

Space–Time Fading Channel Estimation and Symbol Detection in Unknown Spatially Correlated Noise

Aleksandar Dogandžić, *Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

Abstract—We present maximum likelihood (ML) methods for space–time fading channel estimation with an antenna array in spatially correlated noise having unknown covariance; the results are applied to symbol detection. The received signal is modeled as a linear combination of multipath-delayed and Doppler-shifted copies of the transmitted waveform. We consider structured and unstructured array response models and derive the Cramér–Rao bound (CRB) for the unknown directions of arrival, time delays, and Doppler shifts. We also develop methods for spatial and temporal interference suppression. Finally, we propose coherent matched-filter and concentrated-likelihood receivers that account for the spatial noise covariance and analyze their performance.

Index Terms—Array processing, coherent matched-filter detection, noncoherent detection, time-varying channel estimation.

I. INTRODUCTION

TO COMBAT fading and suppress interference in wireless communications, both channel and noise properties need to be estimated. In this paper (see also [1]), we propose algorithms and derive performance measures for space–time estimation of *fast, frequency-selective fading* channels in the presence of spatially correlated noise having *unknown* covariance. This is unlike most previous work, which typically assumes slow-fading channel (neglecting the Doppler effects) [2]–[9] or spatially white noise [10], [11], or both [12]–[16]. (References to several recent papers dealing with spatially correlated noise will be given later.) We consider structured and unstructured array responses and model the signal as a linear combination of parametric basis functions that account for the multipath and Doppler effects. The proposed basis-function model is used to combat fading (due to multipath and Doppler effects), whereas the spatially correlated noise accounts for co-channel interference. We then propose receivers that utilize *both* estimates of the channel and spatial noise covariance and analyze their performance.

The signal and noise models are explained in detail in Section II. First, we present the structured array response model

that allows estimation of directions of arrival (DOAs). The DOA information can be useful for locating E-911 calls and other applications (for example, radar and sonar; see the discussion in Section II). Then, in Section II-A, we present the unstructured array response model, which avoids estimating the DOA parameters, thereby reducing the computational complexity. In Section II-B, we propose several basis-function models that account for fast, frequency-selective fading by suitably discretizing the time-delay and Doppler spreads of the received signal. Compared with the exact signal model, these basis-function models significantly reduce the number of (nonlinear) time-delay and Doppler parameters to be estimated. In Section II-B1, we show that the proposed basis-function formulation is well suited for transmit–receive antenna array systems as well.

In Sections III and IV, we present maximum likelihood (ML) estimation algorithms and Cramér–Rao bounds (CRBs) for the structured array response model. We show that the CRBs for the direction-of-arrival and basis-function parameters (time delays and Doppler shifts) are uncoupled. As a consequence, the CRB expressions for the basis-function parameters are independent of the array-response parameterization (and vice versa). This result has important practical implications (motivating the unstructured array model); see Section IV. Then, in Section V, we derive ML algorithms for the unstructured model. In Sections III and V, we also derive the concentrated likelihood functions for the unknown nonlinear parameters and discuss relationship with previous work. Additionally, in Section V-A, we investigate how the time-delay and Doppler-shift discretizations affect the channel and noise covariance estimates and concentrated likelihood function. In Section VI, we discuss methods for spatial and temporal interference suppression based on decompositions of the concentrated likelihood function.

In Section VII, we propose coherent matched-filter and concentrated-likelihood receivers that incorporate the estimates of both the channel and spatial noise covariance. These receivers exploit the following.

- multipath and Doppler diversity (using the proposed basis-function models; see Section II-B);
- interference-suppression capability of the antenna array (using the spatially correlated noise model for the interference).

We show analytically that the proposed coherent matched-filter receiver outperforms (in terms of error probability) the coherent receiver that does not take into account the spatial noise covariance. In Section VII-B, we propose concentrated-likelihood (with respect to the channel and noise parameters) receiver and derive its recursive implementation (see Section VII-C). Finally, concluding remarks are given in Section VIII.

Manuscript received August 25, 2000; revised October 29, 2001. This work was supported by the Air Force Office of Scientific Research under Grants F49620-99-1-0067 and F49620-00-1-0083, the National Science Foundation under Grant CCR-0105334, and the Office of Naval Research under Grant N00014-01-1-0681. The associate editor coordinating the review of this paper and approving it for publication was Dr. Olivier Besson.

A. Dogandžić is with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA. He is now with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: ald@iastate.edu).

A. Nehorai is with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: nehorai@ece.uic.edu).

Publisher Item Identifier S 1053-587X(02)01342-9.

II. SIGNAL AND NOISE MODELS

Suppose an m -element antenna array receives several scaled, time delayed, and Doppler-shifted copies of a known complex baseband signal $s(t)$. Then, the $m \times 1$ vector signal received by the array at time t becomes

$$\mathbf{y}(t) = \sum_{p=1}^P \mathbf{a}(\boldsymbol{\theta}_{Tp}) \cdot x_{Tp} \cdot \exp(j\omega_{DTp}t) s(t - \tau_{Tp}) + \mathbf{e}(t) \quad (2.1)$$

for $t = 1, \dots, N$, where $\boldsymbol{\theta}_{Tp}$ is the vector of DOA parameters (and may contain additional parameters, such as scattering and polarization coefficients; see [17]), and x_{Tp} , ω_{DTp} , and τ_{Tp} are the complex amplitude, Doppler shift, and time delay for the p th path ($p = 1, 2, \dots, P$). Further, $\mathbf{a}(\boldsymbol{\theta})$ denotes the (single DOA) array response vector, and $\mathbf{e}(t)$ is additive noise. Note that in the above model, it is assumed that the Doppler effect can be modeled by a frequency shift (which is the *narrowband signal assumption*; see e.g., [18, ch. 9, eqs. (19) and (23)]), and the propagation time of the signal across the array is much smaller than the reciprocal of the signal bandwidth, which is the standard *narrowband array assumption*. The time-delay and Doppler spreads of the received signal in (2.1) are equal to $\max\{\tau_{T1}, \tau_{T2}, \dots, \tau_{TP}\} - \min\{\tau_{T1}, \tau_{T2}, \dots, \tau_{TP}\}$ and $\max\{\omega_{DT1}, \omega_{DT2}, \dots, \omega_{DTP}\} - \min\{\omega_{DT1}, \omega_{DT2}, \dots, \omega_{DTP}\}$, respectively. We use the subscript T in the above parameterization to emphasize that these are the “true” parameters, as opposed to the discretized model introduced below (see Section II-B).

The model (2.1) contains a large number of parameters, especially when the number of paths P is large. A way to solve this is to parameterize the time delays τ_{Tp} and Doppler shifts ω_{DTp} by using suitable basis-function expansions, which we discuss in detail in Section II-B. To accommodate basis functions and more general scenarios, we generalize (2.1) to the *structured array model*

$$\mathbf{y}(t) = A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \mathbf{e}(t), \quad t = 1, \dots, N \quad (2.2)$$

where $A(\boldsymbol{\theta})$ is an $m \times r$ array response matrix parameterization, and X is an $r \times d$ matrix of unknown coefficients for the functional representation described by a $d \times 1$ vector of basis functions $\boldsymbol{\phi}(t, \boldsymbol{\eta})$ (where the parameter vector $\boldsymbol{\eta}$ is unknown in general). Note that the true model in (2.1) is a special case of (2.2) with $r = d = P$, $A(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_{T1}) \cdots \mathbf{a}(\boldsymbol{\theta}_{TP})]$, $\boldsymbol{\theta} = [\boldsymbol{\theta}_{T1}^T, \dots, \boldsymbol{\theta}_{TP}^T]^T$, $X = \text{diag}(x_{T1}, x_{T2}, \dots, x_{TP})$, $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = \boldsymbol{\phi}_T(t, \boldsymbol{\eta}_T) = [\exp(j\omega_{DT1}t)s(t - \tau_{T1}), \dots, \exp(j\omega_{DTP}t)s(t - \tau_{TP})]^T$, and $\boldsymbol{\eta} = \boldsymbol{\eta}_T = [\omega_{DT1}, \dots, \omega_{DTP}, \tau_{T1}, \dots, \tau_{TP}]^T$. The structured array basis-function formulation in (2.2) is very general. It has been used in [19] for EEG/MEG source location, in [4] for DOA estimation (assuming completely known basis functions), and in [20] for radar detection (assuming known DOA's and basis functions). This model stems from multivariate statistical analysis; see the discussion in Section III. In Section II-B, we show how it can be used to model fast-fading, frequency-selective communication channels, as well as transmit-receive antenna array systems (see Section II-B1). Additionally, it can be used to model multiple moving point

targets (applicable to radar or sonar array processing); then, $A(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_{T1}) \cdots \mathbf{a}(\boldsymbol{\theta}_{TP})]$ and $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = \boldsymbol{\phi}_T(t, \boldsymbol{\eta}_T)$, where P is now the number of targets.

The noise term $\mathbf{e}(t)$ models interference due to receiver noise, co-channel interference (CCI) in wireless communication channels, or clutter in radar. Throughout this paper, we assume that $\mathbf{e}(t)$ is zero-mean Gaussian, temporally white and spatially correlated with unknown positive definite covariance Σ , constant in time.

Define the vector of channel parameters $\boldsymbol{\rho} = [\text{vec}(\text{Re}\{X\})^T, \text{vec}(\text{Im}\{X\})^T, \boldsymbol{\theta}^T, \boldsymbol{\eta}^T]^T$ and the mean (i.e., noiseless) signal vector at time t as $\boldsymbol{\mu}(t, \boldsymbol{\rho}) = A(\boldsymbol{\theta})X\boldsymbol{\phi}(t, \boldsymbol{\eta})$. Our goal is to estimate $\boldsymbol{\rho}$ and the noise covariance Σ . For the structured array model in (2.2), we need a suitable parameterization of the array response matrix $A(\boldsymbol{\theta})$. For example, we may choose $A(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1) \cdots \mathbf{a}(\boldsymbol{\theta}_r)]$, where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T \cdots \boldsymbol{\theta}_r^T]^T$, discretize the angular spread of the received signals [21], or use a distributed source model [22]. For communication applications in particular, we are interested in synchronizing the receiver (i.e., estimating the time-delay and Doppler parameters $\boldsymbol{\eta}$; see Section II-B) and estimating the channel matrix $H = A(\boldsymbol{\theta})X$. In radar array processing problems, it is of particular interest to find the estimate of $\boldsymbol{\eta}$ (radial velocities and ranges of all the targets) and the targets' direction parameters $\boldsymbol{\theta}$; then, the “true” model (2.1) is the model of choice.

If we are not interested in estimating the DOAs or do not have an appropriate parameterization for the array response matrix, we may use the unstructured array model described later. For synchronization, it is preferable to use the unstructured array model; see the discussion in Section IV.

A. Unstructured Array Model

We describe the unstructured array response model in which the whole array response matrix is assumed to be unknown with an arbitrary (known) rank, as in [23]. The unstructured array model is robust compared with an incorrectly-structured array model.¹ In addition, it avoids nonlinear DOA parameterization.

The unstructured array model can be written as

$$\mathbf{y}(t) = H\boldsymbol{\phi}(t, \boldsymbol{\eta}) + \mathbf{e}(t), \quad t = 1, \dots, N \quad (2.3)$$

with the same noise assumptions as for the structured model in the previous section. Here, H is the channel response matrix of rank r , parameterized as $H = AX$, where A is an $m \times r$ unknown array response [of full rank $r \leq \min(m, d)$], and X is an $r \times d$ unknown basis-function coefficient matrix having full rank r . (We discuss the identifiability of A and X in Section V and Appendix B.) The vector of the unknown channel parameters for this model is $\boldsymbol{\rho}_u = [\text{vec}(\text{Re}\{H\})^T, \text{vec}(\text{Im}\{H\})^T, \boldsymbol{\eta}^T]^T$. The mean signal vector for the unstructured model at time t is then $\boldsymbol{\mu}_u(t, \boldsymbol{\rho}_u) = H\boldsymbol{\phi}(t, \boldsymbol{\eta})$.

Several special cases of the above model have been considered in the literature. Full-rank unstructured channel models [i.e., $r = \min(d, m)$] are used in [11] and [16] (assuming spatially white noise) and [8] and [9] (assuming spatially correlated

¹Structured array model inaccuracies are due to changes in antenna locations, temperature, surrounding environment, gain and phase errors, mutual coupling, quantization and interpolation errors, etc., and can lead to a significant loss in estimation performance.

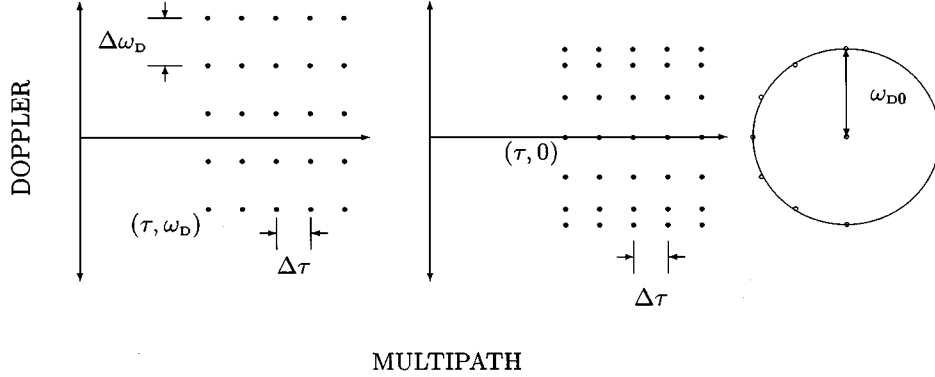


Fig. 1. Rectangular sampling grid in the time-frequency plane with (Left) uniform time-delay and Doppler spread discretizations and (Right) uniform time-delay and Jakes' Doppler spread discretizations.

noise with unknown covariance). In [16], methods are derived for estimating time delays of multiple asynchronous signals and in [11] for estimating time delays and Doppler shifts of a signal in multipath environment [i.e., using $\phi(t, \boldsymbol{\eta}) = \phi_{\mathbf{T}}(t, \boldsymbol{\eta}_{\mathbf{T}})$]. In [8] and [9], methods are derived for time-delay estimation in a *slow, frequency-selective fading* environment. A *flat-fading* scenario, when the channel matrix degenerates to a vector (i.e., $r = d = 1$), is used in [2] and [6] (see also [24] and [25] for related radar work).

B. Basis-Function Models

Now, we propose the following fairly general basis-function structure to model fast, frequency-selective fading

$$\phi(t, \boldsymbol{\eta}) = \phi_{\tau}(t, \boldsymbol{\eta}_{\tau}) \otimes \phi_{\omega_{\text{D}}}(t, \boldsymbol{\eta}_{\omega_{\text{D}}}) \quad (2.4)$$

where $\phi_{\tau}(t, \boldsymbol{\eta}_{\tau})$ (of size $L \times 1$) and $\phi_{\omega_{\text{D}}}(t, \boldsymbol{\eta}_{\omega_{\text{D}}})$ (of size $K \times 1$) are suitable parameterizations of the received signal's time delay (or multipath) and Doppler spreads, and \otimes denotes the Kronecker product. Here, $\boldsymbol{\eta} = [\boldsymbol{\eta}_{\tau}^T, \boldsymbol{\eta}_{\omega_{\text{D}}}^T]^T$, and the number of basis functions is $d = LK$. Observe that (2.4) models the fading effects using a rectangular grid in the time-frequency plane; see Fig. 1 and the following discussion.

