Sinusoidal Modeling and Adaptive Channel Prediction in Mobile OFDM Systems

Ian C. Wong, *Student Member, IEEE* and *Brian L. Evans, *Senior Member, IEEE*

Abstract

We propose a wireless fading channel prediction algorithm for a pilot-symbol aided Orthogonal Frequency Division Multiplexing (OFDM) system. Assuming a doubly selective (time and frequency varying) ray-based physical channel model and equispaced pilot subcarriers in time and frequency, this algorithm performs channel model parameter acquisition using a 2-Step 1D ESPRIT (estimation of signal parameters via rotational invariance techniques) as a first stage, and channel prediction via model extrapolation as a second stage. Since the channel model parameter acquisition has cubic complexity, we also propose a linear complexity channel parameter tracking algorithm based on an improved adaptive ESPRIT algorithm to continuously adapt to the time-varying channel model parameters. We derive the Cramer-Rao Lower Bound (CRLB) and Asymptotic CRLB (ACRLB) for the mean squared error (MSE) in OFDM channel prediction. We show that our proposed OFDM channel prediction algorithm has better MSE performance while maintaining similar computational complexity than previous methods in sparse multipath fading channels characterized by specular scattering, which is most suitable for outdoor mobile macro-cell scenarios. Thus, our method can be seen as a complement to the existing schemes that are more suitable for dense multipath channels with diffuse scattering, which is typical of urban pico-cell and indoor wireless scenarios. We provide simulation results based on the IEEE 802.16e mobile broadband wireless access standard to corroborate our claims.

I. Introduction

OFDM is a common modulation method for high speed data access systems, e.g. IEEE 802.16e mobile wireless broadband access [1] systems. Modern high-performance OFDM transmission strategies,

The authors are with the Wireless Networking and Communications Group, The University of Texas at Austin, Austin, TX USA; Email: {iwong,bevans}@ece.utexas.edu; Phone: 1(512)2321457; FAX:1(512)4715907

The material in this paper was presented in part at the 2005 IEEE Global Telecommunications Conference and the 2005 IEEE Asilomar Conference in Signals, Systems, and Computers.

e.g. adaptive multi-user resource allocation [2], and adaptive multi-antenna precoding [3], require the transmitter to have knowledge of the current channel state information (CSI). CSI can be obtained through feedback of channel estimates from the receiver, or through the transmitter's own estimates in a time division duplex (TDD) reciprocal channel. In high mobility environments, where the Doppler frequency is high and the channel changes rapidly, the CSI used by the transmitter would be outdated due to the processing and feedback delays, causing significant performance degradation [2]. An effective means of overcoming the feedback delay is *channel prediction*.

Channel prediction algorithms for flat-fading channels have been investigated extensively in the past several years. In [4] [5], the channel is modeled as an autoregressive (AR) wide-sense stationary (WSS) stochastic process, and linear minimum mean square error (LMMSE) prediction using previous channel estimates is used to extrapolate the complex-valued fading channel. These methods can be broadly classified as *analytical* prediction approaches, since they model the wireless channel as a general stochastic process without regard to the actual physical scattering mechanism of the channel. On the other hand, several prediction algorithms model the channel as a sum of complex sinusoids [6] [7] [8]. These approaches use frequency estimation techniques to determine the dominant sinusoids in the physical fading process, and extrapolate the fading channel using the estimated model. We classify these approaches as *physical* prediction approaches, since they use ray-based scattering mechanisms to model the wireless channel.

More recently, channel prediction algorithms for doubly selective fading channels have also been studied. In [9], the power of the doubly selective channel is predicted by predicting the individual complex taps in the channel impulse response. In [10], OFDM channel prediction is performed by treating each subcarrier in the OFDM symbol as a flat-fading AR WSS stochastic process, and an LMMSE prediction filter based on previous downsampled channel estimates similar to that in [4] was used. In [11], a Kalman-filtering based channel estimation and unbiased channel power prediction based on [9] was used on the pilot-subcarriers. In [12], decision-directed and adaptive short-term channel prediction on the time-domain channel taps was proposed. Their approach uses an IFFT/FFT pair to derive the time-domain channel taps, perform the prediction, and then return to the frequency domain. These previous approaches are based on the analytical model of the channel, and differ primarily on *where the prediction is performed*, i.e. whether on all the subcarriers, the pilot subcarriers, or the time-domain channel taps. In [13], we compared these different OFDM prediction approaches using a common framework, and have concluded that prediction using the time-domain channel taps has better MSE performance with similar complexity. The use of the physical approach for time and frequency selective channel prediction have also been

considered in [14]. They extended the idea of using sinusoidal parameter estimation to the frequency selective case, using 2D-Unitary ESPRIT [15, Ch. 63] to extract the 2D frequencies.

In [16], a thorough comparison of the non-physical AR and physical sinusoidal modeling for flat-fading channel prediction using simulations and measured channels is performed. Their main conclusion was that sinusoidal modeling is better than AR modeling only for simulated channels based on simple channel models (e.g. Jakes' model), but is actually worse than AR modeling for measured channels. However, in a more recent work [8], the findings in [16] were corroborated only for urban channels, but the reverse was seen for sub-urban channels. They also concluded that AR modeling is more sensitive to channel estimation error, whereas sinusoidal modeling is more sensitive to model mismatch and non-stationarities. Furthermore, since both [14] and [8] used channel sounding measurements, which typically have very high SNR (around 40 dB), whereas practical communication links have lower SNR, the advantage actually tips towards sinusoidal modeling in practical cases, especially when model-order selection and tracking techniques with low complexity are used. Thus, we focus on the sinusoidal modeling approach for doubly selective channels, and propose low-complexity model-order and sinusoidal parameter tracking techniques to overcome the shortcomings of sinusoidal modeling.

In this paper, we propose a channel prediction algorithm for a pilot-symbol aided OFDM system with equispaced pilot subcarriers in time and frequency. Our algorithm is derived assuming a doubly selective ray-based physical channel modeled as a sum of 2D complex sinusoids whose frequencies correspond to the time-delay and Doppler frequency of each ray. We assume that the channel model parameters vary slowly; i.e. these parameters are essentially constant within a small estimation and prediction time window. Thus, using the available pilot subcarriers, we can estimate these parameters reliably and simply extrapolate the model forward to predict the future channel response. Estimating the parameters of a sum of 2D complex sinusoids in noise is a classical problem in radar, sonar, and other array signal processing fields, and many algorithms have been proposed in the past [15, Sec. XII]. However, these previous methods are highly computationally intensive and may not be suitable for cost-effective real-time implementation in mobile wireless communication systems. These approaches are based on *joint estimation*, where the time delay and Doppler frequency of each ray are determined together. Our algorithm is based on the observation that typical outdoor mobile wireless propagation environments involve clusters of scatterers, e.g. a group of large buildings and far away hills. Thus, the numerous propagating rays typically cluster around several mean time delays and angles of arrival [17]. Thus, we propose a *sequential* 2-Step 1D estimation approach, where the time-delays and the corresponding complex amplitudes are first estimated, and then followed by Doppler frequency estimation. In our work,

TABLE I

NOTATION GLOSSARY

Notation	Description	Notation	Description
$\alpha_{r,p}$	Amplitude for rth ray of pth path	P	No. of paths
$\phi_{r,p}$	Phase for rth ray of pth path	P_{max}	Max. no. of paths
$f_{r,p}$	Doppler freq. for rth ray of pth path	R_p	No. of rays for path p
$\omega_{r,p}$	Normalized radian Doppler freq.	$R_{p,max}$	Max. no. of rays per path
τ_p	Time delay for rth ray of pth path	N	No. of subcarriers
φ_p	Normalized radian time delay	N_{cp}	Cyclic prefix length
Δf	Subcarrier frequency spacing	N_f	No. of pilot subcarriers
D_f	Pilot subcarrier spacing	N_t	No. of pilot symbols
D_t	Pilot symbol spacing	N_u	No. of used subcarriers

we chose the ESPRIT [15, Ch. 63.2] algorithm since it has been shown to provide better MSE performance than other comparable methods, and has a closed-form solution, which eases implementation.

Although less computationally complex than the joint estimation approach, the 2-step 1D approach is still much more complex than the AR-modeling-based approaches, since ESPRIT has cubic complexity in the number of samples. Fortunately, since the time-delays and Doppler frequencies are slowly varying, this step needs to be performed only once during an initialization phase, and more computationally efficient tracking techniques can be used for subsequent symbols. Thus, we also propose a linear complexity channel parameter tracking algorithm based on our own improved adaptive ESPRIT algorithm.

We also derive the CRLB and ACRLB for the MSE in OFDM channel prediction. We show via simulation that our proposed algorithm is quite close to the MSE lower bound for practical propagation scenarios and pilot-subcarrier lengths. We also provide simulation results that compare our algorithm with AR modeling-based algorithms, and it is shown that our algorithm provides better MSE performance at a similar computational complexity for sparse multipath channels characterized by specular scattering, which is suitable for outdoor mobile macro-cell scenarios.

II. SYSTEM MODEL

A. OFDM Baseband System Model

[1]We consider an $N = N_u + N_g + 1$ subcarrier OFDM system, where $N_u = N_d + N_f$ subcarriers (N_u is assumed to be even) in the middle of the band are used, which include the N_d data subcarriers and the

[1]Table I summarizes the notation for the most commonly used parameters.

N_f pilot subcarriers, and 1 null subcarrier at the 0th index. The rest of the N_g subcarriers are the guard subcarriers, which are nulled out to allow the signal to naturally decay and create the FFT "brick wall" shaping. Let F_s be the sampling frequency, which gives us a subcarrier spacing of $\Delta f = F_s/N$, and a useful symbol time of $T_u = 1/\Delta f$. We use a cyclic prefix (CP) of duration T_{cp} to avoid inter-symbol interference and maintain the orthogonality of the subcarriers, which gives us a total OFDM symbol time of $T_{sym} = T_{cp} + T_u$. The complex baseband representation for the nth OFDMA symbol is

$$s_n(t) = \sum_{\substack{k=-N_u/2 \\ k \neq 0}}^{N_u/2} X(n,k) e^{j2\pi k \Delta f (t - T_{cp})}, \quad \text{for } nT_{sym} \leq t < (n+1)T_{sym} \tag{1}$$

We assume that the cyclic-prefix time T_{cp} is greater than the maximum delay spread of the channel τ_{max}, and that time and frequency synchronization is perfect, such that the received signal at the kth subcarrier of the nth symbol can be written as

$$Y(n,k) = H(n,k)X(n,k)G(k) + W(n,k) \tag{2}$$

where $H(n,k)$ is the complex channel response, $X(n,k)$ is the complex transmitted symbol we wish to communicate, $G(k) = G_T(k)G_R(k)$ with $G_T(k)$ and $G_R(k)$ as the pulse-shaping transmitter and receiver filter frequency response values at the kth subcarrier frequency, and $W(n,k)$ is a zero-mean, circular-symmetric, complex additive white Gaussian noise (AWGN) with variance σ^2.

B. Doubly Selective Wireless Channel

The continuous-time impulse response of a doubly selective wireless channel can be modeled as the superposition of a discrete number of resolvable paths [17]

$$h_c(t,\tau) = \sum_{p=1}^{P} \gamma_p(t)\delta(\tau - \tau_p(t)) \tag{3}$$

where $\tau_p(t)$ and $\gamma_p(t)$ are the time-varying delay and complex gain of the pth path, and P is the number of paths. We assume that reflectors and scatterers are far enough from the receiver, so that the waves incident on the receiving antenna are *plane waves*; and that the receiver travels in a linear motion and with constant velocity[2]. These assumptions allow $\gamma_p(t)$ to be fitted by a sum of complex sinusoids

$$\gamma_p(t) = \sum_{r=1}^{R_p} \alpha_{r,p}(t) e^{j\phi_{r,p}(t)} e^{j2\pi f_{r,p} t} \tag{4}$$

[2]These assumptions are valid in outdoor macro-cellular propagation scenarios and in considering small time windows of interest, i.e. a few wavelengths.

where R_p is the number of rays contributing to the pth path, and $\alpha_{r,p}(t)$, $\phi_{r,p}(t)$ and $f_{r,p}$ are the real valued amplitude, phase, and Doppler frequency, respectively, for the rth ray in the pth path. Although these parameters are in general dependent on time, we assume that they are constant within a small estimation and prediction time window[3]. We allow these parameters to vary slowly beyond this window, where we use tracking techniques to adapt to the non-stationary channel at a low computational cost.

