

Lattice-Reduction-Aided Broadcast Precoding

Christoph Windpassinger, Robert F. H. Fischer, Johannes B. Huber

Lehrstuhl für Informationsübertragung, Universität Erlangen–Nürnberg,

Cauerstraße 7/LIT, 91058 Erlangen, Germany, <http://www.LNT.de>, email: {windpass,fischer,huber}@LNT.de

Abstract

We describe a low-complexity lattice-reduction-aided precoding scheme for multiuser broadcast scenarios. This scheme fills the gap between low-complexity Tomlinson-Harashima precoding and the sphere-decoder-based system of Peel et al. [1]. Simulation results show that, replacing the closest point search with the Babai approximation [2], still each user will experience the full diversity order supported by the system.

1 Introduction

Recently, the multiuser broadcast precoding problem has received considerable attention. New information-theoretic results on the sum capacity of the multiuser broadcast channel, e.g., [3], [4], [5] have shown that some kind of Costa- or Tomlinson-Harashima precoding [6], [7], [8], [9] is necessary to attain it. Peel et al. [1] have recently introduced a “vector perturbation technique” and showed that the uncoded error rate curves thus obtained exhibit the full diversity order of the system. The key idea was already present in shaping without scrambling [10], [11], which is based on a successive processing to efficiently find the perturbation vector.

However, the technique in [1] requires the use of the rather complex sphere-decoder [12] in order to solve a lattice closest-point problem, and this technique can be viewed as some kind of “maximum-likelihood detection at the transmitter”.

In the present work we consider the use of Babai’s approximate closest-point solution [2] to come up with a much less complex precoding technique, along the lines of [13], [14]. This approximate solution relies on the lattice basis reduction algorithm of A. K. Lenstra, H. W. Lenstra and L. Lovász (LLL) [15], which is sub-optimum but very efficient.

It turns out that while there is some loss in power efficiency with respect to the sphere-decoder based precoding technique, the full diversity is also present with the approximate solution, leading to significant gains in uncoded error rate for high signal-to-noise ratios.

Note that the precoding schemes considered in this work are optimized under the constraint that a standard scalar modulo receiver frontend, as in Tomlinson-Harashima precoding, is used (similar to “shaping without scrambling”), whereas precoding systems based on higher-dimensional lattices, aiming for the 1.53 dB shaping gain, require the use of higher-dimensional lattice quantizers at the receiver, cf. [9].

The paper is structured as follows: In Section 2 we introduce the transmission model. The precoding method

described by Peel et al. [1] is discussed in Section 3. Section 4 shows how precoding is performed using Babai’s approximate solution to the closest-vector problem, and simulation results follow in Section 5. Some concluding remarks are offered in Section 6.

2 Transmission Model and Conventional Precoding

We consider the transmission from a base station (BS) with N_T transmit antennas to $K \leq N_T$ users (mobile stations, MS) with a single receive antenna each, as shown in Fig. 1. The transmission channel is assumed to be frequency-flat, and the received signal of user k is modeled in the equivalent complex baseband as

$$y_k = \sum_{l=1}^{N_T} h_{k,l} x_l + n_k, \quad (1)$$

where $x_l \in \mathbb{C}$ is the signal on the l th transmit antenna, and $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ the noise for the k th user.

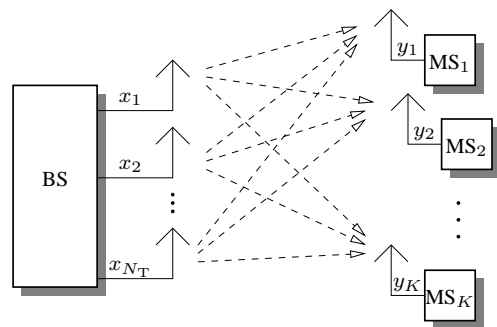


Fig. 1. Illustration of the broadcast setup considered in this work.

We use the vector/matrix notation

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

with $\mathbf{x} = [x_1, \dots, x_{N_T}]^T$, channel matrix $\mathbf{H} = [h_{k,l}]$, $\mathbf{n} = [n_1, \dots, n_K]^T$ and $\mathbf{y} = [y_1, \dots, y_K]^T$, where we

keep in mind that receiver processing is limited to individual components of \mathbf{y} (only diagonal matrix operations are allowed).

