MIMO Wireless Communication

Daniel W. Bliss, Keith W. Forsythe, and Amanda M. Chan

Wireless communication using multiple-input multiple-output (MIMO) systems enables increased spectral efficiency for a given total transmit power. Increased capacity is achieved by introducing additional spatial channels that are exploited by using space-time coding. In this article, we survey the environmental factors that affect MIMO capacity. These factors include channel complexity, external interference, and channel estimation error. We discuss examples of space-time codes, including space-time low-density parity-check codes and space-time turbo codes, and we investigate receiver approaches, including multichannel multiuser detection (MCMUD). The 'multichannel' term indicates that the receiver incorporates multiple antennas by using space-time-frequency adaptive processing. The article reports the experimental performance of these codes and receivers.

Multiple-input multiple-output (MIMO) systems are a natural extension of developments in antenna array communication. While the advantages of multiple receive antennas, such as gain and spatial diversity, have been known and exploited for some time [1, 2, 3], the use of transmit diversity has only been investigated recently [4, 5]. The advantages of MIMO communication, which exploits the physical channel between many transmit and receive antennas, are currently receiving significant attention [6–9]. While the channel can be so nonstationary that it cannot be estimated in any useful sense [10], in this article we assume the channel is quasistatic.

MIMO systems provide a number of advantages over single-antenna-to-single-antenna communication. Sensitivity to fading is reduced by the spatial diversity provided by multiple spatial paths. Under certain environmental conditions, the power requirements associated with high spectral-efficiency communication can be significantly reduced by avoiding the compressive region of the information-theoretic capacity bound. Here, spectral efficiency is defined as the total number of information bits per second per Hertz transmitted from one array to the other.

After an introductory section, we describe the concept of MIMO information-theoretic capacity bounds. Because the phenomenology of the channel is important for capacity, we discuss this phenomenology and associated parameterization techniques, followed by examples of space-time codes and their respective receivers and decoders. We performed experiments to investigate channel phenomenology and to test coding and receiver techniques.

Capacity

We discuss MIMO information-theoretic performance bounds in more detail in the next section. Capacity increases linearly with signal-to-noise ratio (SNR) at low SNR, but increases logarithmically with SNR at high SNR. In a MIMO system, a given total transmit power can be divided among multiple spatial paths (or modes), driving the capacity closer to the linear regime for each mode, thus increasing the aggregate spectral efficiency. As seen in Figure 1, which assumes an optimal high spectral-efficiency MIMO channel (a channel matrix with a flat singular-value distribution), MIMO systems enable high spectral efficiency at much lower required energy per information bit.

The information-theoretic bound on the spectral efficiency is a function of the total transmit power and the channel phenomenology. In implementing MIMO systems, we must decide whether channel estimation information will be fed back to the transmitter so that the transmitter can adapt. Most MIMO communication research has focused on systems without feedback. A MIMO system with an uninformed transmitter (without feedback) is simpler to implement, and at high SNR its spectral-efficiency bound approaches that of an informed transmitter (with feedback).

One of the environmental issues with which communication systems must contend is interference, either unintentional or intentional. Because MIMO systems use antenna arrays, localized interference can be mitigated naturally. The benefits extend beyond those achieved by single-input multiple-output systems, that is, a single transmitter and a multiple-antenna receiver, because the transmit diversity nearly guarantees that nulling an interferer cannot unintentionally null a large fraction of the transmit signal energy.

Phenomenology

We discuss channel phenomenology and channel parameterization techniques in more detail in a later section. Aspects of the channel that affect MIMO system capacity, namely, channel complexity and channel stationarity, are addressed in this paper. The first aspect, channel complexity, is a function of the richness of scatterers. In general, capacity at high spectral efficiency increases as the singular values of the channel matrix increase. The distribution of singular values is a measure of the relative usefulness of various spatial paths through the channel.

Space-Time Coding and Receivers

In order to implement a MIMO communication system, we must first select a particular coding scheme. Most space-time coding schemes have a strong connection to well-known single-input single-output (SISO) coding approaches and assume an uninformed transmitter (UT). Later in the article we discuss space-time low-density parity-check codes, space-time turbo codes, and their respective receivers. Space-time coding can exploit the MIMO degrees of freedom to increase redundancy, spectral efficiency, or some combination

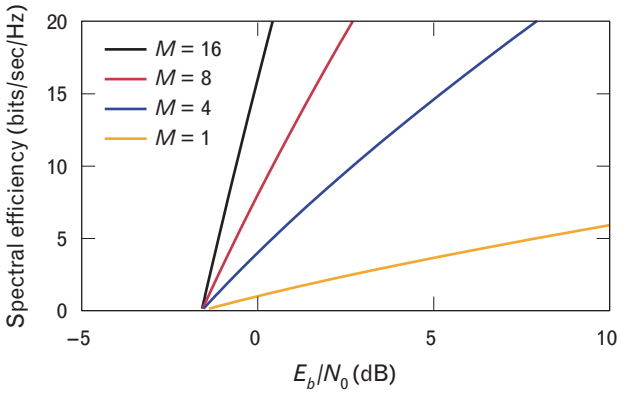


FIGURE 1. Spectral-efficiency bound as a function of noise-spectral-density-normalized energy per information bit (E_b/N_0). The graph compares four different $M \times M$ multiple-input multiple-output (MIMO) systems, assuming channel matrices with flat singular-value distribution.

of these characteristics [11]. Preliminary ideas are discussed elsewhere [6].

A simple and elegant solution that maximizes diversity and enables simple decoupled detection is proposed in Reference 12. More generally, orthogonal space-time block codes are discussed in References 13 and 14. A general discussion of distributing data across transmitters (linear dispersive codes) is given in Reference 15. High SNR design criteria and specific examples are given for space-time trellis codes in Reference 16. Unitary codes optimized for operation in Rayleigh fading are presented in Reference 17. Space-time coding without the requirement of channel estimation is also a common topic in the literature. Many differential coding schemes have been proposed [18]. Under various constraints at the transmitter and receiver, information-theoretic capacity can be evaluated without conditioning on knowledge of the propagation channel [19, 20]. More recently, MIMO extensions of turbo coding have been suggested [21, 22]. Finally, coding techniques for informed transmitter systems have received some interest [23, 24].

Experimental Results

Because information-theoretic capacity and practical performance are dependent upon the channel phenomenology, a variety of experiments were performed. Both channel phenomenology and experimental procedures are discussed in later sections. Experiments were per-

WATER FILLING

WATER FILLING is a metaphor for the solution of several optimization problems related to channel capacity. The simplest physical example is perhaps the case of spectral allocation for maximal total capacity under a total power constraint. Let x_k denote the power received in the kth frequency cell, which has interference (including thermal noise) denoted n_k. If the total received power is constrained to be x, then the total capacity is maximized by solving

$$\max_{\{x_k:\sum_k x_k = x\}} \sum_k \log(1 + x_k/n_k)$$

$$= \max_{\{x_k:\sum_k x_k = x\}} \sum_k \log(n_k + x_k) - \sum_k \log(n_k).$$

Use Lagrange multipliers and evaluate

$$\frac{\partial}{\partial x_k}\left[\sum_j \log(n_j + x_j) - \mu\left(\sum_j x_j - x\right)\right]$$

to find a solution. The solution satisfies $x_k + n_k = \mu^{-1}$ for all nonzero x_k. Figure A illustrates the solution graphically as an example of water filling. The difference between the water level (blue) and the noise level (red) is the power allocated to the signal in each frequency cell. The volume of the water is the total received power of the signal. Note that cells with high levels of interference are not used at all.

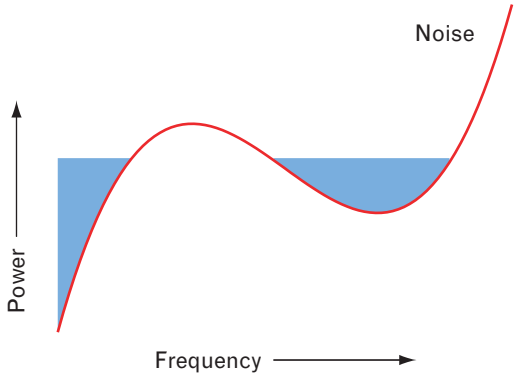


FIGURE A. Notional water-filling example.

A similar solution results when the capacity is expressed by

$$\sum_k \log(1 + g_k x_k)$$

for gains g_k. One can write the gains as $g_k = n_k^{-1}$ and use the water-filling argument above. In this context, cells with low gains may not be used at all.

formed in an outdoor nonstationary environment in a mixed residential, industrial, and light urban settings. Intentional high-power interference was included.

Information-Theoretic Capacity

The information-theoretic capacity of MIMO systems has been widely discussed [7, 25]. The development of the informed transmitter ("water filling") and uninformed transmitter approaches is repeated in this section, along with a discussion of the relative performance of these approaches. (The concept of "water filling" is explained in the sidebar entitled "Water Filling.") In addition, we introduce the topic of spectral-efficiency bounds in the presence of interference, and we discuss spectral-efficiency bounds in frequency-selective environments. Finally, we summarize alternative channel performance metrics.

Informed Transmitter

For narrowband MIMO systems, the coupling between the transmitter and receiver for each sample in time can be modeled by using

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{z}$ is the complex receive-array output,

$$\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$$

is the $n_R \times n_T$ (number of receive by transmit antenna)

channel matrix, **x** is the transmit-array vector, and **n** is zero-mean-complex Gaussian noise.

The capacity is defined as the maximum of the mutual information [26]

$$\mathcal{I}(\mathbf{z},\mathbf{x}|\mathbf{H}) = \left\langle \log_2 \left[\frac{p(\mathbf{z}|\mathbf{x},\mathbf{H})}{p(\mathbf{z}|\mathbf{H})} \right] \right\rangle, \tag{2}$$

over the source conditional probability density $p(\mathbf{x}|\mathbf{H})$ subject to various transmit constraints, where the expectation value is indicated by the notation $\langle \cdots \rangle$. Noting that the mutual information can be expressed as the difference between two conditional entropies

$$\mathcal{I}(\mathbf{z},\mathbf{x}|\mathbf{H}) = h(\mathbf{z}|\mathbf{H}) - h(\mathbf{z}|\mathbf{x},\mathbf{H}), \tag{3}$$

that

$$h(\mathbf{z}|\mathbf{x},\mathbf{H}) = h(\mathbf{n}) = n_R \log_2(\pi e \sigma_n^2),$$

and that $h(\mathbf{z}|\mathbf{H})$ is maximized for a zero-mean Gaussian source **x**, the capacity is given by

$$C = \sup_{\langle \mathbf{x}\mathbf{x}^\dagger \rangle} \log_2 \frac{\left| \sigma_n^2 \mathbf{I}_{n_R} + \mathbf{H} \langle \mathbf{x}\mathbf{x}^\dagger \rangle \mathbf{H}^\dagger \right|}{\left| \sigma_n^2 \mathbf{I}_{n_R} \right|}, \tag{4}$$

where the notation $|\cdots|$ indicates determinant, † indicates Hermitian conjugate, and $\mathbf{I}_{n_R}$ indicates an identity matrix of size n_R. A variety of possible constraints exist for $\langle \mathbf{x}\mathbf{x}^\dagger \rangle$, depending on the assumed transmitter limitations. Here we assume that the fundamental limitation is the total power transmitted. Optimization over the $n_T \times n_T$ noise-normalized transmit covariance matrix, $\mathbf{P} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle / \sigma_n^2$, is constrained by the total noise-normalized transmit power P_o. By allowing different transmit powers at each antenna, we can enforce this constraint by using the form $\mathrm{tr}\{\mathbf{P}\} \le P_o$. The informed transmitter (IT) channel capacity is achieved if the channel is known by both the transmitter and receiver, giving

$$C_{\mathrm{IT}} = \sup_{\mathbf{P};\, \mathrm{tr}(\mathbf{P}) = P_o} \log_2 \left| \mathbf{I}_{n_R} + \mathbf{H}\mathbf{P}\mathbf{H}^\dagger \right|. \tag{5}$$

To avoid radiating negative power, we impose the additional constraint $\mathbf{P} > \mathbf{0}$ by using only a subset of channel modes.

The resulting capacity is given by

$$C_{\mathrm{IT}} = \log_2 \left| \frac{P_o + \mathrm{tr}\{\mathbf{D}^{-1}\}}{n_+} \mathbf{D} \right|. \tag{6}$$

A water-filling argument establishes that the entries d_m in the diagonal matrix

$$\mathbf{D} \in \mathbb{R}^{n_+ \times n_+}$$

contain the n_+ top-ordered eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$. The values d_m must satisfy

$$d_m > \frac{n_+}{P_o + \mathrm{tr}\{\mathbf{D}^{-1}\}}. \tag{7}$$

If Equation 7 is not satisfied for some d_m, it will not be satisfied for any smaller d_m.

In this discussion we assume that the environment is stationary over a period long enough for the error associated with channel estimation to vanish asymptotically. In order to study typical performance of quasistationary channels sampled from a given probability distribution, capacity is averaged over an ensemble of quasistationary environments. Under the ergodic assumption (that is, the ensemble average is equal to the time average), the mean capacity $\langle C_{IT} \rangle$ is the channel capacity.

Uninformed Transmitter

If the channel is not known at the transmitter, then an optimal transmission strategy is to transmit equal power from each antenna $\mathbf{P} = P_o / n_T \, \mathbf{I}_{n_T}$ [7]. Assuming that the receiver can accurately estimate the channel, but the transmitter does not attempt to optimize its output to compensate for the channel, the uninformed transmitter (UT) maximum spectral efficiency bound is given by

$$C_{\mathrm{UT}} = \log_2 \left| \mathbf{I}_{n_R} + \frac{P_o}{n_T} \mathbf{H}\mathbf{H}^\dagger \right|. \tag{8}$$

This is a common transmit constraint, as it may be difficult to provide the transmitter channel estimates. The sidebar entitled "Toy 2 × 2 Channel Model" discusses an example of IT and UT capacities for a simple line-of-sight environment.

Capacity Ratio

At high SNR, C_{IT} and C_{UT} converge. This can be ob-

served in the large P_o limit of the ratio of Equations 6 and 8,

$$\begin{aligned}\frac{C_{\mathrm{IT}}}{C_{\mathrm{UT}}} &\to \frac{\log_2\left|\frac{P_o+\mathrm{tr}\{\mathbf{D}^{-1}\}}{n_{min}}\mathbf{D}\right|}{\log_2\left|\mathbf{I}_{n_{min}}+\frac{P_o}{n_T}\mathbf{D}\right|} \\ &\to \frac{\log_2(P_o)-\log_2(n_{min})+\frac{\log_2|\mathbf{D}|}{n_{min}}}{\log_2(P_o)-\log_2(n_T)+\frac{\log_2|\mathbf{D}|}{n_{min}}} \\ &\to 1,\end{aligned} \tag{9}$$

where the $n_{\min}$ diagonal entries in $\mathbf{D}$ contain all nonzero eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$. If $n_T > n_R$, then the convergence to one is logarithmically slow.

At low SNR the ratio $C_{\mathrm{IT}}/C_{\mathrm{UT}}$ is given by

$$\begin{aligned}\frac{C_{\mathrm{IT}}}{C_{\mathrm{UT}}} &\to \frac{\log_2[(P_o+1/d_{max})d_{max}]}{\log_2\left|\mathbf{I}_{n_R}+\frac{P_o}{n_T}\mathbf{H}\mathbf{H}^\dagger\right|} \\ &= \frac{\log(1+P_o\,\mathrm{maxeig}\{\mathbf{H}\mathbf{H}^\dagger\})}{\mathrm{tr}\left\{\log\left(\mathbf{I}_{n_R}+\frac{P_o}{n_T}\mathbf{H}\mathbf{H}^\dagger\right)\right\}} \\ &\approx \frac{\mathrm{maxeig}\{\mathbf{H}\mathbf{H}^\dagger\}}{\frac{1}{n_T}\mathrm{tr}\{\mathbf{H}\mathbf{H}^\dagger\}},\end{aligned} \tag{10}$$

using Equation 6 with $n_+ = 1$ and Equation 8. Given this asymptotic result, we can make a few observations. The spectral-efficiency ratio is given by the maximum to the average eigenvalue ratio of $\mathbf{H}^\dagger\mathbf{H}$. If the channel is rank one, such as in the case of a multiple-input single-output (MISO) system, the ratio is approximately equal to n_T. Finally, in the special case in which $\mathbf{H}^\dagger\mathbf{H}$ has a flat eigenvalue distribution, the optimal transmit covariance matrix is not unique. Nonetheless, the ratio $C_{\mathrm{IT}}/C_{\mathrm{UT}}$ approaches one.

