Pilot Decontamination Through Pilot Sequence Hopping in Massive MIMO Systems

Jesper H. Sørensen and Elisabeth de Carvalho
Aalborg University, Department of Electronic Systems, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark
E-mail: {jhs,edc}@es.aau.dk

Abstract—This work concerns wireless cellular networks applying massive multiple-input multiple-output (MIMO) technology. In such a system, the base station in a given cell is equipped with a very large number (hundreds or even thousands) of antennas and serves multiple users. Estimation of the channel from the base station to each user is performed at the base station using an uplink pilot sequence. Such a channel estimation procedure suffers from pilot contamination. Orthogonal pilot sequences are used in a given cell but, due to the shortage of orthogonal sequences, the same pilot sequences must be reused in neighboring cells, causing pilot contamination. The solution presented in this paper suppresses pilot contamination, without the need for coordination among cells. Pilot sequence hopping is performed at each transmission slot, which provides a randomization of the pilot contamination. Using a modified Kalman filter, it is shown that such randomized contamination can be significantly suppressed. Comparisons with conventional estimation methods show that the mean squared error can be lowered as much as an order of magnitude at low mobility.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology [1] is finding its way into practical systems, like LTE and its successor LTE-Advanced. It is a key component for these systems’ ability to improve the spectral efficiency. The success of MIMO technology has motivated research in extending the idea of MIMO to cases with hundreds, or even thousands of antennas, at transmitting and/or receiving side. This is often termed massive MIMO. In mobile communication systems, like LTE, the more realistic scenario is to have a massive amount of antennas only at the base station (BS), due to the physical limitations at the user equipment (UE). It has been shown that such a system [2], in theory, can eliminate entirely the effect of small-scale fading and thermal noise, when the number of BS antennas goes to infinity. The only remaining impairment is inter-cell interference, caused by imperfect channel state information (CSI), which is a result of non-orthogonality of training pilots used to gather the CSI. This is often referred to as pilot contamination. It is considered as one of the major challenges in massive MIMO systems [3].

Mitigation of pilot contamination has been the focus of several works recently. These fall into two categories; one with coordination among cells and one without. The first category includes [4], where it is utilized that the desired and interfering signals can be distinguished in the channel covariance matrices, as long as the angle-of-arrival spreads of desired and interfering signals do not overlap. A pilot coordination scheme is proposed to help satisfying this condition. The work in [5] utilizes coordination among base stations to share downlink messages. Each BS then performs linear combinations of messages intended for users applying the same pilot sequence. This is shown to eliminate interference when the number of base station antennas goes to infinity.

The major contribution of this paper is a pilot decontamination, which does not require inter-cell coordination, and is able to exploit past pilot signals. It is based on pilot sequence hopping performed within each cell. Pilot sequence hopping means that every user chooses a new pilot sequence in each transmission slot. Consider a user of interest and the effect of the inter-cell pilot contamination when pilot sequence hopping is applied. At each transmission slot, the pilot signal of the user is contaminated by a different set of interfering users. Hence channel estimation at each transmission slot is affected by a different set of interfering channels. If channel estimation is carried out based solely on the pilot sequence of the current slot, then pilot sequence hopping does not bring any gain. The key in our solution is a channel estimation that incorporates multiple time slots so that it can benefit from randomization of the pilot contamination. Recent work utilizing temporal correlation for channel estimation is found in [12], although not in combination with pilot hopping and not with the purpose of mitigating pilot contamination.

Consider the simple example, where the channel of the UE of interest is time-invariant. Its estimation is performed across multiple time slots. Specifically, the resulting channel estimate is the average of the estimates across the time slots. In the
averaging process, the contamination signal is averaged out. 
Note that, if the contamination signal remains constant across 
the time slots, i.e. there is no hopping, this averaging brings 
no benefit (except an averaging of the receive noise).

When the channel is time-variant and correlated across 
time slots, it remains possible to exploit the information 
about the channel across time slots by an appropriate filtering 
and benefit from contamination randomization. In this paper, 
channel estimation across multiple time slots is performed 
using a modified version of the Kalman filter, which is capable 
of tracking the channel and the channel correlation. The 
level of contamination suppression depends on the channel 
correlation between slots of the UE of interest as well as the 
contaminators. In LTE, channel correlation between time slots 
is large even at medium-high speeds, making the proposed 
solution very efficient.

