Low-Overhead Cyclic Reference Signals for Channel Estimation in FDD Massive MIMO

Rikke Apelfröjd, Wolfgang Zirwas, and Mikael Sternad

***Abstract*— Massive multiple input multiple output (MIMO) transmission and coordinated multipoint transmission are candidate technologies for increasing data throughput in evolving 5G standards. Frequency division duplex (FDD) is likely to remain predominant in large parts of the spectrum below 6 GHz for future 5G systems. Therefore, it is important to estimate the downlink FDD channels from a very large number of antennas, while avoiding an excessive downlink reference signal overhead. We here propose and investigate a three part solution. First, massive MIMO downlinks use a fixed grid of beams. For each user, only a subset of beams will then be relevant, and require estimation. Second, sets of coded reference signal sequences, with cyclic patterns over time, are used. Third, each terminal estimates its most relevant channels. We here propose and compare a linear mean square estimation and a Kalman estimation. Both utilize frequency and antenna correlation, and the later also utilizes temporal correlation. In extensive simulations, this scheme provides channel estimates that lead to an insignificant beamforming performance degradation as compared to full channel knowledge. The cyclic pattern of coded reference signals is found to be important for reliable channel estimation, without having to adjust the reference signals to specific users.**

***Index Terms*— Radio channel estimation, frequency division duplex, antenna arrays.**

I. Introduction

MULTI user multiple input multiple output (MIMO) downlink transmission techniques are becoming increasingly important in the study of future systems. For the past decade, the evolution of multiuser MIMO has moved in two main directions: massive MIMO [1]–[5] and coherent coordinated multipoint (CoMP) joint transmission (JT) (also known as network MIMO) and coordinated beamforming [6]–[8]. Each of these, and combinations of them, have been identified as key enablers for the fifth generation mobile system [9]–[12].

Coherent JT CoMP allows for interference mitigation schemes to reduce intercell interference. This is especially important for boosting performance at the cell edges in intercell interference limited networks, such as heterogeneous networks with frequency reuse 1, [13]–[16]. Delays in the fixed network cause outdated CSI, which can severely reduce gains [17], [18], but channel prediction in combination with robust precoding has shown promising results [19], [20].

In massive MIMO, the number of transmit antennas at a base station is very large. This leads to several advantages. In the special case when the number of simultaneously scheduled users is much smaller than the number of transmit antennas, channel vectors to different users will be almost orthogonal with high probability. Each user will then also experience a large linear beamforming gain from maximum ratio combining (MRC) [3], provided that a constant CSI quality can be ensured. More users can be added to optimize system performance, e.g. the sum throughput [21], but this may come at the cost of cell edge performance.

Massive MIMO requires CSI for channels from a vast number of antennas. Adding coherent JT CoMP to the framework would increase this requirement further. A main challenge with massive MIMO downlinks in frequency division duplex (FDD) systems is therefore to avoid a massive downlink reference signal (RS) overhead. In current systems the RS are often resource orthogonal, i.e. each antenna or beam transmits RSs on different resources. Another option is to let the antennas/beams transmit their RSs on the same set of resources, but the sequence of reference signals over the resources form a code, similar to code division multiple access (CDMA), one code per antenna/beam. These codes can be made to be orthogonal if the number of RS resources is larger than or equal to the number of antennas/beams. Such a pilot scheme is called code orthogonal. The overhead associated with massive MIMO can be reduced by using fewer RS resources than the number of beams/antennas resulting in non-code orthogonal RS sequences that are overlapping in the time-frequency domain i.e. the antennas must transmit RSs on the *same* RS resources, which is suboptimal [22].

Many researchers therefore instead focus on time division duplex (TDD) systems. There, RSs could be transmitted from the scheduled users in uplink time slots. The users are mostly assumed to be few, so orthogonal RS could be used. Then, channel reciprocity can be utilized to obtain estimates of the downlink channels. Channel estimation in TDD is challenging due to imperfections in channel reciprocity (caused by incomplete compensation for uplink-downlink offsets or by outdating of CSI when users are mobile), and also due to limited transmit power at the user, and by hardware impairments and lack of downlink interference estimates [4], [12], [21], [23]. Still,

Manuscript received July 2, 2018; revised October 12, 2018 and December 20, 2018; accepted January 16, 2019. Date of publication January 25, 2019; date of current version May 15, 2019. The associate editor coordinating the review of this paper and approving it for publication was L. Lampe. *(Corresponding author: Mikael Sternad.)*

R. Apelfröjd and M. Sternad are with Signals and Systems, Uppsala University, 751 21 Uppsala, Sweden (e-mail: mikael.sternad@signal.uu.se).

W. Zirwas is with Nokia Bell Labs, 81541 Munich, Germany.

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

Digital Object Identifier 10.1109/TCOMM.2019.2895059

it has great potential as illustrated in [3]. There, a comparison of the plausible operation conditions of FDD and TDD in massive MIMO assuming resource *orthogonal* RSs concludes that TDD is more beneficial than FDD.

However, there is one important argument for why we need to solve the problem of using overlapping downlink references signals from a large number of antennas in FDD massive MIMO: A large part of the spectrum is presently allocated to FDD and will probably remain so for many years to come. It would be unfortunate not to be able to take advantage of the potential massive MIMO gains in these spectral resources. Björnsson *et al.* [5] identified enabling massive MIMO for FDD systems as the "critical question" for future research on the topic of massive MIMO. Solving the joint problem of RS design and channel estimation for massive MIMO and CoMP in FDD systems would also allow backward compatibility, which is a desirable quality for next generation systems [24]. This motivates us to study and develop a strategy that is useful for channel estimation and prediction in wireless systems that may use combinations of small cells, massive MIMO and JT CoMP within a cooperative area.

For this purpose we use the fixed grid of beam concept, where each beam is wideband and is controlled by an effective or virtual antenna port.[1]

A. Contribution

A scheme for downlink channel estimation for massive MIMO in combination with JT CoMP in FDD systems must solve two main problems. First, we have a potentially very large set of channel components that need to be estimated without introducing an unreasonably large overhead. Second, the solution must support a large number of users, with very different conditions in terms of channel gains and fading. We will present a scheme for an FDD implementation where the overhead scales with the number of channels that will be relevant for a terminal, which will typically be in the range 5-30, in systems with hundreds of antenna elements.

The primary key property that we use is that when the channel components have varying average gains, then each user needs to estimate only the strongest channel components as seen from that user. If different users will have different strongest channel components, then estimating only their strongest channel components will lead to an insignificant decrease in the multi-users scheduling gain. Signals from antennas located at different base stations will in general have large differences in received power. For antennas located at the same site, the average channel gains should, on the other hand, be very similar. We therefore need to introduce some system design elements to reduce this similarity between channels for co-localized antennas.

Our proposed framework has four main components:

1) Antennas will be structured into a fixed *grid of beams*. The downlink channel between a user and one antenna port will be denoted a channel component. At any given user position, only fractions of the antenna ports will have strong signal, so only a fraction of the channel components needs to be estimated.
2) Downlink RSs will be transmitted as overlapping RSs using *coded RS sequences*. The codes are designed such that they provide unique RS patterns for each of a potentially very large number of antenna ports within a cooperation area. The size K of the RS blocks (the coded sequence length) is selected proportional to the number of channel components that need to be estimated for a typical user.
3) *Correlation* over time, space and frequency is utilized by a linear least mean squared error (LLMSE) estimator or by a Kalman filter to improve the CSI quality.
4) Use of *cyclic sequences of RS codes* ensures good estimations regardless of the users position, by utilizing time correlation.

The main purpose of this work is to find a solution that is practical and can be implemented. The purpose is not to design optimal RS codes, as these would need to be re-optimized each time a new user is scheduled. The work is an extension of the RS design and channel estimation introduced in [25] and further investigated in [26]. We here extend the solution in [25] beyond that of flat block-fading channels by first utilizing an LLMSE estimation. Second, we introduce a Kalman filter estimate that uses low order autoregressive (AR) models to represent the temporal correlation. This improves the performance, but comes at the cost of added off-line complexity. We therefore investigate a reduced Kalman filter and show that this gives an improvement compared to the LLMSE filter. The AR-models utilized in the Kalman filter need to be estimated. In particular the covariance matrices of these models can prove difficult to estimate with a limited amount of training data. We will address these difficulties and provide simulation results to show that using the Moore-Penrose Pseudo inverse to estimate covariance matrices is a good choice. Furthermore, we add cyclic RSs to the framework in [25] and show that this is important for estimation performance and for user fairness.

B. Related Work

Channel estimation for massive MIMO in FDD has recently gained interest [27]–[34]. Similar to our design, these works assume overlapping RSs and utilize some type of correlations to improve the estimates. In contrast to our design, the works of [27]–[33] focus on optimizing the RSs based on the channel properties of the scheduled user. Such a solution would demand that the reference signals are re-optimized each time a new user is scheduled. In a situation with bursty traffic, this would cause extra feedback overhead and introduce undesirable delays.

