Approaching MIMO Channel Capacity with Reduced-Complexity Soft Sphere Decoding

Renqiu Wang and *Georgios B. Giannakis (contact author)*

Abstract — Hard Sphere-Decoding (SD) has well appreciated merits for near-optimal demodulation of multiuser, block single-antenna, or, multi-antenna transmissions over multi-input multi-output (MIMO) channels. At increased complexity, a soft version of SD, so termed list SD (LSD), has been recently applied to coded layered space-time (LST) systems enabling them to approach MIMO channel capacity. By introducing a novel bit-level multi-stream LST transmitter along with a soft-to-hard decoder conversion, we show how to achieve the near-capacity performance of LSD at reduced complexity, and even outperform it as the size of the block to be decoded (M) increases. Specifically, for binary real LST codes we develop exact max-log based SD schemes with average complexity $O(M^4)$, and various approximate alternatives trading-off performance for average complexity down to $O(M^3)$. These schemes apply directly to the real and imaginary parts of QPSK signalling, and also to QAM signalling after incorporating an appropriate interference estimation and cancellation module. We corroborate our reduced-complexity near-optimal soft SD algorithms with simulations.

Keywords: soft sphere decoding, iterative decoding, space-time, MIMO channel capacity

I. INTRODUCTION

In wireless communications, quite often we wish to estimate the $M \times 1$ information bearing symbol vector $\mathbf{s}$ from the $N \times 1$ data vector $\mathbf{y}$ in the block coding model

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{s}$ has entries belonging to a finite alphabet $\mathcal{S}$, $\mathbf{H}$ is a known $N \times M$ real or complex matrix, and $\mathbf{n}$ is a $N \times 1$ Gaussian noise vector. This problem is encountered in many applications including single-antenna block transmissions, space-time (ST) multi-antenna transmissions, or, in multi-user detection of CDMA transmissions. When $\mathbf{s}$ is drawn from QPSK or rectangular QAM constellations, (1) can be easily transformed to a real model where $\mathbf{s}$, $\mathbf{y}$, $\mathbf{H}$ and $\mathbf{n}$ all belong to the real field. The transformation will be detailed in Section III. Since in this paper we only focus on QPSK and rectangular QAM signalling, we assume that (1) is a real model.

The authors are with the Dept. of ECE, Univ. of Minnesota, Minneapolis, MN 55455, USA; Tel/fax: (612)626-7781/625-4583; Emails: {renqiu, georgios}@ece.umn.edu

When $\mathbf{s}$ is *uncoded* and $\mathbf{n}$ is white Gaussian, the optimal solution of (1) in the sense of minimizing the symbol error rate (SER) is offered by the maximum likelihood (ML) decoder which solves the integer least-squares (ILS) problem:

$$\hat{\mathbf{s}}_{ml} = \arg\min_{\mathbf{s} \in \mathcal{S}^M} ||\mathbf{y} - \mathbf{H}\mathbf{s}||^2. \tag{2}$$

Since the optimal block symbol decoding via exhaustive search in (2) incurs complexity $O(|\mathcal{S}|^M)$ that is exponential in M, a number of alternative algorithms have been developed to achieve near-optimal performance with polynomial complexity. Those include sphere decoding (SD) [1–3], semi-definite-programming (SDP) [4], and the Probabilistic Data Association (PDA) [5].

However, when $\mathbf{s}$ in (1) is *coded* with some kind of error control code (ECC), soft iterative (a.k.a. turbo) detection is capable of approaching the ultimate performance limit dictated by the capacity of the channel $\mathbf{H}$ using maximum *a posteriori* (MAP) decoding. Such an approach has been followed in [6], where a soft, so-called List Sphere Decoding (LSD) algorithm, has been derived to compute the extrinsic information based on a list of candidates obtained inside a preset sphere. Although the LSD scheme can approach the multi-input multi-output (MIMO) capacity of layered space-time (LST) systems, it comes with the limitation that the radius of the sphere can not be reduced during the search. Another weak point is that LSD is not applicable when all candidates in the list have only value $+1$ (or -1) for a certain bit.

In this paper, we first derive a soft-to-hard transformation for binary constellations to convert the max-log based MAP (max-MAP) decoding problem of real block codes to a set of hard SD problems. In addition to providing an *exact* max-MAP decoder, we also derive approximate alternatives to further reduce complexity to the order of a single hard SD. Applying our soft-to-hard decoding schemes to a *single-stream* coded LST system for QPSK signalling, and to a *bit-level multi-stream* coded LST system for QAM signalling, we demonstrate by simulations that MIMO channel capacity can be approached at reduced complexity.

