

Closed-Loop Beam Alignment for Massive MIMO Channel Estimation

Andrew J. Duly, *Member, IEEE*, Taejoon Kim, *Member, IEEE*, David J. Love, *Senior Member, IEEE*, and James V. Krogmeier, *Member, IEEE*

Abstract—Training sequences are designed to probe wireless channels to obtain channel state information for block-fading channels. Optimal training sounds the channel using orthogonal beamforming vectors to find an estimate that optimizes some cost function, such as mean square error. As the number of transmit antennas increases, however, the training overhead becomes significant. This creates a need for alternative channel estimation schemes for increasingly large transmit arrays. In this work, we relax the orthogonal restriction on sounding vectors. The use of a feedback channel after each forward channel use during training enables closed-loop sounding vector design. A misalignment cost function is introduced, which provides a metric to sequentially design sounding vectors. In turn, the structure of the sounding vectors aligns the transmit beamformer with the true channel direction, thereby increasing beamforming gain. This beam alignment scheme for massive MIMO is shown to improve beamforming gain over conventional orthogonal training for a MISO channel.

Index Terms—Adaptive sensing, training sequence, channel estimation, massive MIMO.

I. INTRODUCTION

IN wireless communications, accurate channel state information plays a central role in realizing the gains afforded by coherent communications. In light of this, much literature exists to design training sequences to estimate the channel. For a block-fading channel model, [1] addressed capacity maximizing values for various parameters, including the length of the training phase and the power of the training phase. The error covariance of the minimum mean square error (MMSE) channel estimate is minimized when orthogonal waveforms are transmitted from each element. For orthogonal transmission from each element, the discrete length of the training signal must be at least equal to the number of transmit antennas. For massive MIMO systems, this would lead to training consuming a large fraction of the coherence interval.

Channel estimation with training sequences of length less than the number of transmit antennas was considered for distributed transmit beamforming systems in [2]. The set of training sequences which minimizes the mean square error of the MMSE channel estimate was shown to be similar to those maximizing the expected beamforming gain, which gives a

Manuscript received August 13, 2013; revised February 21, 2014; accepted March 20, 2014. Date of publication April 8, 2014; date of current version August 8, 2014. This work was supported in part by City University of Hong Kong under Grant 7200347. The associate editor coordinating the review of this paper and approving it for publication was M. Matthaiou.

A. J. Duly, D. J. Love, and J. V. Krogmeier are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: andrew.j.duly@ieee.org; djlove@ecn.purdue.edu; jvk@ecn.purdue.edu).

T. Kim is with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: taejokim@cityu.edu.hk).

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/LCOMM.2014.2316157

training signal matrix with orthogonal columns. The optimal full-rank set of receive beamforming vectors for angle of arrival estimation was derived in [3]. The channel gain term was treated as a deterministic unknown, and beamforming vectors were derived to minimize the variance on the angle of arrival estimate.

The literature referenced thus far in this work pertains to a class of open loop schemes in which the training sequence is predetermined. These training sequences cannot be adapted to knowledge gathered of the instantaneous channel. Alternatively, adaptive training schemes use a feedback link from the receiver to the transmitter to guide the design of the training sequence. Closed-loop training across multiple blocks for massive MIMO beamforming systems is considered in [4]. Assuming a temporally correlated channel, their algorithm selected the entire training signal for a given block based on channel estimates from previous blocks. Improvements in average receive SNR were shown by utilizing the training sequences received in previous channel blocks. Line-of-sight channel estimation for large arrays in backhaul cellular networks is presented in [5]. Adaptive subspace sampling, where samples from previous channel uses aid in beamformer design, gave improved beamforming gain over non-adaptive techniques.

In this work, we focus on channel estimation for single-user massive MIMO. These techniques can be extended to multiuser scenarios, a topic of future work. As the number of antennas grows large, optimal channel estimation no longer remains practical. Some of the general properties for massive MIMO systems were shown to hold for arrays between 100 and 1,000 antennas in [6]. We target practical massive MIMO systems and develop a beam alignment scheme to estimate MISO channels on the lower end of that range. We consider beamforming for a MISO channel in a single coherent channel block. A feedback channel allows adaptation of the training sequence after each channel use. In general, current systems do not feed back information after every channel use. The increased complexity, however, may become necessary for massive MIMO systems to shorten the training phase. The training sequence is comprised of sounding vectors, which are sequentially designed in a manner that aligns the estimated channel with the true instantaneous channel direction, treating the channel gain term as a nuisance parameter. This work focuses on frequency division duplexing (FDD) systems, where channel reciprocity cannot be used for transmit beamforming. Nonetheless, this scheme remains applicable for time division duplexing (TDD) systems with no uplink sounding.

