mmWave Channel Estimation via Approximate Message Passing with Side Information

Dror Baron
Dept. Electrical & Computer Eng.
North Carolina State University
Raleigh, NC, USA
barondror@ncsu.edu

Cynthia Rush
Department of Statistics
Columbia University
New York, NY, USA
cynthia.rush@columbia.edu

Yavuz Yapici
Dept. Electrical & Computer Eng.
North Carolina State University
Raleigh, NC, USA
yyapici@ncsu.edu

Abstract—This work considers millimeter-wave channel estimation in a setting where parameters of the underlying mmWave channels are varying dynamically over time and there is a single drifting path. In this setting, channel estimates at time block k can be used as side information (SI) when estimating the channel at block $k + 1$. To estimate channel parameters, we employ an SI-aided (complex) approximate message passing algorithm and compare its performance to a benchmark based on orthogonal matching pursuit.

Index Terms—Approximate message passing, channel estimation, mmWave, side information, spectral estimation.

I. INTRODUCTION

Mobile user demand for wireless data services has been increasing dramatically in recent years. As the conventional sub-6 GHz communications spectrum is packed with existing wireless services, the *millimeter-wave* (mmWave) frequency band has become a key asset for next-generation cellular networks. Along with increasing antenna array sizes at both sides of the transceiver, *compressed sensing* (CS) based algorithms have received great attention in estimating mmWave channels. Owing to the mobility of users and scattering obstacles (moving cars and so on) in the communication environment, parameters underlying mmWave channels vary dynamically over time. These variations can either be estimated from scratch, likely at the expense of significant training overhead, or tracked by making use of dynamic channel characteristics. The focus of our work is to perform channel estimation using *approximate message passing* (AMP) aided by *side information* (SI). Our AMP-SI approach to channel estimation utilizes the dynamic channel structure, leading to improved estimation quality and reducing training overhead.

Approximate Message Passing. We use a class of low-complexity algorithms, referred to as AMP [1]–[4], for channel estimation. AMP was originally introduced in the context of CS [5], [6], where one wishes to recover an unknown sparse vector $\boldsymbol{\beta}$ from noisy linear measurements $\mathbf{y}$ modeled as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \mathbf{n}, \tag{1}$$

where $\mathbf{A}$ is a measurement matrix with more columns than rows, and $\mathbf{n}$ is *independent and identically distributed* (i.i.d.) noise. AMP iteratively estimates $\boldsymbol{\beta}$ using a possibly non-linear denoiser function tailored to prior knowledge about $\boldsymbol{\beta}$. One

key property of AMP is that under some technical conditions on the measurement matrix $\mathbf{A}$ and signal $\boldsymbol{\beta}$, observations at each iteration of the algorithm, referred to as *pseudo-data*, are asymptotically (in the large system limit) equal in distribution to $\boldsymbol{\beta}$ plus i.i.d. Gaussian noise.

AMP with Side Information (AMP-SI). Recently [7], [8], AMP-SI was introduced as an algorithmic framework that incorporates SI into AMP for CS tasks (1). AMP-SI has been empirically demonstrated to have good reconstruction quality, and is easy to use. For example, we have proposed to use AMP-SI for a toy model for channel estimation in emerging mmWave communication systems [9], where the time dynamics of the channel structure allow previous channel estimates to be used as SI when estimating the current channel structure [7]. In Liu et al. [8], the nice empirical performance of AMP-SI was strengthened through a rigorous performance analysis. For these reasons, it is not surprising that our novel approach to channel estimation outperforms a benchmark based on *orthogonal matching pursuit* (OMP) [10] as evidenced in Sec. V, and it is unlikely that other non-AMP based approaches would yield further improvements.

Contributions. Our main insight in this paper is that the channel matrix can be represented sparsely over the domain of angles of arrival and departure. This insight leads us to develop a denoiser within AMP-SI that monitors and estimates paths with continuous angles of arrival and departure. We use 2D spectral estimation within AMP-SI for a simplified problem with a single drifting path, and will address increasingly complicated (and thus realistic) models.

Notation. Let $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ denote the complex conjugate and Hermitian operations, respectively. We use $\mathbf{0}_M$ and $\mathbf{I}_M$ to represent a zero vector of size $M \times 1$ and identity matrix of size $M \times M$, respectively. Next, $[\mathbf{x}]_i$ stands for the i-th element of the vector $\mathbf{x}$ and $[\mathbf{M}]_{ij}$ for the (i, j)-th element of the matrix $\mathbf{M}$. A complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{C}$ is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{C})$, and $\mathcal{U}[a, b]$ stands for the uniform distribution taking values between a and b. Finally, the set of integers $\{1, 2, \ldots, M\}$ is denoted by $[M]$, and the Dirac delta function, $\delta_{k\ell}$, takes the value 1 if $k = \ell$, and 0 otherwise.