The remainder of this paper is organized as follows. Section 
II presents the applied system model and the problem of pilot 
contamination. The proposed solution is described in section 
III and evaluated and compared to existing solutions in section 
IV. Finally, conclusions are drawn in section V.

II. SYSTEM MODEL

In this work we denote scalars in lower case, vectors in 
bold lower case and matrices in bold upper case. A superscript 
‘\text{T}’ denotes the transpose and a superscript ‘\text{H}’ denotes the 
conjugate transpose.

This work treats a cellular system consisting of \( L \) cells 
with \( K \) users in each cell. A massive MIMO scenario is 
considered, where the BS has \( M \) antennas and the UE has 
a single antenna. We restrict our attention to the channel 
estimation performed in a single cell, which we term “the cell 
of interest” and assign the index “0”. The channel between 
the BS in the cell of interest and the \( k \)’th user in the \( \ell \)’th cell 
is denoted \( h_{k\ell} \), where the individual channel coefficients are complex scalars. Note 
that for \( \ell > 0 \), \( h_{k\ell} \) refers to a channel between the BS of 
interest and a UE connected to a different base station. We 
 furthermore restrict our attention to the estimation of a single 
channel coefficient, hence a channel is denoted as the complex 
scalar \( h_{k\ell} \). The work easily extends to vector estimations, in 
which case spatial correlation can be exploited for improved 
performance. A rich scattering environment is assumed, such 
that \( h_{k\ell} \) can be modeled using Clarke’s model [13], hence

\[
\hat{h}_{k\ell} = \frac{1}{\sqrt{N_s}} \sum_{m=1}^{N_s} e^{2\pi f_d t \cos \alpha_m + \phi_m},
\]

where \( N_s \) is the number of scatterers, \( f_d \) is the maximum 
Doppler shift, \( \alpha_m \) and \( \phi_m \) is the angle of arrival and initial 
phase, respectively, of the wave from the \( m \)’th scatterer. Both 
\( \alpha_m \) and \( \phi_m \) are i.i.d. in the interval \([−\pi, \pi)\) and 
\( f_d = \frac{v}{c} f_c \), where \( v \) is the speed of the UE, \( c \) is the speed of light and \( f_c \) 
is the carrier frequency.

In a massive MIMO system, collection of channel state 
information (CSI) is performed using uplink pilot training.

The CSI achieved this way is utilized in both downlink 
and uplink transmissions based on the channel reciprocity 
assumption. We define a pilot training period followed by 
an uplink and a downlink transmission period as a time slot. 
See Fig. 2 for an example of a transmission schedule with 
two time slots. During the \( n \)’th pilot training period, the 
\( k \)’th user in the \( \ell \)’th cell transmits a pilot sequence \( x_{n\ell}^{k\ell} = \left[ x_{n\ell}^{k\ell}(1) x_{n\ell}^{k\ell}(2) \ldots x_{n\ell}^{k\ell}(\tau) \right]^T \), 
where \( \tau \) is the pilot sequence length. Ideally, all pilot sequences in the 
total system are orthogonal, in order to avoid interference. However, this would 
require pilot sequences of at least length \( L \cdot K \), which in 
most practical systems is not feasible. Instead, orthogonality 
within each cell only is ensured, i.e. \( \tau = K \), thereby dealing 
with the potentially strongest sources of interference. As a 
result, all cells use the same set of pilots, potentially causing 
interference from neighboring cells. This is referred to as pilot 
contamination. We define the contaminating set, \( C_{n\ell}^{k\ell} \), as the set 
of all pairs \( i, j \), which identify all UEs applying the same pilot 
sequence in the \( n \)’th time slot as the \( k \)’th user in the \( \ell \)’th cell. 
Hence, \( x_{n\ell}^{ij} = x_{n\ell}^{k\ell} \forall i, j \in C_{n\ell}^{k\ell} \).

