

Training Signal Design for Correlated Massive MIMO Channel Estimation

Mojtaba Soltanalian*, Mohammad Mahdi Naghsh, Nafiseh Shariati, Petre Stoica, and Babak Hassibi

Abstract—In this paper, we propose a new approach to the design of training sequences that can be used for an accurate estimation of multi-input multi-output (MIMO) channels. The proposed method is particularly instrumental in training sequence designs that deal with three key challenges: (i) arbitrary channel and noise statistics that do not follow specific models, (ii) limitations on the properties of the transmit signals, and (iii) signal design for large-scale or massive antenna arrays. Several numerical examples are provided to examine the proposed method.

Index Terms—Channel estimation, massive MIMO, peak-to-average-power ratio (PAR), signal design, training sequence.

I. INTRODUCTION

Accurate knowledge of multi-input multi-output (MIMO) channel state information (CSI) plays an important role in exploiting the full potential of MIMO communication systems. A practical way to acquire instantaneous CSI of the MIMO channel is training-based channel estimation which relies on transmitting an *a priori* known sequence, namely a *training* or *pilot* sequence, to the receiver, and estimating the instantaneous channel coefficients based on the received signal [1]–[11]. This channel estimation scheme is illustrated in Fig. 1.

Massive MIMO is an emerging technological concept in communications where a large number of antennas is employed at the base stations. Massive MIMO systems have exhibited superior performance with respect to different quality metrics, including capacity, transmit energy efficiency, latency, and robustness to jamming [12]–[16]. Signal design for massive MIMO deals with many streams of signals or, equivalently, many design variables. Such a large number of degrees of freedom paves the way for a better quality of signal design, but at the same time, makes the design computationally expensive. Hence, it is no surprise that a considerable attention has been paid to efficient methods that can handle the signal design complexity arising from such large-scale arrays of antennas.

This work is concerned with an efficient design of training signals for channel estimation in massive MIMO systems. We

This work was supported in part by the European Research Council (ERC) and the Swedish Research Council (VR). M. Soltanalian is with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607. M. M. Naghsh is with the Dept. of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan 84156-83111, Iran. P. Stoica is with the Dept. of Information Technology, Uppsala University, Uppsala, Sweden. N. Shariati is with Ericsson Research, Stockholm, Sweden. She contributed to this work while she was with the Department of Signal Processing, KTH Royal Institute of Technology, Stockholm, Sweden. B. Hassibi is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125.

*Corresponding author— Email: mojtaba.soltanalian@gmail.com

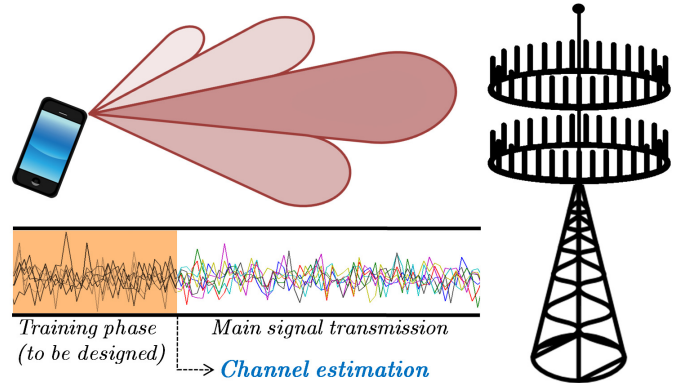


Fig. 1. MIMO training for channel estimation: The optimized training signals are transmitted during the *training phase* in order to estimate the instantaneous channel information based on the received signal and the known input.

TABLE I
NOTATIONS

$\mathbf{x}(k)$	the k^{th} entry of the vector \mathbf{x}
$\ \mathbf{x}\ _n$	the l_n -norm of \mathbf{x} , defined as $(\sum_k \mathbf{x}(k) ^n)^{\frac{1}{n}}$
\mathbf{X}^H	the complex conjugate of a matrix \mathbf{X}
\mathbf{X}^T	the transpose of a matrix \mathbf{X}
$\text{tr}(\mathbf{X})$	the trace of a matrix \mathbf{X}
$\ \mathbf{X}\ _F$	the Frobenius norm of a matrix \mathbf{X}
$\text{vec}(\mathbf{X})$	the vector obtained by column-wise stacking of \mathbf{X}
$\arg(\mathbf{X})$	the phase angle (in radians) of \mathbf{X}
$\exp(\mathbf{X})$	the matrix exponential function defined element-wise as $[\exp(\mathbf{X})]_{k,l} = e^{[\mathbf{X}]_{k,l}}$
$\Re\{\mathbf{X}\}$	the real part of \mathbf{X}
$\mathbf{X} \succ \mathbf{Y}$	$\mathbf{X} - \mathbf{Y}$ is positive definite
\otimes	the Kronecker product
\mathbf{I}_n	the identity matrix of dimension n
\mathbb{C}	the set of complex numbers
$f(n) = \mathcal{O}(n^x)$	$f(n)$ is upper bounded by $c n^x$ for some $0 < c < \infty$
$\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{X})$	the zero-mean random vector \mathbf{x} is distributed according to the circularly symmetric complex Gaussian distribution with covariance \mathbf{X} .

discuss the problem formulation and the contributions of the paper in the following.

