This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

Joint Channel Estimation and Robust Beamforming Design for AF Relaying Using IMM Kalman Filters

Mohammad Amin Maleki Sadr, Benoit Champagne

***Abstract*—This paper addresses the joint problem of recursive channel estimation and robust beamformer design in peer-to-peer communication through a network of relays over time-varying radio channels. Using observed signal samples at the relay and receiver nodes, the Channel State Information (CSI) is estimated centrally by taking advantage of a Markov model for the transmitter-relay and relay-receiver channels, and employing either the Extended Kalman Filter (EKF) or the Cubature Kalman Filter (CKF). Based on the estimated CSI, two robust approaches are conceived for designing the relay beamforming where the aim is to minimize the total transmission power of the relays subject to Signal-to-Interference plus Noise Ratio (SINR) constraints at each of the receiver nodes. Furthermore, the Interacting Multiple Model (IMM) approach for mixing non-stationary and stationary Markov models is employed to extend the time-varying robust beamforming design to non-stationary environments. Through numerical simulations, the recursive CSI estimation methods are shown to be efficient, i.e., unbiased and converging to the Cramer-Rao Lower Bound (CRLB). Furthermore, the results confirm the better performance of the proposed robust relay beamforming design algorithms compared to existing methods in terms of relevant transmission metrics, including relay power consumption and spectral efficiency.**

***Index Terms*—Cramer-Rao-Lower-Bound, Cubature Kalman Filter, Extended Kalman Filter, Imperfect channel state information, Interacting Multiple Models.**

I. Introduction

Cooperative communication has received significant interest in recent years as it offers a promising way of achieving spatial diversity without using Multiple-Input-Multiple-Output (MIMO) processing [1], [2]. However, due to the dynamic nature of the wireless channels and the mobility of users, degradations of the required channel state information (CSI) are inevitable, which in turn negatively impact the communication reliability and transmission rate [3]. Employing relay nodes between the transmitter-receiver pairs can reduce the effects of channel degradations and help mitigate inter-user interference, thereby allowing reliable communication between each transmitter-receiver pair.

Manuscript submitted on August 1st, 2021. It was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors are with the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada (e-mail: mohammadamin.malekisadr@mail.mcgill.ca; benoit.champagne@mcgill.ca).

Among existing relaying schemes, such as Amplify-and-Forward (AF) [4], [5], [6], [7], [8], Decode-and-Forward (DF) [9], and Coded-cooperation-and-Forward (CF) [10], AF is much more attractive due to its relatively low implementation complexity and high security. A network of AF relays can be used to implement distributed peer-to-peer beamforming, i.e., a cooperative scheme wherein each one of the paired transmitter-receiver links in the network is enhanced through spatial processing [11], [12]. Indeed, while interference from other pairs can decrease the Signal-to-Interference plus noise (SINR) at a target receiver, relay beamforming can be exploited to mitigate the interference and improve the SINR. Besides traditional mobile wireless communications, distributed peer-to-peer relay beamforming finds applications in ad hoc Device-to-Device (D2D) networks [13], [14] and internet-of-things (IoT) use cases [15].

A major problem in the practical implementation of the above distributed relay schemes occurs when the various nodes comprising the network (i.e., transmitters, receivers, and especially relays) are moving relative to each other and their environment. In this case, and depending on the velocity of motion, the relevant CSI[1] cannot be well estimated as it contains errors due to limited channel- state feedback quantization and feedback delays. This issue becomes particularly challenging in the case of rapidly changing doubly-selective channels, since both the coherence time and coherence bandwidth can be affected. Thus, designing beamforming weights robust against CSI errors under such conditions is of great practical interest. Besides, optimal allocation of power among different relays also relies on accurate CSI estimation. Hence, optimizing the spatial relay beamforming and power allocation is a challenging task under temporal channel variations.

In the literature on cooperative robust relay beamforming, it is assumed that an initial estimation of the required channel coefficients is available, albeit with some errors. While these errors are unknown, their size (as represented by the variance or other error mode parameters) is constant over time and does not decrease as additional measurements are made. Within this framework, two different types of robust peer-to-peer beamforming design approaches have been proposed, depending on the particular way in which the CSI error is modeled, that is:

[1]Note that in this paper, the CSI stands for the actual channel coefficients.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

stochastic-based methods as in [6], [5]; and worst-case methods as in [16], [17]. In particular, a stochastic-based robust method for adjusting the relays' beamformer weights is proposed in [6], while in [5] and [7], related approaches with reduced complexity are presented for tackling the robust beamformer design problem. One important aspect that has not been considered in these studies is the possibility of reducing CSI uncertainty over time by exploiting new signal observations made at the relays and the receivers. While such an approach allowing dynamic modeling of the channel error has been recently undertaken in [18], the proposed method requires solving a non-convex optimization problem and its convergence is somewhat sensitive to initial conditions.

Parametric models of channel variations can be broadly classified as stationary (or time-invariant) and non-stationary (time-variant). Stationary models are well suited for situations where the channel varies slowly while the delay and Doppler spreads do not change appreciably; in fact, the mobile speed is assumed to remain constant at each time step [19]. In contrast, non-stationary models are typically employed when both the coherence time and the coherence bandwidth of the channel are time-varying. In the case of fast variations, recursive feedback channel models may not satisfy the Lyapunov stability condition [20]. Hence, as time passes, channel estimates derived from these improper models may greatly differ from the true values, since the estimation error can grow unboundedly.

Considering the wide spectrum of possible channel conditions, it becomes extremely important to devise channel estimation techniques that can learn and track the channel parameters of mobile users under a variety of conditions, including both stationary and non-stationary situations. This kind of learning-based techniques has not been considered in previous stochastic or worst-case approaches to robust beamforming design. For instance, in previous work [6], it was assumed that the random channel estimation errors obey an invariable known distribution. Such a model innately cannot track the non-stationary behavior of the wireless channel statistics.

In this paper, we therefore address the joint problem of recursive channel estimation and robust beamformer design in peer-to-peer communication through a network of relays over time-varying radio channels. In effect, cooperative relaying provides a form of space division multiple-access, whereby spatial channel characteristics of the different peer-to-peer links are exploited to mitigate interference and improve SINR. In [8], the authors have shown that this approach can outperform other common schemes, including Time Division (TDMA), Code Division (CDMA) and Frequency Division Multiple Access (FDMA). We assume that there exists a processing (or fusion) center that has access to a subset of the signal samples observed at the relays and the receivers with the ability to fuse information from these two sources. The information provided to the processing center, which is referred to as measurements in the sequel, is used to estimate the channel coefficients, whose evolution is represented by a parametric state-space model. The measurements are a function of beamforming weights which are optimized at each time step[2]. Due to the non-linear nature of the measurements at the relays and receiver nodes with respect to the channel state vector, we have utilized the Cubature Kalman Filter (CKF) [21] as well as the Extended Kalman Filter (EKF) [22] for tracking the channel state information.

While the EKF is attractive due to its simplicity, the CKF is known to have a better performance in non-linear state space models compared to other non-linear approximations of the Kalman filter (KF) [21]. The main contributions and distinguishing features of this paper are summarized as follows:

- Two different robust optimization approaches for beamformer design are proposed in which the total transmission power of the relays is minimized subject to, respectively, probabilistic and approximated mean SINR constraints at the receiver nodes.
- We apply two KF-based methods i.e., EKF and CKF, for estimating the CSI required by the robust beamformers. In contrast to previous studies [6], [5], [16], our proposed approaches are recursive and seek to mitigate the channel errors at each time step dynamically by using sample measurements at the relays and receivers.
- In our approach, the convergence of the channel estimation method does not depend on the initial conditions. In fact, in [18], although the robust approach is adaptive, the beamforming vector design highly depends on having proper initial conditions.
- We propose an innovative approach based on the Interacting Multiple Models (IMM) for Markov models, in order to jointly handle both stationary and non-stationary channel behaviors.
- We evaluate the computational complexity of our proposed approaches in detail. We also derive the Cramer-Rao Lower Bound (CRLB) for CSI estimation in the considered framework.
- The results of extensive numerical simulations show that our proposed methods are unbiased and achieve the CRLB, thereby demonstrating their statistical efficiency in practice.

The remainder of this paper is organized as follows.

[2]We emphasize that in our proposed method, the rate at which the observed samples are forwarded to the processing center is different from the sampling or symbol rate of the communication system. That is, only a subset of the available received samples, as obtained from sub-sampling at a much lower rate consistent with the rate of change of the channel, need to be forwarded to the central processor. Typically, only a few samples per coherence interval will be needed, which represents a very small fraction of the received data. Hence, the proposed scheme will not entail a substantial overhead in practice.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

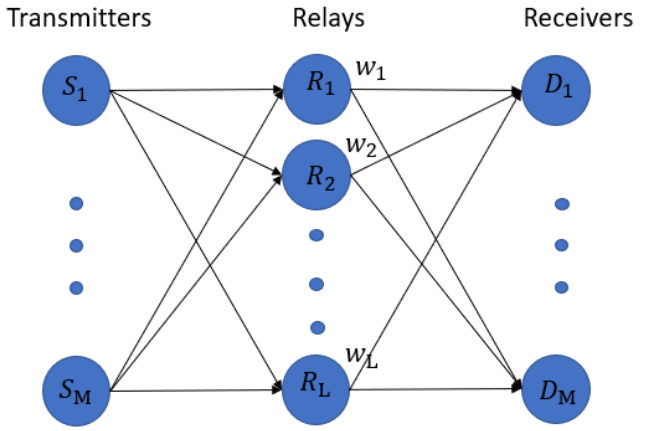

Figure 1: Point-to-point relay system model.

