Channel Tracking and Transmit Beamforming With Frugal Feedback

Omar Mehanna, *Student Member, IEEE*, and Nicholas D. Sidiropoulos, *Fellow, IEEE*

***Abstract*—Channel state feedback is a serious burden that limits deployment of transmit beamforming systems with many antennas in frequency-division duplex (FDD) mode. Transmit beamforming with limited feedback systems estimate the channel at the receiver and send quantized channel state or beamformer information to the transmitter. A different approach that exploits the spatio-temporal correlation of the channel is proposed here. The transmitter periodically sends a *beamformed* pilot signal in the downlink, while the receiver quantizes the corresponding received signal and feeds back the bits to the transmitter. Assuming an autoregressive (AR) channel model, Kalman filtering (KF) based on the sign of innovations (SOI) is proposed for channel tracking, and closed-form expressions for the channel estimation mean-squared error (MSE) are derived under certain conditions. For more general channel models, a novel tracking approach is proposed that exploits the quantization bits in a maximum *a posteriori* (MAP) formulation. Simulations show that close to optimum performance can be attained with only 2 bits per channel dwell time block, even for systems with many transmit antennas. This clears a hurdle for transmit beamforming with many antennas in FDD mode—which was almost impossible with the prior state-of-art.**

***Index Terms*—Beamforming, estimation, Kalman filtering, limited-rate feedback, quantization, time-varying channels.**

I. Introduction

TRANSMIT beamforming can enhance the performance of multiple-input multiple-output (MIMO) systems by exploiting channel state information (CSI) at the transmitter. In the frequency-division duplex (FDD) mode, where the downlink and uplink channels are not reciprocal, the receiver must feedback information about the downlink channel to the transmitter. In systems with many transmit antennas, the feedback overhead can be overwhelming; and the challenge is to limit the feedback to only a few bits that still provide sufficient information about the channel.

Almost all work on transmit beamforming with *limited feedback* addresses this challenge by designing efficient beamformer weight vector quantization algorithms at the receiver. The focus is on designing a common beamformer codebook (known at the transmitter and receiver). At runtime, the receiver estimates the downlink channel, finds the best-matching beamforming vector in the codebook, and feeds back its index to the transmitter [2]. Codebook design can be based on maximizing the average signal-to-noise ratio (SNR) [3], maximizing the average mutual information [4], or minimizing the outage probability [5], and it can be viewed as a vector quantization problem, where the generalized Lloyd algorithm (GLA) can be used to construct the codebook [6]. This codebook-based framework assumes accurate CSI at the receiver, which in turn implies significant downlink pilot overhead. For large codebooks, which are necessary when the number of transmit-antennas is large, the feedback overhead can be significant, and the computational complexity of searching the codebook for the best beamformer can be prohibitive.

Another important issue is that most prior work assumes a Rayleigh block-fading model, according to which the channel remains constant over a block of symbols and changes independently across different blocks. The block-fading assumption overlooks the channel temporal correlation, which can be exploited to decrease the feedback rate [7], [8]. In [7] and [8], the temporal correlation of the channel is exploited by modeling the quantized CSI at the receiver as a finite-state Markov chain, and computing the transition probability of every codebook entry given the previous (one or more) codebook entries. In [7], variable-length Huffman source coding is applied to the transition probabilities of the Markov chain to *compress* the CSI feedback. This approach is not suitable for practical communication systems with limited feedback, which provision a fixed number of feedback bits per CSI slot, as in e.g., LTE [9]. Considering this issue, a different *fixed-length* but *lossy* CSI compression algorithm is proposed in [8], where low-probability transitions between the Markov chain states are truncated. For large-size codebooks, computing the transition probabilities accurately for a large number of Markov states is an elusive task that requires very long training periods. Moreover, the transition probabilities are dependent on the specific channel model—new computations are necessary whenever the model varies significantly.

This paper proposes a different approach for beamforming with limited feedback, that exploits the spatio-temporal channel correlation, and avoids the limitations of codebook-based feedback and Markov chain modeling. The transmitter is assumed to periodically transmit a *beamformed* pilot signal in the downlink, while the receiver *quantizes* the corresponding received signal (2-bit quantization is considered in this paper), and sends the quantization bits to the transmitter through the uplink feedback

Manuscript received March 11, 2014; revised July 31, 2014; accepted September 15, 2014. Date of publication October 28, 2014; date of current version November 12, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ana Perez-Neira. Supported in part by NSF ECCS-1231504, NSF AST-1247885. Conference version of part of the results will appear in Proc. 48th Asilomar Conference on Signals, Systems and Computers, November 2–5, 2014 [1].

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: meha0006; nikos@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2014.2365762

channel. Therefore, instead of estimating the channel at the receiver and sending the quantized CSI to the transmitter as in codebook-based beamforming, the receiver feeds back a quantized (noisy) linear measurement of the channel. The challenge here is whether the transmitter can accurately estimate and track the channel using such few (periodic) feedback bits.

Assuming that the channel can be modeled by an autoregressive (AR) model [10], and that the receiver feeds back the analog-amplitude (un-quantized or finely-quantized) received signal to the transmitter, Kalman filtering (KF) [11] is used in [12] to track the channel at the transmitter. However, sending the analog or finely-quantized received signal back to the transmitter is problematic in terms of uplink rate and transmit power. In this paper, we consider a 2-bit quantization scheme that is based on the sign of innovation (SOI), and demonstrate how the SOI-KF framework of [13] can be extended and used for transmit beamforming with limited feedback if the channel follows an AR model. Moreover, we derive closed-form expressions for the channel estimation mean-squared error (MSE), and very tight closed-form approximations for the achievable average SNR, under certain conditions. Furthermore, for general (non-AR or even unknown) channel models, a novel channel tracking approach is proposed that exploits the quantization bits in a maximum *a posteriori* (MAP) estimation formulation. Simulations confirm that by exploiting the high temporal and/or spatial correlation of the channel, and with very limited feedback rate (i.e., 2-bits per block), the performance achieved using the proposed approaches is close to that attainable with perfect CSI at the transmitter. The performance degrades, however, when the channel correlation is weak. Simulations also show that very large-size codebooks are required for codebook-based beamforming to achieve the same performance as the proposed approaches. Our results advocate for using transmit beamforming for massive MIMO in FDD mode, whereas the focus of massive MIMO has so far been on time-division duplex (TDD) operation, because of the huge feedback overhead associated with CSI feedback [14].

A conference version of part of the results in this paper will appear in [1]. This journal version includes full derivations and proofs, a fleshed-out exposition, and comprehensive simulations and comparisons to the state-of-art.

The rest of the paper is organized as follows. The limited feedback beamforming system model is presented in Section II. Channel estimation approaches with analog-amplitude feedback are provided in Section III, whereas the estimation approaches with the quantized 2-bit feedback are given in Section IV. Performance analysis and closed form results are presented in Section V. Simulations and discussions on the various trade-offs are presented in Section VI, and conclusions are drawn in Section VII. Technical derivations and proofs are deferred to the Appendix.

Notation: Boldface uppercase letters denote matrices, whereas boldface lowercase letters denote column vectors; $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian (conjugate) transpose operators, respectively; $\mathrm{Trace}(\cdot)$, $\|\cdot\|$, $|\cdot|$, $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the trace, the Euclidean norm, the absolute value, the real, and the imaginary operators, respectively; Matlab notations $\mathrm{diag}(\mathbf{x})$ and $\mathrm{Toeplitz}(\mathbf{x})$ denote the diagonal matrix and the Toeplitz matrix that are formed with vector $\mathbf{x}$, respectively; $\mathrm{mod}(x, y)$ returns the modulus after division of x by y; the operator $\odot$ denotes the Hadamard (elementwise) product of two matrices; $\mathbb{E}[\cdot]$ denotes the ensemble average; $\mathcal{CN}(\mathbf{a}, \mathbf{C})$ denotes the complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{C}$; $\mathbf{I}$ denotes the identity matrix; the function $\mathrm{sign}(x) = 1$ if $x \geq 0$ and -1 otherwise; and $Q(x) := \frac{1}{2\pi}\int_x^\infty e^{-u^2/2}du$ is the standard Gaussian tail integral.

II. System Model

Consider a downlink transmit beamforming setting comprising a transmitter with N antennas and a receiver with a single receive antenna. Extensions to account for multiple receive antennas and multiple receivers are discussed at the end of Section V. We consider a time-slotted downlink frame structure, where the duration of each slot is T seconds. We assume that at the beginning of each time slot n, the transmitter sends a unit-power pilot symbol $s(n)$ that is known at the receiver (i.e., downlink pilot rate is $1/T$ symbols/s), followed by data transmission for the remainder of the slot duration. The pilot symbol $s(n)$ is *beamformed* with a unit-norm $N \times 1$ beamforming vector $\mathbf{w}(n)$ (i.e., the weights applied to the N transmit-antenna elements when transmitting $s(n)$ are the conjugate entries of $\mathbf{w}(n)$), whereas the data symbols are *beamformed* with a different unit-norm $N \times 1$ beamforming vector $\tilde{\mathbf{w}}(n)$.

We assume that the complex $N \times 1$ vector that models the frequency-flat channel between the N transmit-antennas and the receive antenna at time slot n, denoted by $\mathbf{h}(n)$, is complex Gaussian distributed with zero mean and covariance matrix $\mathbf{C}_h$, i.e., $\mathbf{h}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_h)$, for all n. The covariance $\mathbf{C}_h$ describes the spatial correlation of the channel, and is assumed to be known at the transmitter and the receiver. The channel vector $\mathbf{h}(n)$ is assumed to be fixed within time slot n, and the random process $\{\mathbf{h}(n)\}$ is assumed to be stationary, ergodic, and *temporally correlated*.[1] A simple model for $\{\mathbf{h}(n)\}$, which allows specifying the temporal correlation of the channel, is the first-order AR model:

$$\mathbf{h}(n) = \sqrt{\alpha}\mathbf{h}(n-1) + \sqrt{1-\alpha}\mathbf{u}(n) \tag{1}$$

where $\mathbf{u}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_h)$, $\mathbf{h}(n-1)$ is statistically independent of $\mathbf{u}(n)$ for all n, and $\alpha \leq 1$ controls the degree of temporal correlation of the channel, $\mathbb{E}[\mathbf{h}(n)\mathbf{h}^H(n-k)] = \alpha^{k/2}\mathbf{C}_h$. The AR model (1) has been widely considered in the literature to model the temporal progression of the channel (see, for example, [15], [16], [12]). Extending (1) to higher orders is straightforward [11, Ch. 13]. The channel is not restricted to the model (1) in this work, but (1) is considered for its analytical tractability. Note that unlike the common assumption in the literature on limited feedback (cf. [2] and references therein), we do *not* assume that the channel is perfectly known at the receiver.

