Kalman Predictions for Multipoint OFDM Downlink Channels

Rikke Apelfröjd

Abstract

Coordinated Multipoint (CoMP) transmission provides high theoretic gains in spectral efficiency with coherent Joint Transmission (JT) to multiple users. However, this requires accurate Channel State Information at the Transmitter (CSIT). Unfortunately, coherent JT CoMP often is accompanied by long system delays, due to e.g. data sharing over backhaul links. Therefore, the CSIT will be outdated.

This report provides a detailed description on how to increase the accuracy of the CSIT by utilizing Kalman filters to predict Orthogonal Frequency Division Multiplexing (OFDM) downlink channels. The small scale fading of these channels are modeled by Auto Regressive (AR) models of finite order. The report includes descriptions on how to estimate these models based on past knowledge of the channel as well as analytical result on the predictability of such models. Different technical design aspects for deploying the Kalman filters in communication, such as pilot patterns, AR model estimations and the location of Kalman filters that predict downlink Frequency Division Duplex (FDD) channels, are also discussed.

The aim of the report is to in detail describe the prediction procedure used in previous work. Some of the results from this previous work are here presented and extended to provide a complete overview. All simulation results are based on measured channels.

The report also includes a description on how to model block-fading channels with a specified channel accuracy that would have been obtained with Kalman predictions. This model can then be used for system simulations.

Contents

1 Introduction

Shadowed areas and interference at cell borders pose challenges for future wireless broadband systems. A potentially powerful remedy would be Coordinated Multipoint (CoMP) transmission, which have presently been added as a study item to the 3GPP LTE standard in Release 11 [1].

CoMP techniques for downlink transmission are often categorized into two groups [2,3]; joint scheduling and/or joint beamforming and Joint Transmission (JT). The second group has the potential to provide the largest gains, see e.g. [3, 4], but is also more sensitive to inaccurate Channel State Information at the Transmitter side (CSIT). Near accurate CSIT is especially important for *coherent* JT CoMP [5, 6], also known a network Multiple-Input Multiple-Output (MIMO).

For Frequency Division Duplex (FDD) systems the CSIT is based on measurements of downlink pilot symbols, i.e. symbols that are known at both transmitter and receiver side. Depending on where the channel estimation is localized, either these measurements or the CSIT is quantized at the receiver side and fed back through the uplink to the base stations. For a centralized CoMP scheme, the CSIT is sent over backhaul links to a Control Unit (CU)[1]. In the CU, CoMP decisions, such as which beamforming weights to use, are made based on the CSIT and sent over backhaul to the cooperating base stations which then serve the users. This process is time consuming and therefore the estimated channel might not be valid at the time the messages are transmitted. This outdating of CSIT is especially severe for high mobility users [5].

The problem of outdated CSIT can to some extent be counteracted by channel predictions. Linear channel predictions through Kalman filtering have been suggested repeatedly in the literature and have shown to increase the accuracy of CSI, see e.g. [7, 8]. If the statistics of the channel (the channel model) is perfectly known, then the Kalman filter provides optimal prediction with respect to the Mean Squared Error (MSE) [9, 10].

In [11] Kalman predictions for Orthogonal Frequency Division Multiplexing (OFDM) MIMO channels where thoroughly investigated. In [12–14] these were further extended and utilized for CoMP systems. Results showed that channel predictions through Kalman filters provided large gains in CSIT accuracy. The predicted channels were then used together with robust linear precoding techniques for coherent downlink JT CoMP. The predictions were shown to be sufficient to ensure high gains for JT CoMP as compared to single cell transmission with reuse 1.

Kalman filters have the benefit that they provide, not only the the optimal prediction, but also the second order statistics of the prediction errors. These are used in the robust precoder design in [12–14] to ensure that poorly predicted channels do not ruin the possibilities of having CoMP gains. They could also be used for other applications. Such an application could be adjust the quantization granularity in adaptive quantization schemes. The channels that are hard to predict would then be given a lower quantization granularity than the channels that are easy to predict. If the feedback capacity is limited by a total number of quantization bits, this type of adaptive quantization scheme allows differently strong channels to be represented with balanced accuracy levels. The quantization errors can then be controlled to have roughly the same order of magnitude as the prediction errors, thereby lowering the additional impact of the quantization error on the precoder performance.

Channel predictions with Kalman filters can, as was shown in [12, 14], push the limit of when users

[1] The CU is a logical entity that is responsible for making the CoMP choices. It may be located at one or more of the base stations within the coordinated cluster of base stations.

may benefit from coherent JT CoMP. However, the benefits from channel predictions are not limited to JT CoMP applications. Near accurate CSIT is also important for multi-user MIMO gains [15]. Moreover, in [16] it was shown that accurate CSI enable the gain of adaptive modulation and coding schemes. The covariance information obtainable by Kalman prediction can also be used to adjust the adaptive schemes so that target bit error rates are attained also in the presence of channel uncertainty [17].

The small scale fading of a radio channel is a result of a user moving through a standing wave pattern. This wave pattern can be described by a sum of damped sinusoids. Many schemes for linear prediction are based on the propagation of ideal sinusoids, see e.g. [18]. In reality the sinusoid components of measured channels are not perfect, but damped. The Kalman filter is based on a state space channel model that includes damped (or ideal) sinusoids and is the optimal linear predictor, under either of these assumptions.

The dampening of the sinusoids limits the predictability of small scale fading channels based on the fading statistics [19]. For prediction horizons beyond these limitations on predictability other solutions must be used. One such solution is to place a predictor antenna in front of the receive antenna in the direction of travel of the user [20]. The predictor antenna then "looks into future". This solution is ideal for vehicles where antennas can be placed on the roof, but impractical for e.g. pedestrian user as it would then be difficult to place the predictor antenna correctly. In the here presented simulations of Kalman predictions without a separate predictor antenna, we therefore focus on predictions for low mobility users such as pedestrians or cyclists.

1.1 Contributions

This report provides a detailed description of the Kalman predictor that is used in [12–14]. In order to provide a clear overview, some of the results in [11] are also included. The measurement based prediction results of [12, 14] are included and presented in more detail. These results include prediction performance for different prediction horizons, with different levels of intracluster noise and with different pilot patterns.

System delays, due to prediction calculations, feedback and backhaul transmission of CSIT, sharing of payload data over backhaul, precoder calculations and transmission of precoding weights over backhaul, will cause outdating of CSIT. The effect of the system delay on the CSIT is here studied, both with and in the absence of Kalman filters.

The small scale fading of channels is represented by Auto Regressive (AR) models. The theoretical results on how to calculate the predictability of such channels are here provided.

For CoMP downlink transmission, base stations are assumed to be partitioned into cooperation clusters. The base stations within a cluster would use CoMP to control the intracluster interference, while the base stations outside a cluster cause intercluster interference. There is a trade-off in the cluster size. A large cluster size will provide less intercluster interference, but cause larger system delays due to e.g. backhaul via multiple links and nodes. A large system delay will in turn cause more outdating of the CSIT, which will cause higher intracluster interference. Fortunately, reasonably sized clusters, with respect to system delays, can be formed if the intercluster interference is limited through different schemes, see. e.g. [21–23]. In Section 5, intercluster interference suppression is shown to be very important for prediction performance.

An issue with CoMP, is that the total number of transmit antennas in the cluster will be large. If all antennas are to transmit downlink pilots on orthogonal time-frequency resources of an OFDM downlink, then the pilots must either be transmitted very sparsely or the pilot overhead will be large. In [24] a code-orthogonal pilot pattern was suggested for multi-user uplink transmission. Pilots from different terminals are then transmitted on the same resources using Code-Division Multiple Access (CDMA) techniques. This

pattern was adjusted for downlink MIMO transmission in [11]. Here, the effect of code-orthogonal and resource-orthogonal pilot patterns are investigated. Based on these investigations, a solution is suggested in which pilots from transmitters at the *same* base station should be code-orthogonal, while transmitters from *different* base stations should use resource-orthogonal pilots.

The simulations results in this report are all base on channel measurement. However, the report also includes an appendix describing how to simulate predicted channels and prediction errors in a block-fading simulator under the assumption that Kalman filters are used for predictions. Such simulated channels were used for the comparison of different CoMP schemes in [5]. The main idea behind them is to ensure that predictions and prediction errors are uncorrelated, as is the case with optimal predictors. A similar concept is used in [25].

1.2 Structure of the report

In Section 2, the mathematical model of short term fading, used in this report, is described. Analytical results on the predictability of such models is provided. The section also includes a description on the Kalman filter and prediction equations. These are briefly motivated, but for a detailed derivation, readers are referred to the extensive literature on the topic, e.g. [9, 10].

The problem of estimating the mathematical models of the small scale fading for MIMO CoMP channels is described in Section 3. The description on how to estimate channel models for MIMO channels in the frequency domain, Section 3.1-3.2, follows [11], with the exception that here the MIMO channels are not assumed to be uncorrelated. The model is then extended to include channels from base stations at multiple sites, Section 3.3.

Section 4 includes discussions on different design issues when estimating the mathematical models of Section 3 and when deploying the Kalman filters. These discussions include

1. The model order for the models of fading channels,
2. The frequency with which the models should be re-estimated,
3. The issues and benefits of jointly predicting several correlated subchannels,
4. How to initiate the Kalman filters,
5. How to design the pilot patterns,
6. Where to locate the channel predictors.

In Section 5 and Appendix B, simulations for different prediction scenarios are presented. These simulations are based on 20 MHz downlink channels that were measured in an urban environment at a carrier frequency of 2.66 GHz, which is typical for 3GPP LTE systems.

The conclusions drawn from the simulations in Section 5 are summarized in Section 6.

Finally, Appendix A includes a description on how to simulate channel predictions in a block-fading model with the assumption that they were obtained by Kalman predictors in a prescribed fading environment.

1.3 Notations and assumptions

Assume that a cluster of N_{BS} base stations, each with N_{tx} antennas, are participating in CoMP transmission. The object is to estimate and predict the multipath OFDM channels between them and a single user with N_{rx} receive antennas. Also, assume N_k pilot-bearing subchannels with negligible Inter Symbol Interference (ISI) and Inter Carrier Interference (ICI). The channel between one transmit antenna and one receiver at one subcarrier can then be represented as a scalar complex-valued gain. The equations are all in discrete time where each time step is defined by the pilot spacing in time.

Unless otherwise specified, capital letters refer to matrices and lower case letters to column vectors. Sub-indices j, n, m, k will indicate the part of a matrix that is associated with base station $j = 1, ..., N_{BS}$, transmit antenna $n = 1, ..., N_{tx}$, receive antenna $m = 1, ..., N_{rx}$ and subcarrier $k = 1, ..., N_K$, respectively.

The conjugate transpose of a matrix A is denoted A^* and $\operatorname{diag}\{A_i\}_{i=1,...,N}$ indicates a block diagonal matrix

$$\begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_N \end{bmatrix}.$$

The trace of a matrix is denoted $\operatorname{tr}\{A\}$, while $\odot$ and $\oslash$ denote element-wise multiplication and element-wise division respectively. An estimate of a vector a at the time t is denoted $\hat{a}(t|t-\tau)$ if it is based on measurements available for all past data up until time $t-\tau$. Expectations, $E[\cdot]$, are over the statistics of noise.

2 Linear Filtering

The Kalman filter was first introduced by Rudolf Kalman [9] in 1960. It is the optimal linear filter with respect to minimizing the Mean Square Error (MSE) of the estimate and can be derived from the Wiener filter [10]. Both the Kalman and Wiener filters are optimal linear filters. Their difference lies in the execution. The Kalman filter is based on the state space model of the system and, although each estimate is based on all available measurements, only the latest measurement is needed to update the estimate in every step. Therefore it is well suited for systems that cannot store large amounts of measurement data, such as mobile phones where memory allocated to baseband signal processing applications is limited. Furthermore, per-subframe processing over time is preferred to processing that uses large sets of past data, due to complexity and implementation reasons. The Kalman filter also provides the covariances of the estimates as part of the updates. For the Wiener filter, these must be calculated separately.

2.1 Short term fading models

Ideally, a set of fading channels can be described as a sum of *perfect* sinusoid representing the standing wave pattern of the environment, due to reflections from discrete reflectors or scatterers. In reality, the fading of the channels is often better described by a sum of *damped* sinusoid and a process noise, see [26]. Use of process noise with constant variance will then assure that the model is wide sense stationary, i.e. that the statistical properties of the channel model does not change over time. This can be mathematically expressed as an Auto-Regressive (AR) model or, more generally, as an Auto-Regressive Moving-Average (ARMA) model, on state space form

$$\begin{aligned} x(t+1) &= Ax(t) + Be(t), \\ h(t) &= Cx(t). \end{aligned} \tag{1}$$

Here, $x(t)$ is a state vector and A, B and C are state matrices. The vector $h(t)$ includes the MIMO channels between a user with N_{rx} antennas and N_{BS} base stations, each with N_{tx} transmit antennas. The process noise $e(t)$ used in this model is assumed white, i.i.d. zero mean and independent with respect to the state vector. It has a covariance matrix

$$Q = E\left[e(t)e(t)^*\right]. \tag{2}$$

The AR or ARMA model of (1)-(2) is a mathematically correct representation of a sum of dampened sinusoid if its poles are all inside the unit circle, i.e. if the system is stable. The ideal case with undamped sinusoids is described if the poles are exactly on the unit circle. The process noise will go to zero if all poles are moved towards the unit circle. This means that the state vector that describes a sum of sinusoids will regenerate itself, with no unpredictable exogenous influence $e(t)$. That, in turn, implies that if the current state is exactly known, then undamped sinusoids with perfectly known properties can be perfectly predicted infinitely into the future.

In this report, the focus will be on AR models, as the complexity with solving these, based on measured training data, is lower than that associated with solving ARMA models.

2.2 Predictability of Auto-Regressive models

The predictability of a channel that follows the model of (1) depends on the state matrices and on the covariance matrix of the process noise. The one step prediction of the channel $h(t)$ in (1) is given by

$$\hat{h}(t+1|t) = C\hat{x}(t+1|t). \tag{3}$$

Here $\hat{x}(t+1|t)$ is the one step prediction of the state vector. Assume that we have full knowledge of the channel up until time t, then $x(t)$ is known. As $e(t)$ is white and independent, the best estimate (that which maximizes the entropy) of $e(t)$ is given by its mean, i.e. $\hat{e}(t) = E[e(t)] = 0$. Then, the best estimate of the state vector is

$$\hat{x}(t+1|t) = Ax(t) + \hat{e}(t) = Ax(t), \tag{4}$$

and for the channel vector it is

$$\hat{h}(t+1|t) = C\hat{x}(t+1|t) = CAx(t). \tag{5}$$

As this prediction maximize the entropy, there is, on average, no prediction that performs better than (5). Hence, if the model (1), (2) exactly describes the evolution of the channel vector $h(t)$, then the covariance matrix of the optimal one step prediction error vector is through (1)-(5) given by

$$\begin{aligned} & E\left[\left(\hat{h}(t+1|t) - h(t+1)\right)\left(\hat{h}(t+1|t) - h(t+1)\right)^*\right] \\ =& E\left[(-CBe(t))\,(-CBe(t))^*\right] \\ =& CBQB^*C^*. \end{aligned} \tag{6}$$

With a similar reasoning the covariance matrix of the optimal τ-step prediction error is given by

$$\begin{aligned} & E\left[\left(\hat{h}(t+\tau|t) - h(t+\tau)\right)\left(\hat{h}(t+\tau|t) - h(t+\tau)\right)^*\right] \\ =& C\left(\sum_{i=0}^{\tau-1} A^i BQB^* \left(A^i\right)^*\right) C^*. \end{aligned} \tag{7}$$

Figure 1 shows the Normalized Mean Squared Error (NMSE) of the theoretical prediction error with optimal prediction and known state vector $x(t)$ as given by (7). The two curves show the achievable performance for predictions of a scalar channel component $h(t)$ with unit variance when the AR model in (1) has model order four with poles in $0.91 \pm 0.35i$ and $0.86 \pm 0.33i$ or in $0.82 \pm 0.29i$ and $0.70 \pm 0.10i$ respectively. Other parameters are stated in Table 1. These two models give rise to one very spiky spectrum (resembling the Jakes' Doppler spectrum) and one flat spectrum as shown in Figure 2. The results are shown as a function of the spatial prediction horizon expressed in number of carrier wavelengths d_λ. This can be translated into a prediction horizon L in time through

$$L = \frac{d_\lambda \cdot c}{v \cdot f_c}, \tag{8}$$

where c is the speed of light, v is the users velocity and f_c is the carrier frequency. This means that if the prediction horizon in time, through the system delay, is fixed, then a fast user will have to use a longer spatial prediction horizon than a slow user and hence experience a worse prediction performance.