II. SYSTEM MODEL

Consider point-to-point downlink communication in mmWave frequency spectrum (Fig. 1), where a *base station* (BS) communicates with a mobile *user equipment* (UE).
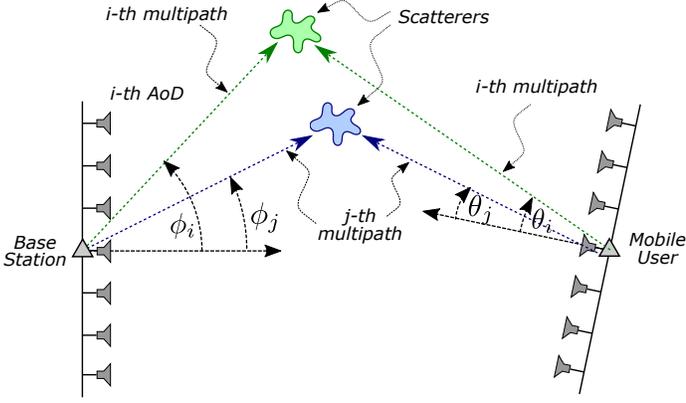
Fig. 1. System model of mmWave communication. Representative i-th and j-th multipaths are shown, along with corresponding angles of arrival $\{\theta_i, \theta_j\}$ and departure $\{\phi_i, \phi_j\}$.

The number of *transmit* and *receive* antennas at the BS and UE are M_t and M_r, respectively, both of which form a *one-dimensional* (1D) *uniform linear array* (ULA).

We study a blockwise transmission strategy that relies on block fading. The downlink channel in the k-th transmission block, denoted by $\mathbf{H}_k$, does not change during the k-th transmission block; the channel matrix takes a new value in the next transmission block. We will assume that across blocks k, there is some dependence or structure, so that an estimate of the block $\mathbf{H}_k$ can be used as *side information* (SI) in estimating $\mathbf{H}_{k+1}$. The relationship between $\mathbf{H}_k$ and $\mathbf{H}_{k+1}$ is detailed in Sec. II-B, and the channel model in Sec. II-A. Our goal will be to estimate $\mathbf{H}_k$ for each block k from received signals, with the signal model specified in Sec. III.

A. Transmission and Channel Models

In our transmission strategy, the BS transmits *pilot* symbols (known to the UE) during the first T_p time slots of each transmission block, which consists of T time slots in total. The UE estimates the channel using these T_p pilot symbols together with the SI, which is obtained from the channel during previous transmission blocks. The BS uses the remaining $T - T_\mathsf{p}$ time slots to transmit payload data (unknown to the UE), which is decoded by the UE using the channel estimate of the current transmission block.

The respective downlink channel, $\mathbf{H}_k$, within the k-th transmission block is given as follows,

$$\mathbf{H}_k = \sqrt{\frac{M_\mathsf{t} M_\mathsf{r}}{L_\mathsf{p}}} \sum_{\ell=1}^{L_\mathsf{p}} \sqrt{\gamma} \alpha_{k,\ell} \, \mathbf{a}(\theta_{k,\ell}, M_\mathsf{r}) \mathbf{a}(\phi_{k,\ell}, M_\mathsf{t})^\mathsf{H}, \quad (2)$$

where L_p is the number of multipath components, γ is the *signal-to-noise ratio* (SNR), $\alpha_{k,\ell}$ is the complex gain of the ℓ-th multipath component assumed to be distributed as $\mathcal{CN}(0,1)$ with uncorrelated gains for different paths, i.e. $\mathbb{E}[\boldsymbol{\alpha}_k (\boldsymbol{\alpha}_k^*)^\mathsf{T}] = \sigma^2 \mathbf{I}_{L_\mathsf{p}}$ where $\boldsymbol{\alpha}_k = [\alpha_{k,1} \, \alpha_{k,2} \ldots \alpha_{k,L_\mathsf{p}}]^\mathsf{T}$. In addition, $\theta_{k,\ell}$ and $\phi_{k,\ell}$ are the *angle-of-arrival* (AoA) and *angle-of-departure* (AoD), respectively, of the ℓ-th multipath component with $\theta_{k,\ell} \sim \mathcal{U}[0, 2\pi]$ and $\phi_{k,\ell} \sim \mathcal{U}[0, 2\pi]$. Note that the distributions for the complex path gain and

AoA/AoD folow from mmWave channel measurement studies [9]. Furthermore, $\mathbf{a}(\theta_{k,\ell}, M_\mathsf{r})$ and $\mathbf{a}(\phi_{k,\ell}, M_\mathsf{t})$ represent the array steering vectors at the receive (UE) and transmit (BS) sides, respectively, where the m-th element of a generic array steering vector $\mathbf{a}(\varphi, M)$ is

$$\left[\mathbf{a}(\varphi, M) \right]_m = \frac{1}{\sqrt{M}} \exp \left\{ j 2\pi \frac{d_\mathsf{a}}{\lambda} (m-1) \sin(\varphi) \right\}, \quad (3)$$

for $m \in [M]$, with an arbitrary phase φ representing AoA/AoD, and M elements representing the number of transmit/receive antenna elements, where d_a is the antenna element spacing of the ULA, and λ is the carrier frequency wavelength.