The pilot signal received by the BS of interest, concerning 
the \( k \)’th user in the \( n \)’th time slot can be expressed as

\[
y_{n\ell}^{k\ell} = h_{n\ell}^{k\ell} x_{n\ell}^{k\ell} + \sum_{i,j \in C_{n\ell}^{k\ell}} h_{n\ell}^{ij} x_{n\ell}^{ij} + z_{n\ell}^{k\ell},
\]

where \( z_{n\ell}^{k\ell} = \left[ z_{n\ell}^{k\ell}(1) z_{n\ell}^{k\ell}(2) \ldots z_{n\ell}^{k\ell}(\tau) \right]^T \) and \( z_{n\ell}^{k\ell}(j) \) are
circularly symmetric Gaussian random variables with zero mean and unit variance for all $j$. Here, only signals leading to contamination are included in the sum term, since any $h_{nl}^{[l]} \mathbf{x}_{nl}^{[l]}$ where $i, j \neq C_{nl}^{[l]}$ are removed when correlating with the applied pilot sequence. Hence, all contributions from the sum term are undesirable and will contaminate the CSI. Without loss of generality, we focus on the channel estimation for a single user in a single cell. Hence, in the remainder of the paper, we omit the superscript $k$ for ease of notation.

### III. Pilot Decontamination

The solution to pilot contamination proposed in this work consists of two components:

1) **Pilot sequence hopping:** This component refers to random shuffling of the pilots applied within a cell. This shuffle occurs between every time slot. The purpose of this component is to decorrelate the contaminating signals. When pilots are shuffled, the set of contaminating users will be replaced by a new set, whose channel coefficients are uncorrelated with those of the previous set.

2) **Kalman filtering:** The autocorrelation of the channel coefficient of the user of interest is high at low mobility. This means that information about the value of the current channel coefficient exists not only in the most recent pilot signal, but also in past pilot signals. This can be extracted using a filter. For this purpose a Kalman filter is desirable due to its recursive structure, which provides low complexity, yet optimal performance. Additionally, since the contaminating signals have been decorrelated, the Kalman filter will suppress the impact of these signals, leading to pilot decontamination.

#### A. Pilot Sequence Hopping

Pilot sequence hopping is a technique where the UEs randomly switch to a new pilot sequence in between time slots. This must be coordinated with the BS, which in practice can be realized by letting the BS send a seed for a pseudorandom number generator to each UE. Random pilot sequence hopping is illustrated in Fig. 3 in the case of $\tau = K = 5$. Note how the identity of the contaminator changes between time slots, as opposed to a fixed pilot sequence schedule, where the contaminator remains the same UE. Consequently, the undesirable part of the pilot signal, i.e. the sum term in (2), varies rapidly between time slots compared to the variation caused by the mobility of a single contaminator in a fixed schedule. In fact, the impact of pilot sequence hopping, from a contamination perspective, can be viewed as a dramatic increase of the mobility of the contaminator. This in turn leads to a lowered autocorrelation, or decorrelation, in the contaminating signal, which is the motivation behind performing pilot sequence hopping.

The level of decorrelation is related to the time between two instances, where the same user acts as a contaminator. We refer to this as the collision distance, and we denote it $t_c$, see Fig. 3. Note that in the case of a fixed pilot schedule, $t_c = 1$. The goal of pilot sequence hopping is to maximize $t_c$, either in an expected sense or maximum sense, i.e. maximization of the minimum value. The latter can be pursued through a minimal level of coordination of pilot sequence schedules among neighboring cells. However, this work is strictly restricted to a framework with no inter-cell coordination, hence, we focus on the expected value of $t_c$. If pilot sequence hopping is performed at random and $\tau = K$, then $t_c$ follows a geometric distribution, such that

$$P(t_c = d) = (1 - p)^{d-1}p, \quad d = 1, 2, \ldots,$$

$$p = \frac{1}{K}, \quad (3)$$

where $P(t_c = d)$ is the probability that the collision distance is $d$ and $p$ is the probability of a given UE being the next contaminator. The expected value of $t_c$, $\mathbb{E}[t_c]$, is then found as

$$\mathbb{E}[t_c] = \sum_{d=1}^{\infty} d(1 - p)^{d-1}p$$

$$= \sum_{d=1}^{\infty} d \left( \frac{K-1}{K} \right)^{d-1} \frac{1}{K}$$

$$= K. \quad (4)$$

Hence, the expected collision distance increases with the number of users/pilots per cell, which follows intuition.

