

MMSE-GDFE Lattice Decoding for Under-determined Linear Channels

Mohamed Oussama Damen*, Hesham El Gamal** and Giuseppe Caire†

Abstract— Recently, the authors established the fundamental role of minimum mean square error generalized decision-feedback equalizer (MMSE-GDFE) lattice decoding in achieving the optimal diversity-vs-multiplexing trade-off of delay limited multiple-input multiple-output (MIMO) channels. This optimality motivates the current work where we exploit this technique in constructing an efficient decoding algorithm for under-determined linear channels. The proposed algorithm consists of an MMSE-GDFE front-end followed by a lattice reduction algorithm with a greedy ordering technique and, finally, a lattice search stage. By introducing flexibility in the termination strategy of the lattice search stage, we allow for trading performance for a reduction in the complexity. The proposed algorithm is shown, through experimental results in MIMO quasi-static channels, to offer significant gains over the state of the art decoding algorithms in terms of performance enhancement and complexity reduction. From one side, when the search is pursued until the best lattice point is found, the performance of the proposed algorithm is shown to be within a small fraction of a dB from the maximum likelihood (ML) decoder while offering a large reduction in complexity compared to the most efficient implementation of ML decoding proposed by Dayal and Varanasi (e.g., an order of magnitude in certain representative scenarios). On the other side, when the search is terminated after the first point is found, the algorithm only requires linear complexity while offering significant performance gains (in the order of several dBs) over the linear complexity algorithm proposed recently by Yao and Wornell, and independently by Windpassinger and Fisher.

I. INTRODUCTION AND PROBLEM FORMULATION

Sphere decoding based on Pohst and Schnorr-Euchner enumeration [1], [2] has received significant research interest in recent years (e.g., see [3], [4] and references therein). In particular, the introduction of the sphere decoder as a space-time ML decoder in [5] opened the door for constructing efficient and sophisticated codes that reap most of the promised theoretical gains of multiple-input multiple-output (MIMO) fading channels (e.g., [6], [7]). In [3], the authors have illustrated the importance of using the minimum mean square error generalized decision-feedback equalizer (MMSE-GDFE) with greedy ordering in reducing the average expected complexity of the sphere decoder. In a more recent work [8], the authors have established the optimality of MMSE-GDFE lattice decoding, when used

with a properly constructed Lattice Space-Time (LAST) code, in terms of achieving the diversity-vs-rate tradeoff characterized by Zheng and Tse [9] for MIMO channels with arbitrary numbers of transmit and receive antennas. In this work, we exploit this decoding strategy in constructing an efficient decoder for under-determined systems (e.g., in a point-to-point MIMO channel, an under-determined system corresponds to the case where the number of symbols sent per channel use is larger than the number of receive antennas). The proposed algorithm consists of three stages, namely 1) an MMSE-GDFE feed-forward filter, 2) a lattice reduction algorithm (e.g., the LLL algorithm [10]) along with a greedy ordering technique, and 3) a lattice search strategy based on Schnorr-Euchner enumeration. We further introduce flexibility in terminating the Schnorr-Euchner enumeration to allow for more freedom in the performance-vs-complexity tradeoff. At one extreme, when the search is terminated after the first point is found, the complexity of the algorithm only grows linearly in the number of variables (assuming very slow fading). This algorithm is shown to offer several dBs of performance gains over the recently proposed algorithm by Yao and Wornell [11] (and independently by Windpassinger and Fisher [12]) in a V-BLAST configuration. At the other extreme, when the search algorithm is allowed to pursue the best lattice point (according to the minimum Euclidean distance criterion), the performance of the algorithm is shown to be within a very small fraction of a dB from the performance of the maximum likelihood (ML) decoder. This variant of our algorithm, however, is shown to offer an order of magnitude of complexity reduction compared to Dayal and Varanasi recent implementation of the generalized sphere decoder (GSD) [13] in a 3×1 quasi-static fading channel [14]. Moreover, the reduction in complexity offered by our algorithm compared with the GSDs in [13], [14] is also expected to increase with the dimensionality of the problem.

