CrossMark

Uncoordinated pilot decontamination in massive MIMO systems

Jesper H. Sørensen[*] (ID) and Elisabeth de Carvalho

Abstract

This work concerns wireless cellular networks applying time division duplexing (TDD) massive multiple-input multiple-output (MIMO) technology. Such systems suffer from pilot contamination during channel estimation, due to the shortage of orthogonal pilot sequences. This paper presents a solution based on pilot sequence hopping, which provides a randomization of the pilot contamination. It is shown that such randomized contamination can be significantly suppressed through appropriate filtering. The resulting channel estimation scheme requires no inter-cell coordination, which is a strong advantage for practical implementations. Comparisons with conventional estimation methods show that the MSE can be lowered as much as an order of magnitude at low mobility. Achievable uplink and downlink rates are increased by 42 and 46%, respectively, in a system with 128 antennas at the base station.

Keywords: Massive MIMO, Pilot contamination, Kalman filter

1 Introduction

Muliple-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 BSs 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 BS antennas goes to infinity.

The category without coordination also includes notable contributions. A multi-cell precoding technique is used in [6] with the objective of not only minimizing the mean squared error of the signals of interest within the cell but also minimizing the interference imposed to other cells. In [7], it is shown that channel estimates can be found as eigenvectors of the covariance matrix of the received signal when the number of BS antennas grows large and the system has "favorable propagation." The work in [8–11] is based on examining the eigenvalue distribution of the received signal to identify an interference free subspace on which the signal is projected. It is

---

*Correspondence: jhs@es.aau.dk
Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

shown that an interference free subspace can be identified when certain conditions are fulfilled concerning the number of BS antennas, user equipment antennas, channel coherence time, and the signal-to-interference ratio. Recently, in [12], a combination of the solutions in [4] and [8–11] was proposed. The resulting solution unites the strengths of these solutions leading to a more robust pilot decontamination.

In this paper, we propose 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 time slot. Thus, in every time slot, the pilot signal of a user is contaminated by a different set of interfering users, which means channel estimation 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 [13], although not in combination with pilot hopping and not with the purpose of mitigating pilot contamination. Random selection of pilot sequences is also explored in [14] and [15]. Both works consider the random access problem in cellular networks. In [14], pilot contamination is avoided through a distributed collision detection algorithm, which enables users with weak channels to detect that they are contaminators of a user with a strong channel and as a result postpone their transmission. The work in [15] considers codeword transmissions that are spread across multiple time slots, each with a different contaminator. This decorrelates the contamination within a single codeword, which improves performance.

When the channel is time-variant and correlated across time slots, it is 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.

This work is an extension of the work in [16], where the concept of pilot sequence hopping in combination with a Kalman channel tracker is introduced. In this paper, the work is extended with more sophisticated mobility models and a generalization of the estimation algorithm, which

allows higher order Kalman process models. Furthermore, the Bayesian Cramer-Rao lower bound is derived for the estimation problem at hand and applied as a benchmark in the numerical evaluations.

The remainder of this paper is organized as follows. Section 2 presents the applied channel and mobility models and the problem of pilot contamination. The proposed solution is described in Section 3 and analyzed in Section 4. Section 5 provides numerical results and a comparison to existing solutions. Finally, conclusions are drawn in Section 6.

2 System model
In this work, we denote scalars in lower case, vectors in bold lower case, and matrices in bold upper case. A superscript "T" denotes the transpose, and a superscript "H" denotes the conjugate transpose.
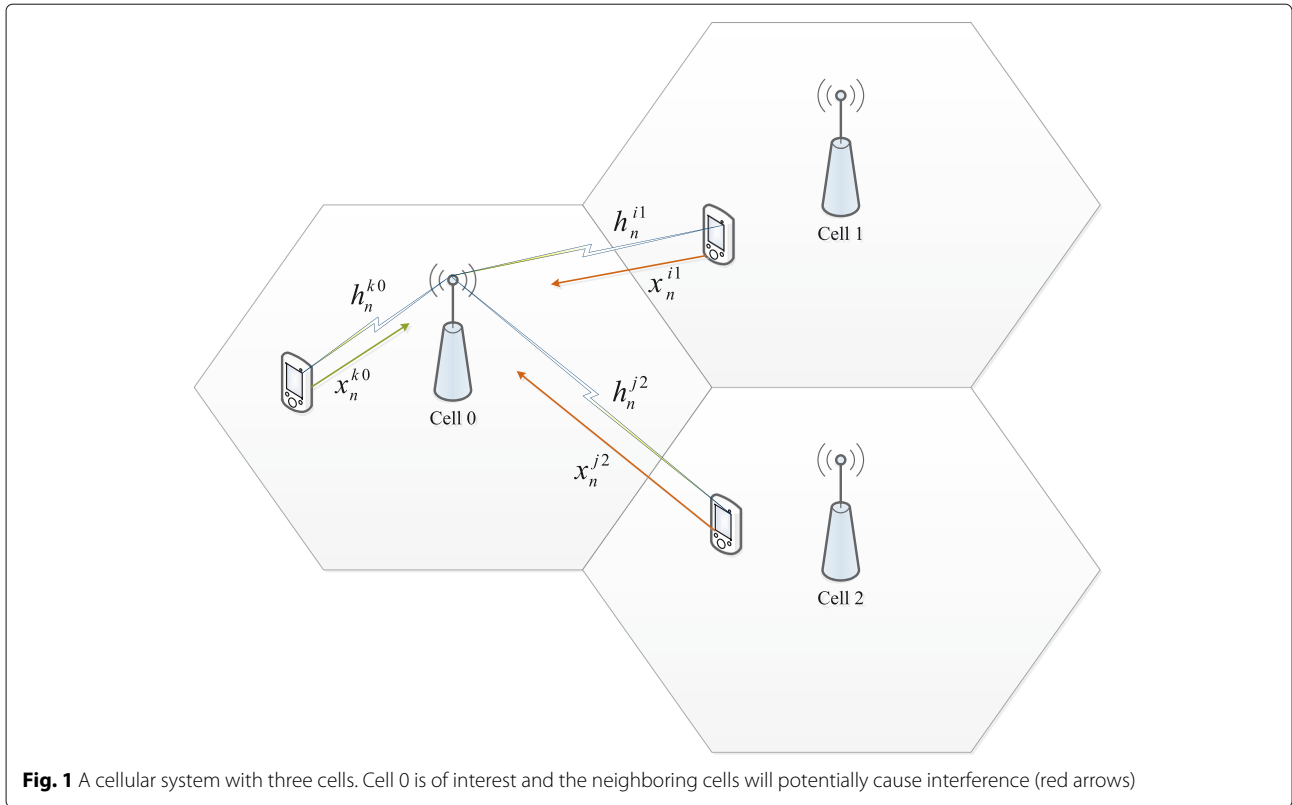
2.1 Channel model
This work treats a cellular system consisting of L cells with K users in each cell, see Fig. 1. A time division duplexing (TDD) massive MIMO scenario is considered, where the BS has M antennas and the UE has a single antenna. We restrict our attention to the channel estimation performed 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 kth user in the ℓth cell is denoted as $\boldsymbol{h}^{k\ell} = \left[h^{k\ell}(1) h^{k\ell}(2) \dots h^{k\ell}(M) \right]$, where the individual channel coefficients are complex scalars. Note that for $\ell > 0$, $\boldsymbol{h}^{k\ell}$ refers to a channel between the BS of interest and a UE connected to a different BS. 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 [17]. Hence, in a time-slotted system, where a time slot has a length of t_s seconds, the channel coefficient in the nth time slot is

$$ h_n^{k\ell} = \frac{1}{\sqrt{N_s}} \sum_{m=1}^{N_s} \exp\left(j2\pi f_d n t_s \cos\alpha_m + \phi_m \right), \qquad (1) $$

where N_s is the number of fixed scatterers associated with all BS/UE pairs, f_d is the maximum Doppler shift, α_m and ϕ_m are the angle of arrival and initial phase, respectively, of the wave from the mth scatterer. Both α_m and ϕ_m are independent and uniformly distributed in the interval $[-\pi, \pi)$, which results from random scatterer locations. Furthermore, $f_d = \frac{v}{c} f_c$, where v is the speed of the UE, c is the speed of light and f_c is the carrier frequency.

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

Fig. 1 A cellular system with three cells. Cell 0 is of interest and the neighboring cells will potentially cause interference (red arrows)

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. We assume all transmissions in the system are synchronized, which represents a worst-case scenario from a pilot contamination perspective, as argued in [2]. During the nth pilot training period, the kth user in the ℓth cell transmits a pilot sequence $\boldsymbol{x}_n^{k\ell} = \left[x_n^{k\ell}(1)x_n^{k\ell}(2)\ldots x_n^{k\ell}(\tau)\right]^T$, where τ is the pilot sequence length. Ideally, all pilot sequences in the entire system are orthogonal, in order to avoid interference. However, this would require pilot sequences of at least length $L \times K$, 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, $\mathcal{C}_n^{k\ell}$, as the set of all pairs i, j, which identify all UEs applying the same pilot sequence in the nth time slot as the kth user in the ℓth cell. Hence, $\boldsymbol{x}_n^{ij} = \boldsymbol{x}_n^{k\ell} \forall i, j \in \mathcal{C}_n^{k\ell}$.