Interference

By extending the discussion in the previous section [8, 27], we can calculate capacity in the presence of uncooperative (worst case) external interference η, in addition to the spatially-white complex Gaussian noise $\mathbf{n}$ considered previously. The mutual information is again given by Equations 2 and 3, where entropy $h(\mathbf{z}|\mathbf{x},\mathbf{H})$ in the presence of the external interference becomes $h(\mathbf{n}+\eta)$,

$$h(\mathbf{z}|\mathbf{x},\mathbf{H}) \le \log_2\left\{\pi e\left|\sigma_n^2\mathbf{I}+\sigma_n^2\mathbf{R}\right|\right\},$$

and $\sigma_n^2\mathbf{R}$ is the spatial-interference covariance matrix. Equality is achieved if and only if the interference amplitudes have a Gaussian distribution. Thus the worst-case *informed* capacity, the maximum-minimum mutual information,

$$C_{\mathrm{int}} = \sup_{p(\mathbf{z}|\mathbf{H})}\inf_{p(\vec{\eta})} I(\mathbf{z},\mathbf{x}|\mathbf{H}), \tag{12}$$

becomes

$$C_{\mathrm{IT,\,int}} = \sup_{\tilde{\mathbf{P}};\mathrm{tr}(\tilde{\mathbf{P}})=P_o}\log_2\left|\mathbf{I}+\tilde{\mathbf{H}}\tilde{\mathbf{P}}\tilde{\mathbf{H}}^\dagger\right|, \tag{13}$$

using

$$\tilde{\mathbf{H}} \equiv (\mathbf{I}+\mathbf{R})^{-1/2}\mathbf{H}. \tag{14}$$

Gaussian interference corresponds to a saddle point of the mutual information at which the maximum-minimum capacity is achieved. The capacity in the presence of Gaussian interference has a form identical to Equation 6 under the transformation $\mathbf{D} \to \tilde{\mathbf{D}}$, where $\tilde{\mathbf{D}}$ contains the eigenvalues of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^\dagger$. The transmitted noise-normalized power covariance matrix $\tilde{\mathbf{P}}$ is calculated by using $\tilde{\mathbf{H}}$. Similarly, the uninformed transmitter spectral-efficiency bound in the presence of noise is given by the same transformation of $\mathbf{H} \to \tilde{\mathbf{H}}$.

In the limit of high spectral efficiency for n_J infinite *J*/*S* jammers, the loss in capacity approaches

$$C_{\mathrm{int}} \to \frac{\min(n_T, n_R - n_J)}{\min(n_T, n_R)}C. \tag{15}$$

In general, the theoretical capacity is not significantly affected as long as the number of antennas is much larger than the number of jammers. This resistance to the effects of jammers is demonstrated experimentally later in the article.

Frequency-Selective Channels

In environments in which there is frequency-selective fading, the channel matrix $\mathbf{H}(f)$ is a function of frequency. Exploiting the orthogonality of frequency channels, the capacity in frequency-selective fading can be calculated by using an extension of Equations 6 and 8. For the uninformed transmitter, this leads to the frequency-selective spectral-efficiency bound

TOY 2 × 2 CHANNEL MODEL

Because the distribution of channel matrix eigenvalues is essential to the effectiveness of multiple-input, multiple-output (MIMO) communication, we employ a toy example for the purposes of introduction, and we discuss the eigenvalue distribution of a 2 × 2 narrowband MIMO system in the absence of environmental scatterers. To visualize the example, we can imagine two receive and two transmit antennas located at the corners of a rectangle. The ratio of channel matrix eigenvalues can be changed by varying the shape of the rectangle. The columns of the channel matrix **H** (in Equation 1 in the main article) can be viewed as the receiver-array response vectors, one vector for each transmit antenna,

$$\mathbf{H} = \sqrt{2}\begin{pmatrix} a_1\mathbf{v}_1 & a_2\mathbf{v}_2 \end{pmatrix},$$

where a_1 and a_2 are constants of proportionality (equal to the root-mean-squared transmit-to-receive attenuation for transmit antennas 1 and 2 respectively) that take into account geometric attenuation and antenna gain effects, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are unit-norm array response vectors. For the purpose of this discussion, we assume $a = a_1 = a_2$, which is valid if the rectangle deformation does not significantly affect overall transmitter-to-receiver distances.

The capacity of the 2 × 2 MIMO system is a function of the channel singular values and the total transmit power. Eigenvalues of $\mathbf{HH}^\dagger$ are given by

$$\mu_{1,2} = 2a^2\left(1 \pm \left\| \mathbf{v}_1^\dagger \mathbf{v}_2 \right\|\right),$$

where the absolute value is denoted by $\|\ldots\|$. The separation between receive array responses can be described in a convenient form in terms of generalized beamwidths [40],

$$b_{12} = \frac{2}{\pi}\arccos\left\{\left\| \mathbf{v}_1^\dagger \mathbf{v}_2 \right\|\right\}.$$

For small angular separations, this definition of beamwidths closely approximates many *ad hoc* definitions for physical arrays. Figure A displays the eigenvalues μ_1 and μ_2 as a function of generalized beamwidth separation. When the transmit and receive arrays are small, indicated by a small separation in beamwidths, one eigenvalue is dominant. As the array apertures become larger, indicated by a larger separation, one array's individual elements can be resolved by the other array. Consequently, the smaller eigenvalue increases. Conversely, the larger eigenvalue decreases slightly.

Equations 6 and 7 in the main article are employed to determine the capacity for the 2 × 2 system. The "water-filling" technique (explained in a previous sidebar) must first determine if both modes in the channel are employed. Both modes are used if the following condition is satisfied,

$$\mu_2 > \frac{2}{P_o + \frac{1}{\mu_1} + \frac{1}{\mu_2}},$$

$$P_o > \frac{1}{\mu_2} - \frac{1}{\mu_1} > \frac{\left\| \mathbf{v}_1^\dagger \mathbf{v}_2 \right\|}{a^2\left(1 - \left\| \mathbf{v}_1^\dagger \mathbf{v}_2 \right\|^2\right)},$$

assuming $\mu_1 > \mu_2$.

If the condition is not satisfied, then only the

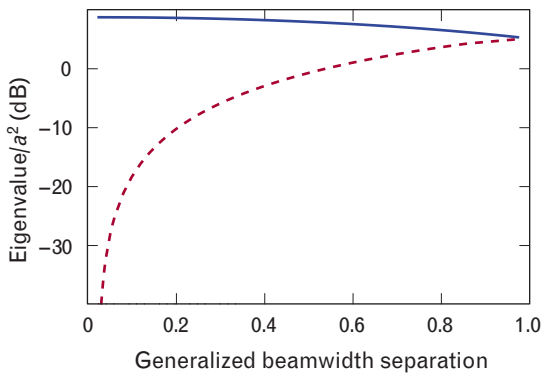


FIGURE A. Eigenvalues of $\mathbf{HH}^\dagger$ for a 2 × 2 line-of-sight channel as a function of antenna separation.

stronger channel mode is employed and the capacity, from Equation 6, is given by

$$C_{\mathrm{IT}} = \log_2(1+\mu_1 P_o),$$
$$= \log_2\left[1+2a^2\left(1+\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|\right)P_o\right];$$

otherwise, both modes are used and the capacity is given by

$$C_{\mathrm{IT}} = \log_2\left|\frac{P_o+\frac{1}{\mu_1}+\frac{1}{\mu_2}}{2}\begin{pmatrix}\mu_1 & 0\\ 0 & \mu_2\end{pmatrix}\right|,$$
$$= \log_2\left\{\left(\frac{\mu_1\mu_2 P_o+\mu_1+\mu_2}{2}\right)^2\frac{1}{\mu_1\mu_2}\right\},$$
$$= 2\log_2\left\{a^2\left(1-\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|^2\right)P_o+1\right\}$$
$$-\log_2\left\{1-\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|^2\right\}.$$

Figure B displays the resulting capacity as a function of a^2P_o (mean single-input single-output SNR) for two beamwidth separations, 0.1 and 0.9. At low values of a^2P_o the capacity associated with small beamwidth separation performs best. In this regime, capacity is linear with receive power, and small beamwidth separation increases the coherent gain. At high values of a^2P_o large beamwidth separation produces a higher capacity as the optimal MIMO system distributes the energy between modes.

The total received power is given by

$$2\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|a^2P_o$$

when using one mode, and

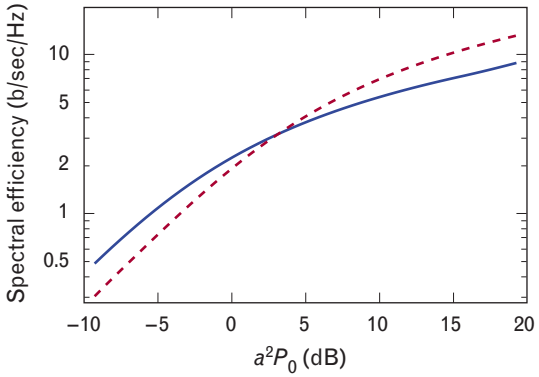


FIGURE B. The informed transmitter capacity of a 2×2 line-of-sight channel, assuming antenna beamwidth separations of 0.1 (solid line) and 0.9 (dashed line).

$$2a^2P_o+\frac{2\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|^2}{\left(1-\left\|\mathbf{v}_1^{\dagger}\mathbf{v}_2\right\|^2\right)},$$

when using two modes, where P_o is the total noise-normalized power. In both cases, the total received power is much larger than a^2P_o.

In complicated multipath environments, small arrays employ scatterers to create virtual arrays of a much larger effective aperture. The effect of the scatterers upon capacity depends on their number and distribution in the environment. The individual antenna elements can be resolved by the larger effective aperture produced by the scatterers. As demonstrated in Figure A, the ability to resolve antenna elements is related to the number of large singular values of the channel matrix and thus the capacity.

$$C_{\mathrm{UT,FS}} = \frac{\int df\, C_{\mathrm{UT}}(P_o;\mathbf{H}(f))}{\int df}$$
$$\approx \frac{\sum_{n=1}^{n_f}\Delta f\log_2\left|\mathbf{I}+\frac{P_o}{n_T}\mathbf{H}(f_n)\mathbf{H}^{\dagger}(f_n)\right|}{\sum_{n=1}^{n_f}\Delta f}$$
$$\approx \frac{1}{n_f}\log_2\left|\mathbf{I}+\frac{P_o}{n_T}\breve{\mathbf{H}}\breve{\mathbf{H}}^{\dagger}\right|, \qquad (16)$$

where the distance between frequency samples is given by Δf and the n_f-bin frequency-partitioned channel matrix is given by

$$\breve{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{H}(f_1) & 0 & 0 & 0 \\ 0 & \mathbf{H}(f_2) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{H}(f_{n_f}) \end{pmatrix}. \quad (17)$$

For the informed transmitter channel capacity, power is optimally distributed amongst both spatial modes and frequency channels. The capacity can be expressed

$$C_{\text{IT, FS}} \approx \max_{\breve{\mathbf{P}}} \frac{1}{n_f} \log_2 \left| \mathbf{I} + \breve{\mathbf{H}} \breve{\mathbf{P}} \breve{\mathbf{H}}^{\dagger} \right|, \quad (18)$$

which is maximized by Equation 6 with the appropriate substitutions for the frequency-selective channel, and diagonal entries in $\mathbf{D}$ in Equation 7 are selected from the eigenvalues of $\breve{\mathbf{H}}\breve{\mathbf{H}}^{\dagger}$. Because of the block diagonal structure of $\mathbf{H}$, the $(n_T \cdot n_f) \times (n_T \cdot n_f)$ space-frequency noise-normalized transmit covariance matrix $\breve{\mathbf{H}}$ is a block diagonal matrix, normalized so that

$$\text{tr}\{\breve{\mathbf{P}}\} \le n_f P_o .$$

Other Performance Metrics

The information-theoretic capacity is not the only possible metric of performance. As an example, another useful performance metric is the outage capacity [16], or the achievable spectral-efficiency bound, assuming a given probability of error-free decoding of a frame. In many practical situations this metric may be the best measure of performance, for example, in the case in which the system can resend frames of data.

Channel Phenomenology

In this section we describe tools for modeling, estimating, and characterizing MIMO channels. These topics are discussed in greater detail elsewhere [25, 28]. First we introduce the standard model and simple modifications to it. Then we discuss the simplest channel characterization, which is mean receive power, followed by a description of channel estimation techniques, methods for determining how much channels have changed, and channel parameterization and estimation techniques.

Standard Model

A variety of techniques are used to simulate the channel matrix [29]. The simplest approach is to assume that all the entries in the channel matrix are sampled from identical independent complex Gaussians $\mathbf{H} \sim \mathbf{G}$. This assumption corresponds to an environment with complicated multipath scattering. While this approach is convenient from the perspective of performing analytic calculations, it may provide a channel eigenvalue distribution that is too flat. At the other extreme, channels can be characterized by a diversity order [30], which is used to indicate an effective cut-off in the eigenvalue distribution induced by spatial correlation. A number of approaches that introduce spatial correlations have been suggested. One approach uses the form

$$\mathbf{H} = \mathbf{M}_L \, \mathbf{G} \, \mathbf{M}_R . \quad (19)$$

The above model results in a $(n_T \cdot n_R) \times (n_T \cdot n_R)$ link-by-link covariance matrix of the Kronecker product form $(\mathbf{M}_L \mathbf{M}_L^{\dagger}) \otimes (\mathbf{M}_R \mathbf{M}_R^{\dagger})^*$ for the entries in the channel matrix $\mathbf{H}$. This product structure can arise from a spherical Green's function model of propagation, provided several additional conditions are met. First, scatterers are concentrated around (but not too close to) the transmitter and receiver. Second, multiple scattering of a particular kind (from transmitter element to transmitter scatterer to receiver scatterer to receiver element) dominates propagation. Third, scatterers are sufficiently separated in angle when viewed by their associated array.

Received Power

It is often convenient to parameterize the incoming signal power in terms of $a^2 P_o$, where a^2 is the mean-squared link attenuation. It can be employed to easily compare performance by using different constraints and environments. This choice corresponds to the typical noise-normalized received power for a single receive and single transmit antenna radiating power $\sigma_n^2 P_o$. However, this choice can be mildly misleading because the total received power will, in general, be much larger than $a^2 P_o$. In general, a^2 is defined by the Frobenius norm squared of the channel matrix normalized by the number of transmitters and receivers,

$$a^2 = \frac{\text{tr}\{\mathbf{H}\mathbf{H}^{\dagger}\}}{n_T \, n_R} . \quad (20)$$

The total received noise-normalized power produced by a set of orthogonal receive beamformers is given by

$\mathrm{tr}\{\mathbf{HPH}^\dagger\}$. The uninformed transmitter rate is maximized by sending equal power to all transmit antennas so that $\mathrm{tr}\{\mathbf{HPH}^\dagger\}$ becomes $P_o/n_T\,\mathrm{tr}\{\mathbf{HH}^\dagger\} = n_R\,a^2P_o$. It is worth noting that $\mathbf{P}$ is not in general optimized by the informed transmitter to maximize received power but to maximize capacity.

The total received power for the capacity-optimized informed transmitter, given an arbitrary channel matrix, is

$$\begin{aligned}\mathrm{tr}\{\mathbf{H}\mathbf{P}_{IT}\mathbf{H}^\dagger\} &= \mathrm{tr}\left\{\left(\frac{P_o+\mathrm{tr}(\mathbf{D}^{-1})}{n_+}\right)\mathbf{D}-\mathbf{I}_{n_+}\right\} \\ &= P_o\,\frac{\mathrm{tr}\{\mathbf{D}\}}{n_+}+\frac{\mathrm{tr}\{\mathbf{D}^{-1}\}\mathrm{tr}\{\mathbf{D}\}-n_+^2}{n_+}\,.\end{aligned} \tag{21}$$

The first term in Equation 21 is bounded from below by

$$\begin{aligned}P_o\,\frac{\mathrm{tr}\{\mathbf{D}\}}{n_+} &\geq P_o\,\frac{\mathrm{tr}\{\mathbf{HH}^\dagger\}}{\min\{n_T,n_R\}} \\ &\geq \max\{n_T,n_R\}\,a^2P_o\,.\end{aligned} \tag{22}$$

The second term in Equation 22 is bounded from below by zero. Consequently, the total received power is greater than or equal to $\max\{n_T,n_R\}\,a^2P_o$.

For very small a^2P_o, far from the nonlinear regime of the Shannon limit, the optimal solution is to maximize received power. This is done by transmitting the best mode only, setting $n_+ = 1$. In this regime the total received power is given by

$$\mathrm{tr}\{\mathbf{HP}_{IT}\mathbf{H}^\dagger\} \rightarrow P_o\,\mathrm{maxeig}\{\mathbf{HH}^\dagger\}\,. \tag{23}$$

This result is bounded from above by $n_T\,n_R\,a^2P_o$, which is achieved if there is only a single nontrivial mode in the channel.

