TOWARDS CLOSING THE CAPACITY GAP ON MULTIPLE ANTENNA CHANNELS

Haris Vikalo

Stanford University
Stanford, CA 94309

Babak Hassibi

California Institute of Technology
Pasadena, CA 91106

ABSTRACT

In recent years, soft iterative decoding techniques have been shown to greatly improve the bit error rate performance of various communication systems. For multiple antenna systems, however, it is not clear what is the best way to obtain the soft-information required of the iterative scheme with low complexity. In this paper, we propose a modification of the Fincke-Pohst (sphere decoder) algorithm to estimate the MAP probability of the received symbol sequence. The new algorithm solves a nonlinear integer least-squares problem and, over a wide range of rates and SNRs, has polynomial-time (often cubic) complexity. The performance of the algorithm, combined with convolutional, turbo, and LDPC codes is demonstrated on several multiple antenna channels.

1. INTRODUCTION

Recently, the pursuit of high-speed wireless data services has generated a significant amount of activity in the communications research community. The physical limitations of the wireless medium present many challenges to the design of reliable communication systems. As has been shown in [1], multiple antenna wireless communication systems are capable of providing data transmission at potentially very high rates. In multiple antenna systems, space-time [2] (along with traditional error-correcting) codes are often employed at the transmitter to induce diversity. Furthermore, to secure high reliability of the data transmission, special attention has to be payed to the receiver design. However, good decoding schemes may result in high complexity of the receiver.

A low-complexity detection scheme for multiple antenna systems in a fading environment has been proposed in [3]. This detection scheme (so-called "nulling-and-canceling"), depending on the adopted criterion, essentially performs zero-forcing or minimum-mean-square-error decision feedback equalization on block transmissions. In [4], a technique referred to as the "sphere decoder" (based on the Fincke-Pohst algorithm [5]) was proposed for lattice code decoding and further adapted for space-time codes in [6]. The sphere decoder provides the maximum-likelihood (ML) estimate of the transmitted signal sequence and so often significantly outperforms nulling and cancelling. Moreover, it was generally believed to require much greater computational complexity than the cubic-time nulling and cancelling techniques. However, in [7] an analytic expression for the expected complexity of the sphere decoder has been obtained where it is shown that, over a wide range of rates and signal-to-noise ratios (SNRs), the expected complexity is polynomial-time (often sub-cubic). This implies that in many cases of interest maximum-likelihood performance can be obtained with complexity similar to nulling and cancelling.

Another area of intense research activity is that of soft iterative decoding. Such techniques have been reported to achieve impressive results for codes with long codeword length. Following the

seminal paper by Berrou et al [8], there have been many results on turbo decoding, with performances approaching the Shannon limit on single-input single-output systems (see [9] and references therein). More recently, low-density parity check (LDPC) codes, long neglected since their introduction by Gallager [10] have also been resurrected (see, e.g., [11],[12]).

Crucial to both turbo and LDPC decoding techniques is the use of the probabilistic ("soft") information about each bit in the transmitted sequence. For multiple antenna systems employing space-time codes it is not clear what is the best way to obtain this soft-information with low complexity. As noted in [13], where turbo-coded modulation for multiple antenna systems has been studied, if soft information is obtained by means of an exhaustive search, the computational complexity grows exponentially in the number of transmit antennas and in the size of the constellation. Hence, for high-rate systems with large number of antennas, the exhaustive search proves to be practically infeasible. Therefore heuristics are often employed to obtain soft channel information (see, e.g., [13]). Recently in [14], the sphere decoder has been employed to obtain a list of bit sequences that are "good" in a likelihood sense. This list is then used to generate soft information, which is subsequently updated by iterative decoder decisions.

In this paper, we propose a modification to the original Fincke-Pohst algorithm to obtain soft information on the transmitted bit sequence. The modified Fincke-Pohst algorithm essentially performs a maximum a posteriori (MAP) search and provides soft information for the iterative decoder (e.g., turbo or LDPC). The soft decoder's output is then fed back to the Fincke-Pohst MAP (FP-MAP), and iterated on. Our method differs from that of [14] in that the sphere decoder is modified (to allow for the introduction of soft information from the iterative decoder), that it performs MAP search, and that it is repeated for each iteration. We will assume that there are more receive (N) than transmit (M) antennas, so that the sphere decoder can be efficiently implemented. To accommodate for $N < M$, one can use an LD code as in [16].