In Section II, the system model and problem formulation are presented. In Section III, two different robust optimization approaches for solving the beamforming problem are proposed. Subsequently, the channel estimation based on EKF and CKF is investigated in Section IV. In Section V, the IMM method for combining the non-stationary and stationary models is discussed. Section VI delves into performance considerations, including the complexity analysis of the proposed algorithms and derivation of a simplified CRLB for channel state vector estimation. Simulation results are presented and discussed in Section VII. Finally, conclusions are drawn in Section VIII.

Notations: In this paper, uppercase and lowercase bold letters are used to denote matrices and vectors, respectively, while superscripts $(.)^*$, $(.)^T$, $(.)^H$ denote, complex conjugate, transpose, Hermitian (conjugate transpose), respectively. The notations $\mathrm{vec}\,(.)$, $\mathrm{Tr}\,(.)$, $\odot$, and $\otimes$ respectively, denote vectorization operator, matrix trace operator, element-wise (or Hadamard) matrix product and Kronecker matrix product. The operator $\mathrm{diag}\,(\mathbf{A})$ represents a column vector which contains the diagonal elements of square matrix $\mathbf{A}$, while $\mathrm{diag}\,(\mathbf{a})$ denotes a diagonal matrix with the elements of vector $\mathbf{a}$ on its main diagonal. The notation $\mathrm{BD}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ represents a block diagonal matrix with matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$ on its main block diagonal. Furthermore $\mathcal{CN}(\mu, \sigma^2)$ denotes a complex circular Gaussian distribution with mean and variance μ and σ^2, respectively.

II. SYSTEM MODEL AND PROBLEM STATEMENT

A. System Model

We consider a network for point-to-point communication between M transmitter-receiver pairs, through a layer of L relays operating in parallel, as illustrated in Fig. 1. The transmitters $\{S_i\}_{i=1}^M$, relays $\{R_r\}_{r=1}^L$, and receivers $\{D_i\}_{i=1}^M$, are all assumed to be equipped with a single antenna. In this paper, we focus on half-duplex relayed communications from the sources S_i to the destinations D_i; however, full-duplex[3] communications between the two sets of users can be achieved by means of time division duplexing (TDD) or frequency division duplexing (FDD) schemes. We let $f_{pr}(k)$, and $g_{rp}(k)$ denote the complex Rayleigh flat fading channel gain coefficients between the pth transmitter and the rth relay, and between the rth relay and the pth receiver at the kth time step, respectively; there exists no direct link between the transmitters and the receivers. For notational simplicity, we shall temporarily drop the time step index k from the channel coefficients (e.g., $f_{pr} \equiv f_{pr}(k)$, $g_{rp} \equiv g_{rp}(k)$); hence, all the equations in this and the next section are for the kth time step[4]. At the given time step, the vector of signal samples observed at the relays is expressed as

$$\mathbf{x} = \sum_{p=1}^{M} \mathbf{f}_p s_p + \mathbf{v}_m^x \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_L]^T \in \mathbb{C}^L$, $\mathbf{v}_m^x = [v_{m1}^x, v_{m2}^x, \ldots, v_{mL}^x]^T$ is a vector of Additive White Gaussian Noise (AWGN) at the relays with complex circular Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma_x^2\mathbf{I})$, $\mathbf{f}_p = [f_{p1}, f_{p2}, \ldots, f_{pL}]^T$ and $\{s_p\}_{p=1}^M$ are the transmitted information symbols. The latter are assumed to be uncorrelated with zero mean and variance $P_p = E\{|s_p|^2\}$. The rth relay multiplies its received signal by a complex weight w_r^*, then retransmits the resulting product to the receivers. Hence, the vector of transmitted signals by the L relays can be written as

$$\mathbf{t} = \mathbf{W}^H \mathbf{x} \tag{2}$$

where $\mathbf{W} = \mathrm{diag}(\mathbf{w})$ and $\mathbf{w} = [w_1, w_2, \ldots, w_L]^T$. The observed signal sample at the jth receiver node is

$$\begin{aligned} y_j &= \mathbf{g}_j{}^T \mathbf{t} + \eta_j, \qquad \forall j \in \mathcal{J} \triangleq \{1, \ldots, M\} \\ &= \mathbf{g}_j^T \mathbf{W}^H \mathbf{f}_j s_j + \mathbf{g}_j{}^T \mathbf{W}^H \sum_{p \neq j}^{M} \mathbf{f}_p s_p + v_m^{y_j} \end{aligned} \tag{3}$$

where $\mathbf{g}_j = [g_{j1}, \ldots, g_{jL}]^T$, $v_m^{y_j} = \mathbf{g}_j^T \mathbf{W}^H \mathbf{v}_m^{\mathbf{x}} + \eta_j$ is the total noise at the jth receiver, and $\eta_j \sim \mathcal{CN}(0, \sigma_\eta^2)$.

[3]Traditionally, the term full-duplex refers to the use of bilateral channels (whether single or shared media) for simultaneous transmission of information in both directions between two end-points [23]. This includes the use of TDD and FDD, which are intrinsically half-duplex (one-way) modes of communication, to emulate full-duplex or two-way communications by interlacing the two different directions of transmission over time or frequency, respectively. In recent years, with advances in RF hardware and DSP technologies, there has been a growing body of research on so-called full-duplex relaying, wherein two-way relayed communications can be achieved simultaneously over the same frequency band through advanced self-interference cancellation techniques [24]. In the above statement, more traditional interpretation of the term full-duplex is implied.

[4]The index k will be reintroduced in Section IV when investigating recursive channel estimation.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

The SINR at the jth receiver node is written as

$$\mathrm{SINR}_j = \frac{P_S^j}{P_N^j + P_I^j}, \qquad j \in \mathcal{J} \tag{4}$$

where P_S^j, P_I^j, P_N^j are the desired signal, interference, and noise power, respectively. By using (2), the total transmit power of the relays is

$$P_T = \mathrm{E}\{\|\mathbf{t}\|^2\} = \mathbf{w}^H \mathbf{D} \mathbf{w} \tag{5}$$

where $\mathbf{D} = \mathrm{diag}([\mathbf{R_x}]_{1,1}[\mathbf{R_x}]_{2,2} \dots [\mathbf{R_x}]_{L,L})$ and $\mathbf{R_x} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\}$ is the correlation matrix of the received signals at the relays. The latter matrix can be expressed as

$$\mathbf{R_x} = \sum_{p=1}^{d} P_p \mathbf{f}_p \mathbf{f}_p^H + \sigma_x^2 \mathbf{I} \tag{6}$$

Using (3), the power of the jth desired signal can be written as

$$P_S^j = \mathrm{E}\left\{\mathbf{g}_j^T \mathbf{W}^H \mathbf{f}_j \mathbf{f}_j^H \mathbf{W} \mathbf{g}_j^*\right\} \mathrm{E}\{|s_j|^2\} = \mathbf{w}^H \mathbf{R}_h^j \mathbf{w} \tag{7}$$

where $\mathbf{h}_j^j$ and $\mathbf{R}_h^j$ are defined as

$$\mathbf{R}_h^j \triangleq P_j \mathbf{h}_j^j \mathbf{h}_j^{jH}, \qquad \mathbf{h}_j^j \triangleq \mathbf{g}_j \odot \mathbf{f}_j$$

Similarly, the interference power at the jth receiver can be written as

$$P_I^j = \mathrm{E}\{\mathbf{g}_j^T \mathbf{W}^H \sum_{p \neq j}^{M} \sum_{q \neq j}^{M} \mathbf{f}_p \mathbf{f}_q^H s_p s_q^* \mathbf{W} \mathbf{g}_k^*\} = \mathbf{w}^H \mathbf{Q}_j \mathbf{w} \tag{8}$$

where

$$\mathbf{Q}_j \triangleq \sum_{p \neq j}^{M} P_j \mathbf{h}_j^p (\mathbf{h}_j^p)^H, \quad \mathbf{h}_j^p \triangleq \mathbf{g}_j \odot \mathbf{f}_p,$$

Similarly, the noise power at the jth receiver is

$$\begin{aligned} P_N^j &= \mathrm{E}\left\{(\mathbf{g}_j^T \mathbf{W}^H \mathbf{v}_m^x + \eta_j)(\mathbf{g}_j^T \mathbf{W}^H \mathbf{v}_m^x + \eta_j)^H\right\} \\ &= \mathbf{w}^H \mathbf{D}_j \mathbf{w} + \sigma_\eta^2 \end{aligned} \tag{9}$$

where $\mathbf{D}_j \triangleq \sigma_x^2 \mathrm{diag}(\mathbf{g}_j^*)\mathrm{diag}(\mathbf{g}_j)$. Hence, the SINR at the jth receiver is

$$\mathrm{SINR}_j = \frac{\mathbf{w}^H \mathbf{R}_h^j \mathbf{w}}{\mathbf{w}^H (\mathbf{Q}_j + \mathbf{D}_j) \mathbf{w} + \sigma_\eta^2} = \frac{A_j}{B_j}, \qquad j \in \mathcal{J} \tag{10}$$

where $A_j \triangleq \mathbf{w}^H \mathbf{R}_h^j \mathbf{w}$ and $B_j \triangleq \mathbf{w}^H (\mathbf{Q}_j + \mathbf{D}_j) \mathbf{w} + \sigma_\eta^2$.

