Report version of

Joint Reference Signal Design and Kalman/Wiener Channel Estimation for FDD Massive MIMO

Rikke Apelfröjd[1], Wolfgang Zirwas[2] and Mikael Sternad[1]

[1] *Signals and Systems, Uppsala University,* [2] *Nokia Bell Labs, Munich, Germany.*
Contact: Box 534, 751 21 Uppsala, Sweden, {rikke.apelfrojd, mikael.sternad}@signal.uu.se

This is an extended report version of an article which is currently under peer review for IEEE Transaction on Communications

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, non-orthogonal 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 can also utilize 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 the non-orthogonal coded reference signals is found to be important for reliable channel estimation, without having to adjust the reference signals to specific users.

1 Introduction

Multiple input multiple output (MIMO) downlink transmission techniques that serve multiple users have been introduced in LTE 4G [1] and are becoming increasingly important in the study of future systems. As these transmit schemes rely on channel state information (CSI) at the transmitter [2, 3], channel estimates of sufficient accuracy become crucial.

For the past decade, evolution of multiuser MIMO has moved in two main directions: massive MIMO [4–8] and coherent coordinated multipoint (CoMP) joint transmission (JT) (also known as network MIMO) and coordinated beamforming [9–11]. Each of these, and combinations of them, have been identified as key enablers for the fifth generation mobile system [12–15]. The channel estimation challenges of downlinks in such scenarios, with a very large number of antennas and radio channels, have motivated our present work.

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, [16–19]. Delays in the fixed network cause outdated CSI, which can severely reduce gains [20, 21]. However, channel prediction in combination with robust precoding has shown promising results [22, 23].

In massive (or large-scale or full dimension) MIMO, the number of transmit antennas at a site is very large. This leads to several advantages. In the special case when the number of simultaneously scheduled users is much smaller (by a factor >10) than the number of transmit antennas, channel vectors to different users will be almost orthogonal with high probability. Each user will also experience a large linear beamforming gain from maximum ratio combining (MRC) [6], provided that a constant CSI quality can be ensured. More users can be added to optimize system performance, e.g. the sum throughput [24], but this may come at the cost of cell edge performance. A combination of massive MIMO and coherent JT CoMP has the potential to increase sum throughput without sacrificing cell edge coverage.

Massive MIMO requires CSI for channels to a vast number of antennas. Adding coherent JT CoMP to the framework would increase this requirement further. A main challenge with massive MIMO in frequency division duplex (FDD) systems is therefore to avoid a massive downlink reference signal (RS) overhead. To acheive this, RS sequences for downlink channel estimation must be overlapping in the time-frequency domain (non-resource orthogonal) and hence suboptimal [25].

Many researchers therefore instead focus on time division duplex (TDD) systems. There, RSs could be transmitted from the scheduled users in uplink time slots who 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. Although channel estimation in TDD is limited due to imperfections in channel reciprocity, limited transmit power at the user, hardware impairments and lack of downlink interference estimates [7, 15, 24, 26], it has great potential as illustrated in [6]. There, a comparison of the plausible operation conditions of FDD and TDD in massive MIMO assuming *orthogonal* RSs concludes that TDD is more beneficial than FDD. These results are based on approximations, but the comparison gives a fair picture in a qualitative sense.

However, there is one important argument for why we need to solve the problem of using non-orthogonal 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. In [8], the authors 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 [27]. 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.

1.1 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 is typically 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 gain, then each user only has to estimate 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 sites 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.

In addition to this, estimation algorithms that utilize the correlation over time, space and frequency are used to improve CSI.

Our proposed framework has four main components:

1. Antennas will be structured into a fixed *grid of beams*, where each beam is wideband and controlled by an ef-

fective or virtual antenna port[1]. 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 non-resource orthogonal 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 *cycling sequences of RS codes* ensures good estimations regardless of the users position, by weighted time-averaging estimation errors that are caused by RS non-orthogonality.

This work is an extension of the RS design and channel estimation introduced in [28]. We here extend the solution in [28] beyond that of flat block-fading channels by first utilizing an LLMSE estimation. Second, we introduce a Kalman filter estimate that use 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 More-Penrose Pseudo inverse to estimate covariance matrices is a good choice. Furthermore, we add cyclic RSs to the framework in [28] and show that this is important in the aspect of user fairness.

1.2 Related work

Channel estimation for massive MIMO in FDD has recently gained interest [29–36]. Similar to our design, these works assume non-resource orthogonal RSs and utilize some type of correlations to improve the estimates. In contrast to our design, the works of [29–35] 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 [29, 30], user specific RS design 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, is proposed in [31]. These concepts would demand user specific RS resources, so the overhead increases with the number of users and the benefit achieved by using non-orthogonal RSs decreases rapidly as the number of users increase.

Authors of [32–34] instead optimize the RSs off-line to improve the average channel estimation for the scheduled users. The required RS overhead would in [32, 33], 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 [34] 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 [37].

These multi-user methods, though they are an improvement compared to the single user case, still require RS re-optimization when new users are scheduled. Our solution introduce 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 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 compared to the TDD scenario, where only the scheduled users can be allowed to transmit RS in order to limite RS overhead.

A somewhat related idea is proposed in [36] 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.

1.3 Layout and notations

Section 2 provides details on the fixt grid of beams and the RS codes that are the key design elements to our solution. Some comments on how intoducing a fixed grid of beams may limit the end performance of massive MIMO and on how our RS code design can be used to improve channel estimates in TDD can be found in Appendices A and B respectively. The estimation algorithms used for evaluations are described in Section 3, with some details on the statistical modelling in Appendix C. Simulations are provided in Section 4, with some supporting results in Appendix D. Section 5 highlights conclusions and suggests areas for further investigation.

[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.

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 $|\mathcal{A}|$.

2 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 is a total of $N_{PRC} = N_{BS} \cdot N_{tx}$ physical radio channels between the serving antennas and each single antenna user[2].

We use sets of K reference symbols, where we may have $K \ll N_{BS} \cdot N_{tx}$. For example, these could be the $K = 40$ CSI RS time-frequency recources that are allocated within each resource block bandwidth of 180 kHz every $T_s \geq 5$ ms in LTE rel. 10 [1]. The K RSs are transmitted in the downlink with a period of T_s, indexed by τ.

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 $K \ll N_{tx}$ reference symbols[3]. However, if we create a fixed grid of N_B beams at each base station, then the resulting $N_{CC} = N_B \cdot N_{BS}$ channel components will have different average channel powers, as seen from one user position.

To achieve massive MIMO gains, each user need only to estimate its strongest channel components. The second important step in our design is hence to design N_{CC} RSs codes of length K such that regardless of where the user is located, it will be able to estimate its strongest channel components sufficiently accurately.

We consider a system design as outlined in Figure 1, where a total of N_{PRC} antennas are arranged into $N_{CC} = N_B * N_{BS}$ beams by digital or analogue beamforming. Each beam transmits a unique RS code in the downlink. From these, the users will be able to estimate their strongest channel components, which then are fed back to the base station. Through the CSI a precoder which directs each message towards the intended user is designed. In this work, we focus on the estimation of CSI.

[2]The extension to multi-antenna users is outside the scope of this paper. We refer the interested reader to [38].

[3]Correlations between channels may improve estimations, as suggested in [29, 30, 32, 34], but the correlation between channels from different antennas may be low and provide only little extra estimation performance.

2.1 Fixed grid of Beams

When using digital beamforming, the channel component from the n'th beam on a resource k at time τ is

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

where $\tilde{h}(k,l,\tau)$ is the physical radio channel from the l'th antenna and $b(n,l)$ are the beamforming weights.

The fixed grid of beams improves the channel estimation by two main contributions. First, and most important, 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 [28], 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 RSs. Second, beamforming reduces the time variations of channels [4], and hence makes them easier to estimate [38].