2. SYSTEM MODEL

We assume a discrete-time block-fading multiple antenna channel model, where the channel is known to the receiver. This is a reasonable assumption for communication systems where the signaling rate is much faster than the pace at which the propagation environment changes, so that the channel may be learned via, e.g., transmitting known training sequences. During any channel use the transmitted signal $\mathbf{s} \in \mathcal{S}^{M \times 1}$ and received signal $\mathbf{x} \in \mathcal{S}^{N \times 1}$ are related by

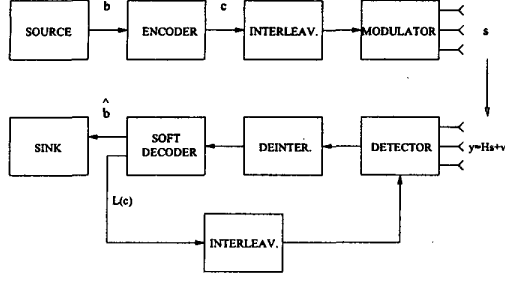$$\mathbf{x} = \sqrt{\frac{\rho}{M}} \mathbf{H} \mathbf{s} + \mathbf{v}, \tag{1}$$

Fig. 1. *System model*

where $\mathbf{H} \in \mathcal{C}^{N \times M}$ is the known channel matrix, and $\mathbf{v} \in \mathcal{C}^{N \times 1}$ is the additive noise vector, comprised of independent, identically distributed complex-Gaussian entries $\mathcal{C}(0,1)$. If we assume that the entries of s and $\mathbf{H}$ have, on the average, unit variance, then ρ is the expected received signal-to-noise ratio (SNR).

An iterative decoding scheme is shown in Figure 1. The information bit sequence b is first encoded with error-correcting code. Upon interleaving, coded bit sequence c is modulated onto symbol vectors s and transmitted across the MIMO channel. Detector in Figure 1, based on modified sphere decoder, accepts at the input the received symbol sequence, along with a priori probabilities of the coded bits; it outputs both the estimated coded bit sequence and probabilities associated with the correct detection of each bit. This probability is often expressed as a log-likelihood ratio (LLR),

$$L(c_i | \mathbf{x}) = \log \frac{P[c_i = +1 | \mathbf{x}]}{P[c_i = -1 | \mathbf{x}]}. \tag{2}$$

[Note: we will represent logical 0 with amplitude level -1, and logical 1 with amplitude level $+1$.]

The computational complexity of a traditional MAP algorithm can be prohibitive for application in multiple antenna systems. Since the sphere decoding algorithm of Fincke and Pohst can supply us with the ML estimate of s with reasonable complexity, one may speculate whether a modification can be devised to yield soft information with low complexity. We show that this can be done and describe how to efficiently approximate the LLRs in the following section.

3. MODIFIED FP ALGORITHM FOR MAP DETECTION

Assuming model (1), MAP detector maximizes the posterior probability $p_{\mathbf{s}|\mathbf{x}}(\mathbf{s}|\mathbf{x})$, i.e., it solves optimization problem

$$\max_{s \in \mathcal{D}_L^m \subset \mathcal{Z}^m} p_{\mathbf{s}|\mathbf{x}}(\mathbf{s}|\mathbf{x}). \tag{3}$$

[The notation $s \in \mathcal{D}_L^m \subset \mathcal{Z}^m$ denotes that the search space in (3) is a (finite) subset of the m-dimensional integer lattice $\mathcal{Z}^m$. This subset is often an L-PAM constellation,

$$\mathcal{D}_L^m = \{-\frac{L-1}{2}, -\frac{L-3}{2}, \dots, \frac{L-3}{2}, \frac{L-1}{2}\},$$

where L is usually a power of 2. Moreover, $m = 2M$ and s is simply the $m = 2M$-dimensional real vector obtained from the M-dimensional complex vector s.]

Using Bayes' rule,

$$\arg \max_{s \in \mathcal{D}_L^m} p_{\mathbf{s}|\mathbf{x}}(\mathbf{s}|\mathbf{x}) = \arg \max_{s \in \mathcal{D}_L^m} p_{\mathbf{x}|\mathbf{s}}(\mathbf{x}|\mathbf{s}) p_{\mathbf{s}}(\mathbf{s}).$$

Since $\mathbf{H}$ is known and noise is zero-mean unit-variance Gaussian, the conditional distribution of x given s is

$$p_{x|s}(x|s) = \frac{1}{\pi} e^{-\|x - Hs\|^2},$$

where $x \in \mathcal{R}^{2N \times 1}$ is the real vector obtained from x and $H = \sqrt{\frac{\rho}{M}} \begin{bmatrix} \mathbf{H}_R & \mathbf{H}_I \\ -\mathbf{H}_I & \mathbf{H}_R \end{bmatrix} \in \mathcal{R}^{2N \times 2M}$. Further, by assuming that the symbols $s_1, s_2, \dots, s_m$ are independent, we note that $p_{\mathbf{s}}(\mathbf{s}) = \prod_{k=1}^m p(s_k) = e^{\sum_{k=1}^m \log p(s_k)}$. Then, for a known channel in AWGN, maximization (3) is equivalent to the optimization

$$\min_{s \in \mathcal{D}_L^m} \left[\|x - Hs\|^2 - \sum_{k=1}^m \log p(s_k) \right] \tag{4}$$