II. SOFT DECODING, COMPLEXITY AND TRADE-OFFS

Suppose that each entry s_m of the symbol vector $\mathbf{s}$ in (1) is obtained by mapping a $M_c \times 1$ binary vector $\mathbf{x}^{(m)}$ with ± 1 entries, and let $\mathbf{x} := [\mathbf{x}^{(1)^T}, \mathbf{x}^{(2)^T}, \ldots, \mathbf{x}^{(M)^T}]^T$, where ^{T} denotes transposition. The MAP decoder for obtaining $\mathbf{x}$ from $\mathbf{y}$ minimizes the bit error rate (BER) by evaluating the log-likelihood ratio (LLR) of the *a posteriori* probability of each bit x_k: $\lambda_D(x_k) := \ln[P(x_k = +1|\mathbf{y})/P(x_k = -1|\mathbf{y})]$. With $\lambda_A(x_k) := \ln[P(x_k = +1)/P(x_k = -1)]$ denoting the *a priori* information of x_k, define $\boldsymbol{\lambda}_A := [\lambda_A(x_1), \ldots, \lambda_A(x_{MM_c})]^T$, and let $\lambda_E(x_k|\mathbf{y}) := \lambda_D(x_k|\mathbf{y}) - \lambda_A(x_k)$ stand for the extrinsic information of x_k. Bayes' theorem and the assumption that the entries of $\mathbf{x}$ are independent yield:

$$\lambda_D(x_k|\mathbf{y}) = \lambda_A(x_k) + \ln \frac{\sum\limits_{\mathbf{x}\in\mathbb{X}_{k,+1}} P(\mathbf{y}|\mathbf{x})\exp\{\frac{1}{2}\mathbf{x}_{[k]}^T\boldsymbol{\lambda}_{A,[k]}\}}{\sum\limits_{\mathbf{x}\in\mathbb{X}_{k,-1}} P(\mathbf{y}|\mathbf{x})\exp\{\frac{1}{2}\mathbf{x}_{[k]}^T\boldsymbol{\lambda}_{A,[k]}\}},$$

where $\mathbb{X}_{k,+1} := \{\mathbf{x}|x_k = +1\}$, $\mathbb{X}_{k,-1} := \{\mathbf{x}|x_k = -1\}$, $\mathbf{x}_{[k]}$ is the sub-vector of $\mathbf{x}$ obtained by omitting its kth element x_k, and likewise $\boldsymbol{\lambda}_{A,[k]}$ is obtained from $\boldsymbol{\lambda}_A$ by omitting its kth element $\lambda_A(x_k)$.

Since $\mathbf{n}$ is white Gaussian, using the max-log approximation [7], we can approximate the extrinsic information of x_k as [6]

$$\begin{aligned}\lambda_E(x_k|\mathbf{y}) \approx &(1/2)\max_{x\in\mathbb{X}_{k,+1}}\{-\frac{1}{\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{s}||^2 + \mathbf{x}_{[k]}^T\boldsymbol{\lambda}_{A,[k]}\}\\ &-(1/2)\max_{x\in\mathbb{X}_{k,-1}}\{-\frac{1}{\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{s}||^2 + \mathbf{x}_{[k]}^T\boldsymbol{\lambda}_{A,[k]}\}.\end{aligned} \tag{3}$$

Eq. (3) shows that carrying out the MAP soft decoding of $\mathbf{x}$ from $\mathbf{y}$ reduces to solving two integer least-squares problems with linear constraints for each $k \in [1, M]$.

A. Soft-to-hard conversion - Scheme 1

Our reduced complexity MAP decoder starts by observing that (3) can be rewritten as

$$\begin{aligned}\lambda_E(x_k|\mathbf{y}) =&(1/2)\max_{\mathbf{x}\in\mathbb{X}_{k,+1}}\{-\frac{1}{\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{s}||^2 + \mathbf{x}^T\boldsymbol{\lambda}_A\}\\ &-(1/2)\max_{\mathbf{x}\in\mathbb{X}_{k,-1}}\{-\frac{1}{\sigma^2}||\mathbf{y}-\mathbf{H}\mathbf{s}||^2 + \mathbf{x}^T\boldsymbol{\lambda}_A\}\\ &-\lambda_A(x_k).\end{aligned} \tag{4}$$

Since in randomly faded ST channels the matrix $\mathbf{H}$ has full column rank almost surely, it follows that in the binary case ($\mathbf{s} = \mathbf{x}$), we can always find a vector $\tilde{\mathbf{y}}$ satisfying

$$2\mathbf{H}^T\tilde{\mathbf{y}} = \sigma^2\boldsymbol{\lambda}_A. \tag{5}$$

Our key observation is that using (5) we can rewrite (4) as

$$\begin{aligned}\lambda_E(x_k|\mathbf{y}) =& -\frac{1}{2\sigma^2}\min_{\mathbf{x}\in\mathbb{X}_{k,+1}}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\mathbf{x}||^2\\ &+\frac{1}{2\sigma^2}\min_{\mathbf{x}\in\mathbb{X}_{k,-1}}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\mathbf{x}||^2\\ &-\lambda_A(x_k).\end{aligned} \tag{6}$$

