MAP Decoding for Multi-Antenna Systems with Non-Uniform Sources: Exact Pairwise Error Probability and Applications

Firouz Behnamfar, Fady Alajaji, *Senior Member, IEEE,* and Tamás Linder, *Senior Member, IEEE*

***Abstract*—We study the maximum *a posteriori* (MAP) decoding of memoryless non-uniform sources over multiple-antenna channels. Our model is general enough to include space-time coding, BLAST architectures, and single-transmit multi-receive antenna systems which employ any type of channel coding. We derive a closed-form expression for the codeword pairwise error probability (PEP) of general multi-antenna codes using moment generating function and Laplace transform arguments. We then consider space-time orthogonal block (STOB) coding and prove that, similar to the maximum likelihood (ML) decoding case, detection of symbols is decoupled in MAP decoding. We also derive the symbol PEP in closed-form for STOB codes. We apply these results in several scenarios. First, we design a binary antipodal signaling scheme which minimizes the system bit error rate (BER) under STOB coding. At a BER of 10^{-6}, this constellation has a channel signal-to-noise ratio (CSNR) gain of 4.7 dB over conventional BPSK signaling for a binary non-uniform source with $p_0 \triangleq P(0) = 0.9$. We next design space-time linear dispersion (LD) codes which are optimized for the source distribution under the criterion of minimizing the union upper bound on the frame error rate (FER). Two codes are given here: one outperforms V-BLAST by 3.5 dB and Alamouti's code by 12.3 dB at an FER of 10^{-2} for a binary source with $p_0 = 0.9$, and the other outperforms V-BLAST by 4.2 dB at an FER of 10^{-3} for a uniform source. These codes also outperform the LD codes of [13] constructed under a different criteria. Finally, the problem of bit-to-signal mapping is studied. It is shown that for a binary source with $p_0 = 0.9$, 64-QAM signaling, and SER = 10^{-3}, a gain of 3.7 dB can be achieved using a better-than-Gray mapping. For a system with one transmit and two receive antennas that uses trellis coding with 16-QAM signaling, a 1.8 dB gain over quasi-Gray mapping and ML decoding is observed when MAP decoding is used for binary sources with $p_0 = 0.9$.**

***Index Terms*—Joint source-channel coding, space-time coding, MAP decoding, pairwise error probability, statistical redundancy, convolutional codes, trellis coding, maximum-ratio combining, multi-antenna fading channels, diversity, wireless communications.**

Paper approved by P. Y. Kam, the Editor for Modulation and Detection of the IEEE Communications Society. Manuscript received May 29, 2006; revised July 23, 2007 and February 7, 2008.

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Premier's Research Excellence Award (PREA) of Ontario. Parts of this paper were presented at The 2003 Information Theory Workshop (ITW'03), March 2003, Paris, France, and The 43rd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2005.

F. Behnamfar is with Nortel Networks, 3500 Carling Avenue, Ottawa, Ontario K2H 8E9, Canada (e-mail: firouz@nortel.com).

F. Alajaji and T. Linder are with the Department of Mathematics and Statistics and the Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada (e-mail: {fady, linder}@mast.queensu.ca). T. Linder is also associated with the Computer and Automation Research Institute of the Hungarian Academy of Sciences, Budapest H-1111, Hungary.

Digital Object Identifier 10.1109/TCOMM.2009.0901.060333

I. Introduction

IDEALLY, a lossless or lossy source coder would compress data into an independent, identically distributed (i.i.d.) nearly uniform bit-stream (for sufficiently long blocklengths). However, most practical source coding methods are not ideal; hence there exists a residual redundancy (in the form of non-uniform distribution and/or memory) at their output which will be present at the input of the channel encoder. For example, the line spectral parameters at the output of codebook-excited linear predictive (CELP) speech vocoders may contain up to 42% of (residual) redundancy due to non-uniformity and memory (see, e.g., [3]). Another example is the bit-stream at the output of vector quantizers with moderate blocklengths. Furthermore, natural data sources, which in certain complexity-constrained applications (e.g., wireless sensor networks) are transmitted uncompressed over the channel, exhibit even higher amounts of redundancy. For example, binary images may contain as much as 80% of redundancy due to non-uniformity; this translates into a probability as high as 97% for having a "0" (as opposed to a "1") in the image bit-stream (see, e.g., [34] and the references therein).

In this paper, we study how exploiting the source non-uniformity at the transmitter and/or the receiver can improve the performance of multi-antenna systems in the presence of quasi-static Rayleigh fading. This scenario allows for the use of channel coding (such as convolutional, Turbo or low-density parity check (LDPC) coding) before the multi-antenna encoding operation, as long as systematic channel codes are used. If such codes are employed, then the resulting bit-stream at the input of the multi-antenna coder will still be non-uniform (albeit to a lesser extent than the original source, depending on the code rate and blocklength). If non-systematic channel codes are used, the resulting bit-stream will be closer to uniform; in this case, a different (and challenging) approach, not considered here, would be to jointly design the channel code and the multi-antenna encoder to exploit the non-uniformity of the original source (refer to [34] for examples of non-systematic Turbo codes that exploit the source non-uniformity in a single-antenna system).

Our contribution is threefold. First, we derive the maximum *a posteriori* (MAP) decoding rule for multi-antenna codewords. We then derive a closed-form expression for the

codeword pairwise error probability (PEP) of general multi-antenna codes (including any space-time and BLAST codes) under MAP decoding. Finally, we explore some applications of the above results and show that there can be a large gain in performing MAP decoding as compared with maximum likelihood (ML) decoding. Knowing the exact PEP in closed form facilitates the derivation of better estimates of the system error rates, since the Chernoff upper bound on the codeword PEP derived in [29] is often too loose to be useful. In fact, numerical results in [7] show that the Chernoff-based union upper bound is significantly ineffective for symbol error rate (SER) and bit error rate (BER) estimation at the error rates of practical interest in wireless communications since the bounds are often larger than 1 at low to medium values of the channel signal-to-noise ratio (CSNR). The exact codeword PEP is hence of vital interest for both analysis and design purposes.

For ML decoding, the main challenge in finding the PEPs of interest under fading is to average $Q(\sqrt{X})$ where $Q(\cdot)$ is the Gaussian error integral and X is a non-negative random variable. A closed-form expression for the codeword PEP of space-time codes of arbitrary structure under slow Rayleigh fading and ML decoding is derived in [19]. The derivation is based on an alternate formula for the $Q(\cdot)$ function [10], which only works for non-negative arguments. As will be seen in the sequel, computing the PEP between a pair of MAP decoded codewords requires finding the expected value of $Q(\sqrt{X}+\lambda/\sqrt{X})$, where λ is a real (positive or negative) number; this is more involved than the ML decoding case. We use singular value decomposition and Laplace transform arguments to derive the above PEP. Other work on the error analysis of space-time coded channels under ML decoding include [31], where an expression for the exact PEP of space-time trellis codes is found and used to derive an upper bound on the BER. Another form of the exact PEP is derived in [27] which is easier to compute in certain cases. The authors have presented simple formulas in closed-form for the exact PEP of space-time codes in [7], where very tight upper and lower bounds on system SER and BER are also derived. To the best of our knowledge, there is no work in the literature on performance analysis or simulation of space-time codes under MAP decoding.

Next, we consider the special case of space-time orthogonal block (STOB) codes and show that for this case, when the symbols input to the space-time encoder are i.i.d. (but not necessarily uniformly distributed), detection of symbols is decoupled (as in the ML decoding case). We then derive the symbol PEP under MAP decoding for STOB codes. The PEP expression is also valid for systems that utilize maximum-ratio combining (MRC).

Finally, we apply the PEP results to three coding scenarios. First, we find the optimal binary antipodal signaling in the sense of minimizing the BER of space-time orthogonal coded systems. We prove that the optimal binary antipodal signaling does not actually depend on the fading distribution and is the same as the one derived for the additive white Gaussian noise (AWGN) channel. Second, we construct space-time linear dispersion (LD) codes for both non-uniform and uniform i.i.d. sources. Unlike [13], where the code design criterion is to maximize the mutual information between the channel input and output, we opt to minimize the union upper bound on the frame error rate (FER) of the code. We note that even for a simple dual-transmit dual-receive system with BPSK modulation, gains up to 4.2 dB can be obtained over V-BLAST for a uniform i.i.d. source at an FER of 10^{-3}. Third, we address the design of bit-to-signal mappings which take the input non-uniformity into account to minimize the BER of two systems: one system uses STOB codes while the other one is a trellis coded system with 16-QAM signaling in a single-transmit multiple-receive antenna setup. We observe that the gains with better-than-Gray mappings can be significant if the source has non-uniform distribution. For example, in a trellis coded system with 2 receive antennas and $p_0 \triangleq P(b=0)=0.9$ (where b is a data bit), at FER = 10^{-3}, a CSNR gain of 0.8 dB can be obtained through MAP decoding (instead of ML decoding) and an additional gain of 1.0 dB can be achieved using a signal mapping which is carefully designed (hence a total gain of 1.8 dB over quasi-Gray mapping and ML decoding is obtained).

MAP decoding for sources with redundancy (due to non-uniform distribution and/or memory) is a form of joint source-channel coding/decoding. It would then be interesting to compare the performance of MAP-decoded schemes with that of tandem coding systems, i.e., systems with separate and independent source compression and channel coding blocks. Most previous coding designs, such as [2], [18], show that independent (tandem) source and channel coding outperforms joint source-channel coding above some threshold CSNR.[1] As can be seen in the simulations of this paper, the CSNR threshold beyond which tandem coding outperforms MAP decoding is quite large. In particular, there are many examples in which joint source-channel coding outperforms tandem coding for the entire CSNR range (or error rates) of interest. Indeed, in a recent information theoretic study [33], it is proved that the error exponent (which is the rate of asymptotic exponential decay of the probability of block error) of optimal joint source-channel coding can be as large as twice the error exponent of optimal tandem systems (which concatenate optimal source coding with optimal channel coding). This implies that for the same probability of error, optimal joint source-channel coding would require half the encoding/decoding delay of the optimal tandem scheme.

The rest of this paper is organized as follows. Section II describes the multi-input multi-output (MIMO) channel model and formulates the MAP decoding rule based on which the exact codeword PEP is derived in Section III. In Section IV, we derive the MAP decoding rule and symbol PEP for the special case of STOB codes. Applications of the PEP formulas in binary signaling, LD code design, and bit-to-signal mapping are presented in Section V. Section VI presents the numerical results and discussions. The paper is concluded in Section VII.