[1]The fixed per-slot channel assumption is mainly intended to simplify the analytical derivations and for simulation purposes; relaxing this assumption has no impact on the proposed channel tracking algorithms.

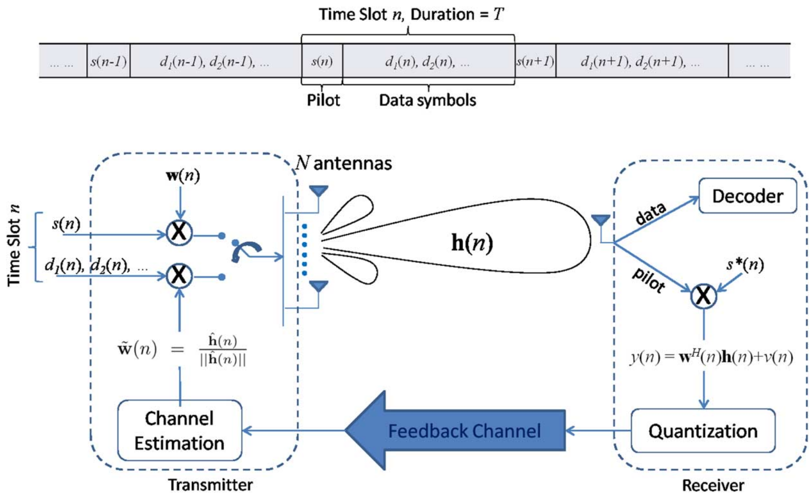

Fig. 1. Downlink frame structure and limited feedback beamforming system model.

The received signal that corresponds to the transmitted pilot $\mathbf{s}(n)$ can be expressed as

$$\bar{y}(n) = \mathbf{w}^H(n)\mathbf{h}(n)s(n) + \bar{v}(n) \tag{2}$$

where the random variable $\bar{v}(n) \sim \mathcal{CN}(0, \sigma_v^2)$ models the additive white Gaussian noise (AWGN), and $\{\bar{v}(n)\}$ are independent and identically distributed (i.i.d.). Multiplying the received signal $\bar{y}(n)$ by $s^*(n)$ (i.e., de-scrambling) at the receiver yields

$$y(n) := s^*(n)\bar{y}(n) = \mathbf{w}^H(n)\mathbf{h}(n) + v(n) \tag{3}$$

where $v(n) \sim \mathcal{CN}(0, \sigma_v^2)$ and $\{v(n)\}$ are i.i.d.

The receiver then passes $y(n)$ through a quantizer, and the output quantization bits are sent to the transmitter through an uplink feedback channel. The challenge at the transmitter is to estimate and track the channel $\mathbf{h}(n)$ using such few (periodic) feedback bits. The transmitter then uses the channel estimate $\hat{\mathbf{h}}(n)$ to design the beamforming vector that is used for data transmission in time slot n as $\tilde{\mathbf{w}}(n) = \frac{\hat{\mathbf{h}}(n)}{\|\hat{\mathbf{h}}(n)\|}$. Assuming that the data symbols are temporally white with zero-mean and unit-variance, and that the AWGN is zero-mean and unit-variance, the average receive-SNR can be expressed as[2] $\gamma = \mathbb{E}[|\tilde{\mathbf{w}}^H(n)\mathbf{h}(n)|^2]$. Several design approaches for the pilot beamforming vector $\mathbf{w}(n)$ are discussed in Section V, and compared in Section VI. The time-slotted downlink frame structure and the proposed limited feedback beamforming system are illustrated in Fig. 1.

In Section III, we first consider the case where the receiver feeds back the complex analog-amplitude (or finely-quantized) signal $y(n)$ to the transmitter at each time slot, yielding a bound on the performance with quantization. The more practical case with *very* limited feedback, where the receiver feeds back only 2 bits to the transmitter at each time slot, is then considered in Section IV.

III. Analog-Amplitude Feedback

Here we assume that the receiver will send the complex analog-amplitude (or finely-quantized) signal $y(n)$ to the transmitter through an uplink feedback channel. Assuming an AR channel model, we first consider a KF approach for estimating and tracking $\mathbf{h}(n)$, followed by a minimum mean-square error (MMSE) approach that can be applied for any channel model.

A. KF Approach

Assuming an AR channel evolution model as (1), in addition to the linear observation model of $y(n)$ as (3), the transmitter can apply the KF iterations to estimate and track $\mathbf{h}(n)$ from $\{y(k)\}_{k=1}^n$ [11, Ch. 13]. KF has been considered for tracking a time-correlated channel in [12], [15], [16].

Define the vector of observations $\mathbf{y}_n := [y(n), y(n-1), \ldots, y(1)]^T$ and the innovation

$$\breve{y}(n) := y(n) - \mathbf{w}^H(n)\tilde{\mathbf{h}}(n) \tag{4}$$

where $\tilde{\mathbf{h}}(n) := \mathbb{E}[\mathbf{h}(n)|\mathbf{y}_{n-1}]$ is the predicted channel vector, which equals $\sqrt{\alpha}\hat{\mathbf{h}}(n-1)$ for the considered AR model. Exploiting that the posterior distribution $p(\mathbf{h}(n)|\mathbf{y}_n)$ is Gaussian for the *linear Gaussian* state and observation models considered, the MMSE estimate of $\mathbf{h}(n)$ can be recursively obtained by the KF equations [11, Ch. 13]:

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{KF}}(n) &= \mathbb{E}\left[\mathbf{h}(n)|\mathbf{y}_n\right] \\ &= \mathbb{E}\left[\mathbf{h}(n)|\mathbf{y}_{n-1}\right] + \mathbb{E}\left[\mathbf{h}(n)|\breve{y}(n)\right] \\ &= \sqrt{\alpha}\hat{\mathbf{h}}_{\mathrm{KF}}(n-1) + \frac{\tilde{\mathbf{M}}(n)\mathbf{w}(n)}{\mathbf{w}^H(n)\tilde{\mathbf{M}}(n)\mathbf{w}(n) + \sigma_v^2}\breve{y}(n)\end{aligned} \tag{5}$$

where the prediction error covariance matrix (ECM) is

$$\begin{aligned}\tilde{\mathbf{M}}(n) &= \mathbb{E}\left[\left(\mathbf{h}(n) - \tilde{\mathbf{h}}(n)\right)\left(\mathbf{h}(n) - \tilde{\mathbf{h}}(n)\right)^H\right] \\ &= \alpha\mathbf{M}_{\mathrm{KF}}(n-1) + (1-\alpha)\mathbf{C}_h\end{aligned} \tag{6}$$

and the estimation ECM is

$$\begin{aligned}\mathbf{M}_{\mathrm{KF}}(n) &= \mathbb{E}\left[\left(\mathbf{h}(n) - \hat{\mathbf{h}}(n)\right)\left(\mathbf{h}(n) - \hat{\mathbf{h}}(n)\right)^H\right] \\ &= \tilde{\mathbf{M}}(n) - \frac{\tilde{\mathbf{M}}(n)\mathbf{w}(n)\mathbf{w}^H(n)\tilde{\mathbf{M}}(n)}{\mathbf{w}^H(n)\tilde{\mathbf{M}}(n)\mathbf{w}(n) + \sigma_v^2}\end{aligned} \tag{7}$$

[2]Feedback delay is not considered in this work. The effect of feedback delay on the throughput has been considered in [8].

For a general (non-AR) channel model, one approach is to approximate the actual channel evolution by the AR model (1), using α that gives the best performance (e.g., α that minimizes the average estimation error or maximizes the average achieved SNR). The performance of this approach is illustrated in Section VI. We next consider a different channel tracking approach that does not require a specific channel evolution model.

B. MMSE Approach

Here we consider a simple and general approach that does not assume a model for $\mathbf{h}(n)$. When estimating $\mathbf{h}(n)$ using the current and prior observations $\{y(k)\}_{k=1}^{n}$, more weight should be given to recent observations, while older observations should be given less weight. Motivated by the exponentially-weighted recursive least-squares (RLS) algorithm [17, Ch. 30], we consider approximating the set of observations $\{y(k) = \mathbf{w}^H(k)\mathbf{h}(k) + v(k)\}_{k=1}^{n}$ with the set $\left\{y(k) = \mathbf{w}^H(k)\mathbf{h}(n) + \lambda^{\frac{k-n}{2}} v(k)\right\}_{k=1}^{n}$, where $0 \leq \lambda \leq 1$. The role of the *forgetting factor* λ is to (exponentially) increase the noise variance of the older observations, implying more uncertainty in the approximate equality of the linear measurement $y(k) = \mathbf{w}^H(k)\mathbf{h}(n)$ as $n-k$ increases.

Define the beamforming matrix $\mathbf{W}_n := [\mathbf{w}(n), \mathbf{w}(n-1), \ldots, \mathbf{w}(1)]^H$ and the diagonal noise covariance matrix $\mathbf{C}_v = \sigma_v^2 \mathrm{diag}([1, \lambda^{-1}, \ldots, \lambda^{-n}])$. Hence, the MMSE estimate of $\mathbf{h}(n)$, assuming the linear Gaussian observations $\{y(k) = \mathbf{w}^H(k)\mathbf{h}(n) + \lambda^{-(n-k)/2} v(k)\}_{k=1}^{n}$, can be obtained as [11, Ch. 12]

$$\hat{\mathbf{h}}_{\mathrm{MMSE}}(n) = \mathbf{C}_h \mathbf{W}_n^H \left(\mathbf{W}_n \mathbf{C}_h \mathbf{W}_n^H + \mathbf{C}_v\right)^{-1} \mathbf{y}_n \quad (8)$$

The matrix $\mathbf{C}_h \mathbf{W}_n^H (\mathbf{W}_n \mathbf{C}_h \mathbf{W}_n^H + \mathbf{C}_v)^{-1}$ can be pre-computed for each n in order to reduce the run-time computational complexity. Note that, because of the exponential decay, only finite-size matrices $\mathbf{W}_n$ and $\mathbf{C}_v$ are needed to compute $\hat{\mathbf{h}}_{\mathrm{MMSE}}(n)$ using (8), as $n \to \infty$. The main challenge in this MMSE approach is to find the value of λ that gives the best performance for each channel model. Performance comparisons between the KF approach and the MMSE approach are considered in Section VI for different channel models.