II. THE SYSTEM MODEL

The proposed algorithm can be used to solve the general problem of separating m sources from observations made by $n \leq m$ sensors through linear Gaussian channels. To simplify the presentation, however, we focus here on the special case of a point-to-point $M \times N$ MIMO quasi-static, flat-fading, and Gaussian channel where the baseband complex model is given by

$$\mathbf{y}_t^c = \sqrt{\frac{\rho}{M}} \mathbf{H}^c \mathbf{x}_t^c + \mathbf{w}_t^c, \quad t = 1, \dots, T \quad (1)$$

* M. O. Damen is with the Department of Electrical and Computer Engineering, University of Alberta, ECERF W2, Edmonton, AB, T6G 2V4, Canada. E-mails: damen@ee.ualberta.ca

** H. El Gamal is with the Department of Electrical Engineering, the Ohio State University, Columbus, OH. E-mail: helgamal@ee.eng.ohio-state.edu

† G. Caire is with the Mobile Communications Group, Institut EU-RECOM, 2229 Route des Cretes, B.P. 193, 06904 SOPHIA ANTIPO-LIS CEDEX, France. E-mail: giuseppe.caire@eurecom.fr

where $\{\mathbf{x}_t^c \in \mathbb{C}^M : t = 1, \dots, T\}$ is the transmitted signal, $\{\mathbf{y}_t^c \in \mathbb{C}^N : t = 1, \dots, T\}$ is the received signal, $\{\mathbf{w}_t^c \in \mathbb{C}^N : t = 1, \dots, T\}$ denotes the channel Gaussian noise, assumed temporally and spatially white with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$, and \mathbf{H}^c is the $N \times M$ channel matrix with the (i, j) -th element h_{ij}^c representing the fading coefficient between the j -th transmit and the i -th receive antenna. The fading coefficients are further assumed to be i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$ and remain fixed for $t = 1, \dots, T$, where T is the duration of a space-time codeword. With the normalization

$$\mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T |\mathbf{x}_t^c|^2 \right] \leq M,$$

the parameter ρ represents the *average* signal-to-noise ratio (SNR) per receiver antenna. The channel matrix \mathbf{H}^c is assumed to be perfectly known at the receiver and completely unknown at the transmitter. Assume that $N \leq M$ and consider the real channel model equivalent to (1)

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \quad (2)$$

where we define $\mathbf{x} = (\mathbf{x}_1^T, \dots, \mathbf{x}_T^T)^T$ with $\mathbf{x}_t^T = [\text{Re}\{\mathbf{x}_t^c\}^T, \text{Im}\{\mathbf{x}_t^c\}^T]^T$, $\mathbf{w} = (\mathbf{w}_1^T, \dots, \mathbf{w}_T^T)^T$ with $\mathbf{w}_t^T = [\text{Re}\{\mathbf{w}_t^c\}^T, \text{Im}\{\mathbf{w}_t^c\}^T]^T$, and

$$\mathbf{H} \triangleq \sqrt{\frac{\rho}{M}} \mathbf{I}_T \otimes \left(\begin{bmatrix} \text{Re}\{\mathbf{H}^c\} & -\text{Im}\{\mathbf{H}^c\} \\ \text{Im}\{\mathbf{H}^c\} & \text{Re}\{\mathbf{H}^c\} \end{bmatrix} \right) \quad (3)$$

is the $2NT \times 2MT$ block-diagonal real channel matrix consisting of the same $2N \times 2M$ diagonal block repeated T times (\mathbf{I}_T is the identity matrix of dimension T here and \otimes denotes the Kronecker product). Assume further that the vector \mathbf{x} belong to a certain lattice code (i.e., $\mathbf{x} = \mathbf{G}\mathbf{u}$ with \mathbf{G} the $m \times m$ lattice generator matrix, $\mathbf{u} \in \mathbb{Z}^m$ with $m = 2MT$, and the lattice code is given by the lattice points inside a carving region \mathcal{R}). For example, with $\mathbf{G} = \mathbf{I}_m$, $T = 1$, and $\mathbf{G}^{-1}\mathcal{R}$ is a (scaled and shifted) hypercube, we obtain the well-known V-BLAST system with uncoded QAM constellations; for $T = M$ and \mathbf{G} a suitably chosen rotation, one obtains a version of a TAST constellation [7].