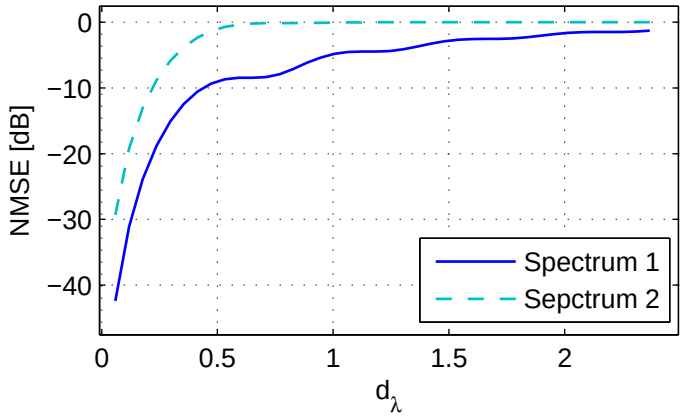
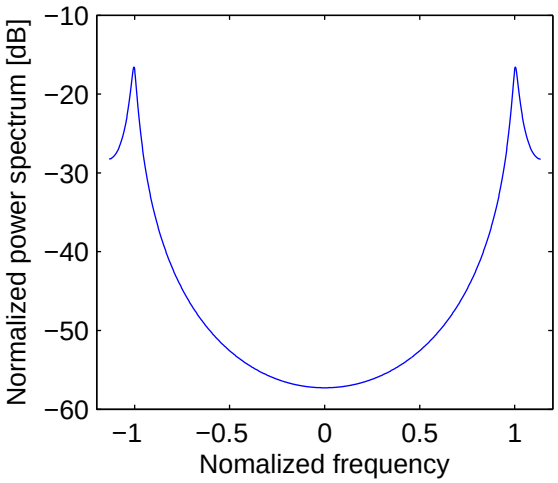


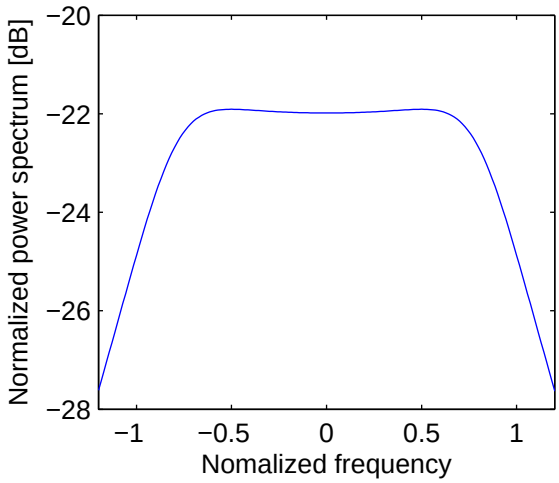
Figure 1: The theoretic predictability of AR models given by (7). The fourth order models have the poles and spectra shown in Figure 2. Prediction horizons d_λ are expressed in terms of number of wavelengths in space.

Table 1: Parameter setting for theoretical results.

Parameter	Value
Carrier frequency	3.7 GHz
Bandwidth	50 MHz
Number of subcarriers	2048
User velocity	50km/h
Pilot spacing in time	345.6 μs
Pilot spacing in frequency	24 kHz



(a) Spectrum 1: "Jakes'-like" spectrum with poles in $0.91 \pm 0.35i$ and $0.86 \pm 0.33i$.



(b) Spectrum 2: "Flat" spectrum with poles in $0.82 \pm 0.29i$ and $0.70 \pm 0.10i$.

Figure 2: The Doppler spectra of the AR models used for Figure 1. Both spectra are normalized to have maximum Doppler frequency 1 and total power 1.

Note also, that as the poles move closer to the unit circle the spectrum gets more distinct peaks and the predictability increases. This is because the signal more closely resembles a finite sum of "pure" sinusoids, which are easier to predict.

The curves in Figure 1 provide an upper bound on predictions performance. Errors in the AR model, or the use of noisy measurements of the channels up until time t, will decrease the performance.

2.3 Kalman filter and predictor

In this section the Kalman equations are stated with a brief motivation. For the detailed derivations of these, please refer to some of the extensive literature on the subject, e.g. [10].

Assume that the statistical properties of the small scale fading can be described by (1)-(2). The filter input is a measurement or estimate of the channels at time t, described by

$$y(t) = \Phi(t)h(t) + n(t), \tag{9}$$

where $\Phi(t)$ is a matrix consisting of zeros and of pilot symbols, i.e. symbols that are known at the receiver and transmitter side, and $n(t)$ is a measurement noise term. The noise generally consists of interference from base stations that are outside the cooperation cluster, thermal noise and additional background noise. The covariance matrix of the measurement noise is

$$R = E\left[n(t)n(t)^*\right]. \tag{10}$$

A rough estimate of this noise covariance matrix can be gained by e.g. using a smoothed estimate of the channel $\hat{h}(t|t+\tau)$ and calculating the residual $y(t) - \Phi(t)\hat{h}(t|t+\tau)$ as an estimate of the noise $n(t)$. However, this estimate might be very coarse, which can potentially affect the estimate of R and

the prediction performance negatively. A more sophisticated method to gain an estimate of the noise variances is described in [27]. A third alternative is to use time-frequency resource blocks without signal transmission, and estimate noise properties from the received signals in these blocks.

The Kalman filter is recursive, so for now we will assume that all measurements $y(0)$ up until $y(t-1)$, as well as an Minimum Mean Squared Error (MMSE) estimate of the state vector, $\hat{x}(t-1|t-1)$, and the corresponding error covariance matrix

$$P(t-1|t-1) = E\left[\left(x(t-1) - \hat{x}(t-1|t-1)\right)\left(x(t-1) - \hat{x}(t-1|t-1)\right)^*\right], \tag{11}$$

are available. For a discussion on how to initiate the filter please see Section 4.4.

As the process noise is assumed zero mean, the one step MMSE prediction of the state vector is given by

$$\hat{x}(t|t-1) = A\hat{x}(t-1|t-1). \tag{12}$$

The corresponding covariance matrix is then

$$\begin{aligned} P(t|t-1) &= E\left[\left(x(t) - \hat{x}(t|t-1)\right)\left(x(t) - \hat{x}(t|t-1)\right)^*\right], \\ &= AE\left[\left(x(t-1) - \hat{x}(t-1|t-1)\right)\left(x(t-1) - \hat{x}(t-1|t-1)\right)^*\right]A^* + BQB^*, \\ &= AP(t-1|t-1)A^* + BQB^*, \end{aligned} \tag{13}$$

where the second equality holds if the estimate of the state vector is uncorrelated with the process noise at time $t-1$. This holds as $\hat{x}(t-1|t-1)$ is estimated based on the measurements $y(0), ..., y(t-1)$ which only depend on $e(0), ..., e(t-2)$. From the covariance matrix (13) we now calculate the Kalman gain matrix K as

$$K = P(t|t-1)J^*\left(R + JP(t|t-1)J^*\right)^{-1} \quad , \tag{14}$$

where $J = \Phi(t)C$. Through this we can calculate an updated estimate of the state vector and the corresponding estimation error as soon as the next measurement, $y(t)$, is available. These are the given by

$$\begin{aligned} \hat{x}(t|t) &= \hat{x}(t|t-1) + K\left(y(t) - J\hat{x}(t|t-1)\right), \\ P(t|t) &= \left(I - KJ\right)P(t|t-1). \end{aligned} \tag{15}$$

If the Kalman gain is high (has large elements), due to the measurement noise $n(t)$ having small variance compared with the one step prediction error, then the updated estimate will be based mostly on the latest measurement $y(t)$ [10]. If the measurement noise has large variance compared to the one step prediction error variance then the Kalman gain will be low and the updated estimate will mostly be based on the one step prediction that uses the model based extrapolation (12).

The τ step MMSE prediction follows the same reasoning as the one step prediction in (12) and is given by

$$\begin{aligned} \hat{x}(t+\tau|t) &= A\hat{x}(t+\tau-1|t), \\ &= A^2\hat{x}(t+\tau-2|t), \\ &= A^\tau\hat{x}(t|t). \end{aligned} \tag{16}$$

Note that if the system is stable then $\hat{x}(t+\tau|t) \to E\left[x(t)\right]$ as $\tau \to \infty$. The covariance matrix for the prediction error can be calculated recursively, beginning with $P(t|t)$ from (15) through

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

for $t_i = 1, ..., \tau$ resulting in

$$P(t+\tau|t) = A^\tau P(t|t) \left(A^*\right)^\tau + \sum_{i=0}^{\tau-1} A^i B Q B^* \left(A^*\right)^i. \tag{18}$$

The τ step prediction of the channel is, through (1) and (16),

$$\hat{h}(t+\tau|t) = C\hat{x}(t+\tau|t) = CA^\tau \hat{x}(t|t), \tag{19}$$

and the corresponding covariance matrix can be calculated from (17) as

$$\begin{aligned} P_h(t+\tau|t) &= E\left[\left(h(t) - \hat{h}(t+\tau|t)\right)\left(h(t) - \hat{h}(t+\tau|t)\right)^*\right], \\ &= CE\left[\left(x(t) - \hat{x}(t+\tau|t)\right)\left(x(t) - \hat{x}(t+\tau|t)\right)^*\right] C^*, \\ &= CP(t+\tau|t)C^*. \end{aligned} \tag{20}$$

2.3.1 The stationary filter

A property of the Kalman filter is that for time invariant models, i.e. for time-invariant matrices A, B, C and Q in (1), (2) and R and $J = \Phi(t)C$ in (14), as $t \to \infty$, the covariance matrices $P(t|t-1)$ and $P(t|t)$ given by (13) and (15) converge as $P(t|t) \to P_f$, $P(t+1|t) \to P_p$. This state is referred to as the stationary state. Combining equations (13)-(15) we then see that the stationary covariance matrix of the one step prediction error then obeys the algebraic Riccati equation

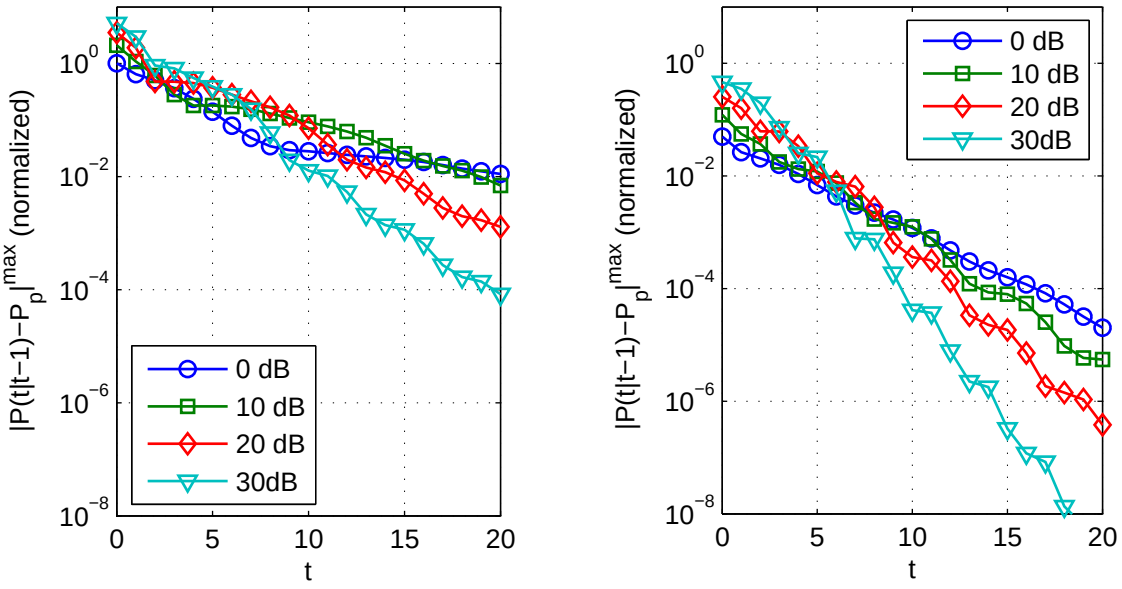
$$P_p = AP_pA^* + BQB^* - AP_pJ^*\left(JP_PJ^* + R\right)^{-1} JP_pA^*. \tag{21}$$

The stationary estimation error covariance matrix is then given by

$$P_f = P_p - P_pJ^*\left(JP_pJ^* + R\right)^{-1} JP_p. \tag{22}$$

The corresponding stationary Kalman filter gain will be a constant matrix, obtained by substituting P_p for $P(t|t-1)$ in (14). The prediction performance of a Kalman predictor will be significantly worse during its initial transient, before it has reached its steady state. How fast the filter converges to the stationary state depends on the placement of the poles of (1) and the SNR of the signal in (9). This is illustrated in Figure 3 where $P(t|t-1)$ is calculated recursively from (13) and compared with its stationary value P_p calculated through (21), for the settings in Table 1. The graphs show the maximum norm of the matrix $(P(t|t-1) - P_p)$ normalized with the maximum norm of P_p as a function of time for different measurement SNR and for the two different systems characterized by the Doppler spectra in Figure 2. The filter with the flat spectrum of 2b shows faster convergence that that with the Jakes'-like spectrum of Figure 2a. Also, the filters converge faster when the SNR is high.

If the symbol matrix $\Phi(t)$ in (9) is cyclic with a period T, then the covariance matrices will instead converge towards a cyclic function of time, i.e. $P(t|t) \to P_f(t) = P_f(t+T)$ and $P(t+1|t) \to P_p(t) = P_p(t+T)$ [11].



(a) Spectrum 1.

(b) Spectrum 2.

Figure 3: The maximum norm of $P(t|t-1) - P_p$ normalized with the maximum norm of P_p. Results are shown for different SNR of the measurement signal (9). The system in (1) is a fourth order AR-model with the poles in $0.91 \pm 0.35i$ and $0.86 \pm 0.33i$ (right) and in $0.82 \pm 0.29i$ and $0.70 \pm 0.10i$ (left) respectively, i.e. with the Doppler spectra of Figure 2. The initial covariance matrix, $P(0|0)$, is given by (50). Note that these results are theoretical.

3 Estimating AR models for multipath CoMP channels

The estimates and predictions of Section 2.3 assume that the statistics of the channel described by (1) is perfectly known. In reality these properties must be estimated. The estimated models will be based on an finite set of past channels which may not be exactly known. This will lead to a residual error not included in the analytical expressions of (15), (17) and (20). In applications where CoMP decisions are based not only on the channel estimate, but also the covariance matrix of the estimation error, such as the robust linear precoder in [12, 13], this can cause problems. It is therefore important that the channel model is precise enough such that the residual error due to model inaccuracies are much smaller than the prediction errors. Alternatively, the covariance matrices used for precoder design must be modified to include also these residual errors.

The Kalman filter can be realized in the time domain or in the frequency domain. Depending on which of these that is chosen, the process noise covariance matrix in (2) will be different. This report focuses on how to estimate AR models for Kalman filtering in the frequency domain. The main results from Chapter 4 of [11] are included in this section. For details on derivations on how to estimate an AR model for Kalman filtering in the time domain approach, please see [11].

For the models in this section we shall make the following assumptions:

1. All MIMO channels between a user and a base station are identically distributed. This is reasonable as the signal propagation is very similar for signals transmitted from closely spaced antennas at the same base station.

2. All subcarriers have identically distributed channels. Generally, higher frequencies interact more with matter and therefore experience higher path loss, but when the carrier frequency $\gg$ the bandwidth, a frequency independent distribution is a good approximation.

3. The correlation between channels on two subcarriers depends only on the separation (in Hz) of the subcarriers, not on their individual carrier frequencies. As for the second assumption, this is a good approximation if the carrier frequency $\gg$ the bandwidth.

4. The small scale fading of the channels from different base stations are independent. This is also a reasonable assumption, as the signals from different base stations will travel very different and widely spaced paths to the user[2].

5. The channels are wide sense stationary. This is approximately true over small time intervals for the short term fading. It is necessary for using the time independent state matrices in (1).

6. All channels are zero mean. Note that for Rician channels with strong Line of Sight (LOS) components, the total channel is $h_{tot}(t) = h(t) + \bar{h}(t)$ where $h(t)$ is a zero mean component which is traced by the Kalman filter and $\bar{h}(t)$ is the mean of $h_{tot}(t)$, which could be estimated and added separately. In practice, we mostly do not track LOS components separately, but let them be included in the AR model as a strong spectral peak close to the Doppler frequency zero.

[2]Note that the same reasoning does not apply to the shadow fading, which is likely to be correlated for different base stations. For example, along a road more than one base station is likely to have good signal strength while all base stations may have poor signal strength in a courtyard.

3.1 State space modeling for one channel component

Let $h_{jnmk}(t)$ denote the fading scalar channel component from the n'th transmitter at base station j to the m'th receiver of the user at pilot-bearing subcarrier number k. We can then fit a n_{AR} order AR model to that particular channel as

$$h_{jnmk}(t) = -a_1 h_{jnmk}(t-1) - ... - a_{n_{AR}} h_{jnmk}(t-n_{AR}) + e_{jnmk}(t). \tag{23}$$

Here, the coefficients $\{a_i\}$ are complex-valued constants, denoted the AR coefficients, and the process noise $e_{jnmk}(t)$ is zero mean, scalar, complex-valued and independent over time. There can, however, be dependency between process noise at different subcarriers, receive antennas and transmit antennas.

The AR model (23) can be transformed to the discrete temporal frequency domain (z transform domain) as

$$h_{jnmk}(z) = \frac{z^{n_{AR}}}{z^{n_{AR}} + a_1 z^{n_{AR}-1} + ... + a_{n_{AR}}} e_{jnmk}(z), \tag{24}$$

or

$$h_{jnmk}(z) = \frac{z^{n_{AR}}}{(z-p_1)(z-p_2)...(z-p_{n_{AR}})} e_{jnmk}(z), \tag{25}$$

where $\{p_i\}$ are the poles of the AR model. Furthermore, the expression in (23) can be rearranged to state space form as

$$\begin{aligned} x_{jnmk}(t+1) &= A_{jnmk} x_{jnmk}(t) + B_{jnmk} e_{jnmk}(t), \\ h_{jnmk}(t) &= C_{jnmk} x_{jnmk}(t), \end{aligned} \tag{26}$$

where the dimensions are given by

$$\begin{aligned} x_{jnmk}(t) &\in \mathbb{C}^{n_{AR} \times 1}, \\ A_{jnmk} &\in \mathbb{C}^{n_{AR} \times n_{AR}}, \\ B_{jnmk} &\in \mathbb{C}^{n_{AR} \times 1}, \\ C_{jnmk} &\in \mathbb{C}^{1 \times n_{AR}}. \end{aligned}$$

Based on the AR coefficients in (23), the state matrices of (26) can be arranged on different standard (canonical) state space forms such as diagonal, controllable or observable form, see e.g. [28]. Here, the diagonal form is chosen, as it will result in lower computational complexity in the Kalman filters, see [11]. The poles $\{p_i\}$ can be obtained from the channel coefficients in (23), see Section 3.2 for details. In the following, assume that all poles are distinct, i.e. $p_i \neq p_l$ for any $l \neq i$. Under that assumption, we can create a state-space representation on diagonal form:

$$\begin{aligned} A_{jnmk} &= \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & p_{n_{AR}} \end{bmatrix}, \\ B_{jnmk} &= \begin{bmatrix} (p_1-p_2)^{-1} (p_1-p_3)^{-1} .... (p_1-p_{n_{AR}})^{-1} \\ \vdots \\ (p_{n_{AR}}-p_1)^{-1} (p_{n_{AR}}-p_2)^{-1} .... (p_{n_{AR}}-p_{n_{AR}-1})^{-1} \end{bmatrix}, \\ C_{jnmk} &= \begin{bmatrix} p_1^{n_{AR}-1} & \cdots & p_{n_{AR}}^{n_{AR}-1} \end{bmatrix}. \end{aligned} \tag{27}$$

If some poles are close to each other ($p_i \approx p_l$), then some of the elements of B_{jnmk} will be very large which may cause computational problems. We might then replace B_{jnmk} and C_{jnmk} with $\frac{1}{\beta}B_{jnmk}$ and βC_{jnmk} where β is a scaling constant chosen to normalize e.g. the largest element in B_{jnmk}. This scaling does not change the solution with respect to (23).

