

# Lattice-Reduction-Aided Detectors for MIMO Communication Systems

Huan Yao and Gregory W. Wornell  
 Dept. EECS and Research Laboratory of Electronics  
 MIT, Cambridge MA 02139, USA

**Abstract**— Lattice-reduction (LR) techniques are developed for enhancing the performance of multiple-input multiple-output (MIMO) digital communication systems. When used in conjunction with traditional linear and nonlinear detectors, LR techniques substantially close the gap to fundamental performance limits with little additional system complexity. Results for individual channels and ensembles are developed, and illustrated in detail for the case of small ( $2 \times 2$ ), uncoded, coherent systems. For example, we show that, relative to the maximum likelihood bound, LR techniques get us within 3dB for any Gaussian channel, and allow us to achieve the same diversity on the Rayleigh fading channel, when sufficiently large constellations are used.

## I. INTRODUCTION

A wide range of wireless communication problems involve multiple-input multiple-output (MIMO) channels. These include the multiuser detection problem and a variety of multiple-antenna transmission problems.

For the system designer, the goal is to achieve an attractive performance-complexity tradeoff. At one end of the spectrum, maximum likelihood detection is optimal, but its complexity generally makes it impractical. A variety of other detectors, both linear and nonlinear, require substantially less complexity, but sacrifice performance significantly. In this paper, we propose very low complexity receiver structures based on lattice-reduction techniques that provide near-optimal performance.

In the system of interest, the transmitted signal vector  $\mathbf{x}$  is to be detected from the received signal vector  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ , where the channel matrix  $\mathbf{H}$  is  $N_r \times N_t$ ,  $N_r \geq N_t$ , and the entries of  $\mathbf{w}$  are independent, circularly-symmetric complex additive white Gaussian noise (AWGN) with density  $\mathcal{CN}(0, 2\sigma_w^2)$ . We consider the case in which the channel matrix  $\mathbf{H}$  is effectively known at the receiver but not at the transmitter. We consider both fixed and random  $\mathbf{H}$  cases. We further restrict our attention to uncoded systems in which the entries of  $\mathbf{x}$  are drawn independently from some constellation.

## II. TRADITIONAL DETECTORS

Lattice reduction is used in conjunction with traditional detector structures, the key features of which we now summarize.

An important performance bound corresponds to *maximum likelihood detection* (MLD), which minimizes the probability of block error. In the case where the noise is AWGN, the minimum distance rule is used,

$$\hat{\mathbf{x}}_{\text{MLD}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2. \quad (1)$$

This work has been supported in part by the U.S. Army Research Laboratory under Collaborative Technology Alliance No. DAAD19-01-2-0011.

In the absence of special structure, MLD requires computing distances to every codeword to find the closest one. Therefore, it has exponential complexity in transmission rate.

By contrast, linear detectors have much lower complexity. They take the form of  $\hat{\mathbf{x}} = f(\mathbf{A}\mathbf{y})$ , where  $\mathbf{A}$  is some matrix and  $f(\cdot)$  is a slicer, which quantizes each entry of  $\mathbf{A}\mathbf{y}$  to the nearest constellation symbol to obtain  $\hat{\mathbf{x}}$ . For familiar constellations such as 4-QAM or 16-QAM, this quantization can be implemented with very little complexity.

The choice  $\mathbf{A} = \mathbf{H}^{-1}$ , where  $\mathbf{H}^{-1}$  denotes the pseudo-inverse  $(\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger$  when  $N_r > N_t$ , corresponds to what is sometimes referred to as *inverse channel detection* (ICD) [3], or, in the case of the multiuser detection problem, the decorrelator. As is well-known, the performance of ICD can suffer dramatically due to noise enhancement if  $\mathbf{H}$  is near singular. Indeed, since  $\mathbf{H}^{-1}\mathbf{y} = \mathbf{x} + \mathbf{H}^{-1}\mathbf{w}$ , the effective noise at the slicer input is  $\mathbf{H}^{-1}\mathbf{w}$ . Other linear detectors include the minimum mean square estimator (MMSE) detector, which offers slightly better performance by mitigating noise enhancement, but is still far from the performance of MLD.

A class of nonlinear detectors that offer better performance with only a modest increase in complexity is that based on successive cancellation. An example is the *Bell Labs Layered Space-Time* (BLAST) receiver [1]. The main steps of the simplest version of BLAST detection are nulling and cancellation.

