On the Proximity Factors of Lattice Reduction-Aided Decoding

Cong Ling, *Member, IEEE*

***Abstract*— Lattice reduction-aided decoding (LRAD) features reduced decoding complexity and near-optimum performance in multi-input multi-output communications. In this paper, a quantitative error-rate analysis of LRAD is presented. To this aim, the proximity factors are defined to measure the worst-case losses in decoding distances associated with the decision region relative to infinite lattice decoding (ILD), namely, closest point search in an infinite lattice. Upper bounds on the proximity factors are derived, which are functions of the dimension n of the lattice alone. The study is then extended to the dual-basis reduction. It is found that the bounds for dual basis reduction may be smaller. Reasonably good bounds are derived in many cases. Proximity factors not only imply the same diversity order in fading channels, but also relate the error probabilities of ILD and LRAD.**

***Index Terms*— coding bounds, dual basis, lattice decoding, lattice reduction, minimum distance, multi-input multi-output communications.**

EDICS: MSP-CODR or SPC-PERF

I. Introduction

The linear multi-input multi-output (MIMO) model covers a range of problems in communications, such as code-division multiple access (CDMA), inter-symbol interference (ISI) channels, linearly precoded orthogonal frequency-division multiplexing (OFDM), multi-antenna fading channels with or without linear encoding, multi-antenna broadcast, and cooperative diversity [1]–[4]. For moderate to large problem sizes, (near-) optimum decoding for MIMO systems represents a challenging problem in communication theory. The naive exhaustive search for maximum-likelihood (ML) decoding suffers from exponential complexity. Suboptimum strategies such as zero-forcing (ZF) and successive interference cancelation (SIC) sometimes incurs heavy performance loss. During the past decades, several improved strategies have been developed, especially in the context of multiuser detection and equalization. Sequential decoding [5], branch-and-bound search [6], and semidefinite relaxation [7] represent the state-of-the-art.

The theory of lattices is a powerful, distinctive approach to fast MIMO decoding [8]–[10]. In digital communications, the signal constellation or codebook is often drawn from a lattice, especially as the spectrally efficient high-order quadrature amplitude modulation (QAM) is increasingly used. Lattice decoding exploits the structure of a lattice to significantly reduce the decoding complexity. Decoding corresponds to the closest vector problem (CVP) for a lattice. In general, solving the CVP may consist of two stages: lattice reduction and a local search usually implemented by sphere decoding. While sphere decoding dramatically lowers the decoding complexity at high signal-to-noise ratio (SNR) [11], it could be computationally intensive at low to moderate SNR's. Furthermore, it is known that its complexity grows exponentially with the system size n for any fixed SNR [12]. Thus, to further widen the decoding speed bottleneck, some performance has to be traded for complexity, i.e., the CVP needs to be solved approximately. The problem of solving CVP approximately was first addressed by Babai in [13], which in essence applies ZF or SIC on a reduced lattice. This technique of approximate lattice decoding is referred to as lattice reduction-aided decoding (LRAD) in communications literature [14], [15], where the expense associated with lattice reduction can be shared if the channel matrix keeps constant during a frame of data.

For various reasons, especially to reduce complexity, the lattice decoder may ignore the boundary, which is often referred to as infinite lattice decoding (ILD). ILD will be used as the benchmark for comparisons in this paper. This is because, with minimum mean-square error (MMSE) preprocessing, ILD remarkably achieves the capacity of a point-to-point Gaussian channel [16], and the optimum diversity and multiplexing tradeoff (DMT) of MIMO fading channels [17], [18]. The limitation of ILD without MMSE was pointed out in [19].

Contributions of this paper: This paper presents a quantitative understanding of approximate lattice decoding. More precisely, we shall develop a systematic approach to characterize its performance in terms of proximity factors, i.e., the worst-case loss in decoding distances associated with the decision region relative to ILD, which relates the error probabilities of ILD and approximate lattice decoding. In this paper, the proximity factors are found to be bounded above by a function of the dimension of the lattice alone. In other words, the output of approximate lattice decoding is in proximity to that of lattice decoding. As an alternative to reducing the primal basis, one can reduce the dual basis. This technique has received less attention thus far. In this paper, we find that in some cases dual basis reduction results in a smaller upper bound on the proximity factor. The derived bounds serve two purposes: one is to bound the performance in its own right, the other is to give insights into approximate lattice decoding. The bounds apply to both fixed and random channels, and hold irrespective of

Copyright (c) 2011 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work was presented in part at the *Int. Symp. Turbo Codes/Int. ITG Conf. Source Channel Coding*, Munich, Germany, Apr. 2006, *IEEE Int. Symp. Inform. Theory*, Seattle, US, July 2006, and *IEEE Int. Conf. Commun.*, Beijing, China, May 2008.

C. Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom (e-mail: cling@ieee.org).

fading statistics.

Relations to prior works: A quantitative analysis was first given by Yao and Wornell [14], who showed that the loss is not greater than 3 dB for 2-dimensional complex Gauss reduction. Our idea of quantifying the performance losses using constant bounds in the general case started in [20] which used the LLL reduction in differential lattice decoding. This technique has been subsequently developed and improved in a series of conference papers [21]–[23], eventually leading to the current paper.

Independently, the full receive diversity of Lenstra, Lenstra and Lovász (LLL) reduction in uncoded MIMO fading channels was shown in [24], [25] by using different approaches. We would like to point out that the diversity order derived in those papers does not fully characterize the error performance, as the SNR gap could be arbitrary for the same diversity order. Proximity factors offer an approach to performance analysis complementary to [18], [24], [25] and could deepen our understanding about approximate lattice decoding. The derivations in those papers made use of the orthogonality defect of a lattice, which could also be used to derive upper bounds on proximity factors. However, the bounds derived in this paper are better (see the extended version of this paper [26]).

The development in this paper makes heavy use of existing results in lattice literature, especially those of Babai [13] and Schnorr [46].

Organization: The rest of the paper is organized as follows. Section II describes the system model. Section III presents the basics of lattices and lattice reduction that are essential to the subsequent development of the theory in this paper. In Section IV, we define the proximity factors of LRAD and show the main results. In Sections V and VI, we derive the proximity factors for reducing the primal and dual basis, respectively. Section VII presents numerical results and then ends with a discussion.

Notation: Matrices and column vectors are denoted by upper and lowercase boldface letters (unless otherwise stated), and the transpose, inverse, pseudoinverse of a matrix $\mathbf{B}$ by $\mathbf{B}^T$, $\mathbf{B}^{-1}$, and $\mathbf{B}^\dagger$, respectively. For convenience, write $\mathbf{B}^{-T} = (\mathbf{B}^{-1})^T$ and $\mathbf{B}^{\dagger T} = (\mathbf{B}^\dagger)^T$. The inner product in the Euclidean space between vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T\mathbf{v}$, and the Euclidean length $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$.

II. MIMO Decoding

We consider a real-valued system model. For complex-valued system, an equivalent real system model with a doubled dimension is used. Reduction of complex lattices without using the equivalent real model is a topic that has been investigated elsewhere [14], [27], [28]. Let $\mathbf{x} = (x_1, ..., x_n)^T$ be the $n \times 1$ data vector, where each symbol x_i is chosen from a finite subset of the integer set $\mathbb{Z}$. With proper scaling and shifting [15], one has the generic $m \times n$ MIMO system model

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{n} \tag{1}$$

where $m \geq n$, $\mathbf{y}, \mathbf{n} \in \mathbb{R}^m$ denote the channel output and noise vectors, respectively, and $\mathbf{B} \in \mathbb{R}^{m \times n}$ is the $m \times n$ full column-rank matrix. The entries of $\mathbf{n}$ are i.i.d. normal with zero mean and variance σ^2. The expression of matrix $\mathbf{B}$ depends on the problem at hand. For example, $\mathbf{B}$ is the channel matrix for uncoded MIMO systems; $\mathbf{B}$ is composed of the channel matrix and the generating matrix of the encoding lattice for space-time block codes [2]; $\mathbf{B}$ involves the product-channel matrix for cooperative communications [4].

For such a system model, the ML decoder is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|^2 \tag{2}$$

where $\mathcal{C}$ stands for the transmitter finite set. Note that the complexity of the standard ML decoding that uses exhaustive search is exponential in n, and also increases with the alphabet size.

There are linear and nonlinear decoders with cubic complexity. In the linear ZF strategy, $\mathbf{y}$ is multiplied on the left by the pseudoinverse of $\mathbf{B}$, to yield the detection rule

$$\hat{\mathbf{x}} = \mathcal{Q}\{\mathbf{B}^\dagger \mathbf{y}\} \tag{3}$$

where $\mathcal{Q}(\cdot)$ denotes the quantization to the nearest integer within the signal boundary. A well-known drawback of ZF is the effect of noise amplification when the channel matrix $\mathbf{B}$ is ill-conditioned. By introducing decision feedback in the detection process, the nonlinear SIC strategy has better performance. One way to do SIC is to perform the QR decomposition $\mathbf{B} = \mathbf{Q}\mathbf{R}$, where $\mathbf{Q}$ has orthogonal columns and $\mathbf{R}$ is an upper triangular matrix [29]. Multiplying (1) on the left with $\mathbf{Q}^\dagger$ we have

$$\mathbf{y}' = \mathbf{Q}^\dagger \mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{n}'. \tag{4}$$

In SIC, the last symbol x_n is estimated first as $\hat{x}_n = \mathcal{Q}(y'_n / r_{n,n})$. Then the estimate is substituted to remove the interference term in y'_{n-1} when x_{n-1} is being estimated. The procedure is continued until the first symbol is detected. That is, we have the following recursion:

$$\hat{x}_i = \mathcal{Q}\left\{ \frac{y'_i - \sum_{j=i+1}^{n} r_{i,j} \hat{x}_j}{r_{i,i}} \right\} \tag{5}$$

for $i = n, n-1, ..., 1$.

ZF and SIC may incur heavy performance loss. For example, in single-user uncoded MIMO fading channels, ZF and SIC are only able to achieve the first-order diversity in an $n \times n$ system [30], i.e., that of a single antenna, which is far below the order n achieved by ML decoding. In fact, the diversity order of SIC cannot be increased even with ordering [31].

The basic idea behind approximate lattice decoding is to use lattice reduction in conjunction with traditional low-complexity decoders. With lattice reduction, the basis $\mathbf{B}$ is transformed into a new basis consisting of roughly orthogonal vectors (this is always possible in a sense defined later)

$$\mathbf{B}' = \mathbf{B}\mathbf{U} \tag{6}$$

where $\mathbf{U}$ is a unimodular matrix, i.e., $\mathbf{U}$ contains only integer entries and the determinant $\det \mathbf{U} = \pm 1$. Indeed, we have the equivalent channel model

$$\mathbf{y} = \mathbf{B}'\mathbf{U}^{-1}\mathbf{x} + \mathbf{n} = \mathbf{B}'\mathbf{x}' + \mathbf{n}, \quad \mathbf{x}' = \mathbf{U}^{-1}\mathbf{x}.$$

Then conventional decoders (ZF or SIC) are applied on the reduced basis. For example, ZF obtains the estimation

$$\hat{\mathbf{x}}' = \mathcal{Q}\left\{(\mathbf{B}')^{\dagger}\mathbf{y}\right\}. \tag{7}$$

This estimate is then transformed back into $\hat{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}}'$. Since the equivalent channel is much more likely to be well-conditioned, the effect of noise enhancement will be moderated. Note that due to the linear transform $\mathbf{U}^{-1}$, it is no longer easy to control the boundary; thus it is typical to quantize to the nearest integer ignoring the signal boundary in (7).

