On Lattice Sequential Decoding for The Unconstrained AWGN Channel

Walid Abediseid, *Member, IEEE* and Mohamed-Slim Alouini, *Fellow, IEEE,*

Abstract

In this paper, the performance limits and the computational complexity of the *lattice sequential decoder* are analyzed for the unconstrained additive white Gaussian noise channel. The performance analysis available in the literature for such a channel has been studied only under the use of the minimum Euclidean distance decoder that is commonly referred to as the *lattice decoder*. Lattice decoders based on solutions to the NP-hard closest vector problem are very complex to implement, and the search for low complexity receivers for the detection of lattice codes is considered a challenging problem. However, the low computational complexity advantage that sequential decoding promises, makes it an alternative solution to the lattice decoder. In this work, we characterize the performance and complexity tradeoff via the error exponent and the decoding complexity, respectively, of such a decoder as a function of the decoding parameter — *the bias term*. For the above channel, we derive the cut-off volume-to-noise ratio that is required to achieve a good error performance with low decoding complexity.

Index Terms

Lattice Coding, Lattice Decoding, Sequential Decoding, Error Exponent, Decoding Complexity.

I. INTRODUCTION

THE theory of *lattices* — a mathematical approach for representing infinite discrete points in Euclidean space [1], has become a powerful tool to analyze many point-to-point and multi-terminal digital and wireless communication systems, particularly, communication systems that can be well described by the *linear Gaussian vector channel* model. This is mainly due to the three facts about channel codes constructed using lattices: they have simple structure, their ability to achieve the fundamental limits (the capacity) of the channel, and most importantly, they can be decoded using efficient decoders called *lattice decoders* [2]. Many researchers have studied the information-theoretic limits of lattice coding and decoding schemes for the linear Gaussian vector channel model [2]–[9].

Poltyrev [3] studied the problem of coding for the unconstrained additive white Gaussian noise (AWGN) channel where the channel input is an infinite lattice. In his setting, the notion of capacity becomes meaningless as infinite rates of transmission are possible. Therefore, another significant measurement was defined that characterizes the performance limits of such coding scheme when decoded using lattice decoders — the *normalized density* of the lattice or equivalently the information density rate of the lattice.

Based on a random lattice coding technique, Poltyrev showed that, using lattice decoding, the average probability of error can be upper bounded as

$$P_{e,\mathrm{av}}(\mu_c) \leq e^{-mE_p(\mu_c)}, \tag{1}$$

where m is the dimension of the lattice code, and $E_p(\mu_c)$ is called the *Poltyrev error exponent* and is shown to be a non-zero, monotonically increasing, positive function for all $\mu_c > 1$. The parameter μ_c, which is called the *volume-to-noise ratio*, to be defined in the sequel, is a quantity that is related to the density of the lattice. Hence, $\mu_c = 1$ has the significance of capacity.

In [4], Loeliger proved that the above upper bound can be achieved using ensembles of *linear* lattices — constructed using linear codes over the ring of p-prime integer numbers, i.e., $\mathbb{Z}_p$, which is usually referred to as Construction A [1]. An important aspect of both Poltyrev's and Loeliger's proofs is based on an important theorem in number theory that is referred to as *Minkowski-Hlawka* theorem [10], [11].

It is clear from the above bound that large lattice codes would be required to approach capacity and therefore more practical decoding methods would be needed. It is well known that lattice decoders that are implemented using sphere decoding algorithms[1] can be considered as a search in a *tree* (see [12], [13] and references therein). Generally speaking, a sphere decoding algorithm explores the tree of all possible lattice points and uses a *path metric* in order to discard paths corresponding

[1]Sphere decoding algorithms were originally implemented to decode signals transmitted via wireless fading channels [13], particularly for the quasi-static multiple-input multiple-output wireless channels as an attempt to reduce the high computational complexity of the optimal maximum-likelihood decoder (see [12]). The latter channel maybe described by the linear Gaussian vector channel model which allows the use of lattice coding, and lattice decoding to analyze the performance limits of such systems.

to points outside the search sphere. Unfortunately, sphere decoding suffers from high computational complexity for low-to-moderate volume-to-noise ratios[2] and for large signal dimensionality in which low error probability is expected [12]. As an alternative to sphere decoding algorithms, *sequential decoders* comprise a set of efficient and powerful decoding techniques able to perform the tree search. These decoders can achieve *near*-optimal performance without suffering the complexity of the sphere decoder for coding rates not too close to capacity[3] [14], [15].

The stack algorithm is a well known algorithm that is used to describe the operation of the sequential decoder [15]. The algorithm was originally constructed as an alternative approach to the maximum-likelihood (ML) decoder for detecting convolutional codes transmitted via discrete memoryless channels. It has been shown in [14] that as long as we operate below the cutoff rate, the decoder can achieve near-ML performance with low decoding complexity.

For the lattice coded/decoded linear Gaussian vector channel model, there is a small body of work that discusses the performance and complexity tradeoff achieved by sequential decoding algorithms. Initial work on this topic was done by Tarokh *et. al.* [17] where sequential decoding is used to decode lattice codes with finite trellis diagram. Shalvi *et. al.* in [18] has considered the use of sequential decoders to decode convolutional lattice codes. These power-limited (finite) lattice codes are generated using lattices combined with special lattice shaping techniques. The convolutional structure of such codes allows the use of the sequential decoders to achieve high data rates with low decoding complexity (this was mainly shown via simulation). However, all previous works lack a thorough theoretical analysis that can describe the systematic approach for tradeoff performance, complexity, and rate (or lattice density) achieved by sequential decoding of infinite lattice codes.

This paper presents a complete performance analysis of the lattice sequential decoder in terms of the achievable error exponent. Moreover, the computational complexity of the decoder is determined via its complexity tail distribution where a new notion of the "cut-off" rate is defined. Both, the error performance and the decoder complexity, are derived as a function of the decoding parameter – the *bias term*. In order to fully characterize the performance of the decoder, we determine for the first time the error exponent achievable by lattice coding and sequential decoding applied to the unconstrained AWGN channel. We derive the error exponent as a function of the bias term which is critical for controlling the amount of computations required at the decoding stage. Achieving low decoding complexity requires increasing the value of the bias term. However, this is done at the expense of increasing erroneous detection. In this work, we follow the footsteps of Poltyrev and use the same definition of capacity for such a channel. We make use of lattice codes drawn from the ensemble of linear lattices, i.e., the Loeliger construction [4].

We analyze in details the computational complexity tail distribution of the lattice sequential decoder. We show that there exists a cut-off volume-to-noise ratio that yields low decoding complexity which is also an increasing function of the bias term. We show that achieving low decoding complexity with good error performance comes at the expense of increasing the cut-off volume-to-noise ratio. Hence, lattice sequential decoders provide a systematic approach for tradeoff performance, complexity, and lattice density.

In contrast to most work in sequential decoding algorithms where the bias term is usually optimized to achieve a good performance-complexity tradeoff, we allow the bias term to vary freely and study the effect of this variation on the performance-complexity tradeoffs achieved by such decoders.

Throughout the paper, we use the following notation. The superscript $^\mathsf{T}$ denotes transpose. For a bounded region $\mathcal{R} \subset \mathbb{R}^m$, $V(\mathcal{R})$ denotes the volume of $\mathcal{R}$. We denote $\mathcal{S}_m(r)$ by the m-dimensional hypersphere of radius r with $V(\mathcal{S}_m(r)) = (\pi r^2)^{m/2}/\Gamma(m/2 + 1)$, where $\Gamma(m) = \int_0^\infty x^{m-1}e^{-x}\,dx$, is the Gamma function. Vectors are represented by bold lowercase letters, and matrices by bold uppercase letters where $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix. The l_2-norm of a vector $\boldsymbol{a}$ is denoted by $\|\boldsymbol{a}\|$. The notation $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{K})$ indicates that $\boldsymbol{v}$ is a real Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{K}$, and $\mathcal{E}\{\cdot\}$ represents the statistical average.

II. CODING WITHOUT RESTRICTION FOR THE AWGN CHANNEL

A. Lattice Properties

A *lattice* is a discrete pointset Λ in a Euclidean space $\mathbb{R}^m$ that is closed under vector addition, i.e., any translate $\Lambda + \boldsymbol{x}$ by a lattice point $\boldsymbol{x} \in \Lambda$ is just Λ again. Let $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m\}$ be a set of linearly independent vectors in $\mathbb{R}^m$. The set Λ of all linear combinations $\boldsymbol{x} = z_1\boldsymbol{g}_1 + z_2\boldsymbol{g}_2 + \cdots + z_m\boldsymbol{g}_m$ with integer coefficients z_i is a lattice, i.e.,

$$\Lambda = \{\boldsymbol{x} = \boldsymbol{G}\boldsymbol{z} : \boldsymbol{z} \in \mathbb{Z}^m\},$$

where $\boldsymbol{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m]$ is an $m \times m$ full-rank generator matrix. Thus, any lattice Λ in $\mathbb{R}^m$ can be seen as a linear transformation of the integer lattice $\mathbb{Z}^m$.

