On Lattice Sequential Decoding for The Unconstrained AWGN Channel

Walid Abediseid and Mohamed-Slim Alouini

Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division
King Abdullah University of Science & Technology (KAUST)
Thuwal, Makkah Province, Saudi Arabia
{walid.abediseid,slim.alouini}@kaust.edu.sa

***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.**

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 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]–[7].

Poltyrev [3] studied the problem of coding for the unconstrained additive white Gaussian noise (AWGN) channel where the channel input being an infinite lattice. In his setting, the notion of capacity becomes meaning less 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

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

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 decreasing, positive function for all $\mu_c > 1$. The parameter μ_c, which is called the *volume-to-noise ratio* (will 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 contrast to Poltyrev, where random coding has been used to prove the above result, Loeliger [4] showed 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 [8], [9].

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 [10], [11] 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 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 to be expected [10]. As an alternative to sphere decoding

[1]Sphere decoding algorithms were originally implemented to decode signals transmitted via wireless fading channels [11], 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 [10]). 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.

[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 be transmitted. Here, for infinite lattice codes, the notion of volume-to-noise ratio is used instead which will be introduced in the sequel.

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] [12], [13].

The stack algorithm is a well-known algorithm that is used to describe the operation of the sequential decoder [13]. 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 [12] 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.* [15] where sequential decoding is used to decode lattice codes with finite trellis diagram. Shalvi *et. al.* in [16] 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 of 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.

The contribution of this paper can be classified into two classes: the performance analysis of the lattice sequential decoder in terms of the achievable error exponent, and the computational complexity of the decoder in terms of the "cut-off" volume-to-noise ratio via the complexity tail distribution. Those 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., 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.

Through out 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 $\Gamma(n) = \int_0^\infty x^{m-1}e^{-x}\,dx$, is the Gamma function. Bold lowercase letters $\boldsymbol{a}$ denote vectors, whose l_2-norm is denoted by $\|\boldsymbol{a}\|$, and bold uppercase letters $\boldsymbol{A}$ denote matrices, where $\boldsymbol{I}_m$ denotes the $m \times m$ identity matrix. 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:

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

[3]The work in [14] 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.

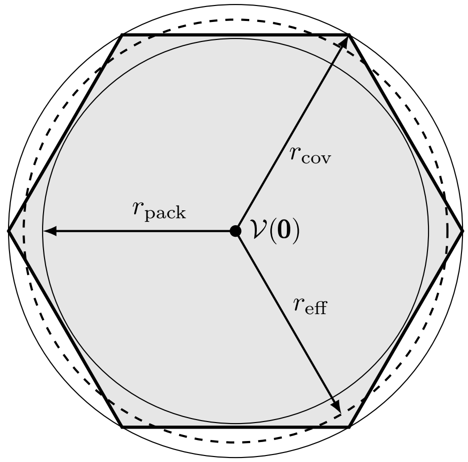

Fig. 1. The packing radius, the effective radius, and the covering radius of the hexagonal lattice.

with the property that $V(a\boldsymbol{G}) = a^m V(\boldsymbol{G})$ for any $a > 0$.

- The covering radius $r_{\rm cov}(\Lambda)$ is the radius of the smallest sphere centered at the origin that contains $\mathcal{V}(\mathbf{0})$. The effective radius $r_{\rm eff}(\Lambda)$ is the radius of the sphere with volume equal to $V(\boldsymbol{G})$. The packing radius $r_{\rm pack}(\Lambda)$ is the radius of the largest sphere centered at the origin inside the Voronoi cell $\mathcal{V}(\mathbf{0})$ (see Fig. 1).
- A sequence of lattices $\{\Lambda_m\}$ of increasing dimension is good for covering [6] if $r_{\rm cov}(\Lambda_m) \rightarrow r_{\rm eff}(\Lambda_m)$.
- **Minkowski-Hlawka Theorem** [8]: Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ 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 $\{\Lambda\}$ of lattices with volume $V(\boldsymbol{G})$ and dimension m such that

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

where the expectation $\mathcal{E}_\Lambda$ is taken over the ensemble of random lattices, $\Lambda^* = \Lambda \backslash \{\mathbf{0}\}$, and $\delta \rightarrow 0$ as $m \rightarrow \infty$. 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 [9].

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 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

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}, \quad (3)$$

where $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_m)$. 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

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

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

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

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

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

From the above analysis, one can notice 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 $1/2 \log(\text{SNR})$ is achievable, where SNR ≥ 0 here represents the average signal-to-noise ratio of the channel[4]. For high SNRs (i.e., for 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

[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 $m \rightarrow \infty$) given by $(2\pi e P_x)^{m/2}$. 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.

