On the Complexity of Decoding Lattices Using the Korkin–Zolotarev Reduced Basis

Amir H. Banihashemi, *Member, IEEE*, and Amir K. Khandani, *Member, IEEE*

***Abstract*—Upper and lower bounds are derived for the decoding complexity of a general lattice L. The bounds are in terms of the dimension n and the coding gain γ of L, and are obtained based on a decoding algorithm which is an improved version of Kannan's method. The latter is currently the fastest known method for the decoding of a general lattice. For the decoding of a point $\boldsymbol{x}$, the proposed algorithm recursively searches inside an n-dimensional rectangular parallelepiped (cube), centered at $\boldsymbol{x}$, with its edges along the Gram–Schmidt vectors of a proper basis of L. We call algorithms of this type recursive cube search (RCS) algorithms. It is shown that Kannan's algorithm also belongs to this category. The complexity of RCS algorithms is measured in terms of the number of lattice points that need to be examined before a decision is made. To tighten the upper bound on the complexity, we select a lattice basis which is reduced in the sense of Korkin–Zolotarev. It is shown that for any selected basis, the decoding complexity (using RCS algorithms) of any sequence of lattices with possible application in communications $(\gamma \geq 1)$ grows at least exponentially with n and γ. It is observed that the densest lattices, and almost all of the lattices used in communications, e.g., Barnes–Wall lattices and the Leech lattice, have equal successive minima (ESM). For the decoding complexity of ESM lattices, a tighter upper bound and a stronger lower bound result are derived.**

***Index Terms*—Coding gain, decoding algorithms, decoding complexity, densest lattices, Korkin–Zolotarev reduction, lattices, successive minima.**

I. Introduction

LATTICES have two main applications in communications: i) efficient signaling over band-limited channels, and ii) vector quantization. In both applications, a finite subset of points of an n-dimensional (n-D) lattice within a bounded supporting region of $\mathbb{R}^n$ is employed. This collection of points is called a *lattice code*.

The major complexity associated with a lattice code is the process of *decoding*, that is, finding the point of the code that has the smallest (Euclidean) distance to an input. Note that as the number of code points is usually a huge number, one cannot use an exhaustive method to implement the decoding operation. Presenting an efficient decoding algorithm for a general lattice is one of the purposes of this paper.

There exist very efficient algorithms for the decoding of well-known lattices with high degree of structure, like the Leech lattice (see [7, pp. 443–448], [10], [29]). Most of these algorithms, however, cannot be applied to a general lattice. There are only two known general-purpose methods to decode a lattice: the trellis approach and the integer programming approach based on the geometry of numbers. The purpose of this paper, which presents part of the results obtained in [3], is to study and analyze the latter approach.

The trellis approach, mainly due to the valuable contributions of Forney [9]–[11], is currently one of the common methods in communications for the decoding of lattices. This approach, which can be applied to any lattice with a finite trellis (including rational lattices), is based on representing the lattice by a trellis diagram which reflects the underlying group structure. Then the Viterbi algorithm [8] is used to decode the trellis.

The trellis structure of lattices and their trellis complexity have been the subject of some recent research [3]–[5], [10], [25]–[27]. In [10], Forney derived lower bounds on the state complexity of the trellis diagrams of lattices, and constructed trellises for some important low-dimensional lattices which either met or nearly met these lower bounds. Subsequently, Tarokh and Blake [25], [26] gave lower bounds which show that for sufficiently large coding gains γ, the average state and edge complexities of any trellis diagram of lattices grow at least exponentially with γ.

Upper bounds on the complexity were derived in [27]. In [3] and [4], these bounds were both improved and generalized, lower bounds on the number of distinct paths in trellis diagrams of lattices were derived, and low-complexity trellises were constructed for some important lattices which either achieved or nearly achieved the lower bounds. The trellis complexity of root lattices and their duals was then investigated to some extent in [3] and [5]. Other relevant results about trellis structure and trellis complexity of block codes can be found in papers in [30], and the references therein.

The problem of lattice decoding also lies at the heart of many integer programming problems [2], [13]–[16]. The main approach to the decoding of lattices in integer programming is based on using a reduced basis for the lattice. The complexity of such decoding algorithms has two parts: i) computing the reduced basis of the lattice, and ii) finding the nearest lattice

Manuscript received December 18, 1995; revised May 27, 1997. This work was supported in part by an Ontario Graduate Scholarship (OGS) and in part by the Information Technology Research Centre of Canada (ITRC). The material in this paper was presented in part at the 39th Annual Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1996, and at the 18th Biennial Symposium on Communications, Queen's University, Kingston, Ont., Canada, June 1996.

A. H. Banihashemi was with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1. He is now with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont., Canada M5S 1A4.

A. K. Khandani is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

Publisher Item Identifier S 0018-9448(98)00009-1.

point using this reduced basis. In the decoding problems encountered in communications, the lattice is fixed, so the basis reduction is performed just once and then the resulting basis is stored for subsequent use. Thus the complexity of solving i) is not of major concern. The fastest lattice decoding algorithm for solving ii) in the context of integer programming appears to be that of Kannan [16].

This work gives a geometrical interpretation of Kannan's algorithm, which clarifies some issues regarding the complexity of the algorithm. Explicit upper and lower bounds on the complexity of Kannan's algorithm for a general lattice are derived. The bounds are in terms of the coding gain and the dimension of the lattice. For lattices with equal successive minima (ESM), a tighter upper bound and a stronger lower bound are obtained. Recalling that extremal lattices (including the densest lattices) belong to the category of ESM lattices, we observe that almost all of the lattices used in channel coding have ESM. It is also proved that lattices A_n^*, D_n^*, and E_n^* have ESM. This means that the lattices used for quantization of uniformly distributed inputs [7, p. 61] are also ESM lattices.

To reduce complexity, we then modify Kannan's algorithm. By pre-computing the covering radii of the lattice and its sublattices, decoding is simplified, especially for the lattices used in communication applications. The modified algorithm employs the Korkin–Zolotarev reduced basis, and solves the decoding problem for an n-D lattice by reducing it to some subproblems of dimensionality $n-1$. Explicit upper and lower bounds on the complexity of the algorithm are derived. Improved complexity results are also obtained for ESM lattices. Using the derived lower bound, it is shown that even with some exponential-time pre-computations (computing the reduced basis and the covering radii), one cannot decode any sequence of lattices with possible application in communications ($\gamma \geq 1$) in polynomial time. The lower bound also indicates that our upper bound results cannot be much improved.

This paper concentrates on the lattices used in signal constellations. However, the problems of lattice-based channel coding and lattice-based vector quantization are closely related. The decoding algorithm discussed here can be used in both of these contexts.

This article is organized as follows. Section II gives an introduction to lattices. Section III explains the concept of coding gain, and also deals with some of the important known ESM lattices. Section IV gives an introduction to the idea of basis reduction, and discusses the Korkin–Zolotarev (K-Z) reduced basis. Section V presents the proposed decoding algorithm and discusses its complexity. Kannan's algorithm is explained and bounds on its complexity are derived. Finally, Section VI contains concluding remarks.

II. Some Definitions and Facts About Lattices

Let $\mathbb{R}^m$ be the m-dimensional real vector space with the standard inner product $\langle .,. \rangle$, and Euclidean length $||\boldsymbol{x}|| = \langle \boldsymbol{x}, \boldsymbol{x} \rangle^{1/2}$. The linear subspace generated by some subset of $\mathbb{R}^m$ is denoted by *span* $(\cdots)$, and its orthogonal complement by *span* $(\cdots)^{\perp}$. A discrete, additive subgroup $L \subset \mathbb{R}^m$ is called a *lattice*. Every lattice L is generated as the integer linear combinations of some set of linearly independent vectors $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n \in L$, where $n \leq m$. The set of vectors $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ is called a *basis* of L. We use the brief notation *span* (L) to denote the real span of the set of basis vectors, i.e., *span* $(L) =$ *span* $(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$.

