Training-Codes for the Noncoherent Multi-Antenna Block-Rayleigh-Fading Channel

Matthias Brehler and Mahesh K. Varanasi

e-mail: {brehler, varanasi}@dsp.colorado.edu
Electrical & Computer Engineering Department
University of Colorado, Boulder, CO 80309

Abstract — **We consider signal design for the noncoherent block-Rayleigh-fading channel, in which neither the transmitter nor the receiver know the fading coefficients. We show that several recently proposed signal designs for this channel can be interpreted as training-based, i.e., part of the coherence interval is used to transmit known symbols to learn the channel and in the remainder a "coherent" information symbol is transmitted. At the receiver, the estimated channel is then used as if correct, which allows to employ efficient implementations (like the sphere-decoder) for the coherent decision rule. We establish conditions for which such a strategy is equivalent to the optimal noncoherent decision rule. Motivated by the good performance of the training schemes even when the coherence time is very small (indeed minimal), we propose the paradigm of training-codes to build signal-constellations for the noncoherent block-Rayleigh-fading channel. Training-codes leverage the advances made in coherent codes and thus can often address the signal-design and efficient-decoding problems simultaneously.**

I. Introduction

THE INFORMATION theoretic work [1,2] on the capacity of the noncoherent multi-antenna block Rayleigh fading channel has spurred several proposals for signaling schemes for this channel [3–6].[1] The detection of all of the proposed schemes suffers from an exponential complexity in the rate of the transmission, because the optimum receiver has to test all signals [3, 13]. Recently however, Tarokh in [14] proposed the so-called generalized noncoherent PSK constellations and noncoherent orthogonal designs that allow both for maximum-likelihood detection with complexity *independent* of the rate. [15] also makes use of the (generalized) orthogonal designs [16–19] to generate noncoherent constellations that enjoy low-complexity maximum-likelihood decoding. In this paper we show that that the schemes of both, [14] and [15], can be more efficiently described as training schemes, which simplifies their generalization and allows for an exact error rate analysis. (The generalized noncoherent PSK constellations of [14] do not fit the training description. We considerably improve upon the original noncoherent PSK constellation in [20].) Consequently, we generalize the training-designs in two ways and considerably improve upon [14, 15]. First, we allow for other, more powerful, codes than the generalized orthogonal designs for the information symbol. Second, we consider not only one but several information symbols and find the optimum energy-allocation between training and data phase. Even though the suggested estimator-detector receiver is not optimal in general, we find "noncoherent" codes, i.e., training-codes, with spectral efficiencies and error rate performances close to the benchmark designs of [6], which we adapt for bit error rate (BER) rather than symbol error rate (SER) as performance measure.

This work was supported in part by NSF Grant CCR-0112977 and by ARO Grant DAAD 19-99-1-029.

[1] We do not consider the differentially coherent schemes of [7–12], because the underlying assumptions on the channel model are incompatible.

II. System Model and Noncoherent Receivers

The description of the N_{T}-transmit, N_{R}-receive antenna system with fading that is constant over D dimensions follows [13]. We write the DN_{R}-length vector of sufficient statistics as

$$\mathbf{y} = \sqrt{\overline{\gamma}}\boldsymbol{\mathcal{S}}_m\mathbf{h} + \boldsymbol{\eta}, \tag{1}$$

where

- $\overline{\gamma} = \frac{E_S}{N_0}$, where E_S is the average transmitted symbol energy,[2]
- $\boldsymbol{\mathcal{S}}_m = \mathbf{I}_{N_{\mathrm{R}}} \otimes \mathbf{S}_m$ is the mth expanded symbol ($1 \leq m \leq M$) and $\mathbf{S}_m$ is $D \times N_{\mathrm{T}}$, normalized such that $M^{-1}\sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{S}_m^\dagger \mathbf{S}_m\right) = 1$,
- $\mathbf{h}$ contains the $N_{\mathrm{T}}N_{\mathrm{R}}$ fading coefficients, which are $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ distributed and normalized such that $\mathrm{E}\left[M^{-1}\sum_{m=1}^{M} \mathbf{h}^\dagger \boldsymbol{\mathcal{S}}_m^\dagger \boldsymbol{\mathcal{S}}_m \mathbf{h}\right] = N_{\mathrm{R}}$, and
- $\boldsymbol{\eta}$ is the additive white Gaussian noise ($\mathcal{CN}(\mathbf{0}, \mathbf{I})$ distributed).

This model is often written as $\mathbf{Y} = \sqrt{\overline{\gamma}}\mathbf{S}_m\mathbf{H} + \mathbf{N}$ (or a transpose thereof), where $\mathbf{y} = \mathrm{vec}\{\mathbf{Y}\}$, $\mathbf{h} = \mathrm{vec}\{\mathbf{H}\}$, and $\boldsymbol{\eta} = \mathrm{vec}\{\mathbf{N}\}$, which is of course equivalent.