Some properties associated with the lattice Λ are of great importance for our analysis:

---

[2]The notion of "signal-to-noise ratio" is usually used for power-constrained channels where only a finite number of codewords or signals can be transmitted. Here, for infinite lattice codes, the notion of volume-to-noise ratio is used instead which will be introduced in the sequel.

[3]The work in [16] considered the application of lattice sequential decoders to various systems that can be described by the linear Gaussian vector channel model, such as the slowly-fading multiple-input multiple-output wireless channel, and the inter-symbol interference channel. In this work, it has been shown that near-sphere decoding performance can be achieved without suffering the high decoding complexity of the sphere decoder.

- The nearest neighbor quantizer $Q(\cdot)$ associated with Λ is defined by

$$Q_\Lambda(\boldsymbol{x}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \|\boldsymbol{\lambda} - \boldsymbol{x}\|.$$

- The Voronoi cell $\mathcal{V}(\boldsymbol{\lambda})$ that corresponds to the lattice point $\boldsymbol{\lambda} \in \Lambda$ is the set of points in $\mathbb{R}^m$ closest to $\boldsymbol{x}$, i.e.,

$$\mathcal{V}(\boldsymbol{\lambda}) = \{\boldsymbol{x} \in \mathbb{R}^m : Q_\Lambda(\boldsymbol{x}) = \boldsymbol{\lambda}\}.$$

  Voronoi cells associated with each lattice point $\boldsymbol{\lambda} \in \Lambda$ are congruent and therefore can be considered as a shift of $\mathcal{V}(\boldsymbol{0})$ by $\boldsymbol{\lambda}$.
- The volume of the Voronoi cell is given by

$$V(\boldsymbol{G}) \triangleq \mathrm{Vol}(\mathcal{V}(\boldsymbol{0})) = \sqrt{\det(\boldsymbol{G}^\mathsf{T}\boldsymbol{G})},$$

  with the property that $V(a\boldsymbol{G}) = a^m V(\boldsymbol{G})$ for any $a > 0$.
- The covering radius $r_{\mathrm{cov}}(\Lambda)$ is the radius of the smallest sphere centered at the origin that contains $\mathcal{V}(\boldsymbol{0})$. The effective radius $r_{\mathrm{eff}}(\Lambda)$ is the radius of the sphere with volume equal to $V(\boldsymbol{G})$. The packing radius $r_{\mathrm{pack}}(\Lambda)$ is the radius of the largest sphere centered at the origin inside the Voronoi cell $\mathcal{V}(\boldsymbol{0})$ (see Fig. 1).
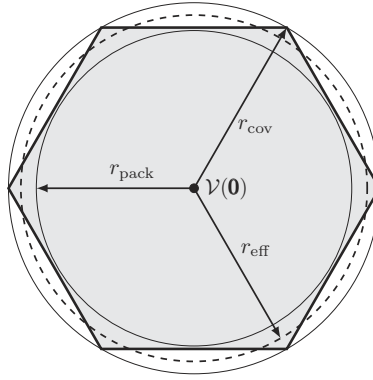

Fig. 1.   The packing radius, the effective radius, and the covering radius of the hexagonal lattice.

- A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is good for covering [8] if $r_{\mathrm{cov}}(\Lambda_m) \to r_{\mathrm{eff}}(\Lambda_m)$.
- **Minkowski-Hlawka Theorem** [10]: Let $f : \mathbb{R}^m \to \mathbb{R}$ be a Riemann integrable function of bounded support (i.e., $f(\boldsymbol{x}) = 0$ if $\|\boldsymbol{x}\|$ exceeds some bound). For any $\delta > 0$, there exist ensembles $\boldsymbol{\Lambda} = \{\Lambda\}$ of lattices with volume $V(\boldsymbol{G})$ and dimension m such that

$$\mathcal{E}_{\boldsymbol{\Lambda}}\left\{\sum_{\boldsymbol{x}\in\Lambda^*} f(\boldsymbol{x})\right\} \le (1+\delta)\frac{1}{V(\boldsymbol{G})}\int_{\mathbb{R}^m} f(\boldsymbol{x})\,d\boldsymbol{x}, \tag{2}$$

  where the expectation $\mathcal{E}_{\boldsymbol{\Lambda}}$ is taken over the ensemble of random lattices, $\Lambda^* = \Lambda\backslash\{\boldsymbol{0}\}$, and $\delta \to 0$ as $m \to \infty$. The above important theorem is sometimes regarded as a pre-Shannon result in information theory. In fact, the Mikowski-Hlawka theorem was originally used for packing lattices to solve the well known sphere-packing problem [11].

B. Poltyrev Error Exponent

Suppose that an m-dimensional lattice point $\boldsymbol{x} = \boldsymbol{G}_c\boldsymbol{z} \in \Lambda_c$ is to be transmitted through the unconstrained AWGN channel, where Λ_c is an infinite lattice code with volume $V_c \triangleq V(\boldsymbol{G}_c)$, that is drawn from the ensemble of linear lattices using the Loeliger construction (see [4] for more details about the construction). The received vector (output of the channel) in this case can be mathematically expressed as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}, \tag{3}$$

where $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_m)$. Due to the unconstrained power condition on the lattice codewords (points), the optimum receiver that minimizes the probability of decoding error can be expressed as

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}\in\mathbb{Z}^m} \|\boldsymbol{y} - \boldsymbol{G}_c\boldsymbol{z}\|^2, \tag{4}$$

which corresponds to searching over the whole lattice Λ_c to find the closest point to the received vector $\boldsymbol{y}$. This is referred to as lattice decoding.

As mentioned in the introduction, Poltyrev studied the problem of coding for the unconstrained AWGN channel with the input alphabet being the whole space $\mathbb{R}^m$. Since infinite power is possible, the notion of capacity becomes meaningless. Instead, the decoding error probability is measured against the normalized per dimension *volume-to-noise ratio* (VNR), μ_c, defined by

$$\mu_c \triangleq \frac{V(\boldsymbol{G}_c)^{2/m}}{2\pi e\sigma^2} = \frac{V_c^{2/m}}{2\pi e\sigma^2}, \tag{5}$$

where $(V_c^{2/m}/2\pi e) \cdot m$ represents the asymptotic (in dimension m) squared radius of a sphere of volume V_c.

Poltyrev showed that the average probability of error (averaged over the ensemble of linear lattice codes Λ_c) is upper bounded by (1) where

$$E_p(\mu_c) = \begin{cases} \frac{1}{2}\left[(\mu_c - 1) - \log\mu_c\right], & 1 < \mu_c \le 2; \\ \frac{1}{2}\log\left(\frac{e\mu_c}{4}\right), & 2 \le \mu_c \le 4; \\ \mu_c/8, & \mu_c \ge 4. \end{cases} \tag{6}$$

From the above analysis, one notices that μ_c can be interpreted as the ratio of the squared radius of a spherical Voronoi cell to the variance of the noise. For small μ_c, i.e., $\mu_c < 1$, the spherical Voronoi cell has radius less than the standard deviation of the noise. In this case, reliable communication is not possible as error is highly likely to occur. As such, $\mu_c = 1$ has the significance of capacity.

Interestingly, Poltyrev showed that if only a *finite* number of lattice points are to be transmitted as codewords with finite power constraint and transmission rate R, then rates R up to $1/2\log(\text{SNR})$ are achievable, where $\text{SNR} \ge 0$ here represents the average signal-to-noise ratio of the channel[4]. For high SNRs (i.e., for $\text{SNR} \gg 1$), $1/2\log(\text{SNR})$ represents the capacity of the AWGN channel, denoted by C. Therefore, the same error probability bound given in (1) and (6) can be used (asymptotically) to characterize the performance of the power-limited lattice coded/decoded AWGN channel by letting (at high SNR) $\mu_c = 2^{2[C-R]}$.

Unfortunately, lattice decoders (usually implemented using sphere decoding algorithms) suffer from high computational complexity for low-to-moderate SNR and for large signal dimensionality in which low error probability is to be expected. As an alternative to lattice decoders, *lattice sequential decoders* comprise a set of efficient and powerful decoding techniques that can achieve *near*-optimal performance without suffering the complexity of the lattice decoder for coding rates not too close to capacity C. In fact, it is well known that sequential decoders can work well (with low decoding complexity) for rates below the cut-off rate R_0 which is only a factor of $4/e$ (1.68 dB) away from capacity C at the high-SNR regime [20]. Therefore, for the unconstrained AWGN channel, $\mu_c = 4/e$ has the significance of the cut-off rate. Here, we call this the *cut-off VNR*, denoted by μ_0.