Note that the model (2.2) with arbitrary X and basis functions of the form (2.4) with $\phi_{\tau}(t, \boldsymbol{\eta}_{\tau}) = [s(t - \tau_{\text{T}1}), s(t - \tau_{\text{T}2}), \dots, s(t - \tau_{\text{T}L})]^T$ and $\phi_{\omega_{\text{D}}}(t, \boldsymbol{\eta}_{\omega_{\text{D}}}) = [\exp(j\omega_{\text{D}1}t), \exp(j\omega_{\text{D}2}t), \dots, \exp(j\omega_{\text{D}K}t)]^T$ captures a more general class of fading channels than (2.1) since, for example, it allows for multipath signals arriving from the same direction. For computational simplicity, we can uniformly discretize the time-delay and Doppler spreads

$$\phi_{\tau}(t, \tau) = [s(t - \tau), s(t - \tau - \Delta\tau), \dots, s(t - \tau - (L - 1)\Delta\tau)]^T \quad (2.5a)$$

$$\phi_{\omega_{\text{D}}}(t, \omega_{\text{D}}) = [\exp[j\omega_{\text{D}}t], \exp[j(\omega_{\text{D}} + \Delta\omega_{\text{D}})t], \dots, \exp[j(\omega_{\text{D}} + (K - 1)\Delta\omega_{\text{D}})t]]^T \quad (2.5b)$$

where τ and ω_{D} are the “global” time delay and Doppler shift, respectively, whereas $\Delta\tau$ and $\Delta\omega_{\text{D}}$ are the corresponding sampling intervals. Since the choices of $\Delta\tau$ and $\Delta\omega_{\text{D}}$, respectively, depend on the bandwidth and duration of the signal $s(t)$, these quantities are known *a priori* [26]. If we also assume that prior knowledge on the delay and Doppler spreads is available,

then we can also determine L and K (and therefore d as well) *a priori*. In this case, $\boldsymbol{\eta}_{\tau}$, $\boldsymbol{\eta}_{\omega_{\text{D}}}$, and $\boldsymbol{\eta}$ reduce to τ , ω_{D} , and $[\omega_{\text{D}}, \tau]^T$, respectively. Basis functions similar to (2.5) have been introduced in [26] and [21] (see also references therein) to model fast-fading, frequency-selective communication channels. Unlike (2.5), it is assumed in [21] and [26] that the receiver is perfectly synchronized to the “global” delay and Doppler shift, which are set to zero. The time-varying channel model proposed in [27] can also be viewed as a special case of (2.4) with $\phi_{\tau}(t, \tau)$ as in (2.5a) (with τ set to zero), $\phi_{\omega_{\text{D}}}(t, \boldsymbol{\eta}_{\omega_{\text{D}}}) = [\exp(j\omega_{\text{D}1}t), \exp(j\omega_{\text{D}2}t), \dots, \exp(j\omega_{\text{D}K}t)]^T$, and $\boldsymbol{\eta}_{\omega_{\text{D}}} = [\omega_{\text{D}1}, \dots, \omega_{\text{D}K}]^T$.

An interesting, physically motivated model for $\phi_{\omega_{\text{D}}}(t, \boldsymbol{\eta}_{\omega_{\text{D}}})$ follows by discretizing the Jakes model [28] with the $K \times 1$ basis-function vector of the form

$$\phi_{\omega_{\text{D}}}(t, \omega_{\text{D}0}) = \begin{bmatrix} \exp[j\omega_{\text{D}0}t], \exp[j\omega_{\text{D}0} \cos(1 \cdot \pi/(K - 1))t] \\ \exp[j\omega_{\text{D}0} \cos(2 \cdot \pi/(K - 1))t], \dots \\ \exp(-j\omega_{\text{D}0}t) \end{bmatrix}^T \quad (2.6)$$

where $\omega_{\text{D}0}$ is proportional to the speed of the mobile (and, consequently, to the Doppler spread), and $\boldsymbol{\eta}_{\omega_{\text{D}}} = \omega_{\text{D}0}$. In general, if a mobile receiver (or mobile transmitter) is surrounded by a rich scattering environment, (2.6) should be used to model the distribution of the received Doppler shifts. Note that in (2.6), the Doppler spread is sampled more densely around $-\omega_{\text{D}0}$ and $\omega_{\text{D}0}$ due to the “bathtub-shaped” Doppler profile [28]. If, however, the Doppler spread is small and around a certain mean value (which may be nonzero, in general), then a simple uniform discretization in (2.5b) should be used.

Fig. 1 illustrates the rectangular grid (2.4) in the *time-frequency plane* for the two sampling schemes considered here. Fig. 1 (left) shows the uniform time-delay and Doppler spread sampling in (2.5), whereas Fig. 1 (right) shows the uniform time-delay and Jakes' Doppler spread sampling in (2.5a) and (2.6).

1) *Transmit-Receive Antenna Arrays*: The measurement model (2.3) can also be used to describe *transmit-receive* antenna array systems [29], [30]. Here, the basis functions describe signals sent by the transmitter array. For example, in a *slow, flat fading* environment (i.e., negligible Doppler effect and

time-delay spread), the basis-function vector $\phi(t, \boldsymbol{\eta})$ simply becomes $\phi(t, \boldsymbol{\eta}) = [s_1(t - \tau), s_2(t - \tau), \dots, s_d(t - \tau)]^T$, where $\boldsymbol{\eta} = \tau$, and $s_1(t), s_2(t), \dots, s_d(t)$ are the signals transmitted by a transmitter array with d elements. Note also that the model (2.3) allows channel matrix H to be of an arbitrary rank, which is of practical importance in transmit–receive antenna array systems [31], [32].

III. MAXIMUM LIKELIHOOD ESTIMATION WITH STRUCTURED ARRAY

We present the ML estimation procedure for the structured array response model in (2.2) and the noise model from Section II. If the DOA parameters $\boldsymbol{\theta}$ and basis-function parameters $\boldsymbol{\eta}$ are known, the above model is known as the generalized multivariate analysis of variance (GMANOVA) [33]. For known basis functions (i.e., known $\boldsymbol{\eta}$), exact and approximate ML methods for DOA estimation are derived in [4]. For spatially and temporally white noise, an ML estimation algorithm for unknown DOAs, time delays, Doppler shifts, and complex amplitudes in the exact signal model (2.1) is derived in [10]. Here, we estimate the DOAs, basis-function parameters (for models from Section II-B), and the unknown spatial covariance. Additionally, we derive CRB expressions for these parameters; see Section IV.

We will present two expressions for the concentrated likelihood function to be maximized with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. These expressions provide new insight into the problem (for example, relationship with Capon spectral estimation) and enable easy derivation of the ML results for unstructured array; see Appendix A. The results in this section are useful for estimating the locations and radial velocities of P moving point targets (for example, in radar) with $r = d = P$, $A(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_{T1}) \cdots \mathbf{a}(\boldsymbol{\theta}_{TP})]$, and $\phi(t, \boldsymbol{\eta}) = \phi_T(t, \boldsymbol{\eta}_T)$.

Define $Y = [\mathbf{y}(1) \cdots \mathbf{y}(N)]$ and $\Phi(\boldsymbol{\eta}) = [\phi(1, \boldsymbol{\eta}) \cdots \phi(N, \boldsymbol{\eta})]$. The ML estimates of X and Σ for known $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are [33, ch. 6.4], [34, ch. 5], [35]

$$\hat{X}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \left[A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \right]^{-1} A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \quad (3.1a)$$

$$\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta}) = \hat{S}_{y|\phi} + \left(I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1} \right) \cdot \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \hat{R}_{y\phi}^* \left(I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1} \right)^* \quad (3.1b)$$

where

$$\hat{S}_{y|\phi} = \hat{R}_{yy} - \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \hat{R}_{y\phi}^* \quad (3.2a)$$

$$\hat{R}_{yy} = \frac{1}{N} Y Y^* \quad (3.2b)$$

$$\hat{R}_{\phi\phi} = \frac{1}{N} \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^* \quad (3.2c)$$

$$\hat{R}_{y\phi} = \hat{R}_{\phi y}^* = \frac{1}{N} Y \Phi(\boldsymbol{\eta})^* \quad (3.2d)$$

$$\hat{T}_A = A(\boldsymbol{\theta}) \left[A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \right]^{-1} A(\boldsymbol{\theta})^* \quad (3.2e)$$

and I_m denotes the identity matrix of size m . Note that $\hat{S}_{y|\phi}$, $\hat{R}_{\phi\phi}$, and $\hat{R}_{y\phi}$ are functions of $\boldsymbol{\eta}$ only, and \hat{T}_A is a function of

both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. To simplify the notation, we omit these dependencies throughout this paper. Further, observe that $\hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1}$ is an adaptive (i.e., estimated) Wiener filter for estimating $\mathbf{y}(t)$ from $\phi(t, \boldsymbol{\eta})$, $t = 1, \dots, N$, and $\hat{S}_{y|\phi}$ is the estimate of its error covariance; see [36, Sec. VII]. In addition, note that $\hat{S}_{y|\phi}^{-1/2} \hat{T}_A \hat{S}_{y|\phi}^{-1/2}$ is the projection matrix onto the column space of $\hat{S}_{y|\phi}^{-1/2} A(\boldsymbol{\theta})$. Here, $\Sigma^{1/2}$ denotes a Hermitian square root of a Hermitian matrix Σ , and $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$; this notation will be used throughout the paper.

If $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ are not known, their ML estimates $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$ are computed by maximizing the concentrated likelihood function $l(\boldsymbol{\theta}, \boldsymbol{\eta}) = |\hat{R}_{yy}| / |\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})|$, obtained by substituting the estimates of X and Σ in (3.1) into the likelihood function (see, e.g., [19, App. A]). This concentrated likelihood function is written in the form of a generalized likelihood ratio (GLR) test statistic (see, e.g., [37] and [38, p. 418] for the definition of GLR) for testing $H_0: X = 0$. To find the ML estimates of X and Σ for unknown $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, substitute $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ in (3.1a) and (3.1b) by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\eta}}$.

The above concentrated likelihood function can be rewritten as (see Appendix A)

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\Phi(\boldsymbol{\eta})[I_N - (1/N) \cdot Y^* W Y] \Phi(\boldsymbol{\eta})^*|}{|\Phi(\boldsymbol{\eta})[I_N - (1/N) \cdot Y^* \hat{R}_{yy}^{-1} Y] \Phi(\boldsymbol{\eta})^*|} = \frac{|\hat{R}_{\phi\phi} - \hat{R}_{y\phi}^* W \hat{R}_{y\phi}|}{|\hat{S}_{\phi|y}|} \quad (3.3)$$

where

$$W = \hat{R}_{yy}^{-1} - \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta})]^{-1} A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \quad (3.4a)$$

$$\hat{S}_{\phi|y} = \hat{R}_{\phi\phi} - \hat{R}_{y\phi}^* \hat{R}_{yy}^{-1} \hat{R}_{y\phi} \quad (3.4b)$$

Here, $\hat{S}_{\phi|y}$ is a function of $\boldsymbol{\eta}$, and W is a function of $\boldsymbol{\theta}$, but we omit these dependencies throughout the paper to simplify the notation. Note that $\hat{R}_{yy}^{-1/2} W \hat{R}_{yy}^{1/2}$ is the projection matrix onto the space orthogonal to the column space of $\hat{R}_{yy}^{-1/2} A(\boldsymbol{\theta})$. Another interesting form of the concentrated likelihood derived in Appendix A is

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})|}{|A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta})|} \quad (3.5)$$

The concentrated likelihood (3.5) is the ratio of the Capon spectral estimate in the direction $\boldsymbol{\theta}$ using the data Y and the Capon spectral estimate in the direction $\boldsymbol{\theta}$ using the projection of Y onto the space orthogonal to the row space of $\Phi(\boldsymbol{\eta})$. This can further be viewed as the overall power arriving from the direction $\boldsymbol{\theta}$, normalized by the power of the noise only, arriving from the same direction $\boldsymbol{\theta}$ [39].

When there is only one path, i.e., $P = 1$ in (2.1) or $L = K = 1$ in (2.2), (2.4), and (2.5), the concentrated likelihood is obtained by replacing $A(\boldsymbol{\theta})$, X , and $\phi(t, \boldsymbol{\eta})$ with $\mathbf{a}(\boldsymbol{\theta})$, x , and $\exp(j\omega_{DT})s(t - \tau)$ in (3.3) and (3.5); see also [25] and [40]. If $s(t) \equiv 1$, the concentrated likelihood further reduces to the expression in [24, (16)]. This scenario, which has been analyzed extensively in [24], implies that matched filtering has

been performed beforehand [i.e., snapshot $\mathbf{y}(t)$ corresponds to the matched-filtered return from the t th pulse], and the DOA and Doppler shift are estimated using the matched-filtered data. This approach, however, does not allow estimation of the time delay.

IV. CRAMÉR–RAO BOUND

We derive general CRB expressions for the DOA and basis-function parameters (time delays and Doppler shifts) in the structured array model (2.2) and the spatially correlated noise with unknown covariance (described in Section II). The results will apply to the unstructured array model as well and justify its use for synchronization; see the discussion below. In radar and sonar, the derived CRB results are useful for analyzing the accuracy of estimating time delays, Doppler shifts, and DOAs of P moving point targets (see Sections II and III).

Define the projection matrices

$$\Pi_A^\perp(\boldsymbol{\theta}, \Sigma) = I_m - \Sigma^{-1/2} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^* \Sigma^{-1} A(\boldsymbol{\theta})]^{-1} \cdot A(\boldsymbol{\theta})^* \Sigma^{-1/2} \quad (4.1a)$$

$$\Pi_{\Phi^*}^\perp(\boldsymbol{\eta}) = I_N - \Phi(\boldsymbol{\eta})^* [\Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^*]^{-1} \Phi(\boldsymbol{\eta}) \quad (4.1b)$$

which span the spaces orthogonal to the column spaces of $\Sigma^{-1/2} A(\boldsymbol{\theta})$ and $\Phi(\boldsymbol{\eta})^*$, respectively. In addition, define $D_A(\boldsymbol{\theta}) = \partial \text{vec}(A(\boldsymbol{\theta})) / \partial \boldsymbol{\theta}^T$ and $D_\Phi(\boldsymbol{\eta}) = \partial \text{vec}(\Phi(\boldsymbol{\eta})) / \partial \boldsymbol{\eta}^T$. We derive the CRBs for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ and show that they are uncoupled (see [40, App. C])

$$\text{CRB}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{1}{2N} [\text{Re}\{D_A(\boldsymbol{\theta})^* [(X \hat{R}_{\phi\phi} X^*)^T \otimes \Sigma^{-1/2} \Pi_A^\perp(\boldsymbol{\theta}, \Sigma) \Sigma^{-1/2}] D_A(\boldsymbol{\theta})\}]^{-1} \quad (4.2a)$$

$$\text{CRB}_{\boldsymbol{\eta}\boldsymbol{\theta}} = 0 \quad (4.2b)$$

$$\begin{aligned} \text{CRB}_{\boldsymbol{\eta}\boldsymbol{\eta}} &= \frac{1}{2} \left[\text{Re} \left\{ \sum_{t=1}^N \frac{\partial \phi(t, \boldsymbol{\eta})^*}{\partial \boldsymbol{\eta}} X^* A(\boldsymbol{\theta})^* \Sigma^{-1} A(\boldsymbol{\theta}) \right. \right. \\ &\quad \cdot X \frac{\partial \phi(t, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} - \frac{1}{N} \sum_{t_2=1}^N \sum_{t_1=1}^N \phi(t_2, \boldsymbol{\eta})^* \\ &\quad \cdot \hat{R}_{\phi\phi}^{-1} \phi(t_1, \boldsymbol{\eta}) \cdot \frac{\partial \phi(t_1, \boldsymbol{\eta})^*}{\partial \boldsymbol{\eta}} X^* A(\boldsymbol{\theta})^* \Sigma^{-1} \\ &\quad \left. \left. \cdot A(\boldsymbol{\theta}) X \frac{\partial \phi(t_2, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \right\} \right]^{-1} \\ &= \frac{1}{2} [\text{Re}\{D_\Phi(\boldsymbol{\eta})^* \\ &\quad \cdot [\Pi_{\Phi^*}^\perp(\boldsymbol{\eta})^T \otimes X^* A(\boldsymbol{\theta})^* \Sigma^{-1} A(\boldsymbol{\theta}) X] \\ &\quad \cdot D_\Phi(\boldsymbol{\eta})\}]^{-1}. \end{aligned} \quad (4.2c)$$

The result (4.2b) is somewhat unexpected since the Fisher information matrix (FIM) for the signal parameters $\boldsymbol{\rho}$ is a full matrix in general; see [40, App. C]. Since the CRB for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ is block-diagonal, $\text{CRB}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ remains the same whether or not $\boldsymbol{\theta}$ is known. Furthermore, $\text{CRB}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ remains the same regardless of the array-response parameterization. Similarly, $\text{CRB}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the same, regardless of whether or not $\boldsymbol{\eta}$ is known and is invariant to basis-function parameterization.

The practical implication of the CRB decoupling is that, if we are primarily interested in synchronization (i.e., estimating $\boldsymbol{\eta}$, as it is often the case in communications), then it is sufficient to use the (tractable) unstructured array model since, asymptotically, the same accuracy in estimating $\boldsymbol{\eta}$ can be achieved for unstructured, structured, or known array response. Therefore, $\text{CRB}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ in (4.2c) is also valid for the unstructured array model in Section II-A, where $A(\boldsymbol{\theta})X$ should be replaced with H . The CRB expressions in (4.2) are derived for a very general measurement model (2.2) (with arbitrarily parameterized array response and basis functions) and are thus applicable to other nonlinear regression problems.

V. MAXIMUM LIKELIHOOD ESTIMATION WITH UNSTRUCTURED ARRAY

We present the ML estimates of the channel and spatial noise covariance matrices for the unstructured array model in (2.3). For real measurements and completely known basis functions (i.e., known $\boldsymbol{\eta}$), this problem has been solved by Stoica and Viberg in [23]. Here (see also [1] and [40]), we extend their results to complex data, provide additional interpretations and alternative derivations, extend the estimation to account for unknown $\boldsymbol{\eta}$, and, finally, in Section VII, apply the results to the receiver design. For spatially and temporally white noise and a full-rank unstructured channel model, estimation algorithms for time delays and Doppler shifts were derived in [11]. Here, we consider a channel model of arbitrary rank (which is denoted by r ; see Section II-A), spatially correlated noise with unknown covariance, and discretized basis-function models that reduce the computational complexity (see Section II-B).

