Signal Processing 186 (2021) 108132

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Coordinated versus uncoordinated channel tracking for high-rate internet of things in multiuser massive MIMO: Algorithms and performance☆

Hediyeh Soltanizadeh, Shahrokh Farahmand*, S. Mohammad Razavizadeh

School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran, Iran

ARTICLE INFO

Article history:
Received 18 August 2020
Revised 10 March 2021
Accepted 18 April 2021
Available online 21 April 2021

Keywords:
Channel tracking
Internet of things
Massive MIMO
Minimum mean square error
Kalman filter
Pilot contamination

ABSTRACT

Incorporation of high-rate internet of things (IoT) service into a massive MIMO framework is investigated. It is revealed that massive MIMO possess the inherent potential to offer such service provided it knows the channels for all devices. Our proposed method is to jointly estimate and track the channels of all devices irrespective of their current activity. Using the dynamical model for devices' channels evolution over time, optimal and sub-optimal trackers are developed for coordinated scenario. Furthermore, we introduce a new paradigm where the BS need not know the pilot access patterns of devices in advance which we refer to as uncoordinated setup. After motivating this scenario, we derive the optimal tracker which is intractable. Then, target tracking approaches are applied to address uncertainties in the measurements and derive sub-optimal trackers. Our proposed approaches explicitly address the channel aging problem and will not require downlink paging and uplink access request control channels which can become bottlenecks in crowded scenarios. The fundamental minimum mean square error (MMSE) gap between optimal coordinated and uncoordinated trackers which is defined as price of anarchy is evaluated and upper-bounded. Stability of optimal trackers is also investigated. Finally, performance of various proposed trackers are numerically compared.

1. Introduction

Mobile broadband (MBB) and massive machine-type communication (mMTC) constitute two main application areas in future generations of cellular networks. mMTC is also referred to as the crowded or overloaded scenario and is mainly driven by internet of things (IoT) applications. To accommodate these two type of services in future networks, promising cellular concepts such as massive MIMO should be adapted to their particular characteristics. Recent efforts on fusion of these concepts are divided into three categories: i) MBB, ii) low-rate IoT, iii) high-rate IoT.

Benefiting from channel hardening property, massive MIMO decouples individual users channels into deterministic ones free from interference, small-scale fading, and noise [20], while applying low-complexity linear beamforming/combining techniques [19,22]. As its major challenge, all users channels should be known at the BS. Orthogonal pilots have been utilized to enable channel estimation [20]. They were proven to be optimal for peer-to-peer ordinary MIMO as well [4]. Each orthogonal pilot should be assigned to one user only to avoid pilot contamination. This is possible for MBB as there are only a few active users. A well-known initial random access procedure for dynamic pilot assignment in MBB is random access to pilots (RAP) which is reminiscent of slotted ALOHA with the difference that collision domain is no longer time but pilot sequences [26]. RAP has been further modified to permit collision resolution [5,23].

The proposed grant-based approaches are not suitable for low-rate IoT. Firstly, crowded scenarios lead to excessive collisions in a slotted ALOHA system greatly reducing its success rate. Secondly, every device has only a few bits and the overall access request/grant procedure incurs too much overhead. Thirdly, there might not be even enough orthogonal pilots for all simultaneously active devices. The proposed remedy is to transmit a pilot pattern followed immediately by the few data bits without any previous handshake with the BS. These schemes are referred to as grant-free and are divided into two broad categories depending on the pilot structure. First category advocates non-orthogonal pilot sequences that are assigned uniquely and permanently to every de-

☆ Parts of this work has appeared in the International Symposium on Telecommunications (IST) 2020 [24].

* Corresponding author.
E-mail addresses: he_soltanizadeh@elec.iust.ac.ir (H. Soltanizadeh), shahrokhf@iust.ac.ir (S. Farahmand), smrazavi@iust.ac.ir (S.M. Razavizadeh).

https://doi.org/10.1016/j.sigpro.2021.108132
0165-1684/

vice. The corresponding decoders mostly rely on compressed sensing [3,8,10,11,28]. Second category recommends sharing orthogonal pilots among several users. Ergodic random access to pilots including data (E-RAPiD) and coded RAPiD (C-RAPiD) are two examples [6]. In addition to the original C-RAPiD [27] which borrowed ideas from collision resolution in slotted ALOHA [7], MBB methods such as [5] has also been modified to incorporate similar collision resolution procedures [13].

Proposed approaches for low-rate IoT pose certain limitations in high-rate IoT. Firstly, neither of the proposed methods can handle large data rates. When using non-orthogonal random access, either payload length following a pilot transmission should be greatly enlarged or many successive access slots should be used. Payload enlargement is limited by the channel coherence time, while many successive access attempts will destroy the sparseness of activity patterns, severely limiting the decoder performance. E-RAPiD will incur excessive large decoding delays as long data packets should be communicated over many coherence times. Thus, real-time applications such as voice, video, robotics surgery, and interactive IoT can not be supported. C-RAPiD will suffer from a lower delay, given that it uses one coherence time, but will have to deal with many active users in every coherence time. This translates to decoding coded messages with many loops and active nodes which will adversely affect message-passing techniques. Two other issues were not addressed by any of the above methods. In crowded scenarios, too many devices need to be paged at every coherence time which jams the downlink control channel. It would be desirable to remove this bottleneck. Secondly, all aforementioned methods assume a block fading model which overlooks channel aging. Finally and most importantly, it is always assumed in the literature that BS knows the pilot-hopping (PH) or pilot access patterns for devices whose channels should be estimated. We provide several motivating reasons as to why this assumption is too restrictive. This assumption is no longer needed in our proposed uncoordinated trackers.

To the best of our knowledge, the only two available channel tracking methods for massive MIMO are presented in [9] and [25]. To be specific, [9] does not transmit pilots and only data is transmitted. In fact, [9] iterates between turbo decoding and channel estimation using previously decoded data as pilots. The proposed approach is limited by the fact that it does not use dynamical models on the channel and does not benefit from optimal sequential estimation procedures such as Kalman filter (KF). Furthermore, error propagation occurs if turbo codes fail to correct all errors. On the other hand, [25] uses pilot-hopping in combination with a KF to track the channel parameters variations. It treats interfering devices as noise and can not benefit from the extra information that their channel estimates will provide. Furthermore, it will not perform satisfactorily if PH pattern and/or pilot transmission times are unknown.

Using a resolvable multipath channel model in the angular domain, an alternative approach to enable pilot sharing was offered in [29]. The angular spread of each device can be further tracked in time [12]. Two major limitations of this approach are: i) must have no local scatterers near BS, ii) does not exploit the resolution capability of small-scale fading, also known as fast fading, which is a consequence of lack of enough scatterers.

1.1. Main contributions

Massive MIMO offers an inherent potential to address all the aforementioned concerns for high-rate IoT. The challenge is to learn all the devices channels. By explicitly considering a dynamic model for channel evolution, we propose joint trackers that simultaneously follow all devices irrespective of their current activity status. To enable high-quality tracking, we decouple pilot transmission from data communication. Thus, in each access slot, one device may transmit pilot only to help improve its available channel estimate, may transmit data only when BS already maintains a good channel estimate for that user, may transmit both pilot and data, or remain completely silent. We also advocate orthogonal pilot sharing, similar to low-rate IoT, but use the different spatial features of various devices to resolve collisions. Our proposed spatial feature for every user is its best available channel estimate from previous access slots. Unlike large-scale path loss [5] or distance to BS [23] which can be the same for many devices in crowded scenarios, the whole channel vector will be different for various users with very high probability. Subsequently, we will propose several low-complexity sub-optimal trackers as well. Our proposed approaches will i) resolve pilot collisions and exploit their information ii) omit the need for separate downlink paging and uplink control channels, iii) explicitly account for channel aging, iv) offer minimum decoding delays compared to E-RAPiD/C-RAPiD, v) decouple pilot transmission and data communication patterns allowing for more flexibility. Our next major contribution, which has not been considered in prior art, is to address the problem of unknown pilot transmission patterns by various devices at the BS. In Section 4 we motivate this scenario and offer the corresponding optimal and sub-optimal trackers. Our chief contributions are:

1. When BS knows pilot transmission patterns, which we refer to as coordinated scenario, optimum joint tracker and two low-complexity sub-optimal trackers are introduced.
2. When BS does not know pilot transmission patterns of devices, which we refer to as uncoordinated scenario, optimal but practically intractable tracker is derived.
3. Borrowing ideas from target tracking community, low-complexity sub-optimal trackers for problem formulated in 2 are introduced. Given that several users are present in pilot collisions, the main target tracking algorithms' assumption that every measurement originates from at most one target is violated. Subsequently, we will modify all these methods to accommodate the pilot contamination model.
4. To further reduce complexity, two novel heuristic trackers based on soft and hard least-squares (LS) and maximum likelihood estimate (MLE) are also derived and their complexities are analyzed.
5. The performance gap between the two optimal trackers, which offers a fundamental performance limit and is referred to as price of anarchy, is evaluated and an upper bound for it is derived. Stability of optimal trackers is also investigated.
6. Performance of the proposed trackers is investigated through extensive simulations.

1.2. Organization

Section 2 formulates the problem. Section 3 presents coordinated channel trackers. Section 4 offers the optimal and two sub-optimal uncoordinated trackers. Section 5 presents the two heuristic low-complexity trackers for the uncoordinated setup. Section 6 provides performance analysis. Section 7 provides numerical results and Section 8 concludes the paper.

2. System model and problem formulation

Let us consider a single cell where a MIMO base station (BS) equipped with M antennas is serving a total of N single-antenna IoT devices. The system operates in a time-division duplex (TDD) mode, where each access slot is divided into 3 segments. In the first segment, a pilot of length τ symbols is transmitted by some of the IoT devices and BS exploits the received signal to track devices channel gains. In the second segment, uplink data are transmitted

followed by downlink data in the third segment. As $N \gg \tau$, there exist more devices than orthogonal pilots. Thus, BS associates the same pilot to a group of K devices where $K := N/\tau$ is an integer. To enable channel tracking, each device transmits the pilot irregularly once in a while according to a random pattern. Different devices may transmit pilots at different rates. In a single random access slot, some devices may transmit data only, some may transmit pilot only, and some transmit both pilot and data while the rest remain totally silent. The objective is to track all IoT devices channels at successive random access slots. Once this challenge is addressed, BS can utilize the obtained channel estimates to perform various beamforming methods at uplink/downlink. It should be noted that BS tracks all devices channels irrespective of whether they have data to transmit or are silent for now. Subsequently, downlink beamforming will be applied whenever downlink data for a particular device is available. On the other hand, uplink beamforming is exploited at every access slot to find those users which have data to transmit to the BS. As the channel tracking process for devices belonging to distinct groups can be decoupled by the orthogonality of their pilots, we focus on channel tracking for users within group one corresponding to pilot sequence one or $\boldsymbol{\phi}_1 \in \mathbf{R}^{\tau\times 1}$ without loss of generality.

We aim to track the flat fading channel gains between the kth device and the BS in the tth time slot, which is represented by $\mathbf{h}_t^{(k)} = [h_t^{(k)}(1), h_t^{(k)}(2)\cdots h_t^{(k)}(M)]^T$ for $k = 1\ldots K$. It is assumed that $\mathbf{h}_t^{(k)}$ includes large-scale path-loss, shadowing and small-scale fading. Furthermore, we assume power allocation is also absorbed into $\mathbf{h}_t^{(k)}$. We implicitly consider a MIMO-OFDM system and focus on one coherence bandwidth. Equivalently, we can assume a narrow band single-carrier system. Instead of adopting a typical block-fading model, we deal with channel aging directly and treat the channels as time-varying dynamical systems. Hence, user k channel gains change from access slot $t-1$ to access slot t according to the given state model:

$$\mathbf{h}_t^{(k)} = \mathbf{A}_t^{(k)}\mathbf{h}_{t-1}^{(k)} + \mathbf{u}_t^{(k)} \tag{1}$$

where $\mathbf{A}_t^{(k)}$s represent known model matrices that determines how fast or slow the channel changes over time slots and $\mathbf{u}_t^{(k)}$ denotes Gaussian process noise with zero mean and covariance matrix $\mathbf{Q}_t^{(k)}$. Process noise is independent across t, k. To simplify matters and without loss of generality, we consider a real system with real channel gains. This amounts to using a one-dimensional constellation such as binary pulse amplitude modulation (B-PAM) and discarding whatever is received at the quadrature channel. While this scheme is not optimum, it serves the purpose of delineating the merits of proposed algorithms. Extension to arbitrary complex constellations and using both in-phase and quadrature components is a worthwhile future research direction. As for initialization, we assume $\mathbf{h}_0^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$. The model in (1) is fairly general and very flexible. Different users channels evolve independent of one another. This is ensured by the independence of initial channels assignment and independence of process noise for various devices.

Each device at a particular access slot decides whether or not to send the pilot with the probability λ_k independently from other devices. Thus, only a subset of devices are active and send pilots within each access slot. We define a binary variable $q_t^{(k)}$ that takes one if the k'th device transmits its pilot at time slot t. Let D_t denote the subset of all active devices in time slot t who transmit $\boldsymbol{\phi}_1$. Then, the received signal at the BS at time slot t is given by $\mathbf{Y}_t \in \mathbf{R}^{M\times\tau}$:

$$\mathbf{Y}_t = \sum_{k=1}^{K} q_t^{(k)}\mathbf{h}_t^{(k)}\boldsymbol{\phi}_1^T + \mathbf{W}_t \tag{2}$$

where $\mathbf{W}_t \in \mathbb{R}^{M\times\tau}$ indicates the additive white Gaussian noise (AWGN) matrix whose entries are independent with zero mean and variance equal σ_w^2. Given $\mathbf{Y}_{1:t}$, our objective is to estimate $\mathbf{h}_t^{(1:K)}$ which amounts to the channels of all users who are allocated to group one. If $q_t^{(k)}$'s are known to the BS, the tracking scheme is coordinated. However, when $q_t^{(k)}$'s are not known at BS, we refer to the scheme as uncoordinated. Our performance criterion is mean square error (MSE). All of our proposed trackers assume an accurate initial acquisition of all devices channels at access slot one. This can be achieved by separating devices in different groups via orthogonal pilots and devices within a group by TDMA to perform an initial training. While extensive, acquisition is performed only once, hence it does not have a significant effect on the overall performance of the trackers. Trackers update all users channel estimates at the pilot transmission phase of each access slot. At the uplink/downlink stages, BS uses the latest channel estimates to perform beamforming towards various devices.