III. THE STACK SEQUENTIAL DECODER

In this section, we briefly introduce the operation of the stack algorithm. This algorithm is an efficient tree search algorithm that attempts to find a "best fit" with the received noisy signal. Before we proceed with the description of such an algorithm, we shall discuss the metric measure for sequential decoding of lattice codes. It is basically based on the *path metric* defined for conventional sequential decoders which is given by [16]

$$\mathcal{M}(\boldsymbol{z}_1^k) = \log\left(\frac{\Pr(\mathcal{H}(\boldsymbol{z}_1^k))f(\boldsymbol{y}_1^k|\mathcal{H}(\boldsymbol{z}_1^k))}{f(\boldsymbol{y}_1^k)}\right), \tag{7}$$

where $\mathcal{H}(\boldsymbol{z}_1^k)$ is the hypothesis that $\boldsymbol{z}_1^k$ form the first k symbols of the transmitted information sequence, and $f(\cdot)$ is the usual probability density function.

Recently, it has been shown that the search for the closest lattice point problem which corresponds to (4) can be efficiently performed using sequential decoders based on the stack algorithm [16]. For our channel model, the path metric given by (7) can be shown to be simplified to (see Appendix A in [16])

$$\mathcal{M}(\boldsymbol{z}_1^k) = bk - \|\boldsymbol{y'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2, \tag{8}$$

where $\boldsymbol{z}_1^k = [z_k, \cdots, z_2, z_1]^T$ denotes the last k components of the integer vector $\boldsymbol{z}$, $\boldsymbol{R}_{kk}$ is the lower $k \times k$ matrix of $\boldsymbol{R}$ that corresponds to the QR decomposition of the code matrix $\boldsymbol{G}_c = \boldsymbol{QR}$, $\boldsymbol{y'} = \boldsymbol{Q}^\top\boldsymbol{y}$, and b is the bias term.

As in the conventional stack decoder [15], to determine a best fit (path), a value is assigned to each node in the tree. This value is called the metric which is given by (8). A flow chart for the operation of the stack decoder is shown in Fig. 2. As the decoder searches nodes, an ordered list of previously examined paths of different lengths is kept in storage. Each stack entry contains a path along with its metric. Each decoding step consists of extending the top (best) path in the stack. The

---

[4]This can be simply done by intersecting the lattice code Λ_c (possibly shifted by a vector $\boldsymbol{u}_0$) with a shaping region $\mathcal{R}$ (a sphere or a Voronoi cell of another lattice), i.e., $\mathcal{C} = (\Lambda_c + \boldsymbol{u}_0) \cap \mathcal{R}$. In this case, the transmission rate is given by $R = \frac{1}{m}\log_2[V(\mathcal{R})/V_c]$, where $V(\mathcal{R})$ is the volume of the shaping region. If we define $mP_x = \frac{1}{|\mathcal{C}|}\sum_{\boldsymbol{x}\in\mathcal{C}}\|\boldsymbol{x}\|^2$ to be the average transmitted power, then one can show that $V(\mathcal{R})$ is asymptotically (as $m \to \infty$) given by $(2\pi eP_x)^{m/2}$. For reliable communication, we must have $V_c > (2\pi e\sigma^2)^{m/2}$. Therefore, rates R up to $\frac{1}{2}\log(P_x/\sigma^2) = \frac{1}{2}\log(\text{SNR})$ is achievable.

determination of the best and next best nodes is simplified in the closest lattice point search problem by using the Schnnor-Euchner enumeration [12] which generates nodes with metrics in ascending order given any node z_1^k. The decoding algorithm terminates when the top path in the stack reaches the end of the tree (refer to [15] for more details about the algorithm).
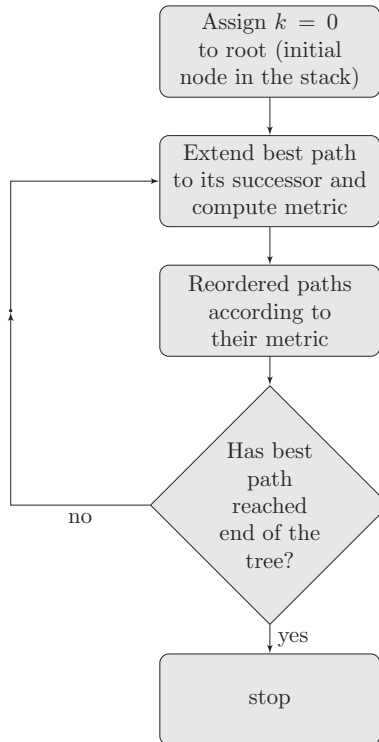

Fig. 2. Flow chart for stack decoding.

The main role of the bias term b used in the algorithm is to control the amount of computations performed by the decoder. In this work, we define the computational complexity of the lattice sequential decoder as the *total number of nodes visited by the decoder* during the search. Also, the bias term is responsible for the excellent performance-rate-complexity tradeoff achieved by such a decoding scheme. The role that the bias parameter plays will be discussed in detail in the subsequent sections.

IV. PERFORMANCE ANALYSIS: AN UPPER BOUND

As mentioned at the introduction, there has been no analysis devoted to sequential decoding applied to the lattice coded unconstrained AWGN channel. In this section, we analyze the performance limits of the stack sequential decoder when lattice coding is applied at the transmitter. We consider the unconstrained AWGN channel as defined by Poltyrev [3]. Finding the exact error performance of such a decoder seems to be difficult. Therefore, we attempt to derive an upper bound on the sequential decoding error probability.

Define $P_e(b)$ as the probability that the sequential decoder makes an erroneous detection at a bias value b (defined in (8)).

Now, due to lattice symmetry, one can assume that the all-zero lattice point $\mathbf{0}$ is transmitted. Then, we have that

$$
\begin{aligned}
P_e(b) &= \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\{\mathbf{0}\text{ was decoded as }\boldsymbol{x}\}\right) \\
&\overset{(a)}{\leq} \Pr\left(\bigcup_{\boldsymbol{z}\in\mathbb{Z}^m\setminus\{\mathbf{0}\}}\{\mathcal{M}(\boldsymbol{z}) > \mathcal{M}_{\min}\}\right) \\
&\overset{(b)}{\leq} \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\{\|\boldsymbol{x}\|^2 - 2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} < bm\}\right) \\
&= \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1 - \frac{bm}{\|\boldsymbol{x}\|^2}\right)\right\}\right), \\
&\overset{(c)}{\leq} \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1 - \frac{bm}{d_{\min}^2(\Lambda_c)}\right)\right\}\right),
\end{aligned}
\tag{9}
$$

where $\Lambda_c^* = \Lambda_c\setminus\{\mathbf{0}\}$, (a) is due to the fact that $\mathcal{M}(\boldsymbol{z}) > \mathcal{M}_{\min}$ is just a necessary condition for $\boldsymbol{x} = \boldsymbol{G}_c\boldsymbol{z}$ to be decoded by the stack decoder, where $\mathcal{M}_{\min} = \min\{0, b - \|\boldsymbol{w}_1^1\|^2, 2b - \|\boldsymbol{w}_1^2\|^2, \ldots, bm - \|\boldsymbol{w}_1^m\|^2\}$ is the minimum metric that corresponds to the transmitted path, (b) follows by noticing that $-(\mathcal{M}_{\min} + \|\boldsymbol{w}\|^2) \leq 0$, and (c) follows from the fact that $\|\boldsymbol{x}\| \geq \min_{\boldsymbol{x}\in\Lambda_c^*}\|\boldsymbol{x}\| \overset{\triangle}{=} d_{\min}(\Lambda_c)$ — the minimum Euclidean distance of the lattice.

It is clear from the above bound that the performance of the lattice sequential decoder depends critically on the shortest distance of the infinite lattice. Unfortunately, calculating the exact minimum distance of a lattice is NP-hard — a problem that is referred to the shortest vector problem [7]. Moreover, finding the exact probability that appears in the RHS of (9) for a particular lattice seems to be difficult. As such, we need to rely on a random technique to further upper bound the average error performance of the decoder. Before doing so, we need to ensure that the lattices in the ensemble are *reasonably good* for channel coding. In order to do this, we need to expurgate the lattice ensemble that appears in (2) appropriately such that the remaining lattices in the expurgated ensemble satisfy a lower bound on the packing radius of the lattice $r_{\text{pack}}(\Lambda_c)$, or equivalently on the minimum Euclidean distance $d_{\min}(\Lambda_c) = 2r_{\text{pack}}(\Lambda_c)$.