Let $L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$ denote the lattice with basis $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$. Its *dimension* (also called *rank*) is n and its *basis matrix* (also called *generator matrix*) is the $n \times m$ matrix B which has the basis vectors $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ as its n rows. The lattice is called *full-dimensional* if $n = m$. The *determinant* of L, denoted by $\det(L)$, is defined as $\det(L) = [\det(BB^T)]^{1/2}$. Geometrically, the determinant of a lattice is the common volume of its *fundamental regions*, where a fundamental region is a building block which, when translated by lattice vectors, partitions the whole space with just one lattice point in each copy. The *Voronoi cell* of a point $\boldsymbol{v} \in L$ is an example of a fundamental region for L. It consists of those points of *span* (L) which are at least as close to $\boldsymbol{v}$ as to any other lattice point. If $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n \in \mathbb{R}^m$ is a basis of lattice L, then $\boldsymbol{b}'_1, \cdots, \boldsymbol{b}'_n$ is also a basis of L if and only if there exists a *unimodular* matrix U (integer matrix with determinant ± 1) such that $UB = B'$. The dimension and the determinant of a lattice are independent of the choice of the basis.

On the space of n-D lattices, the (B, ϵ)-*neighborhood* of a lattice L with the basis matrix $B = [b_{ij}]$ consists of all lattices having a basis $B' = [b'_{ij}]$, such that

$$||B - B'|| \triangleq \max_{i,j} \{|b_{ij} - b'_{ij}|\} < \epsilon$$

where ϵ is an arbitrary positive number.

The *ith successive minimum* $\lambda_i(L)$ of a lattice L is the smallest real number such that there are i linearly independent vectors in L of length at most $\lambda_i(L)$. Clearly, we have

$$\lambda_1(L) \leq \lambda_2(L) \leq \cdots \leq \lambda_n(L). \tag{1}$$

We call a lattice an ESM lattice if its successive minima are equal. Obviously, lattices which are generated by their minimum-length vectors have the ESM property, although to the best of our knowledge the converse to this statement has not been proved. The notation $\lambda(L) = \lambda_1(L)$ is used to denote the length of the shortest nonzero vector(s) in L, which is also equal to the minimum distance between lattice points.

The distance between a vector $\boldsymbol{v} \in$ *span* (L) and the lattice L is defined as the minimum distance between $\boldsymbol{v}$ and the points of L. The *covering radius* $\mu(L)$ of a lattice L is the smallest number such that all vectors $\boldsymbol{v} \in$ *span* (L) are at distance at most $\mu(L)$ from the lattice.

To any ordered lattice basis, say $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n \in \mathbb{R}^m$, one can associate a set of *Gram–Schmidt* (*G-S*) vectors $\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_n \in \mathbb{R}^m$, which are computed using the following recursion:

$$\begin{aligned} \hat{\boldsymbol{b}}_1 &= \boldsymbol{b}_1 \\ \hat{\boldsymbol{b}}_i &= \boldsymbol{b}_i - \sum_{j=1}^{i-1} \mu_{i,j} \hat{\boldsymbol{b}}_j, \qquad \text{for } i = 2, \cdots, n \end{aligned} \tag{2}$$

where the G-S coefficients $\mu_{i,j}$'s are equal to

$$\mu_{i,j} = \frac{\langle \boldsymbol{b}_i, \hat{\boldsymbol{b}}_j \rangle}{\langle \hat{\boldsymbol{b}}_j, \hat{\boldsymbol{b}}_j \rangle}. \tag{3}$$

We have $\mu_{i,i} = 1,\ \forall i$, and $\mu_{i,j} = 0$ for $i < j$. Based on the above relationships, the G-S decomposition may be written in matrix notation as

$$B = [\mu_{i,j}]\hat{B} \tag{4}$$

where $\hat{B}$ has $\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_n$ as its rows and $[\mu_{i,j}]$ is the lower triangular matrix of the G-S coefficients. The vector $\hat{\boldsymbol{b}}_i$ is the projection of $\boldsymbol{b}_i$ on $span\,(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{i-1})^{\perp}$. The vectors $\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_n$ are mutually orthogonal and do not necessarily belong to the lattice. It will often be very helpful to think of the basis vectors as being presented in the orthogonal coordinate system of the G-S vectors. It is also easy to see that

$$\det(L) = \prod_{i=1}^{n} \|\hat{\boldsymbol{b}}_i\|. \tag{5}$$

Using (2) and (3), we see that if $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ have rational coordinates, so do the $\hat{\boldsymbol{b}}_i$'s and they can be computed in polynomial time (with respect to the input size) from $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$.

There exists a lower bound on the length of a shortest nonzero vector of a lattice L in terms of the lengths of its G-S vectors [21, p. 18]:

$$\lambda(L) \geq \min\{\|\hat{\boldsymbol{b}}_1\|, \ldots, \|\hat{\boldsymbol{b}}_n\|\}. \tag{6}$$

Let $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ be a fixed ordered basis of a lattice L. Given $\boldsymbol{v} \in span\,(L)$ and $i \in \{1, \cdots, n\}$, we use the notation $\boldsymbol{v}(i)$ (respectively, $L_i(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$, to denote the orthogonal projection of $\boldsymbol{v}$, respectively $L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$, on the $span\,(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{i-1})^{\perp}$. In particular, $\boldsymbol{v}(1) = \boldsymbol{v}$ and $L_1(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n) = L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$. When no confusion can arise, we use L_i as an abbreviated notation for $L_i(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$. Clearly,

$$span\,(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_i) = span\,(\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_i), \qquad \text{for } 1 \leq i \leq n$$

and $\boldsymbol{b}_i(i), \cdots, \boldsymbol{b}_n(i)$ is a basis of the lattice $L_i(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$.

Throughout the paper, we frequently use important known lattices such as A_n, $A_n^*\ (n \geq 1)$, $D_n, D_n^*(n \geq 3)$, $E_n, E_n^*(n = 6, 7, 8)$, $\mathrm{BW}_n\,(n = 2^m, m = 2, 3, \cdots), K_{12}$, and Λ_{24}. For a comprehensive treatment of their properties, the reader is referred to the excellent encyclopedic book of Conway and Sloane [7].

We assume the bases to be rational. This assumption is made only for computation; all the lemmas, propositions, theorems, and corollaries are valid for a general real basis.

III. Coding Gain, Extremal Lattices, and ESM Lattices

In this section, we explain the concepts of coding gain and extremal lattices. Then we show that many well-known lattices and almost all of the lattices used in communications have ESM. Some connections between the coding gain and the successive minima of a lattice are also mentioned.

A. Coding Gain

In a lattice-based signal constellation, the constellation points belong to a lattice L. As a measure of performance of the corresponding lattice code, *coding gain* is defined as

$$\gamma(L) \triangleq \lambda^2(L)[\det(L)]^{-2/n}. \tag{7}$$

The quantity $\gamma(L)$ is the saving in the average energy due to using the lattice L for the transmission instead of using a rectangular grid of points with integer components (the $\mathbb{Z}^n$ lattice).

Hermite's constant γ_n is defined as the supremum of γ over all n-D lattices. It is known that γ_n is attainable [12, p. 267]. The value of γ_n is explicitly known only for $n \leq 8$. Minkowski's convex body theorem [12, p. 51] implies that $\gamma_n \leq 4\pi^{-1}\Gamma(n/2+1)^{2/n}$, which yields $\gamma_n \leq 2n/3$ for all $n \geq 2$. For simplicity, we will use the inequality $\gamma_n \leq n$, which holds for all values of n. It is also known that for large values of n, we have [7, p. 20],

$$\frac{1}{2\pi e} \leq \frac{\gamma_n}{n} \leq \frac{1.744}{2\pi e}. \tag{8}$$

B. Extremal and ESM Lattices

A lattice L_0 is called *extremal* if $\gamma(L_0)$ is a local maximum; i.e., if in the space of n-D lattices, there exists a neighborhood $\mathcal{N}$ of L_0 such that $\gamma(L) \leq \gamma(L_0)$, for $L \in \mathcal{N}$. Extremal lattices have relatively high coding gains and may be useful in channel coding applications. Clearly, the extremal property of a lattice is invariant under scaling and/or orthogonal transformations of the lattice. The following theorem [12, p. 300] is of great importance:

Theorem 1: Every extremal lattice has ESM.

