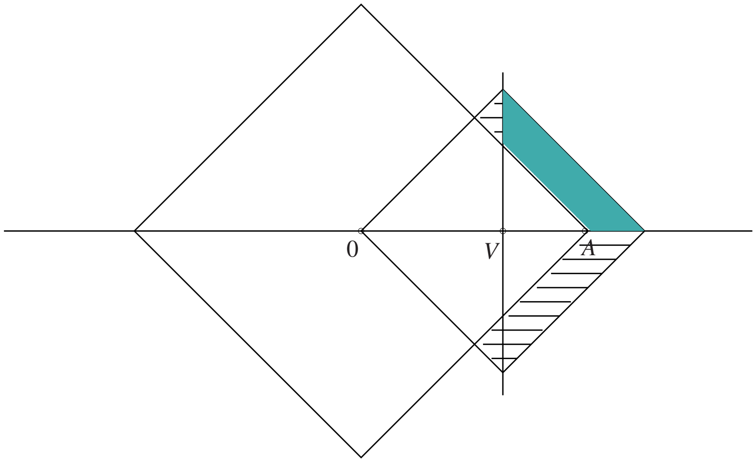


FIGURE 1: The Matérn case with white symmetric exponential noise.

namely, $2^{-n}$ times the volume of the $L_1$ ball of center 0 and radius $v\sigma n/\sqrt{2}$ deprived of the $L_1$ ball of center 0 and radius $(\alpha - \varepsilon - v)^+\sigma n/\sqrt{2}$, that is,

$$V(n) = 2^{-n}\bigg((\sqrt{2}v\sigma)^n \frac{n^n}{n!} - (\sqrt{2}(\alpha - \varepsilon - v)^+\sigma)^n \frac{n^n}{n!}\bigg).$$

Hence,

$$\frac{1}{n}\ln\Big(\sup_{\{x^n :\, |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v)\Big) \ge \frac{1}{n}\ln(2^{n-1}V(n)) \to \ln(\sqrt{2}v\mathrm{e}\sigma) \quad \text{as } n \to \infty.$$

But, from (30),

$$\frac{1}{n}\ln\Big(\sup_{\{x^n :\, |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v)\Big) \le \frac{1}{n}\ln\bigg(\mathrm{vol}B_1^n\bigg(0, \frac{v\sigma n}{\sqrt{2}}\bigg)\bigg) \to \ln(\sqrt{2}v\mathrm{e}\sigma) \quad \text{as } n \to \infty.$$

This completes the proof of (31).

The error exponent associated with this sequence of Matérn point processes thus satisfies the bound

$$\underline{\pi}(\mathrm{Mat}, \mathcal{L}(\Delta), \alpha, \Delta) \ge \inf_{v>0} b(v) + a(v),$$

with $a(v) = v - \ln(v) - 1$ (stemming from $\mathrm{e}^{-vn}(vn)^n/v\Gamma(n)$), and

$$b(v) = \begin{cases} \infty & \text{if } 0 < v < \widetilde{\alpha}/2, \\ (\ln\widetilde{\alpha} - \ln v)^+ & \text{if } \widetilde{\alpha}/2 < v \end{cases}$$

(stemming from $\min(1, \lambda_n \sup_{\{x^n :\, |x^n|_1 = v\sigma n/\sqrt{2}\}} W(x^n, v))$ in (29)). For more details on this derivation, see the long version of this paper [2] and, in particular, the analytical arguments for the AWGN case. This leads to the following expurgated exponent for symmetric exponential white noise:

$$\underline{\pi}(\mathrm{Mat}, \mathcal{L}(\Delta), \alpha, \Delta) \ge \begin{cases} \alpha - \ln(\alpha) - 1 & \text{for } \alpha \le 2, \\ \ln(\alpha) + 1 - 2\ln(2) & \text{for } 2 \le \alpha \le 4, \\ \dfrac{\alpha}{2} - \ln(\alpha) - 1 + 2\ln(2) & \text{for } \alpha \ge 4. \end{cases}$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