Defining

$$\mathbf{F}_m = -\boldsymbol{\mathcal{S}}_m\left(\boldsymbol{\mathcal{S}}_m^\dagger\boldsymbol{\mathcal{S}}_m + \overline{\gamma}^{-1}\boldsymbol{\Sigma}^{-1}\right)^{-1}\boldsymbol{\mathcal{S}}_m^\dagger \text{ and} \tag{2}$$

$$c_m = \ln\left|\overline{\gamma}\boldsymbol{\mathcal{S}}_m^\dagger\boldsymbol{\mathcal{S}}_m + \boldsymbol{\Sigma}^{-1}\right|, \tag{3}$$

the optimum noncoherent receiver for $\mathbf{S}_m$ is found in [13] as

$$\Phi^{\mathrm{ML}} : \widehat{m} = \arg\min_{1\leq m\leq M} \mathbf{y}^\dagger\mathbf{F}_m\mathbf{y} + c_m. \tag{4}$$

[2] Relative to [6, 13] we change our normalization of $\mathbf{S}$ and $\mathbf{h}$ slightly here to make it more obvious that $\overline{\gamma} = E_S/N_0$. However, in [6, 13] we call $\overline{\gamma}$ the SNR, an unfortunate nomenclature, because the received SNR is often defined as $\mathsf{SNR}^{\mathrm{r}} = \frac{E_S}{N_0 D} = \frac{\overline{\gamma}}{D}$. Thus we avoid the term SNR for $\overline{\gamma}$ here and reserve it for the received SNR ($\mathsf{SNR}^{\mathrm{r}}$).

Since the maximum likelihood receiver $\Phi^{\rm ML}$ requires the knowledge of the fading covariance $\mathbf{\Sigma}$ and $\overline{\gamma}$, which may not always be available at the receiver, [13] also defines a generalized likelihood ratio test (GLRT) for $\mathbf{S}_m$, which does not require this information:

$$\begin{aligned}\Phi^{\rm G}:\widehat{m} &= \arg\min_{1\le m\le M} -\mathbf{y}^\dagger\boldsymbol{\mathcal{S}}_m\left(\boldsymbol{\mathcal{S}}_m^\dagger\boldsymbol{\mathcal{S}}_m\right)^{-1}\boldsymbol{\mathcal{S}}_m^\dagger\mathbf{y}\\ &= \arg\min_{1\le m\le M} -\sum_{n=1}^{N_{\rm R}}\mathbf{y}_n^\dagger\mathbf{S}_m\left(\mathbf{S}_m^\dagger\mathbf{S}_m\right)^{-1}\mathbf{S}_m^\dagger\mathbf{y}_n,\end{aligned}$$

where $\mathbf{y}_n$ are the D elements of $\mathbf{y}$ that correspond to receive antenna n. For independent and identically distributed (i.i.d.) fading and unitary signaling (the $N_{\rm T}$ columns of all $\mathbf{S}_m$ are orthogonal) both receivers are equivalent to each other and to the receiver of [3], derived under these assumptions.

III. Training-Codes

Coherent reception is premised on the knowledge of the fading coefficients. In practice, these have to be estimated and so the transmission is split into a training phase and a communications phase [21]. In the latter, the estimated fading coefficients are often used as if correct for a "coherent" modulation scheme. Here we take the point of view that the training symbol together with the information bearing one can jointly be viewed as a symbol of a noncoherent modulation scheme. For instance, the noncoherent orthogonal designs of [14] and the designs of [15] can be more simply interpreted as a training based scheme, as we will show below. More importantly, our viewpoint also allows us to specify new noncoherent space–time codes (and decoding algorithms) by leveraging more powerful existing coherent codes (and decoders) when compared to the orthogonal designs [16–19] used in [14, 15].

To estimate the fading, the symbol $\mathbf{S}_m$ is split into two parts, one part known to the receiver and used as training data and one part that transmits the information:

$$\mathbf{S}_m = \frac{1}{\sqrt{S}}\begin{bmatrix}\sqrt{S\tau}\mathbf{T}\\ \sqrt{1-\tau}\mathbf{B}_{m_1}\\ \vdots\\ \sqrt{1-\tau}\mathbf{B}_{m_S}\end{bmatrix}, \qquad (5)$$

where without loss of generality we let m determine uniquely the S-tuple $(m_1, m_2, \ldots, m_S)$ according to

$$m = \sum_{s=1}^{S}(m_s-1)M_{\rm D}^{s-1}+1$$

and $M = M_{\rm D}^S$. The training symbol $\mathbf{T}$ is $D_{\rm T}\times N_{\rm T}$ and the S data symbols $\left\{\mathbf{B}_{m_s}\right\}_{m_s=1}^{M_{\rm D}}$ are $D_{\rm D}\times N_{\rm T}$, so that $D = D_{\rm T}+SD_{\rm D}$. The fraction of energy of each information symbol that is spent on training is τ ($0\le\tau\le 1$). Similarly to $\boldsymbol{\mathcal{S}}_m = \mathbf{I}_{N_{\rm R}}\otimes\mathbf{S}_m$ we define the expanded training and data symbols $\boldsymbol{\mathcal{T}} = \mathbf{I}_{N_{\rm R}}\otimes\mathbf{T}$ and $\boldsymbol{\mathcal{B}}_m = \mathbf{I}_{N_{\rm R}}\otimes\mathbf{B}_m$. To facilitate the presentation in this section we rewrite the model in terms of sufficient statistics corresponding to the training and data phase, respectively:

$$\mathbf{y}_{\rm T} = \sqrt{\overline{\gamma}\tau}\boldsymbol{\mathcal{T}}\ \mathbf{h}+\boldsymbol{\eta}_{\rm T}, \qquad (6)$$

$$\mathbf{y}_s = \sqrt{\overline{\gamma}\frac{1-\tau}{S}}\boldsymbol{\mathcal{B}}_{m_s}\mathbf{h}+\boldsymbol{\eta}_s, \qquad (7)$$

