A SPACE-TIME CODING APPROACH FOR SYSTEMS EMPLOYING FOUR TRANSMIT ANTENNAS

Constantinos B. Papadias *Gerard J. Foschini*

Wireless Communications Research Department
Bell Laboratories, Lucent Technologies
791 Holmdel-Keyport Rd., Holmdel, NJ 07733, USA
{papadias,foschini}@bell-labs.com

ABSTRACT

We propose a novel transmit diversity scheme for the downlink of communication systems that employ four transmit antennas. The scheme can be seen as a simplified transmission architecture admitting an entire family of space-time codes. It can be also seen as an extension of previously proposed techniques (such as those presented in [1] [2] [3]) to the case of four transmit antennas with complex input data symbols. Our technique has a number of appealing features, namely, it enables a significant portion of the open-loop channel capacity, it requires simple receiver processing (involving typically 2×2 matrix operations in conjunction with single-user or 2-user decoding) and it admits single-user encoding in an overlay fashion.

1. INTRODUCTION

Recent advances in the field of array processing have made it clear that significant gains can be realized in wireless systems with the help of multiple antennas. Denoting by (M, N) a wireless system with M transmit and N receive antenna elements, its Shannon capacity is given by the so-called "log-det" formula [4] – provided that no feedback is sent back to the transmitter. The Bell-labs LAyered Space Time (BLAST) architecture first proposed in [4] is a transceiver super-structure that allows to approach the (M, N) open-loop capacity. In order to achieve the high rates promised by the log-det formula however, multi-dimensional error correction codes are required. Such multi-dimensional coding procedures are generally referred to as "space-time codes (STC's)" (see [2]). A diagonal coding structure (known as D-BLAST) was presented in [4] that is designed in a way to avoid the exponential explosion of complexity with the number of spatial dimensions. Nonetheless, in practice, this scheme is still quite complex and there is great interest in investigating the construction of much simpler alternatives for practical use.

In designing STC's, one appealing approach is that of treating the temporal and spatial dimensions of the code in a *decoupled* fashion. These codes typically operate as follows. The input data stream is first demultiplexed into a number of independent data streams (which we call "sub-streams"). Each of these sub-streams is then separately encoded in time, such as in any spatially single-dimension error-correction coding scheme, independently of the other sub-streams. The encoded sub-streams are then spatially multiplexed on the elements of the transmitter antenna array. By allowing the temporal and spatial parts of the code to be separated, these STC's admit any single-user coding scheme designed for operation over an additive white Gaussian noise (AWGN) channel (such as convolutional, trellis, or Turbo codes). On the other hand, this modularity may result in sacrificing upfront a portion of the total multi-antenna channel capacity.

Some transmitter architectures that allow implementation of decoupled STC's have already been proposed, particularly for the (2,1) case. Alamouti suggested in [1] an open-loop (2,1) technique for TDMA systems. As shown in [5], despite its decoupled nature, this scheme allows the attainment of the (2,1) open-loop capacity. By the phrase "allows the attainment of" we mean that through the use of progressively stronger spatially one-dimensional error correction codes on each of the data sub-streams, the system's data rate can tend toward the (2,1) log-det capacity. A variant of this scheme for the CDMA downlink called Space-Time Spreading (STS) [3] was recently introduced as an optional transmission mode in the cdma-2000 (3GPP2) standard. A similar scheme called space-time transmit diversity (STTD) was introduced in the UMTS W-CDMA standard (3GPP1).

Unfortunately, the (2,1) case seems to be, so far, the only one for which there is an open-loop decoupled space-time code which enables the attainment of the full open-loop capacity. As future systems are likely to employ more than two antennas per sector at the base-station, it is of interest to study STC schemes beyond the (2,1) configuration. In the following, we focus on the (4,1) case as one of immediate interest.