We recall the result in [6, Lemma 1] which states that most lattices in the random ensemble $\boldsymbol{\Lambda}$ that satisfies the Minkowski-Hlawka theorem have good minimum Euclidean distance. In other words, for a lattice Λ_c that is drawn from the random ensemble $\boldsymbol{\Lambda}$ we have that for $0 \leq \zeta < 1$

$$
\Pr(d_{\min}(\Lambda_c) > \zeta r_{\text{eff}}(\Lambda_c)) > 1 - \zeta^m,
\tag{10}
$$

where $r_{\text{eft}}(\Lambda_c)$ is the effective radius of Λ_c. Let $\boldsymbol{\Lambda}_{\text{exp}}$ be the expurgated lattice ensemble that satisfies (10), i.e.,

$$
\boldsymbol{\Lambda}_{\text{exp}}(\zeta) = \{\Lambda_c \in \boldsymbol{\Lambda} : d_{\min}(\Lambda_c) > \zeta r_{\text{eff}}(\Lambda_c), 0 \leq \zeta < 1\}.
\tag{11}
$$

In this case, it is straight forward to show that for a given lattice $\Lambda_c \in \boldsymbol{\Lambda}_{\text{exp}}$, the conditional error probability (9) can be further upper bounded by

$$
\begin{aligned}
&P_e(b|\Lambda_c) \leq \\
&\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1 - \frac{bm}{\zeta^2 r_{\text{eff}}^2(\Lambda_c)}\right)\right\}\bigg|\Lambda_c\right).
\end{aligned}
\tag{12}
$$

Averaging (12) over the expurgated lattice ensemble, we get

$$
\overline{P_e}(b) \leq \mathcal{E}_{\boldsymbol{\Lambda}_{\text{exp}}}\left\{\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1 - \frac{bm}{\zeta^2 r_{\text{eff}}^2(\Lambda_c)}\right)\right\}\bigg|\Lambda_c\right)\right\}.
\tag{13}
$$

Now, in order to use the Minkowski-Hlawka theorem with expurgated ensemble we will need the following relation:

$$
\mathcal{E}_{\boldsymbol{\Lambda}_{\text{exp}}}\{\boldsymbol{X}\} \leq \frac{1}{\Pr(d_{\min}(\Lambda_c) > \zeta r_{\text{eff}}(\Lambda_c))}\mathcal{E}_{\boldsymbol{\Lambda}}\{\boldsymbol{X}\},
\tag{14}
$$

where $\mathcal{E}_{\boldsymbol{\Lambda}}\{\cdot\}$ is the expectation with respect to the ensemble in (2), and $\boldsymbol{X}$ is a nonnegative random variable. Therefore, we have that

$$\overline{P_e}(b) \leq \frac{1}{1-\zeta^m}\mathcal{E}_{\boldsymbol{\Lambda}}\Bigg\{\Pr\Bigg(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\Big\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\Big(1-$$
$$\frac{bm}{\zeta^2 r_{\text{eff}}^2(\Lambda_c)}\Big)\Big\}\Big|\Lambda_c\Bigg)\Bigg\}. \tag{15}$$

As $m \to \infty$ we have

$$r_{\text{eff}}^2(\Lambda_c) = \frac{\Gamma\left(\frac{m}{2}+1\right)^{2/m}}{\pi}V_c^{2/m} \sim \frac{V_c^{2/m}}{2\pi e}\cdot m = \mu_c\sigma^2 m,$$

where μ_c is the VNR, and σ^2 is the noise variance. Note that, asymptotically, as $m \to \infty$, we may let ζ approaches 1 as close as desired. Therefore, the average probability of decoding error can be asymptotically upper bounded by

$$\overline{P_e}(b,\mu_c) \leq \mathcal{E}_{\boldsymbol{\Lambda}}\left\{\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1-\frac{b/\sigma^2}{\mu_c}\right)\right\}\right)\right\}$$
$$= \mathcal{E}_{\boldsymbol{\Lambda}}\left\{\Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{w}} > \|\boldsymbol{x}\|^2\right\}\right)\right\}, \tag{16}$$

where

$$\tilde{\boldsymbol{w}} = \left(1-\frac{b_n}{\mu_c}\right)^{-1}\boldsymbol{w},$$

is a zero-mean Gaussian random vector with elements that are independent, identically, distributed random variables with variance $\tilde{\sigma}^2 = (1-b_n/\mu_c)^{-2}\sigma^2$, and $b_n = b/\sigma^2$ is defined as the normalized bias with respect to the noise variance. It must be noted that the above bound is only valid for all values of b_n such that $1-b_n/\mu_c > 0$, or equivalently for all values of $0 \leq b_n < \mu_c$.

Interestingly, the upper bound (16) corresponds to the probability of decoding error of a received signal $\boldsymbol{y} = \boldsymbol{x}+\tilde{\boldsymbol{w}}$ decoded using the conventional lattice decoder. Therefore, one may observe that the sub-optimality of the sequential decoder can be viewed as a source of channel noise amplification.

Following the footsteps of Poltyrev, the average probability of error can be shown to be upper bounded by

$$\overline{P_e}(b,\mu_c) \leq e^{-mE_b(\mu_c)}, \tag{17}$$

where

$$E_b(\mu_c) = E_p(\tilde{\mu}_c) = \begin{cases} 0, & \tilde{\mu}_c \leq 1; \\ \frac{1}{2}\left[(\tilde{\mu}_c - 1) - \log\tilde{\mu}_c\right], & 1 < \tilde{\mu}_c \leq 2; \\ \frac{1}{2}\log\left(\frac{e\tilde{\mu}_c}{4}\right), & 2 \leq \tilde{\mu}_c \leq 4; \\ \tilde{\mu}_c/8, & \tilde{\mu}_c \geq 4. \end{cases} \tag{18}$$

where

$$\tilde{\mu}_c \triangleq \frac{V(\boldsymbol{G}_c)^{2/m}/2\pi e}{\tilde{\sigma}^2} = \mu_c\left(1-\frac{b_n}{\mu_c}\right)^2. \tag{19}$$

Hence, for sufficiently large m, there exists at least a lattice Λ_c' in the expurgated code ensemble with error probability satisfying

$$P_e(b,\mu_c,\Lambda_c') \leq e^{-mE_b(\mu_c)}. \tag{20}$$

Now, the following important remarks can be made about the above result:

- **Fixed Bias**: In this case, the bias term b is fixed and chosen independent of the VNR μ_c. Note that as μ_c gets large ($\mu_c \gg 1$), one may approximate $\tilde{\mu}_c$ in (19) as

$$\tilde{\mu}_c = \mu_c\left(1-\frac{b_n}{\mu_c}\right)^2 \approx \mu_c\left(1-2\frac{b_n}{\mu_c}\right) = \mu_c - 2b_n. \tag{21}$$

Therefore, the above analysis indicates that fixing the bias term causes a right-shift to the error probability curve (i.e., a reduction in the coding gain). This can be realized from the value of the error exponent for large μ_c which is given by

$$E_b(\mu_c) = \frac{\tilde{\mu}_c}{8} = \frac{\mu_c}{8} - \frac{b_n}{4} = E_p(\mu_c) - \frac{b_n}{4}, \tag{22}$$

where $E_p(\mu_c)$ is the Poltyrev error exponent achieved by the lattice decoder which is defined in (6). Substituting (22) into (20) we get

$$P_e(b, \mu_c) \leq e^{-mE_b(\mu_c)} = e^{-m[E_p(\mu_c) - b_n/4]}$$
$$= \alpha e^{-mE_p(\mu_c)},$$

(23)

where $\alpha = e^{mb_n/4}$. The constant α describes the behavior of the error probability of the sequential decoder for a fixed bias term. Increasing the bias term results in performance reduction compared to the one achieved by the lattice decoder. This reduction is represented by a right-shift to the error probability curve[5], as will be shown in the sequel.

