Joint Antenna Detection and Bayesian Channel Estimation for Non-Coherent User Terminals

Ema Becirovic, Emil Björnson and Erik G Larsson

Becirovic, E., Björnson, E., Larsson, E. G, (2020), Joint Antenna Detection and Bayesian Channel Estimation for Non-Coherent User Terminals, *IEEE Transactions on Wireless Communications*, 19(11), 7081-7096. https://doi.org/10.1109/TWC.2020.3007920

Original publication available at:
https://doi.org/10.1109/TWC.2020.3007920

Joint Antenna Detection and Bayesian Channel Estimation for Non-Coherent User Terminals

Ema Becirovic, *Student Member, IEEE* Emil Björnson, *Senior Member, IEEE* and Erik G. Larsson, *Fellow, IEEE*

***Abstract*—In this paper, we propose a method of improving the channel estimates for non-coherent multi-antenna terminals, which are terminals that cannot control the relative phase between its antenna ports, with channels that can be considered constant over multiple time slots. The terminals have multiple antennas and are free to choose whichever antenna they want to use in each time slot. An unknown phase shift is introduced in each time slot as we cannot guarantee that the terminals are phase coherent across time slots. We compare three different clustering techniques that we use to detect the active antenna. We also compare a set of different statistical and heuristic estimators for the channels and the phase shifts. We evaluate the methods by using correlated Rayleigh fading and three different bounds on the uplink capacity. The accuracy of the capacity bounds are verified with bit-error-rate simulations. With our proposed methods we can have an SNR improvement of approximately 2 dB at 1 bit/s/Hz.**

***Index Terms*—5G mobile communication, antenna detection, channel estimation, massive MIMO**

I. Introduction

CHANNEL estimation is crucial in multiple antenna wireless communication systems for the signals to be directed correctly. Reciprocity-based precoding is superior to feedback-based precoding, and is also the canonical operation mode of TDD massive MIMO [2, Sec. 1.4]. For reciprocity-based precoding to work effectively, and for accurate demodulation of the uplink data, the quality of the uplink channel estimates is important [2]. Since the uplink power can be two orders-of-magnitude lower than the downlink power, the uplink estimation quality can be poor even when the downlink signal-to-noise ratio (SNR) is good.

The slot duration is equal to the minimum coherence time that the system supports. Many users, especially stationary indoor users, will have a substantially longer coherence time. When the channel coherence time exceeds the slot duration, the channel estimates can be improved by averaging (filtering) over multiple slots. Filtering in the frequency domain is also possible, depending on the actual channel coherence bandwidth (delay spread in the time-domain). Several techniques are available for filtering of channel estimates [3]–[6] and also used in state-of-the-art systems.

One of the most common channel-estimate-filtering approaches is to utilize different types of interpolation between estimates [5], [6]. In OFDM-based systems, where the channel is assumed to be changing between symbols, channel estimation can be aided by sending pilot symbols. The pilot symbols are arranged in a pattern across the OFDM-grid, usually a comb pattern or a block pattern. For each of the pilot symbols, simple least-squares channel estimation can be done. Further, the channels corresponding to the symbols in between the pilots need to be estimated. This estimation can be done by interpolation. The interpolators are of different complexity and the simplest ones are the piecewise constant and piecewise linear interpolators. The interpolation can be done, depending on the pilot pattern, either in the time-domain, frequency-domain, or in both.

In 5G, there are two uplink transmission schemes; one is codebook based and the other one is reciprocity based [7, Ch. 7], [8, Ch. 11]. In the codebook based transmission scheme, the choice of precoder is made by the base station which chooses a precoder from a predefined set based on reference signals transmitted by the terminal. A multi-antenna 5G terminal can either have full coherence, partial coherence or be non-coherent. If a terminal has full coherence it can control the relative phase between all of its antenna ports, a terminal with partial coherence can only transmit coherently over pairs of antenna ports, and a non-coherent terminal cannot transmit coherently at all, i.e., it can only use one antenna port at a time. This means that for a non-coherent terminal, the set of precoders that the base station can decide on is limited to using one antenna port at a time. An antenna port is not the same as a physical antenna, but is a terminology used to signify that the receiver cannot distinguish a signal transmitted on a single antenna from a signal transmitted on multiple precoded antennas. That is, a signal transmitted on a single antenna port can be precoded over the physical antennas, to always make use of the entire array aperture and all its power amplifiers.

In the other uplink transmission scheme, reciprocity holds. The terminal transmits one or multiple (per prior agreement with the base station) precoded reference signals, where the choice of precoder(s) is up to the terminal. The base station in turn reports to the terminal which of these reference signals, corresponding to a precoder, that was preferred. The terminal then, in turn, uses the precoder corresponding to this preferred reference signal. In case the terminal transmits only a single reference signal, no feedback is needed from the base station.

A. Contributions

In this paper, we consider non-coherent terminals and reciprocity-based operation, which is one of the important

This work was supported in part by the Swedish Research Council (VR) and in part by ELLIIT. Part of the material in this paper was presented at the IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications, Cannes, France, July 2–5, 2019 [1].

E. Becirovic, E. Björnson, and E. G. Larsson are with the Department of Electrical Engineering (ISY), Linköping University, Linköping, Sweden (e-mail: {ema.becirovic, emil.bjornson, erik.g.larsson}@liu.se).

cases in the 5G standard. We assume the terminal transmits a single reference signal; the choice of precoder is made by the terminal and not known by the base station. As the terminal is non-coherent, this precoder is selected from a finite set of possibilities; the number of possible precoders equals the number of antennas at the terminal. A special case is when each precoder corresponds to transmission (only) through a specific antenna; henceforth, without loss of generality, we will say that the terminal selects an "antenna" instead of a "precoder". Moreover, we assume that in each time slot a phase shift might occur which is due to the fact that the terminals are not required to be phase coherent across time slots which facilitates building energy-efficient hardware, e.g., hardware that is allowed to turn components on and off. The phase shift may also be due to power amplifiers that change the range of operation.

For this scenario, we develop methods for joint estimation of the channel, estimation of the phase shifts, and detection of the active (terminal) antenna, based on pilots received over multiple consecutive slots. Our proposed methods only affect the operation of the base station, not the terminal. Hence, the terminal operates according to the pre-specified standard which does not require it to be phase coherent across slots. We use the well researched and analyzed area of clustering algorithms and propose new algorithms to solve our problem. The main innovation is a new pseudo-metric which takes into account the unknown phase shifts within the clusters. We obtain estimates both for the case of unknown channel statistics (using a maximum-likelihood approach), and for the case of arbitrarily correlated Rayleigh fading with uniformly distributed phase shifts between consecutive slots (using a Bayesian approach). We use three different capacity bounds to evaluate the performance of our proposed methods. We also provide bit-error-rate simulations with modulation and low-density parity-check coding in order to demonstrate the validity of the obtained capacity bounds as proxies for actual link performance in practice.

This paper is a comprehensive extension of our previous conference paper [1]. The main new elements are the derivation of Bayesian estimators that exploit priors on the fading distribution, and the link simulations to validate the capacity bounds.

B. Notation

We use italicized, x, bold lowercase, $\boldsymbol{x}$, bold uppercase, $\boldsymbol{X}$, and calligraphic, $\mathcal{X}$, characters to denote scalars, vectors, matrices and sets respectively. $\boldsymbol{I}_M$ is used to denote the $M \times M$ identity matrix. $\mathbf{0}$ denotes an all-zero vector with appropriate (clear from the context) length. $\boldsymbol{e}_k$ is the k:th unit vector, i.e., a vector with a one in the k:th position and zeros otherwise. $(\cdot)^*$, $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote the conjugate, the transpose, and the conjugate transpose, respectively. We use $\{x_i\}_{i=1}^N$ as shorthand for $x_i, i = 1, \ldots, N$. We use $\hat{x}$ to denote the estimate of x. The imaginary unit is denoted by j. $\mathcal{CN}(\mathbf{0}, \boldsymbol{R})$ is denoting the circularly symmetric complex Gaussian distribution with covariance matrix $\boldsymbol{R}$. $\mathcal{U}(a, b)$ is denoting the uniform distribution between a and b. The expected value and variance are denoted by $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$, respectively. The determinant of a matrix $\boldsymbol{X}$ is denoted by $|\boldsymbol{X}|$.

II. System Model

We consider a single-cell system with M antennas at the base station and one user with K antennas. The channel between the user and the base station is considered constant over many time slots. The base station wants to exploit this property and use the received pilot signal from many time slots to estimate the channel to the user. However, the user is assumed to be non-coherent, as described in the introduction, and is therefore free to use any of its antennas but can only use one at a time. The choice of antenna is unknown to the base station as the decision is made at the terminal. We denote the selected antenna at time t with $S(t) \in \{1, \ldots, K\}$. Let $\mathcal{T}_k$ be the set of time slots where the user selected to use antenna k: $\mathcal{T}_k = \{t : S(t) = k\}$. The union of the sets is equal to the set of all time slots, $\bigcup_{k=1}^K \mathcal{T}_k = \{1, \ldots, T\}$, where T is the number of slots where the channels can be considered constant, i.e., the considered user has a coherence time which is T times longer than the minimum coherence time that the system supports. T is assumed to be known at the base station. The set of time slots of the selected antennas are all disjoint, $\mathcal{T}_k \cap \mathcal{T}_{k'} = \emptyset, k = 1, \ldots, K,\ k' = 1, \ldots, K,\ k \neq k'$, meaning that the user can only activate one antenna at a time. Each time slot consists of τ_{c} symbols, and the user uses codewords that span multiple time slots such that the ergodic capacity is a legitimate performance metric.

At each time slot t, the user first sends a pilot signal $\boldsymbol{\phi}(t) \in \mathbb{C}^{\tau_{\mathrm{p}}}$ of length τ_{p}, $\|\boldsymbol{\phi}(t)\|^2 = 1$, which is known to the base station. The length of the pilot signal, τ_{p}, is assumed to be pre-decided based on, for example, the number of users in the system [9], [10]. The pilot signal is transmitted with an average power of p per symbol. Hence, it is scaled by the square root of the length of the signal, $\sqrt{\tau_{\mathrm{p}}}$. The base station receives

$$\boldsymbol{Y}_{\mathrm{p}}(t) = \sqrt{p\tau_{\mathrm{p}}} e^{\mathsf{j}\,\psi(t)} \boldsymbol{g}_{S(t)} \boldsymbol{\phi}^{\mathsf{H}}(t) + \boldsymbol{W}_{\mathrm{p}}(t),\ t = 1, \ldots, T, \quad (1)$$

where p is the power, $\boldsymbol{g}_k \in \mathbb{C}^M$ is the channel between user antenna k and the base station, and $\psi(t)$ is an unknown phase shift that occurs at time t. Finally, $\boldsymbol{W}_{\mathrm{p}}(t) \in \mathbb{C}^{M \times \tau_{\mathrm{p}}}$ is additive white Gaussian noise with i.i.d. $\mathcal{CN}(0, 1)$ elements. The noise realizations in different time slots are independent. The base station despreads the received signal with the pilot signal as

$$\boldsymbol{Y}_{\mathrm{p}}(t)\boldsymbol{\phi}(t) = \boldsymbol{y}_{\mathrm{p}}(t) = \sqrt{p\tau_{\mathrm{p}}} e^{\mathsf{j}\,\psi(t)} \boldsymbol{g}_{S(t)} + \boldsymbol{w}_{\mathrm{p}}(t),\ t = 1, \ldots, T, \quad (2)$$

where $\boldsymbol{w}_{\mathrm{p}}(t) = \boldsymbol{W}_{\mathrm{p}}(t)\boldsymbol{\phi}(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_M)$. $\boldsymbol{y}_{\mathrm{p}}(t) \in \mathbb{C}^M$ for $t = 1, \ldots, T$ are the signals that the base station uses to estimate the channel between itself and the user.

Furthermore, in each time slot, there is an uplink data phase where we denote the data symbol as $x(t)$, $\mathbb{E}\left\{|x(t)|^2\right\} \leq 1$, in an arbitrary time slot t. When the user sends $x(t)$, the base station receives

$$\boldsymbol{y}_{\mathrm{d}}(t) = \sqrt{p} e^{\mathsf{j}\,\psi(t)} \boldsymbol{g}_{S(t)} x(t) + \boldsymbol{w}_{\mathrm{d}}(t), \quad (3)$$

where $\boldsymbol{w}_{\mathrm{d}}(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_M)$ is additive white Gaussian noise.