2. A (4,1) SCHEME

We will now describe a way to spatially multiplex / spread and transmit four data sub-streams over four transmit antennas. As described above, the transmitted information sequence $b(t)$ is first demultiplexed into four sub-streams $b_i(t)$ $(i = 1, \ldots, 4)$. The sub-streams $\{b_i\}$, each of variance σ_b^2, are assumed to be either left uncoded or separately encoded (whenever we wish to emphasize that they are encoded, we will denote them by $\{\widetilde{b}_i\}$). They are then spread by a total of four "spreading codes" [1], $\mathbf{c}_l$, $(l = 1, 2, 3, 4)$, which are assumed orthonormal, i.e.

$$\mathbf{C}^H\mathbf{C} = \mathbf{I}$$

where $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_4]$ and ^{H} denotes Hermitian transpose. Each code $\mathbf{c}_l$ spans four symbol periods and is represented here as a vertical vector of length Q chips. The baseband signal transmitted from the m-th antenna is denoted by $s_m(t)$ $(m = 1, \ldots, 4)$, whereas its chip-sampled version is denoted by the $Q \times 1$ vector $\mathbf{s}_m$, m=$1, \ldots, 4$. We propose to transmit the four sub-streams in the following fashion from the four Tx antennas:

$$\mathbf{S} = [\mathbf{s}_1\ \mathbf{s}_2\ \mathbf{s}_3\ \mathbf{s}_4] = \begin{bmatrix} b_1\mathbf{c}_1^T + b_2^*\mathbf{c}_2^T + b_3\mathbf{c}_3^T + b_4^*\mathbf{c}_4^T \\ b_2\mathbf{c}_1^T - b_1^*\mathbf{c}_2^T - b_4\mathbf{c}_3^T + b_3^*\mathbf{c}_4^T \\ b_3\mathbf{c}_1^T - b_4^*\mathbf{c}_2^T - b_1\mathbf{c}_3^T - b_2^*\mathbf{c}_4^T \\ b_4\mathbf{c}_1^T - b_3^*\mathbf{c}_2^T + b_2\mathbf{c}_3^T - b_1^*\mathbf{c}_4^T \end{bmatrix}^T$$
$$= \mathbf{C}\begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ b_2^* & -b_1^* & b_4^* & -b_3^* \\ b_3 & -b_4 & -b_1 & b_2 \\ b_4^* & b_3^* & -b_2^* & -b_1^* \end{bmatrix} = \mathbf{CB} \tag{1}$$

where T denotes transpose. In Eq. (1) we have dropped the time index for convenience, so it refers to any given quartet of symbol periods. In observing matrix $\mathbf{B}$, one could think of its vertical dimension as representing "time" and of its horizontal dimension as representing "space". We assume that the channel between the m-th transmit antenna and the receiver antenna is flat-faded, it will hence be represented by a complex scalar h_m. Denoting by $\mathbf{n}$ the $Q \times 1$ additive noise vector at the receiver (assumed to have independent AWGN entries), the received signal is expressed as

$$\mathbf{r} = \mathbf{S}\begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} + \mathbf{n} = \mathbf{CBh} + \mathbf{n} \tag{2}$$

After despreading, we obtain

$$\widetilde{\mathbf{r}} = \mathbf{C}^H\mathbf{r} = \mathbf{Bh} + \mathbf{C}^H\mathbf{n} \tag{3}$$

[1]The terminology "spreading codes" should not be misinterpreted as applicable only to CDMA systems. They are equally applicable to TDMA systems, where they represent orthogonal temporal sequences (windows) each spanning 4 symbol intervals.

By complex-conjugating the second and the fourth entry of $\widetilde{\mathbf{r}}$ in (3), we obtain the following final model for the received signal

$$\mathbf{r}' = \mathbf{Hb} + \mathbf{n}' \tag{4}$$

where $\mathbf{r}' = [\widetilde{\mathbf{r}}(1)\ \widetilde{\mathbf{r}}^*(2)\ \widetilde{\mathbf{r}}(3)\ \widetilde{\mathbf{r}}^*(4)]^T$, $\mathbf{H}$ is defined as

$$\mathbf{H} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ -h_2^* & h_1^* & -h_4^* & h_3^* \\ -h_3 & h_4 & h_1 & -h_2 \\ -h_4^* & -h_3^* & h_2^* & h_1^* \end{bmatrix} \tag{5}$$

and $\mathbf{n}'$ is similarly obtained from $\mathbf{C}^H\mathbf{n}$ by complex-conjugating its second and fourth entry. We are now ready to begin processing the signal $\mathbf{r}'$ as follows.