If the poles are not distinct, then the Jordan form state space model, which is a generalization of the diagonal form, can be used instead [28]. However, from our studies of measured radio channels we have seen that in practice multipoles ($p_i = p_l$) do not occur in low order AR models of their short term fading. Nevertheless, for very spiky spectra the poles may tend to be placed very close to each other.

3.2 Estimating the poles

Multiplying both sides of the expression in (23) with $h^*_{jnmk}(t-\tau)$ for $\tau = 1, ..., n_{AR}$ will provide a set of n_{AR} equations for any given t. Taking the expectation over both sides of these gives

$$\begin{bmatrix} r_{h,jnmk}(1) \\ r_{h,jnmk}(2) \\ \vdots \\ r_{h,jnmk}(n_{AR}) \end{bmatrix} = - \begin{bmatrix} r_{h,jnmk}(0) & r_{h,jnmk}(-1) & \cdots & r_{h,jnmk}(1-n_{AR}) \\ r_{h,jnmk}(1) & r_{h,jnmk}(0) & & r_{h,jnmk}(2-n_{AR}) \\ \vdots & & \ddots & \vdots \\ r_{h,jnmk}(n_{AR}-1) & r_{h,jnmk}(n_{AR}-2) & \cdots & r_{h,jnmk}(0) \end{bmatrix} a, \tag{28}$$

where $a = \begin{bmatrix} a_1 & \cdots & a_{N_{AR}} \end{bmatrix}^T$ and $r_{h,jnmk}(\tau) = E\left[h_{jnmk}(t)h^*_{jnmk}(t-\tau)\right]$ is the autocorrelation of the channel $h_{jnmk}(t)$ over time.

Assume that we have access to a set of channel estimates $\tilde{h}_{jnmk}(t_1), \tilde{h}_{jnmk}(t_1+1), ..., \tilde{h}_{jnmk}(t_2)$ which may e.g. be estimates of the channel components, $\hat{h}_{jnmk}(t|t)$, calculated from (15) or smoothed estimates $\hat{h}_{jnmk}(t|t+\tau)$, see [10]. These will be referred to as the training sequence. From the training sequence we can estimate the autocorrelation function of each channel component over time through

$$\hat{r}_{h,jnmk}(\tau) \approx \frac{1}{t_2 - t_1 - \tau} \sum_{t=t_1+\tau}^{t_2} \tilde{h}_{jnmk}(t)\tilde{h}^*_{jnmk}(t-\tau). \tag{29}$$

As the different subchannels are assumed to follow the same statistical distribution, the estimate of the autocorrelation can be improved upon if all channels over all subcarriers are taken into account. The same argument can be applied for the MIMO channels, leading to an autocorrelation estimate of

$$\hat{r}_{h,j}(\tau) = \frac{1}{N_{tx}N_{rx}N_k} \sum_{n=1}^{N_{tx}} \sum_{m=1}^{N_{rx}} \sum_{k=1}^{N_k} \hat{r}_{h,jnmk}(\tau). \tag{30}$$

The least square estimate of the coefficients in (23) can then be obtained using the Yule-Walker equations [29] and the autocorrelation functions as

$$\hat{a} = - \begin{bmatrix} \hat{r}_{h,j}(0) & \hat{r}_{h,j}(-1) & \cdots & \hat{r}_{h,j}(1-n_{AR}) \\ \hat{r}_{h,j}(1) & \hat{r}_{h,j}(0) & & \hat{r}_{h,j}(2-n_{AR}) \\ \vdots & & \ddots & \vdots \\ \hat{r}_{h,j}(n_{AR}-1) & \hat{r}_{h,j}(n_{AR}-2) & \cdots & \hat{r}_{h,j}(0) \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{h,j}(1) \\ \hat{r}_{h,j}(2) \\ \vdots \\ \hat{r}_{h,j}(n_{AR}) \end{bmatrix}, \tag{31}$$

where $\hat{a} = \begin{bmatrix} \hat{a}_1 & \cdots & \hat{a}_{n_{AR}} \end{bmatrix}^T$ is a vector with the estimates of the parameters $\{a_i\}$ in (23). Now, comparing (24) with (25) we see that through the coefficients the poles $\{p_i\}$ can be extracted by solving the equation

$$p(z) = z^{n_{AR}} + a_1 z^{n_{AR}-1} ... a_{n_{AR}} = 0. \tag{32}$$

Estimating the parameters through (30) and (31) is, in [11], called the autocorrelation method. In Chapter 8 of [11] two other least squares estimation methods, namely the covariance method and the modified covariance method are also described. These are similar to the method above but tend to place the poles closer to the unit circle and provide different end results in terms of prediction performance metrics. Which of the methods to use depends on the channels that are to be predicted. In [11] the different methods were evaluated on measurement data (similar to that described in Section 5.1). The autocorrelation method was then found to provide AR models that generate better prediction performance and, in the investigated cases, the results were closer to the theoretic performance of (20).

3.2.1 Sub-sampling with respect to the prediction horizon

The use of (30) and (31) will find the AR parameters that minimize the variance of the one step prediction error, obtained with the model, $h_{jnmk}(t) - \hat{h}_{jnmk}(t|t-1) = h_{jnmk}(t) + \sum_{i=1}^{n_{AR}} a_i \hat{h}_{jnmk}(t-i)$. However, the end goal is to minimize the *τ'th step* prediction. Then the pole estimation of (30)-(32) might no longer be optimal from a prediction point of view. In fact, it was shown in [11] that the prediction performance can be improved by sub-sampling the model in (23) by a factor τ as

$$h_{jnmk}(t) = -b_1 h_{jnmk}(t-\tau) - ... - b_{n_{AR}} h_{jnmk}(t - \tau \cdot n_{AR}) + \epsilon_{jnmk}(t), \tag{33}$$

where $\epsilon_{jnmk}(t)$ is the residual. Estimation of such an AR model will tend to minimize the τ-step model prediction error. Use of such a subsampled model will therefore align the criterion used for model adjustment more closely with the intended purpose of the model, which is τ-step prediction.

Applying the above reasoning we can then obtain the estimates of the AR parameters $\hat{b} = \begin{bmatrix} \hat{b}_1 & \cdots & \hat{b}_{n_{AR}} \end{bmatrix}$ of the subsampled model (33) through

$$\hat{b} = - \begin{bmatrix} \hat{r}_{h,j}(0) & \cdots & \hat{r}_{h,j}((1-n_{AR})\tau) \\ \hat{r}_{h,j}(\tau) & & \hat{r}_{h,j}((2-n_{AR})\tau) \\ \vdots & \ddots & \vdots \\ \hat{r}_{h,j}((n_{AR}-1)\tau) & \cdots & \hat{r}_{h,j}(0) \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_{h,j}(\tau) \\ \hat{r}_{h,j}(2\tau) \\ \vdots \\ \hat{r}_{h,j}(n_{AR}\tau) \end{bmatrix}, \tag{34}$$

with covariance function values $\hat{r}_{h,j}(\tau)$ given by (30). The poles of the subsampled model, $p_{sub,i}$ are obtained as the roots of the polynomial $p_\tau(z) = z^{n_{AR}} + b_1 z^{n_{AR}-1} + ... b_{n_{AR}}$ and can then be translated to the poles of the model in (23) through

$$p_i = p_{sub,i}^{1/\tau}, \tag{35}$$

for $i = 1, 2, ... n_{AR}$.

With the approach outlined in this section, there is a tendency that the poles are placed very close to each other, especially in extreme cases such as e.g. if the fading statistics follows a perfect Jakes' model. This in turn leads to B_{jnmk} in (27) having some very large elements causing computational errors unless the state matrices are normalized as explained in Section 3.1.

3.3 State space modeling of channels from multiple base stations

As the same estimate of the autocorrelation is used for all subchannels and all OFDM MIMO channels between a base station and a user, the state matrices in (26) are the same for all $k = 1, ..., N_k$ subcarriers, $m = 1, ..., N_{rx}$ receive antennas and $n = 1, ..., N_{tx}$ transmit antennas. In addition, as the fading over the subcarriers are correlated, the prediction errors can be improved by jointly estimating and predicting w subcarriers. The N_k pilot-bearing channels will then be estimated by ${}^{N_k}/_{w}$ parallel filters which we shall index with $q = 1, ..., {}^{N_k}/_{w}$. For simplicity we will assume that the number of subcarriers per filter w is chosen such that N_k is divisible with w.[3]

Let the vector of w channel components from transmit antenna n at base station j to receive antenna m that are to be estimated and predicted by Kalman filter number q be denoted

$$h^q_{jnm}(t) = \left[\begin{array}{ccc} h_{jnm((q-1)w+1)}(t) & \cdots & h_{jnm(qw)}(t) \end{array} \right]^T. \tag{36}$$

We can then set up a state space model that describes the joint statistical properties of these w scalar channel components:

$$\begin{aligned} x^q_{jnm}(t+1) &= \operatorname{diag}\{A_{jnmk}\}_{k=1,...w}\, x^q_{jnm}(t) + \operatorname{diag}\{B_{jnmk}\}_{k=1,...w}\, e^q_{jnm}(t), \\ h^q_{jnm}(t) &= \operatorname{diag}\{C_{jnmk}\}_{k=1,...w}\, x^q_{jnm}(t), \end{aligned} \tag{37}$$

where

$$\begin{aligned} x^q_{jnm}(t) &= \left[\begin{array}{ccc} x_{jnm((q-1)w+1)}(t) & \cdots & x_{jnm(qw)}(t) \end{array} \right],^T \\ e^q_{jnm}(t) &= \left[\begin{array}{ccc} e_{jnm((q-1)w+1)}(t) & \cdots & e_{jnm(qw)}(t) \end{array} \right]^T. \end{aligned} \tag{38}$$

Similar, the AR models for the MIMO channel from the N_{tx} transmit antennas to the N_{rx} receive antennas with w parallel subchannels are given by

$$\begin{aligned} x^q_j(t+1) &= A_j x^q_j(t) + B_j e^q_j(t), \\ h^q_j(t) &= C_j x^q_j(t), \end{aligned} \tag{39}$$

where

$$\begin{aligned} x^q_j(t) &= \left[\begin{array}{cccc} \left(x^q_{j11}(t)\right)^T & \cdots & \left(x^q_{j1N_{rx}}(t)\right)^T & \cdots \left(x^q_{jN_{tx}N_{rx}}(t)\right)^T \end{array} \right]^T, \\ e^q_j(t) &= \left[\begin{array}{cccc} \left(e^q_{j11}(t)\right)^T & \cdots & \left(e^q_{j1N_{rx}}(t)\right)^T & \cdots \left(e^q_{jN_{tx}N_{rx}}(t)\right)^T \end{array} \right]^T, \\ h^q_j(t) &= \left[\begin{array}{cccc} \left(h^q_{j11}(t)\right)^T & \cdots & \left(h^q_{j1N_{rx}}(t)\right)^T & \cdots \left(h^q_{jN_{tx}N_{rx}}(t)\right)^T \end{array} \right]^T, \end{aligned} \tag{40}$$

[3] If for some reason N_k is not divisible with w then an upwards rounded number $\lceil {}^{N_k}/_{w} \rceil$ of Kalman filters may be used where e.g. the last two Kalman filter overlap. Alternatively, the filters may have different values of w, i.e. predict over a different number of subcarriers.

and

$$A_j = \text{diag}\left\{\text{diag}\left\{\text{diag}\left\{A_{jnmk}\right\}_{k=1,...w}\right\}_{m=1,...N_{rx}}\right\}_{n=1,...N_{tx}},$$
$$B_j = \text{diag}\left\{\text{diag}\left\{\text{diag}\left\{B_{jnmk}\right\}_{k=1,...w}\right\}_{m=1,...N_{rx}}\right\}_{n=1,...N_{tx}},$$
$$C_j = \text{diag}\left\{\text{diag}\left\{\text{diag}\left\{C_{jnmk}\right\}_{k=1,..w}\right\}_{m=1,...N_{rx}}\right\}_{n=1,...N_{tx}}.$$

Last, the full system model in (1) that describes the channels from N_{BS} base stations is given by setting

$$\begin{aligned} x^q(t+1) &= Ax^q(t) + Be^q(t), \\ h^q(t) &= Cx^q(t), \end{aligned} \tag{41}$$

with

$$x^q(t) = \begin{bmatrix} x_1^q(t) \\ \vdots \\ x_{N_{BS}}^q(t) \end{bmatrix}, \quad e^q(t) = \begin{bmatrix} e_1^q(t) \\ \vdots \\ e_{N_{BS}}^q(t) \end{bmatrix}, \quad h^q(t) = \begin{bmatrix} h_1^q(t) \\ \vdots \\ h_{N_{BS}}^q(t) \end{bmatrix}, \tag{42}$$

and

$$\begin{aligned} A &= \text{diag}\left\{A_j\right\}_{j=1,...,N_{BS}}, \\ B &= \text{diag}\left\{B_j\right\}_{j=1,...,N_{BS}}, \\ C &= \text{diag}\left\{C_j\right\}_{j=1,...,N_{BS}}. \end{aligned} \tag{43}$$

Note that as the channels from different base stations will experience different short-term fading and shadow fading statistics, the matrices A_j, B_j and C_j must be estimated separately for each set of MIMO channels.

3.4 Estimation of covariance matrices

The co-dependence between the channel components in the MIMO channels of $h_j^q(t)$ are represented by the covariance matrix for the process noise in (1)-(2). In order to calculate this, we first define the covariance matrix of the vector $h_j^q(t)$ of channel components from one base station, in (39),(40), as

$$R_{h,j} = E\left[h_j^q(t)h_j^{q*}(t)\right], \tag{44}$$

Further, let

$$Q_j = E\left[e_j^q(t)e_j^{q*}(t)\right], \tag{45}$$

and

$$\Pi_j = E\left[x_j^q(t)x_j^{q*}(t)\right]. \tag{46}$$

Note that due to the Assumption 3, $R_{h,j}$, Π_j and Q_j are the same for all q, i.e. for all Kalman estimators that operates on different subsets of frequencies. For a wide sense stationary process described by (39), Π_j obeys the Lyapunov equation

$$\Pi_j = A_j\Pi_j A_j^* + B_jQ_jB_j^*. \tag{47}$$

Theorem 4.2.1 in [11] states that given the system in (39), the Lyapunov equation in (47) is satisfied if Q_j is

$$Q_j = R_{h,j} \oslash C_j \left(B_j \mathbf{1} B_j^* \oslash \left(\mathbf{1} - a_j a_j^* \right) \right) C_j^*, \tag{48}$$

where a_j is a vector containing the diagonal elements of A_j in (39) and $\mathbf{1}$ is an all one matrix of appropriate dimensions. Further, as the matrix A_j is diagonal by (26) and (40), the second term on the right hand side of the Lyapunov equation (47) can be rewritten as

$$A_j \Pi A_j^* = a_j a_j^* \odot \Pi_j, \tag{49}$$

where a_j is a column vector containing the diagonal elements of A_j. Then Π_j is by (47) and (49) given by

$$\Pi_j = E\left[x_j^q(t) x_j^{q*}(t) \right] = B_j Q_j B_j^* \oslash \left(\mathbf{1} - a_j a_j^* \right), \tag{50}$$

where $\mathbf{1}$ is an all one matrix of the appropriate dimensions.

While the matrix Q_j and Π_j found by (48) and (50), fulfill the Lyapunov equation (47), they are not guaranteed to provide positive definite matrices. As a covariance matrix by its definition must be positive definite, this introduce a potential error to the state space model. As a result of this error, the convergence of the Kalman filter is not guaranteed, in fact it will tend to diverge when the matrix Q_j is non-positive definite. It is therefore important to find a covariance matrix Q_j that fulfills the requirement of being positive definite.

For a wide sense stationary system, when the covariance matrix $R_{h,j}$ and the state space matrices A_j, B_j and C_j are estimated, then Q_j by (48) will provide a positive definite matrix in some special cases. The first of these is when all the entries of the channel vector $h_j(t)$ are assumed to be identically distributed and hence modeled by identical AR models (23), as in e.g. the case when they represent the different subcarriers between one set of antennas, $N_{tx} = N_{rx} = 1$. The second case is when the entries of the channel vector $h_j(t)$ is are assumed uncorrelated, e.g. when the transmit and receive antennas are sufficiently spaced to ensure that the MIMO channels are uncorrelated in which case the covariance matrix $R_{h,j}$ is block diagonal. The process noise covariance matrix can also be estimated through (48) in a combination of the two above, i.e. when the entries of the channel vector $h_j(t)$ are uncorrelated or identically distributed. A proof to this is provided in Appendix C.