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 [17]. 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 [14]

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

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 well performed using sequential decoders based on the stack algorithm [14]. For our channel model, the path metric given by (7) can be shown to be simplified to (see Appendix A in [14])

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

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 $\boldsymbol{G}_c = \boldsymbol{QR}$, $\boldsymbol{y}' = \boldsymbol{Q}^{\mathsf{T}}\boldsymbol{y}$, and b is the bias term.

The stack sequential decoder is an efficient tree search algorithm that attempts to find a "best fit" with the received noisy signal. As in the conventional stack decoder [13], to determine a best fit (path), values are assigned to each node on the tree. This value is called the metric which is given by (8). 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 determination of the best and next best nodes is simplified in the closest lattice point search problem by using the Schnnor-Euchner enumeration [10] which generates nodes with metrics in ascending order given any node $\boldsymbol{z}_1^k$. The decoding algorithm terminates when the top path in the stack reaches the end of the tree (refer to [13] for more details about the algorithm).

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 details in the subsequent sections.

For the stack algorithm, in order to help distinguish the correct path from all incorrect paths, the value of b should be chosen carefully such that the average metric along the correct path increases, while decreases, on average, along the incorrect path. Assuming the decoder is following the correct path, the average metric at level k is given by

$$\mathcal{E}\{\mu(\boldsymbol{z}_1^k)|\ \boldsymbol{z}_1^k \text{ correct path }\} = (b-\sigma^2)k,$$

which is positive for $b > \sigma^2$.

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. For a given lattice Λ_c, we have

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

where (a) is due to the fact that in general, $\mu_{\min} = \min\{0, b - \|\boldsymbol{w}_1^1\|^2, 2b - \|\boldsymbol{w}_1^2\|^2, \ldots, bm - \|\boldsymbol{w}_1^m\|^2\}$ is the minimum metric that corresponds to the transmitted path, $\mu(\boldsymbol{z}) > \mu_{\min}$ is just a necessary condition for $\boldsymbol{x} = \boldsymbol{Gz}$ to be decoded by the stack decoder, (b) follows by noticing that $-(\mu_{\min} + \|\boldsymbol{w}\|^2) \leq 0$, and (c) follows from the fact that $\|\boldsymbol{x}\| \geq \min_{\boldsymbol{x}\in\Lambda_c^*} \|\boldsymbol{x}\| \triangleq d_{\min}$ (the shortest vector in the lattice).

It is clear from the above analysis that the sequential decoder approaches the performance of the lattice decoder as $b \to 0$. Now, as $m \to \infty$, for a well-constructed lattice code ensemble[5], it is well-known that

$$\frac{d_{\min}^2}{m} \sim \frac{V_c^{2/m}}{2\pi e} = \mu_c \sigma^2.$$

Therefore, the probability of decoding error can by (asymptotically) upper bounded by

$$\begin{aligned} P_e(b|\Lambda_c) &\leq \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2\left(1-\frac{b/\sigma^2}{\mu_c}\right)\right\}\right) \\ &= \Pr\left(\bigcup_{\boldsymbol{x}\in\Lambda_c^*}\left\{2\boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{w}} > \|\boldsymbol{x}\|^2\right\}\right), \end{aligned} \tag{10}$$

where

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

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

Interestingly, the upper bound (10) 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 (averaged over the ensemble of linear lattice codes Λ_c) can be upper bounded by

$$\overline{P_e}(b) \triangleq \mathcal{E}_{\Lambda_c}\{P_e(b|\Lambda_c\}) \leq e^{-mE_b(\mu_c)}, \tag{11}$$

where

$$E_b(\mu_c) = E_p(\tilde{\mu}_c) = \begin{cases} 0, & \tilde{\mu}_c \leq 1; \\ \frac{1}{2}\left[(\tilde{\mu}_c - 1) - \log\tilde{\mu}_c\right], & 1 < \tilde{\mu}_c \leq 2; \\ \frac{1}{2}\log\left(\frac{e\tilde{\mu}_c}{4}\right), & 2 \leq \tilde{\mu}_c \leq 4; \\ \tilde{\mu}_c/8, & \tilde{\mu}_c \geq 4. \end{cases} \tag{12}$$

where

$$\tilde{\mu}_c \triangleq \frac{V(\boldsymbol{G}_c)^{2/m}/2\pi e}{\sigma'^2} = \mu_c\left(1-\frac{\mu_b}{\mu_c}\right)^2 \tag{13}$$

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

$$P_e(b, \Lambda_c^*) \leq e^{-mE_b(\mu_c)}. \tag{14}$$

[5]Codes which are constructed using lattices that satisfy the Minkowski-Hlawka theorem (see [3]–[6] for more details).