*Nulling* : First, the channel matrix is factored as  $\mathbf{H} = \mathbf{Q}\mathbf{R}$ , where  $\mathbf{Q}$  is unitary and  $\mathbf{R}$  is upper triangular. Next, the received signal is preprocessed to obtain  $\mathbf{y}' = \mathbf{Q}^\dagger \mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{w}'$ , where  $\mathbf{w}' = \mathbf{Q}^\dagger \mathbf{w}$  and  $\dagger$  denotes the conjugate transpose operation, so

$$\begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_{N_t} \end{bmatrix} = \begin{bmatrix} r_{11} & \cdots & \cdots & r_{1N_t} \\ 0 & r_{22} & \cdots & r_{2N_t} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{N_t N_t} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_t} \end{bmatrix} + \begin{bmatrix} w'_1 \\ w'_2 \\ \vdots \\ w'_{N_t} \end{bmatrix}. \quad (2)$$

*Cancellation* : Using the preprocessed data (2), the entries of  $\mathbf{x}$  are detected one by one in decreasing order. Specifically, after detecting  $x_k, \dots, x_{N_t}$ , we can subtract their interference out of  $y'_{k-1}$  to detect  $x_{k-1}$ .

Note that if we did not quantize each  $\hat{x}_k$  to the nearest constellation symbol as we proceeded, this form of detection would specialize to ICD. Thus, this quantization serves an important noise-cancellation role.

A major problem with BLAST detection is error propagation. The entry detected first usually has the smallest signal to noise ratio (SNR) and the most error. Unfortunately, detecting later entries correctly vitally depends on having correctly decoded previous entries. For this reason, in an uncoded system, where error correction is not used, the error rate for BLAST detection

is typically dominated by that of the first entry, and thus far from optimal.

To develop a framework within which to introduce lattice reduction, consider MLD, ICD, and BLAST detection in the  $2 \times 2$  (real) example shown in Fig. 1. The transmitted symbols  $x_1$  and  $x_2$  are each integers between  $-N$  and  $N$ , where  $N$  is large, and the channel matrix is, for purpose of illustration,  $\mathbf{H} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$ .

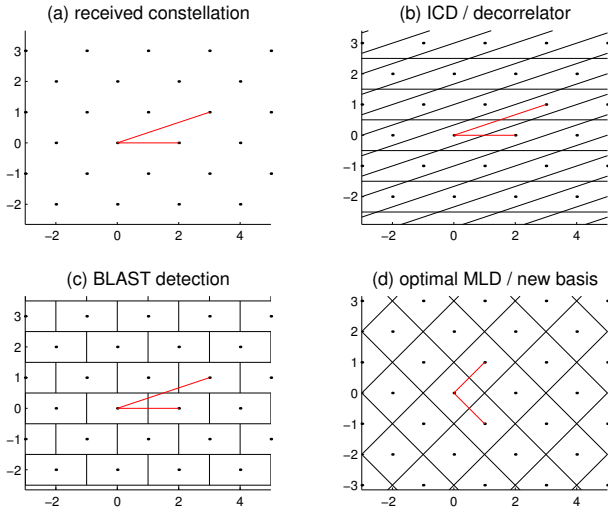


Fig. 1. Comparison of decision boundaries for various detection methods.

The received constellation  $\mathbf{H}\mathbf{x}$  is shown in (a). It can be viewed as a lattice with basis vectors being the two columns of  $\mathbf{H}$ , which are drawn to show the distortion of the lattice. The decision boundaries for ICD, BLAST detection, and MLD are shown in (b), (c), and (d), respectively. For ICD, the decision regions are undesirably elongated and narrow parallelograms; small amounts of noise would lead to detection errors. This is due to the two basis vectors being highly correlated. For BLAST detection, the decision regions are rectangular, because one entry of  $\mathbf{x}$  is detected at a time. While better than ICD, it is still inferior to the optimal decision boundary drawn in (d), whose optimality is apparent by inspection.

In this particular example, if we were to consider the lattice basis vectors to be  $[1 \ 1]^T$  and  $[1 \ -1]^T$  instead of  $[2 \ 0]^T$  and  $[3 \ 1]^T$ , where  $^T$  denotes the transpose operator, then the decision boundaries for ICD and BLAST detection would coincide with those of MLD, and therefore be optimal.

While a basis change cannot always lead to optimum performance, it can in general improve performance. In particular, changing the lattice basis to be more orthogonal and shorter, the sense of which we will make precise later, we can generally obtain better decision boundaries. And the more correlated the columns of  $\mathbf{H}$ , the more significant the improvements. Note that changing lattice basis does not change the lattice. The problem of finding the optimal lattice basis is called the *lattice reduction* (LR) problem.