The optimum or the GLRT receiver continue to be given by (4) and (5) with $\mathbf{S}_m$ given by (5). However, here we are interested in a receiver that estimates the fading first and then uses the estimate as if it were true for the "coherent" reception of each individual data symbol. We consider the minimum mean-square error (MMSE) estimate, because for some key special cases this receiver will also be optimal. The MMSE estimate and the coherent receiver are given by

$$\begin{aligned}\Phi^{\rm E-C}:\ \widehat{\mathbf{h}} &= \sqrt{\overline{\gamma}\tau}\mathbf{\Sigma}\boldsymbol{\mathcal{T}}^\dagger\left(\overline{\gamma}\tau\boldsymbol{\mathcal{T}}\mathbf{\Sigma}\boldsymbol{\mathcal{T}}^\dagger+\mathbf{I}\right)^{-1}\mathbf{y}_{\rm T}, \qquad (8)\\ \widehat{m}_s &= \arg\min_{1\le m_s\le M_{\rm D}}\left\|\mathbf{y}_s-\sqrt{\overline{\gamma}\frac{1-\tau}{S}}\boldsymbol{\mathcal{B}}_{m_s}\widehat{\mathbf{h}}\right\|.\end{aligned}$$

This receiver is in general sub-optimal, because the noise enhanced by the channel estimation error is in general not white and we decide for each data symbol individually, neglecting the correlation in the noise term. In detail, consider the equivalent channel model

$$\mathbf{y}_s = \sqrt{\overline{\gamma}\frac{1-\tau}{S}}\boldsymbol{\mathcal{B}}_{m_s}\widehat{\mathbf{h}}+\widetilde{\boldsymbol{\eta}}_s, \qquad (9)$$

where

$$\widetilde{\boldsymbol{\eta}}_s = \sqrt{\overline{\gamma}\frac{1-\tau}{S}}\boldsymbol{\mathcal{B}}_{m_s}\left(\mathbf{h}-\widehat{\mathbf{h}}\right)+\boldsymbol{\eta}_s. \qquad (10)$$

The channel estimate $\widehat{\mathbf{h}}$ is $\mathcal{CN}\left(\mathbf{0},\mathbf{K}_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}}\right)$ distributed, where

$$\begin{aligned}\mathbf{K}_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}} &= E\left[\widehat{\mathbf{h}}\widehat{\mathbf{h}}^\dagger\right] \qquad (11)\\ &= \mathbf{\Sigma}\tau\boldsymbol{\mathcal{T}}^\dagger\left(\tau\boldsymbol{\mathcal{T}}\mathbf{\Sigma}\boldsymbol{\mathcal{T}}^\dagger+\overline{\gamma}^{-1}\mathbf{I}\right)^{-1}\boldsymbol{\mathcal{T}}\mathbf{\Sigma}\\ &= \mathbf{\Sigma}\left(\mathbf{\Sigma}+\left(\overline{\gamma}\tau\boldsymbol{\mathcal{T}}^\dagger\boldsymbol{\mathcal{T}}\right)^{-1}\right)^{-1}\mathbf{\Sigma}\end{aligned}$$

The noise term $\widetilde{\boldsymbol{\eta}}_s$ of the equivalent channel model has covariance

$$\mathbf{K}_{\widetilde{\boldsymbol{\eta}}_s\widetilde{\boldsymbol{\eta}}_s} = \frac{1-\tau}{S}\boldsymbol{\mathcal{B}}_{m_s}\left(\overline{\gamma}^{-1}\mathbf{\Sigma}^{-1}+\tau\boldsymbol{\mathcal{T}}^\dagger\boldsymbol{\mathcal{T}}\right)^{-1}\boldsymbol{\mathcal{B}}_{m_s}^\dagger+\mathbf{I},$$

which is in general not white (i.e. not a scaled identity) and may depend on $\boldsymbol{\mathcal{B}}_{m_s}$. Note that due to the MMSE estimate $\widetilde{\boldsymbol{\eta}}_s$ is zero-mean (conditioned on $\mathbf{y}_{\rm T}$). Consequently, only for the special case of $S=1$, i.i.d. fading ($\mathbf{\Sigma}=\mathbf{I}$), $N_{\rm T}\mathbf{T}^\dagger\mathbf{T}=\mathbf{I}$, and $N_{\rm T}\mathbf{B}_{m_1}\mathbf{B}_{m_1}^\dagger=\mathbf{I}\ \forall m_1$, the receiver is optimal (maximum-likelihood). For more than one information symbol ($S>1$), the receiver is clearly suboptimal, even if all other conditions are met, because the noise enhancement through the channel estimation error is treated as if independent. Since i.i.d. fading has received by far the most attention in the literature, we restrict ourselves to this case and next discuss possible choices for the training symbol $\mathbf{T}$, the data symbols $\left\{\mathbf{B}_m\right\}_{m=1}^{M_{\rm D}}$, and τ.