Recall that our goal is to estimate $\mathbf{H}_k$ for each block k from received signals and SI. To this end, we will assume throughout that in our definition for $\mathbf{H}_k$ (2) the scalar values M_t, M_r, L_p, and γ are known, and that d_a and λ are also available to the ULA antenna array. Note that the number of paths L_p and SNR γ are relatively stationary for point-to-point communications, and can be separately obtained over many transmission blocks. As a result, estimating the channel block $\mathbf{H}_k$ boils down to estimating the L_p complex Gaussian *random variables* (RVs) $\alpha_{k,\ell}$ and the $2L_\mathsf{p}$ uniform RVs $\theta_{k,\ell}$ and $\phi_{k,\ell}$, from which the arrays $\mathbf{a}(\theta_{k,\ell}, M_\mathsf{r})$ and $\mathbf{a}(\phi_{k,\ell}, M_\mathsf{r})$ and thus $\mathbf{H}_k$ can be recovered.

B. Time Variation for Channel Parameters

We now provide a model for the dynamic relationship between $\mathbf{H}_k$ and $\mathbf{H}_{k+1}$. Our model was discussed by numerous authors (c.f., [11]–[13] and references therein). The complex path gain varies from one transmission block to another following a first order *auto regressive* (AR) process,

$$\boldsymbol{\alpha}_{k+1} = \rho \boldsymbol{\alpha}_k + \mathbf{u}_k^\alpha, \quad (4)$$

where ρ is the correlation coefficient, and $\mathbf{u}_k^\alpha$ is innovation with $\mathbf{u}_k^\alpha \sim \mathcal{CN}(\mathbf{0}_{L_\mathsf{p}}, (1-\rho^2) \mathbf{I}_{L_\mathsf{p}})$. Note that (4) represents variation in the complex path gains (i.e., small-scale fading) with correlation to the previous transmission blocks through ρ. Moreover, both AoA and AoD follow Gaussian innovation processes in the next transmission block,

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \mathbf{u}_k^\theta \qquad \text{and} \qquad \boldsymbol{\phi}_{k+1} = \boldsymbol{\phi}_k + \mathbf{u}_k^\phi, \quad (5)$$

where $\boldsymbol{\theta}_k = [\theta_{k,1} \, \theta_{k,2} \ldots \theta_{k,L_\mathsf{p}}]^\mathsf{T}$ and $\boldsymbol{\phi}_k = [\phi_{k,1} \ldots \phi_{k,L_\mathsf{p}}]^\mathsf{T}$ are the aggregate AoA and AoD vectors, respectively, and $\mathbf{u}_k^\theta \sim \mathcal{N}(\mathbf{0}_{L_\mathsf{p}}, \sigma_\theta^2 \mathbf{I}_{L_\mathsf{p}})$ and $\mathbf{u}_k^\phi \sim \mathcal{N}(\mathbf{0}_{L_\mathsf{p}}, \sigma_\phi^2 \mathbf{I}_{L_\mathsf{p}})$ are corresponding innovation vectors.

III. mmWave Channel Estimation via AMP-SI

In this section, we consider mmWave channel estimation for the scenario described in Sec. II, and explain how SI from previous transmission blocks enhances the estimation quality. To this end, we define an $M_\mathsf{t} \times 1$ complex-valued unit-energy vector $\mathbf{s}_{k,i}$, which represents the pilot symbol transmitted during the i-th time slot within transmission block k. We also assume that $\mathbf{s}_{k,i}$ is selected from an uncorrelated dictionary, i.e., $\mathbb{E}[\mathbf{s}_{k,i} \mathbf{s}_{k',i'}^\mathsf{H}] = \delta_{kk'} \delta_{ii'}$ where $k, k', i, i' \in [M_\mathsf{t}]$.

From signal estimation to matrix estimation. The received signal vector at the UE is given by

$$\mathbf{y}_{k,i} = \mathbf{H}_k \mathbf{s}_{k,i} + \mathbf{n}_{k,i}, \quad (6)$$

where $\mathbf{n}_{k,i}$ is measurement noise that follows $\mathcal{CN}(\mathbf{0}_{M_r}, \mathbf{I}_{M_r})$. Considering pilot transmissions over T_p time slots, the aggregate received signal during the k-th transmission block, $\mathbf{Y}_k = [\mathbf{y}_{k,1}| \cdots |\mathbf{y}_{k,T_p}] \in \mathbb{C}^{M_r \times T_p}$, where $[\mathbf{v}_1| \cdots |\mathbf{v}_{T_p}]$ denotes a matrix obtained by concatenating the column vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{T_p}$, is given by

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{S}_k + \mathbf{N}_k. \tag{7}$$

Note that $\mathbf{S}_k = [\mathbf{s}_{k,1}| \cdots |\mathbf{s}_{k,T_p}]$ and $\mathbf{N}_k = [\mathbf{n}_{k,1}| \cdots |\mathbf{n}_{k,T_p}]$ are $M_t \times T_p$ and $M_r \times T_p$, respectively. Our goal is to estimate $\mathbf{H}_k$ using (7) given observations $\mathbf{Y}_k$ and pilot symbols $\mathbf{S}_k$ and SI from previous transmission blocks.