A. Problem Formulation

We consider a narrowband block fading point-to-point MIMO communication link equipped with n_T transmit and n_R receive antennas. Let $\mathbf{P} \in \mathbb{C}^{B \times n_T}$ be a matrix whose rows comprise the training symbols at each transmitter where

$B \geq n_T$ [2].¹ The training phase can be described by the equation,

$$\mathbf{Y} = \mathbf{H}\mathbf{P}^T + \mathbf{N}, \quad (1)$$

in which $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ denotes the MIMO channel with $\mathbf{H}(k, l)$ being the gain of the MIMO path between receiver k and transmitter l . We assume $\text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R} \in \mathbb{C}^{n_T n_R \times n_T n_R}$ is the channel covariance matrix. Moreover, $\mathbf{N} \in \mathbb{C}^{n_R \times B}$ represents the noise matrix whose columns consist of (possibly colored) Gaussian noise, i.e., $\text{vec}(\mathbf{N}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{M})$ with $\mathbf{M} \in \mathbb{C}^{B n_R \times B n_R}$ denoting the noise covariance matrix. The received data is denoted by $\mathbf{Y} \in \mathbb{C}^{n_R \times B}$.

The key goal is to design the matrix \mathbf{P} in order to produce an accurate estimate of the channel \mathbf{H} . To accomplish this goal, we consider the minimization of the mean-squared error (MSE) of the channel estimate when the minimum mean-square error (MMSE) estimator is used to estimate the instantaneous channel coefficients ([18], [6], [8]), viz.

$$\text{MSE} = \text{tr} \left\{ (\mathbf{R}^{-1} + (\mathbf{P} \otimes \mathbf{I}_{n_R})^H \mathbf{M}^{-1} (\mathbf{P} \otimes \mathbf{I}_{n_R}))^{-1} \right\}. \quad (2)$$

B. Contributions of this work

In the literature, the optimized training sequence is usually found by minimizing the MSE criterion under a *total power*² constraint. Moreover, in order to make the problem tractable (or to obtain a closed-form solution), the covariance matrices associated with the channel and noise are usually assumed to possess special structures, e.g. to satisfy the Kronecker model or a diagonal matrix structure [1], [3], [5], [6], [8]. The contributions of this paper can be summarized as follows:

- The problem formulation is cast in a rather general form; i.e. no special structure is assumed for the covariance matrices of the channel and noise.
- The proposed optimization approach can handle not only the total training power but also more complicated signal constraints that are typically of interest at the transmit-side of the communication system [15], [19]–[21]. Such constraints include constant-modulus, desired peak-to-average-power ratio (PAR), and quantized-phase alphabet restrictions. To the best of our knowledge, such a constrained training sequence design was not addressed in the literature prior to this work.
- We devise an efficient cyclic method to tackle the resulting non-convex design problem. The low computational complexity of the devised optimization approach makes it a good candidate for usage in massive MIMO scenarios [12]–[14], [22].

¹Note that $B \geq n_T$ is a condition for obtaining a meaningful channel estimate [2]. However, although a larger B would be beneficial for obtaining a more accurate estimate of the channel, it was shown in [17] that using a smaller B leads to a larger capacity of the MIMO channel by leaving more time for data transmission. Therefore, it would be typically advisable to use the minimal amount of training symbols, i.e., $B = n_T$.

²Please see (20) and the related discussions for details.

According to the goals set above, we consider the following general form of the design problem

$$\begin{aligned} \min_{\mathbf{P}} \quad & \text{tr} \left\{ (\mathbf{R}^{-1} + (\mathbf{P} \otimes \mathbf{I}_{n_R})^H \mathbf{M}^{-1} (\mathbf{P} \otimes \mathbf{I}_{n_R}))^{-1} \right\} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega \end{aligned} \quad (3)$$

where Ω denotes the constraints imposed on the training sequences.

The rest of this work is organized as follows. The proposed approach is presented in Section II. Several numerical examples are provided in Section III. Finally, Section IV concludes the paper.

II. TRAINING SEQUENCE OPTIMIZATION

In what follows, we propose an iterative cyclic approach to tackle (3). Let $\tilde{\mathbf{P}} = \mathbf{P} \otimes \mathbf{I}_{n_R}$, and note that using the matrix inversion lemma we have

$$\begin{aligned} \boldsymbol{\theta} &\triangleq \left(\mathbf{R}^{-1} + \tilde{\mathbf{P}}^H \mathbf{M}^{-1} \tilde{\mathbf{P}} \right)^{-1} \\ &= \mathbf{R} - \mathbf{R} \tilde{\mathbf{P}}^H \left(\mathbf{M} + \tilde{\mathbf{P}} \mathbf{R} \tilde{\mathbf{P}}^H \right)^{-1} \tilde{\mathbf{P}} \mathbf{R}. \end{aligned} \quad (4)$$

Define

$$\mathbf{X} \triangleq \begin{pmatrix} \mathbf{R} & \mathbf{R} \tilde{\mathbf{P}}^H \\ \tilde{\mathbf{P}} \mathbf{R} & \mathbf{M} + \tilde{\mathbf{P}} \mathbf{R} \tilde{\mathbf{P}}^H \end{pmatrix}, \quad (5)$$

$$\mathbf{U} \triangleq (\mathbf{I}_{n_T n_R} \quad \mathbf{0}_{n_T n_R \times B n_R})^T, \quad (6)$$

and observe that

$$\mathbf{U}^H \mathbf{X}^{-1} \mathbf{U} = \boldsymbol{\theta}^{-1}. \quad (7)$$

The goal is to minimize $\text{tr}\{\boldsymbol{\theta}\}$ with respect to the training matrix \mathbf{P} . To this end, let $f(\mathbf{V}, \mathbf{P}) \triangleq \text{tr}\{\mathbf{V}^H \mathbf{X} \mathbf{V}\}$ (with $\mathbf{V} \in \mathbb{C}^{(n_T n_R + B n_R) \times n_T n_R}$ being an auxiliary variable), and consider the following minimization problem:

$$\begin{aligned} \min_{\mathbf{V}, \mathbf{P}} \quad & f(\mathbf{V}, \mathbf{P}) \\ \text{s.t.} \quad & \mathbf{V}^H \mathbf{U} = \mathbf{I}_{n_T n_R} \\ & \mathbf{P} \in \Omega. \end{aligned} \quad (8)$$