A necessary prerequisite for performing any kind of preequalization is channel state information at the transmitter. Here we assume the current realization of \mathbf{H} to be perfectly known to the transmitter.

For convenience we also use the following real-valued notation:

$$\begin{bmatrix} \Re \mathbf{y} \\ \Im \mathbf{y} \end{bmatrix} = \begin{bmatrix} \Re \mathbf{H} & -\Im \mathbf{H} \\ \Im \mathbf{H} & \Re \mathbf{H} \end{bmatrix} \begin{bmatrix} \Re \mathbf{x} \\ \Im \mathbf{x} \end{bmatrix} + \begin{bmatrix} \Re \mathbf{n} \\ \Im \mathbf{n} \end{bmatrix}, \quad (3)$$

(where the \Re and \Im prefix denote the real and imaginary parts), and we use the subscript $(\cdot)_r$ to denote the vectors and matrices obtained by this separation of real and imaginary parts, obtaining an equivalent $2K$ -dimensional real model of the form

$$\mathbf{y}_r = \mathbf{H}_r \mathbf{x}_r + \mathbf{n}_r. \quad (4)$$

The data symbols to be transmitted to the K users will be denoted as $\mathbf{a} = [a_1, \dots, a_K]^T$, chosen from an M -ary square QAM constellation. E.g., for $M=4$, $a_k \in \{\pm \frac{1}{2} \pm j \frac{1}{2}\}$, or equivalently $a_{r,k} \in \{\pm \frac{1}{2}\}$ ($k = 1, \dots, 2K$).

The most obvious method to perform precoding for this setup is to select $\mathbf{x} = \mathbf{H}^+ \mathbf{a}$, where \mathbf{H}^+ is the right pseudo-inverse of \mathbf{H} , i.e., $\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1}$. In this case, the receiver simply quantizes y_k to the M -ary QAM constellation to recover a_k .

A more power-efficient precoding method is Tomlinson-Harashima precoding (cf., e.g., [16]). This method employs modulo-arithmetics in the precoding stage, and requires a modulo-operation at the receiver before quantizing to the QAM constellation. It is based on a QR-type decomposition of the channel matrix \mathbf{H}_r , and its performance can be increased by using the V-BLAST algorithm [17] to optimize the ordering of the subchannels. Moreover, operating on the real-valued model instead of the complex-valued model is advantageous [18].

However, since all of these schemes perform linear preequalization for at least one of the subchannels, for $K \times N_T$ Rayleigh-fading channels, the average bit error rate curve will show diversity order $N_T - K + 1$, and in particular for $N_T = K$, the diversity order of a *single* Rayleigh-fading channel.

3 Search-Based Broadcast Precoding

In [1] a “vector perturbation” technique was described that is based on the following observation, which is also the basis of shaping without scrambling (cf. the block diagram in Fig. 2) [10], [11]:

Using a modulo-operation at the receiver, i.e.,

$$\tilde{y}_{r,k} = y_{r,k} \bmod A \stackrel{\text{def}}{=} y_{r,k} - A \lfloor (y_{r,k} + \frac{A}{2}) / A \rfloor, \quad (5)$$

where A is chosen such that the points from the signaling constellation can be uniquely recovered from $\tilde{y}_{r,k}$, we consider the “optimum” transmit signal, i.e., the signal which requires minimum transmit power if it is followed by linear preequalization. This is given by

$$\mathbf{x}_r = \mathbf{H}_r^+ (\mathbf{a}_r + \mathbf{p}), \quad (6)$$

with \mathbf{p} chosen such that its influence is eliminated by the modulo frontend, hence $\mathbf{p} \in A\mathbb{Z}^{2K}$, and therefore

$$\mathbf{p} = \arg \min_{\mathbf{p}' \in A\mathbb{Z}^{2K}} \|\mathbf{H}_r^+ (\mathbf{a}_r + \mathbf{p}')\|^2. \quad (7)$$

Thus, instead of linearly preequalizing \mathbf{a}_r , the symbols $\mathbf{a}_r + \mathbf{p}'$ drawn from a virtually periodically extended constellation are used. For unique recoverability, any $A > 2 \max |a_{r,k}|$ is sufficient, e.g., for the 4-QAM constellation $\{\pm \frac{1}{2} \pm j \frac{1}{2}\}$ we can take any $A > 1$. However, it can be shown that the optimum is $A = 2$, and symmetric error regions and a periodic extension of the original signaling constellation results.