The effects of different choices of number of beams N_B, relative to the number of antennas N_{tx} per base station is discussed in Appendix A.

2.2 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, which are arranged in a column vector

$$\bar{h}(\tau) = \begin{bmatrix} \big[h(1,1,\tau)\ ...\ h(K,1,\tau)\big]^T \\ \vdots \\ \big[h(1,N_{CC},\tau)\ ...\ h(K,N_{CC},\tau)\big]^T \end{bmatrix}, \tag{2}$$

Assume that each beam n transmits a RS symbol $\varphi(k,n,\tau)$ on every available RS resource. These may be time dependent, as discussed further in Section 2.2.3. The received downlink signal at the user on the k:th RS resource at the τ:th time instant is then

$$y(k,\tau) = \Sigma_{n=1}^{N_{CC}} \varphi(k,n,\tau) \cdot h(k,n,\tau) + v(k,\tau). \tag{3}$$

Here, $v(k,\tau)$ is the sum of noise and intercluster interference, which we for simplicity shall denote as noise.

The measurements on all RS resources can be collected into a vector

$$\bar{y}(\tau) = \big[y(1,\tau)\ \dots\ y(K,\tau)\big]^T,$$

and the measurement equation is then expressed as

$$\bar{y}(\tau) = \Phi(\tau)\bar{h}(\tau) + \bar{v}(\tau), \tag{4}$$

where

$$\Phi(\tau) = \begin{bmatrix} \text{diag}\{\varphi(k,1,\tau)\}_{k\in[1,K]} \\ \vdots \\ \text{diag}\{\varphi(k,N_{CC},\tau)\}_{k\in[1,K]} \end{bmatrix}^T. \tag{5}$$

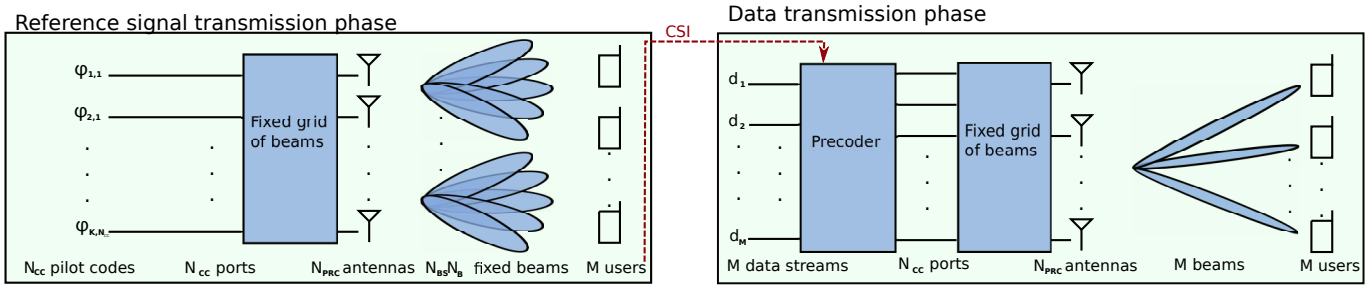

Figure 1: RSs are transmitted in the downlink over a fixed grid of $N_{CC} = N_B \cdot N_{BS}$ beams controlled by N_{CC} antenna ports. Users estimate subsets of the channel components and feed back these estimates over the uplink. During the data transmission phase the N_{CC} beams are precoded, e.g. by MRC or interference mitigation precoding, to direct the signal energy to each user, and to potentially remove interference from the other users.

The measurement noise terms are collected in a vector $\bar{v}(\tau) = \left[v(1,\tau) \ \dots \ v(K,\tau)\right]^T$ which will be assumed to be i.i.d. in time with covariance matrix

$$R_v = E[\bar{v}(\tau)\bar{v}(\tau)^*]. \tag{6}$$

To ensure that a user will be able to separate and estimate its strongest channel components, each beam will have a unique RSs code

$$\bar{\varphi}(n,\tau) = [\varphi(1,n,\tau) \ \dots \ \varphi(K,n,\tau)]^T.$$

In the special case when $N_{CC} \leq K$, these codes could be fully 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.

2.2.1 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 $k = 1,...,K$. The measurement equation in (4) can then be simplified to

$$\begin{aligned}\bar{y}(\tau) &= \underline{\Phi}(\tau)\underline{\bar{h}}(\tau) + \bar{v}(\tau) \\ &= [\bar{\varphi}(1,\tau)\dots\bar{\varphi}(N_{CC},\tau)] \begin{bmatrix} \underline{h}(1,\tau) \\ \vdots \\ \underline{h}(N_{CC},\tau) \end{bmatrix} + \bar{v}(\tau).\end{aligned} \tag{7}$$

Let each antenna port now have its own unique code in accordance with

$$\underline{\Phi}(\tau) = \begin{bmatrix} -1 & 0 & 1 & 1 & -1 & 0 & -1 & -1 & 1 \\ 0 & 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & -1 & -1 & -1 \end{bmatrix}. \tag{8}$$

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 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

$$\hat{\underline{h}}_{\text{rel}}(\tau) = \underline{\Phi}^{\dagger}_{\text{rel}}(\tau)\bar{y}(\tau). \tag{9}$$

The set of relevant channel components, and their number, will depend of the user's location, but for any location, the user will be able to estimate up to $K = 6$ channel components through (9) 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 (9). 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 chose to send no RS on the non-relevant beams. That enhancement is called beam deactivation and is described in [39].

2.2.2 General reference signal codes

The RS matrix (8) was designed such that any subset of up to six column vectors will be linearly independent. Searching for such matrices for higher numbers of K and N, while also adding a constraint of a per antenna RS power budget, would result in an increasingly complicated design problems for integer valued vectors. Instead we set the RS symbol from the n:th beam over the k:th RS resource to be complexed-valued, with a fixed gain and a variable phase

$$\varphi(k,n,\tau) = \exp\left(\theta(k,n,\tau) \cdot j\right). \tag{10}$$

The angles $\theta(k,n,\tau)$ should then be designed to ensure that any subset of K vectors are linearly independent. We here suggest two ways of selecting these. In [28], a *Vandermonde like coding* was used. In this, a real-valued design parameter $\phi(\tau) > 0$, that can be fixed over time or

varying, is selected and the phases are defined as

$$\theta(k, n, \tau) = (k\phi(\tau))^n. \quad (11)$$

How to select the real-valued design parameter $\phi(\tau)$ is an object for investigation which we shall return to in Section 4. Alternatively the phase values $\theta(k, n, \tau)$ may be selected *pseudo-randomly* (but of course known to the receiver) from a uniform distribution over $[-\pi, \pi]$. We here use the Vandermonde like coding (11) in order to ease repeatability of our results, since it is specified by one scalar parameter $\phi(\tau)$.

With both these methods, a few submatrices may still 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 (9) the noise and pilots from the weak channels may be amplified with 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.

In the event that only a finite number of phase angles are available, e.g. if an N-ary phase shift keying (PSK) modulation is used, the phase angles $\theta(k, n, \tau)$ must be rounded to the nearest fixed phase angle. The probability of any two row vectors in $\Phi(\tau)$ being linearly dependent will then increase. However, in an off-line exhaustive search such combinations will be ill-conditioned and hence never selected.

2.2.3 Cycling reference signals

Even with an exhaustive search, some user positions will in general have sets of relevant channel components that form very good RS code submatrices, while others will end up with sets of relevant channel components that form submatrices with rather large eigenvalue spread, causing a system that favours some users over others. However, with an estimation algorithm that utilize temporal correlation, we can improve the fairness amongst user by introducing cycling RSs. This idea was introduced for the uplink in [40] where the number of users exceeded the number of available RS resources.

For the coding (11), 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.

Note that 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. Other estimators can still use the RSs at each time.