However, for an iterative decoding scheme, we also require soft information, i.e., probability that each bit is decoded correctly. Assuming independent bits $c_1, c_2, \dots, c_k$, LLR becomes $L(c_i | \mathbf{x}) =$

$$\underbrace{\log \frac{p[c_i = +1]}{p[c_i = -1]}}_{L_a(c_i)} + \log \underbrace{\frac{\sum_{c \, : \, c_i = +1} p[x|\mathbf{c}] \prod_{j, j \neq i} p[c_j]}{\sum_{c \, : \, c_i = -1} p[x|\mathbf{c}] \prod_{j, j \neq i} p[c_j]}}_{L_e(c_i)}$$

$$= \log \frac{\sum\limits_{s \, : \, c_i = +1} e^{-\|x - Hs\|^2 + \sum_j \log p[s_j]}}{\sum\limits_{s \, : \, c_i = -1} e^{-\|x - Hs\|^2 + \sum_j \log p[s_j]}}$$

$$\tag{5}$$

where $L_a(c_i)$ and $L_e(c_i)$ denote so-called *a priori* and *extrinsic* part of the total soft information, respectively. [Note that, when used in an iterative decoding scheme, it is only $L_e(c_i)$ that is passed to the other decoding block(s) in the scheme.]

Computing (5) over entire signal space $\mathcal{D}_L^m$ is of prohibitive complexity. Instead, we constrain ourselves to those $s \in \mathcal{D}_L^m$ for which argument in (4) is small. [Note that these are the signal vectors whose contribution to the numerator and denominator in (5) is significant.] Applying the idea of Fincke-Pohst algorithm, rather than to search over the entire lattice, we search only over lattice points s that belong to the geometric body described by

$$r^2 \geq (s - \hat{s})^* R^* R(s - \hat{s}) - \sum_{k=1}^n \log p(s_k), \tag{6}$$

where R is lower triangular matrix following QR factorization of H. [Note that, unlike in the original sphere decoder algorithm, this geometric body is no longer a hypersphere.] The search radius r in (6) can be chosen according to the statistical properties of the noise and a priori distribution of s (so that, for instance, algorithm yields chosen average number of points which satisfy (6)).

Let r_{ij} denotes (i, j) entry of R. A necessary condition for s_m to satisfy (6) readily follows,

$$r_{mm}^2 (s_m - \hat{s}_m)^2 - \log p(s_m) \leq r^2. \tag{7}$$

Moreover, for every s_m satisfying (7), we define

$$r_{m-1}^2 = r^2 - r_{mm}^2 (s_m - \hat{s}_m)^2 + \log p(s_m),$$

and obtain a stronger necessary condition for (6) to hold,

$$r_{m-1,m-1}^2 \left(s_{m-1} - \hat{s}_{m-1} + \frac{r_{m-1,m}}{r_{m-1,m-1}} (s_m - \hat{s}_m) \right)^2 - \log p(s_{m-1}) \leq r_{m-1}^2.$$

The procedure continues in this fashion until all the components of vector s are found. Assume that the search yields the set of points $S = \{s^{(1)}, s^{(2)}, \ldots, s^{(l)}\}$. Then the vector $s \in S$ minimizing (4) is the solution to the MAP detection problem. The soft information for each bit c_i can be estimated from (5), by only summing the terms in the numerator and denominator such that $s \in S$.

4. COMPUTATIONAL COMPLEXITY OF FP-MAP ALGORITHM

The computational complexity of FP-MAP algorithm, due to its search strategy, is a random variable. The average complexity of the algorithm is proportional to the expected number of lattice points visited in dimensions $k = 1, 2, \ldots, m$ while solving for (4). Assume that the entries of $\mathbf{H}$ are independent $\mathcal{C}(0,1)$ random variables and, for simplicity that $M = N$ (the more general case of $M \neq N$ can be considered similarly). Then, if we define

$$r'^2 = r^2 + \sum_{j=1}^{m} \log p(s_j).$$

and use the results of [7], the expected number of lattice points in the k-th dimension is given by