In a general setting when the entries of the channel vector are neither assumed block diagonal nor identically distributed, (48) cannot be used to calculate the covariance matrix Q_j. This applies e.g. when the MIMO channels are both correlated and have different dynamics. An alternative method is then to approximate Π_j and Q_j by using the pseudo inverse. The covariance matrices Π_j and Q_j are then estimated based on (44) and (47) by

$$\begin{aligned} \Pi_j &\approx C_j^\dagger R_{hj} (C_j^*)^\dagger, \\ Q_j &\approx B_j^\dagger \left(\Pi_j + A_j \Pi_j A_j^* \right) (B^*)^\dagger. \end{aligned} \tag{51}$$

A second alternative is to define an upper triangular complex-valued matrix M and then set the covariance matrix of the process noise to

$$Q_j = M M^*. \tag{52}$$

The non-zero elements of the upper triangular matrix M can then be used as optimization parameters to minimize a given cost function. Such a cost function may very well be non convex.

An investigation of how different start values impact the resulting estimation error for a specific simulation scenario where six subcarriers (spanning a total bandwidth of 90 kHz) between a single antenna user and six transmit beams were estimated can be found in [35]. A brief overview of the these results are provided in Appendix D. To summaries, we found that when using (52) with a number of different initial values for M, most resulted in worse estimation performance than when (51) was used to estimate Q_j and only one these was equally good as using (51) . We also found that for the given simulation scenario, approximating Q_j by a block diagonal matrix, where any subchannels between the user and two different transmit beam was assumed uncorrelated, also provided good estimation performance, of roughly the same order as it was when using (52) with the best choice of initial value. In the specific scenario, the correlation between the subcarriers are much higher than the correlation between different beams, which explains that the block diagonal structure of Q_j is a sensible choice.

As the optimization through (52) is of very high in computational complexity, making it unfeasible in practice, it is better to either assume a block diagonal matrix or to use the pseudo inverse to approximate Q_j. It should be checked which of the two that are preferable for the specific scenario.

If we choose to estimate the covariance matrix of the process noise covariance matrix Q_j through (48) or (51), it is necessary to first estimate the covariance matrix of the channel vector $h_j(t)$. The same applies if we attempt to calculate Q_j through (52) with a initial value of M that depends on $R_{h,j}$.

The covariance matrix $R_{h,j}$ must be estimated for each base station. This can be done by using the training sequences to create vectors

$$\tilde{h}_j^k(t) = \begin{bmatrix} \begin{pmatrix} \left[\begin{array}{ccc} \tilde{h}_{j11k}(t) & \cdots & \tilde{h}_{j11(k+w-1)}(t) \end{array}\right]^T \\ \vdots \\ \left[\begin{array}{ccc} \tilde{h}_{j1N_{rx}k}(t) & \cdots & \tilde{h}_{j1N_{rx}(k+w-1)}(t) \end{array}\right]^T \end{pmatrix} \\ \vdots \\ \begin{pmatrix} \left[\begin{array}{ccc} \tilde{h}_{jN_{tx}1k}(t) & \cdots & \tilde{h}_{jN_{tx}1(k+w-1)}(t) \end{array}\right]^T \\ \vdots \\ \left[\begin{array}{ccc} \tilde{h}_{jN_{tx}N_{rx}k}(t) & \cdots & \tilde{h}_{jN_{tx}N_{rx}(k+w-1)}(t) \end{array}\right]^T \end{pmatrix} \end{bmatrix}, \tag{53}$$

for $k = 1, ..., N_k + 1 - w$. Then an estimate of the covariance matrix $R_{h,j}$; defined in (44) is calculated by averaging over all pilot-bearing subcarriers and symbols within the time interval $[t_1, t_2]$ through

$$R_{h,j} \approx \frac{1}{(t_2 - t_1)} \frac{1}{(N_k + 1 - w)} \sum_{t=t_1}^{t_2} \sum_{k=1}^{N_k+1-w} \tilde{h}_j^k(t) \tilde{h}_j^{k*}(t). \tag{54}$$

According to Assumption 4 we have $E\left[h_j^q(t) h_i^{q*}(t)\right] = 0$ for $i \neq j$. It then follows that the covariance matrices of the channel vector $h^q(t)$, the process noise vector $e^q(t)$ and the state vector $x^q(t)$ in (41) are given by

$$\begin{array}{rl} R_h & = [h^q(t) h^{q*}(t)] = \operatorname{diag}\left\{R_{h,j}\right\}_{j=1,...,N_{BS}}, \\ Q & = [e^q(t) e^{q*}(t)] = \operatorname{diag}\left\{Q_j\right\}_{j=1,...,N_{BS}}. \\ \Pi & = [x^q(t) x^{q*}(t)] = \operatorname{diag}\left\{\Pi_j\right\}_{j=1,...,N_{BS}}, \end{array} \tag{55}$$

where Q_j are obtained by one of the methods discussed above.

4 Design choices for AR modeling and Kalman prediction

This section includes discussions on some design aspects for AR model estimation and Kalman filtering of OFDM MIMO channels for CoMP applications.

A fading channel is generally not wide sense stationary. The Doppler spectrum which is modeled through (1) depends on the mobility of the users and the shadow fading environment the user is in. As a user moves this will change. However over small time periods, the fading channel can be approximated as wide sense stationary. How long these time periods are depends on the user mobility and the environment. For slow moving users, this is a longer period than for fast moving users. It is also a longer period if the user is moving along a straight road with Line Of Sight (LOS) components than if the user is zig-zagging through an urban environment turning corners of buildings very often. This time interval is one of the system design choices that needs to be decided.

In Table 4.1 of [11], the complexities of the calculations involved in the Kalman filter are stated, under the assumption that the state space model is on diagonal form. The complexity of one single Kalman filter grows as more subcarriers are included in one filter through (37) as w^3 and with the order of the AR model in (23) as $(n_{AR})^2$. However, increasing these parameters might potentially provide more accurate predictions as correlation over subcarriers can be used for noise reduction and a larger model order might provide a more detailed model of the channels. Both the model order of the AR models and the number of jointly predicted subcarriers are hence design parameters that need to be set.

Other design aspects addressed in this section are how to initiate the Kalman filters, how to design the pilot matrix $\Phi(t)$ in (9) and where to locate the Kalman filters for estimating downlink channels: At the terminals or at the network side.

4.1 The time interval for re-adjusting AR models

Let T be a time interval, in number of pilot-bearing OFDM symbols, during which the channel fading can be considered approximately wide sense stationary. Then the AR model can be estimated from the channel estimates of the first αT part of the interval. This AR model can then be applied in Kalman filters for the second $(1-\alpha)\,T$ part of the interval.

The channel estimates used for estimating the AR model would, in turn, be the output from a Kalman filter with an AR model based on the previous estimate. The very first AR model in this sequence might be based on a simple estimate, that is based on the measurement signal (9).

For example, if $\alpha = \frac{1}{2}$ then we might get $\tilde{h}(t)$ by smoothing a rough estimate based on $y(t)$ in time and frequency domain for $t = 1, ..., \frac{T}{2}$. From these we then estimate the AR model according to Sections 3.1-3.3. This AR model is then used in Kalman filters to provide channel prediction $\hat{h}(t+\tau|t)$ for $t = \frac{T}{2}+1, ..., T$ through (19). The Kalman filter also provides channel estimates $\tilde{h}(t)$ e.g. from (15) for $t = \frac{T}{2}+1, ..., T$ which can be used to estimate a new AR model. This new AR model is then used in the Kalman filters for time $t = T+1, ..., \frac{3T}{2}$, and so on.

In practice, it might be good to let the Kalman filters with different AR models overlap such that the next Kalman filter has already reached it's stationary or cyclo-stationary state (see Section 2.3.1) when the previous filter stops operating.

There is a trade off when it comes to the length of T. On one hand, a shorter T will provide a "more" wide sense stationary channel. On the other hand, a longer T lowers the computational overhead that will come from estimating AR models and running time overlapping filters. The time T should therefore be

as large as possible without causing too large prediction performance loss. How long that is differs from user to user and must be estimated for every specific case. Note also that a larger interval αT allows for a more accurate estimate of the autocorrelation in (29), as the number of terms in the sum increases.

4.2 The choice of AR model order

The poles of the AR model are used to fit the power spectrum of the time variability of the estimated channel to the Doppler spectrum of the measured channel, as illustrated in Figure 4. While the "true" Doppler spectrum cannot be measured exactly, we here display a data-based estimate based on a limited time interval, and denote it the measured Doppler spectrum. In this figure we see that as the model order increases, more of the peaks of the signal's measured Doppler spectrum are included in the AR model. When $n_{AR} = 2$, only one peak of the spectrum is included, and it is slightly off compared to the maximum peak of the signal's measured Doppler spectrum. The 6'th order model includes 4 peaks which are more in accordance with the measured spectrum and the 10'th order model manage to include the 8 strongest peaks. Hence a higher model order more accurately represents the spectral properties of the *training sequence* $\tilde{h}_{jnmk}(t_1), \tilde{h}_{jnmk}(t_1+1), ..., \tilde{h}_{jnmk}(t_2)$.

However, it is not necessarily true that a higher model order will provide more accurate *channel predictions*. First, the channel estimates obtained from the training sequence may include some small errors. Second, the accuracy of the estimation of the autocorrelation function in (30) is limited by the length of the training sequence. Third, recall that the channel is not perfectly wide sense stationary. Because of this, no matter how small we make T, there will always be some differences in the spectrum of the training sequence and that of the channel which is to be predicted. Figure 5 shows how the peaks in the spectrum of the measured channel shift from one time interval to another.

Due to this, there will be a limitation in the prediction performance which will be lower than the theoretical limit described in Section 2.2, where the statistics was assumed to be exactly known and constant. Moreover, there will be a finite model order for which this limit is reached. The limit of when the use of a higher model order starts to decrease prediction performance must be found for each specific system.

4.3 The number of jointly predicted subcarriers

In a frequency selective OFDM system, predicting w subcarriers jointly increases the accuracy of the channel estimate $\hat{x}(t|t)$ and, in extension, the prediction performance. This is illustrated in Figure 6. Here, the theoretic performance is investigated for channel estimation $\hat{h}(t|t)$ given a Doppler spectrum as in Figure 2a, with a pilot SNR of 10 dB and the parameters settings of Table 1. The frequency selectivity provided by the subcarriers are of the WINNER II C2 NLOS channel [30], see Table 2. The performance gain from increasing the filter width w is significant, especially for the middle subcarriers. At the filter edges the NMSE increases. This is because the filter edge subcarriers only gain noise reduction from using the subcarriers on one side of them, whereas the middle subcarriers gain noise reduction from subcarriers on both sides.

When increasing w for a given total channel bandwidth containing N_k pilot-bearing subcarriers, the complexity of each Kalman filter that estimates w subcarriers in parallel grows proportional to w^3 for diagonal state-space models [11] while the number of filters (N_k/w) is inversely proportional to w. The total complexity therefore grows as w^2. A complexity case study in Section 8.6 of [11] indicates a quite

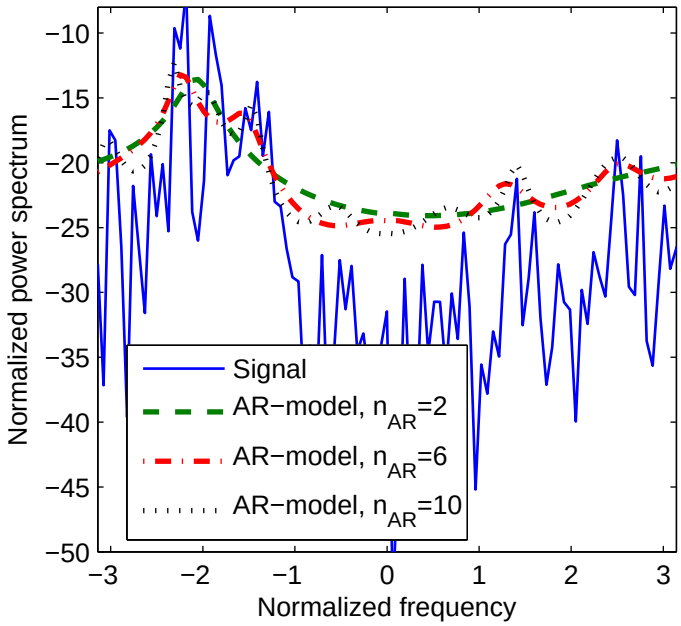


Figure 4: The measured Doppler spectrum of an arbitrary training sequence (from channel measurements) used for AR model estimation (solid line) and the corresponding spectra of the estimated AR models with different model orders.

Table 2: Power delay profile used in the example of Figure 6.

Delay [ns]	0	60	75	145	150	155	150	190
Power [dB]	-6.4	-3.4	-2.0	-3.0	-5.2	-7.0	-1.9	-3.4
Delay [ns]	220	225	230	335	370	430	510	685
Power [dB]	-3.4	-5.6	-7.4	-4.6	-7.8	-7.8	-9.3	-12.0
Delay [ns]	725	735	800	960	1020	1100	1210	1845
Power [dB]	-8.5	-13.2	-11.2	-20.8	-14.5	-11.7	-17.2	-16.7

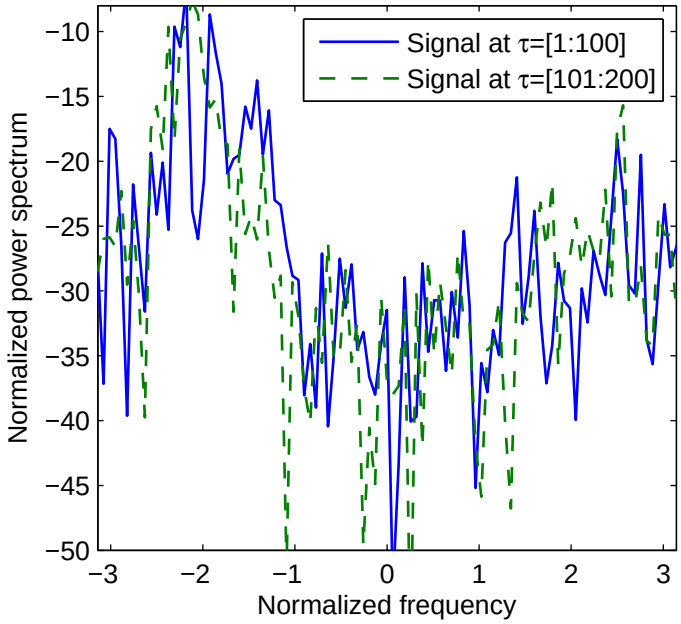


Figure 5: The measured Doppler spectrum of the an arbitrary training sequence in Figure 4 and the measured Doppler spectrum of an equally sized nearby time interval. (One time step here corresponds to 5.3 ms)

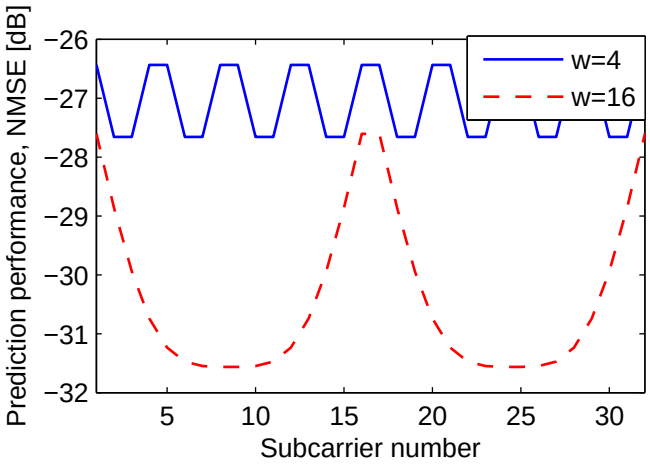


Figure 6: Theoretic Normalized Mean Squared Error (NMSE) as a function of subcarrier number, calculated through (20) with $\tau = 0$, of the filter estimate when $w = 4$ and $w = 16$ subcarriers are predicted jointly. The AR model of Figure 2a is used.

reasonable level of complexity when w is small: With $N_k = 432$, $w = 4$ (108 Kalman filters), $n_{AR} = 4$, and $N_{rx} = 1$, $1.6 \cdot 10^8$ complex arithmetic operations per second would be needed for $N_{BS} \cdot N_{tx} = 3$, and $7 \cdot 10^8$ operations are required for $N_{BS} \cdot N_{tx} = 12$.

The number of jointly predicted subcarriers w might therefore need to be kept low in order to achieve a feasible solution. Then the prediction errors may be reduced in a second step, by Wiener smoothing over estimates from multiple Kalman filters, working on adjacent frequency groups. The true prediction error covariances then differ from those given by (20) due to two effects. First, the AR models are imperfect which increases the errors. Second, Wiener-smoothing over frequency decreases the errors. In our studies, these two effects leave the variance of the prediction error *slightly* lower than that given by (20).

4.4 Initiating the filter

If the AR models are updated every T time steps based on channel estimates obtained by consecutive Kalman filters as described in Section 4.1, then this simplifies the initialization of a new Kalman filter. The initial state vector in (12) and the corresponding covariance matrix (15) can be provided by the current filter as the latest filter estimate with its corresponding covariance. This would decrease the convergence time of the filter as described in Section 2.3.1.

However, if this is not possible, e.g. if the user equipment has been out of reach from a base station, the filter can be initiated by setting the state vector to its mean

$$\hat{x}(0|0) = E\left[x(t)\right] = 0. \tag{56}$$

Then the corresponding covariance matrix is given by

$$\begin{aligned} P(0|0) &= E\left[(\hat{x}(0|0) - x(0))\,(\hat{x}(0|0) - x(0))^*\right], \\ &= E\left[x(0)x^*(0)\right], \\ &= E\left[x(t)x^*(t)\right], \\ &= \Pi. \end{aligned} \tag{57}$$

4.5 Pilot patterns

Recall that, by (9) and (41), the channel model and the pilot measurements $y^q(t)$ at filter q are provided by the equations

$$\begin{aligned} x^q(t+1) &= Ax^q(t) + Be^q(t), \\ h^q(t) &= Cx^q(t), \\ y^q(t) &= \Phi(t)h^q(t) + n^q(t) = J(t)x^q(t) + n^q(t). \end{aligned} \tag{58}$$

These are then used in the Kalman filters through (12)-(20). An important design choice to ensure proper predictions is to choose the structure of the pilot matrix $\Phi(t)$.