Both linear and nonlinear detectors can also be designed with respect to the MMSE criterion. The extension to the MMSE criterion is straightforward by dealing with the augmented channel matrix

$$\mathbf{B}_{\mathrm{a}} = \begin{bmatrix} \mathbf{B} \\ \sigma\mathbf{I} \end{bmatrix}. \tag{8}$$

When applied to the augmented channel matrix, ZF and SIC are equivalent to the standard MMSE and MMSE-SIC, respectively. In fact, MMSE is essential for ILD to achieve the optimum diversity and multiplexing tradeoff for finite constellations [17], [18].

III. Lattices and Lattice Reduction

An n-dimensional lattice in the m-dimensional embedding Euclidean space $\mathbb{R}^m$ is generated as the integer linear combination of n linearly independent vectors [32]:

$$L \triangleq L(\mathbf{B}) = \left\{ \sum_{i=1}^{n} x_i \mathbf{b}_i \mid x_i \in \mathbb{Z}, i = 1, ..., n \right\} \tag{9}$$

where $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_n]$ is called the basis of L.

A lattice L can be generated by infinitely many bases, of which one would like to select one that is in some sense nice or reduced. In many applications, it is advantageous to have the basis vectors as short as possible. Lattice reduction, the shortest vector problem (SVP) and CVP are closely related problems. All bases of the lattice arise by the transformation $\mathbf{B}' = \mathbf{B}\mathbf{U}$, where $\mathbf{U}$ is a unimodular matrix.

The dual lattice L^* of a lattice L is defined as those vectors $\mathbf{u}$, such that the inner product $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{Z}$, for all $\mathbf{v} \in L$ [32]. One might define the dual basis as $\mathbf{B}^{\dagger T}$. In this paper, we follow the definition of the dual basis $\mathbf{B}^* \triangleq \mathbf{B}^{\dagger T}\mathbf{J}$ in [33], where

$$\mathbf{J} \triangleq \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ & \ddots & & \\ 1 & \cdots & 0 & 0 \end{bmatrix}$$

is the flipping matrix, i.e., it reverses the columns of $\mathbf{B}^{\dagger T}$ in the left-right direction. Under this definition, the dual basis satisfy

$$\langle \mathbf{b}_i, \mathbf{b}_k^* \rangle = \delta_{i,n-k+1}. \tag{10}$$

where $\mathbf{b}_k^*$ is the k-th column of $\mathbf{B}^*$ and $\delta_{i,k}$ is the Kronecker delta. We have $\det L^* = 1/\det L$ and $L^{**} = L$.

For the sake of convenience, reduction of the dual basis will be referred to as dual reduction. Dual reduction still gives a nice basis of the primary lattice. To see this, suppose $\mathbf{U}^*$ is the unimodular matrix arising from reducing the dual basis $\mathbf{B}^*$, then the corresponding reduced primary basis is given by

$$\mathbf{B}' = (\mathbf{B}^{\dagger T}\mathbf{J}\mathbf{U}^*)^{\dagger T}\mathbf{J} = \mathbf{B}\mathbf{J}(\mathbf{U}^*)^{\dagger T}\mathbf{J}$$

Since $\mathbf{J}(\mathbf{U}^*)^{\dagger T}\mathbf{J}$ is also a unimodular matrix, $\mathbf{B}'$ must be a basis of the primary lattice. In approximate lattice decoding, we can use $\mathbf{J}(\mathbf{U}^*)^{\dagger T}\mathbf{J}$ as the transformation matrix in (6), whereas anything else remains the same.

Let $\lambda(L)$ be the length of the shortest nonzero vector in L. The Hermite constant is defined as [32]

$$\gamma_n \triangleq \sup_{L} \frac{\lambda^2(L)}{\det^{2/n} L}, \tag{11}$$

where the supremum is taken over all lattices L of dimension n. We shall need some bounds on the Hermite constant in the derivations of proximity factors in the following sections. Hermite showed that γ_n is bounded [32]. Mordell derived the inequality $\gamma_n^{n-2} \leq \gamma_{n-1}^{n-1}$ [34]. The upper bound was improved by Blichfeldt [35]

$$\gamma_n \leq \frac{2}{\pi}\Gamma^{2/n}\left(2 + \frac{n}{2}\right), \tag{12}$$

which has an asymptotic value $n/\pi e$. Even better estimates are known as n goes to infinity [33]:

$$\frac{n}{2\pi e} + o(n) \leq \gamma_n \leq \frac{0.872n}{\pi e} + o(n), \quad n \to \infty. \tag{13}$$

Moreover, the Hermite constant is known exactly for $n \leq 8$:

$$\begin{aligned} &\gamma_1 = 1, \gamma_2^2 = 4/3,\ \gamma_3^3 = 2,\ \ \gamma_4^4 = 4, \\ &\gamma_5^5 = 8, \gamma_6^6 = 64/3, \gamma_7^7 = 64, \gamma_8^8 = 2^8. \end{aligned} \tag{14}$$

The theory of lattice reduction [32] started with Gauss who introduced the concept of a lattice and found short vectors in 2-dimensional lattices. Reduction for general n was proposed by Hermite. More manageable versions were given by Minkowski [36], and Korkin and Zolotarev (KZ) [37], which try to find short basis vectors recursively. For a Minkowski-reduced basis $\mathbf{B}$, $\mathbf{b}_1$ is a shortest nonzero vector in L, and $\mathbf{b}_i$ for $i = 2, \cdots, n$ is the shortest vector in L such that $[\mathbf{b}_1, \cdots \mathbf{b}_i]$ may be extended to a basis of L. In KZ reduction, this is performed in the linear space orthogonal to the vectors already found. In 1982, Lenstra, Lenstra and Lovász proposed the first polynomial time algorithm which finds a vector not longer than $2^{(n-1)/2}$ times the shortest nonzero vector.

Both KZ and LLL reduction can be seen as possible generalizations of the Gauss reduction, and they can be defined in terms of the Gram-Schmidt orthogonalization. One can compute the Gram-Schmidt orthogonalization by the recursion [38]

$$\hat{\mathbf{b}}_i = \mathbf{b}_i - \sum_{j=1}^{i-1} \mu_{i,j}\hat{\mathbf{b}}_j, \quad \text{for } i = 1, ..., n \tag{15}$$

where $\mu_{i,j} = \langle \mathbf{b}_i, \hat{\mathbf{b}}_j \rangle / \|\hat{\mathbf{b}}_j\|^2$. Note that the Gram-Schmidt orthogonalization depends on the order of the basis vectors. In matrix notation, Gram-Schmidt orthogonalization can be written as $\mathbf{B} = \hat{\mathbf{B}}\boldsymbol{\mu}^T$, where $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1, ..., \hat{\mathbf{b}}_n]$, and $\boldsymbol{\mu}$ is a lower-triangular matrix with unit diagonal elements. In relation to the QR decomposition, one has $\mu_{j,i} = r_{i,j}/r_{i,i}$ and $\hat{\mathbf{b}}_i = r_{i,i} \cdot \mathbf{q}_i$ where $\mathbf{q}_i$ is the i-th column of $\mathbf{Q}$.

A. KZ Reduction

A basis $\mathbf{B}$ is said to be KZ-reduced if it satisfies the following conditions [33]:

- $\mathbf{b}_1$ is the shortest nonzero vector of L;
- $|\mu_{i,j}| \leq 1/2$, for $1 \leq j < i \leq n$;
- If $L^{(n-1)}$ denotes that orthogonal projection of L on the orthogonal complement of the linear space spanned by $\mathbf{b}_1$, then the projections $\mathbf{b}_i - \mu_{i,1}\mathbf{b}_1$ of $\mathbf{b}_2, ..., \mathbf{b}_n$ yield a KZ basis of $L^{(n-1)}$. Equivalently, $\hat{\mathbf{b}}_i$ is a shortest nonzero vector of $L^{(n-i+1)}$, orthogonal projection of L on the orthogonal complement of the linear space spanned by $\mathbf{b}_1, ..., \mathbf{b}_{i-1}$.

Finding a KZ basis is polynomial-time equivalent to finding the shortest vector. Hence, the KZ basis is hard to find for lattices of high dimension. When $n = 2$, KZ reduction is the same as Gaussian reduction.

B. LLL Reduction

A basis is LLL reduced if [39]

- $|\mu_{i,j}| \leq 1/2$, for $1 \leq j < i \leq n$;
- $\|\hat{\mathbf{b}}_i\|^2 + \mu_{i,i-1}^2 \|\hat{\mathbf{b}}_{i-1}\|^2 \geq \delta \|\hat{\mathbf{b}}_{i-1}\|^2$, for $1 < i \leq n$.

The first clause is called size reduction, while the second is known as the Lovász condition. The parameter δ takes values in the interval $(1/4, 1]$ and it is common to choose $\delta = 3/4$ for a tradeoff between quality and complexity. It is well known that the standard LLL algorithm requires $O(n^4)$ arithmetic operations for $n \times n$ integer bases [39]. Recently, we established the $O(n^3 \log n)$ average time bound for real-valued bases whose vectors are i.i.d. standard normal [40], which is lower than thought before. A similar result on the average number of iterations was obtained in [41].

Define $\beta \triangleq 1/(\delta - 1/4)$. An LLL-reduced lattice has the following properties:

- $\|\mathbf{b}_1\| \leq \beta^{(n-1)/2} \lambda(L)$;
- $\|\mathbf{b}_1\| \leq \beta^{(n-1)/4} \det^{1/n} L$;
- $\prod_{i=1}^{n} \|\mathbf{b}_i\| \leq \beta^{n(n-1)/4} \det L$.

These inequalities indicate in various senses that the basis vectors are short. These bounds are tight in the worst case in the sense that there exist bases reaching the bounds.

IV. Proximity Factors

In this section, we shall introduce an analytic tool for approximate lattice decoding. The lattice literature is usually concerned with the approximation factor of CVP. For instance, Babai proved that ZF and SIC on an LLL-reduced basis find the closest vector, up to a factor $1 + 2n(9/2)^{n/2}$ and $2^{n/2}$ (for $\delta = 3/4$), respectively [7]. However, such results do not directly translate into how close approximate lattice decoding is to ILD in terms of the decoding distance associated with the decision region, which is more useful in digital communications. To determine the proximity, we examine the decision regions of the ZF and SIC decoders and the associated distances[1].

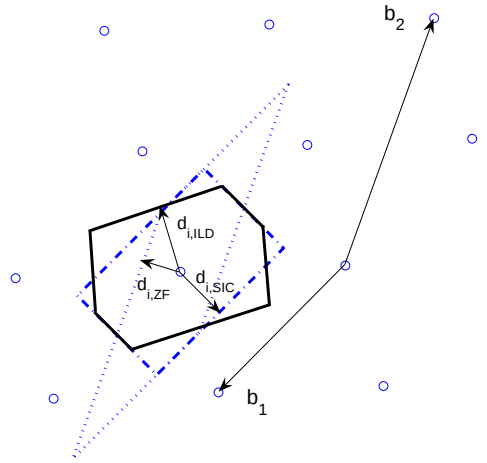


Fig. 1. Decision regions of ZF (dotted), SIC (dash-dotted) and ILD (solid) and the corresponding distances in a 2-dimensional lattice.

A. Definition

We consider a fixed but arbitrary lattice. Since we assume an infinite lattice, the signal space is geometrically uniform. Without loss of generality, let the transmitted lattice point be $\mathbf{x} = \mathbf{0}$. The decision regions $\mathcal{R}$ of various strategies are depicted in Fig. 1 for a 2-dimensional lattice.

The Voronoi cell, corresponding to the decision region of ILD, is a polytope defined by

$$\mathcal{R}_{\text{ILD}} = \{\mathbf{y} : \|\mathbf{y} - \mathbf{v}\| \geq \|\mathbf{y}\|, \quad \forall \mathbf{v} \in L\}.$$

Each facet of the Voronoi cell lies on the perpendicular bisector of the line connecting the origin and the Voronoi neighbor $\mathbf{v}_i \in \mathcal{N}$, where $\mathcal{N}$ represents the Voronoi neighborhood.