Channel Estimation

The Gaussian probability density function for a multivariate, signal-in-the-mean, statistical model of the received signal $\mathbf{Z}$, assuming $\mathbf{T}\in\mathbb{C}^{n_T\times n_s}$ is the transmit sequence, is given by

$$p(\mathbf{Z}\,|\,\mathbf{R},\mathbf{H},\mathbf{T}) = \frac{e^{-\mathrm{tr}[(\mathbf{Z}-\mathbf{HT})^\dagger\,\mathbf{R}^{-1}(\mathbf{Z}-\mathbf{HT})]}}{\pi^{n_s n_R}\,|\mathbf{R}|^{n_s}}\,, \tag{24}$$

where $\mathbf{R}$ is the noise-plus-interference covariance matrix. The maximum-likelihood estimate of $\mathbf{H}$ is given by

$$\hat{\mathbf{H}} = \mathbf{Z}\,\mathbf{T}^\dagger\,(\mathbf{T}\mathbf{T}^\dagger)^{-1}\,, \tag{25}$$

assuming that the reference signals in $\mathbf{T}$ are known and $\mathbf{T}\mathbf{T}^\dagger$ is nonsingular.

The previous channel-estimation discussion explicitly assumed flat fading. However, the frequency-selective channels can be estimated by first estimating a finite impulse-response MIMO channel, which can be transformed to the frequency domain.

A finite impulse-response extension of Equation 1 is given by introducing delayed copies of $\mathbf{T}$ at delays $\delta_1,\delta_2,\ldots,\delta_{n_{taps}}$,

$$\tilde{\mathbf{T}} \equiv \begin{pmatrix}\mathbf{T}(\delta_1)\\ \mathbf{T}(\delta_2)\\ \vdots\\ \mathbf{T}(\delta_{n_{taps}})\end{pmatrix}, \tag{26}$$

so that the transmit matrix has dimension $(n_T\cdot n_{taps})\times n_s$. The resulting wideband channel matrix has the dimension $(n_T\cdot n_{taps})\times n_s$,

$$\begin{aligned}&[\hat{\mathbf{H}}(\delta_1)\quad \hat{\mathbf{H}}(\delta_2)\quad\cdots\quad \hat{\mathbf{H}}(\delta_{n_{taps}})] \\ &\quad = \mathbf{Z}\tilde{\mathbf{T}}^\dagger\,(\tilde{\mathbf{T}}\tilde{\mathbf{T}}^\dagger)^{-1}.\end{aligned} \tag{27}$$

Using this form, an effective channel filter is associated with each transmit-to-receive antenna link. By assuming regular delay sampling, we can use a discrete Fourier transform to construct the explicit frequency-selective form,

$$\begin{aligned}&[\hat{\mathbf{H}}(f_1)\quad \hat{\mathbf{H}}(f_2)\ \cdots\ \hat{\mathbf{H}}(f_{n_{taps}})] \\ &\quad = [\hat{\mathbf{H}}(\delta_1)\quad \hat{\mathbf{H}}(\delta_2)\ \cdots\ \hat{\mathbf{H}}(\delta_{n_{taps}})](\mathcal{F}_{n_{taps}}\otimes\mathbf{I}_{n_T})\,,\end{aligned} \tag{28}$$

where the n-point discrete Fourier transform is represented by $\mathcal{F}_n$ and the Kronecker product is represented by $\otimes$.

Channel-Difference Metrics

A variety of metrics are possible. In investigating channel variations, no one metric will be useful for all situations. As an example, two completely different channels can have the same capacity. Depending upon the issue being investigated, we may wish to think of these matri-

ces as being similar or very different. Here two metrics are discussed. Both metrics are ad hoc, but motivations are provided. The first metric measures differences in channel singular-value distributions. The second metric is sensitive to differences in both the singular-value distribution and the channel eigenvector structure.

Eigenvalue-Based Metric

As was mentioned earlier, MIMO capacity is only a function of the channel singular values. Equivalently, capacity is invariant under channel-matrix transformations of the form

$$\mathbf{H} \rightarrow \mathbf{W}_1 \mathbf{H} \mathbf{W}_2^{\dagger}, \tag{29}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are arbitrary unitary matrices. Consequently, for some applications it is useful to employ a metric that is also invariant under this transformation. Because capacity is a function of the structure of the channel singular-value distribution, the metric should be sensitive to this structure.

The channel capacity is a function of $\mathbf{H}\mathbf{H}^{\dagger}$. A natural metric would employ the distance between the capacity for two channel matrices at the same average total received power, that is, the same $a_2 P_o$,

$$\begin{aligned} \Delta C_{\mathrm{UT}} = & \log_2 \left| \mathbf{I} + \frac{a^2 P_o}{n_T} \overline{\mathbf{H}}_a \overline{\mathbf{H}}_a^{\dagger} \right| \\ & - \log_2 \left| \mathbf{I} + \frac{a^2 P_o}{n_T} \overline{\mathbf{H}}_b \overline{\mathbf{H}}_b^{\dagger} \right|. \end{aligned} \tag{30}$$

However, there are two problems with this definition. First, the difference is a function of P_o. Second, there is degeneracy in $\mathbf{H}$ singular values that gives a particular capacity. To address the first issue, the difference can be investigated in a high SNR limit, giving

$$\begin{aligned} \Delta C_{\mathrm{UT}} & \approx \log_2 \left| \overline{\mathbf{H}}_a \overline{\mathbf{H}}_a^{\dagger} \right| - \log_2 \left| \overline{\mathbf{H}}_b \overline{\mathbf{H}}_b^{\dagger} \right| \\ & = \sum_{m=1}^{\min(n_T, n_R)} \log_2 \lambda_m (\overline{\mathbf{H}}_a \overline{\mathbf{H}}_a^{\dagger}) - \log_2 \lambda_m (\overline{\mathbf{H}}_b \overline{\mathbf{H}}_b^{\dagger}), \end{aligned} \tag{31}$$

where $\lambda_m(\mathbf{X})$ indicates the mth largest eigenvalue of $\mathbf{X}$. To increase the sensitivity to the shape of the eigenvalue distribution, the metric is defined to be the Euclidean difference, assuming that each eigenvalue is associated with an orthogonal dimension, giving

$$\begin{aligned} & \delta^2 (\bar{\mathbf{H}}_a, \bar{\mathbf{H}}_b) \\ & \equiv \sum_{m=1}^{\min(n_T, n_R)} [\log_2 \lambda_m (\bar{\mathbf{H}}_a \bar{\mathbf{H}}_a^{\dagger}) - \log_2 \lambda_m (\bar{\mathbf{H}}_b \bar{\mathbf{H}}_b^{\dagger})]. \end{aligned} \tag{32}$$

Fractional Receiver Loss Metric

In this section we introduce a power-weighted mean $\cos^2\theta$ metric. The metric takes into account both the eigenvalue and eigenvector structure of the channels. It is motivated by the effect of receive-beamformer mismatch on capacity. Starting with Equation 8, the low SNR uninformed transmitter capacity approximation is given by

$$\begin{aligned} C &= \log_2 \left| \mathbf{I} + \frac{P_o}{n_T} \mathbf{H}\mathbf{H}^{\dagger} \right| \\ &\approx \log_2 (e) \, \mathrm{tr} \left\{ \frac{P_o}{n_T} \mathbf{H}\mathbf{H}^{\dagger} \right\} \\ &= \log_2 (e) \frac{P_o}{n_T} \sum_m \mathbf{h}_m^{\dagger} \mathbf{h}_m \\ &= \log_2 (e) \frac{P_o}{n_T} \sum_m \left\| \mathbf{w}_m^{\dagger} \mathbf{h}_m \right\|^2 \\ \mathbf{w}_m &\equiv \frac{\mathbf{h}_m}{\|\mathbf{h}_m\|}, \end{aligned} \tag{33}$$

where $\mathbf{h}_m$ is the column of the channel matrix associated with transmitter m, and $\|\ldots\|$ indicates the l_2 norm. In the low SNR limit, the optimal receive beamformer is given by the matched response given in $\mathbf{w}_m$. If some other beamformer is employed, labeled $\mathbf{w}'_m$, then signal energy is lost, adversely affecting the capacity,

$$C' \approx \log_2 (e) \frac{P_o}{n_T} \sum_m \left\| \mathbf{w}'^{\dagger}_m \, \mathbf{h}_m \right\|^2. \tag{34}$$

One possible reason that a beamformer might use the wrong matched spatial filter is channel nonstationarity. The fractional capacity loss is given by

$$\begin{aligned} \frac{C'}{C} &\approx \frac{\sum_m \left\| \mathbf{w}'^{\dagger}_m \, \mathbf{h}_m \right\|^2}{\sum_m \|\mathbf{h}_m\|^2} = \frac{\sum_m \|\mathbf{h}_m\|^2 \left\| \frac{\mathbf{h}'^{\dagger}_m}{\|\mathbf{h}'_m\|} \frac{\mathbf{h}_m}{\|\mathbf{h}_m\|} \right\|^2}{\sum_m \|\mathbf{h}_m\|^2} \\ &= \frac{\sum_m \|\mathbf{h}_m\|^2 \cos^2 \theta_m}{\sum_m \|\mathbf{h}_m\|^2}, \end{aligned} \tag{35}$$

which is the power-weighted mean $\cos^2\theta_m$ estimate, where $\cos\theta_m$ is defined to be the inner product between the "good" and "bad" unit-norm array responses for the mth transmitter. It is generally desirable for metrics to be symmetric with respect to $\mathbf{H}$ and $\mathbf{H}'$, thus avoiding moral attributions with regard to channel matrices. Using the previous discussion as motivation, a symmetric form is given by

$$\gamma(\mathbf{H},\mathbf{H}') \equiv \frac{\sum_m \|\mathbf{h}_m\| \, \|\mathbf{h}'_m\| \cos^2\theta_m}{\sum_m \|\mathbf{h}_m\| \, \|\mathbf{h}'_m\|}, \qquad (36)$$

where the "power-weighted" expectation is evaluated over transmitters.

Singular Values

The singular-value distribution of $\mathbf{H}$, or the related eigenvalue distribution of $\mathbf{H}\mathbf{H}^\dagger$, is a useful tool for understanding the expected performance of MIMO communication systems. From the discussion earlier, we can see that the channel capacity is a function of channel singular values, but not the singular-vector structure of the channel. Thus channel phenomenology can be investigated by studying the statistics of channel singular-value distributions.

Channel Parameterization

A commonly employed model assumes the channel is proportional to a matrix $\mathbf{G}$, where the entries are independently drawn from a unit-norm complex circular Gaussian distribution. While the distribution is convenient, it does suffer from a singular-value distribution that is overly optimistic for many environments. As was previously discussed, one solution is to introduce spatial correlations using the transformation $\mathbf{F} = b\mathbf{M}_L\mathbf{G}\mathbf{M}_R^\dagger$ [29]. While this approach is limited, it produces simply more realistic channels than the uncorrelated Gaussian model. The spatial correlation matrices can be factored so that $\mathbf{M}_L = \mathbf{U}\mathbf{A}_{\alpha_L}\mathbf{U}^\dagger$ and $\mathbf{M}_R = \mathbf{V}\mathbf{A}_{\alpha_R}\mathbf{V}^\dagger$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and $\mathbf{A}_{\alpha_L}$ and $\mathbf{A}_{\alpha_R}$ are positive-semidefinite diagonal matrices.

Assuming that the number of transmit and receive antennas are equal and have similar spatial correlation characteristics, the diagonal matrices can be set equal, $\mathbf{A}_\alpha = \mathbf{A}_{\alpha_L} = \mathbf{A}_{\alpha_R}$, producing the new random channel matrix $\mathbf{F}$, where

$$\begin{aligned}\mathbf{F} &= b\mathbf{U}\mathbf{A}_\alpha\mathbf{U}^\dagger\mathbf{G}'\mathbf{V}\mathbf{A}_\alpha\mathbf{V}^\dagger \\ &= b\mathbf{U}\mathbf{A}_\alpha\mathbf{G}\mathbf{A}_\alpha\mathbf{V}^\dagger\end{aligned} \qquad (37)$$

and

$$\mathbf{A}_\alpha = \sqrt{n}\,\frac{\mathrm{diag}\{\alpha^0,\alpha^1,\ldots,\alpha^{n-1}\}}{\sqrt{\mathrm{tr}\,(\mathrm{diag}\{\alpha^0,\alpha^1,\ldots,\alpha^{n-1}\}^2)}}, \qquad (38)$$

where b is used to set overall scale, n is given by the size of $\mathbf{A}_\alpha$, and $\mathbf{U}$ and $\mathbf{V}$ indicate random unitary matrices. Used here is the fact that arbitrary unitary transformations do not affect the statistics of the Gaussian matrix. The form of $\mathbf{A}_\alpha$ given here is somewhat arbitrary, but has the satisfying characteristics that as $\alpha \to 0$, a rank-one channel matrix is produced, and as $\alpha \to 1$, a spatially uncorrelated Gaussian matrix is produced. Furthermore, empirically this model provides good fits to experimental distributions. The normalization for $\mathbf{A}_\alpha$ is chosen so that the expected value of $\|\mathbf{F}\|_F^2$ is $b^2 n_T n_R$, where $\|\cdots\|_F^2$ indicates the Frobenius norm.

Channel Parameter Estimation

An estimate for $\hat{\alpha}$ associated with particular transmit and receive locations is given by minimizing the mean-square metric given in Equation 32,

$$\hat{\alpha} = \arg\min\left\langle \delta^2[\hat{\mathbf{H}},\hat{\mathbf{F}}(\alpha)]\right\rangle, \qquad (39)$$

where $\hat{\mathbf{X}}$ indicates the estimated value of $\mathbf{X}$. Here the expectation, denoted by $\langle\cdots\rangle$, indicates averaging is over an ensemble of $\mathbf{F}$ for a given α and an ensemble of $\mathbf{H}$ for given transmit and receiver sites.

It is worth noting that this approach does not necessarily provide an unbiased estimate of α. Estimates of α, using the metric introduced here, are dependent upon the received SNR. Data presented later in the article have sufficiently high SNR such that α can be estimated within ±0.02.

Space-Time Low-Density Parity-Check Codes

This section of the article introduces low-density parity-check (LDPC) codes, which were studied extensively by R.G. Gallager [31]. The significance of modern implementations of LDPC codes rests on iterative decoding algorithms that, for LDPC codes, are applications of techniques formulated for Bayesian belief networks, which are introduced below. This section also discusses a simple application of LDPC to space-time codes.

BAYESIAN BELIEF NETWORKS

Graphical decoding is often based on the Bayesian belief networks popularized by Pearl, in the context of machine learning, in a well-known monograph [1]. An interpretation of various decoding algorithms in terms of Bayesian belief networks is presented elsewhere [2].

To appreciate the use of belief networks for decoding, consider the probability density function denoted $p(x_1, x_2, x_3, x_4)$ in Figure A. This function factors in the manner shown in the figure, expressing simpler variable dependencies than those allowed by the multivariate notation $p(x_1, \ldots, x_l)$. The factorization can be represented by a directed acyclic graph as shown, with directed arrows expressing conditional probabilities of the more general form $p(x|u_1, \ldots, u_l)$.

For decoding purposes, each node in the graph maintains an alphabet (for example, the symbol alphabet for coding applications) and several (probability) distributions over this alphabet. One probability distribution, denoted $\pi(x)$, can be interpreted as a prior density on the alphabet while another (nonnegative, but not a normalized density) distribution, denoted $\lambda(x)$, can be interpreted as a likelihood function on the alphabet.

In addition, each node keeps track of a belief function that is the product of priors and likelihoods: $\pi(x)\,\lambda(x)$. The maximum of the belief function can be used as a decision on the value of the node's alphabet.

To evaluate a consistent set of distribution functions, messages are received and transmitted from each node. Messages that flow from parent to child are denoted $\pi_k^P(u_k)$ and are treated as if they were priors, while messages that flow from child to parent are denoted $\lambda_k^C(x)$ and are treated as if they were likelihoods. At each node, messages received from parents and children are used to update the internal (for that node) prior and likelihood functions $\pi(x)$ and $\lambda(x)$ for the node's alphabet.

Nodes are activated in any order, subject only to the requirement that all incoming messages are available. When a node is activated, it calculates its internal prior and likelihood functions and then makes its messages available to its parent and child nodes. Initial settings of the internal functions are provided (but not shown in the figure) to enable the process to start.

