

Linear Space-Time Codes Optimized For Successive Cancellation Decoding

Badri Varadarajan and John R. Barry

School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia, 30332-0250

Abstract— We optimize linear space-time codes for the case when the receiver uses successive cancellation decoding. Specifically, the proposed codes minimize the perfect cancellation bound on word error probability, which assumes error-free cancellation of previously detected symbols. Assuming perfect cancellation, we prove that to minimize the error probability in each stage of decoding, the encoding matrix must have orthogonal columns, regardless of the channel matrix.

Given the encoding matrix, the average of the perfect cancellation bound over the random channel matrix serves as an upper bound on word error probability. The bound is minimized by numerically optimizing the distribution of data rate and energy among the various inputs to the space-time code. Simulation results for a 4-input, 4-output Rayleigh fading channel show that, at 12 b/s/Hz, optimizing the data rate and energy allocations for a linear complex field code leads to a performance improvement of nearly 9 dB.

I. INTRODUCTION

The performance of communication systems operating over multiple-input, multiple-output (MIMO) fading channels depends strongly on the type of decoder employed. Decoders that minimize error probability (for example, sphere decoders), have high computational complexity when the number of channel inputs is large. One low-complexity alternative is the class of *successive cancellation* (SC) decoders, popularized by the V-BLAST architecture [2]. SC decoders detect one symbol in every stage of decoding, after cancelling the contribution of already detected symbols and nulling out the interference from other undetected symbols.

Over fading channels, SC decoders have low diversity order [2], leading to poor performance at high signal-to-noise energy ratio (SNR). One method to improve the diversity order is to use space-time codes at the transmitter. Generalized layered space-time (LST) codes [6] divide the transmit antennas into groups, each of which is independently space-time encoded. Correspondingly, successive cancellation is extended so that each stage detects an entire space-time coded group instead of just one transmit symbol.

An alternative method, which is the focus of this paper, is to use *linear dispersion* codes [1] at the transmitter. Symbol-by-symbol successive cancellation is run on the effective channel formed by the combination of the linear space-time code and the underlying MIMO channel. If the cancellation of past (detected) symbols is correct, this amounts to transmitting each input symbol across a scalar additive noise channel, allowing one to employ AWGN channel

architectures. Thus, one can transmit uncoded QAM symbols or use a near-capacity code in every stage. In practice, cancellation from past stages is error-prone. However, assuming error-free cancellation gives the useful *perfect cancellation* (PC) bound on actual performance. The PC bound is a decreasing function of the effective SNR at each stage, which in turn depends strongly on the random channel matrix and the encoding matrix used by the space-time code.

We show that to maximize the effective SNR at each stage (and hence minimize the PC bound), the encoding matrix must have orthogonal columns. Remarkably, this holds for every realization of the channel matrix, irrespective of the coding architecture used at each stage, or of the average SNR. Unfortunately, getting the optimum orthogonal-column encoding matrix is an open problem. One possible candidate, though not necessarily optimal, is the recently developed linear complex field codes [8][9], which achieve full rate and full diversity with sphere decoding.

Assuming that the encoding matrix has been chosen, for every coding architecture one can compute the PC bound on the error probability by averaging over the random channel matrix. The average PC bound depends on the bit and energy distributions among the various stages, as well as the average SNR of operation. Consequently, the average PC bound on error probability can be minimized by optimally distributing the total number of bits and the total energy per block among the inputs to different stages. When no space-time code is used (as in the V-BLAST transmitter), optimal bit and energy loading has been done in [3] for uncoded QAM transmission on each stage. Subsequently, in [4] (see also [5]), an upper bound to the word error probability was used to optimize allocations for uncoded QAM transmission using Alamouti-based GLST space-time codes.

However, the techniques of [3] and [4] do not extend directly to any general linear space-time code. To compute the average PC bound, one requires the distribution of effective SNR at every stage. Unfortunately, the distribution is strongly dependent on the encoding matrix, and is difficult to obtain in closed form. In this paper, we propose to circumvent this problem by empirically obtaining discrete approximations of the necessary distribution. The empirical approximation needs to be obtained only once for every encoding matrix, and can be re-used for optimization at different SNR.