For fixed \mathbf{P} , the minimizer \mathbf{V} of (8) can be obtained using Result 35 in [23, p. 354] as

$$\begin{aligned} \mathbf{V}_* &= \mathbf{X}^{-1} \mathbf{U} (\mathbf{U}^H \mathbf{X}^{-1} \mathbf{U})^{-1} \\ &= \begin{pmatrix} \boldsymbol{\theta}^{-1} \\ - \left(\mathbf{M} + \tilde{\mathbf{P}} \mathbf{R} \tilde{\mathbf{P}}^H \right)^{-1} \tilde{\mathbf{P}} \mathbf{R} \boldsymbol{\theta}^{-1} \end{pmatrix} \boldsymbol{\theta} \\ &= \begin{pmatrix} \mathbf{I}_{n_T n_R} \\ - \left(\mathbf{M} + \tilde{\mathbf{P}} \mathbf{R} \tilde{\mathbf{P}}^H \right)^{-1} \tilde{\mathbf{P}} \mathbf{R} \end{pmatrix}. \end{aligned} \quad (9)$$

In order to see why the over-parametrization in (8) is useful, note that at the minimizer $\mathbf{V} = \mathbf{V}_*$ of (8),

$$f(\mathbf{V}_*, \mathbf{P}) = \text{tr}\{\boldsymbol{\theta}\}. \quad (10)$$

Hence, each step of a cyclic optimization of (8) with respect to \mathbf{V} and \mathbf{P} will lead to a decrease of $\text{tr}\{\boldsymbol{\theta}\}$. More precisely, by letting $g(\mathbf{P}) = \text{tr}\{\boldsymbol{\theta}\}$ one can observe that

$$\begin{aligned} g(\mathbf{P}^{(k+1)}) &= f(\mathbf{V}^{(k+2)}, \mathbf{P}^{(k+1)}) \\ &\leq f(\mathbf{V}^{(k+1)}, \mathbf{P}^{(k+1)}) \\ &\leq f(\mathbf{V}^{(k+1)}, \mathbf{P}^{(k)}) = g(\mathbf{P}^{(k)}) \end{aligned} \quad (11)$$

where $\mathbf{P}^{(k+1)}$ and $\mathbf{V}^{(k+2)}$ are obtained by fixing $\mathbf{V} = \mathbf{V}^{(k+1)}$ and $\mathbf{P} = \mathbf{P}^{(k+1)}$ in the criterion, respectively, and k denotes the iteration number.

Next, note that the minimization of $f(\mathbf{V}, \mathbf{P})$ with respect to \mathbf{P} boils down to the following quadratic optimization problem:

$$\begin{aligned} \min_{\mathbf{P}} \quad & \text{tr}\{\mathbf{Q}_1 \tilde{\mathbf{P}} \mathbf{Q}_2 \tilde{\mathbf{P}}^H\} + 2 \Re\{\text{tr}\{\mathbf{Q}_3 \tilde{\mathbf{P}}^H\}\} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega \end{aligned} \quad (12)$$

where

$$\begin{cases} \mathbf{Q}_1 = \tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \\ \mathbf{Q}_2 = \mathbf{R} \\ \mathbf{Q}_3 = \tilde{\mathbf{V}} \mathbf{R} \end{cases} \quad (13)$$

and

$$\tilde{\mathbf{V}} = - \left(\mathbf{M} + \tilde{\mathbf{P}}_* \mathbf{R} \tilde{\mathbf{P}}_*^H \right)^{-1} \tilde{\mathbf{P}}_* \mathbf{R} \quad (14)$$

is the lower block of \mathbf{V} (see (9)), in which $\tilde{\mathbf{P}}_*$ denotes the previous known value of $\tilde{\mathbf{P}}$. The latter optimization problem can be recast as

$$\begin{aligned} \min_{\tilde{\mathbf{s}}} \quad & \tilde{\mathbf{s}}^H (\mathbf{Q}_2 \otimes \mathbf{Q}_1) \tilde{\mathbf{s}} + 2 \Re\{\tilde{\mathbf{s}}^H \text{vec}(\mathbf{Q}_3)\} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega, \\ & \tilde{\mathbf{s}} = \text{vec}(\tilde{\mathbf{P}}) = \text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R}), \end{aligned} \quad (15)$$

or equivalently,

$$\begin{aligned} \min_{\tilde{\mathbf{s}}} \quad & \begin{pmatrix} \tilde{\mathbf{s}} \\ 1 \end{pmatrix}^H \begin{pmatrix} \mathbf{Q}_2 \otimes \mathbf{Q}_1 & \text{vec}(\mathbf{Q}_3) \\ \text{vec}^H(\mathbf{Q}_3) & 0 \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{s}} \\ 1 \end{pmatrix} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega, \\ & \tilde{\mathbf{s}} = \text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R}). \end{aligned} \quad (16)$$

Next we show that one can significantly reduce the dimension of the above optimization problem thanks to the structured sparsity of $\tilde{\mathbf{s}}$ (proving Lemma 1, as well as Lemma 2 below, is straightforward):

Lemma 1. *The vector $\tilde{\mathbf{s}} = \text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R})$ is a sparse vector whose non-zero values occur at locations*

$$k_2 B n_R^2 + k_1 B n_R + k_0 n_R + k_1 + 1, \quad (17)$$

for which

$$\begin{aligned} k_0 &\in \{0, 1, \dots, B-1\}, \\ k_1 &\in \{0, 1, \dots, n_R-1\}, \\ k_2 &\in \{0, 1, \dots, n_T-1\}. \end{aligned} \quad (18)$$