In the earliest of the works above, namely [27], [28], user specific RS designs were suggested. Based on downlink transmission of these RSs, the terminal generated an uplink feedback to the base station which then utilized Kalman filters to acquire CSI. Another single-user scheme, partially based on the use of compressed sensing, was proposed in [29]. These concepts would demand user specific RS resources, so the

[1]When we here use the word fixed, we mean fixed over a slow time scale, e.g. several seconds. However, the fixed grid of beams can change whenever the distribution of users changes significantly, e.g. if an office building is empty during night then the grid of beams can be adjusted such that it transmits little or no energy into that building.

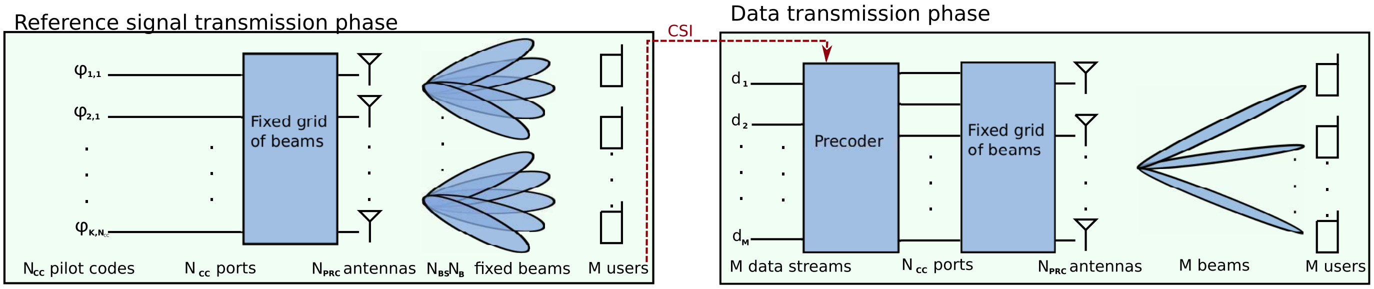


Fig. 1. System setup.

overhead increases with the number of users and the benefit achieved by using overlapping RSs decreases rapidly as the number of users increase.

Jiang *et al.* [30], Tseng *et al.* [31], and Gao *et al.* [32] instead optimize the RSs off-line to improve the average channel estimation for the scheduled users. The required RS overhead would in [30] and [31], increase with the number of active users, which is an undesirable property, as a large part of the massive MIMO gain comes from serving a large number of users simultaneously (due to the logarithmic behaviour of the capacity). The work of [32] assumes sparsity in the channel impulse response (CIR) and correlation between the channels from different antennas. This estimation scheme does not provide gains when CIR are not sparse, which often occurs in real channels [35].

These multi-user methods, though they are an improvement as compared to the single user case, still require RS re-optimization when new users are scheduled, and whenever the shadow fading changes. Our solution instead introduces a fixed grid of beams and cycling between pre-determined sets of RSs. The combination of these will ensure that most *potential* users can estimate their channels. Then, RSs need not be readjusted and fed back and even users not yet scheduled for service can prepare for transmission by estimating their channels based on the downlink RS. This is also a strength as compared to the TDD scenario, where only the scheduled users can be allowed to transmit RS in order to limit RS overhead.

A somewhat related idea is proposed in [34] where RSs are transmitted over a number of beams, lower than the number of transmit antennas at the base station. This work focuses on estimating only the strongest one or two beams, claiming that to be sufficient to obtain close to full sum-rate capacity gain. While this may be reasonable for MRC transmission when the users are few, and inter user interference can be ignored, it will not be adequate when interference mitigation is necessary.

A comparative study between estimation of massive MIMO channels in TDD and in FDD (with grid of beam concepts) has recently been performed in [36]. The authors found that for clusters of closely spaced users, or a hot spot scenario, the performance with the FDD grid of beam concept is reduced. This problem is caused by using a fixed allocation of a low to moderate number of beams that are designed to span the full cell area. The number of beams serving a hot spot area might then become small and inadequate. A potential remedy for that is to rearrange the fixed grid of beam on a slower time scale as suggested in [26].

Regardless of the estimation algorithms, estimation errors will always be present to some extent and these inaccuracies should ideally be accounted for when adjusting the precoding as in e.g. [19]. The estimation inaccuracies can be treated in a similar way as inaccuracies due to feedback limitation, see e.g. [37], [38]. Furthermore, when the estimation error variance varies for different channel estimates, as it will in our proposed scheme, then the feedback rates can be adjusted such that the feedback errors are in the same range as the estimation errors.

Layout and Notations

Section II provides details on the fixed grid of beams and the RS codes that are the key design elements to our solution. The estimation algorithms used for evaluations are described in Section III. Simulations are provided in Section IV and Section V highlights conclusions and suggests areas for further investigation.

An extended report version of this paper is available [39]. This report includes more details on the channel estimation algorithms as well as some comments on how intoducing a fixed grid of beams may limit the end performance of massive MIMO and also on how our RS code design could be used to improve channel estimates in TDD.

We use $\hat{x}(t_1|t_2)$ and $\hat{x}(t_1)$ to denote an estimate of a vector $\bar{x}(t_1)$ at time t_1. The first is based on all past measurements up until time t_2 and the second on one measurement at time t_1 only. The notations $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^\dagger$ represent transpose, hermitian transpose and pseudo inverse respectively. The operator $E[\cdot]$ represents averaging over both time and frequency. The number of elements in a set $\mathcal{A}$ is denoted $|\mathcal{A}|$.

II. Key Design Elements

An important first step for our proposed scheme is to create effective channel components that have *different* path loss and shadow fading. Assuming an OFDM FDD downlink where RSs are transmitted in cooperation clusters of N_{BS} base stations, each equipped with N_{tx} antennas, there are a total of $N_{PRC} = N_{BS} \cdot N_{tx}$ physical radio channels between the serving antennas and each single antenna user. As the path loss and shadow fading of the N_{tx} channels to a user from the antennas located at the same base stations will be similar, it will be very difficult to separate those channels from each other based on $\ll N_{PRC}$ reference symbols. However, if we consider a system, where a total of N_{PRC} antennas are arranged into a fixed grid of beams with $N_{CC} = N_B \cdot N_{BS}$ beams by digital or analogue beamforming as in Figure 1, then the resulting N_{CC} channel components will have different average channel powers, as seen from one user.

We here use a fixed grid of beam as follows: Assume that $\tilde{h}(k, l, \tau)$ is the complex number representing the physical radio channel from the l'th antenna on a resource k at time τ in the frequency domain. These are translated into beams via beamforming weights $b(n, l)$. The channel component from the n'th beam is then

$$h(k, n, \tau) = \sum_{l=1}^{N_{PRC}} b(n, l)\tilde{h}(k, l, \tau) \quad \text{for } n = 1, .., N_{CC}, \quad k = 1, \ldots, K. \tag{1}$$

The fixed grid of beams improves the channel estimation by two main contributions. First, due to being scattered and reflected differently, the beams will have different strengths at any given position as seen from one of the users. As verified by system simulation studies in [25], the number of the strongest (relevant) beams will therefore typically be $\ll N_{CC}$. This improves the possibility to resolve all relevant channels using only $K \ll N_{CC}$ RSs. Second, beamforming reduces the time variations of channels [1], and hence makes them easier to estimate [40].

To achieve massive MIMO gains, each user then needs to estimate its strongest channel components only. We use sets of K reference symbols, where we may have $K \ll N_{CC}$. The K RSs are transmitted in the downlink with a period of T_s, indexed by τ. Each beam transmits a unique RS code in the downlink. From these, the users will be able to estimate their strongest channel components, which are then fed back to the base station. Through the CSI a precoder which directs the signal energy to each user, e.g. by MRC or interference mitigation precoding, can then be designed. In this work, we focus on the estimation of CSI. By designing N_{CC} RSs codes of length K such that any set of up to K of these are linearly independent, any user will be able to estimate its K strongest channel components, although with a bias caused by contamination from the other (weaker) beams.

When the strongest downlink beam channels (CCs) have been obtained for each user who may be scheduled, they are made available on the network side. They may then be used to influence the scheduling of users and the physical resource allocation of the subsequent downlink transmission. They are also used to design a precoder matrix that generates inputs to the fixed grid of beams, see Figure 1. The precoder could be designed for single user maximum ratio combining (MRC), multi-site multi-user coordinated beamforming, or zero-forcing joint transmission. Kalman estimators produce not only point estimates of the CCs but also estimates of their accuracy, which enables the use of robust linear precoding for joint transmission for multiple sites [19], [20].

Good multi-user results will require the use of a sufficiently high number of fixed beams per base station, so that no coverage holes are created by the fixed beamforming. This issue is not specific to the present channel estimation scheme, it applies to any scheme that uses grid-of-beams.