As mentioned in Sec. I, we use AMP-SI to estimate the channel $\mathbf{H}_k$ from the model (7). While (7) is not identical to the CS problem (1), taking the transpose of (7) we have

$$\mathbf{Y}_k^T = \mathbf{S}_k^T \mathbf{H}_k^T + \mathbf{N}_k^T, \tag{8}$$

which is more aligned with the AMP framework. In particular, we could modify (8) by vectorizing $\mathbf{Y}_k^T$ and $\mathbf{H}_k^T$ and composing a measurement matrix $\mathbf{A}$ having $\mathbf{S}_k^T$ repeated on the diagonal. However, this modification is not necessary, because AMP provides favorable results even when applied to multi-dimensional signals [14], [15]. Therefore, we run AMP directly on the multi-dimensional problem.

AMP-SI with 2D denoisers. Consider a fixed block k. We run AMP-SI on the matrix $\mathbf{H}_k^T$ directly, employing a well-chosen 2D (matrix) denoiser. Each AMP iteration will have access to pseudo-data that is asymptotically (in the large system limit) equal in distribution to $\mathbf{H}_k^T$ plus a matrix of i.i.d. Gaussian noise, where the existing AMP theory allows us to calculate a good approximation for the noise variance. Importantly, the 2D denoiser we propose incorporates SI from the previous estimate at block $k − 1$, where this SI is our estimate for the multipath parameters in block $k − 1$, and our knowledge of the dynamics of these parameters per (4)-(5).

Denoising a 2D matrix within AMP, as opposed to a vector, is non-standard. That said, some related art (including by the authors) is encouraging. For example, previous applications of AMP with multi-dimensional denoisers have provided encouraging empirical results in image reconstruction [14], [15]. Beyond empirical results, a rigorous analysis by Ma *et al.* [16] provides performance guarantees for a family of multi-dimensional sliding window denoisers used within AMP.

Spectral estimation in 2D. Our proposed denoiser resembles work by Hamzehei and Duarte [17], [18], who performed *analog denoising* of 1D vectors within AMP in the context of spectral estimation. Our denoisers resemble theirs, except that we perform 2D instead of 1D spectral estimation, with improved performance owing to SI from block $k − 1$. *Our main insight is that H_k is sparse over the <u>continuous</u> (AoA, AoD) domain*. This insight leads us to develop a denoiser that monitors and estimates paths with continuous ϕ and θ. Our work seems most related to Bellili et al. [19], where the authors sparsify a linear inverse problem using Fourier arguments. Our approach expands over theirs by performing 2D continuous spectral estimation within AMP.

To make the details of our denoiser tractable (Sec. IV), we begin with a simplified setting comprised of a *single drifting path*. That is, we assume $L_p = 1$ in channel model (2). While this paper introduces AMP-SI and our 2D spectral estimation denoiser for this simplified version of the problem, we aim to leverage these results and develop a series of denoisers addressing increasingly complicated (and thus realistic) models. Our current work will be extended to multiple paths with birth-death-drift dynamics between blocks [7].

IV. ONE DRIFTING PATH

As mentioned previously, we set $L_p = 1$ in (2), and model the channel in the k-th transmission block, $\mathbf{H}_k$, as

$$\mathbf{H}_k = \sqrt{M_t M_r \gamma}\, \alpha_k\, \mathbf{a}(\theta_k, M_r)\mathbf{a}(\phi_k, M_t)^H. \tag{9}$$

In this section, we introduce an AMP-SI algorithm for completing this parameter estimation task, and discuss some implementation details.

AMP-SI details. The AMP-SI algorithm for estimating $\mathbf{H}_k$ in (9) takes the following form. Initialize the matrix estimate with $\widehat{\mathbf{H}}_k^t = \mathbf{0}$, a zeros matrix, and at iteration $t \geq 0$, compute

$$\mathbf{R}^t = \mathbf{Y}_k^T - \mathbf{S}_k^T \widehat{\mathbf{H}}_k^t + \mathbf{R}^{t-1} \langle div\, \eta_t(\mathbf{V}^t, \mathrm{SI}_{k-1}) \rangle,$$
$$\widehat{\mathbf{H}}_k^{t+1} = \eta_t(\mathbf{V}^t, \mathrm{SI}_{k-1}), \tag{10}$$
$$\mathbf{V}^{t+1} = \mathbf{S}_k^* \mathbf{R}^t + \widehat{\mathbf{H}}_k^t,$$