To exploit the sparsity of $\tilde{\mathbf{s}}$, let $\mathbf{J} \in \mathbb{C}^{B n_T n_R \times B n_T n_R^2}$ comprise the rows of $\mathbf{I}_{B n_T n_R^2}$ that are indexed by (17). Then (16) can be reformulated as

$$\begin{aligned} \min_{\mathbf{s}} \quad & \begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix}^H \begin{pmatrix} \mathbf{J}(\mathbf{Q}_2 \otimes \mathbf{Q}_1) \mathbf{J}^T & \mathbf{J}(\text{vec}(\mathbf{Q}_3)) \\ \text{vec}^H(\mathbf{Q}_3) \mathbf{J}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega, \\ & \mathbf{s} = \mathbf{J} \tilde{\mathbf{s}} = \mathbf{J}(\text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R})). \end{aligned} \quad (19)$$

In the following, we constrain all the columns of \mathbf{P} to have a fixed l_2 -norm equal to \sqrt{B} , which characterizes a fixed-energy scenario at each transmit antenna. Note that, in the previous literature, usually the collective energy of the antennas is assumed to be upper bounded, viz.

$$\text{tr}\{\mathbf{P} \mathbf{P}^H\} \leq B n_T. \quad (20)$$

However, the bounded-energy constraints such as (20) usually are satisfied at equality (turning the bounded-energy constraint to the fixed-energy alternative) due to the fact that the communication system employs all the energy resource to achieve a better signal transmission quality. Moreover, per-antenna energy constraints are practically more useful compared to the collective energy constraints, as the latter might cause an uneven (and sometimes harmful) distribution of power over the antennas [19]. In fact, the per-antenna energy constraint is typically relaxed to a form similar to (20) solely to make the problems convex, or more tractable. A direct consequence of the fixed-energy assumption on the columns of \mathbf{P} is that the vector \mathbf{s} (as well as $\tilde{\mathbf{s}}$) contains a given energy, as (similar to $\tilde{\mathbf{s}}$) the vector \mathbf{s} includes exactly n_R copies of each and every entry of \mathbf{P} . This fact is elaborated on further in Lemma 2 which identifies the locations in $\tilde{\mathbf{s}}$ associated with the entries of \mathbf{P} . Lemma 2 will also be essential to tackle the quadratic program arising from the over-parametrization in (8) and its equivalent forms discussed herein.

Lemma 2. *In the vector $\tilde{\mathbf{s}} = \text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R})$, the entry $\mathbf{P}(k, l)$ appears exactly n_R times at locations*

$$(l-1)B n_R^2 + t B n_R + (k-1)n_R + t + 1, \quad (21)$$

for $t \in \{0, 1, \dots, n_R-1\}$.

Let

$$\mathbf{Q} \triangleq \begin{pmatrix} \mathbf{J}(\mathbf{Q}_2 \otimes \mathbf{Q}_1) \mathbf{J}^T & \mathbf{J}(\text{vec}(\mathbf{Q}_3)) \\ \text{vec}^H(\mathbf{Q}_3) \mathbf{J}^T & 0 \end{pmatrix} \quad (22)$$

and also let $\hat{\mathbf{Q}} = \lambda \mathbf{I} - \mathbf{Q} \succ \mathbf{0}$, where λ is a real number larger than the maximum eigenvalue of \mathbf{Q} . Due to the fixed energy of \mathbf{s} , (19) can be tackled via considering the following equivalent optimization problem:

$$\begin{aligned} \max_{\mathbf{s}} \quad & \hat{\mathbf{s}}^H \hat{\mathbf{Q}} \hat{\mathbf{s}} \\ \text{s.t.} \quad & \mathbf{P} \in \Omega, \\ & \mathbf{s} = \mathbf{J}(\text{vec}(\mathbf{P} \otimes \mathbf{I}_{n_R})), \\ & \hat{\mathbf{s}} = \begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix}. \end{aligned} \quad (23)$$

TABLE II
CYCLIC ALGORITHM FOR CONSTRAINED TRAINING SIGNAL DESIGN

Step 0: Initialize \mathbf{P} using a random matrix in Ω .
Step 1: Compute the minimizer \mathbf{V} of (8) using (9).
Step 2: Update the current design of \mathbf{P} using the power method-like iterations defined in (27), and the appropriate matrix solutions below (27), μ times (or until convergence).
Step 3: Repeat steps 1 and 2 until a stop criterion is satisfied, e.g. $|\text{MSE}^{(k+1)} - \text{MSE}^{(k)}| < \epsilon$ for some given $\epsilon > 0$, where k denotes the outer loop iteration number.

Now let $\widehat{\Omega}$ denote the constraints on $\widehat{\mathbf{s}}$ imposed in (23). Using this notation, we can rewrite (23) as

$$\max_{\widehat{\mathbf{s}} \in \widehat{\Omega}} \widehat{\mathbf{s}}^H \widehat{\mathbf{Q}} \widehat{\mathbf{s}} \quad (24)$$

The above optimization problem is NP-hard in general, particularly so when \mathbf{P} belongs to the unimodular, or q -ary matrix sets [24], [25]. However, an increasing sequence of values of the objective function in (24), and at the same time, a monotonically decreasing sequence of values for (8), can be obtained by an extension of the power method-like iterations originally developed in [25] and [26]. Namely, starting from a current design of $\widehat{\mathbf{s}}$, say $\widehat{\mathbf{s}}^{(0)}$ (or equivalently a current design of the training matrix i.e. $\mathbf{P}^{(0)}$), an increasing objective of (24) can be obtained by updating $\widehat{\mathbf{s}}$ using the nearest-vector problem:

$$\min_{\widehat{\mathbf{s}}^{(h+1)} \in \widehat{\Omega}} \left\| \widehat{\mathbf{s}}^{(h+1)} - \widehat{\mathbf{Q}} \widehat{\mathbf{s}}^{(h)} \right\|_2 \quad (25)$$

where h denotes the inner-loop iteration number. We note that, to obtain an increasing objective of (24), and a decreasing (8), one does not necessarily need to employ (25) until convergence; indeed, (25) can be used for an arbitrary number of times, say μ , or until convergence. We refer the interested reader to find proofs and further details on the properties of power method-like iterations in [25] and [26].