TABLE I
THE FOUR CASES OF CHANNEL ESTIMATION OVER MANY SLOTS

	One antenna	Multiple antennas
Without phase shifts	LS/MMSE (6)/(7)	Section V
With phase shifts	Section IV	Section VI

III. Channel Estimation

We utilize that the channel is constant over multiple slots to improve the channel estimates compared to estimating the channel separately in each slot. The channel estimation can be broken down into four different cases by considering either multi-antenna terminals or single-antenna terminals, and the presence or absence of phase shifts between the slots, see Table I.

In the simplest case, when we have a single antenna and no phase shifts between the time slots, channel estimation is relatively easy. The despreaded signal (2) is then

$$\boldsymbol{y}_{\mathrm{p}}(t) = \sqrt{p\tau_{\mathrm{p}}}\boldsymbol{g} + \boldsymbol{w}_{\mathrm{p}}(t),\ t = 1, \ldots, T. \tag{4}$$

We can take the mean over the channel estimates to obtain a sufficient statistic

$$\boldsymbol{y}' = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_{\mathrm{p}}(t) = \sqrt{p\tau_{\mathrm{p}}}\boldsymbol{g} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{w}_{\mathrm{p}}(t)}_{\boldsymbol{w}'}, \tag{5}$$

where $\boldsymbol{w}' \sim \mathcal{CN}\left(\boldsymbol{0}, \frac{1}{T}\boldsymbol{I}_M\right)$. The least-squares (LS) estimate of $\boldsymbol{g}$ is then

$$\hat{\boldsymbol{g}} = \frac{1}{\sqrt{p\tau_{\mathrm{p}}}T}\sum_{t=1}^{T}\boldsymbol{y}_{\mathrm{p}}(t). \tag{6}$$

If we have full statistical prior information about the distribution of the channel $\boldsymbol{g}$, we can state the minimum-mean-square-error (MMSE) estimate. As an example, for a correlated Rayleigh fading channel with prior $\boldsymbol{g} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R})$, where $\boldsymbol{R}$ is the correlation matrix, the MMSE estimate is [11, Ch. 10] [12, Thm. 3.1]

$$\hat{\boldsymbol{g}} = \sqrt{p\tau_{\mathrm{p}}}\boldsymbol{R}\left(p\tau_{\mathrm{p}}\boldsymbol{R} + \frac{1}{T}\boldsymbol{I}_M\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_{\mathrm{p}}(t)\right). \tag{7}$$

The other three cases in Table I require some more analysis and are discussed in the sections that follow. In Section IV, the case where the terminal has a single antenna and there are phase shifts between the slots is discussed. Next, in Section V, the case where the terminal has multiple antennas and the terminal is free to switch between them but where there are no phase shifts between the slots is treated. Finally, the combination of multiple antennas at the terminal and phase shifts between the slots is analyzed in Section VI.

IV. Single Antenna Terminal with Phase Shifts

In the single antenna terminal case, the despreaded signal (2) is

$$\boldsymbol{y}_{\mathrm{p}}(t) = \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g} + \boldsymbol{w}_{\mathrm{p}}(t),\ t = 1, \ldots, T. \tag{8}$$

In this case, the variables we need to estimate are the phase shifts, $\psi(t)$, $t = 1, \ldots, T$, and the channel, $\boldsymbol{g}$. The LS estimate is a joint minimization of the phase shifts and the channel, and is defined as

$$\left\{\hat{\boldsymbol{g}}, \left\{\hat{\psi}(t)\right\}_{t=1}^{T}\right\}^{\mathrm{LS}} = \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmin}} \sum_{t=1}^{T}\left\|\boldsymbol{y}_{\mathrm{p}}(t) - \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g}\right\|^2. \tag{9}$$

The maximum-likelihood (ML) estimate can also be used to jointly estimate the phase shifts and the channel. The ML estimate is defined as

$$\left\{\hat{\boldsymbol{g}}, \left\{\hat{\psi}(t)\right\}_{t=1}^{T}\right\}^{\mathrm{ML}} = \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmax}}\ p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t=1}^{T} \mid \boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}\right) \tag{10}$$

$$= \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmax}} \prod_{t=1}^{T}\frac{1}{\pi^M}\exp\left(-\|\boldsymbol{y}_{\mathrm{p}}(t) - \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g}\|^2\right) \tag{11}$$

$$= \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmin}} \sum_{t=1}^{T}\left\|\boldsymbol{y}_{\mathrm{p}}(t) - \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g}\right\|^2, \tag{12}$$

where in the first step we utilize that the noise is Gaussian and then simplify to get (12). The joint ML estimate is the same as the LS estimate (9). Neither the the LS nor the ML estimate assumes any priors on the channel or the phase shifts. In what follows, we will derive estimates based on statistical knowledge. There are different combinations of these depending on which priors that are available.

If we assume a correlated Rayleigh fading channel, $\boldsymbol{g} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R})$, and uniformly distributed phase shifts, $\psi(t) \sim \mathcal{U}(-\pi, \pi)$, $t = 1, \ldots, T$ which are independent between slots, we can state the maximum-a-posteriori (MAP) estimate and simplify to get an estimator for the channel and the phase shifts as

$$\left\{\hat{\boldsymbol{g}}, \left\{\hat{\psi}(t)\right\}_{t=1}^{T}\right\}^{\mathrm{MAP}} = \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmax}}\ p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t=1}^{T} \mid \boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}\right)p(\boldsymbol{g})p\left(\{\psi(t)\}_{t=1}^{T}\right) \tag{13}$$

$$= \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmax}} \frac{\exp\left(-\sum_{t=1}^{T}\|\boldsymbol{y}_{\mathrm{p}}(t) - \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g}\|^2 - \boldsymbol{g}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{g}\right)}{2^T\pi^{2M+T}|\boldsymbol{R}|} \tag{14}$$

$$= \underset{\boldsymbol{g}, \{\psi(t)\}_{t=1}^{T}}{\operatorname{argmin}}\ \boldsymbol{g}^{\mathsf{H}}\boldsymbol{R}^{-1}\boldsymbol{g} + \sum_{t=1}^{T}\left\|\boldsymbol{y}_{\mathrm{p}}(t) - \sqrt{p\tau_{\mathrm{p}}}e^{\mathrm{j}\,\psi(t)}\boldsymbol{g}\right\|^2. \tag{15}$$

Here, (13) holds for all distributions, while (14) and (15) hold for our assumed priors.

Furthermore, by only assuming the uniform prior distribution of the phase shifts, $\psi(t) \sim \mathcal{U}(-\pi, \pi)$, $t = 1, \ldots, T$, independent between slots, we can state the likelihood distribution only conditioned on the channel by marginalizing out the

phase shifts. The distribution holds for all prior distributions of the channel and is

$$p\left(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \boldsymbol{g}\right) = \int p\left(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \boldsymbol{g}, \{\psi(t)\}_{t=1}^T\right) p(\{\psi(t)\}_{t=1}^T) d\{\psi(t)\}_{t=1}^T \tag{16}$$

$$= \int \frac{\exp\left(-\sum_{t=1}^T \|\boldsymbol{y}_\mathrm{p}(t) - \sqrt{p\tau_\mathrm{p}} e^{\mathsf{j}\,\psi(t)} \boldsymbol{g}\|^2\right)}{\pi^{M+T} 2^T} d\{\psi(t)\}_{t=1}^T \tag{17}$$

$$= \frac{\exp\left(-\sum_{t=1}^T \|\boldsymbol{y}_\mathrm{p}(t)\|^2 - Tp\tau_\mathrm{p}\|\boldsymbol{g}\|^2\right)}{\pi^{M+T} 2^T} \times \prod_{t=1}^T \int_{-\pi}^{\pi} \exp\left(2\Re\left\{\sqrt{p\tau_\mathrm{p}} e^{\mathsf{j}\,\psi(t)} \boldsymbol{y}_\mathrm{p}(t)^\mathsf{H} \boldsymbol{g}\right\}\right) d\psi(t) \tag{18}$$

$$= \frac{\exp\left(-\sum_{t=1}^T \|\boldsymbol{y}_\mathrm{p}(t)\|^2 - Tp\tau_\mathrm{p}\|\boldsymbol{g}\|^2\right)}{\pi^{M}} \times \prod_{t=1}^T \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(2\sqrt{p\tau_\mathrm{p}} |\boldsymbol{y}_\mathrm{p}(t)^\mathsf{H} \boldsymbol{g}| \cos\left(\psi(t)\right)\right) d\psi(t) \tag{19}$$

$$= \frac{\exp\left(-\sum_{t=1}^T \|\boldsymbol{y}_\mathrm{p}(t)\|^2 - Tp\tau_\mathrm{p}\|\boldsymbol{g}\|^2\right)}{\pi^{M}} \times \prod_{t=1}^T \left(I_0(2\sqrt{p\tau_\mathrm{p}} |\boldsymbol{y}_\mathrm{p}^\mathsf{H}(t) \boldsymbol{g}|)\right), \tag{20}$$

where

$$I_0(z) = \frac{1}{\pi} \int_0^\pi e^{z\cos(\theta)} d\theta$$

is the modified Bessel function of the first kind which can be bounded like [13]

$$\frac{e^z}{1+2z} < I_0(z) < \frac{e^z}{\sqrt{1+2z}}. \tag{21}$$

The marginal distribution in (20) can be used to state an ML estimator of the channel:

$$\hat{\boldsymbol{g}}^{\text{ML-}\boldsymbol{g}} = \operatorname*{argmax}_{\boldsymbol{g}} p\left(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \boldsymbol{g}\right) \tag{22}$$

$$= \operatorname*{argmax}_{\boldsymbol{g}} \frac{1}{\pi^M} \exp\left(-\sum_{t=1}^T \|\boldsymbol{y}_\mathrm{p}(t)\|^2 - Tp\tau_\mathrm{p}\|\boldsymbol{g}\|^2\right) \times \prod_{t=1}^T \left(I_0(2\sqrt{p\tau_\mathrm{p}} |\boldsymbol{y}_\mathrm{p}^\mathsf{H}(t) \boldsymbol{g}|)\right) \tag{23}$$

$$= \operatorname*{argmax}_{\boldsymbol{g}} -Tp\tau_\mathrm{p}\|\boldsymbol{g}\|^2 + \sum_{t=1}^T \log\left(I_0\left(2\sqrt{p\tau_\mathrm{p}} |\boldsymbol{y}_\mathrm{p}^\mathsf{H}(t) \boldsymbol{g}|\right)\right). \tag{24}$$

If the channel is correlated Rayleigh fading, $\boldsymbol{g} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R})$, a MAP estimator for the channel is

$$\hat{\boldsymbol{g}}^{\text{MAP-}\boldsymbol{g}} = \operatorname*{argmax}_{\boldsymbol{g}} p\left(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \boldsymbol{g}\right) p(\boldsymbol{g}) \tag{25}$$

$$= \operatorname*{argmax}_{\boldsymbol{g}} -\boldsymbol{g}^\mathsf{H} \left(Tp\tau_\mathrm{p} \boldsymbol{I}_M + \boldsymbol{R}^{-1}\right) \boldsymbol{g} + \sum_{t=1}^T \log\left(I_0\left(2\sqrt{p\tau_\mathrm{p}} |\boldsymbol{y}_\mathrm{p}^\mathsf{H}(t) \boldsymbol{g}|\right)\right). \tag{26}$$

Again, (22) and (25) hold for all prior distributions while (24) holds if we assume the uniformly distributed phase shifts and (26) holds by further assuming a correlated Rayleigh fading channel.