B. Problem Statement

Our aim is to design the beamforming weights at the relays so that power consumption is minimized while preserving a desired level of QoS at the receiver nodes. However, and in contrast to earlier works, we will seek to exploit *a priori* knowledge of the channel dynamics to improve the performance of the beamformer as the radio channels undergo time variations.

The proposed joint channel estimation and beamforming design procedure is illustrated in Fig. 2, where

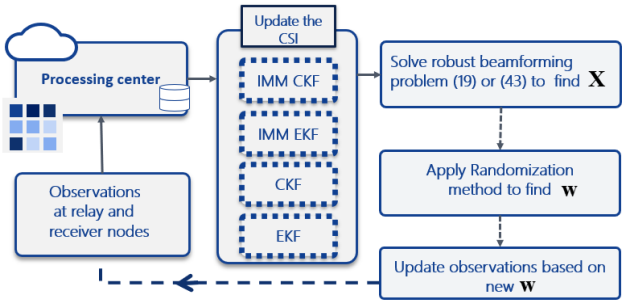

Figure 2: Joint channel estimation and beamforming design at each time step

the designed beamforming weights at the relays are functions of the CSI. At each time step, we first update the CSI, which consists of the relevant channel vectors, by using a non-linear form of Kalman filter (either EKF or CKF), and then use the updated channel vectors to solve for the optimal relaying beamforming weights. Hence, the channel estimation and beamforming design are done jointly, but in an *alternative* manner, which is a common approach for solving complex high-dimensional optimization problems, see e.g. [25], [26]. Interestingly, for the particular relaying problem with time-varying channels under consideration in this work, the alternations are applied over consecutive time steps, which leads to an adaptive procedure. More specifically, the CSI needs to be estimated in order to solve the following optimization problem for the relaying weights

$$\min_{\mathbf{w}} \quad \mathbf{w}^H \mathbf{D} \mathbf{w} \tag{11a}$$

$$\text{s.t.} \quad \mathrm{SINR}_j \geq \gamma_j, \qquad \forall j \in \mathcal{J} \tag{11b}$$

where γ_j is a preselected SINR threshold value at the jth receiver node. Due to rapid changes in CSI, optimization problem (11) should be solved in real-time, i.e. at each specific time step, as the estimates of the various channel vectors are updated. Furthermore, due to unavoidable CSI uncertainties in the estimation, a robust design approach is favoured in lieu of the deterministic formulation (11).

In [6], the authors solved a stochastic version of the robust beamforming design problem by assuming perfect knowledge of the distribution of the channel estimation error (including the mean and covariance of this distribution). However, in the case of time-varying channels with possible non-stationary model parameters, such detailed knowledge about the propagation environment is rarely available. In [18], the authors proposed a joint approach for the prediction of beamforming weights and CSI, by reformulating a non-convex optimization problem as a constrained least squares problem. The major drawbacks of this solution are its high complexity and the need for a reliable initial estimation of the beamforming weights, which may not be possible in practical scenarios.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

In contrast to previous studies [16], [18] the robust approaches for beamformer design proposed in this work are recursive in nature and seek to mitigate the channel errors at each time step dynamically. This is achieved by incorporating KF-based methods within the processing pipeline, in order to better exploit sample measurements at the relays and receivers. Consequently, in our approach, the convergence of the channel estimation method does not strongly depend on the initial conditions. In the presence of CSI uncertainty, the vector of channel coefficients between the pth transmitter and the relay nodes can be expressed as

$$\mathbf{f}_p = \bar{\mathbf{f}}_p + \Delta\mathbf{f}_p, \qquad p \in \mathcal{J} \tag{12}$$

where $\mathbf{f}_p$, $\bar{\mathbf{f}}_p$ and $\Delta\mathbf{f}_p$ stand for the estimated, actual and error vectors, respectively. Similarly, the vector of channel coefficients between the relays and the jth receiver is expressed as

$$\mathbf{g}_j = \bar{\mathbf{g}}_j + \Delta\mathbf{g}_j, \qquad j \in \mathcal{J} \tag{13}$$

where $\mathbf{g}_j$, $\bar{\mathbf{g}}_j$ and $\Delta\mathbf{g}_j$ stand for estimated, actual and error vectors, respectively. In both (12) and (13), the error vectors $\Delta\mathbf{f}_p$ and $\Delta\mathbf{g}_j$ are assumed to follow complex circular Gaussian distributions. In accordance with (12)-(13), the expectation operator in e.g. $\mathcal{CN}(0, \sigma_f^2\mathbf{I})$ and $\mathcal{CN}(0, \sigma_g^2\mathbf{I})$, respectively. The model equations derived in part A remain valid provided that the operator $\mathrm{E}\{.\}$ is interpreted as a conditional expectation, given particular realizations of the CSI errors $\Delta\mathbf{f}_p$ and $\Delta\mathbf{g}_j$. The newly proposed robust beamformer design strategies (given the updated CSI) are described in details in Section III; while the non-linear KF-based channel estimation and tracking (given updated beamformer measurements) is described in Section IV (with extension to IMM in Section V). At each time iteration, as explained above, new beamforming weights are designed using a robust procedure, which is followed by channel vector updating using non-linear KF. Hence, the beamforming weights and channel vectors are jointly enhanced in an alternative manner over time.

III. Robust Beamforming Design Strategy

In this section, two different variations of problem (11) are proposed and developed for making the relay beamforming solution robust against CSI errors.

A. Maximizing First-Order Approximate Mean of SINR Under Imperfect CSI

It has been proved that assuming exact CSI in situations where CSI is afflicted by uncertainty may severely degrade system performance [6]. In this sub-section, we reformulate (11) as a robust optimization problem based on the approximation of the mean method (AMM) for the SINR. Using the approach in [27], the expected value of the SINR in (10) is approximated by a Taylor series around the mean values of A_j and B_j:

$$\begin{aligned}\mathrm{SINR}_j = \mathrm{SINR}_j\left(\bar{A}_j, \bar{B}_j\right) + \tfrac{\partial \mathrm{SINR}_j}{\partial A_j}\left(A_j - \bar{A}_j\right) \\ + \tfrac{\partial\, \mathrm{SINR}_j}{\partial B_j}\left(B_j - \bar{B}_j\right) + \mathrm{h.o.t.}\end{aligned} \tag{14}$$

where $\mathrm{h.o.t.}$ denotes the higher order terms. Taking the expectation of (14), we have

$$\begin{aligned}\mathrm{E}(\mathrm{SINR}_j) = \mathrm{SINR}(\bar{A}_j, \bar{B}_j) + \int\int (\tfrac{\partial \mathrm{SINR}_j}{\partial A_j}\left(A_j - \bar{A}_j\right) \\ + \tfrac{\partial \mathrm{SINR}_j}{\partial B_j}\left(B_j - \bar{B}_j\right)) f\left(A_j, B_j\right) dA_j dB_j + \mathrm{h.o.t.}\end{aligned} \tag{15}$$

where $f\left(A_j, B_j\right)$ denotes the joint probability density function of A_j and B_j. In [27], the first-order approximation of (15) is further simplified as $\mathrm{E}(\mathrm{SINR}_j) \cong \mathrm{SINR}(\bar{A}_j, \bar{B}_j) + \frac{2\mathrm{SINR}(\bar{A}_j, \bar{B}_j)}{\bar{B}_j^2}$, while the second and higher order terms do not have a sensible impact on the final result. It is also shown through numerical analysis that the above expression provides a tight lower-bound of true SINR under the low channel uncertainty error condition.