8.1.2. *White uniform noise.* Let $D$ be uniform on $[-\sqrt{3}\sigma, +\sqrt{3}\sigma]$, which is centered and with variance $\sigma^2$. The differential entropy is $h(D) = \ln(2\sqrt{3}\sigma)$ and

$$\mathbb{E}(f(D)^{-\theta}) = (2\sqrt{3}\sigma)^{\theta},$$

so that $G(\theta) = \theta \ln(2\sqrt{3}\sigma)$ and

$$I(u) = \begin{cases} \infty & \text{if } u \neq \ln(2\sqrt{3}\sigma), \\ 0 & \text{if } u = \ln(2\sqrt{3}\sigma), \end{cases} \tag{32}$$

which is a good and convex rate function.

From Lemma 1,

$$J(u) = \begin{cases} -\infty & \text{for } u < \ln(2\sqrt{3}\sigma), \\ \ln(2\sqrt{3}\sigma) & \text{for } u \geq \ln(2\sqrt{3}\sigma). \end{cases} \tag{33}$$

It follows from (33), (32), and (15) that

$$\underline{\pi}(\text{Poi}, \mathcal{L}(\Delta), \alpha, \Delta) \geq F(\ln(2\sqrt{3}\sigma)) = \ln(\alpha).$$

The right-hand side of the preceding equation is the random coding exponent for white uniform noise.

### 8.2. Colored Gaussian noise

The CGN case is that where $\{\Delta_k\}$ is a stationary and ergodic Gaussian process with spectral density function $g(\beta)$, i.e.

$$\mathbb{E}(\Delta_0 \Delta_k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\beta} g(\beta)\, \mathrm{d}\beta$$

for all $k$. It is well known (see, e.g. [10]) that the differential entropy rate of such a stationary process exists and is given by

$$h(\Delta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln(2\mathrm{e}\pi g(\beta))\, \mathrm{d}\beta. \tag{34}$$

The conditions for the validity of the Gärtner-Ellis theorem hold with

$$G(\theta) = \frac{\theta}{2} \ln(2\pi) - \frac{1}{2} \ln(1-\theta) + \frac{\theta}{2} \bigg( \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(g(\beta))\, \mathrm{d}\beta \bigg),$$

when $\theta < 1$ and $G(\theta) = \infty$ for $\theta > 1$. This yields

$$I(u) = \begin{cases} \infty & \text{if } u \leq (1/4\pi) \int_{-\pi}^{\pi} \ln(2\pi g(\beta))\, \mathrm{d}\beta, \\ u - h(\Delta) & \\ \quad -\dfrac{1}{2} \ln\bigg(2u - \dfrac{1}{2\pi} \displaystyle\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\, \mathrm{d}\beta\bigg) & \text{otherwise}, \end{cases} \tag{35}$$

with $h(\Delta)$ as in (34). This is a good, convex, and continuous rate function.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

From (35) and (13) we obtain

$$J(u) = \begin{cases} -\infty & \text{if } u \le (1/4\pi)\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\,\mathrm{d}\beta, \\ \dfrac{1}{4\pi}\displaystyle\int_{-\pi}^{\pi} \ln(2\pi e g(\beta))\,\mathrm{d}\beta \\ \quad +\dfrac{1}{2}\ln\bigg(2u - \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\,\mathrm{d}\beta\bigg) & \text{otherwise.} \end{cases} \tag{36}$$

This function is continuous.

Theorem 4, (35), and (36) yield

$$\begin{aligned} \underline{\pi}(\mathrm{Poi}, \mathcal{L}(\Delta), \alpha, \Delta) \ge \inf_u \bigg\{ & \bigg(\ln(\alpha) - \frac{1}{2}\ln\bigg(2u - \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\,\mathrm{d}\beta\bigg)\bigg)^+ \\ & + u - \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln(2\pi e g(\beta))\,\mathrm{d}\beta \\ & - \frac{1}{2}\ln\bigg(2u - \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\,\mathrm{d}\beta\bigg)\bigg\}. \end{aligned}$$

Making the substitution

$$v = \sqrt{2u - \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln(2\pi g(\beta))\,\mathrm{d}\beta},$$

we find that the last infimum is

$$\inf_{v \ge 0} \bigg\{ (\ln(\alpha) - \ln(v))^+ + \frac{v^2}{2} - \frac{1}{2} - \ln(v) \bigg\},$$

and, hence, we obtain the same function to optimize as in the AWGN case. So the random coding exponent is that of (39).