Removing the time-dependency of the parameters in (4), substituting it into (3), and taking its Fourier transform, we get the frequency response of the channel within the stationary window $H_c(t,f) = \sum_{p=1}^{P}\sum_{r=1}^{R_p} \alpha_{r,p} e^{j\phi_{r,p}} e^{j2\pi(f_{r,p}t-\tau_p f)}$ which is composed of $R = \sum_{p=1}^{P} R_p$ complex sinusoidal rays. Each ray is characterized by the quadruplet $\{\alpha_{r,p}, \phi_{r,p}, f_{r,p}, \tau_p\}$. We assume that no two rays share the same pair of Doppler frequency and time-delay values $\{f_{r,p}, \tau_p\}$ (otherwise they can be combined to form one ray), but different rays may share the same delay or Doppler frequency. We further assume that the OFDM system with symbol period T_{sym} and subcarrier spacing Δf have proper cyclic extension and sample timing, giving us the sampled channel frequency response

$$H(n,k) = \sum_{p=1}^{P}\sum_{r=1}^{R_p} \alpha_{r,p} e^{j\phi_{r,p}} e^{j2\pi(f_{r,p}nT_{sym}-\tau_p k\Delta f)} \tag{5}$$

Note that practical values of $f_{r,p}$ and τ_p are bounded, i.e. $0 \le f_{r,p} < f_{max}$ and $0 \le \tau_p < \tau_{max}$ where f_{max} is the maximum Doppler frequency and τ_{max} is the maximum delay spread. Hence, (5) has finite support in both Doppler and delay domains.

C. OFDM Pilot Pattern and Least-Squares Channel Estimation

We assume that there are N_f pilot subcarriers inserted every $D_f = \lfloor N_u/N_f \rfloor$ subcarriers within the N_u used subcarriers in every OFDM symbol, where $D_f \le 1/(\Delta f \tau_{max})$ to avoid aliasing in the frequency domain. We also assume that a block of N_t current and previous OFDM symbols (which we call pilot symbols) equally inserted every D_t OFDM symbols, where $D_t \le 1/(2T_{sym}f_{max})$ in order to avoid aliasing in the time domain.

Let $k_q = \left(\frac{N_f}{2} - q\right) D_f$, $q \in \mathcal{Q}$, $\mathcal{Q} \triangleq \{0, \ldots, N_f - 1\}$ denote the set of pilot subcarrier indices across frequency[4], and $n_l = lD_t$, $l \in \mathcal{L}$, $\mathcal{L} \triangleq \{0, \ldots, N_t - 1\}$ denote the set of pilot symbol across time. Using

[3]The fast fading nature of the channel is due to the superposition of these sinusoids. A similar model has been used as simulation models for mobile wireless channels [18] and also for wireless channel modeling and prediction [19] [14]

[4]We implicitly assumed that N_f is even such that index k_q is an integer. A similar pilot subcarrier set can be defined for N_f odd without changing the development of the algorithm.

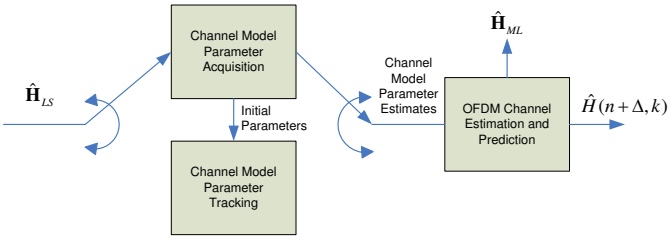


Fig. 1. Top-level block diagram for proposed OFDM channel prediction algorithm

these $N_t \times N_f$ pilot subcarriers[5], we can perform a least squares (LS) estimate of the channel at the pilot locations using the received signal $Y(n_l, k_q)$ and the known pilot symbols $X(n_l, k_q)$, given as

$$\begin{aligned}\hat{H}_{LS}(l,q) &= Y(n_l,k_q)/X(n_l,k_q) \\ &= G(k_q)\sum_{p=1}^{P}\sum_{r=1}^{R_p} c_{r,p}e^{j(\omega_{r,p}l+\varphi_p q)} + \widetilde{W}(l,q), \quad l \in \mathcal{L}, \quad q \in \mathcal{Q} \end{aligned} \tag{6}$$

where $\omega_{r,p} = 2\pi f_{r,p} D_t T_{sym}$, $\varphi_p = 2\pi\tau_p D_f \Delta f$, and $c_{r,p} = \alpha_{r,p}e^{j(\phi_{r,p}-\varphi_p N_f/2)}$. We assume that the pilot symbols $X(n_l, k_q)$ are deterministic complex values with unit magnitude, such that $\widetilde{W}(l,q) = W(n_l,k_q)/X(n_l,k_q)$ is AWGN with variance σ^2. For notational convenience, let

$$\hat{\mathbf{H}}_{LS} = \left[\hat{\mathbf{h}}_0^{0:N_f-1} \ldots \hat{\mathbf{h}}_{N_t-1}^{0:N_f-1}\right] \tag{7}$$

be the $N_f \times N_t$ matrix of the LS estimates, where $\hat{\mathbf{h}}_l^{a:b} = [\hat{H}_{LS}(l,a), \ldots, \hat{H}_{LS}(l,b)]^T$ is the column vector of the estimates on time index n_l and pilot indices $\{k_q\}_{q=a}^{q=b}$.

III. OFDM Channel Prediction Algorithm

The top-level block diagram for our proposed OFDM channel prediction algorithm is shown in Fig. 1. Our channel prediction algorithm first estimates the parameters in our deterministic channel model using the least squares estimates $\hat{\mathbf{H}}_{LS}$ in (7), and then extrapolates this model to predict the future channel $\hat{H}(n+\Delta, k)$. Improved maximum likelihood (ML) channel estimates $\hat{\mathbf{H}}_{ML}$ are also generated as a by-product of our prediction algorithm. After the initial parameter acquisition and channel prediction step, subsequent channel parameters are then tracked using computationally efficient adaptive algorithms.

[5]Although we require equispacing of the pilots in time and frequency, we do not require that the pilots align across time (i.e. the pilots need not be in the same subcarrier index for all time), thus allowing more general pilot patterns used in e.g. [1].

A. Channel model parameter acquisition

Notice that (6) is in the form of a 2D complex sum-of-sinusoids (2D-SoS) in additive white Gaussian noise, if not for the $G(k_q)$ pulse-shaping filter frequency response terms, which can be considered as a deterministic multiplicative noise across the frequency dimension. Fortunately, it is not very difficult to design these filters to have almost flat passband response. Thus, we simply ignore the minimal effect of $G(k_q)$ in the derivation of our algorithm, and treat (6) as a standard 2D-SoS model. However, we explicitly included the $G(k_q)$ factors in the results section for fair comparison with other algorithms.

Estimating the parameters of a 2D-SoS model jointly is a classical problem, and is very well studied in the radar, sonar, and other array signal processing literature (see e.g. [15, Sec. XII] and references therein). Although a straightforward application of these techniques may be used, they are too computationally expensive for cost-effective implementation, primarily because of the required pairing operation of the two frequencies for each ray. Other techniques that exploit shift invariance do not require a pairing operation [15, Ch. 63], but are still highly complex due to the large matrix decomposition and joint diagonalization.

An important observation on the structure of the wireless channel, however, allows us to reduce the computational burden. We decompose (6) as

$$\hat{H}_{LS}(l,q) = \sum_{p=0}^{P-1} g_p(l) e^{j\varphi_p q} + \widetilde{W}(l,q) \tag{8}$$

where the complex gain for the pth propagation is

$$g_p(l) = \sum_{r=0}^{R_p-1} c_{r,p} e^{j\omega_{r,p} l} \tag{9}$$

As mentioned in the introduction section, we exploit the fact that several (R_p) sinusoidal rays actually share the same time-delay value φ_p. This allows us to divide the estimation task into a *time delay (φ_p) estimation step*, and then a *Doppler frequency ($\omega_{r,p}$) estimation step*, which we call the 2-Step 1D approach to channel parameter estimation.

If we treat $g_p(l)$ in (8) as simply the unknown complex amplitudes of the sum of sinusoids with frequencies φ_p, then (8) amounts to 1D-SoS parameter estimation. This is a much simpler problem than its two-dimensional counterpart, and a huge body of literature is available in solving this problem (see e.g. [20]). In this type of estimation problem, once the number of sinusoids P is estimated, the difficulty lies in estimating the frequencies of the complex exponentials (φ_p in this case), since these frequencies enter the model in a non-linear fashion. Although the ML estimates for the frequencies are desirable, techniques to find these exactly or even approximately are highly complex iterative procedures that are not guaranteed to converge [21]. Hence, we opted to base our algorithm on the non-iterative eigen-analysis

technique ESPRIT [15, Ch. 63.2], which has been shown to perform better than other techniques in terms of MSE, while still being amenable to real-time computation. Once these frequencies have been estimated, the ML estimates for the complex amplitudes $g_p(l)$ are easily computed using a standard linear LS approach. We shall refer to this step where the time-delays φ_p and complex multipath taps $g_p(l)$ are estimated as the time-delay estimation step, or *Step 1*. Once the estimates of $g_p(l)$ are generated, we can subsequently use these estimates as the left hand side of (9), and once again we have a 1D-SoS parameter estimation problem. This allows us to use the same algorithm as above and generate the estimates for R_p, $\omega_{r,p}$, and $c_{r,p}$. We shall refer to this step where the Doppler frequencies $\omega_{r,p}$ and complex amplitudes $c_{r,p}$ are estimated as the Doppler frequency estimation step, or *Step 2*. The two-step algorithm is described in more detail next.

1) Time-delay estimation: [6]The time-delay estimation step can be further broken down into 4 substeps: estimating the autocorrelation matrix and computing its eigenvalue decomposition (EVD), estimating the number of paths P, estimating the time-delays φ_p, and estimating complex amplitudes $\{g_p(l)\}_{l\in\mathcal{L}}$ for each path $p = 0, \ldots, \hat{P} - 1$.

a) *Estimate the autocorrelation matrix* $\mathbf{R}^f$ *and compute its EVD*: Since there are N_t OFDM symbols, we can use an average of the frequency autocorrelation estimates for each symbol generated using the modified covariance method (a.k.a. forward-backward) [21],

$$\hat{\mathbf{R}}^f = \frac{1}{N_t}\sum_{l=0}^{N_t-1}\frac{1}{2}\left(\hat{\mathbf{\Psi}}(n_l) + \mathbf{J}\hat{\mathbf{\Psi}}(n_l)^H\mathbf{J}\right) \tag{10}$$

$$\hat{\mathbf{\Psi}}(n_l) = \frac{1}{N_f}\sum_{q=K}^{N_f}\left(\hat{\mathbf{h}}_l^{(q-K+1):q}\right)\left(\hat{\mathbf{h}}_l^{(q-K+1):q}\right)^H \tag{11}$$

and $\mathbf{J}$ is the exchange matrix with ones on the anti-diagonal and zeros elsewhere, and K is the size of the autocorrelation matrix chosen to be greater than the maximum possible number of paths[7] P_{max} and less than N_f. The eigenvalue decomposition (EVD) of $\hat{\mathbf{R}}^f$

$$\hat{\mathbf{R}}^f \overset{\text{EVD}}{\rightarrow} \sum_{k=1}^{K}\hat{\lambda}_k^f\hat{\mathbf{v}}_k\hat{\mathbf{v}}_k^H, \quad \hat{\lambda}_1^f \geq \ldots \geq \hat{\lambda}_K^f \tag{12}$$

is then performed where $\{\hat{\lambda}_k^f\}_{k=1}^K$ are the non-increasingly ordered estimated eigenvalues of the estimated autocorrelation matrix $\hat{\mathbf{R}}^f$.

[6]Note that this step is similar to the time-delay acquisition step of the OFDM channel estimation algorithm in [22, Sec. III B-C], although we assumed a different channel model, and our focus is on its use for *prediction*, not estimation.