It is worth mentioning that if $\mathbf{h}(n)$ is assumed deterministic instead of random, the exponentially-weighted RLS algorithm can be applied to estimate and track $\mathbf{h}(n)$ from $\{y(k)\}_{k=1}^{n}$ [17, Ch. 30]. It is also worth mentioning that if second order statistics are available, i.e., $E[\mathbf{h}(n)\mathbf{h}^H(n-k)]$ for all k, then Wiener filtering (WF) can be applied [11, Ch. 12]. Assuming, for example, that $E[\mathbf{h}(n)\mathbf{h}^H(n-k)] = \rho_k \mathbf{C}_h$ (where $\rho_0 = 1$ and ρ_k is known for $k \geq 1$), the WF channel estimate can be obtained as:

$$\hat{\mathbf{h}}_{\mathrm{WF}}(n) = \mathbf{C}_h \tilde{\mathbf{W}}_n^H \left(\mathbf{W}_n \mathbf{C}_h \mathbf{W}_n^H \odot \Gamma + \tilde{\mathbf{C}}_v\right)^{-1} \mathbf{y}_n \quad (9)$$

where $\tilde{\mathbf{C}}_v := \sigma_v^2 \mathbf{I}$, $\tilde{\mathbf{W}}_n := [\mathbf{w}(n), \rho_1 \mathbf{w}(n-1), \ldots, \rho_n \mathbf{w}(1)]^H$ and $\Gamma := \mathrm{Toeplitz}([\rho_0, \rho_1, \ldots, \rho_n])$.

IV. 2-Bit Quantized Feedback

Sending the complex analog-amplitude (or finely-quantized) signal $y(n)$ via the uplink feedback channel entails a large overhead in terms of the uplink resources (rate, transmit-power). Instead, consider the following 2-bit quantization scheme at the receiver. It is easy to see that the KF channel tracking approach in (5) depends on the innovation $\breve{y}(n)$ defined in (4), i.e., the difference between the current observation and the predicted observation based on past observations. Thus, we consider one-bit quantization for the real part of $\breve{y}(n)$, and one-bit quantization for the imaginary part $\breve{y}(n)$. This can be expressed as

$$b_r(n) = \mathrm{sign}\left[\mathrm{Re}\{y(n)\} - d_r(n)\right] \quad (10)$$

$$b_i(n) = \mathrm{sign}\left[\mathrm{Im}\{y(n)\} - d_i(n)\right] \quad (11)$$

where $d_r(n) := \mathrm{Re}\{\mathbf{w}^H(n)\tilde{\mathbf{h}}(n)\}$, $d_i(n) := \mathrm{Im}\{\mathbf{w}^H(n)\tilde{\mathbf{h}}(n)\}$, and $\tilde{\mathbf{h}}(n)$ is the predicted channel given the past observations. In order to compute $d_r(n)$ and $d_i(n)$ that are required to perform the 2-bit quantization in (10) and (11), the receiver has to know the pilot beamforming vector $\mathbf{w}(n)$, and must compute $\tilde{\mathbf{h}}(n)$ in the same way as the transmitter, as will be discussed later.

After the quantization, the receiver sends the two bits $b_r(n)$ and $b_i(n)$ to the transmitter via the uplink feedback channel. The feedback channel is assumed free of errors, which is a typical assumption in the literature on limited feedback [2]. We use the term 'frugal feedback' to describe this feedback process, where the term 'frugal' carries a double implication: *low on resources* (bits here) but *judiciously allocated*. It is the fact that we quantize $\breve{y}(n)$ that enables the good performance, which is not tenable with 'any two' bits.

Note that with such 2-bit quantization, the downlink pilot rate is only $1/T$ symbols/s, and the uplink feedback rate is only $2/T$ bits/s. The challenge here is whether the transmitter can accurately estimate and track the complex N-dimensional channel $\mathbf{h}(n)$, using only the periodically received pairs of feedback bits $b_r(n)$ and $b_i(n)$. To address this challenge, we first consider a SOI-KF approach (based on [13]) that is suitable for the AR channel model, followed by a novel MAP approach that is applicable for general channel models.

A. SOI-KF Approach

Here we assume the AR channel model in (1), and the binary observation model given by (10) and (11), where $\tilde{\mathbf{h}}(n) = \sqrt{\alpha}\hat{\mathbf{h}}(n-1)$ for the AR model. To estimate and track $\mathbf{h}(n)$ at the transmitter using $\{b_r(k)\}_{k=1}^{n}$ and $\{b_i(k)\}_{k=1}^{n}$, we extend the SOI-KF framework from the real vector space considered in [13] to the complex vector space. To facilitate operating in the more convenient real domain, consider the following definitions:

$$\mathbf{b}_n := [b_r(1), \ldots, b_r(n), b_i(1), \ldots, b_i(n)]^T,$$

$$\mathbf{w}_r(n) := \left[\mathrm{Re}\{\mathbf{w}(n)\}^T, \mathrm{Im}\{\mathbf{w}(n)\}^T\right]^T,$$

$$\mathbf{w}_i(n) := \left[-\mathrm{Im}\{\mathbf{w}(n)\}^T, \mathrm{Re}\{\mathbf{w}(n)\}^T\right]^T,$$

$$\bar{\mathbf{h}}(n) := \left[\mathrm{Re}\{\mathbf{h}(n)\}^T, \mathrm{Im}\{\mathbf{h}(n)\}^T\right]^T,$$

$$\bar{\mathbf{C}}_h := \mathbb{E}\left[\bar{\mathbf{h}}(n)\bar{\mathbf{h}}(n)^H\right]$$

such that $\mathrm{Re}\{\mathbf{w}^H(n)\mathbf{h}(n)\} = \mathbf{w}_r^T(n)\bar{\mathbf{h}}(n)$ and $\mathrm{Im}\{\mathbf{w}^H(n)\mathbf{h}(n)\} = \mathbf{w}_i^T(n)\bar{\mathbf{h}}(n)$.

The distribution $p(\bar{\mathbf{h}}(n)|\mathbf{b}_n)$ is not necessarily Gaussian because the binary observation model is not linear, and hence the exact MMSE estimator, i.e., $\mathbb{E}[\mathbf{h}(n)|\mathbf{b}_n]$, requires multiple nested numerical integrations to compute the posterior distribution $p(\bar{\mathbf{h}}(n)|\mathbf{b}_n)$ [13]. Assuming that $p(\bar{\mathbf{h}}(n)|\mathbf{b}_{n-1}) = \mathcal{N}(\sqrt{\alpha}\hat{\mathbf{h}}(n-1), \breve{\mathbf{M}}(n))$, and utilizing the results of [13], the MMSE estimate $\hat{\mathbf{h}}_{\text{SOI-KF}}(n) := \mathbb{E}[\bar{\mathbf{h}}(n)|\mathbf{b}_n]$ can be obtained using the following KF-like recursive equations (cf. [13]):

$$\begin{aligned}\hat{\mathbf{h}}_{\text{SOI-KF}}(n) = \sqrt{\alpha}\hat{\mathbf{h}}_{\text{SOI-KF}}(n-1) &+ \frac{\sqrt{\frac{2}{\pi}}\breve{\mathbf{M}}(n)\mathbf{w}_r(n)b_r(n)}{\sqrt{\mathbf{w}_r^T(n)\breve{\mathbf{M}}(n)\mathbf{w}_r(n)+\sigma_v^2/2}} \\ &+ \frac{\sqrt{\frac{2}{\pi}}\bar{\mathbf{M}}(n)\mathbf{w}_i(n)b_i(n)}{\sqrt{\mathbf{w}_i^T(n)\bar{\mathbf{M}}(n)\mathbf{w}_i(n)+\sigma_v^2/2}}\end{aligned} \tag{12}$$

where

$$\breve{\mathbf{M}}(n) = \alpha\mathbf{M}_{\text{SOI-KF}}(n-1) + (1-\alpha)\bar{\mathbf{C}}_h \tag{13}$$

$$\bar{\mathbf{M}}(n) = \breve{\mathbf{M}}(n) - \frac{\left(\frac{2}{\pi}\right)\breve{\mathbf{M}}(n)\mathbf{w}_r(n)\mathbf{w}_r^T(n)\breve{\mathbf{M}}(n)}{\mathbf{w}_r^T(n)\breve{\mathbf{M}}(n)\mathbf{w}_r(n)+\sigma_v^2/2} \tag{14}$$

$$\mathbf{M}_{\text{SOI-KF}}(n) = \bar{\mathbf{M}}(n) - \frac{\left(\frac{2}{\pi}\right)\bar{\mathbf{M}}(n)\mathbf{w}_i(n)\mathbf{w}_i^T(n)\bar{\mathbf{M}}(n)}{\mathbf{w}_i^T(n)\bar{\mathbf{M}}(n)\mathbf{w}_i(n)+\sigma_v^2/2} \tag{15}$$

There are two underlying assumptions in the SOI-KF approach: (1) the actual channel model follows an AR model; and (2) the distribution $p(\mathbf{h}(n)|\mathbf{b}_{n-1})$ is Gaussian. Relaxing both assumptions, we next develop a MAP estimation and tracking approach that does not assume a specific channel evolution model, and which can yield superior performance relative to the SOI-KF approach, as we will show in the simulations.

B. 2-Bit MAP Approach

We consider the same exponential weighting idea that is used in Section III-B, where the set of measurements $\{y(k) = \mathbf{w}^H(k)\mathbf{h}(k) + v(k)\}_{k=1}^n$ is approximated and replaced with the set $\left\{y(k) = \mathbf{w}^H(k)\mathbf{h}(n) + \lambda^{\frac{k-n}{2}}v(k)\right\}_{k=1}^n$ for $0 \le \lambda \le 1$. Using this assumption, we formulate a MAP estimation problem for $\mathbf{h}(n)$, given the $2n$ measurement bits $\{b_r(k)\}_{k=1}^n$ and $\{b_i(k)\}_{k=1}^n$ [11, Ch. 11]. Note that without assuming a specific channel model, the predicted channel can be taken to be the same as its most recent estimate, i.e., $\tilde{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1)$.