Let $d_{i,\text{ILD}}$ be the Euclidean distance from point $\mathbf{0}$ to the i-th facet of $\mathcal{R}_{\text{ILD}}$. It is easy to see from Fig. 1 that $d_{i,\text{ILD}} = \|\mathbf{v}_i\|/2$. The decision regions of both ZF and SIC are polyhedra with $2n$ facets and are symmetric with respect to the origin. Specifically, the ZF decision region is just the fundamental parallelogram $\mathcal{P}(\mathbf{B}) = \{\mathbf{Ba}, |a_i| \leq 1/2\}$ centered at $\mathbf{x} = \mathbf{0}$. Meanwhile, the SIC decision region is a rectangle specified by the Gram-Schmidt vectors as $\{\hat{\mathbf{B}}\mathbf{a}, |a_i| \leq 1/2\}$ [42].

Let $d_{i,\text{ZF}}$ and $d_{i,\text{SIC}}$ be the Euclidean distance from point $\mathbf{0}$ to the i-th facet of the decision regions of ZF and SIC, respectively. We are concerned with a spectrum of n distances due to symmetry. It can be seen from Fig. 1 that the distance spectrum of ZF is given by

$$d_{i,\text{ZF}} = \frac{1}{2}\|\mathbf{b}_i\| \sin\theta_i, \quad i = 1, ..., n$$

where θ_i denotes the angle between $\mathbf{b}_i$ and the linear space spanned by the other $n-1$ basis vectors. In Appendix I, we derive a formula to calculate θ_i from $\mathbf{B}$ and its Gram-Schmidt orthogonalization, which will be used in latter sections to bound $d_{i,\text{ZF}}$.

For SIC, the distance spectrum is given by [42]

$$d_{i,\text{SIC}} = \frac{1}{2}\|\hat{\mathbf{b}}_i\| = \frac{1}{2}\|\mathbf{b}_i\| \sin\phi_i, \quad i = 1, ..., n$$

[1]The distances defined here are not the standard distances between lattice points, but those associated with the decision regions.

where ϕ_i is be the angle between $\mathbf{b}_i$ and the hyperplane spanned by $\mathbf{b}_1, ..., \mathbf{b}_{i-1}$. Note that $\|\mathbf{b}_i\| \sin\theta_i$ is the distance between the vector point $\mathbf{b}_i$ and the linear space spanned by the other $n-1$ basis vectors. Since $\hat{\mathbf{b}}_i$ only needs to be orthogonal to $\mathbf{b}_1, ..., \mathbf{b}_{i-1}$, we must have $\theta_i \leq \phi_i$ and hence $d_{i,\text{ZF}} \leq d_{i,\text{SIC}}$.

The minimum decoding distance (also known as *decoding radius*) $d_{\text{ILD}} \triangleq \min_i\{d_{i,\text{ILD}}\} = \lambda(L)/2$ plays an important role. We are motivated to define the *proximity factors* measuring the proximity between the performance of ILD and approximate lattice decoding as follows:

$$\begin{aligned} \rho_{i,\text{ZF}} &\triangleq \sup_{\mathbf{B}\in\mathcal{B}_{\text{Reduced}}} \frac{d_{\text{ILD}}^2}{d_{i,\text{ZF}}^2} = \sup_{\mathbf{B}\in\mathcal{B}_{\text{Reduced}}} \frac{\lambda^2(L)}{\|\mathbf{b}_i\|^2 \sin^2\theta_i} \\ \rho_{i,\text{SIC}} &\triangleq \sup_{\mathbf{B}\in\mathcal{B}_{\text{Reduced}}} \frac{d_{\text{ILD}}^2}{d_{i,\text{SIC}}^2} = \sup_{\mathbf{B}\in\mathcal{B}_{\text{Reduced}}} \frac{\lambda^2(L)}{\|\hat{\mathbf{b}}_i\|^2} \end{aligned} \tag{16}$$

where the supremum is taken over the set $\mathcal{B}_{\text{Reduced}}$ of basis matrices satisfying a certain *reduction criterion* for any n-dimensional lattices L. Furthermore, we define $\rho_{\text{ZF}} \triangleq \max_{1\leq i\leq n} \rho_{i,\text{ZF}}$ and $\rho_{\text{SIC}} \triangleq \max_{1\leq i\leq n} \rho_{i,\text{SIC}}$, which quantify the worst-case loss of approximate lattice decoding in the minimum squared decoding distance (i.e., squared decoding radius). It is noteworthy that the proximity factors are not a function of any particular lattice L, and that without reduction they are unbounded. However, we shall show in the latter Sections that with lattice reduction the proximity factor is bounded by a constant that is a function of n only.

B. Distance Properties of Dual Reduction

By (10), $\mathbf{b}_{n-i+1}^*$ of the dual basis is perpendicular to the $(n-1)$-dimensional hyperplane spanned by vectors $\mathbf{b}_1, ..., \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, ..., \mathbf{b}_n$. Furthermore, since $(\pi/2 - \theta_i)$ is the angle between $\mathbf{b}_i$ and $\mathbf{b}_{n-i+1}^*$, we have

$$\langle \mathbf{b}_i, \mathbf{b}_{n-i+1}^* \rangle = \|\mathbf{b}_{n-i+1}^*\| \cdot \|\mathbf{b}_i\| \cdot \cos(\pi/2 - \theta_i) = \|\mathbf{b}_{n-i+1}^*\| \cdot \|\mathbf{b}_i\| \cdot \sin\theta_i = 1.$$

Therefore, the following relation holds for dual reduction:

$$d_{i,\text{ZF}} = \frac{1}{2}\|\mathbf{b}_i\| \sin\theta_i = \frac{1}{2}\|\mathbf{b}_{n-i+1}^*\|^{-1}. \tag{17}$$

It indicates that we can have large decoding distances in ZF if the vectors of the dual basis are short.

To study the decoding distances of SIC, we need the following proposition, whose proof is given in Appendix II.

Proposition 1: Let $\mathbf{B} = \hat{\mathbf{B}}\boldsymbol{\mu}^T$ and $\mathbf{B}^* = \hat{\mathbf{B}}^*(\boldsymbol{\mu}^*)^T$ be the Gram-Schmidt orthogonalization of the primal basis $\mathbf{B}$ and dual basis $\mathbf{B}^*$, respectively. Then

$$\begin{aligned} \hat{\mathbf{B}}^* &= \hat{\mathbf{B}}^{\dagger T}\mathbf{J}, \\ \boldsymbol{\mu}^* &= \mathbf{J}\boldsymbol{\mu}^{-T}\mathbf{J}. \end{aligned} \tag{18}$$

The property $\hat{\mathbf{B}}^* = \hat{\mathbf{B}}^{\dagger T}\mathbf{J}$ implies a useful relation

$$\hat{\mathbf{b}}_{n-i+1}^* = \frac{\hat{\mathbf{b}}_i}{\|\hat{\mathbf{b}}_i\|^2}, \quad i = 1, 2, ..., n \tag{19}$$

where $\hat{\mathbf{b}}_1^*, ..., \hat{\mathbf{b}}_n^*$ are the Gram-Schmidt vectors of the dual basis $\mathbf{B}^*$. The elegant relation (19) was derived earlier in [33], [43]. The relation $\boldsymbol{\mu}^* = \mathbf{J}\boldsymbol{\mu}^{-T}\mathbf{J}$ is new; it corresponds to flipping $\boldsymbol{\mu}^{-1}$ with respect to the anti-diagonal.

Meanwhile, by (19),

$$d_{i,\text{SIC}} = \frac{1}{2}\|\hat{\mathbf{b}}_i\| = \frac{1}{2}\|\hat{\mathbf{b}}_{n-i+1}^*\|^{-1}. \tag{20}$$

It shows that SIC can have large decoding distances if the reversed dual basis has short Gram-Schmidt vectors.

C. Performance Bounds for Approximate Lattice Decoding

For decoding in an infinite lattice, an error occurs when the noise falls outside of $\mathcal{R}$, and the error probability does not depend on the transmitted vector $\mathbf{x}$. Accordingly, the error probability $P_e(\mathbf{B})$ in an infinite lattice with basis $\mathbf{B}$ equals that for vector $\mathbf{x} = \mathbf{0}$ and is given by

$$P_e(\mathbf{B}) = P(\hat{\mathbf{x}} \neq \mathbf{0} | \mathbf{x} = \mathbf{0}) = P(\mathbf{n} \in \overline{\mathcal{R}}). \tag{21}$$

To keep the results general, we write $\text{SNR} = c/\sigma^2$, where c is a constant depending on the problem. By the symmetry of the Voronoi cell (i.e., there are at least two vectors with the shortest length,) we have the lower bound on the conditional decoding error probability of ILD

$$P_{e,\text{ILD}}(\text{SNR}, \mathbf{B}) \geq 2Q\left(\frac{d_{\text{ILD}}}{\sigma}\right) = 2Q\left(\sqrt{\frac{d_{\text{ILD}}^2 \cdot \text{SNR}}{c}}\right). \tag{22}$$

Meanwhile, the union bound on the conditional error probability of ZF reads

$$P_{e,\text{ZF}}(\text{SNR}, \mathbf{B}) \leq 2\sum_{i=1}^{n} Q\left(\frac{d_{i,\text{ZF}}}{\sigma}\right) \tag{23}$$

where the factor 2 is due to symmetry. The union bound for SIC admits a form similar to (23). Given the same basis matrix $\mathbf{B}$, the conditional error probability of lattice reduction-aided ZF can be bounded above as

$$P_{e,\text{ZF}}(\text{SNR}, \mathbf{B}) \leq 2\sum_{i=1}^{n} Q\left(\frac{d_{\text{ILD}}}{\sqrt{\rho_{i,\text{ZF}}}\sigma}\right) = 2\sum_{i=1}^{n} Q\left(\sqrt{\frac{d_{\text{ILD}}^2 \cdot \text{SNR}}{c \cdot \rho_{i,\text{ZF}}}}\right). \tag{24}$$

since $d_{i,\text{ZF}}^2 \geq \rho_{i,\text{ZF}} \cdot d_{\text{ILD}}^2$ by definition (16) and since $Q(\cdot)$ is a decreasing function. It is worth pointing out that while the distance d_{ILD} is a function of $\mathbf{B}$, $\rho_{i,\text{ZF}}$ is not.

Now, combining (22) and (24), we have

$$P_{e,\text{ZF}}(\text{SNR}, \mathbf{B}) \leq \sum_{i=1}^{n} P_{e,\text{ILD}}\left(\frac{\text{SNR}}{\rho_{i,\text{ZF}}}, \mathbf{B}\right). \tag{25}$$

Since (25) holds for any $\mathbf{B}$, averaging out $\mathbf{B}$ we obtain

$$P_{e,\text{ZF}}(\text{SNR}) \leq \sum_{i=1}^{n} P_{e,\text{ILD}}\left(\frac{\text{SNR}}{\rho_{i,\text{ZF}}}\right) \tag{26}$$

for arbitrary SNR. In particular,

$$P_{e,\text{ZF}}(\text{SNR}) \leq nP_{e,\text{ILD}}\left(\frac{\text{SNR}}{\rho_{\text{ZF}}}\right). \tag{27}$$

The relations (26) and (27) hold irrespective of fading statistics, and similar relations exist for the error probability of SIC[2]. They reveal, in a quantitative manner, that LRAD

[2]Such relations do not necessarily hold for the *symbol error probability* of SIC, due to error propagation.

performs within a constant bound from ILD. The mere effect on the error rate curve is an SNR shift from that of ILD, up to a multiplicative factor n, which obviously does not change the diversity order[3]. In other words, the diversity order is the same as that of ILD. Therefore, existing results on the diversity order of ILD can be extended to LRAD. In particular, since ILD achieves full receive diversity in the uncoded V-BLAST system [24], LRAD also achieves full diversity. This provides an alternative way of showing the diversity order of LRAD given in [24].