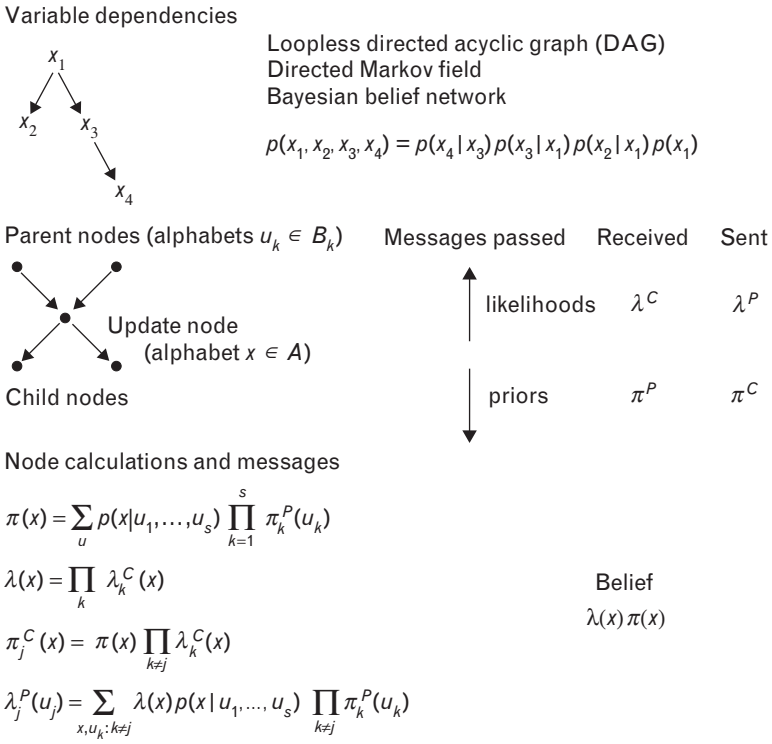
Low-Density Parity-Check Codes

LDPC codes were developed by Gallager, who studied their distance properties and decoding in a well-known monograph [31]. With the advent of graphical decoding techniques, soft-decision decoding of LDPC codes became practical, resulting in renewed interest in these codes. Subsequent developments in code design and decoding have led to codes that achieve levels of performance astonishingly close to the Shannon capacity [32], albeit at the cost of extremely long codewords. However, decoding complexity of LDPC codes scales linearly (with a fixed number of iterations) with the code length, making relatively long codes practical.

LDPC codes are linear block codes defined by a parity-check matrix. Each symbol in the codeword is involved in only a few parity-check equations. Consequently, most entries in the parity-check matrix are zero. Regular LDPC codes have n_C parity-check equations for each symbol, and each parity-check equation involves n_R symbols. Thus, if the dimensionality of the parity-check matrix is $r \times c$, we have $rn_R = cn_C$ for regular LDPC codes. As an example, the LDPC code used for some of the experiments described later satisfies $(r, c) = (512, 1024)$ and $(n_R, n_C) = (8, 4)$. More powerful codes that are not regular are also known [33].

LDPC Decoding

Recently, graphical decoding techniques have motivated practical code design. Bayesian belief networks [34] can be used to formulate decoders for LDPC codes and turbo codes (the sidebar entitled "Bayesian Belief Net-



For loopless graphs, the order of activation does not matter and the process converges. Unfortunately, for decoding applications, interesting graphs have loops, so order of activation matters and convergence is not guaranteed. Typically, nodes are activated in a repetitive pattern for a certain number of iterations until a stopping criterion is met. Symbol decisions are based on the belief function.

References

1. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, Calif., 1988).
2. R.J. McEliece, D.J.C. MacKay, and J.-F. Cheng, "Turbo Decoding as an Instance of Pearl's 'Belief Propagation' Algorithm," *IEEE J. Sel. Areas Commun.* **16** (2), 1998, pp. 140–152.

FIGURE A. Bayesian belief networks provide a framework for representing conditional probabilities in a graphical manner. Each node has a symbol alphabet on which it maintains a belief function that factors as a product of a prior-like function and a likelihood-like function. Beliefs are updated by passing messages among nodes in a manner suggested by the terminology. Initial states and a node update order must be chosen. Only in special cases do the iterations converge to a Bayesian decision, but for many interesting applications, the iterative technique is both practical and effective. Turbo codes and low density parity-check codes have decoders based on this paradigm.

works" provides more information). However, beyond connecting the decoding algorithm of LDPC codes to Bayesian belief networks, a thorough explanation of the steps in this algorithm is outside the scope of this article; we present only a concise summary.

For LDPC codes, Figure 2 shows a graph illustrating data and parity-check dependencies for the codewords. In general, each nonzero entry in the parity-check matrix indicates the edge of a graph connecting a parity-check node (row index) and a codeword symbol (column index). The example in Figure 2 is a single parity-check code on four symbols. The graph shows the symbol nodes $c_1,\ldots, c_4$, the data nodes $z_1,\ldots, z_4$, and the parity-check nodes, labeled by zeroes. Each edge between a parity-check node and a symbol node corresponds to a nonzero entry in the parity-check matrix.

Decoding occurs by treating the graph as a Bayesian belief network using the conditional probabilites $p(z_k \mid c_k)$, which express the likelihood ratios, and

$$p(0 \mid c_{i_1},\ldots,c_{i_l}) = \delta\left(\sum\nolimits_k c_{i_k}\right),$$

which expresses the parity-check relation. The resulting algorithm can be viewed as sweeping through the rows and columns of the parity-check matrix, updating likelihood ratios l_k for each nonzero entry in the matrix. The notation below denotes l_{i_j} as the likelihood ratio stored with the i_j-th (nonzero) entry in a fixed (for the given step) row or column of the parity-check matrix. In this form, the iterative steps of the algorithm are summarized for the simple case of a binary symbol alphabet by the equations:

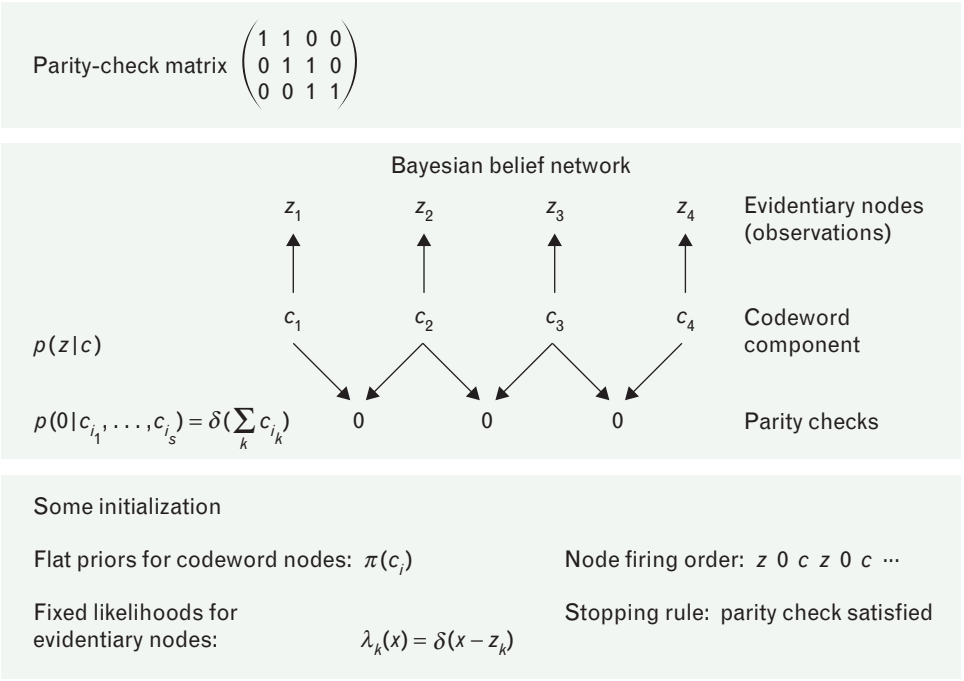


FIGURE 2. Application of Bayesian belief networks to low-density parity-check codes. Soft-decision decoding of low density parity-check codes can be based on Bayesian belief networks. Both the redundancies in codewords c_k and the relationship between the codewords and the data z_k can be represented graphically. The data-codeword relationship is expressed through the probability densities $p(z|c)$, which are assumed to be independent sample-to-sample. Redundancies in the codewords are expressesed in a similar notation as $p(0|c_{i_1},\ldots, c_{i_s})$ where the symbols c_{i_k}, $1 \le k \le s$, are involved in a parity check. In this manner, all depedancies are expressed graphically through conditional probability densities as required for the formalism of Bayesian belief networks.

1. Row sweeps

$$\tanh\left[\frac{l_{i_k}^{(\text{new})}}{2}\right] = \prod_{j\neq k} \tanh\left[\frac{l_{i_j}^{(\text{old})}}{2}\right]$$

2. Column sweeps [column c with log-likelihood ratio $l_c^{(\text{LLR})}$]

$$l_{i_k}^{(\text{new})} = \sum_{j\neq k} l_{i_j}^{(\text{new})} + l_c^{(\text{LLR})}$$

3. Bit decisions (column c)

$$\text{sign}\left[\sum_k l_{i_k} + l_c^{(\text{LLR})}\right].$$

For the code used in the experiments, each row sweep involves eight l_{i_j} per row and each column sweep four l_{i_j} per column.

Each of the row (column) operations is independent of any other row (column operation) and hence can be implemented in any order or in parallel. This allows considerable acceleration of hardware decoders. Decoding can be halted after a fixed number of iterations or after the parity-check equations are satisfied.

Some simplifications that are not possible for nonbinary symbol alphabets are involved in the binary case. In this more general context, the row/column sweeps are expressed by:

1. Row sweeps

$$\pi(\beta_{i_k})\,\mathbf{p}_{i_k}^{(\text{new})} = 2^n \mathbf{U}_n \square_{j\neq k}\left[\mathbf{U}_n\,\pi(\beta_{i_j})\,\mathbf{p}_{i_j}^{(\text{old})}\right]$$

2. Column sweeps

$$\mathbf{p}_{i_k}^{(\text{new})} \propto \left[\square_{j\neq k}\,\mathbf{p}_{i_j}^{(\text{old})}\right]\square\,\mathbf{p}_c^{(\text{LF})}$$

3. Symbol decisions

$$\left(\square_m \mathbf{p}_{i_m}\right)\square\,\mathbf{p}_c^{(\text{LF})}.$$

The components of the vector $\mathbf{p}_k$ express probabilities for the values of the kth symbol, the permutation π_k indicates the effect of a particular nonbinary coefficient in the parity-check equation, $\mathbf{U}_n$ is a Walsh-Hadamard matrix, and the notation $\square$ denotes the Hadamard (component by component) product.

Space-Time Extension LDPC

There are a variety of extensions of LDPC codes to space-time codes, which are introduced and explained in the sidebar entitled "Space-Time Codes." For the experiments described below, only one type of extension was considered.

Each space-time channel transmits one of several possible quadrature phase-shift keying (QPSK) waveforms with slightly offset carrier frequencies. The differential frequencies are sufficiently large to effectively decorrelate the transmitted waveforms over the length of a codeword (1024 bits) even if the data sequences in each channel are identical. These differential frequencies are also large compared to the expected Doppler spreads and small compared to the signal bandwidth.

In the simplest example of such a code, the I and Q components of a transmitter represent, respectively, two different LDPC codewords. Each transmitter sends the same complex baseband sequence (QPSK) shifted in frequency. The transmitter outputs, viewed collectively as a vector at any instant, vary in time and thus effectively probe the environment characterized by the channel matrix. Since the transmitted vector varies significantly over the duration of a codeword, the coding provides spatial diversity. Decoding occurs by forming likelihood ratios based on channel-matrix estimates and then using the iterative decoder described above. Note that the channel matrix can change during the codeword, in which case channel-matrix estimates can vary sample to sample.

The LDPC space-time code just described exhibits full spatial redundancy among all transmitters. Less redundancy, and therefore higher data rates, can be achieved by dividing the transmitters into subsets, each of which is fully redundant yet different from any other subset. For example, the space-time code discussed later, in the section on experiments, uses four transmitters. The first two transmitters send two bits (redundant in I and Q) of a symbol of an LDPC codeword over GF(16). The remaining two transmitters send the other two bits of the same symbol. Decoding is based on likelihood functions built over GF(16) using estimates of the channel matrices. Again, differential frequencies among transmitters enable spatial diversity.

Space-Time Turbo Code and Multichannel Multiuser Detectors

While the theoretical performance is determined by the channel phenomenology, practical MIMO performance requires the selection of a space-time code and an appropriate matched receiver. In this section we discuss the space-time turbo code used in this example. We develop a maximum-likelihood formulation of a multiple-antenna multiuser receiver, and we discuss suboptimal implementations of the receiver. We also introduce minimum-mean-squared-error extensions of the receiver, and we discuss the value and use of training data.

Space-Time Turbo Code

Turbo codes, introduced elsewhere [35], illustrate that codes constructed with simple components, such as with interleavers and convolutional encoders, combined with an iterative decoding process can achieve near-Shannon capacity performance. The iterative decoding process, taking advantage of information exchange among component decoders, provides a feasible way to approach optimal performance. For each component decoder, the best decoding algorithm is the maximum *a posteriori* (MAP) algorithm or the BCJR algorithm [36], which is derived from the MAP principle. Modifications of the MAP algorithm include log-MAP and max-log-MAP [37]. Recently, implementation of turbo decoders has been carried out and high data-rate decoding is possible [38].

A number of space-time extensions of turbo coding have been suggested [21, 22]. The approach used here, which was introduced elsewhere [39], provides a 2-bit/sec/Hz link for a 4×4 MIMO system with independent QPSK waveforms from each transmitter. A single data stream is turbo encoded and the encoded data stream is distributed redundantly amongst the transmitters. The turbo encoder employs a rate-1/3, 16-state convolutional encoder twice with two different 4096-bit random interleavers. The distribution of systematic bits is such

SPACE-TIME CODES

Space-time codes are used with multiple transmitters to provide spatial as well as temporal redundancy in the data received by an array of antennas. There are two basic approaches to space-time coding. In the first approach, the transmitter can be informed of the propagation channel by the receiver and thus adjust its coding accordingly. This approach offers the largest information-theoretic capacity but can be difficult to accomplish in a dynamic environment. The second approach, which is taken here, uses fixed codes of various rates that offer good performance on average (over all channels). These codes share transmitted power equally among all spatial channels.

The number of different types of space-time codes is too large to provide a useful overview here. Instead we briefly describe two important categories of space-time codes that are not treated in the text.

Block Orthogonal Codes

For data $\mathbf{Z}$ and channel matrix $\mathbf{H}$, consider a set of matrix symbols $\mathbf{S}$ contained in S. The information bits are encoded in matrices that are constrained to lie in the class S. This class is defined by the property that $\mathbf{SS}^\dagger$ is proportional to the identity matrix with a fixed (independent of $\mathbf{S}$) proportionality constant. The maximum-likelihood decision for $\mathbf{S}$ is based on finding

$$\begin{aligned}&\underset{\mathbf{S}\in S}{\operatorname{argmin}} \left\| \mathbf{Z} - \mathbf{HS} \right\|^2 \\ &\quad = \underset{\mathbf{S}\in S}{\arg\max}\, \operatorname{Re}\, \operatorname{tr}(\mathbf{ZS}^\dagger \mathbf{H}^\dagger),\end{aligned}$$

which involves a linear function in the entries of $\mathbf{S}$. For some simple classes S, linearity of the likelihood function decouples decisions on the data symbols. For example, consider the Alamouti code [1].

$$S = \left\{ \mathbf{S} : \mathbf{S} = \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix} \right\}.$$

The information symbols s_1 and s_2 are sent redundantly over both channels. The likelihood function is linear in each s_k, decoupling demodulation decisions.

Another example of an orthogonal matrix code is

$$\mathbf{S} = [\mathbf{S}_0 \mathbf{S}_0^*] \text{ with}$$

$$\mathbf{S}_0 = \begin{pmatrix} s_1 & -s_2 & -s_3 & -s_4 \\ s_2 & s_1 & s_4 & -s_3 \\ s_3 & -s_4 & s_1 & s_2 \\ s_4 & s_3 & -s_2 & s_1 \end{pmatrix}.$$

Space-Time Trellis Codes

Figure A provides an example of a space-time trellis code. A pair of bits (I_t^1, I_t^2) at time t enters a convolutional encoder with integer coefficients a_k^p and b_k^p at the pth lag in the kth channel. The input bits are interpreted as the integers 0 or 1. Computations occur modulo 4 and result in an integer value between 0 and 3 for each channel. A fixed mapping between these four integers and the quadrature phase-shift keying (QPSK) alphabet completes the coding and modulation.

The trellis code is defined by the coefficients $\{a_k^p, b_k^p\}$. These are often chosen under one of several design criteria, also shown in the figure. Each codeword is a matrix symbol $\mathbf{C}$. The probability of an error in deciding between two such

that each systematic bit is sent twice on two different transmitters. The parity bits are sent once, distributed randomly amongst the transmitters. The difference in weighting between the systematic and parity bits provides an effective puncturing of the code. Because more energy is dedicated to systematic bits, remodulation errors have a reduced effect on subtraction performance, in principle improving the performance of the iterative multiuser detection for a given bit error rate.

Multichannel Multiuser Detector

The multichannel multiuser detector (MCMUD) algorithm, discussed elsewhere [3, 39, 40], is a minimum-mean-squared-error (MMSE) extension to an iterative implementation of a maximum-likelihood multiple-antenna receiver. The MCMUD algorithm employed for this analysis iteratively combines a blind space-time-frequency adaptive beamformer with a multiuser detector.