The probability that $b_r(k) = 1$ (and similarly for the probability that $b_i(k) = 1$) at time slot n given $\mathbf{h}(n)$ can be expressed in terms of the Q-function as

$$\begin{aligned}p\left[b_r(k)=1|\bar{\mathbf{h}}(n)\right] &= p\left[\mathbf{w}_r^T(k)\bar{\mathbf{h}}(n) + \lambda^{\frac{k-n}{2}}\text{Re}\{v(k)\} \ge d_r(k)\right] \\ &= Q\left(\frac{d_r(k) - \mathbf{w}_r^T(k)\bar{\mathbf{h}}(n)}{\sigma_n(k)}\right)\end{aligned} \tag{16}$$

where $\sigma_n(k) := \lambda^{\frac{k-n}{2}}\sigma_v/\sqrt{2}$. Since the noise samples $\{v(k)\}_{k=1}^n$ are independent, the probability mass function (PMF) of $\mathbf{b}_n$, given $\bar{\mathbf{h}}(n)$, is given as

$$\begin{aligned}p\left[\mathbf{b}_n|\bar{\mathbf{h}}(n)\right] = &\prod_{k=1}^n Q\left(\frac{-b_r(k)\left(\mathbf{w}_r^T(k)\bar{\mathbf{h}}(n) - d_r(k)\right)}{\sigma_n(k)}\right) \\ &\times \prod_{k=1}^n Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\bar{\mathbf{h}}(n) - d_i(k)\right)}{\sigma_n(k)}\right)\end{aligned} \tag{17}$$

Now the MAP estimate can be obtained as

$$\begin{aligned}&\hat{\mathbf{h}}_{\text{MAP}}(n) \\ &= \arg\max_{\bar{\mathbf{h}}(n)} p\left[\mathbf{b}_n|\bar{\mathbf{h}}(n)\right] p\left[\bar{\mathbf{h}}(n)\right] \\ &= \arg\max_{\bar{\mathbf{h}}(n)} \sum_{k=1}^n \Bigg[\log Q\left(\frac{-b_r(k)\left(\mathbf{w}_r^T(k)\bar{\mathbf{h}}(n) - d_r(k)\right)}{\sigma_n(k)}\right) \\ &\qquad + \log Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\bar{\mathbf{h}}(n) - d_i(k)\right)}{\sigma_n(k)}\right)\Bigg] \\ &\quad - \frac{1}{2}\bar{\mathbf{h}}(n)^T\bar{\mathbf{C}}_h^{-1}\bar{\mathbf{h}}(n)\end{aligned} \tag{18}$$

Since the Q-function is log-concave [18, pp. 104], problem (18) is convex and can be solved efficiently using Newton's method [18, Sec. 9.5].

In Newton's method, defining the function $\Phi_n(\mathbf{x})$ as the negative of the objective function in (18) (defined explicitly in (20)), and starting from a feasible initial point $\mathbf{x}$, multiple damped Newton steps of type

$$\mathbf{x} = \mathbf{x} - \beta\left(\nabla^2\Phi_n(\mathbf{x})\right)^{-1}\nabla\Phi_n(\mathbf{x}) \tag{19}$$

are used to find the minimizer of the convex function $\Phi_n(\mathbf{x})$ (where $\beta \ge 0$ is the step-size). Closed form expressions for the gradient vector $\nabla\Phi_n(\mathbf{x})$ and the Hessian matrix $\nabla^2\Phi_n(\mathbf{x})$ are derived in (21) and (22), respectively.

$$\begin{aligned}\Phi_n(\mathbf{x}) := -\sum_{k=1}^n &\Bigg[\log Q\left(\frac{-b_r(k)\left(\mathbf{w}_r^T(k)\mathbf{x} - d_r(k)\right)}{\sigma_n(k)}\right) \\ &+ \log Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\mathbf{x} - d_i(k)\right)}{\sigma_n(k)}\right)\Bigg] \\ &+ \frac{1}{2}\mathbf{x}^T\bar{\mathbf{C}}_h^{-1}\mathbf{x}\end{aligned} \tag{20}$$

$$\begin{aligned}\nabla\Phi_n(\mathbf{x}) = -\sum_{k=1}^n &\frac{1}{\sqrt{2\pi\sigma_n^2(k)}} \\ &\times \left[\frac{b_r(k)e^{-\frac{\left(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k)\right)^2}{2\sigma_n^2(k)}}}{Q\left(\frac{-b_r(k)(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k))}{\sigma_n(k)}\right)}\mathbf{w}_r(k) + \frac{b_i(k)e^{-\frac{\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)^2}{2\sigma_n^2(k)}}}{Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)}{\sigma_n(k)}\right)}\mathbf{w}_i(k)\right] \\ &+ \bar{\mathbf{C}}_h^{-1}\mathbf{x}\end{aligned} \tag{21}$$

$$\begin{aligned}\nabla^2\Phi_n(\mathbf{x})=&\sum_{k=1}^{n}\left[\frac{e^{-\frac{\left(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k)\right)^2}{\sigma_n^2(k)}}}{2\pi\sigma_n^2(k)\left[Q\left(\frac{-b_r(k)(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k))}{\sigma_n(k)}\right)\right]^2}\right.\\&\left.+\frac{b_r(k)\big(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k)\big)e^{-\frac{\left(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k)\right)^2}{2\sigma_n^2(k)}}}{\sqrt{2\pi}\sigma_n^3(k)Q\left(\frac{-b_r(k)(\mathbf{w}_r^T(k)\mathbf{x}-d_r(k))}{\sigma_n(k)}\right)}\right]\\&\times\mathbf{w}_r(k)\mathbf{w}_r^T(k)\\&+\sum_{k=1}^{n}\left[\frac{e^{-\frac{\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)^2}{\sigma_n^2(k)}}}{2\pi\sigma_n^2(k)\left[Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)}{\sigma_n(k)}\right)\right]^2}\right.\\&\left.+\frac{b_i(k)\big(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\big)e^{-\frac{\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)^2}{2\sigma_n^2(k)}}}{\sqrt{2\pi}\sigma_n^3(k)Q\left(\frac{-b_i(k)\left(\mathbf{w}_i^T(k)\mathbf{x}-d_i(k)\right)}{\sigma_n(k)}\right)}\right]\\&\times\mathbf{w}_i(k)\mathbf{w}_i^T(k)\\&+\bar{\mathbf{C}}_h^{-1}\end{aligned}\qquad(22)$$

In order to reduce the complexity of solving (18) exactly, we consider applying only a single iteration of Newton's method (with unit-step $\beta = 1$) to obtain $\hat{\mathbf{h}}(n)$, using $\hat{\mathbf{h}}(n-1)$ as the initial point. The proposed low-complexity approximate MAP (AMAP) estimate can be expressed as

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{AMAP}}(n)&=\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1)\\&-\left(\nabla^2\Phi_n\left(\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1)\right)\right)^{-1}\nabla\Phi_n\left(\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1)\right)\end{aligned}\qquad(23)$$

Intuitively, when the channel is tracked well, the actual channel $\bar{\mathbf{h}}(n)$ at time n is very close to the estimated channel $\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1)$ at time $n-1$, hence a single Newton step is sufficient to obtain a close approximation of the exact MAP estimate (18). For the rest of this paper, references to the *2-bit MAP approach* will mean the AMAP in (23), not the exact MAP in (18).

The complexity of computing $\hat{\mathbf{h}}_{\mathrm{AMAP}}(n)$ using (23) is determined by computing and inverting the $2N \times 2N$ Hessian matrix $\nabla^2\Phi_n(\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1))$. Note that because of the exponential increase of $\sigma_n(k)$ as $n-k$ increases, the number of measurement bits that are required to compute $\nabla^2\Phi_n(\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1))$ and $\nabla\Phi_n(\hat{\mathbf{h}}_{\mathrm{AMAP}}(n-1))$ (and the corresponding terms in the summation), as $n \to \infty$, are finite. The 2-bit MAP approach is computationally more complex than the SOI-KF approach; however, the performance of the 2-bit MAP approach can be better than that of the SOI-KF approach, as shown in Section VI. It is also worth mentioning that, in terms of applications, the proposed 2-bit MAP approach is not restricted to channel tracking—it can be used for general estimation and tracking problems with (very) limited feedback.

V. Performance Analysis

It is clear that the performance of the considered channel tracking schemes depends on the actual channel model and the choice of pilot beamforming vectors $\{\mathbf{w}(n)\}$. In this section we restrict attention to the analytically tractable AR channel model (1).

A *greedy* beamforming design strategy for the KF approach is to use the beamforming vector $\mathbf{w}(n)$ that minimizes $\mathrm{Trace}(\mathbf{M}_{\mathrm{KF}}(n))$ at time n. This has been considered in [19]. From (7), the optimization problem can be expressed as

$$\mathbf{w}(n)=\arg\max_{\mathbf{w},\mathrm{s.t.}\|\mathbf{w}\|^2=1}\mathrm{Trace}\left(\frac{\tilde{\mathbf{M}}(n)\mathbf{w}\mathbf{w}^H\tilde{\mathbf{M}}(n)}{\mathbf{w}^H\tilde{\mathbf{M}}(n)\mathbf{w}+\sigma_v^2}\right)\qquad(24)$$

The objective function in (24) can be expressed as a Rayleigh quotient as

$$\begin{aligned}\mathrm{Trace}\left(\frac{\tilde{\mathbf{M}}(n)\mathbf{w}\mathbf{w}^H\tilde{\mathbf{M}}(n)}{\mathbf{w}^H\tilde{\mathbf{M}}(n)\mathbf{w}+\sigma_v^2}\right)&=\frac{\mathbf{w}^H\tilde{\mathbf{M}}^2(n)\mathbf{w}}{\mathbf{w}^H\left(\tilde{\mathbf{M}}(n)+\sigma_v^2\mathbf{I}\right)\mathbf{w}}\\&=\frac{\mathbf{z}^H\mathbf{B}^{-1/2}\tilde{\mathbf{M}}^2(n)\mathbf{B}^{-1/2}\mathbf{z}}{\mathbf{z}^H\mathbf{z}}\end{aligned}$$

where $\mathbf{B} = \tilde{\mathbf{M}}(n) + \sigma_v^2\mathbf{I}$ and $\mathbf{z} = \mathbf{B}^{1/2}\mathbf{w}$. The optimal $\mathbf{z}$ that maximizes the Rayleigh quotient $\frac{\mathbf{z}^H\mathbf{E}\mathbf{z}}{\mathbf{z}^H\mathbf{z}}$, where $\mathbf{E} := \mathbf{B}^{-1/2}\tilde{\mathbf{M}}^2(n)\mathbf{B}^{-1/2}$, is the eigenvector that corresponds to the maximum eigenvalue of $\mathbf{E}$, denoted $\mathbf{z}^*$. Then the optimal beamforming vector solution to (24) is obtained as $\mathbf{w}(n) = \frac{\mathbf{B}^{-1/2}\mathbf{z}^*}{\|\mathbf{B}^{-1/2}\mathbf{z}^*\|}$.