To illustrate the effect of cycling RSs on user fairness, we assume a system with $N_{cc} = 72$ channel components and $K = 18$ RS resources. We form three RS matrices $(\Phi(\tau))$ through (11) using $\phi(\tau) = \{1, 2, 3\}$. The values of $\phi(\tau)$ are chosen to ensure reasonably low numbers of $\text{cond}(\Phi(\tau)$ (given by $\{2.55, 2.48, 2.27\}$)).

In Figure 2 we study the condition numbers of submatrices to these. 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 have 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 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 good condition. In contrast, if we were to cycle the three then most users (> 99.9 %) would have a condition below 100. When temporal correlation is used (and high enough), then users can use the estimates based on their best RS matirx to improve the subsequent estimates

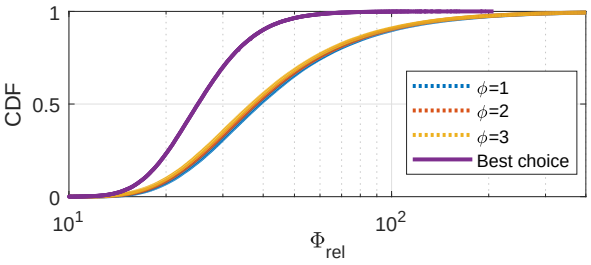

Figure 2: CDF of the condition of the RS submatrix $\Phi_{\text{rel}}(\tau)$ for different sets of 18 relevant channel components (i.e. different user positions) for different $\Phi(\tau)$.

2.2.4 Comments on reference signal contamination

In some of the pioneering work in massive MIMO, reference signal contamination from other cells was pointed out as the one physical factor that limits massive MIMO. With a frequency reuse 1 scheme, reference signals from other base stations will certainly cause some contamination. Due to the logarithmic nature of throughput, the largest gains for massive MIMO, when more antennas are added to the framework, comes from serving more users within each recourse slot, utilizing spacial multiplexing.

Then if gains should be secured for a very large number antennas, a very large number of users must be served. They must then in the TDD framework all transmit uplink reference signals on the same resources. These observations have led to a focus in research on how to minimize reference signal contamination i.e. the influence of non-orthogonality on the estimation performance.

We here briefly discuss the effect of reference signal contamination in our proposed solution. For the sake of clarity we distinguish between inter-cluster reference signal contamination, i.e. the interference from reference signal transmission in different cooperation clusters, and intra-cluster reference signal contamination.

The first of these, the *inter-cluster contamination*, can be kept fairly small. In [41], a scheme was proposed where each site participates in several overlapping cooperation clusters. Each cooperation cluster is allocated different parts of the total bandwidth. For example, if each site participates in six overlapping clusters, then each cluster is allocated one sixth of the total bandwidth[4]. By complementing the overlapping clusters with cluster specific vertical antenna tilting, the proposed scheme was able to lower the inter-cluster interference floor significantly, as the neighbouring clusters did not cause mutual interference. The concept was supported by field trials.

The inter-cluster reference signal contamination, i.e. the inter-cluster interference during the reference signal transmission phase, can be even further reduced by adjusting the coded reference signals discussed in Section 2.2. We may allocate different sets of codes to interfering cooperation clusters in a close vicinity of each other and design those sets of codes such that the closer two cooperation clusters are, the more orthogonal are their codes are. Note that the primary focus remains to ensure that the reference signal codes within each cooperation cluster's set are as orthogonal as possible. These sets can then be designed to influence each other as little as possible as a secondary effect.

The *intra-cluster reference signal contamination* will therefore likely be the main contributor to the total interference from reference signal contamination. As our scheme is designed to utilize a lower number of reference signal resources than the number of available channel components, some reference signal contamination is inevitable. In order to see the influence of this, we must study the residual error in channel estimates for example by comparing with fully orthogonal reference signals.

Beam deactivation (not sending reference signals from unused beams) is an interesting method for reducing intra-cluster contamination [39]. Its effect is not studied in the present paper.

3 Channel estimation

A main feature of our proposed solution 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 to improve the estimates of those used for transmission. 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[5]. 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 shown in Figure 3. First, we assume that perfectly 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 and frequency and to find the set of relevant channels for new users that enter the system. 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 the continuously estimated.

3.1 LLMSE estimation

We can account for the correlation in space and frequency by utilizing 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

$$\hat{h}_{\text{rel}} = R_{h,\text{rel},y} R_y^{-1} \bar{y}, \tag{12}$$

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. These can be estimated either from the sparsly transmited orthogonal RS or from past channel estimates, see Appendix C.1 for details[6].

3.2 Kalman filter

Kalman filters have been found useful for channel estimation and prediction, see e.g. [38, 42, 44–46]. To incorporate the temporal correlation (as well as correlation in frequency and space), we utilize a Kalman filter, which

[4] From a user perspective, this seem like frequency reuse, but every base station utilizes the full bandwidth, it just cooperates with different other base stations on different resources.

[5] 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 [18, 42] and references therein.

[6] As no temporal correlation is assumed to be utilized by this LLMSE estimator, the time index τ is not indicated in this section assuming that the measurements, the noise, the RS matrices and the channel components all have the same time index.

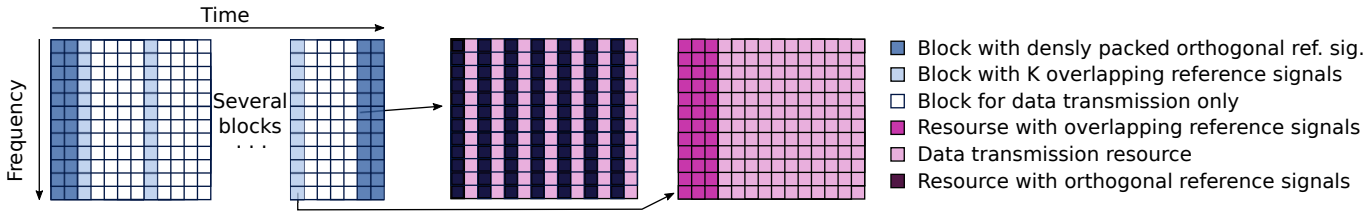

Figure 3: Exempel of RS structure in the downlink. Resources are divided into blocks of 12 subcarriers and 14 OFDM-symbols (as in LTE). Every fifth block contains K RS resources that can be used for channel estimation. On a slow time scale (every 500:th block) two subsequent blocks include 50% resource slots for fully orthogonal RSs. The total RS overhead in this example is approximately 4.5%.

enable us to use channel information from all previous measurements. For this, we model the channel statistics by an auto regressive (AR) model

$$\begin{aligned} x(\tau+1) &= Ax(\tau) + B\bar{u}(\tau), \\ \bar{h}(\tau) &= Cx(\tau). \end{aligned} \tag{13}$$

Here A, B and C are complex-valued state space matrices, $u(t)$ is the process noise and $x(\tau)$ 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 $\bar{u}(\tau)$, of dimension KN_{CC}, given by

$$Q = E[\bar{u}(\tau)\bar{u}^*(\tau)]. \tag{14}$$

The measurement vector in (4) can then be expressed as

$$\bar{y}(\tau) = \Phi(\tau)Cx(\tau) + \bar{v}(\tau). \tag{15}$$

The model (13)-(15) requires several subsequent channel estimates, so it cannot be estimated by the sparsely transmitted orthogonal RSs in Figure 3. 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. The AR modelling is described in Section 3.2.1 below.