### III. LATTICE REDUCTION

A lattice in  $n$  complex dimensions can be described by  $\mathcal{L} = \{\mathbf{s} \mid \mathbf{s} = \mathbf{B}\boldsymbol{\lambda}\}$  where  $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$  is a ma-

trix whose columns are basis vectors for the lattice and  $\boldsymbol{\lambda} = [\lambda_1 \ \lambda_2 \ \cdots \ \lambda_n]^T$  is a vector of complex integer weights, i.e.,  $\lambda_i \in \mathbb{Z} + \mathbb{Z}j$  with  $\mathbb{Z}$  denoting the set of integers.

For any lattice  $\mathcal{L}$  there are many possible bases. Indeed, if  $\mathbf{B}$  is a basis, so is  $\mathbf{B}' = \mathbf{B}\mathbf{P}$  for any matrix  $\mathbf{P}$  such that both  $\mathbf{P}$  and  $\mathbf{P}^{-1}$  have integer entries. Specifically, a point  $\mathbf{s}$  represented by  $\mathbf{x}$  in the basis  $\mathbf{B}$  is represented by  $\mathbf{z} = \mathbf{P}^{-1}\mathbf{x}$  in the basis  $\mathbf{B}'$ , i.e.,  $\mathbf{s} = \mathbf{B}\mathbf{x} = (\mathbf{B}\mathbf{P})(\mathbf{P}^{-1}\mathbf{x}) = \mathbf{B}'\mathbf{z}$ .

The basic idea behind using lattice reduction in conjunction with traditional low-complexity detectors is to operate in a chosen lattice basis that is optimized for those detectors, as shown in Fig. 2.

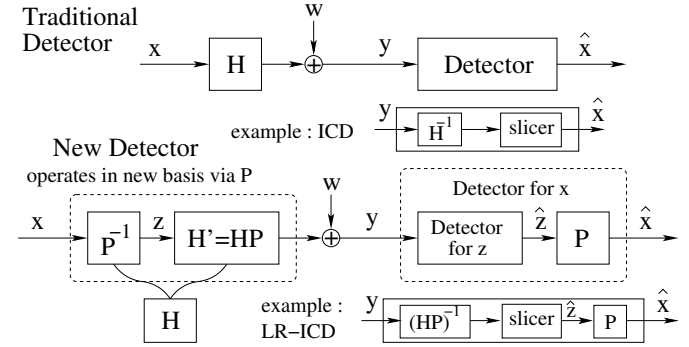


Fig. 2. Using lattice reduction in conjunction with traditional detectors.

In the traditional system, the detector compensates for the original channel  $\mathbf{H}$  to produce  $\hat{\mathbf{x}}$ . In the new system, we perform a basis change via a matrix  $\mathbf{P}$ , specifically

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} = (\mathbf{H}\mathbf{P})(\mathbf{P}^{-1}\mathbf{x}) + \mathbf{w} = \mathbf{H}'\mathbf{z} + \mathbf{w}. \quad (3)$$

With this basis change, the traditional detector is first used to compensate for the new channel  $\mathbf{H}' = \mathbf{H}\mathbf{P}$  to produce  $\hat{\mathbf{z}}$ , then  $\hat{\mathbf{x}}$  is produced via  $\hat{\mathbf{x}} = \mathbf{P}\hat{\mathbf{z}}$ . For example, if ICD is employed, then  $(\mathbf{H}')^{-1}\mathbf{y}$  is quantized to produce  $\hat{\mathbf{z}}$ , from which we obtain  $\hat{\mathbf{x}}$  via  $\hat{\mathbf{x}} = \mathbf{P}\hat{\mathbf{z}}$ .

To choose an appropriate basis changing matrix  $\mathbf{P}$ , we note that ICD and BLAST detection are more effective when the channel matrix is further from being singular. Geometrically, this corresponds to wanting the columns of the new  $\mathbf{H}'$ , which are the new basis vectors of the received constellation lattice, to be less correlated and shorter. Thus, the problem of improving the condition of the effective channel  $\mathbf{H}'$  is one of reducing the lattice basis corresponding to the original channel  $\mathbf{H}$ .

Lattice reduction in high dimensions is a known NP-hard problem. One approach is the polynomial time, sub-optimal LLL algorithm [2]. In this paper, we focus on the  $2 \times 2$  case ( $N_r = N_t = 2$ ) that arises frequently in practice, where we develop important insights and an optimal iterative algorithm.