It is well established [7][9] that the restriction to SC decoding leads to a rate-diversity tradeoff for linear space-time codes. Consequently, an interesting problem in code

This research was supported by in part National Science Foundation grants CCR-0082329 and CCR-0121565.

design for SC decoders is to find the rate that minimizes error probability, at a given data rate and SNR. The bit and energy allocation in the second design step above *implicitly* optimizes the rate by making some of the inputs carry zero bits, effectively reducing the rate of the space-time code.

In Section II, we present the channel model and describe successive cancellation decoding for linear space-time codes. In Section III, we derive the perfect cancellation bound, and show the optimality of using encoding matrices with orthonormal columns. Section IV describes the optimization of the input bit and energy allocations, given the encoding matrix. Section V presents simulation results demonstrating the benefits of optimization. Finally, in Section VI, we summarize our conclusions.

II. SYSTEM MODEL AND THE SUCCESSIVE CANCELLATION ALGORITHM

We consider a t -transmit, r -receive antenna static wireless narrowband MIMO channel, modeled as

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k, \quad (1)$$

where \mathbf{x}_k is the $t \times 1$ channel input and \mathbf{y}_k the $r \times 1$ channel output at time k . The noise \mathbf{n}_k is spectrally and temporally white, so that $\mathbb{E}[\mathbf{n}_{k+l}\mathbf{n}_k^H] = \delta_l N_0 \mathbf{I}_r$. The entries of the $r \times t$ Rayleigh fading channel matrix \mathbf{H} are independent, circularly symmetric, unit-variance Gaussian random variables. We assume that \mathbf{H} is unknown to the transmitter, but known to the receiver. In every block of encoding, a rate- K/N space-time encoder takes in a $K \times 1$ complex vector \mathbf{u} and generates $t \times 1$ transmit vectors \mathbf{x}_i for N consecutive signaling intervals. We restrict attention to *strictly linear* space-time codes which obtain each complex output symbol by some linear combination of the K elements of the input \mathbf{u} . More precisely, consider the $Nt \times 1$ composite transmit vector \mathbf{x} formed by stacking the N transmit vectors in the code-block one below the other, namely $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T]^T$. For a strictly linear space-time code, \mathbf{x} is related to the input vector \mathbf{u} by

$$\mathbf{x} = \mathbf{M}\mathbf{u}, \quad (2)$$

where the $Nt \times K$ *encoding matrix* \mathbf{M} completely specifies the code. We assume that the elements of \mathbf{u} have unit energy, and \mathbf{M} is chosen so that its energy, given by $\|\mathbf{M}\|_F^2$, is equal to one. Thus, the average transmit energy per block is $\mathbb{E}[\|\mathbf{x}\|^2] = 1$. The SNR S is defined as the ratio of the average received signal energy to the average noise energy. It is easy to see that $S = \mathbb{E}[\|\mathbf{x}\|^2]/(NN_0) = 1/(NN_0)$, where N is the length of the space-time code block.

The composite received vector \mathbf{y} of dimension $Nr \times 1$ is obtained by stacking the received vectors one below the other. From the channel model (1), we get

$$\mathbf{y} = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \ddots & \\ \vdots & \ddots & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{H} \end{bmatrix} \mathbf{x} + \mathbf{n} = \tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}, \quad (3)$$

where we have defined the block diagonal matrix $\tilde{\mathbf{H}}$, and \mathbf{n} is the composite noise vector. Substituting the encoding rule (2) in the above and defining $\mathbf{G} = \tilde{\mathbf{H}}\mathbf{M}$, we get the relation

$$\mathbf{y} = \tilde{\mathbf{H}}\mathbf{M}\mathbf{u} + \mathbf{n} = \mathbf{G}\mathbf{u} + \mathbf{n}. \quad (4)$$

Given \mathbf{H} , the successive cancellation (SC) decoding algorithm is based on the so-called QL decomposition $\mathbf{G} = \mathbf{Q}\mathbf{L}$, where \mathbf{Q} has orthonormal columns and \mathbf{L} is a $K \times K$ lower triangular matrix with real, non-negative diagonal entries. The decomposition can be done, for example, using Gram Schmidt orthonormalization of the columns of \mathbf{G} . Now, the SC algorithm obtains the effective received vector $\mathbf{y}' = \mathbf{Q}^H\mathbf{y}$. Substituting for \mathbf{y} from (4) and using $\mathbf{Q}^H\mathbf{G} = \mathbf{L}$, we get

$$\mathbf{y}' = \mathbf{Q}^H\mathbf{y} = \mathbf{L}\mathbf{u} + \mathbf{n}', \quad (5)$$

where we have defined the effective noise vector $\mathbf{n}' = \mathbf{Q}^H\mathbf{n}$. The orthonormality of the columns of \mathbf{Q} ensures that the elements of \mathbf{n}' are independent complex Gaussians of variance N_0 .