II. SYSTEM SETUP

Consider a multiple-input single-output (MISO) wireless channel with M transmit antennas and a single receive antenna. In coherent communications, the wireless channel is considered

known and knowledge of the channel at the receiver aids symbol detection and demodulation. For a given channel use, the input-output relationship is

$$r = (\sqrt{\rho}\mathbf{x})^*(\alpha\mathbf{h}) + n$$

where $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmitted signal with $E[\|\mathbf{x}\|^2] = 1$, ρ denotes the transmit power, $r \in \mathbb{C}$ is the received signal, $n \sim \mathcal{CN}(0, \sigma_n^2)$ the additive noise, $(\cdot)^*$ denotes the Hermitian transpose, and $\alpha\mathbf{h}$ is the MISO wireless channel, comprised of the complex scalar channel gain α and the channel direction vector \mathbf{h} with $\|\mathbf{h}\| = 1$. The channel direction vector \mathbf{h} describes direction only and lacks any sort of gain or attenuation.

Transmit beamforming is commonly used in point-to-point MISO systems, but it is also expected to be critical to multiuser massive MIMO systems. A beamformed transmit signal is defined as $\mathbf{x} = \mathbf{f}s$, where the beamforming vector \mathbf{f} and complex symbol s are constrained to $\|\mathbf{f}\| = 1$ and $E[|s|^2] = 1$. The input-output relationship then simplifies to $r = \sqrt{\rho}\alpha(\mathbf{f}^*\mathbf{h})s + n$, where $|\mathbf{f}^*\mathbf{h}|^2$ denotes the beamforming gain. The optimal beamformer \mathbf{f}_{opt} that maximizes the beamforming gain and achieves $|\mathbf{f}_{\text{opt}}^*\mathbf{h}|^2 = 1$ is $\mathbf{f}_{\text{opt}} = \mathbf{h}$. However, the true channel direction \mathbf{h} is unknown and must be estimated.

To estimate the wireless channel, assume K channel uses are available for this task. If the transmit signal for data transmission is restricted to beamforming, it is natural to restrict the training signal in the same fashion. Let the sounding vector \mathbf{w}_k represent the training signal for the k th channel use. The receive vector for the first k channel uses of the training phase is given by

$$\mathbf{y}_k = \alpha\sqrt{\rho}\mathbf{W}_k^*\mathbf{h} + \mathbf{n}_k \quad (1)$$

where $\mathbf{W}_k = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k]$ is the set of k sounding vectors and \mathbf{y}_k is the $k \times 1$ complex receive sample vector.

To aid in channel estimation, we consider a limited-rate feedback channel from the receiver to the transmitter as shown in Fig. 1. Let \mathcal{W} represent the set of possible sounding vectors, $\mathcal{W} = \{\mathbf{w}^{(\ell)} \in \mathbb{C}^{M \times 1} : \|\mathbf{w}^{(\ell)}\| = 1, \ell = 1, \dots, L\}$. Due to the restricted bandwidth of the feedback channel, we design \mathcal{W} to be a discrete set with $|\mathcal{W}| = L$, where $|\mathcal{W}|$ denotes the cardinality of the set \mathcal{W} , and make this codebook available ahead of time to both the transmitter and receiver. For the k th channel use, the receiver calculates an estimate of \mathbf{h} and decides on the appropriate $\mathbf{w}^{(\ell)} \in \mathcal{W}$ to sound the $(k+1)$ th channel use. The index of this sounding vector is fed back to the transmitter, as described in Fig. 1. Note the feedback link is utilized after every channel use during training, which differs from previous channel estimation schemes, including [4]. This creates considerable feedback overhead per coherence time, which is the tradeoff for fewer pilots.