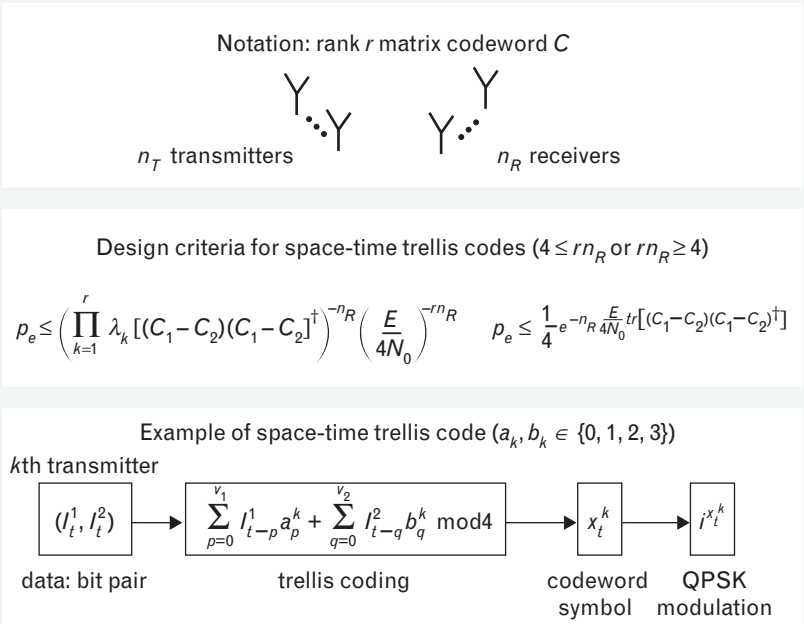


FIGURE A. Space-time trellis codes introduce spatial as well as temporal redundancy in the transmitted data. Code design often involves a pruned search over a class of codes based on a simple figure of merit. For example, the minimum least-squares distance between codewords (represented by space-time matrices C_k) can be maximized. In the example shown, an alphabet consisting of the integers modulo 4 is used for convolutional encoding at each transmitter. The resulting output symbols are mapped to a QPSK alphabet. The coefficients a_k, b_k determine the code. Note that the spectral efficiency is 2 bits/sec/Hz.

symbols can be bounded by (Bhattacharyya bound)

$$p_e \le e^{-\frac{E}{4N_0}(\mathbf{C}_1-\mathbf{C}_2)^\dagger \mathbf{H}^\dagger \mathbf{H}(\mathbf{C}_1-\mathbf{C}_2)}.$$

The approximation $\mathbf{H}^\dagger\mathbf{H} \approx n_R I_{n_T}$ motivates one of the design criteria shown in the figure. Integrating over $\mathbf{H}$ motivates the other.

In both cases r denotes the rank of the matrix difference $\mathbf{C}_1 - \mathbf{C}_2$. Constrained searches over the code coefficients are commonly used to find codes with the smallest possible error between closest codewords under either criterion. When $4 \ge rn_R$, it is important to ensure that the rank of the matrix difference is not too small. When $4 < rn_R$, maximizing the Euclidean distance between the two codewords $\mathbf{C}_k$ becomes important.

Reference

1. S.M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE J. Sel. Areas Commun.*, **16** (8), 1998, pp. 1451–1458.

We present here the results of the maximum likelihood (ML) formulation of MCMUD, employing a quasistatic narrowband MIMO-channel model. The number of receive antennas n_R by number of samples, n_s data matrix, $\mathbf{Z} \in \mathbb{C}^{n_R \times n_s}$, is given by

$$\mathbf{Z} = \mathbf{HT} + \mathbf{N}, \tag{44}$$

where the channel matrix $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ contains the complex attenuation between each transmit antenna and receive antenna; $\mathbf{T} \in \mathbb{C}^{n_T \times n_s}$ is the transmitted sequence; and $\mathbf{N} \in \mathbb{C}^{n_R \times n_s}$ is additive Gaussian interference plus noise. The probability density for a multivariate signal-in-the-mean model is given by

$$p(\mathbf{Z}|\mathbf{R},\mathbf{H},\mathbf{T}) = \frac{e^{-\mathrm{tr}\left\{(\mathbf{Z}-\mathbf{HT})^\dagger \mathbf{R}^{-1}(\mathbf{Z}-\mathbf{HT})\right\}}}{\pi^{n_s n_R}|\mathbf{R}|^{n_s}}, \tag{41}$$

where $\mathbf{R}$ indicates the spatial covariance matrix of the interference plus noise, $|\cdots|$ indicates the determinant of a matrix, † indicates the Hermitian conjugate, and tr indicates the trace of a matrix. Maximizing the probability density with respect to $\mathbf{H}$ is equivalent to minimizing the tr$\{\cdots\}$ in Equation 41,

$$\mathrm{tr}\left\{(\mathbf{Z}-\mathbf{HT})(\mathbf{Z}-\mathbf{HT})^\dagger \mathbf{R}^{-1}\right\}, \tag{42}$$

which is satisfied by

$$\hat{\mathbf{H}} = \mathbf{ZT}^\dagger(\mathbf{TT}^\dagger)^{-1}, \tag{43}$$

assuming $\mathbf{TT}^\dagger$ is not rank deficient. Substituting $\hat{\mathbf{H}}$,

$$p \to \frac{e^{-\mathrm{tr}\left\{\mathbf{Z}\mathbf{P}_\mathbf{T}^\perp \mathbf{Z}^\dagger \mathbf{R}^{-1}\right\}}}{\pi^{n_s n_R}|\mathbf{R}|^{n_s}}, \tag{44}$$

where the matrix $\mathbf{P}_\mathbf{T} \equiv \mathbf{T}^\dagger(\mathbf{TT}^\dagger)^{-1}\mathbf{T}$ projects onto the

row space spanned by $\mathbf{T}$, and $\mathbf{P}_{\mathbf{T}}^{\perp} = \mathbf{I}_{n_s} - \mathbf{P}_{\mathbf{T}}$ projects onto the orthogonal complement of the row space of $\mathbf{T}$. Maximizing with respect to an internal parameter of $\mathbf{R}$ gives

$$\mathrm{tr}\{\mathbf{R}^{-1}\mathbf{Z}\mathbf{P}_T^{\perp}\mathbf{Z}^{\dagger}\mathbf{R}^{-1}\dot{\mathbf{R}}\} - n_s\mathrm{tr}\{\mathbf{R}^{-1}\dot{\mathbf{R}}\} = 0, \quad (45)$$

where $\dot{\mathbf{R}}$ indicates the derivative of $\mathbf{R}$ with respect to some internal parameter. This relationship is satisfied when

$$\hat{\mathbf{R}} = \frac{\mathbf{Z}\mathbf{P}_{\mathbf{T}}^{\perp}\mathbf{Z}^{\dagger}}{n_s}, \quad (46)$$

assuming that $\mathbf{R}$ is not rank deficient. Using these results, the ML statistic for estimating $\mathbf{T}$ is given by

$$\max_{\mathbf{R},\mathbf{H}} p(\mathbf{Z}|\mathbf{R},\mathbf{H},\mathbf{T}) = \left(\frac{\pi e}{n_s}\right)^{-n_s n_R} |\mathbf{Z}\mathbf{P}_{\mathbf{T}}^{\perp}\mathbf{Z}^{\dagger}|^{-n_s}. \quad (47)$$

The determinant of $\mathbf{Z}\mathbf{P}_{\mathbf{T}}^{\perp}\mathbf{Z}^{\dagger}$ is minimized to demodulate the signals for all transmitters jointly.

Although it is theoretically possible to use the statistic $\mathbf{Z}\mathbf{P}_{\mathbf{T}}^{\perp}\mathbf{Z}^{\dagger}$ directly for demodulation, an iterative approach is much more practical. We define $\mathbf{T} \equiv (\mathbf{T}_{\mathbf{A}}^{\dagger}\mathbf{T}_{\mathbf{B}}^{\dagger})^{\dagger}$ to be a partitioned form of $\mathbf{T}$, where the $n_A \times n_s$ matrix $\mathbf{T}_{\mathbf{A}}$ contains the signals associated with a particular subset of n_A transmit antennas and the $(n_T - n_A) \times n_s$ matrix $\mathbf{T}_{\mathbf{B}}$ contains the signals associated with all other transmit antennas. By factoring $\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp} = \mathbf{X}^{\dagger}\mathbf{X}$, the rows of $\mathbf{X}$ form an orthonormal basis for the complement of the row space of $\mathbf{T}_{\mathbf{B}}$ such that $\mathbf{X}\mathbf{X}^{\dagger} = \mathbf{I}$, where the symmetric identity matrix has a dimension of n_s minus the number of rows in $\mathbf{T}_{\mathbf{B}}$. By defining $\mathbf{Z}_{\mathbf{X}} \equiv \mathbf{Z}\mathbf{X}^{\dagger}$ and $\mathbf{T}_{\mathbf{X}} \equiv \mathbf{T}_{\mathbf{A}}\mathbf{X}^{\dagger}$, we can show that

$$\mathbf{Z}\mathbf{P}_{\mathbf{T}}^{\perp}\mathbf{Z}^{\dagger} = \mathbf{Z}_{\mathbf{X}}\mathbf{P}_{\mathbf{T}_{\mathbf{X}}}^{\perp}\mathbf{Z}_{\mathbf{X}}^{\dagger}. \quad (48)$$

The determinant can be factored into terms with and without reference to $\mathbf{T}_{\mathbf{A}}$,

$$\left|\mathbf{Z}_{\mathbf{X}}\mathbf{P}_{\mathbf{T}_{\mathbf{X}}}^{\perp}\mathbf{Z}_{\mathbf{X}}^{\dagger}\right| = \left|\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathbf{X}}^{\dagger}\right|\left|\mathbf{I}_{n_s} - \mathbf{P}_{\mathbf{T}_{\mathbf{X}}}\mathbf{P}_{\mathbf{Z}_{\mathbf{X}}}\right|. \quad (49)$$

Because the first term is free of $\mathbf{T}_{\mathbf{A}}$, demodulation is performed by minimizing the second term. This form suggests an iterative approach, where the signal associated with each transmitter, in turn, is considered to be user $\mathbf{A}$ and is demodulated by minimizing $\left|\mathbf{I}_{n_s} - \mathbf{P}_{\mathbf{T}_{\mathbf{X}}}\mathbf{P}_{\mathbf{Z}_{\mathbf{X}}}\right|$.

If $\mathbf{T}_{\mathbf{A}}$ is a row vector, such that $n_A = 1$, then the second term can be simplified and interpreted in terms of a beamformer

$$\begin{aligned}\left|\mathbf{I}_{n_s} - \mathbf{P}_{\mathbf{T}_{\mathbf{X}}}\mathbf{P}_{\mathbf{Z}_{\mathbf{X}}}\right| &= 1 - (\mathbf{T}_{\mathbf{X}}\mathbf{T}_{\mathbf{X}}^{\dagger})^{-1}\mathbf{T}_{\mathbf{X}}\mathbf{P}_{\mathbf{Z}_{\mathbf{X}}}\mathbf{T}_{\mathbf{X}}^{\dagger}, \\ &= 1 - \frac{\mathbf{w}_{\mathbf{A}}^{\dagger}\mathbf{Z}_{\mathbf{X}}\mathbf{T}_{\mathbf{X}}^{\dagger}}{n_s}, \\ &= 1 - \frac{\mathbf{w}_{\mathbf{A}}^{\dagger}\mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}}{n_s},\end{aligned} \quad (50)$$

where

$$\begin{aligned}\mathbf{w}_{\mathbf{A}} &= \hat{\mathbf{R}}_{\mathbf{X}}^{-1}\hat{\mathbf{H}}_{\mathbf{A}}, \\ \hat{\mathbf{R}}_{\mathbf{X}} &\equiv \frac{1}{n_s}\mathbf{Z}_{\mathbf{X}}\mathbf{Z}_{\mathbf{X}}^{\dagger} = \frac{1}{n_s}\mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{Z}^{\dagger}, \\ \hat{\mathbf{H}}_{\mathbf{A}} &\equiv \mathbf{Z}_{\mathbf{X}}\mathbf{T}_{\mathbf{X}}^{\dagger}(\mathbf{T}_{\mathbf{X}}\mathbf{T}_{\mathbf{X}}^{\dagger})^{-1} \\ &= \mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1}.\end{aligned} \quad (51)$$

The $n_R \times 1$ vector $\mathbf{w}_{\mathbf{A}}$ contains the receive beamforming weights, $\hat{\mathbf{R}}_{\mathbf{X}}$ is the interference-mitigated signal-plus-noise covariance matrix estimate, and $\hat{\mathbf{H}}_{\mathbf{A}}$ is the channel estimate associated with $\mathbf{T}_{\mathbf{A}}$. It is worth noting that the form for $\hat{\mathbf{H}}_{\mathbf{A}}$ is simply the column of $\hat{\mathbf{H}}$, given in Equation 43, associated with $\mathbf{T}_{\mathbf{A}}$.

$$\begin{aligned}\hat{\mathbf{H}} &= \mathbf{Z}\mathbf{T}^{\dagger}(\mathbf{T}\mathbf{T}^{\dagger})^{-1} \\ &= \mathbf{Z}\begin{pmatrix}\mathbf{T}_{\mathbf{A}}^{\dagger} & \mathbf{T}_{\mathbf{B}}^{\dagger}\end{pmatrix}\begin{pmatrix}\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger} & \mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{B}}^{\dagger} \\ \mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{A}}^{\dagger} & \mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}}^{\dagger}\end{pmatrix}^{-1} \\ &\equiv \mathbf{Z}\begin{pmatrix}\mathbf{T}_{\mathbf{A}}^{\dagger} & \mathbf{T}_{\mathbf{B}}^{\dagger}\end{pmatrix}\begin{pmatrix}\mathbf{M}_{1,1} & \mathbf{M}_{1,2} \\ \mathbf{M}_{1,2}^{\dagger} & \mathbf{M}_{2,2}\end{pmatrix}^{-1} \\ &= \mathbf{Z}\begin{pmatrix}\mathbf{T}_{\mathbf{A}}^{\dagger} & \mathbf{T}_{\mathbf{B}}^{\dagger}\end{pmatrix} \\ &\cdot\begin{pmatrix}(\mathbf{M}_{1,1} - \mathbf{M}_{1,2}\mathbf{M}_{2,2}^{-1}\mathbf{M}_{1,2}^{\dagger})^{-1} & \cdots \\ -\mathbf{M}_{2,2}^{-1}\mathbf{M}_{1,2}^{\dagger}(\mathbf{M}_{1,1} - \mathbf{M}_{1,2}\mathbf{M}_{2,2}^{-1}\mathbf{M}_{1,2}^{\dagger})^{-1} & \cdots\end{pmatrix}.\end{aligned} \quad (52)$$

By focusing on the first column and substituting in for $\mathbf{M}_{1,1}$, $\mathbf{M}_{1,2}$, and $\mathbf{M}_{2,2}$, we can find $\hat{\mathbf{H}}_{\mathbf{A}}$.

$$\begin{aligned}\hat{\mathbf{H}}_{\mathbf{A}} &= \mathbf{Z}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger} - \mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{B}}^{\dagger}[\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}}^{\dagger}]^{-1}\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1} \\ &\quad - \mathbf{Z}\mathbf{T}_{\mathbf{B}}^{\dagger}[\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}}^{\dagger}]^{-1}\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{A}}^{\dagger} \\ &\quad \times(\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger} - \mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{B}}^{\dagger}[\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{B}}^{\dagger}]^{-1}\mathbf{T}_{\mathbf{B}}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1} \\ &= \mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1},\end{aligned} \quad (53)$$

which is the same form found in Equation 51.

Demodulation is performed by maximizing the magnitude of the inner product of the beamformer output $\mathbf{w}_{\mathbf{A}}^{\dagger}\mathbf{Z}$ and the interference-mitigated reference signal $\mathbf{T}_{\mathbf{A}}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}$.

Suboptimal Implementation

A variety of suboptimal but computationally more efficient variants are possible. In general, these approximations become increasingly valid as the number of samples in the block increases.

The first computational simplification is found by noting that the normalization term of the channel estimate in Equation 51 can be approximated by

$$\begin{aligned}\hat{\mathbf{H}}_{\mathbf{A}} &= \mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1} \\ &= \mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1/2} \\ &\quad\times(\mathbf{I}-[\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger}]^{-1/2}\mathbf{T}_{\mathbf{A}}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}\mathbf{T}_{\mathbf{A}}^{\dagger}[\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger}]^{-1/2})^{-1} \\ &\quad\times(\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1/2} \\ &\approx \mathbf{Z}\mathbf{P}_{\mathbf{T}_{\mathbf{B}}}^{\perp}\mathbf{T}_{\mathbf{A}}^{\dagger}(\mathbf{T}_{\mathbf{A}}\mathbf{T}_{\mathbf{A}}^{\dagger})^{-1}.\end{aligned} \tag{54}$$

(We did not assume that $\mathbf{T}_{\mathbf{A}}$ is a row vector in the previous discussion.)

The second approximation reduces the computation cost of the projection operator. The operator that projects on the orthogonal complement of the row space of $\mathbf{M}$ is given by

$$\mathbf{P}_{\mathbf{M}}^{\perp} = \mathbf{I} - \mathbf{M}^{\dagger}(\mathbf{M}\mathbf{M}^{\dagger})^{-1}\mathbf{M}. \tag{55}$$

This operator can be approximated by

$$\mathbf{P}_{\mathbf{M}}^{\perp} \approx \prod_{m}\left[\mathbf{I} - \mathbf{M}_{m}^{\dagger}(\mathbf{M}_{m}\mathbf{M}_{m}^{\dagger})^{-1}\mathbf{M}_{m}\right], \tag{56}$$

where m indicates the mth row in the matrix. By repeating the application of this approximate projection operator, we can reduce the approximation error at the expense of additional computational complexity.