### 8.3. White Gaussian noise

The AWGN case is a special case of WN where $f$ is Gaussian with mean 0 and variance $\sigma^2$. In this case, the differential entropy of $f$ is $h(D) = \frac{1}{2}\ln(2\pi e\sigma^2)$, and we have

$$I(u) = \begin{cases} +\infty & \text{for } u \le \frac{1}{2}\ln(2\pi\sigma^2), \\ u - \frac{1}{2}\ln(2e\pi\sigma^2) - \frac{1}{2}\ln(2u - \ln(2\pi\sigma^2)) & \text{otherwise,} \end{cases} \tag{37}$$

which is a good and convex rate function.

It immediately follows from Lemma 1 that

$$J(u) = \begin{cases} -\infty & \text{for } u \le \frac{1}{2}\ln(2\pi\sigma^2), \\ \frac{1}{2}\ln(2\pi e\sigma^2) + \frac{1}{2}\ln(2u - \ln(2\pi\sigma^2)) & \text{otherwise.} \end{cases} \tag{38}$$

We therefore recover the following result, first obtained by Poltyrev for the AWGN case in [16] and revisited in [1]:

$$\underline{\pi}(\mathrm{Poi}, \mathcal{L}(\mathrm{AWGN}), \alpha, \mathrm{AWGN}) \ge \begin{cases} \frac{1}{2}\alpha^2 - \frac{1}{2} - \ln\alpha & \text{if } 1 \le \alpha < \sqrt{2}, \\ \frac{1}{2} - \ln 2 + \ln\alpha & \text{if } \sqrt{2} \le \alpha < \infty. \end{cases} \tag{39}$$

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

This follows from (15), as we will now show. (In fact, we can show that this lower bound is tight; see [16].) Using (37) for $I$ and (38) for $J$ in (15) and using the substitution $v = \sqrt{2(u-(1/2)\ln(2\pi\sigma^2))}$, we obtain the following equivalent optimization problem:

$$\text{Minimize } a(v)+b(v) \text{ over } v\ge 0, \text{ with } a(v)=\frac{v^2}{2}-\frac{1}{2}-\ln(v) \text{ and } b(v)=(\ln\alpha-\ln v)^{+}.$$

This is precisely the optimization problem analyzed in [1]. This gives $(\frac{1}{2}\alpha^2)-\frac{1}{2}-\ln\alpha$ when $1<\alpha<\sqrt{2}$ and $\frac{1}{2}-\ln 2+\ln\alpha$ when $\alpha>\sqrt{2}$.

The next discussion is centered on the expurgated exponent based on the Matérn I process. Fix $\varepsilon>0$. Consider a sequence of Matérn I processes $\widetilde{\mu}^n$. The point process $\widetilde{\mu}^n$ is built from a Poisson process $\mu_n$ of rate $\lambda_n=\mathrm{e}^{nR}$, where $R=\frac{1}{2}\ln(1/2\pi\mathrm{e}\alpha^2\sigma^2)$ for $\alpha>1$, and has exclusion radius $(\alpha-\varepsilon)\sigma\sqrt{n}$. The intensity of this Matérn I point process is

$$\widetilde{\lambda}_n=\lambda_n\mathrm{e}^{-\lambda_n V_B^n((\alpha-\varepsilon)\sigma\sqrt{n})},$$

and it is easy to see that $\widetilde{\lambda}_n/\lambda_n\to 1$ as $n\to\infty$, with $\widetilde{\lambda}_n<\lambda_n$ for all $n$.