In the  $2 \times 2$  case,  $\mathbf{H} = [\mathbf{b}_1 \ \mathbf{b}_2]$ . Let us use  $\tilde{\mathbf{b}}_1$  to denote the component of  $\mathbf{b}_1$  that is orthogonal to  $\mathbf{b}_2$ , and define  $\tilde{\mathbf{b}}_2$  similarly. For BLAST detection, the effective SNR at the point of detecting  $x_1$  and  $x_2$  are  $r_{11} = \|\mathbf{b}_1\|$  and  $r_{22} = \|\tilde{\mathbf{b}}_2\|$ , respectively. Therefore, the best basis is the one with the largest  $\min(\|\mathbf{b}_1\|, \|\tilde{\mathbf{b}}_2\|)$ . For ICD, the corresponding measure is  $\min(\|\tilde{\mathbf{b}}_1\|, \|\mathbf{b}_2\|)$ . It can be shown that the basis  $(\mathbf{u}, \mathbf{v})$ , where  $\mathbf{u}$  is the shortest (non-zero) vector in the lattice and  $\mathbf{v}$  is

the shortest vector that is not a multiple of  $\mathbf{u}$ , is the optimal basis for both detection methods. The proofs are straightforward, but are omitted due to space limitations.

### A. Reduction Algorithm

Given an original set of basis vectors  $(\mathbf{b}_1, \mathbf{b}_2)$  for a lattice with  $\|\mathbf{b}_1\| \leq \|\mathbf{b}_2\|$ , we develop an iterative algorithm to progressively reduce their correlation and converge to the desired basis vectors  $(\mathbf{u}, \mathbf{v})$ .

One intuitive way to reduce the correlation between two lattice basis vectors is to subtract integer copies of one vector out of the other. Let  $\mathbf{b}'_2 = (\mathbf{b}_2 - n\mathbf{b}_1)$  be the replacement for  $\mathbf{b}_2$ . The parameter  $n$  should be chosen so as to minimize the correlation between  $\mathbf{b}_1$  and  $\mathbf{b}'_2$ , i.e.,

$$n^* = \arg \min_{n \in \mathbb{Z} + \mathbb{Z}j} |\langle \mathbf{b}_1, \mathbf{b}_2 \rangle - n\|\mathbf{b}_1\|^2| = \left\lfloor \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|^2} \right\rfloor, \quad (4)$$

where the function  $\lfloor \cdot \rfloor$  rounds its argument to the nearest integer. For complex arguments, real and imaginary parts are rounded separately. And to avoid ambiguity, half integers are rounded to even integers. Note that this choice of  $n$  given by (4) also minimizes the norm of  $\mathbf{b}'_2$ .

The resulting correlation after replacing  $\mathbf{b}_2$  with  $\mathbf{b}'_2$  is

$$\begin{aligned} \langle \mathbf{b}_1, \mathbf{b}'_2 \rangle &= \left\langle \mathbf{b}_1, \left( \mathbf{b}_2 - \left\lfloor \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|^2} \right\rfloor \mathbf{b}_1 \right) \right\rangle \\ &= \left( \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|^2} - \left\lfloor \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|^2} \right\rfloor \right) \cdot \|\mathbf{b}_1\|^2. \end{aligned} \quad (5)$$

Since the rounding errors for real and imaginary parts are each no more than  $1/2$ , we have

$$|\operatorname{Re}\{\langle \mathbf{b}_1, \mathbf{b}'_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2 \text{ and } |\operatorname{Im}\{\langle \mathbf{b}_1, \mathbf{b}'_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2.$$

After replacing  $\mathbf{b}_2$  with the optimal  $\mathbf{b}'_2$ , if this new  $\mathbf{b}_2$  is shorter than  $\mathbf{b}_1$ , we swap them and then check whether further subtraction is possible.

Summarizing, the algorithm is as follows:

- 1) Check the correlation. If  $|\operatorname{Re}\{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$  and  $|\operatorname{Im}\{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$ , stop. Otherwise, replace  $\mathbf{b}_2$  with  $\mathbf{b}_2 - \left\lfloor \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle}{\|\mathbf{b}_1\|^2} \right\rfloor \mathbf{b}_1$  and go to step 2.
- 2) Check their lengths. If  $\|\mathbf{b}_2\| > \|\mathbf{b}_1\|$ , stop. Otherwise, swap them and go to step 1.

When this iterative procedure stops, the resulting basis will have the properties  $\|\mathbf{b}_1\| \leq \|\mathbf{b}_2\|$ ,  $|\operatorname{Re}\{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$  and  $|\operatorname{Im}\{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle\}| \leq \frac{1}{2}\|\mathbf{b}_1\|^2$ . It follows that basis vectors with these properties are the ones we desire, as we show next.