- **Variable Bias**: Now, let the normalized bias b_n to scale linearly with the VNR μ_c as $b_n = (1 - \sqrt{\delta})\mu_c$ where $0 < \delta \leq 1$, then the error exponent in this case can be expressed as

$$E_b(\mu_c) = \begin{cases} 0, & \mu_c \leq 1/\delta; \\ \frac{1}{2}\left[(\delta\mu_c - 1) - \log \delta\mu_c\right], & 1/\delta < \mu_c \leq 2/\delta; \\ \frac{1}{2}\log\left(\frac{e\delta\mu_c}{4}\right), & 2/\delta \leq \mu_c \leq 4/\delta; \\ \delta\mu_c/8, & \mu_c \geq 4/\delta. \end{cases}$$

(24)

It is clear from the above analysis that if $\delta \to 1$ (or $b_n \to 0$) then the performance of the sequential decoder approaches the performance of the lattice decoder. On the other extreme, if $\delta \to 0$ (or $b_n \to \mu_c$) then reliable communication may not be possible under lattice sequential decoding. Fig. 3 shows the error exponent achieved by the lattice sequential decoder for the case of the variable bias term described above. It is clear from Fig. 3 that for high VNR μ_c, the effect of varying δ occurs as a change in the *slope* of the error exponent curve, where at high VNR we have $E_b(\mu_c) = \delta\mu_c/8$. Moreover, the maximum achievable VNR under sequential decoding with normalized bias $b_n = (1 - \sqrt{\delta})\mu_c$ is given by $1/\delta$. Therefore, for $\delta \neq 1$, reliable communication may not be possible at VNR close to capacity ($\mu_c = 1$).
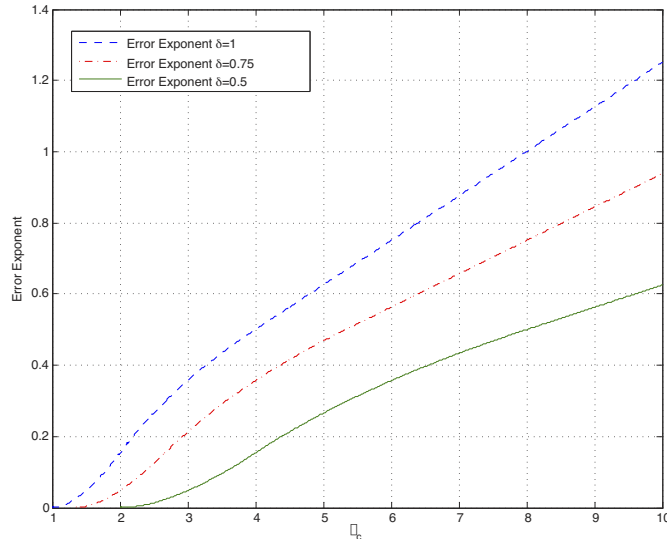

Fig. 3. The achievable error exponent of the lattice sequential decoder when the normalized bias term $b_n = (1 - \sqrt{\delta})\mu_c$ for $\delta = 0.5, 0.75,$ and 1.

The main result that we draw from the above discussion is that, increasing the bias term lower the decoding error probability. However, the loss in the error performance achieved by any sub-optimal decoder is usually compensated by some improvements in the decoding complexity. This fact will be demonstrated next where we consider the computational complexity analysis of the sequential decoder for the unconstrained AWGN channel.

V. THE "CUT-OFF" VOLUME-TO-NOISE RATIO: AN UPPER BOUND ON THE COMPLEXITY DISTRIBUTION

The main use of the sequential decoder is to achieve a low decoding complexity compared to the very complex lattice decoder. As in conventional sequential decoder, one need to back-off from capacity to achieve such improvements. For convolutional codes detected using sequential decoders, a cut-off rate has been defined for such decoding schemes. The cut-off rate R_0 is the rate for which the transmitter should not exceed if one needs to expect a low decoding complexity. If $R > R_0$, then

---

[5]It must be noted that, although the bound (20) shows that the shift is $\alpha = e^{mb_n/4}$, the exact amount of right-shift is less than α as will be shown by the simulation results in Section VI.

the complexity of the sequential decoder increases exponentially with the constraint length of the code [14]. For the power-constrained AWGN channel, the cut-off rate R_0 is $4/e$ (1.68 dB) away from capacity at high SNR [20]. In this section, we study in details the complexity of the sequential decoder for the unconstrained AWGN channel.

Due to the random nature of the channel noise, the computational complexity of the lattice sequential decoder is also random. Therefore, it would be more appropriate to study the complexity behavior of such a decoder via its complexity tail distribution defined as $\Pr(\mathcal{N}(\Lambda_c) \geq L)$, where $\mathcal{N}(\Lambda_c)$ is defined as the *total number of nodes in the tree that have been visited by the decoder during the search* for a given lattice Λ_c, and L is the distribution parameter.

Similar to the power-limited AWGN channel, we define a "cut-off" VNR μ_0 for the unconstrained AWGN channel to be the value of μ_c for which both low decoding complexity and low decoding error probability are possible.

To simplify the analysis, we start by bounding the total number of computations $\mathcal{N}(\Lambda_c)$ from above as follows. First, one should note that all nodes in the tree that have been visited by the sequential decoder must have partial path metrics $\mathcal{M}(\boldsymbol{z}_1^k)$ that exceed the minimum metric $\mathcal{M}_{\min}$ which corresponds to the decoded path. Let $\phi(\boldsymbol{z}_1^k)$ be the indicator function defined by

$$\phi(\boldsymbol{z}_1^k) = \begin{cases} 1, & \mathcal{M}(\boldsymbol{z}_1^k) \geq \mathcal{M}_{\min}; \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

Since $\mathcal{M}(\boldsymbol{z}_1^k) \geq \mathcal{M}_{\min}$ is a sufficient condition for the node to be visited by the decoder, then $\mathcal{N}(\Lambda_c)$ may be upper bounded by

$$\mathcal{N}(\Lambda_c) \leq \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k). \tag{26}$$

Also, the complexity tail distribution can be upper bounded as

$$\Pr(\mathcal{N}(\Lambda_c) \geq L) \leq \Pr(\mathcal{N}(\Lambda_c) \geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m) + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m), \tag{27}$$

where the above upper bound is derived using the well known separation of the typical noise events from the non-typical ones [20]. Next, we would like to upper bound the first term in the RHS of (27).

Given $\|\boldsymbol{w}\|^2 \leq \sigma^2 m$, and by noticing that $-(\mathcal{M}_{\min} + \|\boldsymbol{w}\|^2) \leq 0$, we obtain

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k) \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \tag{28}$$

where $\phi(\boldsymbol{z}_1^k)$ is the indicator function defined in (25), and

$$\phi^{'}(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{w}_1^{'k} - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk + \sigma^2 m; \\ 0, & \text{otherwise}. \end{cases} \tag{29}$$

where $\boldsymbol{w}_1^{'k}$ is the last k components of $\boldsymbol{w}' = \boldsymbol{Q}^\mathsf{T}\boldsymbol{w}$. Now, let

$$\phi_k^{''}(\boldsymbol{z}) = \begin{cases} S_k, & \text{if } \|\boldsymbol{w}' - \boldsymbol{R}\boldsymbol{z}\|^2 \leq bm - \mathcal{M}_{\min}; \\ 0, & \text{otherwise}, \end{cases}$$

where

$$S_k = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \tag{30}$$

then it can be shown that

$$\mathcal{N}(\Lambda_c) \leq \sum_{k=1}^{m} \sum_{\boldsymbol{z} \in \mathbb{Z}^m} \phi_k^{''}(\boldsymbol{z}) \leq \sum_{k=1}^{m} \sum_{\boldsymbol{x} \in \Lambda_c} \tilde{\phi}_k(\boldsymbol{x}),$$

where

$$\tilde{\phi}_k(\boldsymbol{x}) = \begin{cases} S_k, & \text{if } \|\boldsymbol{x}\|^2 - 2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} \leq bm; \\ 0, & \text{otherwise}. \end{cases}$$

Interestingly, the sum that appears in (30) represents the number of partial integer lattice points $\boldsymbol{z}_1^k \in \mathbb{Z}^k$ that are located inside a sphere of squared radius $bk + \sigma^2 m$ centered at the received signal ($\boldsymbol{y} = \boldsymbol{w}$ in our case). One can approximate S_k by the ratio of the volume of the k-dimensional sphere of squared radius $bk + \sigma^2 m$ to the volume of the Voronoi cell of the lattice Λ_k that corresponds to $\boldsymbol{R}_{kk}$, denoted by $V(\boldsymbol{R}_{kk})$, (see [11] for more details), i.e.,

$$S_k \approx \frac{\mathcal{S}_k(\sqrt{bk + \sigma^2 m})}{V(\boldsymbol{R}_{kk})} = \frac{(\pi)^{k/2}}{\Gamma(k/2 + 1)} \frac{[bk + \sigma^2 m]^{k/2}}{\det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2}}. \tag{31}$$

For a given lattice Λ_c, we have

$$
\begin{aligned}
\Pr(\mathcal{N}(\Lambda_c) &\geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m | \Lambda_c) \\
&\leq \Pr(\tilde{\mathcal{N}}(\Lambda_c) \geq L - m, \|\boldsymbol{w}\|^2 \leq \sigma^2 m | \Lambda_c) \\
&\leq \frac{\mathcal{E}_{\boldsymbol{w}}\{\tilde{\mathcal{N}}(\Lambda_c)|\Lambda_c, \|\boldsymbol{w}\|^2 \leq \sigma^2 m\}}{L - m}, \quad \text{for } L > m,
\end{aligned}
\tag{32}
$$

where the last inequality follows from using Markov inequality, and $\tilde{\mathcal{N}}(\Lambda_c)$ is defined as

$$
\tilde{\mathcal{N}}(\Lambda_c) = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k \setminus \{\boldsymbol{0}\}} \phi(\boldsymbol{z}_1^k),
$$

since we have assumed that the all-zero lattice point was transmitted.