Let $\underline{\pi}(\text{Mat}, \mathcal{L}(\text{AWGN}), \alpha, \text{AWGN})$ denote the error exponent (11) associated with this family of Matérn point processes. We prove below that

$$\underline{\pi}(\text{Mat}, \mathcal{L}(\text{AWGN}), \alpha, \text{AWGN}) \ge \frac{\alpha^2}{8} \quad \text{for all } \alpha\ge 2. \tag{40}$$

Take an exclusion radius of $(\alpha-\varepsilon)\sigma\sqrt{n}$. From (19),

$$p_{\mathrm{e}}(n)\le\int_{v\in\mathbb{R}^+}\min(1,\lambda_n\mathrm{vol}(B^n(0,(\alpha-\varepsilon)\sigma\sqrt{n}))^c\cap B^n(y^n(v),(v\sigma\sqrt{n})))\\ \times g_1^n(v\sqrt{n})\sqrt{n}\,\mathrm{d}v,$$

with $y^n(v)=(v\sigma\sqrt{n},0,\dots,0)$. We prove below that

$$\mathrm{vol}(B^n(0,(\alpha-\varepsilon)\sigma\sqrt{n}))^c\cap B^n(y^n(v),(v\sigma\sqrt{n}))\le V_B^n(c(v)\sigma\sqrt{n}), \tag{41}$$

with

$$c(v)=\begin{cases}0 & \text{if } 0<v<\widetilde{\alpha}/2,\\ \sqrt{\left(v^2-\left(v-\dfrac{\widetilde{\alpha}^2}{2v}\right)^2\right)} & \text{if } \widetilde{\alpha}/2<v<\widetilde{\alpha}/\sqrt{2},\\ v & \text{if } \widetilde{\alpha}/\sqrt{2}<v.\end{cases} \tag{42}$$

with $\widetilde{\alpha}=\alpha-\varepsilon$. If $v<\widetilde{\alpha}/2$ then

$$B^n(y^n(v),v\sigma\sqrt{n})\subset B^n(0,\widetilde{\alpha}\sigma\sqrt{n})),$$

so $c(v)=0$ in (41).

For $\widetilde{\alpha}/2<v<\widetilde{\alpha}/\sqrt{2}$, we have to find an upper bound on the volume of the portion of the ball of radius $v\sigma\sqrt{n}$ around the point at distance $v\sigma\sqrt{n}$ from the origin (along some ray) that is outside the ball $B^n(0,\alpha\sigma\sqrt{n})$ (this is depicted by the shaded area in Figure 2). A first upper bound on this volume is the portion of the former ball cut off by the hyperplane perpendicular to the ray and at a distance $\mathrm{d}\sigma\sqrt{n}$ from it (i.e. a distance of $(v+d)\sigma\sqrt{n}$ along this ray from the origin), where $d=\alpha^2/2v-v$ by elementary geometry. The latter portion is in turn included

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

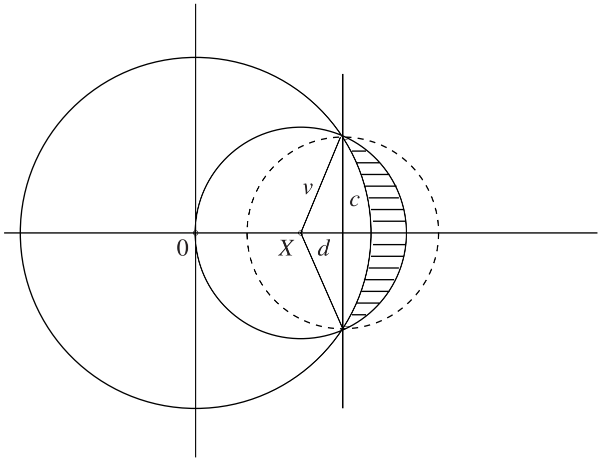


FIGURE 2: The Matérn case with white Gaussian noise.

in a ball of radius $\sigma\sqrt{n}\sqrt{v^2-d^2}$ (which is depicted by the dashed circle in Figure 2). In this figure, the large ball centered on the origin is the exclusion ball of the Matérn construction around the tagged codeword. Its radius is $(\alpha-\varepsilon)\sigma\sqrt{n}$. The point $X$ is the location of the noise added to the tagged codeword. Its norm is $v\sigma\sqrt{n}$. The ball centered on $X$ with radius $v\sigma\sqrt{n}$ is the vulnerability region in the Poisson case. In the Matérn case, the vulnerability region is the shaded lune depicted in the figure. This is the case with $\alpha/2 < v < \alpha/\sqrt{2}$. The area of this lune is upper bounded by that of the ball of radius $c=\sqrt{n(v^2-d^2)}\sigma$, with $d$ as above. This ball is represented by the dashed line disc. Hence, $c(v)=\sqrt{v^2-d^2}$. This completes the proof of (41) and (42).