[7]We make the reasonable assumption that P_{max} is known *à priori*. This value is typically determined by the propagation environment, and the desired accuracy of the channel characterization.

b) *Estimate the number of paths P*: Estimation of the number of paths P is essentially a model-selection problem, wherein the Minimum Description Length (MDL) is the method most often used due to its consistency [21]. We employ the MDL appropriate for the modified covariance averaging technique [23] given as

$$\hat{P} = \arg \min_{1 \le \mu \le K-1} -\log \left(\frac{\left(\prod_{k=\mu+1}^{K} \hat{\lambda}_k^f \right)^{\frac{1}{K-\mu}}}{\frac{1}{K-\mu} \sum_{k=\mu+1}^{K} \hat{\lambda}_k^f} \right)^{N_t(K-\mu)} + \frac{1}{4}\mu(2K-\mu+1)\log N_t \tag{13}$$

c) *Estimate the time-delays $\{\varphi_p\}_{p=1}^{\hat{P}}$*: Let $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_{\hat{P}}]$ denote the matrix whose columns are the eigenvectors associated with the $\hat{P}$ largest eigenvalues of $\hat{\mathbf{R}}^f$, and let $\hat{\mathbf{V}}_1$ and $\hat{\mathbf{V}}_2$ denote the upper and lower $(K-1) \times \hat{P}$ submatrices of $\hat{\mathbf{V}}$. The time delay estimates are

$$\hat{\varphi}_p = \arg(\hat{\epsilon}_p), \quad p = 1, \dots, \hat{P} \tag{14}$$

where $\arg(x)$ is the radian phase angle of the complex number x, and $\{\hat{\epsilon}_p\}_{p=1}^{\hat{P}}$ are the eigenvalues of the $\hat{P} \times \hat{P}$ ESPRIT spectral matrix $\hat{\mathbf{\Phi}}_{\hat{P}}$, given as

$$\hat{\mathbf{\Phi}}_{\hat{P}} = (\hat{\mathbf{V}}_1^H \hat{\mathbf{V}}_1)^{-1} \hat{\mathbf{V}}_1^H \hat{\mathbf{V}}_2 \tag{15}$$

d) *Estimate the complex amplitudes $\{\hat{\mathbf{g}}_p(l)\}_{p=1}^{\hat{P}}$*: Assuming that the time delay estimates $\{\hat{\varphi}_p\}_{p=1}^{\hat{P}}$ are correct, the ML estimate for the $\hat{P} \times N_t$ matrix of complex amplitudes, given as

$$\hat{\mathbf{G}}_{ML} = [\hat{\mathbf{g}}_1 \dots \hat{\mathbf{g}}_{\hat{P}}]^T \tag{16}$$

$$\hat{\mathbf{g}}_p = [\hat{g}_p(1) \quad \hat{g}_p(2) \quad \dots \quad \hat{g}_p(N_t)]^T \tag{17}$$

which can be computed via least squares, i.e.

$$\hat{\mathbf{G}}_{ML} = \left(\hat{\mathbf{E}}^H \hat{\mathbf{E}} \right)^{-1} \hat{\mathbf{E}}^H \hat{\mathbf{H}}_{LS} \tag{18}$$

where $[\hat{\mathbf{E}}]_{q,p} = e^{j\hat{\varphi}_p q}$, $q = 0, \dots, N_f - 1$, $p = 0, \dots, \hat{P} - 1$, is the $N_f \times \hat{P}$ Fourier transform matrix.

2) Doppler frequency estimation: In this second step, we utilize the estimated complex amplitudes $\hat{\mathbf{g}}_p$ (17) for each path p to estimate the remaining parameters in our channel model. We simply replace the left-hand side of (9), $g_p(l)$, with its corresponding estimate $\hat{g}_p(l)$, and proceed similarly as Step 1, since it is also a sinusoidal parameter estimation problem. For each path $p = 1, \dots, \hat{P}$, we repeat the four substeps similar to Section III-A.1 above with slight modifications.

a) *Estimate the autocorrelation function* $\hat{\mathbf{R}}_p^t$ *and compute its EVD*: We estimate the autocorrelation matrix $\hat{\mathbf{R}}_p^t$ across time for path p using the modified covariance method as in (10), but without the averaging since we only have one observation vector, i.e.

$$\hat{\mathbf{R}}_p^t = \frac{1}{2}\left(\hat{\mathbf{\Psi}}(p) + \mathbf{J}\hat{\mathbf{\Psi}}(p)^H\mathbf{J}\right) \tag{19}$$

$$\hat{\mathbf{\Psi}}(p) = \frac{1}{N_t}\sum_{l=I}^{N_t}\left(\hat{\mathbf{g}}_p^{(l-I+1):l}\right)\left(\hat{\mathbf{g}}_p^{(l-I+1):l}\right)^H \tag{20}$$

The EVD of $\hat{\mathbf{R}}_p^t$ gives the eigenvalues $\{\hat{\lambda}_i^t\}_{i=1}^I$ and their corresponding eigenvectors. Similar to Sec. III-A.1.a), we should choose I larger than the maximum number of rays $R_{p,max}$ but less than N_t. In practice, N_t is chosen to balance complexity versus the modeling accuracy, and a good rule of thumb is to choose $I \approx 3/5N_t$ [24].

b) *Estimate the number of rays* R_p: We estimate the number of complex sinusoidal rays R_p contributing to path p also using the MDL derived for sinusoidal estimation described in [25].

c) *Estimate the Doppler frequencies*$\{\omega_{r,p}\}_{r=1}^{\hat{R}_p}$: We estimate the Doppler frequencies using the same ESPRIT algorithm given in (14)-(15).

d) *Estimate the complex amplitudes* $\{c_{r,p}\}_{r=1}^{\hat{R}_p}$: We estimate the amplitudes also via LS

$$\hat{\mathbf{c}}_p = \left(\mathbf{E}_p^H\mathbf{E}_p\right)^{-1}\mathbf{E}_p^H\hat{\mathbf{g}}_\mathbf{p} \tag{21}$$

where $\hat{\mathbf{c}}_p = [\hat{c}_{1,p}\ldots\hat{c}_{\hat{R}_p,p}]^T$ and $[\hat{\mathbf{E}}_p]_{l,r} = e^{j\hat{\omega}_{r,p}l}, l = 0,\ldots,N_t-1, r = 1,\ldots,\hat{R}_p$ is the $N_t \times \hat{R}_p$ estimated Fourier transform matrix.

B. Channel estimation and prediction

Now that we have estimated all the parameters needed in our model, we just plug in these parameters into our model in (5) to find our predicted channel Δ symbols ahead as

$$\hat{H}(n+\Delta,k) = \sum_{p=1}^{\hat{P}}\sum_{r=1}^{\hat{R}_p}\hat{\alpha}_{r,p}e^{j\hat{\phi}_{r,p}}e^{j2\pi(\hat{f}_{r,p}(n+\Delta)T_{sym}-\hat{\tau}_p k\Delta f)} \tag{22}$$

where $\hat{\tau}_p = \frac{\hat{\varphi}_p}{2\pi D_f}$, $\hat{f}_{r,p} = \frac{\hat{\omega}_{r,p}}{2\pi D_t}$, $\hat{\alpha}_{r,p} = \text{mag}\{\hat{c}_{r,p}\}$, and $\hat{\phi}_{r,p} = \arg\{\hat{c}_{r,p}\} + \hat{\varphi}_p\frac{N_f}{2}$. Although we used simple model extrapolation in this case, we can similarly use Bayesian-type prediction by modeling the complex amplitudes $c_{r,p}$ as IID random variables. It can be shown that LMMSE channel prediction is also model extrapolation, with the parameters estimated using Bayesian LMMSE techniques [8].

Note that each of the N_t columns of $\hat{\mathbf{G}}_{ML}$ in (18) is actually the time-domain channel tap estimate for the frequency-selective channel. Hence, we can write

$$\hat{\mathbf{H}}_{ML} = \mathbf{W}\hat{\mathbf{G}}_{ML} \tag{23}$$

where the $N_u \times N_t$ matrix $\hat{\mathbf{H}}_{ML}$ is the ML estimate for the frequency response of the channel at the N_u data carrying subcarriers given that the time-delays are correct, and the $(N_u + 1) \times \hat{P}$ Fourier transform matrix with elements $[\mathbf{W}]_{k,p} = e^{j\frac{\hat{\tau}_p}{D_f}(k-1-\frac{N_u}{2})}$. Thus, our algorithm can be used as an OFDM channel estimation and prediction solution.

C. Channel Model Parameters Tracking

Although less complex than the joint estimation approaches, the 2-step 1D approach has $O(N_f^3 + N_t^3\hat{P})$ principal computational complexity, which is still highly complex for efficient implementation. Fortunately, since the time-delays and Doppler frequencies are slowly varying, this step needs to be performed only once during an initialization phase, and more computationally efficient subspace tracking techniques can be used for subsequent symbols. In this section, we derive a completely adaptive version of the 2-step 1D ESPRIT-based algorithm that has $O\left(N_f P_{max} + N_t \hat{P} R_{p,max}\right)$ complexity that is able to efficiently track the slowly time-varying parameters, and also adapt to the time-varying model order. Since both time-delay and Doppler frequency estimation entail essentially the same operations, we shall limit our discussion to the time-delay estimation step, and explain briefly the modifications necessary for the Doppler frequency estimation step. Similar to our discussion of Step 1 in Sec. III-A.1, we subdivide this step into adaptive autocorrelation estimation and subspace tracking, adaptive model-order estimation, adaptive time-delay estimation, and adaptive complex amplitude estimation. Table II presents a listing of this step together with the complexity of each operation as it appears in the ensuing discussion.

1) Adaptive autocorrelation estimation and subspace tracking: We update the matrix of LS estimates using an exponential window, given as[8]

$$\hat{\mathbf{H}}_{LS}^H(l) = \begin{bmatrix} (1-\beta)^{1/2}\hat{\mathbf{h}}^H(l) \\ \beta^{1/2}\hat{\mathbf{H}}_{LS}^H(l-1) \end{bmatrix} \tag{24}$$

where $\hat{\mathbf{H}}_{LS}^H(l-1)$ is the "old" observation matrix, $\hat{\mathbf{h}}(l)$ is the "new" LS-estimated N_f-length vector, and $0 < \beta < 1$ is the exponential forgetting factor. This smooth exponential update leads to the development of efficient singular value decomposition (SVD) updating algorithms for computing the dominant singular values and right singular vectors of the growing data matrix at each time step [26], which can effectively replace computing the EVD of the frequency autocorrelation matrix in (10), since this is equivalent to tracking the dominant eigenvalues and eigenvectors of $\hat{\mathbf{R}}^f(l)$.

[8]Note that the vertical dimension grows without bound in this equation. It is thus the purpose of various subspace tracking techniques (e.g. [26]) to approximate this "exponential windowing" update with algorithms that do not have the computationally infeasible "growing vertical dimension".

TABLE II

CHANNEL MODEL PARAMETERS TRACKING STEP 1: TIME-DELAY ESTIMATION.