At the receiver, $\mathbf{y}_r = \mathbf{H}_r \mathbf{H}_r^+ (\mathbf{a}_r + \mathbf{p})$, and with the modulo frontend, for each $k = 1, \dots, 2K$ (real representation),

$$\tilde{y}_{r,k} = y_{r,k} \bmod A \quad (8)$$

$$= (a_{r,k} + p_k + n_{r,k}) \bmod A \quad (9)$$

$$= (a_{r,k} + n_{r,k}) \bmod A \quad (10)$$

is observed.

We can write the minimization as

$$\begin{aligned} \min_{\mathbf{p}' \in A\mathbb{Z}^{2K}} \|\mathbf{H}_r^+ (\mathbf{a}_r + \mathbf{p}')\|^2 = \\ \min_{\mathbf{p}' \in A\mathbb{Z}^{2K}} \|\mathbf{H}_r^+ \mathbf{a}_r + \mathbf{H}_r^+ \mathbf{p}'\|^2, \quad (11) \end{aligned}$$

and hence \mathbf{p}' contains the coordinates of the point in the lattice $A\mathbf{H}_r^+ \mathbb{Z}^{2K}$ closest to $-\mathbf{H}_r^+ \mathbf{a}_r$.

Since the search space for this closest point problem has a finite number of dimensions ($2K$), unlike in the applications for which shaping without scrambling was originally considered, the full search can be effectively performed using standard lattice decoding techniques, e.g., the sphere decoder [12], cf. also [11]. The corresponding block diagram is shown in Fig. 3.

4 Lattice-Reduction-Aided Broadcast Precoding

While it has been shown that the average complexity of the sphere decoder is not as bad as the worst case complexity (exponential in the number of dimensions) suggests [19], it is still quite high compared to both linear preequalization, which merely requires multiplication of the transmit data vector with the inverse channel matrix, and Tomlinson-Harashima precoding, which

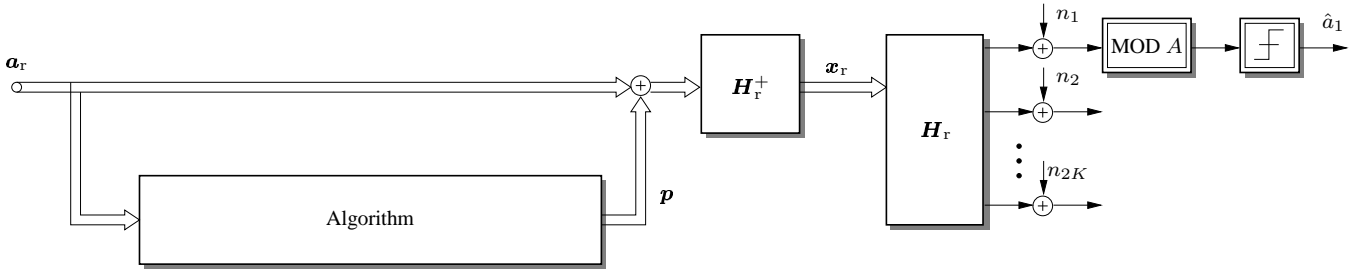


Fig. 2. Basic block diagram of shaping without scrambling. Only one receiver processing path is shown.

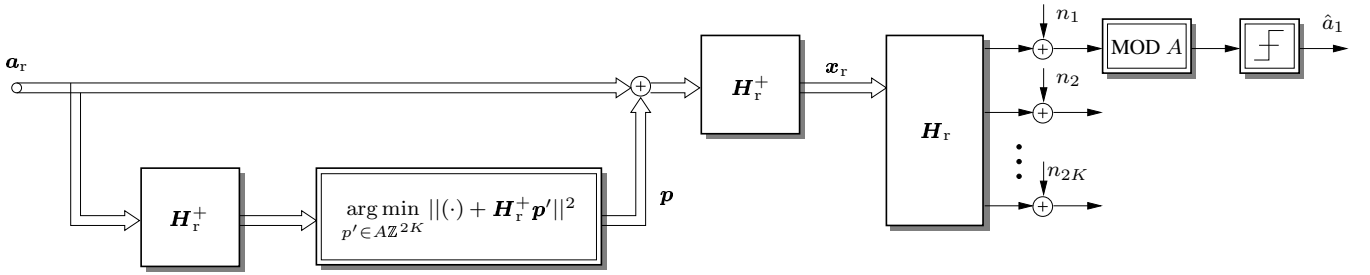


Fig. 3. Illustration of the search-based broadcast precoding scheme by Peel et al. Only one receiver processing path is shown.