As a corollary, it can be concluded that:

Corollary 1: The densest lattices have ESM.

Proof: The coding gain of the densest lattices are globally maximum, and therefore locally maximum. (Another proof of this corollary, independent of Theorem 1, is given in the Appendix). □

Noting that E_6, E_7, and E_8 are the densest lattices in their corresponding dimensions, it follows from Corollary 1 that they are ESM lattices.

Coxeter proved that the lattices A_n and D_n are extremal (see [12, p. 404]). Barnes and Wall constructed another infinite sequence of extremal lattices (BW_n), which is probably the most famous lattice sequence in communications. Two other well-known extremal lattices are the Leech (Λ_{24}) and Coxeter–Todd (K_{12}) lattices. We therefore obtain the following corollary.

Corollary 2: The lattices $A_n\,(n \geq 1)$, $D_n\,(n \geq 3)$, $E_n\,(n = 6, 7, 8)$, $\mathrm{BW}_n\,(n = 2^m, m = 2, 3, \cdots)$, Λ_{24}, and K_{12} are ESM lattices.

Despite these results, the ESM condition does not have a strong impact on the achievable coding gain for a lattice. The best lower bound that can be obtained on the coding gain of a general ESM lattice is trivial: $\gamma \geq 1$ [6]. It is also shown in [6] that, especially in large dimensions, obtaining large coding gains is possible without having ESM.

The following lemma introduces some other classes of ESM lattices.

Lemma 1: The lattices $A_n^*\,(n \geq 1)$, $D_n^*\,(n \geq 3)$, and $E_n^*\,(n = 6,7,8)$ are ESM lattices.

Proof: Since the proof is similar for the three classes of lattices, only the proof for lattices A_n^* is given here. Consider the following $n \times (n+1)$ basis matrix for A_n^* [7, p. 115]:

$$B = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & 0 & 0 & \cdots & -1 & 0 \\ \frac{-n}{n+1} & \frac{1}{n+1} & \frac{1}{n+1} & \cdots & \frac{1}{n+1} & \frac{1}{n+1} \end{pmatrix}.$$

Using this basis matrix, we obtain $\lambda = \sqrt{n/(n+1)}$. It is not difficult to see that the n lattice vectors $\boldsymbol{b}_n$, $\boldsymbol{b}_n + \boldsymbol{b}_1$, $\boldsymbol{b}_n + \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n + \boldsymbol{b}_{n-1}$ are independent and have length λ. □

The above results imply that almost all of the lattices currently used in communications, either in channel coding or in quantization applications, are ESM lattices.

IV. Korkin–Zolotarev (K-Z) Reduced Basis

The algorithm for finding the closest point of the n-D integer lattice $\mathbb{Z}^n$ to an arbitrary point $\boldsymbol{x} \in \mathbb{R}^n$ is particularly simple. For a real number r, let $\lceil r \rfloor \in \mathbb{Z}$ denote the nearest integer to r. It is not difficult to see that $\lceil \boldsymbol{x} \rfloor \triangleq (\lceil x_1 \rfloor, \ldots, \lceil x_n \rfloor)$ is the closest point of $\mathbb{Z}^n$ to $\boldsymbol{x}$. We call this method of decoding the "round-off procedure."

Let L be a lattice in $\mathbb{R}^m$ given by a basis $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$, and suppose that $\boldsymbol{x} \in \mathbb{R}^m$ is an arbitrary point. Let $\boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{x}''$ with $\boldsymbol{x}' \in \mathit{span}\,(L)$ and $\boldsymbol{x}'' \in \mathit{span}\,(L)^{\perp}$. Clearly, the nearest point of L to $\boldsymbol{x}$ is the one nearest to $\boldsymbol{x}'$. Let

$$\boldsymbol{x}' = \sum_{i=1}^{n} \alpha_i \boldsymbol{b}_i.$$

The round-off procedure on the basis $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ in $\mathit{span}\,(L)$ decodes $\boldsymbol{x}'$ to

$$\boldsymbol{y} = \sum_{i=1}^{n} \lceil \alpha_i \rfloor \boldsymbol{b}_i.$$

Geometrically, this is equivalent to employing a parallelotope decision region[1] spanned by vectors $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$, centered at each lattice point. It can be shown that although the round-off procedure is a very efficient polynomial-time algorithm, it obtains the nearest point of the lattice if and only if the basis vectors are mutually orthogonal. Unfortunately, for lattices with $\gamma > 1$, such a basis does not exist (as can easily be proved by contradiction).

The nice properties of orthogonal bases motivate searching for bases of a lattice that are nearly orthogonal. The problem of transforming a given lattice basis into a basis consisting of vectors which are pairwise nearly orthogonal is called *lattice basis reduction.*[2] Reduction theory, which has its historical roots in the 18th century, was mainly motivated by the classical question of finding the minima of positive-definite integral forms. Several distinct notions of reduction have been studied, including those associated with the names Hermite, Minkowski, Korkin–Zolotarev (K-Z), and more recently Lenstra, Lenstra, and Lovász (L^3); see, e.g., [12, pp. 147–164]. After the introduction of the L^3 reduced basis, which can be computed in polynomial time, reduction theory has found many applications in a variety of areas (see, e.g., [2], [13]–[16], [19]–[21], [24, pp. 71–74]). However, it can be shown that for the decoding of lattices, the K-Z reduced basis is a more powerful tool than the L^3 reduced basis [6]. In the following, we explain the K-Z reduced basis, which is used in our decoding algorithms.

Let $L \subset \mathbb{Q}^m$ be a lattice with ordered basis $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$, and corresponding G-S vectors $\hat{\boldsymbol{b}}_1, \cdots, \hat{\boldsymbol{b}}_n$ ($\mathbb{Q}$ is the set of rational numbers). In the lattice decoding algorithm presented in Subsection V-A, one needs to check the distance between a given vector $\boldsymbol{x} \in \mathit{span}\,(L)$ and a subset of the lattice vectors. The upper bound on the number of candidates needed to be checked depends on the lengths of the G-S vectors. Consequently, it is desirable to make these lengths as small as possible by finding a properly reduced basis of the lattice L. By the reduction theory introduced by Korkin and Zolotarev [17], the vectors of a basis can be selected such that the lengths of the corresponding G-S vectors are minimized successively, i.e.,

$$\|\hat{\boldsymbol{b}}_i\| = \lambda(L_i), \qquad \text{for } i = 1, \cdots, n \tag{9}$$

where $\lambda(L_i)$ is the length of a shortest vector of the ith projected lattice L_i. In particular, $\|\hat{\boldsymbol{b}}_1\| = \|\boldsymbol{b}_1\| = \lambda(L)$. K-Z reduced bases are extensively studied in [18].

It can be shown that each lattice has at least one K-Z reduced basis (see [16] or [23]). There is no polynomial-time algorithm known for K-Z reduction. Finding a K-Z reduced basis of a lattice is actually polynomial-time equivalent to finding a shortest vector of the lattice. The fastest known algorithm for K-Z reduction of a basis $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n \in \mathbb{Z}^m$ with $\varphi = \max(\|\boldsymbol{b}_1\|^2, \cdots, \|\boldsymbol{b}_n\|^2)$ and $m = O(n)$ is due to Schnorr [23] and has a theoretical worst case time bound of $\sqrt{n}^{\,n+o(n)} + O(n^4 \log \varphi)$ arithmetic steps on $O(n \log \varphi)$-bit integers. This algorithm is an improved version of Kannan's shortest lattice vector algorithm [16].

Example 1: The following are K-Z reduced bases for the lattices D_4 and E_8 (see the top of the following page).

The following lemma is subsequently of great importance.