Our first proposed method relies on the use of the first term in (15)[5]. Specifically, in light of (11b), we have

$$\mathrm{SINR}_j(\bar{A}_j, \bar{B}_j) = \left(\frac{\mathbf{w}^H \bar{\mathbf{R}}_h^j \mathbf{w}}{\mathbf{w}^H(\bar{\mathbf{Q}}_j + \bar{\mathbf{D}}_j)\mathbf{w} + \sigma_\eta^2}\right) \geq \gamma_j \tag{16}$$

where $\bar{\mathbf{R}}_h^j = P_j(\bar{\mathbf{h}}_j^j(\bar{\mathbf{h}}_j^j)^H + \sigma_{\mathbf{h}_j}^2\mathbf{I})$, $\bar{\mathbf{D}}_j = \sigma_x^2 \mathrm{diag}(\bar{\mathbf{g}}_j\bar{\mathbf{g}}_j^H + \sigma_{\mathbf{g}}^2\mathbf{I})$, $\bar{\mathbf{Q}}_j = \sum_{p\neq j}^{M} P_p(\bar{\mathbf{h}}_j^p(\bar{\mathbf{h}}_j^p)^H + \sigma_{\mathbf{h}}^2\mathbf{I})$, $\bar{\mathbf{h}}_j^j = \mathrm{E}(\mathbf{g}_j \odot \mathbf{f}_j) = \bar{\mathbf{g}}_j \odot \bar{\mathbf{f}}_j$ and $\bar{\mathbf{h}}_j^p = \bar{\mathbf{g}}_j \odot \bar{\mathbf{f}}_p$. Hence, an optimum beamforming weight vector can be obtained by solving the following problem:

$$\min_{\mathbf{w}} \quad \mathbf{w}^H\mathbf{D}\mathbf{w} \tag{17a}$$

$$\text{s.t.} \quad \mathbf{w}^H(\bar{\mathbf{R}}_h^j - \gamma_j(\bar{\mathbf{Q}}_j + \bar{\mathbf{D}}_j))\mathbf{w} \geq \sigma_\eta^2\gamma_j \ \ \forall j \in \mathcal{J} \tag{17b}$$

which is indeed equivalent to

$$\min_{\mathbf{X}} \quad \mathrm{Tr}\left(\mathbf{D}\mathbf{X}\right) \tag{18a}$$

$$\text{s.t.} \quad \mathrm{Tr}(\bar{\mathbf{R}}_h^j - \gamma_j(\bar{\mathbf{Q}}_j + \bar{\mathbf{D}}_j)\mathbf{X}) \geq \sigma_\eta^2\gamma_j \ \ \forall j \in \mathcal{J} \tag{18b}$$

$$\mathrm{Rank}(\mathbf{X}) = \mathbf{1} \tag{18c}$$

where $\mathbf{X} = \mathbf{w}\mathbf{w}^H$. After dropping the rank one constraint in (18c) which is non-convex, this can be rewritten in relaxed form as follows:

$$\min_{\mathbf{X}} \quad \mathrm{Tr}\left(\mathbf{D}\mathbf{X}\right) \tag{19a}$$

$$\text{s.t.} \quad \mathrm{Tr}((\bar{\mathbf{R}}_h^j - \gamma_j(\bar{\mathbf{Q}}_j + \bar{\mathbf{D}}_j))\mathbf{X}) \geq \sigma_\eta^2\gamma_j \ \ \forall j \in \mathcal{J} \tag{19b}$$

Problem (19) is a convex Semi-Definite Program (SDP) which can be easily solved by CVX software [28]. As

[5]The reason for not including the term $\frac{2\mathrm{SINR}(\bar{A}_j, \bar{B}_j)}{\bar{B}_j^2}$ is twofold: the denominator $\bar{B}_j^2$ contains 4th power terms with respect ro $\mathbf{w}$ which makes the problem non-convex; in low uncertainty error regime, this term does not have a significant effect on the overall SINR term, as we show by simulations in Section VII.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

mentioned in [29], the SDP problems in form (19) do not admit a closed-form solution and cannot be solved analytically.

We shall first seek to extract beamforming weights $\mathbf{w}$ from the solution of (19), which provides a lower bound to the solution of (18), since the feasibility region (18b-c) of the non-convex problem is a subset of the relaxed feasibility region (19b). In general, the solution of the relaxed problem (19), denoted as $\mathbf{X}_{opt}$, may have a arbitrary rank. If the rank of $\mathbf{X}_{opt}$ is one, the principal eigenvector of $\mathbf{X}_{opt}$ is the optimal solution to the original problem. Otherwise, if the rank of the matrix $\mathbf{X}_{opt}$ is higher than one, an approximation technique is needed to obtain a rank one solution from the relaxed problem. To this end, we can employ an effective procedure known as randomization in the literature [8], [6], [7]. The idea behind this technique is to generate a candidate set of beamforming vectors from the optimal solution of (19) and conducting finite search over those that fall in the feasibility region. The solution obtained by means of this procedure can provide a tight lower bound to the relaxed problem.

To design a randomization procedure for our problem, let $\mathbf{X}_{opt} = \mathbf{U}\mathbf{V}\mathbf{U}^H$ denotes the eigenvalue decomposition of $\mathbf{X}_{opt}$. The candidate vectors $\mathbf{w}_c$ can be chosen as $\mathbf{w}_c = \mathbf{U}\mathbf{V}^{\frac{1}{2}}\boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, so that $\mathrm{E}(\mathbf{w}_c\mathbf{w}_c^H) = \mathbf{X}_{opt}$. A feasible solution can be obtained by generating a sufficiently large number of realizations of $\mathbf{w}_c$, and then simply choosing the best feasible solution. Then, one way to generate the candidate solution of problem (18) is to scale $\mathbf{w}_c$ by an appropriate factor. This scaling factor is obtained by solving the following linear optimization problem:

$$\begin{aligned} \min_{\lambda} \quad & \lambda \\ \text{s.t.} \quad & \lambda \geq \sigma_\eta^2 \gamma_j \mathcal{T}_j^{-1}, \qquad \forall j \in \mathcal{J} \end{aligned} \tag{20}$$

where $\mathcal{T}_j = \mathrm{Tr}(\bar{\mathbf{R}}_h^j - \gamma_j(\bar{\mathbf{Q}}_j + \bar{\mathbf{D}}_j)\mathbf{w}_c\mathbf{w}_c^H)$. The random vector generation process is repeated until N_{max} feasible candidates have been found. Among these, we select the vector which yields the minimum objective value in the optimization problem (18). The main steps of this randomization method are summarized under Algorithm 1, where in line (6), equation (20) is solved.

B. Statistically Robust Design

In this section, we introduce our stochastic method (SM) as second proposed variation of problem (11), which relies on a stochastic formulation of the QoS constrained to achieve robustness of the solutions. Specifically, we aim to solve the following optimization problem:

$$\min_{\mathbf{w}} \quad \mathbf{w}^H\mathbf{D}\mathbf{w} \tag{21a}$$

$$\text{s.t} \quad \Pr\left(\mathrm{SINR}_j \leq \gamma_j\right) \leq \varepsilon_j, \qquad \forall j \in \mathcal{J} \tag{21b}$$

where ε_j is a maximum threshold outage probability at the jth receiver node. Let us define an auxiliary variable Z_j as follows:

$$Z_j = \mathbf{w}^H \left(\mathbf{R}_h^j - \gamma_j\mathbf{Q}_j - \gamma_j\mathbf{D}_j\right)\mathbf{w} \tag{22}$$

Hence, constraint (21b) can be equivalently expressed as

$$\Pr\left(Z_j \leq \gamma_j\sigma_\eta^2\right) \leq \varepsilon_j \ , \qquad \forall j \in \mathcal{J} \tag{23}$$

It can easily be verified that (22) contains only terms up to order four in $\Delta\mathbf{f}_p$, and $\Delta\mathbf{g}_j$. In [6], It was shown that the third and forth order error terms in a similar expression be neglected. Guided by these observations, the following theorem can be applied to obtain a tractable convex approximation, in the form of second order cone constraints, to the chance constraints (23).

Theorem 1. *Let $\xi_1, \xi_2, ..., \xi_m$ be independent standard Gaussian random variables. Consider the function $Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined via*

$$\begin{aligned} Q(\mathbf{x}, \boldsymbol{\xi}) = -a_0(\mathbf{x}) + \sum_{i=1}^{m} \xi_i a_i(\mathbf{x}) + \sum_{i,j=1}^{m} \xi_i\xi_j a_{i,j}(\mathbf{x}) \\ + \sum_{i,j,k=1}^{m} \xi_i\xi_j\xi_k a_{i,j,k}(\mathbf{x}) + \sum_{i,j,k,l=1}^{m} \xi_i\xi_j\xi_k\xi_l a_{i,j,k,l}(\mathbf{x}) \end{aligned} \tag{24}$$

where $a_0(.)$ is affine and $a_i(.)$, $a_{i,j}(.)$, $a_{i,j,k}(.)$, $a_{i,j,k,l}(.)$ are linear in their arguments. Consider the chance constraint

$$P\left(Q\left(\boldsymbol{x}, \boldsymbol{\xi}\right) \geq 0\right) \leq \varepsilon \tag{25}$$

where $\varepsilon > 0$ is given. Set

$$\bar{q}\left(\varepsilon\right) = \begin{cases} \frac{-\ln\varepsilon + \sqrt{(\ln\varepsilon)^2 - 8\ln\varepsilon}}{4}, & \varepsilon \leq e^{-8} \\ 2, & else \end{cases} \tag{26}$$

and $\bar{Q}\left(\boldsymbol{x}, \boldsymbol{\xi}\right) = Q\left(\boldsymbol{x}, \boldsymbol{\xi}\right) + a_0\left(\mathbf{x}\right)$. Then, the following propositions hold:
(a) For all $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\xi} \in \mathbb{R}^m$, for some function $\mathbf{U} : \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^n$, we have $\bar{Q}(\boldsymbol{x}, \boldsymbol{\xi})^2 = \mathbf{x}^T\mathbf{U}\left(\boldsymbol{\xi}\right)\mathbf{x}$
(b) Let $\mathbf{U} = \mathrm{E}\{\mathbf{U}\left(\boldsymbol{\xi}\right)\} \succ 0$ and

$$c\left(\varepsilon\right) = \begin{cases} \left(\bar{q}(\varepsilon) - 1\right)^2 \exp(\frac{2\bar{q}(\varepsilon)}{\bar{q}(\varepsilon)-1}), & \varepsilon \leq e^{-8} \\ \frac{1}{\sqrt{\varepsilon}}, & else \end{cases} \tag{27}$$

The second-order cone constraint

$$a_0\left(\mathbf{x}\right) \geq c(\varepsilon) \left\| \mathbf{U}^{\frac{1}{2}}\mathbf{x} \right\| \tag{28}$$

serves as a tractable safe[6] approximation of the chance constraint (25).