If $\hat{\mathbf{x}}_{map} = \arg\min_{\mathbf{x}\in\mathbb{X}}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\mathbf{x}||^2$, then (6) can be further simplified as

$$\begin{aligned}\lambda_E(x_k|\mathbf{y}) =& -\frac{\hat{x}_{k,map}}{2\sigma^2}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\hat{\mathbf{x}}_{map}||^2\\ &+\frac{\hat{x}_{k,map}}{2\sigma^2}\min_{\mathbf{x}\in\mathbb{X}_{k,-\hat{x}_{k,map}}}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\mathbf{x}||^2\\ &-\lambda_A(x_k).\end{aligned} \tag{7}$$

Letting $\hat{\mathbf{x}}_k$ denote the "best vector" for which $x_k = -\hat{x}_{k,map}$, it follows that

$$\hat{\mathbf{x}}_k = \arg\min_{\mathbf{x}\in\mathbb{X}_{k,-\hat{x}_{k,map}}}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\mathbf{x}||^2, \quad k = 1, \ldots, M.$$

We can use hard SD to find $\hat{\mathbf{x}}_{map}$ and $\{\hat{\mathbf{x}}_k\}_{k=1}^M$. Thus, to obtain the extrinsic information for the entire vector $\mathbf{x}$ we need one hard SD step with block size M to find $\hat{\mathbf{x}}_{map}$, and M hard SD steps with block size $M-1$ to find $\{\hat{\mathbf{x}}_k\}_{k=1}^M$. Notice that the extrinsic LLR values we obtained are exact under the max-log approximation. During each hard SD step, the initial radius can be chosen effectively as in [2,3]. And once a point is found inside the sphere, the radius can be reduced to the current point. The complexity of soft SD is thus reduced considerably by radius initialization and reduction. Since the average complexity of hard SD at high SNR with block size M is $O(M^3)$ [3], the average complexity of our soft-to-hard SD approach is $O(M^3)+MO((M-1)^3) \approx O(M^4)$.

With our exact max-log based MAP decoder as a starting point, we can further reduce the average complexity of soft SD down to $O(M^3)$ after a few approximations that we detail in the ensuing subsections.

B. Approximation A

Let $\check{\mathbf{x}}_k$ be the vector with identical entries as $\hat{\mathbf{x}}_{map}$ except for the kth element that is sign reversed: $\check{x}_k = -\hat{x}_{k,map}$. Clearly, the probability that $\hat{\mathbf{x}}_k$ equals $\check{\mathbf{x}}_k$ increases as the SNR increases. This motivates us to approximate the extrinsic information of x_k as

$$\begin{aligned}\lambda_E(x_k|\mathbf{y}) \approx& -\frac{\hat{x}_{k,map}}{2\sigma^2}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\hat{\mathbf{x}}_{map}||^2\\ &+\frac{\hat{x}_{k,map}}{2\sigma^2}||\mathbf{y}+\tilde{\mathbf{y}}-\mathbf{H}\check{\mathbf{x}}_k||^2 - \lambda_A(x_k).\end{aligned}$$

Thus, to obtain the extrinsic information for the entire vector $\mathbf{x}$ we need only one hard SD step with block size M to find $\hat{\mathbf{x}}_{map}$. The overall average decoding complexity is now $O(M^3 + M) \approx O(M^3)$.

C. Approximation B

Let $\check{\mathcal{X}}_k^{[2]}$ denote the set of vectors that have one more entry different from $\hat{\mathbf{x}}_{map}$ besides x_k; i.e.,

$$\check{\mathcal{X}}_k^{[2]} := \{\mathbf{x} | x_k = -\hat{x}_{k,map}, \frac{1}{2}\sum_{i=1, i\neq k}^{M} |x_i - \hat{x}_{i,map}| = 1\}.$$

The set $\check{\mathcal{X}}_k^{[2]}$ has M elements. Since at high SNR, $\hat{\mathbf{x}}_k \in \check{\mathcal{X}}_k^{[2]}$ with high probability, we can improve Approximation A as follows:

$$\lambda_E(x_k|\mathbf{y}) \doteq -\frac{\hat{x}_{k,map}}{2\sigma^2}||\mathbf{y} + \tilde{\mathbf{y}} - \mathbf{H}\hat{\mathbf{x}}_{map}||^2 + \frac{\hat{x}_{k,map}}{2\sigma^2}\min_{\mathbf{x}\in\check{\mathcal{X}}_k^{[2]}}||\mathbf{y} + \tilde{\mathbf{y}} - \mathbf{H}\mathbf{x}||^2 - \lambda_A(x_k).$$