Lemma 2: If $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n$ is a K-Z reduced basis of an n-D ESM lattice, then

$$\max_{1 \leq i \leq n} \|\hat{\boldsymbol{b}}_i\| = \|\hat{\boldsymbol{b}}_1\|. \tag{10}$$

Proof: First, we consider an arbitrary lattice L. By the definition of $\lambda_i(L)$, there exist at least i linearly independent vectors of L of length at most $\lambda_i(L)$. Under the projection $L \longrightarrow L_i$, at least one of them, say $\boldsymbol{v}$, has a nonzero projection $\boldsymbol{v}(i)$. Therefore, we have $\lambda(L_i) \leq \|\boldsymbol{v}(i)\|$. This inequality combined with the fact that $\|\boldsymbol{v}(i)\| \leq \|\boldsymbol{v}\| \leq \lambda_i(L)$ results in $\lambda(L_i) \leq \lambda_i(L)$. Combining this with (9), we obtain $\|\hat{\boldsymbol{b}}_i\| \leq$

[1] The *decision region* of a point $\boldsymbol{P}_i$ belonging to a discrete collection of points $\{\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots\} \subset \mathit{span}\,(L)$ consists of those points of $\mathit{span}\,(L)$ which are decoded to $\boldsymbol{P}_i$.

[2] More generally, reduction theory is concerned with selecting a basis with desirable properties.

$$B_{D_4} = \begin{pmatrix} -1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 \end{pmatrix} \qquad B_{E_8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 0 & 1 & 0 & 1 \\ 0 & -1 & -1 & 0 & -1 & 0 & 0 & 1 \\ -1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}.$$

$\lambda_i(L)$. The proof then follows by putting this result together with the facts that for an ESM lattice $\lambda_1 = \lambda_2 = \cdots = \lambda_n$, and for a K-Z reduced basis $\|\hat{\boldsymbol{b}}_1\| = \lambda_1$. □

It is interesting to note that the above result can be equivalently expressed as

$$\lambda(L) = \max\{\|\hat{\boldsymbol{b}}_1\|, \cdots, \|\hat{\boldsymbol{b}}_n\|\}.$$

Comparing this equality with (6) gives further evidence for the strength of the K-Z reduction.

As already mentioned, in applications to communication systems, we assume that the computation devoted to finding a reduced basis is done once (off-line), and we do not consider this computation as a part of the decoding complexity.

In the rest of the paper, the lattices are assumed to be full-dimensional, i.e., $n = m$. This assumption simplifies some of the discussions without loss of generality.

V. Lattice Decoding Problem

In this section, we first discuss Kannan's decoding algorithm and its complexity. Then, in Subsection V-A, we propose some modifications which increase the efficiency of Kannan's algorithm, especially in communication applications. Complexity bounds for the proposed algorithm are derived in Subsection V-B.

Consider the following lattice decoding problem:

(**LDP**) given the vector $\boldsymbol{x} \in \mathbb{Q}^n$and lattice $L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n) \subset \mathbb{Q}^n$, find a lattice vector $\boldsymbol{b} = \sum_{j=1}^{n} \beta_j \boldsymbol{b}_j$ such that $\|\boldsymbol{x} - \boldsymbol{b}\|$ is minimized. (11)

In 1981, Van Emde Boas proved that the LDP is NP-hard [28]. A simpler proof was subsequently given by Kannan in 1987 [16]. More recently, it has been shown by Arora *et al*. that even approximating the solution within any constant factor is NP-hard [1]. Some other relevant results regarding approximate solutions for the LDP can be found in [2], [13], and [18].

The fastest (best upper bound on the complexity) known algorithm for solving the LDP for a general lattice is due to Kannan [16], an improved version of his earlier work in [15]. Prior to [16], Helfrich [14] also made some improvements in the running time of some of the algorithms in [15]. In [16], Kannan uses the same reduced basis as used in this paper,[3] and shows that for some particular i_1 such that

$$\|\hat{\boldsymbol{b}}_{i_1}\| = \max \|\hat{\boldsymbol{b}}_j\|, \qquad j \in \{1, \cdots, n\}$$

there exists a subset of $\mathbb{Z}^{n-i_1+1}$ of cardinality at most $(n+\sqrt{n})^{(n-i_1+1)}$ that contains the values $(\beta_{i_1}, \cdots, \beta_n)$ of the nearest point. Now, if $\boldsymbol{b}'$ solves the LDP for the vector

$$\boldsymbol{x} - \sum_{j=i_1}^{n} \beta_j \boldsymbol{b}_j$$

and the lattice $L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_{i_1-1})$, then

$$\boldsymbol{b}' + \sum_{j=i_1}^{n} \beta_j \boldsymbol{b}_j$$

is a solution candidate of the problem for $\boldsymbol{x}$ and $L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n)$. Therefore, the original problem can be reduced to at most $(n+\sqrt{n})^{(n-i_1+1)}$ subproblems, each of dimensionality $i_1 - 1$. In the following, we present a geometrical interpretation of Kannan's algorithm which provides a better understanding of some complexity issues discussed later.

Let indices $i_0, \cdots, i_k$, where $k \leq n$ is a constant integer, be successively defined by

$$\|\hat{\boldsymbol{b}}_{i_j}\| = \max_{1 \leq m \leq i_{j-1}-1} \|\hat{\boldsymbol{b}}_m\|$$

for $1 \leq j \leq k$, with

$$n + 1 = i_0 > i_1 > \cdots > i_k = 1.$$

Here i_1 is the same index as defined in the last paragraph. A careful inspection of Kannan's algorithm [16] reveals that if $i_j \leq q \leq i_{j-1} - 1$, then the algorithm recursively searches for the candidates $\boldsymbol{b}$ such that the projection length of $(\boldsymbol{b} - \boldsymbol{x})$ along the G-S vector $\hat{\boldsymbol{b}}_q$ is at most

$$\ell_j = \left(\sum_{m=1}^{i_{j-1}-1} \|\hat{\boldsymbol{b}}_m\|^2 \right)^{1/2} \Big/ 2.$$

Thus the algorithm may be thought of as searching among lattice points in a rectangular parallelepiped centered at $\boldsymbol{x}$, with edges pointing parallel to the G-S vectors of the lattice. The edge length of the parallelepiped along $\hat{\boldsymbol{b}}_q$ is $2\ell_j$. (Note that $\ell_1 > \cdots > \ell_k$). For simplicity, we think of such rectangular parallelepipeds as cubes, and we call the algorithms of this type "recursive cube search" (RCS) algorithms. As we will see later, our proposed algorithm also belongs to this category.

[3] The reduced basis used by Kannan has an extra condition on the value of the G-S coefficients $\mu_{i,j}$, i.e., $|\mu_{i,j}| \leq 1/2$ for $1 \leq j < i \leq n$. However, this condition does not affect the G-S orthogonalization of the basis.

The computational complexity of Kannan's algorithm depends on the values of $i_1, \cdots, i_{k-1}$. One can see that $i_1 = 1$ leads to the highest complexity. In this case, the original problem is reduced to at most $(n + \sqrt{n})^n$ 0-dimensional subproblems. Each subproblem is solved by just checking the distance between $\boldsymbol{x}$ and a certain point of the lattice. It can be shown that the number of arithmetic operations for Kannan's algorithm is bounded above by $(2n)^{n+O(1)}$.[4] This bound is obtained based on the amount of computation required for the worst case of $i_1 = 1$, and the fact that the number of arithmetic operations needed to find a candidate and check its distance to the given vector is polynomially bounded by n.

We are also interested in obtaining lower bounds on the complexity of such algorithms. To emphasize the importance of lower bounds, we pose the following question: "Is there any sequence of lattices with possible application in communications $(\gamma \geq 1)$ such that Kannan's algorithm can decode them in polynomial time for an arbitrary $\boldsymbol{x}$?" As we will see later, using the derived lower bound, the answer to this question is negative even if one finds the best possible basis.

In the following, the number of lattice points $N(\cdot)$ that should be checked is defined as the measure of decoding complexity. This can be easily translated to a statement in terms of the required number of arithmetic operations in O-notation. For both Kannan's and our proposed algorithm, $N(\cdot)$ depends not only on the selection of basis, but also on the vector $\boldsymbol{x}$ and the structure of the lattice itself. The notation $N(\boldsymbol{x}, L, B)$ is therefore used for the number of candidates, where B is the generator matrix of the lattice. For the sake of simplicity, we sometimes use the notation N instead of $N(\boldsymbol{x}, L, B)$. We also sometimes use the logarithm of the number of candidates as an index of complexity, referred to hereafter as the *log-complexity* (the base of the logarithm can be selected arbitrarily).