Input:	$\hat{\mathbf{h}}(l)$		
Output:	$\hat{P}, \{\hat{\varphi}_p\}_{p=1}^{\hat{P}}, \hat{\mathbf{g}}_l$		
Section	**Operation**	**Complexity**	**Equation No.**
III-C.1	*Adaptive autocorrelation estimation and subspace tracking*		
	Run subspace tracker $\xrightarrow[\text{Mod Bi-SVD3}]{}$ $\hat{\mathbf{V}}(n), \hat{\boldsymbol{\sigma}}, \mathbf{G}, \hat{\mathbf{h}}^{\perp}$	$10N_fP_{max} + 3P_{max}^3$	Table III
III-C.2	*Adaptive model-order estimation*		
	$\hat{\sigma}^2(l) = \beta\hat{\sigma}^2(l-1) + (1-\beta)\frac{\hat{\mathbf{h}}^{\perp H}\hat{\mathbf{h}}^{\perp}}{N_f - P_{max}}$	N_f	(27)
	$\hat{\lambda_i}^f = \begin{cases} \hat{\sigma_i}^2, & i \in [1, P_{max}] \\ \hat{\sigma}^2(l), & i \in [P_{max}+1, N_f] \end{cases}$	P_{max}	(28)
	$\tilde{N}_t = \frac{1-\beta^l}{1-\beta}$		(29)
	$\hat{P} = \arg \min_{1 \le \mu \le P_{max}} -\log \left(\frac{\left(\prod_{k=\mu+1}^{N_f} \hat{\lambda}_k^f \right)^{\frac{1}{N_f-\mu}}}{\frac{1}{N_f-\mu} \sum_{k=\mu+1}^{N_f} \hat{\lambda}_k^f} \right)^{\tilde{N}_t(N_f-\mu)} + \frac{1}{2}\mu(2N_f - \mu)\log \tilde{N}_t$	N_fP_{max}	(30)
III-C.3	*Adaptive time-delay estimation using adaptive ESPRIT*		
	$\bar{\mathbf{h}}^{\perp} = \hat{\mathbf{h}}^{\perp} / \parallel \hat{\mathbf{h}}^{\perp} \parallel$		
	$\bar{\mathbf{h}}_1^{\perp} = [\mathbf{I}_{N_f-1} \quad \mathbf{0}_{(N_f-1)\times 1}]\bar{\mathbf{h}}^{\perp}$		
	$\bar{\mathbf{h}}_2^{\perp} = [\mathbf{0}_{(N_f-1)\times 1} \quad \mathbf{I}_{N_f-1}]\bar{\mathbf{h}}^{\perp}$		
	$\hat{\mathbf{V}}_1(l-1) = [\mathbf{I}_{N_f-1} \quad \mathbf{0}_{(N_f-1)\times 1}]\hat{\mathbf{V}}(l-1)$		
	$\hat{\mathbf{V}}_2(l-1) = [\mathbf{0}_{(N_f-1)\times 1} \quad \mathbf{I}_{N_f-1}]\hat{\mathbf{V}}(l-1)$		
	$\begin{bmatrix} \boldsymbol{\Upsilon}(l) & * \\ * & * \end{bmatrix} = \mathbf{G}(l) \begin{bmatrix} \boldsymbol{\Upsilon}(l-1) & \hat{\mathbf{V}}_1^H(l-1)\bar{\mathbf{h}}_2^{\perp} \\ \bar{\mathbf{h}}_1^{\perp H}\hat{\mathbf{V}}_2(l-1) & \bar{\mathbf{h}}_1^{\perp H}\bar{\mathbf{h}}_2^{\perp}(l) \end{bmatrix} \mathbf{G}^H(l)$	$O(P_{max}^2)$	(38)
	$\boldsymbol{\Upsilon}(l) \xrightarrow[\text{Extract } \hat{P} \times \hat{P} \text{ top-left submatrix}]{} \boldsymbol{\Upsilon}_{\hat{P}}(l)$		
	$\hat{\mathbf{V}}(l) \xrightarrow[\text{Extract } \hat{P} \text{ bottom-left-most row vector}]{} \hat{\mathbf{v}}_{\hat{P},1}^H(l)$		
	$\hat{\boldsymbol{\Phi}}_{\hat{P}}(l) = \boldsymbol{\Upsilon}_{\hat{P}}(l) + \frac{\hat{\mathbf{v}}_{\hat{P},1}(l)\left(\boldsymbol{\Upsilon}_{\hat{P}}(l)^H \hat{\mathbf{v}}_{\hat{P},1}(l)\right)^H}{1 - \parallel \hat{\mathbf{v}}_{\hat{P},1}(l) \parallel^2}$	$\hat{P}^2$	(33)
	$\hat{\boldsymbol{\Phi}}_{\hat{P}}(l) \xrightarrow[\text{EVD}]{} \sum_{p=1}^{\hat{P}} \hat{\epsilon_p}\mathbf{u}_p\mathbf{u}_p^H$	$O(\hat{P}^3)$	
	$\hat{\varphi}_p = \arg(\hat{\epsilon}_p), \quad p = 1, \ldots, \hat{P}$	$\hat{P}$	
III-C.4	*Adaptive complex amplitude estimation*		
	$\Delta\hat{\varphi}_p(l) = \hat{\varphi}_p(l) - \hat{\varphi}_p(l-1), \quad p = 1, \ldots, \hat{P}$	$\hat{P}$	
	$[\boldsymbol{\Delta}(l)]_{q,p} = e^{j\Delta\hat{\varphi}_p(l)q}$		
	$\hat{\mathbf{Q}}(l) = \boldsymbol{\Delta}(l) \odot \mathbf{Q}(l-1)$	$N_f\hat{P}$	(43)
	$\hat{\mathbf{R}}(l) = \mathbf{R}(l-1)$		
	$\hat{\mathbf{R}}(l)\hat{\mathbf{g}}_l = \hat{\mathbf{Q}}^H(l)\hat{\mathbf{h}}(l) \xrightarrow[\text{Backsubstitution}]{} \hat{\mathbf{g}}_l$	$N_f\hat{P} + \hat{P}^2/2$	(41)

Our approach is based on the Bi-iterative SVD-3 subspace tracker [26, Table III], with the stabilizing modifications proposed in [27, p. 2995]. We chose this algorithm over other algorithms primarily due to its ability to track the eigenvalues (for model-order tracking), its guarantee of orthonormality of the singular vectors (for efficient adaptive ESPRIT implementation, which will be discussed later), and its low computational complexity of $O(N_f P_{max})$. We include the iteration step of the modified Bi-SVD 3 algorithm with the complexity estimates in Table III for the reader's convenience.

The main subspace updating step of the modified Bi-SVD 3 algorithm is given by

$$\left[\hat{\mathbf{V}}(l) \quad \mathbf{q}(l)\right] = \left[\hat{\mathbf{V}}(l-1) \quad \bar{\mathbf{h}}^{\perp}(l)\right] \mathbf{G}^{H}(l) \tag{25}$$

(c.f. Step 9 of Table III) where $\bar{\mathbf{h}}^{\perp}(l) = \frac{\hat{\mathbf{h}}^{\perp}(l)}{\|\hat{\mathbf{h}}^{\perp}(l)\|}$ can be interpreted as the normalized "innovations vector" that is orthogonal to the old dominant subspace spanned by $\hat{\mathbf{V}}(l-1)$ [26], and $\hat{\mathbf{V}}(l)$ is the $N_f \times P_{max}$ "updated" dominant subspace spanning matrix. The key to the low complexity of this algorithm is in the structure of the matrix $\mathbf{G}^{H}(l)$, which is a sequence of $2P_{max}-1$ Givens planar rotations [28], i.e.

$$\mathbf{G}(l) = \prod_{i=1}^{2P_{max}-1} \mathbf{G}_i(l) \tag{26}$$

where each $\mathbf{G}_i(l)$ is a Givens rotation matrix. We do not actually form the matrix $\mathbf{G}(l)$, but instead apply the $2P_{max}-1$ rotations sequentially to the desired matrix. Hence, this can be interpreted as an efficient "sequential rotational update" for the old subspace matrix $\hat{\mathbf{V}}(l-1)$ with complexity $O(N_f P_{max})$ [26].

2) Adaptive model-order estimation: Note that by squaring the estimated P_{max} singular values, we have an estimate of the dominant eigenvalues. However, in order to employ the MDL criterion, we also require an estimate of the rest of the $N_f - P_{max}$ non-dominant eigenvalues. Theoretically, the noise eigenvalues should all be equal to the noise power σ^2. Thus, we estimate these non-dominant eigenvalues by simply equating them to an estimate of the noise power. Note that $\hat{\mathbf{h}}^{\perp}(l)$ in Step 2 of Table III is the component of $\hat{\mathbf{h}}(l)$ that lies in the noise subspace [26]. The noise power estimate can then be derived as

$$\hat{\sigma}^2(l) = \beta\hat{\sigma}^2(l-1) + (1-\beta)\frac{\hat{\mathbf{h}}^{\perp}(l)^H \hat{\mathbf{h}}^{\perp}(l)}{N_f - P_{max}} \tag{27}$$

Therefore, our eigenvalue estimates can be written as

$$\hat{\lambda_i}^{f}(l) = \begin{cases} \hat{\sigma_i}^2(l), i \in [1, P_{max}] \\ \hat{\sigma}^2(l), \ i \in [P_{max}+1, N_f] \end{cases} \tag{28}$$

where $\hat{\sigma}_i(l)$ is the ith element of the singular value estimates given in Step 11 of Table III. Finally, by using the effective window length at time n_l, given as

$$\tilde{N}_t(l) = \frac{1-\beta^l}{1-\beta}, \tag{29}$$

TABLE III

BI-ITERATION SVD SUBSPACE TRACKER BI-SVD 3 [26] WITH MODIFICATIONS INTRODUCED IN [27, P. 2995]. $\mathbf{G}(l)$ IS A SEQUENCE OF $(2P_{max}-1)$ GIVENS ROTATIONS.

Input:	$\hat{\mathbf{h}}(l)$	
Output:	$\hat{\mathbf{V}}(l), \hat{\boldsymbol{\sigma}}(l), \mathbf{G}(l), \hat{\mathbf{h}}^{\perp}(l)$	
Step No.	**Operation**	**Complexity**
1	$\mathbf{d}(l) = \hat{\mathbf{V}}^H(l-1)\hat{\mathbf{h}}(l)$	$N_f P_{max}$
2	$\hat{\mathbf{h}}^{\perp}(l) = \hat{\mathbf{h}}(l) - \hat{\mathbf{V}}(l-1)\mathbf{d}(l)$	$N_f P_{max}$
3	$\mathbf{D}(l) = \mathbf{R}_B(l-1)\boldsymbol{\Theta}_A(l-1)$	$P_{max}^3/3$
4	$\begin{bmatrix} (1-\beta)^{1/2}\mathbf{d}^H(l) \\ \beta^{1/2}\mathbf{D}(l) \end{bmatrix} \xrightarrow{\text{Givens QR}} \begin{bmatrix} \mathbf{R}_B(l) \\ \mathbf{0}_{1\times P_{max}} \end{bmatrix}$	$2P_{max}^3$
5	$\mathbf{d}_R^H(l)\mathbf{R}_B(l) = \mathbf{d}^H(l) \quad \xrightarrow{\text{Back substitution}} \quad \mathbf{d}_R^H(l)$	$P_{max}^2/2$
6	$\mathbf{D}_R(l)\mathbf{R}_B(l) = \mathbf{D}(l) \quad \xrightarrow{\text{Back substitution}} \quad \mathbf{D}_R(l)$	$P_{max}^3/3$
7	$\mathbf{R}_A(l-1)\mathbf{D}_R(l) \quad \xrightarrow{\text{Extract upper triangular portion}} \quad \mathbf{T}$	
8	$\begin{bmatrix} \mathbf{R}_A(l) \\ \mathbf{0}_{1\times P_{max}} \end{bmatrix} = \mathbf{G}(l) \begin{bmatrix} \beta\mathbf{T} + (1-\beta)\mathbf{d}(l)\mathbf{d}_R^H(l) \\ (1-\beta) \parallel \hat{\mathbf{h}}^{\perp}(l) \parallel \mathbf{d}_R^H(l) \end{bmatrix}$	$4P_{max}^2$
9	$\begin{bmatrix} \hat{\mathbf{V}}(l) & \mathbf{q}(l) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{V}}(l-1) & \hat{\mathbf{h}}^{\perp}(l)/ \parallel \hat{\mathbf{h}}^{\perp}(l) \parallel \end{bmatrix} \mathbf{G}^H(l)$	$8N_f P_{max}$
10	$\mathbf{G}^H(l) \quad \xrightarrow{\text{Extract upper left } P_{max} \text{ square matrix}} \quad \boldsymbol{\Theta}_A(l)$	
11	$\hat{\boldsymbol{\sigma}}(l) = \text{diag}(\mathbf{R}_B(l)\boldsymbol{\Theta}_A(l))$	$P_{max}^3/3$

we have the adaptive model order

$$\hat{P}(l) = \arg \min_{1\le\mu\le P_{max}} -\log \left(\frac{\left(\prod_{k=\mu+1}^{N_f} \hat{\lambda}_k^f(l) \right)^{\frac{1}{N_f-\mu}}}{\frac{1}{N_f-\mu} \sum_{k=\mu+1}^{N_f} \hat{\lambda}_k^f(l)} \right)^{\tilde{N}_t(l)(N_f-\mu)} + \frac{1}{2}\mu(2N_f-\mu)\log \tilde{N}_t(l) \qquad (30)$$

Note that a similar adaptive MDL criterion has been developed in [29]. Their approach tracks a varying dimension $\hat{P}(l-1) + P_{aux}$, and increments by P_{aux} or decrements by any number the model order estimate $\hat{P}(l)$ based on the criterion. We track a constant dimension P_{max} of the dominant subspace, and use the adaptive MDL to determine $\hat{P}(l)$. We have verified through simulations that the constant dimension tracking approach is better in terms of speed of convergence. This can be intuitively explained by the fact that there is a delay for the subspace tracker (approximately the effective window length) to be able to provide a reliable estimate of the eigenvalues, especially when the dimension of the dominant subspace has recently changed. Thus, by fixing the dimension of the dominant subspace being tracked, all the eigenvalue estimates for the P_{max} dimensions are equally reliable once the initial window length has been reached. The tracking of a fixed dimensionality is also used by the adaptive ESPRIT algorithm of [30], but they use a simple thresholding approach where the model order is determined by the number

of eigenvalues with magnitudes exceeding the noise estimate by a certain factor. Although quite simple in concept, no clear criteria for choosing the factor is available, and thus extensive simulations are required to assess a good value for a particular scenario. Thus, our approach is advantageous in that it uses well established information theoretic criteria, thus doing away with ad-hoc choosing of threshold factors.