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 (13) as

$$\tilde{\mu}_c = \mu_c\left(1-\frac{\mu_b}{\mu_c}\right)^2 \approx \mu_c\left(1-2\frac{\mu_b}{\mu_c}\right) = \mu_c - 2\mu_b. \tag{15}$$

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

$$E_b(\mu_c) = \frac{\tilde{\mu}_c}{8} = \frac{\mu_c}{8} - \frac{\mu_b}{4} = E_p(\mu_c) - \frac{\mu_b}{4}, \tag{16}$$

where $E_p(\mu_c)$ is the Poltyrev error exponent achieved by the lattice decoder which is defined in (6). Substituting (16) into (14) we get

$$\begin{aligned} P_e(b) \leq e^{-mE_b(\mu_c)} &= e^{-m[E_p(\mu_c)-\mu_b/4]} \\ &= \alpha e^{-mE_p(\mu_c)} \end{aligned} \tag{17}$$

where $\alpha = e^{m\mu_b/4}$. The constant α sheds some lights on the behavior of the error probability of the sequential decoder for 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[6], as will be shown in the sequel.

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

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

It is clear from the above analysis that if $\delta = 1$ ($\mu_b = 0$) then the performance of the sequential decoder approaches the performance of the lattice decoder. On the other extreme, if $\delta = 0$ ($\mu_b = \mu_c$) then reliable communication may not be possible under lattice sequential decoding. Fig. 2 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. 2 that for high VNR μ_c, the affect of varying δ occurs as a change in the *slope* of the error exponent curve, where at high

[6]It must be noted that, although the bound (14) shows that the shift is $\alpha = e^{m\mu_b/4}$, the exact amount of right-shift is less than α as will be shown by the simulation results in Section VI.

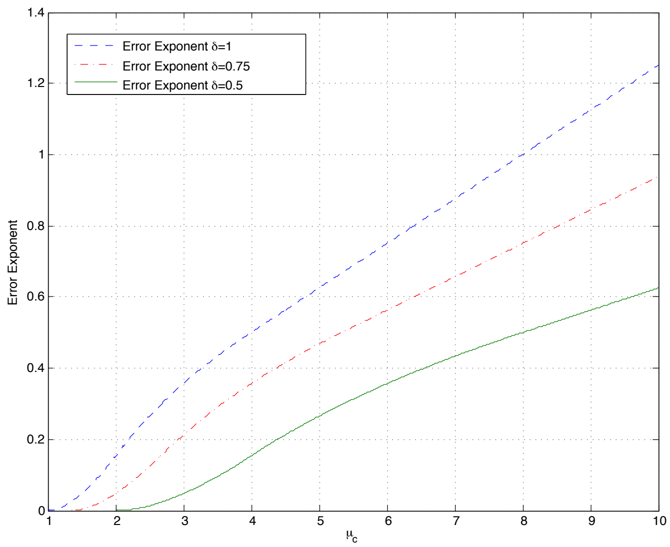

Fig. 2. The achievable error exponent of the lattice sequential decoder when the normalized bias term $\mu_b = (1-\sqrt{\delta})\mu_c$ for $\delta = 0.5, 0.75,$ and 1.

VNR we have $E_b(\mu_c) = \delta\mu_c/8$. Moreover, the achievable VNR under sequential decoding with normalized bias $\mu_b = (1-\sqrt{\delta})\mu_c$ is given by $1/\delta$. Therefore, for $\delta \neq 1$, reliable communication may not be possible at VNR close to capacity ($\mu_c = 1$).

Next, we consider the computational complexity analysis of the sequential decoder for the unconstrained AWGN channel. As was discussed above, increasing the bias term increases the decoding error probability as well. 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.

V. THE "CUT-OFF" VOLUME-TO-NOISE RATIO

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 decoder, a cut-off rate has been defined for such decoding scheme. 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 the complexity of the sequential decoder increases exponentially with the constraint length of the code [12]. For the power-constrained AWGN channel, the cut-off rate R_0 is $4/e$ (1.68 dB) away from capacity at high SNR (see [17] and references therein).