The conditional average of $\tilde{\mathcal{N}}(\Lambda_c)$ with respect to the noise can be further upper bounded as

$$
\begin{aligned}
\mathcal{E}_{\boldsymbol{w}}\{\tilde{\mathcal{N}}(\Lambda_c)|\Lambda_c, \|\boldsymbol{w}\|^2 \leq \sigma^2 m\} \\
\leq \sum_{k=1}^{m} S_k \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{x}\|^2 - 2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} < bm).
\end{aligned}
\tag{33}
$$

Therefore, we have

$$
\begin{aligned}
\Pr(\mathcal{N}(\Lambda_c) &\geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m | \Lambda_c) \\
&\leq \frac{\sum_{k=1}^{m} S_k}{L - m} \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm).
\end{aligned}
\tag{34}
$$

Now, for $L = m + \sum_{k=1}^{m} S_k$, we have that

$$
\begin{aligned}
\Pr(\mathcal{N}(\Lambda_c) &\geq L | \Lambda_c) \\
&\leq \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm) + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m) \\
&\leq \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr\left(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 \left(1 - \frac{bm}{d_{\min}^2(\Lambda_c)}\right)\right) \\
&\qquad\qquad + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m).
\end{aligned}
\tag{35}
$$

At this point, it is worth mentioning that the first term in the RHS of (35) represents the *union bound* of the sequential decoding error probability in (9). Therefore, as will be shown in the sequel, there exists a minimum VNR μ_c, defined as the cut-off VNR μ_0, such that for all $\mu_c > \mu_0$ low decoding error probability and low decoding complexity can be achieved simultaneously.

Similar to the error probability analysis, assuming Λ_c is drawn from the expurgated lattice ensemble $\boldsymbol{\Lambda}_{\text{exp}}$ defined in (11), we have that

$$
\begin{aligned}
\Pr\left(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 \left(1 - \frac{bm}{d_{\min}^2(\Lambda_c)}\right)\right) \\
\leq \Pr\left(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 \left(1 - \frac{bm}{\zeta^2 r_{\text{eff}}^2(\Lambda_c)}\right)\right).
\end{aligned}
$$

where $0 < \zeta < 1$. Using Chernoff bound, as $m \to \infty$

$$
\Pr\left(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 \left(1 - \frac{bm}{\zeta^2 r_{\text{eff}}^2(\Lambda_c)}\right)\right) \leq e^{-\|\boldsymbol{x}\|^2/8\tilde{\sigma}^2},
$$

where $\tilde{\sigma}^2 = (1 - b_n/\mu_c)^{-2}\mu_c$, and $b_n = b/\sigma^2$. Therefore, one may asymptotically upper bound the first term in the RHS of (35) as

$$
\sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm) \leq \sum_{\boldsymbol{x} \in \Lambda_c^*} e^{-\|\boldsymbol{x}\|^2/8\tilde{\sigma}^2}.
\tag{36}
$$

Substituting (36) in (35), and taking the expectation of (35) over the ensemble of expurgated lattices, we obtain

$$
\begin{aligned}
\overline{\Pr(\mathcal{N}(\Lambda_c) \geq L)} &= \mathcal{E}_{\boldsymbol{\Lambda}_{\text{exp}}}\{\Pr(\mathcal{N}(\Lambda_c) \geq L | \Lambda_c)\} \\
&\leq \frac{1}{V_c} \int_{\mathbb{R}^m} e^{-\|\boldsymbol{x}\|^2/8\tilde{\sigma}^2} \, d\boldsymbol{x} + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m).
\end{aligned}
\tag{37}
$$

Evaluating the integral in the above upper bound we get

$$\frac{1}{V_c} \int_{\mathbb{R}^m} e^{-\|\boldsymbol{x}\|^2/8\tilde{\sigma}^2} \, d\boldsymbol{x} = \frac{(8\pi\tilde{\sigma}^2)^{m/2}}{V_c} = \left(\frac{4/e}{\tilde{\mu}_c}\right)^m, \tag{38}$$

where $\tilde{\mu}_c = \mu_c \left(1 - b_n/\mu_c\right)^2$. Therefore, for large m, we can further upper bound (37) for all $0 \le b_n \le \mu_c$ as

$$\overline{\Pr(\mathcal{N}(\Lambda_c) \ge L)} \le \left(\frac{4/e}{\tilde{\mu}_c}\right)^m + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m). \tag{39}$$

Hence, for sufficiently large m, there exists at least a lattice Λ_c' in the ensemble with complexity tail distribution satisfying (39) for all values of $L \ge m + \sum_{k=1}^{m} S_k$. It follows from standard typicality arguments that for any $\epsilon > 0$ there exists m_0 such that for all $m > m_0$

$$\Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m) < \epsilon/2.$$

The first term in the upper bound (39) can be made smaller than $\epsilon/2$ for sufficiently large m, i.e.,

$$\Pr(\mathcal{N}(\Lambda_c') \ge L) \le \epsilon, \quad m \to \infty,$$

if $\tilde{\mu}_c > 4/e$. This result indicates that large computational complexity may be avoided, while maintaining good error performance, at μ_c above the cut-off VNR μ_0 which is given by the roots of the following equation

$$\mu_0 \left(1 - \frac{b_n}{\mu_0}\right)^2 = \frac{4}{e}.$$

Note that when $b_n = 0$ we have $\mu_0 = 4/e$. Under this constraint, solving the above equation for μ_0 we get

$$\mu_0 = \left(b_n + \frac{2}{e}\left[1 + \sqrt{b_n e + 1}\right]\right). \tag{40}$$

It is interesting to note that as $b \to 0$ (the value of the bias that achieves close to lattice decoding performance) we have $\mu_0 \to 4/e$, where $4/e$ represents the gap between the cut-off rate and the capacity of the power-constraint AWGN channel [19], [20]. Since the union bound in (35) provides a good estimate to the decoding error probability at high VNR (i.e., for μ_c greater than the cut-off VNR μ_0) (see [19]), achieving a good error performance for large values of b, where low decoding complexity is expected, comes at the expense of increasing the VNR (or equivalently reducing the coding rate for the case of finite lattice codes).

The analysis above indicates that the total number of computations that is required by the decoder to decode a message while achieving low error probability can be approximated by

$$L \approx m + \sum_{k=1}^{m} \frac{(\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + \sigma^2 m]^{k/2}}{(\det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2})}. \tag{41}$$

In order to see how the complexity is affected by the channel and the decoder parameters, we express the unconstrained AWGN channel by the vector model $\boldsymbol{y} = \sqrt{\mu_c}\boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{x}$ is the transmitted lattice point that is selected randomly from a lattice Λ_c with generator matrix $\boldsymbol{G}_c = \boldsymbol{Q}\boldsymbol{R}$, μ_c is the VNR, and $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The volume of the Voronoi cell of Λ_c is selected such that the VNR at the output of the channel is μ_c. In this case we have $V_c = (2\pi e)^{m/2}$. As a result, we may express L as

$$L \approx m + \sum_{k=1}^{m} \frac{(\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[b_n k + m]^{k/2}}{\mu_c^{k/2} \det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2}}. \tag{42}$$

It is clear from the above equation that as $\mu_c \to \infty$, we have $L \to m$. Therefore, regardless the value of the bias term chosen at the decoding stage, the complexity of the decoder is approximately linear in the code dimension when the VNR is very large. This fact is also verified experimentally as will be shown in the sequel.

In fact, as expected, the total number of computations performed by the decoder at any VNR μ_c beyond μ_0 is bounded from above. This can be seen by substituting (40) in (42). Then, one may upper bound the total number of computations performed by the decoder as

$$\begin{aligned} L &\le m + \sum_{k=1}^{m} \frac{(\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[b_n k + m]^{k/2}}{\mu_0^{k/2} \det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2}} \\ &\approx m + \sum_{k=1}^{m} \frac{(\pi k)^{k/2}}{\Gamma(k/2+1) \det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2}}, \quad b_n \to \infty. \end{aligned} \tag{43}$$

It is clear from the above bound that as b increases (or as $b_n \to \infty$), the computational complexity scales almost linearly with the code dimension m. The simulation results (introduced next) agree with the above analysis.