A. Reference Signal Code Design

Assuming that the channel estimation is performed for each user independently, we can focus on the channel components (1) of a single user. We let each beam n transmit a RS symbol $\varphi(k, n, \tau)$ on every available RS resource. The received downlink signals at the user on the $k \in [1, K]$ RS bearing resources at time τ, $\bar{y}(\tau) = \left[y(1, \tau) \ \ldots \ y(K, \tau)\right]^T$, are then given by

$$\bar{y}(\tau) = \Phi(\tau)\bar{h}(\tau) + \bar{v}(\tau), \tag{2}$$

with the channel vector

$$\bar{h}(\tau) = \begin{bmatrix} \left[h(1, 1, \tau) \ \ldots \ h(K, 1, \tau)\right]^T \\ \vdots \\ \left[h(1, N_{CC}, \tau) \ \ldots \ h(K, N_{CC}, \tau)\right]^T \end{bmatrix}$$

an RS matrix

$$\Phi(\tau) = \left[\text{diag}\{\varphi(k, 1, \tau)\}_{k\in[1,K]} \ldots \text{diag}\{\varphi(k, N_{CC}, \tau)\}_{k\in[1,K]}\right]$$

and a noise vector

$$\bar{v}(\tau) = \left[v(1, \tau) \ \ldots \ v(K, \tau)\right]^T$$

with covariance matrix

$$R_v = E[\bar{v}(\tau)\bar{v}(\tau)^*]. \tag{3}$$

Here $v(k, \tau)$ is the sum of noise and intercluster interference on resource k, assumed to be i.i.d. over time. These terms will henceforth be denoted as noise.

To ensure that a user will be able to separate and estimate its strongest channel components, each beam will have a unique RSs code $\bar{\varphi}(n, \tau) = [\varphi(1, n, \tau) \ldots \varphi(K, n, \tau)]^T$. In the special case when $N_{CC} \leq K$, these codes could be fully orthogonal (resource-orthogonal or code orthogonal). However, in order to have good coverage in the full cooperation area we would like to allow for more beams than there are available RS resources. Then the codes cannot be allowed to be orthogonal. We may then loosen this requirement and instead require any subset of up to K codes out of the N_{CC} codes should be linearly independent.

1) An Introductory Example: To give an intuitive understanding of the concept, we begin with an example assuming only $N_{CC} = 9$ fixed beams and $K = 6$ flat fading RS resources. As the channels are flat fading we have $h(k, n, \tau) = \underline{h}(n, \tau)$ for $k = 1, \ldots, K$, where the underline is used to mark the channel over all subcarriers in the flat fading scenario. The measurement equation in (2) can then be simplified, using a compressed RS matrix $\underline{\Phi}(\tau)$, to

$$\begin{aligned} \bar{y}(\tau) &= \underline{\Phi}(\tau)\underline{\bar{h}}(\tau) + \bar{v}(\tau) \\ &= [\bar{\varphi}(1, \tau) \ldots \bar{\varphi}(N_{CC}, \tau)] \begin{bmatrix} \underline{h}(1, \tau) \\ \vdots \\ \underline{h}(N_{CC}, \tau) \end{bmatrix} + \bar{v}(\tau). \end{aligned} \tag{4}$$

As an example, let each antenna port now have its own unique code in accordance with (5), as shown at the top of the next page. Assume that only three beams are relevant for the user of interest, and let $\underline{\Phi}_{\text{rel}}(\tau)$ be a 6x3 matrix that is formed

$$\underline{\Phi}(\tau) = \begin{bmatrix} e^{2.0j} & e^{-2.3j} & e^{1.7j} & e^{-2.8j} & e^{0.6j} & e^{1.2j} & e^{2.3j} & e^{-1.6j} & e^{3.1j} \\ e^{-2.3j} & e^{-2.8j} & e^{1.2j} & e^{-1.6j} & e^{-1.2j} & e^{-0.6j} & e^{-2.5j} & e^{2.4j} & e^{-3.1j} \\ e^{-0.2j} & e^{-1.7j} & e^{2.4j} & e^{1.7j} & e^{-2.6j} & e^{-2.9j} & e^{1.2j} & e^{1.2j} & e^{0.8j} \\ e^{1.7j} & e^{1.2j} & e^{3.1j} & e^{-0.6j} & e^{1.2j} & e^{-3.1j} & e^{0.7j} & e^{-0.9j} & e^{-0.9j} \\ e^{-2.6j} & e^{-0.5j} & e^{1.0j} & e^{-2.8j} & e^{3.1j} & e^{-0.4j} & e^{2.7j} & e^{1.9j} & e^{0.6j} \\ e^{-0.6j} & e^{-0.5j} & e^{0.1j} & e^{1.5j} & e^{-1.0j} & e^{0.7j} & e^{2.3j} & e^{2.2j} & e^{1.8j} \end{bmatrix}. \quad (5)$$

by the corresponding columns of $\underline{\Phi}(\tau)$. As columns of $\underline{\Phi}(\tau)$ are linearly independent, the submatrix will have full rank and the left pseudo inverse exists. The three relevant channel components can then be collected in a vector $\underline{h}_{\text{rel}}(\tau)$ and estimated through

$$\hat{\underline{h}}_{\text{rel}}(\tau) = \underline{\Phi}^{\dagger}_{\text{rel}}(\tau)\bar{y}(\tau). \quad (6)$$

The subindex rel is used to indicate the part of a vector or a matrix associated with the relevant channels.

The set of relevant channel components will depend on the user's location, but for any location, the user will be able to estimate up to $K = 6$ channel components through (6), as any subset of up to six column vectors of $\underline{\Phi}(\tau)$ will be linearly independent. This can easily be verified by testing all possible subsets in this simple case with only 9 channel components.

The RSs transmitted over the non-relevant channel components will cause a bias in the estimate (6). The size of bias depends on the number of non-relevant channel components, their gains, and the degree of orthogonality between the vectors in $\underline{\Phi}_{\text{rel}}(\tau)$ and the other vectors. The bias is zero in the case of fully orthogonal vectors. It is also zero if we use beam deactivation, described in [26], to send no RS on the non-relevant beams.

In this example, the channels $h(k, n, \tau)$ were assumed to be equal within the time-frequency block where the K RS were transmitted. If this is not the case, we could assume the channels $h(k, n, \tau)$ to be unequal within the block for different k and estimate all of them. However, this will increase the complexity and reduce the performance of the estimator. In the following, the investigated channel estimators will be based on the perfect correlation (flat fading) assumption, but they are then evaluated in a scenario where the correlation is high but not perfect. This situation will be obtained by placing the k RS on adjacent subcarriers within the frequency extent of a LTE resource block. Other sets of codes can then be transmitted in other sections of the bandwidth.

2) General Reference Signal Codes: The RS matrix (5) was designed such that any subset of up to six column vectors will be linearly independent. We wish to create a general such matrix with a per antenna RS power constraint. Such a constrain will ensure that all beams have equal RS power budget.[2] This can e.g. be achieved by setting

$$\varphi(k, n, \tau) = \exp\left(\theta(k, n, \tau) \cdot j\right), \quad (7)$$

where j is the imaginary unit. The angles $\theta(k, n, \tau)$ should then be designed to ensure that any subset of K vectors are linearly independent. There are many selection criteria and methods. In [25], a real-valued design parameter $\phi(\tau) > 0$, that can be fixed over time or varying, is selected and the phase for resource k and beam n is defined as

$$\theta(k, n, \tau) = (k\phi(\tau))^n. \quad (8)$$

We here use the codes by (8) in order to ease repeatability of our results, since it is specified by one scalar parameter $\phi(\tau)$.

How to select the real-valued design parameter $\phi(\tau)$ of (8) is an object for investigation which we shall return to in Section IV. Some selections of $\phi(\tau)$ cause submatrices to be very ill-conditioned. A users whose strongest channel components have RS codes that compose such an ill-conditioned submatrix may then end up with very poor channel estimates. For example, through the inverse in (6) the noise and RSs from the weak channels may be amplified by a poorly conditioned matrix $\underline{\Phi}_{\text{rel}}(\tau)$. In order to ensure that users at any position will be able to estimate their strongest channel components, selecting the phase angles $\theta(k, n, \tau)$ should be done off-line, enabling an exhaustive search.

3) Cycling Reference Signals: Even with an exhaustive search, some sets of relevant channel components form very good RS code submatrices, while a few others will form submatrices with rather large eigenvalue spread, making it more difficult to estimate the channels of those users. However, with an estimation algorithm that does utilize temporal correlation, we can improve the fairness amongst user by introducing cycling RSs. This idea was introduced for the uplink in [41] where the number of users exceeded the number of available RS resources.

For the coding (8), we consider μ different parameters $\phi(\tau)$ that all result in RS matrices with reasonable low condition numbers of their $K \times K$ submatrices. These are then cycled with a period μ over time such that $\phi(\tau) = \phi(\tau + \mu)$. It is then likely that any subset of relevant channels will receive a well conditioned submatrix for at least one of the cycling RS matrices. Over a time period of μ, the user will have at least one good estimate, and a number of reasonably good estimates. Through this, we introduce diversity into our RS coding scheme.