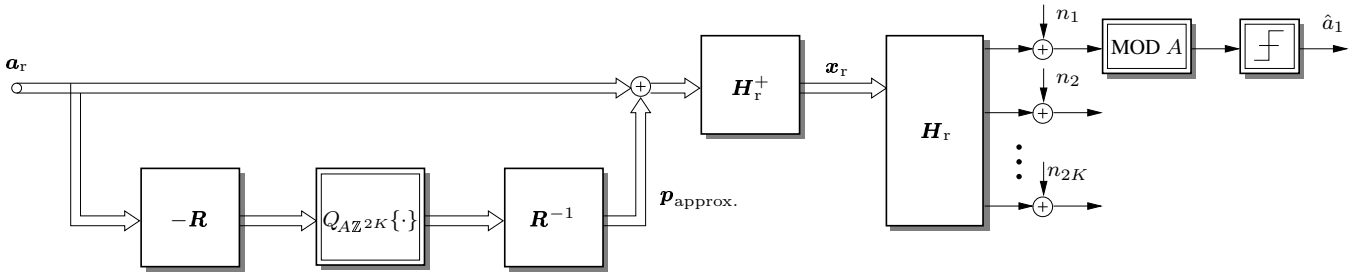


Fig. 4. Illustration of the rounding-off approximation broadcast precoding scheme. Only one receiver processing path is shown.

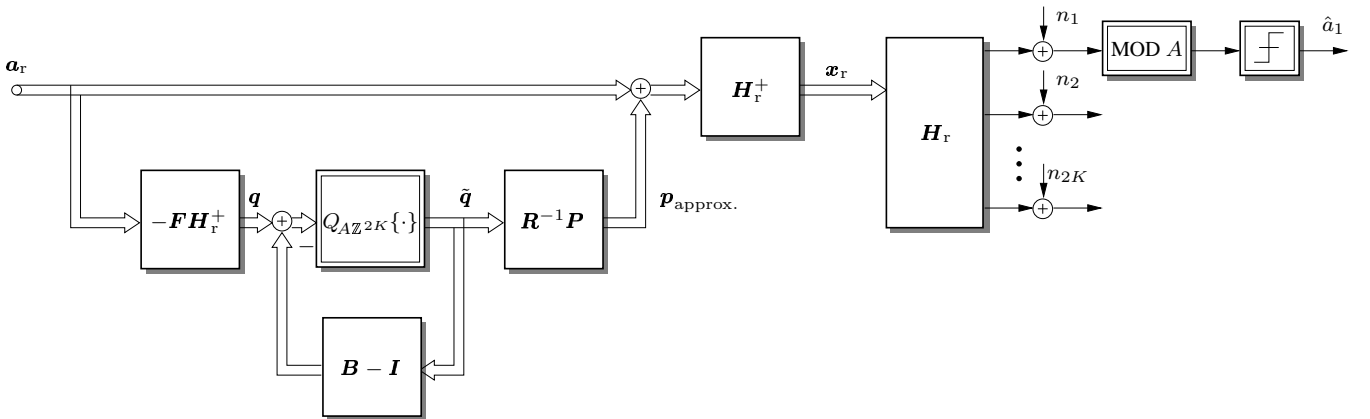


Fig. 5. Illustration of the nearest-plane approximation broadcast precoding scheme. Only one receiver processing path is shown.

