



**HAL**  
open science

# Decoding by Embedding: Correct Decoding Radius and DMT Optimality

Cong Ling, Shuiyin Liu, Laura Luzzi, Damien Stehlé

► **To cite this version:**

Cong Ling, Shuiyin Liu, Laura Luzzi, Damien Stehlé. Decoding by Embedding: Correct Decoding Radius and DMT Optimality. IEEE International Symposium on Information Theory (ISIT'11), Jul 2011, Saint-Petersburg, Russia. 5 p. hal-00648134

**HAL Id: hal-00648134**

**<https://centralesupelec.hal.science/hal-00648134>**

Submitted on 5 Dec 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Decoding by Embedding: Correct Decoding Radius and DMT Optimality

Cong Ling and Shuiyin Liu  
Department of Electrical and  
Electronic Engineering  
Imperial College London  
London, SW7 2AZ, UK  
c.ling, shuiyin.liu06@imperial.ac.uk

Laura Luzzi  
Alcatel-Lucent Chair on Flexible Radio,  
Supélec,  
91192, Gif-Sur-Yvette, France  
laura.luzzi@supélec.fr

Damien Stehlé  
CNRS/ENS Lyon  
Laboratoire LIP  
46 Allée d'Italie F69364  
Lyon Cedex 07, France  
damien.stehle@gmail.com

**Abstract**—In lattice-coded multiple-input multiple-output (MIMO) systems, optimal decoding amounts to solving the closest vector problem (CVP). *Embedding* is a powerful technique for the approximate CVP, yet its remarkable performance is not well understood. In this paper, we analyze the embedding technique from a *bounded distance decoding* (BDD) viewpoint.  $1/(2\gamma)$ -BDD is referred to as a decoder that finds the closest vector when the noise norm is smaller than  $\lambda_1/(2\gamma)$ , where  $\lambda_1$  is the minimum distance of the lattice. We prove that the Lenstra, Lenstra and Lovász (LLL) algorithm can achieve  $1/(2\gamma)$ -BDD for  $\gamma \approx O(2^{n/4})$ . This substantially improves the existing result  $\gamma = O(2^n)$  for embedding decoding. We also prove that BDD of the regularized lattice is optimal in terms of the diversity-multiplexing gain tradeoff (DMT).

## I. INTRODUCTION

Lattice decoding for the linear multiple-input multiple-output (MIMO) channel is a problem of high relevance in multi-antenna, broadcast, cooperative and other multi-terminal communication systems [1]. Maximum-likelihood (ML) decoding for a lattice can be realized efficiently by sphere decoding [2], whose complexity can however grow prohibitively with the dimension  $n$ . The decoding complexity is especially high in the case of coded or distributed systems, where the lattice dimension is usually larger. Thus, the practical implementation of decoders often has to resort to approximate solutions, which mostly fall under two main strategies. One is to reduce the complexity of sphere decoding, while another is lattice reduction-aided decoding. The latter in essence applies zero-forcing (ZF), successive interference cancellation (SIC) or other suboptimal receivers to a reduced basis of the lattice [3]. It is known that regularized lattice-reduction aided decoding can achieve the optimal diversity and multiplexing tradeoff (DMT) [4].

However, lattice-reduction-aided decoding exhibits a widening gap to (infinite) lattice decoding [5], and thus there is a strong demand for computationally efficient suboptimal decoding algorithms that offer improved performance. Several such approaches are emerging, including list decoding, sampling [6] and embedding [7]. It was shown in [6] that the sampling technique can provide a constant improvement in the signal-to-noise ratio (SNR) gain at polynomial complexity. In

sharp contrast, no theoretic improvement has been proved for embedding, despite its remarkable performance in simulation. This is the motivation of this paper.

The decoding problem considered in [7] can be viewed as a variant of the CVP known as  $1/(2\gamma)$ -bounded distance decoding (BDD), where the closest vector is found under the assumption that the noise norm is small compared to the minimum distance  $\lambda_1$  of the lattice, i.e., no more than  $\lambda_1/(2\gamma)$ .

In this paper, we prove that the embedding technique can reduce  $1/(2\gamma)$ -BDD to the  $\gamma$ -unique shortest vector problem (uSVP). Note that the problems are harder for smaller values of  $\gamma$ . On the algorithmic side, we show that  $\gamma$ -uSVP for  $\gamma = O(2^{n/4})$  can be solved by the Lenstra, Lenstra and Lovász (LLL) algorithm. This is a new result of independent interest, which is stronger than the usual bound  $\gamma = O(2^{n/2})$  in literature. Combining the two results, we prove that embedding decoding using the LLL algorithm can solve  $1/(2\gamma)$ -BDD for  $\gamma \approx O(2^{n/4})$ . This is significantly better than the bound  $\gamma = O(2^n)$  proven in [7]. It should be mentioned that these are worst-case bounds; the actual decoding performance is often better.

Moreover, we prove that the regularized BDD is DMT-optimal. This represents a nontrivial extension of the analysis in [4] for  $C$ -approximation algorithms of CVP. Indeed, it is easy to see that  $C$ -approximate algorithms are a special case of BDD, because any decoding technique which provides a  $C$ -approximate CVP solution is also able to solve  $1/(2C)$ -BDD. However, the converse is not necessarily true.