For the lattice sequential decoder, the analysis of the computational complexity is considered a difficult problem. As a start to solve this problem for the the unconstrained AWGN channel, we would like to find a lower bound on the "cut-off" VNR μ_0 — the value of μ_c for which low decoding complexity is possible. Such lower bound appears to be analytically tractable by bounding the *average* decoding complexity, which is defined here as the *total number of nodes in the tree that have been visited by the decoder during the search.*

A. Average Computational Complexity: A Lower Bound

First, one should note that all nodes in the tree that are visited by the sequential decoder prior decoding the message must have partial path metrics $\mu(\boldsymbol{z}_1^k)$ that exceed the minimum metric $\mu_{\min}$ that corresponds to the decoded path. Let $\phi(\boldsymbol{z}_1^k)$ be the indicator function defined by

$$\phi(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \mu(\boldsymbol{z}_1^k) \geq \mu_{\min}; \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

and let $\mathcal{N}(\Lambda_c)$ be a random variable that denotes the total number of visited nodes during the search for a given lattice code Λ_c. Then, $\mathcal{N}(\Lambda_c)$ can be expressed as

$$\mathcal{N}(\Lambda_c) = \sum_{k=1}^{m} \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k). \tag{20}$$

Similar to the performance analysis, due to lattice symmetry, we assume that the all-zero lattice point $\mathbf{0}$ was transmitted. It is clear that

$$\mathcal{N}(\Lambda_c) \geq \sum_{\boldsymbol{z}_1^m \in \mathbb{Z}^m} \phi(\boldsymbol{z}_1^m).$$

Now, since $\|\boldsymbol{w}_1^k\|^2 > 0$ for any $1 \leq k \leq m$, then

$$\begin{aligned} \mu_{\min} &= \min\{0, b - \|\boldsymbol{w}_1^1\|^2, 2b - \|\boldsymbol{w}_1^2\|^2, \ldots, mb - \|\boldsymbol{w}_1^m\|^2\} \\ &\leq \min\{0, b-0, 2b-0, \cdots, mb-0\} = 0. \end{aligned}$$

In this case, the average number of computations (averaged over the statistics of the noise) can be upper bounded as

$$\mathcal{E}_{\boldsymbol{w}}\{\mathcal{N}(\Lambda_c)\} \geq \mathcal{E}_{\boldsymbol{w}}\left\{ \sum_{\boldsymbol{x}\in\Lambda_c} \vartheta_{\boldsymbol{w}}(\boldsymbol{x}) \right\}, \tag{21}$$

where

$$\vartheta_{\boldsymbol{w}}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \|\boldsymbol{w}-\boldsymbol{x}\|^2 \leq bk; \\ 0, & \text{otherwise.} \end{cases}$$

We can further lower bound (21) as follows:

$$\begin{aligned} \mathcal{E}_{\boldsymbol{w}}\left\{ \sum_{\boldsymbol{x}\in\Lambda_c} \vartheta_{\boldsymbol{w}}(\boldsymbol{x}) \right\} &= \sum_{\boldsymbol{x}\in\Lambda_c} \int_{\boldsymbol{w}\in\mathbb{R}^m} \frac{e^{-\|\boldsymbol{w}\|^2/2\sigma^2}}{(2\pi\sigma^2)^{m/2}} \vartheta_{\boldsymbol{w}}(\boldsymbol{x})\, d\boldsymbol{w} \\ &\geq \sum_{\boldsymbol{x}\in\Lambda_c} \int_{\|\boldsymbol{w}\|^2 \leq r_{\rm cov}^2} \frac{e^{-\|\boldsymbol{w}\|^2/2\sigma^2}}{(2\pi\sigma^2)^{m/2}} \vartheta_{\boldsymbol{w}}(\boldsymbol{x})\, d\boldsymbol{w} \\ &\geq \frac{e^{-r_{\rm cov}^2/2\sigma^2}}{(2\pi\sigma^2)^{m/2}} \sum_{\boldsymbol{x}\in\Lambda_c} \int_{\boldsymbol{w}\in\mathcal{V}(\mathbf{0})} \vartheta_{\boldsymbol{w}}(\boldsymbol{x})\, d\boldsymbol{w} \\ &= \frac{e^{-r_{\rm cov}^2/2\sigma^2}}{(2\pi\sigma^2)^{m/2}} V(\mathcal{S}_m(\sqrt{bk})). \end{aligned} \tag{22}$$

The above lower bound is valid for any lattice code Λ_c. Assuming the use of a well-constructed lattice (e.g., Loeliger ensemble [4]), we have that for large m (i.e., as $m \to \infty$), $r_{\rm cov}^2 \to r_{\rm eff}^2 = (V_c^{2/m}/2\pi e)m = \mu_c\sigma^2 m$ and $V(\mathcal{S}_m(\sqrt{bk})) \to$

$(2\pi ebk/m)^{m/2}$. Therefore, asymptotically, the average number of computations can be lower bound as

$$\begin{aligned}\mathcal{E}_{\boldsymbol{w}}\{\mathcal{N}(\Lambda_c)\} &\geq \frac{e^{-\mu_c m/2}}{(2\pi\sigma^2)^{m/2}}(2\pi eb)^{m/2} \\ &\geq \left(\mu_b e^{1-\mu_c}\right)^{m/2}. \end{aligned} \quad (23)$$