Define $\mathcal{P} = \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1/2}$ and

$$\hat{C}_{y\phi} = \hat{R}_{yy}^{-1/2} \mathcal{P} = \hat{R}_{yy}^{-1/2} \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1/2} \quad (5.1)$$

which is the estimated cross-correlation between the vectors $\hat{R}_{yy}^{-1/2} \mathbf{y}(t)$ and $\hat{R}_{\phi\phi}^{-1/2} \phi(t, \boldsymbol{\eta})$, or the estimated *coherence matrix* between $\mathbf{y}(t)$ and $\phi(t, \boldsymbol{\eta})$; see [36, Sect. VII]. For simplicity of notation, we omit the dependence of \mathcal{P} and $\hat{C}_{y\phi}$ on $\boldsymbol{\eta}$. Consider the singular value decomposition (SVD) of $\hat{C}_{y\phi}$

$$\hat{C}_{y\phi} = \hat{U} \hat{\Lambda} \hat{V}^* \quad (5.2a)$$

$$\hat{U}^* \hat{U} = \hat{U} \hat{U}^* = I_m, \quad \hat{V}^* \hat{V} = \hat{V} \hat{V}^* = I_d \quad (5.2b)$$

$$\hat{\Lambda} = \begin{cases} [\hat{\Lambda}(m), 0], & m < d \\ [\hat{\Lambda}(d), 0]^T, & m > d \end{cases} \quad (5.2c)$$

$$\hat{\Lambda}(m) = \text{diag}\{\hat{\lambda}(1), \hat{\lambda}(2), \dots, \hat{\lambda}(m)\} \quad (5.2d)$$

where $\hat{\lambda}(1) \geq \hat{\lambda}(2) \geq \dots \geq \hat{\lambda}(\min(m, d)) \geq 0$. Again, for notational simplicity, we omit the dependence of the above quantities on $\boldsymbol{\eta}$. Note that $0 \leq \hat{\lambda}^2(i) \leq 1$ for $i = 1, 2, \dots, \min(m, d)$ because $0 \leq \hat{C}_{y\phi}^* \hat{C}_{y\phi} \leq I_d$. Then, after simple manipulations, we find that $1/[1 - \hat{\lambda}^2(i)]$, $i = 1, 2, \dots, \min(m, d)$ are the generalized eigenvalues of $\hat{S}_{y|\phi}^{-1}$ and \hat{R}_{yy}^{-1} , sorted in nonincreasing order [since $1/(1-x)$ is an increasing function of x for $0 \leq x \leq 1$]. In addition, note that $\hat{\lambda}(i)$ are the estimated *canonical correlations*, i.e., $\hat{\lambda}(i)$ is the cosine of the angle between the

i th components of estimated *canonical coordinates* of the data and basis functions, which is defined (see [36]) as

$$\hat{\mathbf{y}}_c(t, \boldsymbol{\eta}) = \hat{U}(r)^* \hat{R}_{yy}^{-1/2} \mathbf{y}(t) \quad (5.3a)$$

$$\hat{\boldsymbol{\phi}}_c(t, \boldsymbol{\eta}) = \hat{V}(r)^* \hat{R}_{\phi\phi}^{-1/2} \boldsymbol{\phi}(t, \boldsymbol{\eta}) \quad (5.3b)$$

for $t = 1, 2, \dots, N$. Using the Poincaré separation theorem [38, pp. 64–65], maximizing (3.5) with respect to the unstructured array response $A(\boldsymbol{\theta}) = A$ [of size $m \times r$, and having full rank $r \leq \min(d, m)$] yields

$$l_u(\boldsymbol{\eta}) = \prod_{j=1}^r \frac{1}{1 - \hat{\lambda}^2(j)} \quad (5.4)$$

see also Appendix B. Interestingly, for a full-rank channel, i.e., $r = \min(d, m)$, (5.4) reduces to

$$l_u(\boldsymbol{\eta}) = \frac{|\hat{R}_{\phi\phi}|}{|\hat{S}_{\phi|y}|} = \frac{|\hat{R}_{yy}|}{|\hat{S}_{y|\phi}|} \quad (5.5)$$

which follows from the fact that the determinant of a matrix equals to the product of its eigenvalues. It can also be obtained from (3.3) and (3.5) by replacing $A(\boldsymbol{\theta})$ with I_m . Based on [41, Th. 2.1] and [36, Sect. IV], we can then view $\log[l_u(\boldsymbol{\eta})]$ [with $l_u(\boldsymbol{\eta})$ given in (5.5)] as a measure of (estimated) mutual information between $\mathbf{y}(t)$ and $\boldsymbol{\phi}(t, \boldsymbol{\eta})$.

We now present the ML estimate of the reduced-rank channel matrix $H = AX$. Note that it is not possible to find unique estimates of A and X , but their product is unique; see Appendix B. First, we adopt the following notation: $\hat{U}(r)$ and $\hat{V}(r)$ are the matrices containing the first r columns of \hat{U} and \hat{V} , respectively. In Appendix B, we derive the ML estimates of the channel and the noise covariance for known $\boldsymbol{\eta}$

$$\hat{H}(\boldsymbol{\eta}) = \hat{R}_{yy}^{1/2} \hat{U}(r) \hat{U}(r)^* \hat{R}_{yy}^{-1/2} \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \quad (5.6a)$$

$$= \mathcal{P} \hat{V}(r) \hat{V}(r)^* \hat{R}_{\phi\phi}^{-1/2} \quad (5.6b)$$

$$= \hat{R}_{yy}^{1/2} \hat{U}(r) \hat{\Lambda}(r) \hat{V}(r)^* \hat{R}_{\phi\phi}^{-1/2} \quad (5.6c)$$

$$\begin{aligned} \hat{\Sigma}(\boldsymbol{\eta}) &= \hat{R}_{yy} - \mathcal{P} \hat{V}(r) \hat{V}(r)^* \mathcal{P}^* \\ &= \hat{R}_{yy} - \hat{R}_{yy}^{1/2} \hat{U}(r) \hat{\Lambda}^2(r) \hat{U}(r)^* \hat{R}_{yy}^{1/2}. \end{aligned} \quad (5.6d)$$

The above three expressions for the ML estimate of H are equivalent, but useful. Equation (5.6a) represents the channel estimate computed using only the left singular vectors of $\hat{C}_{y\phi}$; similarly, (5.6b) (which is derived for the real case in [23, (34)]) uses only the right singular vectors of $\hat{C}_{y\phi}$, whereas (5.6c) uses the reduced-rank representation $\hat{U}(r) \hat{\Lambda}(r) \hat{V}(r)^*$ of $\hat{C}_{y\phi}$. Methods for efficiently computing (5.6) when $r \ll \min(m, d)$ are derived in [42].

When $r = \min(m, d)$, i.e., the channel matrix H is of full rank, its ML estimate is simply the estimated Wiener filter $\hat{H}(\boldsymbol{\eta}) = \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1}$, and the ML estimate of the noise covariance simplifies to $\hat{\Sigma}(\boldsymbol{\eta}) = \hat{S}_{y|\phi} = \hat{R}_{yy} - \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \hat{R}_{y\phi}^*$ [which follow from (3.1) by replacing $A(\boldsymbol{\theta})$ with I_m]. The reduced-rank channel estimates in (5.6a) can then be viewed as

a product of an idempotent matrix $\hat{R}_{yy}^{1/2} \hat{U}(r) \hat{U}(r)^* \hat{R}_{yy}^{-1/2}$ and the full-rank channel estimate.

If $\boldsymbol{\eta}$ is unknown, its ML estimate $\hat{\boldsymbol{\eta}}$ is obtained by maximizing the concentrated likelihood in (5.4). To find the ML estimates of H and Σ , replace $\boldsymbol{\eta}$ in (5.6) by $\hat{\boldsymbol{\eta}}$. Note that the following holds: $[Y - \hat{H}(\hat{\boldsymbol{\eta}}) \Phi(\hat{\boldsymbol{\eta}})] \cdot \Phi(\hat{\boldsymbol{\eta}})^* \hat{H}(\hat{\boldsymbol{\eta}})^* = 0$, which can be viewed as orthogonality between fit and residuals: an extension of a well-known univariate result.

For a channel with negligible time-delay and Doppler spreads, i.e., when $L = K = 1$ in (2.5), the channel response matrix reduces to a vector, and the concentrated likelihood for the unstructured array is obtained by replacing $A(\boldsymbol{\theta})$ and $\boldsymbol{\phi}(t, \boldsymbol{\eta})$ with I_m and $\exp(j\omega_D t) s(t - \tau)$ in (3.3); see [25] and [40]. In [25] and [40], we show that the resulting concentrated likelihood function can be viewed as a multivariate extension of the *spectrogram* [43, p. 95].

A. Effects of Time-Delay and Doppler Spread Discretizations

We investigate the effects of time-delay and Doppler spread discretizations on the parameter estimates.

In general, the channel and spatial noise covariance estimates in (5.6) are affected both by the noise and basis-function discretizations. To focus on the discretization effects, we assume here that a large number of snapshots N is available (i.e., $N \rightarrow \infty$) and that the signal-to-noise ratio (SNR) is high; as a result, the noise effects can be neglected. Thus, the imperfections in the estimation will be due only to the time-delay and Doppler spread discretizations in (2.4)–(2.6). We introduce the following notation for the “true” model in (2.1): $\mathbf{y}(t) = H_T \boldsymbol{\Phi}_T(t, \boldsymbol{\eta}_T) + \mathbf{e}(t)$, where $H_T = [x_{T1} \mathbf{a}(\theta_{T1}) \cdots x_{TP} \mathbf{a}(\theta_{TP})]$ is the “true” channel matrix, and $\boldsymbol{\Phi}_T(t, \boldsymbol{\eta}_T) = [\exp(j\omega_{DT1} t) s(t - \tau_{T1}), \dots, \exp(j\omega_{DTP} t) s(t - \tau_{TP})]^T$ are the “true” basis functions. In addition, define $\Phi_T(\boldsymbol{\eta}_T) = [\boldsymbol{\Phi}_T(1, \boldsymbol{\eta}_T) \cdots \boldsymbol{\Phi}_T(N, \boldsymbol{\eta}_T)]$ and $\boldsymbol{\eta}_T = [\omega_{DT1}, \dots, \omega_{DTP}, \tau_{T1}, \dots, \tau_{TP}]^T$. Note that we use the subscript T to differentiate between the true model and the rectangular-grid sampling schemes in (2.4)–(2.6) that we use to fit it. Then, applying the unstructured channel estimation for known $\boldsymbol{\eta}$ in (5.6), we obtain

$$\hat{H} = H_T \hat{R}_{\phi\phi_T}^* \hat{R}_{\phi\phi}^{-1} \quad (5.7a)$$

$$\hat{\Sigma} = \Sigma + H_T \hat{S}_{\phi_T|\phi} H_T^* \quad (5.7b)$$

where $\hat{S}_{\phi_T|\phi} = \hat{R}_{\phi_T\phi_T} - \hat{R}_{\phi\phi_T}^* \hat{R}_{\phi\phi}^{-1} \hat{R}_{\phi\phi_T}$, $\hat{R}_{\phi_T\phi_T} = (1/N) \cdot \Phi_T(\boldsymbol{\eta}_T) \Phi_T(\boldsymbol{\eta}_T)^*$, and $\hat{R}_{\phi\phi_T} = \hat{R}_{\phi_T\phi}^* = (1/N) \cdot \Phi(\boldsymbol{\eta}) \Phi_T(\boldsymbol{\eta}_T)^*$. The above expressions show the distortion in estimating the channel and spatial noise covariance matrices due to the discretization effects. The second term in (5.7b), clearly due to the modeling error, is positive semi-definite; thus, $\hat{\Sigma} \geq \Sigma$, which is intuitively appealing. If we use the “true” model, i.e., $\Phi_T(\boldsymbol{\eta}_T) = \Phi(\boldsymbol{\eta})$, then the channel and noise covariance would be perfectly estimated, i.e., $\hat{H} = H_T$ and $\hat{\Sigma} = \Sigma$. Further, observe that the concentrated likelihood (5.4) becomes

$$l_u(\boldsymbol{\eta}) = \frac{|\Sigma + H_T \hat{R}_{\phi_T\phi_T} H_T^*|}{|\Sigma + H_T \hat{S}_{\phi_T|\phi} H_T^*|} \quad (5.8)$$

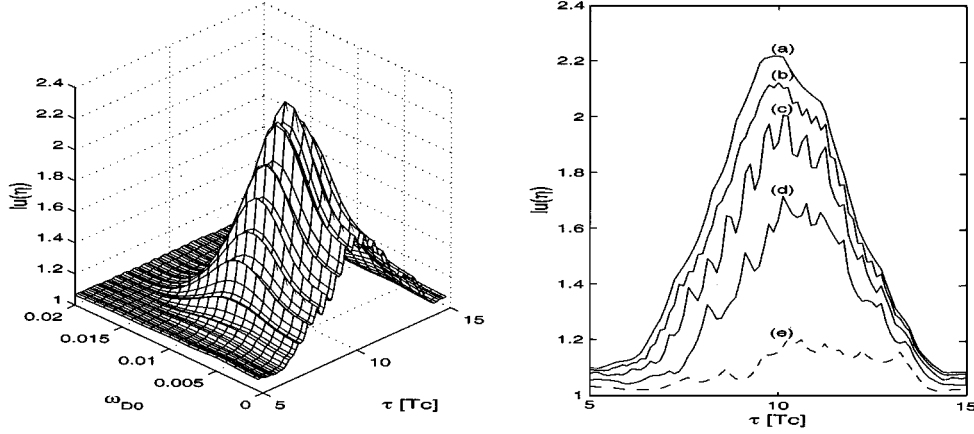


Fig. 2. Concentrated likelihood as a function of (Left) the “global” time delay τ (in multiples of the chip duration T_c) and Doppler parameter ω_{D0} for $K = 2$ discrete Doppler frequencies and $\Delta\tau = T_c/2$. (Right) τ for $K = 2$ and (a) $\Delta\tau = T_c/8$, (b) $\Delta\tau = T_c/4$, (c) $\Delta\tau = T_c/2$, (d) $\Delta\tau = T_c$, and (e) neglected Doppler effect and $\Delta\tau = T_c/2$.

which, for the “true” model [i.e., $\Phi_T(\boldsymbol{\eta}_T) = \Phi(\boldsymbol{\eta})$ and, therefore, $\hat{H} = H_T = H$ and $\hat{\Sigma} = \Sigma$], simplifies to

$$l_u(\boldsymbol{\eta}) = \frac{|\Sigma + H\hat{R}_{\phi\phi}H^*|}{|\Sigma|} = |I_m + \Sigma^{-1/2}H\hat{R}_{\phi\phi}H^*\Sigma^{-1/2}|. \quad (5.9)$$

In the case of transmit–receive antenna array systems in slow flat-fading environment (see Section II-B1), we can choose the basis functions $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = [s_1(t - \tau), s_2(t - \tau), \dots, s_d(t - \tau)]^T$ and therefore design $\hat{R}_{\phi\phi}$, as desired. If the transmitter has knowledge of the channel H and noise covariance Σ , it can transmit waveforms having $\hat{R}_{\phi\phi}$ such that (5.9) is maximized using “water filling,” as in, e.g., [44] [then, the maximized (5.9) can be viewed as a measure of capacity of the channel].

Numerical Example 1—Concentrated Likelihood Function: In this numerical example, we study the concentrated likelihood function in (5.8). We demonstrate the feasibility of estimating the global time delay τ and Doppler parameter ω_{D0} [using the basis-function model in (2.4), (2.5a), and (2.6)], as well as benefits of accounting for the Doppler effect. We have generated a known signal $s(t) = \sum_{k=1}^{100} b_k s_0(t - (k-1)T_c)$ containing 100 chips having rectangular pulse shapes [i.e., $s_0(t) = u(t) - u(t - T_c)$, where $u(t)$ is the Heaviside step function], where the amplitudes b_k are independent, identically distributed (i.i.d.) random variables, taking values ± 1 with equal probability. Let the sampling period be $T_c/8$; thus, the number of samples is $N = 800$. We use a five-element uniform linear array with half-wavelength spacing. The received signal was generated by using the “true” model in (2.1) with $P = 20$ paths. The time delays τ_{Tp} are assumed to be uniformly distributed between $t = 10T_c$ and $t = 14T_c$ (i.e., between samples 80 and 112). The complex amplitudes x_{Tp} are i.i.d. zero-mean complex Gaussian random variables having unit variance. The DOAs θ_{Tp} are uniformly distributed between 0 and 2π , and the corresponding Doppler shifts were computed as $\omega_{DTp} = \omega_{D0} \cos(\theta_{Tp})$, where $\omega_{D0} = 0.01$, which is quite large (guaranteeing a fast-fading environment, which requires that the product of the symbol duration and the Doppler spread is larger than 0.01; see [26]). The spatial noise covariance is

assumed to be of the form $\Sigma = \sigma^2 \cdot (0.9I_5 + 0.1 \cdot 1_{5 \times 5})$, where $1_{5 \times 5}$ denotes the 5×5 matrix of ones. To ensure high SNR [needed for the formulas (5.7) and (5.8) to be valid], we choose the noise level $\sigma^2 = 0.1$. Note that we could choose a larger σ^2 , but then, for high SNR, we would need a signal of longer duration for channel estimation. To fit the received signal, we have used the rectangular-grid basis-function model (2.4), with $\boldsymbol{\phi}_\tau(t, \boldsymbol{\eta}_\tau)$ in (2.5a), whereas the Doppler basis functions $\boldsymbol{\phi}_{\omega_D}(t, \boldsymbol{\eta}_{\omega_D})$ are chosen according to the Jakes model (2.6) with only $K = 2$ discrete Doppler shift frequencies, i.e., $\boldsymbol{\phi}_{\omega_D}(t, \omega_{D0}) = [\exp(-j\omega_{D0}t), \exp(j\omega_{D0}t)]^T$.