The paper is organized as follows: Section II presents the transmission model and lattice decoding. In Section III the decoding radius of embedding decoding is analyzed. The DMT analysis of BDD is given in Section IV. Section V evaluates the performance by computer simulation.

## II. LATTICE CODING AND DECODING

### A. System Model

Consider an  $n_T \times n_R$  flat-fading MIMO system model consisting of  $n_T$  transmitters and  $n_R$  receivers

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (1)$$

where  $\mathbf{X} \in \mathbb{C}^{n_T \times T}$ ,  $\mathbf{Y}$ ,  $\mathbf{N} \in \mathbb{C}^{n_R \times T}$  of block length  $T$  denote the channel input, output and noise, respectively, and  $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$  is the  $n_R \times n_T$  full-rank channel gain matrix with  $n_R \geq n_T$ , all of its elements are i.i.d. complex Gaussian random variables  $\mathcal{CN}(0, 1)$ . The entries of  $\mathbf{N}$  are i.i.d. complex Gaussian with variance  $\sigma^2$  each. The codewords  $\mathbf{X}$  satisfy the average power constraint  $E[\|\mathbf{X}\|_F^2/T] = 1$ . Hence, the signal-to-noise ratio (SNR) at each receive antenna is  $1/\sigma^2$ .

When a lattice space-time block code is employed, the QAM information vector  $\mathbf{x}$  is multiplied by the generator matrix  $\mathbf{G}$  of the encoding lattice. An  $n_T \times T$  codeword matrix  $\mathbf{X}$  is formed by column-wise stacking of consecutive  $n_T$ -tuples of the vector  $\mathbf{s} = \mathbf{G}\mathbf{x} \in \mathbb{C}^{n_T T}$ . By column-by-column vectorization of the matrices  $\mathbf{Y}$  and  $\mathbf{N}$  in (1), i.e.,  $\mathbf{y} = \text{Vec}(\mathbf{Y})$  and  $\mathbf{n} = \text{Vec}(\mathbf{N})$ , the received signal at the destination can be expressed as

$$\mathbf{y} = (\mathbf{I}_T \otimes \mathbf{H}) \mathbf{G}\mathbf{x} + \mathbf{n}. \quad (2)$$

When  $T = 1$  and  $\mathbf{G} = \mathbf{I}_{n_T}$ , (2) reduces to the model for uncoded MIMO communication  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ . Furthermore, by separating real and imaginary parts, we obtain the equivalent  $2n_T \times 2n_R$  real-valued model

$$\begin{bmatrix} \Re \mathbf{y} \\ \Im \mathbf{y} \end{bmatrix} = \begin{bmatrix} \Re \mathbf{H} & -\Im \mathbf{H} \\ \Im \mathbf{H} & \Re \mathbf{H} \end{bmatrix} \begin{bmatrix} \Re \mathbf{x} \\ \Im \mathbf{x} \end{bmatrix} + \begin{bmatrix} \Re \mathbf{n} \\ \Im \mathbf{n} \end{bmatrix}. \quad (3)$$

The QAM constellations  $\mathcal{C}$  can be interpreted as the shifted and scaled version of a finite subset  $\mathcal{A}^{n_T}$  of the integer lattice  $\mathbb{Z}^{n_T}$ , i.e.,  $\mathcal{C} = a(\mathcal{A}^{n_T} + [1/2, \dots, 1/2]^T)$ , where the factor  $a$  arises from energy normalization. For example, we have  $\mathcal{A}^{n_T} = \{-\sqrt{M}/2, \dots, \sqrt{M}/2 - 1\}$  for  $M$ -QAM signalling.

Therefore, with scaling and shifting, we consider the generic  $n \times m$  ( $m \geq n$ ) real-valued MIMO system model

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n}, \quad (4)$$

where  $\mathbf{B} \in \mathbb{R}^{m \times n}$ , can be interpreted as the basis matrix of the decoding lattice. Obviously,  $n = 2n_T T$  and  $m = 2n_R T$ . The data vector  $\mathbf{x}$  is drawn from a finite subset  $\mathcal{A}^n \subset \mathbb{Z}^n$  to satisfy the power constraint.

### B. Lattice Decoding

An  $n$ -dimensional lattice in the  $m$ -dimensional Euclidean space  $\mathbb{R}^m$  ( $n \leq m$ ) is the set of integer linear combinations of  $n$  independent vectors  $\mathbf{b}_1, \dots, \mathbf{b}_n \in \mathbb{R}^m$ :

$$\mathcal{L}(\mathbf{B}) = \left\{ \sum_{i=1}^n x_i \mathbf{b}_i \mid x_i \in \mathbb{Z}, i = 1, \dots, n \right\}.$$

The matrix  $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_n]$  is a basis of the lattice  $\mathcal{L}(\mathbf{B})$ . In matrix form,  $\mathcal{L}(\mathbf{B}) = \{\mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}$ . For any point  $\mathbf{y} \in \mathbb{R}^m$  and any lattice  $\mathcal{L}(\mathbf{B})$ , the distance of  $\mathbf{y}$  to the lattice is  $\text{dist}(\mathbf{y}, \mathcal{L}(\mathbf{B})) = \min_{\mathbf{x} \in \mathbb{Z}^n} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|$ . A *shortest vector* of a lattice  $\mathcal{L}(\mathbf{B})$  is a non-zero vector in  $\mathcal{L}(\mathbf{B})$  with the smallest  $l_2$  norm. The length of the shortest vector, often referred to as the *minimum distance*, of  $\mathcal{L}(\mathbf{B})$  is denoted by  $\lambda_1(\mathbf{B})$ .