Proof. By assumption, for each $\boldsymbol{\xi} \in \mathbb{R}^m$, the function $\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)$ is linear in $\mathbf{x} \in \mathbb{R}^n$. This implies that $\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)^2$ is a non-negative homogeneous quadratic polynomial in

[6]The term *safe* indicates that this approximation does not violate the pre-defined outage probability ε or pre-defined SINR level γ_j when applied to (21b).

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

$\mathbf{x} \in \mathbb{R}^n$, which establishes (a). To prove (b), we invoke [30] (Theorem 5.10), which states that for any $q \geq 2$:

$$\mathrm{E}\{\left|\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)\right|^q\}^{\frac{1}{q}} \leq (q-1)^2 \mathrm{E}\{\left|\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)\right|^2\}^{\frac{1}{2}} \quad (29)$$

This, together with Markov's inequality and the result in (a), implies that for any $q \geq 2$

$$\begin{aligned} \Pr\left(\left|\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)\right| \geq t\right) &\leq t^{-q} \mathrm{E}\{\left|\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)\right|^q\} \\ &= \left[t^{-1}(q-1)^2 \left\|\mathbf{U}^{\frac{1}{2}}\mathbf{x}\right\|\right]^q \end{aligned} \quad (30)$$

By setting $q = \bar{q}(\varepsilon) \geq 2$ and $t \geq c\left(\varepsilon\right) \left\|\mathbf{U}^{\frac{1}{2}}\mathbf{x}\right\|$, it follows from (30) that

$$\begin{aligned} \Pr\left(Q\left(\mathbf{x}, \boldsymbol{\xi}\right) \geq 0\right) &= \Pr\left(\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right) \geq a_0\left(\mathbf{x}\right)\right) \\ &\leq \Pr\left(\left|\bar{Q}\left(\mathbf{x}, \boldsymbol{\xi}\right)\right| \geq a_0\left(\mathbf{x}\right)\right) \\ &\leq \varepsilon \end{aligned} \quad (31)$$

whenever (28) holds. In that sense, the second-order cone constraint (28) is a safe (and tractable) approximation of (25). □

Here, we aim to apply the result of Theorem 1 to the non-convex problem (21). The auxiliary variable Z_j can be simplified as

$$\begin{aligned} Z_j =& \Delta\mathbf{f}^H \mathbf{Q}_{\mathbf{f}} \Delta\mathbf{f} + \Delta\mathbf{g}_j^H \mathbf{Q}_{\mathbf{f},\mathbf{g}} \Delta\mathbf{f} + \Delta\mathbf{f}^H \mathbf{Q}_{\mathbf{f},\mathbf{g}}^H \Delta\mathbf{g}_j \\ &+ \Delta\mathbf{g}_j^T \widehat{\mathbf{Q}}_{\mathbf{f},\mathbf{g}} \Delta\mathbf{f} + \Delta\mathbf{f}^H \widehat{\mathbf{Q}}_{\mathbf{f},\mathbf{g}}^H \Delta\mathbf{g}_j^* + \Delta\mathbf{g}_j^H \mathbf{Q}_{\mathbf{g}} \Delta\mathbf{g}_j \\ &+ \mathbf{c}_{\mathbf{g}}^H \Delta\mathbf{g}_j + \Delta\mathbf{g}_j^H \mathbf{c}_{\mathbf{g}} + \mathbf{c}_{\mathbf{f}}^H \Delta\mathbf{f} + \Delta\mathbf{f}^H \mathbf{c}_{\mathbf{f}} + d_j + \text{h.o.t.} \end{aligned} \quad (32)$$

By neglecting higher order terms in the third and fourth powers of $\Delta\mathbf{f} \triangleq \left[\Delta\mathbf{f}_1^T, \ldots, \Delta\mathbf{f}_M^T\right]^T$ and $\Delta\mathbf{g}_j$, we can reformulate the latter as

$$Z_j = \Delta\mathbf{H}_j^H \tilde{\mathbf{Q}} \Delta\mathbf{H}_j + 2\mathrm{Re}\left\{\tilde{\mathbf{r}}_j \Delta\mathbf{H}_j^H\right\} + d_j \quad (33)$$

where $\tilde{\mathbf{Q}} \triangleq \begin{bmatrix} \mathbf{R}_I & \mathbf{R}_{IQ} \\ \mathbf{R}_{IQ} & \mathbf{R}_Q \end{bmatrix}$, $\mathbf{R}_I \triangleq \begin{bmatrix} \mathbf{Q}_{\mathbf{f}} & \mathbf{Q}_{\mathbf{f},\mathbf{g}}^H \\ \mathbf{Q}_{\mathbf{f},\mathbf{g}} & \mathbf{Q}_{\mathbf{g}} \end{bmatrix}$, $\mathbf{R}_{I,Q} \triangleq \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \widehat{\mathbf{Q}}_{\mathbf{f},\mathbf{g}}^H \mathbf{0} & \end{bmatrix}$, $\mathbf{R}_Q \triangleq \mathbf{0}$, $\Delta\mathbf{H}_j^H \triangleq \left[\Delta\mathbf{h}_j^H \; \Delta\mathbf{h}_j^T\right]$, $\Delta\mathbf{h}_j^T \triangleq \left[\Delta\mathbf{f}^T \;\; \Delta\mathbf{g}_j^T\right]$, $\tilde{\mathbf{r}}_j \triangleq \frac{1}{2}\left[\mathbf{c}^H \mathbf{c}^T\right]$, $\mathbf{c}^H \triangleq \left[\mathbf{c}_{\mathbf{f}}{}^H \mathbf{c}_{\mathbf{g}}{}^H\right]$ and the other matrices $\mathbf{Q}_{\mathbf{f}}, \mathbf{Q}_{\mathbf{g}}, \mathbf{Q}_{\mathbf{f},\mathbf{g}}, \widehat{\mathbf{Q}}_{\mathbf{f},\mathbf{g}}, \mathbf{c}_{\mathbf{f}}, \mathbf{c}_{\mathbf{g}}$, and d_j have been defined in Appendix. From (33), it follows that

$$\begin{aligned} \mathrm{E}\{|Z_j|^2\} = \mathrm{E}\{&\Delta\mathbf{H}_j^H \tilde{\mathbf{Q}} \Delta\mathbf{H}_j \Delta\mathbf{H}_j^H \tilde{\mathbf{Q}} \Delta\mathbf{H}_j\} \\ &+ 4\mathrm{E}\{\Delta\mathbf{H}_j^H \mathbf{r}_j \mathbf{r}_j^H \Delta\mathbf{H}_j\} \end{aligned} \quad (34)$$

We note that $\Delta\mathbf{H}_j \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ can be derived as

$$\boldsymbol{\Sigma} = \mathrm{E}(\Delta\mathbf{H}_j \Delta\mathbf{H}_j^H) = \mathrm{BD}(\sigma_f^2\mathbf{I}, \sigma_g^2\mathbf{I}, \sigma_f^2\mathbf{I}, \sigma_g^2\mathbf{I}) \quad (35)$$

After some algebraic manipulations (34) can be expressed in the equivalent form

$$\mathrm{E}\{|Z_j|^2\} = \mathrm{vec}(\mathbf{X})^H \mathbf{U}_j \mathrm{vec}(\mathbf{X}) \quad (36)$$

where $\mathbf{X} = \mathbf{w}\mathbf{w}^H$ and the expression of $\mathbf{U}_j$ is developed in Appendix II. The affine term in (33) can also be expressed as

$$\begin{aligned} d_j =& \mathrm{vec}(\mathbf{X})^H \mathrm{vec}(P_j \bar{\mathbf{h}}_j^j (\bar{\mathbf{h}}_j^j)^H) \\ &- \gamma_j \sigma_x^2 \mathrm{vec}(\mathbf{X})^H \mathrm{diag}(\mathrm{vec}(\mathbf{I})) \mathrm{vec}(\mathbf{g}_j \mathbf{g}_j^H) \\ &- \gamma_j \mathrm{vec}(\mathbf{X})^H \mathrm{vec}(\sum_{p \neq j}^{M} P_p \bar{\mathbf{h}}_j^p (\bar{\mathbf{h}}_j^p)^H) \end{aligned} \quad (37)$$

At this point, based on Theorem 1, the following set of second-order cone constraints serves as a convex approximation of the chance constraint (23):

$$d_j \geq c(\varepsilon) \left\|\mathbf{U}_j^{\frac{1}{2}} \mathbf{X}\right\|, \quad \forall j \in \mathcal{J} \quad (38)$$