The pilot signal received by the BS of interest, concerning the kth user in the nth time slot can be expressed as

$$\boldsymbol{y}_n^{k0} = h_n^{k0}\boldsymbol{x}_n^{k0} + \sum_{i,j \in \mathcal{C}_n^{k0}} h_n^{ij}\boldsymbol{x}_n^{ij} + \boldsymbol{z}_n^{k0}, \qquad (2)$$

where $\boldsymbol{z}_n^{k0} = \left[z_n^{k0}(1)z_n^{k0}(2)\ldots z_n^{k0}(\tau)\right]^T$ and $z_n^{k0}(j)$ 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 $h_n^{ij}\boldsymbol{x}_n^{ij} \forall i, j \notin \mathcal{C}_n^{k\ell}$ are removed when correlating with

Fig. 2 Scheduling example
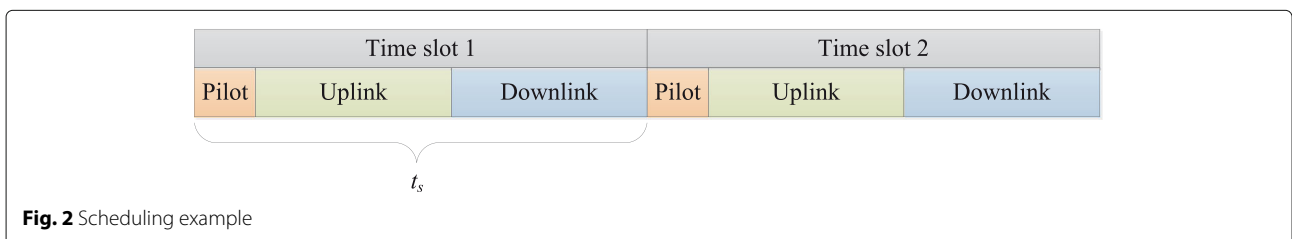
the applied pilot sequence. Hence, all contributions from the sum term are undesirable and will contaminate the CSI. Without loss of generality, we focus on the channel estimation for a single user in a single cell. Hence, in the remainder of the paper, we omit the superscript k for ease of notation.

2.2 Mobility model

In the employed mobility model, we restrict our attention to the consequences of mobility on the small-scale fading characteristics, i.e., f_d. Therefore, consequences like shadowing, varying path loss, and cell handover are disregarded. Since we employ pilot sequence hopping in this work, we can furthermore restrict our attention to the mobility of the UE of interest. This is explained in Section 3.1. We consider three different mobility models:

- *M1*: In this mobility model, the UE moves at a constant speed, v_1, for T_1 seconds.
- *M2*: (Train) This model emulates the mobility experienced in a train. Initially, the speed is zero for $T_{2,1}$ seconds. Then, the speed increases linearly, i.e., with constant acceleration, $\delta_{2,+}$, until a specified maximum speed, v_2. This speed is maintained for $T_{2,2}$ seconds, after which the speed is decreased linearly, with deceleration, $\delta_{2,-}$, until mobility has seized. Finally, the speed is kept at zero for $T_{2,3}$ seconds.
- *M3*: (Car) The third mobility model emulates the behavior of a car for T_3 seconds. The model operates with a vector of possible speeds, $\boldsymbol{v} = [v_0 v_1 \ldots v_{\max}]$, where the individual speeds are uniformly spaced between zero and $v_{\max}$. The initial speed is $v_0 = 0$. In every time slot, the speed is increased with probability p_+ and decreased with probability p_- and remains constant with probability $1 - p_+ - p_-$. Acceleration and deceleration are constant at $\delta_{3,+}$ and $\delta_{3,-}$, respectively. Speed changes always occur to the nearest speed in $\boldsymbol{v}$, and both acceleration from $v_{\max}$ and deceleration from v_0 result in no change.

Examples of all three mobility models are plotted in Fig. 3.

3 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. Since the channel coefficients are time-varying, we are dealing with a tracking problem. For this purpose, a Kalman filter is attractive due to its excellent tracking capability and recursive structure, which provides good performance at low complexity. Since the contaminating signals have been decorrelated, the Kalman filter will suppress the impact of these signals, leading to pilot *decontamination*.

3.1 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. This ensures that the coordination overhead is limited to the initial connection phase, whereby it can be neglected. Random pilot sequence hopping is illustrated in Fig. 4 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 as t_c^k for the kth user, see Fig. 4. Note that in the case of a fixed pilot schedule, $t_c^k = 1$. The goal of pilot sequence hopping is to maximize t_c^k, either in an expected sense or maxmin 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^k. If pilot sequence hopping is performed at random and $\tau = K$, then t_c^k follows a geometric distribution, such that
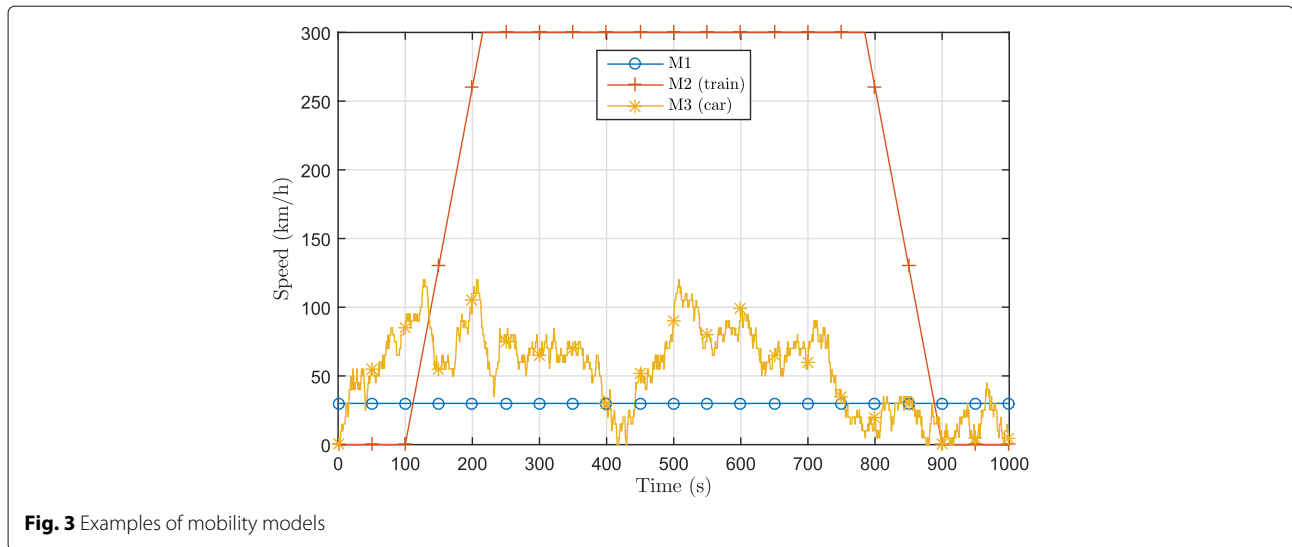
Fig. 3 Examples of mobility models

$$P\left(t_c^k = w\right) = (1-p)^{w-1}p, \qquad w = 1, 2, \ldots, \infty$$

$$p = \frac{1}{K}, \tag{3}$$

where $P\left(t_c^k = w\right)$ is the probability that the collision distance is w and p is the probability of a given UE being the next contaminator. We then have $\mathbb{E}\left[t_c^k\right] = K$, i.e., the expected collision distance increases with the number of users/pilots per cell, which follows intuition. Note that the collision distance is a user-specific measure, which holds for all potential contaminators in the system. Hence, the

analysis still holds when considering systems with more than one neighboring cell.

The maximization of $\mathbb{E}\left[t_c^k\right]$ leads to a decorrelation of the contaminating signals. The benefit of this is reaped using appropriate filtering techniques. For this purpose, we have chosen a modified version of the Kalman filter, which is described next.