Introducing cycling RSs does not introduce any additional overhead, nor does it require that all users are equipped with estimators that can utilize temporal correlation. However, channel estimators that utilize the temporal channel correlation will be able to improve their estimate by combining estimates obtained with different subsequent RS code vectors, thereby reducing the influence of badly conditioned cases.

To illustrate the effect of cycling RSs on user fairness, we assume a system with $N_{cc} = 72$ channel components and

[2]Constraints on signal envelope properties in the time domain are outside the scope of this work, but if the RSs are chosen off-line, the framework could be extended to include this aspect.

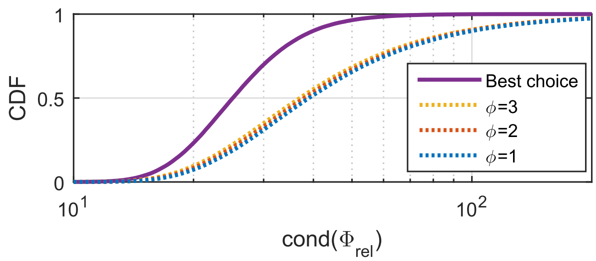


Fig. 2. CDF of the condition number of the RS submatrix $\Phi_{\text{rel}}(\tau)$ for different sets of 18 relevant channel components (i.e. different user positions) for different $\phi(\tau)$ in (8).

$K = 18$ RS resources. We form three RS matrices $(\Phi(\tau))$ through (8) using $\phi(\tau) = \{1, 2, 3\}$. The values of $\phi(\tau)$ are chosen to ensure reasonably low numbers of cond$(\Phi(\tau)$ (given by $\{2.55, 2.48, 2.27\}$).

In Figure 2 we study the condition numbers of submatrices to these. A low condition number indicates a low eigenvalue spread and hence the channels will be easier to estimate correctly. For example, a pseudoinverse estimate (6) will generate less noise amplification. Here, 10^5 sets of $K = 18$ relevant channel components were randomly selected and the condition of the RS submatrix associated with that set of channel components was calculated. These would correspond to user positions with different sets of relevant channels, and in order to ensure user fairness, all such sets need to have reasonable condition. Along with the cdf of the submatrices associated with each of the three $(\Phi(\tau))$, we show a cdf denoted "Best choice", which is the result if for every set of $K = 18$ relevant channel components, we select the $(\Phi(\tau))$ whose submatrix has the lowest condition.

The three different $\phi(\tau)$ have very similar CDF's where approximately 10 % of submatrices, corresponding to 10 % of the potential users, have a condition number of 100 or above. These users would be at an disadvantage if only one of these RS matrices are chosen, as their estimates would likely be worse than for the users with well conditioned matrices. In contrast, if we were to cycle the three, then most users (> 99.9 %) would have a condition number below 100 in at least one out of three RS transmissions, and should be able to gain a good estimate for those times. When temporal correlation is used (and is high enough), users can then use the estimates based on their best RS matrix to improve the subsequent estimates.

III. Channel Estimation

A main feature of our proposed solution is to only estimate a subset of the channel components for each user. These components are referred to as the relevant channel components, subindexed by rel. They may include only the channel components required by the data transmission and hence be selected by some threshold, or they may also include some extra channel components. The relevant channel components for each user can be estimated separately, either directly in the user equipment or in the base stations, based on feedback of measurements from the users.[3] We here assume that each user estimates its own relevant downlink channel components and reports them when required.

For the estimates we assume a RS structure as follows; First, we assume sets of orthogonal RSs are transmitted sparsely, e.g. every 0.5 s. As shadow fading only changes on a long time scale, of at least several of hundreds of ms for pedestrian users, we can use these to estimate the channel correlation in space (i.e. the cross correlation between the channel components) and frequency. In addition, as the relevant channels change with the shadow fading, due to e.g. new building causing new beams to be reflected and hence a change in the set of relevant channels, these sparsely transmitted orthogonal RSs can also be used to find the set of relevant channels for new users that enter the system. Our preferred way to extract the correlation would be to allow each beam a few orthogonal RS over the full carrier bandwidth. Through these the channel components are estimated and then interpolation can be used over all subcarriers to gain a large set of channels from which to estimate the correlation. As the orthogonal RSs are repeated infrequently, they do not introduce a large extra overhead cost. Second, on a faster time scale, e.g. every couple of ms, all beams transmit their individual RS codes on K available time-frequency RS resources on a set of subcarriers with highly correlated fading. From these RSs the channel of these subcarriers are then estimated repeatedly over time.

There are other ways to estimate the relevant CCs. For example the relevant CCs could be estimated from uplink channel estimates. But this has other drawbacks (all potential users would need to send uplink RS) so we prefer to average downlink channel estimates over frequency.

A. LLMSE Estimation

We utilize the correlation over space and frequency by using an LLMSE (or Wiener) filter [43]. This estimator can be used as a start-up estimator, before information about temporal statistics of the channel has been obtained. For this we define a channel vector $\bar{h}_{\text{rel}}$, which consists of the relevant channel components for the given user and $\bar{h}_{\bar{\text{rel}}}$ which consists of the non relevant channel components. The estimates of the relevant channel components are

$$\hat{h}_{\text{rel}} = R_{h,\text{rel},y} R_y^{-1} \bar{y}, \tag{9}$$

where $R_{h,\text{rel},y} = E[\bar{h}_{\text{rel}} \bar{y}^*]$ is the cross covariance matrix between the vector of relevant channel components and the measurement signal and $R_y = E[\bar{y}\bar{y}^*]$ is the covariance matrix of the measurement signal vector. The required covariances can be estimated from the sparsely transmitted orthogonal RS or from past channel estimates, see [39] for details.

B. Kalman Filter

Kalman filters have been found useful for channel estimation and prediction, see e.g. [40], [42], [44]–[46]. To incorporate the temporal correlation (as well as correlation over frequency and space), we utilize a Kalman filter, which enables us to use channel information from *all* previous measurements with lower complexity than an LLMSE filter that attempts

[3]For a discussion on benefits and drawback of placing the downlink channel estimation in the terminal and base stations respectivly, we refer the reader to [15] and [42] and references therein.

the same. For this, we model the channel statistics over time by an auto regressive (AR) model

$$\begin{aligned}\bar{x}(\tau+1) &= A\bar{x}(\tau) + B\bar{u}(\tau),\\ \bar{h}(\tau) &= C\bar{x}(\tau)\\ Q &= E[\bar{u}(\tau)\bar{u}^*(\tau)].\end{aligned} \tag{10}$$

Here A, B and C are complex-valued state space matrices, $\bar{u}(t)$ is the white zero-mean process noise and $\bar{x}(\tau)$ is a state space vector of dimension $\rho K N_{CC}$, where ρ is the model order. The frequency correlation and the spatial/antenna correlation between channel components is modelled through the covariance matrix of the process noise, Q, of dimension KN_{CC}.

The measurement vector in (2) can then be expressed as

$$\bar{y}(\tau) = \Phi(\tau)C\bar{x}(\tau) + \bar{v}(\tau). \tag{11}$$

Estimation of the parameters of the model (10)-(11) requires several subsequent channel estimates, so it cannot be estimated by the sparsely transmitted orthogonal RSs. However, if the channel is first estimated by the LLMSE filter, then the AR models can be estimated after a time window corresponding to the user having moved a few tens of the carrier wavelength, see [42] for details.

For every new measurement (11), the filter can recursively compute the channel estimate through

$$\begin{aligned}\hat{x}(\tau|\tau) &= A\hat{x}(\tau-1|\tau-1) + \mathcal{K}(\tau)(\bar{y}(\tau)\\ &\quad + J(\tau)A\hat{x}(\tau-1|\tau-1)),\end{aligned} \tag{12}$$

$$\begin{aligned}P(\tau|\tau) &= AP(\tau-1|\tau-1)A^* + BQB^*\\ &\quad - \mathcal{K}(\tau)J(\tau)(AP(\tau-1|\tau-1)A^* + BQB^*),\end{aligned} \tag{13}$$

$$\hat{h}(\tau|\tau) = C\hat{x}(\tau|\tau), \tag{14}$$

where, $J(\tau) = \Phi(\tau)C$, $P(\tau|\tau) = E[(\bar{x}(\tau) - \hat{x}(\tau|\tau))(\bar{x}(\tau) - \hat{x}(\tau|\tau))^*$ is the covariance matrix of the state vector estimation error and the matrix $\mathcal{K}(\tau)$, known as the Kalman filter gain, is obtained through

$$\mathcal{K}(\tau) = P(\tau|\tau-1)J(\tau)^*(J(\tau)P(\tau|\tau-1)J(\tau)^* + R_v)^{-1}. \tag{15}$$

If the state transition matrix A is set to an all zero matrix (reflecting that we have no information of the temporal correlation) and we set $h_{\text{rel}} = h$, then the estimate (14) will coincide with the LLMSE estimate (9).