For every new measurement $\bar{y}(\tau)$ by (15), the filter can then recursively compute the channel estimate through

$$\hat{x}(\tau|\tau-1) = Ax(\tau-1|\tau-1), \tag{16}$$

$$P(\tau|\tau-1) = AP(\tau-1|\tau-1)A^* + BQB^*, \tag{17}$$

$$\hat{x}(\tau|\tau) = \hat{x}(\tau|\tau-1) + \mathcal{K}(\tau)(\bar{y}(\tau) + J(\tau)\hat{x}(\tau|\tau-1)), \tag{18}$$

$$P(\tau|\tau) = P(\tau|\tau-1) - \mathcal{K}(\tau)J(\tau)P(\tau|\tau-1), \tag{19}$$

$$\hat{h}(\tau|\tau) = C\hat{x}(\tau|\tau). \tag{20}$$

Here, $J(\tau) = \Phi(\tau)C$, $P(\tau|\tau)$ is the covariance matrix of the state vector estimation error

$$P(\tau|\tau) = E[(x(\tau) - \hat{x}(\tau|\tau))(x(\tau) - \hat{x}(\tau|\tau))^*, . \tag{21}$$

and the matrix $\mathcal{K}(\tau)$, known as the Kalman filter gain, is obtained through

$$\mathcal{K}(\tau) = P(\tau|\tau-1)J(\tau)^*(J(\tau)P(\tau|\tau-1)J(\tau)^* + R_v)^{-1}. \tag{22}$$

If the state 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 (20) coincides with the LLMSE estimate (12).

The Kalman equations (13)-(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 $P(0|0) = E[x(\tau)x^*(\tau)]$.

Provided that the RS matrix is cyclic with $\Phi(\tau) = \Phi(\tau+\mu)$, the filter will converge into a cyclo-stationary filter with $P(\tau|\tau-1) = P(\tau+\mu|\tau+\mu-1)$ within a few cycles. Then, (17),(19) and (22), can be calculated off-line by solving a Riccati equation, see Appendix C.3 for details. This procedure is only needed whenever a new AR-model is estimated.

3.2.1 Estimation of AR model coefficients

The small scale fading of channel components can be modelled by ρ'th order AR processes [47]. The channel components $h(k,n,\tau)$ in (2) can then be written as a sum of their ρ past realisations and a white noise term $u(k,n,\tau)$, denoted process noise, through

$$h(k,n,\tau) = -\Sigma_{i=1}^{\rho} a_i h(k,n,\tau-i) + u(k,n,\tau). \tag{23}$$

The poles of the AR process depend on the temporal correlations of $h(k,n,\tau)$, which in turn depend on the main characteristics of the environment (in terms of main scatterers, presence of line-of-sight component, etc), affected by the shadow fading. As shadow fading changes slowly, it can be assumed to be wide sense stationary over T subsequent blocks of RS resources. For realistic channels, the order of the AR-model should be in the range of 4-6 for pedestrian users, see e.g. [38].

The channel model coefficients $\{a_i\}$ in (23), can then be estimated for temporal blocks of duration $T/2$ and be used for the subsequent $T/2$ duration time block. In Appendix C.2 we describe how to estimate the poles given the autocorrelation function of the channel, and based on these, how to set up the state space model (13).

3.2.2 Estimation of the process noise covariance matrix Q

While equation (13) models the temporal correlation, the frequency correlation and the spatial/antenna correlation between channel components is modelled through the covariance matrix of the process noise, of dimension KN_{CC}, given by (14).

Estimating the process noise covariance matrix can be complicated. For the ideal case where the diagonal state matrices A, B and C perfectly model the time dynamics of the channel then it can be shown, [38], that Q by (13)-(39) is given by

$$Q = R_h \oslash C(B\mathbf{1}B^* \oslash (\mathbf{1} - \bar{a}\bar{a}^*))C^*, \tag{24}$$

where $\bar{a}$ is a column vector containing the diagonal elements of A, $\oslash$ denotes elementwise division, and $\mathbf{1}$ is a matrix of ones. This is a good choice 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 [38]. They also include cases 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 (24) 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

$$\begin{aligned} Q &\approx B^\dagger(\Pi + A\Pi A^*)(B^*)^\dagger, \\ \Pi &\approx C^\dagger R_h (C^*)^\dagger, \end{aligned} \tag{25}$$

where $\Pi = E[x(\tau)x^*(\tau)]$. Equation (25) follows directly from the state space model, and will provide a Q-matrix which would result in 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 positive semidefinite Q as

$$Q = M^*M. \tag{26}$$

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.

There alternatives are compared in Appendix D, and based on these results, the estimate (25) has been used in the channel estimation performance investigation of Section 4.2.

3.3 Reduced complexity Kalman filter

Channel estimation through (16)-(20) provides the optimal (linear) estimate, but the on-line complexity grows with the square of the number of channel components [38]. 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

$$\begin{aligned} x_{\text{rel}}(\tau+1) &= A_{\text{rel}}x_{\text{rel}}(\tau) + B_{\text{rel}}\bar{u}_{\text{rel}}(\tau), \\ \bar{h}_{\text{rel}}(\tau) &= C_{\text{rel}}x_{\text{rel}}(\tau), \end{aligned} \tag{27}$$

where the process noise $\bar{u}_{\text{rel}}(\tau)$ is i.i.d. with covariance matrix

$$Q_{\text{rel}} = E[\bar{u}_{\text{rel}}(\tau)\bar{u}_{\text{rel}}(\tau)^*], \tag{28}$$

which are similar to (13) and (14), with exceptions of the dimensions. The measurement is then

$$\bar{y}(\tau) = \Phi_{\text{rel}}(\tau)\bar{h}_{\text{rel}}(\tau) + \bar{v}(\tau) + \bar{w}(\tau), \tag{29}$$

where $\Phi_{\text{rel}}(\tau)$ is the sub-matrix of the RS matrix $\Phi(\tau)$ that consists of the column vectors corresponding to the channel components. Equation (29) differs from (4) as it includes an additional noise term $\bar{w}(\tau)$. This is the contribution of the non-relevant channel components

$$\bar{w}(\tau) = \Phi_{\bar{\text{rel}}}(\tau)\bar{h}_{\bar{\text{rel}}}(\tau). \tag{30}$$

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

$$\begin{aligned} T(t) &= E[\bar{w}(\tau)\bar{w}^*(\tau-t)] \\ &= E[\Phi_{\bar{\text{rel}}}(\tau)\bar{h}_{\bar{\text{rel}}}(\tau)\bar{h}^*_{\bar{\text{rel}}}(\tau-t)\Phi^*_{\bar{\text{rel}}}(\tau-t)]. \end{aligned} \tag{31}$$

Element (i,j) of $T(t)$ will be

$$T_{i,j}(t) = E[\bar{w}_i(\tau)\bar{w}_j^*(\tau-t)], \tag{32}$$

with

$$\bar{w}_i(\tau) = \sum_{n\in\mathcal{H}_{\bar{\text{rel}}}} \varphi(i,n,\tau)h(i,n,\tau). \tag{33}$$

If we consider the special case where the RS matrix $\Phi(\tau)$ is constant, then each term of the sum in (33) will in itself be a scaled and rotated channel (compare with beamforming). When averaging over different elements $\varphi(k,n,\tau)$ of the RS matrix $\Phi(\tau)$, they will with high probability be only weakly correlated in the sense that $E[\varphi(k,n,\tau)\varphi^*(k+\Delta k, n+\Delta n,\tau)]$ will be close to zero unless $\Delta k = \Delta n = 0$. The RS codes will therefore break up most correlations of channel components, leading to $T(t)$ in (31) being diagonal dominant. To improve the estimate, the Kalman filter can be designed to track also the K noise terms (33).[7]

If cyclic RSs are used then, with the same argument as above, we can say that with high probability $E[\varphi(k,n,\tau)\varphi^*(k,n,\tau+t)] \approx 0$ when $t \neq \mu \cdot m$ where m is an integer. To capture the dynamic, we would then have to keep track of μK noise terms.

[7] Their statistics will then be included in an expanded AR state space model.

However, we may also choose to take a less Bayesian approach and 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 (27)-(28) directly in the Kalman filter (16)-(20) by replacing A, B, C, Q and R_v by A_{rel}, B_{rel}, C_{rel}, Q_{rel} and $R_v + T$ respectively.