SC detection proceeds in K stages. In the i^{th} stage, the decoder obtains the *slicer input* z_i according to

$$z_i = (y'_i - \sum_{j=1}^{i-1} l_{ij}\hat{u}_j), \quad (6)$$

where l_{ij} denotes the $(i, j)^{\text{th}}$ entry of \mathbf{L} . Then z_i is used to obtain the i^{th} symbol decision \hat{u}_i using the conventional AWGN channel model $z_i = l_{ii}u_i + n'_i$. Starting from the first stage $i = 1$, note that past decisions \hat{u}_j are always available for cancellation in (6).

III. THE PERFECT CANCELLATION BOUND AND OPTIMALITY OF ORTHOGONAL ENCODING MATRICES

We now derive an upper bound for the word error probability under successive cancellation for a given value of the random channel matrix \mathbf{H} . Let \mathcal{E}_i denote the event that there is an error in the i^{th} stage, namely that $\hat{u}_i \neq u_i$, and let C_i denote the complementary event $\hat{u}_i = u_i$. It is easy to see that

$$\Pr(\hat{\mathbf{u}} \neq \mathbf{u} | \mathbf{H}) = \sum_{i=1}^K \Pr(\mathcal{E}_i, C_{i-1}, \dots, C_1 | \mathbf{H}) = \sum_{i=1}^K \Pr(\mathcal{E}_i | C_{i-1}, \dots, C_1, \mathbf{H}) \Pr(C_{i-1}, \dots, C_1 | \mathbf{H}). \quad (7)$$

By neglecting the second term in each product above, we get the *perfect cancellation* (PC) bound

$$\Pr(\hat{\mathbf{u}} \neq \mathbf{u} | \mathbf{H}) \leq \sum_{i=1}^K \Pr(\mathcal{E}_i | C_{i-1}, \dots, C_1, \mathbf{H}). \quad (8)$$

Each term on the right hand side is the probability of error in the i^{th} stage, assuming error-free cancellation of symbols from previous stages. Under the assumption, the scaled AWGN channel model $z_i = l_{ii}u_i + n'_i$ holds. Since u_i has unit average energy, the effective SNR at stage i is l_{ii}^2/N_0 .

The effective SNR at various stages — and hence the PC bound on decoding — depends on the choice of the encoding matrix \mathbf{M} . We now show that to maximize the effective SNR,

\mathbf{M} should have orthogonal columns. Let the QL decomposition of \mathbf{M} be $\mathbf{M} = \mathbf{P}\mathbf{K}$, where \mathbf{P} has orthonormal columns and the lower triangular matrix \mathbf{K} has non-negative diagonal entries. For a given channel matrix \mathbf{H} , let the QL decomposition of $\tilde{\mathbf{H}}\mathbf{P}$ be $\mathbf{Q}\mathbf{L}'$. Then, it is easy to show that the unique QL decomposition of $\tilde{\mathbf{H}}\mathbf{M} = \tilde{\mathbf{H}}\mathbf{P}\mathbf{K}$ is $\mathbf{Q}\mathbf{L}'\mathbf{K}$. In particular, $\mathbf{L} = \mathbf{L}'\mathbf{K}$ and the diagonal entries l_{ii} which determine the PC bound (8) are given by $l_{ii} = l_{ii}'k_{ii}$. To maximize the effective SNR, one needs to maximize k_{ii} under the energy constraint $\|\mathbf{K}\|_F = \|\mathbf{M}\|_F = 1$. Clearly, this is done by concentrating all the energy on the diagonal of \mathbf{K} . When \mathbf{K} is diagonal, the columns of \mathbf{M} are proportional to the orthonormal columns of \mathbf{P} and are therefore orthogonal, proving the following result.