By the same arguments as in the Poisson case (see [2, Section 10.3]),

$$\underline{\pi}(\text{Mat}, \mathcal{L}(\text{AWGN}), \alpha, \text{AWGN}) \geq \inf_{v>0} b(v) + a(v),$$

with $a(v) = \frac{1}{2}v^2 - \frac{1}{2} - \ln v$ and

$$b(v) = \begin{cases} \infty & \text{if } 0 < v < \frac{1}{2}\widetilde{\alpha}, \\ \ln\alpha - \dfrac{1}{2}\ln\left(v^2 - \left(v - \dfrac{\widetilde{\alpha}^2}{2v}\right)^2\right) & \text{if } \frac{1}{2}\widetilde{\alpha} < v < \widetilde{\alpha}/\sqrt{2}, \\ (\ln\alpha - \ln v)^+ & \text{if } \widetilde{\alpha}/\sqrt{2} < v. \end{cases}$$

Bound (40) follows when minimizing over $v$ for each $\widetilde{\alpha} \geq 2$ and then letting $\varepsilon$ tend to 0.

The lower bound on $\underline{\eta}(\alpha)$ given in (39) and (40), namely,

$$\pi(\alpha) \geq \begin{cases} \frac{1}{2}\alpha^2 - \frac{1}{2} - \ln\alpha & \text{if } 1 \leq \alpha < \sqrt{2}, \\ \frac{1}{2} - \ln 2 + \ln\alpha & \text{if } \sqrt{2} \leq \alpha < 2, \\ \frac{1}{8}\alpha^2 & \text{if } \alpha \geq 2, \end{cases}$$

was first obtained by Poltyrev [16] (see Equations (32) and (36) therein).

## 9. Mismatched decoding

A scenario of interest in applications is that of *mismatched decoding*, where the decoder has been designed for some noise $\Delta$ but the actual noise is in fact $\widetilde{\Delta}$. In the next theorem, $\Delta$ and $\widetilde{\Delta}$ are real-valued, centered, stationary, and ergodic stochastic processes.

By the same arguments as in the matched case, we obtain the following result.

**Theorem 8.** *For all stationary and ergodic point processes $\mu^n$, all $\Delta$, and actual displacement vectors governed by $\{\widetilde{\Delta}_k\}_k$ (independent from point to point), the probability of error under ML decoding, assuming that the law of the displacements is governed by the law of $\Delta$, satisfies*

$$p_{\mathrm{e}}(n) \le \int_{x^n \in \mathbb{R}^n} \mathbb{P}_0^n((\mu^n - \varepsilon_0)(\mathcal{F}(x^n)) > 0) \widetilde{f}^n(x^n)\, \mathrm{d}x^n.$$

*If $\mu^n$ is a Poisson process of intensity $\lambda_n$ then*

$$p_{\mathrm{e}}(n) \le \int_{u \in \mathbb{R}} (1 - \exp(-\lambda_n W_\Delta^n(u))) \rho_{\widetilde{\Delta}}^n(\mathrm{d}u),$$

*where $\rho_{\widetilde{\Delta}}^n(\mathrm{d}u)$ is the law of the random variable $-(1/n)\ln(\widetilde{f}^n(\widetilde{D}^n))$ on $\mathbb{R}$ and $W_\Delta^n$ is the log-likelihood level volume for $\Delta$.*

The random coding exponent for mismatched decoding is given as follows.