In the case of MMSE pre-processing, caution should be exercised to interpret proximity factors. Because the matrix (8) is not a physically existent channel, proximity factors do not exactly reflect error probability gaps. Nonetheless, using the DMT optimality of bounded distance decoding (BDD) for the regularized lattice (8) [44] and the fact that a finite proximity factor implies BDD[4], we can show that any decoder for the lattice (8) with finite proximity factors achieves the optimal DMT for approximately universal codes [18].

D. Main Results

Table I shows the main results of this paper, i.e., the bounds on proximity factors for reducing the primary basis and for reducing the dual basis. The derivations will be given in the following two sections. Notice that the value $4/3$ is exact for Gaussian reduction, and could be obtained by specializing the bounds for LLL or KZ reduction. In the meantime, the bounds for LLL and KZ reduction in Table I are less explicit, i.e., they are expressed in the parameter β for LLL reduction and in the Hermite and KZ constants for KZ reduction. To give the readers a feeling about the dependence of the bounds on n, Table II shows more explicit bounds for LLL and KZ reduction, where the bounds are however less tight. For LLL reduction, the bounds on proximity factors are exponential; if expressed in dB, they are linear with n. Of note is the fact that the bound 2^n for dual LLL-ZF is much better than the bound $n2^{n(n-1)}$ implied by [24, Lemma 3].

It is seen that reducing the dual basis leads to smaller bounds in many cases. Namely, when the dual basis is reduced, LLL-ZF has a smaller base in its exponential bound, while KZ reduction results in polynomial bounds on proximity factors for both ZF and SIC. It is also interesting that, unlike primal basis reduction, LLL-ZF and LLL-SIC have the same bounds on proximity factors when the dual basis is reduced.

V. Primal Basis Reduction

With lattice reduction, the supremum in (16) is taken over the set of reduced bases of the lattice L. Consequently, the proximity factors will be bounded. In this Section, we shall derive explicit upper bounds for LLL and KZ reductions. The bounds are expressed in closed form and turn out to be functions of n only.

[3]The exact amount of the SNR shift depends on averaging over the channel. The proximity factors only provide an upper bound to the maximal possible shift (after applying the factor n).

[4]BDD is a variant of approximate CVP where correct decoding is guaranteed when the distance between $\mathbf{y}$ and L is less than $\lambda(L)/(2\gamma)$ [44]. It is easy to see that the proximity factor is equal to γ^2. Therefore, any decoder with a finite proximity factor ρ is $\lambda(L)/(2\sqrt{\rho})$-BDD.

TABLE II
More Explicit But Less Tight Upper Bounds on Proximity Factors (n: dimension; $\delta = 3/4$ for LLL reduction)

	LLL-ZF	LLL-SIC	KZ-ZF	KZ-SIC
Primal Reduction	$\left(\frac{9}{2}\right)^n$	2^n	$\left(\frac{9}{4}\right)^n \cdot n^{1+\ln n}$	$n^{1+\ln n}$
Dual Reduction	2^n	2^n	n^3	n^2

A. LLL Reduction

The derivation of the bounds for LLL reduction will be done by adapting the techniques of the original LLL paper [39] and Babai [13]. While the exponential bounds in [13], [39] have the best base, in this subsection we will improve the bounds by a constant factor.

1) SIC:

Theorem 1: For LLL-SIC, the proximity factors are bounded by

$$\rho_{i,\text{SIC}} \leq 1 + \frac{\beta^i - \beta}{4(\beta - 1)}. \tag{28}$$

Proof: From [39] one has

$$\|\hat{\mathbf{b}}_j\|^2 \leq \beta^{i-j}\|\hat{\mathbf{b}}_i\|^2, \quad \text{for } 1 \leq j < i \leq n. \tag{29}$$

Substituting this into the representation of $\mathbf{b}_i$ in Gram-Schmidt vectors

$$\|\mathbf{b}_i\|^2 = \|\hat{\mathbf{b}}_i\|^2 + \sum_{j=1}^{i-1} \mu_{i,j}^2 \|\hat{\mathbf{b}}_j\|^2, \tag{30}$$

we obtain

$$\|\mathbf{b}_i\|^2 \leq \left(1 + \sum_{j=1}^{i-1} \beta^{i-j}/4\right) \|\hat{\mathbf{b}}_i\|^2. \tag{31}$$

Note that (31) was also given in [39] for the special case $\delta = 3/4$.

Let $\mathcal{B}_{\text{LLL}}$ be the set of LLL-reduced basis matrices. Since $\lambda(L) \leq \|\mathbf{b}_i\|$, the loss in the i-th squared Euclidean distance is bounded by

$$\begin{aligned} \rho_{i,\text{SIC}} &= \sup_{\mathbf{B} \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\hat{\mathbf{b}}_i\|^2} \\ &\leq \left(1 + \frac{\beta^i - \beta}{4(\beta - 1)}\right) \sup_{\mathbf{B} \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\mathbf{b}_i\|^2} \\ &\leq 1 + \frac{\beta^i - \beta}{4(\beta - 1)}. \end{aligned} \tag{32}$$

■

It is easy to verify that when $\beta \geq 4/3$ (which is the case for LLL reduction) (32) is smaller than the bound

$$\rho_{i,\text{SIC}} \leq \beta^{i-1} \tag{33}$$

directly obtainable from [39]. Obviously,

$$\rho_{\text{SIC}} \leq 1 + \frac{\beta^n - \beta}{4(\beta - 1)} \leq \beta^{n-1}. \tag{34}$$

TABLE I

UPPER BOUNDS ON PROXIMITY FACTORS (n: DIMENSION; $\beta = 1/(\delta - 1/4), 1/4 < \delta \leq 1$; γ_j: HERMITE CONSTANT; ξ_j: KZ CONSTANT)

	Gauss-ZF	Gauss-SIC	LLL-ZF	LLL-SIC	KZ-ZF	KZ-SIC
Primal Reduction	$\frac{4}{3}$	$\frac{4}{3}$	$\frac{\beta}{9\beta-4}\left(\frac{9\beta}{4}\right)^{n-1} + \frac{8\beta-4}{9\beta-4}$	$1 + \frac{\beta^n-\beta}{4(\beta-1)}$	$1 + \sum_{j=1}^{n-1}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\xi_{j+1}$	ξ_n
Dual Reduction	$\frac{4}{3}$	$\frac{4}{3}$	β^{n-1}	β^{n-1}	$1 + \frac{1}{4}\sum_{j=2}^{n}\gamma_j^2$	$\max_{1\leq j\leq n}\gamma_j^2$

2) ZF: The derivation for LLL-ZF needs the following lemma. It improves Babai's lower bound on θ_i [13] and extends it to arbitrarily values of $\delta \in (1/4, 1)$. As recognized by Babai [13], the lower bound on θ_i describes a geometric feature of an LLL-reduced basis, which is of independent interest (e.g., in Seysen's measure of a lattice [45]). The proof is given in Appendix III.

Lemma 1: If $\mathbf{B}$ is an LLL-reduced basis of lattice $L(\mathbf{B})$, then

$$\sin^2\theta_i \geq \left[\frac{\beta}{9\beta-4}\left(\frac{9\beta}{4}\right)^{n-i} + \frac{8\beta-4}{9\beta-4}\right]^{-1} \times \left[1 + \frac{\beta^i-\beta}{4(\beta-1)}\right]^{-1} \tag{35}$$

for $1 \leq i \leq n$.

It is worth mentioning that the lower bound of Lemma 1 is tight. The equality is achieved when $\|\hat{\mathbf{b}}_j\|^2 = \beta^{i-j}\|\hat{\mathbf{b}}_i\|^2$ and $\mu_{i,j} = -1/2$ for all $1 \leq j < i \leq n$. This can be seen from the proof.

Applying Lemma 1 and the trivial inequality $\|\mathbf{b}_i\| \geq \lambda(L)$ to

$$\rho_{i,\text{ZF}} \leq \sup_{\mathbf{B}\in\mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\mathbf{b}_i\|^2}\frac{1}{\sin^2\theta_i}, \tag{36}$$

we obtain the following theorem:

Theorem 2: For LLL-ZF, the proximity factors are bounded by

$$\rho_{i,\text{ZF}} \leq \left[\frac{\beta}{9\beta-4}\left(\frac{9\beta}{4}\right)^{n-i} + \frac{8\beta-4}{9\beta-4}\right]\left(1 + \frac{1}{4}\cdot\frac{\beta^i-\beta}{\beta-1}\right). \tag{37}$$

Since the upper bound is maximized when $i = 1$, we have

$$\rho_{\text{ZF}} \leq \frac{\beta}{9\beta-4}\left(\frac{9\beta}{4}\right)^{n-1} + \frac{8\beta-4}{9\beta-4}. \tag{38}$$

This is better than the bound $\rho_{\text{ZF}} \leq \left(\frac{9\beta}{4}\right)^{n-1}$ obtained previously in [21].

B. KZ Reduction

1) SIC: To derive the proximity factors for SIC with KZ reduction, we make use of the KZ constant defined in [46] as

$$\xi_n \triangleq \sup_{\mathbf{B}\in\mathcal{B}_{\text{KZ}}} \frac{\|\mathbf{b}_1\|^2}{\|\hat{\mathbf{b}}_n\|^2} \tag{39}$$

where $\mathcal{B}_{\text{KZ}}$ is the set of all KZ-reduced bases of dimension n. The KZ constant is bounded in terms of the Hermite constant by [46]

$$\xi_n \leq \gamma_n\gamma_n^{1/(n-1)}\gamma_{n-1}^{1/(n-2)}...\gamma_2 \tag{40}$$

for $n \geq 2$. Since $\|\mathbf{b}_1\| = \lambda(L)$ in a KZ-reduced lattice, we have

$$\rho_{n,\text{SIC}} = \xi_n.$$

To obtain $\rho_{i,\text{SIC}}$ for $i < n$, note that if $\mathbf{B}$ is a KZ-reduced basis of lattice $L(\mathbf{B})$, then $\mathbf{B}_i = [\mathbf{b}_1, ..., \mathbf{b}_i]$ is a KZ reduced basis of the sublattice $L_i = L(\mathbf{B}_i)$. This follows from the definition of KZ reduction and the fact that $L_i \subseteq L$. Accordingly, we have

Theorem 3: For KZ-SIC, the proximity factors are bounded by

$$\rho_{i,\text{SIC}} = \xi_i.$$

Since the KZ constant ξ_n is a nondecreasing function of n [46], we have $\rho_{\text{SIC}} = \xi_n$, i.e., the KZ constant exactly quantifies the worst-case performance loss of SIC. The KZ constant is exactly known for $n = 2$ and $n = 3$: $\xi_2 = 4/3$ and $\xi_3 = 3/2$ [46]. It is also known that [33], [46]

$$\xi_n \leq n^{1+\ln n}, \tag{41}$$

which is sub-exponential.