We now give precise definitions for the lattice problems that are central to this work.

### - Shortest Vector Problem (SVP):

Given a lattice  $\mathcal{L}(\mathbf{B})$ , find a non-zero vector  $\mathbf{v} \in \mathcal{L}(\mathbf{B})$  of norm  $\lambda_1(\mathbf{B})$ .

### - Approximate Shortest Vector Problem (ApproxSVP):

Given a lattice  $\mathcal{L}(\mathbf{B})$  and an approximation factor  $C \geq 1$ , find a non-zero vector  $\mathbf{v} \in \mathcal{L}(\mathbf{B})$  of norm smaller than  $C\lambda_1(\mathbf{B})$ .

### - $\gamma$ -unique Shortest Vector Problem ( $\gamma$ -uSVP):

Given a lattice  $\mathcal{L}(\mathbf{B})$  such that  $\lambda_2(\mathbf{B}) > \gamma\lambda_1(\mathbf{B})$ , find a non-zero vector  $\mathbf{v} \in \mathcal{L}(\mathbf{B})$  of norm  $\lambda_1(\mathbf{B})$ .

### - $1/(2\gamma)$ -Bounded Distance Decoding ( $\beta$ -BDD):

Given a lattice  $\mathcal{L}(\mathbf{B})$  and a vector  $\mathbf{y}$  such that  $\text{dist}(\mathbf{y}, \mathcal{L}(\mathbf{B})) < 1/(2\gamma)\lambda_1(\mathbf{B})$ , find the lattice vector  $\mathbf{B}\hat{\mathbf{x}} \in \mathcal{L}(\mathbf{B})$  closest to  $\mathbf{y}$ .

A lattice has infinitely many bases. In general, every matrix  $\bar{\mathbf{B}} = \mathbf{B}\mathbf{U}$ , where  $\mathbf{U}$  is an *unimodular* matrix, i.e.,  $\det(\mathbf{U}) = \pm 1$  and all elements of  $\mathbf{U}$  are integers, is also a basis of  $\mathcal{L}(\mathbf{B})$ . The celebrated LLL algorithm [8] is the first polynomial-time algorithm of lattice reduction which finds a vector not much longer than the shortest nonzero vector. Let  $\mathbf{B} = \mathbf{Q}\mathbf{R}$  be the QR decomposition, where  $\mathbf{Q}$  has orthogonal columns and  $\mathbf{R}$  is an upper triangular matrix with nonnegative diagonal elements  $r_{i,i}$  for  $i = 1, \dots, n$ . An LLL-reduced basis  $\mathbf{B}$  has the following properties [8]:

$$r_{j,j} \leq \alpha^{(i-j)/2} r_{i,i} \quad (5)$$

for  $1 \leq j < i \leq n$ , and

$$\alpha^{-(n-1)/2} \lambda_1(\mathbf{B}) \leq \min_{1 \leq i \leq n} r_{i,i} \leq \lambda_1(\mathbf{B}) \quad (6)$$

where  $\alpha = 1/(\delta - 1/4)$ ,  $1/4 < \delta \leq 1$ . We have  $\alpha = 2$  for the most common value  $\delta = 3/4$ .

Babai's nearest plane algorithm [3] or LLL-SIC decoding, combining lattice reduction and SIC, can be viewed as the most basic BDD. The correct decoding radius of SIC is given by [5]

$$R_{\text{SIC}} = \frac{1}{2} \min_{1 \leq i \leq n} r_{i,i}, \quad (7)$$

which means that correct decoding is guaranteed if  $\|\mathbf{n}\| \leq R_{\text{SIC}}$ .

### III. DECODING RADIUS OF EMBEDDING DECODING

The core of the embedding technique is that basis matrix  $\mathbf{B}$  and the received vector  $\mathbf{y}$  are embedded in a higher dimensional lattice. More precisely, we consider the following  $(m+1) \times (n+1)$  basis matrix [9]

$$\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{B} & -\mathbf{y} \\ \mathbf{0}_{1 \times n} & t \end{bmatrix} \quad (8)$$

where  $t > 0$  is a parameter to be determined. The strategy is to reduce CVP to SVP in the following way: for a suitable choice of  $t$  and for sufficiently small noise norm,  $\mathbf{v} = [(\mathbf{B}\mathbf{x} - \mathbf{y})^T \ t]^T$  is the shortest vector in the lattice  $\mathcal{L}(\tilde{\mathbf{B}})$ ; thus an

SVP algorithm will find it, and the message  $\mathbf{x}$  can be easily recovered from the coordinates of this vector in the basis  $\tilde{\mathbf{B}}$ :

$$\text{If } \mathbf{v} = \tilde{\mathbf{B}} \begin{pmatrix} \mathbf{x}' \\ q \end{pmatrix} = \begin{pmatrix} \mathbf{B}\mathbf{x}' - q\mathbf{y} \\ qt \end{pmatrix}, \text{ then } \hat{\mathbf{x}} = \mathbf{x}' \ (q = \pm 1). \quad (9)$$

At the same time,  $t$  should not be too small or too large, otherwise  $[(\mathbf{B}\mathbf{x} - \mathbf{y})^T \ t]^T$  might not be the shortest vector.