To obtain the extrinsic information for the entire vector $\mathbf{x}$ here, we need only one hard SD step with block size M to find $\hat{\mathbf{x}}_{map}$, and M searching steps for the "best vectors" minimizing $||\mathbf{y} + \tilde{\mathbf{y}} - \mathbf{H}\mathbf{x}||^2$ respectively, in the sets $\{\check{\mathcal{X}}_k^{[2]}\}_{k=1}^M$. The overall average decoding complexity turns out to be $O(M^3 + M^2) \approx O(M^3)$.

D. Approximation C

If higher accuracy is desired, then we can approximate each $\hat{\mathbf{x}}_k$ with the "best vector" minimizing $||\mathbf{y} + \tilde{\mathbf{y}} - \mathbf{H}\mathbf{x}||^2$ in the set $\check{\mathcal{X}}_k^{[3]} := \{\mathbf{x}|x_k = -\hat{x}_{k,map}, \frac{1}{2}\sum_{i=1,i\neq k}^M |x_i - \hat{x}_{i,map}| = 1 \text{ or } 2\}$. In this case the overall average decoding complexity will be $O(M^3 + M(M + (M-1)(M-2))) \approx 2O(M^3)$.

III. Approaching MIMO capacity

In this section we will show how to apply our reduced complexity soft SD schemes to a coded LST system with the goal of approaching MIMO channel capacity. We will first discuss the model for QPSK signalling in subsection III-A, and then generalize it to a bit-level multi-stream coded LST model applicable to QAM signalling in subsection III-B. At the receiver end of both systems, the complex block model is always converted to its real counterpart. Let N_t denote the number of transmit antennas, N_r the number of receive antennas, $\mathbf{s}_0$ the $N_t \times 1$ transmitted vector, $\mathbf{y}_0$ the $N_r \times 1$ received vector, $\mathbf{H}_0$ the $N_r \times N_t$ MIMO channel coefficient matrix, and $\mathbf{n}_0$ the $N_r \times 1$ white Gaussian noise vector. By separating the real and imaginary parts of each vector and matrix, we can transform the complex model $\mathbf{y}_0 = \mathbf{H}_0\mathbf{s}_0 + \mathbf{n}_0$ to the real model:

$$\mathbf{y} = \begin{bmatrix}\mathbf{y}_{0,r}\\ \mathbf{y}_{0,i}\end{bmatrix} = \begin{bmatrix}\mathbf{H}_{0,r} & -\mathbf{H}_{0,i}\\ \mathbf{H}_{0,i} & \mathbf{H}_{0,r}\end{bmatrix}\begin{bmatrix}\mathbf{s}_{0,r}\\ \mathbf{s}_{0,i}\end{bmatrix} + \begin{bmatrix}\mathbf{n}_{0,r}\\ \mathbf{n}_{0,i}\end{bmatrix} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{8}$$

where subscripts r and i denote the real and imaginary part. The block size is now $M = 2N_t$. To ensure that $\mathbf{H}$ has full column rank, we clearly need $N_r \geq N_t$. Here, we take $N_r = N_t$ for brevity.

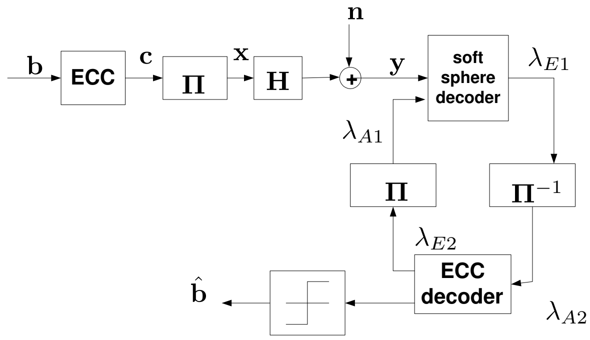


Fig. 1. Real equivalent system model for QPSK signalling

A. QPSK signalling

The real equivalent LST system model for QPSK signalling is depicted in Figure 1. The information bits $\mathbf{b}$ are first encoded by an ECC module to yield $\mathbf{c}$, and then go through a random interleaver Π. Interleaved bits $\mathbf{x}$ are mapped to QPSK symbols. QPSK symbol vectors $\mathbf{s}_0$ are transmitted using the parallel LST scheme known as V-BLAST [8]. At the receiver end, soft iterations between the MIMO channel decoding module and the ECC decoding module are used. In the MIMO channel decoding module, a complex-to-real conversion is performed as in (8), and $\mathbf{s}$ is subsequently replaced by the binary vector $\mathbf{x}$. Our soft-to-hard SD scheme is then applied to compute the extrinsic information of $\mathbf{x}$. The extrinsic information is exchanged between two decoding modules through the interleaver/deinterleaver modules denoted as Π/Π^{-1} in Fig. 1.