Note that there are no guarantees that this greedy beamforming approach yields the best overall estimation/tracking performance for more than one time slot. In fact, we show in the next section via simulations that a different simple beamforming scheme can outperform this greedy beamforming approach, when the channel is spatially correlated (i.e., $\mathbf{C}_h$ is not a diagonal matrix). If $\mathbf{C}_h = \sigma_h^2\mathbf{I}$, and the initial ECM $\mathbf{M}_{\mathrm{KF}}(0) = \nu\mathbf{I}$, $\nu \geq 0$, it is easy to see that the greedy optimization (24) selects a single antenna for each n, with different antennas selected in a round-robin fashion, i.e., the i-th entry of $\mathbf{w}(n)$ is 1 if $\mathrm{mod}(n, N) + 1 = i$ and 0 otherwise. In the sequel, we will refer to this beamforming scheme as *single-antenna beamforming*.

The following proposition gives a closed-form expression for the channel estimation MSE with the KF and SOI-KF approaches (for sufficiently large n), using *single-antenna beamforming*, and assuming $\mathbf{C}_h = \sigma_h^2\mathbf{I}$.

Proposition 1: Consider the AR channel model (1), the linear observation model (3), the *single-antenna beamforming* scheme, and assume that $\mathbf{C}_h = \sigma_h^2\mathbf{I}$ (and that the distribution $p(\mathbf{h}(n)|\mathbf{b}_{n-1})$ is Gaussian for the SOI-KF approach). Then,

$$\begin{aligned}\varepsilon_{\mathrm{KF}}&:=\lim_{n\to\infty}\mathrm{Trace}\left(\mathbf{M}_{\mathrm{KF}}(n)\right)\\&=N\sigma_h^2-\left(\sigma_h^2-\left(\sqrt{c_1^2+c_2}-c_1\right)\right)\frac{1-\alpha^N}{1-\alpha}\end{aligned}\qquad(25)$$

$$\begin{aligned}\varepsilon_{\mathrm{SOI-KF}}&:=\lim_{n\to\infty}\mathrm{Trace}\left(\mathbf{M}_{\mathrm{SOI-KF}}(n)\right)\\&=N\sigma_h^2-\left(\sigma_h^2-\frac{\sqrt{c_4^2+c_3c_5}-c_4}{c_3}\right)\frac{1-\alpha^N}{1-\alpha}\end{aligned}\qquad(26)$$

where

$$c_1 = \frac{(1-\alpha^N)\left(\sigma_h^2+\sigma_v^2\right)}{2\alpha^N},$$
$$c_2 = \frac{(1-\alpha^N)\sigma_h^2\sigma_v^2}{\alpha^N},$$
$$c_3 = \alpha^N - \alpha^{2N}\left(1-\frac{2}{\pi}\right),$$
$$c_4 = \frac{\sigma_v^2}{2}(1-\alpha^N) + \frac{\sigma_h^2}{2}(1-\alpha^N)\left(1-2\alpha^N\left(1-\frac{2}{\pi}\right)\right),$$
$$c_5 = \sigma_v^2\sigma_h^2(1-\alpha^N) + \sigma_h^4\left(1-\alpha^N\right)^2\left(1-\frac{2}{\pi}\right).$$

Proof: See Appendix A.

Remark 1: Note that analogous closed-form results are *not available for general KF or SOI-KF*; what allows these results here is our specific choice of pilot beamforming strategy (single-antenna beamforming), which, as we will show in the simulations, also happens to be the best among several alternatives that we tried.

Using the same assumptions as Proposition 1, and the relations $\mathbf{e}(n) := \mathbf{h}(n) - \hat{\mathbf{h}}_{\rm KF}(n)$, where $\mathbb{E}[\mathbf{e}(n)] = \mathbf{0}$, $\mathbb{E}[\|\mathbf{e}(n)\|^2] = {\rm Trace}(\mathbf{M}_{\rm KF}(n))$, $\mathbb{E}[\hat{\mathbf{h}}_{\rm KF}^H(n)\mathbf{e}(n)] = 0$ (orthogonality principle), and $\mathbb{E}[\|\mathbf{h}(n)\|^2] = N\sigma_h^2 = \mathbb{E}[\|\hat{\mathbf{h}}_{\rm KF}(n)\|^2] + \mathbb{E}[\|\mathbf{e}(n)\|^2]$, a lower bound on the average achieved SNR with the KF approach for large n can be obtained as

$$\begin{aligned}\gamma_{\rm KF} &:= \lim_{n\to\infty}\mathbb{E}\left[\frac{\left|\hat{\mathbf{h}}_{\rm KF}^H(n)\mathbf{h}(n)\right|^2}{\left\|\hat{\mathbf{h}}_{\rm KF}(n)\right\|^2}\right] \\ &= \lim_{n\to\infty}\mathbb{E}\left[\left|\left\|\hat{\mathbf{h}}_{\rm KF}(n)\right\| + \frac{\hat{\mathbf{h}}_{\rm KF}^H(n)\mathbf{e}(n)}{\left\|\hat{\mathbf{h}}_{\rm KF}(n)\right\|}\right|^2\right] \\ &= \lim_{n\to\infty}\mathbb{E}\left[\left\|\hat{\mathbf{h}}_{\rm KF}(n)\right\|^2\right] + \lim_{n\to\infty}\mathbb{E}\left[\frac{\left|\hat{\mathbf{h}}_{\rm KF}^H(n)\mathbf{e}(n)\right|^2}{\left\|\hat{\mathbf{h}}_{\rm KF}(n)\right\|^2}\right] \\ &\geq N\sigma_h^2 - \varepsilon_{\rm KF} \end{aligned} \tag{27}$$

since $\psi(n) := \mathbb{E}\left[\frac{|\hat{\mathbf{h}}_{\rm KF}^H(n)\mathbf{e}(n)|^2}{\|\hat{\mathbf{h}}_{\rm KF}(n)\|^2}\right] \geq 0$. Denoting the i-th entry of $\hat{\mathbf{h}}_{\rm KF}(n)$ as a_i for brevity,

$$\begin{aligned}\lim_{n\to\infty}\psi(n) &= \lim_{n\to\infty}\mathbb{E}\left[\frac{\left|\sum_{i=1}^N a_i^* e_i(n)\right|^2}{\sum_{i=1}^N |a_i|^2}\right] \\ &= \lim_{n\to\infty}\mathbb{E}\left[\frac{\sum_{i=1}^N |a_i|^2 |e_i(n)|^2}{\sum_{i=1}^N |a_i|^2}\right] \\ &\approx \frac{\varepsilon_{\rm KF}}{N}\end{aligned} \tag{28}$$

where the last approximation step in (28) is obtained assuming that $\hat{\mathbf{h}}_{\rm KF}^H(n)$ and $\mathbf{e}(n)$ are independent (they are uncorrelated but not necessarily independent). Hence $\gamma_{\rm KF}$ can be closely approximated as

$$\gamma_{\rm KF} \approx N\sigma_h^2 - \left(\frac{N-1}{N}\right)\varepsilon_{\rm KF} \tag{29}$$

Similarly, a lower bound on the average achieved SNR with the SOI-KF approach at large n can be obtained as

$$\begin{aligned}\gamma_{\rm SOI-KF} &:= \lim_{n\to\infty}\mathbb{E}\left[\frac{\left|\hat{\mathbf{h}}_{\rm SOI-KF}^H(n)\mathbf{h}(n)\right|^2}{\left\|\hat{\mathbf{h}}_{\rm SOI-KF}(n)\right\|^2}\right] \\ &\geq N\sigma_h^2 - \varepsilon_{\rm SOI-KF}\end{aligned} \tag{30}$$

and a close approximation is obtained as

$$\gamma_{\rm SOI-KF} \approx N\sigma_h^2 - \left(\frac{N-1}{N}\right)\varepsilon_{\rm SOI-KF} \tag{31}$$

The approximations (29) and (31), are evaluated in Section VI.

It is easy to verify in Proposition 1 that if $\alpha \to 1$ (i.e., the channel is time-invariant), then $\varepsilon_{\rm KF}, \varepsilon_{\rm SOI-KF} \to 0$ and $\gamma_{\rm KF}, \gamma_{\rm SOI-KF} \to N\sigma_h^2$. In other words if the channel is time-invariant, then the estimation error will go to zero, and the average SNR will reach the case with perfect CSI at the transmitter, as $n \to \infty$. It is also easy to check that $\varepsilon_{\rm KF}$ and $\varepsilon_{\rm SOI-KF}$ are increasing functions in N, σ_h^2, and σ_v^2. An empirical observation made in our simulations is worth mentioning: we noticed that ${\rm Trace}(\mathbf{M}_{\rm KF}(n))$ converges to the limit in (25) for $n \geq 2N$, while ${\rm Trace}(\mathbf{M}_{\rm SOI-KF}(n))$ converges to the limit in (26) for $n \geq 4N$.

A generalization to *single-antenna beamforming* is the case where the beamforming vector $\mathbf{w}(n)$ is selected as one of the columns of a $N \times N$ unitary matrix $\mathbf{U}$ in a round-robin fashion, i.e., $\mathbf{w}(n)$ is the i-th column of $\mathbf{U}$ if ${\rm mod}(n,N)+1 = i$. We will refer to this scheme as *unitary beamforming*, and note that single-antenna beamforming is a special case of unitary beamforming with $\mathbf{U} = \mathbf{I}$. Based on extensive numerical tests, we conjecture that the closed-form expressions for $\varepsilon_{\rm KF}$ and $\varepsilon_{\rm SOI-KF}$ in (25) and (26), respectively, are also applicable for the general case of unitary beamforming, using any unitary matrix $\mathbf{U}$. Moreover, we conjecture the optimality of the unitary beamforming scheme in terms of minimizing $\varepsilon_{\rm KF}$ and $\varepsilon_{\rm SOI-KF}$ (and maximizing $\gamma_{\rm KF}$ and $\gamma_{\rm SOI-KF}$), if $\mathbf{C}_h = \sigma_h^2\mathbf{I}$.