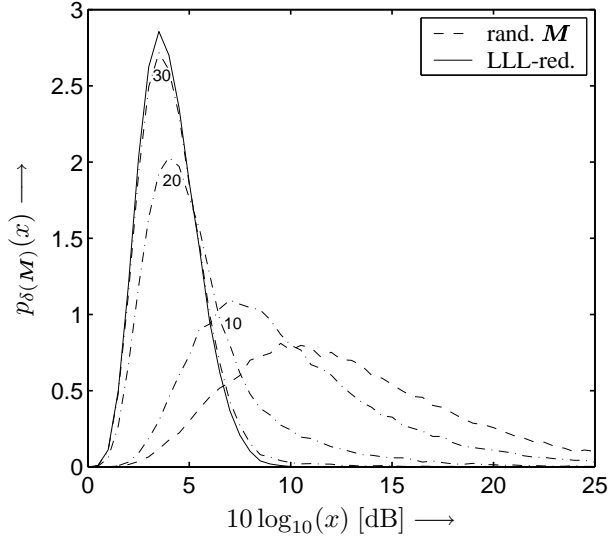


Fig. 6. Pdf of orthogonality defect of random 4×4 matrices and their LLL-reduced counterparts (restricting the LLL main loop to 10, 20, 30 iterations).

is based on two subsequent matrix operations (and the modulo-operation), and which is thus only marginally more complex than linear preequalization.

Here we suggest to use the closest point approximation [2], similar to [13], [14], to obtain a simple but efficient method for broadcast precoding:

Starting from \mathbf{a}_r and \mathbf{H}_r , we use the LLL algorithm [15] on the columns of \mathbf{H}_r^+ to obtain

$$\mathbf{H}_r^+ = \mathbf{W}\mathbf{R}, \quad (12)$$

where $\mathbf{W} \in \mathbb{R}^{2N_T \times 2K}$ is the LLL-reduced basis with approximately orthogonal columns, and $\mathbf{R} \in \text{GL}_{2K}(\mathbb{Z})$, i.e., an integer matrix with $|\det(\mathbf{R})| = 1$, describes this transform. This algorithm has polynomial complexity, and to illustrate its effectiveness we have plotted the pdfs of the orthogonality defect¹ of random matrices of size 4×4 with complex-Gaussian distributed entries in Figure 6, together with those of the corresponding LLL-reduced matrices. In addition we show the orthogonality defects achieved if the LLL-algorithm (which is an iterative algorithm) is aborted after 10, 20, 30 iterations of its main loop (i.e., if we want to keep the preprocessing time limited to a small constant value). Thus, particularly matrices with large orthogonality defects can be avoided, and less noise enhancement is suffered by linear equalization based on these LLL-reduced matrices (if the matrices could be made truly orthogonal, no noise enhancement would result at all).

If we take the ‘‘rounding-off’’ approximation from [2], the solution of (7) is given by

$$\mathbf{p}_{\text{approx.}} = -\mathbf{R}^{-1}Q_{AZ^{2K}}\{\mathbf{R}\mathbf{a}_r\}, \quad (13)$$

¹Orthogonality defect of matrix $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_K]$ with columns \mathbf{m}_k :

$\delta(\mathbf{M}) \stackrel{\text{def}}{=} \frac{\prod_{k=1}^K \|\mathbf{m}_k\|}{|\det(\mathbf{M})|}$, with $\delta(\mathbf{M}) \geq 1$, and $\delta(\mathbf{M}) = 1 \Leftrightarrow \mathbf{M}$ orthogonal

where we have used $Q_{AZ^{2K}}\{\cdot\}$ to denote component-wise rounding of a $2K$ -dimensional vector to the scaled integer lattice AZ^{2K} .

Consequently, the transmit signal is given as

$$\mathbf{x}_r = \mathbf{H}_r^+(\mathbf{a}_r - \mathbf{R}^{-1}Q_{AZ^{2K}}\{\mathbf{R}\mathbf{a}_r\}). \quad (14)$$

The simple structure of this scheme is illustrated by the block diagram in Fig. 4.

We also consider a variant of the nearest plane algorithm [2] for the solution of the closest point problem. This approximation is identical to decision-feedback equalization, i.e., consists of a successive quantization taking into account previous quantized values.