In Fig. 2 (left), we show the concentrated likelihood in (5.8) as a function of the “global” time delay τ and Doppler parameter ω_{D0} , where the time-delay sampling period is $\Delta\tau = T_c/2$, and the Doppler effect is modeled using the Jakes model (2.6) with only $K = 2$ discrete Doppler frequencies. As expected, this function has a peak at values of τ and ω_{D0} close to their true values $10T_c$ and 0.01.

In Fig. 2 (right), the curves (a)–(d) show the concentrated likelihood in (5.8) as a function of τ for the true value of ω_{D0} (i.e., $\omega_{D0} = 10^{-2}$) and the time delay sampling periods $\Delta\tau \in \{T_c/8, T_c/4, T_c/2, T_c\}$ [where the corresponding sizes of $\boldsymbol{\phi}_\tau(t, \tau)$ are $L = 4T_c/\Delta\tau \in \{32, 16, 8, 4\}$], and $K = 2$ discrete Doppler frequencies in (2.6). Curve (e) shows the concentrated likelihood (5.8) for $\Delta\tau = T_c/2$ when the Doppler effect is neglected [i.e., $\boldsymbol{\phi}_{\omega_D}(t, \omega_{D0}) = 1$, and therefore, $\boldsymbol{\phi}(t, \boldsymbol{\eta}) = \boldsymbol{\phi}_\tau(t, \tau)$]. Note that the above functions have local maxima at multiples of $\Delta\tau$ and are relatively flat, i.e., the “global” time delay does not need to be estimated accurately to capture most of the signal energy with the discretized basis functions. In addition, denser discretization yields higher and sharper concentrated likelihood, as expected. More importantly, accounting for the Doppler effect results in a significantly higher and sharper concentrated likelihood compared with the case when it is neglected. [Note the difference between the concentrated likelihood functions (d) and (e), which employ the same number of basis functions $d = 8$.] Higher values of the concentrated likelihood also enable better detection performance, which is demonstrated by Example 2 in Section VII-A2.

Thus, the proposed discretized basis-function models are sufficiently accurate for communications. However, they may be unacceptable in radar, where the accuracy of estimating time delays and Doppler shifts is crucial; then, we should use the ‘‘true’’ model with $\phi(t, \boldsymbol{\eta}) = \phi_T(t, \boldsymbol{\eta}_T)$.

VI. SUPPRESSION OF SPATIAL AND TEMPORAL INTERFERENCE

We develop methods to filter out spatial and temporal interference from the received data in the presence of spatially correlated noise with unknown covariance. We show that the concentrated likelihood function expressions in (3.3) and (3.5) can be suitably decomposed into a product of two terms: the concentrated likelihood for the received data and interference only and the concentrated likelihood for the filtered data and signal (both having been filtered to suppress the interference). These results may be useful for multiuser interference suppression, which we briefly discuss in Section VI-B. Throughout this section, we will use the following notation for the concentrated likelihood function in (3.3) and (3.5): $l(\boldsymbol{\theta}, \boldsymbol{\eta}|Y, A(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta}))$, meaning that it is computed for the data set Y , array response $A(\boldsymbol{\theta})$, and basis-function matrix $\Phi(\boldsymbol{\eta})$. This notation allows us to state the obtained results compactly.

A. Spatial Interference

Consider first the case where the array response is $A(\boldsymbol{\theta}) = [A_i, A_s(\boldsymbol{\theta})]$, where we have the following.

- A_i is the known array response of an interference.
- $A_s(\boldsymbol{\theta})$ is the array response matrix of useful signal.

Then, the concentrated likelihood function in (3.5) can be written as

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}|Y, A(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta})) = \frac{\begin{vmatrix} A_i^* \hat{S}_{y|\phi}^{-1} A_i & A_i^* \hat{S}_{y|\phi}^{-1} A_s(\boldsymbol{\theta}) \\ A_s(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A_i & A_s(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A_s(\boldsymbol{\theta}) \end{vmatrix}}{\begin{vmatrix} A_i^* \hat{R}_{yy}^{-1} A_i & A_i^* \hat{R}_{yy}^{-1} A_s(\boldsymbol{\theta}) \\ A_s(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A_i & A_s(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A_s(\boldsymbol{\theta}) \end{vmatrix}} \quad (6.1a)$$

$$= \frac{|A_i^* \hat{S}_{y|\phi}^{-1} A_i|}{|A_i^* \hat{R}_{yy}^{-1} A_i|} \cdot \frac{|A_s(\boldsymbol{\theta})^* [\hat{S}_{y|\phi}^{-1} - \hat{S}_{y|\phi}^{-1} A_i (A_i^* \hat{S}_{y|\phi}^{-1} A_i)^{-1} A_i^* \hat{S}_{y|\phi}^{-1}] A_s(\boldsymbol{\theta})|}{|A_s(\boldsymbol{\theta})^* [\hat{R}_{yy}^{-1} - \hat{R}_{yy}^{-1} A_i (A_i^* \hat{R}_{yy}^{-1} A_i)^{-1} A_i^* \hat{R}_{yy}^{-1}] A_s(\boldsymbol{\theta})|} \quad (6.1b)$$

$$= \frac{|A_i^* \hat{S}_{y|\phi}^{-1} A_i|}{|A_i^* \hat{R}_{yy}^{-1} A_i|} \cdot \frac{|A_s(\boldsymbol{\theta})^* \Psi_{A_i} (\Psi_{A_i}^* \hat{S}_{y|\phi} \Psi_{A_i})^{-1} \Psi_{A_i}^* A_s(\boldsymbol{\theta})|}{|A_s(\boldsymbol{\theta})^* \Psi_{A_i} (\Psi_{A_i}^* \hat{R}_{yy} \Psi_{A_i})^{-1} \Psi_{A_i}^* A_s(\boldsymbol{\theta})|} \quad (6.1c)$$

where Ψ_{A_i} is an arbitrary full-rank matrix such that its columns span the space orthogonal to the column space of A_i ; thus, $\Psi_{A_i}^* A_i = 0$. Here, (6.1b) follows by applying the formula for the determinant of a partitioned matrix [34, result v, p. 8] to

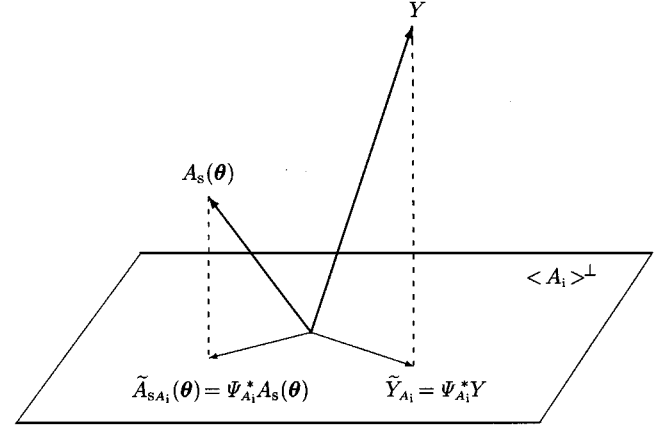


Fig. 3. Spatial interference is filtered out by projecting the measurements Y and the array response of the useful signal $A_s(\boldsymbol{\theta})$ onto the subspace orthogonal to the spatial interference.

(6.1a), and (6.1c) follows then by using [35, Lemma 1] (see also [38, p. 77]), stating

$$\Sigma^{-1} - \Sigma^{-1} A_i (A_i^* \Sigma^{-1} A_i)^{-1} A_i^* \Sigma^{-1} = \Psi_{A_i} (\Psi_{A_i}^* \Sigma \Psi_{A_i})^{-1} \Psi_{A_i}^* \quad (6.2)$$

We can, for example, choose Ψ_{A_i} to be the matrix whose columns are the $m - \text{rank}(A_i)$ eigenvectors of $\Pi_{A_i}^\perp = I_m - A_i (A_i^* A_i)^{-1} A_i^*$ with corresponding eigenvalues equal to 1; then, $\Pi_{A_i}^\perp = \Psi_{A_i} \Psi_{A_i}^*$.

Equation (6.1c) shows that the concentrated likelihood in (3.5) can be decomposed into a product of the concentrated likelihood function for the data set Y and interference array response A_i only, and the concentrated likelihood function for the filtered data set $\tilde{Y}_{A_i} = \Psi_{A_i}^* Y$ and filtered array response of the useful signal $\tilde{A}_{s,A_i}(\boldsymbol{\theta}) = \Psi_{A_i}^* A_s(\boldsymbol{\theta})$; see Fig. 3. This result can be compactly stated as $l(\boldsymbol{\theta}, \boldsymbol{\eta}|Y, A(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta})) = l(\boldsymbol{\eta}|Y, A_i, \Phi(\boldsymbol{\eta})) \cdot l(\boldsymbol{\theta}, \boldsymbol{\eta}|\tilde{Y}_{A_i}, \tilde{A}_{s,A_i}(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta}))$. Note that the first term in the above product does not depend on $A_s(\boldsymbol{\theta})$; hence, it is also independent of the DOA parameters $\boldsymbol{\theta}$. As a result, to find the ML estimate of $\boldsymbol{\theta}$, we need to maximize the second term only, and for that purpose, it is sufficient to use the filtered data and array response matrices \tilde{Y}_{A_i} and $\tilde{A}_{s,A_i}(\boldsymbol{\theta})$.

The above result can be used to implement a (suboptimal) algorithm for sequential DOA estimation, where previously estimated DOAs would be used to construct the array response of the interference A_i .

B. Temporal Interference

Consider now the case where the basis function matrix is $\Phi(\boldsymbol{\eta}) = [\Phi_i^T, \Phi_s(\boldsymbol{\eta})^T]^T$, where we have the following.

- Φ_i is the known basis-function matrix of an interference.
- $\Phi_s(\boldsymbol{\eta})$ is the basis-function matrix of useful signal.

Define the projection matrices spanning the column spaces of Y^* and Φ_i^* as $\Pi_{Y^*} = (1/N) \cdot Y^* \hat{R}_{yy}^{-1} Y = Y^* (Y Y^*)^{-1} Y$ and $\Pi_{\Phi_i^*} = \Phi_i^* (\Phi_i \Phi_i^*)^{-1} \Phi_i$ and the corresponding complementary projection matrices as $\Pi_{Y^*}^\perp = I_N - \Pi_{Y^*}$ and $\Pi_{\Phi_i^*}^\perp = I_N - \Pi_{\Phi_i^*}$. In addition, define $L_A = I_N - (1/N) \cdot Y^* W Y = \Pi_{Y^*}^\perp + Y^* (Y Y^*)^{-1} A(\boldsymbol{\theta}) \cdot [A(\boldsymbol{\theta})^* (Y Y^*)^{-1} A(\boldsymbol{\theta})]^{-1} \cdot A(\boldsymbol{\theta})^* (Y Y^*)^{-1} Y$. Let Ψ_A be a matrix of size $m \times (m - r)$

such that $\Psi_A^* A(\boldsymbol{\theta}) = 0$, i.e., Ψ_A spans the space orthogonal to the column space of $A(\boldsymbol{\theta})$. Here, L_A and Ψ_A are functions of $\boldsymbol{\theta}$, but we omit this dependence to simplify the notation. In addition, define Ψ_{Φ_i} as the matrix whose columns are the $N - \text{rank}(\Phi_i)$ eigenvectors of $\Pi_{\Phi_i}^\perp$ with corresponding eigenvalues equal to 1; then, $\Pi_{\Phi_i}^\perp = \Psi_{\Phi_i} \Psi_{\Phi_i}^*$. Using the identity (6.2) [where Σ , A_i , and Ψ_{A_i} are replaced with $Y Y^*$, $A(\boldsymbol{\theta})$, and Ψ_A], we obtain

$$\Pi_{Y^*} - Y^*(Y Y^*)^{-1} A(\boldsymbol{\theta}) [A(\boldsymbol{\theta})^* (Y Y^*)^{-1} A(\boldsymbol{\theta})]^{-1} \cdot A(\boldsymbol{\theta})^* (Y Y^*)^{-1} Y = Y^* \Psi_A (\Psi_A^* Y Y^* \Psi_A)^{-1} \Psi_A^* Y$$

and then, $I_N - L_A = Y^* \Psi_A (\Psi_A^* Y Y^* \Psi_A)^{-1} \Psi_A^* Y = \Pi_{Y^* \Psi_A}$, which is the projection matrix onto the column space of $Y^* \Psi_A$. Further, the following identities easily follow by applying the projection-matrix decomposition formula:²

$\Pi_{[Y^*, \Phi_i^*]} = \Pi_{Y^*} + \Pi_{Y^*}^\perp \Phi_i^* (\Phi_i \Pi_{Y^*}^\perp \Phi_i^*)^{-1} \Phi_i \Pi_{Y^*}^\perp = \Pi_{\Phi_i^*} + \Pi_{\Phi_i^*}^\perp Y^* (Y \Pi_{\Phi_i^*}^\perp Y^*)^{-1} Y \Pi_{\Phi_i^*}^\perp$, and similarly, $\Pi_{[Y^* \Psi_A, \Phi_i^*]} = I_N - L_A + L_A \Phi_i^* (\Phi_i L_A \Phi_i^*)^{-1} \Phi_i L_A = \Pi_{\Phi_i^*} + \Pi_{\Phi_i^*}^\perp Y^* \Psi_A (\Psi_A^* Y \Pi_{\Phi_i^*}^\perp Y^* \Psi_A)^{-1} \Psi_A^* Y \Pi_{\Phi_i^*}^\perp$. Applying the formula for determinant of a partitioned matrix and using the above definitions and identities, the concentrated likelihood function in (3.3) can be written as in the equation shown at the bottom of the page. This equation shows that the concentrated likelihood in (3.3) can be decomposed into a product of the concentrated likelihood function for the data set Y and interference basis-function matrix Φ_i only, and the concentrated likelihood function for the filtered data set $\tilde{Y}_{\Phi_i} = Y \Psi_{\Phi_i}$ and filtered basis-function matrix of the useful signal $\tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta}) = \Phi_s(\boldsymbol{\eta}) \Psi_{\Phi_i}$; see Fig. 4. This decomposition can be written as $l(\boldsymbol{\theta}, \boldsymbol{\eta} | Y, A(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta})) = l(\boldsymbol{\theta} | Y, A(\boldsymbol{\theta}), \Phi_i) \cdot l(\boldsymbol{\theta}, \boldsymbol{\eta} | \tilde{Y}_{\Phi_i}, A(\boldsymbol{\theta}), \tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta}))$. The first term in the above product does not depend on $\Phi_s(\boldsymbol{\eta})$; hence it is also independent of $\boldsymbol{\eta}$. Thus, to find the ML estimate of $\boldsymbol{\eta}$, we need to maximize the second term only.

Note that the above decomposition also applies to the concentrated likelihood function for full-rank unstructured channel

²For the proof of this result, see, for example, [45, Th. A.45].

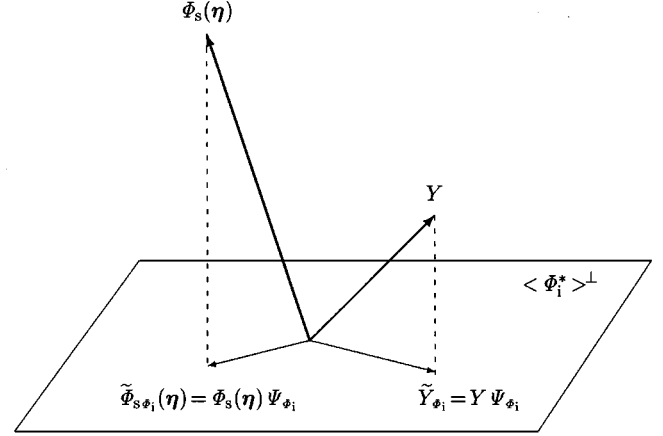


Fig. 4. Temporal interference is filtered out by projecting the measurements Y and the basis functions of the useful signal $\Phi_s(\boldsymbol{\eta})$ onto the subspace orthogonal to the temporal interference.