Thus, at $\mu_b e^{1-\mu_c} > 1$, i.e., at VNR $\mu_c < \log_e(\mu_b e)$, the average number of computations performed by the sequential decoder must go to infinity *exponentially* as m is increased. Therefore, in order to ensure a very low decoding complexity as the bias b increases, the cut-off VNR μ_0 should be chosen to satisfy the following lower bound

$$\mu_0 \geq 1 + \log_e(\mu_b). \quad (24)$$

As argued in section III, a low decoding complexity is to be expected as long as $\mu_b = b/\sigma^2 > 1$. Since $\mu_0 \geq 1 + \log_e(\mu_b)$, the cut-off VNR is *strictly* greater than 1, where $\mu_c = 1$ is the capacity of the unconstrained AWGN channel. Therefore, one has to stay away from capacity if low decoding complexity is to be expected. The above lower bound on μ_0 can be further improved by considering the decoding complexity tail distribution as will be shown in the sequel.

B. An Upper Bound on The Complexity Distribution

Although sequential decoding algorithms are simple to describe, the analysis of the decoder's computational complexity is considered difficult. This is due to the fact that the amount of computations performed by the decoder attempting to decode a message is random. Therefore, the complexity of such a decoder is usually analyzed through its *computational distribution* defined by $\Pr(\mathcal{N}(\Lambda_c) \geq L)$, where L is the distribution parameter. In this section we would like to find an upper bound on the decoding complexity distribution and to find the values of μ_0 that lead to low decoding complexity. Consider now the following analysis:

First, the complexity tail distribution can be upper bounded as

$$\begin{aligned}\Pr(\mathcal{N}(\Lambda_c) \geq L) \leq &\Pr(\mathcal{N}(\Lambda_c) \geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m) + \\ &\Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m), \end{aligned} \quad (25)$$

where the above upper bound is derived using the well-known separation of the typical noise events from the non-typical ones [18]. Next, we would like to upper bound the second term in the RHS of (25).

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

$$\sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi(\boldsymbol{z}_1^k) \leq \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \quad (26)$$

where $\phi(\boldsymbol{z}_1^k)$ is the indicator function defined in (19), and

$$\phi^{'}(\boldsymbol{z}_1^k) = \begin{cases} 1, & \text{if } \|\boldsymbol{w'}_1^k - \boldsymbol{R}_{kk}\boldsymbol{z}_1^k\|^2 \leq bk + \sigma^2 m; \\ 0, & \text{otherwise.} \end{cases} \quad (27)$$

Now, let

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

where

$$S_k = \sum_{\boldsymbol{z}_1^k \in \mathbb{Z}^k} \phi^{'}(\boldsymbol{z}_1^k), \quad (28)$$

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 (28) 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 m-dimensional sphere of squared radius $bk + \sigma^2 m$ to the volume of the Voronoi cell of the lattice as (see [9] for more details)

$$S_k \approx \frac{(16\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + \sigma^2 m]^{k/2}}{(\det(\boldsymbol{R}_{kk}^{\mathsf{T}}\boldsymbol{R}_{kk})^{1/2})}. \quad (29)$$

For a given lattice Λ_c, we have

$$\begin{aligned}&\Pr(\mathcal{N}(\Lambda_c) \geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m|\Lambda_c) \\ &\quad \leq \Pr(\mathcal{N}(\Lambda_c) \geq L - m, \|\boldsymbol{w}\|^2 \leq \sigma^2 m|\Lambda_c) \\ &\quad \leq \frac{\mathcal{E}_{\boldsymbol{w}}\{\tilde{\mathcal{N}}(\Lambda_c)|\Lambda_c, \|\boldsymbol{w}\|^2 \leq \sigma^2 m\}}{L - m}, \quad \text{for } L > m, \end{aligned} \quad (30)$$

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

$$\begin{aligned}&\mathcal{E}_{\boldsymbol{w}}\{\tilde{\mathcal{N}}(\Lambda_c)|\Lambda_c, \|\boldsymbol{w}\|^2 \leq \sigma^2 m\} \\ &\quad \leq \sum_{k=1}^{m} S_k \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(\|\boldsymbol{x}\|^2 - 2(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{w} < bm). \end{aligned} \quad (31)$$

Therefore, we have

$$\begin{aligned}&\Pr(\mathcal{N}(\Lambda_c) \geq L, \|\boldsymbol{w}\|^2 \leq \sigma^2 m|\Lambda_c) \\ &\quad \leq \frac{\sum_{k=1}^{m} S_k}{L - m} \sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(2(\boldsymbol{x})^{\mathsf{T}}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm). \end{aligned} \quad (32)$$