Intuitively, the beamforming vectors that are used for learning/tracking the channel should provide complementary views of the entire channel vector $\mathbf{h}(n)$. For example, the $N \times N$ matrix $[\mathbf{w}(n), \mathbf{w}(n-1), \ldots, \mathbf{w}(n-N+1)]$ should be full-rank if $\mathbf{C}_h = \sigma_h^2\mathbf{I}$. Thus, the beamforming vectors $\{\mathbf{w}(n)\}$ that are used for pilots $\{s(n)\}$ for channel tracking should be different than the beamforming vectors $\{\tilde{\mathbf{w}}(n)\}$ that are used for data transmission. Choosing $\mathbf{w}(n) = \tilde{\mathbf{w}}(n) = \frac{\hat{\mathbf{h}}(n)}{\|\hat{\mathbf{h}}(n)\|}$, which is the case considered in [12], yields poor performance. This point is further elaborated in Section VI.

A. Comparing With Codebook-Based Beamforming

As mentioned earlier, the state-of-the-art in transmit beamforming with limited feedback is focused on designing a common beamformer codebook (known at the transmitter and the receiver). The setup assumes that the receiver will accurately estimate the downlink channel, search the codebook, and feed back the index of the best beamformer in the codebook to the transmitter [2]. In [6], it is stated that for beamforming over i.i.d. Rayleigh fading channels with beamformer codebook of

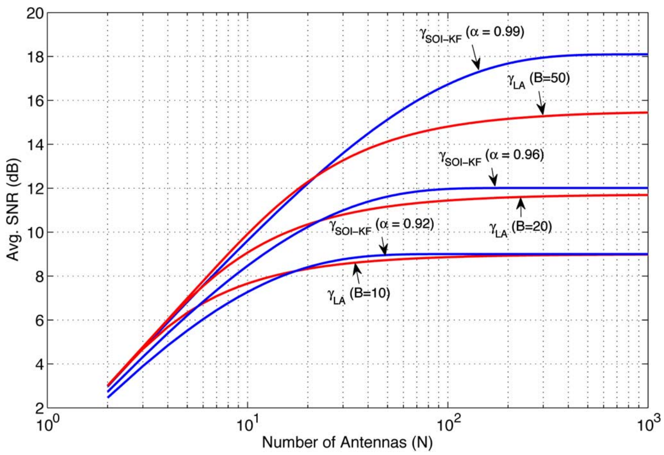

Fig. 2. Comparison between $\gamma_{\mathrm{SOI-KF}}$ with $\alpha \in \{0.92, 0.96, 0.99\}$ and γ_{LA} with $B \in \{10, 20, 50\}$ as N increases.

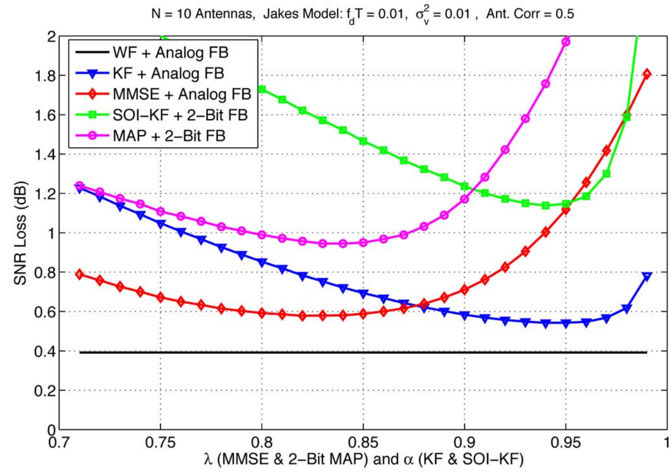

Fig. 3. Performance comparison for the considered beamforming approaches with $N = 10$ transmit-antennas, and using Jake's channel model with $f_dT = 0.01$.

size 2^B designed by the GLA, the achieved average SNR γ_{LA} can be closely approximated as

$$\gamma_{\mathrm{LA}} = N\sigma_h^2 - (N-1)\sigma_h^2 2^{-B/(N-1)} \tag{32}$$

Note that expression (32) is obtained ignoring the temporal correlation of the channel and assuming perfect CSI at the receiver (unlike the case for $\gamma_{\mathrm{SOI-KF}}$).

Fig. 2 plots the lower bound on $\gamma_{\mathrm{SOI-KF}}$ from (30) and γ_{LA} from (32) as N increases, assuming $\mathbf{C}_h = \mathbf{I}$, $\sigma_v^2 = 0.001$, $\alpha \in \{0.92, 0.96, 0.99\}$, and $B \in \{10, 20, 50\}$. The figure shows the increase of $\gamma_{\mathrm{SOI-KF}}$ as N increases and as α increases (i.e., channel becomes more correlated across time). The figure also shows that a large number of feedback bits B (i.e., large codebook) is required for codebook-based beamforming to achieve the same performance as the SOI-KF approach, which is obtained using only 2 feedback bits per channel dwell time block of length T. The number of bits B required for γ_{LA} to achieve $\gamma_{\mathrm{SOI-KF}}$ increases as N or α increases. For example, the figure shows that $\gamma_{\mathrm{SOI-KF}}$ (with $\alpha = 0.99$) outperforms γ_{LA} with $B = 10$ feedback bits for $N \geq 3$, and outperforms γ_{LA} with $B = 50$ feedback bits for $N \geq 22$.

Exploiting the channel temporal correlation to reduce the feedback rate, [7] and [8] propose modeling the quantized CSI at the receiver using a finite-state Markov chain. As shown in Fig. 2, at least $B = 50$ bits are needed to achieve the same SNR performance that is achieved with only 2 feedback bits using the SOI-KF approach when $N = 20$ and $\alpha = 0.99$, for example. This means that at least 2^{50} Markov states need to modeled and 2^{100} transition probabilities must be computed in order to apply the compression techniques in [8] and [7], which is clearly computationally prohibitive.

Before moving to the numerical results, two practice-oriented remarks are in order.

- *Variable-length quantization*. To further decrease the feedback rate to 1 bit per T, the receiver can send only the bits that correspond to the real measurements, $\{b_r(n)\}$, in even time slots, while the bits that correspond to the imaginary measurements, $\{b_i(n)\}$, are sent in odd time slots (or vice versa). On the other hand, the estimation performance can be improved by increasing the feedback quantization bits (at the cost of higher feedback rate) using the *iteratively quantized Kalman filter* approach introduced in [20], where the quantization bits are iteratively formed using the sign of the difference between the observation $y(n)$ and its estimate based on past observations along with previous bits of the current observation.
- *Multiple receive antennas*. Extending this work to a setting with more than one receive antennas (or multiple receivers) is straightforward if the receive antennas are uncorrelated. A separate estimation/tracking problem can be set up for the channel vector that corresponds to each receive antenna.

VI. Numerical Results

To test the performance of the proposed beamforming and feedback techniques, we consider the widely used Jake's channel model [21] in Figs. 3, 4, 5, and 6. According to Jake's model, the spatio-temporal correlation matrix can be expressed as $\mathbb{E}[\mathbf{h}(n)\mathbf{h}^H(n-k)] = \rho_k \mathbf{C}_h$, for $k \geq 0$, where $\rho_k := J_0(2\pi f_d T k)$, J_0 is the 0-th-order Bessel function, and f_d denotes the Doppler frequency. The *unitary beamforming* scheme that is described in Section V is used for all figures. The SNR loss, defined as the ratio of the average SNR achieved with perfect CSI at the transmitter (i.e., $\mathbb{E}[\|\mathbf{h}(n)\|^2]$) to the average SNR achieved with the estimated channel (i.e., $\mathbb{E}\left[\frac{|\hat{\mathbf{h}}^H(n)\mathbf{h}(n)|^2}{\|\hat{\mathbf{h}}(n)\|^2}\right]$), is used to measure and compare the performance of the proposed techniques.

The setup for Fig. 3 considers a transmitter with $N = 10$ antennas, Doppler frequency $f_d = 10$ Hz, time slot duration $T = 1$ ms (same performance for any values of f_d and T that satisfy $f_dT = 0.01$), spatial correlation matrix $\mathbf{C}_h = \sigma_h^2$ Toeplitz $([0.5^0, 0.5^1, \ldots, 0.5^9])$, where $\sigma_h^2 = 0.1$, and observation noise variance $\sigma_v^2 = 0.01$. The figure illustrates the trade-off between the SNR loss of the KF and SOI-KF approaches and α, the trade-off between the SNR loss of the MMSE and 2-bit MAP approaches and the forgetting factor λ, and the SNR loss using the WF (9) (which requires additional knowledge of $\rho_k =$

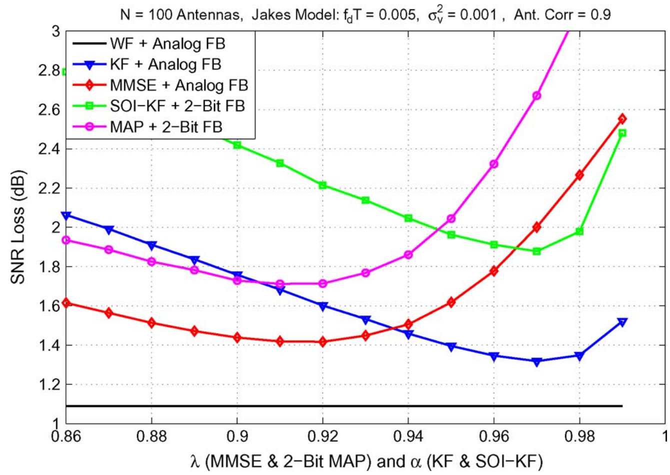

Fig. 4. Performance comparison for the considered beamforming approaches with $N = 100$ transmit-antennas, and using Jake's channel model with $f_d T = 0.005$.