In conclusion, lattice sequential decoders allow for a systematic approach for trading off performance, VNR, and complexity. For a fixed VNR, increasing the bias term allows to achieve low decoding complexity but at the expense of poor performance. In order to improve the performance without affecting the complexity, one need to increase the VNR μ_c or equivalently to increase the lattice density, to recover the performance loss.

VI. SIMULATION RESULTS

In our simulation we consider the unconstrained AWGN channel with m channel uses that is described by the vector model $\boldsymbol{y} = \sqrt{\mu_c}\boldsymbol{x} + \boldsymbol{w}$, where $\boldsymbol{x}$ is the transmitted lattice point that is selected randomly from a lattice Λ_c with generator matrix $\boldsymbol{G}_c$, μ_c is the VNR, and $\boldsymbol{w}$ is an AWGN vector with elements that are independent identically distributed, zero-mean Gaussian random variables with unit variance. We consider the Loeliger ensemble of mod-p lattices, where p is a prime. First, we generate the set of all lattices given by $\Lambda_c = \kappa(\mathsf{C} + p\mathbb{Z}^m)$, where κ is a scaling coefficient chosen such that the Voronoi cell volume $V_c = (2\pi e)^{m/2}$, $\mathbb{Z}_p$ denotes the field of mod-p integers, and $\mathsf{C} \subset \mathbb{Z}_p^m$ is a linear code over $\mathbb{Z}_p$ with generator matrix in systematic form $[\boldsymbol{I}\ \boldsymbol{P}^{\mathsf{T}}]^{\mathsf{T}}$, where $\boldsymbol{P}$ is the parity-check matrix. In the following, we select a lattice code at random with $p = 1001$ and fix the code for all simulation results.

Fig. 4 and Fig. 5 demonstrate the great advantage of using the lattice sequential decoder as an alternative to the optimal lattice sphere decoder. The performance and the complexity of both decoders are plotted for a lattice code of length $m = 30$. As depicted in Fig. 5, there is a significant complexity reduction achieved by using the sequential decoder over the lattice decoder, especially for low-to-moderate VNR, at the expense of very low error performance loss (a fraction of dB), for a bias value $b = 1$ (see Fig. 4).

Fig. 6 and Fig. 7 show the effect of increasing the bias term on the average error probability for the case of fixed and variable bias values, respectively. In Fig. 6, we choose fixed bias values (independent of μ_c) and plot the average error probability versus the VNR μ_c in dB. We also plot the performance of the optimal lattice decoder implemented via the sphere decoder algorithm [12] to measure the price of using the sequential decoder in terms of the performance loss. It is clear from the figure that increasing the bias term causes a right-shift to the sequential decoder error probability curve, while maintaining the rate at which the curve decays,[6] particularly at high VNR values. This basically agrees with the derived theoretical results provided
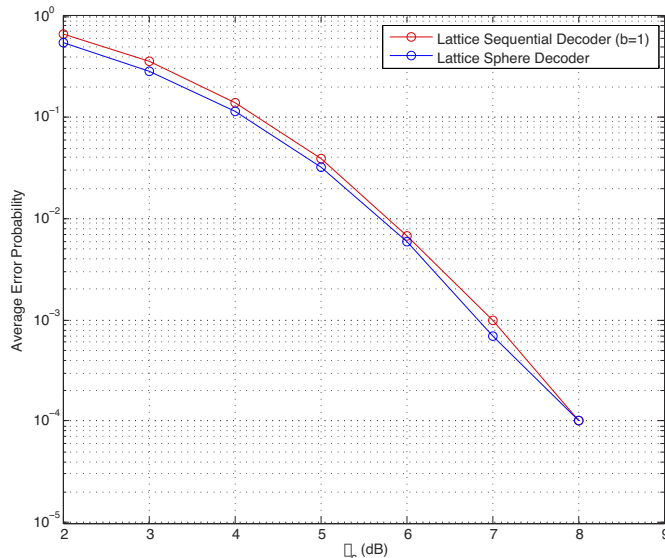

Fig. 4. Performance comparison between the sphere decoder and the sequential decoder with $b = 1$ for a lattice code of dimension $m = 30$.

in (22) and (23). On the other hand, if we let b to scale linearly with VNR as $(1 - \sqrt{\delta})\mu_c$, where $0 \le \delta \le 1$ (see (24)), then according to the error exponent analysis, we expect that the rate of decay (slope) of the error probability curve would decrease[7] as we decrease δ. This is depicted in Fig. 7, which also agrees with the derived theoretical results.

---

[6]The asymptotic rate of decay of the error probability curve maybe defined as

$$\text{slope} \triangleq \lim_{\mu_c \to \infty} \frac{-\log_e P_e(\mu_c)}{\log_e \mu_c}.$$

Now, for the case of fixed bias, using (23) we get

$$\text{slope} = \lim_{\mu_c \to \infty} \frac{m\mu_c}{8\log_e \mu_c} - \lim_{\mu_c \to \infty} \frac{b_n}{4\log_e \mu_c} = \lim_{\mu_c \to \infty} \frac{m\mu_c}{8\log_e \mu_c},$$

which indicates that the slope of the error probability is the same for any finite b.

[7]In this case, the rate of decay (slope) of the error probability curve can be shown to be equal to $\delta[m\mu_c/8\log_e \mu_c]$ which depends on b via δ.
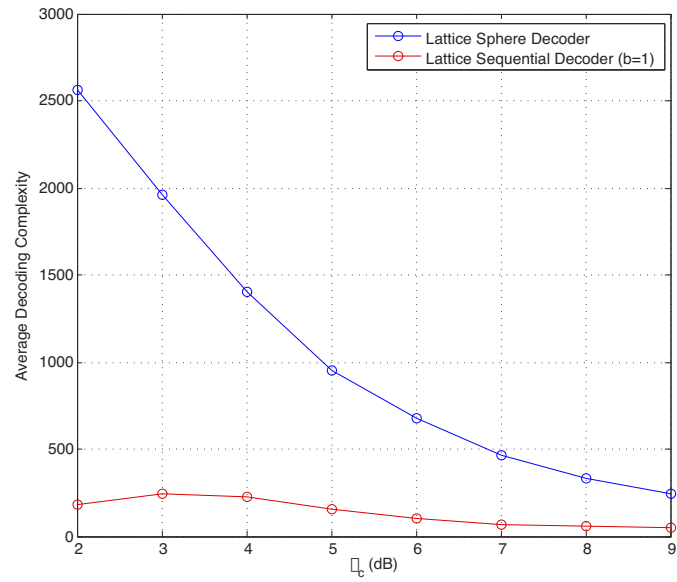
Fig. 5. Average computational complexity comparison between the lattice sphere decoder and the sequential decoder with bias term $b = 1$ for a lattice code of dimension $m = 30$.
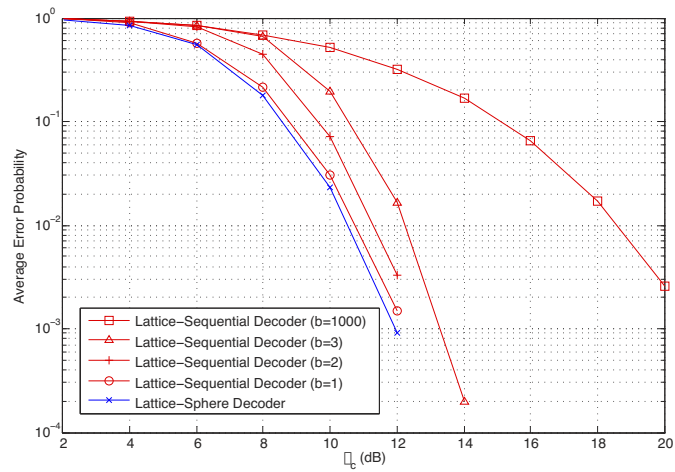

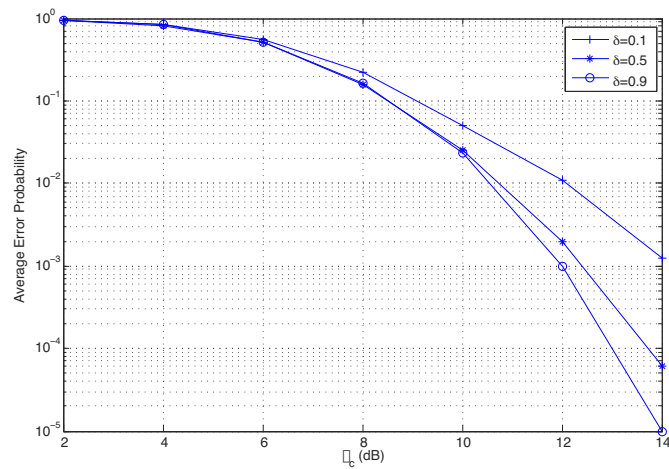Fig. 6. Comparison of the lattice sequential decoder's performance for various values of (fixed) bias term.