One advantage of applying our soft SD scheme in the iterative decoding process is that the extrinsic information from the ECC decoding module at the last iteration can be used to generate initial estimates of $\hat{\mathbf{x}}_{k,map}$ and $\{\hat{\mathbf{x}}_k\}_{k=1}^M$. These initial estimates will provide tight initial radii for the hard SD steps. The complexity of each hard SD is thus further reduced.

B. QAM signalling

The soft-to-hard SD schemes we developed in Section II are only applicable when $\mathbf{s} = \mathbf{x}$ in (3). For QAM

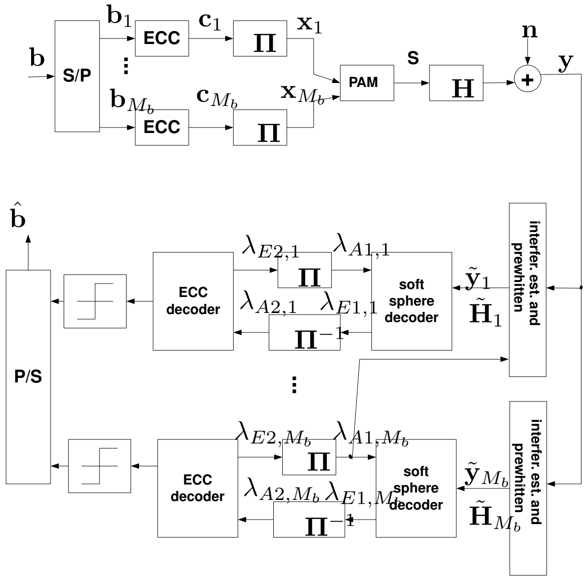


Fig. 2. Real equivalent system model for QAM signalling

signalling, after the complex-to-real conversion in (8), $\mathbf{s}$ is a $2N_t \times 1$ vector composed of PAM symbols. If M_b denotes the number of bit levels in a PAM symbol, then s_k can be expressed as a linear combination of M_b bits $\{x_{i,k}\}_{i=1}^{M_b}$; i.e. $s_k = \sum_{i=1}^{M_b} 2^{i-1} x_{i,k}$. Hence, the PAM vector $\mathbf{s}$ can be expressed using a linear combination of M_b binary vectors $\{\mathbf{x}_i\}_{i=1}^{M_b}$ as: $\mathbf{s} = \sum_{i=1}^{M_b} 2^{i-1}\mathbf{x}_i$, where $\mathbf{x}_i$ is the binary vector in the ith bit level. Let $\mathbf{x}$ denote the entire binary vector $[\mathbf{x}_1^T, \ldots, \mathbf{x}_{M_b}^T]^T$, and let $\mathbf{H}_{eq}$ stand for the equivalent block coding matrix for $\mathbf{x}$ given by $\mathbf{H}_{eq} = [\mathbf{H}, 2\mathbf{H}, \ldots, 2^{M_b-1}\mathbf{H}]$. Eq. (1) can now be rewritten as $\mathbf{y} = \mathbf{H}_{eq}\mathbf{x} + \mathbf{n}$. However, we cannot find a vector $\tilde{\mathbf{y}}$ satisfying $2\mathbf{H}_{eq}^T\tilde{\mathbf{y}} = \sigma^2 \boldsymbol{\lambda}_A$, because $\mathbf{H}_{eq}$ is rank deficient. Therefore, the soft SD scheme we developed for the binary case cannot be applied directly to QAM signalling.

Because different $x_{i,k}$ bits in s_k will be received with generally different SNRs, we adopt a bit-level multi-stream coded LST transmission for QAM signalling as depicted in Figure 2. At the transmitter, the stream of information bits $\mathbf{b}$ is first divided into M_b substreams: $\{\mathbf{b}_i\}_{i=1}^{M_b}$. Each substream is coded with an ECC and scrambled through a random interleaver. From each substream, we take one interleaved bit to form a PAM symbol consisting of M_b bits. Two PAM symbols are further combined to form a QAM symbol. The QAM symbol vectors are then transmitted using V-BLAST. At the receiver end, the iterative decoding scheme is performed in a layered fashion per bit level. In the MIMO channel decoding module, the complex MIMO block model is first converted to the real block model as in (8). When decoding one bit level, the interference from other bit levels will be treated as Gaussian noise, and based on the *a priori* information the mean and covariance matrix of the equivalent noise will be estimated as in [5, 9]. This reduces the decoding of each bit level to an equivalent QPSK decoding problem in the presence of colored Gaussian noise. After prewhitening the noise, our soft SD scheme can be readily applied with the extrinsic information exchanged through the interleaver/deinterleaver between the ECC decoding module and the MIMO channel decoding module, as before.