In the rest of this paper we assume $\mathbf{A}_t^{(k)}$s and $\mathbf{Q}_t^{(k)}$s are known to the trackers for all k, t. Whether we really know these parameters in practice is an important implementation concern. We will address this issue's various aspects in three steps.

i) From a theoretical perspective, the state model parameters are assumed known and given in a Kalman filter (KF). However in an extended version of KF, which is usually referred to as adaptive KF, these parameters are also estimated from data which is the more practical approach. Performance of KF with perfect knowledge of state model parameters is always an upper bound on the performance of adaptive KF. Given this is the first manuscript with target tracking algorithms proposed for massive MIMO, we perform a proof of concept and investigate the upper bound on true performance which is given by a tracker with known state model parameters. MSE/rate for any adaptive versions with unknown state parameters will be lower/upper bounded by that of the proposed methods in this work.

ii) In wireless channels, time variations occur as a combination of three phenomenon: small-scale fading, shadowing and path-loss changes over time. All three variations can be incorporated as follows. First, we use Jakes model [15] for small-scale fading and find the state model parameters so that the proposed auto-regressive coefficients best fits Jakes model. Then, we model path-loss changes using each specific users velocity which is made up of its speed and direction. If we exploit certain localization techniques, velocity estimation is feasible in practice. Then, state model parameters variations due to path-loss can be modeled and incorporated. Third, we include the effect of shadowing by adding to the eigenvalues of $\mathbf{Q}_k$. Finally, any real deviations of this model from the true channel variations model can be incorporated by further increasing $\mathbf{Q}_k$ eigenvalues to represent more model uncertainty.

iii) We can extend this work to the setup where $\mathbf{A}_t^{(k)}$s and $\mathbf{Q}_t^{(k)}$s are also estimated from data. Main challenge will be to prove problem identifiability for both coordinated and uncoordinated cases. If the problem is proven to be well-behaving, when $q_t^{(k)} = 1$ we can estimate $\mathbf{A}_t^{(k)}$s and $\mathbf{Q}_t^{(k)}$s from measurements. However, when $q_t^{(k)} = 0$, there is no measurement to use for $\mathbf{A}_t^{(k)}$ and $\mathbf{Q}_t^{(k)}$ estimation in that particular t, k. In that case, we have to use the same values of $\mathbf{A}_t^{(k)}$ and $\mathbf{Q}_t^{(k)}$ as in the previous time slot. A combination of ii) and iii) is also viable where we use ii) when $q_t^{(k)} = 0$ and iii) when $q_t^{(k)} \neq 0$.

Next section deals with coordinated channel tracking where three algorithms with varying degrees of complexity-performance trade-off are presented.

3. Tracking with coordinated pilot access

First, we provide the optimum tracker which is a Kalman filter that tracks an aggregated state containing the joint state of all the K devices. Then, two sub-optimal alternatives with lower complexity are offered. We refer to all three as coordinated because BS knows the pilot transmission pattern $q_t^{(k)}$s.

3.1. Joint coordinated Kalman filter (JC-KF)

While (1) suggests that channels for different devices evolve independently and thus independent individual trackers might be optimal, the optimum tracker can not be decoupled across devices. This comes naturally as measurements in (2) do include collisions which introduce coupling between various devices. First, we remove the effect of the devices in other groups by multiplying measurements with the orthonormal pilot of group 1:

$$\mathbf{y}_t := \mathbf{Y}_t\boldsymbol{\phi}_1 = \sum_{k=1}^{K} q_t^{(k)}\mathbf{h}_t^{(k)} + \mathbf{w}_t. \tag{3}$$

Here, $\mathbf{y}_t \in \mathbb{R}^{M\times 1}$ represents the measurement vector corresponding to group one received at the M antennas at time t and $\mathbf{w}_t := \mathbf{W}_t\boldsymbol{\phi}_1$ denotes the corresponding noise vector which is still Gaussian and independent across entries with variance σ_w^2. We use the general covariance $\mathbf{R}_t$ in place of $\sigma_w^2\mathbf{I}$ to allow for correlated measurement noise also. It is needed for measurement noises to be independent over time. Upon introducing the aggregate state vector $\mathbf{h}_t := [\mathbf{h}_t^{(1)^T}, \mathbf{h}_t^{(2)^T}, \dots, \mathbf{h}_t^{(K)^T}]^T$ of size MK, $\mathbf{y}_t$ can be written as

$$\mathbf{y}_t = \begin{bmatrix} q_t^{(1)}\mathbf{I}_M \mid & q_t^{(2)}\mathbf{I}_M \mid & \cdots \mid & q_t^{(K)}\mathbf{I}_M \end{bmatrix}\mathbf{h}_t + \mathbf{w}_t := \mathbf{B}_t\mathbf{h}_t + \mathbf{w}_t. \tag{4}$$

Furthermore, (1) can be written in a joint state format as

$$\mathbf{h}_t = \mathbf{A}_t\mathbf{h}_{t-1} + \mathbf{u}_t, \tag{5}$$

where $\mathbf{A}_t$ of size $KM \times KM$ is block diagonal with $\mathbf{A}_t^{(k)}$s on the diagonals and $\mathbf{u}_t := [\mathbf{u}_t^{(1)^T}, \mathbf{u}_t^{(2)^T}, \dots, \mathbf{u}_t^{(K)^T}]^T$ is zero-mean with a block diagonal covariance matrix $\mathbf{Q}_t$ with $\mathbf{Q}_t^{(k)}$s on the diagonal. It can be observed that state equation in (5) is linear with all variables being jointly Gaussian. When $\mathbf{B}_t$s are known, measurement model is also linear with all variables being jointly Gaussian. Therefore, the optimum MMSE tracker is a Kalman filter run on the aggregate state. As in a typical KF, JC-KF begins with an initial estimate obtained from acquisition stage and then iteratively performs prediction and correction steps [17].

Prediction:

$$\hat{\mathbf{h}}_{t|t-1} = \mathbf{A}_t\hat{\mathbf{h}}_{t-1|t-1}, \quad \mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t^T + \mathbf{Q}_t. \tag{6}$$

Correction: First define $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{B}_t^T\left(\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^T + \mathbf{R}_t\right)^{-1}$. Then,

$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{B}_t\hat{\mathbf{h}}_{t|t-1}), \qquad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{B}_t)\mathbf{P}_{t|t-1}. \tag{7}$$

The JC-KF serves as a benchmark to compare all other tracking algorithms that will be introduced later on, including coordinated and uncoordinated methods, as it offers the smallest possible MSE. Before proceeding, we present two remarks.

remark 1. The covariance matrix $\mathbf{P}_{t|t}$ is of size $KM \times KM$ and contains the covariance matrix for channel estimate of device k on its k'th diagonal block. Furthermore, it contains the cross-covariance between users k and j channel estimates on the (k, j)'th block.

remark 2. Depending on the various devices activity patterns that forms $\mathbf{B}_t$ in (4), the MMSE for user k can be different from user j. Note that MMSE for user k is given by the trace of $\mathbf{P}_{k,k}$ which is the k'th $M \times M$ block on the diagonal of $\mathbf{P}_{t|t}$. This MMSE is independent of the measurements, and only a function of $\mathbf{B}_t$. It provides the fundamental MMSE limit that is achievable by the joint activity pattern encoded in $\mathbf{B}_t$.

3.2. Coordinated Kalman filter with collisions discarded

The chief limitation of JC-KF lies in its high complexity as its aggregated state is of size MK and the corresponding covariance matrix is if size $(MK)^2$. For large K, JC-KF can become prohibitively complex. To address this challenge we introduce two sub-optimal alternatives which utilize independent Kalman filters for various devices. It can be seen that if we discard $\mathbf{y}_t$ whenever it suffers a collision, coupling is removed. Therefore, we introduce coordinated independent KF (CI-KF) as follows. For each device k, run the following iterations independently.

Prediction:

$$\hat{\mathbf{h}}_{t|t-1}^{(k)} = \mathbf{A}_t^{(k)}\hat{\mathbf{h}}_{t-1|t-1}^{(k)}, \quad \mathbf{P}_{t|t-1}^{(k)} = \mathbf{A}_t^{(k)}\mathbf{P}_{t-1|t-1}^{(k)}\mathbf{A}_t^{(k)^T} + \mathbf{Q}_t^{(k)}. \tag{8}$$

Correction: If only user k transmitted $\boldsymbol{\phi}_1$ perform correction as below. Otherwise, skip to the prediction for next access slot. For correction, define $\mathbf{K}_t = \mathbf{P}_{t|t-1}^{(k)}\left(\mathbf{P}_{t|t-1}^{(k)} + \mathbf{R}_t\right)^{-1}$. Then,

$$\hat{\mathbf{h}}_{t|t}^{(k)} = \hat{\mathbf{h}}_{t|t-1}^{(k)} + \mathbf{K}_t(\mathbf{y}_t - \hat{\mathbf{h}}_{t|t-1}^{(k)}), \qquad \mathbf{P}_{t|t}^{(k)} = (\mathbf{I} - \mathbf{K}_t)\mathbf{P}_{t|t-1}^{(k)}. \tag{9}$$

If there is a collision, the corresponding $\mathbf{y}_t$ is discarded and only prediction is performed for all devices. If only a single device accessed the pilot, the channel estimate corresponding to that user is both predicted and corrected, while other devices channel estimates are only predicted. CI-KF operates satisfactorily particularly when number of collisions are a few. This occurs either if K is small or if K is large but BS schedules users so that collisions are minimized. However, CI-KF MSE is lower bounded by that of JC-KF since CI-KF can not exploit the information contained in the collisions. CI-KF needs $\mathcal{O}(KM^3)$ arithmetic operations per access slot while JC-KF needs $\mathcal{O}(K^2M^3)$. Complexity gap is considerable when K is large.

3.3. Coordinated Kalman filter via belief propagation

CI-KF will skip many measurements when collisions are abound and will perform poorly. To overcome this limitation, we introduce a second sub-optimal KF that is derived using belief propagation (BP) on a factor graph. This BP-based KF (BP-KF), exploits collisions while its complexity is of the same order as CI-KF. BP-KF algorithm is concisely presented here, however its derivation is relegated to the Appendix A.

1. Fix $\mathbf{h}_0^{(k)} \sim \mathcal{N}(\hat{\mathbf{h}}_0^{(k)}, \mathbf{P}_{0|0}^{(k)})$ for all $k = 1, \dots, K$ which are obtained from the initial acquisition stage. Set $t = 0$.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step independently for all devices. This amounts to calculating $\hat{\mathbf{h}}_{t|t-1}^{(k)} = \mathbf{A}_t^{(k)}\hat{\mathbf{h}}_{t-1|t-1}^{(k)}$ and $\mathbf{P}_{t|t-1}^{(k)} = \mathbf{A}_t^{(k)}\mathbf{P}_{t-1|t-1}^{(k)}\mathbf{A}_t^{(k)^T} + \mathbf{Q}_t^{(k)}$.
4. For those users k who are present in the current collision form the fictitious measurement $\hat{\mathbf{y}}_t^{(k)}$ and its covariance matrix $\hat{\mathbf{R}}_t^{(k)}$ given as

$$\hat{\mathbf{y}}_t^{(k)} := \mathbf{y}_t - \sum_{j\neq k} q_t^{(j)}\hat{\mathbf{h}}_{t|t-1}^{(j)}, \qquad \hat{\mathbf{R}}_t^{(k)} := \mathbf{R}_t + \sum_{j\neq k} q_t^{(j)}\hat{\mathbf{P}}_{t|t-1}^{(j)} \tag{10}$$

5. Correct the channel estimate for the users that participate in the collision as follows. For user k assume the measurement model $\hat{\mathbf{y}}_t^{(k)} := \mathbf{h}_t^{(k)} + \hat{\mathbf{w}}_t^{(k)}$ where $\hat{\mathbf{w}}_t^{(k)}$ is the fictitious noise with zero-mean and covariance $\hat{\mathbf{R}}_t^{(k)}$, then run the correction step.

Define $\mathbf{K}_t^{(k)} = \mathbf{P}_{t|t-1}^{(k)}\left(\mathbf{P}_{t|t-1}^{(k)} + \hat{\mathbf{R}}_t^{(k)}\right)^{-1}$. Then, perform

$$\hat{\mathbf{h}}_{t|t}^{(k)} = \hat{\mathbf{h}}_{t|t-1}^{(k)} + \mathbf{K}_t^{(k)}\left(\hat{\mathbf{y}}_t^{(k)} - \hat{\mathbf{h}}_{t|t-1}^{(k)}\right), \qquad \mathbf{P}_{t|t} = \left(\mathbf{I} - \mathbf{K}_t^{(k)}\right)\mathbf{P}_{t|t-1}^{(k)}. \tag{11}$$

6. Set $t \leftarrow t+1$ and go to step 2.

It is interesting to observe how BP-KF deals with collisions. Firstly, BP-KF amounts to applying JC-KF but after each iteration it discards the cross-correlation terms which are the off diagonal blocks in the joint covariance matrix. This renders the covariance matrix block diagonal leading to independent Kalman filters across users. Second observation is that the aforementioned fictitious measurement compensates for other users presence in the collision by subtracting the best available estimate of their channels but enlarges the covariance of the fictitious measurement noise to account for errors in other users channel estimates which results in (10). BP-KF complexity per access slot is $\mathcal{O}(KM^3)$.

3.4. Comparison with prior art

Comparing [25] against our methods, we draw several conclusions. First, [25] treats the desired user channel as an FIR filter in time and tracks all the channel coefficients allowing for tracking the complete channel response in the frequency domain, while we track one coherence bandwidth only. Secondly, [25] allows for $\mathbf{A}_t^{(k)}$s to be unknown, while we treat them as known. On the other hand, [25] fails to utilize the information obtained from tracking other devices as JC-KF does. Indeed, if there are no collisions at all, our CI-KF and tracker in [25] will be the same. Another major limitation of [25] is that it knows the pilot access pattern of the desired device, hence it amounts to a coordinated approach according to our notation in spite of the term uncoordinated in their title. Therefore, [25] can not deal with unknown pilot transmission pattern for the desired user at the BS. Still, comparisons with our proposed coordinated trackers will be performed in the simulations.

4. Tracking with uncoordinated pilot access

As its main advantage, coordination allows for a simple tracker with polynomial complexity such as JC-KF to become the MMSE estimate. On the negative side, it diminishes system flexibility. To enable coordination, devices should either use a fixed random pilot access pattern, or periodically notify the BS of changes in access pattern. Changing, or adaptive, access pattern can be necessary from several perspectives. For devices that run on a battery or harvest energy, a fixed pilot access pattern is not justified as they should adapt pilot access rate to their available energy levels. Furthermore, a single device might transmit data with various reliability requirements over time. To increase reliability, the device may decide to increase its pilot access rate, while decreasing it when data is tolerant to errors. However, periodic notifications on new access policies from all IoT devices can jam the control channel and lead to a control bottleneck. To ensure flexibility in pilot access while avoiding excessive control signaling, we offer a novel viewpoint not considered before. The main idea is to allow all devices to individually choose their pilot access slots at will without any coordination with the BS. However, BS will be tasked with the additional burden of detecting pilot access patterns. Therefore, we assume $q_t^{(k)}$s are not known to the BS.

As BS does not know in advance which devices are present in a collision, the tracking task becomes significantly more complicated. This challenge has existed for a long time in target tracking community and is referred to as the measurement origin uncertainty problem, the data association problem, or data assignment problem. While the MMSE optimal tracker is prohibitively complex, many sub-optimal trackers have been developed. Unfortunately, target tracking solutions can not be readily applied to our framework as tracking community always dealt with measurement origin uncertainty, but they always assumed that every single measurement is at most generated by one target. This assumption is no longer valid in our scenario as a combination of devices are present in a collision. We correspondingly modify the available target tracking algorithms to accommodate this new assumption. In this Section, we first present the optimal uncoordinated tracker. Then, we present three sub-optimal trackers which are global nearest neighbor (GNN) [2], probabilistic data association filter (PDAF) [2], and multiple hypotheses tracker (MHT) [21].