*Proposition 1:* Given a two dimensional lattice with basis vectors  $\mathbf{u}$  and  $\mathbf{v}$ . If  $\|\mathbf{u}\| \leq \|\mathbf{v}\|$ ,  $|\operatorname{Re}\{\langle \mathbf{u}, \mathbf{v} \rangle\}| \leq \frac{1}{2}\|\mathbf{u}\|^2$ , and  $|\operatorname{Im}\{\langle \mathbf{u}, \mathbf{v} \rangle\}| \leq \frac{1}{2}\|\mathbf{u}\|^2$ , then

- 1)  $\mathbf{u}$  is the shortest (non-zero) vector in the lattice.
- 2)  $\mathbf{v}$  is the shortest vector that is not a multiple of  $\mathbf{u}$ .

*Proof:* 1) Since  $(\mathbf{u}, \mathbf{v})$  is a lattice basis, any vector  $\mathbf{s}$  in the lattice can be written as  $\mathbf{s} = a\mathbf{u} + b\mathbf{v}$ , with  $a, b \in \mathbb{Z} + \mathbb{Z}j$ .

$$\begin{aligned} \|\mathbf{s}\|^2 &= \|a\mathbf{u} + b\mathbf{v}\|^2 \\ &= |a|^2\|\mathbf{u}\|^2 + |b|^2\|\mathbf{v}\|^2 + 2\operatorname{Re}\{a^\dagger b\langle \mathbf{u}, \mathbf{v} \rangle\} \\ &\geq (a_r^2 + a_i^2 + b_r^2 + b_i^2 - |a_r b_r + a_i b_i| - |a_i b_r - a_r b_i|)\|\mathbf{u}\|^2 \\ &\geq \|\mathbf{u}\|^2 \quad \text{when } a_r, a_i, b_r, b_i \text{ are not all 0,} \end{aligned} \quad (6)$$

where  $a_r = \operatorname{Re}\{a\}$ ,  $a_i = \operatorname{Im}\{a\}$ ,  $b_r = \operatorname{Re}\{b\}$ ,  $b_i = \operatorname{Im}\{b\}$ . The last step uses the identities, for any real numbers  $a, b, c, d$ ,

$$\begin{aligned} a^2 + b^2 + c^2 + d^2 &\geq |ac| + |bd| + |bc| + |ad| \\ |ac| + |bd| &\geq |ac + bd| \text{ and } |bc| + |ad| \geq |bc - ad| \end{aligned}$$

The necessary and sufficient conditions for the equalities to hold are  $|a|=|b|=|c|=|d|$ ,  $abcd \geq 0$  and  $abcd \leq 0$ , respectively.

2) Any vector  $\mathbf{s}$  in the lattice that is not a multiple of  $\mathbf{u}$  can be written as  $\mathbf{s} = a\mathbf{u} + b\mathbf{v}$ ,  $a, b \in \mathbb{Z} + \mathbb{Z}j$ , and  $b \neq 0$ .

$$\begin{aligned} \|\mathbf{s}\|^2 &= \|a\mathbf{u} + b\mathbf{v}\|^2 \\ &= |b|^2 (\|\mathbf{v}\|^2 - \|\mathbf{u}\|^2) \\ &\quad + \underbrace{|a|^2\|\mathbf{u}\|^2 + |b|^2\|\mathbf{u}\|^2 + 2\operatorname{Re}\{a^\dagger b\langle \mathbf{u}, \mathbf{v} \rangle\}} \\ &\geq |b|^2 (\|\mathbf{v}\|^2 - \|\mathbf{u}\|^2) + \|\mathbf{u}\|^2 + (\|\mathbf{v}\|^2 - \|\mathbf{v}\|^2) \\ &= (|b|^2 - 1) \cdot (\|\mathbf{v}\|^2 - \|\mathbf{u}\|^2) + \|\mathbf{v}\|^2 \\ &\geq \|\mathbf{v}\|^2 \quad \text{because } b \neq 0 \end{aligned} \quad (7)$$

It is clear that the procedure will end. In particular, after each iteration, the lengths of both basis vectors decrease (at least one decreases strictly); otherwise, the procedure ends. Since lattices are discrete, there can only be finitely many vectors shorter than the original ones. Thus, the procedure must end. ■

In the remainder of the paper we examine the effects of using lattice reduction with traditional detectors. Let us use LR-ICD and LR-BLAST to refer to the detection schemes that combine lattice reduction with ICD and BLAST detection respectively.

## IV. GAUSSIAN CHANNELS

In this section we develop individual channel results, i.e., for a fixed channel matrix  $\mathbf{H}$ .

### A. Complexity

The incremental complexity inherent in the use of lattice reduction is determined by the number of iterations required to reduce the basis. The worst-case complexity is unbounded. In particular, we are able to construct infinite sequences of channel matrices (based on Fibonacci sequences) that take an increasing numbers of iterations to finish. However, the fraction of all possible  $2 \times 2$  channels that require more than one or two iterations is actually very small, so in practice the complexity increase is negligible.