It is shown in [22] that for i.i.d. fading a training symbol $\mathbf{T}$ with orthonormal columns minimizes the total expected estimation error $E\left[\left(\mathbf{h}-\widehat{\mathbf{h}}\right)^{\dagger}\left(\mathbf{h}-\widehat{\mathbf{h}}\right)\right]$. Since conveniently this is also the choice that is necessary to make $\Phi^{\mathrm{E-C}}$ optimal, we assume $N_{\mathrm{T}}\mathbf{T}^{\dagger}\mathbf{T}=\mathbf{I}$ in the following discussion (in addition to i.i.d. fading).

As for the choice of the data symbols $\left\{\mathbf{B}_{m_s}\right\}_{m_s=1}^{M_{\mathrm{D}}}$, for one transmit antenna and one information dimension ($N_{\mathrm{T}}=1$, $S=D_{\mathrm{D}}=1$) the usual PSK constellation fulfills of course $\mathbf{B}_m^{\dagger}\mathbf{B}_m=1$ and thus the usual coherent decision rule is optimum if the MMSE channel estimate is used in place of the true channel.[3] For $S=1$ and constellations consisting of unitary matrices (for $N_{\mathrm{T}}>1$), possibly drawn from (finite) groups [11] or generated through a Cayley transform [12], the detector-estimator receiver $\Phi^{\mathrm{E-C}}$ is optimal (provided $\mathbf{\Sigma}=\mathbf{I}$ and $N_{\mathrm{T}}\mathbf{T}^{\dagger}\mathbf{T}=\mathbf{I}$). The same is true for the space–time block codes based on the *square* orthogonal designs when used with training if the underlying constellation is PSK. Interestingly, the orthogonal designs with training when viewed as noncoherent codes are precisely the so-called noncoherent orthogonal designs proposed in [14] (with $\mathbf{T}$ corresponding to the identity matrix) and also correspond to the designs of [15] (with the training and the information interleaved, but a simple permutation renders the designs equivalent to (5)). Note however, that without the training interpretation, the proofs of optimality in [14] and [15] are not particularly insightful and somewhat cumbersome. Moreover, the orthogonal designs with training clearly permit low-complexity (estimator-detector) decoding since the detector is the same as in the coherent case–a fact that is stated in [14] but without this insight.[4] Even before [14], the use of good coherent unitary constellations combined with an identity block–to be viewed as training–has been suggested as a noncoherent constellation in [9, Section V.B]. However, this observation was subsequently not used to generate specific constellations in [9] and appears to have been mostly overlooked. Moreover, the benefit of simplified decoding when square orthogonal designs are employed, although implicitly contained in [9], was only brought to attention in [14]–but without the insight gained from the training interpretation. Here we present the training interpretation rigorously, which allows us to exactly state the performance degradation for the above mentioned conditions. In addition we present performance comparisons between training-codes and truly noncoherent designs below.

Unfortunately, there are only three square, "full-rate" orthogonal designs and only the 2×2 Alamouti scheme [16] can be used with complex symbols (i.e. PSK symbols, for optimum decoding). The 4×4 and 8×8 are real designs only and thus can be used with just BPSK, if $\Phi^{\mathrm{E-C}}$ were to be optimal.[5] In addition to these full-rate designs, there is a generalized complex square design for four transmit antennas that can be used to transmit three symbols in four dimensions (see [18, 19] for an easier representation of this rate $3/4$ code than the one given in [17]). Note that this complex 4×4 design with training was not identified in [14] as a noncoherent orthogonal design (it is however used in [15]). Finally, $\Phi^{\mathrm{E-C}}$ would be optimal for the "non-square" codes of [17] only if they are used in transposed form (for example, the code for $N_{\mathrm{T}}=3$ and $D=4$ would be used for $N_{\mathrm{T}}=4$ and $D=3$). However, they do not lend themselves to simplified decoding nor do they guarantee full transmit diversity (but do give a diversity order of three, in the example).

For the case that the noise terms $\widetilde{\eta}_s$ are all white, it is not hard to find the optimum energy allocation τ and perform an exact performance analysis relative to the coherent reception of $\frac{1}{\sqrt{S}}\mathbf{B}_{m_s}$. The noise covariances become

$$\mathbf{K}_{\widetilde{\boldsymbol{\eta}}_s\widetilde{\boldsymbol{\eta}}_s}=\left(\frac{1-\tau}{N_{\mathrm{T}}S\overline{\gamma}^{-1}+S\tau}+1\right)\mathbf{I}$$

and the covariance of the channel estimate is

$$\mathbf{K}_{\widehat{\mathbf{h}}\widehat{\mathbf{h}}}=\frac{\overline{\gamma}\tau}{\overline{\gamma}\tau+N_{\mathrm{T}}}\mathbf{I}.$$

Thus the SNR (equivalently E_S/N_0 or E_b/N_0) degradation of training-aided noncoherent detection versus the coherent detection of $\frac{1}{\sqrt{S}}\mathbf{B}_{m_s}$ is

$$\xi=\frac{\overline{\gamma}\tau(1-\tau)}{\overline{\gamma}\tau+N_{\mathrm{T}}}\left(\frac{1-\tau}{N_{\mathrm{T}}S\overline{\gamma}^{-1}+S\tau}+1\right)^{-1}. \qquad (12)$$