Each transmit antenna n at each base station j are here assumed to be allocated pilot resources on the jointly estimated subcarriers according to a diagonal pilot matrix $\Phi_{jn}(t) \in \mathbb{C}^{w \times w}$. Diagonal elements set to zero, indicate that the transmitter shall not transmit any pilots on the corresponding subcarrier at time slot t. The pilot $\Phi(t)$ in (58), for all antennas at all base stations, is then formulated as

$$\Phi(t) = \left[\begin{array}{cccccc} \Phi_{11}(t) & \cdots & \Phi_{1N_{tx}}(t) & \cdots & \Phi_{N_{BS}N_{tx}}(t) \end{array}\right].$$

The allocation of the pilots have been discussed for the uplink channels in [24] and for downlink MIMO channels in Chapter 6.4 of [11]. The pilots transmitted from different transmit antennas may be code-orthogonal, or resources-orthogonal.

If the pilots are resource-orthogonal then only one transmit antenna will transmit on any give resource slot. As an example, assume a CoMP Multiple-Input Single output (MISO) system with Binary Phase Shift Keying (BPSK) with symbols $\{-1, 1\}$ and $N_{BS} = 2$ base stations each with $N_{tx} = 2$ transmit antennas. A single antenna user ($N_{rx} = 1$) predicts the channels over $w = 4$ subcarriers jointly. Then the following resource-orthogonal pilot pattern could be used for Φ_{jn}, for base station $j = \{1, 2\}$ and antenna $n = \{1, 2\}$:

$$\begin{aligned} \Phi_{11}(t) &= \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \quad \Phi_{12}(t) = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \\ \Phi_{21}(t) &= \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 0 \end{bmatrix}, \quad \Phi_{22}(t) = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 0 & \\ & & & 1 \end{bmatrix}. \end{aligned} \tag{59}$$

One benefit from this is that the energy of pilots from one transmitter will not leak over into the pilots of another base station, for the base stations within the cluster. A second benefit is that the problem of estimating channels from different base stations are then not correlated, so they can be implemented as separate parallel Kalman filters, lowering the computational complexity. The drawback is that if the total number of transmit antennas are large, then the pilots of each transmitter will have to be placed very sparse in time and/or frequency to limit pilot overhead. Therefore, the received signals at consecutive received pilots might have low correlation. This will decrease the noise reduction abilities of the Kalman filter, that work by utilizing such correlations.

If the pilots are code-orthogonal then some of the transmitters may transmit on the same resource. They will then transmit with a symbol pattern such that they are orthogonal over a number of dependent subcarriers. For the example above, the code-orthogonal pilot matrices might then be

$$\begin{aligned} \Phi_{11}(t) &= \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad \Phi_{12}(t) = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & 1 & \\ & & & -1 \end{bmatrix}, \\ \Phi_{21}(t) &= \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & -1 & \\ & & & -1 \end{bmatrix}, \quad \Phi_{22}(t) = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & 1 \end{bmatrix}. \end{aligned} \tag{60}$$

If each sub-channel $h_{jnmk}(t)$ estimated by the Kalman filter, only differs from a nearby sub-channel $h_{jnm(k+\kappa)}(t)$ by a known phase rotation φ, then the channel measurements based on code-orthogonal pilots can be perfectly separated if $w \geq N_{BS} \cdot N_{tx}$. For the pilot matrices in (60), the measurement signal

in (9) is then

$$y(t) = \Phi(t)h(t) + n(t) = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ e^{i\cdot\varphi} & -e^{i\cdot\varphi} & e^{i\cdot\varphi} & -e^{i\cdot\varphi} \\ e^{i\cdot 2\varphi} & e^{i\cdot 2\varphi} & -e^{i\cdot 2\varphi} & -e^{i\cdot 2\varphi} \\ e^{i\cdot 3\varphi} & -e^{i\cdot 3\varphi} & -e^{i\cdot 3\varphi} & e^{i\cdot 3\varphi} \end{bmatrix}}_{\Theta(t)} \underbrace{\begin{bmatrix} h_{1111}(t) \\ h_{1211}(t) \\ h_{2111}(t) \\ h_{2211}(t) \end{bmatrix}}_{\tilde{h}(t)} + n(t). \tag{61}$$

As $\Theta(t)$ has full rank it is invertible. Multiplying both sides of (61) with this inverse gives an estimate of the vector of channels $\tilde{h}(t)$ in (61) at subcarrier $k = 1$:

$$\tilde{y}(t) = \Theta^{-1}(t)y(t) = \tilde{h}(t) + \tilde{n}(t). \tag{62}$$

Here $\tilde{n}(t)$ is a vector of independent Gaussians and each element of the measurement vector $\tilde{y}(t)$ only relates to channels from one transmit antenna. However, if φ has an unknown component $\Delta\varphi$ then there will be leakage of energy between the different channels. This energy comes in as an extra noise term in (62) and results in lowering the effective SNR of the measurement[4]. If some of the channel gains are much larger than the others, then the effective SNR of the channel measurements of the weaker channels will be significantly lowered even if $\Delta\varphi$ is small. However, if the channels gains are equally strong and $\Delta\varphi$ is small (relative to π) then the effective SNR is only little affected.

Note that to make a fair comparison between code-orthogonal pilots and resource-orthogonal pilots the matrices in (60) should be re-scaled to ensure that the total transmit power is equal.

A third alternative is to let transmitters located at the same base station transmit code-orthogonal pilots and transmitters at different base stations transmit resource-orthogonal pilots. For the above example this leads to the matrices

$$\begin{aligned} \Phi_{11}(t) &= \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \quad \Phi_{12}(t) = \begin{bmatrix} 1 & & & \\ & -1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}, \\ \Phi_{21}(t) &= \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad \Phi_{22}(t) = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & -1 \end{bmatrix}. \end{aligned} \tag{63}$$

This is a sensible alternative as channels from the same base stations normally have equally strong channel gains on average, while channels from different base station vary a lot.

With this scheme, the pilots transmitted from each antenna can be more densely placed than with a fully resource-orthogonal pilot pattern. For example, assume a CoMP cooperation system with 9 base stations, each with 4 transmit antennas, and a pilot spacing of 14×5 OFDM symbols in time, corresponding to 5 subframes in an LTE downlink, or 5 ms. Further, assume 3GPP release 10 CSI specific reference signal positions [31] are used, see Figure 7, and that these are repeated every 5 subframe. Then, with fully resource-orthogonal pilots only one pilot per transmit antenna can be used for every resource block

[4] A similar problem of loss of orthogonality in uplinks in CDMA transmission systems is called the "near-far problem".

(12 subcarriers), spaced by 5 subframes (5 ms) in time. With the combination of resource-orthogonal and code-orthogonal pilots, each antenna can transmit 4 pilots per resource block, every fifth subframe.

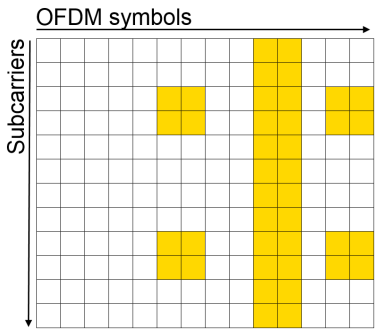


Figure 7: The structure of CSI specific pilots (filled) within a subframe in time (14 OFDM symbols or 1 ms) and one resource block width in frequency (12 subcarriers of 180 kHz) in 3GPP release 10. These pilots may be repeated every fifth subframe [31].

If $w < N_{BS}N_{tx}$, then orthogonality cannot be kept over the subcarrier even if the phase correlation between the subcarriers is perfectly known. Then different pilots can be used in different time slots in a cyclical recurrent pattern in order to preserve orthogonality and obtain invertibility similar to (62) over a whole time-frequency pattern, that comprise $\geq N_{BS}N_{tx}$ time-frequency pilot positions. For details, please see [24] and Section 6.4.2 of [11].

4.6 Location of the Kalman filters

Channel estimators or predictors for FDD downlinks may be physically located at the terminals (users) or on the network side. In the former case, the estimators or predictions are fed back to the network over uplink control channels. In the latter case, measurements need to be fed back, and the predictors operate based on these measurements.The feedback load per predicted resource block per user depends on the detailed system design in these two alternatives. There are different advantages to each of these alternatives.

In [32] the performance of JT CoMP was evaluated when channels where predicted at the user side and network side respectively. In a centralized setup, with a central CU calculating the joint beamformers, the placement did not affect the CoMP gains. However, in a distributed CoMP scenario, [32] found that CoMP gains were increased if predictions were performed on the network side.

Another benefit of locating the Kalman filters at the network side is that this often allows for more sophisticated equipments, so a higher computational complexity might be allowed and so the number of jointly predicted subchannels, w, can be set high. In addition, the prediction quality of the users will vary less as the difference in the different users equipment will affect the predictions less.

On the other hand, if the channel estimators are located at the user side, adaptive quantization schemes based on prediction performance may be used. Then, if a user feeds back e.g. two predicted channels where one of them is very accurate and the other is very inaccurate, the more accurate channel may be quantized with higher granularity than the inaccurate channel. The distribution of the feedback bits are then based on the accuracy of the channels described by (20). That way, quantization errors will affect

the performance less. Note that as the accuracy (the diagonal elements of the covariance matrix (20)) are fed back on a slow time scale, they do not impose a large feedback overhead.

Moreover, scheduling and link adaptation decisions are today based on quite coarse Channel Quality Index (CQI) that is fed back to the network side. As the number of users considered for scheduling is often much larger than the number of users that is actually scheduled, the total feedback will be reduced if the full CSIT is fed back only for the users that are actually scheduled. If the channel estimators are located at the user side, and if a small extra reporting delay is acceptable between scheduling and transmission, then the users may report complex channel gains only for the users/data streams that have actually been scheduled. A similar two-step feedback concept has been proposed in [33].

If instead the predictor were placed on the network side, measurements would need to be reported from all uses that may potentially be scheduled and for all resource blocks that may potentially be used. This is because Kalman (and Wiener) predictors need a time-history of past measurements in order to reach the steady state and perform reliable predictions, as described in Section 2.3.1.

Finally, if the users have multiple receive antennas then they may apply a receiver filter in order to increase their data rate. If Kalman filters are located at the users, then these receiver filters could be based of the latest updated channel filter estimates provided by (15). If the Kalman filters are not located at the users, then an additional channel estimation method must be applied for tuning the receiver at the terminal.

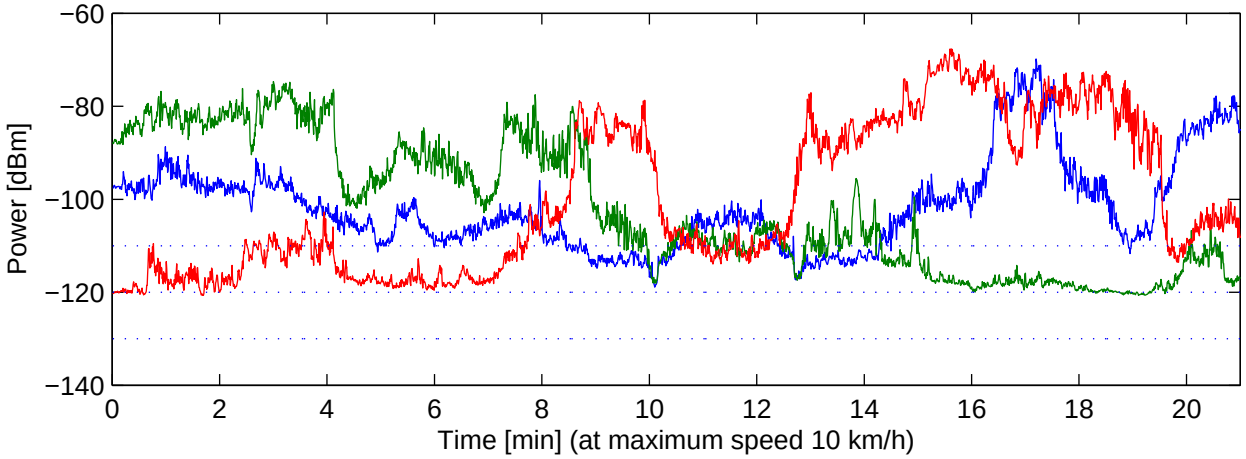


Figure 8: The variation of the power of the received signals that were transmitted from base station 1 (blue), base station 2 (green) and base station 3 (red). The time-axis is based on the upsampled system with pilot spacing $\tau = 1.3$ ms assuming pedestrian velocities. Three noise floors of -130 to -110 dBm indicated with dotted lines are used in the simulations, here and in [12, 14]. In [13], a set of noise levels {-120 dBm, -110 dBm, -100 dBm}, i.e. 10 dB higher, is used.

5 Measurement based predictions

5.1 Channel measurements

All simulations in this section are based on channel sounding measurements carried out by Ericsson Research. Three omnidirectional single antenna base stations, located at different sites with 350-600 m distance, were used to transmit channel sounding pilots at a carrier frequency of 2.66 GHz. These were measured by a vehicle driving at a velocity of $10 - 30$ km/h in an outdoor urban environment in central Kista, Stockholm. Resource-orthogonal pilots were transmitted every 5.3 ms and on every subcarrier à 15 kHz, with pilots from each base station on every third subcarrier. The received signal powers from the base stations are plotted in Figure 8. The measurements are of high quality and can hence be assumed to represent the true complex channel gains in space. For more details on the measurement setup, please see [34].

The channels are upsampled 25 times over time giving a pilot spacing of 0.21 ms at the velocity of the measurement. This corresponds to a pilot spacing of 1.3 ms at pedestrian velocities of up to 5 km/h. The upsampling uses the fast Fourier transform to ensure that no extra frequency components are added.

The upsampled channels are used to create the measurement signals (9).

5.2 Simulation assumptions

The downlink channels from the $N_{BS} = 3$ single antenna base stations are predicted for the entire measurement route in Figure 8.

In Chapter 8 of [11], which used measurements similar to those in this report, it was concluded that when the autocorrelation method described in Section 3.2 was used, a fourth order AR model sufficed.

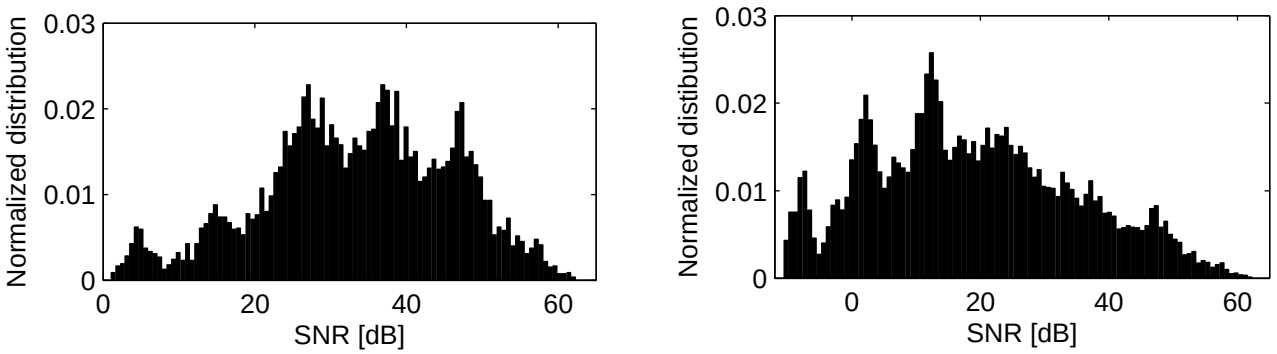


Figure 9: SNR distributions for the strongest base station (left) and all base stations (right) of the measurement signal in (9) over three measurement where the three noise floors in Figure 8 were used.

Therefore, the fading statistics in time and frequency, will here be represented by fourth order AR-models. These are estimated periodically every 800 pilots in time ($\sim$ 1 s) based on noise-free channel data, i.e. on perfect CSIT, from the past 800 pilots in time. The timing interval discussed in Section 4.1 is hence $T = 1600$ ($\sim$ 2 s) with $\alpha = 0.5$. From studying the measured data, we have found that this time interval is appropriate with respect to the long term fading. It is short enough to ensure that the statistics of the Doppler spectrum stays fairly constant within the interval, but also long enough to provide appropriate prediction performance statistics and CoMP performance statistics for each interval. For high mobility users the interval might need to be shorter.

Chapter 8 of [11], also investigated how the prediction performance was affected by estimating the AR model using subsampled channels as described in Section 3.2.1. The investigations showed that performance increased when using the subsampled channels. Therefore, the poles of the AR models are here estimated as in Section 3.2.1, using a subsampling factor corresponding to the different prediction horizons used (for the filter estimate, no subsampling is used).

Signal measurements with an appropriate range of SNRs are created by using (9) with a transmit power of 1 (i.e. the non zero pilot symbols of $\Phi(t)$ will have unit power) and additive white Gaussian noise of three different power levels, σ_n^2, see Figure 8. The covariance matrix of (10) is then given by $R(t) = \sigma_n^2 I$. The noise is i.i.d. over subcarriers. On average over all three noise levels and for all three base stations, the median SNR is 24 dB at the investigated positions. The SNR distribution, shown in Figure 9, is similar to that obtained when applying the intercluster interference mitigation framework of [14, 21, 22]. That proposal forms overlapping static clusters that use different time-frequency allocations and further controls interference by using different antenna downtilts and transmit powers to the outside and to the inside of each cluster.