4.1. Optimal tracker

The optimal, in the MMSE sense, joint tracker is given by

$$\begin{aligned}\hat{\mathbf{h}}_t :&= \mathrm{E}[\mathbf{h}_t|\mathbf{y}_{1:t}] = \mathrm{E}[\mathrm{E}[\mathbf{h}_t|\mathbf{y}_{1:t}, \mathbf{q}_{1:t}]|\mathbf{y}_{1:t}] \\ &= \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})\ \mathrm{E}[\mathbf{h}_t|\mathbf{y}_{1:t}, \mathbf{q}_{1:t}]\end{aligned} \tag{12}$$

Conditioned on $\mathbf{q}_{1:t}$, the inner expected value is simply evaluated by a joint KF that assumes $\mathbf{B}_{1:t}$s are given according to that specific $\mathbf{q}_{1:t}$ pattern. To compute the outer expectation, we should enumerate over all possible instances of $\mathbf{q}_{1:t}$. It is worth mentioning that each $\mathbf{q}_t$ can assume 2^K values and $\mathbf{q}_{1:t}$ assumes 2^{tK} different values. Thus, to evaluate optimal MMSE tracker, 2^{Kt} joint KFs should be run in parallel and then combined by the weights $p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})$. Indeed, the optimal MMSE tracker is distributed as a Gaussian mixture with exponentially increasing number of mixtures over time. The weights $p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})$ are evaluated as follows:

$$p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{q}_{1:t})p(\mathbf{q}_{1:t})}{p(\mathbf{y}_{1:t})}, \tag{13}$$

$$\begin{aligned}p(\mathbf{q}_{1:t}) &= \prod_{i=1}^{t} p(\mathbf{q}_i) = \prod_{i=1}^{t}\prod_{k=1}^{K} \lambda_k^{q_i^{(k)}}(1-\lambda_k)^{1-q_i^{(k)}} \\ p(\mathbf{y}_{1:t}|\mathbf{q}_{1:t}) &= p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{q}_{1:t})p(\mathbf{y}_{1:t-1}|\mathbf{q}_{1:t}) \\ &= p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \mathbf{q}_{1:t})p(\mathbf{y}_{1:t-1}|\mathbf{q}_{1:t-1}), \\ &= \prod_{i=1}^{t} p(\mathbf{y}_i|\mathbf{y}_{1:i-1}, \mathbf{q}_{1:i}) \\ &= \prod_{i=1}^{t} \mathcal{N}\left(\mathbf{y}_i; \mathbf{B}_i(\mathbf{q}_i)\hat{\mathbf{h}}_{i|i-1}, \mathbf{B}_i(\mathbf{q}_i)\mathbf{P}_{i|i-1}\mathbf{B}_i^T(\mathbf{q}_i) + \mathbf{R}_i\right).\end{aligned} \tag{14}$$

In (14), we run a joint KF assuming a given $\mathbf{q}_{1:t}$ and evaluate how good the given measurements are predicted by this choice of $\mathbf{q}_{1:t}$. This process is repeated for all possible choices of $\mathbf{q}_{1:t}$. Once weights are obtained from (13), they are normalized to one to account for the unknown $p(\mathbf{y}_{1:t})$ in the denominator in (13) and then plugged into (12) to compute the mean and covariance matrix for $\hat{\mathbf{h}}_t$. The overall complexity of optimal tracker is $\mathcal{O}(K^3M^3 2^{Kt})$.

4.2. Global nearest neighbor

Optimal tracker enumerates all possible hypotheses on the origin of measurements and then assigns a corresponding weight to each hypothesis. The weight is assigned according to how well that particular hypothesis is predicted by the measurements. Unfortunately, number of hypotheses grow exponentially in t preventing the applicability of the optimal tracker. One simple remedy is to

greedily pick the hypothesis that predicts the data best at the current time and fix it for future access slots. This algorithm is referred to as GNN.

GNN operates as follows. At time t, it is assumed that $\mathbf{q}_{1:t-1}$ are selected correctly. Thus, one only needs to select the best possible $\mathbf{q}_t$ which assumes 2^K values. Unlike the optimal tracker, GNN performs hard assignment meaning that it discards all hypotheses except the best greedy one. GNN algorithm is briefly described below.

1. Fix $\mathbf{h}_0 \sim \mathcal{N}(\hat{\mathbf{h}}_0, \mathbf{P}_{0|0})$ which are obtained from the initial acquisition stage. Set $t = 0$.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step on the joint state. This amounts to calculating $\hat{\mathbf{h}}_{t|t-1} = \mathbf{A}_t \hat{\mathbf{h}}_{t-1|t-1}$ and $\mathbf{P}_{t|t-1} = \mathbf{A}_t \mathbf{P}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{Q}_t$.
4. Consider all possible hypotheses on $\mathbf{q}_t$. Then find the one that maximizes $p(\mathbf{q}_t | \mathbf{q}^*_{1:t-1}, \mathbf{y}_{1:t})$ and denote it by $\mathbf{q}^*_t$. This is done as follows:

$$\begin{aligned} &p\left(\mathbf{q}_t \middle| \mathbf{q}^*_{1:t-1}, \mathbf{y}_{1:t}\right) \\ &= \frac{p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}, \mathbf{q}^*_{1:t-1}, \mathbf{q}_t\right) p\left(\mathbf{q}_t \middle| \mathbf{y}_{1:t-1}, \mathbf{q}^*_{1:t-1}\right)}{p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}, \mathbf{q}^*_{1:t-1}\right)} \\ &\propto p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}, \mathbf{q}^*_{1:t-1}, \mathbf{q}_t\right) p(\mathbf{q}_t) \\ &\propto \mathcal{N}\Big(\mathbf{y}_t; \mathbf{B}_t(\mathbf{q}_t)\hat{\mathbf{h}}_{t|t-1}, \mathbf{B}_t(\mathbf{q}_t)\mathbf{P}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t) + \mathbf{R}_t\Big) \times \\ &\prod_{k=1}^K \lambda_k^{q_t^{(k)}} (1-\lambda_k)^{q_t^{(k)}}. \end{aligned}$$

   Note that $\hat{\mathbf{h}}_{t|t-1}, \mathbf{P}_{t|t-1}$ are computed from the previous iterations of the single KF that assumes $\mathbf{q}^*_{1:t-1}$.
5. Set $\mathbf{B}^*_t := \mathbf{B}_t(\mathbf{q}^*_t)$, then run the correction step assuming $\mathbf{B}^*_t$ was the true model. First, define $\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{B}_t^{*T}\left(\mathbf{B}^*_t \mathbf{P}_{t|t-1}\mathbf{B}_t^{*T} + \mathbf{R}_t\right)^{-1}$. Then, perform

$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{B}^*_t \hat{\mathbf{h}}_{t|t-1}), \qquad \mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \mathbf{B}^*_t)\mathbf{P}_{t|t-1}. \tag{15}$$

6. Set $t \leftarrow t+1$ and go to step 2.

GNN tracker assumes a Gaussian distribution for $\hat{\mathbf{h}}_{t|t}$ to simplify calculations in contrast to the optimal tracker which is distributed as Gaussian mixture model (GMM). Furthermore, it uses a single-step greedy approach to select the best possible $\mathbf{q}_t$. Once the greedy optimum is chosen it becomes fixed for future random access slots and its optimality is never re-evaluated again.

4.3. Multiple hypothesis tracker

GNN keeps track of only one hypothesis over time, while the optimal tracker keeps account of all possible hypotheses. To fill the gap between these two, one might suggest to keep track of a fixed N_h hypotheses instead of one. This idea will give rise to MHT. To elaborate, assume $\Omega^i_{t-1} := \{\mathbf{q}_1(i), \mathbf{q}_2(i), \ldots, \mathbf{q}_{t-1}(i)\}$ denotes the i'th hypothesis that is kept by MHT at time $t-1$. The MHT is briefly described below.

1. Fix $\mathbf{h}_0 \sim \mathcal{N}(\hat{\mathbf{h}}_0, \mathbf{P}_{0|0})$ which are obtained from the initial acquisition stage. Set $t = 0$. Set $\Omega^i_0 = \emptyset$ for all $i = 1, 2, \ldots, N_h$. Set $p(\Omega^i_0) = 1/N_h$ for all i.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step on the joint state for all hypotheses. This amounts to calculating $\hat{\mathbf{h}}^{(\Omega^i_{t-1})}_{t|t-1} = \mathbf{A}_t \hat{\mathbf{h}}^{(\Omega^i_{t-1})}_{t-1|t-1}$ and $\mathbf{P}^{(\Omega^i_{t-1})}_{t|t-1} = \mathbf{A}_t \mathbf{P}^{(\Omega^i_{t-1})}_{t-1|t-1} \mathbf{A}_t^T + \mathbf{Q}_t$ for the i'th hypothesis Ω^i_{t-1}.
4. For every i, augment Ω^i_{t-1} with all possible choices on $\mathbf{q}_t$. Since $\mathbf{q}_t$ can take 2^K different values, each Ω^i_{t-1} is expanded into 2^K hypotheses. Mathematically, $\tilde{\Omega}^{(i,j)}_t := \left\{\Omega^i_{t-1}, \mathbf{q}_t(j)\right\}$ where $j = 1, 2, \ldots, 2^K$ and $i = 1, 2, \ldots, N_h$.
5. Next, evaluate the probability of each hypotheses as follows.

$$\begin{aligned} &p(\tilde{\Omega}^{(i,j)}_t | \mathbf{y}_{1:t}) \\ &= \frac{p(\mathbf{y}_t | \tilde{\Omega}^{(i,j)}_t, \mathbf{y}_{1:t-1})\ p(\tilde{\Omega}^{(i,j)}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})} \\ &\propto \mathcal{N}\Big(\mathbf{y}_t; \mathbf{B}_t(\mathbf{q}_t(j))\hat{\mathbf{h}}^{(\Omega^i_{t-1})}_{t|t-1}, \mathbf{B}_t(\mathbf{q}_t(j))\mathbf{P}^{(\Omega^i_{t-1})}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t(j)) + \mathbf{R}_t\Big) \\ &p(\Omega^i_{t-1} | \mathbf{y}_{1:t-1}) \end{aligned} \tag{16}$$

6. Select the N_h largest values in (16) among $N_h 2^K$ hypotheses and discard the rest. Then, assign these N_h hypotheses to $\Omega^1_t, \Omega^2_t, \ldots, \Omega^{N_h}_t$.
7. Normalize the weights as

$$p(\Omega^{\hat{i}}_t | \mathbf{y}_{1:t}) = \frac{p(\tilde{\Omega}^{(i,j)}_t | \mathbf{y}_{1:t})}{\sum_{i=1}^{N_h} p(\tilde{\Omega}^{(i,j)}_t | \mathbf{y}_{1:t})}$$

   where right hand side probabilities are given by (16) when $\Omega^{\hat{i}}_t = \tilde{\Omega}^{(i,j)}_t$ is the $\hat{i}$'th largest probable hypothesis.
8. For $\hat{i} = 1, 2, \ldots, N_h$ run the correction step as follows

$$\begin{aligned} \mathbf{K}^{(\Omega^{\hat{i}}_t)}_t &:= \mathbf{P}^{(\Omega^i_{t-1})}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t(j))\Big(\mathbf{B}_t(\mathbf{q}_t(j))\mathbf{P}^{(\Omega^i_{t-1})}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t(j)) + \mathbf{R}_t\Big)^{-1}, \\ \hat{\mathbf{h}}^{(\Omega^{\hat{i}}_t)}_{t|t} &= \hat{\mathbf{h}}^{(\Omega^i_{t-1})}_{t|t-1} + \mathbf{K}^{(\Omega^{\hat{i}}_t)}_t \Big(\mathbf{y}_t - \mathbf{B}_t(\mathbf{q}_t(j))\hat{\mathbf{h}}^{(\Omega^i_{t-1})}_{t|t-1}\Big), \\ \mathbf{P}^{(\Omega^{\hat{i}}_t)}_{t|t} &= \Big(\mathbf{I} - \mathbf{K}^{(\Omega^{\hat{i}}_t)}_t \mathbf{B}_t(\mathbf{q}_t(j))\Big)\mathbf{P}^{(\Omega^i_{t-1})}_{t|t-1}. \end{aligned}$$

9. As the estimate at time t, select the hypothesis with the largest weight and assign the corresponding $\hat{\mathbf{h}}^{(\Omega^{\hat{i}}_t)}_{t|t}$ as MHT tracker output.
10. Set $t \leftarrow t+1$ and go to step 2.

When $N_h = 1$, MHT reduces to GNN. Choice of N_h provides a trade-off between complexity and performance. Like GNN, MHT relies on hard assignments as it discards all the unfavorable hypotheses.

4.4. Probabilistic data association filter

A characteristic of optimal tracker which is missing in GNN and MHT is its soft (probabilistic) assignment of hypotheses. Therefore, all hypotheses do have an impact on the final estimate but unfavorable ones have a smaller effect than favorable ones. The impact of each hypotheses is determined by its corresponding weight. PDAF utilizes the same idea of soft assignment but does so only for the current measurement. To elaborate further, we recall that MMSE is given by $\hat{\mathbf{h}}_{t|t} = \mathrm{E}[\mathbf{h}_t | \mathbf{y}_{1:t}]$ and the corresponding probability density function is evaluated as follow:

$$p(\mathbf{h}_t | \mathbf{y}_{1:t}) = \sum_{\mathbf{q}_t} p(\mathbf{h}_t, \mathbf{q}_t | \mathbf{y}_{1:t}) = \sum_{\mathbf{q}_t} p(\mathbf{q}_t | \mathbf{y}_{1:t})\ p(\mathbf{h}_t | \mathbf{q}_t, \mathbf{y}_{1:t}) \tag{17}$$

We have conditioned on $\mathbf{q}_t$ only compared to the optimal uncoordinated tracker which was conditioned on $\mathbf{q}_{1:t}$ (12). The summation is over all 2^K possible hypotheses on $\mathbf{q}_t$. The weights on the right hand side of (17) can be written as

$$\begin{aligned} p(\mathbf{q}_t | \mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t | \mathbf{q}_t, \mathbf{y}_{1:t-1})\ p(\mathbf{q}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})} \\ &\propto p(\mathbf{y}_t | \mathbf{q}_t, \mathbf{y}_{1:t-1})\ \lambda_k^{q_t^{(k)}} (1-\lambda_k)^{(1-q_t^{(k)})}. \end{aligned} \tag{18}$$

The first term in (18) is further expanded as

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{q}_t, \mathbf{y}_{1:t-1}) &= \int p(\mathbf{y}_t | \mathbf{h}_t, \mathbf{q}_t, \mathbf{y}_{1:t-1})\ p(\mathbf{h}_t | \mathbf{q}_t, \mathbf{y}_{1:t-1})\ d\mathbf{h}_t \\ &= \int p(\mathbf{y}_t | \mathbf{h}_t, \mathbf{q}_t)\ p(\mathbf{h}_t | \mathbf{y}_{1:t-1})\ d\mathbf{h}_t \end{aligned} \tag{19}$$

$$\approx \mathcal{N}\Big(\mathbf{y}_t; \mathbf{B}_t(\mathbf{q}_t)\hat{\mathbf{h}}_{t|t-1}, \mathbf{B}_t(\mathbf{q}_t)\mathbf{P}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t) + \mathbf{R}_t\Big)$$