(5.5) since it follows from (3.3) by setting $A(\boldsymbol{\theta}) = I_m$. Thus, (5.5) can be decomposed as

$$l(\boldsymbol{\eta} | Y, I_m, \Phi(\boldsymbol{\eta})) = l(Y, I_m, \Phi_i) \cdot l(\boldsymbol{\eta} | \tilde{Y}_{\Phi_i}, I_m, \tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta})) \\ = \frac{|\Phi_i \Phi_i^*|}{|\Phi_i \Pi_{Y^*}^\perp \Phi_i^*|} \cdot \frac{|\tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta}) \tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta})^*|}{|\tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta}) \Pi_{(\tilde{Y}_{\Phi_i})^*}^\perp \tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta})^*|}.$$

The second term in the above expression can be used to implement the noncoherent concentrated-likelihood receiver (as is discussed in Section VII-B1) with multiuser interference suppression capability. For one basis function [i.e., $d = 1$, implying that $\tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta})$ is a row vector] and one receive antenna (i.e., $m = 1$, implying that \tilde{Y}_{Φ_i} is a row vector), and after the monotonic transformation $1 - 1/l(\boldsymbol{\eta} | \tilde{Y}_{\Phi_i}, I_m, \tilde{\Phi}_{s\Phi_i}(\boldsymbol{\eta}))$, this term reduces to the noncoherent detector in [46].

VII. SYMBOL DETECTION

Recently, several authors have proposed incorporating *known* spatial noise covariance into the receiver design. In [47]–[49], maximum likelihood sequence estimators (MLSE) accounting

$$l(\boldsymbol{\theta}, \boldsymbol{\eta} | Y, A(\boldsymbol{\theta}), \Phi(\boldsymbol{\eta})) = \frac{\begin{vmatrix} \Phi_i L_A \Phi_i^* & \Phi_i L_A \Phi_s(\boldsymbol{\eta})^* \\ \Phi_s(\boldsymbol{\eta}) L_A \Phi_i^* & \Phi_s(\boldsymbol{\eta}) L_A \Phi_s(\boldsymbol{\eta})^* \end{vmatrix}}{\begin{vmatrix} \Phi_i \Pi_{Y^*}^\perp \Phi_i^* & \Phi_i \Pi_{Y^*}^\perp \Phi_s(\boldsymbol{\eta})^* \\ \Phi_s(\boldsymbol{\eta}) \Pi_{Y^*}^\perp \Phi_i^* & \Phi_s(\boldsymbol{\eta}) \Pi_{Y^*}^\perp \Phi_s(\boldsymbol{\eta})^* \end{vmatrix}} \\ = \frac{|\Phi_i L_A \Phi_i^*|}{|\Phi_i \Pi_{Y^*}^\perp \Phi_i^*|} \cdot \frac{|\Phi_s(\boldsymbol{\eta}) \cdot [L_A - L_A \Phi_i^* (\Phi_i L_A \Phi_i^*)^{-1} \Phi_i L_A] \cdot \Phi_s(\boldsymbol{\eta})^*|}{|\Phi_s(\boldsymbol{\eta}) \cdot [\Pi_{Y^*}^\perp - \Pi_{Y^*}^\perp \Phi_i^* (\Phi_i \Pi_{Y^*}^\perp \Phi_i^*)^{-1} \Phi_i \Pi_{Y^*}^\perp] \cdot \Phi_s(\boldsymbol{\eta})^*|} \\ = \frac{|\Phi_i L_A \Phi_i^*|}{|\Phi_i \Pi_{Y^*}^\perp \Phi_i^*|} \cdot \frac{|\Phi_s(\boldsymbol{\eta}) \cdot [\Pi_{\Phi_i^*}^\perp - \Pi_{\Phi_i^*}^\perp Y^* \Psi_A (\Psi_A^* Y \Pi_{\Phi_i^*}^\perp Y^* \Psi_A)^{-1} \Psi_A^* Y \Pi_{\Phi_i^*}^\perp] \cdot \Phi_s(\boldsymbol{\eta})^*|}{|\Phi_s(\boldsymbol{\eta}) \cdot [\Pi_{\Phi_i^*}^\perp - \Pi_{\Phi_i^*}^\perp Y^* (Y \Pi_{\Phi_i^*}^\perp Y^*)^{-1} Y \Pi_{\Phi_i^*}^\perp] \cdot \Phi_s(\boldsymbol{\eta})^*|} \\ = \frac{|\Phi_i L_A \Phi_i^*|}{|\Phi_i \Pi_{Y^*}^\perp \Phi_i^*|} \cdot \frac{|\Phi_s(\boldsymbol{\eta}) \Psi_{\Phi_i} \cdot [I_N - \Psi_{\Phi_i}^* Y^* \Psi_A (\Psi_A^* Y \Psi_{\Phi_i} \Psi_{\Phi_i}^* Y^* \Psi_A)^{-1} \Psi_A^* Y \Psi_{\Phi_i}] \cdot \Psi_{\Phi_i}^* \Phi_s(\boldsymbol{\eta})^*|}{|\Phi_s(\boldsymbol{\eta}) \Psi_{\Phi_i} \cdot [I_N - \Psi_{\Phi_i}^* Y^* (Y \Psi_{\Phi_i} \Psi_{\Phi_i}^* Y^*)^{-1} Y \Psi_{\Phi_i}] \cdot \Psi_{\Phi_i}^* \Phi_s(\boldsymbol{\eta})^*|}$$

for the known spatial noise covariance were proposed; lately, they have been extended to account for temporally correlated CCI following the vector autoregressive (VAR) model [50]. Here, we derive coherent matched-filter and concentrated-likelihood receivers that utilize *estimated* channel response and spatial noise covariance matrices (using results of Sections III and V) *as well as* the basis-function signal models from Section II-B. We also discuss the relationship between these receivers.

Define the data sets Y_K (of size $m \times N_K$) and Y_R (of size $m \times N_R$) containing the known (training or previously detected) sequence and unknown symbol to be detected, respectively. Then, for J symbols, we associate possible basis-function matrices with the symbol to be detected as Φ_{Rj} for $j = 0, 1, 2, \dots, J-1$. In the following discussion, we assume that the channel synchronization has been done, i.e., $\boldsymbol{\eta}$ is known (hence, for notational simplicity, we also omit functional dependence on $\boldsymbol{\eta}$ throughout this section).

A. Coherent Matched-Filter Receiver

We derive a coherent matched-filter receiver that accounts for spatial noise covariance and analyze its performance. In Section VII-A1, we derive analytical expressions for error probability when binary signaling is employed and show that the proposed receiver outperforms the coherent receiver that does not take into account the spatial noise covariance. In Section VII-A2, we analyze the impact of the basis-function discretization (proposed in Section II-B) on the error probability.

We use the training data Y_K and ML methods in Sections III and V to estimate the channel and spatial covariance matrices (denoted by \hat{H}_K and $\hat{\Sigma}_K$, respectively) and plug them into the likelihood function for the unknown symbol (using Y_R). Then, the coherent matched-filter receiver is based on choosing the j that maximizes the decision statistic

$$\varpi_j = 2 \operatorname{Re}\{\operatorname{tr}[Y_R^* \hat{\Sigma}_K^{-1} \hat{H}_K \Phi_{Rj}]\} - \operatorname{tr}[\Phi_{Rj}^* \hat{H}_K^* \hat{\Sigma}_K^{-1} \hat{H}_K \Phi_{Rj}] \quad (7.1)$$

for $j = 0, 1, \dots, J-1$. For the binary case (i.e., $J = 2$), the decision is made by computing

$$\begin{aligned} & \operatorname{sign}(\varpi_1 - \varpi_0) \\ &= \operatorname{sign}(\operatorname{Re}\{\operatorname{tr}[2Y_R^* \hat{\Sigma}_K^{-1} \hat{H}_K (\Phi_{R1} - \Phi_{R0}) \\ & \quad - (\Phi_{R1} + \Phi_{R0})^* \hat{H}_K^* \hat{\Sigma}_K^{-1} \hat{H}_K (\Phi_{R1} - \Phi_{R0})]\}) \end{aligned} \quad (7.2)$$

where the second term disappears if antipodal signaling is employed and intersymbol interference (ISI) is negligible (then $\Phi_{R1} = -\Phi_{R0}$) or 1 and 0 bases have the same covariance matrices (i.e., $\Phi_{R0} \Phi_{R0}^* = \Phi_{R1} \Phi_{R1}^*$). Then, the detector is simply $\operatorname{sign}(\operatorname{Re}\{\operatorname{tr}[Y_R^* \hat{\Sigma}_K^{-1} \hat{H}_K (\Phi_{R1} - \Phi_{R0})]\})$.

If \hat{H}_K and $\hat{\Sigma}_K$ are the unstructured channel and noise estimates from equations (5.6) applied to Y_K , then we derive (see

[40, App. D]) the following expressions needed to compute ϖ_j in (7.1).

$$\begin{aligned} & \operatorname{tr}[Y_R^* \hat{\Sigma}_K^{-1} \hat{H}_K \Phi_{Rj}] \\ &= \sum_{t=1}^N \hat{\boldsymbol{y}}_{Rc}(t)^* [I_r - \hat{\Lambda}_K^2(r)]^{-1} \hat{\Lambda}_K(r) \hat{\boldsymbol{\phi}}_{Rjc}(t) \\ &= \sum_{i=1}^r \frac{\hat{\lambda}_K(i)}{1 - \hat{\lambda}_K^2(i)} \cdot \sum_{t=1}^N \hat{\boldsymbol{y}}_{Rc,i}(t)^* \cdot \hat{\boldsymbol{\phi}}_{Rjc,i}(t) \end{aligned} \quad (7.3a)$$

$$\begin{aligned} & \operatorname{tr}[\Phi_{Rj}^* \hat{H}_K^* \hat{\Sigma}_K^{-1} \hat{H}_K \Phi_{Rj}] \\ &= \sum_{i=1}^r \frac{\hat{\lambda}_K^2(i)}{1 - \hat{\lambda}_K^2(i)} \cdot \sum_{t=1}^N \hat{\boldsymbol{\phi}}_{Rjc,i}(t)^* \cdot \hat{\boldsymbol{\phi}}_{Rjc,i}(t) \end{aligned} \quad (7.3b)$$

where $\hat{\lambda}_K(i)$ are the canonical correlations, and $\hat{\boldsymbol{y}}_{Rc}(t)$ and $\hat{\boldsymbol{\phi}}_{Rjc}(t)$ are the canonical coordinates estimated from the training data Y_K . Here, $\hat{\boldsymbol{y}}_{Rc}(t) = \hat{U}_K(r)^* \hat{R}_{Kyy}^{-1/2} \boldsymbol{y}_R(t)$, and $\hat{\boldsymbol{\phi}}_{Rjc}(t) = \hat{V}_K(r)^* \hat{R}_{K\phi\phi}^{-1/2} \boldsymbol{\phi}_{Rj}(t)$, where $\hat{C}_{Kyy} = \hat{R}_{Kyy}^{-1/2} \hat{R}_{Kyy} \hat{R}_{Kyy}^{-1/2} = \hat{U}_K \hat{\Lambda}_K \hat{V}_K^*$, $\hat{R}_{Kyy} = Y_K Y_K^* / N_K$, $\hat{R}_{K\phi\phi} = \Phi_K \Phi_K^* / N_K$, and $\hat{R}_{Kyy} = Y_K \Phi_K^* / N_K$, which easily follow from the results of Section V. Thus, for the unstructured channel, the coherent matched-filter reception can be viewed as matching between the estimated canonical coordinates that is suitably weighted by the corresponding canonical correlations.

For spatially white noise (i.e., substituting $\hat{\Sigma}_K = \sigma^2 I$ in the above equations), exactly known channel parameters in the “true” model (2.1) (i.e., $\hat{H}_K = H_T$), and negligible Doppler effect (i.e., slow fading, implying $\omega_{DT1} = \omega_{DT2} = \dots = \omega_{DTP} = 0$), the above receiver reduces to the space-time RAKE receiver in [51, (111)]. In addition, for spatially white noise, structured array model with uniformly discretized angular spread and basis functions in (2.5), the above receiver reduces to the coherent matched-filter receiver in [21]. Clearly, these methods are not suited for situations involving CCI since they are based on spatially and temporally white noise assumptions. Note, however, that the spatially correlated and temporally white noise assumption used in this paper may also be too simplistic to account for strong CCI.

Recently, more realistic RAKE receivers, which account both for spatial and temporal interference, were proposed; see [52], [53], and references therein. Unlike in this paper, the spatio-temporal noise model in [52] is parametric (thus sensitive to modeling inaccuracy) and complex, requiring exact knowledge of channel parameters of all interfering users. In [53], a simpler, space-time separable noise model is used with a parametric model for the temporal covariance matrix.

1) *Probability of Error for the Binary Case:* We show how accounting for the spatial noise covariance can significantly improve the performance of the coherent matched-filter receiver in (7.2). To concentrate on the noise correlation effects, we adopt simplifying assumptions that the training sequence is long (i.e., $N_K \rightarrow \infty$), and there are no basis-function discretization effects. Then, the estimates of the channel and spatial noise covariance obtained from the training sequence are equal to their

exact values, i.e., $\hat{H}_K = H$ and $\hat{\Sigma}_K = \Sigma$. We show here that the detector (7.2) outperforms the detector that assumes spatially white noise [i.e., with $\hat{\Sigma}_K = \sigma^2 I_m$ substituted in (7.2)] in terms of error probability. The expected value of $\varpi_1 - \varpi_0$ [see (7.2)] and its variance under hypothesis H_j (symbol j is received) is

$$\begin{aligned} E[\varpi_1 - \varpi_0 | H_j] \\ = (-1)^{j+1} \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma^{-1} H (\Phi_{R1} - \Phi_{R0})] \end{aligned} \quad (7.4a)$$

$$\begin{aligned} \text{var}[\varpi_1 - \varpi_0 | H_j] \\ = 2 \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma^{-1} H (\Phi_{R1} - \Phi_{R0})] \end{aligned} \quad (7.4b)$$

for $j = 0, 1$, where (7.4b) is derived in Appendix C. Assuming that the symbols 1 and 0 are equiprobable, the average error probability is

$$P_e = Q\left(\sqrt{\frac{1}{2} \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma^{-1} H (\Phi_{R1} - \Phi_{R0})]}\right) \quad (7.5)$$

where $Q(x) = (1/\sqrt{2\pi}) \cdot \int_x^\infty \exp(-t^2/2) dt$. For the spatially white noise detector, the expected values and variances of its output are $\pm \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* H (\Phi_{R1} - \Phi_{R0})]$ and $2 \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma H (\Phi_{R1} - \Phi_{R0})]$, respectively. Then, the error probability is

$$P_{ew} = Q\left(\frac{\text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* H (\Phi_{R1} - \Phi_{R0})]}{\sqrt{2 \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma H (\Phi_{R1} - \Phi_{R0})]}}\right). \quad (7.6)$$

A straightforward application of the Cauchy–Schwarz inequality [38, p. 54] yields the inequality shown at the bottom of the page. Using this inequality to compare (7.5) and (7.6), we get

$$P_e \leq P_{ew} \quad (7.7)$$

where the equality holds if the noise is indeed spatially white.