Luzzi *et al.* [7] chose  $t = \frac{1}{2\sqrt{2}\alpha^{n/2}} \min_{1 \leq i \leq n} r_{i,i}$  and used the LLL algorithm to find the shortest vector in the lattice  $\mathcal{L}(\tilde{\mathbf{B}})$ . Their scheme, under the term *augmented lattice reduction* (ALR), was shown to achieve the correct decoding radius

$$R_{\text{ALR}} = \frac{1}{2\sqrt{2}\alpha^{n-\frac{1}{2}}} \lambda_1(\mathbf{B}). \quad (10)$$

In the following subsections, we will improve this bound.

### A. Correct Decoding Radius for General $t$

In [10], it is proved that by choosing  $t = \text{dist}(\mathbf{y}, \mathbf{B})$ , the embedding technique can reduce  $1/(2\gamma)$ -BDD to  $\gamma$ -uSVP. In this subsection, we will show that one can achieve the same correct decoding radius by setting  $t \triangleq \frac{1}{2\gamma} \lambda_1(\mathbf{B})$ , thus bypassing the assumption of  $\text{dist}(\mathbf{y}, \mathbf{B})$  in [10].

*Theorem 1 (Decoding Radius for Embedding):* Applying  $\gamma$ -uSVP ( $\gamma \geq 1$ ) to the extended lattice (8) with parameter  $t$  ( $0 < t < \lambda_1(\mathbf{B})/\gamma$ ) and computing the estimate (9) guarantees a correct decoding radius

$$R_{\text{Emb}} = \sqrt{\frac{t}{\gamma} \lambda_1(\mathbf{B}) - t^2} \quad (11)$$

whose maximum is

$$R_{\text{Emb}} = \frac{1}{2\gamma} \lambda_1(\mathbf{B}) \quad (12)$$

obtained by setting  $t \triangleq \frac{1}{2\gamma} \lambda_1(\mathbf{B})$ .

The proof of Theorem 1 uses the following lemma.

*Lemma 1:* Let  $\tilde{\mathbf{B}}$  be the matrix defined in (8), and let  $0 < t < \frac{1}{\gamma} \lambda_1(\mathbf{B})$ , with  $\gamma \geq 1$ . Suppose that

$$\|\mathbf{y} - \mathbf{B}\mathbf{x}\| \leq \sqrt{\frac{t}{\gamma} \lambda_1(\mathbf{B}) - t^2},$$

then  $\mathbf{v} = \begin{pmatrix} \mathbf{B}\mathbf{x} - \mathbf{y} \\ t \end{pmatrix} = \begin{pmatrix} -\mathbf{n} \\ t \end{pmatrix}$  is a  $\gamma$ -unique shortest vector of  $\mathcal{L}(\tilde{\mathbf{B}})$ .

*Proof:* Let  $\tilde{\mathbf{B}}$  be the matrix defined in (8), and let  $\mathbf{w}$  be an arbitrary nonzero vector in  $\mathcal{L}(\mathbf{B})$ . Any vector in  $\mathcal{L}(\tilde{\mathbf{B}})$  that is not a multiple of  $\mathbf{v}$  can be represented by  $\mathbf{w}' = \mathbf{w} + q\mathbf{v}$ , with  $q \in \mathbb{Z}$  and  $\mathbf{w} \in \mathcal{L}(\mathbf{B})$ . We will show that  $\|\mathbf{w}'\| \geq \gamma\|\mathbf{v}\|$ . The norm of  $\mathbf{w}'$  can be written as

$$\|\mathbf{w}'\| = \sqrt{\|\mathbf{w} - q\mathbf{n}\|^2 + (qt)^2}.$$

If  $\|q\mathbf{n}\| \leq \lambda_1(\mathbf{B})$ , using the triangular inequality, we have the lower bound

$$\begin{aligned} \|\mathbf{w}'\| &\geq \sqrt{(\lambda_1(\mathbf{B}) - q\|\mathbf{n}\|)^2 + (qt)^2} \\ &= \sqrt{\lambda_1(\mathbf{B})^2 - 2q\lambda_1(\mathbf{B})\|\mathbf{n}\| + q^2\|\mathbf{n}\|^2 + q^2t^2} \\ &\geq \frac{\lambda_1(\mathbf{B})t}{\sqrt{\|\mathbf{n}\|^2 + t^2}}. \end{aligned}$$