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

$$\Pr(\mathcal{N}(\Lambda_c) \geq L|\Lambda_c) \leq \sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm) + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m) \tag{33}$$

Using Chernoff bound,

$$\Pr\left(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} \geq \|\boldsymbol{x}\|^2 - bm\right) \leq \begin{cases} e^{-\|\boldsymbol{x}\|^2/8\sigma^2} e^{bm/4\sigma^2}, & \|\boldsymbol{x}\|^2 > bm; \\ 1, & \|\boldsymbol{x}\|^2 \leq bm, \end{cases} \tag{34}$$

Let

$$f(\boldsymbol{x}) = \begin{cases} 0, & \|\boldsymbol{x}\|^2 > bm; \\ 1, & \|\boldsymbol{x}\|^2 \leq bm, \end{cases}$$

and

$$g(\boldsymbol{x}) = \begin{cases} e^{-\|\boldsymbol{x}\|^2/8\sigma^2} e^{bm/4\sigma^2}, & \|\boldsymbol{x}\|^2 > bm; \\ 0, & \|\boldsymbol{x}\|^2 \leq bm. \end{cases}$$

Therefore, one may upper bound the first term in the RHS of (33) as

$$\sum_{\boldsymbol{x}\in\Lambda_c^*} \Pr(2(\boldsymbol{x})^\mathsf{T}\boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm) \leq \sum_{\boldsymbol{x}\in\Lambda_c^*} f(\boldsymbol{x}) + \sum_{\boldsymbol{x}\in\Lambda_c^*} g(\boldsymbol{x}). \tag{35}$$

By taking the expectation of (33) over the ensemble average of random lattices, we obtain

$$\begin{aligned} \overline{\Pr(\mathcal{N}(\Lambda_c) \geq L)} &= \mathcal{E}_{\Lambda_c}\{\Pr(\mathcal{N}(\Lambda_c) \geq L|\Lambda_c)\} \\ &\leq \frac{1}{V_c}\int_{\mathbb{R}^m} f(\boldsymbol{x})\, d\boldsymbol{x} + \frac{1}{V_c}\int_{\mathbb{R}^m} g(\boldsymbol{x})\, d\boldsymbol{x} + \\ &\quad \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m). \end{aligned} \tag{36}$$

Evaluating the integrals in the above upper bound we get

$$\begin{aligned} \frac{1}{V_c}\int_{\mathbb{R}^m} f(\boldsymbol{x})\, d\boldsymbol{x} &= \frac{\pi^{m/2}(bm)^{m/2}}{V_c\Gamma(m/2+1)} \\ &\sim \frac{\pi^{m/2}(bm)^{m/2}}{V_c\left(\frac{m}{2e}\right)^{m/2}} = \left(\frac{\mu_b}{\mu_c}\right)^{m/2}, \end{aligned} \tag{37}$$

and

$$\begin{aligned} \frac{1}{V_c}\int_{\mathbb{R}^m} g(\boldsymbol{x})\, d\boldsymbol{x} &= \frac{e^{bm/4\sigma^2}}{V_c} \int_{\|\boldsymbol{x}\|^2 > bm} e^{-\|\boldsymbol{x}\|^2/8\sigma^2}\, d\boldsymbol{x} \\ &= \left(\frac{4e^{\mu_b/2-1}}{\mu_c}\right)^{m/2} \Pr(\|\boldsymbol{\nu}\|^2 > bm), \end{aligned} \tag{38}$$

where $\boldsymbol{\nu} \sim \mathcal{N}(0, 4\sigma^2\boldsymbol{I})$. Now, the probability $\Pr(\|\boldsymbol{\nu}\|^2 > bm)$ can be tightly upper bounded as

$$\begin{aligned} \Pr(\|\boldsymbol{\nu}\|^2 > bm) &\leq \begin{cases} \left(\frac{be}{4\sigma^2}\right)^{m/2} e^{-bm/8\sigma^2}, & b > 4\sigma^2; \\ 1, & b \leq 4\sigma^2, \end{cases} \\ &= \begin{cases} \left(\frac{\mu_b e}{4}\right)^{m/2} e^{-\mu_b m/8}, & \mu_b > 4; \\ 1, & \mu_b \leq 4. \end{cases} \end{aligned} \tag{39}$$

Therefore, for large m, we can further upper bound (36) as follows:

For $0 \leq \mu_b \leq 4$:

$$\overline{\Pr(\mathcal{N}(\Lambda_c) \geq L)} \leq \left(\frac{\mu_b}{\mu_c}\right)^{m/2} + \left(\frac{4e^{\mu_b/2-1}}{\mu_c}\right)^{m/2} + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m). \tag{40}$$