II. System Model and the MAP Decoding Rule

The MIMO communication system considered here employs K transmit and L receive antennas. The input to

[1] An opposite behavior is however observed in [34], where joint source-channel coding based on Turbo coding (with significantly longer block lengths) outperforms tandem coding for high CSNRs.

the system is a stream of i.i.d. symbols which can have non-uniform distribution. The baseband constellation signals are denoted by $\{c_k\}_{k=1}^{2^p}$ where p is a positive integer. We will assume that the average signal energy is normalized as $\sum_k |c_k|^2 p_k = 1$, where p_k is the prior probability of signal or symbol c_k. We assume that every block of input symbols is encoded into a codeword matrix $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, ..., \boldsymbol{s}_w)$, where $\boldsymbol{s}_t = (s_{1,t}, s_{2,t}, ..., s_{K,t})^T$ is simultaneously transmitted, w is the codeword length in symbol periods, and ^{T} denotes transposition.[2] The channel is assumed to be Rayleigh flat fading, so that the complex path gain from transmit antenna i to receive antenna j, denoted by $H_{j,i}$, has a zero-mean unit-variance complex Gaussian distribution, denoted by $\mathcal{CN}(0,1)$, with i.i.d. real and imaginary parts. We assume that the receiver, but not the transmitter, has perfect knowledge of the path gains. Moreover, we assume that the channel is quasi-static, meaning that the path gains remain constant during a codeword transmission, but vary in an i.i.d. fashion from one codeword interval to the other. The additive noise at the j^{th} receive antenna at time t, $N_{j,t}$, is assumed to be $\mathcal{CN}(0,1)$ distributed with i.i.d. real and imaginary parts. We will assume that the input, fading coefficients, and channel noise are independent of each other.

Based on the above, for a CSNR of γ_s at each receive branch and at time t, the signal at receive antenna j can be written as $R_{j,t} = \sqrt{\frac{\gamma_s}{K}} \sum_{i=1}^{K} H_{j,i} s_{i,t} + N_{j,t}$, or in matrix form,

$$\boldsymbol{r}_t = \sqrt{\frac{\gamma_s}{K}} \boldsymbol{H} \boldsymbol{s}_t + \boldsymbol{n}_t, \tag{1}$$

where $\boldsymbol{r}_t = (R_{1,t}, R_{2,t}, ..., R_{L,t})^T$, $\boldsymbol{H} = \{H_{j,i}\}$, and $\boldsymbol{n}_t = (N_{1,t}, N_{2,t}, ..., N_{L,t})^T$.

Let us denote the received signals corresponding to $\boldsymbol{S}$ by $\boldsymbol{R} = (\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_w)$ and the *a priori* probability of codeword $\boldsymbol{S}$ by $p(\boldsymbol{S})$. Assuming that perfect channel state information is available, in MAP decoding one aims to maximize $P(\boldsymbol{S}|\boldsymbol{R}, \boldsymbol{H})$ over the codebook. The MAP decoding rule is hence given by

$$\begin{aligned}
&\arg\max_{\boldsymbol{S}} \quad P\{\boldsymbol{S}|\boldsymbol{R}, \boldsymbol{H}\} = \arg\max_{\boldsymbol{S}} \quad P\{\boldsymbol{R}|\boldsymbol{S}, \boldsymbol{H}\} p(\boldsymbol{S}) \\
&= \arg\max_{\boldsymbol{S}} \quad P\left\{ \boldsymbol{R} - \sqrt{\frac{\gamma_s}{K}} \boldsymbol{H}\boldsymbol{S} \middle| \boldsymbol{S}, \boldsymbol{H} \right\} p(\boldsymbol{S}) \\
&= \arg\max_{\boldsymbol{S}} \quad p(\boldsymbol{S}) \prod_{j,t} \exp\left\{ - \left| R_{j,t} - \sqrt{\frac{\gamma_s}{K}} \sum_i H_{j,i} s_{i,t} \right|^2 \right\} \\
&= \arg\min_{\boldsymbol{S}} \quad \left\{ -\ln\left(p(\boldsymbol{S})\right) + \sum_t \sum_j \left| R_{j,t} - \sqrt{\frac{\gamma_s}{K}} \sum_i H_{j,i} s_{i,t} \right|^2 \right\}.
\end{aligned} \tag{2}$$

III. General Space-Time Codes: The Codeword Pairwise Error Probability under MAP Decoding

The codeword PEP between $\boldsymbol{S}$ and $\hat{\boldsymbol{S}}$ is defined as the probability that $\boldsymbol{S}$ has a larger MAP metric in (2) than $\hat{\boldsymbol{S}}$ given that $\boldsymbol{S}$ is transmitted. Therefore, we have the equation at the top of the next page, where $d_{i,t} = s_{i,t} - \hat{s}_{i,t}$. The codeword PEP is therefore equal to (3), where $\langle x, y \rangle = \Re\{x\}\Re\{y\} + \Im\{x\}\Im\{y\}$ and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ indicate real and imaginary parts, respectively,

$$\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}} = \frac{\gamma_s}{K} \sum_t \sum_j \left| \sum_i H_{j,i} d_{i,t} \right|^2$$

and

$$\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}} = \frac{1}{2} \ln \frac{p(\boldsymbol{S})}{p(\hat{\boldsymbol{S}})}.$$

To compute the expectation (3) in closed-form, we determine the probability density function (pdf) of $\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}$, convert (3) into a linear combination of the derivatives of $\mathcal{L}\{Q(\sqrt{x})\}$, where $\mathcal{L}(\cdot)$ is the Laplace transform operator, and then evaluate these derivatives. First, we note that

$$\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}} = \frac{\gamma_s}{K} \sum_{t=1}^{w} \sum_{j=1}^{L} \left| \sum_{i=1}^{K} H_{j,i} d_{i,t} \right|^2 = \frac{\gamma_s}{K} \sum_{j=1}^{L} \boldsymbol{h}_j^\dagger \mathbf{U} \boldsymbol{h}_j, \tag{4}$$

where $u_{k,i} = \sum_t d_{i,t} d^*_{k,t}$, * denotes complex conjugation, and $\boldsymbol{h}_j$ is the transpose of the j^{th} row of $\boldsymbol{H}$. Since $\boldsymbol{U}$ is Hermitian (i.e., $\boldsymbol{U}^\dagger = \boldsymbol{U}$, where † represents complex conjugate transposition) and nonnegative definite, it can be decomposed as $\boldsymbol{U} = \boldsymbol{V}^\dagger \boldsymbol{D} \boldsymbol{V}$, where $\boldsymbol{D}$ is a nonnegative definite diagonal matrix having the eigenvalues of $\boldsymbol{U}$ on its main diagonal. $\boldsymbol{V}$ is a unitary matrix (i.e., $\boldsymbol{V}^\dagger \boldsymbol{V} = \mathbf{I}_K$) and its columns are the unit-norm eigenvectors of $\boldsymbol{U}$. Therefore, from (4), we have

$$\begin{aligned}
\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}} &= \frac{\gamma_s}{2K} \sum_{j=1}^{L} \boldsymbol{h}_j^\dagger \boldsymbol{V}^\dagger \boldsymbol{D} \boldsymbol{V} \boldsymbol{h}_j = \frac{\gamma_s}{2K} \sum_{j=1}^{L} \boldsymbol{x}_j^\dagger \boldsymbol{D} \boldsymbol{x}_j \\
&= \sum_{j=1}^{L} \sum_{i=1}^{K} \frac{\gamma_s}{2K} \lambda_i |X_{ji}|^2,
\end{aligned} \tag{5}$$

where $\boldsymbol{x}_j = \boldsymbol{V}\boldsymbol{h}_j$, X_{ji} is the i^{th} element of $\boldsymbol{x}_j$, and $\lambda_i = \boldsymbol{D}_{i,i}$. As $\boldsymbol{V}$ is unitary, the entries of $\boldsymbol{x}_j$ are i.i.d. $\mathcal{CN}(0,1)$ and the moment generating function (MGF) of $\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}$ is

$$\Phi_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}(s) = \prod_{k=1}^{Z} \frac{1}{\left(1 - \lambda_k \frac{\gamma_s}{2K} s\right)^{Ln_k}}, \tag{6}$$

where Z is the number of distinct non-zero λ_k's each of multiplicity n_k (with appropriate re-ordering of the eigenvalues). The pdf of a random variable Θ is the inverse Laplace transform ($\mathcal{L}^{-1}$) of $\Phi_\Theta(-s)$. In order to find $\mathcal{L}^{-1}\left\{\Phi_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}(-s)\right\}$, we convert (6) into a sum and then use the linearity of the Laplace transform. Letting $p_k = \frac{2K}{\gamma_s \lambda_k}$, we can write the partial-fraction expansion of (6) as

$$\Phi_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}(-s) = \prod_{k=1}^{Z} \frac{p_k^{Ln_k}}{(s+p_k)^{Ln_k}} = \sum_{k=1}^{Z} \sum_{i=0}^{Ln_k - 1} \frac{\alpha_{i+1,k}}{(s+p_k)^{i+1}}, \tag{7}$$

where

$$\alpha_{Ln_k - i, k} = \frac{1}{i!} \left\{ \frac{d^i}{ds^i} \left[(s+p_k)^{Ln_k} \Phi_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}(-s) \right] \right\} \Bigg|_{s=p_k}, \quad i = 0, ..., Ln_k - 1. \tag{8}$$

[2]Note that one can interpret w as the frame length and hence this model is general enough to include space-time trellis codes.

$$P(\boldsymbol{S}\to\hat{\boldsymbol{S}}|\boldsymbol{H}) = P\left\{\sum_t\sum_j\left|-\sqrt{\frac{\gamma_s}{K}}\sum_i H_{j,i}d_{i,t}+N_{j,t}\right|^2-\ln p(\hat{\boldsymbol{S}})<\sum_t\sum_j|N_{j,t}|^2-\ln p(\boldsymbol{S})\right\},$$

$$\begin{aligned}P(\boldsymbol{S}\to\hat{\boldsymbol{S}}) &= E_{\boldsymbol{H}}\left\{P\left\{\sqrt{2}\sum_t\sum_j\langle N_{j,t},\sum_i H_{j,i}d_{i,t}\rangle>\frac{1}{\sqrt{2}}\Delta_{\boldsymbol{S},\hat{\boldsymbol{S}}}+\frac{\sqrt{2}}{\Delta_{\boldsymbol{S},\hat{\boldsymbol{S}}}}\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}\right\}\right\}\\ &= E_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}\left\{Q\left(\sqrt{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}+\frac{1}{\sqrt{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}}\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}\right)\right\},\end{aligned} \tag{3}$$