3.2 Modified Kalman filter

The problem of estimating a time-varying channel based on pilot signals, also termed channel tracking, can be solved using the Kalman filter [18]. The Kalman filtering
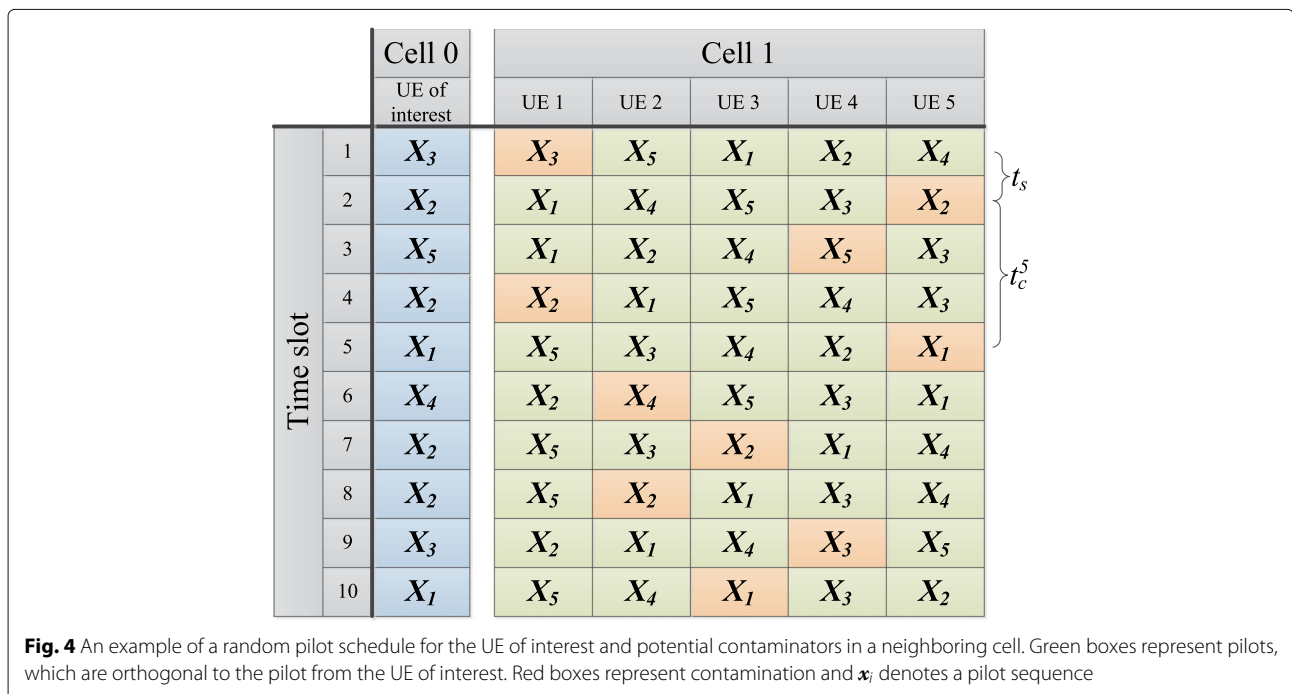

Fig. 4 An example of a random pilot schedule for the UE of interest and potential contaminators in a neighboring cell. Green boxes represent pilots, which are orthogonal to the pilot from the UE of interest. Red boxes represent contamination and $\boldsymbol{x}_i$ denotes a pilot sequence

framework consists of a *process equation* and a *measurement equation*. The process equation expresses how the variable under estimation develops over time. We already chose such a model in (1); however, Clarke's model does not fit the structure of the Kalman filter. As a result, we choose an autoregressive (AR) model as an approximation of the model in (1). An AR model is said to have order $d + 1$ if it expresses the current and d previous values as a function of the $d + 1$ previous values. As d increases, the approximation of Clarke's model is increasingly valid. In our context, if we define $\boldsymbol{h}_n = \left[h_n \ldots h_{n-d} \right]^T$, the process equation for the Kalman filter is expressed as

$$\boldsymbol{h}_n = \boldsymbol{A}_n \boldsymbol{h}_{n-1} + \boldsymbol{v}_n^p, \tag{4}$$

$$\boldsymbol{A}_n = \begin{bmatrix} a_n^1 \ \cdots \ a_n^{d+1} \\ \boldsymbol{I}_d \quad \boldsymbol{0}_{d\times 1} \end{bmatrix}, \tag{5}$$

where $\boldsymbol{I}_d$ is the $d \times d$ identity matrix, $\boldsymbol{0}_{d\times 1}$ is a $d \times 1$ vector of zeros, and $\boldsymbol{v}_n^p = \left[v_n^p(1) \ldots v_n^p(d+1) \right]^T$ is the *process noise*, which is zero mean circularly symmetric Gaussian with covariance matrix $Q_n \boldsymbol{I}_{d+1}$, where

$$Q_n = \gamma_n^0 - \sum_{j=1}^{d+1} a_n^j \gamma_n^j. \tag{6}$$

Here, γ_n^j is the autocovariance of the channel coefficient at a lag of j time slots, and γ_n^0, i.e., channel power, is assumed known. Since $\gamma_n^j = \gamma_n^{-j}$, we can find γ_n^j for $j > 0$ by solving the Yule-Walker equations

$$\begin{bmatrix} \gamma_n^1 \\ \gamma_n^2 \\ \vdots \\ \gamma_n^{d+1} \end{bmatrix} = \begin{bmatrix} \gamma_n^0 & \gamma_n^{-1} & \cdots & \gamma_n^{-d} \\ \gamma_n^1 & \gamma_n^0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_n^{-1} \\ \gamma_n^d & \cdots & \gamma_n^1 & \gamma_n^0 \end{bmatrix} \times \begin{bmatrix} a_n^1 \\ a_n^2 \\ \vdots \\ a_n^{d+1} \end{bmatrix}. \tag{7}$$

The corresponding measurement equation for the Kalman filter is expressed based on (2) as follows:

$$\boldsymbol{y}_n = \boldsymbol{X}_n \boldsymbol{h}_n + \boldsymbol{v}_n^m, \tag{8}$$

$$\boldsymbol{X}_n = \begin{bmatrix} x_n(1) & 0 & \ldots & 0 \\ \vdots & \vdots & \tau \times d & \vdots \\ x_n(\tau) & 0 & \ldots & 0 \end{bmatrix}, \tag{9}$$

where $\boldsymbol{v}_n^m$ is the *measurement noise*, which is zero mean circularly symmetric Gaussian with covariance matrix $\sigma_o^2 \boldsymbol{I}_\tau + \sigma_c^2 \boldsymbol{X}_n \boldsymbol{X}_n^H$. Here, σ_o^2 and σ_c^2 are noise power and total contamination power (average over time), respectively, which are both assumed known.

In a conventional Kalman filter, $\boldsymbol{A}_n$ is assumed constant and known. However, this cannot be assumed in our case; thus, the varying elements, $a_n^j, j = 1, \ldots, d + 1$, must be tracked along with the channel coefficients. For this purpose, we must modify the conventional Kalman filter to

include an AR model tracker. First, we state the conventional Kalman filter [18] in our context, where the AR coefficients are assumed known.

For all n :

$$\boldsymbol{e}_n = \boldsymbol{y}_n - \boldsymbol{X}_n \boldsymbol{A}_{n-1} \hat{\boldsymbol{h}}_{n-1}, \tag{10}$$

$$\boldsymbol{R}_n = \boldsymbol{X}_n \boldsymbol{P}_n \boldsymbol{X}_n^H + \sigma_o^2 \boldsymbol{I}_\tau + \sigma_c^2 \boldsymbol{X}_n \boldsymbol{X}_n^H, \tag{11}$$

$$\boldsymbol{K}_n = \boldsymbol{P}_n \boldsymbol{X}_n^H \boldsymbol{R}_n^{-1}, \tag{12}$$

$$\hat{\boldsymbol{h}}_n = \boldsymbol{A}_n \hat{\boldsymbol{h}}_{n-1} + \boldsymbol{K}_n \boldsymbol{e}_n, \tag{13}$$

$$\boldsymbol{F}_n = \boldsymbol{I}_{d+1} - \boldsymbol{K}_n \boldsymbol{X}_n, \tag{14}$$

$$\boldsymbol{P}_{n+1} = \boldsymbol{A}_n \boldsymbol{F}_n \boldsymbol{P}_n \boldsymbol{A}_n^H + Q \boldsymbol{I}_{d+1}, \tag{15}$$

where $\boldsymbol{I}_\tau$ is the $\tau \times \tau$ identity matrix and $\hat{\boldsymbol{h}}_n$ is the estimate of $\boldsymbol{h}_n$.

For the tracking of the AR coefficients, an approach similar to the one in [19] is taken. In [19], the inclusion of a first-order 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 $\boldsymbol{y}_{n-1}$. In this work, we extend this approach to higher order AR models taking all observations until $\boldsymbol{y}_n$ into account.

The approach is based on the principle of gradient descent. The gradient, ∇_n, with respect to $\boldsymbol{A}_n$ of the cost function, the mean squared error (MSE), is derived and used to adjust $\boldsymbol{A}_n$ in the direction of decreasing MSE. Note that this iterative numerical method is attractive since an analytical minimization of the cost function is complex to perform at every iteration. Furthermore, the minimization may only be feasible under certain conditions. Gradient descent is therefore a more robust and computationally simple solution. The gradient of the MSE is

$$\begin{aligned} \nabla_n(i,j) &= \frac{\partial}{\partial A_{n-1}(i,j)} \mathbb{E}\left[|\boldsymbol{e}_n|^2 \right] \\ &= -\Re\left[\left(\boldsymbol{q}_{n-1,i,j}^H \boldsymbol{A}_{n-1}^H \boldsymbol{X}_n^H + \hat{\boldsymbol{h}}_{n-1}^H \boldsymbol{\Gamma}_{i,j}^H \boldsymbol{X}_n^H \right) \boldsymbol{e}_n \right. \\ &\quad + \left. \boldsymbol{e}_n^H \left(\boldsymbol{X}_n \boldsymbol{A}_{n-1} \boldsymbol{q}_{n-1,i,j} + \boldsymbol{X}_n \boldsymbol{\Gamma}_{i,j} \hat{\boldsymbol{h}}_{n-1} \right) \right], \end{aligned}$$