### B. Performance

These new detection methods realize decision regions (and thus performance) much closer to that of MLD, as we now develop.

Fig. 3 shows a comparison of the decision regions for MLD and LR-ICD. It is drawn for a  $2 \times 2$  real example for illustration purpose. The MLD decision region is a hexagon, and that of LR-ICD is a parallelogram. These regions also coincide with what are referred to as the *Voronoi cell* and *unit cell* of the lattice, respectively.

The minimum distances  $d_{\min}$  from the received constellation point to the decision boundaries are drawn. Recall that  $d_{\min}$

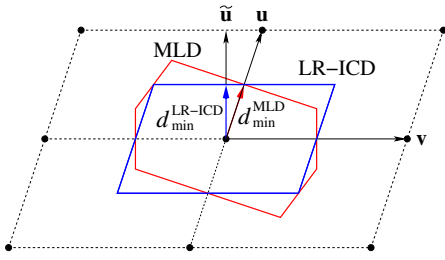


Fig. 3. Comparison of the decision regions for MLD and LR-ICD. Minimum distances to the decision boundaries are also compared.

is the minimum amount of noise needed for an error to occur, and parameterizes the error probability at high SNR in white Gaussian noise as  $2Q(d_{\min}/\sigma_w)$ , where  $\sigma_w^2$  is the noise variance per dimension and  $Q(x) = \int_x^\infty (1/\sqrt{2\pi}) \exp\{-x^2/2\}$ . We see that for LR-ICD,  $d_{\min}$  is shorter, so the performance is worse. This is a result of the basis vectors not being orthogonal. We now develop a precise bound on the ratio of  $d_{\min}^{\text{MLD}}$  to  $d_{\min}^{\text{LR-ICD}}$  to quantify the worst SNR gap to the MLD bound.

Generalizing Fig. 3 to the complex case, we see that

$$d_{\min}^{\text{MLD}} = \frac{1}{2}\|\mathbf{u}\| \text{ and } d_{\min}^{\text{LR-ICD}} = \frac{1}{2}\|\tilde{\mathbf{u}}\|.$$

where

$$\begin{aligned} \|\tilde{\mathbf{u}}\|^2 &= \|\mathbf{u}\|^2 - \left\| \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2} \mathbf{v} \right\|^2 \\ &= \|\mathbf{u}\|^2 - \frac{\text{Re}\{\langle \mathbf{u}, \mathbf{v} \rangle\}^2}{\|\mathbf{v}\|^2} - \frac{\text{Im}\{\langle \mathbf{u}, \mathbf{v} \rangle\}^2}{\|\mathbf{v}\|^2} \\ &\geq \|\mathbf{u}\|^2 \left( 1 - \frac{1}{4} - \frac{1}{4} \right) = \frac{1}{2}\|\mathbf{u}\|^2. \end{aligned} \quad (8)$$

Therefore,  $d_{\min}^{\text{LR-ICD}} \geq (1/\sqrt{2})d_{\min}^{\text{MLD}}$ , which corresponds to a maximum SNR loss of 3dB. Note that this bound is tight; the worst case is achieved by, for example,  $\mathbf{u} = [1 \ 0]^T$ , and  $\mathbf{v} = (\frac{1}{2} + \frac{1}{2}j) [1 \ 1]^T$ . However, for many channel matrices the ratio is much closer to one.

For LR-BLAST,  $d_{\min}^{\text{LR-BLAST}} = \frac{1}{2} \min(\|\mathbf{u}\|, \|\tilde{\mathbf{v}}\|) \geq \frac{1}{2}\|\tilde{\mathbf{u}}\|$ , so it always performs at least as well as LR-ICD. Comparing to MLD,  $d_{\min}^{\text{LR-BLAST}} = d_{\min}^{\text{MLD}}$ , when  $\|\tilde{\mathbf{v}}\| \geq \|\mathbf{u}\|$ , which happens quite often in the  $2 \times 2$  case. However, the worst case is still  $d_{\min}^{\text{LR-BLAST}} = d_{\min}^{\text{LR-ICD}} = (1/\sqrt{2})d_{\min}^{\text{MLD}}$ .

In summary, LR can improve the performance of detection to within 3dB from optimal in terms of  $d_{\min}$ . The actual gap depends on how well the particular channel can be reduced.