Theorem 1. To maximize the effective SNR of a successive cancellation decoder, the encoding matrix of a linear space-time code should have orthogonal columns, irrespective of the average SNR and the channel matrix.

According to Theorem 1, the optimum encoding matrix must be of the form $\mathbf{M} = \mathbf{P}\mathbf{D}$, where \mathbf{P} is an $Nt \times K$ matrix with orthonormal columns, and \mathbf{D} is a $K \times K$ diagonal matrix. Finding the optimum orthonormal component \mathbf{P} is an interesting open problem. The diagonal component \mathbf{D} determines the energy distribution among the various symbols. The i^{th} symbol is transmitted with energy $E_i = d_{ii}^2$.

IV. OPTIMIZATION OF BIT AND ENERGY ALLOCATIONS

In this section, we will assume that the matrix \mathbf{P} in $\mathbf{M} = \mathbf{P}\mathbf{D}$ is fixed, and discuss ways of minimizing the average PC bound by optimally distributing data rate and energy among the various inputs to the space-time code. First, we will change notation slightly and let $\tilde{\mathbf{H}}\mathbf{P} = \mathbf{Q}\mathbf{L}$. Now, the perfect cancellation assumption yields an effective AWGN channel in stage i , whose error probability $\Pr(E_i | C_{i-1}, \dots, C_1, \mathbf{H})$ depends on the bit rate b_i and the effective SNR $S_i = E_i l_{ii}^2 / N_0 = NSE_i l_{ii}^2$ (using $1/N_0 = NS$). Averaging over \mathbf{H} amounts to averaging over l_{ii} , giving

$$\Pr(\hat{\mathbf{u}} \neq \mathbf{u}) \leq \sum_{i=1}^K E_i l_{ii} g(b_i, NSE_i l_{ii}^2). \quad (9)$$

The error function $g(b_i, S_i)$ quantifies the variation of error probability with bit rate b_i and effective SNR S_i , and depends on the architecture used. For example, if each transmit symbol u_i were independently drawn from a square QAM constellation and detected by conventional slicing, the error function is closely approximated by

$$g(b_i, S_i) = 4(1 - 2^{-b_i/2})\mathcal{Q}\left(\sqrt{\frac{3S_i}{2^{b_i} - 1}}\right). \quad (10)$$

On the other hand, when each stage is equipped with an encoder-decoder pair that achieves capacity over an AWGN channel, errors occur only in the case of an outage. In this case,

$$g(b_i, S_i) = 1 \text{ if } \log_2(1 + S_i) < b_i, \text{ and } 0 \text{ otherwise.} \quad (11)$$

For brevity, we will hereafter focus on uncoded square-QAM transmission. To obtain the average PC bound (9), we need to average the error function (10) over l_{ii} . Getting a closed form expression for the probability density function of l_{ii} is a difficult open problem. Instead, we suggest a discrete numerical approximation. For a given trial value of the channel matrix \mathbf{H} , the value of l_{ii} is computed, and rounded off to the nearest discrete value in a finite set \mathcal{L} . Given a sufficient number of randomly generated trial values of \mathbf{H} and a sufficiently large set \mathcal{L} , the discrete pdf on \mathcal{L} so obtained, say $p_i(l)$, closely approximates the unknown pdf of l_{ii} .

Using the discrete approximation to average out the perfect cancellation bound (9), we get the approximate PC bound on $\Pr(\hat{\mathbf{u}} \neq \mathbf{u})$, namely

$$f(\{b_i\}, \{E_i\}, S) = \sum_{i=1}^K \sum_{l \in \mathcal{L}} p_i(l) g(b_i, NSE_i l^2). \quad (12)$$

The approximate PC bound depends on the distribution of the bit rates $\{b_i\}$ and energies $\{E_i\}$, setting up the following optimization problem.

Problem Statement: Given the SNR S and the $Nt \times K$ encoding matrix \mathbf{P} of a linear space-time code transmitting B bits per block, find the bit rate and energy distributions $\{b_i\}, \{E_i\}$ that minimize the upper bound $f(\{b_i\}, \{E_i\}, S)$ subject to the constraints

$$\sum_{i=1}^K b_i = B, b_i \in \{0, 2, 4, 6, \dots\}, \text{ and } \sum_{i=1}^K E_i = 1, E_i \geq 0. \quad (13)$$

We propose the following numerical solution.