III. THE PROPOSED ALGORITHM

Before we outline the proposed algorithm, we briefly discuss our implementation of the MMSE-GDFE filter. Define the augmented channel matrix

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{I}_m \end{bmatrix}$$

and its QR decomposition,

$$\tilde{\mathbf{H}} = \tilde{\mathbf{Q}}\mathbf{R}$$

where $\tilde{\mathbf{Q}} \in \mathbb{R}^{(n+m) \times m}$ has orthonormal columns and $\mathbf{R} \in \mathbb{R}^{m \times m}$ is upper triangular with positive diagonal elements. Further, let $\mathbf{Q} = \mathbf{H}\mathbf{R}^{-1}$ be the upper $n \times m$ part of $\tilde{\mathbf{Q}}$. Then the MMSE-GDFE backward and forward filters are

given by $\mathbf{B} = \mathbf{R}$ and $\mathbf{F} = \mathbf{Q}^T$, respectively. Note that one has $\mathbf{B}^T\mathbf{B} = \tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \mathbf{I}_m + \mathbf{H}^T\mathbf{H}$ by construction, implying that matrix \mathbf{B} is always invertible for any finite SNR.

We are now ready to describe the proposed algorithm in the following steps.

1. Apply the MMSE-GDFE front-end filtering on the received signal to obtain

$$\begin{aligned} \mathbf{y}' &\triangleq \mathbf{F}\mathbf{y} \\ &= \mathbf{B}\mathbf{G}\mathbf{u} + \mathbf{F}\mathbf{w} - [\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{G}\mathbf{u} \\ &= \mathbf{B}\mathbf{G}\mathbf{u} + \mathbf{n} \end{aligned} \quad (4)$$

where \mathbf{n} is the equivalent noise which contains the self noise term $[\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{G}\mathbf{u}$ resulting from MMSE-GDFE stage.

2. Apply the LLL reduction algorithm [10] on matrix $\mathbf{B}\mathbf{G}$ to obtain $\mathbf{S} = \mathbf{B}\mathbf{G}\mathbf{U}$ with \mathbf{U} a unimodular matrix and \mathbf{S} has reduced columns. Alternatively, one can apply other lattice reduction algorithms, such as Korkine-Zolotareff reduction [4], depending on the allowed preprocessing complexity and the given dimension.

3. Order the columns of \mathbf{S} using V-BLAST greedy ordering [3].

4. Apply lattice decoding on \mathbf{S} and \mathbf{y}' using the Schnorr-Euchner enumeration and a finite radius as in [3] to obtain $\hat{\mathbf{u}}$. In this stage, one can incorporate any predefined termination strategy or metric computation inside the Schnorr-Euchner enumeration using the analogy with sequential decoding [3].

5. Finally, the estimated codeword is obtained as $\hat{\mathbf{c}} = \mathbf{G}\mathbf{U}\hat{\mathbf{u}}$.

The power of the MMSE-GDFE front-end is manifested in the fact that (4) is a full rank linear system of m equations and m unknown since \mathbf{B} is always invertible (and has all its eigenvalues larger than 1). Since the noise in (4), $\mathbf{F}\mathbf{w} - [\mathbf{B} - \mathbf{F}\mathbf{H}]\mathbf{G}\mathbf{u}$, is non-Gaussian and biased, minimum Euclidean distance decoding of (4) is different from ML decoding; however, as shown in [3], [8] and illustrated in the simulation results section, minimum Euclidean decoding of (4) yields a performance very close to that of ML decoding. Intuitively, the MMSE-GDFE front-end filtering plays two important roles in decoding MIMO systems, i) neutralize the faded eigenvalues of the wireless channel (i.e., giving a better lattice generator matrix $\mathbf{B}\mathbf{G}$) and ii) ‘‘compress’’ the additive noise (i.e., bringing the preprocessed vector \mathbf{y}' in (4) closer to the lattice points generated by $\mathbf{B}\mathbf{G}$). Furthermore, one can see that the proposed algorithm does not utilize the available information on the carving region \mathcal{R} , and hence, if the output codeword does not belong to the transmitter codebook the receiver will declare an error (This is the reason behind referring to the proposed algorithm as a *lattice decoder*). Of course, one can alternatively project the decoded lattice point $\hat{\mathbf{c}}$ on the transmitter codebook when it falls outside it (when \mathcal{R} has a simple form); however, we have observed that the latter operation gives very marginal improvement compared to simply declaring an error when the decoded lattice point falls outside \mathcal{R} . The reason for not exploiting the information about the carving region is that, in general, the image of the region