4 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 [48]. Three 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 drawn within a circle with a 500 m radius centred at the cluster center. 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 1. Other settings were set to the default values in the Quadriga channel generator, see [49].

Table 1: Simulation parameters used in the Quadriga channel simulator, see [49].

Scenario	WINNER_UMa _C2 _NLOS [50]
Carrier frequency	2.53 GHz
Subcarrier spacing	15 kHz
# subcarriers	144
RS spacing	5 ms
BS hight	32 m
Antenna tilt	$-8°$
# antennas/BS	32
Antenna spacing	0.5 wavelengths

4.1 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 $n = 0, ..., 7$, thus forming eight horizontal beams per base station, yielding a total of $8 \cdot 3 \cdot 3 = 72$ channel components.

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 4 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.

In [18], 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 4 indicate that 15-20 channel components would then need to be estimated.

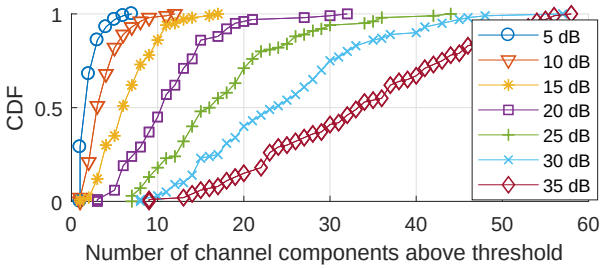

Figure 4: CDF of the number of relevant channel components (CCs) at different user positions when using different thresholds in dB, relative to the power of the strongest channel.

4.2 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, similar to the structure shown in Figure 3. The channels are frequency selective, the correlation between the fading at each edge of the 90 kHz resource blocks is around 0.9-0.98.

The channel components are assigned cyclic RS codes using (10)-(11) 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 (4), 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 respectivley. The channel statistics, represented by (34)-(36) for the LLMSE filter and by (13)-(14) and (6), (27)-(28) and (31) for the Kalman filter, was based on the first part of this. We assume that the measurements of the non-relevant channel components are very noisy and hard to estimate based on the sparsely transmitted orthogonal RSs. Therefore, we set the cross correlation matrix (35) to an all zero estimate and the covariance matrix (36) to a diagonal matix, as discussed in Section 3.1.

Based on these, each user also find a set of 16 relevant channels.[8]. Here the number 16 is fixed, the number of relevant channels is here not determined by a power threshold. These were then estimated using a) pseudo inversion of the reduced RS matrix (9), b) the LLMSE estimate (12) and c) the Kalman filter estimate (20), based the reduced model (27)-(33).

We use the approximation that the noise term $\bar{w}(\tau)$ in (29) is i.i.d. over time. In total 24 subbands of 90 kHz

[8] Without use of any source coding, the corresponding uplink feedback rate would, assuming feedback every 5 ms and 10 bits per channel component, be 160 bits/5 ms=32 kbits/s for each 90 kHz subband, for a terminal moving at 3 km/h. This is high, but not unreasonably high. At lower mobility, the feedback repetition rate and the corresponding data rate can be reduced proportionally. Source coding and vector quantization can reduce the feedback overhead drastically [51].

each (6 subcarriers) are tracked by parallel Kalman filters. In (27) each channel component is modeld by a 4:th order AR model. The process noise covariance matrix Q in (14) is estimated through (25). In Appendix D, we evaluate this choice of Q and compare it to an optimize solution based on (26).

For comparison, and as a lower bound, 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 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[9].

4.2.1 Estimations Performance

Figure 5 shows the NMSE as a function of the channel component number, averaged over subcarriers and users. Here, the inversion through (9) 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 comes at 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 (9) may be sufficient for the strongest channel components, it quickly degrades for weaker channels. Comparing channel inversion through (9) to the LLMSE filter, we see that by utilizing the space and frequency correlations estimations 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 requires interference mitigation.

Most of the previously suggested approaches to channel estimation for FDD massive MIMO is optimize RSs for a specific set of users, see Section 1.2. To relate our results to those methods, we can compare the channel estimates or the overlapping RSs and 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. 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 non-orthogonal 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.

[9]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 %.

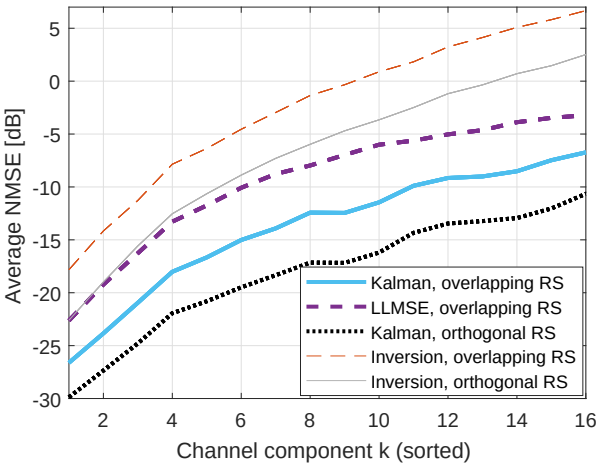

Figure 5: Average NMSE sorted after the RS strength of the channel component.

4.2.2 Effects of using a reduced Kalman filter

Figure 6 illustrates how much we loose by only estimating the $N_{rel} = 16$ relevant channel components by the reduced Kalman filter with the model (27)-(28), rather than estimating all channel components ($N_{rel} = 72$) by the model (13)-(15). As the off-line complexity related to calculating the covariance matrix $P(t|t)$ in (19) 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, 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 (29) in the state vector. Such investigations are left for future work.

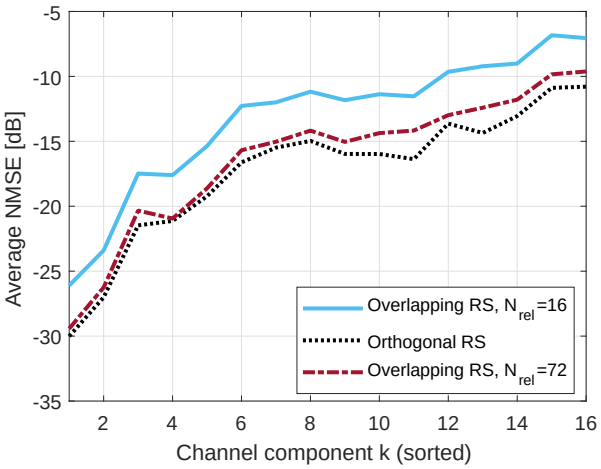

Figure 6: Average NMSE, as achieved by the Kalman filter, sorted after the received RS power of the channel component.

4.2.3 Capacity for MRC

The effect of the estimation errors in terms of MRC beamforming gain is illustrated in Figure 7. The results in this figure is based on the average value over all users, i.e. we assume that the estimation performance are given by the average Kalman estimation NMSE presented in Figure 5 and the SNR per channel component is the inverse of the NMSE achieved by inversion with orthogonal RSs (the thin solid line in Figure 5). The Shannon capacity show in the figure, is based on maximum transmit beamforming to one single user, by combining K of the 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 and an extra gain in accuracy will only translate into a very small capacity gain. Second, 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 7.

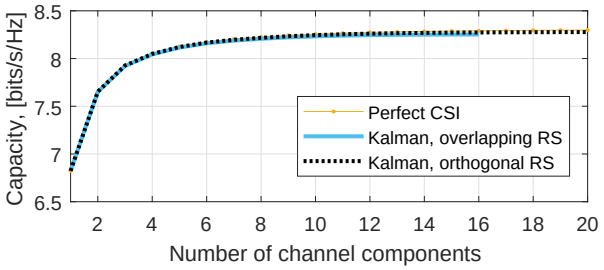

Figure 7: An illustration of the single users maximum ratio beamforming gain, based on Shannon capacity, as a function of the included number of channel components. CSI is based on the Kalman estimates in Figure 5. Note that the curves overlap.