Consequently, optimization problem (21) can be simplified as

$$\min_{\mathbf{X}} \quad \mathrm{Tr}(\mathbf{D}\mathbf{X}) \quad (39a)$$

$$\text{s.t.} \quad d_j \geq c(\varepsilon) \left\|\mathbf{U}_j^{\frac{1}{2}} \mathbf{X}\right\|, \quad \forall j \in \mathcal{J} \quad (39b)$$

$$\mathrm{Rank}(\mathbf{X}) = 1 \quad (39c)$$

The rank constraint in (39) is not convex. By dropping this constraint and using SDP relaxation the problem becomes convex and can be solved efficiently using CVX software [28]. The relaxed SDP problem is

$$\min_{\mathbf{X}} \quad \mathrm{Tr}(\mathbf{D}\mathbf{X}) \quad (40a)$$

$$\text{s.t.} \quad d_j \geq c(\varepsilon) \left\|\mathbf{U}_j^{\frac{1}{2}} \mathbf{X}\right\|, \quad \forall j \in \mathcal{J} \quad (40b)$$

Moreover, the randomization procedure described for problem (19) can be used in a similar way to extract the desired beamfoming weights from the solution of problem (40). The only difference between the the application of this procedure to (19) and (40) lies in the determination of the parameter λ. For (40), the following optimization problem should be solved:

$$\begin{aligned} \min_{\lambda} \quad & \lambda \\ \text{s.t.} \quad & \lambda a_j - c(\varepsilon) b_j \geq 0, \quad \forall j \in \mathcal{J} \end{aligned} \quad (41)$$

where

$$\begin{aligned} a_j =& \mathrm{vec}(\mathbf{w}_c \mathbf{w}_c^H)^H \mathrm{vec}(P_j \bar{\mathbf{h}}_j^j (\bar{\mathbf{h}}_j^j)^H) \\ &- \gamma_j \sigma_x^2 \mathrm{vec}(\mathbf{w}_c \mathbf{w}_c^H) \mathrm{diag}(\mathrm{vec}(\mathbf{I})) \mathrm{vec}(\mathbf{g}_j \mathbf{g}_j^H) \\ &- \gamma_j \mathrm{vec}(\mathbf{w}_c \mathbf{w}_c^H)^H \mathrm{vec}(\sum_{p \neq j}^{M} P_p \bar{\mathbf{h}}_j^p (\bar{\mathbf{h}}_j^p)^H) \\ b_j =& \left\|\mathbf{U}_j^{\frac{1}{2}} \mathbf{w}_c \mathbf{w}_c^H\right\| \end{aligned}$$

The proposed method can also be implemented as per Algorithm 1, but this time solving problem (40) in step 6.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

Algorithm 1 Robust beamforming design with randomization method

Input: $\mathbf{X_{opt}}$
Output: The rank one solution of $\mathbf{X_{opt}}$, $\mathbf{w}$

Compute SVD decomposition $\mathbf{X} = \mathbf{U}\mathbf{V}\mathbf{U}^H$ after solving (19) or (40)
$i = 1$
while $i \leq N_{max}$ **do**
  Generate a complex Gaussian random vector $\boldsymbol{\xi} \sim \mathcal{CN}(0, \mathbf{I})$
  Generate a candidate vector as $\mathbf{w}_c^i = \mathbf{U}\mathbf{V}^{1/2}\boldsymbol{\xi}$
  Solve the optimization problem (20) or (41) and obtain λ^i
  if the optimization problem is infeasible **then**
    discard and return to step 4
  else
    Store $\mathbf{w}_c^i$ and the corresponding λ^i and relay transmit power $\lambda^i \left(\mathbf{w}_c^i \mathbf{D} \mathbf{w}_c^i\right)$
    $i = i + 1$
  end if
end while
Select $\lambda^{opt} = \lambda^i$ and $\mathbf{w}_c^{opt} = \mathbf{w}_c^i$ in which λ^i and $\mathbf{w}_c^i$ correspond to the minimum relay transmit power
Output best candidate vector is $\mathbf{w}^{opt} = \sqrt{\lambda^{opt}}\mathbf{w}_c^{opt}$ and the minimum objective function $\lambda^{opt}\left(\mathbf{w}_c^{opt}\mathbf{D}\mathbf{w}_c^{opt}\right)$

IV. Channel Estimation

In practice, when considering (19) and (40), the required CSI for both types of channels, i.e., from the transmitters to relays and from relays to receivers, is not readily available. In this Section, we focus on the estimation of the unknown channel coefficients and their covariance matrices using a KF formalism.

The observation equations at time step k can be expressed as

$$\mathbf{z}_m^j(k) = \mathbf{h}\left(\mathbf{f}(k), \mathbf{g}_j(k)\right) + \mathbf{v}_m^j(k), \quad \forall j \in \mathcal{J} \tag{42}$$

where $\mathbf{f}(k) = \left[\mathbf{f}_1^T(k), \ldots, \mathbf{f}_M^T(k)\right]^T$, $\mathbf{z}_m^j(k) = [\mathbf{x}(k), y_j(k)]^T$, $\mathbf{v}_m^j(k) = [\mathbf{v}_m^{\mathbf{x}}(k)^T, \mathbf{v}_m^{y_j}(k)]^T$, $\mathbf{h}(\cdot,\cdot)$ is a non-linear function of states, and the subscript m denotes the measured values. We note that at each time step, once we have solved the optimization problems in (19) or (40), the new (i.e., updated) weight vector causes the measurements in (42) to be updated, which justifies the presence of a the feedback loop in Fig. 2. By invoking the Markov model, the recursive update for the channel coefficients from transmitters to relays and from relays to jth receiver are respectively formulated as

$$\mathbf{f}\left(k+1\right) = \alpha \mathbf{f}\left(k\right) + \mathbf{v}_s^{\mathbf{f}}(k) \tag{43}$$

$$\mathbf{g}_j\left(k+1\right) = \beta \mathbf{g}_j\left(k\right) + \mathbf{v}_s^{\mathbf{g}_j}(k), \quad \forall j \in \mathcal{J} \tag{44}$$

where $\alpha = J_0(2\pi F_R T_s)$ and $\beta = J_0(2\pi F_D T_s)$ are the temporal correlation coefficients and $J_0(\cdot)$ is the Bessel function of the first kind of order zero. For a constant symbol duration T_s, the parameters F_R and F_D are the Doppler frequency shifts, which characterize the effect of mobility at the relays and receiver nodes, respectively. We note that these correlation coefficients should have a magnitude less than unity, i.e. $|\alpha|, |\beta| < 1$, to ensure system stability. The collection of state variables at time step k is represented by $\boldsymbol{\xi}_j(k) = \left[\mathbf{f}^T(k) \;\; \mathbf{g}_j^T(k)\right]^T$ which is an $(LM + L) \times 1$ column vector. In light of (43)-(44), the state equations at time step k are

$$\boldsymbol{\xi}_j\left(k+1\right) = \mathbf{F}\boldsymbol{\xi}_j\left(k\right) + \mathbf{v}_s^j(k), \quad \forall j \in \mathcal{J} \tag{45}$$

where $\mathbf{F} = \mathrm{BD}\left(\alpha \mathbf{I}_{LM}, \beta \mathbf{I}_L\right)$, and $\mathbf{v}_s^j(k) = [\mathbf{v}_s^{\mathbf{f}}(k)^T, \mathbf{v}_s^{\mathbf{g}_j}(k)^T]^T$ is the state noise which has a Gussian distribution with covariance matrix $\mathbf{Q} = \sigma_s^2 \mathbf{I}$. Here, the subscript s denotes the state equation. In (42), the observations are not a linear function of the channel coefficients, so we need to utilize the non-linear versions of the KF for CSI estimation. In the following sub-sections, we specifically consider the use of the EKF and CKF for overcoming the non-linearity of the observation equations.