If  $\|q\mathbf{n}\| > \lambda_1(\mathbf{B})$ , we can also obtain the same bound because

$$\|\mathbf{w}'\| \geq qt > \frac{\lambda_1(\mathbf{B})t}{\|\mathbf{n}\|} \geq \frac{\lambda_1(\mathbf{B})t}{\sqrt{\|\mathbf{n}\|^2 + t^2}}.$$

We need to make sure that  $\|\mathbf{w}'\| > \gamma\|\mathbf{v}\|$ , so

$$\frac{\lambda_1(\mathbf{B})t}{\sqrt{\|\mathbf{n}\|^2 + t^2}} > \gamma\sqrt{\|\mathbf{n}\|^2 + t^2}$$

which implies that

$$\begin{aligned} \|\mathbf{n}\|^2 = \|\mathbf{B}\mathbf{x} - \mathbf{y}\|^2 &< \frac{t}{\gamma} \lambda_1(\mathbf{B}) - t^2 \\ &= -\left(t - \frac{\lambda_1(\mathbf{B})}{2\gamma}\right)^2 + \left(\frac{\lambda_1(\mathbf{B})}{2\gamma}\right)^2 \\ &\leq \left(\frac{\lambda_1(\mathbf{B})}{2\gamma}\right)^2. \end{aligned}$$

where the equality holds if  $t = \frac{\lambda_1(\mathbf{B})}{2\gamma}$ . ■

Due to the well known fact that the LLL algorithm can solve  $\gamma$ -uSVP with  $\gamma = \alpha^{n/2}$  for the basis (8) of dimension  $n+1$  [8], one can obtain the correct decoding radius

$$R_{\text{Emb}} = \frac{1}{2\alpha^{n/2}} \lambda_1(\mathbf{B}) \quad (13)$$

by choosing  $t = t_0 \triangleq \frac{1}{2\alpha^{n/2}} \lambda_1(\mathbf{B})$ . This decoding radius improves the bound (10) from [7]. Yet, there is still room to improve. The reason is that the estimate  $\gamma = \alpha^{n/2}$  is pessimistic for  $\gamma$ -uSVP. In fact,  $\alpha^{n/2}$  is just the approximation factor for ApproxSVP achieved by LLL. Any algorithm solving  $\gamma$ -ApproxSVP necessarily solves  $\gamma$ -uSVP, while the converse is not true.

### B. Correct Decoding Radius Achieved by LLL

In this subsection, we will show that LLL can in fact solve  $\gamma$ -uSVP with a smaller  $\gamma$ .

*Lemma 2 (LLL for uSVP):* The LLL algorithm can solve  $\gamma$ -uSVP for  $\gamma = \max_{1 \leq i \leq n-1} \{\sqrt{\gamma_i}\} \alpha^{n/4}$  in an  $n$ -dimensional lattice  $\mathcal{L}(\mathbf{B})$ , where  $\gamma_i$  is the Hermite constant for  $i$ -dimensional lattices.

*Proof:* Suppose that  $\mathbf{B}$  is an LLL-reduced basis, and that  $\lambda_2(\mathbf{B}) > \max_{1 \leq i \leq n-1} \{\sqrt{\gamma_i}\} \alpha^{n/4} \lambda_1(\mathbf{B})$ . We will prove that the first vector output by LLL,  $\mathbf{b}_1$ , is the shortest vector  $\mathbf{v}$ . By contradiction, suppose that  $\mathbf{b}_1 \neq \pm\mathbf{v}$ . Note that  $\mathbf{b}_1$  cannot be a multiple of  $\mathbf{v}$ , or  $\mathbf{B}$  would not be a basis. We may write

$$\mathbf{v} = \sum_{i=1}^k x_i \mathbf{b}_i,$$

where  $x_i$  is an integer and  $k$  is the largest  $i$  such that  $x_i$  is not zero. Then we have  $\lambda_1(\mathbf{B}) = \|\mathbf{v}\| \geq r_{k,k}$ , where  $\mathbf{B} = \mathbf{QR}$  is the QR decomposition of  $\mathbf{B}$ . Using the assumption that  $\mathbf{b}_1 \neq \pm\mathbf{v}$ , we have that  $k > 1$ . On the other hand, we have the following bound for the second minimum  $\lambda_2(\mathbf{B})$

$$\lambda_2(\mathbf{B}) \leq \lambda_1(\mathcal{L}[\mathbf{b}_1, \dots, \mathbf{b}_{k-1}]), \quad k > 1.$$

In fact  $\lambda_2(\mathbf{B})$  must be smaller than the norm of the shortest nonzero vector in the sublattice spanned by  $\{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\}$ , since these vectors are linearly independent with  $\mathbf{v}$ . The fact that  $k > 1$  ensures that there are non-zero vectors in  $\mathcal{L}(\{\mathbf{b}_1, \dots, \mathbf{b}_{k-1}\})$ .