Computational complexity increases with w, so we use a low value of $w = 4$ when code-orthogonal pilot patterns are used. To make a fair comparison (to keep the computational complexity equal) the number of jointly estimated subcarriers per Kalman filter is set to $w = 12$ when resource-orthogonal pilot patterns are used and channels from different base stations are predicted by parallel filters (this can be done only when using resource-orthogonal pilots). For both pilot patterns all pilots have the same transmit power, as in (59) and (60). That means that SNR per pilot is equal for both patterns, but that the pilot transmit power per base station, summed over all subcarriers, is three times higher for the code-orthogonal pilots.

Table 3: Comparison of average NMSE for Kalman predictions and outdated CSIT for prediction horizons of $\tau = 4$, 8, 12 and 18. Results are averaged (in dB) over all base stations and subcarriers. Noise floors -110 dBm, -120 dBm and -130 dBm and resource-orthogonal pilots are used.

τ	σ_n^2 [dBm]	Predictions	Outdated CSI
4	-110	-12.8 dB	-10.5 dB
	-120	-15.3 dB	-12.5 dB
	-130	-17.6 dB	-14.0 dB
8	-110	-11.0 dB	-6.9 dB
	-120	-12.9 dB	-7.9 dB
	-130	-14.8 dB	-8.6 dB
12	-110	-9.6 dB	-4.4 dB
	-120	-11.2 dB	-5.0 dB
	-130	-12.8 dB	-5.3 dB
18	-110	-7.9 dB	-1.8 dB
	-120	-9.2 dB	-2.1 dB
	-130	-10.3 dB	-2.2 dB

The channels are predicted for 432 subcarriers using prediction horizons of $\tau = 0$, 4, 8, 12 and 18 pilots in time. These correspond to distances $d_\lambda \approx 0$, 0.06, 0.13, 0.19 and 0.28 carrier wavelengths at 2.66 GHz. The results for prediction distances d_λ are scalable and could be interpreted as predictions for time horizons through (8). For example the prediction horizons correspond to time horizons of 0, 5, 10, 15 and 23 ms at the pilot spacing of 1.3 ms, pedestrian velocities ≤ 5 km/h and carrier frequency 2.66 GHz.

The prediction performance will be evaluated using the Normalized Mean Squared Error (NMSE) for the channel from the k'th subcarrier at the j'th base station

$$NMSE_{jk} = \frac{\sum_{t=1}^{(1-\alpha)T} |h_{jk}(t) - \hat{h}_{jk}(t|t-\tau)|^2}{\sum_{t=1}^{T} |h_{jk}(t)|^2}, \tag{64}$$

The NMSE are then sorted into three categories - that of the weakest base station, the second strongest base station and the strongest base station. These categories are defined as the NMSE of the channels that have the lowest average channel gain, the second lowest channel gain and the highest channel gain on average over the 0.5 s evaluation interval. Note that which of the three physical base station that falls into which category varies over time.

In [16] an NMSE of > -8 dB was found to be required to achieve most of the multi-user scheduling gains that are obtainable with perfect channel estimates. This NMSE level is here used as a loose limit of "acceptable NMSE".

In the sections below the average NMSE is presented in tables for some particular cases. For the full table of results, please see the Appendix B.

5.3 Comparing outdated CSI and predicted CSI

First, let us study the impact of using Kalman filters to predict the channel $h(t)$ instead of using the outdated filter estimate $\hat{h}(t-\tau|t-\tau)$. Table 3 shows the average NMSE over all time intervals and base

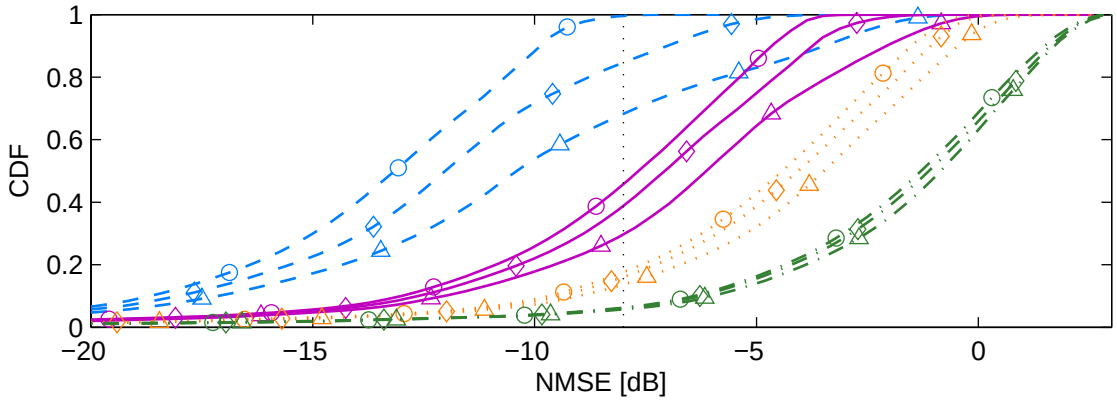


Figure 10: CDF of NMSE when using *outdated CSIT* for prediction horizons $\tau = 4$ (blue dashed lines), $\tau = 8$ (purple solid lines), $\tau = 12$ (orange dotted lines) and $\tau = 18$ (green dashed-dotted lines). Results for noise floors of -130 dBm (circles), -120 dBm (diamonds) and -110 dBm (upwards triangles). Resource-orthogonal pilots are used for the channel estimation. The performance limit of -8 dB is marked as a black dotted vertical line. Note that the CSIT quality increases towards the left.

stations at the different noise floors and prediction horizons. These are compared with the NMSE of the outdated channels, i.e. that which we would get if $\hat{h}(t|t-\tau)$ in (64) is replaced with $\hat{h}(t-\tau|t-\tau)$.

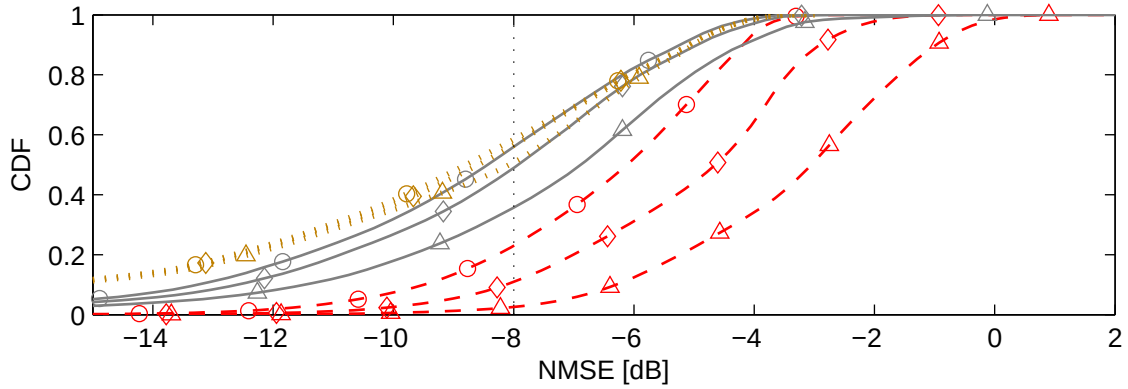
Clearly, as the prediction horizon increase so does the importance of using predictions. For example, at a horizon of $\tau = 12$ the average NMSE of the predictions falls below the limit of -8 dB for all the noise floors but the average NMSE of the outdated channels is above the -8 dB limit.

In Figure 10 the CDF's of the NMSE of the outdated CSIT over all subcarriers and all $(1-\alpha)T$ sampled time intervals are presented. Here, we see that even at a prediction horizon $\tau = 8$, a majority of the channels have an NMSE higher than -8 dB. In fact, it is only at the lowest prediction horizon that most of the CSIT falls withing an acceptable NMSE.

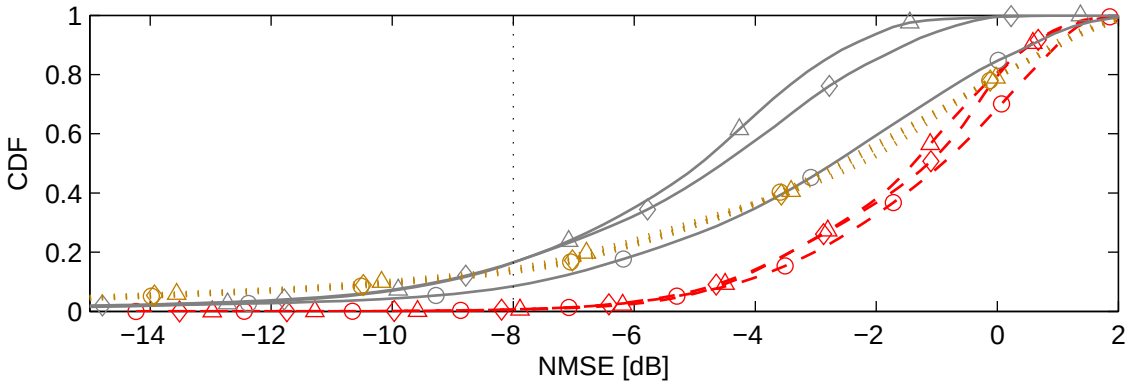
The CDF's for prediction horizons $\tau = 8$ and $\tau = 18$ are shown in more detail in Figure 11. Here, we see that not only is most of the CSIT NMSE above the -8 dB limit, there is very little difference between the NMSE of the CSIT for the strongest and weakest base station. Both the CSIT of the channels of the strongest and second strongest base station experience poor NMSE. Hence, the possibility of having gains from coherent JT CoMP are very small.

5.4 The effect of resource-orthogonal versus code-orthogonal pilot patterns

In this section the effect of the pilot pattern, discussed in Section 4.5, is studied. Table 4 shows the average NMSE for different prediction horizons when resource-orthogonal pilots (ROP) and code-orthogonal (COP). For the resource-orthogonal pattern, each base station transmits pilots on every third subcarrier and for the code-orthogonal pattern the first three matrices of (60) are used. For all noise floors here the NMSE of the predictions with resource-orthogonal pilots are below those of predictions with code-orthogonal pilots. This is the case despite the fact that COP were allowed to use three times more pilot power per base station than ROP in this experiment. However, as the number of total transmit antennas are low (three), the subcarriers used by each base station have high correlation even for



(a) $\tau = 8$



(b) $\tau = 18$

Figure 11: CDF of NMSE of the *outdated CSIT* for $\tau = 8$ (upper) and $\tau = 18$ step predictions (lower) sorted into groups of noise floor $\sigma_n^2 = -130$ dBm (circles), $\sigma_n^2 = -120$ dBm (diamonds) and $\sigma_n^2 = -110$ dBm (triangles). The CDF's are shown for the strongest (brown), second strongest (gray) and weakest (red) base station. The performance limit of -8 dB is marked as a black dotted vertical line. Resource-orthogonal pilots are used. Note that the CSIT quality increases towards the left.

Table 4: Results for resource-orthogonal pilots (ROP) and code-orthogonal pilots (COP) at prediction horizons of $\tau = 0$ (filter estimate), 4 and 8. Results, in terms of NMSE, are averaged (in dB) over all subcarriers. Noise floors -110 dBm and -120 dBm.

τ	σ_n^2 [dBm]	NMSE (all) ROP	NMSE (all) COP	NMSE (weakest) ROP	NMSE (weakest) COP
0	-110	-17.8 dB	-16.8 dB	-7.1 dB	-5.6 dB
	-120	-23.9 dB	-19.6 dB	-12.7 dB	-6.9 dB
4	-110	-12.8 dB	-12.4 dB	-5.9 dB	-5.2 dB
	-120	-15.3 dB	-13.9 dB	-9.4 dB	-6.6 dB
8	-110	-11.0 dB	-10.8 dB	-4.8 dB	-4.4 dB
	-120	-12.9 dB	-12.2 dB	-7.4 dB	-5.9 dB

resources-orthogonal pilots and the noise reduction abilities of the Kalman filter is fairly high. Therefore, the benefits of using resources-orthogonal pilots, i.e. that no energy leak between the pilots, are more dominant in this example than the drawbacks.

From studying Table 4 there are three important observations:

- The loss in prediction performance from using code-orthogonal pilots instead of resource-orthogonal pilots increases as the noise floor decreases. This is because the benefit of code-orthogonal pilots is that they allow for more noise reduction in the filter estimate (15). However, as the noise floor is lowered, the need for noise reduction also becomes less important.

- The performance differences between the two pilot patterns decrease as the prediction horizon increases. The pilot pattern only affects the prediction performance through the filter estimate (15). As the prediction horizon increases, the main errors are due to the theoretical predictability of the channel, see Section 2.2, and the accuracy of the AR models, see Section 4.

- The drawback with using code-orthogonal pilots is largest for the predictions of the weakest base station's channels.

The third observation is further validated from studying Figure 12 where the average NMSE is plotted for each base station over the entire measurement route. Comparing this with Figure 8 we see that the largest difference between the two pilot patterns occur when the weakest base station (that here have largest NMSE) is much weaker than the strongest base station. When the channel gains differ only by a few decibel, as in the interval between 10 and 13 minutes, the different pilot patterns affect the prediction performance very little. This is in accordance with the reasoning in Section 4.5. If channels have very different gains then the effective SNR of (62) will decrease most for the signal from the weakest base station.

Since the amplitudes of the channel gains often vary a lot for signals transmitted from different base stations, but not for signals transmitted from different antennas at the same base station, the third pilot pattern described in Section 4.5 is a good choice for CoMP systems that need to accommodate a large number of transmit antennas in each coordination cluster of base stations. That is, transmitters located at the same base station use code-orthogonal pilots and transmitters located at different base stations use resource orthogonal pilots, as in the example (63).

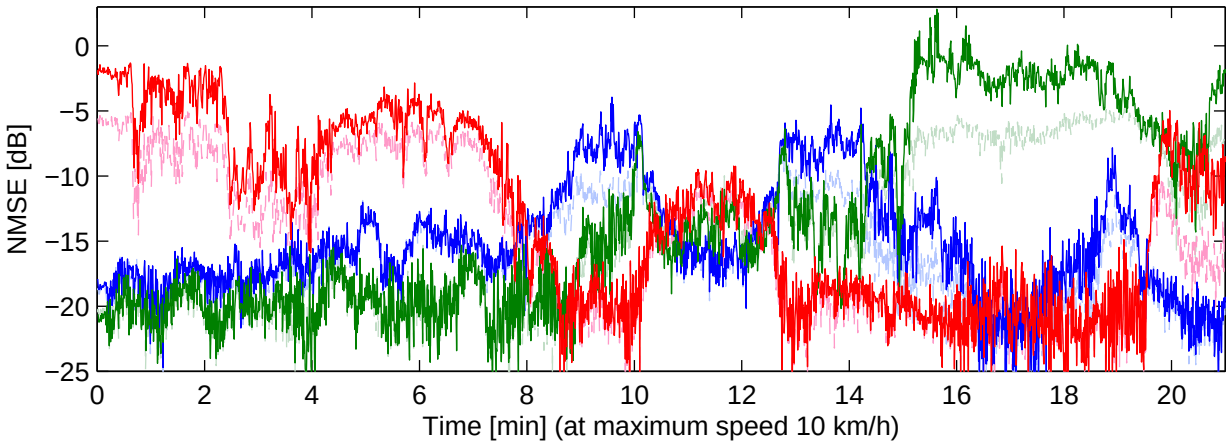


Figure 12: Average NMSE over all subcarriers with a prediction horizon of $\tau = 4$ and a -120 dBm noise floor. Predictions are over the channels from base station 1 (blue), base station 2 (green) and base station 3 (red). Dark colors indicate that code-orthogonal pilots are used and light colors that resource-orthogonal pilots are used.

5.5 The influence of intercluster interference

In this section we study the effect of the the intercluster interference, or the noise term in (9), on prediction performance. Figure 13 shows CDF's of the NMSE for different prediction horizons. The results are sorted into groups of pilot SNR. Note that these groups do not include all data as the SNR might both be above 30 dB and below 0 dB, see Figure 8. Results show that as the prediction horizon increases, the benefit from having a large SNR decreases, compare with the second bullet observation in Section 5.4. However, the gain of having a large SNR is still significant, e.g. the NMSE increases by $5 - 2$ dB when going from SNR $\in [20, 30]$ to SNR $\in [10, 20]$ at $\tau = 18$. Note especially that when comparing the CDF for SNR $\in [20, 30]$ with that for SNR $\in [10, 20]$, at $\tau = 18$, the first CDF crosses the -8 dB limit at 96% while the second CDF crosses this NMSE level at 52%.

In Figure 14 the CDF's are instead shown for the different noise floors. Each CDF now includes a large range of pilot SNR's and the CDF's are therefore wider than those of Figure 13. Lowering the noise floor from -110 dBm to -130 dBm results in that the fraction of the channel predictions reaching the limit of -8 dB increases from 49% to 70% and from 69% to 93% for $\tau = 18$ and $\tau = 8$ respectively.

5.6 Effect of the system delays on the prediction NMSE

The required prediction horizon is given by the system delays caused by feedback, backhaul and computational delays. In Section 5.3 we saw that already at short system delays, of e.g. $\tau = 8$, the outdated CSIT does not provide an acceptable NMSE. However, as shown in Section 2.2, there is a performance limit also for predicted channels. From Figure 14 we see that, depending on the noise floor, some channel predictions will also not reach the acceptable NMSE at $\tau = 8$. In this section, we will study which these channels are.