Maximizing ξ over τ leads to

$$\widehat{\tau}_{\overline{\gamma}}=\begin{cases}\dfrac{\sqrt{S\left(1+\frac{N_{\mathrm{T}}S}{\overline{\gamma}}\right)\left(1+\frac{N_{\mathrm{T}}}{\overline{\gamma}}\right)}-1-\frac{N_{\mathrm{T}}S}{\overline{\gamma}}}{S-1} & \text{for } S>1,\\[2ex] \dfrac{1}{2} & \text{for } S=1,\end{cases} \qquad (13)$$

which depends on $\overline{\gamma}$ for $S>1$ and thus the transmitter would have to adapt the energy distribution between training and communicating to the power level at the receiver. We propose to simplify the scheme by fixing the asymptotically (for $\overline{\gamma}\to\infty$) optimal $\widehat{\tau}_{\infty}$ for all values of $\overline{\gamma}$. With this choice the SNR penalty ξ is easily obtained and the equivalent E_S/N_0 (for the detection of one coherent symbol $\frac{1}{\sqrt{S}}\mathbf{B}_{m_s}$) becomes

$$\left(\frac{E_S}{N_0}\right)_{\mathrm{eq.coherent}}=\begin{cases}\dfrac{\overline{\gamma}S}{1+\frac{N_{\mathrm{T}}\sqrt{S}}{\overline{\gamma}}}\left(\dfrac{\sqrt{S}-1}{S-1}\right)^2 & \text{for } S>1,\\[2ex] \dfrac{\overline{\gamma}S}{4}\dfrac{1}{1+\frac{N_{\mathrm{T}}S}{\overline{\gamma}}} & \text{for } S=1.\end{cases} \qquad (14)$$

[3] Note that there are fast algorithms that permit optimum decoding of PSK for $S>1$ [23, 24].

[4] Note that although the description of $\Phi^{\mathrm{E-C}}$ in (8) requires the knowledge of $\overline{\gamma}$ at the receiver, an actual implementation for $N_{\mathrm{T}}\mathbf{B}_m^{\dagger}\mathbf{B}_m=\mathbf{I}$ does not, as is easily seen when expanding the norm that is minimized. In other words, we can simplify $\widehat{\mathbf{h}}$ to $\boldsymbol{\mathcal{T}}^{\dagger}\mathbf{y}_{\mathrm{T}}$ without losing optimality.

[5] The Hamiltonian constellations, introduced in a differentially coherent context in [9, Section V.C], lift the constraint to the PSK/BPSK constellations at the price that the information symbols cannot be chosen independently anymore.

Thus the schemes of [14] and [15] can be exactly analyzed, because the error rate performance of coherent detection of orthogonal designs is essentially the same as that of maximal-ratio-combining and thus known [25]. Consequently, if the noise terms $\widetilde{\eta}_s$ are all white, it is immediately clear that the training based scheme achieves the same order of transmit diversity as the corresponding coherent system that uses the information matrices $\left\{\mathbf{B}_{m_s}\right\}_{m_s=1}^{M_{\mathrm{D}}}$. Based on simulations, this seems to continue to be true for cases when $\Phi^{\mathrm{E-C}}$ is not optimal. Note that the optimization over τ is related to the optimization of power between training and data phase in the information theoretic context of [22].

Although we stress the cases in which $\Phi^{\mathrm{E-C}}$ is the optimum receiver, in many cases it is not. Examples of this arise for $S > 1$ or for instance when the Alamouti scheme with M-QAM ($M > 4$) or any generalized, non-square, complex orthogonal design is employed for the data symbol. Similarly, $\Phi^{\mathrm{E-C}}$ is not optimal if an treaded algebraic space time code (TAST) of [26] is used to generate the $\mathbf{B}_{m_s}$. Specifically, we suggest to construct a noncoherent constellation $\{\mathbf{S}_m\}_{m=1}^{M}$ from, for example, a TAST block code for four transmit antennas [26]. For $S = 1$ such a code would have the structure

$$\mathbf{S}_m = \frac{1}{2}\begin{bmatrix} \mathbf{T} \\ \mathbf{B}_m \end{bmatrix}, \tag{15}$$

where $\mathbf{T}$ is a arbitrary 4×4 unitary matrix and $\mathbf{B}_m$ is codeword from the TAST code $\boldsymbol{\mathcal{T}}_{4,L,R}$, as detailed in [26] ($L$ designates the number of layers in the code and R the number of symbols per channel use). While only implementations that require testing all M hypotheses are known for the optimum noncoherent receiver Φ^{ML}, the training perspective immediately suggests the estimator-detector $\Phi^{\mathrm{E-C}}$ as a sub-optimum– albeit well motivated– alternative, whose complexity can be lowered by sphere decoding (cf. [27] and [28]) due to the special structure of the data symbol part of the $\left\{\mathbf{S}_m\right\}_{m=1}^{M}$ [26]. Consequently, the leveraging of coherent designs for training-aided noncoherent signaling addresses both, the signal design problem for the noncoherent multi-antenna channel (that had occupied various researchers before [3–6]) and the decoding complexity problem (that has so far only been addressed in [14, 15]) at once.