Taking the inverse Laplace transform of the right hand side of (7), we have

$$f_{\frac{1}{2}\Delta^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}}(x)=\sum_{k=1}^{Z}\sum_{i=0}^{Ln_k-1}\frac{\alpha_{i+1,k}}{i!}x^i e^{-\frac{2K}{\gamma_s\lambda_k}x},\qquad x\geq 0. \tag{9}$$

The next step is simply using (9) to evaluate (3). This yields

$$P(\boldsymbol{S}\to\hat{\boldsymbol{S}})=\sum_{k=1}^{Z}\sum_{i=1}^{Ln_k}\frac{\alpha_{i,k}}{(i-1)!}\int_0^\infty y^{i-1}e^{-\frac{2K}{\gamma_s\lambda_k}y}Q\left(\sqrt{y}+\frac{\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}}{\sqrt{y}}\right)dy. \tag{10}$$

We note that the integral in (10) is the Laplace transform of $y^{i-1}Q\left(\sqrt{y}+\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}/\sqrt{y}\right)$ evaluated at $s=\delta_k^{-2}\triangleq 2K/\gamma_s\lambda_k$. We know that if $f(t)$ and $F(s)$ are Laplace transform pairs ($F(s)=\mathcal{L}\{f(t)\}$), so are $t^nf(t)$ and $(-1)^n\frac{d^n}{ds^n}F(s)$. Therefore, we need to find the $i-1^{\text{st}}$ derivative of $\mathcal{L}\left\{Q\left(\sqrt{y}+\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}/\sqrt{y}\right)\right\}$. Using integration by parts ($\int_0^\infty udv=uv|_0^\infty-\int_0^\infty vdu$) together with the identity (see [1])

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}}\int_0^\infty\exp\left\{-\frac{1}{2}(2s+1)\left(x+\frac{a}{x\sqrt{2s+1}}\right)^2\right\}dx\\ &\quad=\frac{1}{2\sqrt{2s+1}}\exp\left\{-(a+|a|)\sqrt{2s+1}\right\},\end{aligned}$$

we can show that $F_{\text{MAP}}(s)$, the Laplace transform of $Q\left(\sqrt{y}+\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}/\sqrt{y}\right)$, is equal to (11) shown at the top of the next page, where $\text{sgn}(x)=\frac{|x|}{x}$ if $x\neq 0$ and 0 otherwise. The term in the sum in (10) is simply

$$\frac{\alpha_{i,k}}{(i-1)!}(-1)^{i-1}\frac{d^{i-1}}{ds^{i-1}}F_{\text{MAP}}(s)\triangleq\alpha_{i,k}\delta_k^2\pi(i,\delta_k,\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}).$$

We use the Leibniz's formula for the i^{th} derivative of a product [1, Eq. 3.3.8], and a formula for the i^{th} derivative of a composite function [12] as well as induction to find the i^{th} derivative of (11) (see equation (33) in the Appendix). The result is the following expression for the exact codeword PEP of MAP decoded space-time codes

$$P(\boldsymbol{S}\to\hat{\boldsymbol{S}})=\sum_{k=1}^{Z}\sum_{i=1}^{Ln_k}\delta_k^{2i}\alpha_{i,k}\pi(i,\delta_k,\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}), \tag{12}$$

where $\pi(n,\delta,\lambda)$ is given in (13) on the next page and $\sum_{i=L}^{U}z_i\triangleq 1$ if $L>U$. The above formula is also valid for MAP decoding of codewords in single-input multi-output systems under slow Rayleigh fading.

IV. STOB Codes and MAP Decoding

A. The MAP Decoding Rule

Let $\boldsymbol{c}=(c_1,...,c_\tau)^T$ be a vector of τ consecutive symbols input to the STOB encoder and $\boldsymbol{S}=(\boldsymbol{s}_1,...,\boldsymbol{s}_w)$ be the space-time codeword corresponding to it. In the case of STOB codes, we have $w=g\tau$, where g is the coding gain and $\boldsymbol{S}\boldsymbol{S}^\dagger=g\|\boldsymbol{c}\|^2\mathbf{I}_K$. As an example, for the code $\mathcal{G}^3$ in [28], $w=8$, $\tau=4$, and $g=2$, and for Alamouti's code [4], $g=1$ and $w=\tau=2$. It can be shown that (1) can be re-written as [20]

$$\bar{\boldsymbol{r}}^j=\sqrt{\frac{\gamma_s}{K}}\bar{\boldsymbol{H}}^j\boldsymbol{c}+\bar{\boldsymbol{n}}^j\qquad j=1,\cdots,L, \tag{14}$$

where $\bar{R}_{j,t}=R_{j,t}$ and $\bar{N}_{j,t}=N_{j,t}$ for $1\leq t\leq w/2$, $\bar{R}_{j,t}=R^*_{j,t}$ and $\bar{N}_{j,t}=N^*_{j,t}$ for $w/2<t\leq w$, and $\bar{\boldsymbol{H}}^j$ is derived from the j^{th} row of $\boldsymbol{H}$ via negation and complex conjugation of some of its entries (see [20, equations (10) and (19)] for two examples). It is clear that $\bar{N}_{j,t}$ are i.i.d. $\mathcal{CN}(0,1)$. The matrix $\bar{\boldsymbol{H}}^j$ has orthogonal columns, i.e., $\bar{\boldsymbol{H}}^{j\dagger}\bar{\boldsymbol{H}}^j=gY_j\mathbf{I}_\tau$, where $Y_j=\sum_i|H_{j,i}|^2$. Therefore, (14) can be multiplied from the left by $\bar{\boldsymbol{H}}^{j\dagger}$ to yield

$$\tilde{\boldsymbol{r}}^j\triangleq\bar{\boldsymbol{H}}^{j\dagger}\bar{\boldsymbol{r}}^j=g\sqrt{\frac{\gamma_s}{K}}Y_j\boldsymbol{c}+\tilde{\boldsymbol{n}}^j, \tag{15}$$

where $\tilde{\boldsymbol{n}}^j\triangleq\bar{\boldsymbol{H}}^{j\dagger}\bar{\boldsymbol{n}}^j$. Note that each entry of $\tilde{\boldsymbol{r}}^j$ is associated with only *one* symbol. Therefore, if we show that the noise vector $\tilde{\boldsymbol{n}}^j$ is composed of i.i.d. random variables, then we can detect symbol i by only considering the i^{th} entry of the vectors $\tilde{\boldsymbol{r}}^j$, $1\leq j\leq L$.

In order to find the distribution of the noise vector $\tilde{\boldsymbol{n}}^j$, we consider two noise samples $\tilde{N}_{k,\tau_1}$ and $\tilde{N}_{q,\tau_2}$ at two arbitrary symbol intervals τ_1 and τ_2, and for two arbitrary receive antennas k and q. Notice that $\tilde{N}_{k,\tau_1}$ and $\tilde{N}_{q,\tau_2}$ are weighted sums of normal random variables, and hence they have Gaussian distribution. Also, it is straightforward to verify that $E\{\tilde{N}_{k,\tau_1}\}=E\{\tilde{N}_{q,\tau_2}\}=0$. Hence, the correlation of these noise samples is

$$\begin{aligned} F_{\text{MAP}}(s) &= \int_0^\infty \underbrace{Q\left(\sqrt{y}+\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}/\sqrt{y}\right)}_{u}\underbrace{e^{-ys}dy}_{dv} \\ &= \frac{1}{s\sqrt{2\pi}}\int_0^\infty \left\{\frac{d}{dy}\int_{\sqrt{y}+\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}/\sqrt{y}}^\infty e^{-t^2/2}dt\right\}e^{-ys}dy \\ &= -\frac{1}{s\sqrt{2\pi}}\int_0^\infty \left(\frac{1}{2\sqrt{y}}-\frac{\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}}{2y\sqrt{y}}\right)\exp\left\{-\frac{1}{2}\left(y+\Lambda^2_{\boldsymbol{S},\hat{\boldsymbol{S}}}/y+2\Lambda_{s,\hat{s}}\right)\right\}dy \\ &= \frac{1-\text{sgn}(\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}})}{2s}-\frac{1}{2}\left(\frac{1}{s\sqrt{2s+1}}+\frac{\text{sgn}(\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}})}{s}\right)e^{-(\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}+|\Lambda_{\boldsymbol{S},\hat{\boldsymbol{S}}}|\sqrt{2s+1})}, \end{aligned} \tag{11}$$

$$\begin{aligned} \pi(n,\delta,\lambda) &= \frac{1-\text{sgn}(\lambda)}{2}-\frac{1}{2}e^{-\left(\lambda+|\lambda|\sqrt{2\delta^{-2}+1}\right)}\times \\ &\quad \sum_{k=0}^{n-1}\frac{(-1)^{n+k-1}}{(2+\delta^2)^{n-k-1}}\left(-\text{sgn}(\lambda)+\frac{\delta}{\sqrt{2+\delta^2}}\sum_{m=0}^{k}\binom{2m}{m}\frac{1}{(2\delta^2+4)^m}\right)\times \\ &\quad \sum_{l=1}^{n-k-1}\frac{|\lambda|^l(\delta^2+2)^{l/2}}{l!\delta^l}\sum_{p=0}^{l-1}\binom{l}{p}(-1)^{l+p}\prod_{q=0}^{n-k-2}(l-p-2q) \end{aligned} \tag{13}$$

$$\begin{aligned} E\left\{\tilde{N}_{k,\tau_1}\tilde{N}^*_{q,\tau_2}\right\} &= E\left\{\left(\sum_{i=1}^{w}\bar{H}^{k*}_{i,\tau_1}\bar{N}_{k,i}\right)\left(\sum_{j=1}^{w}\bar{H}^{q*}_{j,\tau_2}\bar{N}_{q,j}\right)^*\right\} \\ &= \sum_i\sum_j \bar{H}^{k*}_{i,\tau_1}\bar{H}^{q}_{j,\tau_2}E\left\{\bar{N}_{k,i}\bar{N}^*_{q,j}\right\} \end{aligned}$$