3. LINEAR RECEIVER PROCESSING

We first perform channel-matched filtering in order to collect sufficient statistics for demodulation:

$$\mathbf{r}_{\text{mf}} = \mathbf{H}^H\mathbf{r}' = \begin{bmatrix} \gamma & 0 & \alpha & 0 \\ 0 & \gamma & 0 & -\alpha \\ -\alpha & 0 & \gamma & 0 \\ 0 & \alpha & 0 & \gamma \end{bmatrix}\mathbf{b} + \mathbf{H}^H\mathbf{n}' \tag{6}$$
$$= \Delta_4\ \mathbf{b} + \mathbf{n}_{\text{mf}}$$

where

$$\gamma = \mathbf{h}^H\mathbf{h} = \sum_{m=1}^{4} |h_m|^2$$
$$\alpha = 2j\text{Im}(h_1^*h_3 + h_4^*h_2) \tag{7}$$

Notice both the particular sparse structure of the matrix Δ_4, as well as the fact that γ is real and α is imaginary. These result in Δ_4 being in general full rank $(\det(\Delta_4) = (\gamma^2 + \alpha^2)^2)$. Moreover, due to symmetries in Δ_4, the post-matched-filtering signal model in (6), which is viewed as a 4-input / 4-output system, can be perfectly decoupled into the following set of two 2-input / 2-output systems. Namely, by grouping the entries of $\mathbf{r}_{\text{mf}}$ in two pairs, we obtain:

$$\begin{cases} \begin{bmatrix} r_{\text{mf},1} \\ r_{\text{mf},3} \end{bmatrix} = \Delta_2 \begin{bmatrix} b_1 \\ b_3 \end{bmatrix} + \begin{bmatrix} n_{\text{mf},1} \\ n_{\text{mf},3} \end{bmatrix} \\[2ex] \begin{bmatrix} r_{\text{mf},4} \\ r_{\text{mf},2} \end{bmatrix} = \Delta_2 \begin{bmatrix} b_4 \\ b_2 \end{bmatrix} + \begin{bmatrix} n_{\text{mf},4} \\ n_{\text{mf},2} \end{bmatrix} \end{cases} \tag{8}$$

where

$$\Delta_2 = \begin{bmatrix} \gamma & \alpha \\ -\alpha & \gamma \end{bmatrix} \tag{9}$$

(notice that Δ_2^H=Δ_2). It is important to emphasize that the two pairs of equations in (8) are completely decoupled. Moreover, they correspond to two identical (2,2) signal models that share the same 2×2 channel matrix Δ_2 and that have identically distributed, but statistically independent, 2×1 additive noise vectors. As will be shown later, this can lead to complexity / computation savings at the receiver. The fact that α is in general non-zero, represents interference across the sub-streams inherent with the proposed transmission scheme. A scheme that would result in a similar model but with $\alpha = 0$ would be of course more desirable, as it would allow the attainment of the open-loop capacity of the considered (4,1) system. However, it has been shown (through the theory of orthogonal designs – see [6]) that no such scheme of block-length 4 (as here) exists for the case of 4 transmit antennas. We now describe how the matched-filtered outputs can be further processed in a linear fashion.