5 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 estimations were 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 time independent noise. This comes at a 5 dB performance loss in estimation NMSE, but estimations are still sufficiently accuracy 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 matrices. However, we have also shown that this problem can be essentially eliminated, *without* having to adjust RS patterns to users, by introducing cycling RSs.

Open issues

A straightforward way to improve both the estimator performance and the power efficiency would be by not sending RSs in the beams that are weak to all presently active user. Such beam deactivation schemes are under current investigation.

Furthermore, 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. and have been used to advantafe in a large set-up with 288 beams, in [39].

Our results showed a significant gap in estimation performance of the reduced Kalman filter and that which estimated all channel components. This gap may be reduced if the interfering noise terms $\bar{w}$ from the non-relevant channels, in (33), are tracked along with the relevant channels.

A source of estimation errors stems from estimating the covariance matrices (39), (34)-(36) and (28). 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.

A Comments on limitations when utilizing a fixed grid of beams

There is a trade-off involved in the choice of number of beams per base station. More beams provide more de-

grees of freedom (up to $N_B = N_{tx}$), but with too many beams, the sum of energy from non-relevant antenna ports will be large compared with that from the relevant antenna ports resulting in worse CSI. We here provide some comments on to what extent the fixed grid of beams may limit massive MIMO gains for different N_B. In this section, we focus on the case when channel vectors of different users are approximately orthogonal and MRC can be used to achieve close to optimal performance.

First, note that equation (1) provides a linear transformation of the radio channels from the physical antennas. In the special case when $N_B = N_{tx}$ and the beamforming matrix β, with elements $\beta_{n,j} = b(n,j)$, has full rank, then the linear transformation implies no restrictions.

Using $N_B > N_{tx}$ would reduce the RS power allocated to each beam, and thus reduce the estimation accuracy. If instead, we have $N_B < N_{tx}$, estimation accuracy is improved but we loose gains in that the scalability of massive MIMO grows with the size of N_B rather than with N_{tx}. For TDD massive MIMO, if the number of served users M are proportional to N_{tx} and $M \ll N_{tx}$, then gains grow linearly with N_{tx}. This translates to the $M \propto N_B$ and $M \ll N_B$ for the fixed grid of beams. This does not mean that the sum-rate will be a factor N_B/N_{tx} smaller, because most of the energy from the antennas will be present in the strongest beams as shown in the results of Figure 7. It only means that beamforming gains in terms of SNR do not grow linearly with N_{tx} if N_B is kept constant.

Note that also TDD massive MIMO is limited. Assuming that CSI is acquired through the uplink, all scheduled users must transmit in the uplink RSs. Therefore, similar to the RS downlink problem, when the number of transmit antennas are limited by the RS resources, the number of users that can be served is limited by the available RS resources. That is fine for a scenario where relatively few users with high data rate requirements, but is very limiting if there are many potential users with low and bursty data rate requirements.

B Coded reference signals and channel estimation: benefits for TDD

In this paper we focus on a reference signal coding and estimation scheme for FDD systems. However, some of our results also apply to TDD systems. In a TDD system, pilot resources are limited in the uplink. As massive MIMO gains rely on multi user gains, see Appendix A, this practically puts a limitation on the massive MIMO gains for TDD.

In order to increase the number of users that can be scheduled, we may allow them to transmit uplink reference signals on the same resources. If we where to then introduce fixed grid of beams at base stations, this would lead to less crosstalk in uplinks as users at different locations are less likely to have the same strong beams. Moreover, we can introduce cycling coded reference signals to further improve channel estimates to different users. This was the topic of [40].

C Statistical modeling

C.1 Crosscovariance matrices

We let $\bar{h}_{\text{rel}}$ and $\bar{h}_{\bar{\text{rel}}}$ be vectors consisting of the relevant and non-relevant channel components respectively and define three covariance matrices

$$R_{h,\text{rel}} = E[\bar{h}_{\text{rel}}\bar{h}^*_{\text{rel}}], \tag{34}$$

$$R_{h,\text{rel},\bar{\text{rel}}} = E[\bar{h}_{\text{rel}}\bar{h}^*_{\bar{\text{rel}}}], \tag{35}$$

$$R_{h,\bar{\text{rel}}} = E[\bar{h}_{\bar{\text{rel}}}\bar{h}^*_{\bar{\text{rel}}}]. \tag{36}$$

We assume that the sparsely transmitted fully orthogonal RSs are sufficient to estimate the covariance matrix of the relevant channels (34)[10]. The estimates of the weak components $\bar{h}_{\bar{\text{rel}}}$ may be too noisy to reliably estimate the covariance matrices (35)-(36). It should however be possible to estimate the average gains of the non-relevant channels down to some low power threshold. Then, a reasonable approach is to use an all zero matrix as estimate of (35) and a diagonal matrix with the powers of the non-relevant channels, down to the threshold, on the diagonal as an estimate of (36).

With the covariance matrices (34)-(36) and R_v by (6), we obtain the covariance matrices used in (12) through

$$\begin{aligned} R_{h,\text{rel},y} &= E[\bar{h}_{\text{rel}}\bar{y}^*] \\ &= E[\bar{h}_{\text{rel}}(\Phi_{\text{rel}}\bar{h}_{\text{rel}} + \Phi_{\bar{\text{rel}}}\bar{h}_{\bar{\text{rel}}} + \bar{v})^*] \\ &= R_{h,\text{rel}}\Phi^*_{\text{rel}} + R_{h,\text{rel},\bar{\text{rel}}}\Phi^*_{\bar{\text{rel}}}, \end{aligned} \tag{37}$$

and

$$\begin{aligned} R_y &= E[\bar{y}\bar{y}^*] \\ &= E[(\Phi_{\text{rel}}\bar{h}_{\text{rel}} + \Phi_{\bar{\text{rel}}}\bar{h}_{\bar{\text{rel}}} + \bar{v})(\Phi_{\text{rel}}\bar{h}_{\text{rel}} + \Phi_{\bar{\text{rel}}}\bar{h}_{\bar{\text{rel}}} + \bar{v})^*] \\ &= \Phi_{\text{rel}}R_{h,\text{rel}}\Phi^*_{\text{rel}} + \Phi_{\bar{\text{rel}}}R_{h,\bar{\text{rel}}}\Phi^*_{\bar{\text{rel}}} + \Phi_{\text{rel}}R_{h,\text{rel},\bar{\text{rel}}}\Phi^*_{\bar{\text{rel}}} \\ &\quad + \Phi_{\bar{\text{rel}}}R^*_{h,\text{rel},\bar{\text{rel}}}\Phi^*_{\text{rel}} + R_v. \end{aligned} \tag{38}$$

Here, Φ_{rel} and $\Phi_{\bar{\text{rel}}}$ are the subsets of the RS matrix Φ that contain the RS codes for the relevant channel components and non-relevant channel components respectively.

C.2 State space modeling

As the channel statistics depends on the large scale fading, the model (13)-(14) must be updated repeatedly (e.g. every 0.5 s for pedestrians). We here provide a brief summary of how to estimate the autoregressive channel model (13)-(14) (or (27)-(28) below) from a set of training data. For a detailed description we refer to [42].

[10] A user can further improve the estimates of the channel components off-line by e.g. using a two dimensional smoothing filter over all time and frequency slots. Although these estimates are of no use for the data transmission, as the estimated channels will be severely outdated, they will be useful as training data for the AR modelling.

First, we estimate the covariance matrix of the channel vector

$$R_h = E[\bar{h}(\tau)\bar{h}^*(\tau)], \tag{39}$$

and the autocorrelation functions of each channel component

$$R_{\tau,h}(n,t) = E[h(k,n,\tau)h^*(k,n,\tau+t)], \tag{40}$$

by averaging over all available training data. Note that the autocorrelation functions are independent of k as the fading on different time-frequency resources are approximately identically distributed if they are all well within a shadow fading stationary interval.