3) Adaptive time-delay estimation using adaptive ESPRIT: Fast adaptive ESPRIT algorithms have already been investigated in the past, and could potentially be used in our framework in a straightforward manner. In [30], a class of fast recursive ESPRIT algorithms for adaptive source localization based on subspace tracking and adaptive rank reduction was developed. Their fastest method has complexity $O(Nr)$, where N is the number of observations, and r is the number sources being estimated. However, it has been observed by other researchers [27] [31], and validated by our own simulations, that the $O(Nr)$ method in [30] becomes unstable for sinusoidal frequencies that are very close, which is not uncommon in our scenario. More stable adaptive ESPRIT algorithms with similar complexity have also been proposed in [31], but these have the disadvantage of not being able to adaptively track the time-varying model order. Thus, we develop our own fast adaptive ESPRIT algorithm for time-delay estimation of $O(Nr)$ complexity that is both stable and model-order adaptive.

Let $\hat{\mathbf{V}}_{\hat{P}(l)}(l)$ be the matrix composed of the dominant (leftmost) $\hat{P}$ eigenvector columns of $\hat{\mathbf{V}}(l)$, the output of the subspace tracker from Table III. Perform the partitions

$$\hat{\mathbf{V}}_{\hat{P}}(l) = \begin{bmatrix} \hat{\mathbf{V}}_{\hat{P},1}(l) \\ \hat{\mathbf{v}}^H_{\hat{P},1}(l) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{v}}^H_{\hat{P},2}(l) \\ \hat{\mathbf{V}}_{\hat{P},2}(l) \end{bmatrix} \tag{31}$$

where $\hat{\mathbf{V}}_{\hat{P},1}(l)$ and $\hat{\mathbf{V}}_{\hat{P},2}(l)$ are the $(N_f - 1) \times \hat{P}$ upper and lower submatrices of $\hat{\mathbf{V}}_{\hat{P}}(l)$; $\hat{\mathbf{v}}^H_{\hat{P},1}(l)$ and $\hat{\mathbf{v}}^H_{\hat{P},2}(l)$ are the bottommost and topmost row vectors of the same matrix. We can write the adaptive equivalent of the ESPRIT spectral matrix (15) as

$$\hat{\mathbf{\Phi}}_{\hat{P}}(l) = \left(\hat{\mathbf{V}}^H_{\hat{P},1}(l)\hat{\mathbf{V}}_{\hat{P},1}(l)\right)^{-1} \hat{\mathbf{V}}^H_{\hat{P},1}(l)\hat{\mathbf{V}}_{\hat{P},2}(l) \tag{32}$$

Our initial objective is to derive an $O(N_f L_{max})$ updating scheme for $\hat{\mathbf{\Phi}}_{\hat{P}}(l)$, where we proceed similarly as in [31, Sec. 3.1]. Since $\hat{\mathbf{V}}_{\hat{P}}(l)$ has orthonormal columns, we have $\hat{\mathbf{V}}^H_{\hat{P},1}(l)\hat{\mathbf{V}}_{\hat{P},1}(l) = \mathbf{I}_{\hat{P}} - \hat{\mathbf{v}}_{\hat{P},1}(l)\hat{\mathbf{v}}^H_{\hat{P},1}(l)$, which is simply a rank-one modification of the $\hat{P} \times \hat{P}$ identity matrix. Using the matrix inversion lemma [32], we have $\left(\hat{\mathbf{V}}^H_{\hat{P},1}(l)\hat{\mathbf{V}}_{\hat{P},1}(l)\right)^{-1} = \mathbf{I}_{\hat{P}} + \frac{\hat{\mathbf{v}}_{\hat{P},1}(l)\hat{\mathbf{v}}^H_{\hat{P},1}(l)}{1-\|\hat{\mathbf{v}}_{\hat{P},1}(l)\|^2}$. Using this in (32), we have

$$\hat{\mathbf{\Phi}}_{\hat{P}}(l) = \mathbf{\Upsilon}_{\hat{P}}(l) + \frac{\hat{\mathbf{v}}_{\hat{P},1}(l)\left(\mathbf{\Upsilon}_{\hat{P}}(l)^H \hat{\mathbf{v}}_{\hat{P},1}(l)\right)^H}{1 - \| \hat{\mathbf{v}}_{\hat{P},1}(l) \|^2} \tag{33}$$

$$\mathbf{\Upsilon}_{\hat{P}}(l) = \hat{\mathbf{V}}^H_{\hat{P},1}(l)\hat{\mathbf{V}}_{\hat{P},2}(l) \tag{34}$$

Note that the spectral matrix (33) is simply an $O(\hat{P}^2)$ rank-one update to (34), but computing (34) directly requires $O(N_f\hat{P}^2)$. In [31, Sec. 3.1], they proceed to use subspace trackers that employ rank-one updates to derive an $O(N_fP_{max})$ update for (34). Unfortunately, these subspace trackers do not track the actual eigenvalues, and are therefore unable to estimate and track the time-varying model order $\hat{P}$. Thus, we proceed to derive a novel $O(N_fP_{max})$ recursion for (34).

The main subspace updating step of the modified Bi-SVD 3 is given by (cf. Step 9 of Table III)

$$\begin{bmatrix}\hat{\mathbf{V}}(l) & \mathbf{q}(l)\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{V}}(l-1) & \overline{\hat{\mathbf{h}}^{\perp}(l)}\end{bmatrix}\mathbf{G}^H(l) \tag{35}$$

where $\overline{\hat{\mathbf{h}}^{\perp}(l)} = \frac{\hat{\mathbf{h}}^{\perp}(l)}{\|\hat{\mathbf{h}}^{\perp}(l)\|}$. Pre-multiply both sides of (35) by $[\mathbf{I}_{N_f-1} \quad \mathbf{0}_{(N_f-1)\times 1}]$ and $[\mathbf{0}_{(N_f-1)\times 1} \quad \mathbf{I}_{N_f-1}]$ to get the $(N_f-1)\times(P_{max}+1)$ top submatrix and the $(N_f-1)\times(P_{max}+1)$ bottom submatrix

$$\begin{bmatrix}\hat{\mathbf{V}}_1(l) & \mathbf{q}_1(l)\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{V}}_1(l-1) & \overline{\hat{\mathbf{h}}_1^{\perp}(l)}\end{bmatrix}\mathbf{G}^H(l) \tag{36}$$

$$\begin{bmatrix}\hat{\mathbf{V}}_2(l) & \mathbf{q}_2(l)\end{bmatrix} = \begin{bmatrix}\hat{\mathbf{V}}_2(l-1) & \overline{\hat{\mathbf{h}}_2^{\perp}(l)}\end{bmatrix}\mathbf{G}^H(l) \tag{37}$$

respectively. Taking the Hermitian transpose of both sides of (36) and pre-multiplying it to (37), we have

$$\begin{bmatrix}\boldsymbol{\Upsilon}(l) & * \\ * & *\end{bmatrix} = \mathbf{G}(l)\begin{bmatrix}\boldsymbol{\Upsilon}(l-1) & \hat{\mathbf{V}}_1^H(l-1)\overline{\hat{\mathbf{h}}_2^{\perp}(l)} \\ \overline{\hat{\mathbf{h}}_1^{\perp H}(l)}\hat{\mathbf{V}}_2(l-1) & \overline{\hat{\mathbf{h}}_1^{\perp H}(l)\hat{\mathbf{h}}_2^{\perp}(l)}\end{bmatrix}\mathbf{G}^H(l) \tag{38}$$

$$\boldsymbol{\Upsilon}(l) = \hat{\mathbf{V}}_1^H(l)\hat{\mathbf{V}}_2(l) \tag{39}$$

where '$*$' represents unused quantities. Observe that once we initialize (39), we no longer need to perform this $O(N_fP_{max}^2)$ matrix multiplication, since we can simply update it as a single $P_{max}\times P_{max}$ matrix. Furthermore, since $\mathbf{G}(l)$ represents a sequence of $2P_{max}-1$ Givens plane rotations [28], we can update $\boldsymbol{\Upsilon}_{P_{max}}(l)$ in just $8P_{max}$ operations. Also notice that by partitioning $\hat{\mathbf{V}}_1(l) = [\hat{\mathbf{V}}_{\hat{P},1}(l) \quad *]$ and $\hat{\mathbf{V}}_2(l) = [\hat{\mathbf{V}}_{\hat{P},2}(l) \quad *]$,

$$\boldsymbol{\Upsilon}_{P_{max}}(l) = \begin{bmatrix}\boldsymbol{\Upsilon}_{\hat{P}}(l) & * \\ * & *\end{bmatrix} \tag{40}$$

where $\boldsymbol{\Upsilon}_{\hat{P}}(l)$ is given by (34). Thus, by extracting the $\hat{P}\times\hat{P}$ upper-left submatrix of the left hand side of (38) and using it in (33), we arrive at an efficient adaptive ESPRIT spectral matrix updating algorithm with principal complexity $O(N_fP_{max})$. Finally, we extract the radian phase of each of the complex eigenvalues of the small $\hat{P}\times\hat{P}$ matrix $\hat{\boldsymbol{\Phi}}_{\hat{P}}(l)$ (33) (requiring $O(\hat{P}^3)$ operations), giving an estimate of the time-delays $\{\hat{\varphi}_p\}_{p=1}^{\hat{P}}$.

4) Adaptive complex amplitude estimation: Given the time delay estimates $\{\hat{\varphi}_p\}_{p=1}^{\hat{P}}$ (which is assumed correct), the maximum-likelihood estimate for the $\hat{P}$-length complex amplitude vector $\hat{\mathbf{g}}(l)$ of complex amplitudes is also the LS solution (cf. (18)), which can be computed by back substitution of

$$\mathbf{R}(l)\hat{\mathbf{g}}(l) = \mathbf{Q}^H(l)\hat{\mathbf{h}}(l) \tag{41}$$

where $\mathbf{Q}(l)\mathbf{R}(l) = \hat{\mathbf{E}}(l)$ is the "skinny" QR[9] decomposition [28] of the $N_f \times \hat{P}$ Fourier transform matrix $\hat{\mathbf{E}}(l)$ (see (18)) formed using the estimated time-delays. Computing (41) from scratch requires $O(N_f\hat{P}^2)$, and thus we propose a recursive updating solution to approximate the QR decomposition.

Let $\Delta\hat{\varphi}_i(l) = \hat{\varphi}_i(l) - \hat{\varphi}_i(l-1)$ denote the difference between the time-delay estimates at two consecutive time steps, and let $\mathbf{\Delta}(l)$ be an $N_f \times \hat{P}$ perturbation matrix with elements $[\mathbf{\Delta}(l)]_{q,p} = e^{j\Delta\hat{\varphi}_p(l)q}$. Suppose that we have at hand the previous "skinny" QR decomposition of $\mathbf{E}(l-1) = \mathbf{Q}(l-1)\mathbf{R}(l-1)$. Then,

$$\begin{aligned} \mathbf{E}(l) &= \mathbf{\Delta}(l) \odot \mathbf{E}(l-1) \\ &= \mathbf{\Delta}(l) \odot \mathbf{Q}(l-1)\mathbf{R}(l-1) \end{aligned} \tag{42}$$

where $\odot$ is the Hadamard (elementwise) product. It is reasonable to assume that the time-delay differences across consecutive time steps are quite small, i.e. $|\Delta\hat{\varphi}_p(l)| \approx 0$. Hence, an approximation to the QR factors $\mathbf{E}(l) = \mathbf{Q}(l)\mathbf{R}(l)$ can be given as

$$\mathbf{Q}(l) \approx \hat{\mathbf{Q}}(l) = \mathbf{\Delta}(l) \odot \mathbf{Q}(l-1) \tag{43}$$

which is now only approximately orthogonal, and $\mathbf{R}(l) = \mathbf{R}(l-1)$, which is still upper triangular. This update can be performed in $O(N_f\hat{P})$ operations. In our numerical experiments, we saw that the normalized MSE[10] is negligible ($< 10^{-2}$) up until around 10% of the previous delay value. We can simply compute the QR decomposition directly when the deviation goes beyond this.

The above discussion assumes that the model order is static between iterations, which is the case most often encountered since the time delays are slowly varying. However, it is conceivable that there are small model order changes (e.g. 1 or 2) from one iteration to the next. We can use the same framework above in conjunction with efficient column deleting and appending procedures for the QR decomposition [28], which have complexity $O(\hat{P}^2)$ and $O(N_f\hat{P})$ respectively.