Another property of lattice reduction is that it monotonically improves detection performance. In particular, for both LR-ICD and LR-BLAST, each iteration of the reduction algorithm improves the decision region and increases  $d_{\min}$ . The more correlated the original basis vectors are, the greater the ultimate improvement. This behavior is illustrated by the following example channel matrices

$$\mathbf{H}_1 = \begin{bmatrix} 6 & 7 \\ 8 & -9 \end{bmatrix} \text{ and } \mathbf{H}_2 = \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix}$$

whose resulting SNR gaps are listed in Table I. We see that little improvement is obtained for  $\mathbf{H}_1$ , which has nearly orthogonal

columns, while a large improvement in dB is obtained for  $\mathbf{H}_2$ , which has highly correlated columns.

TABLE I  
SNR GAPS TO MLD PERFORMANCE FOR VARIOUS DETECTORS

	ICD	BLAST	LR-ICD	LR-BLAST
$\mathbf{H}_1$	0.31 dB	0.00 dB	0.31 dB	0.00 dB
$\mathbf{H}_2$	18.1 dB	17.0 dB	0.00 dB	0.00 dB

## V. RAYLEIGH FADING CHANNELS

In this section we develop results for ensembles of channels, i.e., for a random channel matrix  $\mathbf{H}$ . We focus on the Rayleigh fading case in which the entries of  $\mathbf{H}$  are independent  $\mathcal{CN}(0, 1)$  random variables, independent of the Gaussian noise.

### A. Complexity

Since the incremental complexity is dependent on the realized channel, we plot in Fig. 4 on both linear and logarithmic scales the empirical distribution of the number of iterations needed in the Rayleigh fading environment. Note that over 99% of the bases are reduced in two iterations or less, and that it becomes increasingly unlikely to need more iterations.

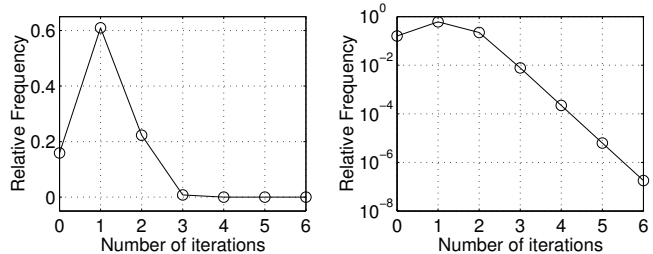


Fig. 4. Distribution of number of iterations needed for  $2 \times 2$  lattice reduction.

### B. Performance

In Rayleigh fading, the the average error probability  $P_e$  decays according to  $P_e \sim 1/\text{SNR}^\nu$  at high SNR, where  $\nu$  is termed the *order of diversity* in the system. The diversity order is a reflection of the systems tolerance of and robustness to the channel being near singular (i.e., in a deep fade).

In the  $2 \times 2$  case, lattice reduction improves the diversity  $\nu$  achieved by ICD and BLAST detection to that of MLD. To see this, the average symbol error rate (SER) curves for the various detection methods are plotted in Fig. 5 for 16-QAM. The top two curves are for ICD and BLAST detection. In the high SNR regime, they both have diversity 1. Note that for BLAST, if there were no error propagation, the diversity for the entry detected second would have been 2. However, its actual diversity is only 1 due to error propagation from the entry detected first, which itself experiences only diversity 1.

The lowest curve is for MLD. The two curves immediately above it and parallel to it correspond to LR-ICD and LR-BLAST. In the high SNR regime, all three evidently have diversity two. This shows the improvement in diversity provided

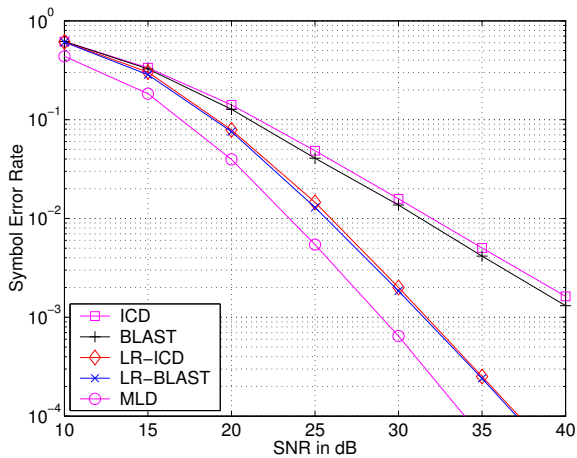


Fig. 5. Symbol error rate curves for various detection methods in the  $2 \times 2$  complex case. The constellation used is 16-QAM.

by using lattice reduction. Notice, with lattice reduction, the relative benefits of BLAST detection over ICD is smaller; this is a result of the basis vectors becoming more orthogonal.