Using Minkowski's first theorem [11], we obtain

$$\begin{aligned} \lambda_2(\mathbf{B}) &\leq \sqrt{\gamma_{k-1}} \det(\mathcal{L}[\mathbf{b}_1, \dots, \mathbf{b}_{k-1}])^{1/(k-1)} \\ &= \sqrt{\gamma_{k-1}} \left( \prod_{i=1}^{k-1} r_{i,i} \right)^{1/(k-1)} \\ &\leq \sqrt{\gamma_{k-1}} r_{k,k} \left( \prod_{i=1}^{k-1} \alpha^{(k-i)/2} \right)^{1/(k-1)} \\ &= \sqrt{\gamma_{k-1}} \alpha^{k/4} r_{k,k} \\ &\leq \max_{1 \leq i \leq n-1} \{ \sqrt{\gamma_i} \} \alpha^{n/4} \lambda_1(\mathbf{B}), \end{aligned}$$

where the inequality  $r_{i,i} \leq \alpha^{(k-i)/2} r_{k,k}$  for  $1 \leq i < k$  follows from (5). The reason why we use  $\max_{1 \leq i \leq n-1} \{ \sqrt{\gamma_i} \}$  instead of  $\gamma_{n-1}$  in the last step is that it is not known whether  $\gamma_i$  is an increasing function. The last statement is a contradiction because we assumed  $\lambda_2(\mathbf{B}) > \max_{1 \leq i \leq n-1} \{ \sqrt{\gamma_i} \} \alpha^{n/4} \lambda_1(\mathbf{B})$ . Therefore,  $\mathbf{b}_1 = \pm\mathbf{v}$ .  $\blacksquare$

Lemma 2 leads to the following result:

*Theorem 2 (Decoding Radius of Embedding using LLL):*

Applying the LLL algorithm to the embedding problem can achieve the correct decoding radius

$$R_{\text{LLL-Emb}} = \frac{1}{2 \max_{1 \leq i \leq n} \{ \sqrt{\gamma_i} \} \alpha^{(n+1)/4}} \lambda_1(\mathbf{B}) \quad (14)$$

by choosing  $t = t_0 \triangleq \frac{\lambda_1(\mathbf{B})}{2 \max_{1 \leq i \leq n} \{ \sqrt{\gamma_i} \} \alpha^{(n+1)/4}}$ .

This is exponentially better than (10). Since the LLL algorithm has polynomial complexity with respect to  $n$ , the embedding decoder also has polynomial complexity (assuming  $\lambda_1(\mathbf{B})$  has been found in the pre-processing stage).

#### IV. DMT ANALYSIS OF BDD

In this section we will prove that, similarly to LLL reduction-aided ZF and SIC decoding, BDD (including embedding decoding) is optimal from the point of view of DMT [12] when a suitable left preprocessing is employed.

In the present discussion, we suppose for the sake of simplicity that  $m = n$ . Following Jaldén and Elia's notation in [4], we consider the equivalent normalized channel model where the noise variance is equal to 1:

$$\mathbf{y}' = \mathbf{B}'\mathbf{x} + \mathbf{n}',$$

where  $\mathbf{B}' = \sqrt{\rho}\mathbf{B}$ ,  $n'_i = \sqrt{\rho}n_i \sim \mathcal{N}(0, 1)$ ,  $\forall i = 1, \dots, n$ . Here  $\rho = 1/\sigma^2$  denotes the SNR. Moreover, we consider the equivalent regularized system

$$\mathbf{y}_1 = \mathbf{R}\mathbf{x} + \mathbf{n}_1, \quad (15)$$

where

$$\begin{pmatrix} \mathbf{B}' \\ \mathbf{I}_{n \times n} \end{pmatrix} = \mathbf{QR}, \quad \mathbf{y}_1 = \mathbf{Q}^\dagger \begin{pmatrix} \mathbf{y}' \\ \mathbf{0}_{n \times 1} \end{pmatrix}.$$

From the point of view of receiver architecture, this amounts to performing left preprocessing before decoding, by using a maximum mean square error generalized decision-feedback equalizer (MMSE-GDFE). We can show that DMT-optimality holds for all instances of BDD by following the same reasoning of the original proof in [4].

*Theorem 3:* For any constant  $\eta > 0$ , the regularized  $\eta$ -BDD is DMT-optimal.

*Proof:* Let  $d_{\text{ML}}(r)$  be the optimal diversity gain corresponding to a multiplexing gain  $r \in \{0, \dots, \min(n_T, n_R)\}$ . Using the same notation as [4], we consider the constellation  $\Lambda_r \cap \mathcal{R}$ , where the lattice  $\Lambda_r = \rho^{-\frac{rT}{n}} \mathbb{Z}^n$  is scaled according to the SNR, and  $\mathcal{R}$  is a fixed shaping region<sup>1</sup>. Let  $\mathcal{B} \subset \mathcal{R}$  be a ball of fixed radius  $R$ , where  $R$  is chosen in such a way that  $\mathbf{d}_1 + \mathbf{d}_2 \in \mathcal{R}$ ,  $\forall \mathbf{d}_1, \mathbf{d}_2 \in \mathcal{B}$ . Let

$$\nu_r = \min_{\substack{\mathbf{d} \in \mathcal{B} \cap \Lambda_r \\ \mathbf{d} \neq \mathbf{0}}} \frac{1}{4} \|\mathbf{B}'\mathbf{d}\|^2.$$