[9]A regular QR decomposition of $\mathbf{E} \in \mathbb{C}^{m\times n}, m > n$ with $\mathbf{Q} \in \mathbb{C}^{m\times m}$ and $\mathbf{R} \in \mathbb{C}^{m\times n}$ can be also be written in partitioned form as $\mathbf{E} = [\mathbf{Q}_1 \quad \mathbf{Q}_2]\begin{bmatrix}\mathbf{R}_1 \\ \mathbf{0}\end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1$ where $\mathbf{Q}_1 \in \mathbb{C}^{m\times n}$ has orthogonal columns and $\mathbf{R}_1 \in \mathbb{C}^{n\times n}$ is a square upper-triangular matrix. $\mathbf{Q}_1\mathbf{R}_1$ is termed the "skinny" QR decomposition

[10]We define the NMSE as $\mathbb{E}\{\| \hat{\mathbf{g}}(l) - \tilde{\mathbf{g}}(l) \|^2 / \| \hat{\mathbf{g}}(l) \|^2\}$ where $\hat{\mathbf{g}}(l)$ is the solution to (41) using the actual QR decomposition computation, and $\tilde{\mathbf{g}}(l)$ the solution when using the update approximation.

D. Doppler Frequency Tracking and Channel Prediction

In tracking Doppler frequencies, we can use the same procedure for each path $p = 1, \ldots, \hat{P}$, with $\hat{\mathbf{g}}_p = [\hat{g}_p(l), \ldots, \hat{g}_p(l-(N_t-1))]^T$ as the N_t-length observation vector for the pth path, and track the $R_{p,max}$-dimension dominant subspace. The principal computational complexity can be analyzed similar to the time-delay estimation step, and is thus omitted due to space constraints. It is given by $O(N_t \hat{P} R_{p,max})$.

IV. Mean-square Error Bounds

In this section, we derive the CRLB of the prediction MSE for OFDM channel prediction[11]. We then derive a simple closed-form expression for the asymptotic CRLB for large N_t and N_f to provide better insight into the effect of various parameters on the MSE lower bound.

For notational convenience, we collect the parameters to be estimated in (6) into length R vectors, e.g.

$$\boldsymbol{\alpha} = [\alpha_{1,1}, \ldots, \alpha_{R_1,1}, \ldots, \alpha_{R_P,1}, \ldots, \alpha_{R_P,P}]^T \tag{44}$$

and re-index each parameter by its location in the vector, e.g. $\alpha_i = [\boldsymbol{\alpha}]_i, i = 1, \ldots, R$. Let $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_R^T]^T$ with $\boldsymbol{\theta}_i = [\alpha_i, \phi_i, \omega_i, \varphi_i]^T$ denote the $4R$-length vector of channel parameters . We rewrite the physical channel model (5) as

$$H(n,k,\boldsymbol{\theta}) = \sum_{i=1}^{R} H_i(n,k,\boldsymbol{\theta}_i) \tag{45}$$

with $H_i(n,k,\boldsymbol{\theta}_i) = \alpha_i e^{j\left(\phi_i + \varphi_i \frac{N_f}{2}\right)} e^{j\left(\frac{\omega_i}{D_t} n - \frac{\varphi_i}{D_f} k\right)}$. The dependence of the channel response on the parameter vector $\boldsymbol{\theta}$ is made explicit in our notation. Note that (45) is a continuous function of the unknown parameters; hence, by the invariance principle of the ML estimate, the CRLB is [33, Ch. 6.4]

$$e(n,k) = \frac{\partial H(n,k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^H \mathbf{I}_{2D}^{-1}(\boldsymbol{\theta}) \frac{\partial H(n,k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{46}$$

$$\frac{\partial H(n,k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[\frac{\partial H_1(n,k,\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}^H, \ldots, \frac{\partial H_{R-1}(n,k,\boldsymbol{\theta}_{R-1})}{\partial \boldsymbol{\theta}_{R-1}}^H \right]^H \tag{47}$$

$$\frac{\partial H_i(n,k,\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}_i} = H_i(n,k,\boldsymbol{\theta}_i) \left[\frac{1}{\alpha_i}, j, j\frac{n}{D_t}, -j\frac{k-\bar{k}}{D_f} \right]^H \tag{48}$$

[11]For simplicity, we omit the effect of the pulse-shaping filter frequency response, but this can be included in a straightforward manner as nuisance parameters.

and where $\bar{k} = \frac{N_f}{2} D_f$ and $\mathbf{I}_{2D}^{-1}(\boldsymbol{\theta})$ is the CRLB matrix for 2D sinusoidal parameter estimation, which can be found in [34, Appendix A]. References [19] and [7] derived a similar MSE bound as (46) by using a different and perhaps more tedious approach (first-order Taylor expansion of the MSE function).

Although the bound in (46) can be used to study the effects of various parameters on the prediction MSE, its expression is not readily interpretable. Furthermore, since the bound depends on the actual parameter vector $\boldsymbol{\theta}$, extensive Monte-Carlo simulations that generate realizations of this parameter vector assuming a certain probability distribution is required to assess the performance. And with the objective of investigating the effect of various channel prediction parameter configurations, this requires performing a Monte-Carlo simulation for each parameter configuration. Approximately $8R^2 N_t N_f + O(R^3)$ operations are required to compute the CRB for each channel realization. This is a heavy computational burden even for the case of offline analysis, esp. when R, N_t, and N_f are large. Thus, we derive a simple closed-form expression for the lower bound on prediction MSE in the asymptotic case of large N_t and N_f. Consider the asymptotic MSE bound given as

$$\begin{aligned} \tilde{\epsilon}(n,k) &= \lim_{\min(N_t,N_f)\to\infty} \epsilon(n,k) \\ &= \frac{\partial H(n,k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^H \tilde{\mathbf{I}}_{2D}^{-1} \frac{\partial H(n,k,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \end{aligned} \tag{49}$$

where $\tilde{\mathbf{I}}_{2D}^{-1} = \text{diag}\{\mathbf{K}_1, \ldots, \mathbf{K}_R\}$ with

$$\mathbf{K}_i = \begin{bmatrix} \frac{\sigma^2}{2N_t N_f} & 0 & 0 & 0 \\ 0 & \frac{7\sigma^2}{2\alpha_i^2 N_t N_f} & \frac{-3\sigma^2}{\alpha_i^2 N_t^2 N_f} & \frac{-3\sigma^2}{\alpha_i^2 N_t N_f^2} \\ 0 & \frac{-3\sigma^2}{\alpha_i^2 N_t^2 N_f} & \frac{6\sigma^2}{\alpha_i^2 N_t^3 N_f} & 0 \\ 0 & \frac{-3\sigma^2}{\alpha_i^2 N_t N_f^2} & 0 & \frac{6\sigma^2}{\alpha_i^2 N_t N_f^3} \end{bmatrix} \tag{50}$$

is the block diagonal asymptotic CRB for 2D superimposed complex exponential parameter estimation [35]. After further simplification, we have

$$\tilde{\epsilon}(n,k) = \sigma^2 R \left(\underbrace{\frac{4}{N_t N_f}}_{(a)} - \underbrace{\frac{6n}{N_t^2 N_f D_t}}_{(b)} + \underbrace{\frac{6n^2}{N_t^3 N_f D_t^2}}_{(c)} + \underbrace{\frac{6\left(k-\bar{k}\right)}{N_t N_f^2 D_f}}_{(d)} + \underbrace{\frac{6\left(k-\bar{k}\right)^2}{N_t N_f^3 D_f^2}}_{(e)} \right) \tag{51}$$

where the individual terms correspond to

(a) Amplitude and phase estimation error variance

(b) Doppler frequency and phase estimation error cross covariance

(c) Doppler frequency estimation error variance

(d) Time delay and phase estimation error cross covariance

(e) Time delay estimation error variance

Using this simple expression for the lower bound on MSE, we could easily deduce the impact of the various parameters on the MSE, e.g.

1) The bound increases linearly with increasing noise variance σ^2 and number of 2D sinusoidal rays R. This is intuitively satisfying, and agrees with previous results that dense multipath channel environments, i.e. R large, are the hardest to predict [7].
2) The contribution to the overall MSE from the error variances corresponding to the estimate of frequencies ω_i and φ_i grow quadratically with n and $|k|$, emphasizing the importance of estimating these accurately.
3) In general, N_t and N_f and the downsampling factors D_t and D_f should be chosen as large as possible to decrease the MSE bound, but are of course subject to limitations imposed by other system considerations such as complexity, training overhead, and the time window for which the constant-parameter sinusoidal model is valid.

V. Simulation results and comparisons

We provide simulation results for an outdoor mobile OFDM system based on the IEEE 802.16e standard [1] with $N = 512$ subcarriers, $N_u = 426$ used subcarriers, $N_{cp} = 32$ cyclic prefix length, $B = 5$ MHz bandwidth, $F_c = 2.6$ GHz carrier frequency, $F_s = 5.712$ Ms/s sampling frequency, and $v = 75$ kph mobile velocity. The OFDM symbol period is given as $T_{sym} = (N + N_{cp})/F_s = 95.24\mu s$, and the subcarrier spacing is $\Delta f = F_s/N = 11.16$ kHz. We also simulate the square-root raised cosine transmit and receive pulse-shaping FIR filters with $2\times$ interpolation, roll-off of $0.15 < (1 - N_u/N) = (1 - 426/512)$, and filter length of 16 taps [17]. The effect of this pulse-shaping filter on the algorithm is minimal, since the passband is almost flat, and the FIR filter simply causes a constant shift on the time delays.

We simulate the delay spread of the channel using various ITU and COST-207 power delay profiles [17] as illustrated in Fig. 2(a), with τ_{max} ranging from 2.5 μs to 7 μs. For each tap in the power delay profile, we simulate the fading channel using the modified Jakes' method [18] with $R_p = 16$ rays per path, which was shown to be sufficient to match the desired second-order statistics [18]. Note that in the modified Jakes' method of [18], the Doppler frequencies are generated randomly per realization as $f_{r,p} = \pm f_{max} \cos((2\pi r + \theta - \pi)/(4R_p)), r = 1, \ldots, R_p$ where $f_{max} = v/(3 \times 10^8/F_c) \approx 180$ Hz is the maximum Doppler frequency, and $\theta \sim U[-\pi, \pi)$. Thus, the frequencies are *not* equispaced, and difficult cases of two frequencies being close together is actually not uncommon, esp. when the number of rays chosen is large (see Fig. 2(b)).

We assume a frame-based transmission with frame length of $T_{frame} = 2$ ms (equivalent to $D_t = \lfloor T_{frame}/T_{sym} \rfloor = 21$ OFDM symbols), where a frame preamble with $N_f = 142$ pilot subcarriers inserted

every $D_f = N_u/N_f = 3$ subcarriers transmitted at the beginning of each frame. In order to determine a suitable value for N_t, we recall from the previous section that N_t should be chosen such that it is within the time-window of the validity of the sinusoidal model. In [9], a rule-of-thumb for the duration of validity of this model is given by

$$T = \sqrt{\frac{\lambda r_{min}}{3v^2}} \tag{52}$$

where $\lambda = 3 \times 10^8/F_c \approx 0.115$ m is the wavelength, v is the velocity in m/s, and r_{min} is the separation of the mobile to its nearest relevant scatterer in m. Assuming a macro-cell outdoor mobile scenario where the dominant scatterers are huge structures like faraway buildings and hills, $r_{min} = 500$ m is a conservative estimate [9], which gives us $T \approx 210$ ms, which is approximately 105 frames. Thus, we set $N_t = 100$ as the default number of pilot symbols[12]. The autocorrelation matrix dimension for time-delay estimation in (11) is chosen as $K = \lceil \frac{3}{5} N_f \rceil$, and similarly for Doppler frequency estimation in (19) as $I = \lceil \frac{3}{5} N_t \rceil$, according to the rule-of-thumb proposed in [24]. Note that these are the default parameters, and are varied accordingly to study the effects of various parameters on the system performance.