The first equality in (19) is valid because given $\mathbf{h}_t, \mathbf{q}_t$, measurement $\mathbf{y}_t$ is independent of past measurements $\mathbf{y}_{1:t-1}$. In addition, given only past measurements $\mathbf{y}_{1:t-1}$, $\mathbf{q}_t$ is independent of $\mathbf{h}_t$. The approximation occurs because $p(\mathbf{h}_t|\mathbf{y}_{1:t-1})$ is given by the optimal tracker which is too complex and thus we approximate it using the predictions on the previous step of PDAF. Upon setting $p(\mathbf{h}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{N}(\mathbf{h}_t; \hat{\mathbf{h}}_{t|t-1}, \mathbf{P}_{t|t-1})$ and $p(\mathbf{y}_t|\mathbf{h}_t, \mathbf{q}_t) = \mathcal{N}(\mathbf{y}_t; \mathbf{B}_t(\mathbf{q}_t)\mathbf{h}_t, \mathbf{R}_t)$ and integrating over $\mathbf{h}_t$ the approximation ensues. Same approximation is utilized in evaluating the second term in (17). Specifically,

$$\begin{aligned} p(\mathbf{h}_t|\mathbf{q}_t, \mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t|\mathbf{h}_t, \mathbf{q}_t, \mathbf{y}_{1:t-1})\ p(\mathbf{h}_t|\mathbf{q}_t, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{q}_t, \mathbf{y}_{1:t-1})} \\ &\propto p(\mathbf{y}_t|\mathbf{h}_t, \mathbf{q}_t)\ p(\mathbf{h}_t|\mathbf{y}_{1:t-1}) \\ &\approx \mathcal{N}(\mathbf{y}_t; \mathbf{B}_t(\mathbf{q}_t)\mathbf{h}_t, \mathbf{R}_t)\ \mathcal{N}(\mathbf{h}_t; \hat{\mathbf{h}}_{t|t-1}, \mathbf{P}_{t|t-1}) \\ &\propto \mathcal{N}\Big(\mathbf{h}_t; \hat{\mathbf{h}}_{t|t}(\mathbf{q}_t), \mathbf{P}_{t|t}(\mathbf{q}_t)\Big)\Big). \end{aligned} \tag{20}$$

where

$$\begin{aligned} \mathbf{K}_t(\mathbf{q}_t) &= \mathbf{P}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t)\Big(\mathbf{B}_t(\mathbf{q}_t)\mathbf{P}_{t|t-1}\mathbf{B}_t^T(\mathbf{q}_t) + \mathbf{R}_t\Big)^{-1} \\ \hat{\mathbf{h}}_{t|t}(\mathbf{q}_t) &= \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t(\mathbf{q}_t)(\mathbf{y}_t - \mathbf{B}_t(\mathbf{q}_t)\hat{\mathbf{h}}_{t|t-1}), \\ \mathbf{P}_{t|t}(\mathbf{q}_t) &= (\mathbf{I} - \mathbf{K}_t(\mathbf{q}_t)\mathbf{B}_t(\mathbf{q}_t))\mathbf{P}_{t|t-1} \end{aligned}$$

$\mathbf{K}_t(\mathbf{q}_t)$ denotes the Kalman gain assuming the hypothesis $\mathbf{q}_t$ is true. Similarly, $\hat{\mathbf{h}}_{t|t}(\mathbf{q}_t)$ and $\mathbf{P}_{t|t}(\mathbf{q}_t)$ are the corresponding mean and covariance after correction step conditioned on $\mathbf{q}_t$. Indeed, (20) illustrates an alternative derivation of the correction step for the Kalman filter. To summarize, the weights in (17) are evaluated via (18) and (19) for all possible 2^K hypotheses over $\mathbf{q}_t$ and normalized to one. We also use the approximation (20) in (17) and will arrive at a Gaussian mixture. PDAF estimate of mean and covariance is given by the Gaussian mixture in (17). It is straightforward to show that this mean and covariance are given by

$$\begin{aligned} \hat{\mathbf{h}}_{t|t} &= \sum_{\mathbf{q}_t} p(\mathbf{q}_t|\mathbf{y}_{1:t})\hat{\mathbf{h}}_{t|t}(\mathbf{q}_t), \\ \mathbf{P}_{t|t} &= \sum_{\mathbf{q}_t} p(\mathbf{q}_t|\mathbf{y}_{1:t})\Big(\mathbf{P}_{t|t}(\mathbf{q}_t) + \hat{\mathbf{h}}_{t|t}(\mathbf{q}_t)\hat{\mathbf{h}}_{t|t}^T(\mathbf{q}_t)\Big) - \hat{\mathbf{h}}_{t|t}\hat{\mathbf{h}}_{t|t}^T. \end{aligned} \tag{21}$$

PDAF is briefly described as follows.

1. Fix $\mathbf{h}_0 \sim \mathcal{N}(\hat{\mathbf{h}}_0, \mathbf{P}_{0|0})$ which are obtained from the initial acquisition stage. Set $t = 0$.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step on the joint state. This amounts to calculating $\hat{\mathbf{h}}_{t|t-1} = \mathbf{A}_t\hat{\mathbf{h}}_{t-1|t-1}$ and $\mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t^T + \mathbf{Q}_t$.
4. Consider all possible 2^K hypotheses on $\mathbf{q}_t$. Compute the weights as in (17) using (18),(19). Then, compute the 2^K correction steps for different $\mathbf{q}_t$ according to (20).
5. Compute the mean and covariance of PDAF according to (21).
6. Set $t \leftarrow t + 1$ and go to step 2.

5. Low-complexity channel tracking with uncoordinated pilot access

While linear in t, GNN, MHT, and PDAF complexities are exponential in K limiting the values of K for which these algorithms are tractable. This section offers two heuristic remedies. The first maintain polynomial complexity while the second has a random complexity whose worst case can be exponential, however its average complexity was observed to be similar to the first method as verified by numerical comparisons.

5.1. Soft and hard least squares

We can rewrite (3) as follows:

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{q}_t + \mathbf{w}_t \tag{22}$$

where $\mathbf{H}_t := [\mathbf{h}_t^{(1)} | \mathbf{h}_t^{(2)} | \cdots | \mathbf{h}_t^{(K)}]$. If $\mathbf{H}_t$ was known, we could have estimated $\mathbf{q}_t$ via an integer constrained least-squares (LS). Subsequently, two difficulties arise. First, integer constraints on $\mathbf{q}_t$ make the problem significantly harder to solve. Secondly, we do not know $\mathbf{H}_t$. To address these challenges, we apply two approximation. First, we relax the integer constraint $\mathbf{q}_t \in \{0, 1\}^K$ into a convex hypercube $\mathbf{q}_t \in [0, 1]^K$. Furthermore, we replace the unknown $\mathbf{H}_t$ with its best available estimate $\hat{\mathbf{H}}_t := [\hat{\mathbf{h}}_{t|t-1}^{(1)} | \hat{\mathbf{h}}_{t|t-1}^{(2)} | \cdots | \hat{\mathbf{h}}_{t|t-1}^{(K)}]$. Upon utilizing these approximations, unconstrained LS is immediately obtained as

$$\hat{\mathbf{q}}_t = \big(\hat{\mathbf{H}}_t^T\hat{\mathbf{H}}_t\big)^{-1}\hat{\mathbf{H}}_t^T\mathbf{y}_t \tag{23}$$

For soft LS, we project the obtained $\hat{\mathbf{q}}_t$ into the hypercube $[0, 1]^K$ which simply amounts to rounding values of $\hat{\mathbf{q}}_t$ out of the [0,1] interval to either 0 or 1 whichever is closer and keeping the values inside the interval intact. Given the convex quadratic cost for LS, this optimization then projection approach is equivalent to solving the constrained LS with a hypercube constraint.

For hard LS, we project the obtained $\hat{\mathbf{q}}_t$ into the set $\{0, 1\}^K$ which amounts to rounding each value to either 0 or 1. Note that hard LS makes hard decisions on if a particular user is present or absent in a collision while soft LS weighs each user corresponding to its unconstrained LS estimate. The overall algorithm is summarized as follows.

1. Fix $\mathbf{h}_0^{(k)} \sim \mathcal{N}(\hat{\mathbf{h}}_0^{(k)}, \mathbf{P}_{0|0}^{(k)})$ for all $k = 1, \ldots, K$ which are obtained from the initial acquisition stage. Set $t = 0$.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step for the aggregate state as $\hat{\mathbf{h}}_{t|t-1} = \mathbf{A}_t\hat{\mathbf{h}}_{t-1|t-1}$ and $\mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t^T + \mathbf{Q}_t$.
4. Reorder vector $\hat{\mathbf{h}}_{t|t-1}$ in a matrix form according to $\hat{\mathbf{H}}_t := [\hat{\mathbf{h}}_{t|t-1}^{(1)} | \hat{\mathbf{h}}_{t|t-1}^{(2)} | \cdots | \hat{\mathbf{h}}_{t|t-1}^{(K)}]$. Then solve (23) to obtain $\hat{\mathbf{q}}_t$ via soft/hard LS.
5. Form
$$\hat{\mathbf{B}}_t := \big[\hat{q}_t^{(1)}\mathbf{I}_M \mid \hat{q}_t^{(2)}\mathbf{I}_M \mid \cdots \mid \hat{q}_t^{(K)}\mathbf{I}_M\big]$$
and run the correction step as
$$\mathbf{K}_t := \mathbf{P}_{t|t-1}\hat{\mathbf{B}}_t^T\big(\hat{\mathbf{B}}_t\mathbf{P}_{t|t-1}\hat{\mathbf{B}}_t^T + \mathbf{R}_t\big)^{-1},$$
$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \hat{\mathbf{B}}_t\hat{\mathbf{h}}_{t|t-1}), \qquad \mathbf{P}_{t|t} = \big(\mathbf{I} - \mathbf{K}_t\hat{\mathbf{B}}_t\big)\mathbf{P}_{t|t-1}.$$
6. Set $t \leftarrow t + 1$ and go to step 2.

The complexity of solving (23) is of order $\mathcal{O}(K^3 + MK^2)$ which is approximated as $\mathcal{O}(MK^2)$ given that $M \gg K$ in practice. Given that the complexity of subsequent KF correction in step 5 is of order $\mathcal{O}(K^2M^3)$, it will dominate the overall complexity. Hence, the overall soft/hard LS complexity will be of order $\mathcal{O}(K^2M^3)$ per time step.

5.2. Locally optimum maximum likelihood

While very simple complexity-wise, the soft/hard LS algorithms in the previous section might perform poorly due to their underlying approximations. One can improve their performance by evaluating the maximum-likelihood estimate (MLE) instead of LS. The joint density of $\mathbf{y}_{1:t}$ parameterized by $\mathbf{q}_t$ is written as

$$\begin{aligned} p(\mathbf{y}_{1:t}; \mathbf{q}_t) &= p(\mathbf{y}_t|\mathbf{y}_{1:t-1}; \mathbf{q}_t)p(\mathbf{y}_{1:t-1}; \mathbf{q}_t) \\ &= p(\mathbf{y}_t|\mathbf{y}_{1:t-1}; \mathbf{q}_t)p(\mathbf{y}_{1:t-1}) \end{aligned} \tag{24}$$

The second equality follows because $\mathbf{y}_{1:t-1}$ does not depend on $\mathbf{q}_t$. To maximize the joint density over $\mathbf{q}_t$, we should maximize $p(\mathbf{y}_t|\mathbf{y}_{1:t-1};\mathbf{q}_t)$ which is distributed as GMM. To ensure tractability, we assume that the MLE filter has correctly found $\mathbf{q}_{1:t-1}$. This assumption has been made by all the other sub-optimal filters of Sections 4 and 5 as well. Then, one can write (22) as

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{q}_t + \mathbf{w}_t = \hat{\mathbf{H}}_t\mathbf{q}_t + \underbrace{(\mathbf{H}_t - \hat{\mathbf{H}}_t)\mathbf{q}_t + \mathbf{w}_t}_{\mathbf{e}_t} \tag{25}$$

where given $\mathbf{y}_{1:t-1}$, the first term yields the mean and the second term, which is $\mathbf{e}_t$, yields the zero-mean Gaussian noise with its covariance given by

$$\begin{aligned}
\mathbf{C_{ee}} &= \mathbb{E}\Big[\mathbf{e}_t\mathbf{e}_t^T|\mathbf{y}_{1:t-1}\Big] \\
&= \mathbb{E}\Big\{\Big[(\mathbf{H}_t - \hat{\mathbf{H}}_t)\mathbf{q}_t + \mathbf{w}_t\Big]\Big[(\mathbf{H}_t - \hat{\mathbf{H}}_t)\mathbf{q}_t + \mathbf{w}_t\Big]^T\Big\} \\
&= \mathbb{E}\Bigg\{\Bigg[\sum_{k=1}^{K} q_t^{(k)}(\mathbf{h}_t^{(k)} - \hat{\mathbf{h}}_{t|t-1}^{(k)}) + \mathbf{w}_t\Bigg] \\
&\quad \Bigg[\sum_{\ell=1}^{K} q_t^{(\ell)}(\mathbf{h}_t^{(\ell)} - \hat{\mathbf{h}}_{t|t-1}^{(\ell)}) + \mathbf{w}_t\Bigg]^T\Bigg\} \\
&= \sum_{k=1}^{K}\sum_{\ell=1}^{K} q_t^{(k)}q_t^{(\ell)}\Big(\mathbf{P}_{t|t-1}\Big)_{k,\ell} + \mathbf{R}_t
\end{aligned}$$

where $\big(\mathbf{P}_{t|t-1}\big)_{k,\ell}$ corresponds to the k,ℓ block of the joint covariance. MLE maximizes the log-likelihood given by

$$\hat{\mathbf{q}}_t = \arg\max_{\mathbf{q}_t\in\{0,1\}^K} -\frac{1}{2}\log|\mathbf{C_{ee}}| - \frac{1}{2}(\mathbf{y}_t - \hat{\mathbf{H}}_t\mathbf{q}_t)^T\mathbf{C_{ee}}^{-1}(\mathbf{y}_t - \hat{\mathbf{H}}_t\mathbf{q}_t) \tag{26}$$

Due to the integer constraints, MLE is difficult to compute. We apply coordinate ascent (CA). Given that each CA step improves ML objective, which is bounded above, its convergence to a local optimum is guaranteed. We begin with $\mathbf{q}_t = \mathbf{0}$ and each time vary a coordinate k, which is $q_t^{(k)}$, between zero and one with all the other coordinates fixed. Then, select the choice that yields a higher objective. Then, we move on to the next coordinate. In the worst-case the algorithm converges in 2^K steps. Finally, it is notable that if we added a prior on $\mathbf{q}_t$ in (24) we arrived at an sub-optimal maximum a-posteriori (MAP) estimate which yields the same computational complexity as MLE. The overall algorithm is summarized as follows.