A. Extended Kalman Filtering Approach

The EKF linearizes the nonlinear measurement functions locally using the Taylor series expansion at the current best estimate of the state. By proceeding in this way, Kalman's original theory can be adapted to nonlinear systems. The Jacobian matrix of (42), $\mathbf{H}_j(k) = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\xi}_j(k)}$, is represented as

$$\mathbf{H}_j(k) = \begin{bmatrix} \mathbf{s} \otimes \mathbf{I}_L & \mathbf{0}_L \\ \mathbf{s} \otimes \mathbf{g}_j^T(k-1)\mathbf{W}^H & \sum_{p=1}^{M} s_p \mathbf{f}_p^T(k-1)\mathbf{W}^H \end{bmatrix} \tag{46}$$

where $\mathbf{s} = \left[s_1, s_2, ..., s_M\right]$ and $j \in \mathcal{J}$. In this way, the observation equations could be regarded as

$$\mathbf{z}_m^j\left(k\right) = \mathbf{H}_j\left(k\right)\boldsymbol{\xi}_j\left(k\right) + \mathbf{v}_m^j(k) \tag{47}$$

By using the extended Kalman filtering approach, we have

$$\boldsymbol{\xi}_j\left(k\right) = \boldsymbol{\xi}_j\left(k-1\right) + \boldsymbol{\Omega}\left(k\right)\left(\mathbf{z}_m^j(k) - \mathbf{z}^j\left(k\,|k-1\right)\right) \tag{48}$$

where $\boldsymbol{\Omega}\left(k\right)$ is the Kalman gain at time step k. The predicted measurement and Kalman filter gain are given by

$$\mathbf{z}^j\left(k\,|k-1\right) = \mathbf{h}(\alpha \mathbf{f}_p\left(k-1\right), \beta \mathbf{g}_j\left(k-1\right)) \tag{49}$$

$$\boldsymbol{\Omega}\left(k\right) = \mathbf{P}\left(k|k-1\right)\boldsymbol{\Upsilon}_k(\boldsymbol{\xi}_j(k))\mathbf{S}(k)^{-1} \tag{50}$$

where $\Upsilon_k(\boldsymbol{\xi}_j(k)) = \mathbf{H}_j^H\left(k, \mathbf{F}^T\boldsymbol{\xi}_j(k)\right) = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\xi}_j(k)}\big|_{\boldsymbol{\xi}_j(k)=\mathbf{F}^T\boldsymbol{\xi}_j(k)}$. Introducing index $l = k-1$, the covariance matrices can be obtained from

$$\mathbf{S}(k) = \boldsymbol{\Upsilon}_k(\boldsymbol{\xi}_j(l))^H \mathbf{P}\left(k|l\right)\boldsymbol{\Upsilon}_k(\boldsymbol{\xi}_j(l)) + \mathbf{R}(l) \tag{51}$$

$$\mathbf{P}\left(k|l\right) = \mathbf{F}^T\mathbf{P}\left(l|l\right)\mathbf{F} + \mathbf{Q}(l) \tag{52}$$

$$\mathbf{P}(k|k) = \boldsymbol{\psi}_k\mathbf{P}(k|l)\boldsymbol{\psi}_k^H + \boldsymbol{\Omega}(k)\mathbf{R}(l)\boldsymbol{\Omega}(k)^H \tag{53}$$

where $\boldsymbol{\psi}_k = \{\mathbf{I} - \boldsymbol{\Omega}(k)\boldsymbol{\Upsilon}_k(\boldsymbol{\xi}_j(k))^H\}$. The recursive algorithm for CSI estimation starts with an initial channel vector estimate $\boldsymbol{\xi}_j(0)$ with associated covariance matrix $\mathbf{P}(0|0)$, and then proceeds by udpating the channel vector estimate $\boldsymbol{\xi}(k)$ and related quantities at each time step k. Although the EKF is popular due to its simplicity, it suffers from the following limitations:

- It diverges in 'highly' nonlinear systems, pointing to its limited approximation capability.
- It often yields an underestimated error covariance.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

- Its application is limited to differentiable functions.

Because of these drawbacks, we next consider the CKF method for the purpose of CSI estimation.

B. Cubature Kalman Filter Approach

The CKF method provides an approximation to the optimal Bayesian filter; it has been shown to attain a better performance than EKF and the Unscented Kalman Filter (UKF) [21]. A single iteration of the CKF algorithm for the CSI estimation problem under consideration is presented in Algorithm 2. The parameter n, which denotes the number of cubature points is set to 5000 in our simulation study.

The worst case complexity of EKF for our estimation problem is given by $O((LM+L)^2)$ [22], and the worst case complexity of CKF is on the same order as EKF. In general, the CKF method exhibits faster convergence towards the optimal solution since it does not need to calculate the Jacobian of the non-linear functions. We emphasize that in the current application, the complexity order of channel estimation is negligible compared to that of the power allocation problem.

V. Multiple Models Channel Estimations

In this section, we introduce a channel estimation method based on the mixture of multiple Kalman filter models in order to extend the robust beamformer design to non-stationary environments. Consider a situation where the statistical properties of the radio environment vary too rapidly. In this case, the difference between the measurement vector and its prediction remains large and the stationary model cannot track the CSI. To overcome this limitation, we need to employ a modeling approach that can cover both stationary and non-stationary conditions. In the literature, the Interacting Multiple Models (IMM) method has been utilized in many applications involving the interaction of multiple processes with different dynamic behaviors, e.g., maneuvering target tracking [31], etc. The IMM algorithm runs several Kalman filters in parallel, wherein the individual filters are initialized at each step using a mixture of results from the previous steps [22]. As seen in Fig. 3, which illustrates the main processing steps of the IMM algorithm, the overall state estimation at the output is a mixture of the individual filter estimates. The IMM algorithm processes all the models simultaneously and switches between them according to their updated weights. In effect, this method resolves the dynamic model uncertainty by using multiple models simultaneously and sensibly combining their outputs.

Assume that at each time step, the output of the true time-varying model may be the result of one among N possible models with different dynamics ($N = 2$ in our case). A closed-form solution at time step k requires running N^k permutation of Kalman filters for each possible history of states, which is not feasible. In the IMM approach, at time step k, the state estimate is computed under $N \geq 2$ models, with each filter using a

Algorithm 2 Single iteration of CKF algorithm for CSI estimation

For ease of notation, we let $l = k-1$. We also indicate the time step index by means of subscripts (e.g. $\boldsymbol{\xi}^j\,(k|k) \rightarrow \boldsymbol{\xi}^j_{k|k}$).

$j = 1$
while $j \leq 2n$ **do**
  Evaluate the cubature points:

$$\boldsymbol{\chi}^j_{l|l} = \boldsymbol{\xi}_{l|l} + \sqrt{\mathbf{P}_{l|l}}\boldsymbol{\zeta}_j$$

  where $\boldsymbol{\zeta}_j = \sqrt{n}\mathbf{e}_j$, and $\{\mathbf{e}_j\}_{j=0}^{2n-1}$ defines a complete symmetric set of cubature points[7].
  Evaluate the propagated cubature points through the dynamic model:

$$\boldsymbol{\chi}^j_{k|l} = \mathbf{F}\boldsymbol{\chi}^j_{l|l} + \mathbf{v}_s$$

  $j = j + 1$
end while
Estimate the predicted state and error covariance matrix:

$$\boldsymbol{\xi}_{k|l} = \frac{1}{2n}\sum_{i=1}^{2n}\boldsymbol{\chi}^i_{k|l}$$

$$\mathbf{P}_{k|l} = \tfrac{1}{2n}\sum_{i=1}^{2n}((\boldsymbol{\chi}^i_{k|l} - \boldsymbol{\xi}_{k|l})(\boldsymbol{\chi}^i_{k|l} - \boldsymbol{\xi}_{k|l})^H)$$

$i = 1$
while $i \leq 2n$ **do**
  Form the cubature points:

$$\boldsymbol{\chi}^i_{k|l} = \sqrt{\mathbf{P}_{k|l}}\mathbf{F}\boldsymbol{\zeta_i} + \boldsymbol{\xi}_{k|l}$$

  Propagate cubature points through the measurement model:

$$\mathbf{Z}^i_{k|l} = \boldsymbol{h}(\boldsymbol{\chi}^i_{k|l})$$

  $i = i + 1$
end while
Estimate the predicted measurement:

$$\mathbf{z}\,(k|l) = \frac{1}{2n}\sum_{i=1}^{2n}\mathbf{Z}^i_{k|l}$$

Estimate the innovations covariance matrix:

$$\mathbf{S}_k = \tfrac{1}{2n}\sum_{i=1}^{2n}(\mathbf{Z}^i{}_{k|l} - \mathbf{z}_{k|l})(\mathbf{Z}^i{}_{k|l} - \mathbf{z}_{k|l})^H + \mathbf{R}_k$$

Estimate the cross-covariance matrix:

$$\mathbf{C}_k = \tfrac{1}{2n}\sum_{i=1}^{2n}(\boldsymbol{\chi}^i_{k|l} - \boldsymbol{\xi}_{k|l})(\mathbf{Z}^i{}_{k|l} - \mathbf{z}_{k|l})^H$$

Estimate the cubature Kalman gain:

$$\boldsymbol{\Omega}_k = \mathbf{C}_k\mathbf{S}_k^{-1}$$

Estimate the updated state:

$$\boldsymbol{\xi}_{k|k} = \boldsymbol{\xi}_{k|l} + \boldsymbol{\Omega}_k(\mathbf{z}_m - \mathbf{z}_{k|l})$$

Estimate the error covariance:

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|l} - \boldsymbol{\Omega}_k\mathbf{S}_k\boldsymbol{\Omega}_k^T$$

[7]The cubature points $\mathbf{e}_j$ can be obtained by permutating and changing the sign of the generator in all possible ways. For example, $\mathbf{e}_j \in \mathbb{R}^2$ is taken from the following set of points: $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}-1\\0\end{bmatrix}, \begin{bmatrix}0\\-1\end{bmatrix}\right\}$

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

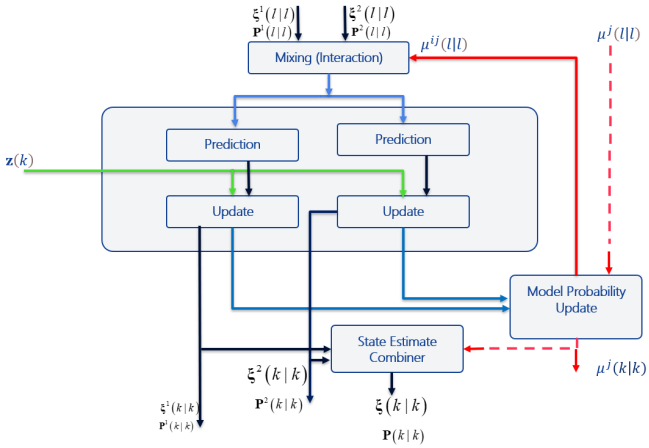

Figure 3: Block diagram of a single step of the IMM algorithm for $N = 2$ interacting models.

different combination of the previous model conditional estimates. The model switching process is assumed to follow a Markov chain with known transition probability matrix $[p_{ij}]$. Application of the IMM algorithm requires three components. The first is a set of N Kalman filters, one for each models or modes of operation. The second is a probability vector, that contains the probabilities that the ith model is in effect at the current time step, k. The third is a transition probability matrix that specifies how probable it is to jump from model i at time k to model j at time $k+1$. A single iteration of the IMM filtering method is presented as Algorithm 3. In our implementation of this algorithm, we set $N = 2$, where our aim is to allow for the combination of a stationary model and a non-stationary one.