It is also insightful to examine the empirical cumulative density of  $d_{\min}^2$  for these detectors, which is depicted in Fig. 6. Relative to the original ICD and BLAST detection (dashed curves), it is evident that with lattice reduction (solid curves), the probability of having small  $d_{\min}^2$  is substantially reduced. Furthermore, comparing the LR curves to the MLD curve (dotted), we see that the performance gap is much less than the worst case 3dB SNR loss. This is because channels yielding these larger losses are rare.

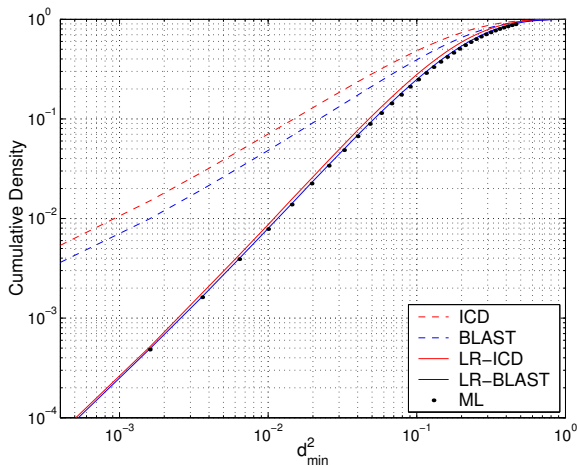


Fig. 6. Comparisons of the cumulative density of  $d_{\min}^2$ .

Note that Fig. 6 reflects the diversity behavior seen in Fig. 5 from a different angle. Indeed, the SER is related to a kind of “outage” probability, the probability of  $d_{\min}^2$  being less than a threshold which is inversely related to SNR.

One feature in Fig. 5 that is not captured by Fig. 6 is the gap between the LR curves and the MLD curve. This is because the detection performance is also affected by the number of nearest neighbors and, indirectly, the size of the constellation. In a finite constellation, some points have fewer nearest neighbors,

for example, the edge points. In some extreme cases, it is even possible for a point to have all its nearest neighbors distance  $d_{\min}$  away to be outside the constellation, resulting in the effective  $d_{\min}$  to be actually greater. For these reasons, LR based detection, which treats the constellation as an infinite lattice, is slightly further sub-optimal compared to MLD, which takes advantage of the finite size of the constellation. However, as the constellation gets larger, these differences diminishes. This finite constellation effect is apparent from extending the constellation to 64-QAM and 256-QAM from the original 16-QAM, at the 25dB SNR level. The corresponding SER curves for MLD are plotted in Fig. 7 together with the corresponding SER of LR-BLAST. We can see that as the constellation gets larger, the gap between MLD and LR-BLAST becomes smaller.

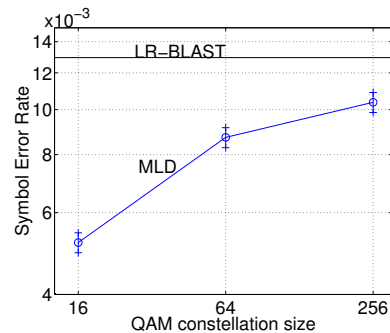


Fig. 7. The gap between SER of MLD and LR-BLAST diminishes with increasing constellation size. The noise level is such that SNR is 25dB for the 16-QAM constellation.

## VI. SUMMARY AND FUTURE WORK

In this paper, we proposed new coherent detection methods for MIMO communication systems. These methods significantly improve the performance of traditionally employed low-complexity detectors, in particular, ICD and BLAST detection, by incorporating lattice reduction. We studied the small ( $2 \times 2$ ), uncoded, coherent systems in detail. We presented an iterative lattice reduction algorithm for optimal decoding and studied its complexity. We showed that the number of iterations needed is typically low and it is increasingly unlikely to need more. We also showed that, relative to optimal MLD, LR techniques is sub-optimal by no more than 3dB in terms of SNR for any Gaussian channel, and allows us to achieve the same diversity on the Rayleigh fading channel, assuming sufficiently large constellations are used.

The proofs and simulations in this paper are limited to the  $2 \times 2$  case. However, for higher dimensional cases, we believe lattice reduction could be very useful as well. Work in this direction is in progress.

## REFERENCES

- [1] G. J. Foschini, *Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multiple Antennas*. Bell Laboratories Technical Journal, Vol. 1, No. 2, Autumn, 1996, pp. 41-59.
- [2] L. Lovasz, *An Algorithmic Theory of Numbers, Graphs, and Convexity*. Philadelphia, PA : Society for Industrial and Applied Mathematics, 1986.
- [3] J. G. Proakis, *Digital Communications, 4th ed.*, New York, NY : McGraw-Hill, 2000.