![Fig. 3.](image-url) "An example of a random pilot schedule for the UE of interest and potential contaminants in a neighboring cell. Green boxes represent pilots, which are orthogonal to the pilot from the UE of interest. Red boxes represent contamination and $\mathbf{x}_i$ denotes a pilot sequence."

#### Example:

To help the understanding of the benefit from pilot sequence hopping, consider the ideal case of a constant channel between BS and UE of interest and a single contaminating neighboring cell. Noise is disregarded in this example, since attention is on decontamination. Moreover, we assume an infinite amount of orthogonal pilot sequences and
an infinite amount of users per cell, such that \( \tau = K = \infty \) and \( \mathbb{E}[t_c] = \infty \), which means contaminating signals in all time slots are independent. For simplicity, we assume \( x_{n}^H x_n = 1 \), such that the estimate in time slot \( n \) is

\[
\hat{h}_n = h + h'_n, \tag{5}
\]

where \( h'_n \) is the channel of the contaminator in time slot \( n \). Now consider a new estimator, \( \tilde{h}_n \), which is the average of all estimates until time slot \( n \). Hence, we have

\[
\tilde{h}_n = \frac{1}{n} \sum_{i=1}^{n} h'_i. \tag{6}
\]

In this case, the error in the estimate is solely composed of the average of the contaminating signals, which are independent and have variance \( \sigma^2 \). Hence, the variance of the estimation error is \( \frac{\sigma^2}{n} \). If pilot sequence hopping had not been performed, the variance of the estimation error had remained \( \sigma^2 \), since \( h'_n \) would be constant. Note that the MSE goes towards zero for \( n \to \infty \), when pilot sequence hopping is performed. This is a result of the fact that a pilot signal in the infinite past carries as much information about the current channel as the most recent pilot signal, in the ideal example of a constant channel. Note also that for finite \( \tau \) (and \( K \)) and thereby finite \( \mathbb{E}[t_c] \), the variance of the estimation error is lower bounded by \( \frac{\sigma^2}{n} \), since only a maximum of \( K \) independent estimates can be achieved. In a more practical example with a time-varying channel, the amount of information carried in a pilot signal decays over time. It is, however, still possible to extract such information using appropriate filtering techniques. For this purpose we have chosen a modified version of the Kalman filter, which is described next.

**B. Modified Kalman Filter**

A conventional Kalman filter can be used to track the state, \( b_n \), of a system based on observations, \( y_n \), where

\[
y_n = C_n b_n + d_n, \tag{7}
\]

and \( C_n \) is the measurement matrix of the system and \( d_n \) is measurement noise. Moreover, the evolution of the system state must follow

\[
b_n = A_n b_{n-1} + v_n, \tag{8}
\]

where \( A_n \) is the state transition matrix and \( v_n \) is the process noise. In a conventional application of the Kalman filter, \( A_n \) is assumed constant and known.

The problem of estimating a time-varying channel based on pilot signals, also termed channel tracking, can be solved using the Kalman filter. The observations as expressed in (2) follow the linear model in (7), where the observation matrix is the transmitted pilot sequence and the tracked state is the channel coefficient. The evolution of the channel coefficient as expressed by Clarke’s model does not follow the model in (8). However, it can be transformed into an autoregressive (AR) model with a finite number of coefficients, which follows the form of (8). If the instantaneous velocity of the user of interest, and thereby the autocorrelation function, are known, the AR coefficients can be found using the Yule-Walker equations [14]. However, this cannot be assumed in our case, hence the AR coefficients must be tracked along with the channel state. For this purpose, we must modify the conventional Kalman filter to include an AR model tracker. A 1st order AR model is applied, since experiments tell us this adequately captures the autocorrelation of the system. Therefore, only a single AR coefficient, \( a_n \), must be tracked.

First we state the conventional Kalman filter [15] in our context, where the AR coefficient is assumed known.