2) ZF:

Theorem 4: For KZ-ZF, the proximity factors are bounded by

$$\rho_{i,\text{ZF}} \leq \xi_i + \sum_{j=1}^{n-i}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\xi_{i+j}. \tag{42}$$

Proof: By (60), we have

$$\begin{aligned}\rho_{i,\text{ZF}} &= \sup_{\mathbf{B}\in\mathcal{B}_{\text{KZ}}} \frac{\lambda^2(L)}{\|\mathbf{b}_i\|^2\sin^2\theta_i} \\ &\leq \sup_{\mathbf{B}\in\mathcal{B}_{\text{KZ}}} \frac{\lambda^2(L)}{\sum_{j=i}^{n} r_j^2\|\hat{\mathbf{b}}_j\|^2}.\end{aligned} \tag{43}$$

where r_j is defined in Appendix I. Applying (65) yields

$$\rho_{i,\text{ZF}} \leq \sup_{\mathbf{B}\in\mathcal{B}_{\text{KZ}}} \lambda^2(L)\left[\|\hat{\mathbf{b}}_i\|^{-2} + \sum_{j=1}^{n-i}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\|\hat{\mathbf{b}}_{i+j}\|^{-2}\right].$$

Using the facts that $\lambda^2(L) = \|\hat{\mathbf{b}}_1\|^2$ and $\|\hat{\mathbf{b}}_1\|^2 \leq \xi_j\|\hat{\mathbf{b}}_j\|^2$ for a KZ-reduced basis, we obtain (42). ■

The maximum is attained when $i = 1$, and

$$\rho_{\text{ZF}} \leq 1 + \sum_{j=1}^{n-1}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\xi_{j+1}. \tag{44}$$

Using the upper bound (41), we obtain a more explicit form

$$\rho_{\text{ZF}} \leq \left(\frac{9}{4}\right)^{n-1}\cdot n^{1+\ln n}. \tag{45}$$

VI. Dual Basis Reduction

In this Section, we shall derive upper bounds on the proximity factors when the dual rather than the primal basis is reduced.

A. Dual LLL Reduction

1) SIC:

Theorem 5: For SIC with dual LLL reduction, the proximity factors are bounded by

$$\rho_{i,\text{SIC}} \leq \beta^{i-1}. \tag{46}$$

Proof: A scrutiny into the derivation for primal-basis reduction reveals that the following condition is crucial:

$$|\mu_{i,i-1}| \leq 1/2, \quad 1 < i \leq n. \tag{47}$$

A basis satisfies the Lovász condition and (47) is called an *effectively* LLL-reduced basis [43]. Effective LLL reduction produces a basis with the same set of Gram-Schmidt vectors. Accordingly, with SIC, it yields the same output as the standard LLL reduction. Moreover, if a basis is effectively LLL-reduced, so is its dual basis [43]. This is because the inverse of the coefficient matrix $\boldsymbol{\mu}$ admits the form [47]

$$\boldsymbol{\mu}^{-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\mu_{2,1} & 1 & \cdots & 0 & 0 \\ \times & -\mu_{3,2} & 1 & \cdots & 0 \\ & & & \ddots & \\ \times & \cdots & \times & -\mu_{n,n-1} & 1 \end{bmatrix}$$

and the Lovász condition can be shown to be satisfied for the dual basis.

Although other off-diagonal elements may have absolute values larger than $1/2$, it does not really cause a problem, since we still have

$$\begin{aligned} \rho_{i,\text{SIC}} &= \sup_{\mathbf{B}^* \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\hat{\mathbf{b}}_i\|^2} \\ &\leq \beta^{i-1} \sup_{\mathbf{B}^* \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\mathbf{b}_1\|^2} \\ &\leq \beta^{i-1}. \end{aligned} \tag{48}$$

■

Note that this is the same as (33) for primal LLL reduction. Accordingly, we expect similar performance for primal and dual LLL reductions in the case of SIC.

2) ZF:

Theorem 6: For ZF with dual LLL reduction, the proximity factors are bounded by

$$\rho_{i,\text{ZF}} \leq \left(1 + \frac{\beta}{4} \cdot \frac{\beta^{n-i} - 1}{\beta - 1}\right) \beta^{i-1}. \tag{49}$$

Proof: The derivation of the bound on $\rho_{i,\text{ZF}}$ is analogous to that for primal LLL reduction. Again, using (60) and noting that expression (43) still applies if KZ-reduced bases are replaced by dual-LLL-reduced bases, we have

$$\rho_{i,\text{ZF}} \leq \sup_{\mathbf{B}^* \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2}.$$

Incorporating the lower bound (69) in Appendix IV, we deduce

$$\rho_{i,\text{ZF}} \leq \sup_{\mathbf{B}^* \in \mathcal{B}_{\text{LLL}}} \lambda^2(L) \left(\|\hat{\mathbf{b}}_i\|^{-2} + \frac{1}{4} \sum_{j=1}^{n-i} \|\hat{\mathbf{b}}_{i+j}\|^{-2} \right). \tag{50}$$

Besides, the primal basis is effectively LLL-reduced when the dual basis is LLL-reduced. Thus, $\|\hat{\mathbf{b}}_i\|^2 \geq \beta^{-(i-1)} \|\mathbf{b}_1\|^2$. We obtain

$$\begin{aligned} \rho_{i,\text{ZF}} &\leq \sup_{\mathbf{B}^* \in \mathcal{B}_{\text{LLL}}} \frac{\lambda^2(L)}{\|\mathbf{b}_1\|^2} \left(1 + \frac{1}{4} \sum_{j=1}^{n-i} \beta^j\right) \beta^{i-1} \\ &= \left(1 + \frac{1}{4} \sum_{j=1}^{n-i} \beta^j\right) \beta^{i-1}, \end{aligned} \tag{51}$$

which leads to (49). ■

Since the right-hand side of (49) is an increasing function of i for $\beta \geq 4/3$, we have

$$\rho_{\text{ZF}} \leq \beta^{n-1}. \tag{52}$$

We emphasize that the worst case bound (52) is the same as dual LLL-SIC and is close to that for SIC with primal LLL reduction.

B. Dual KZ Reduction

1) SIC: It is shown in [33] that if a reversed dual basis is KZ reduced, then $\gamma_i \|\hat{\mathbf{b}}_i\| \geq \lambda(L)$. Accordingly, we have

Theorem 7: For SIC with dual KZ reduction, the proximity factors are bounded by

$$\rho_{i,\text{SIC}} \leq \gamma_i^2. \tag{53}$$

Although the Hermite constant is very likely to be an increasing function, it has never been proved [33]. Hence, we bound the overall proximity factor for SIC by

$$\rho_{\text{SIC}} \leq \max_{1 \leq i \leq n} \gamma_i^2. \tag{54}$$

Since $\gamma_n \leq 2n/3$, we have $\rho_{\text{SIC}} \leq (2n/3)^2 < n^2$. A better bound can be obtained by employing Blichfeldt's bound on the Hermite constant:

$$\rho_{\text{SIC}} \leq \left(\frac{2}{\pi}\right)^2 \Gamma^{4/n} \left(2 + \frac{n}{2}\right). \tag{55}$$

We claim that γ_i^2 is never greater than the bound on the KZ constant (40). We prove this by induction. When $i = 1$, $\xi_1 = \gamma_1^2 = 1$, and the claim is true. Suppose it is true when $k = i-1$, from which we deduce $\gamma_{i-1} \leq \gamma_{i-1}^{1/(i-2)} ... \gamma_2$. When $k = i$ we have

$$\frac{\gamma_i^2}{\gamma_i \gamma_i^{1/(i-1)} \gamma_{i-1}^{1/(i-2)} ... \gamma_2} \leq \frac{\gamma_i}{\gamma_i^{1/(i-1)} \gamma_{i-1}} = \frac{\gamma_i^{(i-2)/(i-1)}}{\gamma_{i-1}}.$$

By Mordell's inequality $\gamma_i^{i-2} \leq \gamma_{i-1}^{i-1}$ [34], this is not greater than one. Thus the claim is true for $k = i$ as well.

Therefore, reducing the reversed dual basis again has a smaller bound on $\rho_{i,\text{SIC}}$ than reducing the primal basis.

2) ZF: Substituting $\gamma_j \|\hat{\mathbf{b}}_j\| \geq \lambda(L)$ for dual KZ reduction [33] into (50), we arrive at

Theorem 8: For ZF with dual KZ reduction, the proximity factors are bounded by

$$\rho_{i,\mathrm{ZF}} \leq \gamma_i^2 + \frac{1}{4}\sum_{j=i+1}^{n} \gamma_j^2. \tag{56}$$

We claim that the right-hand side of (56) decreases with i. To see this, it is sufficient to show $\gamma_i^2 \geq \frac{3}{4}\gamma_{i+1}^2$. We apply Mordell's inequality together with the inequality $\gamma_n \leq (2/\sqrt{3})^{n-1}$, yielding

$$\gamma_{i+1} \leq \gamma_i^{i/(i-1)} = \gamma_i \gamma_i^{1/(i-1)} \leq \gamma_i \cdot 2/\sqrt{3}.$$

Thus the claim is proven.

Accordingly, the right-hand side of (56) attains the maximum when $i = 1$, and

$$\rho_{\mathrm{ZF}} \leq 1 + \frac{1}{4}\sum_{j=2}^{n} \gamma_j^2. \tag{57}$$

Using the bound $\gamma_j \leq 2j/3$ for $j \geq 2$, we can obtain

$$\rho_{\mathrm{ZF}} \leq \frac{8}{9} + \frac{1}{9} \cdot \frac{n(n+1)(2n+1)}{6}. \tag{58}$$

The right side of (58) is never larger than n^3, and will be dominated by $n^3/27$ as n goes large.

VII. Numerical Results and Discussion

To visualize the behaviors of proximity factors, we plot the best upper bounds in Table I. In Fig. 2, we show the upper bounds on the proximity factors for KZ reduction and for LLL reduction with best performance (i.e., $\delta = 1$). For KZ reduction, we employ (44) and (40) in conjunction with the bound (12) on the Hermite constant and the exact value (14) for $n \leq 8$. It is clear that the bounds on proximity factors for KZ reduction are smaller than those for LLL reduction. This is expected since KZ reduction is a stronger notion of lattice reduction.

Fig. 3 shows the upper bounds on ρ_{ZF} with primal and dual LLL reduction. Obviously, the bound for reducing the dual basis is smaller for the purpose of ZF.

In Fig. 4, we show the upper bounds on the proximity factors for KZ reduction of the primal as well as dual bases. Obviously, for ZF, the bound for reducing the dual basis is much smaller. Meanwhile, we can see that, for SIC, the gain due to dual KZ reduction is not significant. The reason is that as far as the Gram-Schmidt vectors are concerned, the primal and dual KZ reduction are in fact not too much different, as also recognized in [10]. Because the product $\prod_{i=1}^{n} \|\hat{\mathbf{b}}_i\| = \det L$ is invariant under lattice reduction, choosing short Gram-Schmidt vectors in the beginning will force long Gram-Schmidt vectors in the end, and vice versa.

Compared with primal basis reduction, the dual-basis reformulation is more natural in the following senses.

For ZF, we have $\mathbf{B}^\dagger \mathbf{y} = \mathbf{x} + \mathbf{B}^\dagger \mathbf{n}$. Then, the noise variance associated with the i-th output is proportional to $\|\mathbf{b}_i^*\|^2$. Obviously, it makes sense to minimize $\|\mathbf{b}_i^*\|^2$.

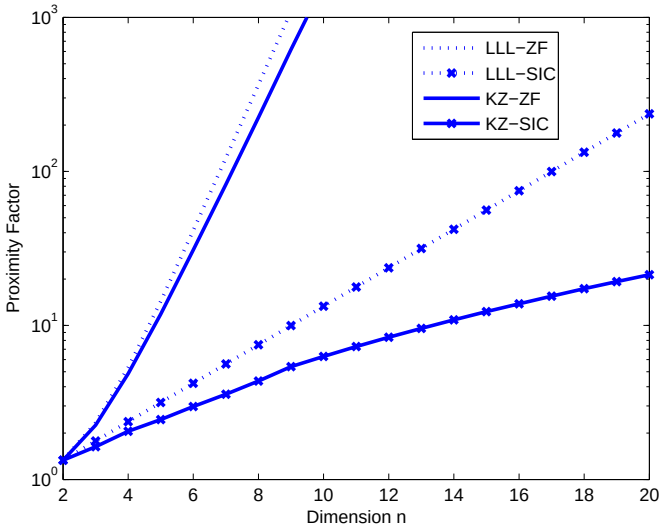


Fig. 2. Upper bounds on the proximity factors ρ_{ZF} and ρ_{SIC} for KZ reduction and for LLL reduction with $\delta = 1$.