In Figure 15 the statistics in Figure 14 are sorted into groups of strongest, second strongest and weakest

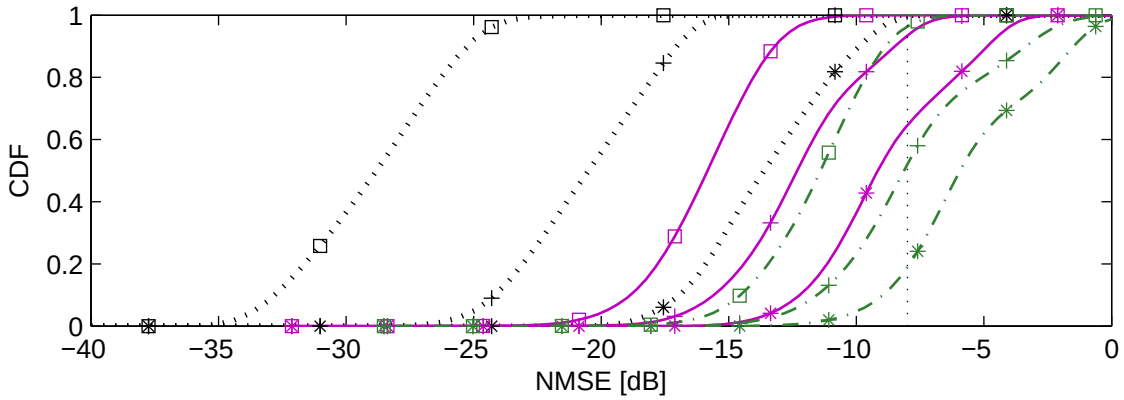


Figure 13: CDF of NMSE sorted into groups of pilot SNR in the intervals $[20, 30]$ dB (squares), $[10, 20]$ dB (pluses) and $[0, 10]$ dB (stars). Prediction horizons are $\tau = 0$ (black dotted lines), $\tau = 8$ (purple solid lines) and $\tau = 18$ (green dashed-dotted lines). The performance limit of -8 dB is marked as a black dotted vertical line. Resource-orthogonal pilots are used. Note that the CSIT quality increases towards the left.

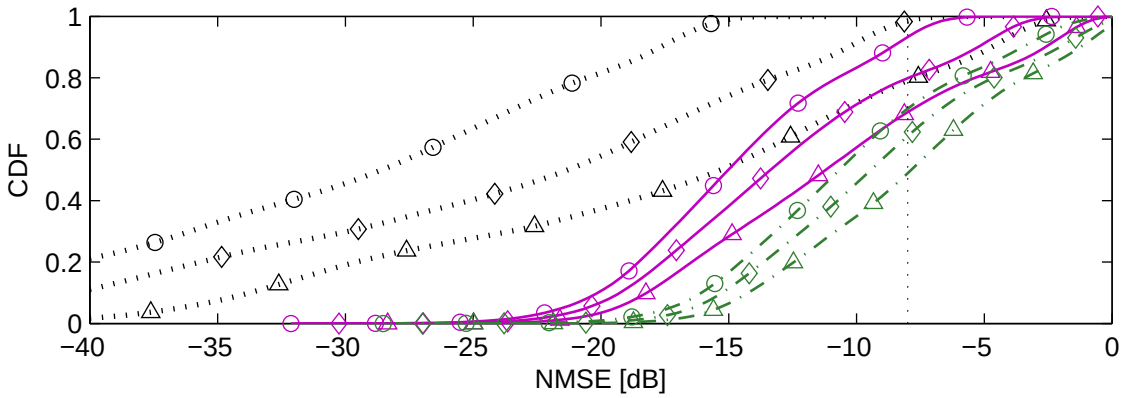
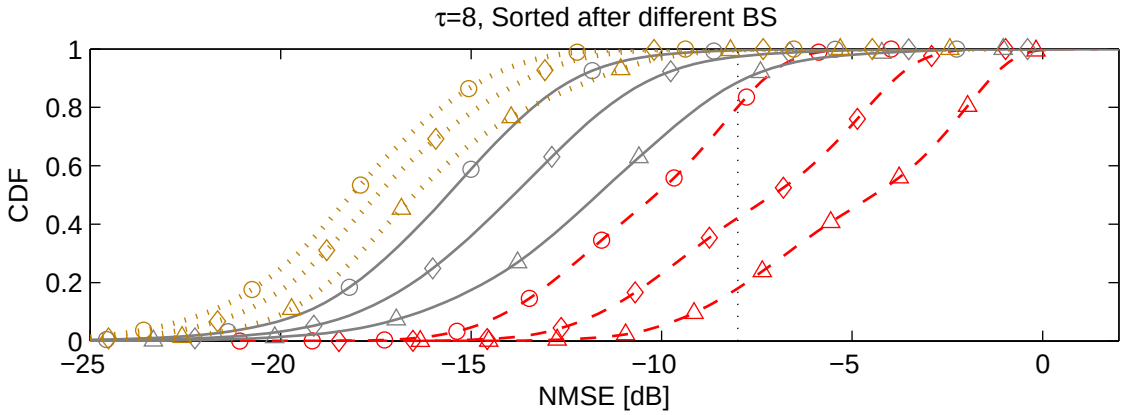


Figure 14: CDF of NMSE sorted into groups of noise floor $\sigma_n^2 = -130$ dBm (circles), $\sigma_n^2 = -120$ dBm (diamonds) and $\sigma_n^2 = -110$ dBm (triangles). Prediction horizons are $\tau = 0$ (black dotted lines), $\tau = 8$ (purple solid lines) and $\tau = 18$ (green dashed-dotted lines). The performance limit of -8 dB is marked as a black dotted vertical line. Resource-orthogonal pilots are used. Note that the CSIT quality increases towards the left.

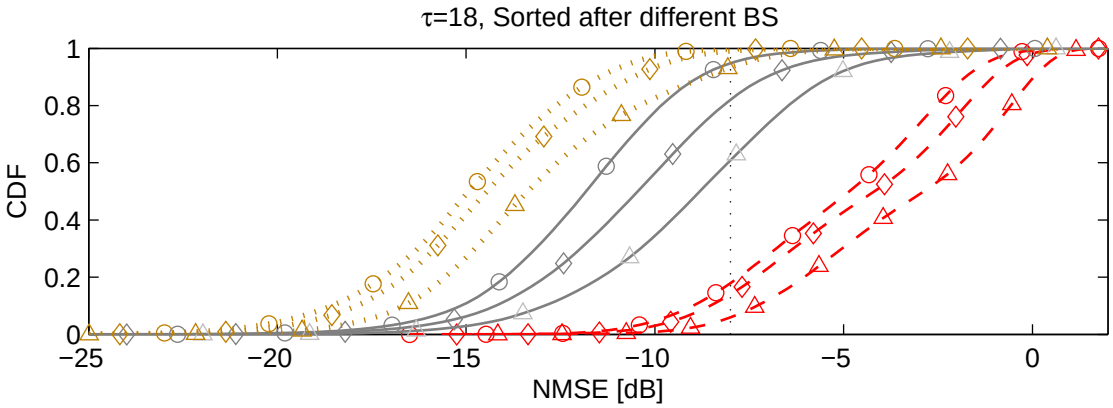
base station for the prediction horizons $\tau = 8$ and $\tau = 18$. Here, we see that for the prediction horizon $\tau = 8$, which can be translated into a system delay of approximately 10 ms for pedestrian users with a pilot spacing 1.3 ms at 2.66 GHz, the channels that do not reach the -8 dB target are mainly those from the weakest base station. In Figure 8 we see that these have received powers that are mostly very much lower than the strongest. In a coherent JT CoMP, they will therefore not cause a lot of interference, provided that the precoding design does not allow for them to become "rotten" apples spoiling the precoding design. However, this can be avoided by taking the uncertainties (20) into account in the precoding design, as was done in [12].

However, as the system delay increases, see the CDF's for $\tau = 18$, most of the channels from the weakest base station and many of the channels for the second strongest base stations, especially for the -110 dBm noise floor, do not reach the -8 dB NMSE limit. This means that some of the useful channels for coherent JT CoMP are very inaccurate, potentially lowering the CoMP gains. Therefore it is very important to keep the system delays short. Note also that at the lowest noise level (-130 dBm) most of the second strongest base stations channels have prediction errors with an NMSE below -8 dBm. Hence, the reduction of inter cluster interference is of equal importance to the reduction of delays in transmission control loops.

In [12] it was shown that the gains from using coherent JT CoMP compared to single cell transmission did decrease when using the predictions at $\tau = 18$ instead of those at $\tau = 8$. However, as the strongest and, for most times, also the second strongest base station still have acceptable NMSE, CoMP gains were still significant (in the range of 40%). This was not the case when using the outdated CSIT in Figure 11.



(a) $\tau = 8$



(b) $\tau = 18$

Figure 15: CDF of NMSE sorted into groups of noise floor $\sigma_n^2 = -130$ dBm (circles), $\sigma_n^2 = -120$ dBm (diamonds) and $\sigma_n^2 = -110$ dBm (triangles). The CDF's are shown for the strongest (brown), second strongest (gray) and weakest (red) base station. The performance limit of -8 dB is marked as a black dotted vertical line. Resource-orthogonal pilots are used. Note that the CSIT quality increases towards the left.

6 Conclusion

From the measurement based prediction performance in Section 5, the following conclusions are drawn

- Even at, for JT CoMP, short system delays of 10 ms and for low mobility users (moving at pedestrian velocities and at a carrier frequency of 2.66 GHz), the outdating of CSIT is large. This can potentially destroy all gains for coherent JT CoMP. The effects are even more severe for longer prediction horizons or higher user mobility.
- Prediction with Kalman filters provide CSIT with acceptable NMSE to achieve significant coherent JT CoMP gains, even at large system delays, e.g. 24 ms for users moving at pedestrian velocities, at a carrier frequency of 2.66 GHz.
- As the weak channels are often hard to predict, it is important that these are not allowed to dictate the precoder for coherent JT CoMP. A robust precoder as in [12] can ensure that these channels do not become "rotten apples" that destroy the potential CoMP gains.
- As prediction performance does not decrease noticeably for code-orthogonal pilots when the channels have approximately equal channel gain, code-orthogonal pilots can be used for transmitters located at the same base stations. This reduces the difficulty of allocating pilot resources with sufficient time-frequency sampling density within cooperation clusters where base stations each have many antennas.
- Prediction performance decrease significantly for the weak channels when code-orthogonal pilots and when the channels have very different channel gain. If the weak channels are necessary for the application, then resource-orthogonal pilots should be used for transmitters located at different base stations, unless the
- There are theoretical limitations for the prediction performance of fading channels, as shown in Section 2.2. However, if system delays are kept short, e.g. < 10 ms, then predictions through Kalman filters can ensure coherent JT CoMP gains even at vehicular of e.g. 60 km/h users at a carrier frequency of 500 MHz. It is hence important to keep the system delays low to achieve gains from coherent JT CoMP.
- It is of equal importance to design the CoMP cluster such that intercluster interference is limited. As the system delays and pilot overhead increase with increased cluster sizes, the intercluster interference must be kept low also in clusters of limited size. The scheme of e.g. [21] can ensure this.

Acknowledgments

The author would like to thank Ericsson Research for providing the channel measurements. The research was funded by the Swedish Research Council via the framework program Dynamic Multipoint Wireless Transmission.

Appendix A: Simulating Kalman predictions for block fading channels

The results in Section 5 of this report are all based on measured channel. However, sometimes channels and channel predictions need to be generated for simulations.

If a block fading channel model is used, then each channel component $h_{jnmk}(t)$ is independent for all j, n, m and k. We wish to provide channel estimates and predictions with appropriate statistical properties of the prediction errors for such a simulation environment. At a given time instant t, the prediction error statistics should be similar to one that would have been obtained if Kalman predictions have been applied. (The time correlation of channel predictions and prediction errors are however not of interest in a block-fading simulation environment.) Then, the simulated channels $h_{jnmk}(t)$ and channel predictions $\hat{h}_{jnmk}(t+\tau|t)$ should satisfy the conditions

$$\begin{aligned} E\left[h_{jnmk}(t)h^*_{jnmk}(t)\right] &= \sigma^2_{h,j}, \\ E\left[\left(h_{jnmk}(t)-\hat{h}_{jnmk}(t|t-\tau)\right)\hat{h}^*_{jnmk}(t|t-\tau)\right] &= 0, \\ E\left[\left(h_{jnmk}(t)-\hat{h}_{jnmk}(t|t-\tau)\right)\left(h_{jnmk}(t)-\hat{h}_{jnmk}(t|t-\tau)\right)^*\right] &= \sigma^2_{\delta h,j}, \end{aligned} \tag{65}$$

where the channel variance $\sigma^2_{h,j}$ is given by some pathloss model possibly in combination with a shadow fading model and/or an antenna power gain factor. For a given assumed fading statistics, the variances of the channel errors $\sigma^2_{\delta h,j}$ are given by the diagonal elements of (20) which can be calculated through (17) with the assumption (21) that the filter is stationary. For block fading channels, the w jointly predicted subcarriers in each filter are assumed to be be flatfading, i.e. perfectly correlated over subcarriers. The covariance matrix of the channels $R_{h,j}$ used to calculate Q in (48) and (55) is given by

$$R_{h,j} = \mathrm{diag}\left\{\sigma^2_{h,j}\right\}_{i=1,\ldots,w\cdot N_{tx}\cdot N_{rx}}.$$

The poles should be placed to create a desired Doppler spectrum, e.g. one of the spectra in Figure 2, re-scaled to the appropriate terminal velocity and carrier frequency.

For zero mean channels, the expected values of (65) are met if the prediction error $\delta h_{jnmk}(t) = h_{jnmk}(t) - \hat{h}_{jnmk}(t|t-\tau)$ and the channel prediction $\hat{h}_{jnmk}(t|t-\tau)$ are modeled as mutually uncorrelated circular symmetric, zero mean complex Gaussian with variances $\sigma^2_{\delta h,j}$ and $\sigma^2_{h,j} - \sigma^2_{\delta h,j}$ respectively.

In Appendix 4.E of [11] a different approach for simulating channel predictions is described. However, that approach only approximately meets the requirements of (65).

Appendix B: Average NMSE statistics

The average NMSE for all investigations in Section 5 are shown in Table 5.

Table 5: Comparisons of the NMSE for channel predictions and outdated channel CSIT averaged over all base stations and the weakest base station only. Results are presented for different prediction horizons and noise floors with resource-orthogonal pilots (ROP) and code-orthogonal pilots (COP). Note that some combinations have not been investigated and therefore do not have an average value.

τ	σ_n^2 [dBm]	NMSE (all)			NMSE (weakest)	
		Predicted		Outdated	Predicted	
		ROP	COP	ROP	ROP	COP
0	-110	-17.8 dB	-16.8 dB	-17.8 dB	-7.1 dB	-5.6 dB
	-120	-23.9 dB	-19.6 dB	-23.9 dB	-12.7 dB	-6.9 dB
	-130	-30.9 dB	-	-30.9 dB	-20.0 dB	-
4	-110	-12.8 dB	-12.4 dB	-10.5 dB	-5.9 dB	-5.2 dB
	-120	-15.3 dB	-13.9 dB	-12.5 dB	-9.4 dB	-6.6 dB
	-130	-17.6 dB	-	-14.0 dB	-13.3 dB	-
8	-110	-11.0 dB	-10.8 dB	-6.9 dB	-4.8 dB	-5.2 dB
	-120	-12.9 dB	-12.2 dB	-7.9 dB	-7.4 dB	-6.6 dB
	-130	-14.8 dB	-	-8.6 dB	-10.3 dB	-
12	-110	-9.6 dB	-	-4.4 dB	-4.0 dB	-
	-120	-11.2 dB	-	-5.0 dB	-5.9 dB	-
	-130	-12.8 dB	-	-5.3 dB	-8.2 dB	-
18	-110	-7.9 dB	-	-1.8 dB	-3.0 dB	-
	-120	-9.2 dB	-	-2.1 dB	-4.1 dB	-
	-130	-10.3 dB	-	-2.2 dB	-5.4 dB	-

Appendix C: Process noise covariance matrix for identically distributed or uncorrelated channels

When all elements in the channel vector that can be described by the AR model in (39) are identically distributed or uncorrelated, then equation (48) will provide a positive semi definite covariance matrix Q_j.

In order to prove this, we need the following lemma:

Lemma 1 Assume that a wide sense stationary process obeys

$$\begin{aligned} x(t+1) &= \mathcal{A}x(t) + \mathcal{B}w(t), \\ \vartheta(t) &= \mathcal{C}x(t), \end{aligned} \tag{66}$$

where $w(t) \in \mathbb{C}^{K\times 1}$is a vector of zero mean circular symmetric Gaussian noise with covariance matrix $\mathcal{Q} = E\left[w(t)w^*(t)\right]$, $\vartheta(t) \in \mathbb{C}^{K\times 1}$, $x(t) \in \mathbb{C}^{Kn_{AR}\times 1}$ with $\Pi = E\left[x(t)x^*(t)\right]$ and

$$\begin{aligned}\mathcal{A} &= \mathrm{diag}\{diag\{\mathbf{a}\}...diag\{\mathbf{a}\}\} \in \mathbb{C}^{Kn_{AR}\times Kn_{AR}},\\ \mathcal{B} &= \mathrm{diag}\{\mathbf{b}...\mathbf{b}\} \in \mathbb{C}^{Kn_{AR}\times K},\\ \mathcal{C} &= \mathrm{diag}\{\mathbf{c}..\mathbf{c}\} \in \mathbb{C}^{K\times Kn_{AR}},\end{aligned} \tag{67}$$

where $\mathbf{a} \in \mathbb{C}^{n_{AR}\times 1}$ $\mathbf{b} \in \mathbb{C}^{n_{AR}\times 1}$ and $\mathbf{c} \in \mathbb{C}^{1\times n_{AR}}$ obeys the diagonal stet-space structure in (27) with at least one non-zero pole. Then

$$\gamma = \mathbf{c}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)\mathbf{c}^*$$

is a real-valued positive constant.