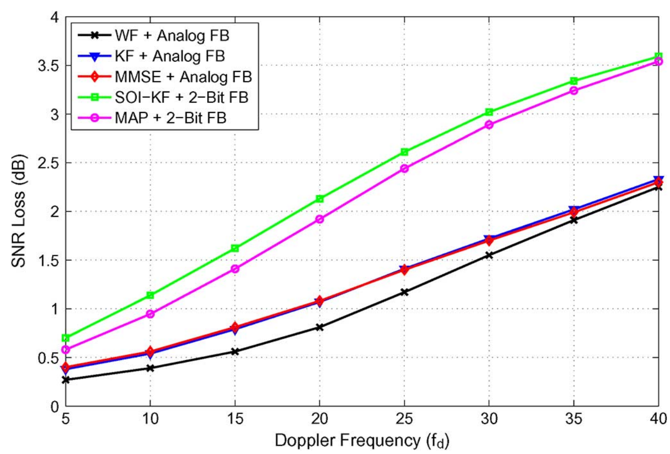

Fig. 5. SNR loss as the Doppler frequency increases in Jakes channel model with $N = 10$.

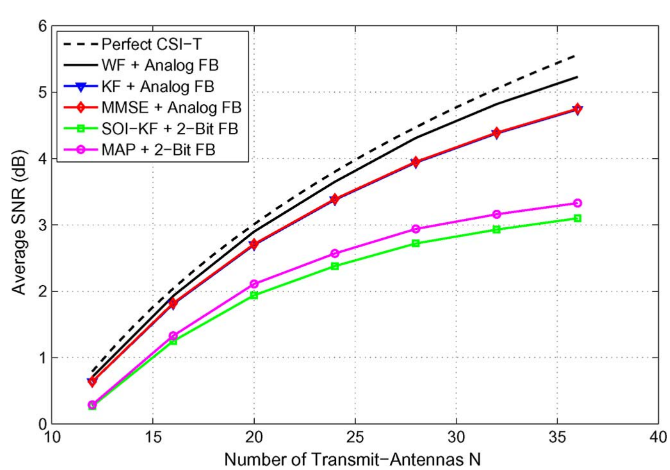

Fig. 6. Average SNR increase as N increases in Jakes channel model with $f_d = 10$ Hz.

$J_0(2\pi f_d T k)$ for all k) as a baseline. The SNR loss plots are obtained via 1000 Monte-Carlo simulation runs, where each run includes 400 time slots.

TABLE I
SNR LOSS COMPARISON OF DIFFERENT BEAMFORMING TECHNIQUES

	Unitary	Single-Ant.	Random	Greedy	KF Est.
WF	0.39	0.39	0.60	0.84	3.29
KF	0.54	0.54	0.83	0.96	4.24
MMSE	0.57	0.58	0.92	1.07	4.79
SOI-KF	1.14	1.10	1.40	1.42	4.45
2-Bit MAP	0.94	0.92	1.19	1.27	4.04

Interestingly, Fig. 3 shows that the difference between the average receive-SNR achieved using the proposed 2-bit MAP approach with only 2 feedback bits every T seconds (at the optimal $\lambda^* = 0.83$), and the *Genie* receive-SNR achieved with perfect CSI at the transmitter, is less than 1 dB. The figure also shows that the average receive-SNR achieved using the proposed 2-bit MAP approach (at $\lambda^* = 0.83$) is 0.2 dB larger than that achieved using the SOI-KF approach (at $\alpha^* = 0.94$), and is only 0.6 dB less than that achieved using WF (9). In other words, the cost of quantizing the received signal $y(n)$ into 2 feedback bits, as compared to the analog-amplitude $y(n)$ feedback, is less than 0.6 dB. Note that in the case of analog-amplitude feedback, it is assumed that $y(n)$ is perfectly known at the transmitter (in addition to the knowledge of $\{\rho_k\}$); accounting for additional uplink (or quantization) errors in the analog feedback case will further decrease the 0.6 dB difference. Another observation from the figure is that the MMSE approach (at $\lambda^* = 0.83$) and the KF approach (at $\alpha^* = 0.94$) are very close in performance. It is worth mentioning that in practice, the optimal values of α or λ for a range of channel models can be pre-computed offline and stored in a lookup table. At runtime, using the current channel statistics or estimated channel parameters (e.g., Doppler frequency), a suitable value of α or λ can be retrieved from the lookup table and applied in the channel tracking algorithm, without performing any expensive computations.

Table I uses the same setup as Fig. 3, and reports the SNR loss (in dB) with different beamforming schemes at $\lambda^* = 0.83$ and $\alpha^* = 0.94$. The considered beamforming schemes, which correspond to the columns of the table, are (in order): (i) the unitary beamforming scheme described in Section V; (ii) the single-antenna beamforming scheme described in Section V; (iii) a random beamforming scheme where $\mathbf{w}(n)$ is a normalized Gaussian random vector for each n; (iv) the greedy beamforming scheme where $\mathbf{w}(n)$ is obtained by solving (24); and (v) the case where $\mathbf{w}(n)$ corresponds to the most recent channel estimate using the KF approach (i.e., $\mathbf{w}(n) = \hat{\mathbf{h}}_{\mathrm{KF}}(n-1)/\|\hat{\mathbf{h}}_{\mathrm{KF}}(n-1)\|$). The table shows that the performance of the unitary beamforming is almost identical to that of the single-antenna beamforming (small difference within the sample averaging error), which is superior to other considered beamforming schemes. The table also verifies that the greedy beamforming scheme using (24) is not optimal, and that using $\mathbf{w}(n) = \hat{\mathbf{h}}_{\mathrm{KF}}(n-1)/\|\hat{\mathbf{h}}_{\mathrm{KF}}(n-1)\|$ yields poor performance, as discussed in Section V.

In Fig. 4, a large system with $N = 100$ antennas is considered, with Doppler frequency $f_d = 5$ Hz, spatial correlation matrix $\mathbf{C}_h = \sigma_h^2$ Toeplitz $([0.9^0, 0.9^1, \ldots, 0.9^{99}])$, where $\sigma_h^2 = 0.01$, and observation noise variance $\sigma_v^2 = 0.001$. Similar to Fig. 3, Fig. 4 illustrates the trade-off between the SNR loss and the parameters λ and α, and confirms that the proposed 2-bit MAP approach with only 2 feedback bits every T seconds is applicable even with large N. At the optimal $\lambda^* = 0.91$, the SNR achieved with 2-bit MAP approach is 1.7 dB less than the case with perfect CSI at the transmitter, 0.6 dB less than WF with analog-signal feedback, and 0.2 dB higher than the SOI-KF approach (at the optimal $\alpha^* = 0.97$). The results shown in this figure help pave the way for using massive MIMO systems in FDD mode [14], by exploiting the high spatio-temporal channel correlation.

Fig. 5 considers the same setup and network parameters as Fig. 3. The SNR loss that corresponds to the different considered estimation/tracking techniques is plotted versus the Doppler frequency, using the numerically optimized λ and α. The SNR loss is increasing with f_d as expected. The figure shows that the SNR loss due to the 2-bit quantization (i.e., 2-bit MAP and SOI-KF approaches), as compared to the case with analog-signal feedback (i.e., KF, MMSE, and WF approaches), is small for small f_d, and increases as f_d increases. The figure also shows that the 2-bit MAP approach outperforms the SOI-KF approach for the considered f_d range, and that the MMSE and KF approaches are very close in performance.

In Fig. 6, the average achieved SNR using the numerically optimized λ and α is plotted as a function of N, considering a setup with $f_d = 10$ Hz, $\mathbf{C}_h = \sigma_h^2\mathbf{I}$, $\sigma_h^2 = 0.1$, and $\sigma_v^2 = 0.01$. The figure shows that the average SNR is increasing with N as expected, and that the gap between the average SNR achieved with 2-bit quantization (using the 2-bit MAP and SOI-KF approaches) and the average SNR achieved with analog-signal feedback (using the KF, MMSE, and WF approaches), is increasing as N increases. The figure also shows that the 2-bit MAP approach outperforms the SOI-KF approach for the considered range of N, and that the MMSE and KF approaches are very close in performance. Using the average SNR expression (32) achieved using GLA for the codebook-based beamforming framework (assuming perfect CSI at the receiver), it can be shown that at least $B = 40$ bits are required to achieve the same performance as the 2-bit MAP approach when $N = 16$ (1.33 dB), and at least $B = 45$ bits are required when $N = 36$ (3.33 dB). Computing the transition probabilities for the finite-state Markov chain model, as considered in [7] and [8], is clearly prohibitive in these cases.

Fig. 7 considers the AR channel model (1), with $N = 10$, $\mathbf{C}_h = \sigma_h^2\mathbf{I}$, $\sigma_h^2 = 0.1$, and $\sigma_v^2 = 0.01$. The SNR loss for the considered techniques is plotted versus α, where the numerically optimized λ is used for the MMSE and 2-bit MAP approaches. The figure also plots the analytical approximations for the KF and SOI-KF approaches using (29) and (31), respectively. Note that for the AR model (1), the performance of the KF (5) and the WF (9) are identical for large n [11]. The figure shows the decrease of the SNR loss as α increases as expected. The figure also shows that the SOI-KF approach outperforms the 2-bit MAP approach for the considered AR channel model, and that the performances of the MMSE and KF approaches are very close. Moreover, the figure shows that the approximations derived in (29) and (31) are very tight, particularly for large α. Considering the average SNR achieved using GLA for the codebook-based beamforming, it can be shown using (32) that

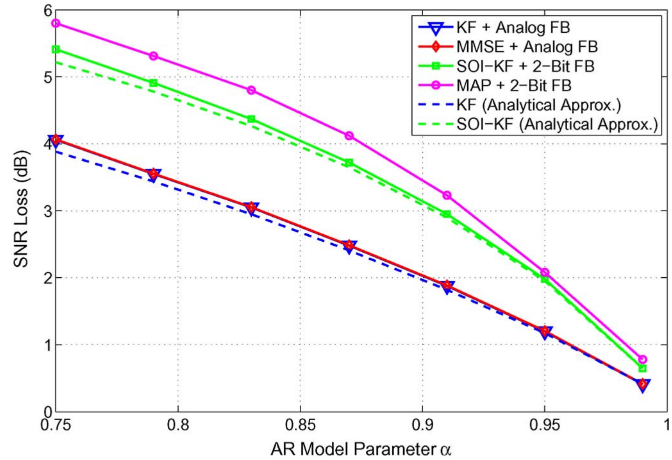

Fig. 7. Average SNR increase as α increases using the AR model (1) with $N = 10$.

at least $B = 12$ bits and $B = 25$ bits are required to achieve the same performance of the SOI-KF approach when $\alpha = 0.95$ and $\alpha = 0.99$, respectively.