Assume the channel direction vector belongs to a discrete set \mathcal{H} . In practice, \mathcal{H} approximates the true channel subspace. The maximum likelihood channel estimator is given as

$$\begin{aligned} \hat{\mathbf{h}}^{(k)} &= \arg \max_{\mathbf{h} \in \mathcal{H}} p(\mathbf{y}_k | \mathbf{h}, \mathbf{W}_k, \alpha) \\ &= \arg \min_{\mathbf{h} \in \mathcal{H}} \|\mathbf{y}_k - \alpha\sqrt{\rho}\mathbf{W}_k^*\mathbf{h}\|^2 \end{aligned}$$

Given this channel direction estimator, after K channel uses of training the beamforming vector for data transmission is set to $\mathbf{f} = \hat{\mathbf{h}}^{(K)}$.

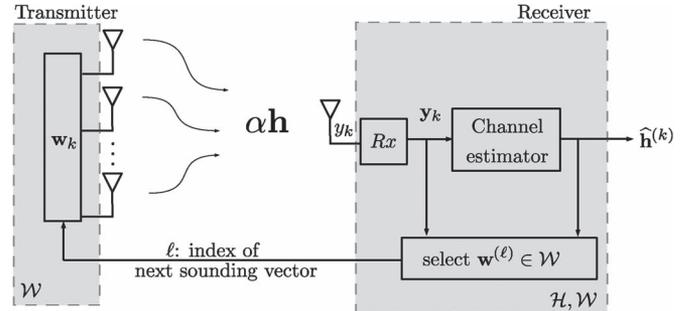


Fig. 1. Block diagram illustrating the proposed training scheme. The receiver selects the sounding vector from the codebook \mathcal{W} for the next channel use and feeds back the codeword index to the transmitter.

In addition to estimating the channel direction, the channel gain term α must also be estimated. We treat α as a nuisance parameter and resort to composite estimation techniques. A generalized maximum likelihood estimator replaces the true channel gain by its maximum likelihood estimate,

$$\hat{\alpha}_{\mathbf{h}} = \arg \max_{\alpha \in \mathbb{C}} p(\mathbf{y}_k | \mathbf{h}, \mathbf{W}_k, \alpha) = \frac{\mathbf{h}^*\mathbf{W}_k\mathbf{y}_k}{\sqrt{\rho}\|\mathbf{W}_k^*\mathbf{h}\|^2}.$$

This result can be shown by minimizing the squared norm $\|\mathbf{y}_k - \alpha\sqrt{\rho}\mathbf{W}_k^*\mathbf{h}\|^2$ over all $\alpha \in \mathbb{C}$. We write the channel magnitude as a complex scalar to incorporate the phase invariance of the channel direction estimate. Beamforming gain is invariant to the phase of the channel direction, e.g., $e^{j\phi}\hat{\mathbf{h}}^{(k)}$ and $\hat{\mathbf{h}}^{(k)}$ give the same beamforming gain. If \mathcal{H} is a discrete set, a given $\mathbf{h} \in \mathcal{H}$ could provide large beamforming gain but give a low likelihood due to its phase. This issue is resolved if we consider α to be complex. The generalized ML channel estimator is given by substituting $\hat{\alpha}_{\mathbf{h}}$ into $p(\mathbf{y}_k | \mathbf{h}, \mathbf{W}_k, \alpha)$

$$\hat{\mathbf{h}}^{(k)} = \arg \max_{\mathbf{h} \in \mathcal{H}} \frac{|\mathbf{y}_k^*\mathbf{W}_k^*\mathbf{h}|^2}{\|\mathbf{W}_k^*\mathbf{h}\|^2} = \arg \max_{\mathbf{h} \in \mathcal{H}} d_k(\mathbf{h}) \quad (2)$$

where $d_k(\mathbf{h}) = |\mathbf{y}_k^*\mathbf{W}_k^*\mathbf{h}|^2 / \|\mathbf{W}_k^*\mathbf{h}\|^2$. In essence, this channel estimator finds the channel vector $\mathbf{h} \in \mathcal{H}$ whose projection onto \mathbf{W}_k most closely aligns with the *direction* of \mathbf{y}_k . The normalizing term $\|\mathbf{W}_k^*\mathbf{h}\|^2$ removes any dependence on vector magnitude. If one interprets each $\mathbf{h} \in \mathcal{H}$ as a beam, the channel estimator selects the beam whose projection onto \mathbf{W}_k most closely aligns with \mathbf{y}_k . Beam alignment is performed by properly selecting successive sounding vectors.