Let $\mathcal{S}$ denote the region of the space that an RCS algorithm searches to solve the LDP. As a rough approximation to the complexity measure N, one can consider its average value $\bar{N}(L, B)$, averaged over all vectors $\boldsymbol{x}$ which are uniformly distributed in a fundamental region of L. It can be seen that this is equal to

$$\bar{N}(L, B) = \frac{\mathrm{vol}\,(\mathcal{S})}{\det(L)} \tag{12}$$

where $\mathrm{vol}\,(\mathcal{S})$ is the volume of $\mathcal{S}$ and $\det(L)$ is the volume of a fundamental region of L.

Using N as the measure of complexity, the complexity of Kannan's algorithm for the decoding of a general n-D lattice for an arbitrary $\boldsymbol{x}$ is upper-bounded by $(n+\sqrt{n})^n$. This upper bound can be improved for ESM lattices.

[4]Kannan's result in [16] is slightly different, however. It is claimed in [16, Theorem 4.5] that the number of arithmetic operations performed by the algorithm is $O(n^n)$. The authors believe that this is an underestimate, since for the worst case of $i_1 = 1$, even the number of candidates, i.e., $(n + \sqrt{n})^n$, cannot be upper-bounded by Cn^n, for any positive constant C. The mistaken component in the proof turns out to be the wrong assumption that the maximum of $\{(i-1)/(n+\sqrt{n})\}^{(i-1)}$ for $1 \leq i \leq n$ is attained at $i = n$. It is not difficult, however, to see that the maximum is 1, and is obtained for $i = 1$.

Theorem 2: For an n-D ESM lattice L with coding gain γ and K-Z reduced basis B, and for any given vector $\boldsymbol{x} \in \mathbb{R}^n$, the complexity of Kannan's algorithm satisfies

$$N(\boldsymbol{x}, L, B) \leq (\sqrt{n}+1)^n \gamma^{n/2}. \tag{13}$$

Proof: Using the notations already used in describing Kannan's algorithm, we have an upper bound of

$$(2\ell_j / \|\hat{\boldsymbol{b}}_q\|) + 1$$

on the number of possible values for each integer β_q, where $i_j \leq q \leq i_{j-1} - 1$. For ESM lattices, using Lemma 2, we have $k = 1$ $(i_1 = 1)$, and each integer β_q, $q = 1, \cdots, n$, takes at most $(2\ell_1 / \|\hat{\boldsymbol{b}}_q\|) + 1$ different values. This corresponds to the following upper bound on N:

$$\begin{aligned} N &\leq \prod_{q=1}^{n} \left(\frac{\sqrt{\sum_{m=1}^{n} \|\hat{\boldsymbol{b}}_m\|^2}}{\|\hat{\boldsymbol{b}}_q\|} + 1 \right) \\ &\leq \prod_{q=1}^{n} \left(\frac{\sqrt{n}\|\hat{\boldsymbol{b}}_1\|}{\|\hat{\boldsymbol{b}}_q\|} + 1 \right) \\ &\leq (\sqrt{n}+1)^n \frac{\|\hat{\boldsymbol{b}}_1\|^n}{\prod_{q=1}^{n} \|\hat{\boldsymbol{b}}_q\|}. \end{aligned} \tag{14}$$

For the last two steps, we have used the fact that $\|\hat{\boldsymbol{b}}_m\| \leq \|\hat{\boldsymbol{b}}_1\|$ for $m = 1, \cdots, n$. The proof then follows by applying (5), the fact that for a K-Z reduced basis $\|\hat{\boldsymbol{b}}_1\| = \lambda$, and the definition of coding gain. □

In fact, for ESM lattices, the algorithm searches among the lattice points in the cube centered at $\boldsymbol{x}$, with edges of length $2\ell_1$ oriented along the G-S vectors.

In the next subsection, we modify Kannan's algorithm by reducing the length of each edge of the search cube. Therefore, all the lower bounds derived on the complexity of the modified algorithm are also valid for Kannan's algorithm. The bounds, which are in terms of coding gain and dimension, imply that for any sequence of lattices with possible application in communications $(\gamma \geq 1)$, and any selected basis, the complexity of Kannan's algorithm grows at least exponentially with n and γ.

A. Modified Kannan's Algorithm

Consider the lattice decoding problem defined in (11). Based on the definition of covering radius, the candidates for the nearest vector of L to a given vector $\boldsymbol{x}$ are the lattice points inside the sphere of radius $\mu(L)$, centered at $\boldsymbol{x}$. However, a good algorithm for finding the lattice points inside a sphere does not exist. The proposed approach for solving the LDP is to consider the lattice points inside a properly selected cube, centered at $\boldsymbol{x}$. To search inside the cube, we devise the following RCS algorithm.

Modified Recursive Cube Search Algorithm: Let $\boldsymbol{b}$ be a candidate for the nearest vector of $L(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n)$ to a given vector

$$\boldsymbol{x}=\sum_{i=1}^{n}\alpha_i\boldsymbol{b}_i,\ \alpha_i\in\mathbb{Q},\quad \forall i$$

and let

$$\boldsymbol{b}=\sum_{i=1}^{n}\beta_i\boldsymbol{b}_i,\ \beta_i\in\mathbb{Z},\quad \forall i.$$

The candidates can be checked by enumerating the coefficients β_i from β_n to β_1, successively. This can be done efficiently by noticing that we just need to search for $\boldsymbol{b}$ among the lattice points such that

$$\begin{aligned}&\|\boldsymbol{b}-\boldsymbol{x}\|\le\mu(L)\Longrightarrow\\&\|(\boldsymbol{b}-\boldsymbol{x})(n)\|=|\beta_n-\alpha_n|\|\hat{\boldsymbol{b}}_n\|\le\mu(L)\end{aligned}\tag{15}$$

where $(\boldsymbol{b}-\boldsymbol{x})(n)$ is the projection of $\boldsymbol{b}-\boldsymbol{x}$ along the G-S vector $\hat{\boldsymbol{b}}_n$. The last inequality in (15) enables us to enumerate β_n for the lattice points under consideration. Now, if $\boldsymbol{b}'$ solves the LDP for the vector $\boldsymbol{x}-\beta_n\boldsymbol{b}_n$ and the lattice $L(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_{n-1})$, then $\boldsymbol{b}'+\beta_n\boldsymbol{b}_n$ is a solution candidate of the problem for $\boldsymbol{x}$ and $L(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n)$. Therefore, the original problem can be reduced to several subproblems, each of dimensionality $n-1$.

Let $\mu_i\triangleq\mu(L(\boldsymbol{b}_1,\cdots,\boldsymbol{b}_i))$, for $i=1,\cdots,n$, and let $\mathcal{S}(\boldsymbol{x})$ denote the region of $\mathbb{R}^n$ that the algorithm searches to solve the LDP for a given vector $\boldsymbol{x}$. It is not difficult to see that $\mathcal{S}(\boldsymbol{x})$ is a cube centered at $\boldsymbol{x}$, with edges of length $2\mu_1,\cdots,2\mu_n$ oriented along the G-S vectors $\hat{\boldsymbol{b}}_1,\cdots,\hat{\boldsymbol{b}}_n$, respectively. The volume of $\mathcal{S}(\boldsymbol{x})$ is therefore equal to

$$\mathrm{vol}\,(\mathcal{S})=2^n\prod_{i=1}^{n}\mu_i.\tag{16}$$

Note that, in general, inequalities of the form

$$\mu_i\le\mu_n=\mu(L),\qquad \text{for } i=1,\cdots,n$$

do not hold. Therefore, to reduce complexity, one can select the edge length of the search cube in the direction of $\hat{\boldsymbol{b}}_i$ to be the minimum of $2\mu_i$ and $2\mu(L)$. All the bounds derived later in Subsection V-B remain valid for this case.

Using the modified RCS algorithm, it appears that we only need to enumerate relatively few candidate integer n-tuples $(\beta_1,\cdots,\beta_n)$. To derive a proper upper bound on the number of candidates, we first prove the following proposition.