3.1. Zero-Forcing processing

A straightforward way of mitigating the interference in $\mathbf{r}_{\mathrm{mf}}$ due to α in (6), is to use a decorrelating (zero forcing – ZF) receiver. This is expected to be computationally simple but comes at the price of some noise enhancement. Mathematically, the ZF receiver operates on the matched-filter outputs as follows

$$\mathbf{r}_{\mathrm{zf}} = \Delta_4^{-1}\mathbf{r}_{\mathrm{mf}} = \mathbf{b} + \Delta_4^{-1}\mathbf{n}_{\mathrm{mf}} \qquad (10)$$

Due to the decoupling expressed in (8), the ZF operation also decouples as follows

$$\begin{bmatrix} r_{\mathrm{zf},1} \\ r_{\mathrm{zf},3} \end{bmatrix} = \Delta_2^{-1} \begin{bmatrix} r_{\mathrm{mf},1} \\ r_{\mathrm{mf},3} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_3 \end{bmatrix} + \Delta_2^{-1} \begin{bmatrix} n_{\mathrm{mf},1} \\ n_{\mathrm{mf},3} \end{bmatrix}$$
$$\begin{bmatrix} r_{\mathrm{zf},4} \\ r_{\mathrm{zf},2} \end{bmatrix} = \Delta_2^{-1} \begin{bmatrix} r_{\mathrm{mf},4} \\ r_{\mathrm{mf},2} \end{bmatrix} = \begin{bmatrix} b_4 \\ b_2 \end{bmatrix} + \Delta_2^{-1} \begin{bmatrix} n_{\mathrm{mf},4} \\ n_{\mathrm{mf},2} \end{bmatrix} \qquad (11)$$

Eq. (11) is equivalent to Eq. (10). This stems from that fact that, if we denote the matrix

$$W_{\mathrm{zf},2}^H = \Delta_2^{-1} = \begin{bmatrix} \beta & \delta \\ -\delta & \beta \end{bmatrix} \qquad (12)$$

(where $\beta = \gamma/(\gamma^2+\alpha^2)$ and $\delta = -\alpha/(\gamma^2+\alpha^2)$), then

$$\Delta_4^{-1} = \begin{bmatrix} \beta & 0 & \delta & 0 \\ 0 & \beta & 0 & -\delta \\ -\delta & 0 & \beta & 0 \\ 0 & \delta & 0 & \beta \end{bmatrix}$$

This has a beneficial impact on computational complexity and receiver implementation simplicity, since only the inversion of a single 2×2 matrix (instead of a 4×4 matrix) is required. Moreover, we note that $\Delta_2^{-1} = 1/(\gamma^2+\alpha^2)\Delta_2^T$. In terms of performance, it can be shown that the above ZF scheme can attain (in the sense mentioned in Section 1) the following capacity:

$$C_{\mathrm{zf}} = \log_2\left(1 + \frac{\rho}{4}\left(\frac{\gamma^2+\alpha^2}{\gamma}\right)\right) \qquad \text{[bps/Hz]} \qquad (13)$$

ρ being the received signal to noise ratio (SNR) defined as $\rho{=}4\sigma_b^2/\sigma_n^2$ (σ_n^2 is the variance of each entry of $\mathbf{n}$).

3.2. MMSE processing

A better compromise between signal recovery and noise amplification (and hence better performance) can be achieved with minimum mean squared error (MMSE) processing. This is achieved by the 4×4 setting $W_{\mathrm{ms},4}$ which minimizes the following MMSE criterion:

$$\min_{W_{\mathrm{ms},4}} \quad \|W_{\mathrm{ms},4}^H\mathbf{r}_{\mathrm{mf}} - \mathbf{b}\|^2 \qquad (14)$$

The minimization of (14) yields the Wiener solution

$$W_{\mathrm{ms},4}^H = \Delta_4^H\left(\Delta_4\Delta_4^H + \frac{\sigma_n^2}{\sigma_b^2}\Delta_4\right)^{-1} \qquad (15)$$

Similarly to the ZF case, the post-MMSE-processed signal $\mathbf{r}_{\mathrm{ms}}{=}W_{\mathrm{ms},4}^H\mathbf{r}_{\mathrm{mf}}$ is decomposable as

$$\begin{bmatrix} r_{\mathrm{ms},1} \\ r_{\mathrm{ms},3} \end{bmatrix} = W_{\mathrm{ms},2}^H \begin{bmatrix} r_{\mathrm{mf},1} \\ r_{\mathrm{mf},3} \end{bmatrix}$$
$$\begin{bmatrix} r_{\mathrm{ms},4} \\ r_{\mathrm{ms},2} \end{bmatrix} = W_{\mathrm{ms},2}^H \begin{bmatrix} r_{\mathrm{mf},4} \\ r_{\mathrm{mf},2} \end{bmatrix} \qquad (16)$$