MMSE Extension

Because of the effects of delay and Doppler-frequency spread, the model given in Equation 40 for the received signal is incomplete for many environments, adversely affecting the performance of the spatial-beamformer interpretation of the ML demodulator. Because turbo codes require relatively long block lengths to be effective, they are particularly sensitive to Doppler offsets. Extending the beamformer to include delay and Doppler corrects this deficiency. With this approach, the spatial-beamformer interpretation presented in Equation 50 is formally the same, but all projectors are extended to include delay and Doppler spread. The data matrix is replaced with

$$\begin{aligned}\mathbf{Z}_{STF} \equiv [&\mathbf{Z}_{\mathbf{X}}^{\dagger}(\delta t_1, \delta f_1)\,\mathbf{Z}_{\mathbf{X}}^{\dagger}(\delta t_1, \delta f_2) \\ &\cdots \mathbf{Z}_{\mathbf{X}}^{\dagger}(\delta t_1, \delta f_{n_{\delta f}}) \cdots \mathbf{Z}_{\mathbf{X}}^{\dagger}(\delta t_{n_{\delta t}}, \delta f_{n_{\delta f}})]^{\dagger},\end{aligned} \tag{57}$$

which is a $(n_R \cdot n_{\delta f} \cdot n_{\delta t}) \times n_s$ matrix that includes possible signal distortions. The new channel estimate has dimension $(n_R \cdot n_{\delta f} \cdot n_{\delta t}) \times n_T$, but $\mathbf{T}$ remains the same. The MMSE beamformer is given by

$$\begin{aligned}\mathbf{w}_{\mathrm{STF}} &= \operatorname{argmin}\left\langle \left\| \mathbf{w}_{\mathrm{STF}}^{\dagger}\mathbf{Z}_{\mathrm{STF}} - \mathbf{T}_{\mathbf{X}} \right\|^2 \right\rangle \\ &= (\mathbf{Z}_{\mathrm{STF}}\mathbf{Z}_{\mathrm{STF}}^{\dagger})^{-1}\mathbf{Z}_{\mathrm{STF}}\,\mathbf{T}_{\mathbf{X}}^{\dagger}.\end{aligned} \tag{58}$$

Figure 3 shows a diagram for this demodulator (MCMUD).

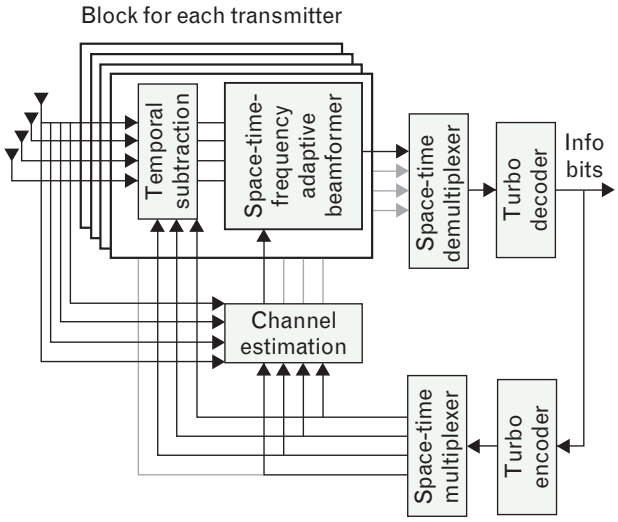


FIGURE 3. Diagram of a multichannel multiuser detector (MCMUD) space-time turbo-code receiver. The receiver iteratively estimates the channel and demodulates the signal. The space-time frequency-adaptive beamformer component compensates for spatial, delay, and frequency-offset correlations. By iteratively decoding the signal, previous signal estimates can be used to temporally remove contributions from other transmitters, which is a form of multiuser detection.

Training Data

In principle, there is no need for training data, because the channel and information can be estimated jointly. Furthermore, the use of training data competes directly with information bits. For reasonably stationary channels, the estimate for the previous frame can be employed as an initial estimate for the demodulator. However, for more quickly moving channels some training data is useful. Here, a small amount of training data is introduced within a frame (20%). This provides an initial channel estimate for the space-time-frequency adaptive beamformer.

In the experiment, knowledge of the encoded signal is used to provide that training data. Because the number of training samples is relatively small, it is useful to use a small number of temporal and frequency taps during the first iteration. Larger dimension space-time-frequency processing is possible by using estimates of the data.

Phenomenological Experiment

This section presents channel-complexity and channel-stationarity experimental results for MIMO systems. We introduce the experiments and then discuss channel mean attenuation and channel complexity. We then discuss the variation of MIMO channels as a function of time and as a function of frequency.

Experimental System

The employed experimental system is a slightly modified version of the system used previously at Lincoln Laboratory [3, 41]. The transmit array consists of up to eight arbitrary waveform transmitters. The transmitters can support up to a 2-MHz bandwidth. These transmitters can be used independently, as two groups of four coherent transmitters, or as a single coherent group of eight transmitters. The transmit systems can be deployed in the laboratory or in vehicles. When operating coherently as a multiantenna transmit system, the individual transmitters can send independent sequences by using a common local oscillator. Synchronization between transmitters and receiver and transmitter geolocation is provided by GPS receivers in the transmitters and receivers.

The Lincoln Laboratory array receiver system is a high-performance sixteen-channel receiver system that can operate over a range of 20 MHz to 2 GHz, supporting a bandwidth up to 8 MHz. The receiver can be deployed in the laboratory or in a stationary "bread truck."

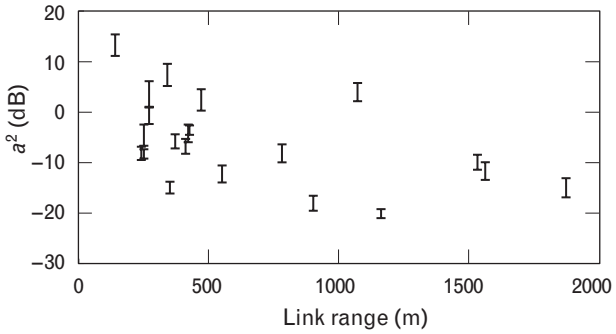


FIGURE 4. Scatter plot of the peak-normalized mean-squared single-input single-output (SISO) link attenuation a^2 versus link range for the outdoor environment near the PCS frequency allocation. The error bars indicate a range of plus or minus one standard deviation of the estimates at a given site.

MIT Campus Experiment

The experiments were performed during July and August 2002 on and near the MIT campus in Cambridge, Massachusetts. These outdoor experiments were performed in a frequency allocation near the PCS band (1.79 GHz). The transmitters periodically emitted 1.7-sec bursts containing a combination of channel-probing and space-time-coding waveforms. A variety of coding and interference regimes were explored for both moving and stationary transmitters. The space-time-coding results are discussed later in the article [39, 40]. Channel-probing sequences using both four and eight transmitters were employed.

The receive antenna array was placed on top of a tall one-story building (at Brookline Street and Henry Street), surrounded by two- and three-story buildings. The transmit array was located on the top of a vehicle within two kilometers of the receive array. Different four- or eight-antenna subsets of the sixteen-channel receiver were used to improve statistical significance. The receive array had a total aperture of less than 8 m, arranged as three subapertures of less than 1.5 m each.

The channel-probing sequence supported a bandwidth of 1.3 MHz with a length of 1.7 msec repeated

ten times. All four or eight transmitters emitted nearly orthogonal signals simultaneously.

Attenuation

Figure 4 displays the peak-normalized mean-squared SISO attenuation averaged over transmit and receive antenna pairs for a given transmit site for the outdoor environment. The uncertainty in the estimate is evaluated by using a bootstrap technique.

Channel Complexity

We present channel complexity by using three different approaches: variation in a^2 estimates, eigenvalue cumulative distribution functions (CDF), and α estimate CDFs. Table 1 is a list of transmit sites used for these results. The table includes the distance (range) between transmitter and receiver, the velocity of the transmitter, the number of transmit antennas, and the estimated α for the transmit site. Uncertainty in α is determined by using the bootstrap technique [42]. The CDF values reported here are evaluated over appropriate entries from Table 1. The systematic uncertainty in the estimation of α caused by estimation bias, given the model, is less than 0.02.

Figure 5 displays CDFs of $a^2 = \mathrm{tr}\{\mathbf{H}\mathbf{H}^\dagger\}/(n_T n_R)$ estimates normalized by mean a^2 for each transmit site. CDFs are displayed for narrowband SISO, 4×4, and 8×8 MIMO systems. Because of the spatial diversity, the variation in mean antenna-pair received power decreases dramatically as the number of antenna pairs increases, as we would expect. This reduction in variation demonstrates one of the most important statistical effects that MIMO links exploit to improve communication link robustness. For example, if we wanted to

Table 1. List of Transmit Sites

Site	*Location*	*Range (m)*	*Velocity (m/sec)*	*Number of antennas*	α
1	Henry and Hasting	150	0.0	8	0.79 ± 0.01
2	Brookline and Erie	520	0.0	8	0.80 ± 0.01
3	Boston University (BU)	430	0.0	8	0.78 ± 0.01
4	BU at Storrow Drive	420	0.0	4	0.72 ± 0.01
5	Glenwood and Pearl	250	10.0	4	0.85 ± 0.01
6	Parking lot	20	0.1	4	0.78 ± 0.02
7	Waverly and Chestnut	270	0.2	4	0.67 ± 0.02
8	Vassar and Amherst	470	0.7	4	0.68 ± 0.02
9	Chestnut and Brookline	140	0.1	4	0.70 ± 0.02
10	Harvard Bridge	1560	11.6	4	0.69 ± 0.02
11	BU Bridge	270	2.7	4	0.83 ± 0.04
12	Vassar and Mass Ave	1070	7.6	4	0.59 ± 0.01
13	Peters and Putnam	240	9.1	4	0.87 ± 0.05
14	Glenwood and Pearl	250	5.2	4	0.76 ± 0.02
15	Brookline and Pacific	780	7.2	4	0.86 ± 0.03
16	Pearl and Erie	550	0.1	4	0.71 ± 0.04
17	Storrow Drive and BU Bridge	410	9.2	4	0.85 ± 0.03
18	Glenwood and Magazine	370	0.0	4	0.78 ± 0.02

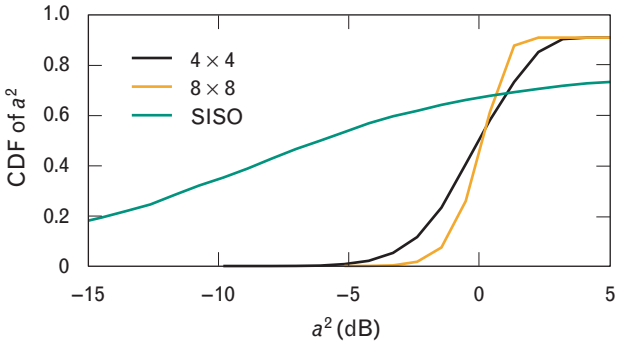


FIGURE 5. Cumulative distribution function (CDF) of channel a^2 estimates, normalized by the mean a^2 for each site, for SISO, 4 × 4, and 8 × 8 MIMO systems.

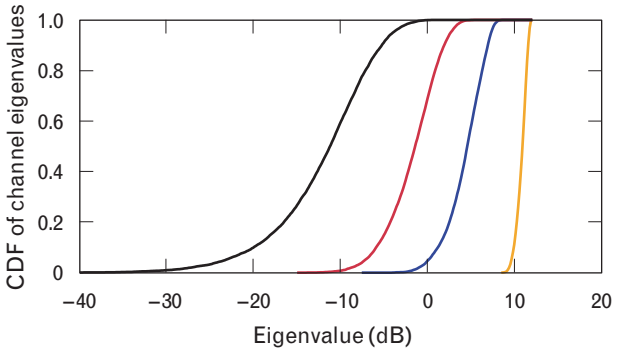


FIGURE 6. CDF of narrowband channel eigenvalue distributions for 4 × 4 MIMO systems.

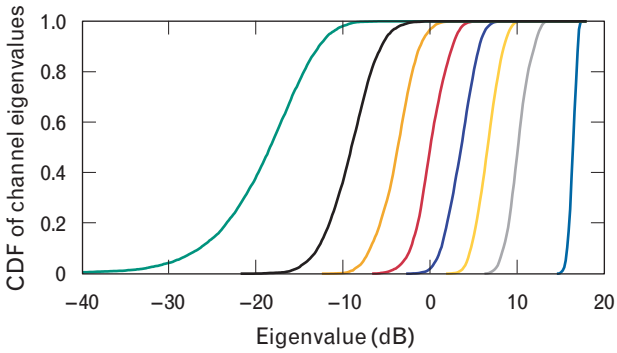


FIGURE 7. CDF of narrowband channel eigenvalue distributions for 8 × 8 MIMO systems.

operate with a probability of 0.9 to close the link, we would have to operate the SISO link with an excess SISO SNR (a^2P_o) margin of over 15 dB. The MIMO systems received the added benefit of array gain, which is not accounted for in the figure.

Figures 6 and 7 present CDFs of eigenvalues for 4 × 4 and 8 × 8 mean-squared-channel-matrix-element-normalized narrowband channel matrices, eig{$\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger$}. The CDFs are evaluated over all site lists. Some care must be taken in interpreting these figures because eigenvalues are not independent. Nonetheless, the steepness of the CDFs is remarkable. We might interpret this to indicate that optimized space-time codes should operate with a relatively high probability of success.

Figure 8 shows the CDFs for α estimates. The mean values of α for each environment are 0.76 for 4 × 4 sites and 0.79 for 8 × 8 sites, where the form $x \pm y$ indicates the estimated value x with statistical uncertainty y estimated by using a bootstrap uncertainty estimation technique. While we might expect smaller variation in the 8 × 8 systems because of the much larger number of paths, this effect may have been exaggerated in Figure 8 because of the limited number of 8 × 8 sites available in the experiment.

Channel Stationarity

Figure 9 displays the temporal variation of eigenvalues of $\bar{\mathbf{H}}\bar{\mathbf{H}}^\dagger$ for stationary and moving transmitters. In this figure the normalization is fixed, allowing for overall shifts in attenuation. As we would expect, the eigenvalues of the moving transmitter vary significantly more than those of the stationary environment. However, the eigenvalues of the stationary transmitter do vary somewhat. While the transmitters and receivers are physically stationary, the environment does move. This effect is particularly noticeable near busy roads. Furthermore, while the multiple antennas are driven with the same local oscillator, given the commercial grade transmitters, there are always some small relative-frequency offsets. The example variation is given for transmit sites 7 and 14 from Table 1.

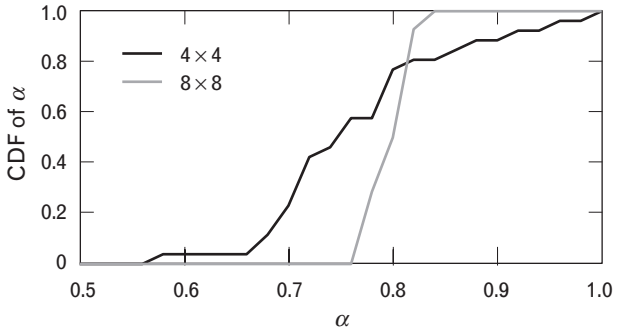


FIGURE 8. CDF of α estimates for 4 × 4 and 8 × 8 MIMO systems.

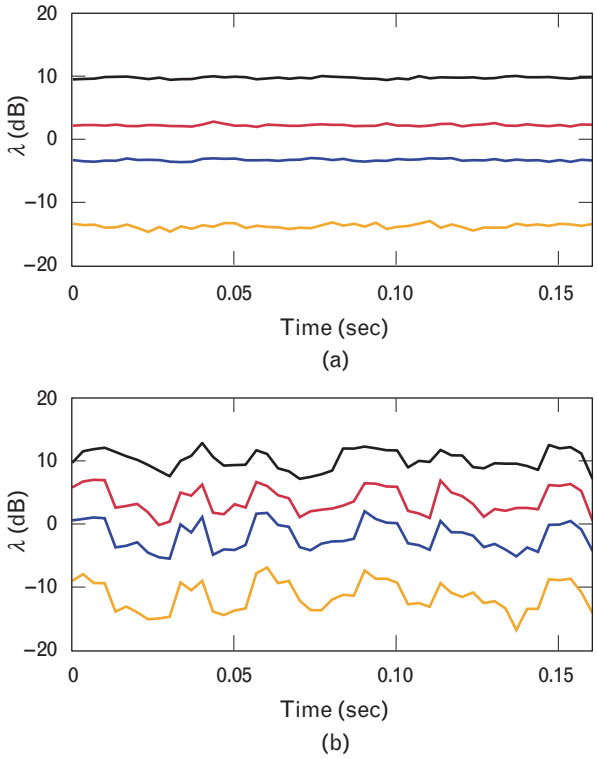


FIGURE 9. Eigenvalues (λ) of $\bar{\mathbf{H}}\bar{\mathbf{H}}^{\dagger}$ as a function of time for (a) stationary and (b) moving transmitters. The same overall attenuation, estimated at $t = 0$, is used for all time samples.