The Kalman equations (12)-(14) require initial values of the estimate $\hat{x}(\tau|\tau)$ and of the corresponding error covariance matrix. These are here set to $\hat{x}(0|0) = 0$ and $P(0|0) = E[\bar{x}(\tau)\bar{x}^*(\tau)]$.

Provided that the RS matrix is cyclic with $\Phi(\tau) = \Phi(\tau+\mu)$, the filter will converge to a cyclo-stationary filter with $P(\tau|\tau-1) = P(\tau+\mu|\tau+\mu-1)$ within a few cycles. Then, (13) can be calculated off-line by solving a Riccati equation, see [40] for details.

1) Estimation of the Process Noise Covariance Matrix Q: The frequency correlation and the spatial/antenna correlation between channel components is modelled through the covariance matrix Q of (10). Estimating this process noise covariance matrix can be complicated. For the ideal case where the state model is on diagonal form and perfectly models the time dynamics of the channel, then it can be shown, [40], that

$$Q = R_h \oslash C(B\mathbf{1}B^* \oslash (\mathbf{1} - \bar{a}\bar{a}^*))C^*, \tag{16}$$

where $R_h = E[\bar{h}(\tau)\bar{h}^*(\tau)] = CE[\bar{x}(\tau)\bar{x}^*(\tau)]C^*$, $\bar{a}$ is a column vector containing the diagonal elements of A, $\oslash$ denotes elementwise division, and $\mathbf{1}$ is a matrix of ones. This works also when using imperfectly estimated state space matrices in a set of special circumstances. These include the case when all channel components can be assumed to be identically distributed, e.g. MIMO channels as in [40] and when the channel components are uncorrelated, e.g. for different site antennas as in [42].

However, the channel components defined by (1) are in general neither identically distributed nor uncorrelated. Nor can we expect our estimates of the state space matrices to perfectly fit the data. Under such general conditions, the solution to (16) may provide an estimate of the process noise covariance matrix Q, which is non-positive definite. Such an error will destroy the convergence of a Kalman filter.

In order to ensure a positive definite matrix, we may instead approximate Q by using the pseudo inverse. There are various ways to do this. We have obtained good results by using

$$\begin{aligned}Q &\approx B^\dagger(\Pi + A\Pi A^*)(B^*)^\dagger,\\ \Pi &\approx C^\dagger R_h(C^*)^\dagger,\end{aligned} \tag{17}$$

where $\Pi = E[\bar{x}(\tau)\bar{x}^*(\tau)]$. Equation (17) follows directly from the state space model, and will provide a channel vector with similar statistical properties as the real channel matrix, however it may not be the best estimate.

An alternative is to define an upper triangular matrix M and form a Q as

$$Q = M^*M, \tag{18}$$

which by definition will be positive semidefinite. The non-zero elements of M can then be optimized for a given criterion, e.g. minimizing the MSE of the channel estimate. Most such optimization criteria will be non-convex and there is a risk that an optimization algorithm will find a local minimum as opposed to a the global minimum.

These alternatives are compared in Appendix, and based on those results, the estimate (17) has been used in the channel estimation performance investigation of Section IV-B.

C. Reduced Complexity Kalman Filter

Channel estimation through (12)-(14) provides the optimal (linear) estimate, but the on-line complexity grows with the square of the number of channel components [40]. Also, for a large number of channel components, the off line complexity related to solving the Ricatti equation may make Kalman filtering infeasible. In order to reduce complexity, we can choose to estimate only the relevant channel components. We therefore introduce a reduced state space model

$$\begin{aligned}\bar{x}_{\text{rel}}(\tau+1) &= A_{\text{rel}}\bar{x}_{\text{rel}}(\tau) + B_{\text{rel}}\bar{u}_{\text{rel}}(\tau),\\ \bar{h}_{\text{rel}}(\tau) &= C_{\text{rel}}\bar{x}_{\text{rel}}(\tau),\\ Q_{\text{rel}} &= E[\bar{u}_{\text{rel}}(\tau)\bar{u}_{\text{rel}}(\tau)^*],\end{aligned} \tag{19}$$

TABLE I
SIMULATION PARAMETERS USED IN THE QUADRIGA CHANNEL SIMULATOR, SEE [48]

Scenario	WINNER_UMa _C2 _NLOS [49]
Carrier frequency	2.53 GHz
Subcarrier spacing	15 kHz
# subcarriers	144
RS spacing	5 ms
Base station hight	32 m
Antenna tilt	$-8°$
# antennas/base station	32
Antenna spacing	0.5 wavelengths

which is similar to (10), with exceptions of the dimensions. The measurement is then

$$\bar{y}(\tau) = \Phi_{\text{rel}}(\tau)\bar{h}_{\text{rel}}(\tau) + \bar{v}(\tau) + \bar{w}(\tau), \tag{20}$$

where $\Phi_{\text{rel}}(\tau)$ is the sub-matrix of the RS matrix $\Phi(\tau)$ that consists of the column vectors corresponding to the relevant channel components. Equation (20) differs from (2) as it includes an additional noise term $\bar{w}(\tau)$. This is the contribution of the non-relevant channel components

$$\bar{w}(\tau) = \Phi_{\bar{\text{rel}}}(\tau)\bar{h}_{\bar{\text{rel}}}(\tau). \tag{21}$$

The noise vector $\bar{w}(\tau)$ is not white and we define a correlation matrix

$$\begin{aligned} T(t) &= E[\bar{w}(\tau)\bar{w}^*(\tau - t)] \\ &= E[\Phi_{\bar{\text{rel}}}(\tau)\bar{h}_{\bar{\text{rel}}}(\tau)\bar{h}^*_{\bar{\text{rel}}}(\tau - t)\Phi^*_{\bar{\text{rel}}}(\tau - t)]. \end{aligned} \tag{22}$$

To reduce complexity we may simply approximate the extra noise term as i.i.d. with correlation matrix $T = E[\bar{w}\bar{w}^*]$. With this approximation we can use the model (19) directly in the Kalman filter (12) by replacing A, B, C, Q and R_v by A_{rel}, B_{rel}, C_{rel}, Q_{rel} and $R_v + T$ respectively. This would cause information loss, but as we shall see in the simulation section, this still provides good channel estimates.

IV. EVALUATION BY SIMULATION

To validate our concept we set up a system level simulation using the Matlab based, open source, Quadriga channel simulator, developed by the Fraunhofer Heinrich Hertz Institute [47]. Three base station sites spaced by 500 m, each with three sectored base stations, were used and define a cooperation cluster. A number of 100 individual users were randomly placed within a circle with a 500 m radius centred at the cluster center, providing 100 potential user positions. For these users, channels were then generated while the users moved for 29 seconds with a velocity of 3 km/h, using the settings defined in Table I. Other settings were set to the default values in the Quadriga channel generator, see [48].

A. Relevant Channel Components

The beamforming weights in (1) are set to $b(n, l) = \exp(2j\alpha_n l/\sqrt{8})$ where j is the imaginary unit and $\alpha_n = (67.5 - 15n)\pi/180$ for $n = 0, \ldots, 7$, thus forming eight horizontal beams per base station, yielding a total of $8 \cdot 3 \cdot 3 = 72$ channel components per subcarrier.

To achieve beamforming gains and/or mitigate interference, the power ratio between the strongest channel component and the other channel components is of importance. Figure 3 shows the CDF of the number of channel components that would be relevant if a transmit scheme utilizes only channel components with power above a threshold relative to the the strongest channel component.

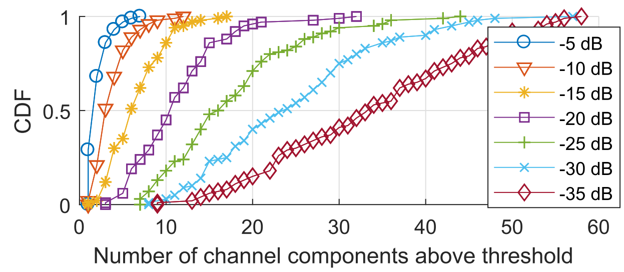


Fig. 3. CDF of the number of relevant channel components (CCs) at different user positions when using different thresholds (in dB) relative to the power of the strongest channel.

In [15] and also in [50] and [51], the use of a threshold of 20 to 25 dB was shown to provide good CoMP performance through interference mitigation. Assuming a threshold of 20 dB, results in Figure 3 indicate that 15-20 channel components would then need to be estimated.

A more lightly loaded system, in which interference mitigation is of less importance, would need fewer relevant channel components per user, as will be shown in Section IV-B.3. In this paper we focus on the noise limited scenario and choose to focus on 16 relevant channel components which is a conservative choice. For interference limited scenarios we refer the reader to [26].