where $W_{\mathrm{ms},2}^H{=}\Delta_2^H\left(\Delta_2\Delta_2^H + (\sigma_n^2/\sigma_b^2)\Delta_2\right)^{-1}$. Note that, as expected, as $\rho \to \infty$, the solution $W_{\mathrm{ms},2}^H$ converges to the ZF solution $W_{\mathrm{zf},2}^H = \Delta_2^{-1}$. Again, the decomposition in (16) allows the attainment of MMSE performance using only one 2×2 matrix inversion. The fact that $\Delta_2^H{=}\Delta_2$ leads to further computational simplifications. The corresponding capacity of the MMSE receiver can be found to be given by

$$C_{\mathrm{ms}} = \log_2\left(1 + \frac{W_1^H\Omega W_1}{W_1^H\Phi W_1 + 4/\rho\left(W_1^H\Delta_2 W_1\right)}\right) \qquad (17)$$

where

$$\Omega = \begin{bmatrix} \gamma \\ -\alpha \end{bmatrix}\begin{bmatrix} \gamma \\ -\alpha \end{bmatrix}^H$$
$$\Phi = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix}\begin{bmatrix} \alpha \\ \gamma \end{bmatrix}^H \qquad (18)$$

and $W_1^H = [1\ \ 0]\ W_{\mathrm{ms},2}^H$.

4. NON-LINEAR RECEIVER PROCESSING

Having in mind the signal model in (8), one might opt to use a non-linear multi-user detector instead of the linear MUD's presented in the previous section. In the following we choose to focus on maximum likelihood MUD. A pre-whitened version of the signal model (8) is

$$\begin{bmatrix} r_{\mathrm{pw},1} \\ r_{\mathrm{pw},3} \end{bmatrix} = \Lambda \begin{bmatrix} b_1 \\ b_3 \end{bmatrix} + \begin{bmatrix} n_{\mathrm{pw},1} \\ n_{\mathrm{pw},3} \end{bmatrix} \qquad (19)$$

(and similarly for the second pair of sub-streams), where

$$\Lambda = \begin{bmatrix} \lambda & \kappa \\ -\kappa & \lambda \end{bmatrix} \qquad (20)$$

with $\lambda{=}(\mu\gamma{+}\nu\alpha)/(\mu^2{+}\nu^2)$, $\kappa{=}(\mu\alpha{-}\nu\gamma)/(\mu^2{+}\nu^2)$,

$$\begin{cases} \mu = \left(1/\sqrt{2}\right) \sqrt{\gamma + \sqrt{\gamma^2+\alpha^2}} \\ \nu = \left(j\text{sign}(\alpha/j)/\sqrt{2}\right) \sqrt{\gamma - \sqrt{\gamma^2+\alpha^2}} \end{cases} \quad (21)$$

$n_{\text{pw},1}$ and $n_{\text{pw},3}$ are mutually independent i.i.d. Gaussian variables of variance σ_n^2 each. Since the noise vector in (19) is jointly Gaussian with covariance matrix $\sigma_n^2\mathbf{I}_2$, the maximum likelihood (ML) multi-user detector for (19) solves the following optimization problem

$$\min_{\{b_1,b_3\}\in\mathcal{A}\times\mathcal{A}} \left\| \begin{bmatrix} r_{\text{pw},1} \\ r_{\text{pw},3} \end{bmatrix} - \Lambda \begin{bmatrix} b_1 \\ b_3 \end{bmatrix} \right\|^2 \quad (22)$$

where $\|\mathbf{x}\|^2{=}\mathbf{x}^H\mathbf{x}$, and $\mathcal{A}$ is the alphabet shared by all the sub-streams. When the sub-streams are encoded, the corresponding optimization problem is

$$\min_{\{\tilde{b}_1,\tilde{b}_3\}\in\mathcal{B}\times\mathcal{B}} \sum_{i=1}^{L} \left\| \begin{bmatrix} r_{\text{pw},1}(i) \\ r_{\text{pw},3(i)} \end{bmatrix} - \Lambda \begin{bmatrix} \tilde{b}_1(i) \\ \tilde{b}_3(i) \end{bmatrix} \right\|^2 \quad (23)$$

where the search is over all encoded sequences $\{\tilde{b}_1\}$, $\{\tilde{b}_3\}$, assumed of length L, and taking values in $\mathcal{B}$.