VII. Conclusions

We proposed a new approach for channel tracking and transmit beamforming with (very) limited feedback. Instead of putting the burden of channel estimation and codebook search on the receiver, we shift the bulk of the work to the transmitter. Using separate beamforming weight vectors for pilot and payload transmission, the transmitter sends a single pilot symbol per channel dwell time block, while the receiver simply sends back a coarsely quantized 2-bit version of the received pilot signal (or the corresponding innovation, in the case of AR modeling). For channel tracking, we proposed a novel 2-bit MAP algorithm, as a 'universal' complement to an extended version of the SOI-KF framework, which we advocate when the channel can be modeled as an AR process. In the AR case, we derived closed-form expressions for the resulting channel MSE, and very tight approximations for the corresponding SNR, assuming circular single-antenna beamforming for the pilots. Careful simulations confirmed that by exploiting the spatio-temporal correlation of the channel, the performance achieved using the proposed *frugal feedback* approaches is close to that attainable with perfect CSI at the transmitter. Simulations also showed that very large-size codebooks are required for codebook-based beamforming to achieve the same performance as the proposed approaches. Our results help pave the way for using transmit beamforming for massive MIMO in FDD instead of TDD mode.

Appendix

A. Proof of Proposition 1

We first focus on the KF approach. It is easy to see from (6) and (7) that $\tilde{\mathbf{M}}(n)$ and $\mathbf{M}_{\mathrm{KF}}(n)$ are diagonal matrices for sufficiently large n when *single-antenna beamforming* is used. Let $\{X_0, X_1, \ldots, X_{N-1}\}$ denote the sorted (ascendingly) diagonal entries of $\mathbf{M}_{\mathrm{KF}}(n)$, and $\{\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{N-1}\}$ denote the sorted (ascendingly) diagonal entries of $\tilde{\mathbf{M}}(n)$, for large n. Since the channel entries are i.i.d. ($\mathbf{C}_h = \sigma_h^2 \mathbf{I}$), it is easy to see that the values of $\{X_0, X_1, \ldots, X_{N-1}\}$ (and $\{\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{N-1}\}$) are the same for any sufficiently large n (i.e., $n \to \infty$) because the KF estimator will be based on present and infinite past observations—only the location of X_i (and $\tilde{X}_i$) in the diagonal of $\mathbf{M}_{\mathrm{KF}}(n)$ (resp. $\breve{\mathbf{M}}(n)$) differs for different n.

From (6), we have the relation $\tilde{X}_k = \alpha X_k + (1-\alpha)\sigma_h^2$, for $k = 0, \ldots, N-1$. Assume that antenna i is used to send $s(n)$ at time n (i.e., the i-th entry of $\mathbf{w}(n)$ equals 1). Prior to time n, antenna i was last accessed at time $n-N$ with the *single-antenna beamforming*, and thus the i-th diagonal entry of $\breve{\mathbf{M}}(n)$ is the largest entry $\tilde{X}_{N-1}$. From (7), only the i-th diagonal entry of $\breve{\mathbf{M}}(n)$ is affected by the recursion in (7), yielding the smallest diagonal entry X_0 of $\mathbf{M}_{\mathrm{KF}}(n)$, whereas the rest of the diagonal entries of $\breve{\mathbf{M}}(n)$ are duplicated in $\mathbf{M}_{\mathrm{KF}}(n)$. These relations can be expressed as

$$X_k = \alpha X_{k-1} + (1-\alpha)\sigma_h^2 = \alpha^k X_0 + (1-\alpha)\sigma_h^2 \sum_{i=0}^{k-1} \alpha^i = \alpha^k X_0 + (1-\alpha^k)\sigma_h^2 \quad (33)$$

for $k = 1, \ldots, N-1$, whereas from (7),

$$X_0 = \tilde{X}_{N-1} - \frac{\tilde{X}_{N-1}^2}{\tilde{X}_{N-1} + \sigma_v^2} = \frac{\tilde{X}_{N-1}\sigma_v^2}{\tilde{X}_{N-1} + \sigma_v^2} \quad (34)$$

From (33)

$$\tilde{X}_{N-1} = \alpha X_{N-1} + (1-\alpha)\sigma_h^2 = \alpha^N X_0 + (1-\alpha^N)\sigma_h^2 \quad (35)$$

Substituting with $\tilde{X}_{N-1}$ from (35) in (34), we obtain the quadratic equation in X_0:

$$X_0^2 + \left(\sigma_v^2 + \sigma_h^2\right)\left(\frac{1-\alpha^N}{\alpha^N}\right) X_0 - \sigma_v^2 \sigma_h^2 \left(\frac{1-\alpha^N}{\alpha^N}\right) = 0 \quad (36)$$

The only positive solution for (36) is $X_0 = -c_1 + \sqrt{c_1^2 + c_2}$, where c_1 and c_2 are defined in Proposition 1. Finally, using (33),

$$\varepsilon_{\mathrm{KF}} = \sum_{k=0}^{N-1} X_k = X_0 \sum_{k=0}^{N-1} \alpha^k + N\sigma_h^2 - \sigma_h^2 \sum_{k=0}^{N-1} \alpha^k = N\sigma_h^2 + \left(X_0 - \sigma_h^2\right)\frac{1-\alpha^N}{1-\alpha} \quad (37)$$

which proves (25).

The proof of (26) for the SOI-KF approach follows along the same lines. Note that the $2N \times 2N$ matrix $\mathbf{M}_{\mathrm{SOI-KF}}(n)$ is diagonal for sufficiently large n, where the upper-left $N \times N$ sub-matrix (which corresponds to the real part) is identical to the lower-right $N \times N$ sub-matrix (which corresponds to the imaginary part). Focusing only on the upper-left sub-matrix, and defining $\{Y_0, Y_1, \ldots, Y_{N-1}\}$ and $\{\tilde{Y}_0, \tilde{Y}_1, \ldots, \tilde{Y}_{N-1}\}$ as the sorted diagonal entries of the upper-left sub-matrix of $\mathbf{M}_{\mathrm{SOI-KF}}(n)$ and $\breve{\mathbf{M}}(n)$, respectively, an expression for Y_0 in this case can be obtained from (15) as

$$Y_0 = \tilde{Y}_{N-1} - \frac{\frac{2}{\pi}\tilde{Y}_{N-1}^2}{\tilde{Y}_{N-1} + \sigma_v^2/2} \quad (38)$$

Substituting with $\tilde{Y}_{N-1} = \alpha^N Y_0 + (1-\alpha^N)\sigma_h^2/2$ in (38), we obtain a quadratic equation in Y_0, which is solved to obtain the only positive solution $Y_0 = \frac{-c_4 + \sqrt{c_4^2 + c_3 c_5}}{2c_3}$, where c_3, c_4 and c_5 are defined in Proposition 1. Then,

$$\varepsilon_{\mathrm{SOI-KF}} = 2\sum_{k=0}^{N-1} Y_k = N\sigma_h^2 + (2Y_0 - \sigma_h^2)\frac{1-\alpha^N}{1-\alpha} \quad (39)$$

which proves (26).

References

[1] O. Mehanna and N. D. Sidiropoulos, "Frugal channel tracking for transmit beamforming," presented at the 48th Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 2–5, 2014.

[2] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[3] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[4] V. K. N. Lau, Y. Liu, and T.-A. Chen, "On the design of MIMO block-fading channels with feedback-link capacity constraint," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 62–70, Jan. 2004.

[5] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. A. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[6] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1853–1863, May 2006.

[7] C. Simon and G. Leus, "Feedback reduction for spatial multiplexing with linear precoding," in *Proc. 32nd Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2007, vol. 3, pp. III-33–III-36.

[8] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Limited feedback beamforming over temporally-correlated channels," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1959–1975, May 2009.

[9] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures*, 3GPP TS 36.213 V9.2.0 LTE, Jun. 2010.

[10] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993, vol. I, Estimation Theory.

[12] M. Sadek, A. Tarighat, and A. H. Sayed, "Exploiting spatio-temporal correlation for rate-efficient transmit beamforming," in *Proc. 38th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2004, pp. 2027–2031.

[13] A. Ribeiro, G. B. Giannakis, and S. I. Roumeliotis, "SOI-KF: Distributed Kalman filtering with low-cost communications using the sign of innovations," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4782–4795, Dec. 2006.

[14] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.

[15] Z. Liu, X. Ma, and G. B. Giannakis, "Spacetime coding and Kalman filtering for time-selective fading channels," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 183–186, Feb. 2012.

[16] R. Bosisio, M. Nicoli, and U. Spagnolini, "Kalman filter of channel modes in time-varying wireless systems," in *Proc. 30th Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, vol. 3, pp. III-785–III-788.

[17] A. H. Sayed, *Fundamentals of Adaptive Filtering*. New York, NY, USA: Wiley, 2003.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[19] F. Jiang, J. Chen, and A. Swindlehurst, "Linearly reconfigurable Kalman filtering for a vector process," presented at the 38th Int. Conf. Acoustics, Speech, Signal Process. (ICASSP), Vancouver, Canada, May 26–31, 2013.

[20] E. J. Msechu, S. I. Roumeliotis, A. Ribeiro, and G. B. Giannakis, "Decentralized quantized Kalman filtering with scalable communication cost," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3727–3741, Aug. 2008.

[21] W. C. Jakes, *Microwave Mobile Communications*. New York, NY, USA: Wiley, 1974.

Omar Mehanna (S'05) received the B.Sc. degree in Electrical Engineering from Alexandria University, Egypt, in 2006, the M.Sc. degree in Electrical Engineering from Nile University, Egypt, in 2009, and the Ph.D. degree in Electrical Engineering from University of Minnesota in 2014. He is currently a Senior Systems Engineer at Qualcomm Technologies, Inc. His current research interests are in cognitive radio and coordinated multi-point communications.

Nicholas D. Sidiropoulos (F'09) received the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece, and M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland—College Park, in 1988, 1990 and 1992, respectively. He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota—Minneapolis (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the University of Minnesota—Minneapolis (2011–). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement. He received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009), and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008). He received the Distinguished Alumni Award of the Department of Electrical and Computer Engineering, University of Maryland, College Park (2013).