As the $\bar{N}_{j,t}$ are zero-mean i.i.d., the above double sum is zero unless $k=q$ and $i=j$, in which case it equals

$$\begin{aligned} &\sum_{i=1}^{w}\bar{H}^{k*}_{i,\tau_1}\bar{H}^{k}_{i,\tau_2}E\left\{|\bar{N}_{k,i}|^2\right\} \\ &\quad = \sum_{i=1}^{w}\bar{H}^{k*}_{i,\tau_1}\bar{H}^{k}_{i,\tau_2} \qquad (16) \\ &\quad = \begin{cases} \sum_i |\bar{H}^k_{i,\tau_1}|^2 = gY_k & \text{if } \tau_1=\tau_2 \\ 0 & \text{otherwise,} \end{cases} \qquad (17) \end{aligned}$$

where (16) follows from the fact that $\bar{N}_{j,t}$ is unit-variance and (17) follows from the orthogonality of $\bar{\boldsymbol{H}}$. Therefore, we have just shown that

$$\tilde{N}_{k,i} \sim \text{i.i.d. } \mathcal{CN}(0, gY_k). \tag{18}$$

For MAP decoding, $\tilde{\boldsymbol{R}}$ can be used instead of $\boldsymbol{R}$ because $\tilde{\boldsymbol{R}}$ is an invertible function of $\boldsymbol{R}$. The detection rule is given by

$$\begin{aligned} c_t &= \arg\max_c P(c|\{\tilde{R}_{l,t}\}_{l=1}^L, \boldsymbol{H}) \\ &= \arg\max_c f(\{\tilde{R}_{l,t}\}_{l=1}^L|c,\boldsymbol{H})\cdot p(c) \\ &= \arg\max_c \prod_{l=1}^{L} f_{\tilde{N}_{l,t}}(\{\tilde{R}_{l,t}-g'Y_lc\}_{l=1}^L)\cdot p(c) \qquad (19) \\ &= \arg\max_c \left\{\ln\left(p(c)\right)-\sum_{l=1}^{L}\frac{|\tilde{R}_{l,t}-g'Y_lc|^2}{gY_l}\right\}, \qquad (20) \end{aligned}$$

where $g'=g\sqrt{\frac{\gamma_s}{K}}$, and (19) and (20) are because $\tilde{N}_{k,t}$ are i.i.d. and Gaussian, respectively, as indicated in (18).

B. The Exact Symbol Pairwise Error Probability

1) The Conditional PEP: Without loss of generality, we consider MAP decoding for the k^{th} symbol period. The error probabilities may be determined using the MAP detection metric given in (20). The receiver should evaluate this metric for the symbols c_i and c_j given that c_i is transmitted (hence $\tilde{R}_{l,k}=g'Y_lc_i+\tilde{N}_{l,k}$) and decide in favor of the one which yields a larger result. Let us denote by $P(c_i\rightarrow c_j)$ the probability that c_j is preferred over c_i (i.e., c_j has a larger metric than c_i in (20)) when c_i is sent. From (20), we want to determine the probability of the following event

$$\begin{aligned} \ln(p(c_i))-\frac{1}{2g}\sum_{l=1}^{L}\frac{|\tilde{R}_{l,k}-g'Y_lc_i|^2}{Y_l} &\le \ln(p(c_j)) \\ -\frac{1}{2g}\sum_{l=1}^{L}\frac{|\tilde{R}_{l,k}-g'Y_lc_j|^2}{Y_l}&, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\sum_{l=1}^{L}\frac{\sqrt{2}\langle c_j-c_i,\tilde{N}_{l,k}\rangle}{|c_j-c_i|} \\ &\quad \ge \frac{1}{|c_j-c_i|}\sqrt{\frac{K}{2\gamma_s}}\ln\frac{p(c_i)}{p(c_j)}+g|c_j-c_i|\sqrt{\frac{\gamma_s}{2K}}\sum_{l=1}^{L}Y_l. \end{aligned} \tag{21}$$

From (18), it follows that $\sqrt{2}\langle c_j-c_i,\tilde{N}_{l,k}\rangle/|c_j-c_i|$ is i.i.d. $\mathcal{CN}(0,gY_l)$. Hence the sum on the left hand side of (21) is $\mathcal{CN}(0,g\sum_{l=1}^{L}Y_l)$. Therefore, the probability of the event in (21), which is the PEP conditioned on the path gains, is given

by

$$P(c_i \to c_j|\boldsymbol{H}) = Q\left(\sqrt{\frac{g\gamma_s}{2K}}|c_i - c_j|\sqrt{Y} + \sqrt{\frac{K}{2g\gamma_s}}\frac{1}{|c_i - c_j|}\ln\frac{p(c_i)}{p(c_j)}\frac{1}{\sqrt{Y}}\right), \tag{22}$$

where

$$Y = \sum_{l=1}^{L} Y_l = \sum_{k=1}^{K}\sum_{l=1}^{L} |H_{l,k}|^2 \tag{23}$$

is the sum of the squared magnitudes of all the path gains.

2) The Symbol PEP in Closed Form: To find the unconditional symbol PEP, we should average (22) with respect to Y. Defining random variables ζ_i as $\zeta_{(j-1)K+i} = \Re\{H_{j,i}\}$ and $\zeta_{KL+(j-1)K+i} = \Im\{H_{j,i}\}$ for $i = 1, ..., K$ and $j = 1, ..., L$, we note that $\zeta_k \sim$ i.i.d. $\mathcal{N}(0, \frac{1}{2})$, and we can write Y as

$$Y = \sum_{i,j} |H_{j,i}|^2 = \sum_{i,j} \Re\{H_{j,i}\}^2 + \Im\{H_{j,i}\} = \sum_{i=1}^{2n} \zeta_i^2,$$

where $n = KL$. Using the moment generating function of normal random variables yields the probability density function of Y as

$$f_Y(y) = \frac{1}{(n-1)!} y^{n-1} e^{-y}, \qquad y \geq 0.$$

Hence, Y has a scaled chi-square distribution with $2n$ degrees of freedom. The average of (22) can then be written as

$$P(c_i \to c_j) = \frac{1}{(n-1)!\delta_{ij}^{2n}} \int_0^\infty y^{n-1} e^{-y/\delta_{ij}^2} Q\left(\sqrt{y} + \frac{\lambda_{ij}}{\sqrt{y}}\right) dy, \tag{24}$$

where $\delta_{ij} = \sqrt{\frac{g\gamma_s}{2K}}|c_i - c_j|$ and $\lambda_{ij} = \frac{1}{2}\ln\frac{p(c_i)}{p(c_j)}$. We note that the the above integral is the same as the one in (10). Therefore, we obtain

$$P(c_i \to c_j) = \frac{1}{(n-1)!\delta_{ij}^{2n}} (-1)^{n-1} \left.\frac{d^n}{ds^n} F_{\text{MAP}}(s)\right|_{s=\delta_{ij}^{-2}}.$$

Therefore, we obtain the PEP between a pair of space-time orthogonal block coded symbols under MAP decoding as

$$P(c_i \to c_j) = \pi(n, \delta_{ij}, \lambda_{ij}), \tag{25}$$

with $\pi(n, \delta_{ij}, \lambda_{ij})$ given in (13). Note that the above formula also holds for MAP decoding of symbols under MRC, as the received signal in those systems has the same form as in (15) with $K = 1$.

When the c_i's are equally likely, MAP decoding reduces to ML decoding and we have $\lambda_{ij} = \frac{1}{2}\ln\frac{p(c_i)}{p(c_j)} = 0$. Hence, the first sum in (13) is non-zero only for $k = n - 1$ and we have

$$P(c_i \to c_j) = \frac{1}{2}\left(1 - \frac{\delta_{ij}}{\sqrt{2 + \delta_{ij}^2}} \sum_{k=0}^{n-1} \binom{2k}{k} \frac{1}{(2\delta_{ij}^2 + 4)^k}\right),$$

which agrees with the result derived in [6].

V. Applications

A. SER and BER of STOB Coded or MRC Systems

In [7], we established tight algorithmic bounds on the SER and BER of STOB coded MIMO systems with arbitrary signaling schemes and bit-to-signal mappings under slow Rayleigh fading and ML decoding. To compute these bounds, one needs to calculate the probability of symbol pairwise error events (i.e., the PEP in (25)) as well as the probability of the intersection of pairs of such error events. Closed-form expressions for these probabilities were derived in [7]. The work in [7] is further extended to the MAP decoding case in [8]. Alternatively, one can use the geometric approach in [25], [24] to compute the exact values of the SER and BER of STOB coded systems. Although the method in [25], [24] is implemented for the AWGN channel only, it can be extended to the STOB coded case using the pairwise error probabilities given in [8].

B. The Optimum Binary Antipodal Signaling for STOB Codes/MRC

In this section we consider binary antipodal signaling and optimize it in the sense of minimizing the BER given by

$$\text{BER} = P(c_1 \to c_2) \cdot p(c_1) + P(c_2 \to c_1) \cdot p(c_2). \tag{26}$$

Normally, one should use the averaged PEP in (26) with $c_1 = a$ and $c_2 = -b$ and find the optimal a and b via differentiation. However, this can be a tedious job in view of the PEP given in (25). Therefore, we use the PEPs at the *receiver* side, i.e., given $\boldsymbol{H}$, to find the solution in an easier way. The optimal constellation derived in this way will not depend on $\boldsymbol{H}$, justifying our approach.

Let us assume that $p(c_1) = p$, and the bits 0 and 1 are mapped to $c_2 = -b$ and $c_1 = a$, respectively. Letting $\beta = \sqrt{\frac{2g\gamma_s}{K}}$, $\sqrt{A} = \beta\frac{(a+b)}{2}\sqrt{Y}$, and $B = \frac{1}{2}\ln\frac{1-p}{p}$, with Y as defined below (22), we can write the BER conditioned on $\boldsymbol{H}$ as

$$\text{BER}_Y = pQ\left(\sqrt{A} - \frac{B}{\sqrt{A}}\right) + (1-p)Q\left(\sqrt{A} + \frac{B}{\sqrt{A}}\right). \tag{27}$$

It is easy to verify that the BER is a strictly decreasing function of A (for any fixed B). Hence, given a constellation energy E_s and p, in order to minimize the BER, one has to maximize A. Note that A is a scaled distance between the constellation points; therefore, signaling schemes with the same distance between their signals have identical performance. It is clear that the constellation with constant average signal energy E_s which maximizes A is the zero-mean constellation, because a constellation with a non-zero mean can simply be shifted to reduce its energy without performance loss.

From the zero-mean condition, we have $b = \frac{p}{1-p}a$, and the average energy condition requires that

$$pa^2 + (1-p)b^2 = E_s.$$

The above two equalities result in

$$(-b, a) = \sqrt{E_s}\left(-\sqrt{\frac{p}{1-p}}, \sqrt{\frac{1-p}{p}}\right),$$

which is therefore the optimal binary antipodal constellation. The above constellation is identical to the antipodal signaling result in [16] for the case of the AWGN channel. As mentioned in Section II, we set $E_s = 1$.