$$E_p(k, L, \rho) = \sum_n \frac{1}{L^k} \sum_{n=0}^{k} C_S(k, n, L) \gamma \left(\frac{\frac{12\rho n}{m(L^2-12)} r'^2}{1 + \frac{12\rho n}{m(L^2-12)}}, \frac{k}{2} \right),$$

$$(8)$$

where $\gamma(\eta, \xi)$ denotes the incomplete gamma function of order ξ and argument η, and $C_S(k, n, L) =$

$$\sum_{\substack{l_0, l_1, \ldots, l_i \geq 0 \\ l_0 + l_1 + \cdots + l_i = k}} \binom{k}{l_0, l_1, \ldots, l_i} \{\prod_{j=0}^{i} \phi_j^{l_j}(x)\}_n,$$

where $i = L/2 - 1$, $\{\phi(x)\}_n$ denotes coefficient of x^n in the polynomial $\phi(x)$, and the function $\phi_j(x)$ is defined as

$$\phi_j(x) = \sum_{q=0}^{L-j-1} \psi_q x^{q^2}, \quad \psi_q = \begin{cases} 2, & \text{if } 1 \leq q \leq j \\ 1, & \text{otherwise} \end{cases}$$

The number of computations per k-dimensional lattice point visited is $2k + 17$, so that the expected complexity is given by

$$C(m, L, \rho) = \sum_{k=1}^{m} (2k + 17) E_p(k, L, \rho).$$

$$(9)$$

The above expression for the expected complexity can be readily evaluated (especially for moderate values of L). This is done in [7], where it is shown that for a wide range of m, L and ρ, the sphere decoder algorithm has complexity comparable to cubic-time methods such as nulling and cancelling. As a general principle, for a fixed m, the complexity decreases by increasing the SNR ρ or by decreasing the value of L. This has important implications for the design of space-time turbo and LDPC codes that lend themselves to efficient iterative decoding.

5. SIMULATION RESULTS

Figure 2 shows the BER performance of the system with $M = 4$ transmit and $N = 4$ receive antennas, $16 - QAM$ constellation and parallel concatenated turbo code with rate $R = 1/2$ and length 9216 information bits. Constituent convolutional codes have memory length 2 and generating polynomials $G_1(D) = 1 + D^2$ (feedforward) and $G_2(D) = 1 + D + D^2$ (feedback). For each iteration of the FP-MAP, turbo (inner) decoder performs 8 iterations of its own. Figure 3 shows the BER performance of the
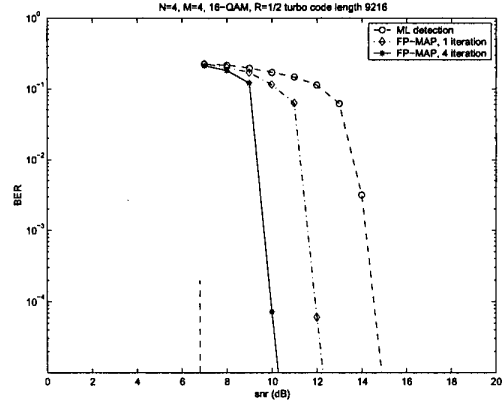

Fig. 2. *Rate 1/2 turbo code, 4 × 4 system, 16-QAM*

system of same size, only with convolutional code (memory 2) in place of the turbo code. Figure 4 shows the BER performance of the same system (4 × 4), with 4-QAM constellation and 8/9 LDPC code of length 1088, column weight 4. When LDPC decoder receives soft information from FP-MAP, it performs 8 iterations before passing what it inferred about coded bits back to FP-MAP. In Figure 2-Figure 4, dashed vertical line denotes capacity limits of the MIMO channel. Turbo coded scheme in Figure 2 gets approximately 3.3dB away from capacity. It outperforms convolutional codes employed on the same system by approximately 3dB. The rate of the system is 8 bits per channel use. LDPC code, on the other hand, is about 4.5dB away from capacity of the system in
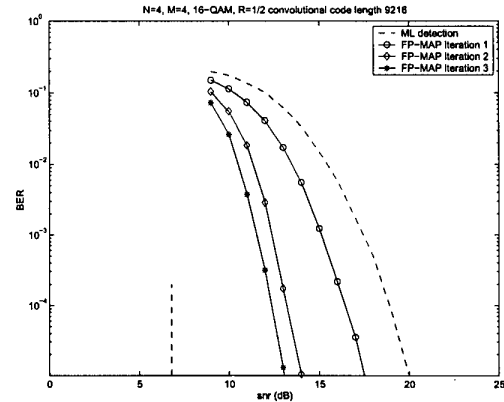