For $\mu_b > 4$:

$$\overline{\Pr(\mathcal{N}(\Lambda_c) \geq L)} \leq \left(\frac{\mu_b}{\mu_c}\right)^{m/2} + \left(\frac{\mu_b e^{\mu_b/4}}{\mu_c}\right)^{m/2} + \Pr(\|\boldsymbol{w}\|^2 > \sigma^2 m). \tag{41}$$

Hence, for sufficiently large m, there exists at least a lattice Λ_c^* in the ensemble with complexity tail distribution satisfying both (40) and (41) for the corresponding normalized bias term μ_b for all values of $L = m + \sum_{k=1}^{m} S_k$.

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 two terms in the upper bound (40) and (41) can be made smaller than $\epsilon/2$ for sufficiently large m, i.e.,

$$\Pr(\mathcal{N}(\Lambda_c^*) \geq L) \leq \epsilon, \quad m \to \infty,$$

if $\mu_c \geq 4e^{\mu_b/2-1}$ for all $0 \leq \mu_b < 4$, and if $\mu_c \geq \mu_b e^{\mu_b/4}$ when $\mu_b \geq 4$, respectively. This result indicates that large computational complexity may be avoided at μ_c below the cut-off VNR μ_0 that is given by

$$\mu_0 = \begin{cases} 4e^{\mu_b/2-1}, & \text{for } 0 \leq \mu_b \leq 4; \\ \mu_b e^{\mu_b/4}, & \text{for } \mu_b > 4. \end{cases} \tag{42}$$

In this case, the total number of computations performed by the decoder can be approximated by

$$L \approx m + \sum_{k=1}^{m} \frac{(16\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[bk + \sigma^2 m]^{k/2}}{(\det(\boldsymbol{R}_{kk}^\mathsf{T}\boldsymbol{R}_{kk})^{1/2})}. \tag{43}$$

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 $\boldsymbol{G}_c = \boldsymbol{QR}$, 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

$$L \approx m + \sum_{k=1}^{m} \frac{(16\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[\mu_b k + m]^{k/2}}{\mu_c^{k/2} (\det(\boldsymbol{R}_{kk}^{\mathsf{T}} \boldsymbol{R}_{kk})^{1/2})}. \tag{44}$$

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.

To see now how the decoding complexity behaves as we increase the bias term, we select the VNR $\mu_c > \mu_0 > \mu_b e^{\mu_b/4}$ when $\mu_b > 4$. Then, one may upper bound the total number of computations performed by the decoder as

$$L \leq m + \sum_{k=1}^{m} \frac{(16\pi)^{k/2}}{\Gamma(k/2+1)} \frac{[\mu_b k + m]^{k/2}}{\mu_b^{k/2} e^{\mu_b k/8} (\det(\boldsymbol{R}_{kk}^{\mathsf{T}} \boldsymbol{R}_{kk})^{1/2})}. \tag{45}$$

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

By taking a closer look at the upper bound (32) one can note that the term

$$\sum_{\boldsymbol{x} \in \Lambda_c^*} \Pr(2(\boldsymbol{x})^{\mathsf{T}} \boldsymbol{w} > \|\boldsymbol{x}\|^2 - bm)$$

represents an upper bound to the decoding error probability (9) derived using the *union bound*. It is well-known that the union bound can provide 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 [17]). Therefore, achieving a good error performance for large values of b, where low decoding complexity is to be expected, comes at the expense of increasing the VNR (or equivalently reducing the coding rate for the case of finite lattice codes).

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 [17]. 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

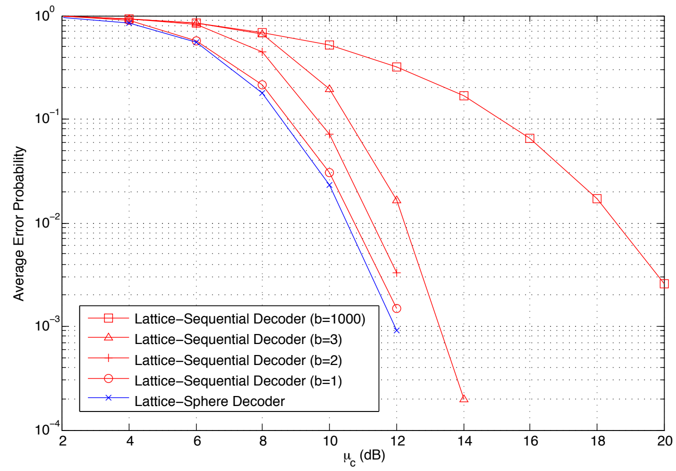

Fig. 3. Comparison of the lattice sequential decoder's performance for various values of (fixed) bias term.