where we interpret $\mathbf{R}^t$ as a *residual*, $\widehat{\mathbf{H}}_k^{t+1}$ is our current estimate of $\mathbf{H}_k$, and $\mathbf{V}^t$ is the *pseudo-data*, which is equal in distribution to $\mathbf{H}_k^T$ plus i.i.d. Gaussian noise with variance $\tau_t^2 \approx ||\mathbf{R}^t||^2/(M_r T_p)$. Our *denoiser*, denoted $\eta(\cdot, \cdot)$, takes as inputs the pseudo-data $\mathbf{V}^t$ and SI from the previous block, denoted SI_{k-1}; the form of $\eta(\cdot, \cdot)$ is specified below. Finally, the residual uses the normalized *divergence* of the denoiser,

$$\langle div\, \eta_t(\mathbf{V}, \mathrm{SI}) \rangle = \frac{1}{M_r M_t} \sum_{i=1}^{M_t} \sum_{j=1}^{M_r} \frac{\partial}{\partial \mathbf{V}_{ij}} [\eta_t(\mathbf{V}, \mathrm{SI})]_{ij}. \tag{11}$$

We highlight that the conjugate operator is applied element-wise to $\mathbf{S}_k$ when computing $\mathbf{V}^t$ as part of complex AMP [20].

Candidate Denoisers. Given the distributional properties of the pseudo-data, namely $\mathbf{V}^t \approx \mathbf{H}_k + \tau_t \mathbf{G}$ where $\mathbf{G}$ has i.i.d. complex Gaussian entries, there are at least two plausible denoising styles for AMP-SI (10).

The first denoising style we consider is *conditional expectation*, where $\widehat{\mathbf{H}}_k^{t+1}$ is calculated using

$$\widehat{\mathbf{H}}_k^{t+1} = \mathbb{E}[\mathbf{H}_k | \mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G}, \mathrm{SI}_{k-1} = \widehat{\mathbf{H}}_{k-1}],$$

where we have explicitly stated that the SI at time $k − 1$ takes the form of an estimate of the channel at the previous block, $\widehat{\mathbf{H}}_{k-1}$. Within each AMP iteration, conditional expectation provides a *minimum mean squared error* (MMSE) estimator of $\widehat{\mathbf{H}}_k^{t+1}$ given the pseudo-data, $\mathbf{V}^t$, and SI, SI_{k-1}. Under some technical conditions, for large scale linear inverse problems [3], upon convergence, AMP with conditional expectation denoisers yields the *overall* MMSE signal estimator. Unfortunately, with our current understanding of the model, the conditional expectation denoiser appears computationally intractable when $L_p > 1$, so we did not consider it further.

The second denoising style we consider is *maximum a posteriori* (MAP), where we compute the triple $(\widehat{\theta}_k, \widehat{\Phi}_k, \widehat{\alpha}_k)$ that maximizes the posterior,

$$f(\theta_k, \Phi_k, \alpha_k | \mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G}, \mathrm{SI}_{k-1} = (\widehat{\theta}_{k-1}, \widehat{\Phi}_{k-1}, \widehat{\alpha}_{k-1})),$$

where $f(\cdot)$ denotes a generic density, and then use the estimated triple $(\widehat{\theta}_k, \widehat{\Phi}_k, \widehat{\alpha}_k)$ to produce an estimate of $\widehat{\mathbf{H}}_k$. In contrast to the conditional expectation denoiser, MAP signal estimation is sub-optimal in terms of MSE in individual AMP iterations, because conditional expectation is the MMSE estimator, and thus minimizes the noise variance for the next iteration, whereas MAP differs from conditional expectation. Additionally, MAP denoisers may not achieve the overall MMSE. Despite MAP having these estimation-theoretic drawbacks, we will see that it offers computational advantages in our analog denoising problem.

For the MAP denoiser, we need to further study the posterior distribution.. First, by Bayes' rule,

$$f(\theta_k, \Phi_k, \alpha_k | \mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G}, \mathrm{SI}_{k-1} = (\widehat{\theta}_{k-1}, \widehat{\Phi}_{k-1}, \widehat{\alpha}_{k-1}))$$
$$= f(\theta_k | \widehat{\theta}_{k-1}) f(\Phi_k | \widehat{\Phi}_{k-1}) f(\alpha_k | \widehat{\alpha}_{k-1}) \times \qquad (12)$$
$$\frac{f(\tau_t \mathbf{G} = \mathbf{V}^t - \mathbf{H}_k \mid \theta_k, \Phi_k, \alpha_k)}{f(\mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G} \mid \mathrm{SI}_{k-1} = (\widehat{\theta}_{k-1}, \widehat{\Phi}_{k-1}, \widehat{\alpha}_{k-1}))},$$