The autocorrelation (40) is used to estimate the AR model coefficients $\{a_i\}_{i=1,..,\rho}$ of (23). This can be achieved by multiplying both sides of (23) by $h^*(k,n,\tau+t)$ for $t = 1, ..., \rho$. Taking the expected values, we get a set of linear equations known as the Yule-Walker equations [54]. By solving these the coefficients are found.

Then, as with the channel components in (2), we stack the process noise terms into a vector

$$\bar{u}(\tau) = \begin{bmatrix} \begin{bmatrix} u(1,1,\tau) & \dots & u(K,1,\tau) \end{bmatrix}^T \\ \vdots \\ \begin{bmatrix} u(1,N_{CC},\tau) & \dots & u(K,N_{CC},\tau) \end{bmatrix}^T \end{bmatrix}. \tag{41}$$

Next, based on the AR model for each resource k and channel component n, separate state space matrices, $A(k,n)$, $B(k,n)$ and $C(k,n)$, are set up on diagonal form, to describe $h(k,n,\tau)$ as $x_{k,n}(\tau+1) = A(k,n)x_{k,n}(\tau) + B(k,n)u_{k,n}(\tau)$; $h(k,n,\tau) = C(k,n)x_{k,n}(\tau)$. The state space matrices of (13) are then given by

$$\begin{aligned} A &= \text{diag}\{\text{diag}\{A(k,n)\}_{k=1,...,K}\}_{n=1,...,N_{CC}}, \\ B &= \text{diag}\{\text{diag}\{B(k,n)\}_{k=1,...,K}\}_{n=1,...,N_{CC}}, \\ C &= \text{diag}\{\text{diag}\{C(k,n)\}_{k=1,...,K}\}_{n=1,...,N_{CC}}. \end{aligned}$$

We here limit the state space vector to only contain the subcarriers within one RS recource block, where the fading on different subcarriers is highly correlated assuming that estimations for other blocks are carried out in parallel[11].

C.3 Covariance matrix of the stationary Kalman filter

To solve for the one step prediction error covariance matrix $P(t+1|t)$ of the state vector $x(t)$ when cyclic reference signals with a cycle of μ are used, a time invariant augmented state space model is created. The state space matrices of the augmented state space model are given by

$$A_{\text{aug}} = \begin{bmatrix} 0 & \dots & 0 & A \\ 0 & \dots & 0 & A^2 \\ & \vdots & & \\ 0 & \dots & 0 & A^\mu \end{bmatrix}, \tag{42}$$

$$B_{\text{aug}} = \begin{bmatrix} B & 0 & \dots & 0 \\ AB & B & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ A^{\mu-1}B & A^{\mu-2}B & \dots & B \end{bmatrix}, \tag{43}$$

and

$$J_{\text{aug}} = \text{diag}\{C\Phi(\tau)\}_{\tau=0,...,\mu-1}, \tag{44}$$

with a state vector for the augmented system given by

$$x_{\text{aug}}(\tau) = \left[x(\tau)^T ... x(\tau+\mu-1)^T\right]^T, \tag{45}$$

which is updated every μ time step. The noise covariance matrices of the augmented system are given by $Q_{\text{aug}} = \text{diag}\{Q\}_{j=1,...,\mu}$, for the process noise, and by $R_{v,\text{aug}} = \text{diag}\{R_v\}_{j=1,...,\mu}$, for the measurement noise. Through these the one step prediction error covariance matrix for the augmented system can be found off-line by solving the algebraic Riccati equation

$$\begin{aligned} P_{\text{aug}} &= E[\tilde{x}_{\text{aug}}(\tau+1|\tau)\tilde{x}^*_{\text{aug}}(\tau+1|\tau)] \\ &= A_{\text{aug}}P_{\text{aug}}A^*_{\text{aug}} + B_{\text{aug}}Q_{\text{aug}}B_{\text{aug}} - \\ &\quad - F_{\text{aug}}P_{\text{aug}}J^*_{\text{aug}}R^{-1}_{e,\text{aug}}J_{\text{aug}}P_{\text{aug}}F^*_{\text{aug}}, \end{aligned} \tag{46}$$

where

$$R_{e,\text{aug}} = J_{\text{aug}}P_{\text{aug}}J^*_{\text{aug}} + R_{v,\text{aug}}, \tag{47}$$

and

$$\tilde{x}_{\text{aug}}(\tau+1|\tau) = x_{\text{aug}}(\tau+1) - \hat{x}_{\text{aug}}(\tau+1|\tau). \tag{48}$$

The one step prediction error covariance matrix $P(\tau+1|\tau)$ by (17) with $\tau = i\mu$, where i is an integer, is given by the upper $(\rho K N_{CC}) \times (\rho K N_{CC})$ matrix of the one step prediction error covariance matrix of the augmented system (46). From this we can calculate the estimation error covariance matrices for $\tau = \{i\mu, i\mu+1, ..., 2i\mu-1\}$ and the one step prediction error covariance matrices for $\tau = \{i\mu+1, ..., 2i\mu-1\}$ through (17), (19) and (22). For more details and proof, see Appendix 4.G of [38].

D Estimation of the process noise covariance matrix

To validate our choice of calculating the covariance matrix of the process noise Q through (25) we here compare this to an alternative where Q is instead calculated through finding a triangular matrix M that relates to Q by (26).

The matrix M is iteratively optimized using the interior point method[12]. For each new M, the covariance matrix Q is calculated and used in the Kalman filter by (16)-(20). 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

[11] Provided a large coherence bandwidth, an extension over a larger frequency range would provide improvements of the channel estimates at the cost of higher computational complexity in the Kalman filters. Such an extension is straightforward, see [42].

[12] 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.

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(11) 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.

Diag The matrix Q is first calculated through (25). 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 (25). Then the initial value of M is found through Cholesky decomposition of Q.

Figure 8 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 (24), 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.

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.

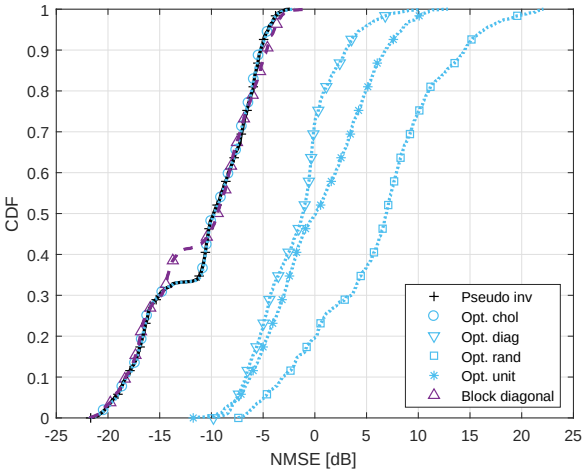

Figure 8: The CDFs of the NMSE provided by the reduced Kalman filter using estimates of the process noise matrix Q through (25)), denoted "Pseudo inv.", through optimiziation of an upper triangular matrix M in (26), denoted "Opt.", with different initaial values for M, and by using a block diagonal channel covariance matrix R_h to find a block diagonal Q through (24), denoted "Block diagonal".

References

[1] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE-Advanced Pro and The Road to 5G*. Academic Press, 2016.

[2] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Com. Mag.*, vol. 42, pp. 60–67, 2004.

[3] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel state information in MIMO downlink channels," *IEEE Trans. on Wireless Com.*, vol. 11, pp. 3037–3046, 2012.

[4] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. on Inf. Theory*, vol. 50, pp. 1893–1909, 2004.

[5] T. Marzetta, "How much training is required for multiuser MIMO?," in *14th Asilomar Conf. on Sig., Sys. and Comp.*, (Pacific Grove, USA), Oct.-Nov. 2006.