From the V-BLAST algorithm [17] applied to \mathbf{W} obtained from the LLL algorithm as above, we get

$$\mathbf{FWP} = \mathbf{B}. \quad (15)$$

Here $\mathbf{B} = [b_{k,l}] \in \mathbb{R}^{2K \times 2K}$ is a lower triangular matrix with unit diagonal ($b_{k,l} = 1$ for $k = l$ and $b_{k,l} = 0$ for $k < l$), $\mathbf{F} \in \mathbb{R}^{2K \times 2N_T}$ a matrix with orthogonal rows and \mathbf{P} a $2K \times 2K$ permutation matrix corresponding to the optimized decision order. (If the decision order is not optimized, i.e., $\mathbf{P} = \mathbf{I}$, these matrices can be obtained from a QL-decomposition of \mathbf{W}). The algorithm sets

$$\mathbf{q} = [q_1, \dots, q_{2K}]^T = -\mathbf{F}\mathbf{H}_r^+\mathbf{a}_r \quad (16)$$

$$\tilde{q}_1 = q_1 \quad (17)$$

and calculates for $k = 2, \dots, 2K$

$$\tilde{q}_k = Q_{AZ}\{q_k - \sum_{l=1}^{k-1} b_{k,l}\tilde{q}_l\}. \quad (18)$$

(Note that $Q_{AZ}\{x\} = A\lfloor \frac{1}{A}(x + \frac{1}{2}) \rfloor$). Finally, we obtain

$$\mathbf{p}_{\text{approx.}} = \mathbf{R}^{-1}\mathbf{P}\tilde{\mathbf{q}}. \quad (19)$$

The straightforward implementation of this scheme is shown in Fig. 5. Except for the upfront calculation of the LLL reduced basis \mathbf{W} , the complexity is similar to that of, e.g., Tomlinson-Harashima precoding.

It is worth noting that while in the detection case the effect of neglecting the boundary region of the constellation has a negative impact on the performance [14], [20], all lattice points are equally valid in this situation.

5 Simulation Results

We now present simulation results that show the performance of the various schemes described above. We assume $h_{k,l} \sim \mathcal{CN}(0, 1)$, i.e., N_T independent Rayleigh fading channels from the base station antennas to each of the user’s receive antenna.

In Fig. 7 bit error rate (BER) curves over the average received energy per information bit \bar{E}_b divided by the one-sided noise power spectral density N_0 for a system with $N_T = 4$ and $K = 4$ are shown, where 4-QAM signals (left plot) and 16-QAM signals (right plot)

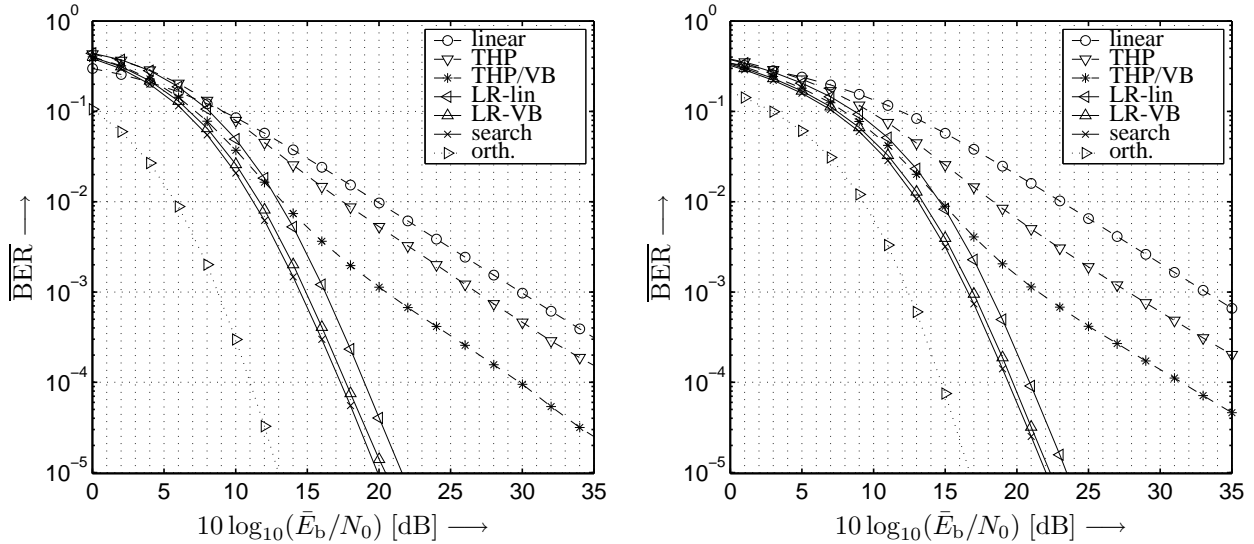


Fig. 7. Simulation results for the different broadcast precoding schemes in a $N_T = 4, K = 4$ system using 4-QAM (left) and 16-QAM (right).