1. Fix $\mathbf{h}_0^{(k)} \sim \mathcal{N}(\hat{\mathbf{h}}_0^{(k)}, \mathbf{P}_{0|0}^{(k)})$ for all $k = 1,\dots,K$ which are obtained from the initial acquisition stage. Set $t = 0$.
2. At time instant $t \geqslant 1$ perform the following steps:
3. Perform prediction step for the aggregate state as $\hat{\mathbf{h}}_{t|t-1} = \mathbf{A}_t\hat{\mathbf{h}}_{t-1|t-1}$ and $\mathbf{P}_{t|t-1} = \mathbf{A}_t\mathbf{P}_{t-1|t-1}\mathbf{A}_t^T + \mathbf{Q}_t$.
4. Reorder vector $\hat{\mathbf{h}}_{t|t-1}$ in a matrix form according to $\hat{\mathbf{H}}_t := [\hat{\mathbf{h}}_{t|t-1}^{(1)} \,|\, \hat{\mathbf{h}}_{t|t-1}^{(2)} \,|\cdots|\, \hat{\mathbf{h}}_{t|t-1}^{(K)}]$. Then solve (26) to obtain $\hat{\mathbf{q}}_t$ via CA which yields a local optimum of (26).
5. Form
$$\hat{\mathbf{B}}_t := \Big[\hat{q}_t^{(1)}\mathbf{I}_M \;|\quad \hat{q}_t^{(2)}\mathbf{I}_M \;|\quad \cdots \quad|\quad \hat{q}_t^{(K)}\mathbf{I}_M\Big]$$
and run the correction step as
$$\mathbf{K}_t := \mathbf{P}_{t|t-1}\hat{\mathbf{B}}_t^T\Big(\hat{\mathbf{B}}_t\mathbf{P}_{t|t-1}\hat{\mathbf{B}}_t^T + \mathbf{R}_t\Big)^{-1},$$
$$\hat{\mathbf{h}}_{t|t} = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \hat{\mathbf{B}}_t\hat{\mathbf{h}}_{t|t-1}), \qquad \mathbf{P}_{t|t} = \big(\mathbf{I} - \mathbf{K}_t\hat{\mathbf{B}}_t\big)\mathbf{P}_{t|t-1}.$$
6. Set $t \leftarrow t+1$ and go to step 2.

At every MLE iteration $\mathbf{C_{ee}}$ should be inverted which amounts to a complexity of $\mathbf{O}(M^3)$. At worst-case, MLE tracker needs 2^K iterations per time step. Hence, its worst-case complexity is of the order $\mathbf{O}(M^3 2^K)$ per time step which is similar to target-tracking based methods. However, simulations suggest that MLE tracker converges much faster than 2^K iterations and its average complexity may be polynomial in K as well as in M.

6. Performance analysis

We will examine the MSE difference between the optimal coordinated versus uncoordinated filters which offers the fundamental performance limit. This gap is referred to as price of anarchy (PoA) because it determines the increased MSE when users change pilot access patterns at will and do not notify the BS. For an ordinary KF, covariance update is independent of measurements and can be carried offline. Proceeding with the JC-KF, we have

$$\begin{aligned}
\mathrm{MSE}_c &= \mathbb{E}\Big[\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(c)}\|^2\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{E}\Big[\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(c)}\|^2|\mathbf{q}_{1:t},\mathbf{y}_{1:t}\Big]|\mathbf{q}_{1:t}\Big]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))|\mathbf{q}_{1:t}\Big]\Big] \\
&= \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t})\ \mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))
\end{aligned} \tag{27}$$

The inner most expectation, which is conditioned on $\mathbf{q}_{1:t}, \mathbf{y}_{1:t}$ amounts to JC-KF and its MMSE is easily obtained by the trace of the KF covariance matrix tailored to that particular pattern of $\mathbf{q}_{1:t}$ and is independent of $\mathbf{y}_t$. Each measurement pattern $\mathbf{q}_{1:t}$ yields a corresponding MSE given by $\mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))$. The overall MSE, averaged over $\mathbf{q}_{1:t}$ is given by the last equation.

Let us focus on the uncoordinated optimum tracker and its MSE:

$$\mathrm{MSE}_u = \mathbb{E}\Big[\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2|\mathbf{y}_{1:t}\Big]\Big] \tag{28}$$

Given $\mathbf{y}_{1:t}$, $\hat{\mathbf{h}}_{t|t}^{(u)}$ is only a function of measurements and hence non-random. Only $\mathbf{h}_t$ is random in the inner expectation and distributed as the Gaussian mixture in (12). One can write

$$\mathbb{E}\Big[\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2|\mathbf{y}_{1:t}\Big] \tag{29}$$

$$= \int \|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2 \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})\mathcal{N}(\mathbf{h}_t;\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}),\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))\ d\mathbf{h}_t$$

$$\begin{aligned}
&= \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t}) \int \|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) + \hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2 \\
&\quad \times\mathcal{N}(\mathbf{h}_t;\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}),\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))\ d\mathbf{h}_t
\end{aligned} \tag{30}$$

$$= \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})\Big[\mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t})) + \|\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2\Big]. \tag{31}$$

In going from (30) to (31), we have expanded the $\|.\|^2$ and used the fact that we are integrating with respect to $\mathcal{N}(\mathbf{h}_t;\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}),\mathbf{P}_{t|t}(\mathbf{q}_{1:t}))$ and hence the term $\|\mathbf{h}_t - \hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t})\|^2$ will be given by the trace of the corresponding covariance matrix. The term $\|\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2$ is constant and the cross-term is zero. Next, we replace (29) into (28) to obtain the final MSE

$$\begin{aligned}
\mathrm{MSE}_u &= \mathbb{E}_{\mathbf{y}_{1:t}}\Bigg[\sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})\Big[\mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t})) \\
&\qquad + \|\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2\Big]\Bigg] \\
&= \sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t})\ \mathrm{trace}(\mathbf{P}_{t|t}(\mathbf{q}_{1:t})) \\
&\qquad + \mathbb{E}_{\mathbf{y}_{1:t}}\Bigg[\sum_{\mathbf{q}_{1:t}} p(\mathbf{q}_{1:t}|\mathbf{y}_{1:t})\|\hat{\mathbf{h}}_{t|t}(\mathbf{q}_{1:t}) - \hat{\mathbf{h}}_{t|t}^{(u)}\|^2\Bigg]
\end{aligned} \tag{32}$$

The first term in (32) equals (27). The second term yields MSE difference between the optimal coordinated and uncoordinated trackers. It is the price of anarchy (PoA). It is desirable to characterize PoA analytically such as proving its boundedness and obtaining an upper bound that tells us how much performance we lose due to lack of coordination. These questions are difficult to address generally. However, for some special cases, they can be answered. First, we derive a single step upper bound on PoA. That is we assume both filters used the same initial estimate $\hat{\mathbf{h}}_0, \hat{\mathbf{P}}_{0,0}$ and check the negative effect of anarchy at time-step $t = 1$. The following theorem ensues whose proof is relegated to the Appendix B.

Theorem 1. *Assuming the same initial estimate $\hat{\mathbf{h}}_0, \hat{\mathbf{P}}_{0,0}$ for both the coordinated and uncoordinated optimum trackers, the single-step PoA is upper bounded by*

$$\text{PoA} \leqslant \sum_{\mathbf{q}_1} p(\mathbf{q}_1) \Bigg[\hat{\mathbf{h}}_{1|0}^T \mathbf{B}(\mathbf{q}_1)^T \mathbf{K}(\mathbf{q}_1)^T \mathbf{K}(\mathbf{q}_1) \mathbf{B}(\mathbf{q}_1) \hat{\mathbf{h}}_{1|0} \tag{33}$$

$$+\text{trace}\big(\mathbf{R} + \mathbf{B}(\mathbf{q}_1)\mathbf{P}_{1|0}\mathbf{B}(\mathbf{q}_1)^T\big) \tag{34}$$

$$- \sum_{\tilde{\mathbf{q}}_1} P(\tilde{\mathbf{q}}_1) \hat{\mathbf{h}}_{1|0}^T \mathbf{B}(\mathbf{q}_1)^T \mathbf{K}(\mathbf{q}_1)^T \mathbf{K}(\tilde{\mathbf{q}}_1) \mathbf{B}(\tilde{\mathbf{q}}_1) \hat{\mathbf{h}}_{1|0} \Bigg]. \tag{35}$$

6.1. Mean square (MS) sense stability of optimal trackers

Unfortunately, Theorem 1 can not determine if PoA is bounded for large t because the presented upper bound grows unbounded as t increases. First, we consider a dynamical system which is stable. In the time-invariant parameters case, this means that all eigenvalues of $\mathbf{A}_t := \mathbf{A}$ have absolute value less than one. The following theorem ensues. The proofs are trivial [1].

Theorem 2. *When the aggregate (or joint) dynamical system in (5) is time-invariant and stable in the MS sense, following conclusions can be drawn:*

1. *Covariance matrix for the state remains bounded and converges to the unique positive semi-definite solution of Lyapunov equation $\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q}$.*
2. *JC-KF is MS sense stable meaning that its covariance matrix remain bounded.*
3. *Optimal uncoordinated tracker is MS sense stable meaning that its covariance matrix remain bounded.*
4. *PoA remains bounded.*

When the dynamical system of the state is stable, optimal filters are guaranteed to be stable without the need for any observability condition. Stability is guaranteed by the fact that optimal MMSE tracker will have a smaller trace of covariance than the unobserved dynamical system as exploiting measurements optimally can only improve MSE. As the original covariance for the unobserved system is bounded so does the MSE for the optimal tracker [1]. Subsequently, PoA which is the difference between MSE of optimal trackers will be bounded as well.

It is well-known that if a time-varying dynamical system is MS sense unstable but uniformly completely observable and controllable, the corresponding Kalman filter is guaranteed to be stable [1,14,16]. This is a sufficient condition but not necessary as weaker detectability condition can be used instead. Our final theorem yields sufficient conditions for the stability of JC-KF.

Theorem 3. *If the dynamical system in (5) is MS sense unstable but uniformly completely controllable, then JC-KF is MS sense stable if the prior $p(\mathbf{q}_{1:t})$ is non-zero only for those $\mathbf{q}_{1:t}$ combinations which maintain uniform complete observability.*

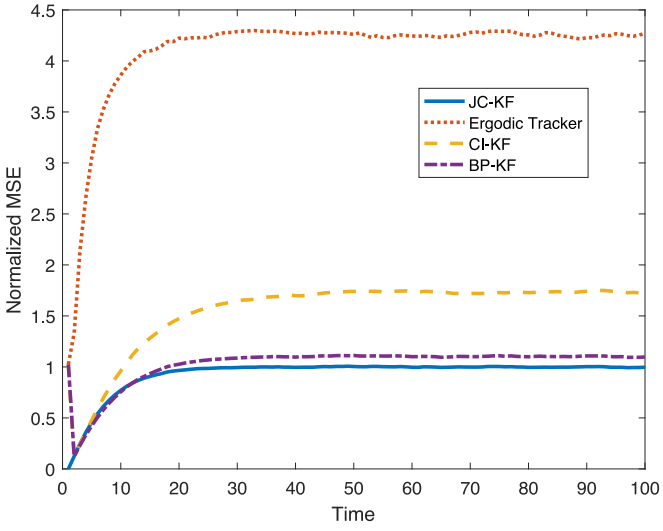

Fig. 1. Normalized MSE of coordinated trackers with $K = 6$, $M = 256$.

Proof. Results trivially from (27). □

Given the condition in Theorem 3, the optimal coordinated tracker is stable. It is of interest to determine the stability or lack of it for the optimal uncoordinated tracker under the same conditions. This will lead to boundedness / unboundedness conclusion for PoA when observability/controllability conditions hold. However, this is a challenging task and demands further investigation on its own. We leave it as an open problem to be addressed in future works.

7. Numerical results

We pursue four distinct goals in our simulations. First, we draw a comparison between coordinated and uncoordinated schemes to determine PoA numerically. Secondly, we check the effect of number of antennas on performance to decide how an ordinary MIMO fairs against massive MIMO. Third, we check the data rates that can be achieved via the proposed tracking schemes. Finally, we compare against existing alternatives. Certain parameters are fixed for all simulations. They include $\tau = 16$, $T = 100$, $\rho = 0.95$, $\mathbf{Q}_t^{(k)} = (1-\rho^2)\mathbf{I}_M$, and $\mathbf{A}_t^{(k)} = \rho\mathbf{I}_M$ for all t, k. Furthermore, we set $\mathbf{R}_t = \mathbf{I}_M$, $\lambda_k = (K-1)/K$ for all $k = 1, 2, \ldots, K$. While we consider setups with $K = 6$ devices, we only plot the results pertaining to first device as all parameters are selected symmetrically and channels are selected independently and at random for all devices. We normalize the channel mean-square error (MSE) for all the trackers by the trace of the covariance matrix $(\mathbf{P}_{t|t})_{1,1}$ which corresponds to the optimal coordinated tracker given by JC-KF.

7.1. Performance of coordinated methods

We consider $K = 6$ and plot the normalized MSE versus timeslot in Fig 1 for $M = 256$ antennas. We obtained the same trend for $M = 16$. This figure depicts a comparison of JC-KF, CI-KF, and BP-KF. As expected, JC-KF performs best with a normalized MSE of one. BP-KF performs slightly worse than JC-KF and given its lower complexity offers a desirable complexity-performance trade-off. As for CI-KF, given the large activation ratio $\lambda = 5/6$, more than one device are active in almost all time slots. Thus, CI-KF discards all measurements and is always predicting which amounts to dead reckoning and its performance deteriorates over time as expected. Ergodic tracker which is the algorithm proposed by [25] performs poorly compared to all proposed methods. The main reason be-

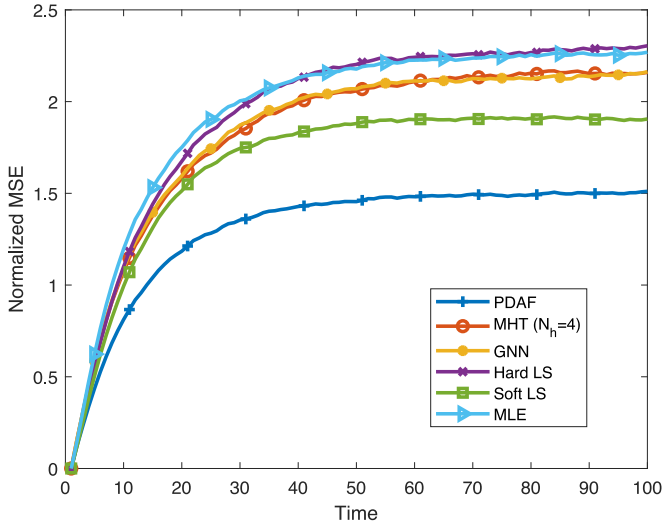

Fig. 2. Normalized MSE of uncoordinated trackers with $K = 6$, $M = 16$.

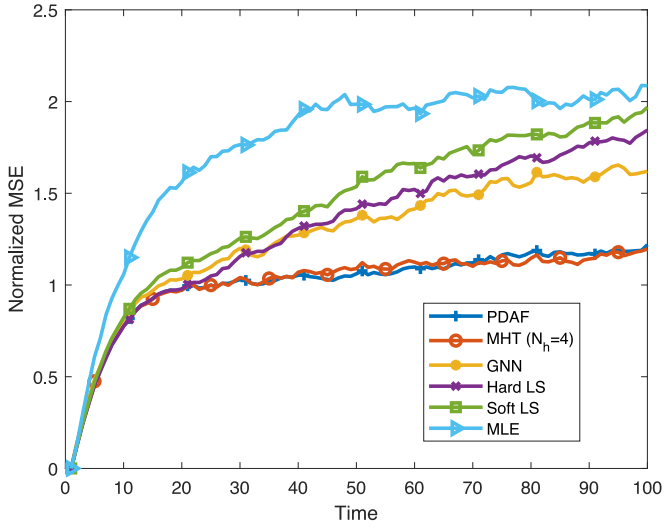

Fig. 3. Normalized MSE of uncoordinated trackers with $K = 6$, $M = 150$.