$$\boldsymbol{\Gamma}_{i,j} = \frac{\partial \boldsymbol{A}_n}{\partial A_n(i,j)},$$

$$\boldsymbol{\Gamma}_{i,j}(k,\ell) = \begin{cases} 1, & \text{if } i = k \text{ and } j = \ell, \\ 0, & \text{elsewhere.} \end{cases}$$

$$\tag{16}$$

Here, $\boldsymbol{q}_{n,i,j} = \frac{\partial \hat{\boldsymbol{h}}_n}{\partial A_n(i,j)}$ and is found by differentiating (13) with respect to $A_n(i,j)$, such that

$$\boldsymbol{q}_{n,i,j} = \boldsymbol{F}_n \left(\boldsymbol{A}_n \boldsymbol{q}_{n-1,i,j} + \boldsymbol{\Gamma}_{i,j} \hat{\boldsymbol{h}}_{n-1} \right) + \boldsymbol{M}_{n,i,j} \boldsymbol{e}_n. \tag{17}$$

where $\boldsymbol{M}_{n,i,j} = \frac{\partial \boldsymbol{K}_n}{\partial A_n(i,j)}$, which is found by differentiating (12) with respect to $A_n(i,j)$; hence,

$$\boldsymbol{M}_{n,i,j} = \boldsymbol{F}_n \boldsymbol{S}_{n,i,j} \boldsymbol{X}_n^H \boldsymbol{R}_n^{-1}. \tag{18}$$

We introduced $S_{n,i,j} = \frac{\partial P_n}{\partial A_n(i,j)}$, which is a differentiation of (15) with respect to $A_n(i,j)$, giving us

$$S_{n+1,i,j} = \Gamma_{i,j} F_n P_n A_n^H + A_n F_n S_{n,i,j} F_n^H A_n^H$$
$$+ A_n F_n P_n \Gamma_{i,j}^H + U_n I_{d+1} \tag{19}$$

Finally, we have $U_n = \frac{\partial Q_n}{\partial A_n(i,j)}$, which can be found analytically by solving for Q_n in (6) and (7) and differentiating with respect to $A_n(i,j)$. Using ∇_n, we can then adjust A_n as follows:

$$A_n = A_{n-1} - \mu[\nabla_n]_{-\nu}^{+\nu}, \tag{20}$$

where μ is a parameter adjusting the convergence speed and the brackets denote truncations. The truncation to ν is for avoiding dramatic adjustments in situations with a high slope. The need for this will be explained in Section 5. In addition to the truncation, we enforce $|z_j| < 0.999$, where z_j are the roots of the polynomial $z^{d+1} - \sum_{j=1}^{d+1} d_n^j z^{d+1-j}$. This ensures a stationary AR process.

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

For all n :

$$e_n = y_n - X_n A_{n-1} \hat{h}_{n-1},$$
$$R_n = X_n P_n X_n^H + \sigma_o^2 I_\tau + \sigma_c^2 X_n X_n^H,$$
$$\nabla_n(i,j) = -\Re \Big[\Big(q_{n-1,i,j}^H A_{n-1}^H X_n^H + \hat{h}_{n-1}^H \Gamma_{i,j}^H X_n^H \Big) e_n$$
$$+ e_n^H \Big(X_n A_{n-1} q_{n-1,i,j} + X_n \Gamma_{i,j} \hat{h}_{n-1} \Big) \Big],$$
$$A_n = A_{n-1} - \mu[\nabla_n]_{-\nu}^{+\nu},$$
$$K_n = P_n X_n^H R_n^{-1},$$
$$\hat{h}_n = A_n \hat{h}_{n-1} + K_n e_n,$$
$$M_{n,i,j} = F_n S_{n,i,j} X_n^H R_n^{-1},$$
$$q_{n,i,j} = F_n \Big(A_n q_{n-1,i,j} + \Gamma_{i,j} \hat{h}_{n-1} \Big) + M_{n,i,j} e_n,$$
$$P_{n+1} = A_n F_n P_n A_n^H + Q_n I_{d+1},$$
$$S_{n+1,i,j} = \Gamma_{i,j} F_n P_n A_n^H + A_n F_n S_{n,i,j} F_n^H A_n^H$$
$$+ A_n F_n P_n \Gamma_{i,j}^H + U_n I_{d+1}. \tag{21}$$

In the following subsection, we derive the lower bound on the MSE of an estimate of the channel coefficients. It serves as a benchmark in the numerical evaluations in Section 5.

4 Analysis

Initially, we present a simplified analysis of a toy example, in order 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 since attention is on decontamination. Moreover, for this toy example only, we assume an infinite amount of orthogonal pilot sequences and an infinite amount of users per cell, such that $\tau = K = \infty$ and $\mathbb{E}[t_c^k] = \infty$, 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

$$\hat{h}_n = h + h_n', \tag{22}$$

where h_n' is the channel of the contaminator in time slot n. We define the MSE of this estimate as $\mathbb{E}\left[\left(h - \hat{h}_n\right)^2\right]$.

Now, consider a new estimator, $\bar{\hat{h}}_n$, which is the average of all estimates until time slot n. Hence, we have

$$\bar{\hat{h}}_n = h + \frac{1}{n} \sum_{i=1}^{n} h_i'. \tag{23}$$

In this case, the error in the estimate is solely composed of the average of the contaminating signals, which are independent and have variance σ_c^2. Hence, $\mathbb{E}\left[\left(h - \bar{\hat{h}}_n\right)^2\right] = \frac{\sigma_c^2}{n+\sigma_c^2}$, if prior knowledge on h is a standard Gaussian. If pilot sequence hopping had not been performed, the MSE had remained $\frac{\sigma_c^2}{1+\sigma_c^2}$ 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 τ (and K) and thereby finite $\mathbb{E}[t_c^k]$, the MSE is lower bounded by $\frac{\sigma_c^2}{K+\sigma_c^2}$ 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. We can account for this in a more elaborate Bayesian analysis, which is described next.

4.1 Bayesian analysis

Given a set of observations, $Y_n = [y_1, \dots, y_n]$, we are interested in deriving the distribution, in particular the variance, of the resulting estimate, $\hat{h}_n$, of the most recent channel coefficient, h_n. Here, $y_k = x_k h_k + v_k^m$, which through least squares estimation gives us a scalar observation $y_k^{ls} = (x_k^H x_k)^{-1} x_k^H y_k = h_k + v_k^m$, where v_k^m is the residual scalar noise, which is zero mean Gaussian with variance $\frac{\sigma_o^2}{\tau} + \sigma_c^2$. We define $y_n^{ls} = \left[y_1^{ls}, \dots, y_n^{ls}\right]$ and can then express the conditional probability density function of the vector of channel coefficients $h_n = [h_1, \dots, h_n]^T$ using Bayes' theorem as

$$f\left(h_n | y_n^{ls}\right) = \frac{f\left(y_n^{ls} | h_n\right) f(h_n)}{f\left(y_n^{ls}\right)}, \tag{24}$$

where $f(\boldsymbol{h}_n)$ is a multivariate Gaussian with mean vector $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\Sigma_{i,j} = \gamma_n^{i-j} = B_0\left(2\pi f_d t_s (i-j)\right)$ and B_0 is the zeroth-order Bessel function of the first kind. Furthermore, $f\left(\boldsymbol{y}_n^{ls}|\boldsymbol{h}_n\right)$ is a multivariate Gaussian with mean vector $\boldsymbol{h}_n$ and covariance matrix $\boldsymbol{C} = \left(\frac{\sigma_o^2}{\tau} + \sigma_c^2\right)\boldsymbol{I}_n$. It is well known that combining a Gaussian prior and a Gaussian likelihood provides a Gaussian posterior. This Gaussian posterior can be expressed as

$$f\left(\boldsymbol{h}_n|\boldsymbol{y}_n^{ls}\right) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{V}|}} e^{-\frac{(\boldsymbol{h}_n-\boldsymbol{\mu}_h)^{\mathrm{H}}\boldsymbol{V}^{-1}(\boldsymbol{h}_n-\boldsymbol{\mu}_h)}{2}}, \quad (25)$$

$$\boldsymbol{V}^{-1} = \boldsymbol{C}^{-1} + \boldsymbol{\Sigma}^{-1}, \quad (26)$$

$$\boldsymbol{\mu}_h = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_n^{ls}. \quad (27)$$

Equations (26) and (27) provide the optimal coefficients of a Bayesian filter and the corresponding covariance. The lower right corner element of $\boldsymbol{V}$ is then the variance of a causal filter estimating the most recent channel coefficient, h_n.

With Eqs. (25) to (27) as a starting point, we can analyze filters based on different assumptions on the underlying model of the channel.