While the moving-transmitter eigenvalues fluctuate more than those of the stationary transmitter, the values are remarkably stable in time. Conversely, an example of the time variation of the power-weighted mean $\cos^2\theta$ metric (from Equation 36), displayed in Figure 10, varies significantly for the moving transmitter within 10 msec. This variation indicates that the eigenvector structure varies significantly, while the distribution of eigenvalues tends to be more stable. In the example, the stationary transmitter is located at site 7, and the moving transmitter is located at site 14. Over the same period the stationary transmitter is relatively stable. Figures 11 and 12 display CDFs for stationary and moving transmitters. The significant variation of the moving transmitter is an indication that implementing an informed transmitter MIMO system would be very challenging for the moving transmitter, but might be viable for some stationary MIMO systems.

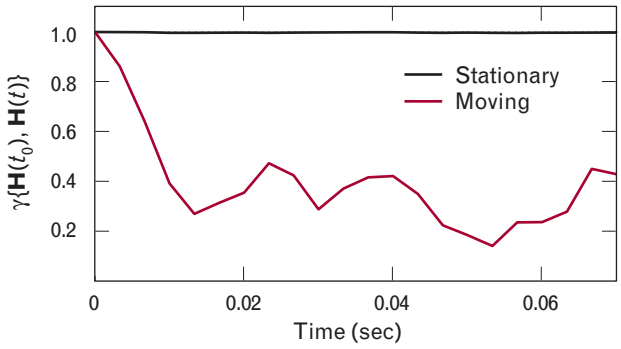


FIGURE 10. Example time variation of power-weighted mean $\cos^2\theta$, $\gamma\{\bar{\mathbf{H}}(t_0),\bar{\mathbf{H}}(t)\}$, for stationary and moving 4×4 MIMO systems.

Frequency-Selective Fading

Figure 13 gives an example of the frequency variation of the power-weighted mean $\cos^2\theta$. The variation is indicated by using the metric presented in Equation 36. In the example, the stationary transmitter is located at site

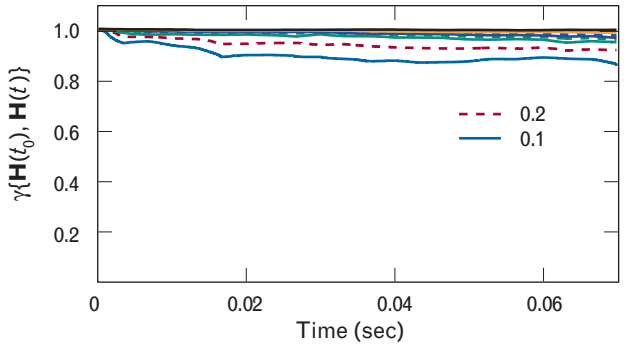


FIGURE 11. CDF of time variation of power-weighted mean $\cos^2\theta$, $\gamma\{\bar{\mathbf{H}}(t_0),\bar{\mathbf{H}}(t)\}$, for a stationary 4×4 MIMO system. The graph displays contours of CDF probabilities of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Because there is little variation, all curves are compressed near a γ value of 1.0

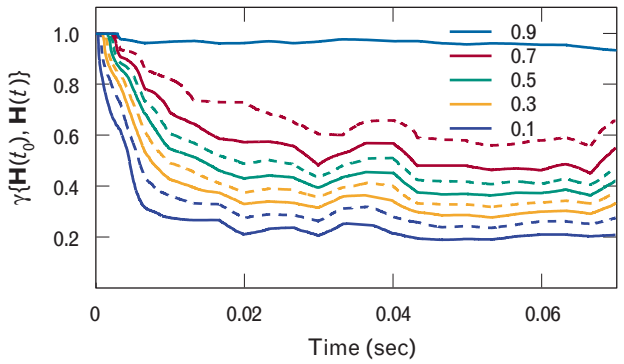


FIGURE 12. CDF of time variation of power-weighted mean $\cos^2\theta$, $\gamma\{\bar{\mathbf{H}}(t_0),\bar{\mathbf{H}}(t)\}$, for a moving 4×4 MIMO system. Contours of CDF probabilities of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 are displayed.

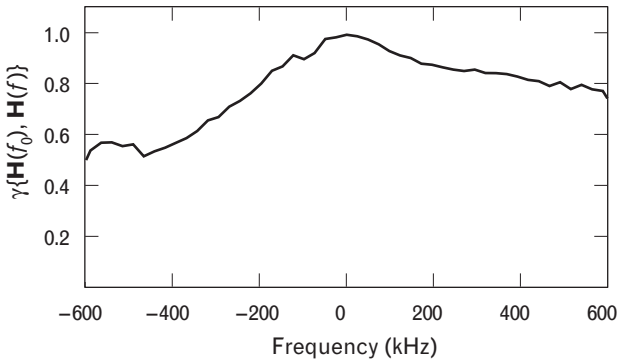


FIGURE 13. Example of frequency-selective variation of the power-weighted mean $\cos^2\theta$, $\gamma\{\bar{\mathbf{H}}(f_0),\bar{\mathbf{H}}(f)\}$.

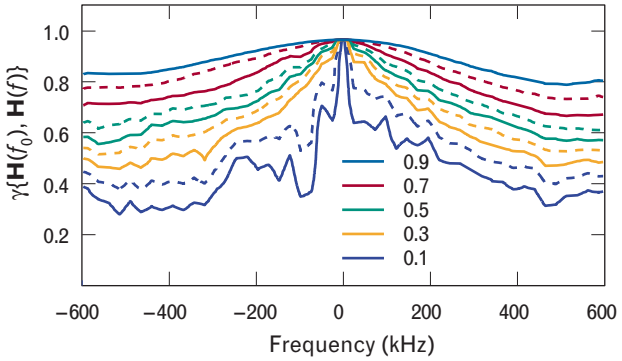


FIGURE 14. CDF of frequency-selective variation of the power-weighted mean $\cos^2\theta$, $\gamma\{\bar{\mathbf{H}}(f_0),\bar{\mathbf{H}}(f)\}$. The graph displays contours of CDF probabilities of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9.

7. Relatively small frequency offsets induce significant changes in $\gamma\{\bar{\mathrm{H}}(f_0),\bar{\mathrm{H}}(f)\}$. Figure 14 shows the CDF of the frequency-selective channel variation. This sensitivity indicates that there is significant resolved delay spread and that, to safely operate with the narrowband assumption, bandwidths less than 100 kHz should be employed. We note that delay spread, and the resulting frequency-selective fading, are both a function of environment and link length. Consequently, some care must be taken in interpreting this result.

Space-Time Low-Density Parity-Check-Code Experiments

A low-density parity-check code over GF(16) provides the basis of the example of experimental and simulated results shown in Figure 15. The code used is half rate with length 1024. The MIMO wireless link is realized with four cohered transmitters located on a stationary van several hundred meters away from an array of receivers situated on a one-story building. The environment consists predominantly of two- and three-story residential buildings and some commercial buildings of similar heights in an urban setting. Propagation delay spreads are typically several microseconds and Doppler spreads are at most a few hundred hertz. There is typically no identifiable line-of-sight component in the propagation. The signal has a pulse-shaped QPSK modulation and bandwidth of about 100 kHz. Coding provides a spectral efficiency of 2 bits/sec/Hz.

The receiver consists of sixteen channels fed by low-gain elements with wide azimuth beamshapes. The elements are oriented in various directions, not necessarily pointing at the sources. For the example below, four element subarrays are chosen at random to provide multichannel receivers. In other words, C(16, 4) 4×4

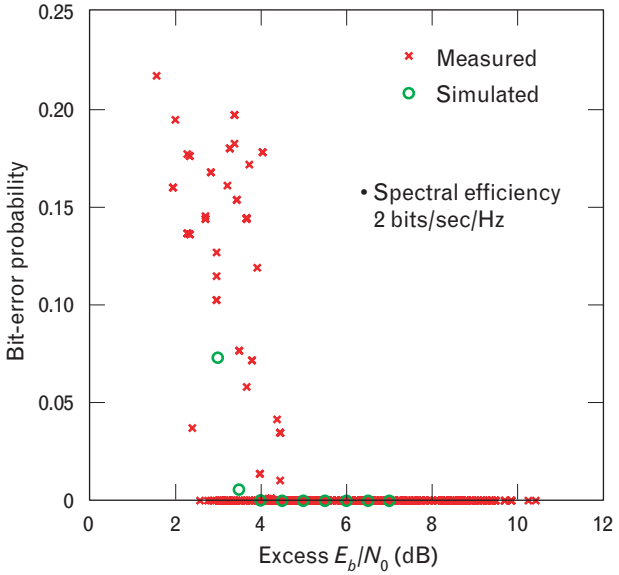


FIGURE 15. Measured and simulated results in bit error rate probability for a space-time low-density parity-check (LDPC) code at a spectral efficiency of 2 bits/sec/Hz. Bit error rates are evaluated for an ensemble of 4×4 MIMO systems. The estimated channel matrices are used in the simulation to model propagation. Each estimated channel matrix suppports a theoretical capacity that can be in excess of 2 bits/sec/Hz. The matrix is scaled until it supports a capacity of exactly 2 bits/sec/Hz. The resulting scale factor is used to evaluate the excess (beyond Shannon) E_b/N_0 associated with the (unscaled) channel matrix. Agreement between measured and simulated results are good to about 1 dB. About 4 to 5 dB excess E_b/N_0 is required to reliably complete the link.

MIMO links can be evaluated. The channel transfer matrices have a random structure that varies from subarray to subarray.

Figure 15 shows symbol-error probability as a function of excess E_b/N_0, which is related to the excess spectral efficiency (beyond 2 bits/sec/Hz) predicted by a capacity bound, given the measured channel transfer matrix. For this example, in a comparatively stationary environment, the channel transfer matrices are used both for the simulated results and for the computation of excess E_b/N_0. As the figure shows, about 4.5 dB excess E_b/N_0 is required to complete the link at 2 bits/sec/Hz. Simulations agree to within about 1 dB.

Space-Time Turbo-Code Experiments

In this section we present the experimental performance of a space-time turbo code. We begin by discussing the experimental parameters, and then we summarize the performance of the MIMO system with stationary transmitter and receiver in a dynamic environment. Additionally, for an even more complicated environment, we describe performance results with a mobile transmitter and multiple strong interferers.

Experimental Parameters

Outdoor experiments were performed in a frequency allocation near the PCS band, using a sixteen-channel receiver. A variety of coding and interference regimes were explored for both moving and stationary transmitters. Channel-probing sequences and four- and eight-transmitter space-time codes were transmitted. This section reports on the outdoor performance results of space-time turbo codes for 4×4 MIMO systems. The outdoor experiments were performed during July and August 2002 on and near the MIT campus. The receive antenna array was placed on top of a one-story building (at Brookline Street and Henry Street) surrounded by two- and three-story buildings.

For the examples discussed in this article, quadrature-phase-shift-key (QPSK) signals were transmitted on four antennas at 123×10^3 chips per second for a total data rate of 246 kb/sec, using the space-time code discussed earlier. A 160-kHz spectral limit was enforced by using a root-raised-cosine pulse shaping. Total transmit power was approximately 100 mW, radiated from 0-dBi antennas. We discuss two examples with different transmit locations. In both examples the link does not have line of sight. Different four-antenna subsets of the sixteen-channel receiver were used to improve statistical significance.

Example 1. In this example, the transmitter was located in the parking lot at Boston University (University Road and Storrow Drive) with about a half-kilometer separation between the transmitter and receiver. Figure 16 shows the geometry of the experiment. Traffic on Storrow Drive is typically heavy and the posted speed limit of 45 mph is generally misinterpreted as the minimum allowed speed. While the transmitter is stationary, the environment is nonstationary because of the traffic.

Example 2. The transmitter was moving at 10 m/sec at a range of 500 m. Figure 17 shows the geometry of the experiment. To simulate the effects of local oscillator errors, we introduced artificial frequency offsets at the transmitters. These errors were within ±80 Hz.

Two wideband jammers were transmitting at a range of 100 m. Each jammer was received at a jammer-to-noise ratio (JNR) of approximately 25 dB. Figure 18 shows the eigenvalues of the noise-normalized interference-plus-noise spatial covariance matrix. The "noise" eigenvalues of the jammer spatial eigenvalue distribution

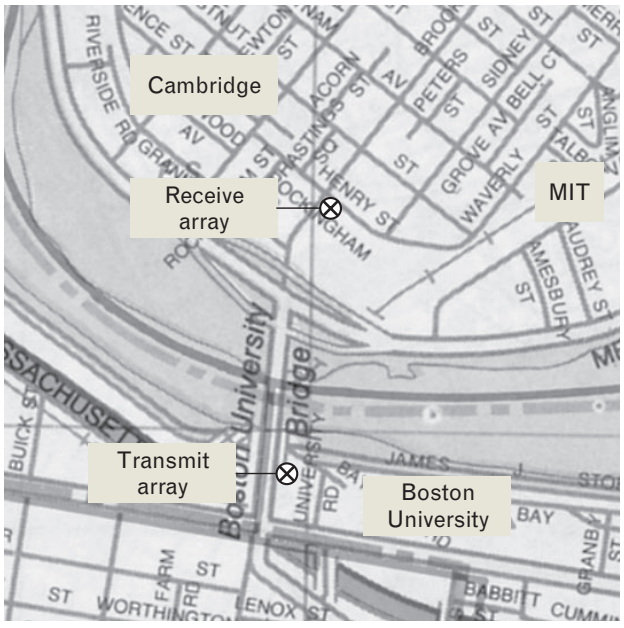


FIGURE 16. Example 1: map of MIMO communication experiment near the MIT campus, including the locations of the transmitter and receiver.

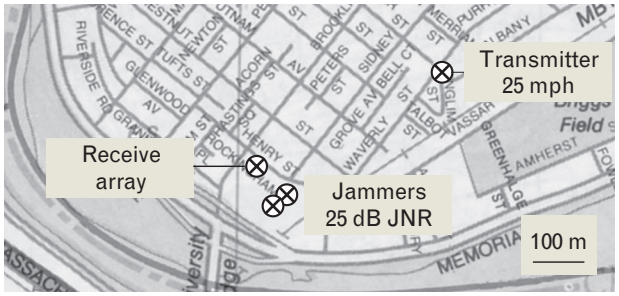


FIGURE 17. Example 2: map of MIMO communication experiment near the MIT campus, including the locations of the transmitter, receiver, and jammers at a fixed jammer-to-noise ratio (JNR).

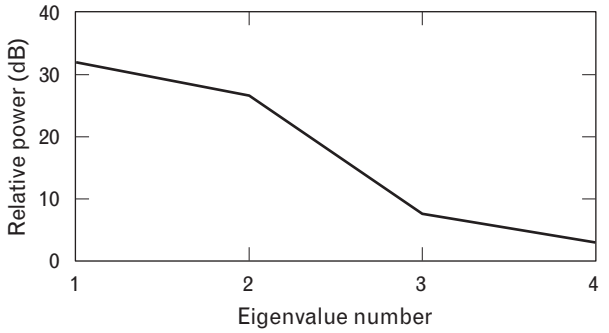


FIGURE 18. Eigenvalue distribution of the noise-normalized interference-plus-noise spatial covariance matrix.

are slightly higher than we would naively expect, given the 0-dB noise normalization. This behavior is probably an indication of either delay spread or nonstationarity in the received jammer signal. Either of these explanations presents additional challenges to the receiver.

Both the delay and the Doppler spread affect the design and performance of the receiver. Here a space-time-frequency adaptive processor is employed. The number of delay and frequency taps in the adaptive processor depends upon the phenomenology. Delay spread was found to be less than ±4 μsec. For the stationary environment, in quiet regions (no nearby traffic), no Doppler spread was detected. For the stationary transmitter near heavy traffic in experimental example 1, the Doppler spread was found to be within ±150 Hz. For the moving transmitter in experimental example 2, the Doppler spread was found to be within ±180 Hz.

Experimental Example 1

Bit error rates for various detector alternatives are reported as a function of mean single-input single-output (SISO) SNR, a^2P_o. Here, a^2 is the mean-squared link attenuation. Figure 19 displays the results for four detection variations:

1. Training-data-based adaptive spatial beamforming (three turbo iterations).
2. Coarse training-data-based space-frequency beamforming (one turbo iteration; Doppler taps: {–1, 0, 1}).
3. Space-time-frequency beamforming employing decision-directed channel estimation without multiuser detection (three turbo iterations; Doppler taps: {–4/3, –2/3, 0, 2/3, 4/3}; temporal taps: {–1/2, 0, 1/2}).
4. MCMUD with space-time-frequency beamforming (three turbo iterations; Doppler taps: {–4/3, –2/3, 0, 2/3, 4/3}; temporal taps: {–1/2, 0, 1/2}), and where Doppler taps are represented in resolution cells (60 Hz).