\mathcal{R} after applying the LLL algorithm does not enjoy any particular structure. Therefore, accounting for this *distorted* image will entail significant increase in the decoding complexity. Moreover, the loss in performance resulting from this sub-optimality (i.e., lattice decoding) was found to be very marginal in all the cases we considered (as evident in the simulation results in the following section) and the complexity reductions resulting from using the LLL are, in most cases, significant (especially for under-determined systems). Note also that the MMSE-GDFE is only applied on the quasi-static channel matrix \mathbf{H} and not on the lattice generator matrix of the input codebook \mathbf{G} . The fact that \mathbf{H} in (2) is block diagonal with the same block containing the real representation of \mathbf{H}^c repeated T times means that the MMSE-GDFE is only operating on the spatial dimension in this case.

Finally, a brief comment about the termination strategy is in order. It is straightforward to see that the first point found by the Schnorr-Euchner enumeration is the Babai point for the lattice in (4) after applying LLL algorithm (one should note that the Babai point is a function of the generator matrix used to find it). If the search is terminated at this point, the complexity is only linear in the number of variables (assuming the fading channel is very slow such that the overhead of MMSE-GDFE and LLL can be ignored). Recently, Yao and Wornell, and independently Windpassinger and Fischer have proposed a decoding algorithm with a comparable complexity to this linear complexity variant of the proposed scheme (we will refer to this algorithm as “the YWWF decoder” in the sequel). In this algorithm, the output codeword is the solution of zero-forcing decision-feedback equalizer (ZF-DFE), or the Babai point, of the LLL reduction of $\mathbf{H}\mathbf{G}$ in (2) for systems with $m \leq n$ [11], [12]¹ (i.e., the difference between the YWWF decoder and the linear complexity variant of our scheme is the MMSE-GDFE front-end). Unlike our algorithm, one can easily see that the YWWF decoder does not extend to under-determined systems. Moreover, we show in the next section our algorithm, with the MMSE-GDFE front-end, largely outperforms the YWWF decoder while maintaining the linear complexity when $m \leq n$.

IV. PERFORMANCE, COMPLEXITY AND NUMERICAL RESULTS

In this section, we present representative simulation results that illustrate the two main advantages of the proposed algorithm, namely, 1) the proposed class of algorithms offer significant complexity reduction over state of the art decoders with comparable performance (with a small fraction of a dB) and 2) the proposed class of algorithms offer significant performance gains over state of the art decoders with comparable complexity. First, with no termination strategy, we demonstrate that the performance of the proposed decoder is virtually indistinguishable from the ML decoder while offering at least an order of magnitude reduction in complexity compared to the most efficient

implementation of the GSD available in the literature (i.e., Dayal and Varanasi scheme [14]). In this comparison, we will use the same set-up as [14] to ensure fairness. In the second scenario, we demonstrate the excellent performance of the linear complexity variant of the proposed algorithm by comparing it to the YWWF decoder in a V-BLAST configuration. To the best of the authors’ knowledge, the YWWF decoder represents the state of the art performance for receivers with linear complexity.

Following the set-up in [14], Fig. 1 reports the performance of two variants of the threaded algebraic space-time (TAST) constellations in a 3×1 point-to-point MIMO channel. In the rate-1 code, we use a 64-QAM constellation and transmit one symbol per channel use (i.e., the space-time constellation matrices have only with one thread containing information). In the rate-3 code, we use a 4-QAM constellation and transmit three symbols per channel use (i.e., the space-time constellation matrices have three threads containing information). As observed in [14], one obtains a small but sizable performance gain when using rate-3 TAST code in a 3×1 configuration at this SNR’s range. For both variants, we can observe that the performance of the proposed MMSE-GDFE lattice decoder is less than 0.1 dB away from the ML decoder. The main disadvantage of the rate-3 code is that it corresponds to an under-determined system with 6 excess unknowns. Therefore, when using the **same decoder** to decode both rate-1 and rate-3 codes, one can measure the average complexity increase with the excess dimensions by comparing the ratio of both decoding complexities. If we define

$$\gamma = \frac{\text{Average complexity of decoding rate-3 code}}{\text{Average complexity of decoding rate-1 code}}, \quad (5)$$