We can also perform the marginalization with respect to the channel distribution. We can do this if we do not know the distribution of the phase shifts. The likelihood distribution conditioned on the phase shifts is

$$p(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \{\psi(t)\}_{t=1}^T) = \int p(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \boldsymbol{g}, \{\psi(t)\}_{t=1}^T) p(\boldsymbol{g}) d\boldsymbol{g} \tag{27}$$

$$= \int \frac{1}{\pi^{2M} |\boldsymbol{R}|} \exp\left(-\|\boldsymbol{Y} - \sqrt{p\tau_\mathrm{p}} \boldsymbol{g} \boldsymbol{v}^\mathsf{T}\|_F^2 - \boldsymbol{g}^\mathsf{H} \boldsymbol{R}^{-1} \boldsymbol{g}\right) d\boldsymbol{g} \tag{28}$$

$$= \int \frac{\exp\left(-\|\boldsymbol{Y}\|_F^2 - (\boldsymbol{g}-\boldsymbol{m})^\mathsf{H} \boldsymbol{A}^{-1} (\boldsymbol{g}-\boldsymbol{m}) + \boldsymbol{m}^\mathsf{H} \boldsymbol{A}^{-1} \boldsymbol{m}\right)}{\pi^{2M} |\boldsymbol{R}|} d\boldsymbol{g} \tag{29}$$

$$= \frac{|\boldsymbol{A}| \pi^M}{\pi^{2M} |\boldsymbol{R}|} \exp\left(-\|\boldsymbol{Y}\|_F^2 + \boldsymbol{m}^\mathsf{H} \boldsymbol{A}^{-1} \boldsymbol{m}\right) \times \underbrace{\left(\frac{1}{|\boldsymbol{A}| \pi^M} \int \exp\left(-(\boldsymbol{g}-\boldsymbol{m})^\mathsf{H} \boldsymbol{A}^{-1} (\boldsymbol{g}-\boldsymbol{m})\right) d\boldsymbol{g}\right)}_{=1} \tag{30}$$

$$= \frac{\exp\left(-\|\boldsymbol{Y}\|_F^2 + p\tau_\mathrm{p} (\boldsymbol{Y}\boldsymbol{v}^*)^\mathsf{H} (Tp\tau_\mathrm{p} \boldsymbol{I} + \boldsymbol{R}^{-1})^{-1} (\boldsymbol{Y}\boldsymbol{v}^*)\right)}{\pi^M |Tp\tau_\mathrm{p} \boldsymbol{R} + \boldsymbol{I}|}, \tag{31}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_\mathrm{p}(1), \ldots, \boldsymbol{y}_\mathrm{p}(t), \ldots, \boldsymbol{y}_\mathrm{p}(T)]$, $\boldsymbol{v} = \left[e^{\mathsf{j}\,\psi(1)}, \ldots, e^{\mathsf{j}\,\psi(t)}, \ldots, e^{\mathsf{j}\,\psi(T)}\right]^\mathsf{T}$, and $\boldsymbol{A} = (Tp\tau_\mathrm{p} \boldsymbol{I} + \boldsymbol{R}^{-1})^{-1}$ and $\boldsymbol{m} = \sqrt{p\tau_\mathrm{p}} \boldsymbol{A} \boldsymbol{Y} \boldsymbol{v}^*$ are helpful variable changes. In this case, given a uniform distribution of the phase shifts, the ML and MAP estimators for the phase shifts are the same:

$$\left\{\left\{\hat{\psi}(t)\right\}_{t=1}^T\right\}^{\text{ML-}\psi} = \left\{\left\{\hat{\psi}(t)\right\}_{t=1}^T\right\}^{\text{MAP-}\psi} = \operatorname*{argmax}_{\{\psi(t)\}_{t=1}^T} p(\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T \mid \{\psi(t)\}_{t=1}^T) \tag{32}$$

$$= \operatorname*{argmax}_{\{\psi(t)\}_{t=1}^T} \boldsymbol{v}^\mathsf{T} \boldsymbol{Y}^\mathsf{H} (Tp\tau_\mathrm{p} \boldsymbol{I} + \boldsymbol{R}^{-1})^{-1} \boldsymbol{Y} \boldsymbol{v}^*. \tag{33}$$

The different channel and phase shift estimators based on statistical knowledge are summarized in Table II. We derived estimators with the correlated Rayleigh fading channel prior, and the independent and uniformly distributed phase shift prior. The correlated Rayleigh fading prior is widely used in communication literature and models practical channels well. The independent and uniform distribution of the phase shifts

TABLE II
THE DIFFERENT CHANNEL AND PHASE SHIFT ESTIMATORS BASED ON THE STATISTICS. THE EQUATION NUMBERS REFERS TO THE CASES WHEN THE CHANNEL IS RAYLEIGH FADING, $\boldsymbol{g} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R})$, AND THE PHASE SHIFTS HAVE UNIFORM DISTRIBUTION $\psi(t) \sim \mathcal{U}(-\pi, \pi), t = 1, \ldots, T$, AND ARE INDEPENDENT.

	ML			MAP		
joint	$\underset{\boldsymbol{g},\{\psi(t)\}}{\operatorname{argmax}}$	$p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \boldsymbol{g}, \{\psi(t)\}\right)$	(12)	$\underset{\boldsymbol{g},\{\psi(t)\}}{\operatorname{argmax}}$	$p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \boldsymbol{g}, \{\psi(t)\}\right) p\left(\boldsymbol{g}\right) p\left(\{\psi(t)\}\right)$	(15)
$\int p(\{\psi(t)\}) d\{\psi(t)\}$	$\underset{\boldsymbol{g}}{\operatorname{argmax}}$	$p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \boldsymbol{g}\right)$	(24)	$\underset{\boldsymbol{g}}{\operatorname{argmax}}$	$p(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \boldsymbol{g}) p(\boldsymbol{g})$	(26)
$\int p(\boldsymbol{g}) d\boldsymbol{g}$	$\underset{\{\psi(t)\}}{\operatorname{argmax}}$	$p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \{\psi(t)\}\right)$	(33)	$\underset{\{\psi(t)\}}{\operatorname{argmax}}$	$p\left(\{\boldsymbol{y}_{\mathrm{p}}(t)\} \mid \{\psi(t)\}\right) p\left(\{\psi(t)\}\right)$	(33)

is a very weak prior, i.e., it provides little structure. Other priors on the phase shifts are certainly possible. For instance, if the estimation problems are rewritten in terms of phase differences between two consecutive slots we can utilize priors on the phase shifts which are not independent between slots. A classical way of modeling phase drifts is to assume that the phase difference, θ, between two consecutive slots has the following distribution, see [14, Chapter 4],

$$p(\theta) = \frac{1}{2\pi B(a,a)} \left(\frac{\pi+\theta}{2\pi}\right)^{a-1} \left(\frac{\pi-\theta}{2\pi}\right)^{a-1}, -\pi \leq \theta \leq \pi.$$

The constant a parameterizes the distribution, and $a = 0$ gives the uniform distribution. This distribution facilitates the evaluation of (16) in terms of a Bessel function.

A. Solutions

The solutions of the optimization problems in (9), (12), (15), (24), (26) and (33) cannot be found in closed form. Additionally, the problems are all non-convex which means that we need to solve the problems in a heuristic or sub-optimal manner. In the subsections that follow we present the algorithms we used to find solutions to the problems.

1) Iterative ML: Here, we present the solver used to solve (9) and (12). If we assume that we know the channel $\boldsymbol{g}$ perfectly, the optimal phase shifts $\psi(t), t = 1, \ldots, T$, in the LS/ML sense, would be

$$\hat{\psi}(t) = \arg(\boldsymbol{g}^{\mathsf{H}} \boldsymbol{y}_{\mathrm{p}}(t)), \quad t = 1, \ldots, T. \tag{34}$$

Similarly, if we assume that we know the phase shifts perfectly, the optimal channel $\boldsymbol{g}$, in the LS/ML sense, would be

$$\hat{\boldsymbol{g}} = \frac{\sum_{t=1}^{T} e^{-\mathrm{j}\,\psi(t)} \boldsymbol{y}_{\mathrm{p}}(t)}{T\sqrt{p\tau_{\mathrm{p}}}}. \tag{35}$$

Using these two facts we present the iterative ML algorithm, Algorithm 1, which sequentially updates the phase shifts and channel estimates using (34) and (35). The algorithm is considered converged if the objective value in (12) changes less than a predefined value between two iterations.

The iterative ML algorithm is a block coordinate descent method [15, Ch. 9]. Such coordinate descent algorithms will converge to a local optimum, albeit a different local optimum for different initial values. Performance guarantees can only be proven if the problem is convex, which our problem is not. The algorithm has complexity $\mathcal{O}(IMT)$ where I is the number of iterations, see Table III.

Algorithm 1 The iterative ML algorithm.

Input: $\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t=1}^{T}$, p, τ_{p}
Output: $\hat{\boldsymbol{g}}$, $\left\{\hat{\psi}(t)\right\}_{t=1}^{T}$

1: $\hat{\boldsymbol{g}} = \dfrac{1}{T\sqrt{p\tau_{\mathrm{p}}}} \displaystyle\sum_{t=1}^{T} \boldsymbol{y}_{\mathrm{p}}(t)$
2: **while** not converged **do**
3: $\quad \hat{\psi}(t) = \arg(\hat{\boldsymbol{g}}^{\mathsf{H}} \boldsymbol{y}_{\mathrm{p}}(t)), \quad t = 1, \ldots, T$
4: $\quad \hat{\boldsymbol{g}} = \dfrac{\displaystyle\sum_{t=1}^{T} e^{-\mathrm{j}\,\hat{\psi}(t)} \boldsymbol{y}_{\mathrm{p}}(t)}{T\sqrt{p\tau_{\mathrm{p}}}}$
5: **end while**

2) Iterative MAP: Just as with the joint ML problem, we can solve the joint MAP problem with an iterative approach, where sequentially the phase shift estimates are updated with (34) and the channel estimate is updated with

$$\hat{\boldsymbol{g}} = \sqrt{p\tau_{\mathrm{p}}}(Tp\tau_{\mathrm{p}}\boldsymbol{I}_M + \boldsymbol{R}^{-1})^{-1} \sum_{t=1}^{T} e^{-\mathrm{j}\,\psi(t)} \boldsymbol{y}_{\mathrm{p}}(t). \tag{36}$$

This algorithm is presented in Algorithm 2. The algorithm is considered converged if the objective value in (15) changes less than a predefined value between two iterations.

Similar to the iterative ML algorithm, the iterative MAP algorithm is a block coordinate descent method and performance guarantees can not be proven since the problem is non-convex. The algorithm has complexity $\mathcal{O}(IM(T+M))$ where I is the number of iterations, see Table III. The added complexity compared to the iterative ML (Algorithm 1) is due to the matrix-vector multiplication on line 4 in the iterative MAP algorithm. The iterative MAP requires two matrix inversions to be pre-computed with complexity $\mathcal{O}(M^3)$. These inversions are done once every time the large-scale properties of the channel change which does not happen during the transmission of a codeword since we use ergodic capacity as a performance metric.

3) Pairwise: The pairwise algorithm is a sequential algorithm that solves the LS problem in (9), meaning that it can be performed while the pilot measurements from the slots are received. "Pairwise" refers to that the algorithm is considering only two measurements at a time. It is described in Algorithm 3. The pairwise algorithm is non-iterative and has complexity $\mathcal{O}(TM)$ which is lower than both the iterative ML and iterative MAP, see Table III.

TABLE III
THE NUMBER OF COMPLEX ARITHMETIC OPERATIONS IN THE CHANNEL ESTIMATION ALGORITHMS. THE MOST COMPLEX ALGORITHM IS THE ITERATIVE MAP, ALGORITHM 2, DUE TO THE MATRIX-VECTOR MULTIPLICATION ON LINE 4. THE PAIRWISE ALGORITHM, ALGORITHM 3, IS THE LEAST COMPLEX BECAUSE IT IS NOT ITERATIVE. "PRE-COMPUTE" REFERS TO WHAT CAN BE PRE-COMPUTED ONCE EVERY TIME THE LARGE SCALE PROPERTIES OF THE CHANNEL CHANGE AND "ONCE" REFERS TO WHAT IS COMPUTED ONCE PER ALGORITHM INSTANCE. THE COLUMN "OVERALL" REFERS TO THE ASYMPTOTIC SCALING OF THE ALGORITHMS.