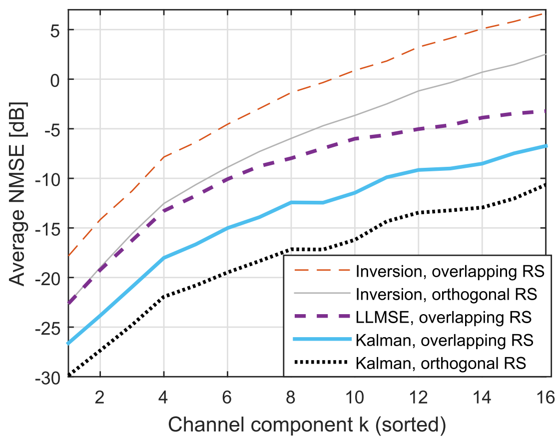
B. Channel Estimation Performance

We assume a RS structure with resource blocks of 90 kHz $\times$ 1 ms (six subcarriers $\times$ 14 OFDM symbols á 71 μs). Every 5 ms, a set of $K = 18$ RSs are transmitted within each subband over three subsequent, identical fading OFDM symbols (giving an overhead of 4.3%). The channels are frequency selective, the normalized correlation between the fading channels at each edge of the 90 kHz resource blocks is around 0.9-0.98.

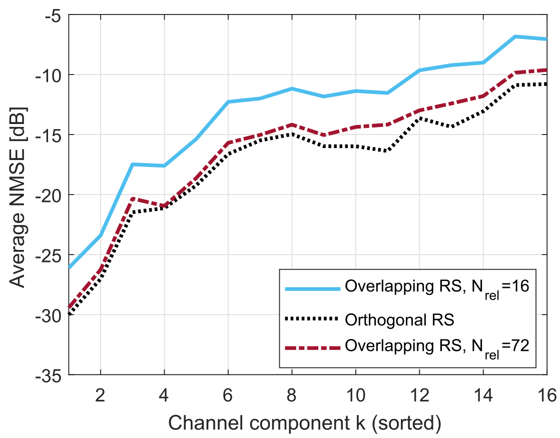
The channel components are assigned cyclic RS codes using (7)-(8) with unit power and a RS cycle of $\mu = 3$ and $\phi(1) = 1$, $\phi(2) = 2$, $\phi(3) = 3$. A measurement signal was simulated through (2), with i.i.d. circular symmetric Gaussian noise $v(k, \tau)$ with the covariance matrix $R_v = 10^{-12} \cdot I$. The resulting SNR is then in the range of of 8-43 dB for the strongest channel component, depending on the user position.

For each user, the time series is divided into two parts of 24 and 5 second respectively. The channel statistics was estimated during the first part.

Based on sparsely transmitted orthogonal RS (every 500 OFDM symbol), each user also finds a set of 16 relevant channels. Here the number 16 is fixed, the number of relevant channels is here not determined by a power threshold. These relevant channel components were then estimated using a) pseudo inversion of the reduced RS matrix (6), b) the LLMSE estimate (9) and c) the Kalman filter estimate (14), based on the reduced model (19).



(a) 100 user positions, $N_{rel} = 16$.



(b) Ten user positions.

Fig. 4. Average NMSE sorted after the RS strength of the channel component.

We use the approximation that the noise term $\bar{w}(\tau)$ in (20) is i.i.d. over time. In total 24 subbands of 90 kHz each (6 subcarriers) are tracked by parallel Kalman filters. In (19), each channel component is modelled by a 4:th order AR model. The process noise covariance matrix Q in (10) is estimated through (17). In Appendix, we evaluate this choice of Q and compare it to an optimized solution based on (18).

For comparison, and as a lower bound, Kalman estimates using fully orthogonal RSs are also presented. In that set-up, each of the 72 beams is assigned RSs for all 144 subcarriers for one out of 72 subsequent OFDM symbols. The per RS symbol power is here set such that the total RS power budget is equal to that of the overlapping references signals. This is an unrealistic set-up, as it would cause an infeasible overhead.[4]

1) Estimation Performance: Figure 4a shows the resulting NMSE as a function of the channel component number, averaged over subcarriers and users. Here, the inversion through (6) with orthogonal RSs represents the best we can do when no correlation is utilized. Comparing this with using overlapping RSs, we may note that the price of reducing the RS overhead from 100% to only a few percent is a loss of 5 dB in estimation performance. The same loss can be seen when comparing estimations of the the Kalman filter when overlapping and orthogonal RSs respectively are used.

While channel estimation by (6) may be sufficient for the strongest channel components, it quickly degrades for weaker channels. Comparing channel inversion through (6) to the LLMSE filter, we see that by utilizing the space and frequency correlations, the estimates are greatly improved, especially for the weaker channel components.

A further improvement of approximately 5 dB can be achieved by utilizing temporal correlation by introducing the Kalman filter, which is especially important for highly loaded system that require interference mitigation.

Most of the previously suggested approaches to channel estimation for FDD massive MIMO optimize RSs for a specific set of users, see Section I-B. To relate our results to those methods, we can compare the channel estimates for the overlapping RSs and for the orthogonal RSs when the Kalman filter is used. In an extreme situation, where the union of the sets of relevant channel components of the scheduled users includes no more than K (here 18) channel components, then an optimization of RSs would result in the estimation performance of the orthogonal RSs (the optimized RSs would then be orthogonal).

In a more realistic situation, where the union of the sets of relevant channel components increases with an increasing number of scheduled users, the estimation performance would move towards that of reduced Kalman estimation with overlapping RSs. Hence, our flexible solution, which does not require RSs to be re-optimized every time a new user is scheduled, should in the worst case scenario result in a 5 dB estimation performance degradation, as compared to cases with optimized RS.

2) Effects of Using a Reduced Kalman Filter: Figure 4b illustrates how much we loose by only estimating the $N_{rel} = 16$ relevant channel components by the reduced Kalman filter with the model (19), rather than estimating all channel components ($N_{rel} = 72$) by the model (10)-(11). As the off-line complexity related to calculating the covariance matrix $P(t|t)$ in (13) through solving the Ricatti equation grows with N_{CC}^3, this investigation is only performed for ten of the 100 user positions.

We can see that when all channel components are estimated, based on overlapping RSs with 4.3 % RS overhead, then the 16 strongest channel components can be estimated almost as well as if orthogonal RSs were used. Parts of the gain from estimating all channel components could likely be achieved by including the interference term $w(t)$ in (20) in the state vector of the reduced Kalman estimator. Such investigations are left for future work.

3) Capacity When Using MRC Beamforming: The effect of the estimation errors in terms of MRC beamforming gain is illustrated in Figure 5. The results in this figure are based on the average values over all users, i.e. we assume that the estimation performances are given by the average Kalman estimation NMSE presented in Figure 4a. As an estimate of the corresponding the SNR per channel component, we use the inverse of the NMSE achieved by inversion with orthogonal

[4]For example if we assume an OFDM-symbol duration of 71μs then we would need to transmit RSs for 5.1 ms of the total 5 ms interval leading to an overhead of 102 %.

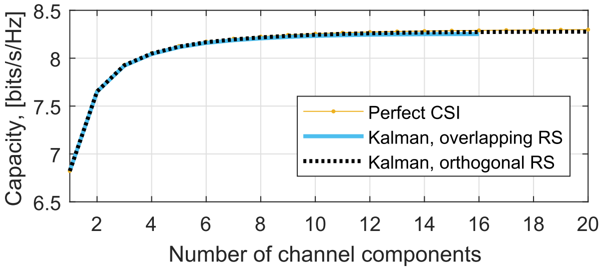


Fig. 5. An illustration of the single users maximum ratio beamforming gain, based on Shannon capacity, as a function of the included number of channel components. CSI accuracy is based on the Kalman estimates in Figure 4a. Note that the curves overlap.

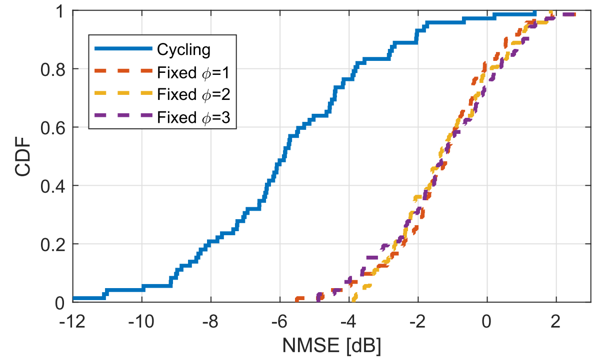


Fig. 6. Comparison of the NMSE distributions of channel estimates based on cycling and non-cycling (fixed) RSs, as in Figure 2. The statistics represent the 72 i.i.d. channels.

RSs (the thin solid line in Figure 4a). The resulting Shannon capacity show in the figure, is based on maximum ratio transmit beamforming to one single user, by combining the strongest fixed beams.