III. SOUNDING VECTOR SELECTION

As one might expect, the sounding vectors have direct influence on the performance of the channel estimator in (2). In this section, we address how to select the sounding vector for each channel use. With feedback after every channel use, the sounding vector for the $(k+1)$ th channel use can be chosen as a function of the previous k receive samples, \mathbf{y}_k . Let us define the misalignment event after $k+1$ channel uses as $\mathcal{M}_i = \{\hat{\mathbf{h}}^{(k+1)} \neq \mathbf{h} | \mathbf{y}_k, \mathbf{W}_{k+1}, \mathbf{h} = \mathbf{h}_i, \hat{\alpha}_{\mathbf{h}_i}\}$. This misalignment event occurs when the estimated channel direction $\hat{\mathbf{h}}^{(k+1)}$ is not the true channel direction $\mathbf{h} = \mathbf{h}_i$. Averaging the probability of the misalignment event over all channel directions gives

the misalignment cost function for the $(k+1)$ th channel use evaluated at channel use k

$$\begin{aligned} P_{k+1|k}^{\text{mis}}(\mathbf{w}_{k+1}) &= \sum_{i=1}^N \Pr(\mathcal{M}_i) p_i^{(k)} \\ &= \sum_{i=1}^N \Pr\left(d_{k+1}(\mathbf{h}_i) < \max_{j \neq i} \left[\{d_{k+1}(\mathbf{h}_j)\} \right] \middle| \mathcal{V}_i\right) p_i^{(k)} \quad (3) \end{aligned}$$

where $p_i^{(k)} = p(\mathbf{h}_i | \mathbf{y}_k, \mathbf{W}_k, \hat{\alpha}_{\mathbf{h}_i})$ are the updated priors after k channel uses and $\mathcal{V}_i = \{\mathbf{h} = \mathbf{h}_i, \mathbf{y}_k, \mathbf{W}_k, \hat{\alpha}_{\mathbf{h}_i}\}$. This cost function measures the penalty of choosing the wrong channel direction as a function of \mathbf{w}_{k+1} . The previous k receive samples are known, and the receive sample for the $(k+1)$ th channel use, y_{k+1} , is treated as a random variable as described in Section III-A. In what follows, the misalignment cost function is adopted to develop a sounding vector selection criterion.

A. Binary Channel Codebook

We approximate, in this subsection, the channel using a binary channel codebook, $\mathcal{H} = \{\mathbf{h}_1, \mathbf{h}_2\}$. Despite being an extremely coarse discretization of the channel space, the exact misalignment cost function can be derived for the binary codebook case

$$\begin{aligned} P_{k+1|k}^{\text{mis}}(\mathbf{w}_{k+1}) &= \Pr(d_{k+1}(\mathbf{h}_1) < d_{k+1}(\mathbf{h}_2) | \mathcal{V}_1) p_1^{(k)} \\ &\quad + \Pr(d_{k+1}(\mathbf{h}_2) < d_{k+1}(\mathbf{h}_1) | \mathcal{V}_2) p_2^{(k)} \quad (4) \end{aligned}$$

which is the weighted sum of pairwise error probability (PEP) terms. The $\text{PEP}(\mathbf{h}_2 \rightarrow \mathbf{h}_1 | \mathcal{V}_2)$ is the probability \mathbf{h}_1 is chosen assuming \mathbf{h}_2 is true

$$\begin{aligned} \text{PEP}(\mathbf{h}_2 \rightarrow \mathbf{h}_1 | \mathcal{V}_2) &= \Pr\left(\frac{|\mathbf{y}_{k+1}^* \mathbf{W}_{k+1}^* \mathbf{h}_1|^2}{\|\mathbf{W}_{k+1}^* \mathbf{h}_1\|^2} > \frac{|\mathbf{y}_{k+1}^* \mathbf{W}_{k+1}^* \mathbf{h}_2|^2}{\|\mathbf{W}_{k+1}^* \mathbf{h}_2\|^2} \middle| \mathcal{V}_2\right). \quad (5) \end{aligned}$$

Since $\mathbf{y}_{k+1}^* \mathbf{W}_{k+1}^* \mathbf{h}_i / \|\mathbf{W}_{k+1}^* \mathbf{h}_i\|$ is a Gaussian random variable when conditioned on \mathbf{h}_i , the pairwise error probability is the probability the magnitude of one Gaussian random variable exceeds the magnitude of another. We now calculate the PEP considering $d_{k+1}(\mathbf{h}_1)$ and $d_{k+1}(\mathbf{h}_2)$ are functions of the same scalar noise term, n_{k+1} .