where we have used the independence of the RVs $(\theta_k, \Phi_k, \alpha_k)$ and the fact that the pseudo-data at iteration t, given by $\mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G}$, is independent of $(\widehat{\theta}_{k-1}, \widehat{\Phi}_{k-1}, \widehat{\alpha}_{k-1})$ given $(\theta_k, \Phi_k, \alpha_k)$. In the numerator (12), the densities are Gaussian,

$$f(\alpha_k | \widehat{\alpha}_{k-1}) \sim \mathcal{CN}(\rho \widehat{\alpha}_{k-1}, 1 - \rho^2), \qquad (13)$$
$$f(\theta_k | \widehat{\theta}_{k-1}) \sim \mathcal{N}(\widehat{\theta}_{k-1}, \sigma_\theta^2), \quad f(\Phi_k | \widehat{\Phi}_{k-1}) \sim \mathcal{N}(\widehat{\phi}_{k-1}, \sigma_\phi^2).$$

The denominator is a normalization constant that does not affect MAP optimization.

MAP Denoiser. Focusing on the MAP denoiser, the form of the conditional distribution given in (12) suggests,

$$\underset{(\theta_k, \Phi_k, \alpha_k) \in \mathbb{R}^2 \times \mathbb{C}}{\arg\max} f(\theta_k, \Phi_k, \alpha_k | \mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G}, \mathrm{SI}_{k-1})$$
$$= \underset{(\theta_k, \Phi_k, \alpha_k) \in \mathbb{R}^2 \times \mathbb{C}}{\arg\max} \Big\{ \log f(\alpha_k | \widehat{\alpha}_{k-1}) + \log f(\theta_k | \widehat{\theta}_{k-1})$$
$$+ \log f(\phi_k | \widehat{\phi}_{k-1}) + \log f(\tau_t \mathbf{G} = \mathbf{V}^t - \mathbf{H}_k \mid \theta_k, \Phi_k, \alpha_k) \Big\}. \qquad (14)$$

We simplify the above using (13),

$$\log f(\alpha_k | \widehat{\alpha}_{k-1}) = -\log(\pi(1 - \rho^2)) - \frac{1}{1 - \rho^2} |\alpha_k - \rho \widehat{\alpha}_{k-1}|^2,$$

$$\log f(\theta_k | \widehat{\theta}_{k-1}) = -\frac{1}{2} \log(2\pi \sigma_\theta^2) - \frac{1}{2\sigma_\theta^2} (\theta_k - \widehat{\theta}_{k-1})^2,$$

where $\log f(\phi_k | \widehat{\phi}_{k-1})$ is similar to $\log f(\theta_k | \widehat{\theta}_{k-1})$, and

$$\log f(\tau_t \mathbf{G} = \mathbf{V}^t - \mathbf{H}_k \mid \theta_k, \Phi_k, \alpha_k)$$
$$= \log \prod_{i=1}^{M_r} \prod_{j=1}^{M_t} f(\tau_t [\mathbf{G}]_{ij} = [\mathbf{V}^t]_{ij} - [\mathbf{H}_k]_{ij} \mid \theta_k, \Phi_k, \alpha_k)$$
$$= -M_t M_r \log(\pi) - \frac{1}{\tau_t^2} ||\mathbf{V}^t - \mathbf{H}_k||_F^2,$$

where for a matrix $\mathbf{M} \in \mathbb{C}^{n \times m}$ we have $||\mathbf{M}||_F^2 = \sum_{i=1}^n \sum_{j=1}^m |[\mathbf{M}]_{ij}|^2$. Plugging into (14), we find

$$\underset{(\theta, \Phi, \alpha) \in \mathbb{R}^2 \times \mathbb{C}}{\arg\max} f(\theta, \Phi, \alpha | \mathbf{V}^t = \mathbf{H}_k + \tau_t \mathbf{G})$$
$$= \underset{(\theta, \Phi, \alpha_R, \alpha_I) \in \mathbb{R}^4}{\arg\min} C_t(\theta, \Phi, \alpha_R, \alpha_I), \qquad (15)$$

where

$$C_t(\theta, \Phi, \alpha_R, \alpha_I) = \frac{(\theta - \widehat{\theta}_{k-1})^2}{2\sigma_\theta^2} + \frac{(\phi - \widehat{\phi}_{k-1})^2}{2\sigma_\phi^2}$$
$$+ \frac{|\alpha_R + j\alpha_I - \rho \widehat{\alpha}_{k-1}|^2}{1 - \rho^2} + \frac{1}{\tau_t^2} ||\mathbf{V}^t - \mathbf{H}_k||_F^2, \qquad (16)$$

and we have written $\alpha \in \mathbb{C}$ as $\alpha = \alpha_R + j\alpha_I$ with $\alpha_R, \alpha_I \in \mathbb{R}$.