Algorithm		Add/sub.	Mult.	Other	Overall
Iterative ML	Pre-compute	0	2	1 div. + 1 sqrt.	$\mathcal{O}(1)$
	Once	$(T-1)M$	M	-	$\mathcal{O}(TM)$
	Each iteration	$2TM-T-M$	$2TM+T+M$	T exp. + T arg.	$\mathcal{O}(TM)$
Iterative MAP	Pre-compute	M	M^2+2	2 inv. $\mathcal{O}(M^3)$+ 1 sqrt.	$\mathcal{O}(M^3)$
	Once	$(T+M-2)M$	M^2	-	$\mathcal{O}((T+M)M)$
	Each iteration	$2TM-T+M^2-2M$	$2TM+M^2+T$	T exp. + T arg.	$\mathcal{O}((T+M)M)$
Pairwise	Pre-compute	0	2	1 div. + 1 sqrt.	$\mathcal{O}(1)$
	Once	$3TM-2T-2M+1$	$3TM+T-M-1$	$(T-1)$ exp. + $(2T-1)$ arg.	$\mathcal{O}(TM)$

Algorithm 2 The iterative MAP algorithm.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T$, p, τ_p, R

Output: $\hat{\boldsymbol{g}}$, $\left\{\hat{\psi}(t)\right\}_{t=1}^T$

1: $\hat{\boldsymbol{g}} = \sqrt{p\tau_\mathrm{p}}(Tp\tau_\mathrm{p}\boldsymbol{I}_M + \boldsymbol{R}^{-1})^{-1}\sum_{t=1}^T \boldsymbol{y}_\mathrm{p}(t)$

2: **while** not converged **do**

3: $\quad \hat{\psi}(t) = \arg(\hat{\boldsymbol{g}}^\mathsf{H}\boldsymbol{y}_\mathrm{p}(t)), \quad t=1,\ldots,T$

4: $\quad \hat{\boldsymbol{g}} = \sqrt{p\tau_\mathrm{p}}(Tp\tau_\mathrm{p}\boldsymbol{I}_M + \boldsymbol{R}^{-1})^{-1}\sum_{t=1}^T e^{-\mathrm{j}\,\hat{\psi}(t)}\boldsymbol{y}_\mathrm{p}(t)$

5: **end while**

Algorithm 3 The pairwise algorithm.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T$, p, τ_p

Output: $\hat{\boldsymbol{g}}$, $\left\{\hat{\psi}(t)\right\}_{t=1}^T$

1: $\boldsymbol{y}_\mathrm{new}(1) = \boldsymbol{y}_\mathrm{p}(1)$

2: **for** $t=2$ to T **do**

3: $\quad \boldsymbol{y}_\mathrm{new}(t) = e^{-\mathrm{j}\arg((\boldsymbol{y}_\mathrm{new})^\mathsf{H}(t-1)\boldsymbol{y}_\mathrm{p}(t))}\boldsymbol{y}_\mathrm{p}(t)$

4: **end for**

5: $\hat{\boldsymbol{g}} = \dfrac{1}{T\sqrt{p\tau_\mathrm{p}}}\sum_{t=1}^T \boldsymbol{y}_\mathrm{new}(t)$

6: $\hat{\psi}(t) = \arg(\hat{\boldsymbol{g}}^\mathsf{H}\boldsymbol{y}_\mathrm{p}(t))$, $t=1,\ldots,T$

4) Gradient Descent: The marginalized ML and MAP in (24) and (26) can be solved (to find a local solution) using standard gradient descent [15, Ch. 3]. Furthermore, the bounds in (21) can be used to speed up the solver.

V. ANTENNA DETECTION

When the user terminal switches between multiple antennas and no phase shifts are occurring between these time slots, the despreaded signal (2) is

$$\boldsymbol{y}_\mathrm{p}(t) = \sqrt{p\tau_\mathrm{p}}\boldsymbol{g}_{S(t)} + \boldsymbol{w}_\mathrm{p}(t),\ t=1,\ldots,T. \tag{37}$$

Here the channel estimation problem boils down to the base station finding the set of slots in which the terminal used each antenna, i.e., the base station should find $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^K$, where $\hat{\mathcal{T}}_1,\ldots,\hat{\mathcal{T}}_K$ are disjoint sets, $\hat{\mathcal{T}}_k \cap \hat{\mathcal{T}}_{k'} = \emptyset, k=1,\ldots,K,\ k'=1,\ldots,K,\ k\neq k'$ and the union of them contains all the time slots: $\bigcup_{k=1}^K \hat{\mathcal{T}}_k = \{1,\ldots,T\}$. The channel estimates can then simply be stated as LS estimates

$$\hat{\boldsymbol{g}}_k = \frac{\sum_{t\in\mathcal{T}_k}\boldsymbol{y}_\mathrm{p}(t)}{|\hat{\mathcal{T}}_k|\sqrt{p\tau_\mathrm{p}}},\ k=1,\ldots,K. \tag{38}$$

It is more complicated to deal with the case when the base station has side information. For example, the MMSE estimator cannot be stated without knowing the performance of the antenna detection, i.e., to state the MMSE estimator, we would need to know all the outcomes of the antenna detection and the probabilities of these events. However, one can still assume that the detection algorithm performed perfectly, $\hat{\mathcal{T}}_k = \mathcal{T}_k, k=1,\ldots,K$, and estimate the channels with methods based on this assumption.

The antenna detection is done by clustering the measurements based on which antenna that was used. This is made possible because each antenna has a constant channel realization within the slots considered, so it is only noise that affects the clustering and detection performance. We study three different clustering algorithms, which are described below.

A. K-means Clustering

The first clustering algorithm is the K-means clustering algorithm [16, Ch. 5]; it is presented in Algorithm 4. The general form can be directly applied to our problem without phase shifts. K-means iteratively finds the K centroids, $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_K$, of the T input points, $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^T$, while trying to minimize the Euclidian distance between the points in the cluster and the centroids,

$$\min_{\{\hat{\mathcal{T}}_k\}_{k=1}^K}\sum_{k=1}^K\sum_{t\in\hat{\mathcal{T}}_k}\left\|\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}} - \boldsymbol{c}_k\right\|^2 \tag{39}$$

Algorithm 4 The K-means algorithm.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$, K, p, τ_p

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Initialize $\{\boldsymbol{c}_k\}_{k=1}^{K}$ with appropriate algorithm

2: $\hat{\mathcal{T}}_k = \left\{t : \left\|\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}} - \boldsymbol{c}_k\right\| \leq \left\|\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}} - \boldsymbol{c}_{k'}\right\|, k' = 1, \ldots, K\right\}, \ k = 1, \ldots, K$

3: **while** $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$ changed **do**

4: $\quad \boldsymbol{c}_k = \frac{1}{|\hat{\mathcal{T}}_k|\sqrt{p\tau_\mathrm{p}}} \sum_{t\in\hat{\mathcal{T}}_k} \boldsymbol{y}_\mathrm{p}(t), \ k = 1, \ldots, K$

5: $\quad \hat{\mathcal{T}}_k = \left\{t : \left\|\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}} - \boldsymbol{c}_k\right\| \leq \left\|\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}} - \boldsymbol{c}_{k'}\right\|, k' = 1, \ldots, K\right\}, \ k = 1, \ldots, K$

6: **end while**

by taking the mean of all the points in each cluster [16, Ch. 5],

$$\boldsymbol{c}_k = \frac{1}{|\hat{\mathcal{T}}_k|\sqrt{p\tau_\mathrm{p}}} \sum_{t\in\hat{\mathcal{T}}_k} \boldsymbol{y}_\mathrm{p}(t), \ k = 1, \ldots, K. \tag{40}$$

The K-means algorithm will converge to different local optima depending on the initialization of the centroids. This initialization can be done by randomly choosing them from the set of the points to be clustered or by a more sophisticated algorithm like the K-means++ algorithm [17]. K-means has a complexity of $\mathcal{O}(TMKI)$, where I is the number of iterations, see Table IV. Note that line 4 in Algorithm 4 gives the LS channel estimate. Hence, the K-means algorithm cyclically detects the antennas and estimates the channel.

B. Expectation-Maximization (EM) Algorithm

The clustering problem can alternatively be formulated as a Gaussian mixture model fit:

$$\begin{aligned}&\{\hat{\boldsymbol{g}}_k\}_{k=1}^{K}, \{\hat{\eta}_k\}_{k=1}^{K} \\ &= \underset{\{\boldsymbol{g}_k\}_{k=1}^{K}, \{\eta_k\}_{k=1}^{K}}{\operatorname{argmax}} \prod_{t=1}^{T} \sum_{k=1}^{K} \eta_k p(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{g}_k)\end{aligned} \tag{41}$$

where

$$p(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{g}_k) = \frac{\exp(-\left\|\boldsymbol{y}_\mathrm{p}(t) - \sqrt{p\tau_\mathrm{p}}\boldsymbol{g}_k\right\|^2)}{\pi^M} \tag{42}$$

is the circularly symmetric complex Gaussian probability density function with covariance matrix $\boldsymbol{I}_M$ evaluated in point $\boldsymbol{y}_\mathrm{p}(t) - \sqrt{p\tau_\mathrm{p}}\boldsymbol{g}_k$ and η_k denotes the probability of any point belonging to cluster k. Equation (41) produces the ML estimate of the mixture parameters: $\boldsymbol{g}_k$ and η_k. Therefore, a possibility is to use the EM algorithm which iteratively finds the ML estimate of the mixture parameters [16, Ch. 6]. The algorithm is presented in Algorithm 5, where a slight modification is made to the original algorithm as we utilize the knowledge of the noise covariance matrix. The $\gamma_{t,k}$ variable denotes the probability of the point $\boldsymbol{y}_\mathrm{p}(t)$ belonging to cluster k and can be interpreted as "fuzzy", or soft, clusters. We can either utilize the soft clusters or create hard clusters by assigning points $\boldsymbol{y}_\mathrm{p}(t)$ to the cluster k with the highest probability, $\hat{\mathcal{T}}_k = \{t : \gamma_{t,k} \geq \gamma_{t,k'}, k' = 1, \ldots, K\}, \ k = 1, \ldots, K$. The EM algorithm will converge to a local optimum. The K-means algorithm is a special case of the EM algorithm. The complexity of the EM algorithm is of the same order as K-means, see Table IV. Note that line 6 in Algorithm 5 can be interpreted as a channel estimate. Hence, the EM algorithm cyclically detects the antennas and estimates the channel.

Algorithm 5 The EM algorithm, where $p(\boldsymbol{x} \mid \boldsymbol{m})$ is the circularly symmetric complex Gaussian probability density function with covariance matrix $\boldsymbol{I}$ evaluated in point $\boldsymbol{x} - \boldsymbol{m}$.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$, K, p, τ_p

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Initialize $\{\boldsymbol{c}_k\}_{k=1}^{K}$ with appropriate algorithm

2: $\eta_k = \frac{1}{K}, \ k = 1, \ldots, K$

3: **while** $\{\gamma_{t,k}\}_{t=1,k=1}^{T,K}$ changed **do**

4: $\quad \gamma_{t,k} = \frac{\eta_k p\left(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{c}_k\right)}{\sum_{k'=1}^{K} \eta_{k'} p\left(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{c}_{k'}\right)}, \ t = 1, \ldots, T, \ k = 1, \ldots, K$

5: $\quad \eta_k = \frac{1}{T}\sum_{t=1}^{T} \gamma_{t,k}, \ k = 1, \ldots, K$

6: $\quad \boldsymbol{c}_k = \frac{\sum_{t=1}^{T} \gamma_{t,k}\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}\sum_{t=1}^{T} \gamma_{t,k}}, \ k = 1, \ldots, K$

7: **end while**

8: $\hat{\mathcal{T}}_k = \{t : \gamma_{t,k} \geq \gamma_{t,k'}, k' = 1, \ldots, K\}, \ k = 1, \ldots, K$

C. Spectral Clustering

The last clustering algorithm to be presented is spectral clustering. The spectral clustering is done with the Ng-Jordan-Weiss algorithm [18] and is presented in Algorithm 6. The affinity matrix, $\boldsymbol{A}$, is describing the similarity between the points and ρ is deciding how fast the similarity should decrease with increased distance between the points. This matrix describes a fully connected graph where the nodes are the measurements and the edges are similarities. The goal of the spectral clustering algorithm is to cut the graph such that the intra-cluster similarities are large while the inter-cluster similarities are small [19]. One of the properties of spectral clustering is that the resulting clusters are not necessarily convex, as compared to e.g., K-means which gives convex clusters. The complexity of the spectral clustering algorithm is dominated by the calculation of the affinity matrix, $\mathcal{O}(T^2M)$ and the eigenvalue decomposition, $\mathcal{O}(T^3)$, see Table IV.

TABLE IV
THE NUMBER OF COMPLEX ARITHMETIC OPERATIONS OF THE CLUSTERING ALGORITHMS 4, 5, 6 AND 10. THE LAST ALGORITHM IS THE INITIALIZATION STRATEGY USED IN THE SIMULATIONS WHICH WILL BE PRESENTED IN SECTION VIII. "PRE-COMPUTE" REFERS TO WHAT CAN BE PRE-COMPUTED ONCE EVERY TIME THE LARGE SCALE PROPERTIES OF THE CHANNEL CHANGE AND "ONCE" REFERS TO WHAT IS COMPUTED ONCE PER ALGORITHM INSTANCE. THE COLUMN "OVERALL" REFERS TO THE ASYMPTOTIC SCALING OF THE ALGORITHMS.