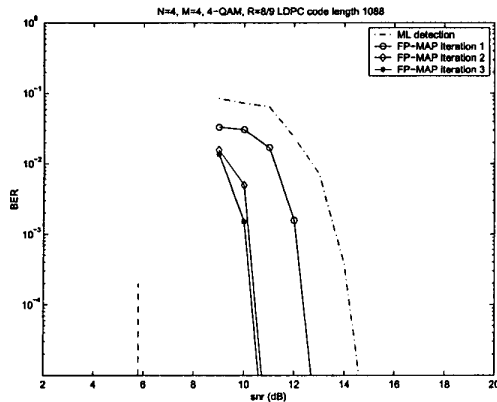**Fig. 3.** *Rate 1/2 convolutional code, 4 × 4 system, 16-QAM*

Fig. 4. *Rate 8/9 LDPC code, 4 × 4 system, 4-QAM*

which it is employed; the data rate in this system is 7.1 bits per channel use. Though the LDPC code is outperformed by the turbo code, it proves to be an interesting alternative, especially in light of complexity exponents $\log_m C(m, L, \rho)$ shown in Figure 6. As indicated, complexity of the detection in the system employing (high rate) LDPC code is approximately cubic, while the complexity in the system with (1/2 rate) turbo code is significantly higher. [Note that the performance of LDPC code is likely affected by imposing a short codelength.]

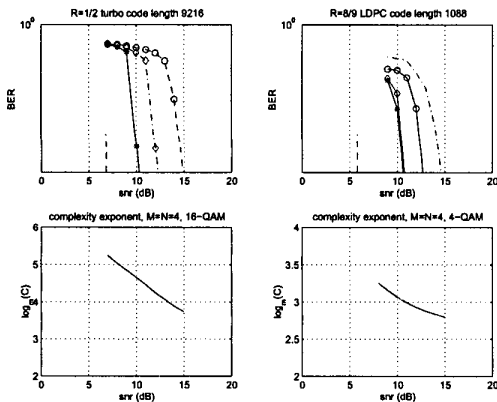**Fig. 5.** *Complexity exponent $\log_m C$*

6. CONCLUSION

In this paper, we developed a modification of the Fincke-Pohst (sphere decoder) algorithm to perform MAP detection and efficiently estimate soft information. When combined with soft iterative decoding schemes, the proposed detection algorithm provides performances of multiple antenna systems which are close to capacity. This was demonstrated on systems employing both turbo and LDPC codes. When combined with soft iterative decoding schemes, proposed algorithm allows one to get close to capacity (e.g., approximately 3.3dB in 4 × 4, 16-QAM system employing turbo code).

We developed an analytic expression for expected complexity of the algorithm. Over a wide range of rates and SNRs, the algorithm has polynomial-time (often cubic) complexity. The algorithm is faster in systems employing the high-rate error-correcting codes. Although the simulations we presented were for $M = N$, once can still use the methodology for $M > N$, provided one uses, for instance, an LD code.

7. REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.

[2] V. Tarokh V, N. Seshadri, and A. R. Calderbank, "Space-time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. on Info. Theory*, Vol. 44, No. 2, pp. 744-765, March 1998.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. Journal*, pp. 41-59, 1996.

[4] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," *IEEE Trans. on Information Theory*, vol. 45, pp. 1639-1642, July 1997.

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

[6] M. O. Damen, A. Chkeif, and J. C. Belfiore, "Lattice codes decoder for space-time codes," *IEEE Communications Letters*, vol. 4, pp. 161-163, May 2000.

[7] B. Hassibi and H. Vikalo, "Expected complexity of the sphere decoder algorithm," *submitted to the IEEE Trans. on Signal Processing*, 2001.

[8] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding. Turbo codes," in *Proc. Int. Conf. Communications*, May 1993, pp. 1064-1070.

[9] J. P. Woodard and L. Hanzo, "Comparative study of turbo decoding techniques: an overview," *IEEE Trans. on Vehicular Technology*, vol. 49, no. 6, November 2000.

[10] R. G. Gallager, *Low-density parity-check codes*, Cambridge, MA: MIT Press, 1963.

[11] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low-density parity-check codes," *Electron. Lett.*, vol. 32, pp. 1645-1646, 1996.

[12] D. J. C. MacKay, "Good error correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, March 1999.

[13] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: system model, decoding approaches, and practical considerations," *IEEE Journal on Sel. Areas in Commun.*, vol. 19, no. 5, May 2001.

[14] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," submitted to *IEEE Journal on Selec. Areas in Commun.*, 2001.

[15] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Info. Theory*, pp. 284-287, March 1974.

[16] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," submitted to *IEEE Trans. Info. Theory*, 2000. Download at http://mars.bell-labs.com.