A. Acquisition Stage

We first investigate the performance of the acquisition part of our algorithm (Sec. III-A) without the pulse-shaping filters, and assuming that the time-delays and Doppler frequencies are indeed stationary within the acquisition window, in order to compare the performance with the bounds that are derived under this assumption. We compare this algorithm to using standard linear prediction on the downsampled time-domain channel taps in (16), similar to the approach in [10]. We used the Burg method to compute the AR coefficients in our simulations, with a filter order of $p = \lceil 3/5N_t \rceil$ to match the complexity of our algorithm. We call this method the Burg Prediction method. Following the convention of previous work [9] [8], we normalize the prediction length in terms of the number of wavelengths λ. For example, predicting 1 frame ahead is equivalent to $L = vT_{frame}/\lambda$ wavelengths. Fig. 3(a) shows the prediction normalized mean square error (NMSE) results for predicting 2λ ahead using the Vehicular-A power delay profile. Fig. 3(b) shows the NMSE comparisons for varying the prediction horizon with SNR$= 7.5$ dB. Our proposed algorithm outperforms the Burg Prediction method for all SNR values and prediction lengths, and the advantage is more pronounced as the SNR and the prediction lengths increase. We also plot the CRLB and ACRLB for comparison, and our method is quite close to these bounds under these

[12]Although $N_t N_f = 14200$ may seem huge, if we use 4 bytes (e.g. single-precision floating point) per channel data point, the memory requirement is just 50 kB which is not excessive given the advanced DSPs used in current base stations.

somewhat idealized channel assumptions. This is consistent with the results in [6] [8] for single-carrier flat-fading channel prediction.

Next, we study the performance of the acquisition part of our algorithm under more realistic scenarios. First, we increase the mobile velocity to 100 kph such that the time-delays and Doppler frequencies be slowly linearly varying, i.e. $\tau_p(t) = \tau_p + C_\tau t$ and $f_{r,p}(t) = f_{r,p} + C_f t$ where C_τ and C_f are small constants. This models the scenario where the acquisition window is larger than the recommended rule of thumb (52), causing the geometry of the scattering field experienced by the mobile to change significantly enough so as to affect the time-delays and Doppler frequencies [9]. We also model the pulse-shaping filters in these experiments.

We compare our methods with the FFT-based implementation of the MMSE algorithm given in [12, Sec. III-C], which we call (FFT-MMSE). In this method, an IFFT is performed on the pilot subcarriers, and linear prediction filters are used for each time-domain tap for the first N_{cp} taps corresponding to the length of the cyclic-prefix, thus capturing most of the energy of the delay-spread channel. An FFT is then performed on the predicted channel to get the frequency domain channel response. Since a straightforward FFT is used in FFT-MMSE, we allowed all the N_u used subcarriers to be used as its pilot subcarriers. However, we still retain the pilot pattern assumed in the system model in the implementation of our algorithm, hence granting a slight advantage to FFT-MMSE. We compared the acquisition part of our proposed algorithm with that of [12] using 1000 channel realizations, the results of which are shown in Fig. 4. The dashed-x's correspond to [12], and the solid-squares correspond to our proposed algorithm. The four NMSE curves for each algorithm from lowest to highest correspond to predicting $\{.25\lambda, .5\lambda, \lambda, 2\lambda\}$ (equivalent to predicting $\{.5, 1, 2, 4\}$ frames ahead) respectively for the various SNRs. We considered the following six cases:

1) TU-Sparse, $R_p = 16$, $N_t = 100$, $N_f = 142$: This is the baseline case from which we compare the other cases, with $N_{cp} = 32$. Note that our algorithm outperforms FFT-MMSE in all SNRs and prediction lengths. Also note that our method is less sensitive to prediction length compared to FFT-MMSE, and this can be observed in all the other cases as well.
2) BU-Sparse, $R_p = 16$, $N_t = 100$, $N_f = 142$: We increase the delay spread in this case, such that the cyclic prefix length is also increased to $N_{cp} = 64$. We can see that our method is insensitive to the increased delay spread, since it picks out the dominant multipaths which remains at $P = 6$. On the other hand, FFT-MMSE is adversely affected, especially for long prediction lengths. This is because the increased N_{cp} increases the noise that creeps into the time-domain estimates from the IFFT operation, and the noise propagation makes the MSE for long prediction lengths worse.
3) TU-Dense, $R_p = 16$, $N_t = 100$, $N_f = 142$: We increase the density of the multipath taps to $P = 12$ in this

case, while retaining the maximum delay spread. We can see that we perform worse than FFT-MMSE at low SNR and short prediction lengths, primarily due to the inability of the MDL to detect the weaker paths embedded in noise. However, we still outperform FFT-MMSE significantly at medium to high SNR and long prediction lengths.

4) TU-Sparse, $R = 32$, $N_t = 100$, $N_f = 142$: In this case, we increase the number of rays per multipath tap to $R_p = 32$. Our method is adversely affected in this case, and is characteristic of sinusoidal modeling based channel prediction approaches (see e.g. [7]). This is due to the presence of more closely-spaced sinusoids, whose parameters are much harder to estimate. However, we still outperform FFT-MMSE in medium to high SNRs and long prediction lengths.
5) TU-Sparse, $R_p = 16$, $N_t = 60$, $N_f = 142$: In this case, we decrease the number of time-domain pilots to 60. We see that the detrimental effect on MSE is minimal, and is seen only for low SNRs. As long as a sufficient number of pilots is chosen to prevent statistical overfitting, the performance is just as good as further increasing N_t.
6) TU-Sparse, $R_p = 16$, $N_t = 100$, $N_f = 42$: In this case, we decreased the number of frequency domain pilots. We can see that the FFT-MMSE is unchanged because we use all N_u pilots for this method all the time, whereas our method is affected because the decreased number N_f increases the error in time-delay estimation.

Note that in all of the above cases, the performance of our algorithm and the FFT-MMSE algorithm, in contrast to Fig. 3(a) and the results in [12], saturates at high SNR and deviates from the ACRLB. This is primarily due to the non-stationarity of the frequency parameters, causing the MSE to be dominated by the model mismatch at high SNR, and also breaking the assumptions for which the CRLB bounds were derived. This makes choosing the appropriate acquisition time window crucial to the performance of the algorithm. This also implies that when the number of rays R_p is high, it become very difficult to predict the channel, since we need N_t to be large, but not too large such that the stationarity no longer holds. This phenomenon has been similarly observed in [7].

There are several reasons our method outperforms FFT-MMSE, although the linear MMSE is known to be optimal for Gaussian channels. First, note that the FFT-MMSE is itself an approximation, since the IFFT is performed to get the time-domain channel taps. This is due to the unknown channel information in the guard subcarriers at the edges, and more importantly, due to the non-sample spaced nature of the channel, which causes the time-domain channel energy to be greater than the length of the CP (see Fig.2(a)). Furthermore, since our channel is simulated using a small number of rays, the channel statistics cannot be assumed Gaussian. Also, we subjected both methods to realistic autocorrelation estimation procedures, and thus the second-order statistics are not fully known, which once again deviates from the

optimality of MMSE. More importantly, the non-stationarity of the delays and Doppler frequencies is also a deviation from the assumptions used in [12].

B. Tracking stage

In order to assess the performance of the tracking portion of our algorithm, we ran the acquisition stage of our algorithm for the first $N_t = 100$ frames, followed by 500 frames using the tracking stage, where we predict one frame (2 ms) ahead. We use $\beta = 0.99$, and the default simulation parameters using the Vehicular-A channel model. In order to simulate the non-stationary time-varying channel, we allow both time-delays and Doppler frequencies to vary linearly over time similar to the previous section. Fig. 5(a) shows the time-delay tracking results for one channel realization for all 6 paths at an SNR = 15 dB. Fig. 5(b) shows the Doppler frequency tracking results for the first 4 Doppler frequencies of the 6th (weakest) path for the same channel realization. We can observe that our algorithm is able to track the time-varying nature of the time-delays and Doppler frequencies.

We compare the performance of our proposed algorithm with the algorithm proposed in [12] in training-based mode. After designing the prediction filter using the block FFT-MMSE method, conventional LMS and RLS adaptive filters are then used to track the non-stationary channel. Note that we investigate only the *tracking* capability of these standard adaptive filters versus our adaptive ESPRIT based tracking methods. We ran this algorithm with $P = I = 60$ to match the complexity of our algorithm. Figure 6 shows the MSE performance for 20 channel realizations, with 500 frames per realization, for mobile velocities 30 and 75 kph (0.14λ and 0.36λ). It can be seen that our algorithm outperforms [12] except for the case of very low SNR, where the MDL model-order selection is unable to detect the weakest paths, resulting in higher MSE when compared to using N_{cp} time domain taps.

C. Complexity Analysis

Table IV summarizes the computational complexity of the proposed adaptive channel prediction algorithm. We compare this with the adaptive algorithm using LMS and RLS in [12], initialized with the reduced complexity FFT-MMSE prediction used in the previous experiments [12]. Typically, the parameters corresponding to the time delay estimation step have relations $N > N_f \approx K \gg P_{max} \geq \hat{P}$, the parameters corresponding to the time delay estimation step have relations $N_t \approx I \gg R_{p,max} \geq \hat{R}_p$, and $N_{cp} > \hat{P}$, where N_{cp} is the cyclic prefix length. We also have $\hat{R} = \sum_p \hat{R}_p$ as the total number of estimated rays. In the initialization stage, the proposed algorithm is more complex than that of [12]. In the tracking stage, our algorithm is slightly more complex than LMS, but is much less complex than

TABLE IV

COMPUTATIONAL COMPLEXITY

Stage	**Proposed Algorithm**	**Algorithm from [12]**
Initialization	$O(N_f^3 + \hat{P}N_t^3)$	$O(N \log_2 N + N_{cp}N_t^2)$
Tracking	$O(N_f P_{max} + N_t \hat{P} R_{p,max})$	LMS - $O(N_{cp}N_t)$ RLS - $O(N_{cp}N_t^2)$
Prediction	$O(N\hat{P} + \hat{R})$	$O(N \log_2 N + N_{cp}N_t)$

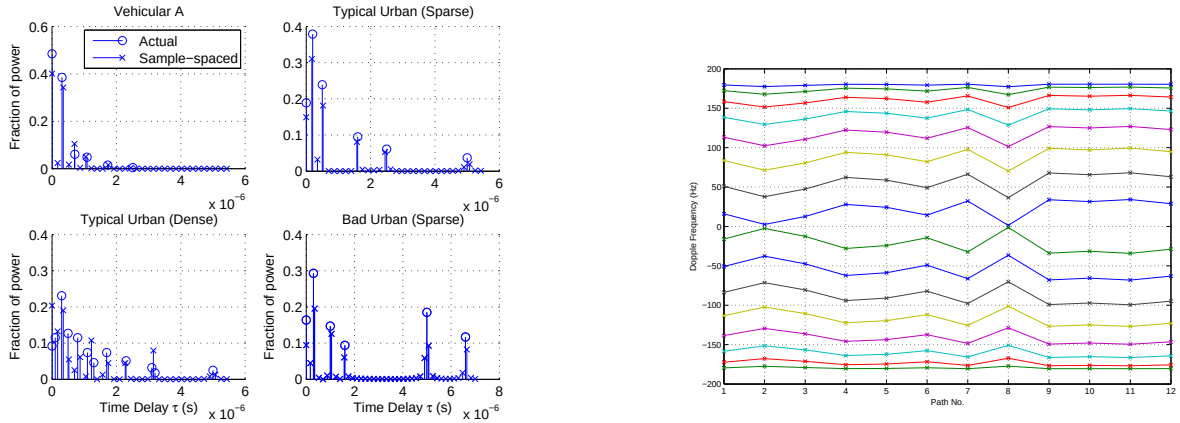
RLS. In the prediction stage, [12] is more complex than our algorithm. The initialization needs to be performed only once, whereas prediction and tracking are performed for each OFDM symbol.

VI. DISCUSSIONS AND CONCLUSION

We have proposed a novel OFDM channel prediction algorithm assuming a doubly selective ray-based physical channel model, where the model parameters are stationary within the initial acquisition and prediction time window, and is slowly time-varying beyond this window. The proposed algorithm performs channel model parameter acquisition using a 2-Step 1D ESPRIT algorithm as a first stage, and channel prediction via model extrapolation as a second stage. Since the channel model parameter acquisition has cubic complexity, we also proposed a linear complexity channel parameter tracking algorithm based on an improved adaptive ESPRIT algorithm to continuously adapt to the time-varying channel model parameters. We also derived the CRLB and asymptotic CRLB for the MSE in OFDM channel prediction.

The key assumption in our algorithm is the specular scattering wireless channel model, where time-delays are shared by the propagating rays. The use of this model can be more easily justified in outdoor macro-cell propagation scenarios (e.g. suburban and rural environments), where a few distinct scatterers contribute most to the fading channel. We have shown through extensive simulations that our proposed OFDM channel prediction algorithm has better MSE performance while maintaining similar computational complexity than previous methods in these scenarios, especially for medium-high SNRs and long prediction lengths. Thus, our method can be seen as a complement to the existing prediction methods that are more suitable for dense multipath channels with diffuse scattering.