We now discuss the decoding process in detail. When decoding the jth bit level, we rewrite (1) as

$$\begin{aligned}\mathbf{y} =& \mathbf{H}\mathbf{s} + \mathbf{n} \\ =& 2^{j-1}\mathbf{H}\mathbf{x}_j + \sum_{i=1, i\neq j}^{M_b} 2^{i-1}\mathbf{H}\mathbf{x}_i + \mathbf{n} \\ =& \mathbf{H}_j\mathbf{x}_j + \mathbf{n}_j, \end{aligned} \tag{9}$$

where $\mathbf{H}_j := 2^{j-1}\mathbf{H}$ is the equivalent block coding matrix for the jth bit level, and $\mathbf{n}_j = \sum_{i=1,i\neq j}^{M_b} 2^{i-1}\mathbf{H}\mathbf{x}_i + \mathbf{n}$ is the equivalent noise encompassing both interference, and the actual noise.

Let $\mathbf{p}_i$ denote the probability vector of the binary vector $\mathbf{x}_i$ with each entry $p_{i,k} := P(x_{i,k} = +1)$. Given $\boldsymbol{\lambda}_{A,i}$, the *a priori* information vector for $\mathbf{x}_i$, we obtain

$$\mathbf{p}_i = 1/2 + (1/2)\tanh(\boldsymbol{\lambda}_{A,i}/2).$$

Since the entries of $\{\mathbf{x}_i\}_{i=1}^{M_b}$ are assumed independent, we can estimate the mean and the covariance matrix of $\mathbf{n}_j$ using $\mathbf{p}_i$. However, if soft information from lower power bit levels is used to decode a certain bit level, we have observed that the decoding error probability becomes worse in our simulations. In fact, as iterations increase at high SNR, the number of errors actually blows up. In a 64-QAM test using turbo ECC with frame length 9000 in a 4×4 transmit/receive antenna setup, at SNR=13dB, the number of errors per iteration we counted in one frame was (0, 2554, 2310, 3398) for 4 iterations.

This "error-blow-up" problem disappears if we modify the estimation method in the following way: when decoding the jth bit level with transmit power $2^{2(j-1)}$, we use only the soft information $\mathbf{p}_i$ from the bit levels with higher power ($i = j+1, \ldots, M_b$), and we assume $p_{i,k} = 0.5$ ($\forall k, i = 1, \ldots, j-1$) for the bit levels with lower power, as if we had no *a priori* information about these levels. Specifically, we use

$$\bar{\mathbf{n}}_j := \mathrm{E}(\mathbf{n}_j) = \sum_{i=1, i\neq j}^{M_b} \mathbf{H}_i \bar{\mathbf{x}}_i = \sum_{i=j+1}^{M_b} \mathbf{H}_i (2\mathbf{p}_i - \mathbf{1}) \quad (10)$$

$$\begin{aligned}\mathbf{C}_j &:= \mathrm{Cov}(\mathbf{n}_j) = \mathrm{Cov}(\mathbf{n}) + \sum_{i=1, i\neq j}^{M_b} \mathbf{H}_i \mathrm{Cov}(\mathbf{x}_i) \mathbf{H}_i^T \\ &= \mathrm{diag}(\sigma^2 \mathbf{1}) + \sum_{i=j+1}^{M_b} \mathbf{H}_i \mathrm{diag}[4\mathbf{p}_i.(\mathbf{1} - \mathbf{p}_i)] \mathbf{H}_i^T \\ &\quad + \sum_{i=1}^{j-1} \mathbf{H}_i \mathrm{diag}[\mathbf{1}] \mathbf{H}_i^T, \end{aligned} \quad (11)$$

where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with a vector in parentheses as its diagonal, and $\mathbf{1}$ stands for the vector with all entries equal to 1.

Before applying our soft SD scheme to the jth bit level, by subtracting $\bar{\mathbf{n}}_j$ from both sides of (9) and left-multiplying them with $\mathbf{C}_j^{-1/2}$, we obtain the equivalent zero-mean white Gaussian noise model:

$$\tilde{\mathbf{y}}_j = \tilde{\mathbf{H}}_j \mathbf{x}_j + \tilde{\mathbf{n}}_j, \quad (12)$$

where $\tilde{\mathbf{y}}_j := \mathbf{C}_j^{-1/2}(\mathbf{y} - \bar{\mathbf{n}}_j)$, $\tilde{\mathbf{H}}_j := \mathbf{C}_j^{-1/2}\mathbf{H}_j$, and $\tilde{\mathbf{n}}_j := \mathbf{C}_j^{-1/2}(\mathbf{n}_j - \bar{\mathbf{n}}_j)$. With the noise block $\tilde{\mathbf{n}}_j$ being zero-mean white Gaussian with identity covariance matrix, our soft SD scheme can then be applied to compute the extrinsic information of $\mathbf{x}_j$.