Algorithm		Add/sub.	Mult.	Other	Overall
K-means	Pre-compute	0	1	1 div. + 1 sqrt.	$\mathcal{O}(1)$
	Once	$TK(2M-1)$	$TM(K+1)$	Cost for init. + $T\binom{K}{2}$ comp.	$\mathcal{O}(TMK)$
	Each iteration	$TK(2M-1)+M(T-1)$	$TM(1+K)+MK$	K div. + $T\binom{K}{2}$ comp.	$\mathcal{O}(TMK)$
EM	Pre-compute	0	1	1 div. + 1 sqrt. + $\frac{1}{\pi^M}$	$\mathcal{O}(1)$
	Once	0	0	1 div. + Cost for init.	$\mathcal{O}(1)$
	Each iteration	$3TK(M+1)-MK-T-2K$	$K(3MT+2T+2)$	$TK+KM$ div. + TK exp. + $T\binom{K}{2}$ comp.	$\mathcal{O}(TMK)$
Spectral clustering	Pre-compute	0	1	1 div.	$\mathcal{O}(1)$
	Affinity matrix	$T(T/2-1)\times(2M-1)$	$T(T/2-1)(M+1)$	$T(T/2-1)$ div. + $T(T/2-1)$ exp.	$\mathcal{O}(T^2+TM)$
	Eigenvalue	-	-	$\mathcal{O}(T^3)$	$\mathcal{O}(T^3)$
	Other	$T(T-1+K-1)$	$2M^2+KT$	$KT+T$ div. + $2T$ sqrt. + K-means (Y iterations)	$\mathcal{O}(T^2+M^2+YTK^2)$
Init.	Euclidean	$(K-1)T(2M-1)$	$(K-1)TM$	$(K-1)(2T-1)$ comp.	$\mathcal{O}(TMK)$
	Pseudo-metric	$(K-1)T(3M-2)$	$(K-1)T(3M+1)$	$(K-1)T$ arg. $(K-1)T$ exp. $(K-1)(2T-1)$ comp.	$\mathcal{O}(TMK)$

Algorithm 6 The spectral clustering algorithm [18].

Input: $\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t=1}^{T}$, K, p, τ_{p}, ρ

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Create $\boldsymbol{A}\in\mathbb{R}^{T\times T}$

2: $A_{i,j}=\exp\left(-\dfrac{\|\boldsymbol{y}_{\mathrm{p}}(i)-\boldsymbol{y}_{\mathrm{p}}(j)\|^2}{\rho p\tau_{\mathrm{p}}}\right)$, $i=1,\ldots,T$, $j=1,\ldots,T$, $i\neq j$

3: $A_{i,i}=0,\ i=1,\ldots,T$

4: Create $\boldsymbol{D}\in\mathbb{R}^{T\times T}$

5: $D_{i,j}=0,\ i=1,\ldots,T,\ j=1,\ldots,T,\ i\neq j$

6: $D_{i,i}=\sum_{j=1}^{T}A_{i,j},\ i=1,\ldots,T$

7: Create $\boldsymbol{L}=\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}}\in\mathbb{R}^{T\times T}$

8: Find the eigenvectors $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_K\in\mathbb{R}^T$ corresponding to the K largest eigenvalues of $\boldsymbol{L}$ and put them as columns of $\boldsymbol{V}\in\mathbb{R}^{T\times K}$

9: Create $\boldsymbol{U}\in\mathbb{R}^{T\times K}$

10: $U_{i,j}=\dfrac{V_{i,j}}{\left(\sum_{j'=1}^{K}V_{i,j'}^2\right)^{\frac{1}{2}}},\ i=1,\ldots,T, j=1,\ldots,K$

11: Cluster the rows of $\boldsymbol{U}$ with Algorithm 4 ($p=1$, $\tau_{\mathrm{p}}=1$) to obtain $\hat{\mathcal{T}}_k$

VI. JOINT ANTENNA DETECTION AND CHANNEL ESTIMATION

In the fourth case of channel estimation in this paper, the user terminal has multiple antennas and there are phase shifts between the time slots. In this case, we have to both detect the active antennas and estimate the channels and phase shifts.

We consider the pseudo-metric

$$d(\boldsymbol{x},\boldsymbol{y})=\min_{\theta}\|\boldsymbol{x}-e^{\mathrm{j}\,\theta}\boldsymbol{y}\|=\|\boldsymbol{x}-e^{-\mathrm{j}\arg(\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y})}\boldsymbol{y}\| \quad (43)$$

as a substitute for the Euclidean distance when assigning points to clusters in K-means, when calculating probabilities in the EM algorithm and when determining the similarity between two points in spectral clustering. The pseudo-metric, $d(\boldsymbol{x},\boldsymbol{y})$, has all the properties that a metric has except for $d(\boldsymbol{x},\boldsymbol{y})=0 \implies \boldsymbol{x}=\boldsymbol{y}$ [20, Ch. 4], i.e., a pseudo-metric is non-negative, symmetric and fulfills the triangle inequality. This metric is considered because the distance between a vector and the phase shifted version of the same vector is zero, $d(\boldsymbol{x},e^{\mathrm{j}\,\theta}\boldsymbol{x})=0$ for any θ, which is suitable for our application as we would like two measurements corresponding to the same terminal antenna to be "close" so that the clustering algorithm puts them in the same cluster. Other metrics that fulfill this property are also possible but no superior metric has been found, see Section VIII for a discussion.

The modified, "phase agnostic", versions of Algorithms 4, 5 and 6 are found in Algorithms 7, 8 and 9, respectively. The phase agnostic versions of the algorithms are more complex than their counterparts without phase shifts due to the slightly higher complexity of the pseudo-metric compared to the Euclidean distance, and the use of Algorithm 1 on lines 4 and 6 in Algorithms 7 and 8, respectively. The convergence of the phase agnostic versions of the algorithms is generally slower.

After detecting the clusters, $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$, either the iterative ML (Algorithm 1) or pairwise (Algorithm 3) algorithm can be used to give the channel and phase shift estimates. This is done by using the estimation algorithms K times with $\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t\in\hat{\mathcal{T}}_k}, k=1,\ldots,K$ as measurements. We can also

Algorithm 7 The phase agnostic K-means algorithm.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$, K, p, τ_p

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Initialize $\{\boldsymbol{c}_k\}_{k=1}^{K}$ with appropriate algorithm

2: $\hat{\mathcal{T}}_k = \left\{t : d\left(\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}}, \boldsymbol{c}_k\right) \leq d\left(\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}}, \boldsymbol{c}_{k'}\right), k' = 1, \ldots, K\right\}, \; k = 1, \ldots, K$

3: **while** $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$ changed **do**

4: Use Algorithm 1 with measurements $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t\in\hat{\mathcal{T}}_k}$ and let the resulting $\hat{\boldsymbol{g}}$ be $\boldsymbol{c}_k$, $k = 1, \ldots, K$. The resulting phase shift estimates can be discarded.

5: $\hat{\mathcal{T}}_k = \left\{t : d\left(\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}}, \boldsymbol{c}_k\right) \leq d\left(\frac{\boldsymbol{y}_\mathrm{p}(t)}{\sqrt{p\tau_\mathrm{p}}}, \boldsymbol{c}_{k'}\right), k' = 1, \ldots, K\right\}, k = 1, \ldots, K$

6: **end while**

use the iterative MAP estimator (Algorithm 2), however, this is not the true MAP estimate because of potential errors in the clustering.

Note that the problem is a joint detection and estimation problem, and we treat it as such. With the EM and K-means algorithms, we do not estimate the channel assuming that the antenna detection result is perfect. Rather, we jointly minimize a cost function with respect to all unknowns through a cyclic optimization procedure that alternates between detection of active antennas and estimation of the channels. The channel estimates can be extracted as a byproduct, from intermediate variables in the K-means (line 4) and EM algorithms (line 6). For the spectral clustering algorithm the channel estimation is performed explicitly after the antenna detection is finished.

VII. Achievable Rate

To evaluate the performance of our channel estimation methods, we use the uplink achievable rate, which is a good measure of performance given that we study long codewords and that suitable channel coding techniques are used. This is the case in practice, as practical systems use channel codes and operate close to capacity. In this paper, we do not study the downlink. However, if we were to study the downlink achievable rate, the causality of the system would have to be considered.

The achievable rate will include the effects of antenna misdetection and channel estimation errors. We will not study these effects separately because for a communication link in general a smaller channel estimation mean-square error does not necessarily imply a higher rate as elaborated for example in [21].

To calculate the uplink achievable rate, we consider the uplink received signal in (3). We multiply the received signal with the conjugate transpose of the receive combining vector $\boldsymbol{v}(t)$ to form

$$\boldsymbol{v}^{\mathsf{H}}(t)\boldsymbol{y}_\mathrm{d}(t) = \sqrt{p}e^{\mathrm{j}\,\psi(t)}\boldsymbol{v}^{\mathsf{H}}(t)\boldsymbol{g}_{S(t)}x(t) + \boldsymbol{v}^{\mathsf{H}}(t)\boldsymbol{w}_\mathrm{d}(t). \quad (44)$$

Algorithm 8 The phase agnostic EM algorithm, where $p(\boldsymbol{x} \mid \boldsymbol{m}) = \frac{1}{\pi^M}\exp(-d(\boldsymbol{x}, \boldsymbol{m})^2)$.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$, K, p, τ_p

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Initialize $\{\boldsymbol{c}_k\}_{k=1}^{K}$ with appropriate algorithm

2: $\eta_k = \frac{1}{K}, \; k = 1, \ldots, K$

3: **while** $\{\gamma_{t,k}\}_{t=1,k=1}^{T,K}$ changed **do**

4: $\gamma_{t,k} = \frac{\eta_k p\left(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{c}_k\right)}{\sum_{k'=1}^{K}\eta_{k'}p\left(\boldsymbol{y}_\mathrm{p}(t) \mid \sqrt{p\tau_\mathrm{p}}\boldsymbol{c}_{k'}\right)}$, $t = 1, \ldots, T, \; k = 1, \ldots, K$

5: $\eta_k = \frac{1}{T}\sum_{t=1}^{T}\gamma_{t,k}, \; k = 1, \ldots, K$

6: Use Algorithm 1 with $\{\gamma_{t,k}\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$ as measurements and let the resulting $\hat{\boldsymbol{g}}$ be $\boldsymbol{c}_k^{\mathrm{un}}$, $k = 1, \ldots, K$. The resulting phase shift estimates can be discarded.

7: $\boldsymbol{c}_k = \frac{\boldsymbol{c}_k^{\mathrm{un}}}{\sum_{t=1}^{T}\gamma_{t,k}}, \; k = 1, \ldots, K$

8: **end while**

9: $\hat{\mathcal{T}}_k = \{t : \gamma_{t,k} \geq \gamma_{t,k'}, k' = 1, \ldots, K\}, \; k = 1, \ldots, K$

Algorithm 9 The phase agnostic spectral clustering algorithm.

Input: $\{\boldsymbol{y}_\mathrm{p}(t)\}_{t=1}^{T}$, K, p, τ_p, ρ

Output: $\left\{\hat{\mathcal{T}}_k\right\}_{k=1}^{K}$

1: Create $\boldsymbol{A} \in \mathbb{R}^{T\times T}$

2: $A_{i,j} = \exp\left(-\frac{d\left(\boldsymbol{y}_\mathrm{p}(i), \boldsymbol{y}_\mathrm{p}(j)\right)^2}{\rho p\tau_\mathrm{p}}\right)$, $i = 1, \ldots, T, \; j = 1, \ldots, T, \; i \neq j$

3: $A_{i,i} = 0, \; i = 1, \ldots, T$

4: Create $\boldsymbol{D} \in \mathbb{R}^{T\times T}$

5: $D_{i,j} = 0, \; i = 1, \ldots, T, \; j = 1, \ldots, T, \; i \neq j$

6: $D_{i,i} = \sum_{j=1}^{T}A_{i,j}, \; i = 1, \ldots, T$

7: Create $\boldsymbol{L} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}} \in \mathbb{R}^{T\times T}$

8: Find the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K \in \mathbb{R}^T$ corresponding to the K largest eigenvalues of $\boldsymbol{L}$ and put them as columns of $\boldsymbol{V} \in \mathbb{R}^{T\times K}$

9: Create $\boldsymbol{U} \in \mathbb{R}^{T\times K}$

10: $U_{i,j} = \frac{V_{i,j}}{\left(\sum_{j'=1}^{K}V_{i,j'}^2\right)^{\frac{1}{2}}}$

11: Cluster the rows of $\boldsymbol{U}$ with Algorithm 4 ($p = 1$, $\tau_\mathrm{p} = 1$) to obtain $\hat{\mathcal{T}}_k$

$$C \geq R^{\text{UaF}} = \log_2 \left(1 + \frac{p \left| \mathbb{E} \left\{ e^{\mathrm{j}\,\psi(t) - \mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)} \right\} \right|^2}{\mathbb{V} \left\{ e^{-\mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{w}_{\text{d}}(t) \right\} + p\, \mathbb{V} \left\{ e^{\mathrm{j}\,\psi(t) - \mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)} \right\}} \right) \tag{45}$$

$$C \geq R^{\text{UaF N}} = \log_2 \left(1 + \frac{p \left| \mathbb{E} \left\{ e^{\mathrm{j}\,\psi(t) - \mathrm{j}\,\hat{\psi}(t)} \frac{\hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2} \right\} \right|^2}{\mathbb{V} \left\{ e^{-\mathrm{j}\,\hat{\psi}(t)} \frac{\hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{w}_{\text{d}}(t)}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2} \right\} + p\, \mathbb{V} \left\{ e^{\mathrm{j}\,\psi(t) - \mathrm{j}\,\hat{\psi}(t)} \frac{\hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2} \right\}} \right) \tag{46}$$

We study three different capacity bounds; the use-and-then-forget (UaF) bound [2, Sec. 2.3.4] [12, Thm. 4.4] with two different receive combining vectors (UaF and UaF N), and the estimation bound (EB) in [22, Lemma 3]. By studying three lower bounds, we get a clearer picture of the true capacity as the greatest lower bound in each point is the most telling.