From Figure 5 we see that there is next to no beamforming gain from using orthogonal RSs for our particular scenario. The reasons for this are as follows: First, the estimation quality is already very good for the strongest channel components when using Kalman estimation based on overlapping RS and an extra gain in accuracy will only translate into a very small capacity gain. Second, MRC beamforming gains are robust to estimation errors. Third, as capacity grows logarithmically with SNR, adding extra channel component, that have low SNR as compared to the strongest channel component, to the beam provides very little extra gain; note the saturation of the curves in Figure 5.

A separate study of interference supression by regularized zero forcing precoding in a CoMP cluster with massive MIMO antennas under a similar simulation set-up as here, but with 288 beams, is presented in [26]. Each beam is then treated as one antenna element of a regular low order MIMO system and the channel components rather than the channels of each antenna set up the channel matrix. This study also includes a beam deactivation algorithm to ensure that no RS power would be wasted on beams that are weak to all users within the system. The results presented there, with 10 % RS overhead, show that performance loss, in terms of payload spectral efficiency, was approximately 12% compared to when perfect CSI was used.

4) Effect of Using Cycling Reference Signals: A small separate experiment focuses on the effects of the use of cycling RSs. To more clearly distinguish the contribution of cycling RS, we here generate a set of channel components that all have the same average SNR. For this purpose, 72 i.i.d. flat fading channels with first order AR statistics were first generated, using filters with a single pole in $0.999e^{0.1j}$ (generating very slow time variations). Separate measurements were then created, through (2), (7) and (8) as in Figure 2 in Section II-A.3 using $K = 18$ RS carrying resources, by using cycling reference signals with $\phi(t) = \{1, 2, 3\}$ and non-cycling (fixed) reference signals with $\phi(t) = 1$, $\phi(t) = 2$ and $\phi(t) = 3$. White Gaussian noise that would correspond to an average SNR of 10 dB for orthogonal RSs was added. All channels ($N_{rel} = 72$) were then estimated using a Kalman filter that was based on accurate estimates of the AR statistics, the SNR and the (flat-fading) subcarrier correlations. The resulting NMSE (averaged over 1000 samples, which correspond to time-series with approximately 3 fading dips) are shown in Figure 6. We here see a clear improvement in the estimation accuracy from using the cycling reference signals versus any of the fixed reference signal patterns.

V. Conclusions

We have here proposed a joint reference signal design and channel estimation scheme that enables sufficiently accurate channel estimation for massive MIMO gains in FDD systems. Our solution begins with introducing fixed beams, over which RSs are transmitted. These break up the i.i.d. statistics of channels for the same base station, and we used this to our advantage by only estimating a subset of relevant channel components. The estimated relevant channel components can then be used for precoding, e.g. through MRC or zero forcing for data transmission.

Channel estimation performance was evaluated using a Kalman filter and an LLSME filter. The Kalman filter comes with the added benefit that channel predictions are straightforward to implement, but requires high complexity off-line calculations. At a user speed of 3 km/h, we obtained significant differences between the two estimation algorithms, which can be especially important in interference limited scenarios. We have shown that with a full order Kalman filter based on overlapping RSs, we achieve almost as good estimation performance as with orthogonal RSs for the relevant channel components.

We have also introduced a reduced Kalman filter which only estimates the relevant channel components, assuming that the interference from the other channel components can be regarded as noise with a time-independent covariance matrix. This comes at a 5 dB performance loss in estimation NMSE in our evaluation scenario, but estimations are still sufficiently accurate to ensure almost the full capacity gain provided by MRC beamforming.

As different users will have different sets of relevant channels, it is of importance that the RS code design ensures that no user positions will result in estimation performance loss due to poorly conditioned RS matrices. We have here shown that well conditioned RS matrices can easily be constructed for a majority of the users, but that some will still experience badly conditioned submatrices. However, we have also shown that this problem can be essentially eliminated, *without* having to adjust RS patterns to users, by introducing cycling RSs.

Results on how the channel estimation performance would translate into sum-rate or some other end-performance, when interference mitigation schemes are used, is a subject for future investigations. The proposed scheme has recently been used to advantage in a large set-up with 288 beams including a beam deactivation scheme, in [26].

Open Issues: Our results showed a significant gap in estimation performance between the reduced Kalman filter and that which estimated all channel components. This gap may be reduced if the interfering noise terms $\bar{w}$ generated by the non-relevant channels are tracked along with the relevant channels.

A source of estimation errors stems from estimating the covariance matrix of the process noise. Given limited channel data and a large set of channels this may prove difficult. In [52] a regularization term is utilized to improve covariance matrix estimation. Such regularization can often be derived through a Bayesian estimation approach with some prior information [53]. A natural extension of the work presented here is to use such an approach to improve the accuracy of the covariance matrices and thereby the channel estimations.

Another natural extension of the work presented here is to investigate Kalman channel predictions, as these are needed for CoMP in combination with massive MIMO.

Appendix
Estimation of the Process Noise Covariance Matrix

To validate our choice of calculating the covariance matrix of the process noise Q through (17) we here compare this to an alternative where Q is instead calculated through finding a triangular matrix M that relates to Q by (18).

The matrix M is here iteratively optimized using the interior point method.[5] For each new M, the covariance matrix Q is calculated and used in the Kalman filter by (14). We optimize M based on minimizing the resulting NMSE, which is calculated over the 1000 time sample in the evaluation interval.

For each new M the Riccati equation must be solved. This is a time consuming process, so to keep the complexity low we have for the purpose of optimizing Q set the number of relevant channels to six and only perform the evaluation for two user positions. We also lower this complexity by using time fixed RS by (8) with $\phi(\tau) = 1$ for all τ.

As the optimization problem is non convex, there are various values of M that will result in a low minimum. To find as many of them as possible we repeat the optimization problem for different starting values of M. These are

- Unit — The initial value of M is set to a unit matrix.
- Rand — The initial value of M is set to an upper triangular matrix whose elements are drawn from a random Gaussian distribution of complex numbers with unit variance.

[5]Here, Matlab's function fmincon is used to find the optimum. The diagonal elements of M are constrained to positive numbers and all other non-zero elements are unbounded.

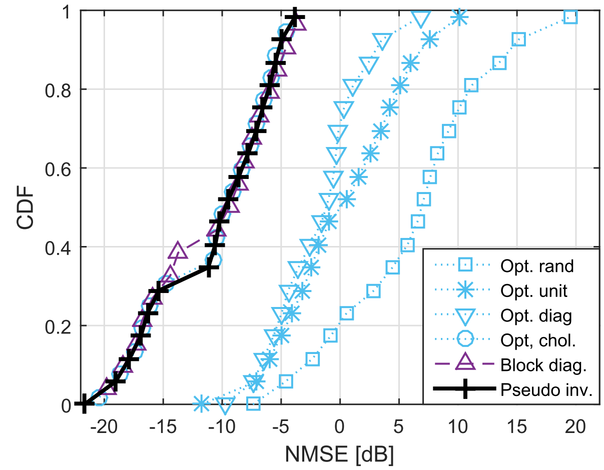


Fig. 7. The CDFs of the NMSE provided by the reduced Kalman filter using estimates of the process noise matrix Q through (17)), denoted "Pseudo inv.", through optimiziation of an upper triangular matrix M in (18), denoted "Opt.", with different initial values for M, and by using a block diagonal channel covariance matrix R_h to find a block diagonal Q through (16), denoted "Block diag.".

- Diag — The matrix Q is first calculated through (17). Then the initial value of M is a diagonal matrix whose diagonal elements are given by the squared root of the diagonal elements of Q.
- Chol — The matrix Q is first calculated through (17). Then the initial value of M is found through Cholesky decomposition of Q.

Figure 7 shows the resulting NMSE as a CDF over the 144 subcarriers with different initial values of M. For comparisons there is also an option in which the channels from different beams were assumed to be uncorrelated by setting the covariance matrix of the channel R_h to a block diagonal matrix and then calculate Q through (16), providing a block diagonal covariance matrix Q. The results show that the initial values of M have a significant impact on the NMSE. As the optimization algorithm is very slow, it is an infeasible option for any realistic scenario. However, we see that both the pseudo inverse and the block diagonal versions of Q provide a low NMSE and can be used successfully for these kinds of data. They provide NMSE performance close to that of an optimized $Q = MM^*$ with the best performing initial value choice ("Chol") for M.

When studying the cross correlations between the channels in further detail, it was clear that the cross correlation between different subcarriers of the same beam had a cross correlation above 0.9 while the channels that belonged to different beams had a cross correlation of less than 0.25. While the cross correlation between the beams is still significant, it does give the channel covariance matrix a block diagonal dominant structure, which may be why the block diagonal structure works so well. Why the pseudo inverse works so well is difficult to say, but to date, we have not been able to find any option that works significantly better.

References

[1] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[2] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. 14th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct./Nov. 2006, pp. 359–363.

[3] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, 2015.

[4] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[5] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[6] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE VTS 53rd Veh. Technol. Conf.*, Rhodes, Greece, May 2001, pp. 1745–1749.