The proposed Kalman filter with a first-order AR model is based on a Markovian assumption, where a channel coefficient only depends on the previous channel coefficient, see Fig. 5. Under this assumption, the posterior of h_n simplifies to the following recursive expression:

$$f\left(h_n|\boldsymbol{y}_n^{ls}\right) = \frac{f(y_n^{ls}|h_n)f\left(h_n|\boldsymbol{y}_{n-1}^{ls}\right)}{f\left(y_n^{ls}|\boldsymbol{y}_{n-1}^{ls}\right)}, \quad (28)$$

$$f\left(h_n|\boldsymbol{y}_{n-1}^{ls}\right) = \int_{h_{n-1}} f(h_n|h_{n-1})f\left(h_{n-1}|\boldsymbol{y}_{n-1}^{ls}\right) dh_{n-1}, \quad (29)$$

where $f\left(h_o|y_o^{ls}\right) = f(h_o)$ since no observation is made at time zero, and $f(h_o)$ is a standard Gaussian, such that the recursion is terminated. From the applied model, we have 0

$$f\left(y_n^{ls}|h_n\right) = \frac{1}{\sqrt{2\pi\left(\frac{\sigma_o^2}{\tau}+\sigma_c^2\right)}} e^{-\frac{\left(y_n^{ls}-h_n\right)^2}{2\left(\frac{\sigma_o^2}{\tau}+\sigma_c^2\right)}}, \quad (30)$$

$$f(h_n|h_{n-1}) = \frac{1}{\sqrt{2\pi(1-a^2)}} e^{-\frac{(h_n-ah_{n-1})^2}{2(1-a^2)}}, \quad (31)$$

$$f\left(h_{n-1}|\boldsymbol{y}_{n-1}^{ls}\right) = \frac{1}{\sqrt{2\pi\sigma_{h_{n-1}}^2}} e^{-\frac{\left(h_{n-1}-\mu_{h_{n-1}}\right)^2}{2\sigma_{h_{n-1}}^2}}, \quad (32)$$

where $a = B_0\left(2\pi f_d t_s\right)$. The integral in (29) can be viewed as the scaling factor in a product of the involved Gaussian distributions, such that

$$f\left(h_n|\boldsymbol{y}_{n-1}^{ls}\right) = \frac{1}{\sqrt{2\pi\left(a^2\sigma_{h_{n-1}}^2+1-a^2\right)}} e^{-\frac{\left(h_n-a\mu_{h_{n-1}}\right)^2}{2\left(a^2\sigma_{h_{n-1}}^2+1-a^2\right)}}, \quad (33)$$

where a change of variable has been performed in (32) from h_{n-1} to ah_{n-1} in order to conform to its representation in (31). We can then express the scaled product of Gaussian distributions in (28) as follows:

$$f\left(h_n|\boldsymbol{y}_n^{ls}\right) = \frac{1}{\sqrt{2\pi\sigma_{h_n}^2}} e^{-\frac{(h_n-\mu_{h_n})^2}{2\sigma_{h_n}^2}}, \quad (34)$$

$$\mu_{h_n} = \frac{\left(a^2\sigma_{h_{n-1}}^2+1-a^2\right)y_n^{ls} + \left(\frac{\sigma_o^2}{\tau}+\sigma_c^2\right)a\mu_{h_{n-1}}}{a^2\sigma_{h_{n-1}}^2+1-a^2+\frac{\sigma_o^2}{\tau}+\sigma_c^2}, \quad (35)$$

$$\sigma_{h_n}^2 = \frac{\left(a^2\sigma_{h_{n-1}}^2+1-a^2\right)\left(\frac{\sigma_o^2}{\tau}+\sigma_c^2\right)}{a^2\sigma_{h_{n-1}}^2+1-a^2+\frac{\sigma_o^2}{\tau}+\sigma_c^2}. \quad (36)$$

We are primarily interested in the evolution of the variance and thereby the MSE of the estimator. Figure 6 shows evaluations of (36) for different levels of mobility, $t_s = 0.5$ ms, $\sigma_c^2 = 0.6$, $\sigma_o^2 = 0.2$, and $\tau = 96$. It shows how the
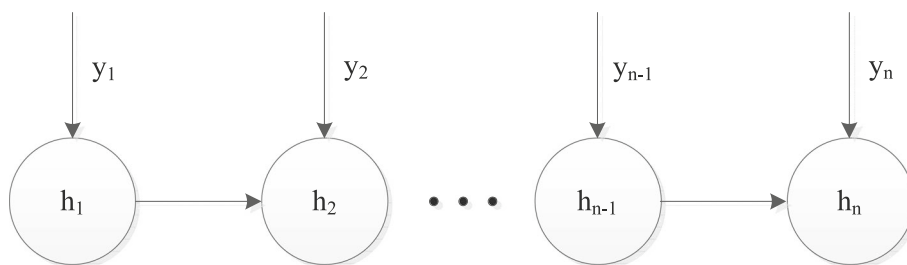

Fig. 5 Bayesian network representing the Markovian assumption applied in the Kalman filter
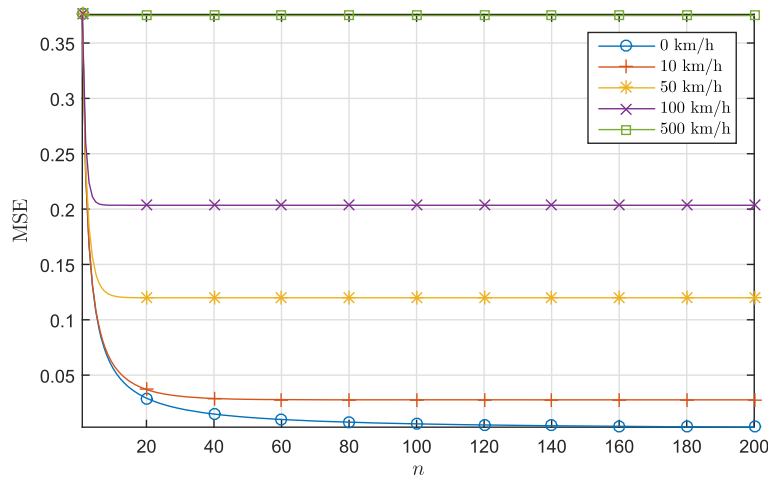
Fig. 6 MSE as a function of the number of pilot signals included in the filter. The results hold for the assumption of a first-order AR model and an optimal filter

MSE converges faster for higher mobilities since the information in past pilot signals decays faster. Note that these results act as a lower bound for the modified Kalman filter since we in (36) assume perfect knowledge of the AR coefficient, a. Any estimation error in the AR coefficient tracker will lead to increased MSE.

In the following subsection, we derive the lower bound on the MSE for arbitrary AR model order. Along with the bound for first-order AR models in (36), it serves as a benchmark in the numerical evaluations in Section 5.

4.2 Cramer-Rao lower bound

The Cramer-Rao lower bound (CRLB) [20] expresses the lower bound on the error covariance of any unbiased estimator based on a set of observations. In our context, the observations are $Y = [y_1, \ldots, y_n]$, and the estimation error covariance is $\mathbb{E}\left[\left(\hat{h} - h\right)\left(\hat{h} - h\right)^H\right]$, where $h = [h_1, \ldots, h_n]^T$. We follow a Bayesian framework, well suited to the tracking of time-varying parameters; thus, we employ the Bayesian CRLB. Having chosen Clarke's model as the channel model, it follows that the a priori distribution of the parameter h is well approximated as a Gaussian distribution. Furthermore, we adapt a compact formulation for the case of complex parameters (see [20, p. 529]). The complex Bayesian CRLB is expressed as

$$\mathbb{E}\left[\left(\hat{h} - h\right)\left(\hat{h} - h\right)^H\right] \geq J^{-1}, \tag{37}$$

where J is the Fisher information matrix. The matrix inequality means that $\mathbb{E}\left[\left(\hat{h} - h\right)\left(\hat{h} - h\right)^H\right] - J^{-1}$ is positive semidefinite. The Fisher information matrix is given by

$$J_{ij} = \mathbb{E}\left[-\frac{\partial^2 \log f_{Y,h}(Y,h)}{\partial h_i \partial h_j}\right], \qquad i,j = 1, \ldots, n. \tag{38}$$

Here, $f_{Y,h}(Y,h)$ is the joint probability density function of the observations and the channel coefficients. This can be expressed as

$$f_{Y,h}(Y,h) = f_{Y|h}(Y|h) f_h(h), \tag{39}$$

$$f_h(h) = c_1 \exp\left(-h^H \Sigma^{-1} h\right), \tag{40}$$

$$f_{Y|h}(Y|h) = \prod_{i=1}^{n} c_2 \exp\left(-\left(y_i - h_i x_i\right)^H C^{-1}\left(y_i - h_i x_i\right)\right), \tag{41}$$

where c_1 and c_2 are constants with independence from Y and h, Σ^{-1} is the inverse of the $n \times n$ covariance matrix of h, and C^{-1} is the inverse of the $\tau \times \tau$ observation error

Table 1 Simulation parameters

| Parameter	Value	Description
σ_o^2	0.2	Noise variance
L	7	Number of cells
K	96	Users per cell
τ	96	Pilot length
μ	10^{-5}	Convergence speed
ν	100	Derivative cap
f_c	1.8 GHz	Carrier frequency
N_s	20	Number of scatterers
t_s	0.5 ms	Time between pilots
$\hat{h}_0$	0	Initial estimate
q_0	0	Initial differentiated estimate
P_1	0	Initial error covariance
S_1	0	Initial differentiated error covariance
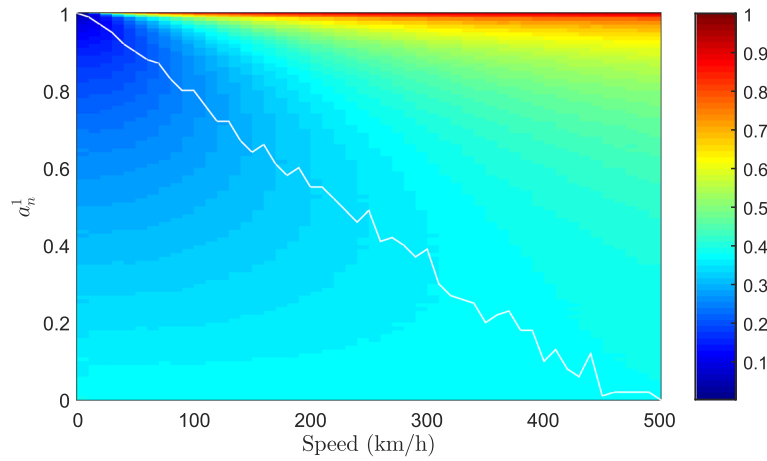
Fig. 7 MSE as a function of the autoregressive model coefficient and the speed of the UE. The coefficient with minimum MSE is marked with a white curve

covariance matrix. The CRLB at time n is the corresponding submatrix of J^{-1}; it gives a lower bound on channel estimation at time n accounting for the past observations.