Lemma 1: Consider two Gaussian random variables

$$\begin{aligned} X &= \mu_x + q_x n \sim \mathcal{CN}(\mu_x, |q_x|^2) \\ Y &= \mu_y + q_y n \sim \mathcal{CN}(\mu_y, |q_y|^2) \end{aligned}$$

where μ_x, μ_y, q_x, q_y are all constant complex scalars and $n \sim \mathcal{CN}(0, 1)$. Then

$$\begin{aligned} \Pr(|X|^2 > |Y|^2) &= \begin{cases} 1 - Q_1\left(\left|\frac{\mu_y q_y^* - \mu_x q_x^*}{|q_y|^2 - |q_x|^2}\right|, \left|\frac{\mu_x q_y - q_x \mu_y}{|q_y|^2 - |q_x|^2}\right|\right) & \text{if } |q_y|^2 > |q_x|^2 \\ Q_1\left(\left|\frac{\mu_y q_y^* - \mu_x q_x^*}{|q_y|^2 - |q_x|^2}\right|, \left|\frac{\mu_x q_y - q_x \mu_y}{|q_y|^2 - |q_x|^2}\right|\right) & \text{if } |q_y|^2 < |q_x|^2 \\ \frac{1}{2} \left[1 + \text{erf}\left(\frac{|\mu_y|^2 - |\mu_x|^2}{2\sqrt{2}|\mu_y^* q_y - \mu_x^* q_x|}\right) \right] & \text{if } |q_y|^2 = |q_x|^2 \end{cases} \quad (6) \end{aligned}$$

where $Q_1(\cdot, \cdot)$ is the first-order Marcum Q function [7].

Proof: See Appendix. ■

Assuming \mathbf{h}_2 is true, the pairwise error probability in (5) is calculated using Lemma 1 by setting $\mu_x = (\mathbf{y}_k^* \mathbf{W}_k^* \mathbf{h}_1 + \sqrt{\rho} \hat{\alpha}_{\mathbf{h}_2}^* \mathbf{h}_2^* \mathbf{w}_{k+1} \mathbf{w}_{k+1}^* \mathbf{h}_1) / \|\mathbf{W}_{k+1}^* \mathbf{h}_1\|$, $q_x = \mathbf{w}_{k+1}^* \mathbf{h}_1 / \|\mathbf{W}_{k+1}^* \mathbf{h}_1\|$, $\mu_y = (\mathbf{y}_k^* \mathbf{W}_k^* \mathbf{h}_2 + \sqrt{\rho} \hat{\alpha}_{\mathbf{h}_2}^* |\mathbf{w}_{k+1}^* \mathbf{h}_2|^2) / \|\mathbf{W}_{k+1}^* \mathbf{h}_2\|$, and $q_y = \mathbf{w}_{k+1}^* \mathbf{h}_2 / \|\mathbf{W}_{k+1}^* \mathbf{h}_2\|$.

The sounding vector for the next channel use is chosen as

$$\mathbf{w}_{k+1} = \arg \min_{\mathbf{w}^{(\ell)} \in \mathcal{W}} P_{k+1|k}^{\text{mis}}(\mathbf{w}^{(\ell)}). \quad (7)$$

B. N-ary Channel Codebook

We now extend our scope to the N -ary scenario. Let $\mathcal{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N\}$. The misalignment cost function expression in (3) cannot be simplified in terms of the pairwise error probabilities, as was the case for the binary channel codebook. Using a union bound argument, we can place an upper bound on the misalignment cost function

$$P_{k+1|k}^{\text{mis}}(\mathbf{w}_{k+1}) \leq \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N \text{PEP}(\mathbf{h}_i \rightarrow \mathbf{h}_j | \mathcal{V}_i) p_i^{(k)}. \quad (8)$$

Calculating the misalignment cost function requires evaluating $|\mathcal{H}|(|\mathcal{H}| - 1)$ PEP terms for each sounding vector codeword in \mathcal{W} . For larger codebooks, the complexity of this algorithm becomes significant, as $|\mathcal{H}|(|\mathcal{H}| - 1)|\mathcal{W}|$ PEP terms must be calculated for each channel use.