Inspired by Lemma 2, we let $\{p'_{k,l}(t)\}_{t=1}^{n_R}$ denote the entries of $\widehat{\mathbf{Q}} \widehat{\mathbf{s}}^{(h)}$ occurring exactly in the same locations as of $\mathbf{P}^{(h+1)}(k, l)$ in $\widehat{\mathbf{s}}^{(h+1)}$. Then (25) can be tackled by minimizing the criterion

$$\begin{aligned} & \sum_{k=1}^B \sum_{l=1}^{n_T} \sum_{t=1}^{n_R} \left| \mathbf{P}^{(h+1)}(k, l) - p'_{k,l}(t) \right|^2 \quad (26) \\ &= \text{const}_1 - 2\Re \left\{ \sum_{k=1}^B \sum_{l=1}^{n_T} \mathbf{P}^{*(h+1)}(k, l) \left(\sum_{t=1}^{n_R} p'_{k,l}(t) \right) \right\} \\ &= \text{const}_2 + \sum_{k=1}^B \sum_{l=1}^{n_T} \left| \mathbf{P}^{(h+1)}(k, l) - \sum_{t=1}^{n_R} p'_{k,l}(t) \right|^2. \end{aligned}$$

Consequently, (25) is equivalent to the nearest-matrix problem:

$$\min_{\mathbf{P}^{(h+1)} \in \Omega} \left\| \mathbf{P}^{(h+1)} - \mathbf{P}_{\Sigma}^{(h)} \right\|_F \quad (27)$$

where $\mathbf{P}_{\Sigma}^{(h)}(k, l) = \sum_{t=1}^{n_R} p'_{k,l}(t)$ for all k, l , as given in (26). For obtaining specific solutions to the nearest training matrix problem above, we consider various signal constraints Ω .

- For $\Omega =$ set of matrices satisfying the fixed-energy constraint:

$$\mathbf{P}^{(h+1)} = \Gamma \left(\mathbf{P}_{\Sigma}^{(h)} \right) \quad (28)$$

where the operator $\Gamma(\cdot)$ scales the columns of the matrix argument to achieve an l_2 -norm of \sqrt{B} for each column.

- For $\Omega =$ set of unimodular matrices:

$$\mathbf{P}^{(h+1)} = \exp \left(j \arg \left(\mathbf{P}_{\Sigma}^{(h)} \right) \right). \quad (29)$$

- For $\Omega =$ set of matrices with discrete-phase entries (from $\{2k\pi/q : k = 0, 1, \dots, q-1\}$):

$$\mathbf{P}^{(h+1)} = \exp \left(j \Phi \left(\arg \left(\mathbf{P}_{\Sigma}^{(h)} \right) \right) \right) \quad (30)$$

where the operator $\Phi(\cdot)$ yields the closest q -ary phase matrix with entries from the set $\{2k\pi/q : k = 0, 1, \dots, q-1\}$.

- For $\Omega =$ set of matrices satisfying a PAR constraint: Such a nearest-matrix problem can be solved efficiently via a recursive algorithm introduced in [27].

The proposed method is summarized in Table II. The algorithm has a per-iteration complexity of $\mathcal{O}(\max\{(Bn_R)^{2.38}, Bn_R(n_T n_R)^2, \mu(Bn_T n_R)^2\})$, where the first two terms account for the computation of $\widehat{\mathbf{V}}$, and the last term accounts for the complexity associated with μ power method-like iterations that are performed within each outer-loop iteration of the method. Note that the suggested algorithm yields a monotonically decreasing MSE criterion, and is guaranteed to converge due to the non-negativity (and thus lower boundedness) of the MSE. The next section provides numerical examples that investigate further the computational efficiency of the method.

Remark (Real-Time Applications): The ideas discussed herein can be employed in cognitive real-time scenarios, e.g. for wireless channels in which the channel characteristics are time-varying particularly due to relative motions of users/scatterers. The time intervals for which the channels can be assumed to be time-invariant depend on the Doppler spreads of the channels (see e.g. [28], [29] for details). Note that if the changes in \mathbf{R} and \mathbf{M} are small, then the proposed optimization method can be used (by starting from the previous estimate of \mathbf{P}) to search for the optimal \mathbf{P} under the new channel/noise condition. Such an approach would typically lead to a quick convergence of the method to quality solutions. ■

III. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed method with respect to the MSE metric, as well as computation time, using different scenarios.