Performance improves with receiver complexity; the algorithm, however, must bootstrap up in complexity iteratively. Starting with the highest complexity on the first iteration increases the probability of converging to the wrong solution. Because the channel contains significant Doppler spread, the spatial beamformer performs poorly. With the relatively long block lengths of the turbo code, Doppler beamforming is required in this environment. We note that experimental performance is essentially the same as was found in simulations. Furthermore, the experimental performance is

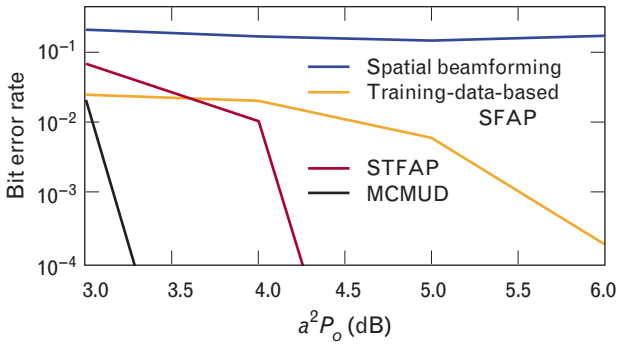


FIGURE 19. Bit error rate of 4 × 4, 2-bit/sec/Hz space-time turbo code as a function of mean SISO signal-to-noise ratio (SNR) (a^2P_o) for a Boston University transmit location, using adaptive spatial beamforming, coarse training-data-based space-frequency adaptive beamforming (SFAP), space-time-frequency adaptive beamforming (STFAP) employing decision-directed channel estimation, and MCMUD with space-time-frequency adaptive beamforming.

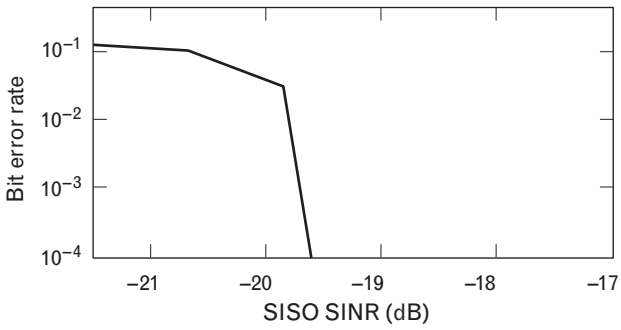


FIGURE 20. Bit error rate of 4 × 4, 2-bit/sec/Hz space-time turbo code using the MCMUD receivver with space-time-frequency beamforming as a function of mean SISO signal-to-interference-plus-noise ratio (SINR).

similar to the simulated performance of the best space-time codes.

Experimental Example 2

The experimental data includes the effects of two high-power wideband jammers, a moving transmitter, and local oscillator errors. Experimental performance of this space-time turbo code for a stationary transmitter in the absence of interference is discussed elsewhere [39].

Figure 20 shows the bit error rate of the space-time turbo code using the MCMUD receiver. The bit error rate is displayed in terms of the mean SISO signal-to-interference-plus-noise ratio (SINR). This is the average SINR at a given receive antenna, assuming that all power of the transmit array is transmitted from a single transmit antenna. We note that this experimental system in this difficult environment operates at an SINR that is 25 dB better than the information-theoretic SISO bound, and operates probably at least 35 dB better than a practical SISO system. Furthermore, there is only approximately a 3-dB loss in a^2P_o performance compared to the performance in an environment without jammers. The effectiveness of the receiver is due in part to its ability to compensate for delay and frequency spread. The MCMUD employs a space-time-frequency adaptive processor that uses a four-antenna receiver with temporal and frequency taps that cover a range of ±4 microseconds and ±200 Hz.

Summary

In this article we addressed information-theoretic, phenomenological, coding, and receiver issues for MIMO communication. Performance bounds assuming either an informed transmitter or an uninformed transmitter were presented for flat-fading, frequency-selective, and jammed environments. A channel phenomenology parameterization was introduced. Experimental phenomenological results were reported, the results indicating that the observed channels can be typically characterized by high degrees of complexity. Furthermore, for environments with transmitters on moving vehicles, the channel varies significantly on a time scale less than 10 msec. Two space-time coding techniques were introduced, one based on LDPC and the other on turbo codes. Experimental demodulation performance results were presented for a variety of environments, including those with wideband jammers. In the presence of the jammer, the MIMO system (using the MCMUD receiver) operated dramatically better than SISO systems.

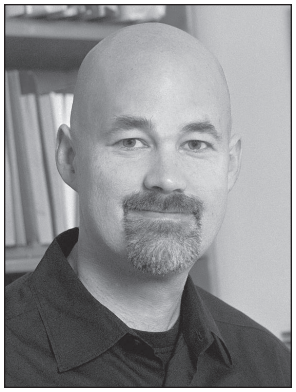
Acknowledgments

The authors would like to thank Peter Wu of Lincoln Laboratory for his help developing the space-time turbo code, and Naveen Sunkavally of MIT and Nick Chang of the University of Michigan for their help with the experiment. The authors would also like to thank the excellent Lincoln Laboratory staff involved in the MIMO experiment, in particular Sean Tobin, Jeff Nowak, Lee Duter, John Mann, Bob Downing, Peter Priestner, Bob Devine, Tony Tavilla, and Andy McKellips. We also thank Ali Yegulalp of Lincoln Laboratory and Vahid Tarokh of Harvard University for their thoughtful comments, and Dorothy Ryan of Lincoln Laboratory for her helpful comments. Finally, the authors would like to thank the MIT New Technology Initiative Committee for their support.

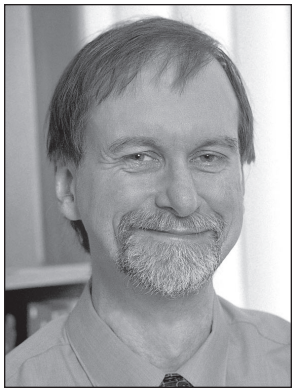
REFERENCES

1. W.C. Jakes, *Microwave Mobile Communications* (Wiley, New York, 1974).
2. R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays* (Wiley, New York, 1980).
3. K.W. Forsythe, D.W. Bliss, and C.M. Keller, "Multichannel Adaptive Beamforming and Interference Mitigation in Multiuser CDMA Systems," *Thirty-Third Asilomar Conf. on Signals, Systems & Computers* **1**, *Pacific Grove, Calif., 24–27 Oct. 1999,* pp. 506–510.
4. A. Wittneben, "Basestation Modulation Diversity for Digital SIMULCAST," *Proc. IEEE Vehicular Technology Conf., St. Louis, Mo., 19–22 May 1991,* pp. 848–853.
5. V. Weerackody, "Diversity for Direct-Sequence Spread Spectrum Using Multiple Transmit Antennas," *Proc. IEEE Int. Communications Conf.* **3**, *Geneva, 23–26 May, 1993,* pp. 1775–1779.
6. G.J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multi-Element Antennas," *Bell Labs Tech. J.* **1** (2), 1996, pp. 41–59.
7. I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecommun.* **10** (6), 1999, pp. 585–595.
8. D.W. Bliss, K.W. Forsythe, A.O. Hero, and A.L. Swindlehurst, "MIMO Environmental Capacity Sensitivity," *Thirty-Fourth Asilomar Conf. on Signals, Systems & Computers* **1**, *Pacific Grove, Calif., 29 Oct.–1 Nov. 2000,* pp. 764–768.
9. D.W. Bliss, K.W. Forsythe, and A.F. Yegulalp, "MIMO Communication Capacity Using Infinite Dimension Random Matrix Eigenvalue Distributions," *Thirty-Fifth Asilomar Conf. on Signals, Systems & Computers* **2**, *Pacific Grove, Calif., 4–7 Nov. 2001,* pp. 969–974.
10. T.L. Marzetta and B.M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Trans. Inf. Theory* **45** (1),1999, pp. 139–157.
11. L. Zheng and D.N.C. Tse, "Diversity and Freedom: A Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. Inf. Theory* **49** (9), 2003, pp. 1076–1093.
12. S.M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE J. Sel. Areas Commun.* **16** (8), 1998, pp. 1451–1458.
13. V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space–Time Block Codes from Orthogonal Designs," *IEEE Trans. Inf. Theory* **45** (5), 1999, pp. 1456–1467.
14. G. Ganesan and P. Stoica, "Space–Time Block Codes: A Maximum SNR Approach," *IEEE Trans. Inf. Theory* **47** (4), 2001, pp. 1650–1656.
15. B. Hassibi and B. Hochwald, "High-Rate Linear Space-Time Codes," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* **4**, *Salt Lake City, Utah, 7–11 May 2001,* pp. 2461–2464.
16. V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space–Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. Inf. Theory* **44** (2), 1998, pp. 744–765.
17. B.M. Hochwald and T.L. Marzetta, "Unitary Space–Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading," *IEEE Trans. Inf. Theory* **46** (2), 2000, pp. 543–564.
18. B.M. Hochwald and W. Sweldens, "Differential Unitary Space-Time Modulation," *IEEE Trans. Com.* **48** (12) 2000, pp. 2041–2052.
19. K.W. Forsythe, "Capacity of Flat-Fading Channels Associated with a Subspace-Invariant Detector," *34th Asilomar Conf. on Signals, Systems and Computers* **1**, *Pacific Grove, Calif., 29 Oct.–1 Nov. 2000,* pp. 411–416.
20. K.W. Forsythe, "Performance of Space-Time Codes over a Flat-Fading Channel Using a Subspace-Invariant Detector," *36th Asilomar Conf. on Signals, Systems and Computers* **1**, *Pacific Grove, Calif., 3–6 Nov. 2002,* pp. 750–755.
21. A. Stefanov and T.M. Duman, "Turbo Coded Modulation for Wireless Communications with Antenna Diversity," *Proc. IEEE Vehicular Technology Conf.* **3**, *Amsterdam, 19–22 Sept. 1999,* pp. 1565–1569.
22. Y. Liu, M.P. Fitz, and O.Y. Takeshita, "Full Rate Space–Time Turbo Codes," *IEEE J. Sel. Areas Commun.* **19** (5), 2001, pp. 969–980.
23. H. Sampath and A.J. Paulraj, "Joint Transmit and Receive Optimization for High Data Rate Wireless Communication Using Multiple Antennas," *Conf. Record Thirty-Third Asilomar Conf. on Signals, Systems & Computers* **1**, *Pacific Grove, Calif., 24–27 Oct. 1999,* pp. 215–219.
24. N. Sharma and E. Geraniotis, "Analyzing the Performance of the Space-Time Block Codes with Partial Channel State Feedback," *Proc. Wireless Communications and Networking Conf.* **3**, *Chicago, 23–28 Sept., 2000*, pp. 1362–1366.
25. D.W. Bliss, K.W. Forsythe, A.O. Hero, and A.F. Yegulalp, "Environmental Issues for MIMO Capacity," *IEEE Trans. Signal Process.* **50** (9), 2002, pp. 2128–2142.
26. T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
27. F. R. Farrokhi, G. J. Foschini, A. Lozano, and R.A. Valenzuela, "Link-Optimal Space–Time Processing with Multiple Transmit and Receive Antennas," *IEEE Comm. Lett.* **5** (3), 2001, pp. 85–87.
28. D.W. Bliss, A.M. Chan, and N.B. Chang, "MIMO Wireless Communication Channel Phenomenology," *IEEE Trans. Antennas Propag.,* **52** (8), 2004, pp. 2073–2082.
29. D. Gesbert, H. Bölcskei, D.A. Gore, and A.J. Paulraj, "Performance Evaluation for Scattering MIMO Channel Models," *Thirty-Fourth Asilomar Conf. on Signals, Systems & Computers* **1**, *Pacific Grove, Calif., 29 Oct.–1 Nov. 2001,* pp. 748–752.
30. H. Bölcskei and A.J. Paulraj, "Performance of Space-Time Codes in the Presence of Spatial Fading Correlation," *Thirty-Fourth Asilomar Conference on Signals, Systems & Computers* **1**, *Pacific Grove, Calif., 29 Oct.–1 Nov. 2000,* pp. 687–693.
31. R.G. Gallager, *Low-Density Parity-Check Codes* (MIT Press, Cambridge, Mass., 1963).
32. S.-Y. Chung, G.D. Forney, Jr., T.J. Richardson, and R. Urbanke, "On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit," *IEEE Commun. Lett.* **5** (2), 2001, pp. 58–60.
33. T.J. Richardson, M.A. Shokrollahi, and R.L. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes," *IEEE Trans. Inf. Theory* **47** (2), 2001, pp. 619–637.
34. R.J. McEliece, D.J.C. MacKay, and J.-F. Cheng, "Turbo Decoding as an Instance of Pearl's 'Belief Propagation' Algorithm," *IEEE J. Sel. Areas Commun.* **16** (2), 1998, pp. 140–152.
35. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes," *Proc. IEEE Int. Communications Conf.* **2**, *Geneva, 23–26 May 1993,* pp. 1064–1070.
36. L.R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal De-

coding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inf. Theory* **20** (2), 1974, pp. 284–287.
37. P. Robertson, E. Villebrun, and P. Hoeher, "Comparison of Optimal and Sub-Optimal MAP Decoding Algorithms Operating in the Log Domain," *Proc. IEEE Int. Communications Conf.* **2,** *Seattle, 18–22 June 1995,* pp. 1009–1013.
38. P. H.–Y. Wu and S.M. Pisuk, "Implementation of a Low Complexity, Low Power, Integer-Based Turbo Decoder," *Proc. IEEE Global Telecommunications Conf., San Antonio, Tex., 25–29 Nov. 2001*, pp. 946–951.
39. D.W. Bliss, P.H. Wu, and A.M. Chan, "Multichannel Multiuser Detection of Space-Time Turbo Codes: Experimental Performance Results," *Thirty-Sixth Asilomar Conference on Signals, Systems & Computers* **2,** *Pacific Grove, Calif., 3–6 Nov. 2002,* pp. 1343–1348.
40. D.W. Bliss, "Robust MIMO Wireless Communication in the Presence of Interference Using Ad Hoc Arrays," *MILCOM 2003*, **2**, Oct. 2003, pp. 1382–1385.
41. C.M. Keller and D.W. Bliss, "Cellular and PCS Propagation Measurements and Statistical Models for Urban Multipath on an Antenna Array," *Proc. 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop, Cambridge, Mass., 16–17 Mar. 2000,* pp. 32–36.
42. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, 1982).
43. K.W. Forsythe, "Utilizing Waveform Features for Adaptive Beamforming and Direction Finding with Narrowband Signals," *Linc. Lab. J.* **10** (2), 1997, pp. 99–126.
44. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, Calif., 1988).



DANIEL W. BLISS
is a a staff member in the Advanced Sensor Techniques group. He received M.S. and Ph.D. degrees in physics from the University of California at San Diego, and a B.S.E.E. in electrical engineering from Arizona State University. Previously employed by General Dynamics, he designed avionics for the Atlas-Centaur launch vehicle, and performed research and development of fault-tolerant avionics. As a member of the superconducting magnet group, he performed magnetic field calculations and optimization for high-energy particle-accelerator superconducting magnets. His doctoral work was in high-energy particle physics, searching for bound states of gluons, studying the two-photon production of hadronic final states, and investigating innovative techniques for lattice-gauge-theory calculations. At Lincoln Laboratory he focuses on multiantenna adaptive signal processing, primarily for communication systems, and on parameter estimation techniques and bounds, primarily for geolocation. His current research includes ultrawide bandwidth communication, geolocation techniques using vector sensor arrays, multiple-input multiple-output (MIMO) radar concepts, algorithm development for multichannel multiuser detectors, MIMO communication channel phenomenology, and information theoretic bounds on MIMO communication systems.



KEITH W. FORSYTHE
is a senior staff member in the Advanced Sensor Techniques group. He received S.B. and S.M. degrees, both in mathematics, from MIT. In 1978 he joined Lincoln Laboratory, where he has worked in the areas of spread-spectrum communication, adaptive sensor-array processing, and synthetic aperture radar (SAR) imaging. His work on spread-spectrum systems includes electromagnetic modeling (geometric theory of diffraction) of antennas mounted on an airframe, error-correction coding, jam-resistant synchronization techniques, and digital matched-filter design and performance. In the area of adaptive sensor-array processing he helped develop a number of signal-processing algorithms that exploit waveform features to achieve levels of performance (beamforming, direction finding, geolocation, and other forms of parameter estimation) beyond those attainable by nonexploitive techniques. His work on SAR imaging involves techniques for resolution enhancement and interference rejection for foliage-penetration systems.

AMANDA M. CHAN
is an associate staff member in the Advanced Sensor Techniques group. She received MSEE and BSEE degrees in electrical engineering from the University of Michigan. Her interests are in channel phenomenology. She has previously worked with implementation of synthetic aperture geolocation of cellular phones. Most recently, she has worked on the implementation of MIMO channel parameterization.