Then Lemma 1 of [4] holds, that is

$$\limsup_{\rho \rightarrow \infty} \frac{\log P\{\nu_r \leq 1\}}{\log \rho} \leq -d_{\text{ML}}(r).$$

Let  $\zeta > 0$  and choose  $\delta$  such that  $\frac{2\zeta T}{n} > \delta > 0$ . We have  $\Lambda_r = \rho^{\frac{\zeta T}{n}} \Lambda_{r+\zeta}$ . As in the original proof,  $\exists \rho_1$  such that  $\forall \rho \geq \rho_1$ ,  $\mathcal{R} \subseteq \frac{1}{2} \rho^{\frac{\zeta T}{n}} \mathcal{B}$ . As in Theorem 1 from [4], we want to show that the conditions

$$\nu_{r+\zeta} \geq 1, \quad \|\mathbf{n}'\|^2 \leq \rho^\delta \quad (16)$$

are sufficient for the regularized  $\eta$ -BDD to decode correctly for sufficiently large SNR. We need a lower bound for

$$d_{\mathbf{R}}^2 = \min_{\hat{\mathbf{x}} \in \Lambda_r \setminus \{\mathbf{0}\}} \frac{1}{4} \|\mathbf{R}\hat{\mathbf{x}}\|^2 = \min_{\hat{\mathbf{x}} \in \Lambda_r \setminus \{\mathbf{0}\}} \frac{1}{4} \left( \|\mathbf{B}'\hat{\mathbf{x}}\|^2 + \|\hat{\mathbf{x}}\|^2 \right).$$

Let  $\varphi(\hat{\mathbf{x}}) = \|\mathbf{B}'\hat{\mathbf{x}}\|^2 + \|\hat{\mathbf{x}}\|^2$ . Let  $\hat{\mathbf{x}} \in \Lambda_r \setminus \{\mathbf{0}\}$  be any lattice point.

- If  $\hat{\mathbf{x}} \notin \frac{1}{2} \rho^{\frac{\zeta T}{n}} \mathcal{B}$ ,  $\varphi(\hat{\mathbf{x}}) \geq \|\hat{\mathbf{x}}\|^2 > \frac{1}{4} R^2 \rho^{\frac{2\zeta T}{n}}$ .
- If  $\hat{\mathbf{x}} \in \frac{1}{2} \rho^{\frac{\zeta T}{n}} \mathcal{B} \cap \Lambda_r = \frac{1}{2} \rho^{\frac{\zeta T}{n}} \mathcal{B} \cap \rho^{\frac{\zeta T}{n}} \Lambda_{r+\zeta}$ , then  $\hat{\mathbf{x}} \rho^{-\frac{\zeta T}{n}} \in \frac{1}{2} \mathcal{B} \cap \Lambda_{r+\zeta}$  and so  $\frac{1}{4} \left\| \mathbf{B}'\hat{\mathbf{x}} \rho^{-\frac{\zeta T}{n}} \right\|^2 \geq 1$  since by the hypothesis (16),  $\nu_{r+\zeta} \geq 1$ . Therefore  $\varphi(\hat{\mathbf{x}}) \geq \|\mathbf{B}'\hat{\mathbf{x}}\|^2 \geq 4\rho^{\frac{2\zeta T}{n}}$ .

In conclusion,  $\exists k > 0$  such that  $d_{\mathbf{R}}^2 \geq k\rho^{\frac{2\zeta T}{n}}$ .

<sup>1</sup>Note that the generator matrix of the lattice code has been absorbed into  $\mathbf{R}$ , hence  $\Lambda_r$  is just a scaled version of  $\mathbb{Z}^n$ .

Now consider the transmitted codeword  $\mathbf{x} \in \Lambda_r \cap \mathcal{R}$ . The regularized  $\eta$ -BDD decoder is able to decode correctly provided that  $\|\mathbf{y}_1 - \mathbf{R}\mathbf{x}\| < \eta d_{\mathbf{R}}$ . We have

$$\|\mathbf{y}_1 - \mathbf{R}\mathbf{x}\| = \|\mathbf{y}' - \mathbf{B}'\mathbf{x}\|^2 + \|\mathbf{x}\|^2 = \|\mathbf{n}'\|^2 + \|\mathbf{x}\|^2 \leq \rho^\delta + c,$$

where  $c = \max_{\mathbf{r} \in \mathcal{R}} \|\mathbf{r}\|^2$  is a constant. Therefore under the conditions (16), the regularized  $\eta$ -BDD decoder is able to decode correctly provided that  $\rho^\delta + c < \eta k \rho^{\frac{2\zeta T}{n}}$ . But  $\delta < \frac{2\zeta T}{n}$ , so  $\exists \bar{\rho}$  such that  $\forall \rho \geq \bar{\rho}$ ,  $\rho^\delta + c < \eta k \rho^{\frac{2\zeta T}{n}}$ . Then as in Theorem 1 from [4] we can conclude that

$$P\{\hat{\mathbf{x}}_{\eta\text{-BDD}} \neq \mathbf{x}\} \leq P\{\nu_{r+\zeta} < 1\} + P\{\|\mathbf{n}'\|^2 > \rho^\delta\},$$

and the second term is negligible for  $\rho \rightarrow \infty$ . So we can say, similarly to the original proof, that

$$\limsup_{\rho \rightarrow \infty} \frac{\log P\{\hat{\mathbf{x}}_{\eta\text{-BDD}} \neq \mathbf{x}\}}{\log \rho} \leq -d_{\text{ML}}(r + \zeta)$$

and then use the right continuity of  $d_{\text{ML}}(r)$ .