To illustrate the advantages of accounting for the spatial noise covariance, we consider a simple example with a two-element antenna array. We then use the following parameterizations of Hermitian positive definite matrices $H(\Phi_{R1} - \Phi_{R0})(\Phi_{R1} - \Phi_{R0})^* H^*$ and Σ .

$$\begin{aligned} H(\Phi_{R1} - \Phi_{R0})(\Phi_{R1} - \Phi_{R0})^* H^* &= \begin{bmatrix} \kappa_1^2 & \rho_\kappa^* \kappa_1 \kappa_2 \\ \rho_\kappa \kappa_1 \kappa_2 & \kappa_2^2 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \sigma_1^2 & \rho_\sigma^* \sigma_1 \sigma_2 \\ \rho_\sigma \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}. \end{aligned}$$

Note that in the case of antipodal signaling and negligible ISI (i.e., $\Phi_{R1} = -\Phi_{R0} = \Phi$), we have $H(\Phi_{R1} - \Phi_{R0})(\Phi_{R1} - \Phi_{R0})^* H^* = 4H\Phi\Phi^*H^*$, which is proportional to the correlation matrix of the mean (noiseless) response of the received data. Then, κ_1^2 and κ_2^2 are positive real numbers proportional to the variances of the mean responses at the two antennas, whereas ρ_κ is the correlation between the mean responses (and, thus, $|\rho_\kappa| \leq 1$). Similarly, σ_1^2 and σ_2^2 are the noise variances at the two antennas, whereas ρ_σ is the noise correlation between them (and, thus, $|\rho_\sigma| \leq 1$). Now, the probability of error for the detector (7.2) (which takes into account the spatial noise covariance) follows from (7.5) as

$\Phi_{R0})^* H^* = 4H\Phi\Phi^*H^*$, which is proportional to the correlation matrix of the mean (noiseless) response of the received data. Then, κ_1^2 and κ_2^2 are positive real numbers proportional to the variances of the mean responses at the two antennas, whereas ρ_κ is the correlation between the mean responses (and, thus, $|\rho_\kappa| \leq 1$). Similarly, σ_1^2 and σ_2^2 are the noise variances at the two antennas, whereas ρ_σ is the noise correlation between them (and, thus, $|\rho_\sigma| \leq 1$). Now, the probability of error for the detector (7.2) (which takes into account the spatial noise covariance) follows from (7.5) as

$$P_e = Q\left(\sqrt{\frac{1}{2(1-|\rho_\sigma|^2)} \cdot \left[\frac{\kappa_1^2}{\sigma_1^2} + \frac{\kappa_2^2}{\sigma_2^2} - 2 \text{Re}\{\rho_\sigma^* \rho_\kappa\} \cdot \frac{\kappa_1}{\sigma_1} \frac{\kappa_2}{\sigma_2}\right]}\right) \quad (7.8)$$

whereas the error probability for the detector that assumes spatially white noise follows from (7.6) as

$$P_{ew} = Q\left(\frac{\kappa_1^2 + \kappa_2^2}{\sqrt{2 \cdot (\sigma_1^2 \kappa_1^2 + \sigma_2^2 \kappa_2^2 + 2 \text{Re}\{\rho_\sigma^* \rho_\kappa\} \cdot \sigma_1 \sigma_2 \kappa_1 \kappa_2)}}\right). \quad (7.9)$$

We now show that significant improvements can be achieved by the proposed detector in (7.2) (compared with the detector that assumes spatially white noise) if the noise levels at the two antennas are disproportionate. Consider the case where the noise variance at one of the antennas is very large; without loss of generality, we assume here that $\sigma_2^2 \rightarrow \infty$. Clearly, the detector assuming spatially white noise cannot adapt to this situation as its probability of error [in (7.9)] approaches $P_{ew} = Q(0) = 1/2$ as $\sigma_2^2 \rightarrow \infty$ (as in a random detector). However, the proposed detector (7.2) achieves probability of error $P_e|_{\sigma_2^2 \rightarrow \infty} = Q(\kappa_1/\sigma_1 \cdot 1/\sqrt{2 \cdot (1-|\rho_\sigma|^2)})$; see (7.8). Furthermore, it holds that $P_e|_{\sigma_2^2 \rightarrow \infty} \leq P_{e1} = Q(\kappa_1/\sigma_1 \cdot 1/\sqrt{2})$, where P_{e1} is the error probability for the coherent receiver employing antenna 1 only. Here, strict inequality holds for $|\rho_\sigma| \neq 0$, implying that the proposed coherent matched-filter detector successfully exploits the noise correlation between the antennas. (Equality holds for $\rho_\sigma = 0$ when it effectively discards the measurements from the second antenna and achieves the performance of the single-antenna receiver.)

To further study the effects of noise and mean response correlations (ρ_σ and ρ_κ) on the error probability, consider the simple case where $\kappa_1 = \kappa_2 = \kappa$ and $\sigma_1 = \sigma_2 = \sigma$. Then, (7.8) and (7.9) further simplify to $P_e = Q(\sqrt{(\kappa/\sigma)^2 \cdot [1 - \text{Re}\{\rho_\sigma^* \rho_\kappa\}]/[1 - |\rho_\sigma|^2]})$ and $P_{ew} = Q(\sqrt{(\kappa/\sigma)^2 \cdot 1/[1 + \text{Re}\{\rho_\sigma^* \rho_\kappa\}]})$. Fig. 5(a) shows P_e as a function of $\text{Re}\{\rho_\sigma^* \rho_\kappa\}$ and the SNR term $20 \cdot \log_{10}(\kappa/\sigma)$ for various spatial correlations: $|\rho_\sigma| \in \{0.1, 0.4, 0.7\}$. (Note that $-|\rho_\sigma| \leq \text{Re}\{\rho_\sigma^* \rho_\kappa\} \leq |\rho_\sigma|$, as shown in the figure.) It can be seen that the presence of noise correlation between

$$\frac{\text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* H (\Phi_{R1} - \Phi_{R0})]^2}{\text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma H (\Phi_{R1} - \Phi_{R0})] \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma^{-1} H (\Phi_{R1} - \Phi_{R0})]} \leq 1.$$

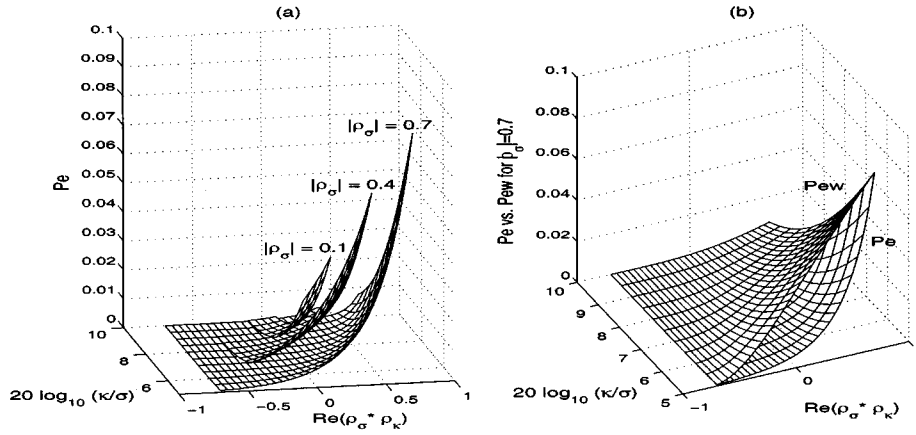


Fig. 5. Error probabilities of the coherent matched-filter receivers as functions of the product of the mean and noise correlation coefficients ($\text{Re}\{\rho_\sigma^* \rho_\kappa\}$) and SNR term $[20 \cdot \log_{10}(\kappa/\sigma)]$ when $\kappa_1 = \kappa_2 = \kappa$ and $\sigma_1 = \sigma_2 = \sigma$. (a) Probability of error for the receiver that accounts for spatially correlated noise (P_e) with varying levels of noise correlation. (b) Comparison of probabilities of error for receivers assuming arbitrary (P_e) and white spatial noise covariances (P_{ew}) when $|\rho_\sigma| = 0.7$.

the antennas improves the error probability; indeed, P_e is smaller for larger values of $|\rho_\sigma|$. Further, for fixed $|\rho_\sigma|$, it is desirable that $\text{Re}\{\rho_\sigma^* \rho_\kappa\}$ is as small as possible (the smallest P_e is achieved for $\text{Re}\{\rho_\sigma^* \rho_\kappa\} = -|\rho_\sigma|$). As expected, P_e drops sharply as the SNR term increases.

Fig. 5(b) illustrates the advantage of accounting for spatially correlated noise by comparing P_e and P_{ew} , where the spatial correlation is fixed at $|\rho_\sigma| = 0.7$. As expected (and proved theoretically above), P_e is always smaller than (or equal to) P_{ew} . From Fig. 5(b), it is interesting to observe that the two detectors approach the same performance as $|\text{Re}\{\rho_\sigma^* \rho_\kappa\}|$ approaches $|\rho_\sigma|$ ($= 0.7$ in this example).

2) *Effects of Time-Delay and Doppler Spread Discretizations on the Probability of Error:* We study the influence of time-delay and Doppler spread discretizations on the error probability [for the coherent detector (7.2)]. To concentrate on the discretization effects, we assume that the training sequence is long (thus neglecting the noise effects; see also Section V-A). In addition, we assume that binary antipodal signaling is employed and that ISI can be neglected; thus, the basis-function matrix of the received symbol can be written as $\Phi_{R1} = -\Phi_{R0} = \Phi$ (of size $d \times N_R$). The estimates of the channel and noise matrices obtained from the training sequence follow from (5.7) (where N should be replaced with N_R in the definitions of $\hat{S}_{\phi_T|\phi}$, $\hat{R}_{\phi_T\phi_T}$, and $\hat{R}_{\phi\phi_T}$, and, as before, H_T and Φ_T are the channel and basis-function matrices for the “true” model). Then, for equiprobable symbols, the probability of error easily follows as

$$P_e = Q \left(\sqrt{2N_R} \cdot \frac{\text{tr}[\hat{H}^* \hat{\Sigma}^{-1} \hat{H} \hat{R}_{\phi\phi}]}{\sqrt{\text{tr}[\hat{H}^* \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \hat{H} \hat{R}_{\phi\phi}]}} \right) = Q \left(\sqrt{2N_R} \cdot \frac{\text{tr}[\hat{\Sigma}^{-1} \cdot (\Sigma - \hat{\Sigma} + H_T \hat{R}_{\phi_T\phi_T} H_T^*)]}{\sqrt{\text{tr}[\hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \cdot (\Sigma - \hat{\Sigma} + H_T \hat{R}_{\phi_T\phi_T} H_T^*)]}} \right) \quad (7.10)$$

where we have used the identity $H_T \hat{R}_{\phi\phi_T}^* \hat{R}_{\phi\phi}^{-1} \hat{R}_{\phi\phi_T} H_T^* = \Sigma - \hat{\Sigma} + H_T \hat{R}_{\phi_T\phi_T} H_T^*$.

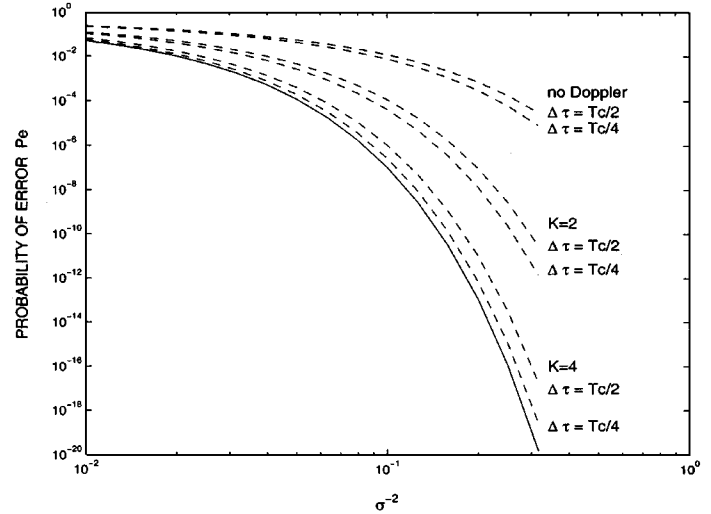


Fig. 6. Probability of error as a function of the inverse of the noise level (σ^{-2}) for $\Delta\tau = T_c/2$ and $\Delta\tau = T_c/4$ with Doppler effect neglected or modeled using $K = 2$ and $K = 4$ basis functions in the Jakes model. Full-line curve shows the performance of the ideal receiver.

Numerical Example 2—Discretization Effects and Probability of Error: In this example, we study the error probability expression in (7.10), which takes into account the discretization effects. Assume that the received symbol is generated using the model in Numerical Example 1 in Section V-A. Recall that we have chosen $\Sigma = \sigma^2 \cdot (0.9I_5 + 0.11I_{5 \times 5})$, where σ^2 can be viewed as the noise level. In Fig. 6, we show the error probability as a function of σ^{-2} for various levels of discretization of the time delays and Doppler shifts. The time-delay spread is discretized using $\Delta\tau = T_c/2$ and $\Delta\tau = T_c/4$, whereas the Doppler effect is neglected or modeled with $K = 2$ and $K = 4$ basis functions in the Jakes model (2.6). The full-line curve shows the performance of the ideal receiver (7.5) that uses the “true” model (requiring the exact knowledge of all time delays and Doppler shifts). Observe that neglecting the Doppler effect severely deteriorates the detection performance, whereas $\Delta\tau = T_c/4$ and $K = 4$ comes very close to the performance of the ideal receiver. These results demonstrate

that using the simple basis-function model from Section II-B and estimated channel and spatial noise covariance matrices in (5.7), detection performance comparable with that of the ideal receiver can be achieved.

B. Concentrated-Likelihood Receiver

Here (see also [1]), we propose an adaptive receiver structure that utilizes simultaneously the data containing several known symbols (which may be previously detected or symbols from the training sequence) *and* the currently received symbol to estimate the channel and noise parameters. Then, the concentrated likelihood (with respect to the channel and noise parameters) is used to detect the new symbol; hence, we have the name *concentrated-likelihood receiver*. An important special case when no training is available is discussed in Section VII-B1. For simplicity, we assume a full-rank unstructured channel [as in Section V, with $r = \min(m, d)$], which allows for an easy recursive computation of the concentrated likelihood, as shown in Section VII-C. In the case of transmit–receive antenna arrays (which is described in Section II-B1), this assumption is not restrictive since we can choose a subset of the set of transmit antennas so that the channel is of full rank. The optimal choice of this subset (that maximizes channel capacity) has been considered recently in [54], and the resulting channel is always of full rank; see [54, Lemma 1].

Define the data matrix containing both the known symbols and currently-received symbol $Y = [Y_K, Y_R]$ with $N = N_K + N_R$ snapshots. Then, the corresponding basis-function matrix for the known symbols and the received symbol j together is $\Phi_j = [\Phi_K, \Phi_{Rj}]$, $j \in \{0, 1, 2, \dots, J-1\}$. Note that when Y_K and Y_R are sequential in time, then $\Phi_{Rj} = [\Phi_{RKj}, \Phi_{RUj}]$, where Φ_{RKj} is the basis-function matrix covering the part of the time where there is an overlap between the known and unknown bases due to the delay spread, and Φ_{RUj} is the basis for the unknown symbol only. In CDMA applications, Φ_{RKj} usually covers only a small part of the symbol duration.

Under the above assumptions, the two forms of the concentrated likelihood function easily follow from (5.5) as

$$l_j = \frac{|\hat{R}_{\phi_j \phi_j}|}{|\hat{S}_{\phi_j | y}|} = \frac{|\hat{R}_{\phi_j \phi_j}|}{|\hat{R}_{\phi_j \phi_j} - \hat{R}_{y \phi_j}^* \hat{R}_{yy}^{-1} \hat{R}_{y \phi_j}|} \quad (7.11a)$$

$$= \frac{|\hat{R}_{yy}|}{|\hat{S}_{y | \phi_j}|} = \frac{|\hat{R}_{yy}|}{|\hat{R}_{yy} - \hat{R}_{y \phi_j} \hat{R}_{\phi_j \phi_j}^{-1} \hat{R}_{y \phi_j}^*|} \quad (7.11b)$$

where $\hat{R}_{\phi_j \phi_j} = (1/N) \cdot \Phi_j \Phi_j^* = (1/N) \cdot (\Phi_K \Phi_K^* + \Phi_{Rj} \Phi_{Rj}^*)$, $\hat{R}_{yy} = (1/N) \cdot Y Y^* = (1/N) \cdot (Y_K Y_K^* + Y_R Y_R^*)$, and $\hat{R}_{y \phi_j} = (1/N) \cdot Y \Phi_j^* = (1/N) \cdot (Y_K \Phi_K^* + Y_R \Phi_{Rj}^*)$; see also (3.2) and (3.4b). The receiver detects the new symbol by choosing the j that maximizes the concentrated likelihood in (7.11) with respect to $j = 0, 1, 2, \dots, J-1$.

We have shown in [40] that the coherent matched-filter receiver for full-rank unstructured channel is approximately

equivalent to the above concentrated-likelihood receiver, assuming long training sequence (i.e., $N_K \gg N_R$). In particular, we have shown that under this assumption

$$\begin{aligned} \ln l_j - \ln |N \hat{R}_{yy}| + \ln |N_K \hat{S}_{Ky|\phi}| \\ \approx -\frac{1}{N_K} \text{tr} \left\{ \hat{S}_{Ky|\phi}^{-1} [Y_R - \hat{H}_K \Phi_{Rj}] [Y_R - \hat{H}_K \Phi_{Rj}]^* \right\} \end{aligned} \quad (7.12)$$

where $\hat{S}_{Ky|\phi} = \hat{\Sigma}_K = (1/N_K) \cdot [Y_K Y_K^* - Y_K \Phi_K^* (\Phi_K \Phi_K^*)^{-1} \Phi_K Y_K^*]$, and $\hat{H}_K = Y_K \Phi_K^* (\Phi_K \Phi_K^*)^{-1}$ (which are the ML estimates of the channel and noise covariance matrices for the full-rank unstructured channel, computed using the data with known symbols only). The coherent matched-filter decision statistic in (7.1) follows directly from (7.12). To gain more insight into the proposed concentrated-likelihood receiver, in [40], we also consider the unstructured channel model with negligible delay and Doppler spreads [i.e., $L = K = 1$ in (2.5)]. Then, the decision statistic becomes closely related to the sample matrix inversion (SMI) beamformer [55]; see [40].