C. Linear Dispersion Code Design for MAP Decoding

Linear dispersion (LD) codes (introduced in [13]) constitute an important class of space-time block codes. Every entry of an LD codeword is a weighted sum of the baseband signals with the weights chosen such that the mutual information between the channel input and output is maximized given the number of transmit and receive antennas. An LD codeword is written as

$$\boldsymbol{S} = \sum_{m=1}^{M} (\alpha_m \boldsymbol{A}_m + j\beta_m \boldsymbol{B}_m),$$

where $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ are $K \times w$ matrices (similar to [13], we assume that $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ have real entries), $c_m = \alpha_m + j\beta_m$ is a symbol to be encoded ($j = \sqrt{-1}$), and M is the block length in symbols (i.e., the number of data symbols to be encoded at a time).

Instead of maximizing the mutual information, here we opt to design LD codes via minimizing the union upper bound on the frame (block) error rate which is given by

$$\sum_{\boldsymbol{S}} p(\boldsymbol{S}) \sum_{\hat{\boldsymbol{S}} \neq \boldsymbol{S}} P(\boldsymbol{S} \to \hat{\boldsymbol{S}}).$$

This bound can be computed using the codeword PEP formula given in (12).

There are a maximum of two distinct eigenvalues for a system with two transmit antennas. It can be verified that the following cases are possible:

- Only one non-zero eigenvalue ($Z = 1$ in (6)):
  In this case the eigenvalues can be equal ($n_k = 2$) or one of them is zero ($n_k = 1$). For the first case, we have
  $$\alpha_{2L,1} = p_1^{2L},$$
  and $\alpha_{i,k} = 0$ otherwise, and for the second case, we have
  $$\alpha_{L,1} = p_1^{L},$$
  and $\alpha_{i,k} = 1$ for $i < L$ or $k = 2$.
- Two distinct non-zero eigenvalues ($Z = 2$ in (6)):
  We have $n_1 = n_2 = 1$ and
  $$\alpha_{L-p,1} = \frac{(-1)^p \prod_{j=1}^{p}(L+j-1)}{p!(p_2 - p_1)^{L+p}} p_1^p p_2^L, \quad p = 0, 1, \cdots, L-1$$
  $$\alpha_{L-p,2} = \frac{(-1)^p \prod_{j=1}^{p}(L+j-1)}{p!(p_1 - p_2)^{L+p}} p_1^L p_2^p, \quad p = 0, 1, \cdots, L-1$$

Our design method is as follows: to guarantee maximum throughput and to make sure that the performance is always better than V-BLAST, we begin with the $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ matrices which correspond to V-BLAST and are given by [11], [13]

$$\boldsymbol{A}_{K(\tau-1)+k} = \boldsymbol{B}_{K(\tau-1)+k} = \boldsymbol{c}_\tau \boldsymbol{d}_k^T, \quad \tau = 1, ..., w, k = 1, ..., K,$$

where $\boldsymbol{c}_\tau$ and $\boldsymbol{d}_k$ are w-dimensional and k-dimensional column all-zero vectors except for a 1 in the τ^{th} and k^{th} entries, respectively. We then improve the code by adding zero-mean Gaussian noise to the $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ matrices, normalizing the $\boldsymbol{A}_m$ and $\boldsymbol{B}_m$ matrices to satisfy the power constraint which, in the uniform-source case, is given by [13, equation(18)]

$$\sum_m (\text{tr}\boldsymbol{A}_m^\dagger \boldsymbol{A}_m + \text{tr}\boldsymbol{B}_m^\dagger \boldsymbol{B}_m) = 2wK, \tag{28}$$

and updating the code if the new FER union bound decreases. We have chosen the variance of the additive noise to decrease according to

$$\sigma_n^2 = 0.25\left(1 - \frac{i}{I_{\max}}\right)^3, i = 1, 2, ..., I_{\max},$$

where i is the iteration number. This regime is chosen following [32] due to its fast convergence rate and good results.

Note that this new code is still a linear dispersion code. Therefore:

1) Similar to [13], we have noticed that the performance of the resulting codes is not sensitive to the design CSNR. Therefore, to avoid numerical problems resulting from the addition of very small numbers, we set the design CSNR at 5 to 10 dB, depending on the number of antennas.
2) The design criterion in [13] is to maximize the mutual information between channel input and output under the assumption that the real and imaginary parts of the signal set have $\mathcal{N}(0, \frac{1}{2})$ distribution, which may be far from the particular signaling scheme and non-uniform distribution to be used. In our method, we optimize the code for the particular signaling scheme and prior probabilities which are going to be used. Obviously, the design method works as well with the assumption of having $\mathcal{N}(0, 1)$ distribution for the signals.
3) The power constraint (28) can be made more restrictive. For example, one could use $\boldsymbol{A}_m^\dagger \boldsymbol{A}_m = \boldsymbol{B}_m^\dagger \boldsymbol{B}_m = \frac{w}{M}\mathbf{I}_K$ for $m = 1, ..., M$. It is noted in [13] that this power constraint generally leads to lower error rates, but we have used (28) in our design to "relax" the condition as much as possible and let the search algorithm converge to any local minimum which satisfies the power constraint.
4) Obviously, the search method is random and may converge to a local minimum. The variance of the additive noise is large at the initial loops to allow large improvements, but it reduces with the iterations to allow convergence and small refinement. We have observed that many small changes are made at lower noise variances.
5) In order to have the possibility of finding better minima, we run the algorithm twice with the second round initialized with the results of the first. This allows "escaping" from a bad local minimum at the beginning of the second round, when the variance of the additive noise is large.

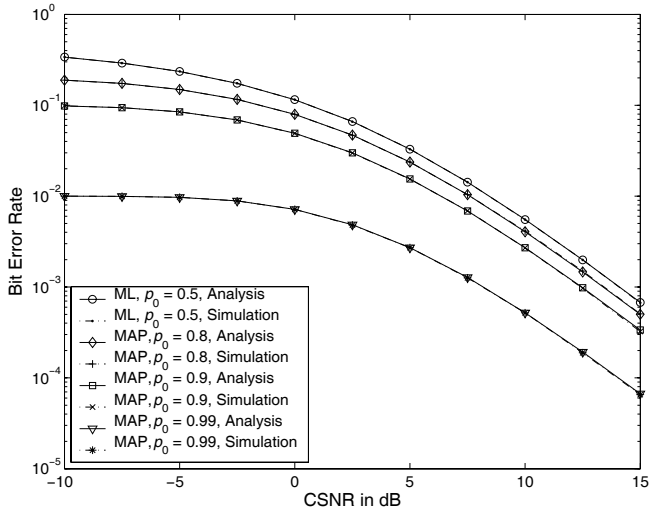

Fig. 1. Results for BPSK signaling, $K = 2, L = 1$, and the $\mathcal{G}^2$ STOB code.

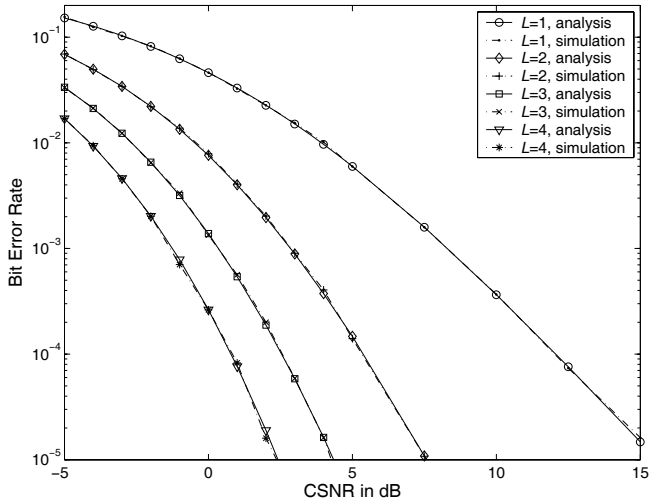

Fig. 2. Results for BPSK signaling, $K = 3$, L varying, and the $\mathcal{G}^3$ STOB code.

VI. Numerical Results

A. Binary Antipodal Signaling

It suffices to study the BER of binary antipodal signaling to show the exactness of our symbol PEP formula in (25). We simulate the transmission of an i.i.d. bit sequence over MIMO channels. The length of the bit-sequence is $\max(\frac{100}{\text{BER}}, 10^6)$ bits.

We consider a system with two transmit and one receive antennas (which uses Alamouti's code [4]) in Figure 1 with various values of p_0. Another system with three transmit and various numbers of receive antennas (and the code $\mathcal{G}^3$ in [28]) is considered in Figure 2. It is observed that the analysis and simulation curves coincide everywhere. In Figure 3, we compare four systems: two systems with BPSK signaling and ML or MAP decoding, and two systems with optimum signaling and ML or MAP decoding. These systems are indicated by ML BPSK, MAP BPSK, ML optimum, and MAP optimum, respectively. The source is an i.i.d. bit-stream with

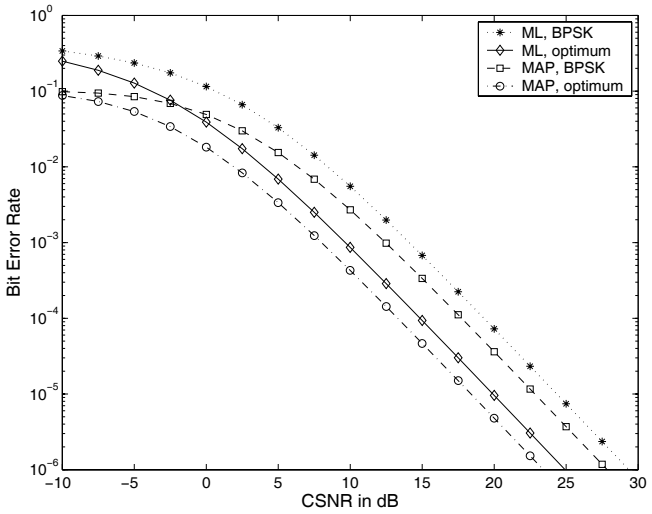

Fig. 3. Comparison between BPSK and optimum signaling schemes for $p_0 = 0.9$, $K = 2, L = 1$, and the $\mathcal{G}^2$ STOB code.

$p_0 = 0.9$. At a BER of 10^{-6}, the gain of using the modified constellation and MAP detection is about 6 dB over the ML decoded system with BPSK modulation. For the same BER, the gain over the MAP decoded BPSK system is 4.7 dB. We also observe that if the CSNR is high enough, optimum signaling and ML detection outperforms BPSK and MAP detection.