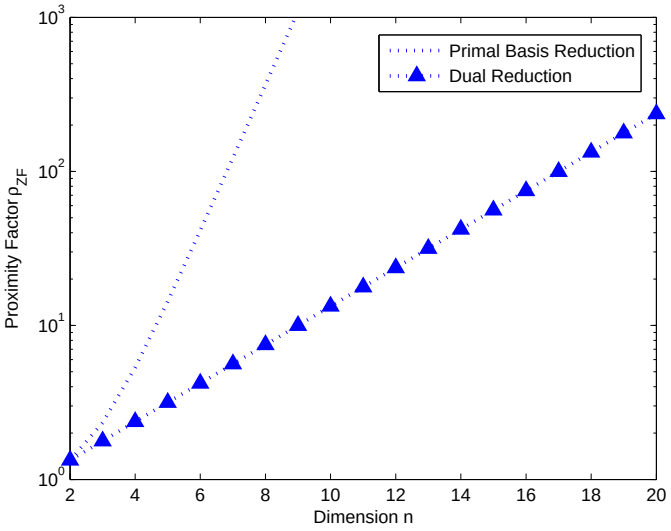


Fig. 3. Upper bounds on the proximity factors ρ_{ZF} for LLL reduction with $\delta = 1$.

Most existing reduction notions aim to find short vectors one after another; hence when applied to the dual basis, they can be viewed as various greedy heuristics of finding the optimal bases for decoding purposes. In particular, the dual Minkowski reduction can be viewed as a greedy algorithm to approximate the ZF criterion, while the dual KZ reduction as a greedy algorithm to approximate the SIC criterion, since they will successively find the shortest vector and the shortest Gram-Schmidt vector for the dual basis, respectively.

To verify the benefit of dual reduction for LLL-ZF, we simulate the BER performance of an 8×8 MIMO system with i.i.d. complex standard normal channel coefficients. The complex basis matrix $\mathbf{H}$ is converted into its real equivalent by

$$\mathbf{B} = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix},$$

and then $\mathbf{B}$ is reduced. The signal constellation at each antenna is 64QAM with Gray mapping. With such a signal

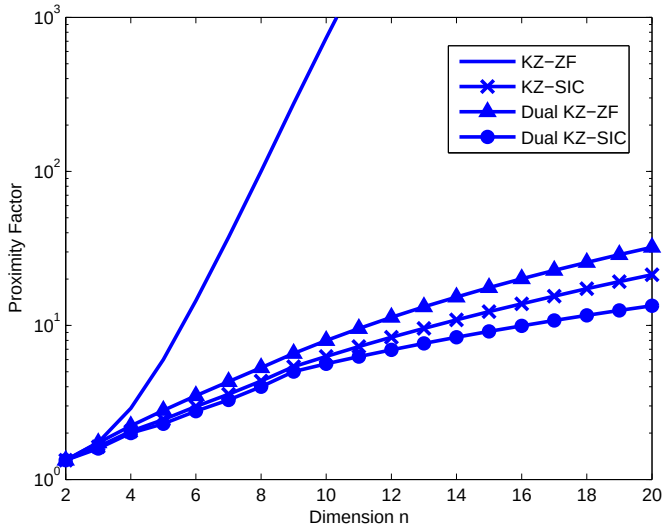


Fig. 4. Upper bounds on the proximity factors for KZ reduction of the primal and dual bases.

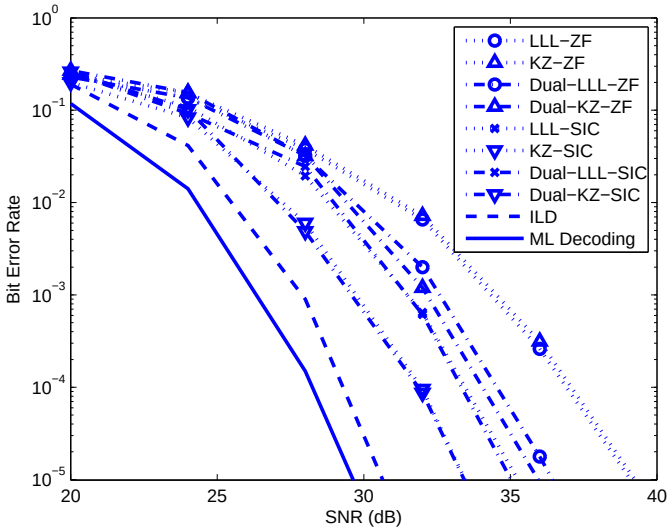


Fig. 5. Simulated BER with LLL reduction ($\delta = 3/4$) and KZ reduction for 64QAM over an 8×8 complex-valued MIMO fading channel.

model, the SNR at each receive antenna is defined as SNR $= n_T E_{x \in 64\text{QAM}}[x^2]/\sigma^2$, where n_T is the number of transmit antennas. Fig. 5 shows the simulated BER with LLL reduction and KZ reduction. When the decoded lattice point falls outside of the signal boundary, we simply round it componentwise back to the 64QAM alphabet. The simulated BER performance is consistent with theoretic analysis. In particular, it is seen that dual reduction exhibits about 3 dB gain for ZF at the BER of 10^{-5}. In the simulation of SIC, though, little difference in the BER is observed between primal and dual reduction. This is not surprising given their close proximity factors (cf. Table I). Lattice decoding ignoring the boundary is also simulated, whose performance is about 1.5 dB worse than ML decoding at the BER of 10^{-5}.

We assess the tightness of the bounds by means of numerical experimentation. Each time, a basis matrix $\mathbf{B}$ is randomly generated and reduced, and then the minimum distance of the decision region (for ZF and SIC, respectively) as well as the shortest vector is found. For each value of n, the

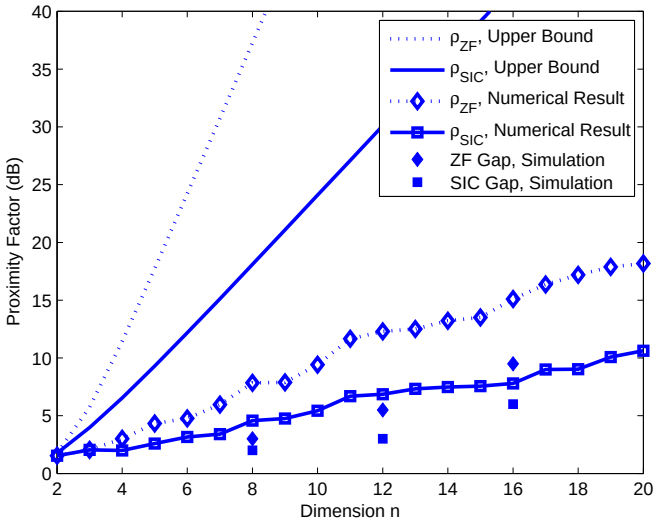


Fig. 6. Comparison of the theoretic upper bounds and numerical results of the proximity factors for LLL reduction ($\delta = 3/4$).

empirical proximity factor is obtained as the maximum over an ensemble of 10000 i.i.d. Gaussian matrices $\mathbf{B}$. While there is no guarantee that this maximum reaches the worst-case bound, it should be a reasonable indicator of the theoretic proximity factor. In particular, it can serve as an *experimental lower bound* on the theoretic proximity factor. Fig. 6 shows the results for (primal-basis) LLL reduction with $\delta = 3/4$. The numerical results (in dB) still grow linearly, but at a lower slope. Fig. 6 also shows the real SNR gap to ML decoding observed in computer simulation for $n = 8, 12, 16$; the SNR gap to ILD is a little smaller. To estimate the real SNR gap, the BER performance is simulated, and curves like those in Fig. 5 are obtained. It can be seen that the real SNR gap is smaller than the experimental proximity factor, due to the average over the MIMO channel. Fig. 7 demonstrates the numerical results for dual LLL reduction. Figs. 8 and 9 show the results for KZ and dual-KZ reductions, respectively. The similar trend can be seen. All the experiments are consistent with theoretic analysis.

In summary, we have shown that LRAD is in proximity to ILD. We have derived analytic bounds on proximity factors that quantify the worst-case gap. The same diversity order as that of ILD in MIMO fading channels comes as a direct consequence. According to the analysis, dual lattice reduction could give smaller upper bounds, especially in ZF where performance gain has been observed in computer simulation.

Finally, we point out several open issues of LRAD.

Firstly, there is room to tighten the bounds derived in this paper. In Table I, the bounds are tight for Gauss reduction and for KZ-SIC with primal reduction. The bound is very likely to be tight for KZ-SIC with dual reduction; in fact it will be tight as long as the Hermite constant is an increasing function which is very likely to be true. The bound is quite good for KZ-ZF with dual reduction (cf. Fig. 4). However, the bounds are not tight for LLL reduction and KZ-ZF with primal reduction. For LLL reduction, we have applied the

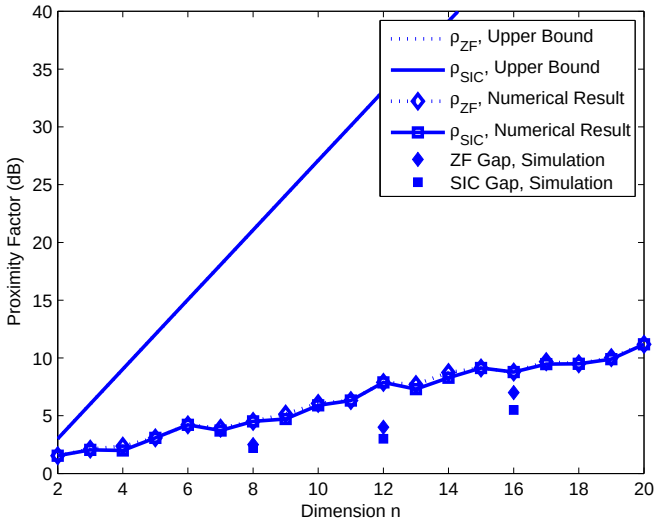


Fig. 7. Comparison of the theoretic upper bounds and numerical results of the proximity factors for dual LLL reduction ($\delta = 3/4$).

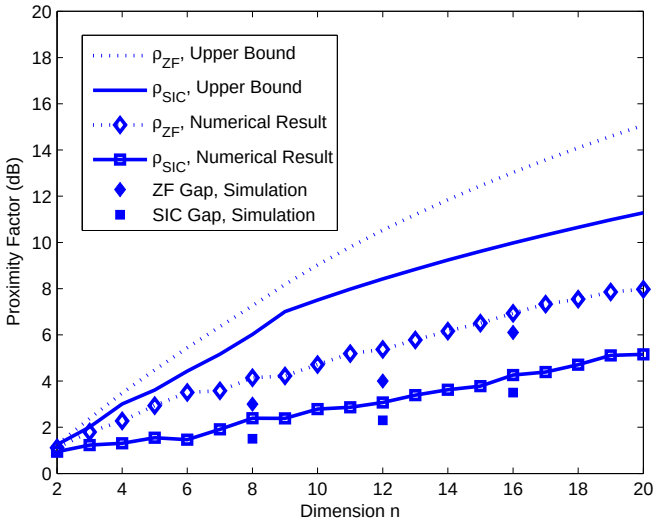


Fig. 9. Comparison of the theoretic upper bounds and numerical results of the proximity factors for dual KZ reduction.