**Theorem 9.** *Assume that $\mu^n$ is a Poisson process with normalized logarithmic intensity $-h(\Delta) - \ln(\alpha)$ with $\alpha > 1$ and that the decoder uses ML decoding under the assumption that the law of the displacement vectors is that of $\Delta$, while the actual displacement vectors are governed by $\{\widetilde{\Delta}_k\}$ (independent from point to point). Suppose that Assumption H-SEN holds for both $\Delta$ and $\widetilde{\Delta}$. Then the associated error exponent is bounded from below by*

$$\inf_u \{F(u) + \widetilde{I}(u)\},$$

*where $\widetilde{I}(u)$ is the rate function of $\rho_{\widetilde{\Delta}}^n$ and*

$$F(u) = (\ln(\alpha) + h(\Delta) - J(u))^+,$$

*where $J(u) = \sup_{s \le u}(s - I(s))$ is the volume exponent for $\Delta$.*

## Acknowledgements

The work of the first author was supported by NSF grants CCF-0500234, CNS-0627161, and CCF-0635372, by Marvell Semiconductor, and by the University of California MICRO program during the initial stages of this project. He is currently supported by the ARO MURI grant W911NF-08-1-0233, 'Tools for the Analysis and Design of Complex Multi-Scale Networks', by the NSF grant CNS-0910702, and by the NSF Science & Technology Center grant CCF-0939370, 'Science of Information'.

This work started when the second author was a Visiting Miller Professor at UC Berkeley.

The work of both authors was supported by the IT-SG-WN Associate Team of the Inria@SiliconValley programme. Both authors would also like to thank the Isaac Newton Institute for providing a wonderful environment during Spring 2010 as part of the program "Stochastic Processes in Communications Sciences", where much of the final part of this work was done.

https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press

## References


[1] Anantharam, V. and Baccelli, F. (2008). A Palm theory approach to error exponents. In *Proc. IEEE Internat. Symp. Inf. Theory*, IEEE, pp. 1768–1772.

[2] Anantharam, V. and Baccelli, F. (2010). Information-theoretic capacity and error exponents of stationary point processes under random additive displacements. Preprint. Available at http://uk.arxiv.org/abs/1012.4924v1.

[3] Ashikhmin, A. E., Barg, A. and Litsyn, S. N. (2000). A new upper bound on the reliability function of the Gaussian channel. *IEEE Trans. Inf. Theory* **46,** 1945–1961.

[4] Barron, A. (1985). The strong Ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem, *Ann. Prob.*, **13,** 1292–1303.

[5] Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley, New York.

[6] Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.

[7] Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston, MA.

[8] El Gamal, A. and Kim, Y.-H. (2011). *Network Information Theory*. Cambridge University Press.

[9] Gallager, R. G. (1968). *Information Theory and Reliable Communication*. John Wiley, New York.

[10] Gray, R. M. (2006). *Toeplitz and Circulant Matrices: A Review*. NOW Publishers, Delft.

[11] Han, T. S. (2003). *Information-Spectrum Methods in Information Theory*. Springer, Berlin.

[12] Kallenberg, O. (1983). *Random Measures*, 3rd edn. Akademie-Verlag, Berlin.

[13] Kieffer, J. C. (1974). A simple proof of the Moy–Perez generalization of the Shannon–McMillan theorem. *Pacific J. Math.* **51,** 203–206.

[14] Matérn, B. (1960). *Spatial Variation*, 2nd edn. Springer, Berlin.

[15] Møller, J. (1994). *Lectures on Random Voronoĭ Tessellations* (Lecture Notes Statist. **87**). Springer, New York.

[16] Poltyrev, G. (1994). On coding without restrictions for the AWGN channel. *IEEE Trans. Inf. Theory* **40,** 409–417.

[17] Shannon, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.* **27,** 379–423, 623–656.

[18] Shannon, C. E. (1959). Probability of error for optimal codes in a Gaussian channel. *Bell Sys. Tech. J.* **38,** 611–656.

[19] Varadhan, S. R. S. (1984). *Large Deviations and Applications*. SIAM, Philadelphia, PA.


https://doi.org/10.1239/aap/1427814578 Published online by Cambridge University Press