IV. Noncoherent Designs for BER as Performance Measure

A signal design method based on the minimization of the asymptotic union bound on symbol error rate (SER) is proposed in [6]. The constellations generated there outperform all previous ones for communicating noncoherently over the multi-antenna channel and the design criterion can be used to improve previous methods. Since in this paper we compare different constellations with respect of average bit error rate (BER), we adapt the design criterion of [6] to minimize an asymptotic bound on the BER, rather than on the SER. Thus the signal design algorithm has to minimize

$$P = M^{-1} \sum_{m=1}^{M} \sum_{\substack{k=1 \\ k \neq m}}^{M} b_{km} P_{k,m}\,, \tag{16}$$

where b_{km} is the Hamming distance between the binary representation of $m-1$ and $k-1$ and

$$P_{k,m} = \frac{\overline{\gamma}^{-N_{\mathrm{T}}N_{\mathrm{R}}} \binom{2N_{\mathrm{T}}N_{\mathrm{R}}-1}{N_{\mathrm{T}}N_{\mathrm{R}}}}{|\mathbf{\Sigma}| \left| {N_{\mathrm{T}}}^{-1}\mathbf{I} - N_{\mathrm{T}}\mathbf{S}_k^{\dagger}\mathbf{S}_m\mathbf{S}_m^{\dagger}\mathbf{S}_k \right|^{N_{\mathrm{R}}}} \tag{17}$$

is the asymptotic pairwise error probability that the receiver Φ^{ML} prefers signal $\mathbf{S}_k$ over $\mathbf{S}_m$ when $\mathbf{S}_m$ is transmitted and $N_{\mathrm{T}}\mathbf{S}_l^{\dagger}\mathbf{S}_l = \mathbf{I}\ \forall l$ [13]. Obviously, we assign to signal $\mathbf{S}_l$ the bits corresponding to the binary representation of $l-1$. Relative to [6] only the inclusion of the factor b_{km} changes and the same numerical methods can be employed to minimize P over the Grassmann manifold. Whenever possible we design the signal sets to directly minimize P over all signals. For larger constellations, when this becomes numerically infeasible, we resort to the successive update algorithm of [6]. The resulting constellations have exponential encoding and decoding complexity (in the transmission rate) but serve as a benchmark for the other designs.

V. Numerical Examples

Figure 1 compares the numerically optimized designs, training, and the (improved) noncoherent PSK constellations of [14] ([20]) at a fixed BER $P_b = 10^{-3}$ for one transmit and receive antenna, $S = 1$, and $D = 2$ dimensions. The "training-code" that uses the first dimension for a training symbol and transmits a usual QAM symbol in the second performs very close to the optimized designs and only for 1.5 bits/dimension are the improved PSK constellations slightly better.

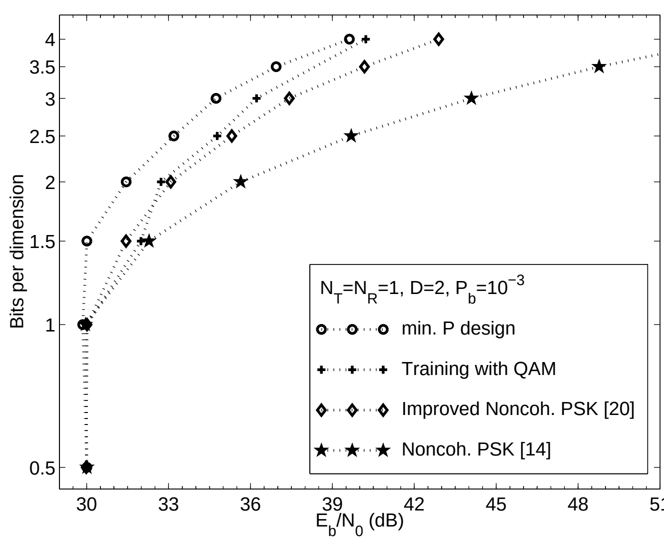

Fig. 1. Spectral efficiency over E_b/N_0 for different constellations when the bit error rate is fixed at 10^{-3}.

Figure 2 compares the different schemes for two transmit and receive antennas in $D = 4$ dimensions. The performance of two different training-codes is displayed: to guarantee transmit diversity the Alamouti scheme [16] with QAM symbols is employed (for BPSK and 4-QAM this corresponds to the constellations of [14, 15]). For comparison, also the performance of transmitting uncoded QAM symbols from the two antennas in the $D_{\mathrm{D}} = 2$ dimensions is displayed. From the figure we see that the training combined with the Alamouti scheme is within 1.5 dB of the signals that are optimized to minimize the asymptotic bound on the BER. The uncoded transmission of QAM symbols does worse, but the gap to the Alamouti scheme becomes closer for increasing rates. Unfortunately, the figure also verifies that the generalization of the (improved) noncoherent PSK constellation to multiple transmit antennas is of no practical interest. Since the generalization is basically a repetition or delay diversity code, this is not too surprising.