MAP Denoiser Implementation Details. Now we discuss the details of implementing the MAP denoiser of (15) in the AMP algorithm in (10), meaning we take

$$\eta_t(\mathbf{V}^t, \mathrm{SI}_{k-1}) = \underset{(\theta, \Phi, \alpha_R, \alpha_I) \in \mathbb{R}^4}{\arg\min} C_t(\theta, \Phi, \alpha_R, \alpha_I), \qquad (17)$$

where we have defined the function $C_t(\cdot)$ in (16). In performing the 4D optimization in (17), it is possible to explicitly solve for the minimizing pair $(\alpha_R, \alpha_I) \in \mathbb{R}^2$ for any given pair $(\theta, \Phi) \in \mathbb{R}^2$, because

$$\frac{\partial C_t}{\partial \alpha_R}(\theta, \Phi, \alpha_R, \alpha_I) = \frac{2(\alpha_R - \rho \mathrm{Re}(\widehat{\alpha}_{k-1}))}{1 - \rho^2}$$
$$- \frac{2\kappa}{\tau_t^2} \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} \Big[\mathrm{Re}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Re}([\mathbf{V}^t]_{ij} - \kappa \alpha [\mathbf{a}_{\theta,\phi}]_{ij})$$
$$- \mathrm{Im}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Im}([\mathbf{V}^t]_{ij} - \kappa \alpha [\mathbf{a}_{\theta,\phi}]_{ij}) \Big],$$

using the shorthand $[\mathbf{a}_{\theta,\phi}]_{ij} := [\mathbf{a}(\theta, M_r)]_i [\mathbf{a}(\phi, M_t)^{\mathrm{H}}]_j$, $\kappa := \sqrt{M_t M_r} \gamma$, and $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ indicate real and imaginary parts, along with the fact that for any $i \in [M_r]$ and $j \in [M_r]$,

$$\frac{\partial |[\mathbf{V}^t]_{ij} - [\mathbf{H}_k]_{ij}|^2}{\partial \alpha_R} = -2\kappa \mathrm{Re}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Re}([\mathbf{V}^t]_{ij} - \kappa \alpha [\mathbf{a}_{\theta,\phi}]_{ij})$$
$$- 2\kappa \mathrm{Im}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Im}([\mathbf{V}^t]_{ij} - \kappa \alpha [\mathbf{a}_{\theta,\phi}]_{ij}).$$

For any (θ, ϕ), the minimizing $\widehat{\alpha}_R$ takes the form

$$\widehat{\alpha}_R = \Big[\frac{1}{1 - \rho^2} + \frac{\kappa^2}{\tau_t^2} \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} \Big| [\mathbf{a}_{\theta,\phi}]_{ij} \Big|^2 \Big]^{-1} \Big[\frac{\rho \mathrm{Re}(\widehat{\alpha}_{k-1})}{1 - \rho^2} +$$
$$\frac{\kappa}{\tau_t^2} \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} \mathrm{Re}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Re}([\mathbf{V}^t]_{ij}) + \mathrm{Im}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Im}([\mathbf{V}^t]_{ij}) \Big].$$

We can similarly show

$$\widehat{\alpha}_I = \Big[\frac{1}{1 - \rho^2} + \frac{\kappa^2}{\tau_t^2} \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} \Big| [\mathbf{a}_{\theta,\phi}]_{ij} \Big|^2 \Big]^{-1} \Big[\frac{\rho \mathrm{Im}(\widehat{\alpha}_{k-1})}{1 - \rho^2} +$$
$$\frac{\kappa}{\tau_t^2} \sum_{i=1}^{M_r} \sum_{j=1}^{M_t} \mathrm{Re}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Im}([\mathbf{V}^t]_{ij}) - \mathrm{Im}([\mathbf{a}_{\theta,\phi}]_{ij}) \mathrm{Re}([\mathbf{V}^t]_{ij}) \Big].$$

In implementing the denoiser in (17) within AMP, for any pair $(\widehat{\phi}, \widehat{\theta})$ we solve for the optimal $\widehat{\alpha}$ using the estimates of $\widehat{\alpha}_R$ and $\widehat{\alpha}_I$ given just above. Optimal $(\widehat{\phi}, \widehat{\theta})$ are computed
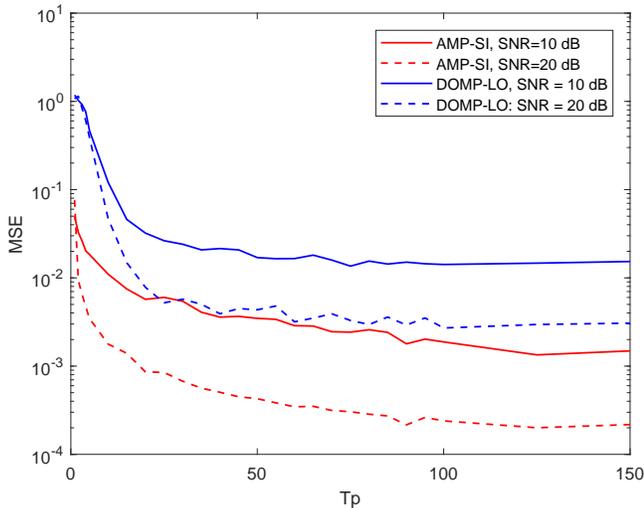
Fig. 2. MSE as function of number of pilots T_p per transmission block. (AMP-SI used 300 iterations.)

using a grid search over values within four standard deviations of the SI. The remaining consideration in implementing (10) is to compute the divergence of the denoiser (11). While it is difficult to compute the divergence analytically, our implementation in Sec. V approximates it numerically.