A. Simulation Settings

To generate covariance matrices related to a spatially correlated scenario, we exploit the exponential model which is an appropriate model whenever a control over correlation is required [31]. For the sake of comparison with the existing

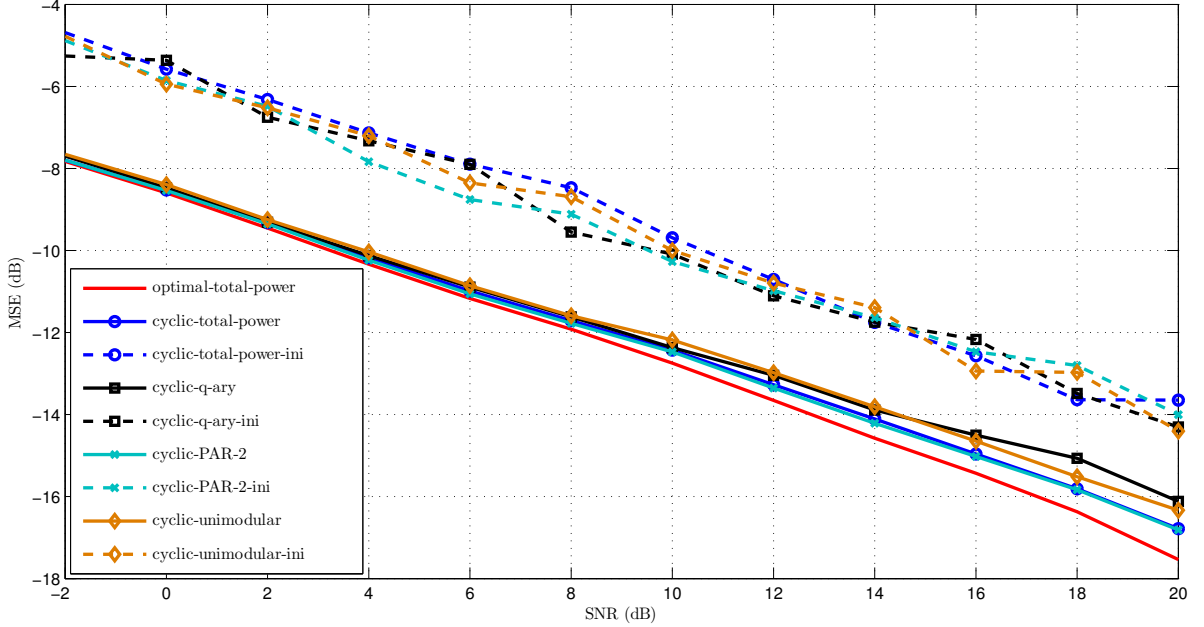


Fig. 2. MSE comparison of different schemes for a 4×4 MIMO channel where $B = 4$ and $\gamma = 16$. All the curves labeled as *cyclic(.)* represent the MSE values obtained using the proposed algorithm whereas the *optimal-total-power* curve shows the optimal performance corresponding to the design in [6]. In order to demonstrate the improvement of MSE values through the iterations, the initial MSE values are also shown using the labels *cyclic(.)-ini*.

methods in the literature, we assume that both the channel matrix \mathbf{R} and the noise matrix \mathbf{M} follow the Kronecker model which is a broadly used simplifying assumption in the MIMO literature [32]. Interestingly, closed-form solutions for the optimal training design problem with a finite collective energy (also known as *total-power* constraint) have been derived in the literature under certain statistical assumptions; see e.g. [6]. Among these assumptions, the Kronecker structure for the channel and noise covariance matrices is one of the essential conditions to be met in order to make the analysis tractable. The exploited model can be described as follows. For a generic covariance matrix \mathbf{C} , we let $[\mathbf{C}]_{k,l} = \rho^{l-k}$ for $k \leq l$, and $[\mathbf{C}]_{k,l} = [\mathbf{C}]_{l,k}^*$ for $l < k$, with $|\rho| < 1$ denoting the correlation coefficient. Note that a larger ρ represents a stronger correlation among the antennas in the array [31]. Specifically, we assume $\mathbf{R} = (\mathbf{R}_T^T \otimes \mathbf{R}_R)$, where $\rho_{rt} = 0.9e^{-j\theta_{rt}}$ and $\rho_{rr} = 0.65e^{-j\theta_{rr}}$ are used to generate covariance matrices \mathbf{R}_T and \mathbf{R}_R at the transmit side and the receive side, respectively. Moreover, we let $\mathbf{M} = (\mathbf{M}_T^T \otimes \mathbf{M}_R)$ where $\mathbf{M}_R = \mathbf{R}_R$, and $\rho_{mt} = 0.8e^{-j\theta_{mt}}$ is the correlation coefficient used to generate the temporal covariance matrix \mathbf{M}_T . The phase arguments $(\theta_{rt}, \theta_{rr}, \theta_{mt})$ appearing above were chosen randomly as $(0.8349\pi, 0.4289\pi, 0.5361\pi)$.

We also normalize \mathbf{R} and \mathbf{M} such that $\text{tr}\{\mathbf{R}\} = 1$ and $\text{tr}\{\mathbf{M}\} = \kappa$, and define the training signal-to-noise ratio (SNR) as $\text{SNR} \triangleq \gamma/\kappa$, where $\gamma = \|\mathbf{P}\|_F^2$ denotes the *total training energy*. Then for a given γ , different values of κ realize different SNR values. We consider $\gamma = Bn_T$, and set the stop threshold of the outer-loop iterations in Table II as $\epsilon = 10^{-5}$.

B. MSE Metric

We begin by using the proposed cyclic method to design training sequences with various signal constraints, namely, total-power (per antenna), unimodularity, quantized-phase (with $q = 64$), and low-PAR (viz. $\text{PAR} \leq 2$). We consider a 4×4 MIMO channel with $B = 4$, and show the performance of the suggested algorithm using the MSE as the figure of merit. As discussed earlier, the quality assessment of the training signals with respect to the MSE metric is accomplished by considering a statistical scenario for which a closed-form optimal solution exists. In this case, we use the closed-form solution of [6] for total-power constraint as a lower bound for comparison. The results are shown in Fig. 2. For each specific signal constraint, we have used the proposed method 30 times (using different random initializations) and have reported the average of the obtained MSE values along with the average of the associated MSE values at initialization. It can be observed from Fig. 2 that the proposed method performs nearly optimally in all cases. It is also interesting to note that the ultimate MSE values for different signal constraints appear to be very similar although some constraints are more strict than the others. This phenomenon has also been observed in some recent publications such as [25], [33] and [34]. The minor differences between the MSE values achieved by the cyclic method and that of [6] can be explained by the fact that the solution from [6] yields the MSE at the global optimum for total-power energy constraint which is more relaxed than per-antenna energy constraints taken into account by the cyclic method.