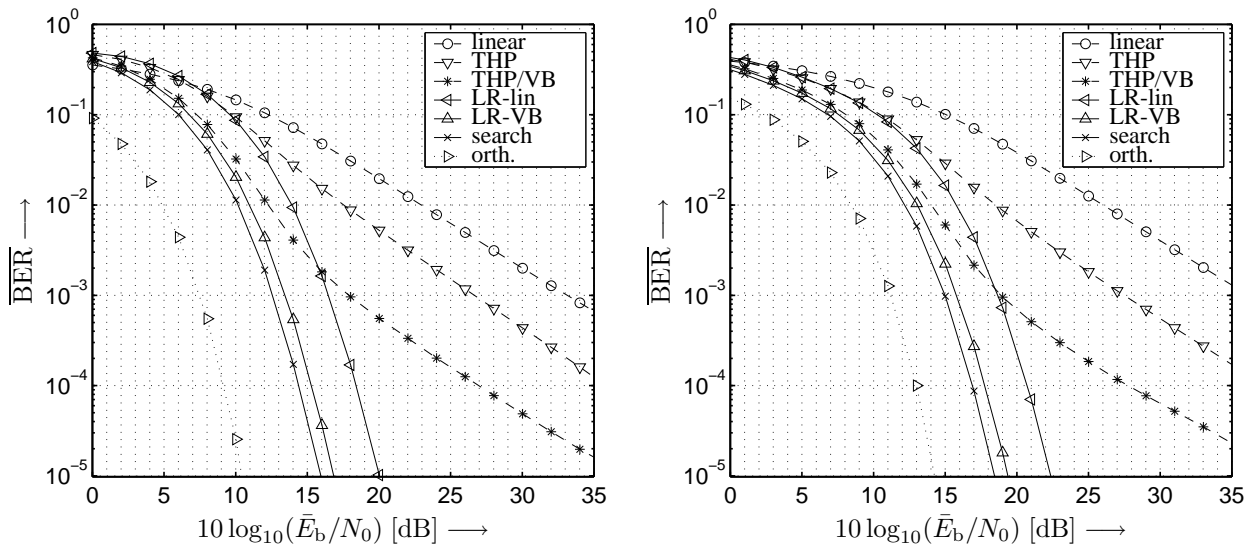


Fig. 8. Simulation results for the different broadcast precoding schemes in a $N_T = 8, K = 8$ system using 4-QAM (left) and 16-QAM (right).

are transmitted to all of the users independently. All precoding methods were normalized for constant transmit power. To show the diversity inherent in this type of system we have also included the error rate curve for a similar system where the signals do not interfere (“orth.”). The worst performance of all precoding schemes considered is achieved by linear pre-equalization (“linear”); straightforward Tomlinson-Harashima precoding (“THP”) works slightly better (in the very-low-SNR region linear pre-equalization benefits from the absence of the modulo frontend), and Tomlinson-Harashima precoding based on the V-BLAST permutation of the subchannels (“THP/VB”) improves still some more. The error rate curves for all three of these techniques exhibit the diversity order 1, however.

The result obtained from the full search using the sphere decoder (“search”) improves significantly over linear pre-equalization as well as Tomlinson-Harashima

precoding, particularly exhibiting the full diversity order of 4.

Strikingly, both approximation-based schemes, lattice-reduction-aided precoding with linear (“LR-lin”) and V-BLAST nearest-plane (“LR-VB”), also exhibit the full diversity order 4, and in particular the nearest-plane approximation shows only little loss in power efficiency with respect to the full search, at significantly lower complexity.

The same holds true for a system with $N_T = 8$ and $K = 8$, shown in Fig. 8, where the V-BLAST based Tomlinson-Harashima precoding system improves quite substantially upon the non-permutation optimized version, but again eventually settles to diversity order 1. Even for this larger system the loss of the suboptimum V-BLAST nearest-plane (“LR-VB”) precoding compared to the full search (“search”) is not substantial.

6 Conclusions

The simulations conducted show that for system dimensions of practical interest, the expensive search for the precoding symbol in [1] can be avoided using the Babai approximation, at only little cost in performance.

The uncoded error rate of the resulting system improves significantly over linear preequalization as well as Tomlinson-Harashima precoding, and exhibits the full diversity offered by the communication channel, at the same time the processing overhead is similar to that of Tomlinson-Harashima precoding.