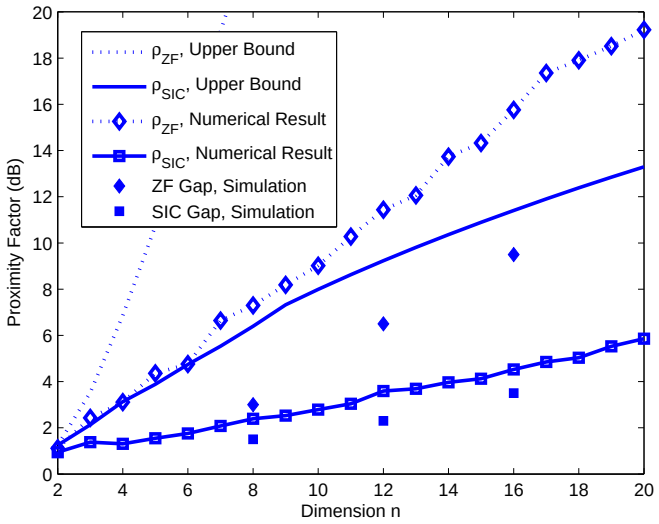


Fig. 8. Comparison of the theoretic upper bounds and numerical results of the proximity factors for KZ reduction.

trivial bound $\lambda^2(L)/\|\mathbf{b}_i\|^2 \leq 1$ or $\lambda^2(L)/\|\mathbf{b}_1\|^2 \leq 1$. But this is likely to loosen the bound, since $\beta^{-(n-1)} \leq \lambda_i^2(L)/\|\mathbf{b}_i\|^2$ and $\beta^{-(n-1)} \leq \lambda^2(L)/\|\mathbf{b}_1\|^2$ for LLL reduction [39].

Secondly, we emphasize that the worst-case bounds should be carefully interpreted. The average-case performance may be a more meaningful measure if the basis $\mathbf{B}$ is random. A smaller proximity factor does not necessarily guarantee better average performance. For example, KZ reduction must have smaller approximation factors than LLL reduction, but their BER performance appears to be very close in MIMO fading channels [15]. Hence an average-case analysis will complement the worse-case analysis. In simulations of LLL reduction, we observed that the SNR gap (in dB) indeed widens linearly with the dimension, but at a lower slope. Recently, Nguyen and Stehlé showed that the average behavior of LLL reduction is still exponential, but at a slower rate [48].

Thirdly, the analysis of dual reduction naturally raises the issue of simultaneously reducing the primal and dual bases. Methods of simultaneous reduction exist, notably those of Seysen [28], [45] and Howgrave-Graham [49]. More details can be found in the extended version of this paper [26].

Appendix I
The Angle θ_i of a Basis

We derive a formula for angel θ_i of a basis $\mathbf{B}$, which is used to compute the distance of ZF. Following Babai [13], let $\mathbf{a} = \sum_{j\neq i} \alpha_j \mathbf{b}_j$ ($\alpha_j \in \mathbb{R}$ for $j \neq i$) be an arbitrary vector on the hyperplane spanned by vectors $\mathbf{b}_1, \cdots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \cdots, \mathbf{b}_n$. The vector $\mathbf{z} = \mathbf{a} - \mathbf{b}_i$ will attain its minimum length if and only if it is perpendicular to the hyperplane. We express $\mathbf{z}$ in terms of the Gram-Schmidt vectors as $\mathbf{z} = \sum_{j=1}^{n} r_j \hat{\mathbf{b}}_j$. Set $\alpha_i = -1$. Then we have the expression [39]

$$r_j = \alpha_j + \sum_{t=j+1}^{n} \alpha_t \mu_{t,j}.$$

We can see

$$\|\mathbf{b}_i\|^2 \sin^2\theta_i = \min \|\mathbf{z}\|^2 = \min \sum_{j=1}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2 \qquad (59)$$

where the minimum is taken over the coefficients α_j. It is not difficult to see that $r_j = 0$ for $j < i$ if $\mathbf{z}$ is orthogonal to the hyperplane. Accordingly, we may exclude the $j < i$ terms, to yield

$$\|\mathbf{b}_i\|^2 \sin^2\theta_i = \min \sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2. \qquad (60)$$

This was also done by Babai [13]. Here we have explicitly shown that excluding the $j < i$ terms does not weaken the bound.

For notational convenience, define $\boldsymbol{\alpha} = [\alpha_i, \cdots, \alpha_n]^T$, $\mathbf{r} = [r_i, \cdots, r_n]^T$, $\mathbf{M}$ be $(n-i+1) \times (n-i+1)$ bottom-right submatrix of $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$ be the diagonal matrix $\text{diag}(\|\hat{\mathbf{b}}_i\|^2, \|\hat{\mathbf{b}}_{i+1}\|^2, \cdots, \|\hat{\mathbf{b}}_n\|^2)$, and $\mathbf{A} = \mathbf{M}\boldsymbol{\Lambda}\mathbf{M}^T$. We have the expressions $\mathbf{r} = \mathbf{M}^T\boldsymbol{\alpha}$ and

$$\sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2 = \mathbf{r}^T\boldsymbol{\Lambda}\mathbf{r} = \boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha}.$$

We want to find the minimum of $\boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha}$ subject to the constraint $\alpha_i = -1$. This can be achieved by using the Lagrangian multiplier method. Define the objective function

$$f = \boldsymbol{\alpha}^T\mathbf{A}\boldsymbol{\alpha} + s(\alpha_i + 1).$$

Nulling the partial derivative of f with respect to $\boldsymbol{\alpha}$, we have

$$2\mathbf{A}\boldsymbol{\alpha} + s\mathbf{e}_1 = 0$$

where the unit vector $\mathbf{e}_1$ has all zero elements except that the first is one. From this we obtain $\boldsymbol{\alpha} = -\frac{s}{2}\mathbf{A}^{-1}\mathbf{e}_1$. Furthermore, since we know $\alpha_i = -\frac{s}{2}(\mathbf{A}^{-1})_{1,1} = -1$, we have $s = 2/(\mathbf{A}^{-1})_{1,1}$. Therefore,

$$f_{min} = \frac{s^2}{4}\mathbf{e}_i^T\mathbf{A}^{-1}\mathbf{e}_i = \frac{s^2}{4}(\mathbf{A}^{-1})_{1,1} = \frac{1}{(\mathbf{A}^{-1})_{1,1}}, \quad (61)$$

and from (59) we deduce

$$\sin^2\theta_i = \frac{1}{\|\mathbf{b}_i\|^2(\mathbf{A}^{-1})_{1,1}}. \quad (62)$$

We have expressed θ_i in terms of $\mathbf{b}_i$ and the Gram-Schmidt orthogonalization of $\mathbf{B}$. Given $\mathbf{B}$, the angle θ_i can easily be calculated.

APPENDIX II
PROOF OF PROPOSITION 1

Proof: By definition, we have that for matrix $\mathbf{B}$ with full column rank

$$\mathbf{B}^{\dagger T} = [(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T]^T = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-T}.$$

Substituting $\mathbf{B} = \hat{\mathbf{B}}\boldsymbol{\mu}^T$, we arrive at

$$\mathbf{B}^{\dagger T} = \hat{\mathbf{B}}\boldsymbol{\mu}^T(\boldsymbol{\mu}\hat{\mathbf{B}}^T\hat{\mathbf{B}}\boldsymbol{\mu}^T)^{-T} = \hat{\mathbf{B}}(\hat{\mathbf{B}}^T\hat{\mathbf{B}})^{-T}\boldsymbol{\mu}^{-1} = \hat{\mathbf{B}}^{\dagger T}\boldsymbol{\mu}^{-1}.$$

Hence the reversed dual basis can be written as

$$\mathbf{B}^* = \mathbf{B}^{\dagger T}\mathbf{J} = \hat{\mathbf{B}}^{\dagger T}\boldsymbol{\mu}^{-1}\mathbf{J}.$$

As $\mathbf{J}\mathbf{J} = \mathbf{I}$, it can be rewritten into

$$\mathbf{B}^* = \hat{\mathbf{B}}^{\dagger T}\mathbf{J} \cdot \mathbf{J}\boldsymbol{\mu}^{-1}\mathbf{J}.$$

Since $\mathbf{J}\boldsymbol{\mu}^{-1}\mathbf{J}$ is upper triangular with unit diagonal elements, and since the columns of $\hat{\mathbf{B}}^{\dagger T}\mathbf{J}$ are orthogonal to each other, this is exactly the Gram-Schmidt orthogonalization of $\mathbf{B}^*$. Therefore, we have $\hat{\mathbf{B}}^* = \hat{\mathbf{B}}^{\dagger T}\mathbf{J}$ and $(\boldsymbol{\mu}^*)^T = \mathbf{J}\boldsymbol{\mu}^{-1}\mathbf{J}$. As $\mathbf{J} = \mathbf{J}^T$, $\boldsymbol{\mu}^* = \mathbf{J}\boldsymbol{\mu}^{-T}\mathbf{J}$ holds. ■

APPENDIX III
PROOF OF LEMMA 1

Babai showed the bound [13]

$$\sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2 \geq (2/9)^{n-i}\|\hat{\mathbf{b}}_i\|^2. \quad (63)$$

Here we derive a better lower bound by using a different approach, which leads to a better bound for the angle θ_i. Basically, we examine the maximum of the first element of matrix $\mathbf{A}^{-1} = \mathbf{M}^{-T}\boldsymbol{\Lambda}^{-1}\mathbf{M}^{-1}$ for a size-reduced (lower-triangular) matrix $\mathbf{M}$. To do so, we need the following lemma:

Lemma 2: The absolute value of the (i,j)-th entry of the inverse $\mathbf{M}^{-1}$ of an $l \times l$ size-reduced matrix $\mathbf{M}$ is not greater than

$$\frac{1}{3} \cdot \left(\frac{3}{2}\right)^{i-j}, \quad 1 \leq j < i \leq l, \quad (64)$$

and this is achieved when all off-diagonal elements of $\mathbf{M}$ are equal to $-1/2$.

Proof: We prove it by induction on the dimension l. Note that $\mathbf{M}^{-1}$ is lower triangular, with unit diagonal elements.

Lemma 2 is obviously true when $l = 2$, since the off-diagonal element of $\mathbf{M}^{-1}$ is $-m_{2,1}$, and the maximum $1/2$ is achieved when $m_{2,1} = -1/2$.

Suppose Lemma 2 is true for $l = k-1$. When $l = k$, we partition the matrix $\mathbf{M}$ into the form

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{k-1} & \mathbf{0} \\ \mathbf{h}^T & 1 \end{bmatrix}.$$

Using the formula for the inverse of a partitioned matrix [50], $\mathbf{M}^{-1}$ can be expressed as

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{M}_{k-1}^{-1} & \mathbf{0} \\ -\mathbf{h}^T\mathbf{M}_{k-1}^{-1} & 1 \end{bmatrix}.$$

Here, the (i,j)-th element of $\mathbf{M}_{k-1}^{-1}$ is given by (64) for $1 \leq j < i \leq k-1$. Since they are all positive, each element of $-\mathbf{h}^T\mathbf{M}_{k-1}^{-1}$ is maximized when $\mathbf{h}$ is an all $-1/2$ vector.

Consider the first element of $-\mathbf{h}^T\mathbf{M}_{k-1}^{-1}$, given by

$$\frac{1}{2}\left[1 + \sum_{j=1}^{k-2}\frac{1}{3}\cdot\left(\frac{3}{2}\right)^j\right] = \frac{1}{3}\cdot\left(\frac{3}{2}\right)^{k-1},$$

which verifies (64). The other elements of $-\mathbf{h}^T\mathbf{M}_{k-1}^{-1}$ can be verified in the same way. ■

Denote by $\mathbf{p}$ the first column of $\mathbf{M}^{-1}$. Then we have

$$\begin{aligned}(\mathbf{A}^{-1})_{1,1} &= \mathbf{p}^T\boldsymbol{\Lambda}^{-1}\mathbf{p} \\ &\leq \|\hat{\mathbf{b}}_i\|^{-2} + \sum_{j=1}^{n-i}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\|\hat{\mathbf{b}}_{i+j}\|^{-2}. \end{aligned} \quad (65)$$

Note that the only condition for (65) is that the basis is size-reduced, i.e., $|\mu_{i,j}| \leq 1/2$ for $i < j$. The choice of the parameter δ would have no effect on this condition.