By introducing the $n \times n$ matrix $\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}_{i,j} = \mathbb{E}\left[-\frac{\partial^2 \log f_{Y|h}(Y|h)}{\partial h_i \partial h_j}\right]$ and combining (38) through (41), we get

$$
\begin{aligned}
J &= \boldsymbol{\Omega} + \boldsymbol{\Sigma}^{-1}, \\
\boldsymbol{\Omega}_{i,j} &= \begin{cases} \boldsymbol{x}_i^H \boldsymbol{C}^{-1} \boldsymbol{x}_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \\
\boldsymbol{C} &= \sigma_o^2 \boldsymbol{I}_\tau + \sigma_c^2 \boldsymbol{x}_n \boldsymbol{x}_n^H, \\
\boldsymbol{\Sigma}_{i,j} &= B_0\left(2\pi f_d t_s (i-j)\right),
\end{aligned}
\tag{42}
$$

Working with the inverse of $\boldsymbol{\Sigma}$ may cause numerical problems at low mobility, where $\boldsymbol{\Sigma}$ is near-singular. This can be avoided by utilizing the matrix inversion lemma. We thus have

$$
\begin{aligned}
J^{-1} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\, (\boldsymbol{I}_n \otimes \boldsymbol{x}_i)^H &\left(\left(\boldsymbol{I}_n \otimes \boldsymbol{C}^{-1}\right)^{-1}\right. \\
&\left.+ (\boldsymbol{I}_n \otimes \boldsymbol{x}_i)\, \boldsymbol{\Sigma}\, (\boldsymbol{I}_n \otimes \boldsymbol{x}_i)^H\right)^{-1} (\boldsymbol{I}_n \otimes \boldsymbol{x}_i)\, \boldsymbol{\Sigma},
\end{aligned}
\tag{43}
$$

where $\otimes$ denotes the Kronecker product. Furthermore, expression (43) allows a continuity with the case of no mobility for which the channel estimate of time n is the result of an average (see Eq. (23)).
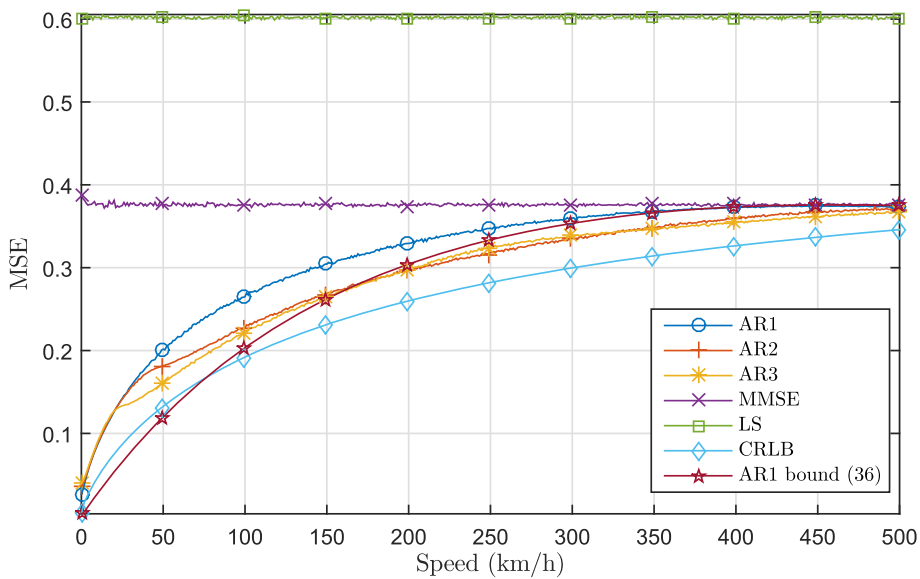

Fig. 8 Comparison between the proposed scheme and conventional solutions with respect to mean squared error as a function of mobility. The M1 mobility model is applied, and SIR is 2.2 dB
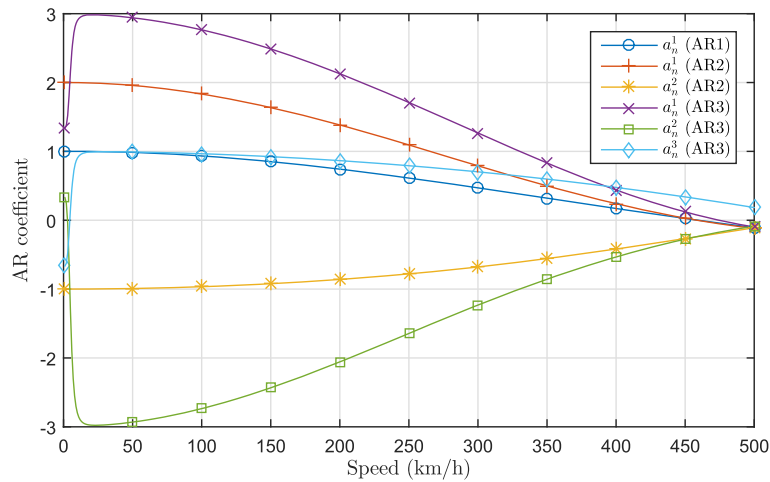
Fig. 9 AR model coefficients, found by solving (7), as a function of the speed of the UE

5 Numerical results

The proposed scheme has been simulated and compared to 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 (44) and (45), respectively. An overview of the parameters, which are common for all simulations, is given in Table 1. The choice of μ is based on experiments showing that this is a good compromise between convergence speed and limitation of noise-induced variance in the estimate. Throughout all simulations, we assume that contaminating signals have zero autocorrelation between time slots, which is justified by the choice of $K = 96$, such that $\mathbb{E}\left[t_c^k\right] = 96$. All simulation results are averages of 100 iterations of a scenario as specified by the mobility models in Section 2.2.

$$\hat{h}_n^{ls} = \left(x_n^H x_n\right)^{-1} x_n^H y_n, \tag{44}$$

$$\hat{h}_n^{mmse} = x_n^H \left(x_n x_n^H + \sigma_n^2 I_\tau + \sigma_c^2 x_n x_n^H\right)^{-1} y_n. \tag{45}$$

Initially, results are shown for the conventional Kalman filter expressed in Eqs. (10) through (15), using a first-order AR model as the process equation. MSE, defined as $\mathbb{E}\left[\left(h_n - \hat{h}_n\right)^2\right]$, as a function of the user mobility, v, and the AR coefficient, a_n^1, is shown in Fig. 7. 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 3.2. Moreover, it is seen that the derivative of the MSE with respect to a_n^1 may
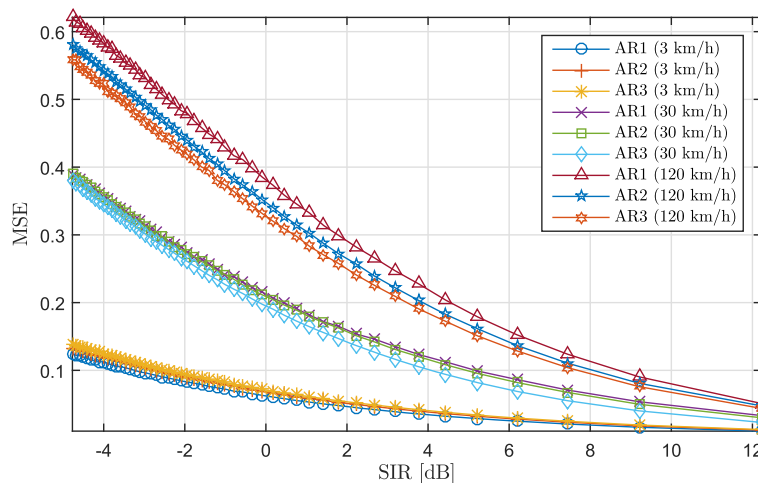