The first bound we study is the UaF bound where the receive combining vector is the channel estimate scaled by a deterministic constant, in this case M, $\boldsymbol{v}(t) = \frac{e^{\mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}}{M}$. This is a receive combining vector that is widely used in massive MIMO literature as the capacity bound using this receive combining vector can be found in closed form in many cases and is optimal in the case of perfect channel state information. This receive combining vector, together with the UaF capacity bounding technique gives us (45) (at the top of this page).

In the second bound we still use the UaF bounding technique, however the receive combining vector is different here and is the channel estimate scaled by the inverse of its squared norm, $\boldsymbol{v}(t) = \frac{e^{\mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2}$. The resulting bound is showed in [12, Section 4.2.1] to be tighter than UaF in some cases, especially in the cases with less channel hardening. The bound is called UaF N and is given by (46) (at the top of this page).

To adhere to the system model for the third bound we consider, EB, in [22, Lemma 3] our receive combining vector is the channel estimate scaled by its norm $\boldsymbol{v}(t) = \frac{e^{\mathrm{j}\,\hat{\psi}(t)} \hat{\boldsymbol{g}}_{\hat{S}(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|}$. The system model in [22] is a downlink multiuser MIMO system with linearly precoded transmission to multiple single-antenna users. To apply the EB bound from [22] to our system model, we state the conjugate of (44) with the notation from [22] under the braces (without the block indices):

$$\underbrace{\boldsymbol{y}_{\text{d}}^{\mathsf{H}}(t)\boldsymbol{v}(t)}_{=y_k} = \underbrace{e^{-\mathrm{j}\,\psi(t)} \boldsymbol{g}_{S(t)}^{\mathsf{H}}}_{=\mathbf{h}_k^{\mathsf{H}}} \underbrace{\underbrace{\sqrt{p}}_{=\sqrt{\mathcal{E}_k}} \underbrace{\boldsymbol{v}(t)}_{=\mathbf{v}_k} \underbrace{x(t)^*}_{=s_k}}_{=\mathbf{x}} + \underbrace{\boldsymbol{w}_{\text{d}}^{\mathsf{H}}(t)\boldsymbol{v}(t)}_{=z_k}.$$

Unit precoding vectors are assumed in [22], i.e., $\|\mathbf{v}_k\|^2 = 1$, which is why we use the normalized receive combining vector in our model. Also, z_k should be independent of $\mathbf{v}_k$ which is also fulfilled in our model with the normalized $\boldsymbol{v}(t)$. This can be verified by noting that the probability distribution of $\boldsymbol{w}_{\text{d}}^{\mathsf{H}}(t)\boldsymbol{v}(t)$ conditioned on $\boldsymbol{v}(t)$ does not depend on $\boldsymbol{v}(t)$ for any unit-norm $\boldsymbol{v}(t)$. This gives our last capacity bound:

$$C \geq R^{\text{EB}} = \mathbb{E} \left\{ \log_2 \left(1 + p \left| \frac{e^{\mathrm{j}(\psi(t) - \hat{\psi}(t))} \hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|} \right|^2 \right) \right\} - \frac{1}{\tau_{\text{c}}} \log_2 \left(1 + \tau_{\text{c}} p\, \mathbb{V} \left\{ \frac{e^{\mathrm{j}(\psi(t) - \hat{\psi}(t))} \hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}} \boldsymbol{g}_{S(t)}}{\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|} \right\} \right), \tag{45}$$

where τ_{c} is the size of one slot.

VIII. Numerical Simulations

As the expectations and variances in the achievable rate expressions cannot be calculated we do numerical simulations to see how our detection and estimation methods perform using these rates. We also evaluate how much that can be gained by having prior knowledge of the correlation matrices.

We use three different forms of Rayleigh fading, $\boldsymbol{g}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R}), k = 1, \ldots, K$:

1) Uncorrelated Rayleigh fading, $\boldsymbol{R} = \boldsymbol{I}_M$.
2) Correlated Rayleigh fading with the approximate Gaussian local scattering (AGLS) model [12, Sec. 2.6] with angular standard deviation (ASD), $\sigma_\varphi = 10°$, nominal angle, $\varphi = 30°$ and half-wavelength antenna spacing, where the element in row j and column k of the correlation matrix $\boldsymbol{R}$ is
$$\boldsymbol{R}_{j,k} = e^{2\pi\,\mathrm{j}\,0.5(j-k)\sin(\varphi)} e^{\frac{\sigma_\varphi^2}{2}(2\pi 0.5(j-k)\cos(\varphi))^2}, \quad j = 1, \ldots, M,\ k = 1, \ldots, M. \tag{46}$$
3) Correlated Rayleigh fading with the exponential correlation model from [23], assuming a uniform linear array where the channel correlation between adjacent antennas is equal to a constant r:
$$\boldsymbol{R} = \begin{bmatrix} r^0 & r^1 & \cdots & r^{M-1} \\ r^1 & r^0 & \cdots & r^{M-2} \\ \vdots & \vdots & \ddots & \vdots \\ r^{M-1} & r^{M-2} & \cdots & r^0 \end{bmatrix}. \tag{47}$$
We use $r = 0.9$; r being real-valued corresponds to the signal from the terminal impinging broadside onto the array.

The phase shift $\psi(t)$ is assumed to be uniformly distributed between $-\pi$ and π and independent between time slots. When having multiple antennas, the terminal is assumed to have

four antennas, $K = 4$, and each antenna observes a random realization with the same distribution. This corresponds to a rich scattering environment around the terminal. The selected antenna is selected randomly using a uniform distribution in each time slot. In some cases, the antenna switching is not entirely random but based on some side information that is available to the terminal. In principle, the terminal could communicate additional information through its choice of which antenna that is active, similar to the spatial modulation concept [24]. A possible future extension of the paper could consider corresponding algorithms and trade-offs. Throughout the simulations we use $\tau_{\mathrm{c}} = 200$ as in [12, Sec. 4.1.3]. Furthermore, the simulations depicted are with $M = 100$, $T = 20$, and $\tau_{\mathrm{p}} = 10$.

Both the K-means and EM algorithm are initialized[1] by starting with a random measurement as the first cluster center. After that, the measurement furthest (in Euclidean distance or the pseudo-metric d) from all the existing cluster centers are added until there are K cluster centers as opposed to the K-means++ algorithm which adds cluster centers with a probability that is proportional to the distance to the existing cluster centers. Our initialization algorithm can be found in Algorithm 10 and its complexity is given in Table IV. We observed that our initialization method performs better, compared to K-means++, on average considering we only have one initialization. The results might change slightly if we can start the algorithms with many more initializations and save the best choice. The scaling parameter, ρ, in the spectral clustering is set to M to prevent the distances from growing large at high SNR or large number of antennas. Consider two pilot measurements, from slot t_1 and t_2, without phase shifts. The expected squared distance between these measurements given that the terminal used the same antenna and different antennas in slot t_1 and t_2 would be $\mathbb{E}\left\{\|\boldsymbol{y}_{\mathrm{p}}(t_1) - \boldsymbol{y}_{\mathrm{p}}(t_2)\|^2 \mid S(t_1) = S(t_2)\right\} = 2M$ and $\mathbb{E}\left\{\|\boldsymbol{y}_{\mathrm{p}}(t_1) - \boldsymbol{y}_{\mathrm{p}}(t_2)\|^2 \mid S(t_1) \neq S(t_2)\right\} = 2M(p\tau_{\mathrm{p}} + 1)$, respectively. Both of these distances scale with M. We chose the scaling parameter such that the number of antennas will not affect the clustering. In [18], it is suggested that one can search over many different scaling parameters, and choose the one that gives the tightest clusters. We limit the number of iterations in K-means and EM and set the maximum number of iterations to 1000. The K-means algorithm, EM algorithm, and spectral clustering algorithm (either the standard version or the phase agnostic version) are denoted by KM, EM and SC, respectively, in the figure legends.

The problems in (24) (ML-$\boldsymbol{g}$) and (26) (MAP-$\boldsymbol{g}$) are solved with a gradient descent method where the modified Bessel function of the first kind either has been lower (L) or upper (U) bounded. The starting points of the gradient descent is 5 iterations of either the iterative ML or iterative MAP algorithms.

First, we study the case where the terminal has a single antenna, like in Section IV, which means that we need to estimate both the channel and the phase shifts. Fig. 1, where achievable rate is shown as a function of SNR, shows that the pairwise algorithm is clearly inferior to the other methods, especially for low SNRs. We see that the MAP based estimators do not perform noticeably better when the channel has uncorrelated Rayleigh fading. This behavior is expected, as uncorrelated Rayleigh fading is a weak prior. However, with correlated Rayleigh fading, we see that the ML based estimators provide worse estimates than the MAP based estimates. This is clear, as more information is used when doing MAP estimation compared to ML estimation. The loss of not having priors is not huge; it is less than 1 dB at 1 bit/s/Hz. From the numerical simulations, the iterative ML (Algorithm 1) and iterative MAP (Algorithm 2) do not require many iterations to converge. At $p = -15$ dB less than 10 iterations are required on average and at $p = 0$ dB 5 iterations are required on average.