Let the superscript $^{(t)}$ index time, T denote the number of received vectors in a coded frame, and subscripts $_1$ and $_2$ denote the index of the MIMO decoding module and the ECC decoding module, respectively. The steps of the iterative decoding process are as follows:

1) Initialization: $\mathbf{p}_j^{(t)} = 0.5\mathbf{1}$ and $\boldsymbol{\lambda}_{A,j}^{(t)} = \mathbf{0}$, $j = 1, \ldots, M_b, t = 1, \ldots, T$.
2) One iteration:
   a) $j = M_b$.
   b) In the MIMO channel decoding module:
      i) $t = 1$.
      ii) Convert the QAM model (1) to the QPSK model as (9), and estimate $\bar{\mathbf{n}}_j^{(t)}$ and $\mathbf{C}_j^{(t)}$ according to (10) and (11).
      iii) Prewhiten the noise $\mathbf{n}_j^{(t)}$ as in (12) and decode $\mathbf{x}_j^{(t)}$ with soft SD scheme and output extrinsic information vector $\boldsymbol{\lambda}_{E1,j}^{(t)}$.
      iv) $t = t + 1$; if $t \leq T$, return to (ii).
   c) Deinterleave $\boldsymbol{\lambda}_{E1,j}^{(t)}, t = 1, \ldots, T$ to obtain the *a priori* information $\boldsymbol{\lambda}_{A2,j}$ for the ECC decoding module.
   d) In the ECC decoding module: use a soft decoding scheme depending on the ECC used. Output the extrinsic information $\boldsymbol{\lambda}_{E2,j}$.
   e) Interleave $\boldsymbol{\lambda}_{E2,j}$ to obtain the *a priori* information $\boldsymbol{\lambda}_{A1,j}^{(t)}$ and $\mathbf{p}_j^{(t)}$, $t = 1, \ldots, T$ for the MIMO channel decoding module.
   f) $j = j - 1$; if $j \geq 1$, return to (b).
3) Return to (2) until a desired performance is achieved, or, the number of iterations reaches a certain number.

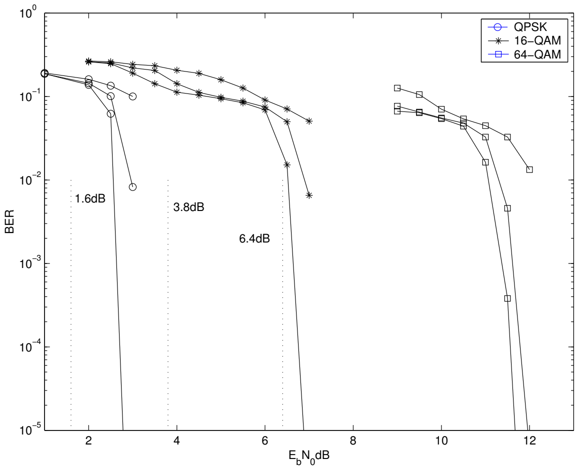
IV. SIMULATIONS

In this section, we present simulations using a parallel concatenated (turbo) ECC with rate $R = 1/2$, as in [6]. Each constituent convolutional code has memory 2, feedback polynomial $G_r(D) = 1 + D + D^2$, and feedforward polynomial $G(D) = 1 + D^2$. To maintain comparable decoding complexity with [6], for QPSK signalling, we choose the interleaver size to be 9000, and the number of inner iterations for the ECC decoding module to be 10. For 16-QAM signalling, we choose the interleaver size of each bit level to be 4500, and the number of inner iterations for the ECC decoding module of bit level 2 and bit level 1 to be 5 and 15, respectively. For 64-QAM signalling, we choose the interleaver size of each bit level to be 3000, and the number of inner iterations for the ECC decoding module of bit level 3, 2 and 1 to be 4, 8 and 18, respectively. As in [6], we generate independent Rayleigh flat fading channels between transmit/receive antennas and assume perfect channel estimation at the receiver end.