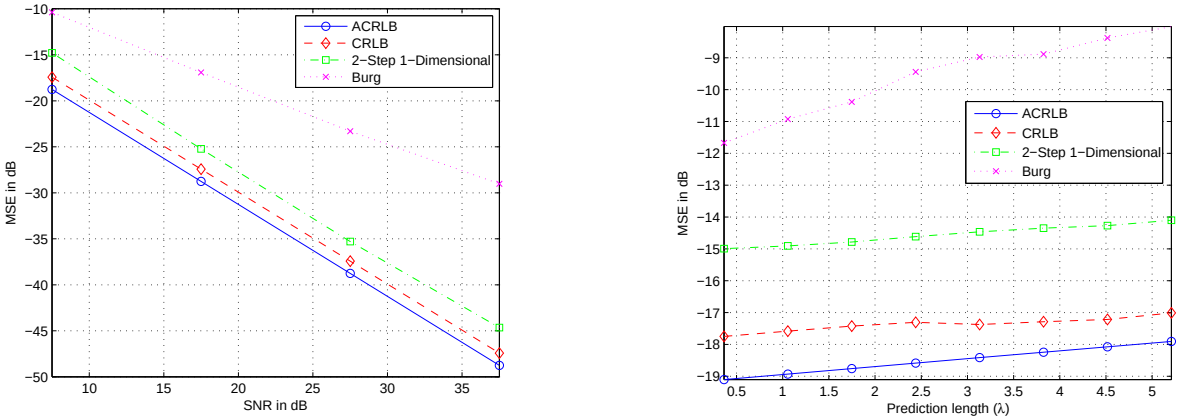
In the scenarios where there are mixed specular and diffuse scattering, which is typical of certain urban pico-cell and indoor wireless scenarios, a more general superposition of a specular and AR model may be more suitable. This has been confirmed for measured channels, esp. in suburban areas in [8], where a joint sinusoidal and AR model showed slight performance benefits over AR modeling alone. They noticed



(a) Power delay profiles

(b) Doppler frequencies

Fig. 2. Power delay profiles and sample Doppler frequencies used in wireless channel simulations.



(a) Normalized MSE vs. SNR

(b) Normalized MSE vs. prediction length.

Fig. 3. Normalized MSE performance of our algorithm versus linear prediction on the time-domain channel taps.

that sinusoidal modeling is most prone to modeling errors and non-stationarities, and *"model selection based on frequency tracking might help to improve the overall performance"*. Since only static block based algorithms were used in [8], we believe that using our adaptive frequency tracking algorithms with adaptive model order estimation can improve the performance of the joint model even further.

REFERENCES

[1] *Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE Std. 802.16e-2005, Feb. 2006.

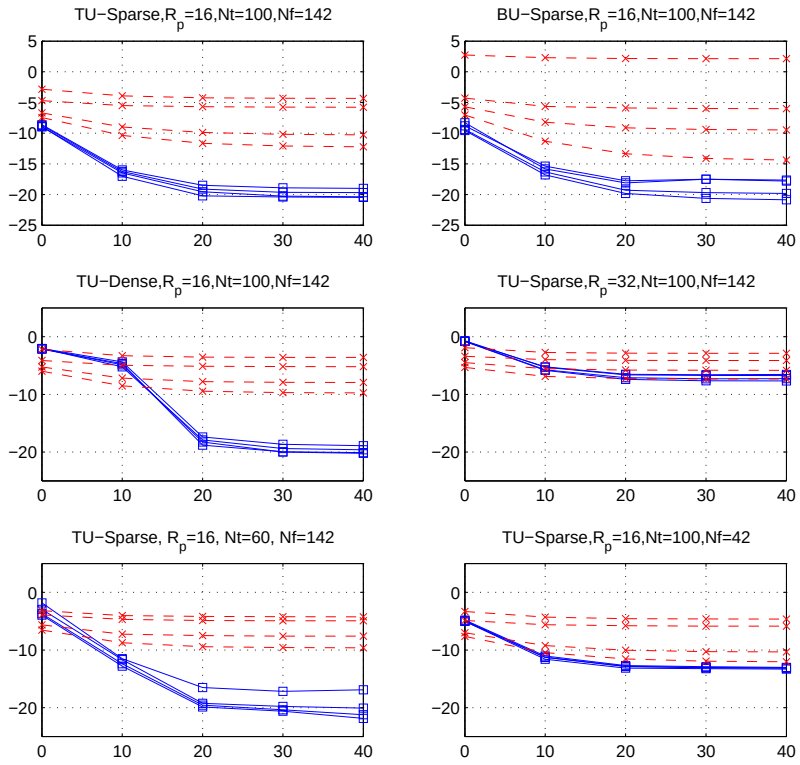
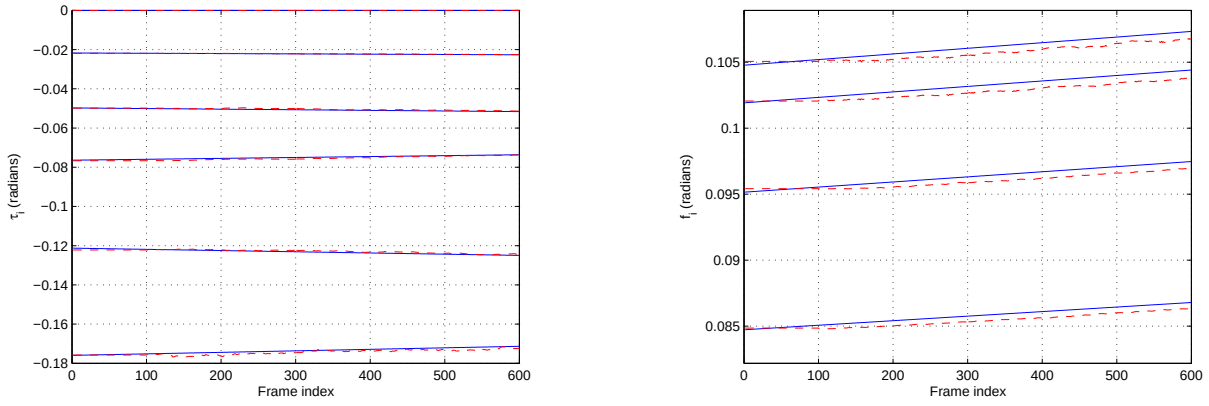


Fig. 4. Normalized mean-squared error (NMSE) vs. SNR performance. The dashed-x's correspond to [12], and the solid-squares correspond to our proposed algorithm. The four NMSE curves for each algorithm from lowest to highest correspond to predicting $\{.25\lambda, .5\lambda, \lambda, 2\lambda\}$ respectively.



(a) Time-delays

(b) Doppler frequencies

Fig. 5. Tracking results for the time-delays $\{\tau_p\}_{p=1}^{P}$ and first 4 Doppler frequencies for path $p = 6$ with SNR $= 15$ dB. Solid lines are the actual values whereas dotted lines are the estimated values.

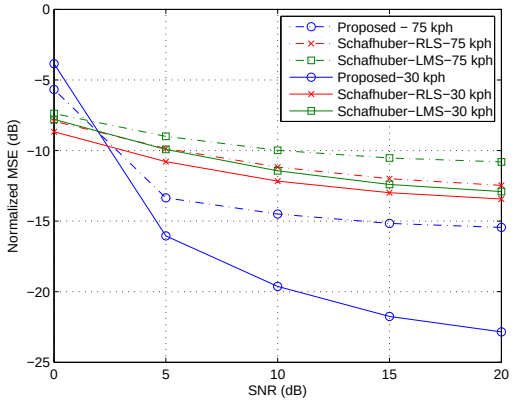


Fig. 6. Prediction MSE results for the complete acquisition-tracking algorithm. We compare our results with [12].

[2] I. C. Wong and B. L. Evans, "Optimal OFDMA subcarrier, rate, and power allocation for ergodic rates maximization with imperfect channel knowledge," in *Proc. IEEE Int. Conf. on Acoust., Sp., and Sig. Proc.*, Honolulu, HI, April 2007.

[3] P. Xia, S. Zhou, and G. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 202–213, Jan. 2004.

[4] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-Range Prediction of Fading Signals," *IEEE Signal Processing Mag.*, vol. 17, no. 3, pp. 62–75, May 2000.

[5] G. Oien, H. Holm, and K. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 758–769, May 2004.

[6] J. K. Hwang and J. H. Winters, "Sinusoidal modeling and prediction of fast fading processes," in *Proc. IEEE Global Telecommunications Conference*, Nov. 1998, pp. 892–897.

[7] P. D. Teal, "Real time characterisation of the mobile multipath channel," Ph.D. dissertation, Australian National University, 2002. [Online]. Available: http://thesis.anu.edu.au/uploads/approved/adt-ANU20020722.085502/public/02whole.pdf

[8] M. Chen, T. Ekman, and M. Viberg, "New Approaches for Channel Prediction Based on Sinusoidal Modeling," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. Article ID 49 393, 13 pages, 2007.

[9] T. Ekman, "Prediction of mobile radio channels," Ph.D. dissertation, Uppsala University, 2002. [Online]. Available: http://www.signal.uu.se/Publications/pdf/a023.pdf

[10] A. Forenza and R. W. Jr. Heath, "Link Adaptation and Channel Prediction in Wireless OFDM Systems," in *Proc. IEEE Midwest Symp. on Circuits and Sys.*, Aug. 2002, pp. 211–214.

[11] M. Sternad and D. Aronsson, "Channel estimation and prediction for adaptive OFDM downlinks [vehicular applications]," in *Proc. IEEE Vehicular Technology Conference*, vol. 2, Oct. 2003, pp. 1283–1287.

[12] D. Schafhuber and G. Matz, "MMSE and Adaptive Prediction of Time-Varying Channels for OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 593–602, Mar. 2005.

[13] I. C. Wong, A. Forenza, R. Heath, and B. Evans, "Long range channel prediction for adaptive OFDM systems," in *Proc. IEEE Asilomar Conf. on Sig., Sys., and Comp.*, vol. 1, Nov. 2004, pp. 732–736.

[14] S. Semmelrodt and R. Kattenbach, "A 2-D fading forecast of time-variant channels based on parametric modeling

techniques," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Comm.*, Lisbon, Portugal, 2002.

[15] V. Madisetti and D. B. Williams, *The digital signal processing handbook*. CRC Press ; IEEE Press, 1998.

[16] S. Semmelrodt and R. Kattenbach, "Investigation of different fading forecast schemes for flat fading radio channels," in *Proc. IEEE Vehicular Technology Conference*, vol. 1, 2003, pp. 149–153.

[17] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Kluwer Academic, 2001.

[18] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, June 2003.

[19] S. Barbarossa and A. Scaglione, "Theoretical bounds on the estimation and prediction of multipath time-varying channels," in *Proc. IEEE Int. Conf. on Acoust., Speech, and Sig. Proc.*, vol. 5, June 2000, pp. 2545–2548.

[20] G. Bretthorst, *Bayesian spectrum analysis and parameter estimation*. Springer-Verlag New York, 1988.

[21] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[22] B. Yang, K. Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[23] G. Xu, I. R.H. Roy, and T. Kailath, "Detection of number of sources via exploitation of centro-symmetry property," *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 102–112, Jan. 1994.

[24] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, 1982.

[25] V. Reddy and L. Biradar, "SVD-based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis," *IEEE Trans. Signal Processing*, vol. 41, no. 9, pp. 2872–2881, Sept. 1993.

[26] P. Strobach, "Bi-iteration SVD subspace tracking algorithms," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1222–1240, May 1997.

[27] S. Ouyang and Y. Hua, "Bi-iterative least-square method for subspace tracking," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2984–2996, Aug. 2005.

[28] G. H. Golub and C. F. V. Loan, *Matrix computations*. Johns Hopkins University Press, 1983.

[29] B. Yang, "An extension of the PASTd algorithm to both rank and subspace tracking," *IEEE Signal Processing Lett.*, vol. 2, no. 9, pp. 179–182, Sept. 1995.

[30] P. Strobach, "Fast recursive subspace adaptive ESPRIT algorithms," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2413–2430, Sept. 1998.

[31] R. Badeau, G. Richard, and B. David, "Fast adaptive ESPRIT algorithm," in *IEEE Workshop on Statistical Signal Processing*, Bordeaux, France, July 17-20 2005.

[32] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985.

[33] L. L. Scharf and C. Demeure, *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*. Addison-Wesley Pub. Co., 1991.

[34] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 2267–2280, Sept. 1992.

[35] C. Rao, L. Zhao, and B. Zhou, "Maximum likelihood estimation of 2-D superimposed exponential signals," *IEEE Trans. Signal Processing*, vol. 42, no. 7, pp. 1795–1802, July 1994.