Step 1. From the matrix \mathbf{P} , use Monte Carlo simulation to obtain the approximate discretized pdf of the effective scaling factor l_{ii} for all stages $i = 1, 2, \dots, K$.

Step 2. For all valid bit sequences $\{b_i\}$ with non-negative even integer values summing to B , use standard constrained gradient descent to find the values of $\{E_i\}$, which minimize (12) while satisfying (13).

Step 3. Pick the combination of $\{b_i^*\}$ and $\{E_i^*\}$ that has the lowest average PC bound (12).

Intuitively, since the first few stages typically have lower effective SNR, it is better to allocate lesser bit rate and/or higher energy to these. In particular, it is sometimes optimum to allocate zero bits to the first few symbols. This amounts to not using the corresponding inputs of the space-time code, or equivalently reducing the effective number of inputs and hence the rate of the space-time code. Thus, the bit-loading procedure described above also optimally trades off rate for diversity.

We emphasize that to optimize the bit and energy allocations, the transmitter needs to know only the statistical properties of the channel and not the channel itself. Also, though the optimization metric (13) depends on the SNR S , note that the numerical estimation of the pdf in Step 1 needs to be performed just once for every \mathbf{P} , and can be re-used for optimization at different SNR.

One final remark relates to detection order at the receiver. Though the transmitter is optimized assuming that the receiver does not do any ordering, the receiver can use its knowledge of the channel to obtain a detection order that reduces the probability of error. The V-BLAST ordering algorithm maximizes the lowest value of l_{ii} among all the detection stages. However, taking the bit and energy allocation into account, we suggest that the symbol error probability $g(b_i, NSE_{i,l_{ii}})$ from (10) should be optimized instead of l_{ii} . To achieve this, the decoder should use l_{ii} to compute $g(b_i, NSE_{i,l_{ii}})$ for all undetected symbols, and pick the symbol least susceptible error for subsequent detection.

V. SIMULATION RESULTS

In this section, we present simulation results to illustrate the benefits of the design process described in this paper. We compare space-time codes transmitting 12 b/s/Hz over a 4-transmit, 4-receive antenna Rayleigh fading channel. We start with a linear complex field (LCF) code [9] of length $N = 2$ and $K = 8$ inputs per block, with parameters $\alpha = e^{0.5}$ and $\beta = e^{2.0}$ (see [9] for details). The encoding matrix has orthogonal columns and satisfies the optimality criterion in Theorem 1. Without the bit-loading step, all the inputs to this code would be drawn from an 8-QAM alphabet and transmitted with equal energy. This leads to poor performance with SC at the receiver, as seen from the curve labeled “unoptimized” in Fig. 1. However, optimization at SNR = 27.5 dB yields bit and energy allocations of $\{b_i^*\} = \{0, 0, 0, 4, 4, 4, 6, 6\}$ and $\{E_i^*\} = \{0, 0, 0, 0.2278, 0.1402, 0.1353, 0.2718, 0.2249\}$, respectively. Using the optimized allocations leads to a performance improvement of nearly 9 dB at 1% word error probability. Note that optimization leads to three zero bit allocations, thus reducing the rate of the space-time code to $5/2$ from its original value of $8/2 = 4$. The rate reduction results in increased diversity gain, as seen from the steeper error curve of the optimized code.