Fig. 10 Comparison between AR models with different orders with respect to mean squared error as a function of the signal-to-interference ratio
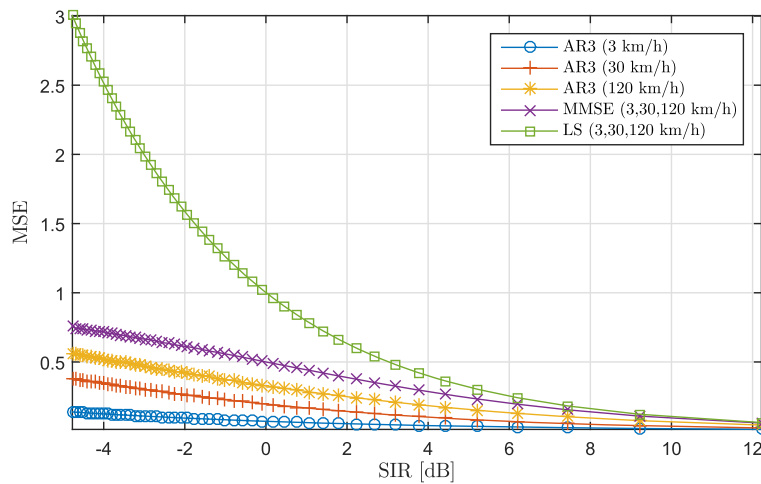
Fig. 11 Comparison between the proposed scheme and conventional solutions with respect to mean squared error as a function of the signal-to-interference ratio

attain very high values at high a_n^1. This can cause undesirably high variance in the estimate of the optimal a_n^1, which motivates the use of a derivative cap, ν.

Figure 8 shows a comparison of the proposed estimator, the LS estimator, and the MMSE estimator, with respect to MSE as a function of the speed of the UE using mobility model M1. The simulations apply $T_1 = 100$ s and $\sigma_c^2 = 0.6$. For the proposed estimator, AR models of first, second, and third order are included. In all three cases, the initial AR model coefficients are numerically optimized through a parameter sweep. The results were $a_0^1 = 0.3$ for the first-order model, $a_0^1 = 0.8$ and $a_0^2 = 0.2$ for the second-order model, and $a_0^1 = 0.7$, $a_0^2 = -0.3$, and $a_0^3 = -0.2$ for the third-order model. The results show that a significant performance improvement is achieved

at medium and high levels of mobility when increasing the AR model order from one to two. Further increasing to a third-order model yields a more mixed result. At medium mobility, a significant gain is achieved, whereas at higher mobility, the performances of the second- and third-order models are quite close and take turns in being the better. At low mobility, no gain is achieved when increasing the order of the AR model. At first glance, this may be surprising, but when looking at the coefficients found from solving the Yule-Walker equations, see Fig. 9, it is evident that low mobility presents a particularly challenging tracking task for a third-order AR model, due to large dynamics. Compared to the conventional methods of LS and MMSE, the proposed scheme offers a performance in an entirely different league. The
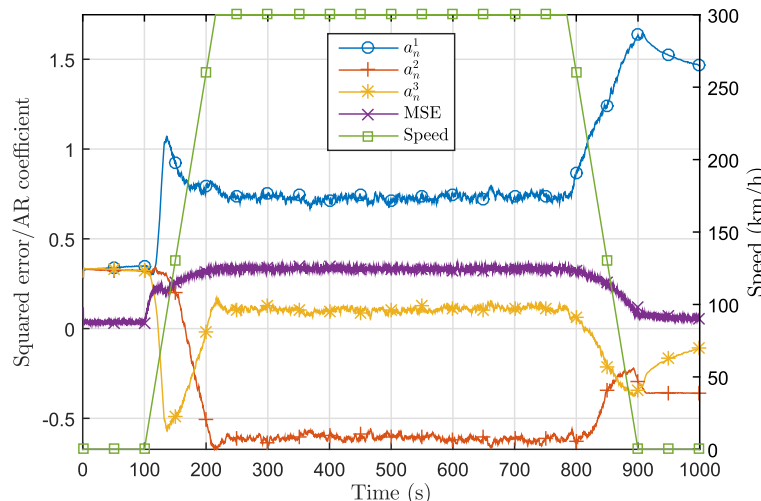

Fig. 12 Example of a simulation with a third-order AR model and the M2 mobility model. SIR is 2.2 dB
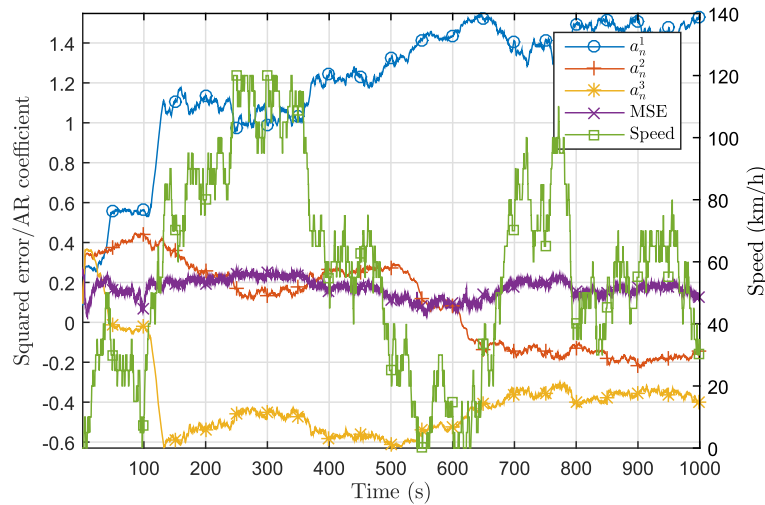
Fig. 13 Example of a simulation with a third-order AR model and the M3 mobility model. SIR is 2.2 dB

gain decreases as speed increases, but only at unusually high speeds is the gain insignificant. There is still a gap to the Cramer-Rao lower bound, although much of it has been closed when comparing to the conventional methods. When comparing to the bound associated with a first-order AR model, as expressed in (36), it is seen that the lack of performance gain at high mobility is largely explained by the choice of process model for the Kalman filter.

A different perspective is given in Fig. 10. Here, the MSE is plotted as a function of the signal-to-interference ratio (SIR), at typical mobility levels as defined by 3GPP [21]. Again, it is seen how increasing the order of the AR model is an advantage at medium and high mobility, but not at lower speeds. Figure 10 shows that this holds in a wide range of SIR.

From the same perspective, Fig. 11 shows a comparison of the conventional methods and the proposed scheme with a third-order model. Here, it is seen that the performance improvement is achieved in the entire SIR range. Decreasing the SIR is particularly penalizing the LS method, while MMSE and the proposed method are better able to cope with the increased interference.

Next, we look at the more sophisticated mobility models, M2 and M3. For M2, we use parameters $T_{2,1} = T_{2,3} = 100$ s, $T_{2,2} = 800$ s, $\delta_{2,+} = 2.6$ m/s^2, $\delta_{2,-} = -2.6$ m/s^2, and $v_2 = 300$ km/h. An example of a simulation with a single sequence of channel realizations is shown in Fig. 12. For M3, we use parameters $T_3 = 1000$ s, $v = [05\ldots 120]$ km/h, $p_+ = p_- = 0.00025$, $\delta_{3,+} = 2.5$ m/s^2, and $\delta_{3,-} = -4$ m/s^2. Figure 13 shows a simulation example with this mobility model. It is seen that with
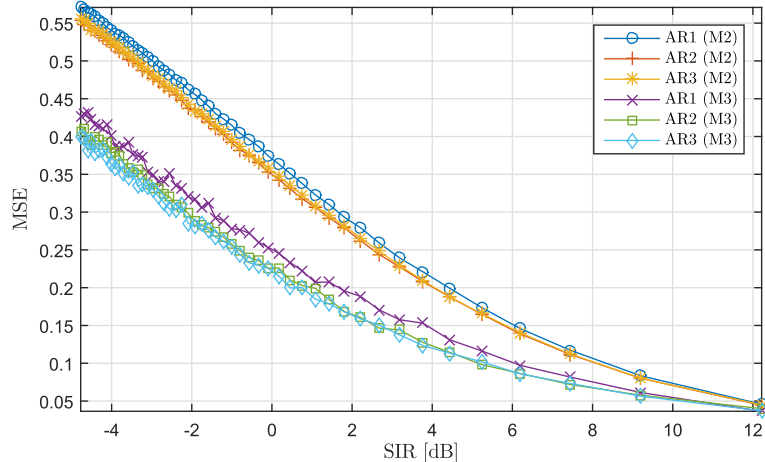

Fig. 14 Comparison of different AR model orders when using M2 and M3 mobility models
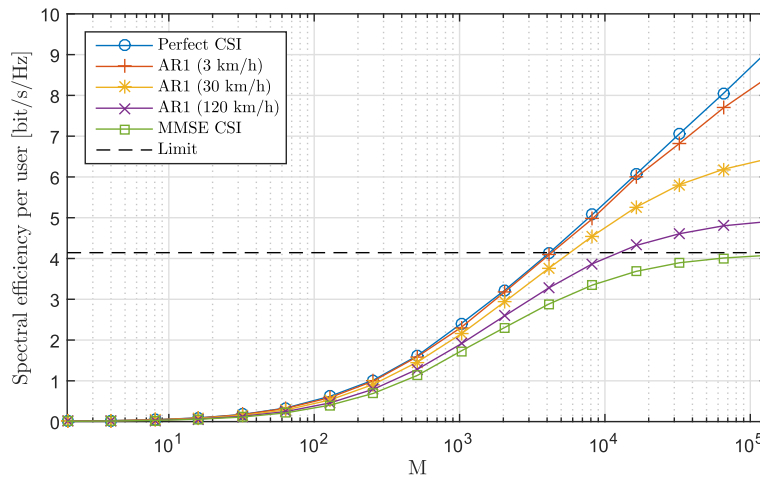
Fig. 15 The achievable downlink rate as a function of the number of antennas, *M*, at the BS

both mobility models, the AR parameter tracker is able to adjust to the varying speed and thereby adapt to the varying amount of information available in past pilot signals. In general, we expect to see a low MSE when the speed is low and vice versa. The figures confirm that this is in fact achieved. Figure 14 shows a comparison of the different AR model orders. It shows that increasing the order also provides performance enhancements in a wide range of SIR when considering M2 and M3.