C. Low Complexity Beam Alignment

In general, the size of the channel codebook and sounding vector codebook scales with the number of transmit antennas to maintain a fixed quantization error. As a result, the complexity to select a sounding vector at each channel use will scale with M . To make the beam alignment algorithm tractable for moderately sized arrays, we approximate (8). To do this, we note that as the number of channel uses increases, the beam alignment algorithm typically prefers a single codeword corresponding to a high $p_i^{(k)}$. Based on this observation, we can simplify the misalignment cost function to only consider the two most likely codewords. Letting $\mathbf{h}_{(1)}$ represent the most likely channel codeword and $\mathbf{h}_{(2)}$ represent the second most likely channel codeword, we can approximate (8) as

$$\begin{aligned} P_{k+1|k}^{\text{mis}}(\mathbf{w}_{k+1}) &\approx \text{PEP}(\mathbf{h}_{(1)} \rightarrow \mathbf{h}_{(2)} | \mathcal{V}_{(1)}) p_{(1)}^{(k)} \\ &\quad + \text{PEP}(\mathbf{h}_{(2)} \rightarrow \mathbf{h}_{(1)} | \mathcal{V}_{(2)}) p_{(2)}^{(k)}. \end{aligned}$$

In this manner, the beam alignment algorithm calculates $2|\mathcal{W}|$ PEP terms, a significant reduction from the union bound which calculates $|\mathcal{H}|(|\mathcal{H}| - 1)|\mathcal{W}|$ PEP terms.

IV. SIMULATIONS

In this section, we present numerical results for shortened training intervals, e.g., $K < M$. Consider a line-of-sight channel for a uniform linear array. Current cellular deployments place basestation antennas on high towers. In many cases, the channel between the base station and the user lacks any rich scattering, making it appropriate to model as a line-of-sight channel. Fig. 2 shows the average beamforming gain

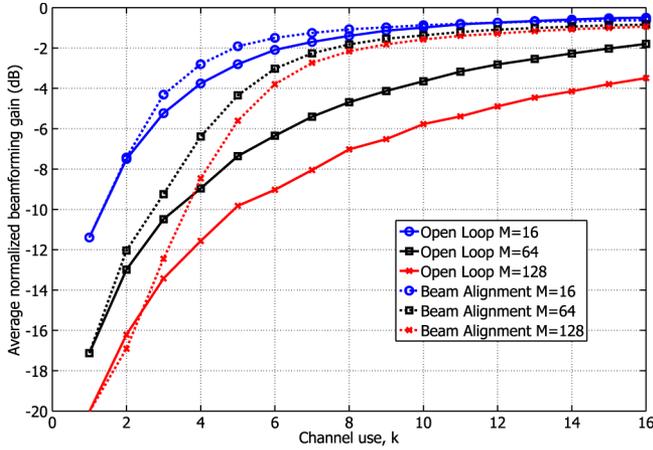


Fig. 2. Average normalized beamforming gain $|\mathbf{h}^* \hat{\mathbf{h}}^{(k)}|^2$ for a line-of-sight channel.

as a function of channel use for SNR = 0 dB. We model the channel to be on the array manifold, $\mathbf{h} = \mathbf{a}(\theta)$, where $[\mathbf{a}(\theta)]_m = (1/\sqrt{M})e^{j\pi(m-1)\sin(\theta)}$. The channel gain term α is randomly chosen as a sum of M independent zero-mean unit-variance complex Gaussian random variables. To construct \mathcal{H} , the angle θ was uniformly quantized over $\theta \in [-\pi, \pi]$. Results were averaged over 5000 iterations and used a channel codebook of size $|\mathcal{H}| = 2M$ and a sounding vector codebook of size $|\mathcal{W}| = M$. Sounding vectors are chosen by minimizing the misalignment cost function using an approximation for the updated priors expression. An open-loop scheme is also presented, with predetermined sounding vectors designed to be the first K columns of the identity matrix. The open-loop scheme then estimates $\alpha\mathbf{h}$ using an MMSE estimator, and quantizes its estimate to the channel codebook \mathcal{H} .