Proof: As the system (66) is stationary it obeys the Lyapunov equation

$$\Pi = \mathcal{A}\Pi\mathcal{A}^* + \mathcal{B}\mathcal{Q}\mathcal{B}^* = \begin{bmatrix}\mathbf{a}\\ \vdots \\ \mathbf{a}\end{bmatrix}\begin{bmatrix}\mathbf{a}^* & \cdots & \mathbf{a}^*\end{bmatrix} \odot \Pi + \mathcal{B}\mathcal{Q}\mathcal{B}^*$$

where the second equality comes from the fact that $\mathcal{A}_i$ is diagonal. We can rewrite this into

$$\begin{aligned}\Pi &= \mathcal{B}\mathcal{Q}\mathcal{B}^* \oslash \left[\mathbf{1} - \begin{bmatrix}\mathbf{a}\\ \vdots \\ \mathbf{a}\end{bmatrix}\begin{bmatrix}\mathbf{a}^* & \cdots & \mathbf{a}^*\end{bmatrix}\right],\\ &= \begin{pmatrix} q_{11}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & q_{12}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & \cdots & q_{1K}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)\\ q_{12}^*\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & q_{22}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & \cdots & q_{2K}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)\\ \vdots & \vdots & \ddots & \vdots\\ q_{1K}^*\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & q_{2K}^*\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right) & \cdots & q_{KK}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)\end{pmatrix}.\end{aligned} \tag{68}$$

As the covariance matrix Π is positive semi-definite, any of the principal submatrices $q_{kk}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)$ of the right hand side of (68) must also be positive definite and as $\mathcal{Q}$ is a covariance matrix, that implies that its diagonal elements q_{kk} must be real valued positive constants. Hence $\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)$ is a positive semi definite matrix. That implies that for any vector $\mathbf{c}$

$$\mathbf{c}\left(\mathbf{b}\mathbf{b}^* \oslash \left(\mathbf{1} - \mathbf{a}\mathbf{a}^*\right)\right)\mathbf{c}^* \geq 0.$$

To see that this can be turned into a strict equality, we assume the diagonal form of (27). Then

$$\gamma = \sum_{n_1=1}^{n_{AR}}\sum_{n_2=2}^{n_{AR}} \prod_{\substack{n_3=1\\ n_3\neq n_1}}^{n_{AR}} \prod_{\substack{n_4=1\\ n_4\neq n_2}}^{n_{AR}} \frac{\left(p_{n_1}p_{n_2}^*\right)^{n_{AR}-1}\left(\left(p_{n_1}-p_{n_3}\right)\left(p_{n_2}-p_{n_4}\right)^*\right)^{-1}}{1-p_{n_1}p_{n_2}^*}.$$

As at least one pole is assumed non-zero,$\gamma > 0$.

Theorem 1 Assume that all elements in the channel vector $h_j(t)$ are identically distributed or uncorrelated. That is, the elements in the channel vector can be arranged such that

$$h_j(t) = \left[\begin{array}{cccc} h_{j2}^T(t) & h_{j3}^T(t) & \cdots & h_{j\mathcal{N}}^T(t) \end{array} \right]^T,$$

where $h_{j\nu}(t)$ is a vector consisting of a total of $\mathcal{M}_\nu$ identically distributed elements $\{h_{j\nu\mu}(t)\}_{\mu=1,...,\mathcal{M}_\nu}$ that can be modeled by identical stable AR models

$$\begin{aligned} x_{j\nu\mu}(t+1) &= \text{diag}\left\{\mathbf{a}_{j\nu}\right\} x_{j\nu\mu}(t) + \mathbf{b}_{j\nu} e_{j\nu\mu}(t), \\ h_{j\nu\mu}(t) &= \mathbf{c}_{j\nu} x_{j\nu\mu}(t). \end{aligned} \tag{69}$$

and any element in $h_{j\nu}(t)$ is uncorrelated with any element in $h_{j\theta}(t)$ for $\nu \neq \theta$. Then the process noise covariance matrix estimation by (48) will provide a positive semi definite matrix.

Proof: Let $h_j(t)$ follow the model (39), then the covariance matrix of the channel vector and the state space matrices, have the structure

$$\begin{aligned} A_j &= \text{diag}\left\{\text{diag}\left\{\text{diag}\left\{\mathbf{a}_{j1}\right\}, ..., \text{diag}\left\{\mathbf{a}_{j1}\right\}\right\}, ..., \text{diag}\left\{\text{diag}\left\{\mathbf{a}_{j\mathcal{N}}\right\}, ..., \text{diag}\left\{\mathbf{a}_{j\mathcal{N}}\right\}\right\}\right\}, \\ B_j &= \text{diag}\left\{\text{diag}\left\{\mathbf{b}_{j1}, ..., \mathbf{b}_{j1}\right\}, ..., \text{diag}\left\{\mathbf{b}_{j\mathcal{N}}, ..., \mathbf{b}_{j\mathcal{N}}\right\}\right\}, \\ C_j &= \text{diag}\left\{\text{diag}\left\{\mathbf{c}_{j1}, ..., \mathbf{c}_{j1}\right\}, ..., \text{diag}\left\{\mathbf{c}_{j\mathcal{N}}, ..., \mathbf{c}_{j\mathcal{N}}\right\}\right\}, \\ R_{hj} &= \text{diag}\left\{R_{hj,1}, ..., R_{hj,\mathcal{N}}\right\}, \\ \Pi_j &= \text{diag}\left\{\Pi_{j1}, ...\Pi_{j\mathcal{N}}\right\}. \end{aligned}$$

Here, $R_{hj,\nu} = E\left[h_{j\nu}(t)h_{j\nu}^*(t)\right]$ are positive semi definite matrices. By (48) we then have that Q_j is a block diagonal matrix with

$$Q_j = \text{diag}\left\{\frac{1}{\gamma_{j\nu}} R_{hj,\nu}\right\}_{\nu=1,...,\mathcal{N}},$$

where $\gamma_{j\nu} = \mathbf{c}_{j\nu}\left(\mathbf{b}_{j\nu}\mathbf{b}_{j\nu}^* \oslash \left(\mathbf{1} - \mathbf{a}_{j\nu}\mathbf{a}_{j\nu}^*\right)\right)\mathbf{c}_{j\nu}^*$ is a real-valued positive scalar due to Lemma 1. This implies that $\frac{1}{\gamma_{j\nu}} R_{h,\nu}$ are positive semi definite matrices leading to Q_j being positive definite.

Appendix D: Estimating the process noise covariance matrix for correlated and differently distributed channels

The different methods of estimating the covariance matrix Q_j that where discussed in subsection 3.4 have been evaluated by simulations in [35]. Here, we have assumed an OFDM massive MIMO CoMP system with nine base stations located at three sites spaced by 500 m, each with three sectored base stations. The base stations were each equipped with 32 antennas. In the evaluation discussed below a single antenna user was positioned at a distance of 407 m, 383 m, 738 m from the three sites. Then channels were simulated using the quadriga channel simulator [36]. Each set of 32 antennas was used to create eight beams, giving a total of 72 beams. Measurements were created by transmitting overlapping non code orthogonal pilots from each of the 72 beams over a total of 18 time-frequency resources (consisting of six frequency selective subcarriers spanning 90 kHz by three subsequent OFDM symbols). From the perspective of the user,

different beams will in this situation have very different strength. In [35], this is utilized to lower the complexity of the estimation. Instead of estimating the channels from all 72 beams, each user only has to estimate the subset of the channels from those beams that are strongest, as seen from that user.

To evaluate the different alternatives of estimating the process noise covariance matrix, we therefore introduced a reduced channel vector consisting of the channels on the six adjacent subcarriers from the six strongest beams only. The subchannels to these beams were then modeled using fourth order AR models where the state space matrices where estimated according to Section 3. The process noise covariance matrix Q_j is estimated both by the pseudo inverse as in (51) and also by optimizing the elements in the upper triangular matrix M of (52) . For the later alternative, we use four different starting points for M as follows

chol The initial matrix M is found by cholesky decomposition of the solution to (51) .

diag The initial matrix M is a diagonal matrix with diagonal elements, given by the squared root of the diagonal elements of the solution Q_j to (51) .

rand The elements of the initial matrix M are randomly drawn from a zero mean complex Gaussian distribution with unit variance.

unit The initial matrix M is given by the unit matrix.

The upper triangular matrix M (52) is optimized by using the matlab optimization function **fmincon**. As a optimization criterion we evaluate the NMSE that would be achieved for a Q_j given by each M through (52) and aim to minimize this. For further details regarding the simulations, please refer to [35].

Figure 16 shows the results of Kalman estimations when the process noise covariance matrix Q_j is estimated by the pseudo inverse in (51) and by optimizing the elements in the upper triangular matrix M of (52) with respect to the NMSE. For comparisons there is also an option in which the channels from different beams were assumed to be uncorrelated by setting the covariance matrix R_{hj} in (54) to a block diagonal matrix and then calculate Q_j through (48).

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 unfeasible option for any realistic scenario.

However, we see that both the pseudo inverse and the block diagonal versions of Q_j 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.

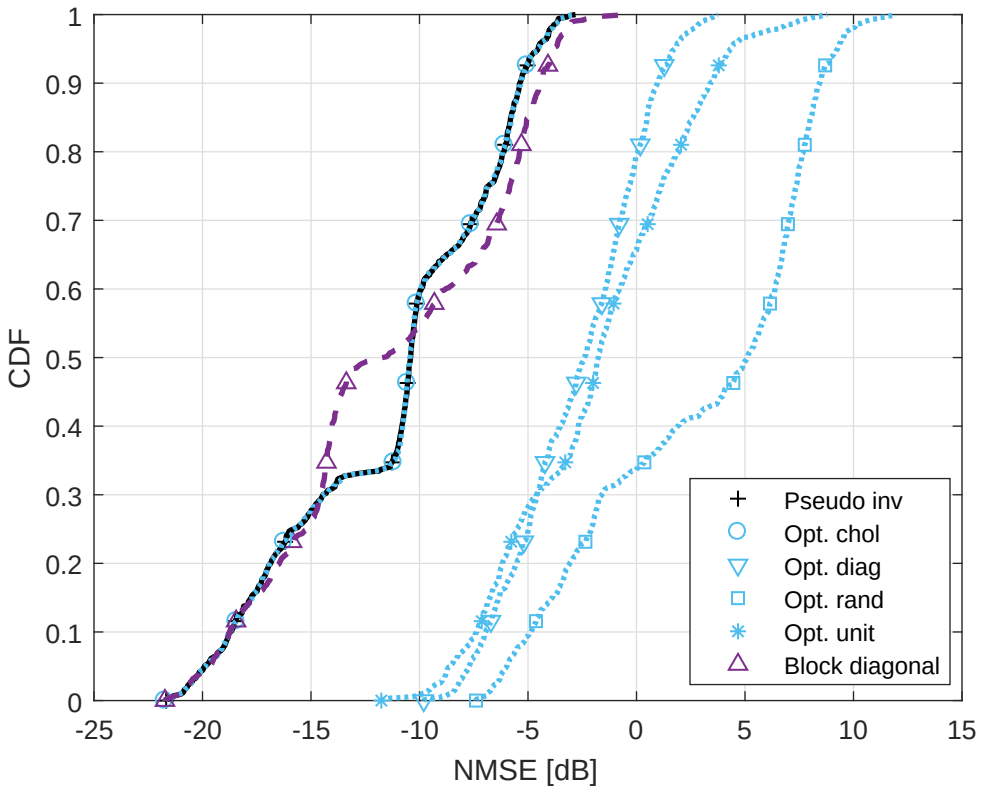


Figure 16: CDF of the NMSE when the process noise covariance matrix Q_j is estimated by the pseudo inverse in (51), solid lines, by optimizing the elements in the uppertriangular matrix M of (52) with different initial values, dotted lines and when the beams are assumed to be uncorrelated, dashed lines.

References

[1] J. Lee, Y.Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Com. Mag.*, vol 50, pp. 44-50, 2012.

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

[3] 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.

[4] W. Lee, I. Lee, J.S. Kwak, B. Ihm and S. Han, "Multi-BS MIMO cooperation: challenges and practical solutions in 4G systems," *IEEE Wireless Com.*, 19, pp. 89-96, 2012.

[5] J. Li, A. Papadogiannis, R. Apelfröjd, T. Svensson and M. Sternad, "Performance analysis of coordinated multi-point transmission schemes with imperfect CSI," in *Proc. of IEEE PIMRC 2012*, 9-12 Sept. 2012.

[6] L. Thiele, M. Olbrich, M. Kurras and B. Matthiesen, "Channel aging effects in CoMP transmission: Gains from linear channel prediction," *45th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, USA, Nov. 6-9 2011.

[7] L. Lindbom, "Simplified Kalman estimation of fading mobile radio channels: high performance at LMS computational load," In *Proc. of ICASSP*, Minneapolis, USA, vol. 3, pp. 352–355, Apr. 1993.

[8] C. Komnimakis, C. Fragouli, A.H. Sayeed, R.D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. on Signal Processing*, vol 50(5), pp. 10651076, May 2002.

[9] R. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME (J. Basic Engineering)*, vol 82 D, pp. 35-45, 1960.

[10] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, Chapter 9, ISBN 0-13-022464-2, 2000.

[11] D. Aronsson, *Channel Estimation and Prediction for MIMO OFDM Systems - Key Design and Performance Aspects of Kalman-based Algorithms*, Ph. D. Thesis, Uppsala University, 2011. http://www.signal.uu.se/Publications/ptheses.html. Accessed 7 Nov. 2013.

[12] 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 Networking*, June 2014.

[13] R. Apelfröjd, M. Sternad and D. Aronsson, "Measurement-based evaluation of robust linear precoding for downlink CoMP," in *Proc. of IEEE ICC 2012*, Ottawa, Canada, 10-15 June 2012.

[14] ARTIST4G D1.4, "Interference avoidance techniques and system design, Artist4G technical deliverable," June 2012, https://ict-artist4g.eu/projet/deliverables. Accessed 7 Nov. 2013.

[15] 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.

[16] M. Sternad, S. Falahati, T. Svensson and D. Aronsson, "Adaptive TDMA/OFDMA for wide area coverage and vehicular velocities," *in Proc. IST Mobile and Vehicular,* Dresden, June 19-23, 2005.

[17] S. Falahati, A. Svensson, M. Sternad and T. Ekman, "Adaptive modulation system for predicted wireless channels," IEEE Trans. on Com., vol 52, pp. 307-316, Feb. 2004.

[18] K. Manolakis, S. Jaeckel, V. Jungnickel and V. Braun, "Channel prediction by doppler-delay analysis and benefits for base station cooperation," In *Proc. of IEEE VTC Spring 2013*, Dresden, Germany, June 2013.

[19] K. J. Åström and B. Wittenmark, *Computer-Controlled Systems*, Prentice Hall, ISBN 0-13-172784-2, 1990 (second edition).

[20] M. Sternad, M. Grieger, R. Apelfröjd, T. Svensson, D. Aronsson and A. Belén Martinez, "Using "Predictor Antennas" for Long-Range Prediction of Fast Fading for Moving Relays," in *Proc. of IEEE WCNC 2012*, 4G Mobile Radio Access Networks Workshop , Paris, April 2012.

[21] W. Mennerich and W. Zirwas, "Reporting effort for cooperative systems applying interference floor shaping," in *Proc. of IEEE PIMRC 2011*, Toronto, Canada, 11-14 Sept. 2011.

[22] W. Zirwas and W. Menerich, "The importance of interference floor shaping for CoMP systems," in *Proc. of International OFDM Workshop*, Hamburg, Germany, 31 Aug. 2011.

[23] C.T.K. Ng and H Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J Sel Areas Com.*, vol 28(9), pp. 1446–1454, 2010.

[24] D. Aronsson and M. Sternad, "OFDMA uplink channel prediction to enable frequency-adaptive multiuser scheduling," *EUSIPCO 2007,* Poznan, Poland, September 2007.

[25] T. Wild, "Comparing downlink coordinated multi-point schemes with imperfect channel knowledge," in *Proc. of IEEE VTC Fall 2011*, San Francisco, CA, USA, 5-8 Sept. 2011.

[26] T. Ekman, *Prediction of Mobile Radio Channels*, Chapter 1, Ph. D. Thesis, Uppsala University, 2002. http://www.signal.uu.se/Publications/ptheses.html. Accessed 5 May 2014.

[27] T. Wigren, "Recursive noise floor estimation in WCDMA," *IEEE Trans. on Vehicular Techn.*, vol. 59, pp. 2615-2620, 2010.

[28] T. Kailath, *Linear Systems*, Prentice Hall, Chapter 2, ISBN 0-13-536961-4, 1980.

[29] T. Söderström and P. Stoica, *System Identification*, Chapter 8, Prenctice Hall, ISBN 0-13-881236-5, 1989.

[30] IST-4-027756 WINNER II, D1.1.2v1.2: "WINNER II channel models," http://www.ist-winner.org/deliverables.html. Accessed 16 April 2014.

[31] E. Dahlman, S. Parkvall and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*, Academic Press, ISBN: 978-0-12-385489-6, 2011.

[32] R. Fritzsche, E. Ohlmer and G. Fettweis, "Where to predict the channel in cooperative cellular networks with backhaul delays," in *Proc. of the 9th International ITG Conference on Systems, Communications and Coding (SCC'13)*, Munich, Germany, Jan. 2013.

[33] N. Seifi, M. Viberg, R.W. Heath, J. Zhang and M. Coldrey, "Multi-mode transmission in network MIMO downlink with incomplete CSI," *EURASIP Jour. on Advances in Signal Processing*, vol. 2011, Article ID 743916, doi:10.1155/2011/743916.

[34] J. Medbo, I. Siomina, A. Kangas and J. Furuskog, "Propagation channel impact on LTE positioning accuracy - A study based on real measurements of observed time difference of arrival," in *Proc. of IEEE PIMRC 2009*, Tokyo, Japan, Sept. 2009.

[35] R. Apelfröjd, W. Zirwas and M. Sternad, "Joint Reference Signal Design and Kalman/Wiener Channel Estimation for FDD Massive MIMO". Available on http://www.signal.uu.se/Publications/pdf/r1701.pdf.

[36] Fraunhofer Heinrich Hertz Institute, "Quasi deterministic radio channel generator user manual and documentation", Fraunhofer Heinrich Hertz Institute, Berlin 2016. Available on http://quadriga-channel-model.de/#Documentation.