Proposition 1: Let $\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n$ be an ordered basis of the lattice L, with G-S orthogonalization $\hat{\boldsymbol{b}}_1,\cdots,\hat{\boldsymbol{b}}_n$. Then, we have

$$\mu(L)\le d\triangleq\frac{1}{2}\sqrt{\sum_{i=1}^{n}\|\hat{\boldsymbol{b}}_i\|^2}\tag{17}$$

where the inequality holds with equality if and only if there exists an orthogonal basis for L with the lengths of its vectors equal to $\|\hat{\boldsymbol{b}}_i\|, i=1,\cdots,n$.

Proof: To each lattice point $\boldsymbol{b}$, we assign a cubic suboptimum decision region centered at $\boldsymbol{b}$ with its edges along the G-S coordinates $\hat{\boldsymbol{b}}_1,\cdots,\hat{\boldsymbol{b}}_n$. The edge lengths are selected as $\|\hat{\boldsymbol{b}}_1\|,\cdots,\|\hat{\boldsymbol{b}}_n\|$, respectively. It is not difficult to see that these cubic regions partition $\mathbb{R}^n$ (each of them is a fundamental region for the lattice). Using the fact that the maximum distance between a lattice point and the points of its decision region is d and the definition of covering radius, the inequality follows immediately.

It is easy to see that if the basis is orthogonal, the inequality in (17) holds with equality. In this case, the Voronoi cells for the lattice points coincide with the aforementioned cubic decision regions. We sketch a proof to show that if $\mu(L)=d$, then there exists a basis of L which is orthogonal. Suppose that such a basis does not exist. Consider an arbitrary lattice point $\boldsymbol{b}$ and call its corresponding cubic decision region $\mathcal{R}(\boldsymbol{b})$. Except for the vertices of $\mathcal{R}(\boldsymbol{b})$, we have $\|\boldsymbol{b}-\boldsymbol{v}\|<d$, for every vector $\boldsymbol{v}\in\mathcal{R}(\boldsymbol{b})$. Let $\boldsymbol{p}$ be an arbitrary vertex of $\mathcal{R}(\boldsymbol{b})$. Since we are assuming that L is not rectangular, and because of the congruence of the structure, there should exist a decision region $\mathcal{R}(\boldsymbol{b}')$ adjacent to $\mathcal{R}(\boldsymbol{b})$ which has $\boldsymbol{p}$ at its intersection with $\mathcal{R}(\boldsymbol{b})$, but $\boldsymbol{p}$ is not a vertex of $\mathcal{R}(\boldsymbol{b}')$, and therefore $\|\boldsymbol{b}'-\boldsymbol{p}\|<d$. It can be seen that if such a region does not exist, in other words, if all the adjacent cubic regions have $\boldsymbol{p}$ as their vertex, then for the whole structure all the adjacent cubes coincide in their vertices and, consequently, the lattice is rectangular and has an orthogonal basis. This shows that there exists no vector $\boldsymbol{v}\in\mathbb{R}^n$ such that the distance between $\boldsymbol{v}$ and the lattice is greater than or equal to d, which results in $\mu(L)<d$. □

It can be concluded from the proof of Proposition 1 that for any given $\boldsymbol{x}\in\mathbb{R}^n$, there exists a unique $\boldsymbol{b}\in L$ such that

$$\boldsymbol{x}=\boldsymbol{b}+\sum_{i=1}^{n}\eta_i\hat{\boldsymbol{b}}_i,\qquad -\frac{1}{2}\le\eta_i<\frac{1}{2},\ \forall i$$

where $\boldsymbol{x}$ is located inside the cubic decision region of point $\boldsymbol{b}$.

In the following, we derive lower and upper bounds on $N(\boldsymbol{x},L,B)$ for the modified RCS algorithm.

Proposition 2: For any basis B of a lattice L and any $\boldsymbol{x}\in\mathbb{R}^n$, the number of candidate points $N(\boldsymbol{x},L,B)$ of the modified RCS algorithm satisfies

$$\prod_{i=2}^{n}\left(\frac{2\mu_i}{\|\hat{\boldsymbol{b}}_i\|}-1\right)<N(\boldsymbol{x},L,B)\le\prod_{i=1}^{n}\left(\frac{2\mu_i}{\|\hat{\boldsymbol{b}}_i\|}+1\right).\tag{18}$$

Proof: We enumerate the β_i's from β_n to β_1, successively. This means that for each $\beta_i,\ i=1,\cdots,n$, the values taken by $\beta_{i+1},\cdots,\beta_n$ have been already selected. Starting from β_n, using (15) and the fact that β_n is an integer, we obtain the lower bound of $(2\mu_n/\|\hat{\boldsymbol{b}}_n\|)-1$ and the upper bound of $(2\mu_n/\|\hat{\boldsymbol{b}}_n\|)+1$ on the number of possible values for β_n. Using similar bounds for every $i=1,\cdots,n$, the inequality follows. Note that for $i=1$, $\mu_1=\|\hat{\boldsymbol{b}}_1\|/2$, and the number of possible values for β_1 is at least 1. □

Remark: Applying (5) and Proposition 1 to the upper bound in inequality (18), we obtain

$$N(\boldsymbol{x}, L, B) \leq \frac{\prod_{i=1}^{n} \left(\sqrt{\sum_{j=1}^{i} \|\hat{\boldsymbol{b}}_j\|^2} + \|\hat{\boldsymbol{b}}_i\| \right)}{\det(L)}. \tag{19}$$

For a fixed lattice L, $\det(L)$ has a fixed value, and the bound in (19) is just a function of the lengths of the G-S vectors. This justifies the selection of a K-Z reduced basis for the decoding algorithm. Note that although the $\|\hat{\boldsymbol{b}}_i\|$'s are minimized successively for a K-Z reduced basis, this selection does not necessarily result in minimizing the upper bound in (19).

In an efficient implementation of the algorithm, one can simply update the distances from point to point, and keep only the nearest vector $\boldsymbol{b}^\star \in L$ to $\boldsymbol{x}$ found so far. Noting that the coefficient matrix in (4) is triangular, a more efficient implementation is possible if one uses the G-S coordinates to represent the basis vectors. In this work, however, our main concern is complexity bounds. From this point of view, it is clear that the number of required arithmetic operations for finding a candidate lattice point and checking its distance to the given vector is polynomially bounded by n.

Having selected the values $\beta_{k+1}, \cdots, \beta_n$, one can also use branch-and-bound, and prune any further search of vectors of the form

$$\boldsymbol{b} = \boldsymbol{b}' + \sum_{i=k+1}^{n} \beta_i \boldsymbol{b}_i, \qquad \text{with } \boldsymbol{b}' \in L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k)$$

if $\|(\boldsymbol{b}-\boldsymbol{x})(k+1)\| \geq \|\boldsymbol{b}^\star - \boldsymbol{x}\|$, where $(\boldsymbol{b}-\boldsymbol{x})(k+1)$ is the orthogonal projection of $\boldsymbol{b}-\boldsymbol{x}$ on *span* $(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k)^\perp$. One can also stop enumerating β_1 when $\|\boldsymbol{b}-\boldsymbol{x}\|$ starts increasing. Moreover, there is a quick certificate for the closest vector of a lattice L to a given vector $\boldsymbol{x}$, i.e., if for a candidate $\boldsymbol{b}$, $\|\boldsymbol{b}-\boldsymbol{x}\| \leq \lambda(L)/2$, then $\boldsymbol{b}$ is the closest vector of L to $\boldsymbol{x}$. This condition can be checked for each candidate, and if it is satisfied one can stop the algorithm.[5]

It is conceivable that by embedding the above modifications in the algorithm, not every lattice vector in the cubic region $\mathcal{S}(\boldsymbol{x})$ must be examined. This could affect the derived lower bounds on the complexity of RCS algorithms. However, this effect is presumably small, especially for the asymptotics of the algorithm, and in any case it cannot be mathematically analyzed in a reasonable way.