then a straightforward implementation of the GSD, as outlined in [13], would result in $\gamma = \mathcal{O}(4^6)$. In fact, even with the modification proposed in [14], Dayal and Varanasi could only bring this number down to $\gamma = 460$ at an SNR of 30 dB. In Table IV, we report the values of complexity ratio γ for the proposed algorithm at different SNRs, where one can see the significant reduction in complexity (i.e., from 460 to 12 at an SNR of 30 dB). Also, note that a complexity ratio of $\gamma = 460$ in [14] was obtained as a ratio of the complexity of their generalized sphere decoder for the rate-3 TAST code [14] over that of the sphere decoder for the rate-1 TAST code (which is greater than the complexity of MMSE-GDFE lattice decoder for the rate-1 TAST code used as the denominator of (5)), and therefore, the reduction in complexity obtained by our algorithm compared to the GSD [13], [14] is more than the factor $460/12 = 38.33$ in this scenario. Based on experimental observations, we also expect this gain in complexity reduction to increase with the excess dimension $m - n$. In fact, based on this result, one is tempted to conjecture that the complexity of the proposed algorithm grows polynomially in the number of variables **even in an under-determined** configuration (i.e., a cubic complexity in the number of variables would result in $\gamma = \mathcal{O}(27)$ in this example). If this claim is true, then this will be the first known algorithm with near-

¹In [11], [12], $\mathbf{G} = \mathbf{I}_m$.

TABLE I

AVERAGE COMPLEXITY RATIO OF THE PROPOSED ALGORITHM FOR RATE-3 TAST CODE OVER RATE-1 TAST CODE (WITH THE SAME RATE) IN A 3×1 MIMO SYSTEM

SNR (dB)	22	24	26	28	30
γ	41	31	23	16	12

ML performance and polynomial complexity in an under-determined configuration. Unfortunately, we don't have a rigorous proof for this claim at the moment. Finally, we stress that the near-ML performance of the proposed algorithm, with complete search, was observed in many other scenarios. We, however, opted not to report these results for brevity.

To illustrate our second point, Fig. 2 compares the performance of the linear complexity variant of our algorithm with that of the YWWF decoder [11], [12] in a 4×4 V-BLAST MIMO system with a 4-QAM constellation. From the figure, one can see the 3 dB gain offered by the proposed algorithm compared to the YWWF decoder. In fact, quite surprisingly, the performance of the proposed linear complexity algorithm is within a fraction of a dB from that of ML decoder, whereas the algorithm in [11], [12] is more than 3 dB away from the performance ML detection. The comparison is done for a system with $m \leq n$ because the YWWF decoder does not extend to under-determined configurations. Finally, we only reported numerical results in two scenarios for brevity and the conclusions drawn from these two examples were found to hold in many other scenarios as well.

V. CONCLUSION

In this paper, we have proposed a new decoding algorithm for linear Gaussian channels that combines MMSE-GDFE filtering, lattice reduction techniques with greedy ordering and lattice decoding using Schnorr-Euchner enumeration. The proposed algorithm offers near-ML decoding and extends naturally to under-determined systems (more unknowns than observations). In under-determined systems, the proposed algorithm was shown to offer significant complexity reduction compared to the most efficient implementation of ML decoding available in the literature. By using a dynamic termination strategy for the Schnorr-Euchner enumeration stage, we developed a generalization of the basic MMSE-GDFE lattice decoder that allows for graceful performance-vs-complexity tradeoff. In the lowest complexity implementation, the complexity of this algorithm collapses to a linear function in the number of variables (assuming very slow fading). This linear complexity variant was shown to significantly outperform the YWWF linear complexity decoder in a V-BLAST configuration. Although our focus was devoted to a point-to-point MIMO configuration, it is straightforward to see that the proposed algorithm can be utilized in other interesting scenarios such as an overloaded multiuser code division multiple access

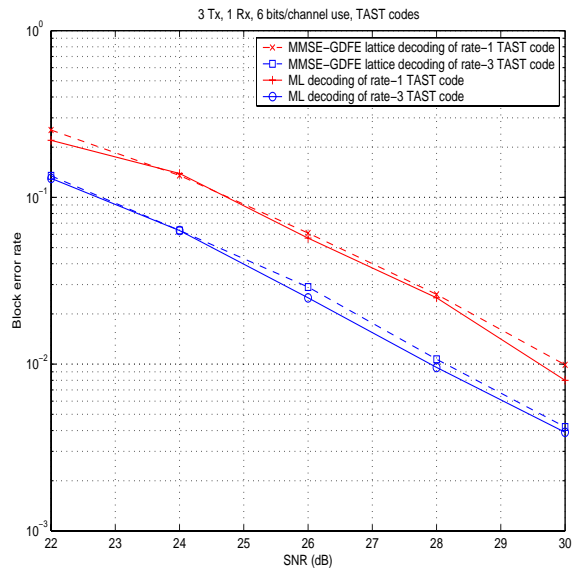


Fig. 1. Performance of TAST codes under MMSE-GDFE lattice decoding and ML detection with $M = 3$ and $N = 1$.