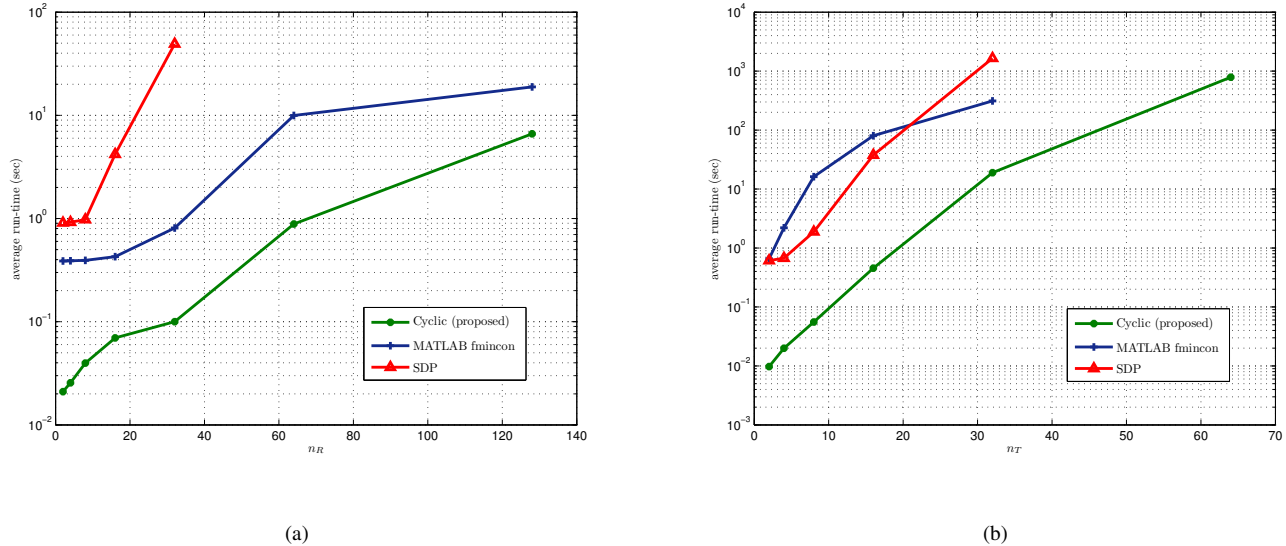


Fig. 4. Comparison of computation times required by the proposed cyclic method, the general purpose MATLAB optimization function `fmincon`, and the SDP-based approach of [8] for training signal design. The run-time values are averaged over 10 trials with different random initializations.

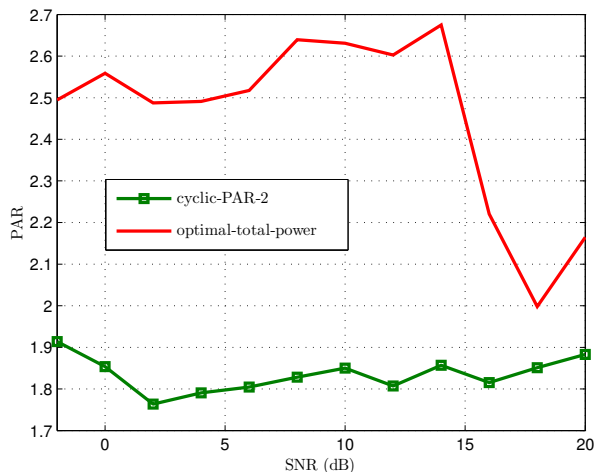


Fig. 3. Comparison of the PAR values obtained from the proposed cyclic method in Fig. 2 and those of [6].

Next, we focus on the PAR constrained example above to illustrate how one of our main goals, i.e. satisfying the signal constraints, is achieved by comparing the resulting PAR values obtained by the proposed algorithm with that of the sequences suggested in [6]. Fig. 3 shows that the proposed cyclic method not only provides MSE values close to the optimal, but it also keeps the PAR values below the specified threshold level, i.e. $\text{PAR} \leq 2$. On the other hand, there is no control over the PAR values associated with the optimal training sequences in [6].

C. Computational Efficiency

Finally, we investigate the computational efficiency of the proposed method. To this end, the required computation

time of the proposed method is compared with that of the general purpose MATLAB function `fmincon` as well as the semidefinite programming (SDP)-based approach of [8] in the *total-power* constraint case. Note that the computational method of [8] can be used only for identity M ; hence, we set M to identity in this example to perform the comparison. In the first experiment, we set the number of transmitters n_T to a fix number 2, and let n_R belong to the set $\{2, 4, 8, 16, 32, 64, 128\}$. In the second, we fix $n_R = 2$ and choose n_T from $\{2, 4, 8, 16, 32, 64\}$. We set $B = n_T$ and compute the average of the required computation time for 10 different trials (with random initializations). The results are shown in Fig. 4(a) and Fig. 4(b), respectively. Note that the curves associated with the SDP-based approach of [8] and MATLAB `fmincon` are truncated at certain values of n_R or n_T due to prohibitive computational burden associated with these methods. The results leading to Fig. 4 were obtained using a standard PC with Intel Core i5 CPU 760 @2.80GHz, and 8GB memory.