It can be shown that the Shannon capacity achieved through the above ML detection (in the limit of infinitely long random codes) is given by

$$C_{\text{ml}} = \frac{1}{2}\log_2\det\left(\mathbf{I}_2 + \frac{\rho}{4}\Delta_2\right) \quad (24)$$

Interestingly, it can be also shown that this is the maximum allowable capacity within our proposed scheme. Notice, however, that this is still short of the full (4,1) log-det open-loop capacity, which is given by

$$C_o = \log_2\left(1 + \frac{\rho\gamma}{4}\right) \quad (25)$$

5. NUMERICAL RESULTS

In the following, we use the expressions (13), (17), (24), (25) to evaluate outage capacities achievable within our scheme. Figure 1 shows 10% outage capacities of the ZF, MMSE, and ML schemes as compared to the maximum open-loop (4,1) capacity. The channel coefficients $\{h_m\}$, $m = 1, \ldots, 4$, are independently chosen Gaussian random variables of unit variance (corresponding to Rayleigh-distributed amplitudes). The outage values are computed based on one thousand independent random runs. The ML receiver for our scheme is capable of achieving, depending on the SNR, between 95.84% and 99.52% of the open-loop capacity! The corresponding range for the MMSE detector is 89.69%–99.21% and for the ZF detector 66.83%–89.34%. Notice that at low SNR's, both the MMSE and the ML capacities approach closely the open-loop capacity. Also,

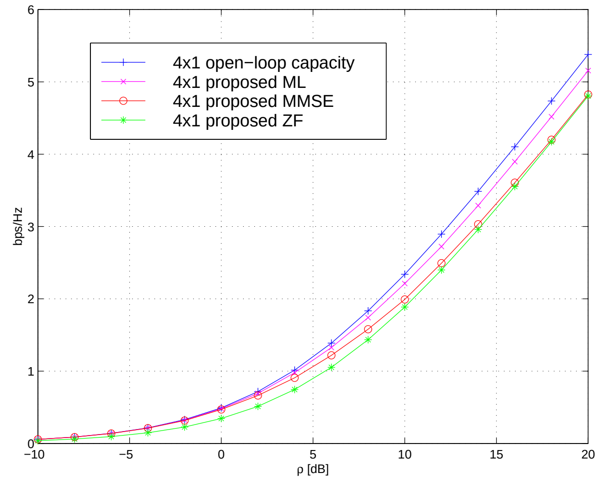

Figure 1: Outage capacities of the proposed scheme compared to the (4,1) open-loop capacity.

notice that at high SNR's, the two linear schemes are capable of achieving a large percentage of the open-loop capacity (on the order of about 90%).

REFERENCES

[1] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451-1458, Oct. 1998.

[2] V. Tarokh, N. Seshadri, A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Information Theory*, vol. 44, No. 2, pp. 744-765, March 1998.

[3] B. Hochwald, L. Marzetta, C. Papadias, "A transmitter diversity scheme for wideband CDMA systems based on space-time spreading," *IEEE Journal on Selected Areas in Communications*, Wideband CDMA II, vol. 19, No. 1, Jan. 2001.

[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, No. 2, pp. 41-59, Autumn 1996.

[5] C. Papadias, "On the spectral efficiency of space-time spreading schemes for multiple antenna CDMA systems," *33rd Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, pp. 639-643, Oct. 24-27, 1999.

[6] A. V. Geramita and J. Seberry, *Orthogonal Designs: Quadratic Forms and Hadamard Matrices*, New York: Marcel Dekker, 1979.