[7] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2002, p. 50.

[8] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[9] V. Jungnickel *et al.*, "ojd, and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[10] Z. Ma, Z. Zhang, Z. Ding, P. Fan, and H. Li, "Key techniques for 5G wireless communications: Network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, no. 4, pp. 1–20, 2015.

[11] "5G energy performance," Ericsson, Stockholm, Sweden, White Papers 28423-3265, Apr. 2015.

[12] L. Lu, G. Y. Li, A. Lee Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[13] D. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[14] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 65–71, Jun. 2012.

[15] V. D'Amico *et al.*, "Interference avoidance techniques and system design," EU Project Adv. Radio Interface Technol. 4G Syst., ARTIST4G Project Report D1.4, Tech. Rep. EU FP7-ITC-2009, 2012.

[16] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[17] *Technical Specification Group Radio Access Network; Coordinated Multi-Point Operation for LTE Physical Layer Aspects, (Release 11)*, document 3GPP TR 36.819, 2011.

[18] L. Thiele, M. Olbrich, M. Kurras, and B. Matthiesen, "Channel aging effects in CoMP transmission: Gains from linear channel prediction," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 1924–1928.

[19] R. Apelfröjd and M. Sternad, "Design and measurement-based evaluations of coherent JT CoMP: A study of precoding, user grouping and resource allocation using predicted CSI," *EURASIP J. Wireless Commun. Netw.*, vol. 2014, no. 1, p. 100, Jun. 2014.

[20] R. Apelfröjd and M. Sternad, "Robust linear precoder for coordinated multipoint joint transmission under limited backhaul with imperfect CSI," in *Proc. 11th Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014, pp. 138–143.

[21] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.

[22] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[23] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *Proc. IEEE 60th Veh. Technol. Conf.*, Los Angeles, CA, USA, Sep. 2004, pp. 4265–4269.

[24] "5G radio access," Ericsson, Stockholm, Sweden, White Papers 28423-3204, Apr. 2016.

[25] W. Zirwas, M. B. Amin, and M. Sternad, "Coded CSI reference signals for 5G—Exploiting sparsity of FDD massive MIMO radio channels," in *Proc. 20th Int. ITG Workshop Smart Antennas (WSA)*, Munich, Germany, Mar. 2016, pp. 1–8.

[26] W. Zirwas, M. Sternad, and R. Apelfröjd, "Key solutions for a massive MIMO FDD system," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Montreal, QC, Canada, Oct. 2017, pp. 1–7.

[27] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[28] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[29] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.

[30] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "On the achievable rates of FDD massive MIMO systems with spatial channel correlation," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Shanghai, China, Oct. 2014, pp. 276–280.

[31] C.-C. Tseng, J.-Y. Wu, and T.-S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block ℓ_1-minimization," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1055–1067, Mar. 2016.

[32] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 601–617, Feb. 2015.

[33] B. Tomasi, A. Decurninge, and M. Guillaud, "SNOPS: Short non-orthogonal pilot sequences for downlink channel state estimation in FDD massive MIMO," in *Proc. IEEE Globecom Workshops*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[34] M. Kurras, L. Thiele, T. Haustein, W. Lei, and C. Yan, "Full dimension MIMO for frequency division duplex under signaling and feedback constraints," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Budapest, Hungary, Aug./Sep. 2016, pp. 1985–1989.

[35] T. Ekman, "Prediction of mobile radio channels: Modeling and design," Ph.D. dissertation, Uppsala Univ., Uppsala, Sweden, Oct. 2002. [Online]. Available: http://www.signal.uu.se/Publications/ptheses.html

[36] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive MIMO performance—TDD versus FDD: What do measurements say?" *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2247–2261, Apr. 2018.

[37] J. Choi, K. Lee, D. J. Love, T. Kim, and R. W. Heath, Jr., "Advanced limited feedback designs for FD-MIMO using uniform planar arrays," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[38] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[39] R. Apelfröjd, W. Zirwas, and M. Sternad, "Joint reference signal design and Kalman/Wiener channel estimation for FDD massive MIMO," Signals Syst. Group, Dept. Eng. Sci., Uppsala Univ., Uppsala, Sweden, Tech. Rep. R1707, 2018. [Online]. Available: http://www.signal.uu.se/Publications/pdf/r1701.pdf

[40] D. Aronsson, "Channel Estimation and Prediction for MIMO OFDM Systems: Key design and performance aspects of Kalman-based algorithms," Ph.D. dissertation, Dept. Eng. Sci., Uppsala Univ., Uppsala, Sweden, Mar. 2011. [Online]. Available: http://www.signal.uu.se/Publications/ptheses.html

[41] D. Aronsson and M. Sternad, "Kalman predictor design for frequency-adaptive scheduling of FDD OFDMA uplinks," in *Proc. IEEE 18th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Athens, Greece, Sep. 2007, pp. 1–5.

[42] R. Apelfröjd, "Kalman predictions for multipoint OFDM downlink channels," Signals Syst. Group, Dept. Eng. Sci., Uppsala Univ., Uppsala, Sweden, Tech. Rep. C1402, 2018. [Online]. Available: http://www.signal.uu.se/Publications/pdf/c1402.pdf

[43] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[44] L. Lindbom, "Simplified Kalman estimation of fading mobile radio channels: High performance at LMS computational load," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Pacific Grove, CA, USA, Apr. 1993, pp. 352–355.

[45] C. Komninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.

[46] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunström, "Towards systems beyond 3G based on adaptive OFDMA transmission," *Proc. IEEE*, vol. 95, no. 12, pp. 2432–2455, Dec. 2007.
[47] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.
[48] Fraunhofer Heinrich Hertz Institute. (2016). Quasi deterministic radio channel generator user manual and documentation. [Online]. Available: http://quadriga-channel-model.de/
[49] J. Meinilä *et al.* (2010). *Celtic/cp5-026 D5.3: WINNER + Final Channel Models*. [Online]. Available: http://docplayer.net/34577643-D5-3-winner-final-channel-models.html
[50] A. M. Ahmadian, W. Zirwas, R. S. Ganesan, and B. Panzner, "Low complexity Moore-Penrose inverse for large CoMP areas with sparse massive mimo channel matrices," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Sep. 2016, pp. 1–7.
[51] R. S. Ganesan, W. Zirwas, B. Panzner, K. I. Pedersen, and K. Valkealahti, "Integrating 3D channel model and grid of beams for 5 Gm MIMO system level simulations," in *Proc. IEEE 84th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2016, pp. 1–6.
[52] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 815–830, Oct. 2014.
[53] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

Rikke Apelfröjd received the M.S. degree in engineering physics and the Ph.D. degree in electrical engineering from Uppsala University, Sweden, in 2011 and 2018, respectively. Her Ph.D. thesis Channel Estimation and Prediction for 5G Applications was supervised by Prof. M. Sternad and co-supervised by Prof. A. Ahlén. Her research interests include channel estimation, coordinated multipoint transmission, prediction antennas, and 5G communication.

Wolfgang Zirwas received the Diploma degree in communication technologies from the Technical University of Munich in 1987. He joined the Siemens Central Research Lab for Communication, Munich, in 1987. Since 1996, he has been contributing to and partially also lead national and EU projects, like ATMmobil, COVERAGE, 3GeT, ScaleNet, WINNER, ARTIST4G, METIS, and Fantastic5G. He is currently a part of Nokia Bell Labs, where he is in research with focus on the physical layer for 3GPP new radio systems. He has authored and co-authored 120 conference and journal papers, contributed to four books, and more than 200 patent applications. His research interests include multi-hop, massive MIMO, cooperative multipoint transmission, and OFDM. Since 2010, his research interest is on interference mitigation for cellular mobile radio 5G networks and its related enablers like channel estimation and prediction. He served in the TCP of European Wireless Conference, PIMRC, WSA, and WDN-CN2012. He served as an Associate Editor for the *Scientific World Journal*.

Mikael Sternad received the Ph.D. degree in automatic control from Uppsala University in 1987. He is currently a Professor in automatic control with the Department of Engineering Sciences, Uppsala University, Sweden.

He has acted as a Project Leader of the National 4G Research Project Wireless IP and the Swedish Research Council Framework Project Dynamic Multipoint Transmission, and several VINNOVA projects. He was engaged in the EU WINNER which formed the basis for the 4G wireless standardization effort, the Artist4G Project, and the EU FP7 Project METIS. His main research interests is signal processing applied to mobile broadband communication problems, such as channel prediction schemes for fast link adaptation, and scheduling and coordinated multipoint transmission.

He is also working on sound field control, acoustic zones and personal audio, 3D sound and recording, and rendering for virtual reality applications. From 2001 to 2005, he was a Chairman of Dirac Research AB which is active in these fields, where he is also a Co-Founder. His research interest is robust and adaptive MIMO feedforward control algorithms. These can be applied to wireless transmission (network MIMO, or CoMP), and to sound field control using multiple loudspeakers.