Fig. 7. Comparison of the lattice sequential decoder's performance when the bias term varies with the VNR as $b_n = (1 - \sqrt{\delta})\mu_c$ for several values of δ.
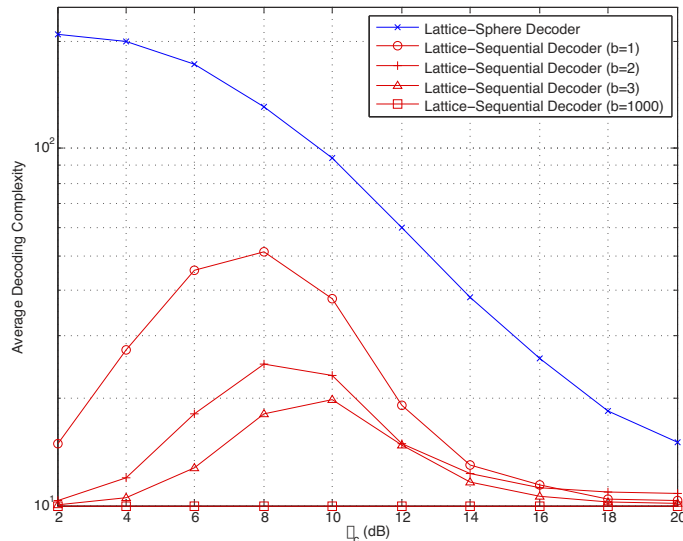
Fig. 8. The average computational complexity achieved by the sequential decoder for different values of the bias term.

Finally, Fig. 8 shows the effect of increasing the bias term on the average computational complexity (defined as the total number of visited nodes during the search). For comparison, we also include in Fig. 8 the average complexity of the sphere decoder for the same lattice code. The average complexity is plotted versus the VNR in dB. It is clear that for all values of b the sequential decoder has much lower complexity compared to the lattice (sphere) decoder, especially for low-to-moderate VNRs. The reason for the bell-like shape of the average complexity that occurs at low-to-moderate VNRs is due to the fact that with high-probability the received signal is close to the edge of the Voronoi cell. This basically requires the decoder to visit more nodes in the tree before decoding the message. As the VNR decreases or increases, the received signal becomes closer to a wrong lattice point or to the transmitted lattice point, respectively, which allows the decoder to decode the message without visiting many nodes. This leads to the very low average complexity as depicted in Fig. 8. The result also shows that as we increase the bias term, the average complexity significantly reduces, especially for low-to-moderate VNR values. As $b \to \infty$, the number of computations becomes equal to m (the signal dimension) for all VNR values. This agrees with the derived theoretical results.

In conclusion, simulation results indicate that increasing the bias term in the decoding algorithm significantly reduces the complexity but at the expense of losing performance.

VII. CONCLUSION

In this paper, we have analyzed the performance limits and the computational complexity of the lattice sequential decoder applied to the unconstrained AWGN channel. The error probability has been analyzed following the footsteps of Poltyrev by deriving the error exponent of the sequential decoder as a function of the VNR and the decoding parameter—the bias term. The bias term is responsible for the performance-complexity tradeoff achieved by the decoder. It has been shown (analytically and via simulation) that, if the bias term is fixed and independent of the VNR, then increasing the bias term causes only a right-shift to the error probability curve (occurs as a loss in the coding gain). However, if the bias term is scaled linearly with the VNR, the rate at which the error probability curve decays gets affected accordingly. It has also been shown that increasing the bias term significantly reduces the average number of computations required by the decoder to decode a message. However, the price of the complexity improvements comes at the expense of a loss in the performance. Hence, a fundamental trade-off exists between the error performance, the decoding complexity, and the VNR.

By revealing the tradeoff between performance, complexity and lattice density, it introduces the concept of lattice density into lattice decoding for the first time making it a promising area in lattice applications for digital and wireless communications. An interesting venue for future work is to derive bounds on the moments of sequential decoding complexity. As shown in this paper, since the decoding complexity is random, there exists a non-zero probability that the amount of computations performed by the decoder may become excessive causing a buffer overflow which is considered an important metric for the design of a sequential decoder. Therefore, studying these moments (e.g., the variance of the decoding complexity) is important to obtain estimates on the probability of buffer overflow [21].

References

[1] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*, 3rd ed. Springer Verlag NewYork, 1999.

[2] R. deBuda, "The upper bound of a new near-optimal code," *IEEE Trans. on Inform. Theory*, vol. IT-21, no. 7 pp. 441-445, July 1975.

[3] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 409-417, Mar. 1994.

[4] H. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1767-1773, Nov. 1997.

[5] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. on Inform. Theory*, vol. 44, no. 1, pp. 273-278, Jan. 1998.

[6] A. Ingber and R. Zamir, "Expurgated infinite constellations at finite dimensions," in *Proc. IEEE Int'l Symp. Inform. Theory*, (ISIT'12) MA, USA, July 2012.

[7] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. on Inform. Theory*, vol. 48, no. 8, pp. 22012214, Aug. 2002.

[8] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3401-3416, Oct. 2005.

[9] L.-C. Choo, C. Ling, and K.-K. Wong, "Achievable rates for lattice coding over the Gaussian wiretap channel," in *Proc. IEEE Physical Layer Security Workshop in Conjunction with IEEE Int'l Comm. Conf. (ICC'11)*, Kyoto, Japan, June 2011.

[10] H. Minkowski, "Zur Geometrie der Zahlen," *Math. Z.*, vol. 49, pp. 285-312, 1944.

[11] C. A. Rogers, *Packing and Covering*, Cambridge, UK: Cambridge Uni. Press, 1964.

[12] B. Hassibi and H. Vikalo, "On sphere decoding algorithm. Part I: expected complexity," *IEEE Trans. Sign. Proc.*, vol. 53, no. 8, pp. 2389-2401, Aug. 2005.

[13] J. Boutros, N. Gresset, L. Brunel, and M. Fossorier, "Soft-input soft-output lattice sphere decoder for linear channels," in *Proc. IEEE Global Comm. Conf. (GLOBECOM03)*, San Francisco, USA, Dec. 2003.

[14] I. M. Jacobs and E. R. Berlekamp, "A lower bound to the distribution of computation for sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 167-174, April 1976.

[15] F. Jelinek, "A fast sequential decoding algorithm using a stack," *IBM J. Res. Dev.* , vol. 13, pp. 675-685, Nov.1969.

[16] A. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 933-953, Mar. 2005.

[17] V. Tarokh, A. Vardy, and K. Zeger, "Sequential decoding of lattice codes," in *Proc. IEEE Int'l Symp. Inform. Theory*, (ISIT'97), Ulm, Germany, June 1997.

[18] O. Shalvi, N. Sommer, and M. Feder, "Signal codes: Convolutional lattice codes," *IEEE Trans. on Inform. Theory*, vol. 57, no. 8, pp. 5203-5226, Aug. 2011.

[19] G. D. Forney Jr., M. D. Trott, and S. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. on Inform. Theory*, vol. 46, no. 3, pp. 820-850, May 2000.

[20] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.

[21] J. E. Savage, "The computation problem with sequential decoding", M. I. T Lincoln Lab., Lexington, Mass., Tech. Rept. 371, Feb. 1965.
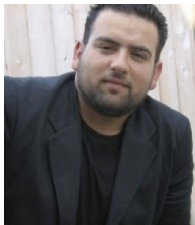
Walid Abediseid (S'04, M'11) was born in Etobicoke, Ontario, Canada. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from the University of Ottawa, Canada, in 2004 and 2007, respectively. He then received his Ph.D. from the Department of Electrical and Computer Engineering, University of Waterloo in 2011. He is a postdoctoral fellow at King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia, since December 2011. His research interests include coding and information theory, MIMO wireless communication systems, lattice applications for digital and wireless communications, detection and estimation.

Dr. Abediseid was a recipient of Research In Motion Graduate Scholarship from 2008 to 2009, and the NSERC Alexander Graham Bell Canada Graduate Scholarship — one of Canada's most prestigious graduate awards from 2009 to 2011.

Mohamed-Slim Alouini (S'94, M'98, SM'03, F09) was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas AM University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.