We pause here to note that extending the MAP denoising technique just outlined to two drifting paths, i.e., $L_p = 2$ in our channel model (2), seems feasible but computationally expensive, and it is unclear if these approaches will work when going beyond, i.e., $L_p \geq 3$. Extending our results here to these regimes will be pursued in future work.

V. Numerical Results

We demonstrate the performance of the proposed AMP-SI algorithm by presenting numerical results based on Monte Carlo simulations. As a benchmark, we used *Dirichlet orthogonal matching pursuit with local optimization* (DOMP-LO) [21], which uses Dirichlet kernels while estimating the unknown mmWave channel. Our communications setting used $M_t = 16$, $M_r = 8$, $\rho = 0.995$, $\sigma_\theta^2 = \sigma_\phi^2 = 1$, and $\gamma \in \{10, 20\}$ dB, which represent a mmWave channel with reasonable time variation. Fig. 2 shows numerical results using 300 AMP-SI iterations; DOMP-LO used a dictionary size of 32, which yields the best performance. It can be seen that the empirical MSE declines as more pilots, T_p, are used. Moreover, a 20 dB SNR outperformed 10 dB, resulting in lower curves. Overall, the empirical MSE obtained by our AMP-SI approach was roughly an order of magnitude less than that of DOMP-LO. We also compared our setting to one without SI by increasing the variance of the drift. Larger variance reduced the estimation quality; we omit the details for brevity. Finally, our future work will evaluate the spectral efficiency of the proposed algorithm along with hybrid/digital beamforming schemes in comparison to training length, and compare the performance gains over existing methods (e.g., [22], [23]).

References

[1] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[2] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices," *Journal of Statistical Mechanics: Theory and Experiment*, no. 8, 2012.

[3] A. Montanari, "Graphical models concepts in compressed sensing," in *Compressed Sensing*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012, pp. 394–438.

[4] S. Rangan, P. Schniter, and A. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, 2019.

[5] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[6] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[7] A. Ma, Y. Zhou, C. Rush, D. Baron, and D. Needell, "An approximate message passing framework for side information," *IEEE Trans. Signal Proc.*, vol. 67, no. 7, pp. 1875–1888, Apr. 2019.

[8] H. Liu, C. Rush, and D. Baron, "An analysis of state evolution for approximate message passing with side information," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2019.

[9] M. Samimi and T. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.

[10] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, pp. 40–44, Nov. 1993.

[11] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[12] V. Va, H. Vikalo, and R. Heath, "Beam tracking for mobile millimeter wave communication systems," in *IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 743–747.

[13] S. Jayaprakasam, X. Ma, J. Choi, and S. Kim, "Robust beam-tracking for mmWave mobile communications," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2654–2657, Dec. 2017.

[14] J. Tan, Y. Ma, and D. Baron, "Compressive imaging via approximate message passing with image denoising," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2085–2092, Apr. 2015.

[15] C. Metzler, A. Maleki, and R. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, Sept. 2016.

[16] Y. Ma, C. Rush, and D. Baron, "Analysis of approximate message passing with non-separable denoisers and Markov random field priors," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7367–7389, Nov. 2019.

[17] S. Hamzehei and M. Duarte, "Compressive parameter estimation via approximate message passing," in *2015 IEEE Int. Conf. on Acoustics, Speech and Signal Process.* IEEE, 2015, pp. 3327–3331.

[18] ——, "Compressive direction-of-arrival estimation off the grid," in *2016 50th Asilomar Conference on Signals, Systems and Computers.* IEEE, 2016, pp. 1081–1085.

[19] F. Bellili, F. Sohrabi, and W. Yu, "Generalized approximate message passing for massive MIMO mmwave channel estimation with Laplacian prior," *IEEE Trans. on Comm.*, vol. 67, no. 5, pp. 3205–3219, 2019.

[20] A. Maleki, L. Anitori, Z. Yang, and R. Baraniuk, "Asymptotic analysis of complex lasso via complex approximate message passing (camp)," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290–4308, July 2013.

[21] C. K. Anjinappa, Y. Zhou, Y. Yapıcı, D. Baron, and I. Güvenç, "Channel estimation in mmWave hybrid MIMO system via off-grid Dirichlet kernels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, Hawaii, Dec. 2019.

[22] V. Boljanovic, H. Yan, and D. Cabric, "Tracking sparse mmWave channel under time varying multipath scatterers," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2018, pp. 1274–1279.

[23] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1244–1248, 2019.