For an LLL-reduced basis, $\|\hat{\mathbf{b}}_{i+j}\|^2 \geq \beta^{-j}\|\hat{\mathbf{b}}_i\|^2$. Hence,

$$\begin{aligned}(\mathbf{A}^{-1})_{1,1} &\leq \left[1 + \sum_{j=1}^{n-i}\left(\frac{1}{3}\right)^2\left(\frac{3}{2}\right)^{2j}\beta^j\right]\|\hat{\mathbf{b}}_i\|^{-2} \\ &= \left[\frac{\beta}{9\beta-4}\left(\frac{9\beta}{4}\right)^{n-i} + \frac{8\beta-4}{9\beta-4}\right]\|\hat{\mathbf{b}}_i\|^{-2}, \end{aligned} \quad (66)$$

and correspondingly,

$$\sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2 \geq \left[\frac{\beta}{9\beta - 4} \left(\frac{9\beta}{4} \right)^{n-i} + \frac{8\beta - 4}{9\beta - 4} \right]^{-1} \|\hat{\mathbf{b}}_i\|^2. \tag{67}$$

When $\beta = 2$, the new bound is asymptotically tighter by a factor of 7 than Babai's lower bound (67).

Substituting (31) and (66) into (62), we obtain Lemma 1. Note that we have also incorporated the better estimate (31) of the length of $\mathbf{b}_i$.

Appendix IV
Lower Bound for Dual Size Reduction

We derive the counterparts of (65) and (67) when the reversed dual basis is size-reduced. The notation in Appendix III is followed. The difference is that, by the following lemma, all off-diagonal elements of $\mathbf{M}^{-1}$ lie in interval $[-1/2, 1/2]$.

Lemma 3: If the dual basis is size-reduced, then the off-diagonal elements of $\boldsymbol{\mu}^{-1}$ lie in the interval $[-1/2, 1/2]$.

Proof: By definition, the off-diagonal elements of $\boldsymbol{\mu}^*$ lie in $[-1/2, 1/2]$ if the reversed dual basis $\mathbf{B}^*$ is size-reduced. By Proposition 1, $\boldsymbol{\mu}^{-1} = \mathbf{J}(\boldsymbol{\mu}^*)^T\mathbf{J}$. This proves Lemma 3. ■

Consequently,

$$\begin{aligned} (\mathbf{A}^{-1})_{1,1} &= \mathbf{p}^T \boldsymbol{\Lambda}^{-1} \mathbf{p} \\ &\leq \|\hat{\mathbf{b}}_i\|^{-2} + \frac{1}{4} \sum_{j=1}^{n-i} \|\hat{\mathbf{b}}_{i+j}\|^{-2}, \end{aligned} \tag{68}$$

and

$$\sum_{j=i}^{n} r_j^2 \|\hat{\mathbf{b}}_j\|^2 \geq \left(\|\hat{\mathbf{b}}_i\|^{-2} + \frac{1}{4} \sum_{j=1}^{n-i} \|\hat{\mathbf{b}}_{i+j}\|^{-2} \right)^{-1}. \tag{69}$$

In contrast, when the primal basis is size-reduced, the off-diagonal elements of $\mathbf{M}^{-1}$ can only be bounded by (64), which can be much larger than $1/2$. This analysis clearly shows that it is the dual basis, rather than the primal basis, that should be reduced for the purpose of ZF.

Acknowledgment

The author would like to thank W. H. Mow and N. Howgrave-Graham for helpful discussions.

References

[1] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2389–2402, Oct. 2003.

[2] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The Golden code: A 2 x 2 full-rate space-time code with nonvanishing determinants," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1432–1436, Apr. 2005.

[3] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector perturbation technique for near-capacity multiantenna multiuser communications-Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, pp. 537–544, Mar. 2005.

[4] S. Yang and J.-C. Belfiore, "Towards the optimal amplify-and-forward cooperative diversity scheme," *IEEE Trans. Inform. Theory*, vol. 53, pp. 3114–3126, Sept. 2007.

[5] A. D. Murugan, H. E. Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inform. Theory*, vol. 52, pp. 933–953, Mar. 2006.

[6] J. Luo, K. Pattipati, P. Willett, and G. Levchuk, "Fast optimal and suboptimal any-time algorithms for CDMA multiuser detection based on branch and bound," *IEEE Trans. Commun.*, vol. 52, pp. 632–642, Apr. 2004.

[7] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1–13, Aug. 2006.

[8] W. H. Mow, "Maximum likelihood sequence estimation from the lattice viewpoint," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1591–1600, Sept. 1994.

[9] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1639–1642, July 1999.

[10] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2201–2214, Aug. 2002.

[11] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, pp. 2806–2818, Aug. 2005.

[12] J. Jaldén and B. Ottersen, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, pp. 1474–1484, Apr. 2005.

[13] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, 1986.

[14] H. Yao and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. Globecom'02*, Taipei, China, Nov. 2002, pp. 17–21.

[15] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Information Theory Workshop*, Paris, France, Mar. 2003, pp. 345–348.

[16] U. Erez and R. Zamir, "Achieving (1/2)log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2293–2314, Oct. 2004.

[17] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 968–985, June 2004.

[18] J. Jaldén and P. Elia, "LR-aided MMSE lattice decoding is DMT optimal for all approximately universal codes," in *Proc. Int. Symp. Inform. Theory (ISIT'09)*, Seoul, Korea, 2009.

[19] M. Taherzadeh and A. K. Khandani, "On the limitations of the naive lattice decoding," *IEEE Trans. Inform. Theory*, submitted for publication. [Online]. Available: http://arxiv.org/abs/0709.0035

[20] C. Ling, W. H. Mow, K. H. Li, and A. C. Kot, "Multiple-antenna differential lattice decoding," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 1281–1289, Sept. 2005.

[21] C. Ling, "Towards characterizing the performance of approximate lattice decoding in MIMO communications," in *Proc. Int. Symp. Turbo Codes and ITG Conf. Source Channel Coding*, Munich, Germany, Apr. 2006.

[22] ——, "Approximate lattice decoding: Primal versus dual lattice reduction," in *Proc. Int. Symp. Inform. Theory (ISIT'06)*, Seattle, WA, July 2006.

[23] ——, "Improved upper bounds for approximate lattice decoding with dual-basis reduction," in *Proc. Int. Conf. Commun. (ICC'08)*, Beijing, China, May 2008.

[24] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding," *IEEE Trans. Inform. Theory*, vol. 53, pp. 4801–4805, Dec. 2007.

[25] X. Ma and W. Zhang, "Performance analysis for V-BLAST systems with lattice-reduction aided linear equalization," *IEEE Trans. Commun.*, vol. 56, pp. 309–318, Feb. 2008.

[26] C. Ling, "On the proximity factors of lattice reduction-aided decoding," Technical Report, Imperial College London, Apr. 2007. [Online]. Available: http://www.commsp.ee.ic.ac.uk/ cling/

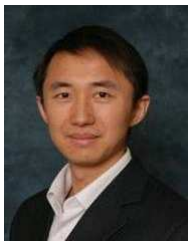
[27] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection," *IEEE Trans. Signal Processing*, vol. 57, pp. 2701–2710, July 2009.

[28] D. Seethaler, G. Matz, and F. Hlawatsch, "Low-complexity MIMO data detection using Seysen's lattice reduction algorithm," in *Proc. IEEE ICASSP 2007*, Honolulu, USA, Apr. 2007, pp. III–53–III–56.

[29] J.-K. Zhang, A. Kavcic, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inform. Theory*, vol. 51, pp. 154–172, Jan. 2005.

[30] N. Prasad and M. K. Varanasi, "Analysis of decision feedback detection for MIMO rayleigh-fading channels and the optimization of power and rate allocation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1009–1025, June 2004.

[31] Y. Jiang and M. K. Varanasi, "The effect of ordered detection and antenna selection on diversity gain of decision feedback detector," in *Proc. IEEE ICC 2007*, Glasgwo, Scotland, UK, June 2007, pp. 5383–5388.
[32] P. M. Gruber and C. G. Lekkerkerker, *Geometry of Numbers*. Amsterdam, Netherlands: Elsevier, 1987.
[33] J. C. Lagarias, W. H. Lenstra, and C. P. Schnorr, "Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice," *Combinatorica*, vol. 10, no. 4, pp. 333–348, 1990.
[34] L. J. Mordell, "Observation on the minimum of a positive quadratic form in eight variables," *Bull. London Math. Soc.*, vol. 19, pp. 1–6, 1944.
[35] H. F. Blichfeldt, "A new principle in the geometry of numbers, with some applications," *Trans. Am. Math. Soc.*, vol. 16, pp. 227–235, 1914.
[36] H. Minkowski, *Geometrie der Zahlen*, Leipzig, Germany, 1896.
[37] A. Korkine and G. Zolotareff, "Sur les formes quadratiques," *Math. Annalen*, vol. 6, pp. 366–389, 1873.
[38] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.
[39] A. K. Lenstra, J. H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.
[40] C. Ling and N. Howgrave-Graham, "Effective LLL reduction for lattice decoding," in *Proc. Int. Symp. Inform. Theory (ISIT'07)*, Nice, France, June 2007.
[41] J. Jaldén, D. Seethaler, and G. Matz, "Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems," in *Proc. ICASSP'08*, Las Vegas, NV, US, 2008, pp. 2685–2688.
[42] S. Verdu, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.
[43] N. Howgrave-Graham, "Finding small roots of univariate modular equations revisited," in *Proc. the 6th IMA Int. Conf. Crypt. Coding*, 1997, pp. 131–142.
[44] C. Ling, S. Liu, L. Luzzi, and D. Stehlé, "Decoding by embedding: Correct decoding radius and DMT optimality," submitted to ISIT 2011. [Online]. Available: http://www.commsp.ee.ic.ac.uk/ cling/
[45] M. Seysen, "Simultaneous reduction of a lattice basis and its reciprocal basis," *Combinatorica*, vol. 13, pp. 363–376, 1993.
[46] C. P. Schnorr, "A hierarchy of polynomial time lattice basis reduction algorithms," *Theor. Comput. Sci.*, vol. 53, no. 2-3, pp. 201–224, 1987.
[47] N. Howgrave-Graham, NTRU Cryptosystems, Burlington, MA, Dec. 2006, private communication.
[48] P. Nguyen and D. Stehlé, "LLL on the average," *Lecture Notes in Computer Science*, vol. 4076, pp. 238–256, 2006.
[49] N. Howgrave-Graham, "Isodual reduction of lattices," 2007, Cryptology ePrint Archive: Report 2007/105.
[50] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.



Cong Ling received the B.S. and M.S. degrees in electrical engineering from the Nanjing Institute of Communications Engineering, Nanjing, China, in 1995 and 1997, respectively, and the Ph.D. degree in electrical engineering from the Nanyang Technological University, Singapore, in 2005.

He is currently a Lecturer in the Electrical and Electronic Engineering Department at Imperial College London. His research interests are in the general area of wireless communications, especially network information theory, coding, and lattice methods. Before joining Imperial College, he had been on the faculties of Nanjing Institute of Communications Engineering and King's College.

Dr. Ling is an Associate Editor of IEEE Transactions on Vehicular Technology. He has also served on the program committees of several international conferences including IEEE Information Theory Workshop (2006, 2009), Globecom, and ICC.