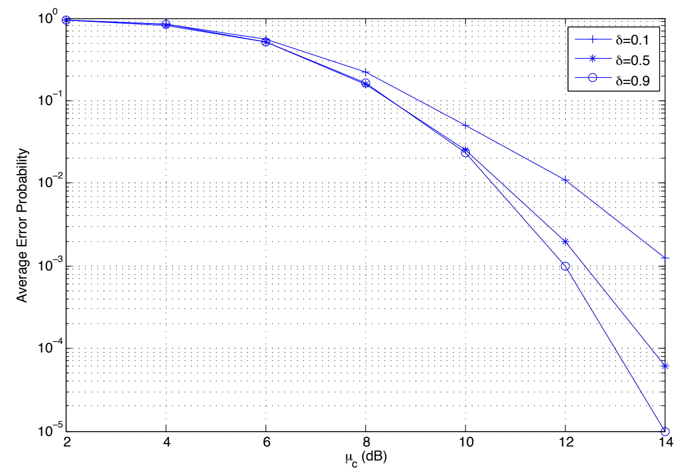

Fig. 4. Comparison of the lattice sequential decoder's performance when the bias term varies with the VNR as $\mu_b = (1 - \sqrt{\delta})\mu_c$ for several values of δ.

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 $m = 10$ and $p = 47$ and fix the code for all simulation results.

Fig. 3 and Fig. 4 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. 3, 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 [10] to measure the price of using the sequential decoder in terms of the performance

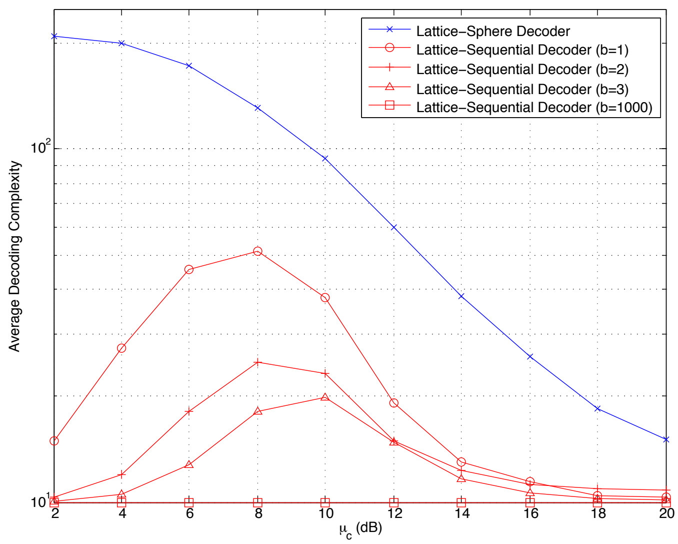

Fig. 5. The average computational complexity achieved by the sequential decoder for different values of the bias term.

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,[7] particularly at high VNR values. This basically agrees with the derived theoretical results provided in (16) and (17). On the other hand, if we let b to scale linearly with VNR as $(1-\sqrt{\delta})\mu_c$, where $0 \leq \delta \leq 1$ (see (18)), then according to the error exponent analysis, we expect that the rate of decay (slope) of the error probability curve would decrease[8] as we decrease δ. This is depicted in Fig. 4, which also agrees with the derived theoretical results.

Finally, Fig. 5 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. 5 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. 5. 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.

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. Poltyerv, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 409-417, March 1994.

[4] H. Loeliger, "Averaging bounds for lattices and linear codes", *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1767-11773, 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] 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.

[7] L.-C. Choo, C. Ling, and K.-K. Wong, "Achievable rates for lattice coding over the Gaussian wiretap channel," in Proceeding *IEEE Physical Layer Security Workshop in Conjunction with IEEE International Communication Conference (ICC'11)*, Kyoto, Japan, June 2011.

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

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

[7] 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 (17) we get

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

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

[8] 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.

[10] B. Hassibi and H. Vikalo, “On sphere decoding algorithm. Part I: expected complexity,” *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2389-2401, Aug. 2005.

[11] J. Boutros, N. Gresset, L. Brunel, and M. Fossorier, “Soft-input soft-output lattice sphere decoder for linear channels,” in Proceeding *IEEE Global Communications Conference (GLOBECOM03)*, San Francisco, USA, Dec. 2003.

[12] 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.

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

[14] 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, March 2005.

[15] V. Tarokh, A. Vardy, and K. Zeger, “Sequential decoding of lattice codes”, in Proceeding *IEEE International Symp. of Inform.. Theory* (ISIT’97), Ulm, Germany, June 1997.

[16] 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.

[17] 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.

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