It is evident from Fig. 4 that, in both cases, the proposed method yields significant improvements in the computation speed compared to its counterparts. However, the case with $n_T = 2$ and variable n_R represents a more practical scenario for massive MIMO configurations [35]. This is due to the fact that, in order to keep track of the CSI, the training signals (designed at the base station) will usually be transmitted by the user—which typically has a small number of antennas (leading to a reduced training overhead). Upon receiving the training signal, the channel estimation occurs at the base station. Then, thanks to channel reciprocity, both user and the base station can use the channel estimate while communicating with each other. In light of the latter remark, it is also interesting to observe that the design of training sequences appears to be easier in the case with $n_T = 2$ and variable n_R . This is

presumably due to a smaller number of free variables (i.e. B_{NT}) in this case. The computational complexity results illustrate the applicability of the proposed method to currently available prototypes of massive MIMO; see e.g. Argos [36] with 64 antennas at the base station.

IV. CONCLUSIONS

The problem of designing training signals for correlated MIMO channel estimation was considered, and a cyclic method based on a novel over-parametrization of the original MSE minimization problem was introduced. The proposed approach can be used for arbitrary channel and noise covariance matrices, and moreover, for the design of transmit signals in constrained cases such as with transmitter limitations or specific communication schemes. It was shown that the suggested approach can be implemented efficiently from a computational point of view—a fact that paves the way for the method to be employed in massive MIMO scenarios.

REFERENCES

- [1] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.
- [2] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [3] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 884–893, 2006.
- [4] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 453–466, March 2005.
- [5] Y. Liu, T. Wong, and W. Hager, "Training signal design for estimation of correlated MIMO channels with colored interference," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1486–1497, Apr. 2007.
- [6] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1807–1820, March 2010.
- [7] N. Shariati and M. Bengtsson, "Robust training sequence design for spatially correlated MIMO channels and arbitrary colored disturbance," in *IEEE 22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*. Toronto, ON, Canada: IEEE, 2011, pp. 1939–1943.
- [8] N. Shariati, J. Wang, and M. Bengtsson, "Robust training sequence design for correlated MIMO channel estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 107–120, Jan. 2014.
- [9] T.-H. Chang, W.-C. Chiang, Y. Hong, and C.-Y. Chi, "Training sequence design for discriminatory channel estimation in wireless MIMO systems," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6223–6237, December 2010.
- [10] H. D. Tuan, H. H. Kha, H. H. Nguyen, and V.-J. Luong, "Optimized training sequences for spatially correlated MIMO-OFDM," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2768–2778, 2010.
- [11] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 787–801, 2014.
- [12] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [13] L. Lu, G. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, October 2014.
- [14] 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 Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
- [15] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, November 2014.
- [16] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.
- [17] B. Hassibi and B. Hochwald, "Optimal training in space-time systems," in *Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, vol. 1. IEEE, 2000, pp. 743–747.
- [18] S. M. Kay, "Fundamentals of statistical signal processing: Estimation theory," Upper Saddle River, NJ, USA, 1993.
- [19] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 3, pp. 1059–1071, March 2013.
- [20] C. Studer and E. G. Larsson, "PAR-aware large-scale multi-user MIMO-OFDM downlink," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 303–313, February 2013.
- [21] M. Soltanalian, *Signal Design for Active Sensing and Communications*. Uppsala, Sweden: Uppsala Dissertations from the Faculty of Science and Technology, Acta Universitatis Upsaliensis, 2014.
- [22] H. Ngo, M. Matthaiou, and E. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, in press, 2014.
- [23] P. Stoica and R. Moses, *Spectral Analysis of Signals*. New Jersey: Prentice Hall, 2005. [Online]. Available: <http://user.it.uu.se/~ps/ref.html>
- [24] S. Zhang and Y. Huang, "Complex quadratic optimization and semidefinite programming," *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 871–890, 2006.
- [25] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1221–1234, March 2014.
- [26] M. Soltanalian, B. Tang, J. Li, and P. Stoica, "Joint design of the receive filter and transmit sequence for active sensing," *IEEE Signal Processing Letters*, vol. 20, no. 5, pp. 423–426, 2013.
- [27] J. Tropp, I. Dhillon, R. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 188–209, January 2005.
- [28] T. S. Rappaport, *Wireless communications: principles and practice*, 2nd ed. Prentice Hall, 2002.
- [29] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.
- [30] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2572–2583, 2004.
- [31] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Communications Letters*, vol. 5, no. 9, pp. 369–371, September 2001.
- [32] N. Shariati and M. Bengtsson, "How far from Kronecker can a MIMO channel be? Does it matter?" in *11th European Wireless Conference 2011*. Vienna, Austria: VDE, 2011, pp. 1–7.
- [33] M. M. Naghsh, M. Soltanalian, and M. Modarres-Hashemi, "Radar code design for detection of moving targets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2762–2778, 2014.
- [34] M. M. Naghsh, M. Modarres-Hashemi, S. Shahbazpanahi, M. Soltanalian, and P. Stoica, "Unified optimization framework for multi-static radar code design using information-theoretic criteria," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5401–5416, November 2013.
- [35] H. Q. Ngo, *Massive MIMO: Fundamentals and System Designs*. Linköping University Electronic Press, 2015, vol. 1642.
- [36] C. Shepard, H. Yu, N. Anand, E. Li, T. L. Marzetta, R. Yang, and L. Zhong, "Argos: Practical many-antenna base stations," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 53–64.
- [37] "Massive MIMO info point," <http://massivemimo.eu/>, accessed: 2014-10-31.