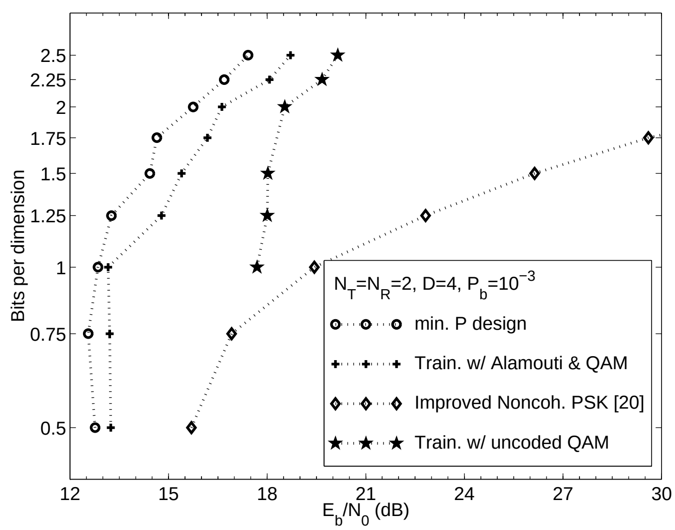

Fig. 2. Spectral efficiency over E_b/N_0 for different constellations when the bit error rate is fixed at 10^{-3}.

While for $N_{\mathrm{T}} = 2$ the Alamouti scheme performs very favorably (despite the fact that it is not optimal from a capacity perspective [29]), for larger number of transmit antennas the price for the simplicity of the orthogonal designs becomes high. In Figure 3 we compare training with the complex orthogonal design for $N_{\mathrm{T}} = 4$ transmit antennas of [18, 19] with the training code suggested in (15). Since the orthogonal design transmits only 3 QAM symbols while the selected TAST code $\mathcal{T}_{4,3,3}$[6] transmits 12, we need 256-QAM symbols for the orthogonal design and QPSK for the TAST training code to achieve a spectral efficiency of 3 bits/dimension. Note that to the best of our knowledge no other noncoherent constellations for $N_{\mathrm{T}} = 4$ have been suggested with such a high spectral efficiency. As one may expect, the training-TAST code considerably outperforms the noncoherent orthogonal design. For comparison, the performance of both schemes with perfect channel knowledge is given (lines with diamond markers). Interestingly, the gap between perfect channel knowledge and channel estimation is quite close to the analytic gap derived for unitary $\left\{\mathbf{B}_m\right\}_{m=1}^{M}$ above (3 dB asymptotically since the comparison is relative to $(1-\tau)\mathbf{B}_m$).

[6] We chose $\phi = \mathrm{e}^{j/2}$, $\theta = \mathrm{e}^{j\pi/8}$, and the $\mathcal{G}_2$ rotation of [30].

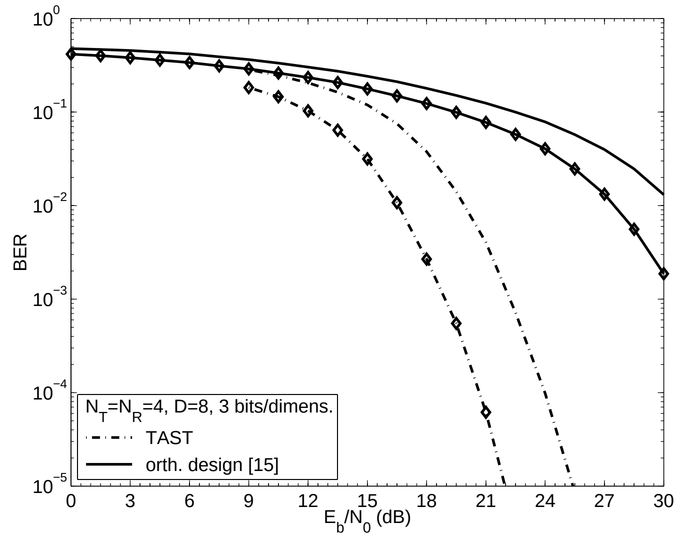

Fig. 3. BER comparison of TAST with 3 layers and 4-QAM with the rate $3/4$-complex orthogonal design and 256-QAM. Plain line-styles give the performance with for estimating the channel, line-styles with diamond markers the performance when a genie provides perfect channel estimates.

So far we have considered $S = 1$ only. In Figure 4 we compare codes for $D = 8$ and $N_{\mathrm{T}} = 2$. [4] suggests a unitary code-design that systematically generates the signal-set by starting with an initial signal and successively rotating it. The original constellation of [4] for $D = 8$, $N_{\mathrm{T}} = 2$ did not guarantee full transmit diversity and is improved upon in [6]. We compare this improved systematic design with a training scheme that transmits a unitary matrix as training symbol and 3 Alamouti symbols with QPSK as information carrying symbols. For comparison we show the performance for the (asymptotically) optimal choice of τ and $\tau = 1/(S+1)$, which ensures that the training symbol has the same energy as an Alamouti symbol. For $S = 3$ the difference in negligible. Since the systematic constellation of [4, 6] contains 2304 matrices, the training scheme has slightly higher spectral efficiency (4096 matrices, 1.5 bits/dimension versus roughly 1.4 bits/dimension). Training requires about 2 dB more E_b/N_0 that the improved systematic design. Note however, that this comparison is on the basis of frame error rate (FER) because the mapping from frames to bits is not clear for the systematic design (giving a slight disadvantage to the training code). More importantly, the training scheme excels in terms of its simplicity for en- and particularly de-coding and easily scales to longer frames and/or higher spectral efficiencies.