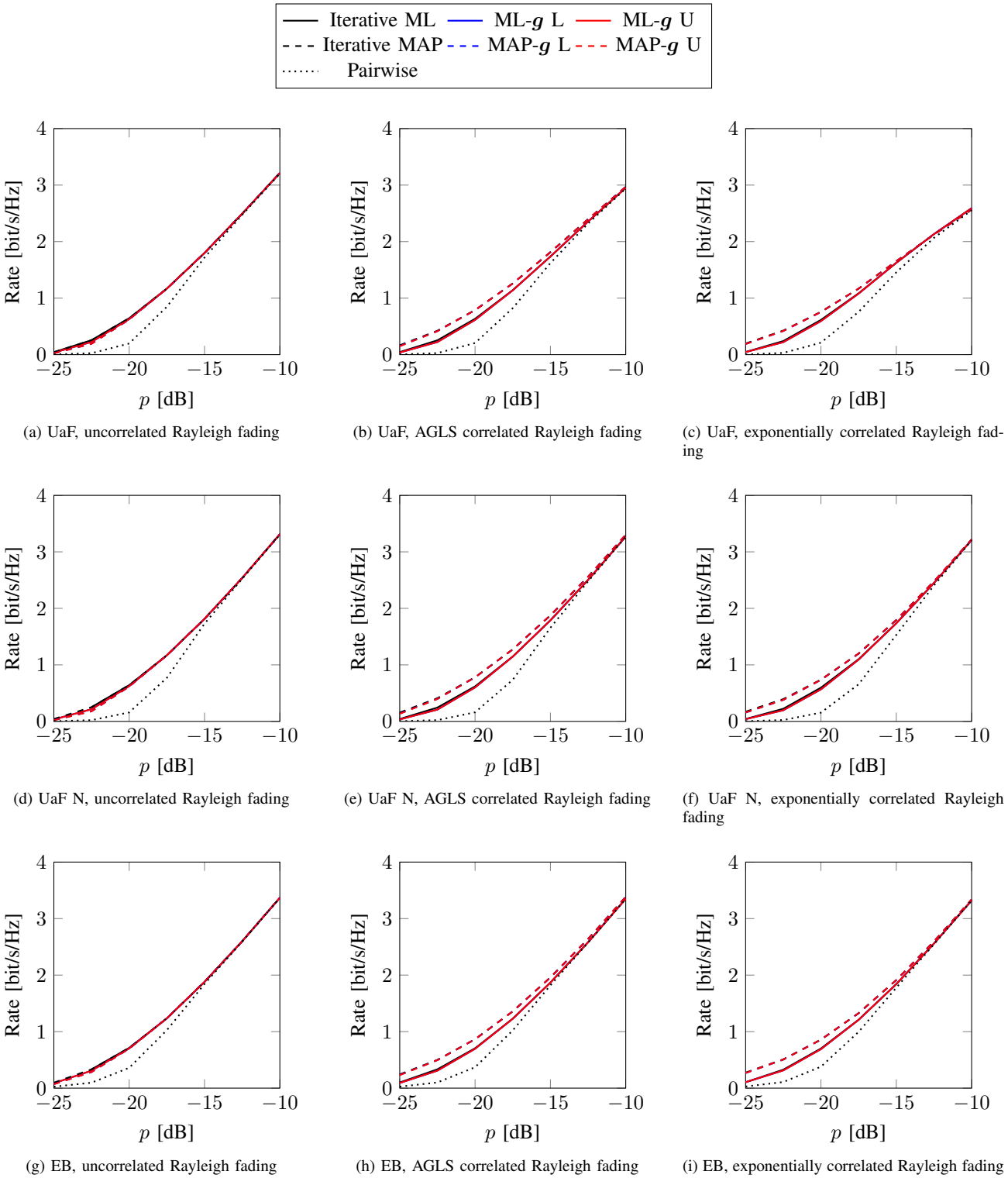
Second, we study how the clustering algorithms perform when there are no phase shifts between the time slots, i.e., the channel estimation case in Section V. The results are seen in Fig. 2, where achievable rate is shown as a function of SNR. In the figure, what is most clear is that the UaF bound is the worst of the three used bounds; both the UaF N and the EB give higher achievable rates. The UaF bound can be calculated in closed form in some system models and is tight when the channel hardens which is why it is commonly used in the massive MIMO literature. Here, we do not have much channel hardening due to occasional missed detection of the antenna clusters. By scaling the receive combining vector by the inverse of its norm squared we obtain a much better bound, UaF N. However, the bound that is the best in this case is EB, which is because the second term in (45) becomes very small when τ_{c} is much larger than the number of users which is the case here ($200 \gg 1$). EB almost attains the same achievable rate as when the channel is perfectly known, $\log_2(1 + Mp)$, at high SNR. Furthermore, we can see that the clustering algorithms have similar performance, however, spectral clustering performs better in lower SNR and the K-

Algorithm 10 The furthest distance initialization for the K-means and EM algorithm where $f(\boldsymbol{x}, \boldsymbol{y})$ is either $d(\boldsymbol{x}, \boldsymbol{y})$ or $\|\boldsymbol{x} - \boldsymbol{y}\|$.

Input: $\{\boldsymbol{y}_{\mathrm{p}}(t)\}_{t=1}^{T}$, K
Output: $\{\boldsymbol{c}_k\}_{k=1}^{K}$

```
Randomly choose c_1 from {y_p(t)}_{t=1}^T
for k = 2 to K do
    for t = 1 to T do
        a_t = ∞
        for k' = 1 to k − 1 do
            a_t = min(a_t, f(c_{k'}, y_p(t)))
        end for
    end for
    Set c_k = y_p( argmax_{t∈{1,...,T}} a_t )
end for
```

[1]During the course of this work, it has come to our knowledge that initializing these algorithms appropriately is very important. In the earlier version of this work [1], the K-means++ algorithm [17] was used, and as can be seen by comparing the figures, performed much worse than when using the current initialization method.



(a) UaF, uncorrelated Rayleigh fading

(b) UaF, AGLS correlated Rayleigh fading

(c) UaF, exponentially correlated Rayleigh fading

(d) UaF N, uncorrelated Rayleigh fading

(e) UaF N, AGLS correlated Rayleigh fading

(f) UaF N, exponentially correlated Rayleigh fading

(g) EB, uncorrelated Rayleigh fading

(h) EB, AGLS correlated Rayleigh fading

(i) EB, exponentially correlated Rayleigh fading

Fig. 1. Simulation of achievable rates (UaF, UaF N, and EB) with one antenna at the terminal and phase shifts between the time slots (scenario in Section IV), with different fading models (uncorrelated, AGLS correlated, and exponentially correlated Rayleigh fading). The channel is constant for $T = 20$ slots, the base station has $M = 100$ antennas and the pilot length is $\tau_{\mathrm{p}} = 10$.

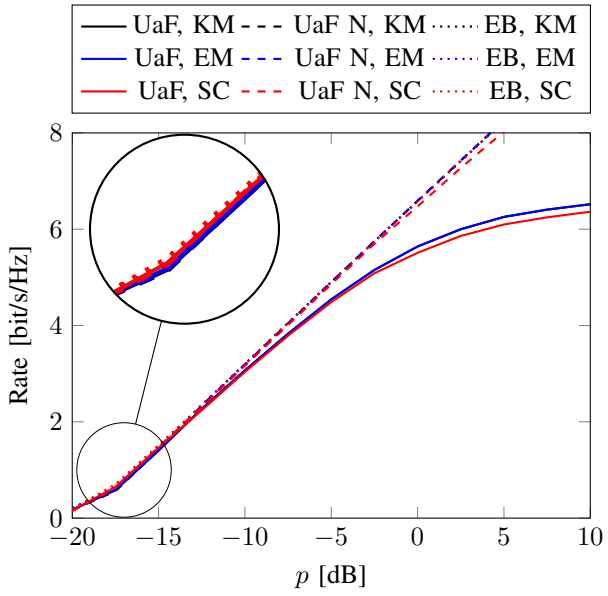


Fig. 2. Simulation of achievable rates with multiple terminal antennas ($K = 4$) and no phase shifts between the time slots, with MAP channel estimates and uncorrelated Rayleigh fading channels. The channels are constant for $T = 20$ slots, the base station has $M = 100$ antennas and, the pilot length is $\tau_{\mathrm{p}} = 10$.

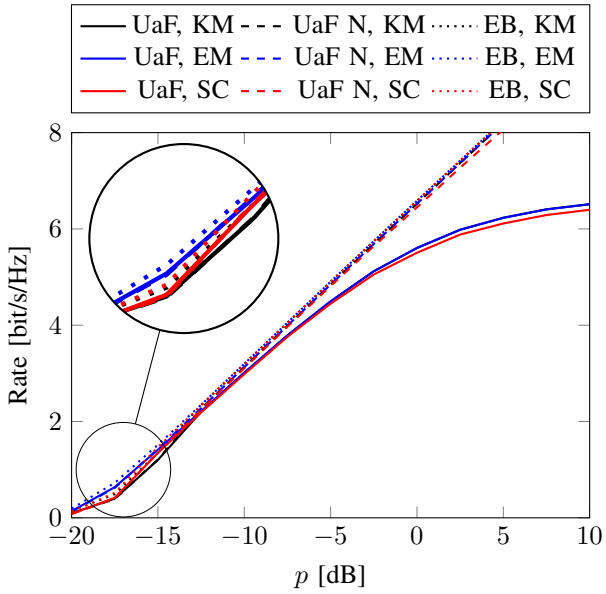


Fig. 3. Simulation of achievable rates with multiple terminal antennas ($K = 4$) and phase shifts between time slots, with iterative MAP and uncorrelated Rayleigh fading channels and uniformly distributed, independent, phase shifts. The channels are constant for $T = 20$ slots, the base station has $M = 100$ antennas, and the pilot length is $\tau_{\mathrm{p}} = 10$.

means and EM algorithms perform better in higher SNR.

Third, we combine the phase shifts and the antenna switching and study the case in Section VI. The results can be seen in Fig. 3, which shows achievable rate as a function of SNR. Here, we can see that the clustering algorithms are close in performance. However, the EM algorithm is consistently better than the other two. We can also conclude that our method of handling the phase shifts in the clustering algorithms is close to optimal. This conclusion is drawn from the fact that the curves in Figs. 2 and 3 are very close.

Lastly, to answer how much we gain from utilizing multiple time slots when estimating the channel, we compare the case where the base station estimates the channel separately in each slot to our case with phase shifts and multiple antennas. Other benchmarks such as non-coherent transmission are possible but such transmission schemes do not use pilots and therefore the comparison to our scheme is irrelevant since our methods are an add-on to a pilot-based scheme. The result is shown in Fig. 4. In this figure, the achievable rate when having perfect channel knowledge is also considered. We see that the rates are improved the most at relatively low SNR; there is an improvement of approximately 2 dB at 1 bit/s/Hz. In this figure we also evaluated the clustering performance by including a genie-aided antenna detector. We see that clustering with our phase agnostic EM algorithm comes very close to the genie-aided detector.

Although not included in the manuscript, more simulation scenarios have been evaluated.

- Particularly, we saw that our methods will always give an improvement compared to re-estimating the channel in each slot, irrespective of what τ_{p} is.
- Moreover, we observed that, while the K-means and

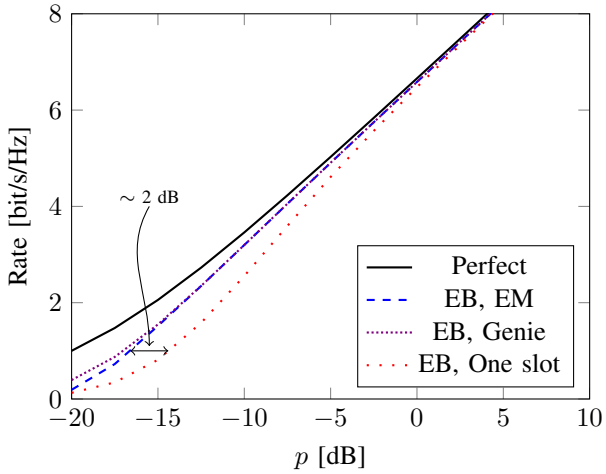


Fig. 4. Comparison of achievable rates when using only a single slot for channel estimation and the proposed method of utilizing multiple slots. The terminal has ($K = 4$) antennas, the base station has $M = 100$ antennas and the pilot length is $\tau_{\mathrm{p}} = 10$ The channels are uncorrelated Rayleigh fading and constant for $T = 20$ slots. The phase shifts between the slots are assumed to be independent and uniformly distributed. Here, the estimation is done by iterative MAP when estimating over many slots and the ordinary MAP when estimating the channel in one slot. The curve denoted by "Perfect" is the case where the channel is exactly known at the base station and "Genie" is denoting a genie-aided antenna detector. An improvement of around 2 dB can be seen at 1 bit/s/Hz.

EM algorithm will always give an improvement over re-estimating the channel in each slot, the spectral clustering algorithm is sensitive to the number of slots and might degrade the performance if $T < 20$ compared to the baseline.

- We have also evaluated the chordal distance, $d_{\mathrm{C}}(\boldsymbol{x},\boldsymbol{y}) = \sqrt{1-\frac{|\boldsymbol{x}^{\mathsf{H}}\boldsymbol{y}|^2}{\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2}}$, as an alternative for the pseudo-metric d in the phase agnostic clustering algorithms. The chordal distance performs slightly worse for K-means while no noticeable difference could be seen for spectral clustering. This is because the chordal distance only considers the separation in angle, while the separation in magnitude is lost.
- We also considered the case where the channels from the terminal antennas are correlated with the Kronecker model [25] and the exponential correlation model similar to (47) at the terminal side. Perhaps somewhat unexpectedly, we saw a slight performance increase in the low SNR region when the channels from the terminal antennas were correlated and the signal is impinging broadside or the phase agnostic clustering algorithms are used. We conjecture that when the channels from the terminal antennas are correlated, the antenna detection might be less crucial.
- Simulations evaluating channel mean-square error and antenna detection accuracy have also been performed. These results show the same trends as the achievable rate.

To shed some more light on, and validate, how the capacity bounds hold up in practice, in terms of using realistic modulation and channel coding, we continue our numerical analysis by considering coded bit-error-rates. We assume that the terminal is sending quadrature phase-shift keying (QPSK) modulated symbols and uses a low-density parity-check (LDPC) code with rate 1/2, which means that the transmission rate is 1 bit/s/Hz. The LDPC code has length 1000 and is defined in [26][2]. The codes in [26] are optimized for AWGN channels, but they work well in massive MIMO settings because of channel hardening, as shown in [27]. In the simulation we use the receive combining vector $\boldsymbol{v}(t) = \frac{e^{\mathsf{j}\,\hat{\psi}(t)}\hat{\boldsymbol{g}}_{\hat{S}(t)}}{M}$.