The closed-loop beam alignment algorithm is able to increase the average beamforming gain over orthogonal training for a small number of channel uses. At $k = 5$, there is a 0.4 dB gain for $M = 16$, a 3 dB gain for $M = 64$, and a 5 dB gain for $M = 128$. The beam alignment scheme gives considerable improvement for $K/M \ll 1$, and the improvement over open-loop grows with M .

V. CONCLUSION

This work developed a closed-loop beam alignment scheme which, through the use of feedback after each channel use, sequentially designs sounding vectors to probe the channel in an efficient manner. This scheme provides a practical channel estimation algorithm for massive MIMO, where large scale arrays make optimal training impractical. A generalized maximum likelihood detector was developed to jointly perform the channel estimation and channel quantization. Since beamforming gain is only concerned with the direction of the channel, the detector replaces the channel gain by its maximum likelihood estimate. Sounding vectors are selected to minimize the misalignment cost function, a metric updated with knowledge

of the previous receive samples in a Bayesian framework. The closed-loop beam alignment scheme shows improved beamforming gain over conventional orthogonal training signals, especially for the $K < M$. Additional analysis is required to determine the necessary amount of training as the number of transmit antennas increases.

APPENDIX

Consider two Gaussian random variables, $X = \mu_x + q_x n$ and $Y = \mu_y + q_y n$. It should be mentioned that the expression $\Pr(|X|^2 > |Y|^2)$ in Appendix B of [7] is not applicable when both X and Y are functions of the same scalar noise term $n \sim \mathcal{CN}(0, 1)$, in which for a scalar noise term n , $E[|X - \mu_x|^2]E[|Y - \mu_y|^2] - |E[(X - \mu_x)(Y - \mu_y)^*]|^2 = 0$.

The inequality inside the probability expression, $|\mu_x + q_x n|^2 > |\mu_y + q_y n|^2$, can be simplified by completing the square and rearranging terms for three separate cases (which depend on the magnitudes of q_x and q_y)

$$\Pr(|\mu_x + q_x n|^2 > |\mu_y + q_y n|^2) = \begin{cases} \Pr\left(|Z|^2 < \left|\frac{\mu_x q_y - q_x \mu_y}{|q_y|^2 - |q_x|^2}\right|^2\right) & \text{if } |q_y|^2 > |q_x|^2 \\ \Pr\left(|Z|^2 > \left|\frac{\mu_x q_y - q_x \mu_y}{|q_y|^2 - |q_x|^2}\right|^2\right) & \text{if } |q_y|^2 < |q_x|^2 \\ \Pr\left(V < \frac{|\mu_y|^2 - |\mu_x|^2}{2}\right) & \text{if } |q_y|^2 = |q_x|^2 \end{cases}$$

where $Z = n + ((\mu_y q_y^* - \mu_x q_x^*) / (|q_y|^2 - |q_x|^2))$ is a complex Gaussian random variable and $V = \text{Re}\{\mu_y^* q_y - \mu_x^* q_x\} \text{Re}\{n\} - \text{Im}\{\mu_y^* q_y - \mu_x^* q_x\} \text{Im}\{n\}$ is a real Gaussian random variable. $|Z|^2$ is a noncentral chi-squared random variable with $\Pr(|Z|^2 < z) = 1 - Q_1(|\lambda|, \sqrt{z})$ and V is the sum of two independent real Gaussian random variables, with $\Pr(V < v) = 1/2(1 + \text{erf}(v/\sqrt{2|\mu_y^* q_y - \mu_x^* q_x|^2}))$. We then conclude the result in (6).

REFERENCES

- [1] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [2] J. Zhang, T. Yang, and Z. Chen, "Under-determined training and estimation for distributed transmit beamforming systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1936–1946, Apr. 2013.
- [3] D. Fuhrmann, "Adaptive sensing of target signature with unknown amplitude," in *Proc. 42nd Asilomar Conf.*, Oct. 2008, pp. 218–222.
- [4] D. J. Love, J. Choi, and P. Bidigare, "A closed-loop training approach for massive MIMO beamforming systems," in *Proc. Conf. Inf. Sci. Syst.*, Mar. 2013, pp. 1–5.
- [5] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.
- [6] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [7] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill, 2001.