For all \( n \):

\[
e_n = y_n - x_{n-1} a_n \hat{h}_{n-1}, \tag{9}
\]

\[
R_n = x_n p_n x_n^H + \sigma^2 \mathbb{I}, \tag{10}
\]

\[
k_n = p_n x_n^H R_n^{-1}, \tag{11}
\]

\[
\hat{h}_n = a_n \hat{h}_{n-1} + k_n e_n, \tag{12}
\]

\[
p_{n+1} = a_n^2 (1 - k_n) + (1 - a_n)^2, \tag{13}
\]

where \( \sigma^2 \) and \( \sigma_n^2 \) are noise power and total contamination power (average over time), respectively, which are both assumed known, \( \mathbb{I} \) is the \( \tau \times \tau \) identity matrix and \( \hat{h}_n \) is the estimate of \( h_n \).

For the tracking of the AR coefficient, an approach similar to the one in [16] is taken. In [16] the inclusion of an AR coefficient tracker is presented for a Kalman predictor, i.e. a filter with the purpose of predicting the channel, \( h_n \), based on all observations until \( y_{n-1} \). In this work, we extend this approach to take all observations until \( y_n \) into account.

The approach is based on calculating the partial derivative with respect to \( a_n \) of the cost function, the mean squared error (MSE), and using this to adjust \( a_n \) in the direction of decreasing MSE. The partial derivative of the MSE is

\[
\nabla_n = \frac{\partial}{\partial a_n} \mathbb{E}[e_n^2] = - (q_{n-1}^H x_{n-1} + \hat{h}_{n-1}^H x_{n}^H) e_n, \tag{14}
\]

where \( q_n = \frac{\partial h_n}{\partial a_n} \) and is found by differentiating (12) with respect to \( a_n \), such that

\[
q_n = (1 - k_n) a_n (a_n q_{n-1} + \hat{h}_{n-1}) + m_n e_n. \tag{15}
\]

Here, \( m_n = \frac{\partial x_n}{\partial a_n} \), which is found by differentiating (11) with respect to \( a_n \), hence

\[
m_n = (1 - k_n) s_n x_n^H R_n^{-1}. \tag{16}
\]
Finally, we introduced \( s_n = \frac{\partial p_n}{\partial a_n} \), which is a differentiation of (13) with respect to \( a_n \), giving us

\[
s_{n+1} = a_n^2 (1 - k_n x_n) s_n (1 - x_n^H R_n^H) - 2a_n k_n x_n p_n. \tag{17}
\]

Using \( \nabla_n \), we can adjust \( a_n \) as follows

\[
a_n = [a_{n-1} - \mu (\nabla_n)^{+\nu}]_{0}, \tag{18}
\]

where \( \mu \) is a parameter adjusting the convergence speed and the brackets denote truncations. The inner truncation involving \( \nu \) is to avoid dramatic adjustments in situations with a high slope and the outer truncation is to obey \( 0 \leq a_n \leq 1 \). The need for \( \nu \) will be explained in section IV.

We can now state the modified Kalman filtering algorithm including an AR coefficient tracker:

For all \( n \):

\[
e_n = y_n - x_n a_n^{-1} \hat{h}_{n-1},
\]

\[
R_n = x_n p_n x_n^H + \sigma_n^2 I + \sigma_n^2 x_n x_n^H,
\]

\[
\nabla_n = -(q_n^H a_n - \hat{h}_{n-1} x_n^H + h_n^H x_n^H) e_n,
\]

\[
a_n = [a_{n-1} - \mu (\nabla_n)^{-\nu}]_{0},
\]

\[
k_n = p_n x_n^H R_n^{-1},
\]

\[
\hat{h}_n = a_n \hat{h}_{n-1} + k_n e_n,
\]

\[
m_n = (1 - k_n x_n) s_n x_n^H R_n^{-1},
\]

\[
q_n = (1 - k_n x_n) (a_n q_{n-1} + \hat{h}_{n-1}) + m_n e_n,
\]

\[
p_{n+1} = a_n^2 (1 - k_n x_n) p_n + (1 - a_n)^2,
\]

\[
s_{n+1} = a_n^2 (1 - k_n x_n) s_n (1 - x_n^H R_n^H) - 2a_n k_n x_n p_n. \tag{19}
\]