The implemented decoder is an AWGN decoder which is not optimal but performs well in massive MIMO settings because of channel hardening. The AWGN decoder considers a symbol s is sent with power ρ and corrupted by AWGN w with variance σ^2. The received signal is $y = \rho s + w$. The log-likelihood ratio of the i:th bit of the symbol s, i.e., s^i is

$$\log\left(\frac{p(y|s^i=0)}{p(y|s^i=1)}\right) = \log\left(\frac{\sum_{s\in\mathcal{S}_0^i}\exp\left(\frac{|y-\rho s|^2}{\sigma^2}\right)}{\sum_{s\in\mathcal{S}_1^i}\exp\left(\frac{|y-\rho s|^2}{\sigma^2}\right)}\right), \quad (48)$$

where $\mathcal{S}_0^i$ and $\mathcal{S}_1^i$ are the sets of symbols with a 0 or 1 in the i:th position, respectively. In our setting the symbol $x(t)$ is transmitted over the channel and received at the base station, see (3). The received signal is combined with the receive combining vector and interpreted as a symbol that passed through an AWGN channel where the effective channel gain is $\frac{\sqrt{p}}{M}\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2 s$ and the effective noise variance is

$$\sigma^2 = \frac{p}{M^2}\,\mathbb{E}\left\{\left|\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\| - e^{\mathsf{j}\,\psi(t)-\mathsf{j}\,\hat{\psi}(t)}\hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}}\boldsymbol{g}_{S(t)}\right|^2\right\} + \frac{1}{M^2}\,\mathbb{V}\left\{e^{-\mathsf{j}\,\hat{\psi}(t)}\hat{\boldsymbol{g}}_{\hat{S}(t)}^{\mathsf{H}}\boldsymbol{w}_d(t)\right\}. \quad (49)$$

Thus, the log-likelihood ratio for the i:th bit in the symbol is

$$\log\left(\frac{\sum_{s\in\mathcal{S}_0^i}\exp\left(-\frac{\left|\boldsymbol{v}^{\mathsf{H}}(t)\boldsymbol{y}_{\mathrm{d}}(t)-\frac{\sqrt{p}}{M}\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2 s\right|^2}{\sigma^2}\right)}{\sum_{s\in\mathcal{S}_1^i}\exp\left(-\frac{\left|\boldsymbol{v}^{\mathsf{H}}(t)\boldsymbol{y}_{\mathrm{d}}(t)-\frac{\sqrt{p}}{M}\|\hat{\boldsymbol{g}}_{\hat{S}(t)}\|^2 s\right|^2}{\sigma^2}\right)}\right), \quad (50)$$

While σ^2 is a deterministic number, it is not available in closed form and we used Monte-Carlo simulations to calculate it prior to running the performance simulations.

The result of the bit-error-rate simulations can be seen in Fig. 5, where the vertical lines are the achievable rates predicted by the UaF bound at 1 bit/s/Hz. We can see that even though we do not have as much channel hardening as in conventional massive MIMO in [27], the code, which is optimized for AWGN channels, also performs well in our scenario as the gaps between the vertical lines and the waterfalls are small. The gaps between the vertical lines and the waterfalls are due to the limited code length, shaping loss of QPSK modulation, the AWGN assumption, and the fact that we have a lower bound on the capacity. As can be seen in the figure, the predicted gaps between the capacity bounds are kept fairly well in the bit-error results, which further validates the results in the paper. Additionally, the loss from having phase shifts does not degrade the performance of the multiple-antenna case much more than the single-antenna case which again indicates that our method of handling the phase shifts in the clustering algorithms is close to optimal.

IX. Conclusions

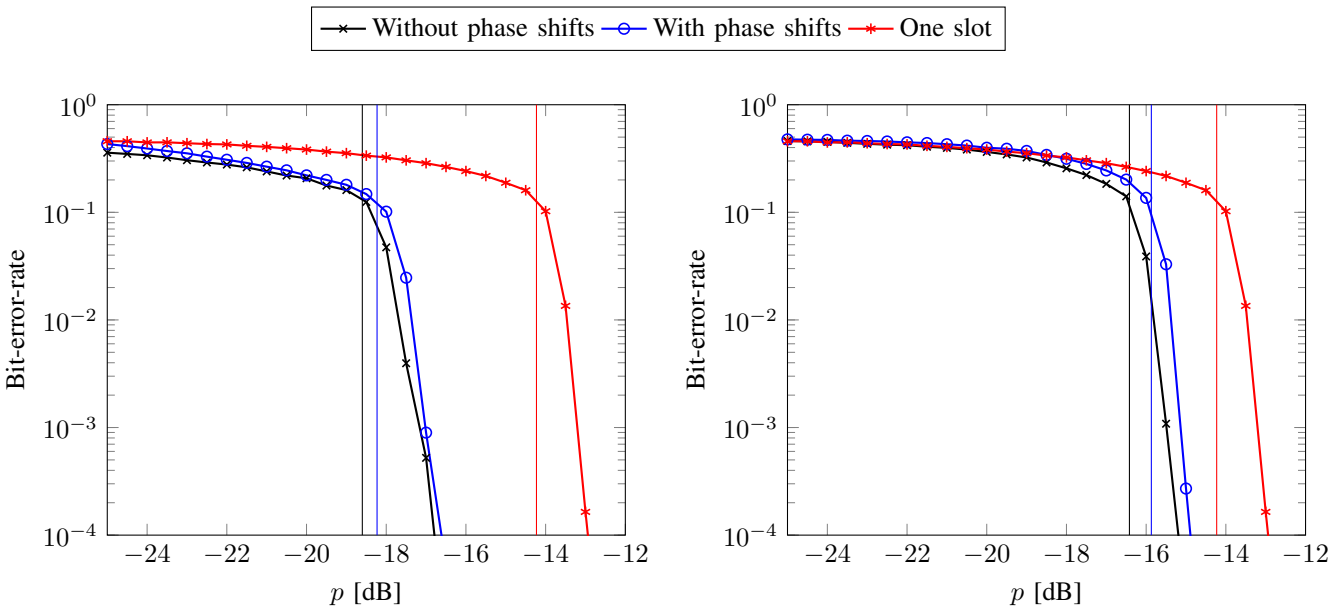
In this paper we considered the problem of filtering the channel estimates between time slots of a non-coherent 5G user terminal. Our suggested methods detect the selected antenna and estimate the phase shift and channels in each time slot. We saw that, in most cases, our phase agnostic EM algorithm is the best for the antenna detection and that our joint MAP algorithm best estimates the channels and the phase shifts. We showed that the uplink data rate can be improved by considering more time slots than only considering one time slot for channel estimation and we verified our capacity bounds with a bit-error-rate simulation.

The scheme is also applicable in multi-user scenarios. The users are all acting by sending orthogonal pilots and data. If their coherence time is longer than one system coherence block the methods presented in the paper can be used to improve their channel estimates.

[2]We use the code with maximum variable node degree 9 in Table I in [26].

References

[1] E. Becirovic, E. Björnson, and E. G. Larsson, "Joint antenna detection and channel estimation for non-coherent user terminals," in *IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2019, pp. 1–5.



(a) Single antenna, $K = 1$

(b) Multiple antennas, $K = 4$

Fig. 5. Simulation of the coded bit-error-rate with uncorrelated Rayleigh fading channels. The clustering of the measurements is solved with spectral clustering (Algorithms 6 and 9), and estimation is done by the Iterative MAP (Algorithm 2) when applicable (with phase shifts) and the ordinary MAP otherwise. At least 10 frame errors, i.e., code word errors, are counted per SNR point. The vertical lines are approximate achievable rates (UaF). The base station is equipped with 100 antennas, $M = 100$.

[2] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.

[3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1207–1215, July 2000.

[4] Z. Tang, R. C. Cannizzaro, G. Leus, and P. Banelli, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2226–2238, May 2007.

[5] M. K. Ozdemir and H. Arslan, "Channel estimation for wireless OFDM systems," *IEEE Communications Surveys Tutorials*, vol. 9, no. 2, pp. 18–48, February 2007.

[6] Y. Liu, Z. Tan, H. Hu, L. J. Cimini, and G. Y. Li, "Channel estimation for OFDM," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1891–1908, Fourth quarter 2014.

[7] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer*. Academic Press, 2018.

[8] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[9] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, 2014.

[10] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2363–2378, May 2017.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing*, ser. Prentice-Hall signal processing series. Prentice Hall PTR, 1993.

[12] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[13] Z.-H. Yang and Y.-M. Chu, "On approximating the modified Bessel function of the first kind and Toader-Qi mean," *Journal of Inequalities and Applications*, vol. 2016, no. 1, p. 40, February 2016. [Online]. Available: https://doi.org/10.1186/s13660-016-0988-1

[14] H. L. Van Trees, K. L. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I : Detection, Estimation, and Filtering Theory*. John Wiley & Sons, Incorporated, 2013.

[15] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Science & Business Media, 2006.

[16] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*, 3rd ed. Elsevier, 2012.

[17] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[18] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[19] U. von Luxburg, "A tutorial on spectral clustering," *Stat Comput*, vol. 17, no. 4, pp. 395–416, December 2007.

[20] J. L. Kelley, *General Topology*, ser. The University series in higher mathematics. Van Nostrand, 1955.

[21] E. Björnson, "Pilot contamination is not captured by the MSE," https://ma-mimo.ellintech.se/2018/09/13/pilot-contamination-is-not-captured-by-the-mse/, (accessed Apr. 9, 2020).

[22] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3258–3268, May 2018.

[23] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," vol. 5, no. 9, pp. 369–371, 2001.

[24] R. Y. Mesleh, H. Haas, S. Sinanovic, C. W. Ahn, and S. Yun, "Spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2228–2241, 2008.

[25] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1211–1226, August 2002.

[26] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[27] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

Ema Becirovic (S'18) received the M.Sc. degree from Linköping University, Sweden, in 2018, where she is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Division for Communication Systems. Her research is mainly focused on massive MIMO and massive machine-type communications.

Emil Björnson (S'07-M'12-SM'17) received the M.S. degree in engineering mathematics from Lund University, Sweden, in 2007, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011. From 2012 to 2014, he held a joint post-doctoral position at the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France, and the KTH Royal Institute of Technology. He joined Linköping University, Sweden, in 2014, where he is currently an Associate Professor and a Docent with the Division of Communication Systems.

He has authored the textbooks *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013) and *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017). He is dedicated to reproducible research and has made a large amount of simulation code publicly available. He performs research on MIMO communications, radio resource allocation, machine learning for communications, and energy efficiency. Since 2017, he has been on the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING since 2016.

He has performed MIMO research for over ten years and has filed more than twenty MIMO related patent applications. He has received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, the 2016 Best Ph.D. Award from EURASIP, the 2018 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 EURASIP Early Career Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, and the 2019 IEEE Signal Processing Magazine Best Column Award. He also co-authored papers that received Best Paper Awards at the conferences, including WCSP 2009, the IEEE CAMSAP 2011, the IEEE WCNC 2014, the IEEE ICC 2015, WCSP 2017, and the IEEE SAM 2014.

Erik G. Larsson (S'99–M'03–SM'10–F'16) received the Ph.D. degree from Uppsala University, Uppsala, Sweden, in 2002. He is currently Professor of Communication Systems at Linköping University (LiU) in Linköping, Sweden. He was with the KTH Royal Institute of Technology in Stockholm, Sweden, the George Washington University, USA, the University of Florida, USA, and Ericsson Research, Sweden. His main professional interests are within the areas of wireless communications and signal processing. He co-authored *Space-Time Block Coding for Wireless Communications* (Cambridge University Press, 2003) and *Fundamentals of Massive MIMO* (Cambridge University Press, 2016). He is co-inventor of 19 issued U.S. patents.

Currently he is an editorial board member of the *IEEE Signal Processing Magazine*, and a member of the *IEEE Transactions on Wireless Communications* steering committee. He served as chair of the IEEE Signal Processing Society SPCOM technical committee (2015–2016), chair of the *IEEE Wireless Communications Letters* steering committee (2014–2015), General respectively Technical Chair of the Asilomar SSC conference (2015, 2012), technical co-chair of the IEEE Communication Theory Workshop (2019), and member of the IEEE Signal Processing Society Awards Board (2017–2019). He was Associate Editor for, among others, the *IEEE Transactions on Communications* (2010-2014) and the *IEEE Transactions on Signal Processing* (2006-2010).

He received the IEEE Signal Processing Magazine Best Column Award twice, in 2012 and 2014, the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, the IEEE ComSoc Best Tutorial Paper Award in 2018, and the IEEE ComSoc Fred W. Ellersick Prize in 2019.