When the additive noise at the receiver is strong (i.e., at low γ_s), the second term in the argument of the $Q(\cdot)$ function in (22) has the dominant effect; hence MAP decoding with BPSK signals is more effective than ML decoding with optimum signals. In less noisy channel conditions (high γ_s), the first term in the argument of the $Q(\cdot)$ function becomes dominant and hence ML decoding with optimum signaling outperforms MAP decoding with symmetric signals. As previously mentioned, MAP decoding with optimum signaling is always better than other systems.

B. Tandem versus Joint Source-Channel Coding

Figure 4 compares a MAP decoded system with two tandem systems (under an identical overall rate) for a dual-transmit single-receive channel with 16-QAM modulation. The input bit-stream is i.i.d. with $p_0 = 0.89$ so that the source entropy is $H(X) = 0.5$. The tandem systems consist of 4$^{\text{th}}$ order Huffman coding followed by one of a 16-state or 64-state rate-1/2 convolutional coding blocks. The convolutional codes are non-systematic and chosen from [15]. The length of the input bit-stream is 2×10^6 bits. The test is repeated 500 times and the average BER is reported. It is observed that the tandem system breaks down (due to error propagation in the Huffman decoder) for CSNR < 25 dB. The MAP-decoded system outperforms the 16-state tandem coded system for BER $> 2 \times 10^{-7}$. It also outperforms the 64-state tandem coded system for BER $> 3 \times 10^{-7}$. Therefore, the jointly coded system has superior performance for the BER range of interest in many systems in practice. This is a typical behavior of to-date joint source-channel coding designs (that do not employ Turbo/LDPC codes): they are superior to tandem systems for

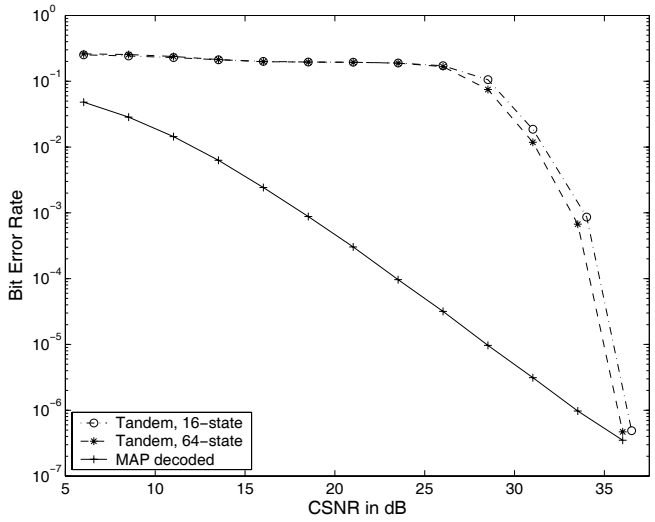

Fig. 4. Comparison between tandem and MAP-decoded schemes. 16-QAM signaling, $K = 2, L = 1$, $p_0 = 0.89$, and the $\mathcal{G}^2$ STOB code.

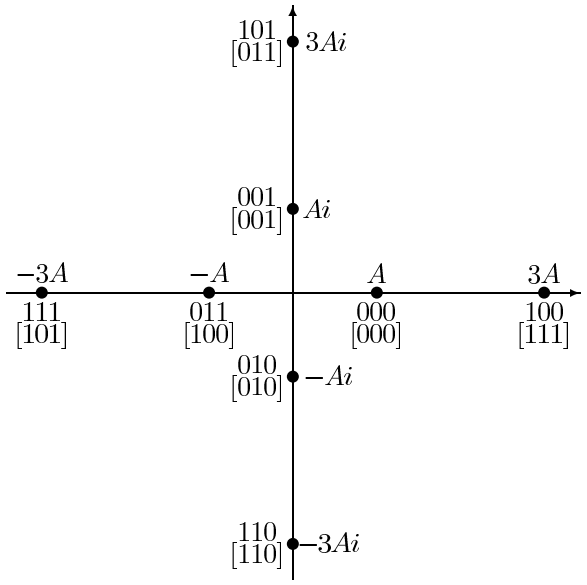

Fig. 5. The star-QAM signaling scheme with quasi-Gray and M1 (in brackets) mappings.

a (sometimes wide) range of CSNRs below a certain threshold [2], [3], [5], [17]. A similar study as in [34] can also be done, where it is shown that joint source-channel Turbo codes outperform tandem systems that use source coding (Huffman coding) and classical (channel-coding based) Turbo codes.

C. Constellation Mapping with STOB Codes

We next demonstrate that a large gain can be achieved via signal mappings designed according to the source non-uniform distribution over Gray and quasi-Gray mappings. The M1 mapping is introduced in [26] and is designed for the transmission of non-uniform binary sources over single antenna AWGN channels. It minimizes the SER union bound for single antenna Rayleigh fading channels with M-ary PSK and square QAM signaling. Here we use the guidelines in

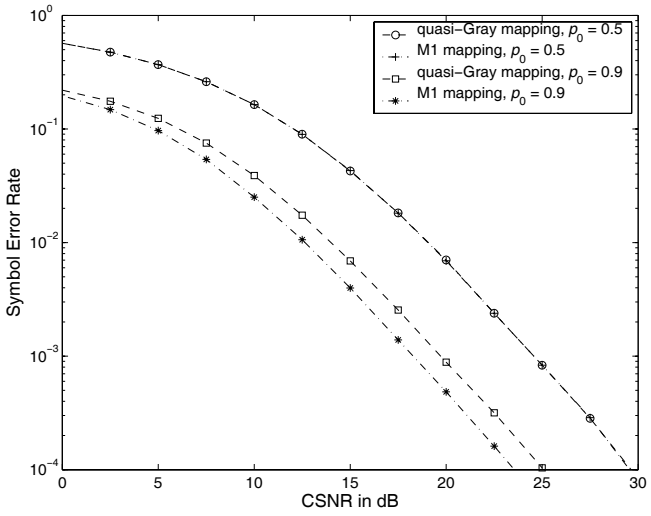

Fig. 6. Comparison between the Gray and the M1 mappings for Star 8-QAM modulation, $K = 2, L = 1$, and the $\mathcal{G}^2$ STOB code.

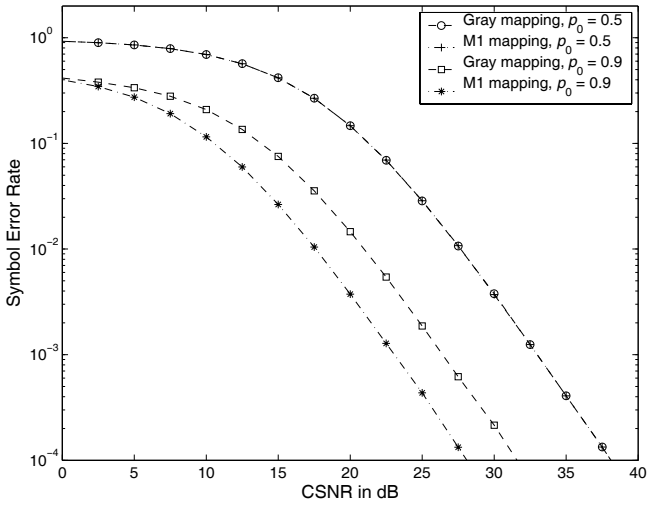

Fig. 7. Comparison between the Gray and the M1 mappings for 64-QAM modulation, $K = 2, L = 1$, and the $\mathcal{G}^2$ STOB code.

[26] with the symbol PEP formula in (25) to obtain the M1 mapping for the Star 8-QAM (shown in Figure 5 together with the Gray signal labeling) and 64-QAM (shown in [26, Fig. 9] also together with the Gray signal labeling) signal sets. Figures 6 and 7 compare the SER curves for the Gray and M1 mappings for the Star 8-QAM and 64-QAM signal sets, respectively. These figures show that the M1 map performs very well for STOB coded MIMO channels, even for small signal sets such as Star 8-QAM. The gain of the M1 mapping over Gray mapping is 1.4 dB for Star 8-QAM and 3.7 dB for 64-QAM at SER = 10^{-3}. The gain due to source redundancy is 10.4 dB for 64-QAM signaling.

D. Linear Dispersion Code Design

Figures 8 and 9 demonstrate the FER and BER performance of our LD code search method for a non-uniform binary i.i.d. source with $p_0 = 0.9$, respectively, for a dual-transmit dual-receive antenna system.

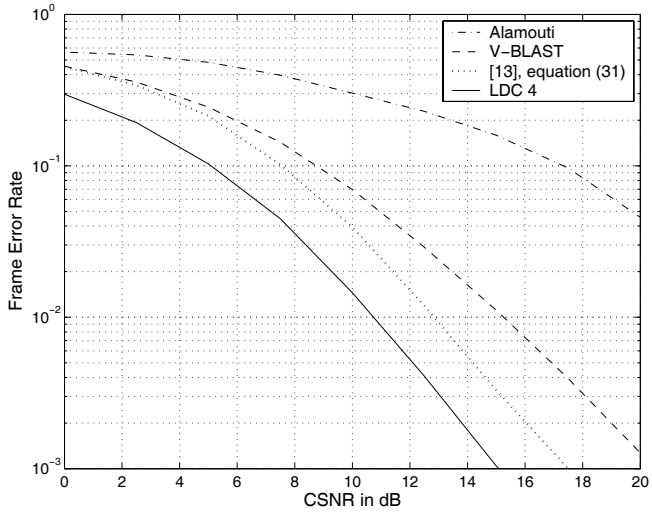

Fig. 8. Comparison between V-BLAST, $\mathcal{G}^2$, the LD code of [13, eq. (31)], and the new code; $K = 2, L = 2$. Q-PSK modulation (16-QAM for $\mathcal{G}^2$) and MAP decoding with $p_0 = 0.9$.