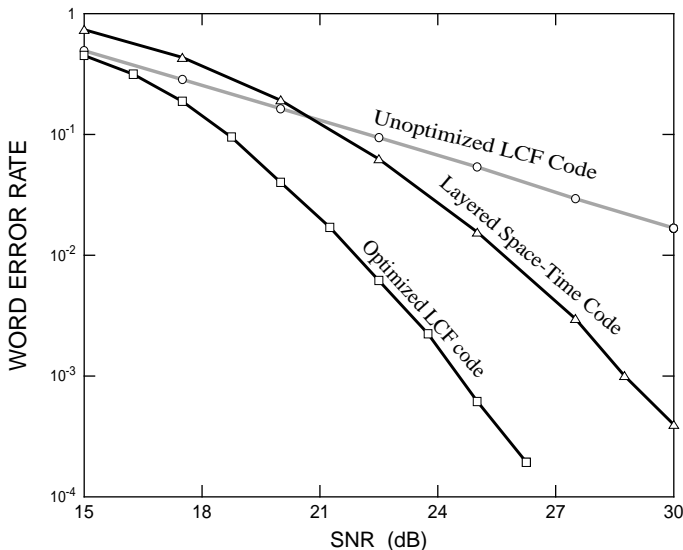


Fig. 1. Comparing space-time codes transmitting 12 b/s/Hz over 4-input, 4-output Rayleigh fading channel.

We also compare the optimized LCF code with a layered space-time (LST) code consisting of two parallel Alamouti codes transmitting symbols from a unit-energy 64-QAM alphabet, so as to maintain a total data rate of 12 b/s/Hz. In every stage of successive cancellation decoding of the LST code, a group of two Alamouti-coded symbols is detected. Remarkably, the optimized LCF code outperforms the LST code by 4.5 dB, despite the fact that layered space-time code uses the more sophisticated group detection instead of symbol detection at every stage.

VI. CONCLUSIONS

We optimized linear space-time codes for successive cancellation decoding. We showed that to maximize the effective SNR at every stage of SC decoding, the encoding matrix of the space-time code should have orthogonal columns. Further, given such an encoding matrix, we optimized the bit and energy distribution among the space-time code inputs, so as to minimize the perfect cancellation bound on word error probability. The distributions necessary for optimization were estimated empirically. Simulation results showed the benefits of the bit and energy loading, which also implicitly determines the optimum rate of a linear space-time code at a given SNR and data rate. Further, the bit-loading procedure can be extended to optimally design codes for other code architectures in each stage, and also to MMSE-based successive cancellation decoders.

VII. REFERENCES

- [1] B. Hassibi and B. M. Hochwald, “High-rate codes that are linear in space and time,” *IEEE Trans. on Info. Theory*, vol. 48, no. 7, July 2002.
- [2] G. Golden, G. Foschini, R. Venezuela, P. Wolniansky, “Detection Algorithm and initial laboratory results using V-BLAST space-time communication architecture,” *Elec. Lett.*, vol. 35-1, pp. 14-16, Jan. 1999.
- [3] N. Prasad and M. Varanasi, “Analysis of Decision Feedback Detection for MIMO Rayleigh Fading Channels and Optimum Allocation of Transmitter Powers and QAM Constellations,” *39th Annual Allerton Conf. on Comm. Control, and Comput.*, Monticello, IL, Oct. 2001.
- [4] N. Prasad and M. Varanasi, “Optimum Efficiently Decodable Layered Space-Time Block Codes,” *Proc. Asilomar Conf. on Signals, Systems and Computers*, Monterey, CA, Nov. 2001.
- [5] L. H. C. Jason, Meixia Tao and R. S. Cheng, “Optimal power allocation scheme on generalized layered space-time coding systems,” *Proc. IEEE Intl. Conf. on Comm., ICC 2001*, vol. 6, pp. 1706-1710, June 2001.
- [6] V. Tarokh, A. Naguib, N. Seshadri and A. Calderbank, “Combined array processing and space-time coding,” *IEEE Trans. on Info. Theory*, vol. 44, no. 2, pp. 1121-1128, Jan. 1999.
- [7] M. O. Damen and H. El Gamaal, “On the Diversity-vs.-Rate Tradeoff in MIMO Systems,” *Proceedings of Info. Theory Workshop ITW 2003*, Paris, pp. 357-360, Mar 31-Apr 4, 2003.
- [8] M. O. Damen, A. Tewfik and J. C. Belfiore, “A construction of a space-time code based on number theory,” *IEEE Trans. on Info. Theory*, vol. 48, no. 3, pp. 753-760, March 2002.
- [9] X. Ma and G. B. Giannakis, “Layered Space-Time Complex Field Coding: Full-Diversity with full rate, and trade-offs,” *Proc. of the 2nd Sensor Array and Multichannel SP Workshop*, Rosslyn, VA, pp. 442-446, August 2002.