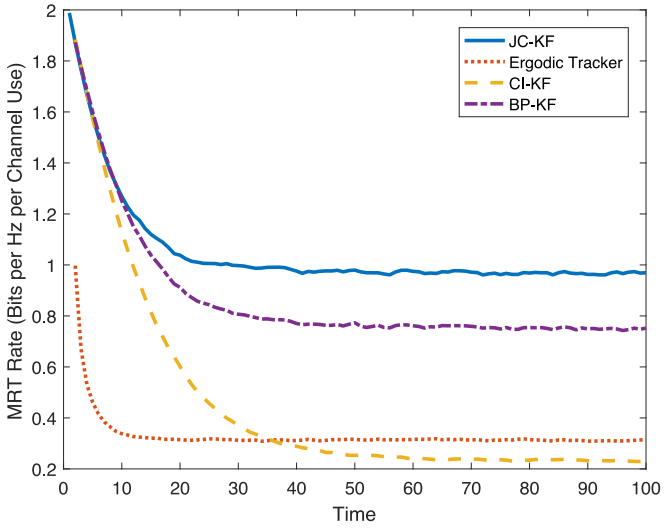

Fig. 4. MRT data rate of coordinated trackers with $K = 6$, $M = 16$.

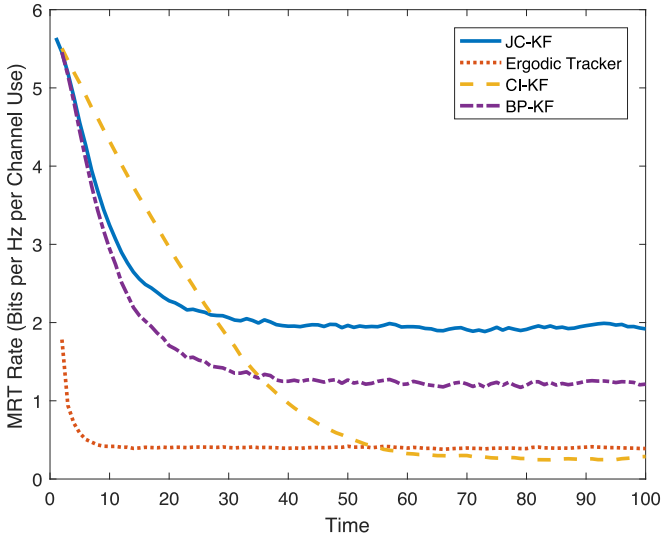

Fig. 5. MRT data rate of coordinated trackers with $K = 6$, $M = 256$.

hind its poor performance can be attributed to the fact that Ergodic tracker does not explicitly account for interference from other users and simply treats them as noise.

7.2. Performance of uncoordinated methods

For $K = 6$ devices and $M = 16$, $M = 150$ antennas, Normalized MSE (NMSE) for all uncoordinated trackers are plotted in Fig. 2 and Fig. 3 respectively. The optimal uncoordinated tracker is too complex to be practical and thus overlooked. As for PoA, it is definitely lower than the best performing uncoordinated algorithm which is suggested to be PDAF by the figures. A major difference emerges between $M = 16$ and $M = 256$. For $M = 16$, ordinary MIMO, NMSE is about 1.5 for PDAF which means that worst case PoA equals half the coordinated MMSE. The PDAF NMSE for $M = 150$ is about 1.1 which means that worst case PoA is about one tenth of the coordinated MMSE. This result is remarkable in the sense that if we go into the massive MIMO regime, a practical uncoordinated algorithm like PDAF gets very close to the performance of optimum coordinated tracker. However, the gap is considerably larger for ordinary MIMO. Same conclusion is valid for all other uncoordinated trackers as their NMSE is considerably smaller in the massive MIMO regime. Fig. 2 suggests that Soft LS, MHT (4 best hypotheses), and GNN perform slightly better than Hard LS and MLE. For the massive MIMO setup in Fig. 3, PDAF and MHT perform best and close to one another. Afterwards, GNN performs better than hard/soft LS and MLE. MLE performs the poorest and this weak performance can be attributed to non-convexity of ML objective. CA on MLE can only reach a local optimum which is revealed via this simulation to maintain a worse performance than LS global optimum.

7.3. Achievable data rates: MRT

The final aim in any communication system is to endow each individual user with the highest possible data-rate that can be secured with the available CSI. Here, we use a simple MRT beamformer which selects the beamforming weights to match the channel estimates obtained through various algorithms. Fig. 4 and Fig. 5 show a comparison of the JC-KF, CI-KF, and BP-KF with respect to the capacity for user one for the cases of $K = 6$, $M = 16$ and $K = 6$, $M = 256$, respectively. It can be observed that joint tracking of all K users does indeed greatly improve the data rate compared to decoupled sub-optimal individual trackers. CI-KF performs dead reckoning as mentioned before and its performance deterio-

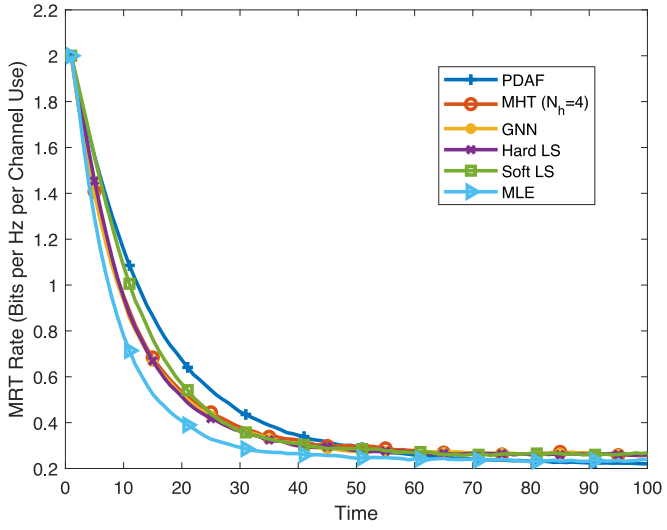

Fig. 6. MRT data rate of uncoordinated trackers with $K = 6,\ M = 16$.

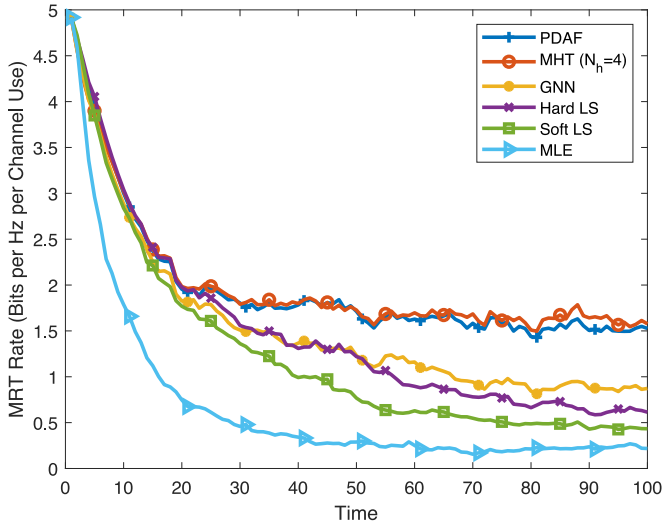

Fig. 7. MRT data rate of uncoordinated trackers with $K = 6,\ M = 150$.

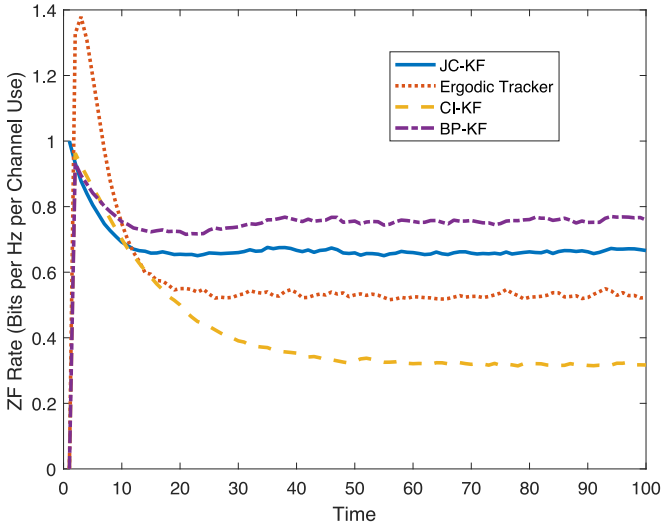

Fig. 8. ZF data rate of coordinated trackers with $K = 6,\ M = 16$.

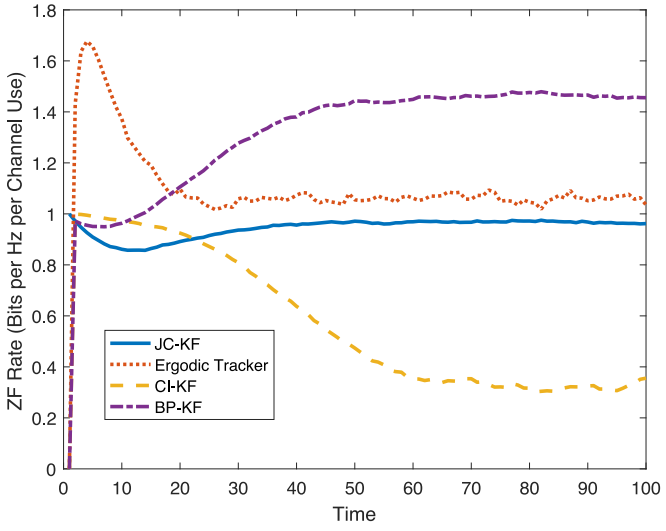

Fig. 9. ZF data rate of coordinated trackers with $K = 6,\ M = 256$.

rates over time. Ergodic tracker also performs poorly given its high MSE. It is observed that BP-KF can achieve a satisfactory rate. Given its lower complexity compared to JC-KF, BP-KF strikes a desirable trade-off. JC-KF can obtain a rate twice as high in the massive regime compared to ordinary MIMO per user.

MRT beamformer data rates for uncoordinated trackers are plotted in Fig. 6 and Fig. 7 for $K = 6,\ M = 16$ and $K = 6,\ M = 150$ respectively. It is observed that for ordinary MIMO there is no significant gap between data rates of various algorithms and they all perform poorly. However, in the massive MIMO setting, performance gap is considerable with PDAF and MHT performing best and locally optimal ML performing worst. In ordinary MIMO, achievable rates decay quickly from 2 bits per Hz per channel use to 0.2 bits per Hz per channel use for all trackers which is an unsatisfactory rate. This is a 10 fold decrease in rate. However, in the massive MIMO of Fig. 7 the rate for best performing algorithms, which are PDAF and MHT, suffers a 3 fold decrease from 5 bits per Hz per channel use to 1.5 bits per Hz per channel use. Thus, we conclude that the combination of massive MIMO and uncoordinated tracking offers a practical and affordable solution to address pilot contamination with the added flexibility that BS needs not even know the pilot transmission patterns of various devices. Unfortunately, this reasonable performance comes at a cost of still considerable complexity. Low-complexity alternatives are not that efficient and the best of the three which is Hard LS achieves a rate of 0.5 bits per Hz per channel use which is a 10 fold decrease from perfect CSI case. Further investigation of low-complexity alternatives to PDAF and MHT are well-warranted in massive MIMO

7.4. Achievable data rates: zero-forcing

To study the effect of various beamformers on data rate, zero-forcing (ZF) beamformer has been considered for coordinated trackers with $M = 16, 256$ antennas and uncoordinated trackers with $M = 16, 150$ antennas. The results are plotted in Figs. 8, 9, 10, 11 respectively. Given the very nonlinear fashion in which CSI error propagate in ZF, the results are very different than MRT. For ordinary MIMO, BP-KF performs best followed by JC-KF and Ergodic tracker. We know that JC-KF incurs the lowest MSE. Yet, it does not boast the highest rate. This can make for another justification on why rate should also be considered alongside MSE. A comparison of Figs. 4 and 8 reveals that the best procured rate is lower for ZF compared to MRT. This can be attributed to the fact that orthogonality of ZF beamformer breaks down in the presence of CSI error and hence it can not operate as robustly as MRT. For massive

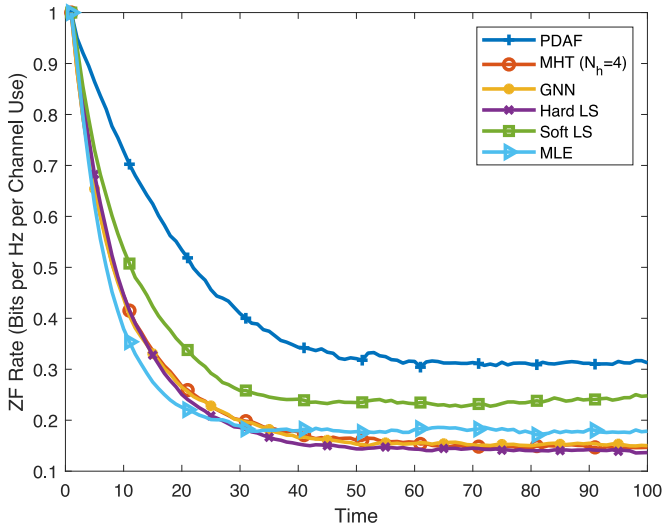

Fig. 10. ZF data rate of uncoordinated trackers with $K = 6$, $M = 16$.

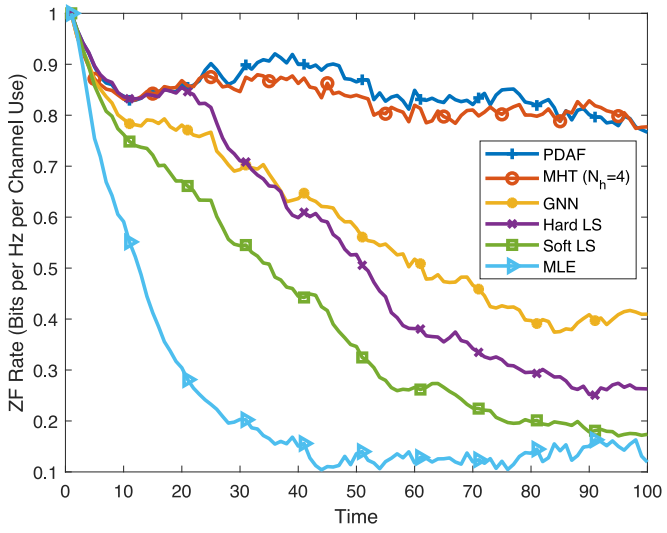

Fig. 11. ZF data rate of uncoordinated trackers with $K = 6$, $M = 150$.

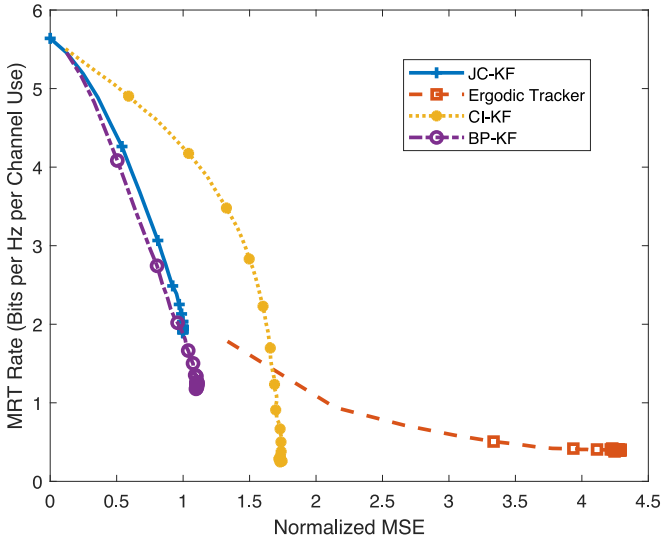

Fig. 12. MRT data rate versus MSE for coordinated trackers with $M = 256$.