B. Complexity Bounds for the Modified RCS Algorithm

1) Upper Bounds on Complexity: Using Proposition 1, it is clear that Kannan's algorithm searches in a larger region of the space as compared to the modified algorithm, and consequently has a larger number of candidate points to check.[6] This implies that the upper bounds derived for Kannan's algorithm are also valid for the proposed algorithm. Tighter bounds, however, can be found as follows.

Theorem 3: For an n-D lattice L with a K-Z reduced basis B, and for any $\boldsymbol{x} \in \mathbb{R}^n$

$$N(\boldsymbol{x}, L, B) \leq \prod_{i=1}^{n} (\sqrt{i}+1) {\gamma_n}^{n/2}. \tag{20}$$

Proof: From the proof of Lemma 2, we know that for a K-Z reduced basis, $\|\hat{\boldsymbol{b}}_j\| \leq \lambda_j(L)$ for $j = 1, \cdots, n$. Putting this together with (19), we obtain

$$\begin{aligned} N &\leq \frac{\prod_{i=1}^{n} \left(\sqrt{\sum_{j=1}^{i} \lambda_j^2(L)} + \lambda_i(L) \right)}{\det(L)} \\ &\leq \frac{\prod_{i=1}^{n} (\sqrt{i}+1)\lambda_i(L)}{\det(L)} \end{aligned} \tag{21}$$

where for the last step, we have used (1). Combining inequality (21) with a result of Minkowski which implies $\lambda_1 \cdots \lambda_n \leq \det(L){\gamma_n}^{n/2}$ [12, p. 195] completes the proof. □

As a corollary of Theorem 3, we obtain the following upper bound on complexity.

Corollary 3: For an n-D lattice L with a K-Z reduced basis B, and for any $\boldsymbol{x} \in \mathbb{R}^n$

$$N(\boldsymbol{x}, L, B) \leq (\sqrt{n}+1)^n {\gamma_n}^{n/2}. \tag{22}$$

For ESM lattices, the bound in (20) can be improved as follows. The two bounds coincide for the densest lattices.

Theorem 4: For an n-D ESM lattice L with coding gain γ and K-Z reduced basis B, and for any $\boldsymbol{x} \in \mathbb{R}^n$

$$N(\boldsymbol{x}, L, B) \leq \prod_{i=1}^{n} (\sqrt{i}+1) \gamma^{n/2}. \tag{23}$$

Proof: The result follows by applying $\lambda_1 = \cdots = \lambda_n$ to (21), and using the definition of coding gain. □

As a corollary of the above theorem, we obtain the same upper bound as given in Theorem 2 on the complexity of the algorithm for ESM lattices. Using $\gamma_n \leq n$, inequality (22) corresponds to a bound $(2n)^{n+O(1)}$ for the required number of arithmetic operations, and to a log-complexity of $n \log n + O(n)$. Although this bound cannot be improved for the densest lattices, for most ESM lattices better complexity bounds can be found based on (13).

Example 2: Consider the Barnes–Wall lattices BW_n $(n = 2^m, m \geq 2)$, with coding gain $\gamma = \sqrt{n/2}$. Substituting this quantity in (13), we obtain $N(\mathrm{BW}_n) \leq (1+\sqrt{n})^n (n/2)^{n/4}$. It is not difficult to see that the corresponding log-complexity is $(3n \log n)/4 + O(n)$.

2) Lower Bounds on Complexity: As we already know, solving the LDP for a general lattice is NP-hard. Combining this fact with the widely believed conjecture of NP $\neq$ P implies that no proposed algorithm can solve the LDP for a general lattice in polynomial time. Now, one might ask the following question: "Is it possible to solve the LDP in polynomial time in communication applications?" When posing this problem, one might have the idea of doing some precomputations (e.g.,

[5] This point was suggested by one of the reviewers.

[6] Note that even if we use the upper bounds on μ_i's given by (17) instead of the μ_i's themselves, in general, the algorithm still searches in a smaller cube and therefore has a lower complexity compared to Kannan's method. In this case, the difference between complexities could be especially large for the decoding of lattices with $\max_i \|\hat{\boldsymbol{b}}_i\| = \|\hat{\boldsymbol{b}}_1\|$ (including ESM lattices).

finding an appropriate basis and/or computing the covering radii of the lattice and its sublattices). In the following, we show that for lattices with $\gamma \geq 1$, there does not exist any basis such that using the proposed algorithm, and therefore Kannan's algorithm, one can solve the LDP in polynomial time. To show this, we first prove the following theorem.

Theorem 5: For an n-D lattice L with any basis B

$$\bar{N}(L,B) \geq 0.866[1.333\gamma(L)]^{n/2}. \tag{24}$$

Proof: For the modified RCS algorithm, using (12) and (16), we have

$$\bar{N}(L,B) = \frac{2^n \prod_{i=1}^{n} \mu_i}{\det(L)}. \tag{25}$$

A result due to Ryškov [22] implies that for an n-D lattice, $\mu/\lambda \geq \sqrt{n/(2n+2)}$. Applying this inequality for dimensions $i = 1, \cdots, n$ to (25), and using the fact that $\lambda(L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_i)) \geq \lambda(L)$, for $i = 1, \cdots, n$, we obtain

$$\bar{N}(L,B) \geq \prod_{i=1}^{n} \sqrt{\frac{2i}{i+1}}\, \frac{\lambda^n(L)}{\det(L)} \geq (1.1547)^{n-1}[\gamma(L)]^{n/2}. \tag{26}$$

For the last step, we have used the definition of $\gamma(L)$, and the fact that $\sqrt{2i/(i+1)}$ is a uniformly increasing function of i with the value of 1.1547 at $i = 2$. The proof then immediately follows from (26). □

Note that for a K-Z reduced basis $\lambda(L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_i)) = \lambda(L)$, for $i = 1, \cdots, n$. However, this cannot improve the lower bound in (24). The following corollary is a consequence of Theorem 5.

Corollary 4: For an n-D lattice L with any basis B, there exists some $\boldsymbol{x} \in \mathbb{R}^n$ such that

$$N(\boldsymbol{x}, L, B) \geq 0.866[1.333\gamma(L)]^{n/2}. \tag{27}$$

Inequality (27) shows that for lattices with sufficiently large coding gains, for a general given vector $\boldsymbol{x}$, the complexity of the modified RCS algorithm grows at least exponentially with n and γ (note that $n \geq \gamma$).

As a complement to Theorem 4, we derive the following lower bound on the decoding complexity of ESM lattices which is stronger than Corollary 4.

Theorem 6: For an n-D ESM lattice L with coding gain γ and K-Z reduced basis B, and for any $\boldsymbol{x} \in \mathbb{R}^n$

$$N(\boldsymbol{x}, L, B) > 6.464(0.023\gamma)^{n/2}. \tag{28}$$

Proof: Starting from the lower bound in (18), we first multiply it by $2\mu_1/\|\hat{\boldsymbol{b}}_1\| = 1$, then use (5), Lemma 2, Ryškov's inequality for dimensions $i = 1, \cdots, n$, and finally the fact that for a K-Z reduced basis $\lambda(L(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_i)) = \lambda(L)$, for $i = 1, \cdots, n$, to obtain

$$N(\boldsymbol{x}, L, B) > \prod_{i=2}^{n} \left(\sqrt{\frac{2i}{i+1}} - 1 \right) \frac{\lambda^n(L)}{\det(L)}. \tag{29}$$

Applying the definition of coding gain γ, and substituting $i = 2$ in all the terms in the above product complete the proof (note that $\sqrt{2i/(i+1)}$ is a uniformly increasing function of i). □

For the densest lattices and for large values of n, combining Theorem 6 with the lower bound in (8) results in a log-complexity of at least $(n/2)\log n + O(n)$ for the decoding algorithm.

VI. Concluding Remarks

Solving the lattice decoding problem (LDP) is the major obstacle associated with using lattices in communication applications. There exist very efficient algorithms for solving the LDP in the case of lattices with strong algebraic structure. However, this is not the case for a general lattice. In this paper, we have obtained some results regarding the complexity of solving the LDP for a general lattice. These results relate the decoding complexity to the coding gain and the dimension of the lattice, and are obtained based on a decoding approach which is an improved version of Kannan's algorithm. Improved complexity results have been obtained for ESM lattices.