IV. NUMERICAL RESULTS

The proposed scheme (Estimator) has been simulated and compared to the scheme from [16] (Predictor) and the conventional solutions of least squares (LS) estimation and minimum mean squared error (MMSE) estimation based on a single time slot. The expressions for the LS and MMSE estimators are given in (20) and (21), respectively. An overview of the parameters, which are common for all simulations, is given in Table I. The choice of \( \mu \) is based on experiments showing that this is a good compromise between convergence speed and robustness towards variance. Throughout all simulations, we assume that all users have equal and constant mobility. Moreover, we assume that contaminating signals have zero autocorrelation between time slots, which is justified by the choice of \( K = 96 \), such that \( \mathbb{E}[t_c] = 96 \), cf. (4).

\[
\hat{h}_n^{ls} = (x_n^H x_n)^{-1} x_n^H y_n, \tag{20}
\]

\[
\hat{h}_n^{mmse} = x_n^H (x_n x_n^H + \sigma_n^2 I + \sigma_n^2 x_n x_n^H)^{-1} y_n. \tag{21}
\]

Initially, results are shown for the conventional Kalman filter expressed in equations (9) through (13). MSE as a function of the user mobility, \( v \), and the AR coefficient, \( a_n \), is shown in Fig. 4. From this figure, it is evident how important it is to have an accurate AR model, which suits the current mobility of the UE of interest. This stresses the need for the modification of the Kalman filter, as proposed in section III-B. Moreover, it is seen that the derivative of the MSE with respect \( a_n \) may attain very high values at low \( a_n \). This can cause undesirably high variance in the estimate of the optimal \( a_n \), which motivates the use of a derivative cap, \( \nu \).

Fig. 5 shows a comparison of the simulated estimators with respect to MSE as a function of user mobility when \( \sigma_n^2 = 0.6 \). For both the predictor and the scheme proposed in this work, results where the optimal value of \( a_n \) is assumed to be known, have been included. This highlights the performance of the tracker. It is evident that the tracker provides a very

\begin{table}[h]
\centering
\caption{SIMULATION PARAMETERS}
\begin{tabular}{|c|c|c|}
\hline
Parameter & Value & Description \\
\hline
\hline
\( \sigma_n^2 \) & 0.2 & Noise variance \\
\hline
\( L \) & 7 & Number of cells \\
\hline
\( K \) & 96 & Users per cell \\
\hline
\( \tau \) & 96 & Pilot length \\
\hline
\( \mu \) & \( 10^{-5} \) & Convergence speed \\
\hline
\( \nu \) & 100 & Derivative cap \\
\hline
\( f_c \) & 1.8 GHz & Carrier frequency \\
\hline
\( N_s \) & 20 & Number of scatterers \\
\hline
\( t_s \) & 0.5 ms & Time between pilots \\
\hline
\( \sigma_0 \) & 0.5 & Initial AR coefficient \\
\hline
\( h_0 \) & 0 & Initial estimate \\
\hline
\( q_0 \) & 0 & Initial differentiated estimate \\
\hline
\( p_1 \) & 0 & Initial error covariance \\
\hline
\( s_1 \) & 0 & Initial differentiated error covariance \\
\hline
\end{tabular}
\end{table}
good estimate of the optimal AR coefficient. Moreover, it is seen that the proposed scheme outperforms LS and MMSE and performs as well as the predictor at low mobility. At high mobility, the proposed scheme outperforms LS and the predictor, while matching the performance of MMSE.

A different perspective is given in Fig. 6. Here the MSE is plotted as a function of the signal-to-interference ratio (SIR), at typical mobility levels as defined by 3GPP [17]. This figure shows how the proposed scheme is able to suppress even very strong contamination at typical mobility.

![Fig. 6. Comparison between the proposed scheme and conventional solutions plotted as a function of the signal-to-interference ratio (SIR), with respect to means squared error as a function of mobility.](image1)

Numerical evaluations show that the proposed solution can suppress a significant portion of the contamination at low and moderate levels of mobility. Even at high mobility, i.e. car speeds of 100 to 130 km/h, the proposed solution can provide a noticeable gain over conventional estimation methods.

**References**