## V. EXPERIMENTS AND SUMMARY

In this section we evaluate the performance of embedding decoding proposed in Section III through numerical simulations. For comparison purposes, the performances of lattice reduction aided MMSE-SIC decoding and ML decoding are also shown. We assume perfect channel state information at the receiver. Monte Carlo simulation was used to estimate the bit error rate with Gray mapping and LLL reduction ( $\delta=0.75$ ).

In the simulation, we further enhance embedding decoding by making use of all intermediate lattice vectors during the execution of LLL. Such vectors are generated when size reduction is performed; we can obtain one new vector in each size reduction. We can integrate this into LLL, and the complexity will be of the same order. The size check in LLL is on the lengths of Gram-Schmidt vectors. It is preferable to choose a bit smaller  $t$  so that the last column in (8) can be used as many times as possible. Hence, we choose

$$t_{\text{List-Emb}} = \frac{1}{2\sqrt{\gamma_n} \alpha^{(n+1)/4}} \min_{1 \leq i \leq n} r_{i,i}. \quad (17)$$

The advantage is that the knowledge of  $\lambda_1$  is not required, while the performance is actually a little better due to a larger list.

Fig. 1 shows the bit error rate for an uncoded system with  $n_T = n_R = 10$ , 64-QAM. We found that list MMSE embedding is sufficient to obtain near-optimum performance for uncoded systems with  $n_T = n_R = 10$ ; the SNR loss is less than 1.2 dB.

In summary, we have investigated the decoding radius of embedding decoding through the relation between BDD and uSVP. With the knowledge of  $\lambda_1(\mathbf{B})$  which may be obtained by pre-processing, this yields a polynomial-complexity algorithm achieving a correct decoding radius exponentially larger than previously proved. Moreover, we proved that BDD with MMSE-GDFE left processing is DMT-optimal. Due to space

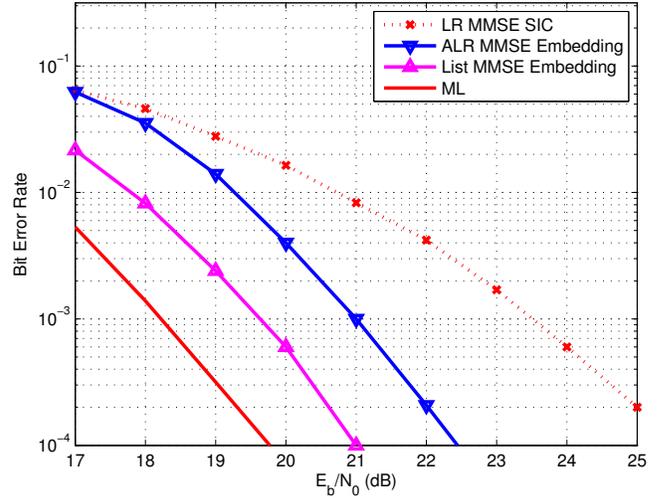


Figure 1. Bit error rate vs. average SNR per bit for the uncoded  $10 \times 10$  system using 64-QAM.

limitation, a rigorous approach that does not require the exact value of  $\lambda_1(\mathbf{B})$  while still retaining polynomial complexity will be reported in the journal version.

## REFERENCES

- [1] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1591–1600, Sep. 1994.
- [2] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1639–1642, Jul. 1999.
- [3] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.
- [4] J. Jaldén and P. Elia, "DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models," *IEEE Trans. Inf. Theory*, vol. 56, pp. 4765–4780, Oct. 2010.
- [5] C. Ling, "On the proximity factors of lattice reduction-aided decoding," *IEEE Trans. Signal Process.*, vol. 59, pp. 2795–2808, Jun. 2011.
- [6] S. Liu, C. Ling, and D. Stehlé, "Randomized lattice decoding: Bridging the gap between lattice reduction and sphere decoding," *IEEE Int. Symp. Inform. Theory (ISIT'10)*, Jun. 2010.
- [7] L. Luzzi, G. R.-B. Othman, and J.-C. Belfiore, "Augmented lattice reduction for MIMO decoding," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 2853–2859, Sep. 2010.
- [8] A. K. Lenstra, J. H. W. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.
- [9] R. Kannan, "Minkowski's convex body theorem and integer programming," *Math. Oper. Res.*, vol. 12, pp. 415–440, Aug. 1987.
- [10] V. Lyubashevsky and D. Micciancio, "On bounded distance decoding, unique shortest vectors, and the minimum distance problem," in *Crypto'09*, Aug. 2009, pp. 577–594.
- [11] H. Minkowski, *Geometrie der Zahlen*, Leipzig, Germany, 1896.
- [12] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49 n.5, pp. 1073 – 1096, 2003.