Finally, we evaluate how the improvements of the channel estimates translate into increased achievable uplink and dowlink rates in a TDD massive MIMO system with *M* antennas at the BS. The achievable rates of such a system were derived in [22, p. 4]. We apply the proposed scheme for each individual channel coefficient in the system and evaluate the resulting achievable uplink and downlink rates. Furthermore, we evaluate the achievable rates in a system with perfect CSI at the BS and a system with CSI achieved with a conventional MMSE estimator at the BS. These act as upper and lower bounds, respectively, to the proposed scheme. The results are shown in Figs. 15 (downlink) and 16 (uplink). It is evident that the proposed scheme provides a significant improvement in achievable rates compared to the system with MMSE estimation. At low mobility, it even comes fairly close to a system with perfect CSI. Although the improvements are more visible at very high and impractical values of *M*, significant relative improvements are also achieved at lower values of *M*. As an example, the achievable uplink and downlink rates are increased by 42 and 46%, respectively, for $M = 128$, compared to MMSE estimation. Another important observation is that the system with MMSE
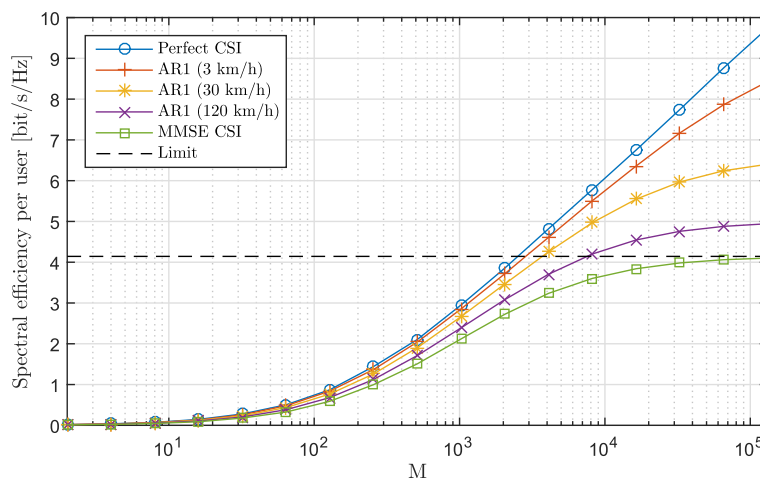

Fig. 16 The achievable uplink rate as a function of the number of antennas, *M*, at the BS

estimation converges to a limit as M goes to infinity. The limit is found as $\log_2\left(1 + \frac{1}{\sigma_c^2}\right)$ and is the well-known limitation caused by pilot contamination. As in the case of perfect CSI, the proposed scheme is able to break this limit, which demonstrates its ability to mitigate pilot contamination.

6 Conclusions

We have presented a solution to pilot contamination in channel estimation, which is a major challenge in TDD massive MIMO systems. It is based on a combination of a pilot sequence hopping scheme and a modified Kalman filter. The pilot sequence hopping scheme involves random shuffling of the assigned pilot sequences within a cell, which ensures decorrelation in the time dimension of the contaminating signals. This is essential since it enables subsequent filtering to suppress the contamination. For this filtering, the Kalman filter has been chosen, due to its ability to track a time-varying state. However, a conventional Kalman filter is not able to adapt to changes in the underlying model, which is necessary when users have unknown and varying levels of mobility. For this problem, we have presented a modified Kalman filter, which can adapt the underlying model based on a minimization of the mean squared error.

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.

Authors' contributions
Both authors have contributed to the design and analyis of the proposed methods as well as the writing of this manuscript. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. T Brown, E De Carvalho, P Kyritsi, *Practical Guide to the MIMO Radio Channel*. (John Wiley & Sons, Hoboken, 2012)
2. TL Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas. Wirel. Commun. IEEE Trans. **9**(11), 3590–3600 (2010). doi:10.1109/TWC.2010.092810.091092
3. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays. Signal Proc. Mag. IEEE. **30**(1), 40–60 (2013). doi:10.1109/MSP.2011.2178495
4. H Yin, D Gesbert, M Filippou, Y Liu, A coordinated approach to channel estimation in large-scale multiple-antenna systems. Sel. Commun. IEEE J. **31**(2), 264–273 (2013). doi:10.1109/JSAC.2013.130214
5. A Ashikhmin, T Marzetta, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium On*. Pilot contamination precoding in multi-cell large scale antenna systems, (2012), pp. 1137–1141. doi:10.1109/ISIT.2012.6283031
6. J Jose, A Ashikhmin, TL Marzetta, S Vishwanath, Pilot contamination and precoding in multi-cell TDD systems. Wirel. Commun. IEEE Trans. **10**(8), 2640–2651 (2011). doi:10.1109/TWC.2011.060711.101155
7. HQ Ngo, EG Larsson, in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference On*. EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays, (2012), pp. 3249–3252. doi:10.1109/ICASSP.2012.6288608
8. RR Muller, L Cottatellucci, M Vehkapera, Blind pilot decontamination. Sel. Topics Signal Proc. IEEE J. **8**(5), 773–786 (2014). doi:10.1109/JSTSP.2014.2310053
9. RR Mueller, M Vehkaperae, L Cottatellucci, in *Smart Antennas (WSA), 2013 17th International ITG Workshop On*. Blind pilot decontamination (VDE Verlag, Offenbach, 2013), pp. 1–6
10. RR Muller, M Vehkapera, L Cottatellucci, in *Signals, Systems and Computers, 2013 Asilomar Conference On*. Analysis of blind pilot decontamination, (2013), pp. 1016–1020. doi:10.1109/ACSSC.2013.6810444
11. L Cottatellucci, RR Müller, M Vehkaperä, in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*. Analysis of pilot decontamination based on power control, (2013), pp. 1–5. doi:10.1109/VTCSpring.2013.6691891
12. H Yin, L Cottatellucci, D Gesbert, RR Müller, G He, Robust pilot decontamination based on joint angle and power domain discrimination. IEEE Trans. Signal Process. **64**(11), 2990–3003 (2016). doi:10.1109/TSP.2016.2535204
13. J Choi, DJ Love, P Bidigare, Downlink training techniques for FDD massive mimo systems: open-loop and closed-loop training with memory. Sel. Topics Signal Proc. IEEE J. **8**(5), 802–814 (2014). doi:10.1109/JSTSP.2014.2313020
14. E Björnson, E de Carvalho, JH Sørensen, EG Larsson, P Popovski, A random access protocol for pilot allocation in crowded massive mimo systems. IEEE Trans. Wireless Commun. **16**(4), 2220–2234 (2017). doi:10.1109/TWC.2017.2660489
15. E de Carvalho, E Björnson, EG Larsson, P Popovski, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Random access for massive mimo systems with intra-cell pilot contamination, (2016), pp. 3361–3365. doi:10.1109/ICASSP.2016.7472300
16. JH Sørensen, E De Carvalho, in *IEEE Global Telecommunications Conference, GLOBECOM*. Pilot decontamination through pilot sequence hopping in massive mimo systems, (2014)
17. R Clarke, A statistical theory of mobile-radio reception. Bell Syst. Tech. J. **47**(6), 957–1000 (1968)
18. S Haykin, *Adaptive Filter Theory*. (Pearson Education, Limited, Harlow, 2013)
19. K-Y Han, S-W Lee, J-S Lim, K-M Sung, Channel estimation for OFDM with fast fading channels by modified Kalman filter. Consum. Electron. IEEE Trans. **50**(2), 443–449 (2004). doi:10.1109/TCE.2004.1309406
20. SM Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. (Prentice-Hall, Inc., Upper Saddle River, 1993)
21. 3GPP: Spatial channel model for multiple input multiple output (MIMO) simulations. TR 25.996, 3rd Generation Partnership Project (3GPP) (September 2012). http://www.3gpp.org/ftp/Specs/html-info/25996.htm. Accessed 13 Sept 2017
22. J Hoydis, S ten Brink, M Debbah, Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? Sel. Areas Commun. IEEE J. **31**(2), 160–171 (2013). doi:10.1109/JSAC.2013.130205