1) *Noncoherent Concentrated-Likelihood Receiver*: Consider now a concentrated-likelihood receiver that utilizes only the data containing the currently received symbol (i.e., no training data is available). The detection statistic for this *noncoherent concentrated-likelihood receiver* equals (7.11), but since training data is not used, we have $Y = Y_R$, $\Phi_j = \Phi_{Rj}$, $\hat{R}_{\phi_j \phi_j} = (1/N) \cdot \Phi_j \Phi_j^* = (1/N) \cdot \Phi_{Rj} \Phi_{Rj}^*$, $\hat{R}_{yy} = (1/N) \cdot Y Y^* = (1/N) \cdot Y_R Y_R^*$, and $\hat{R}_{y \phi_j} = (1/N) \cdot Y \Phi_j^* = (1/N) \cdot Y_R \Phi_{Rj}^*$. The absence of known data, however, imposes constraints that are useful for waveform design. To see which waveforms are desirable, consider the high-SNR case, and assume that symbol j has been sent. Then, the generalized likelihood ratio test statistic to decide between symbols i and j is

$$\left. \frac{l_j}{l_i} \right|_{j \text{ sent}} = \frac{|\Sigma + (1/N) \cdot H[\Phi_j \Phi_j^* - \Phi_j \Phi_i^* (\Phi_i \Phi_i^*)^{-1} \Phi_i \Phi_j^*] H^*|}{|\Sigma|} \quad (7.13)$$

Clearly, we wish to maximize the above expression when $j \neq i$. If we have full control over the choice of Φ_j (e.g., in the case of transmit–receive antenna array systems in slow flat-fading environment; see Section II-B-1), then we should choose $\Phi_j \Phi_i^* = 0$ for $j \neq i$ to maximize the above expression. In addition, to have equal separation for each pair of symbols, we may choose $(1/N) \cdot \Phi_j \Phi_j^* = \hat{R}_{\phi \phi} = \text{constant}$ for all $j = 0, 1, \dots, J-1$. Note that the above two conditions are closely related to the recently proposed unitary space–time codes [56] and are tailored for the case where the channel is unknown to the receiver. In the case of a single transmitter and unstructured channel model with negligible delay and Doppler spreads [i.e., when $L = K = 1$ in (2.5)], the above conditions simply demand that the symbol waveforms are mutually orthogonal and have equal energy. Observe that the waveform design metric in the case of a noncoherent receiver should be based on the following matrix (for codes i and j): $\Phi_j \Phi_j^* - \Phi_j \Phi_i^* (\Phi_i \Phi_i^*)^{-1} \Phi_i \Phi_j^*$, unlike the coherent case (which is discussed in [30]), where it is based on $(\Phi_i - \Phi_j) \cdot (\Phi_i - \Phi_j)^*$. These two matrices are fundamentally

different, e.g., antipodal signaling (i.e., $\Phi_j = -\Phi_i$) would make the noncoherent matrix go to zero, implying that it is clearly inadmissible for noncoherent detection (as expected).

The above detector can be viewed as a multivariate extension (accounting for multiple receive antennas and spatially correlated noise) of the multiuser detector in [57]. For one basis function (i.e., $d = 1$) and one receive antenna (i.e., $m = 1$), it further reduces to the standard noncoherent detector (see, for example, [58, sec. 5.4]).

C. Recursive Implementation

We derive a recursive algorithm for computing the concentrated likelihood function for a full-rank unstructured channel in (5.5), which allows for a fast implementation of concentrated-likelihood receivers in Section VII-B; see (7.11).

Define the data set and basis-function matrices $Y(t) = [Y(t-1), \mathbf{y}(t)]$ and $\Phi(t) = [\Phi(t-1), \phi(t)]$, both containing t snapshots. We denote by $l(t)$ the concentrated likelihood in (5.5) computed for $Y = Y(t)$ and $\Phi = \Phi(t)$. We also define $P_{yy}(t) = [Y(t)Y(t)^*]^{-1}$, $P_{\phi\phi}(t) = [\Phi(t)\Phi(t)^*]^{-1}$, $Q(t) = P_{yy}(t)Y(t)\Phi(t)^*$, and $P_{\phi|y}(t) = [\Phi(t)\Phi(t)^* - \Phi(t)Y(t)^*Q(t)]^{-1}$. Then, $l(t)$ can be computed using the following recursive steps (see [40, App. F]), as shown in (7.14a) and (7.14b) at the bottom of the page, and the updates of the following quantities prepare for the next step, as shown in (7.15a)–(7.15d) at the bottom of the page. An analogous recursive procedure can be derived to compute the concentrated likelihood function based on (7.11b). It would essentially reduce to swapping $\mathbf{y}(t)$ and $\phi(t)$ in the recursion defined by (7.14) and (7.15).

VIII. CONCLUDING REMARKS

We developed maximum likelihood methods for space–time fading channel estimation in spatially correlated noise with unknown covariance. Several basis-function models were proposed to account for the multipath and Doppler effects. Two array response models were used: structured and unstructured.

We computed the Cramér–Rao bound expressions for the unknown direction-of-arrival and basis-function parameters, showed that they are uncoupled, and discussed practical implications of decoupling. We derived coherent matched-filter and concentrated-likelihood receivers that account for the unknown spatial noise covariance. We analyzed the effects of time-delay and Doppler spread discretizations on the performance of the proposed coherent matched-filter receiver and demonstrated its superiority over the corresponding receiver that does not take into account the spatial noise covariance. We also derived a computationally efficient recursive implementation of the concentrated-likelihood receiver and discussed methods for spatial and temporal interference suppression.

Further research will include developing estimation and detection methods that account for temporal noise correlation as well, analyzing the proposed concentrated-likelihood receiver (in the multiuser and transmit–receive antenna array scenarios in particular), and implementing iterative interference cancellation schemes using the results of Section VI.

APPENDIX A TRANSFORMATIONS OF THE CONCENTRATED LIKELIHOOD FUNCTION

We present herein the derivations of concentrated likelihood function expressions in (3.5) and (3.3). These derivations are similar to those in [20] and are given here for completeness of presentation and because this reference is not easily available.

We start with the following expression for the concentrated likelihood function.

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|\hat{R}_{yy}|}{|\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})|}, \quad (\text{A.1})$$

which easily follows by substituting the estimates of X and Σ in (3.1) into the likelihood function for model (2.2); see also Section III. Recall the definition of the matrix $\mathcal{P} = \hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1/2}$ from Section V; then, $\hat{R}_{y\phi}\hat{R}_{\phi\phi}^{-1}\hat{R}_{y\phi}^* = \mathcal{P}\mathcal{P}^*$. Using definitions

$$\gamma(t) = \frac{1}{1 + \mathbf{y}(t)^*P_{yy}(t-1)\mathbf{y}(t)} \quad (7.14a)$$

$$l(t) = l(t-1) \cdot \frac{1 + \phi(t)^*P_{\phi\phi}(t-1)\phi(t)}{1 + \gamma(t) \cdot [\phi(t)^* - \mathbf{y}(t)^*Q(t-1)] \cdot P_{\phi|y}(t-1) \cdot [\phi(t) - Q(t-1)^*\mathbf{y}(t)]} \quad (7.14b)$$

$$P_{yy}(t) = P_{yy}(t-1) - \gamma(t) \cdot P_{yy}(t-1)\mathbf{y}(t)\mathbf{y}(t)^*P_{yy}(t-1) \quad (7.15a)$$

$$Q(t) = Q(t-1) + P_{yy}(t)\mathbf{y}(t) \cdot [\phi(t)^* - \mathbf{y}(t)^*Q(t-1)] \quad (7.15b)$$

$$P_{\phi|y}(t) = P_{\phi|y}(t-1) - \gamma(t) \cdot \frac{P_{\phi|y}(t-1)[\phi(t) - Q(t-1)^*\mathbf{y}(t)] \cdot [\phi(t)^* - \mathbf{y}(t)^*Q(t-1)] \cdot P_{\phi|y}(t-1)}{1 + \gamma(t) \cdot [\phi(t)^* - \mathbf{y}(t)^*Q(t-1)] \cdot P_{\phi\phi}(t-1) \cdot [\phi(t) - Q(t-1)^*\mathbf{y}(t)]} \quad (7.15c)$$

$$P_{\phi\phi}(t) = P_{\phi\phi}(t-1) - \frac{P_{\phi\phi}(t-1)\phi(t)\phi(t)^*P_{\phi\phi}(t-1)}{1 + \phi(t)^*P_{\phi\phi}(t-1)\phi(t)} \quad (7.15d)$$

(3.2a) and (3.1b) and the formula for the determinant of a partitioned matrix (see [34, property v., p. 8]), we derive the following useful identities.

$$\begin{aligned} |\hat{R}_{yy}| &= |\hat{S}_{y|\phi} + \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \hat{R}_{y\phi}^*| = |\hat{S}_{y|\phi} + \mathcal{P}\mathcal{P}^*| \\ &= |\hat{S}_{y|\phi}| \cdot |I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P}| \end{aligned} \quad (\text{A.2a})$$

$$\begin{aligned} |\hat{\Sigma}(\boldsymbol{\theta}, \boldsymbol{\eta})| &= |\hat{S}_{y|\phi} + (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1}) \mathcal{P}\mathcal{P}^* (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1})^*| \\ &= |\hat{S}_{y|\phi}| \cdot |I_d + \mathcal{P}^* (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1})^* \\ &\quad \cdot \hat{S}_{y|\phi}^{-1} (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1}) \mathcal{P}|. \end{aligned} \quad (\text{A.2b})$$

Further, observe that

$$\begin{aligned} \mathcal{P}^* (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1})^* \hat{S}_{y|\phi}^{-1} (I_m - \hat{T}_A \hat{S}_{y|\phi}^{-1}) \mathcal{P} \\ = \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P} \end{aligned} \quad (\text{A.3})$$

which easily follows by using the following identity: $\hat{T}_A \hat{S}_{y|\phi}^{-1} \hat{T}_A = \hat{T}_A$ [see also (3.2e)]. Then, using (A.2) and (A.3), the concentrated likelihood (A.1) can be rewritten as

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{|I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P}|}{|I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P}|}. \quad (\text{A.4})$$

We now apply formula for the determinant of a partitioned matrix [34, property v., p. 8] as

$$\begin{aligned} &\begin{vmatrix} I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} & \mathcal{P}^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \\ A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \mathcal{P} & A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \end{vmatrix} \\ &= |A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})| \cdot |I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P}| \\ &= |\mathcal{Q}| \cdot |A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) - A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \mathcal{P} \mathcal{Q}^{-1} \mathcal{P}^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})| \end{aligned} \quad (\text{A.5})$$

where \mathcal{Q} is defined as

$$\mathcal{Q} = I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P}. \quad (\text{A.6})$$

For simplicity of notation, we omit the dependence of \mathcal{Q} on $\boldsymbol{\eta}$. Applying the matrix inversion lemma (see, e.g., [34, p. 8]) to $\hat{R}_{yy}^{-1} = [\hat{S}_{y|\phi} + \mathcal{P}\mathcal{P}^*]^{-1}$, gives

$$\hat{R}_{yy}^{-1} = \hat{S}_{y|\phi}^{-1} - \hat{S}_{y|\phi}^{-1} \mathcal{P} \mathcal{Q}^{-1} \mathcal{P}^* \hat{S}_{y|\phi}^{-1}. \quad (\text{A.7})$$

Now, using (A.4) and (A.5), we obtain

$$\begin{aligned} l(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \frac{|A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})|}{|A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) - A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \mathcal{P} \mathcal{Q}^{-1} \mathcal{P}^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})|} \\ &= \frac{|A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta})|}{|A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta})|} \end{aligned} \quad (\text{A.8})$$

where the second equality [and (3.5)] follows from (A.7).

To prove (3.3), we rewrite (A.4) as

$$l(\boldsymbol{\theta}, \boldsymbol{\eta}) = \frac{\left[|I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P}|^{-1} \right]}{|\mathcal{Q}^{-1}|}. \quad (\text{A.9})$$

Applying the matrix inversion lemma [34, p. 8] to (A.6) and using $\hat{R}_{yy} = \hat{S}_{y|\phi} + \mathcal{P}\mathcal{P}^*$ yields $\mathcal{Q}^{-1} = I_d - \mathcal{P}^* \hat{R}_{yy}^{-1} \mathcal{P}$, and thus, using $\mathcal{P} = \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1/2}$, we get $|\mathcal{Q}^{-1}| = |\hat{S}_{\phi|y}| / |\hat{R}_{\phi\phi}|$, where $\hat{S}_{\phi|y}$ was defined in (3.4b). Applying the matrix inversion lemma to $\hat{S}_{y|\phi}^{-1} = [\hat{R}_{yy} - \mathcal{P}\mathcal{P}^*]^{-1}$ and using $[I_d - \mathcal{P}^* \hat{R}_{yy}^{-1} \mathcal{P}]^{-1} = \mathcal{Q}$ yields

$$\hat{S}_{y|\phi}^{-1} = [\hat{R}_{yy} - \mathcal{P}\mathcal{P}^*]^{-1} = \hat{R}_{yy}^{-1} + \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} \mathcal{P}^* \hat{R}_{yy}^{-1} \quad (\text{A.10})$$

and then, $A^* \hat{S}_{y|\phi}^{-1} \mathcal{P}$ can be simplified as

$$\begin{aligned} A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \mathcal{P} &= A(\boldsymbol{\theta})^* \left[\hat{R}_{yy}^{-1} + \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} \mathcal{P}^* \hat{R}_{yy}^{-1} \right] \mathcal{P} \\ &= A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} + A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} (I_d - \mathcal{Q}^{-1}) \\ &= A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} \end{aligned} \quad (\text{A.11})$$

where the second equality follows by using $\mathcal{P}^* \hat{R}_{yy}^{-1} \mathcal{P} = I_d - \mathcal{Q}^{-1}$. Finally, we apply (A.6), (3.2e), (A.10), (A.11), and the matrix inversion lemma to simplify

$$\begin{aligned} &I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P} \\ &= \mathcal{Q} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \left[A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} A(\boldsymbol{\theta}) \right]^{-1} A(\boldsymbol{\theta})^* \hat{S}_{y|\phi}^{-1} \mathcal{P} \\ &= \mathcal{Q} - \mathcal{Q} \mathcal{P}^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) \\ &\quad \cdot \left[A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) + A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} \mathcal{P}^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) \right]^{-1} \\ &\quad \cdot A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} \mathcal{Q} \\ &= \left[\mathcal{Q}^{-1} + \mathcal{P}^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) \left[A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} A(\boldsymbol{\theta}) \right]^{-1} A(\boldsymbol{\theta})^* \hat{R}_{yy}^{-1} \mathcal{P} \right]^{-1} \\ &= [I_d - \mathcal{P}^* W \mathcal{P}]^{-1} \end{aligned} \quad (\text{A.12})$$

where the last equality follows after using $\mathcal{Q}^{-1} = I_d - \mathcal{P}^* \hat{R}_{yy}^{-1} \mathcal{P}$ and the definition of W in (3.4a). Then

$$\begin{aligned} &\left[|I_d + \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \mathcal{P} - \mathcal{P}^* \hat{S}_{y|\phi}^{-1} \hat{T}_A \hat{S}_{y|\phi}^{-1} \mathcal{P}|^{-1} \right] \\ &= |I_d - \mathcal{P}^* W \mathcal{P}| \\ &= \left| \hat{R}_{\phi\phi} - \hat{R}_{y\phi}^* W \hat{R}_{y\phi} \right| / \left| \hat{R}_{\phi\phi} \right| \end{aligned} \quad (\text{A.13})$$

where we used $\mathcal{P} = \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1/2}$ in the last equality. Substituting $|\mathcal{Q}^{-1}| = |\hat{S}_{\phi|y}| / |\hat{R}_{\phi\phi}|$ and (A.13) into (A.9), we obtain (3.3).