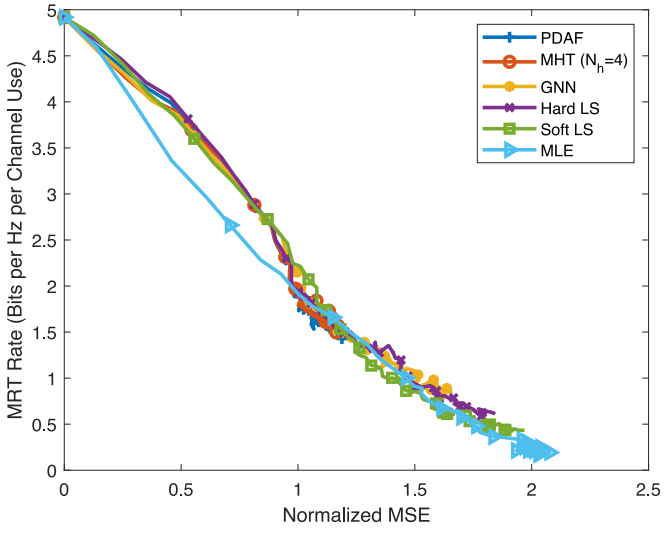

Fig. 13. MRT data rate versus MSE of uncoordinated trackers with $M = 150$.

MIMO of Fig. 9 it is observed that Ergodic tracker performs satisfactorily when ZF is employed and it is indeed better than JC-KF. Again, it is worth mentioning that JC-KF is the optimum tracker in terms of MSE. Yet, its MSE gains fail to show in ZF data rates. Similar to ZF with ordinary MIMO, BP-KF performs the best. All uncoordinated trackers perform poorly as in Fig. 10 while PDAF and MHT offer the best performance in the massive MIMO regime of Fig. 11.

7.5. MRT rate versus MSE

It is very interesting to determine how rate changes versus MSE. This relation is both nonlinear and not one-to-one because both the size of error (MSE) as well as its direction are important in determining the achievable rate. Figs. 12 and 13 plot MRT rate versus MSE for coordinated and uncoordinated scenarios with $M = 256$ and $M = 150$ antennas respectively.

7.6. Complexity analysis and PoA bound

Finally, we compare the performance of running various algorithms in practice. All codes were written in MATLAB and run on a Laptop with Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz processors and 8 Giga Bytes of RAM. The per time-slot complexity of various coordinated algorithms versus number of antennas are plotted in Fig. 14. Similarly, per time-slot complexity of various uncoordinated algorithms are plotted in Fig. 15. Complexity of heuristic methods of Section 5 are plotted separately in Fig. 16. A comparison of Figs. 14 and 15 reveals that lack of coordination with BS in transmitting pilots can incur a large overhead. Specifically, MHT and PDAF incur an overhead 30 times that of optimum coordinated tracker or JC-KF for $M = 150$ in order to also detect collision patterns. Note the difference in x-axis scale between the two figures. JC-KF itself is very complex compared to low-complexity coordinated trackers which enjoy a less than 10 fold decrease in complexity for $M = 256$ antennas. However, the complexity gap is considerably smaller between various algorithms when M is small. Fig. 16 reveals that at a cost in performance, uncoordinated tracking's complexity can be reduced to that of JC-KF. The performance loss can be severe in ordinary MIMO as depicted in previous figures. Theorem 1 is investigated numerically in Fig. 17 where the exact PoA is plotted versus its derived upper-bound. Both are normalized by the square of true channel gains. It can be observed

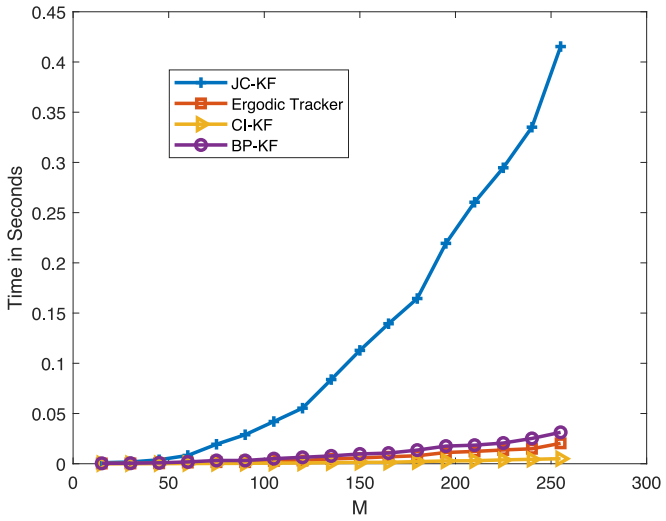

Fig. 14. Complexity of coordinated trackers versus M.

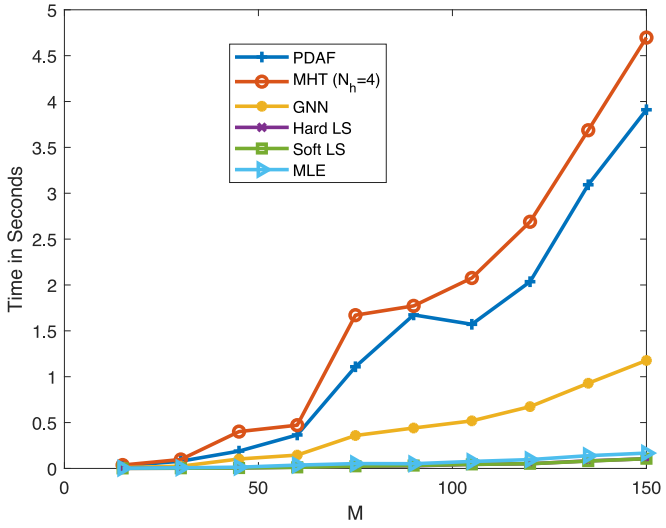

Fig. 15. Complexity of uncoordinated trackers versus M.

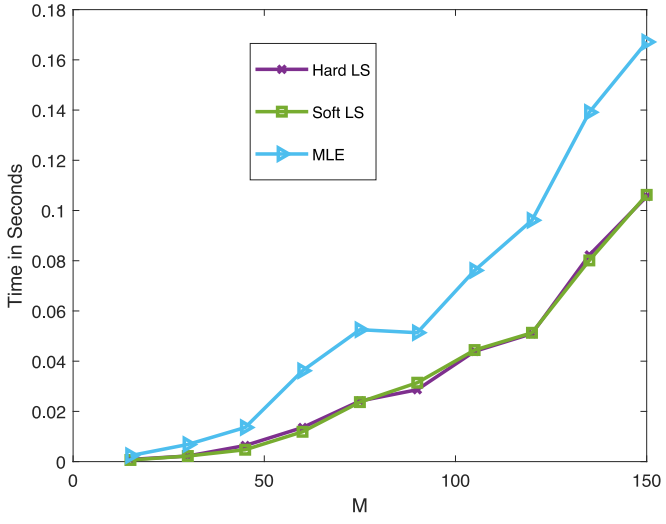

Fig. 16. Complexity of heuristic trackers in Section 5 versus M.

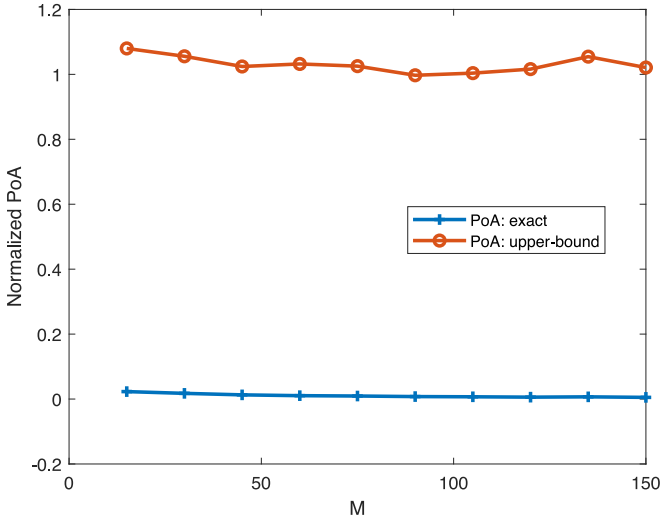

Fig. 17. Exact PoA versus its upper bound in Theorem 1.

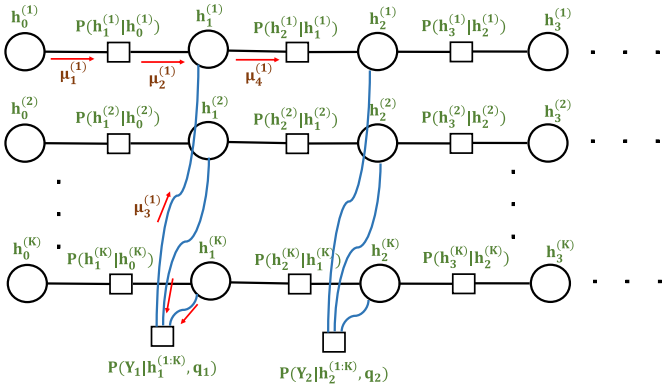

Fig. 18. Factor graph for our filtering scenario and belief schedules.

that the bound is not very tight and this motivates future research to obtain tighter bounds.

8. Conclusion

High-rate IoT setup in massive MIMO was considered and a novel method to track all devices channels simultaneously was advocated. Utilizing a dynamical model for IoT devices channel evolution over time, optimal and various sub-optimal trackers were proposed for coordinated and uncoordinated scenarios. Fundamental performance gap between coordinated and uncoordinated trackers was evaluated analytically. Finally, the performance of various trackers were investigated through extensive simulations.

Declaration of Competing Interest

The authors declare they have no conflict of interests.

Appendix A. Derivation of BP-KF

To derive BP-KF, we use the factor graph notion as provided in [18]. In their seminal paper, it was shown that KF can alternatively be viewed as belief propagation, or more generally a message passing algorithm on a factor graph. Following the same procedure, in the coordinated setup, factor graph for our scheme is depicted in Fig. 18.

To proceed, we derive the four messages, or beliefs, denoted as $\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}, \mu_4^{(1)}$ in Fig. 18. Then, we generalize these messages

to other users and future time slots. Note that given the joint Gaussianity of the overall model, belief messages are Gaussian themselves. Therefore, we only need to include the mean and covariance in the beliefs.

$$\begin{aligned}
\mu_1^{(1)}(\mathbf{h}_0^{(1)}) &= p(\mathbf{h}_0^{(1)}) = \mathcal{N}(\mathbf{h}_0^{(1)}; \hat{\mathbf{h}}_0^{(1)}, \mathbf{P}_{00}^{(1)}) \\
\mu_2^{(1)}(\mathbf{h}_1^{(1)}) &= \int p(\mathbf{h}_1^{(1)}|\mathbf{h}_0^{(1)})\mu_1^{(1)}(\mathbf{h}_0^{(1)})\ d\mathbf{h}_0^{(1)} \\
&= \int p(\mathbf{h}_1^{(1)}|\mathbf{h}_0^{(1)})p(\mathbf{h}_0^{(1)})\ d\mathbf{h}_0^{(1)} \\
&= \int \mathcal{N}(\mathbf{h}_1^{(1)}; \mathbf{A}_1^{(1)}\mathbf{h}_0^{(1)}, \mathbf{Q}_1^{(1)})\ \mathcal{N}(\mathbf{h}_0^{(1)}; \hat{\mathbf{h}}_0^{(1)}, \mathbf{P}_{00}^{(1)})\ d\mathbf{h}_0^{(1)} \\
&= \mathcal{N}\Big(\mathbf{h}_1^{(1)}; \mathbf{A}_1^{(1)}\hat{\mathbf{h}}_0^{(1)}, \mathbf{A}_1^{(1)}\mathbf{P}_{00}^{(1)}\mathbf{A}_1^{(1)^T} + \mathbf{Q}_1^{(1)}\Big) : \\
&= \mathcal{N}\Big(\mathbf{h}_1^{(1)}; \hat{\mathbf{h}}_{1|0}^{(1)}, \mathbf{P}_{10}^{(1)}\Big)
\end{aligned}$$

These beliefs are exactly same as those for ordinary KF. In evaluating $\mu_3^{(1)}, \mu_4^{(1)}$ differences appear:

$$\begin{aligned}
\mu_3^{(1)}(\mathbf{h}_1^{(1)}) &= \int p(\mathbf{y}_1|\mathbf{h}_1, \mathbf{q}_1) \prod_{j:q_1^{(j)}=1,\ j\neq 1,} \mu_2^{(j)}(\mathbf{h}_1^{(j)}) \quad d\mathbf{h}_1^{(j)} \\
&= \int \mathcal{N}\Big(\mathbf{y}_1; \sum_{k=1}^{K} q_1^{(k)}\mathbf{h}_1^{(k)}, \mathbf{R}_1\Big) \\
&\quad \prod_{j:q_1^{(j)}=1,\ j\neq 1,} \mathcal{N}\Big(\mathbf{h}_1^{(j)}; \hat{\mathbf{h}}_{1|0}^{(j)}, \mathbf{P}_{10}^{(j)}\Big) \quad d\mathbf{h}_1^{(j)} \\
&= \mathcal{N}\Big(\mathbf{y}_1 - \sum_{j=2}^{K} q_1^{(j)}\hat{\mathbf{h}}_{1|0}^{(j)}; \mathbf{h}_1^{(1)}, \mathbf{R}_1 + \sum_{j=2}^{K} q_1^{(j)}\hat{\mathbf{P}}_{1|0}^{(j)}\Big) : \\
&= \mathcal{N}\big(\hat{\mathbf{y}}_1^{(1)}; \mathbf{h}_1^{(1)}, \hat{\mathbf{R}}_1^{(1)}\big)
\end{aligned}$$

The above expression is valid only if $q_1^{(1)} = 1$, that is user 1 participates in the collision. Otherwise, there is no connection between variable node $\mathbf{h}_1^{(1)}$ and factor node $p(\mathbf{y}_1|\mathbf{h}_1, \mathbf{q}_1)$ and there will be no $\mu_3^{(1)}$ and we will have $\mu_4^{(1)} = \mu_2^{(1)}$. Finally, $\mu_4^{(1)}$ is computed as in an ordinary KF.

$$\begin{aligned}
\mu_4^{(1)}(\mathbf{h}_1^{(1)}) &= \mu_2^{(1)}(\mathbf{h}_1^{(1)})\mu_3^{(1)}(\mathbf{h}_1^{(1)}) \\
&= \mathcal{N}\Big(\mathbf{h}_1^{(1)}; \hat{\mathbf{h}}_{1|0}^{(1)}, \mathbf{P}_{10}^{(1)}\Big)\mathcal{N}\big(\hat{\mathbf{y}}_1^{(1)}; \mathbf{h}_1^{(1)}, \hat{\mathbf{R}}_1^{(1)}\big) \\
&= \mathcal{N}\Big(\mathbf{h}_1^{(1)}; \hat{\mathbf{h}}_{1|1}^{(1)}, \mathbf{P}_{1|1}^{(1)}\Big)
\end{aligned}$$

where the parameters on the last line are given by (11). Same derivation can be extended in a straightforward manner to other users and future time slots which completes the proof.