The LLL basis reduction required for Babai's closest point approximation is necessary only once per block, and for reasonable block sizes introduces merely a negligible additional overhead. Altogether the computational structure of the resulting precoder is straightforward and simple to implement.

Acknowledgements

This work was supported by the Fraunhofer Institut für Nachrichtentechnik / Heinrich Hertz Institut, Berlin, BMBF grant 01BU150.

References

- [1] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst. A vector-perturbation technique for near-capacity multi-antenna multi-user communication. *Submitted to IEEE Transactions on Communications*, June 2003. Available at <http://mars.bell-labs.com/papers/mod-precoding/>.
- [2] L. Babai. On Lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6(1):1–13, 1986.
- [3] W. Yu, J. M. Cioffi. Sum capacity of Gaussian vector broadcast channels. Submitted to *IEEE Trans. Information Theory*, Nov. 2001.
- [4] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates and sum-capacity of Gaussian MIMO broadcast channels. *IEEE Transactions on Information Theory*, 49(10):2658–2668, October 2003.
- [5] P. Viswanath and D. Tse. Sum capacity of the multiple antenna Gaussian broadcast channel and uplink-downlink duality. *IEEE Transactions on Information Theory*, 49(8):1912–1921, August 2003.
- [6] M. H. M. Costa. Writing on dirty paper. *IEEE Transactions on Information Theory*, 29(3):439–441, May 1983.
- [7] U. Erez, S. Shamai, and R. Zamir. Capacity and lattice-strategies for cancelling known interference. In *Proc. ISITA*, Honolulu, Hawaii, Nov. 2000.
- [8] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber. MIMO Precoding for Decentralized Receivers. In *Proc. IEEE ISIT 2002*, p. 496, Lausanne, Switzerland June/July 2002.
- [9] R. F. H. Fischer, C. Windpassinger, and J. B. Huber. Modulo-Lattice Reduction in Precoding Schemes. In *Proc. IEEE ISIT 2003*, Yokohama, Japan, June 2003.
- [10] R. F. H. Fischer, W. H. Gerstaecker, and J. B. Huber. Dynamics Limited Precoding, Shaping, and Blind Equalization for Fast Digital Transmission over Twisted Pair Lines. *IEEE JSAC*, pp. 1622–1633, Dec. 1995.
- [11] R. F. H. Fischer, C. Stierstorfer and C. Windpassinger. Precoding and Signal Shaping for Transmission over MIMO Channels. In *Proc. 2003 Canadian Workshop on Information Theory*, pp. 83–87, Waterloo, Ontario, Canada, May 2003.
- [12] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest point search in lattices. *IEEE Transactions on Information Theory*, 48(8):2201–2214, Aug. 2002.
- [13] H. Yao, G. W. Wornell. Lattice-Reduction-Aided Detectors for MIMO Communication Systems, in *Proceedings of IEEE Globecom 2002*, Taipei, Taiwan, November 2002.
- [14] C. Windpassinger, R. F. H. Fischer. Low-Complexity Near-Maximum-Likelihood Detection and Precoding for MIMO Systems using Lattice Reduction, in *Proceedings of IEEE ITW 2003*, Paris, France, March 2003.
- [15] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients, *Math. Ann.*, 261:515–534, 1982.
- [16] R. F. H. Fischer. *Precoding and Signal Shaping for Digital Transmission*, John Wiley & Sons, 2002.
- [17] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky. Simplified processing for high spectral efficiency wireless communication employing multi-element arrays. *IEEE Journal on Selected Areas in Communications*, 17:1841–1852, November 1999.
- [18] R. F. H. Fischer and C. Windpassinger. Real- vs. complex-valued equalization in V-BLAST systems. *Electronics Letters*, 39(5):470–471, March 2003.
- [19] B. Hassibi, H. Vikalo. On the expected complexity of integer least-squares problems, in *Proceedings of IEEE ICASSP 2002*, Orlando, Florida, USA, May 2002.
- [20] C. Windpassinger, L. Lampe, and R. F. H. Fischer. From Lattice-Reduction-Aided Detection Towards Maximum-Likelihood Detection in MIMO Systems, in *Proceedings of WOC 2003*, Banff, Canada, July 2003.