APPENDIX B

DERIVATION OF UNSTRUCTURED ARRAY ML ESTIMATION RESULTS

It is easy to show that the columns of \hat{U} defined in (5.2) are also the (normalized) eigenvectors corresponding to the eigenvalues of $\hat{R}_{yy}^{1/2} \hat{S}_{y|\phi}^{-1} \hat{R}_{yy}^{1/2}$, which equal $1/[1 - \hat{\lambda}^2(j)]$, $j = 1, \dots, m$. Since $0 \leq \hat{\lambda}^2(j) \leq 1$ and $1/(1 - x)$ is an increasing function of x for $0 \leq x \leq 1$, the

eigenvalues of $\hat{R}_{yy}^{1/2} \hat{S}_{y|\phi}^{-1} \hat{R}_{yy}^{1/2}$ are sorted in nonincreasing order. Now, choose $A = \hat{R}_{yy}^{1/2} \hat{U} C_A$, where C_A is an $m \times r$ matrix of full rank r . Then, the concentrated likelihood in (3.5) becomes

$$\begin{aligned} l(A, \boldsymbol{\eta}) &= \frac{|C_A^* \hat{U}^* \hat{R}_{yy}^{1/2} \hat{S}_{y|\phi}^{-1} \hat{R}_{yy}^{1/2} \hat{U} C_A|}{|C_A^* C_A|} \\ &= \frac{|C_A^* [I_m - \hat{\Lambda} \hat{\Lambda}^T]^{-1} C_A|}{|C_A^* C_A|} \end{aligned}$$

which is maximized for $C_A = C_{A0} = [I_r, 0]^T$ and equals exactly the expression in (5.4); see also [23, eq. (35)] and [20, p. 25]. In general, the above expression is maximized for $C_A = C_{A0} \Xi_A$, where Ξ_A is an arbitrary $r \times r$ matrix of full rank, and then, $\hat{A} = \hat{R}_{yy}^{1/2} \hat{U}(r) \Xi_A$.

We now show that although the ML estimate \hat{A} is not unique, $\hat{A} \cdot \hat{X}$ is. Using (3.1a), the channel estimate in (5.6) easily follows as

$$\begin{aligned} \hat{H} &= \hat{A} \cdot \hat{X} \\ &= \hat{R}_{yy}^{1/2} \hat{U}(r) \left[\hat{U}(r)^* \hat{R}_{yy}^{1/2} \hat{S}_{y|\phi}^{-1} \hat{R}_{yy}^{1/2} \hat{U}(r) \right]^{-1} \\ &\quad \cdot \hat{U}(r)^* \hat{R}_{yy}^{1/2} \hat{S}_{y|\phi}^{-1} \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \\ &= \hat{R}_{yy}^{1/2} \hat{U}(r) \hat{U}(r)^* \hat{R}_{yy}^{-1/2} \hat{R}_{y\phi} \hat{R}_{\phi\phi}^{-1} \end{aligned} \quad (\text{B.1})$$

which does not depend on Ξ_A . In addition, the ML estimate of Σ in (5.6d) then follows from $\hat{\Sigma} = (1/N) \cdot [Y - \hat{H} \Phi(\boldsymbol{\eta})] \cdot [Y - \hat{H} \Phi(\boldsymbol{\eta})]^*$.

Another way to derive (5.6) is by substituting $A(\boldsymbol{\theta})$, X , and $\Phi(\boldsymbol{\eta})$ by I_m , A , and $X \Phi(\boldsymbol{\eta})$; then, (3.3) becomes

$$\begin{aligned} l(X, \boldsymbol{\eta}) &= \frac{|X \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^* X^*|}{\left| X \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^* X^* - \frac{1}{N} X \Phi(\boldsymbol{\eta}) Y^* \hat{R}_{yy}^{-1} Y \Phi(\boldsymbol{\eta})^* X^* \right|} \\ &= \frac{|X \hat{R}_{\phi\phi} X^*|}{|X \hat{S}_{\phi|y} X^*|} \end{aligned} \quad (\text{B.2})$$

which should be maximized with respect to X . Now, choose $X = E_X \hat{V}^* \hat{R}_{\phi\phi}^{-1/2}$, where E_X is an arbitrary $r \times d$ matrix of rank r . Then, (B.2) reduces to

$$\begin{aligned} l(X, \boldsymbol{\eta}) &= \frac{|E_X E_X^*|}{\left| E_X \hat{V}^* \hat{R}_{\phi\phi}^{-1/2} \hat{S}_{\phi|y} \hat{R}_{\phi\phi}^{-1/2} \hat{V} E_X^* \right|} \\ &= \frac{|E_X E_X^*|}{\left| E_X [I_d - \hat{\Lambda}^T \hat{\Lambda}] E_X^* \right|} \end{aligned}$$

which is maximized for $E_X = E_{X0} = [I_r, 0]$ and equals the expression in (5.4). In general, the above expression is maximized for $E_X = \Xi_X E_{X0}$, where Ξ_X is an arbitrary $r \times r$ matrix of full rank, and then, $\hat{X} = \Xi_X \hat{V}(r)^* \hat{R}_{\phi\phi}^{-1/2}$. Further, (3.1a) gives $\hat{A} = Y \Phi(\boldsymbol{\eta})^* \hat{X}^* \cdot [\hat{X} \Phi(\boldsymbol{\eta}) \Phi(\boldsymbol{\eta})^* \hat{X}^*]^{-1} = \hat{R}_{y\phi} \hat{X}^* \cdot [\hat{X} \hat{R}_{\phi\phi} \hat{X}^*]^{-1}$ and $\hat{H} = \hat{A} \hat{X} = \mathcal{P} \hat{V}(r) \hat{V}(r)^* \hat{R}_{\phi\phi}^{-1/2}$, which is exactly equal to the channel estimate in (5.6b). Then, the second expression for $\hat{\Sigma}$ in (5.6d) easily follows.

APPENDIX C

PROBABILITY OF ERROR FOR COHERENT RECEIVER

We derive the variance formula (7.4b). Note that $\text{var}[\varpi_1 - \varpi_0] = \text{var}[2\text{Re}\{z\}] = \text{var}(z + z^*)$, where $z = \text{tr}[(\Phi_{R1} - \Phi_{R0})^* H^* \Sigma^{-1} E]$, and $E = [e(1), e(2) \cdots e(N)]$. In addition, define $D = (\Phi_{R1} - \Phi_{R0})^* H^*$. Now, using a well-known property of the vec operator [34, result iii, p. 12], we have

$$z = \text{vec}(D^T)^T \cdot (I_N \otimes \Sigma^{-1}) \cdot \text{vec}(E) \quad (\text{C.1})$$

and thus

$$\begin{aligned} E[zz^*] &= \text{vec}(D^T)^T \cdot (I_N \otimes \Sigma^{-1}) \cdot \text{vec}(D^*) \\ &= \text{tr}[D \Sigma^{-1} D^*]. \end{aligned} \quad (\text{C.2})$$

In addition, note that $E[z^2] = E[(z^*)^2] = 0$, which follows from the fact that $E[e(t)e(t)^T] = 0$, $t = 1, 2, \dots, N$, which is a property of circular complex Gaussian distribution. Finally

$$\text{var}[\varpi_1 - \varpi_0] = \text{var}(z + z^*) = 2E[zz^*] = 2\text{tr}[D \Sigma^{-1} D^*] \quad (\text{C.3})$$

which proves (7.4b).

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers, Associate Editor Dr. O. Besson, and Dr. T. A. Thomas from Motorola Laboratories, Schaumburg, IL, for their helpful comments.

REFERENCES

- [1] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation in unknown spatially correlated noise," in *Proc. 37th Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sept. 1999, pp. 948–957.
- [2] D. M. Dlugos and R. A. Scholtz, "Acquisition of spread spectrum signals by an adaptive array," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1253–1270, Aug. 1989.
- [3] M. Cedervall and A. Paulraj, "Joint channel and space-time parameter estimation," in *Proc. 30th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1996, pp. 375–379.
- [4] M. Viberg, P. Stoica, and B. Ottersten, "Maximum likelihood array processing in spatially correlated noise fields using parameterized signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 996–1004, Apr. 1997.
- [5] M. Haardt, C. Brunner, and J. A. Nossek, "Efficient high-resolution 3-D channel sounding," in *Proc. 48th Veh. Technol. Conf.*, Ottawa, ON, Canada, May 1998, pp. 164–168.
- [6] A. Jakobsson, A. L. Swindlehurst, D. Astély, and C. Tidedstav, "A blind frequency domain method for DS-CDMA synchronization using antenna arrays," in *Proc. 32nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 1998, pp. 1848–1852.
- [7] Y.-F. Chen and M. D. Zoltowski, "Joint angle and delay estimation for DS-CDMA with application to reduced dimension space-time RAKE receivers," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2933–2936.
- [8] D. Astély, A. Jakobsson, and A. L. Swindlehurst, "Burst synchronization on unknown frequency selective channels with co-channel interference using an antenna array," in *Proc. 49th Veh. Technol. Conf.*, Houston, TX, May 1999, pp. 2363–2367.
- [9] G. Seco, A. L. Swindlehurst, and D. Astély, "Exploiting antenna arrays for synchronization," in *Signal Processing Advances in Communications: Trends in Single- and Multi-User Systems*, G. B. Giannakis et al., Eds. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. 2, ch. 10.
- [10] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus, and K. I. Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 434–450, Mar. 1999.

- [11] A. Jakobsson, A. L. Swindlehurst, and P. Stoica, "Subspace-based estimation of time delays and Doppler shifts," *IEEE Trans. Signal Processing*, vol. 46, pp. 2472–2483, Sept. 1998.
- [12] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Trans. Signal Processing*, vol. 45, pp. 2477–2484, Oct. 1997.
- [13] A. J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.
- [14] M. Chenu-Tournier, A. Ferreol, and P. Larzabal, "Low complexity blind space-time identification of propagation parameters," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2873–2876.
- [15] N. Bertaux, P. Larzabal, C. Adnet, and E. Chaumette, "A parameterized maximum likelihood method for multipaths channels estimation," in *Proc. 2nd IEEE Workshop Signal Process. Adv. Wireless Commun.*, Annapolis, MD, May 1999, pp. 391–394.
- [16] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449–462, Feb. 1998.
- [17] B. Hochwald and A. Nehorai, "Polarimetric modeling and parameter estimation with applications to remote sensing," *IEEE Trans. Signal Processing*, vol. 43, pp. 1923–1935, Aug. 1995.
- [18] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1971, pt. III.
- [19] A. Dogandžić and A. Nehorai, "Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays," *IEEE Trans. Signal Processing*, vol. 48, pp. 13–25, Jan. 2000.
- [20] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Lincoln Lab., Mass. Inst. Technol., Lexington, Tech. Rep. 848, Apr. 1989.
- [21] E. N. Onggosanusi, A. M. Sayeed, and B. D. Van Veen, "Canonical space-time processing for wireless communications," *IEEE Trans. Commun.*, vol. 48, pp. 1669–1680, Oct. 2000.
- [22] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. Signal Processing*, vol. 43, pp. 2144–2153, Sept. 1995.
- [23] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, pp. 3069–3078, Dec. 1996.
- [24] A. L. Swindlehurst and P. Stoica, "Maximum likelihood methods in radar array signal processing," *Proc. IEEE*, vol. 86, pp. 421–441, Feb. 1998.
- [25] A. Dogandžić and A. Nehorai, "Estimating range, velocity, and direction with a radar array," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2773–2776.
- [26] A. M. Sayeed and B. Aazhang, "Joint multipath-Doppler diversity in mobile wireless communications," *IEEE Trans. Commun.*, vol. 47, pp. 123–132, Jan. 1999.
- [27] G. B. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, pp. 1969–1986, Oct. 1998.
- [28] W. C. Jakes, Ed., *Microwave Mobile Communications*. New York: Wiley, 1974.
- [29] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [30] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
- [31] D. Chizhik, G. J. Foschini, and R. A. Valenzuela, "Capacities of multi-element transmit and receive antennas: Correlations and keyholes," *Electron. Lett.*, vol. 36, pp. 1099–1100, June 2000.
- [32] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "MIMO wireless channels: Capacity and performance prediction," in *Proc. Globecom Conf.*, San Francisco, CA, Nov. 2000, pp. 1083–1088.
- [33] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics*. New York: North-Holland, 1979.
- [34] E. F. Vonesh and V. M. Chinchilli, *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, 1997.
- [35] C. G. Khatri, "A note on a MANOVA model applied to problems in growth curve," *Ann. Inst. Statist. Math.*, vol. 18, pp. 75–86, 1966.
- [36] L. L. Scharf and J. K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, Mar. 1998.
- [37] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, pt. I.
- [38] C. R. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. New York: Wiley, 1973.
- [39] D. R. Brillinger, "A maximum likelihood approach to frequency-wavenumber analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1076–1085, Oct. 1985.
- [40] A. Dogandžić and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," Dept. Elect. Eng. Comput. Sci., Univ. Illinois, Chicago, Rep. UIC-EECS-00-9, Aug. 2000.
- [41] J. M. Cioffi, P. H. Algoet, and P. S. Chow, "Combined equalization and coding with finite-length decision feedback equalization," in *Proc. Globecom Conf.*, San Diego, CA, Dec. 1990, pp. 1664–1668.
- [42] Y. B. Hua, M. N. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Processing*, vol. 49, pp. 457–469, Mar. 2001.
- [43] L. Cohen, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [44] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Euro. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [45] C. R. Rao and T. Toutenburg, *Linear Models: Least Squares and Alternatives*, 2nd ed. New York: Springer-Verlag, 1999.
- [46] M. L. McCloud and L. L. Scharf, "Interference estimation with applications to blind multiple-access communication over fading channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 947–961, May 2000.
- [47] M. Stojanovic, J. Catipovic, and J. G. Proakis, "Adaptive multichannel combining and equalization for underwater acoustic communication," *J. Acoust. Soc. Amer.*, vol. 94, pp. 1621–1631, Sept. 1993.
- [48] G. E. Bottomley and K. Jamal, "Adaptive arrays and MLSE equalization," in *Proc. 45th Veh. Technol. Conf.*, Chicago, IL, July 1995, pp. 50–54.
- [49] K. J. Molnar and G. E. Bottomley, "Adaptive array processing MLSE receivers for TDMA digital cellular/PCS communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1340–1351, Oct. 1998.
- [50] D. Astély and B. Ottersten, "MLSE and spatio-temporal interference rejection combining with antenna arrays," in *Proc. EUSIPCO*, Rhodes, Greece, Sept. 1998, pp. 1341–1344.
- [51] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, pp. 49–83, Nov. 1997.
- [52] D. Astély and A. Artamo, "Uplink spatio-temporal interference rejection combining for WCDMA," in *Proc. IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Taoyuan, Taiwan, R.O.C., Mar. 2001, pp. 326–329.
- [53] J. Vidal, M. Cabrera, and A. Agustin, "Full exploitation of diversity in space-time MMSE receivers," in *Proc. 52nd Veh. Technol. Conf.*, Boston, MA, Sept. 2000, pp. 2497–2502.
- [54] D. A. Gore, R. U. Nabar, and A. Paulraj, "Selecting an optimal set of transmit antennas for a low rank matrix channel," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Istanbul, Turkey, June 2000, pp. 2785–2788.
- [55] I. S. Reed, J. D. Mallet, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, pp. 853–863, 1974.
- [56] 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, pp. 543–564, Mar. 2000.
- [57] E. Visotsky and U. Madhow, "Noncoherent multiuser detection for CDMA systems with nonlinear modulation: A Non-Bayesian approach," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1352–1367, May 2001.
- [58] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.



Aleksandar Dogandžić (S'96–M'01) received the Dipl.Eng. degree summa cum laude in electrical engineering from the University of Belgrade, Belgrade, Yugoslavia, in 1995 and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of Illinois at Chicago (UIC) in 1997 and 2001, respectively.

In August 2001, he joined the Department of Electrical and Computer Engineering, Iowa State University, Ames, as an Assistant Professor. His research interests are in statistical signal processing and its applications to wireless communications, radar, biomedicine, and nondestructive evaluation of materials.

Dr. Dogandžić received the Distinguished Electrical Engineering M.S. Student Award by the Chicago Chapter of the IEEE Communications Society in 1996. He received the Aileen S. Andrew Foundation Graduate Fellowship in 1997 and the UIC University Fellowship in 2000.



Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

After graduation, he worked as a Research Engineer for Systems Control Technology, Inc., Palo Alto, CA. From 1985 to 1995, he was with the Department of Electrical Engineering, Yale University, New Haven, CT, where he became an Associate Professor in 1989. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois at Chicago (UIC), as a Full Professor. From 2000 to 2001, he was Chair of the Department's Electrical and Computer Engineering (ECE) Division, which is now a new department. He holds a joint professorship with the ECE and Bioengineering Departments at UIC. His research interests are in signal processing, communications, and biomedicine.

Dr. Nehorai is Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is also a Member of the Publications Board of the IEEE Signal Processing Society and on the Editorial Board of *Signal Processing*. He has previously been an Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, the IEEE JOURNAL OF OCEANIC ENGINEERING, and *Circuits, Systems, and Signal Processing*. He served as Chairman of the Connecticut IEEE Signal Processing Chapter from 1986 to 1995 and is currently the Chair and a Founding Member of the IEEE Signal Processing Society's Technical Committee on Sensor Array and Multichannel (SAM) Processing. He was the co-General Chair of the First IEEE SAM Signal Processing Workshop, held in 2000, and will serve in this position also in 2002. He was co-recipient, with P. Stoica, of the 1989 IEEE Signal Processing Society's Senior Award for Best Paper. He received the Faculty Research Award from UIC College of Engineering in 1999. In 2001, he was named University Scholar of the University of Illinois. He has been a Fellow of the Royal Statistical Society since 1996.