VI. Conclusions

Inspired by the information theoretic results of [1, 2], the research on signal design for the noncoherent multi-antenna channel has so far focused on constructing sets of unitary matrices that optimize a performance measure related to the bit

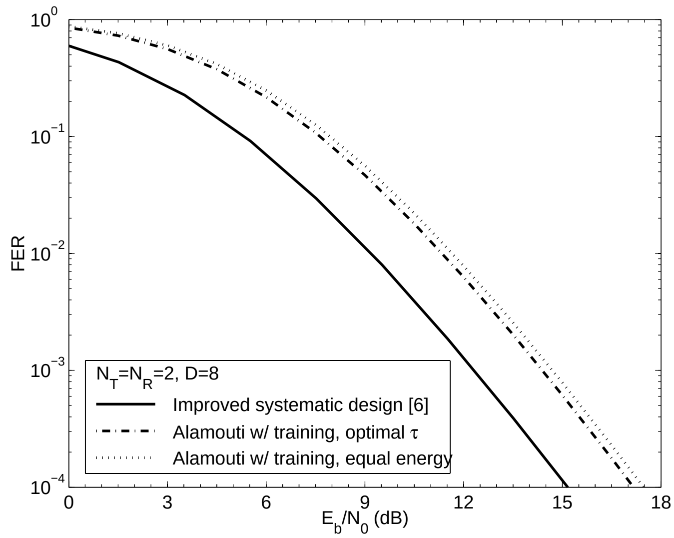

Fig. 4. FER comparison of Alamouti with training and the improved systematic designs of [6]. Note that the the training scheme contains 4096 matrices (frames, corresponding to a SE of 1.5 bits/dimension) while the systematic design has 2304 matrices (i.e. the SE is roughly 1.4 bits/dimension).

or symbol error rate of the set. The resulting constellations are often unwieldy and in most cases displayed a decoding complexity that is exponential in the rate (because all symbols have to be tested). The recent designs of [14, 15] tackle the complexity problem and independently suggest noncoherent schemes based on the generalized orthogonal designs. We show that these–essentially identical–designs fall within the much broader class of training codes introduced here. Training codes are easily constructed according to the principles laid-out in this paper. They scale to longer coherence intervals (frame lengths) and/or higher spectral efficiencies and leverage the significant advances made in coherent space–time code designs. Simultaneously, low-complexity detection (decoding) algorithms are suggested by the training perspective.

References

[1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[2] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[3] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[4] B. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space–time constellations," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[5] D. Agrawal, T. J. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2618–2626, Sept. 2001.

[6] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for noncoherent space–time communication on the block-Rayleigh-fading channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1186–1194, May 2002.

[7] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1169–1174, July 2000.

[8] B. L. Hughes, "Differential space–time modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2567–2578, Nov. 2000.

[9] B. Hochwald and W. Sweldens, "Differential unitary space–time modulation," *IEEE Trans. Commun.*, vol. 48, pp. 2041–2052, Dec. 2000.

[10] H. Jafarkhani and V. Tarokh, "Multiple transmit antenna differential detection from generalized orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2626–2631, Sept. 2001.

[11] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2335–2367, Sept. 2001.

[12] B. Hassibi and B. M. Hochwald, "Cayley differential unitary space–time codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1485–1503, June 2002, Special Issue on Shannon Theory: Perspective, Trends, and Applications.

[13] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space–time receivers," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2383–2399, Sept. 2001.

[14] V. Tarokh and I.-M. Kim, "Existence and construction of noncoherent unitary space–time codes," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3112–3117, Dec. 2002.

[15] W. Zhao, G. Leus, and G. B. Giannakis, "Algebraic design of unitary constellations for uncoded and trellis coded modulation of noncoherent space–time systems," submitted to *IEEE Trans. Inform. Theory*, July 2002, also to appear in *Proc. of Intl. Conf. on Communications*, 2003.

[16] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[17] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[18] G. Ganesan and P. Stoica, "Space–time block codes: A maximum SNR approach," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1650–1656, May 2001.

[19] O. Tirkkonen and A. Hottinen, "Square-matrix embeddable space–time block codes for complex signal constellations," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 384–395, Feb. 2002.

[20] M. Brehler and M. K. Varanasi, "On signal design for the noncoherent multi-antenna block-rayleigh-fading channel," submitted to *IEEE Trans. Inform. Theory*, May 2002.

[21] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[22] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," submitted to *IEEE Trans. Inform. Theory*, Apr. 2000.

[23] K. M. Mackenthun, Jr., "A fast algorithm for multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1471–1474, Feb./Mar./Apr. 1994.

[24] W. Sweldens, "Fast block noncoherent decoding," *IEEE Commun. Letters*, vol. 5, no. 4, pp. 132–134, Apr. 2001.

[25] G. Bauch, "Turbo processing in transmit antenna diversity systems," *Ann. Telecommun.*, Aug. 2001, Special Issue on Turbo Codes and Related Topics.

[26] H. El Gamal and M. O. Damen, "Universal space–time coding," submitted to *IEEE Trans. Inform. Theory*, Jan. 2002.

[27] M. O. Damen, A. Chkeif, and J.-C. Belfiore, "Lattice code decoder for space–time codes," *IEEE Commun. Letters*, vol. 4, no. 5, pp. 161–163, May 2000.

[28] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[29] S. Sandhu and A. Paulraj, "Space–time block codes: A capacity perspective," *IEEE Commun. Letters*, vol. 4, no. 12, pp. 384–386, Dec. 2000.

[30] X. Giraud, E. Boutillon, and J.-C. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 938–952, May 1997.