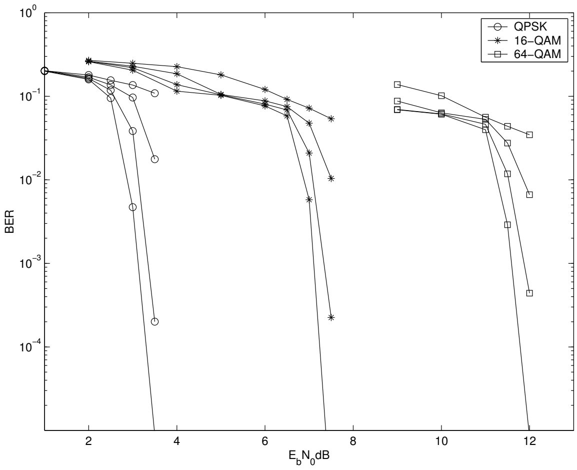
Simulation 1: Figure 3(a) depicts average BER performance in a 8×8 transmit/receive antennas setup when using our soft-to-hard SD Scheme 1. We also tested 2×2 and 4×4 configurations with results almost identical to 8×8 (they are omitted here due to space limitation). We performed 3 outer iterations between the ECC decoding module and the MIMO channel decoding module. Increasing the number of outer iterations further, did not improve performance. Let $(\mathrm{E_b/N_0})_{min}$ denote the SNR required to reach BER=10^{-5}. An interesting observation is that $(\mathrm{E_b/N_0})_{min}$ does not change with N_t for any constellation used in our system.

Compared with [6], our system achieves almost identical performance for QPSK signalling, at reduced complexity. For 16-QAM and 64-QAM, [6] performs better in 2×2 and 4×4 setups, but our scheme outperforms [6] in the 8×8 setup by about 0.5dB. Since our $(\mathrm{E_b/N_0})_{min}$ does not change with N_t, while $(\mathrm{E_b/N_0})_{min}$ increases as N_t increases in [6], we expect our gain to increase as N_t increases.

The reason that our system yields to the system in [6] for 16-QAM and 64-QAM signalling in the 2×2 and 4×4 configurations is because we use the same ECC for different bit levels even though there is 6dB difference



(a)



(b)

Fig. 3. (a) Soft-to-hard Scheme 1 in a 8×8 setup; (b) Approximate Scheme B in a 8×8 setup

between the powers of adjacent bit levels. As a result, the actual SNR for the lowest power level is much lower than other levels. For example, at average SNR=6dB in a 16-QAM system, the SNR for the lowest power level is only 2dB, even if perfect interference cancellation is assumed. However, for QPSK signalling, $(\mathrm{E_b}/\mathrm{N_0})_{min}$ is about 2.6dB. So, to have 2.6dB SNR at the lowest power level, the average SNR for 16-QAM is about 6.6dB and 11.2dB for 64-QAM, which matches well with the simulation results.

Simulation 2: Figure 3(b) depicts average BER with the same parameters as Simulation 1, except that we use our approximate scheme B in the MIMO channel decoding module. Although with this approximate scheme we have about 0.5 dB loss relative to scheme 1, we can still achieve the same performance as [6] for 16-QAM and 64-QAM. However, since the complexity of approximate scheme B is comparable with that of hard SD, our decoding complexity is much less than that of [6].

V. Conclusions

In this paper, we derived a near-optimal soft sphere decoding scheme for binary real block codes by converting soft decoding to a set of hard sphere decoding problems. In addition to an exact max-log based MAP decoder enjoying average complexity $O(M^4)$, we developed approximate alternatives to reduce average complexity down to $O(M^3)$, where M is the decoding block size. Applying our decoding schemes to a bit-level multi-stream coded LST system, we demonstrated that error performance can approach that dictated by MIMO channel capacity for both QPSK and rectangle QAM signalling. With 16-QAM and 64-QAM our soft SD scheme 1 outperforms the LSD system in [6] when the number of transmit/receive antennas is large. Even our approximate decoding scheme B having average complexity as low as $O(M^3)$ can achieve the same performance as LSD for 16-QAM and 64-QAM as the number of transmit/receive antennas increases.

Acknowledgment: This work was supported by the ARL/CTA Grant No. DAAD19-01-2-011.

References

[1] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[2] A. Chan and I. Lee, "A new reduced-complexity sphere decoder for multiple antenna systems," in *Proc. of Intl. Conf. on Communications*, New York City, N.Y., April 28-May 2, 2002, vol. 1, pp. 460–464.

[3] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *Proc. of 35th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2001, pp. 1051–1055.

[4] W.K. Ma, T. N. Davidson, K. M. Wong, Z.Q. Luo, and P. C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. on Signal Processing*, vol. 50, no. 4, pp. 912–922, Apr. 2002.

[5] J. Luo, K.R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Communications Letters*, vol. 5, no. 9, pp. 361–363, Sep. 2001.

[6] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. on Communications*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[7] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and suboptimal MAP decoding algorithms operating in the log domain," in *Proc. Intl. Conf. on Communications*, Seattle, Washington, June 1995, pp. 1009–1013.

[8] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[9] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques and Software*, Artech House, Dedham, MA, 1993.