Appendix B. Proof of Theorem 1

$$\begin{aligned}
\text{PoA} &= \mathbb{E}_{\mathbf{y}_1}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\|\hat{\mathbf{h}}_{1|1}(\mathbf{q}_1) - \hat{\mathbf{h}}_{1|1}^{(u)}\|^2\Big] \\
&= \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big(\|\hat{\mathbf{h}}_{1|1}(\mathbf{q}_1)\|^2 + \|\hat{\mathbf{h}}_{1|1}^{(u)}\|^2 - 2\hat{\mathbf{h}}_{1|1}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|1}^{(u)}\Big)\Big] \\
&= \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big(\|\hat{\mathbf{h}}_{1|1}(\mathbf{q}_1)\|^2 - \|\hat{\mathbf{h}}_{1|1}^{(u)}\|^2\Big)\Big] \qquad (36)
\end{aligned}$$

Indeed, $\hat{\mathbf{h}}_{1|1}(\mathbf{q}_1)$ amounts to a correction step assuming $\mathbf{q}_1$ is the true model, while $\hat{\mathbf{h}}_{1|1}$ is the soft combination of all possible $\mathbf{q}_1$ as derived for the optimal uncoordinated tracker. They can further be written as

$$\begin{aligned}
\hat{\mathbf{h}}_{1|1}(\mathbf{q}_t) &= \hat{\mathbf{h}}_{1|0} + \mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big) \\
\hat{\mathbf{h}}_{1|1}^{(u)} &= \sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big[\hat{\mathbf{h}}_{1|0} + \mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_{1:t} - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big] \\
&= \hat{\mathbf{h}}_{1|0} + \sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big) \qquad (37)
\end{aligned}$$

Applying this to the PoA in (36), we get

$$\begin{aligned}
\text{PoA} &= \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_t|\mathbf{y}_1)\Big(\hat{\mathbf{h}}_{1|0}^T\mathbf{K}(\mathbf{q}_1)\big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\big)\Big)\Big] \\
&\quad + \mathbb{E}\Big[\sum_{\mathbf{q}_t} p(\mathbf{q}_1|\mathbf{y}_1)\Big\|\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big\|^2\Big] \\
&\quad - \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\hat{\mathbf{h}}_{1|0}^T\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big] \\
&\quad - \mathbb{E}\Big[\Big\|\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big\|^2\Big] \\
&= \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big\|\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big\|^2\Big] \\
&\quad - \mathbb{E}\Big[\Big\|\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big\|^2\Big] \qquad (38)
\end{aligned}$$

It is well-known that $\|.\|^2$ is convex, ($\frac{\partial\|X\|^2}{\partial X^2} = 2I$). Hence, Jensen inequality ensure that 38 is positive since it is equal to the expectation of positive entity ($\|\sum_i \lambda_i X_i\|^2 \leqslant \sum_i \lambda_i \|X_i\|^2$).

First term in 38 yields

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big\|\mathbf{K}(\mathbf{q}_1)\Big(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\Big)\Big\|^2\Big] \\
&= \mathbb{E}\Big[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\Big(\mathbf{y}_1^T\mathbf{K}^T(\mathbf{q}_1)\mathbf{K}(\mathbf{q}_1)\mathbf{y}_1 + \hat{\mathbf{h}}_{1|0}^T\mathbf{B}(\mathbf{q}_1{}^T)\mathbf{K}(\mathbf{q}_1{}^T) \\
&\quad \mathbf{K}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\mathbf{B}(\mathbf{q}_1) - 2\hat{\mathbf{h}}_{1|0}\mathbf{B}(\mathbf{q}_1{}^T)\mathbf{K}(\mathbf{q}_1)^T\mathbf{K}(\mathbf{q}_1)\Big)\Big] \qquad (39)
\end{aligned}$$

Now, let us evaluate each term independently.

$$\begin{aligned}
\mathbb{E}[p(\mathbf{q}_1|\mathbf{y}_1)] &= \int p(\mathbf{q}_1|\mathbf{y}_1)\ p(\mathbf{y}_1)d\mathbf{y}_1 = P(\mathbf{q}_1) \\
\mathbb{E}[p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{y}_1] &= \int \mathbf{y}_1\ p(\mathbf{y}_1|\mathbf{q}_1)\ p(\mathbf{q}_1)\ \ d\mathbf{y}_1 \\
&= \int\int \mathbf{y}_1\ p(\mathbf{y}_1|\mathbf{h}_1, \mathbf{q}_1)\ p(\mathbf{h}_1|\mathbf{q}_1)\ p(\mathbf{q}_1)\ \ d\mathbf{y}_1\ \ d\mathbf{h}_1 \\
&= p(\mathbf{q}_1)\ \int p(\mathbf{h}_1)\ \int \mathbf{y}_1\ \mathcal{N}(\mathbf{y}_1; \mathbf{B}(\mathbf{q}_1)\mathbf{h}_1, \mathbf{R}_1)\ \ d\mathbf{y}_1\ \ d\mathbf{h}_1 \\
&= p(\mathbf{q}_1)\ \int \mathbf{B}(\mathbf{q}_1)\mathbf{h}_1 p(\mathbf{h}_1)\ \ d\mathbf{h}_1 = p(\mathbf{q}_1)\ \mathbf{B}(\mathbf{q}_1)\ \mathbb{E}[\mathbf{h}_1] = 0 \qquad (40)
\end{aligned}$$

The last line is zero because $\mathbb{E}[\mathbf{h}_1] = \mathbb{E}[\mathbf{A}\mathbf{h}_0 + \mathbf{u}_1] = 0$. Let us focus on the second moment now.

$$\begin{aligned}
&\mathbb{E}\big[p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{y}_1\mathbf{y}_1^T\big] \\
&= p(\mathbf{q}_1)\int p(\mathbf{h}_1)\int \mathbf{y}_1\ \mathbf{y}_1^T\ \mathcal{N}(\mathbf{y}_1; \mathbf{B}(\mathbf{q}_1)\mathbf{h}_1, \mathbf{R}_1)\ d\mathbf{y}_1\ d\mathbf{h}_1 \\
&= p(\mathbf{q}_1)\int p(\mathbf{h}_1)\big[\mathbf{B}(\mathbf{q}_1)\ \mathbf{h}_1\mathbf{h}_1^T\ \mathbf{B}(\mathbf{q}_1)^T + \mathbf{R}\big]\ \ d\mathbf{h}_1
\end{aligned}$$

$$= p(\mathbf{q}_1)\ \left[\mathbf{R} + \mathbf{B}(\mathbf{q}_1)\ \mathbb{E}\left[\mathbf{h}_1\mathbf{h}_1^T\right]\ \mathbf{B}(\mathbf{q}_1)^T\right]$$

$$\mathbb{E}\left[\mathbf{h}_1\mathbf{h}_1^T\right] = \mathbb{E}\left[(\mathbf{A}\mathbf{h}_0 + \mathbf{u}_1)(\mathbf{A}\mathbf{h}_0 + \mathbf{u}_1)^T\right] = \mathbf{A}\mathbf{P}_0\mathbf{A}^T + \mathbf{Q} \tag{41}$$

Combining these results, the first term in 38 can be written as follows

$$\mathbb{E}\left[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\left\|\mathbf{K}(\mathbf{q}_1)\left(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\right)\right\|^2\right]$$
$$= \sum_{\mathbf{q}_1} p(\mathbf{q}_1)\hat{\mathbf{h}}_0\mathbf{A}^T\mathbf{B}(\mathbf{q}_1{}^T)\mathbf{K}(\mathbf{q}_1{}^T) \times \mathbf{K}(\mathbf{q}_1)\mathbf{B}(\mathbf{q}_1)\mathbf{A}\hat{\mathbf{h}}_0$$
$$+ \sum_{\mathbf{q}_1} p(\mathbf{q}_1)\mathrm{trace}\left[\mathbf{R} + \mathbf{B}(\mathbf{q}_t)\left(\mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q}\right)\mathbf{B}(\mathbf{q}_1)^T\right]$$

Second term in 38 can be simplified as follows

$$\mathbb{E}\left[\left\|\sum_{\mathbf{q}_t} p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{K}(\mathbf{q}_1)\left(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\right)\right\|^2\right]$$
$$\geqslant \left\|\mathbb{E}\left[\sum_{\mathbf{q}_1} p(\mathbf{q}_1|\mathbf{y}_1)\mathbf{K}(\mathbf{q}_1)\left(\mathbf{y}_1 - \mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\right)\right]\right\|^2$$
$$= \left\|\sum_{\mathbf{q}_1} p(\mathbf{q}_1)\mathbf{K}(\mathbf{q}_1)\mathbf{B}(\mathbf{q}_1)\hat{\mathbf{h}}_{1|0}\right\|^2$$
$$= \sum_{\mathbf{q}_1}\sum_{\tilde{\mathbf{q}}_1} p(\mathbf{q}_1)p(\tilde{\mathbf{q}}_1)\hat{\mathbf{h}}_{1|0}^T\mathbf{B}(\mathbf{q}_1)^T\mathbf{K}(\mathbf{q}_1)^T\mathbf{K}(\tilde{\mathbf{q}}_1)\mathbf{B}(\tilde{\mathbf{q}}_1)\hat{\mathbf{h}}_{1|0}$$
$$= \sum_{\mathbf{q}_1}\sum_{\tilde{\mathbf{q}}_1} p(\mathbf{q}_1)p(\tilde{\mathbf{q}}_1)\hat{\mathbf{h}}_0^T\mathbf{A}^T\mathbf{B}(\mathbf{q}_1)^T\mathbf{K}(\mathbf{q}_1)^T\mathbf{K}(\tilde{\mathbf{q}}_1)\mathbf{B}(\tilde{\mathbf{q}}_1)\mathbf{A}\hat{\mathbf{h}}_0$$

We used Jensen inequality on the first line. Replacing into (38) completes the proof.

CRediT authorship contribution statement

Hediyeh Soltanizadeh: Software, Formal analysis, Writing - original draft. **Shahrokh Farahmand:** Conceptualization, Software, Validation, Writing - review & editing, Supervision. **S. Mohammad Razavizadeh:** Methodology, Writing - review & editing, Project administration.

References

[1] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Dover Publications, 1979.
[2] Y. Bar-Shalom, F. Daum, J. Huang, The probabilistic data association filter, IEEE Control Syst. Mag. 29 (6) (2009) 82–100.
[3] Y.D. Beyene, R. Jantti, K. Ruttik, Random access scheme for sporadic users in 5g, IEEE Trans. on Wireless Communications 16 (3) (2017) 1823–1833.
[4] M. Biguesh, A.B. Gershman, Training-based MIMO channel estimation: A study of estimator trade-offs and optimal training signals, IEEE Trans. on Signal Processing 54 (3) (2006) 884–893.
[5] E. Björnson, E.d. Carvalho, J.H. Sørensen, E.G. Larsson, P. Popovski, A random access protocol for pilot allocation in crowded massive MIMO systems, IEEE Transactions on Wireless Communication 16 (4) (2017) 2220–2234.
[6] E. de Carvalho, E. Bjornson, J.H. Sorensen, P. Popovski, E.G. Larsson, Random access protocols for massive MIMO, IEEE Commun. Mag. 55 (9) (2017) 216–222.
[7] E. Casini, R.D. Gaudenzi, O.d.R. Herrero, Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks, IEEE Trans. on Wireless Communications 6 (4) (2007) 1408–1419.
[8] Z. Chen, F. Sohrabi, W. Yu, Sparse activity detection for massive connectivity, IEEE Trans. on Signal Processing 66 (7) (2018) 1890–1904.
[9] S. Dahiya, A.K. Singh, Channel estimation and channel tracking for correlated block-fading channels in massive MIMO systems, Digital Communications and Networks 4 (2) (2018) 138–147.
[10] Y. Du, B. Dong, X. Wang, Z. Liu, P. Gao, S. Li, Efficient multi-user detection for uplink grant-free NOMA: Prior information aided adaptive compressive sensing perspective, IEEE J. Sel. Areas Commun. 35 (12) (2017) 2812–2828.
[11] Y. Du, C. Cheng, B. Dong, Z. Chen, X. Wang, J. Fang, S. Li, Block-sparsity based multiuser detection for uplink grant-free NOMA, IEEE Trans. on Wireless Communications 17 (12) (2018) 7894–7909.
[12] X. Gao, L. Dai, W. Zhang, T. Xie, X. Dai, Z. Wang, Fast channel tracking for terahertz beamspace massive MIMO systems, IEEE Trans. Veh. Technol. 66 (7) (2017) 5689–5696.
[13] H. Han, Y. Li, X. Guo, A graph-based random access protocol for crowded massive MIMO systems, IEEE Transactions on Wireless Communication 16 (11) (2017).
[14] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Dover Publications, 1970.
[15] W.C. Jakes, Microwave Mobile Communications, 2nd Edition, Wiley-IEEE Press, 1994.
[16] T. Karvonen, Stability of Linear and non-Linear Kalman Filters, University of Helsinki, 2014 Master's thesis.
[17] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, 1993.
[18] F.R. Kschischang, B.J. Frey, H.A. Loeliger, Factor graphs and the sum-product algorithm, IEEE Trans. Inf. Theory 47 (2) (2001) 498–519.
[19] E.G. Larsson, O. Edfors, F. Tufvesson, T.L. Marzetta, Massive MIMO for next generation wireless systems, IEEE Commun. Mag. 52 (2) (2014).
[20] T.L. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas, IEEE Transactions on Communication 9 (11) (2010) 3590–3600.
[21] D. Reid, An algorithm for tracking multiple targets, IEEE Trans. on Automatic Control 24 (6) (1979) 843–854.
[22] F. Rusek, D. Persson, B.K. Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays, IEEE Signal Process Mag 30 (1) (2013) 40–60.
[23] L. Sanguinetti, A.A. D'Amico, M. Morelli, M. Debbah, Random access in uplink massive MIMO systems: How to exploit asynchronicity and excess antennas, IEEE Global Communications Conference (GLOBECOM) (2016).
[24] H. Soltanizadeh, S. Farahmand, S.M. Razavizadeh, Uncoordinated channel tracking for high-rate internet of things in multiuser massive MIMO, Proc. of International Symposium on Telecommunications (IST), Tehran, Iran (2020).
[25] J.H. Sørensen, E. de Carvalho, Uncoordinated pilot decontamination in massive MIMO systems, EURASIP J Wirel Commun Netw (156) (2017).
[26] J.H. Sørensen, E.d. Carvalho, P. Popovski, Massive MIMO for crowd scenarios: A solution based on random access, in: Proc. IEEE Globecom Workshops, Austin, TX, USA, 2014.
[27] J.H. Sørensen, E.D. Carvalho, C. Stefanovic, P. Popovski, Coded pilot random access for massive MIMO systems, IEEE Trans. on Wireless Communications 17 (12) (2018) 8035–8046.
[28] M.N. Tehrani, S. Farahmand, Massive random access in massive MIMO via opportunistic thresholding, in: Proc. of International Conference on Communications (ICC), Dublin, Ireland, 2020.
[29] J. Zhao, H. Xie, F. Gao, W. Jia, s. Jin, H. Lin, Time varying channel tracking with spatial and temporal BEM for massive MIMO systems, IEEE Trans. Wireless Commun. 17 (8) (2018).