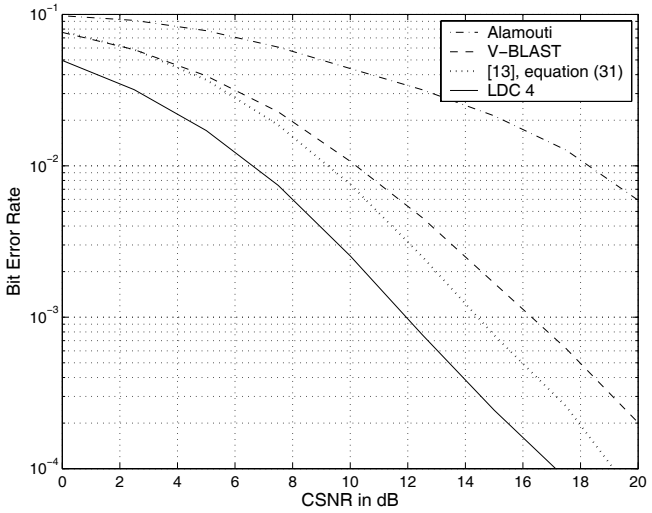

Fig. 9. Comparison between V-BLAST, $\mathcal{G}^2$, the LD code of [13, eq. (31)], and the new code; $K = 2, L = 2$. Q-PSK modulation (16-QAM for $\mathcal{G}^2$) and MAP decoding with $p_0 = 0.9$.

As mentioned in Section V-C, we start from a V-BLAST structure, so $M = 4$. It is observed that the new code significantly outperforms V-BLAST and its gain over V-BLAST continues to grow as the CSNR increases. We believe that this behavior is due to the larger diversity order of LD codes over V-BLAST, as the LD codes send each signal over all transmit antennas while in V-BLAST each signal is sent once from only one transmit antenna, so it experiences only one fading coefficient.

The new code, which will hereafter be referred to as LDC4, is found as

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.65782454 & -1.30819861 \\ -0.10725738 & 0.20531537 \end{bmatrix},$$
$$\boldsymbol{A}_2 = \begin{bmatrix} 0.25968977 & 0.60373361 \\ 0.97828103 & -0.51361949 \end{bmatrix},$$
$$\boldsymbol{A}_3 = \begin{bmatrix} -0.04899691 & 0.56160243 \\ -0.81070085 & -0.35572984 \end{bmatrix},$$
$$\boldsymbol{A}_4 = \begin{bmatrix} -0.70479974 & 0.22760978 \\ 0.04377761 & 0.68613179 \end{bmatrix},$$

and

$$\boldsymbol{B}_1 = \begin{bmatrix} 1.19165069 & -0.11729547 \\ 0.25381635 & 1.74228787 \end{bmatrix},$$
$$\boldsymbol{B}_2 = \begin{bmatrix} -0.98310451 & -1.02226963 \\ -0.26612417 & -1.03714274 \end{bmatrix},$$
$$\boldsymbol{B}_3 = \begin{bmatrix} -0.37841861 & 2.14539824 \\ -0.44476603 & 1.10551397 \end{bmatrix},$$
$$\boldsymbol{B}_4 = \begin{bmatrix} -0.98806533 & 0.27960017 \\ 0.58266695 & 1.40274305 \end{bmatrix}.$$

In order to have the same information rate (of 4 bits/channel use) in our comparison, we use 16-QAM signaling for Alamouti's $\mathcal{G}^2$ code, and Q-PSK signaling otherwise. At FER = 10^{-2}, the CSNR gain for LDC4 is 12.3 dB over $\mathcal{G}^2$, 3.5 dB over V-BLAST, and 1.1 dB over the code of [13, eq. (31)]. At BER = 10^{-3}, the CSNR gain is 13 dB over $\mathcal{G}^2$, 4.3 dB over V-BLAST, and 2.4 dB over the code of [13, eq. (31)], respectively.

When the source is uniform, Alamouti's code offers however a superior performance for 2 transmit antennas and a small number of receive antennas. For this case ($p_0 = 0.5$), we again considered a 2×2 system. When initialized with Alamouti's code and Q-PSK modulation, our search algorithm could not find a code with lower FER union bound. However, when we initialized our algorithm with V-BLAST and BPSK, the algorithm converged to

$$\boldsymbol{A}_1 = \begin{bmatrix} 0.74342773 & -0.08692446 \\ 0.40936511 & 0.50797666 \end{bmatrix},$$
$$\boldsymbol{A}_2 = \begin{bmatrix} -0.36902827 & 0.54303939 \\ 0.75611440 & -0.00025825 \end{bmatrix},$$
$$\boldsymbol{A}_3 = \begin{bmatrix} 0.30229420 & 0.83952155 \\ -0.45246174 & 0.09078508 \end{bmatrix},$$
$$\boldsymbol{A}_4 = \begin{bmatrix} -0.47303478 & -0.03771225 \\ -0.21804135 & 0.85410106 \end{bmatrix},$$

and $\boldsymbol{B}_i = \boldsymbol{0}_{2\times 2}, i = 1, ..., 4$ (because the signals are real). At FER = 10^{-3}, the CSNR gain in using the above code is 4.2 dB over V-BLAST and 1.2 dB over the LD code of [13, eq. (31)]. Alamouti's code, however, demonstrates a 1.4 dB CSNR gain over our code above.

E. Bit-to-Signal Mapping and Trellis Coding

The system considered here has one transmit and two receive antennas. The frame length is 120 bits and the test is repeated 200000 times. A frame error is counted when the decoded and transmitted symbol streams do not exactly match.

Figure 10 demonstrates the performance of a rate-$3/4$ 8-state 16-QAM trellis coded system. We optimize the bit-to-signal mapping for a fixed convolutional encoder structure using the symbol PEP formula in (25) and the guidelines of [26], which result in the M1 mapping of [26, Figure 8]. We compare the M1 and Gray mapped systems for the same

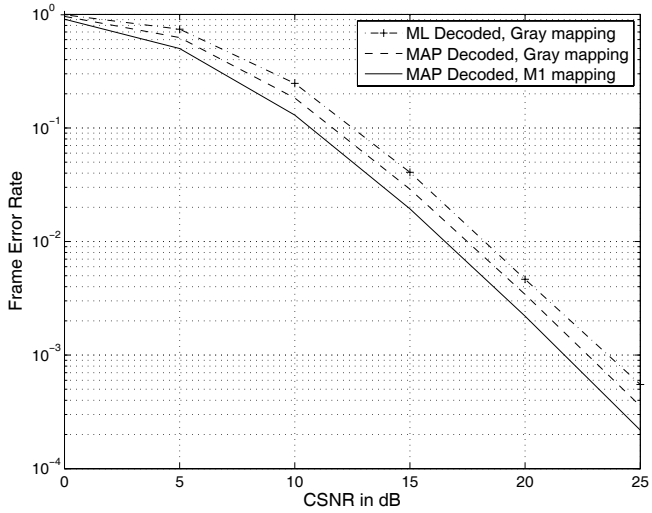

Fig. 10. Comparison between M1 and Gray mappings for the rate 3/4 8-state 16-QAM trellis coded system with $L = 2$ receive antennas. The input is an i.i.d. bit-stream with $p_0 = 0.9$.

encoder structure specified by $(h_0, h_1, h_2) = (11, 2, 4)$. It is observed that, since the M1 mapping is more energy-efficient, it achieves a 1 dB CSNR gain over Gray mapping with MAP decoding at FER = 10^{-3}, and an additional 0.8 dB CSNR gain over ML decoding; i.e., 1.8 dB gain in CSNR over the conventional which use Gray mapping and ML decoding.

VII. Conclusion

In this paper, we addressed the maximum *a posteriori* decoding of non-uniform i.i.d. sources in a multiple-antenna setting. We derived closed-form expressions for the codeword pairwise error probability of general multi-antenna codes as well as the PEP of symbols undergoing space-time orthogonal block coding. It was shown that, similar to ML decoding, detection of symbols is decoupled under MAP decoding. We also explored some applications of the PEP formulas. For example, we proved that the binary antipodal signaling scheme which minimizes the bit error rate in AWGN channels also minimizes the BER under STOB coding. Moreover, we designed space-time LD codes which were optimized for the source distribution. Two typical codes were given which outperformed V-BLAST and the LD codes of [13] by a wide margin. We also addressed the issue of bit-to-signal mapping in STOB coded scenarios as well as trellis-coded MRC systems. Another application is the establishment of tight Bonferroni-type bounds for the SER and BER of MIMO systems which employ MAP decoding (see [8]).

Extensions of this work may include computation of the exact SER and BER of STOB coded channels under MAP decoding following the approach of [25], [24] (which is implemented for the AWGN channel) using the PEP formulas in [8] and optimization of bit-to-signal mapping for STOB coded channels. Optimal quaternary constellation design could also be studied as in [21] for STOB coded channels using the error bounds of [8]. Another direction is to use (12) to find trellis encoders and signal mappings which, when used with the MAP decoding rule in (2), will reduce the FER and/or BER of a trellis-coded system. A possible approach is to approximate the union upper bound, similar to the work in [17] for AWGN channels. The study of (12) to identify key parameters and to derive design criteria for space-time codes could be another extension of this work. It would also be interesting to find the codeword PEP for the case where the source has memory in addition to non-uniformity.

Appendix

Here we derive the n^{th} derivative of (11) with respect to s. First, we present a lemma which can be proved easily via induction.

Lemma– The following hold

a) $\frac{d^n}{ds^n}\left(\frac{1}{s}\right) = \frac{(-1)^n n!}{s^{n+1}}$

b) $\frac{d^n}{ds^n}(2s+1)^{\frac{p}{2}} = (2s+1)^{\frac{p}{2}-n}\prod_{i=0}^{n-1}(p-2i)$

c) $\frac{d^n}{dy^n}e^{-(a+by)} = (-1)^n b^n e^{-(a+by)}$.