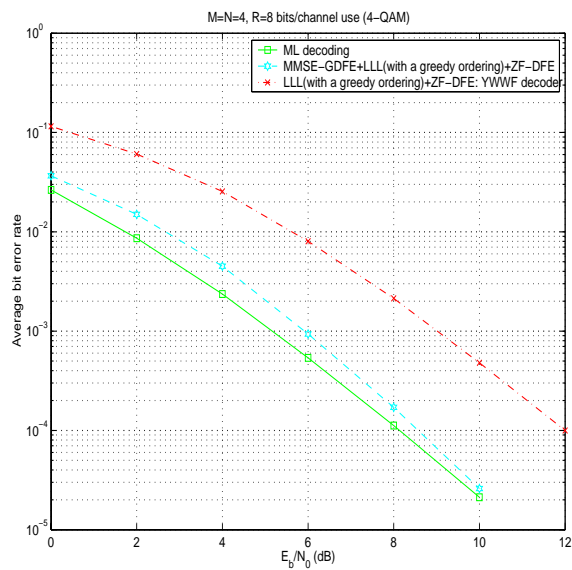


Fig. 2. Performance of linear complexity decoding algorithms in a 4×4 V-BLAST system with a 4-QAM constellation.

(CDMA) system, for example.

One of the most interesting venue for future work is to pursue a theoretical complexity analysis of the proposed algorithm. This analysis is now well motivated by the experimental results which suggest that the average complexity of the proposed algorithm only grows polynomially with the number of variables for a wide range of SNRs, m and n even in an under-determined configuration while maintaining near-optimal performances. If this claim holds against the test of the theoretical analysis, the proposed algorithm will prove to be the first known near-optimal decoding technique that enjoys polynomial complexity in under-determined linear Gaussian channels.

REFERENCES

- [1] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, pp. 463–471, Apr. 1985.
- [2] C. P. Schnorr and M. Euchner, "Lattice basis reduction: improved practical algorithms and solving subset sum problems," *Mathematical Programming*, Vol. 66, pp. 181-191, 1994.
- [3] M. O. Damen, H. El Gamal, and G. Caire, "On maximum likelihood decoding and the search of the closest lattice point" *IEEE Trans. Info. Theory*, vol. 49, pp. 2389-2401, Oct. 2003.
- [4] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, Aug. 2002.
- [5] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice codes decoder for space-time codes," *Commun. Letters*, vol. 4, pp. 161–163, May 2000.
- [6] B. Hassibi and B. Hochwald, "High rate codes that are linear in space and time," *IEEE Trans. Info. Theory*, vol. 48, pp. 1804–824, July 2002.
- [7] H. El Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Info. Theory*, vol. 49, pp. 1097-1119, May 2003.
- [8] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-vs-multiplexing tradeoff of MIMO channels," submitted to *IEEE Trans. Inform. Theory*, Nov. 2003.
- [9] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Info. Theory*, vol. 49, pp. 1073–1096, May 2003.
- [10] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.*, Vol. 261, pp. 515–534, 1982.
- [11] H. Yao and G. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Conf. on Commun.*, Taipei, Taiwan, Nov. 2002.
- [12] C. Windpassinger and R.F.H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Inform. Theory Workshop*, Paris, France, Mar. 2003.
- [13] M. O. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Generalized sphere decoder for asymmetrical space-time communication architecture," *Electron. Lett.*, Vol. 36, p. 166, Jan. 2000.
- [14] P. Dayal and M.K. Varanasi, "A fast generalized sphere decoder for optimum decoding of under-determined MIMO systems," in *Proc. 41th Annual Allerton Conf. on Comm. Control, and Comput.*, Monticello, IL, Oct., 2003.