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

[7] E. Björnsson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive mimo systems with non-ideal hardware: Energy efficiency, estimation and capacity limits," *IEEE Trans. on Inf. Theory*, vol. 60, pp. 7112–7139, Nov. 2014.

[8] E. Björnsson, E. Larsson, and T. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Com. Mag.*, vol. 54, pp. 114 – 123, Feb. 2016.

[9] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. of IEEE VTC 2001*, (Rhodes, Greece), May 2001.

[10] S. A. Jafar and A. J. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. of IEEE International Symposium on Inf. Theory*, 2002.

[11] 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 Jour. on Select. Areas Com.*, vol. 28, pp. 1380–1408, 2010.

[12] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, and T. Svensson, "The role of small cells, coordinated multipoint and massive MIMO in 5G," *IEEE Com. Mag.*, vol. 52, pp. 44–51, May 2014.

[13] Z. Ma, Z. Q. Zhang, Z. G. Ding, P. Z. Fan, and H. C. Li, "Key techniques for 5G wireless communications: network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, 2015.

[14] Ericsson AB, "5G energy performance," *Ericsson White Papers*, Apr. 2015.

[15] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Jour. of Sel. Topics in Sig. Proc.*, vol. 8, pp. 742–758, 2014.

[16] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced deployment: Scenarios and operational challenges," *IEEE Wireless Com. Mag.*, vol. 50, pp. 148–155, 2012.

[17] X. Tao, X. Xu, and Q. Cui, "An overview of cooperative communications," *IEEE Wireless Com. Mag.*, vol. 8, pp. 65–171, 2012.

[18] EU FP7-ITC-2009, ARTIST4G project report D1.4, "Interference avoidance techniques and system design," 2012.

[19] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance and field trial results," *IEEE Com. Magazine*, vol. 49, pp. 102–111, 2011.

[20] 3GPP TR 36.819 v11.0.0, "Technical specification group radio access network; Coordinated multi-point operation for LTE physical layer aspects, (release 11)," 2011.

[21] L. Thiele, M. Olbrich, M. Kurras, and B. Matthiesen, "Channel aging effects in CoMP transmission: Gains from linear channel prediction," in *45th Asilomar Conf. on Sig., Sys. and Comp.*, (Pacific Grove, USA), Nov. 2011.

[22] 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 Jour. on Wireless Com. and Netw.*, vol. 100, June 2014.

[23] R. Apelfröjd and M. Sternad, "Robust linear precoder for coordinated multipoint joint transmission under limited backhaul with imperfect channel state information.," in *Proc. of IEEE ISWCS 14*, (Barcelona, Spain), Aug. 2014.

[24] E. Björnsson, E. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?," *IEEE Trans. on Wireless Com.*, vol. 15, pp. 1293–1308, Feb. 2016.

[25] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. on Inf. Theory*, vol. 49, pp. 951–963, 2003.

[26] J.-C. Guey and L. D. Larsson, "Modeling and evaluation of MIMO systems exploiting channel reciprocity in TDD mode," in *Proc. of IEEE VTC Fall 2004*, (Los Angeles, USA), Sept. 2004.

[27] Ericsson AB, "5G radio access," *Ericsson White Papers*, Apr. 2016.

[28] W. Zirwas, M. B. Amin, and M. Sternad, "Coded CSI reference signals for 5G - exploiting sparsity of massive MIMO radio channels," in *20th International ITG Works. on Smart Antennas (WSA 2016)*, (Munich), March 2016.

[29] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO system: Open-loop and closed-loop training with memory," *IEEE Jour. of Sel. Topics in Sig. Proc.*, vol. 8, pp. 802 – 814, 2014.

[30] S. Noh, M. D. Zolowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Jour. on Sel. Topics in Sign. Proc.*, vol. 8, pp. 787–801, 2014.

[31] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for massive MIMO systems," *IEEE Trans. on Com.*, 2017. In press, DOI 10.1109/TCOMM.2017.2691700.

[32] 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. of IEEE ICCC 2014*, (Shanghai, China), Oct. 2014.

[33] 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. on Com.*, vol. 64, pp. 1055–1067, 2016.

[34] 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. on Com.*, vol. 64, pp. 601–617, 2015.

[35] B. Tomasi, A. Decurninge, and M. Guillaud, "SNOPS: Short non-orthogonal pilot sequences for downlink channel state estimation in FDD massive MIMO," in *IEEE Globecom Works.*, (Washington DC, USA), Dec. 2016.

[36] M. Kurras, L. Thiele, T. Haustein, and C. Y. W. Lei, "Full dimension MIMO for frequency division duplex under signaling and feedback constraints," in *Sig. Proc. Conf. (EUSIPCO)*, (Budapest, Hungary), Aug./Sept. 2016.

[37] T. Ekman, *Predictions of Mobile Radio Channels.* PhD thesis, Uppsala University, Oct. 2002. http://www.signal.uu.se/Publications/ptheses.html.

[38] D. Aronsson, *Channel estimation and prediction for MIMO OFDM systems - Key design and performance aspects of Kalman-based algorithms.* PhD thesis, Uppsala University, March 2011. http://www.signal.uu.se/Publications/ptheses.html.

[39] W. Zirwas, M. Sternad, and R. Apelfröjd, "Key solutions for a massive MIMO FDD system," in *Proc. of IEEE PIMRC 17*, (Montreal, Canada), Sept. 2017.

[40] D. Aronsson and M. Sternad, "Kalman predictor design for frequency-adaptive scheduling of FDD OFDM uplinks," in *Proc. of IEEE PIMRC 07*, (Athens, Greece), Sept. 2007.

[41] W. Mennerich, M. Grieger, W. Zirwas, and G. Fettweis, "Interference mitigation framework for cellular mobile radio networks," *International Journal of Antennas and Propagation*, 2013.

[42] R. Apelfröjd, "Kalman predictions for multipoint OFDM downlink channels," 2014. http://www.signal.uu.se/Publications/pdf/c1402.pdf.

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

[44] L. Lindbom, "Simplied Kalman estimation of fading mobile radio channels: high performance at LMS computational load," in *Proc. of IEEE ICASSP 93*, (Pacific Grove, USA), Apr. 1993.

[45] C. Komnimakis, C. Fragouli, A. Sayed, and R. Wesel, "Multi-input multi-output fading channel tracking and equalization using kalman estimation," *IEEE Trans. on Sig. Proc.*, vol. 50, pp. 1065–1076, 2002.

[46] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunström, "Towards system beyond 3G based on adaptive OFDMA transmission," *Proc. of IEEE*, vol. 95, pp. 2432 – 2455, 2007.

[47] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. on Wirless Com.*, vol. 4, pp. 1650 – 1662, 2005.

[48] S. Jaekel, L. Raschkowski, K. Borner, and L. Thiele, "Quadriga: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. on Antennas and Propagation*, vol. 62, pp. 3242–3256, 2014.

[49] Fraunhofer Heinrich Hertz Institute, "Quasi deterministic radio channel generator user manual and documentation," 2016.

[50] J. Meinilä, P. Kyösti, L. Hentilä, T. Jämsä, E. Suikkanen, E. Kunnari, and M. Narandzic, "Winner+ final channel models," 2010.

[51] T. Eriksson and T. Ottosson, "Compression of feedback for adaptive transmission and scheduling," *Proceedings of the IEEE*, vol. 95, pp. 2314 – 2321, 2007.

[52] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE Jour. of Sel. Topics in Sig. Proc.*, vol. 8, pp. 815 – 830, 2014.

[53] S. M. Kay, *Fundamentals of Statistical Signal Processing. Volume 1: Estimation Theory*. Upper Saddle River: Prentice Hall, 1993.

[54] P. Stoica and T. Söderström, *System Identification*. Upper Saddle River: Prentice Hall, 2001.