The derivative of the first term in (11) may be found via (a). As for the product term, we use the Leibniz's formula [1, Eq. 3.3.8] to treat the two terms separately. The formula is

$$\frac{d^n}{dx^n}uv = \sum_{i=0}^{n}\binom{n}{i}\frac{d^i u}{dx^i}\frac{d^{n-i}v}{dx^{n-i}}. \tag{29}$$

We apply (29) with $u = \frac{a}{s} + \frac{1}{s\sqrt{2s+1}}$ and $v = e^{-\left(a+b\sqrt{2s+1}\right)}$. Using (29) again to find the i^{th} derivative of the second term in u with $u_1 = \frac{1}{s}$ and $v_1 = \frac{1}{\sqrt{2s+1}}$, and applying (a) and (b) with $p = -1$ results in

$$\begin{aligned}
&\frac{d^i}{ds^i}\left(\frac{a}{s} + \frac{1}{s\sqrt{2s+1}}\right) \\
&\quad = \frac{(-1)^i i! a}{s^{i+1}} + \sum_{j=0}^{i}\frac{(-1)^i i!}{(i-j)!}\frac{\prod_{l=1}^{i-j}(2l-1)}{s^{j+1}(2s+1)^{i-j+\frac{1}{2}}} \\
&\quad = \frac{(-1)^i i! a}{s^{i+1}} + \frac{(-1)^i i!}{s^{i+1}}\sum_{k=0}^{i}\binom{2k}{k}\frac{s^k}{2^k(2s+1)^{k+\frac{1}{2}}}.
\end{aligned} \tag{30}$$

As for the exponential term, we use a result from [12] which states that for $f(x) = F(y)$, $y = \phi(x)$, we have

$$\begin{aligned}
\frac{d^n}{dx^n}f(x) &= \sum_{i=1}^{n}\frac{U_i}{i!}F^{(i)}(y), \qquad \text{where} \\
U_i &= \sum_{k=0}^{i-1}\binom{i}{k}(-1)^k y^k \frac{d^n}{dx^n}y^{i-k}.
\end{aligned} \tag{31}$$

Letting $F(y) = e^{-(a+by)}$ and $y = \sqrt{2s+1}$, we use (31), (b), and (c) to get

$$\begin{aligned}
&\frac{d^{n-i}}{ds^{n-i}}e^{-\left(a+b\sqrt{2s+1}\right)} = e^{-\left(a+b\sqrt{2s+1}\right)} \\
&\quad \sum_{j=1}^{n-i}\frac{(2s+1)^{\frac{j}{2}-n+i}b^j}{j!}\sum_{k=0}^{j-1}\binom{j}{k}(-1)^{j+k} \\
&\quad \prod_{l=0}^{n-i-1}(j-k-2l).
\end{aligned} \tag{32}$$

Using (30) and (32) in (29) with $a = \lambda_{ij}$ and $b = |\lambda_{ij}|$ yields the n^{th} derivative of (11) as shown in (33).

$$\frac{d^n}{ds^n}F_{\text{MAP}}(s) = \frac{(1-\text{sgn}(\lambda_{ij}))(-1)^n n!}{2s^{n+1}} - \frac{1}{2}e^{-\left(\lambda_{ij}+|\lambda_{ij}|\sqrt{2s+1}\right)} \times \sum_{k=0}^{n}\binom{n}{k}\frac{(-1)^k k!}{s^{k+1}}\left(-\text{sgn}(\lambda_{ij}) + \sum_{m=0}^{k}\binom{2m}{m}\frac{s^m}{2^m(2s+1)^{m+\frac{1}{2}}}\right) \times \sum_{l=1}^{n-k}\frac{|\lambda_{ij}|^l}{l!(2s+1)^{n-k-\frac{l}{2}}}\sum_{p=0}^{l-1}\binom{l}{p}(-1)^{l+p}\prod_{q=0}^{n-k-1}(l-p-2q). \quad (33)$$

REFERENCES

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York, NY: Dover Publications, fifth ed., 1966.

[2] F. Alajaji, N. Phamdo, N. Farvardin, and T. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inform. Theory,* vol. 42, pp. 230-239, Jan. 1996.

[3] F. Alajaji, N. Phamdo, and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325-336, Sept. 1996.

[4] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

[5] S. Al-Semari, F. Alajaji, and T. Fuja, "Sequence MAP decoding of trellis codes for Gaussian and Rayleigh channels," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1130-1140, July 1999.

[6] G. Bauch, J. Hagenauer, and N. Seshadri, "Turbo processing in transmit antenna diversity systems," *Annals Telecommun.*, vol. 56, pp. 455-471, Aug. 2001.

[7] F. Behnamfar, F. Alajaji, and T. Linder, "Tight error bounds for space-time orthogonal block codes under slow Rayleigh flat fading," *IEEE Trans. Commun.*, vol. 53, pp. 952-956, June 2005.

[8] F. Behnamfar, F. Alajaji, and T. Linder, "Tight bounds on the SER and BER of space-time orthogonal block coded channels under MAP decoding," in *Proc. International Symposium on Information Theory and its Applications (ISITA'04)*, Parma, Italy, Oct. 2004, pp. 874-879.

[9] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, 1991.

[10] J. W. Craig, "A new, simple, and exact result for calculating the probability of error for two-dimensional signal constellations," in *Proc. IEEE Military Commun. Conf.*, McLean, VA, Nov. 1991, pp. 571-575.

[11] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Electron. Lett.*, vol. 35, pp. 14-16, Jan. 1999.

[12] I. Gradstein and I. Ryshik, *Tables of Integrals, Series, and Products*. San Diego, CA: Academic Press, fifth ed., 1994.

[13] B. Hassibi and B. Hochwald, "High rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1804-1824, July 2002.

[14] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. New York: Kluwer Academic Publishers, 1994.

[15] R. Johannesson and K. Zigangirov, *Fundamentals of Convolutional Coding.* New York, NY: IEEE Press, 1999.

[16] I. Korn, J. Fonseka, and S. Xing, "Optimal binary communication with nonequal probabilities," *IEEE Trans. Commun.*, vol. 51, pp. 1435-1438, Sept. 2003.

[17] J. Kroll and N. Phamdo, "Analysis and design of trellis codes optimized for a binary symmetric Markov source with MAP detection," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2977-2987, Nov. 1998.

[18] J. Lim and D. L. Neuhoff, "Joint and tandem source-channel coding with complexity and delay constraints," *IEEE Trans. Commun.*, vol. 51, pp. 757-766, May 2003.

[19] H. Lu, Y. Wang, P. V. Kumar, and K. Chugg, "Remarks on space-time codes including a new lower bound and an improved code," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2752-2757, Oct. 2003.

[20] A. Naguib, N. Seshadri, and A. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Mag.*, vol. 46, pp. 76-92, May 2000.

[21] H. Nguyen and T. Nechiporenko, "Quaternary signal sets for digital communications with nonuniform sources," *IEICE Trans. Fundamentals*, vol. E89-A, pp. 832-835, Mar. 2006.

[22] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

[23] M. K. Simon, "Evaluation of average bit error probability for space-time coding based on a simpler exact evaluation of pairwise error probability," *J. Commun. Networks*, vol. 3, pp. 257-264, Sept. 2001.

[24] L. Szczecinski, C. Gonzalez, and S. Aissa, "Exact expression for the BER of rectangular QAM with arbitrary constellation mapping," *IEEE Trans. Commun.*, vol. 54, pp. 389-392, Mar. 2006.

[25] L. Szczecinski, S. Aissa, C. Gonzalez, and M. Bacic, "Exact evaluation of bit and symbol error rates for arbitrary two-dimensional modulation and non-uniform signaling in AWGN channel", *IEEE Trans. Commun.*, vol. 54, pp. 1049-1056, June 2006.

[26] G. Takahara, F. Alajaji, N. C. Beaulieu, and H. Kuai, "Constellation mappings for two-dimensional signaling of non-uniform sources," *IEEE Trans. Commun.*, vol. 51, pp. 400-408, Mar. 2003.

[27] G. Taricco and E. Biglieri, "Exact pairwise error probability of space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, pp. 510-513, Feb. 2002.

[28] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.

[29] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, Mar. 1998.

[30] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55-67, Jan. 1982.

[31] M. Uysal and C. N. Georghiades, "Error performance analysis of space-time codes over Rayleigh fading channels," *J. Commun. Networks*, vol. 2, pp. 351-356, Dec. 2000.

[32] K. Zeger, J. Vaisey, and A. Gersho, "Globally optimal vector quantizer design by stochastic relaxation," *IEEE Trans. Signal Processing*, vol. 40, pp. 310-322, Feb. 1992.

[33] Y. Zhong, F. Alajaji, and L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1450-1468, Apr 2006.

[34] G.-C. Zhu, F. Alajaji, J. Bajcsy, and P. Mitran, "Transmission of non-uniform memoryless sources via non-systematic turbo codes," *IEEE Trans. Commun.*, vol. 52, pp. 1344-1354, Aug. 2004.

Firouz Behnamfar received the B.Sc. and M.Sc. degrees from Isfahan University of Technology, Isfahan, Iran, and the Ph.D. degree from Queens University at Kingston, ON, Canada, all in electrical engineering, in 1994, 1997, and 2004, respectively. From September 2004 to August 2006, he was a Postdoctoral Fellow and an Adjunct Assistant Professor in the Department of Mathematics and Statistics, Queens University. He is now with Nortel Networks, Ottawa, Canada, where he is involved in system design for 4G wireless networks such as WiMAX and LTE. His professional interests include communication systems, wireless communications, and information theory.

Dr. Behnamfar has won a number of awards including the Outstanding (rank 1) Thesis in Applied Science Award from Queens University in 2004, a postdoctoral fellowship (PDF) award, and an industrial research and development fellowship (IRDF) award, both by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Fady Alajaji (S'90–M'94–SM'00) was born in Beirut, Lebanon, on May 1, 1966. He received the B.E. degree with distinction from the American University of Beirut, Lebanon, and the M.Sc. and Ph.D. degrees from the University of Maryland, College Park, all in electrical engineering, in 1988, 1990 and 1994, respectively. He held a postdoctoral appointment in 1994 at the Institute for Systems Research, University of Maryland.

In 1995, he joined the Department of Mathematics and Statistics at Queen's University, Kingston, Ontario, where he is currently a Professor of Mathematics and Engineering. Since 1997, he has also been cross-appointed in the Department of Electrical and Computer Engineering at the same university. His research interests include information theory, joint source-channel coding, error control coding, data compression and digital communications.

Dr. Alajaji currently serves as Editor for Source and Source-Channel Coding for the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as co-chair of the 1999 Canadian Workshop on Information Theory, as co-chair of the Technical Program Committee (TPC) of the 2004 Biennial Symposium on Communications and as a TPC member of several international conferences and workshops. He received the Premier's Research Excellence Award from the Province of Ontario.

Tamás Linder (S'92-M'93-SM'00) was born in Budapest, Hungary, in 1964. He received the M.S. degree in electrical engineering from the Technical University of Budapest in 1988, and the Ph.D degree in electrical engineering from the Hungarian Academy of Sciences in 1992.

He was a post-doctoral researcher at the University of Hawaii in 1992 and a Visiting Fulbright Scholar at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign during 1993-1994. From 1994 to 1998 he was a faculty member in the Department of Computer Science and Information Theory at the Technical University of Budapest. From 1996 to 1998 he was also a visiting research scholar in the Department of Electrical and Computer Engineering, University of California, San Diego. In 1998 he joined Queen's University where he is now a Professor of Mathematics and Engineering in the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition.

Dr. Linder received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellor's Research Award of Queen's University in 2003. He was an Associate Editor for Source Coding of the IEEE TRANSACTIONS ON INFORMATION THEORY in 2003–2004.