It has been shown that Kannan's algorithm, which is currently the fastest known algorithm for solving the LDP for a general lattice, is a search method inside a rectangular parallelepiped (cube) with edges oriented along the Gram–Schmidt vectors of the lattice. Explicit lower and upper bounds on the complexity of Kannan's algorithm have been derived.

The proposed algorithm solves the LDP recursively by reducing the dimension of the problem by one in each step. It employs a Korkin–Zolotarev (K-Z) reduced basis of the lattice. To increase the efficiency for the decoding of lattices in communications, it also uses the knowledge of the covering radii of the lattice and its sublattices. It has been shown that the algorithm searches in a cube similar to Kannan's, except that the edges of the cube are shorter for the proposed algorithm. Explicit lower and upper bounds have been derived on the complexity of the algorithm in terms of the coding gain and the dimension of the lattice.

It was proved in [26] that the trellis decoding complexity of lattices grows exponentially with coding gain. Our lower bounds prove a parallel result for RCS algorithms, i.e., the decoding complexity of any sequence of lattices with coding gain $\gamma \geq 1$ increases exponentially with dimension and coding gain. This suggests that RCS algorithms are not going to be attractive for decoding dense lattices in high dimensions. The lower bound also indicates that our upper-bound results cannot be much improved.

The densest lattices and most of the lattices used in communications have equal successive minima. Upper and lower bounds of the forms $(1+\sqrt{n})^n \gamma^{n/2}$ and $6.464(0.023\gamma)^{n/2}$, respectively, have been established on the decoding complexity of ESM lattices. The lower bound indicates that for any given vector, the decoding complexity of ESM lattices with sufficiently large coding gain grows exponentially with dimension and coding gain. Using the above bounds, we have obtained $n\log n + O(n)$ and $(n/2)\log n + O(n)$ as upper and lower bounds on the decoding log-complexity of the densest lattices, respectively. It has also been shown that

tighter upper bounds in terms of dimension can be found for many interesting sequences of ESM lattices.

Finally, the results of this work along with the bounds of [4] on the trellis complexity of lattices can be used to compare the RCS and trellis methods for any particular lattice with a finite trellis diagram.

APPENDIX
AN INDEPENDENT PROOF OF COROLLARY 1

Corollary 1: The densest lattices have ESM.

Proof: Suppose L to be an arbitrary n-D lattice with successive minima $\lambda_1, \cdots, \lambda_n$. A famous result of Minkowski implies that $\lambda_1 \cdots \lambda_n \leq \det(L)\gamma_n^{n/2}$ [12, p. 195]. Combining this inequality with the fact that $\lambda_1 \leq \lambda_i$ for $1 \leq i \leq n-1$, results in $\lambda_1^{n-1}\lambda_n \leq \det(L)\gamma_n^{n/2}$. Dividing both sides of the last inequality by λ_1^n and using (7), we obtain $\lambda_n/\lambda_1 \leq \{\gamma_n/\gamma(L)\}^{n/2}$. For the densest lattices, we have $\gamma(L) = \gamma_n$, and the inequality results in $\lambda_n \leq \lambda_1$. Comparing this with (1) proves the corollary. □

ACKNOWLEDGMENT

The authors wish to thank Dr. G. D. Forney for his helpful comments which greatly improved the presentation of the paper. They also wish to thank the anonymous referees for their useful comments and suggestions. In particular, the paper was considerably improved, both in presentation and content, because of the constructive comments and criticism of one of the referees. We are also grateful to Dr. I. F. Blake for helpful discussions and suggestions.

REFERENCES

[1] S. Arora, L. Babai, J. Stern, and Z. Sweedyk, "The hardness of approximate optima in lattices, codes, and systems of linear equations," in *Proc. 34th IEEE Symp. on Foundations of Computer Science*, 1993, pp. 724–733.

[2] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica 6*, pp. 1–13, 1986.

[3] A. H. Banihashemi, "Decoding complexity and trellis structure of lattices," Ph.D. dissertation, E&CE Dept., Univ. of Waterloo, Waterloo, Ont., Canada, 1997.

[4] A. H. Banihashemi and I. F. Blake, "Trellis complexity and minimal trellis diagrams of lattices," submitted to *IEEE Trans. Inform. Theory*, Oct. 1996.

[5] ———, "On the trellis complexity of root lattices and their duals," submitted to *IEEE Trans. Inform. Theory*, Apr. 1997.

[6] A. H. Banihashemi and A. K. Khandani, "Lattice decoding using the Korkin-Zolotarev reduced basis," Tech. Reps. UW-E and CE 95-12, Elec. Comput. Eng. Dept., Univ. of Waterloo, Waterloo, Ont., Canada, 1995.

[7] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd ed. New York: Springer-Verlag, 1993.

[8] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, 1973.

[9] ———, "Coset codes—Part II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, Sept. 1988.

[10] ———, "Density/length profiles and trellis complexity of lattices," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1753–1772, Nov. 1994.

[11] G. D. Forney, Jr. and M. D. Trott, "The dynamics of group codes: State spaces, trellis diagrams, and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, Sept. 1993.

[12] P. M. Gruber and C. G. Lekkerkerker, *Geometry of Numbers*, 2nd ed. Amsterdam, The Netherlands: Elsevier, 1987.

[13] J. Hastad, "Dual vectors and lower bounds for the nearest lattice point problem," *Combinatorica*, 8, pp. 75–81, 1988.

[14] B. Helfrich, "Algorithms to construct Minkowski reduced and Hermite reduced lattice bases," *Theor. Comput. Sci.*, vol. 41, pp. 125–139, 1985.

[15] R. Kannan, "Improved algorithms on integer programming and related lattice problems," in *Proc. 15th Annu. ACM Symp. on Theory of Computing*, 1983, pp. 193–206.

[16] ———, "Minkowski's convex body theorem and integer programming," *Math. of Operations Res.*, vol. 12, pp. 415–440, Aug. 1987.

[17] A. Korkin and G. Zolotarev, "Sur les formes quadratiques," *Math. Ann.*, vol. 6, pp. 366–389, 1873.

[18] J. C. Lagarias, H. W. Lenstra, and C. P. Schnorr, "Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal," *Combinatorica*, 10, pp. 333–348, 1990.

[19] H. W. Lenstra, "Integer programming with a fixed number of variables," *Math. of Operations Res.*, vol. 8, pp. 538–548, Nov. 1983.

[20] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Annalen*, vol. 261, pp. 513–534, 1982.

[21] L. Lovász, "An algorithmic theory of numbers, graphs and convexity," in *NSF-CBMS Regional Conference Series in Applied Mathematics 50*. Philadelphia, PA: SIAM, 1986.

[22] S. S. Ryškov, "Density of an (r, R)-system," *Mat. Zametki*, vol. 16, no. 3, pp. 447–454, Sept. 1974; English translation in *Math. Notes of the Academy of Sciences of the USSR*, vol. 16, pp. 855–858, 1975.

[23] C. P. Schnorr, "A hierarchy of polynomial time lattice basis reduction algorithms," *Theor. Comput. Sci.*, vol. 53, pp. 201–224, 1987.

[24] A. Schrijver, *Theory of Linear and Integer Programming*. New York: Wiley, 1986.

[25] V. Tarokh and I. F. Blake, "Trellis complexity versus the coding gain of lattices I," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1796–1807, Nov. 1996.

[26] ———, "Trellis complexity versus the coding gain of lattices II," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1808–1816, Nov. 1996.

[27] V. Tarokh and A. Vardy, "Upper bounds on trellis complexity of lattices," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1294–1300, July 1997.

[28] P. van Emde Boas, "Another NP-complete partition problem and the complexity of computing short vectors in a lattice," Rep. 81-04, Dept. of Math., Univ. of Amsterdam, Amsterdam, The Netherlands, 1981.

[29] A. Vardy, "Even more efficient bounded-distance decoding of the hexacode, the Golay code, and the Leech lattice," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1495–1499, Sept. 1995.

[30] "Special Issue on Codes and Complexity," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1649–2057, Nov. 1996.