VI. Performance Considerations

A. Complexity Analysis

In this section, we evaluate the computational complexity of our proposed optimization algorithms and compare it with that of the non-robust and existing robust methods. Let us define a mixed conic matrix variable $\mathbf{X}$ as $\mathbf{X} = [\mathbf{X}_q, \mathbf{X}_s]^T$, where $\mathbf{X}_q$ belongs to a quadratic cone $\mathcal{K}^Q$ and $\mathbf{X}_s$ belongs to a semi-definite cone $\mathcal{K}^S$. Generally, a mixed semi-definite programming (SDP) and second-order cone programming (SOCP) optimization problem for a linear objective function f, can be formulated as a standard linear conic problem with the following structure:

$$\min_{\mathbf{X}} \quad f(\mathbf{X}) \tag{54a}$$

$$\text{s.t.} \quad \mathbf{A}_i \circ \mathbf{X} + b_i = 0 \quad \forall i = 1, 2, \dots M \tag{54b}$$

$$\mathbf{X} \in \mathcal{K}^Q \times \mathcal{K}^S \tag{54c}$$

where $\mathbf{A}_i \circ \mathbf{X}$ stands for the inner product of matrices $\mathbf{A}_i$ and $\mathbf{X}$. Generally, the worst-case complexity for solving mixed SD/SOCP using the interior point method (IPM)

Algorithm 3 Single iteration of IMM filter for mixing stationary and non-stationary models ($N = 2$)

For ease of notation, we let $l = k-1$ and we show the time index by a subscript (e.g. $\boldsymbol{\xi}^j(k|k) \rightarrow \boldsymbol{\xi}^j_{k|k}$).

Input: The previous sufficient statistics $\left\{\boldsymbol{\xi}^j_{l|l}, \mathbf{P}^j_{l|l}, \mu^j_l\right\}_{j=1}^N$

Output: The current sufficient statistics $\left\{\boldsymbol{\xi}^j_{k|k}, \mathbf{P}^j_{k|k}, \mu^j_k\right\}_{j=1}^N$

1: $i = 1, j = 1$
2: **while** $i \leq N$ **do**:
3: **while** $j \leq N$ **do**:
4: Mixing:

- Calculate the mixing probabilities $\mu^{ij}_{l|l}$ as:

$$\mu^{ij}_{l|l} = \frac{p_{ji}\mu^j_l}{\sum_{m=1}^{N} p_{li}\mu^m_l}$$

- Calculate the mixed estimates $\boldsymbol{\xi}^{0i}_{l|l}$ and covariance $\mathbf{P}^{0i}_{l|l}$ as

$$\boldsymbol{\xi}^{0i}_{l|l} = \sum_{j=1}^{N} \mu^{ji}_{l|l}\boldsymbol{\xi}^{0j}_{l|l}$$

$$\mathbf{P}^{0i}_{l|l} = \sum_{j=1}^{N} \mu^{ji}_{l|l}\mathbf{P}^{0j}_{l|l} + (\boldsymbol{\xi}^{0j}_{l|l} - \boldsymbol{\xi}^{0i}_{l|l})(\boldsymbol{\xi}^{0j}_{l|l} - \boldsymbol{\xi}^{0i}_{l|l})^H$$

5: $j = j + 1$
6: **end while**
7: Mode Matched Prediction Update: For ith model calculate the predicted estimate $\boldsymbol{\xi}^i_{l|l}$ and the covariance $\mathbf{P}^i_{l|l}$

$$\boldsymbol{\xi}^i_{l|l} = \mathbf{F}^i\boldsymbol{\xi}^{0i}_{l|l}$$

$$\mathbf{P}^i_{l|l} = \mathbf{F}^i\mathbf{P}^{0i}_{l|l}{\mathbf{F}^i}^T + \mathbf{Q}_l$$

8: Mode Matched Measurement Update: For ith model calculate:

- the updated estimate $\boldsymbol{\xi}^i_{k|k}$ and covariance $\mathbf{P}^i_{k|k}$

$$\boldsymbol{\xi}^i_{k|k} = \boldsymbol{\xi}^i_{k|l} + \boldsymbol{\Omega}^i_k(\mathbf{z}^i_m - \mathbf{z}^i_{k|l})$$

$$\mathbf{P}^i_{k|k} = \mathbf{P}^i_{k|l} - \boldsymbol{\Omega}^i_k\mathbf{S}^i_k\left(\boldsymbol{\Omega}^i_k\right)^H$$

$$\mathbf{z}^i_{k|l} = \mathbf{H}^i\boldsymbol{\xi}^i_{k|l}$$

$$\mathbf{S}^i_k = \mathbf{H}^i{\mathbf{P}^i}_{k|l}{\mathbf{H}^i}^H + \mathbf{R}_l$$

$$\boldsymbol{\Omega}^i_k = \mathbf{P}^i_{k|l}{\mathbf{H}^i}^H(\mathbf{S}^i_k)^{-1}.$$

- the updated mode probability μ^i_k as

$$\mu^i_k = \frac{\mathcal{N}(\mathbf{z}_k; \mathbf{z}^i_{k|k-1}, \mathbf{S}^i_k)\sum_{j=1}^{N} p_{ji}\mu^j_{k-1}}{\sum_{m=1}^{N}\left(\mathcal{N}(\mathbf{z}_k; \mathbf{z}^m_{k|l}, \mathbf{S}^m_k)\sum_{j=1}^{N} p_{jm}\mu^j_m\right)}$$

9: Output Estimate Calculation: Calculate the overall estimate $\boldsymbol{\xi}_{k|k}$ and covariance $\mathbf{P}_{k|k}$ as

$$\boldsymbol{\xi}_{k|k} = \sum_{i=1}^{N} \mu^i_k\boldsymbol{\xi}^i_{k|k}$$

$$\mathbf{P}_{k|k} = \sum_{i=1}^{N} \mu^i_k(\mathbf{P}^i_{k|k} + (\boldsymbol{\xi}^i_{k|k} - \boldsymbol{\xi}_{k|k})(\boldsymbol{\xi}^i_{k|k} - \boldsymbol{\xi}_{k|k})^H)$$

10: $i = i + 1$
11: **end while**

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

[32], in terms of required floating point operations (flops), is given by

$$O(n'(m^2\sum_{i=1}^{N_{soc}} n_i^{soc} + m^2\sum_{i=1}^{N_{sd}} (n_i^{sd})^2 + m\sum_{i=1}^{N_{sd}} (n_i^{sd})^3 + m^3)\log(1/\delta)) \tag{55}$$

where δ denotes the accuracy of the optimal solution, n' denotes the number of iterations, m is the number of equality constraints, N_{sd} and N_{soc} are the numbers of SDP and SOCP constraints, respectively and n_i^{sd} and n_i^{soc} are the dimensions of the ith SDP cone and the second order cone, respectively. In Table I, we compute the values of n', m, n_i^{soc}, n_i^{sd}, N_{sd}, N_{soc} of the problems (19) and (40) and of the optimization problems in [8], [16], [6], [18], respectively. In order to obtain the computational complexity of each optimization problem, one needs to rewrite it in the standard form of (54). In order to better illustrate the meaning of the various component terms presented in Table I, let us consider a network with $L = 15$ relay nodes and $M = 3$ source-destination pairs, and set the number of iterations to and the desired accuracy of the optimal solution to $n' = 10$ and $\delta = 0.1\%$, respectively. In Table II, we compare the computational complexity of different methods for this special choice of parameters. It is observed from this Table that the complexity of the proposed AMM method based on (19) is similar to that of the the non-robust method in [8], due to its SDP problem structure, while the complexity of our SM method based on (40) exceeds that of AMM by a small margin. The proposed methods clearly exhibit the lowest complexity, and by several orders of magnitudes, when compared to the worst case method [16], SD/SOCP method [6] and adaptive approach [18]. We note that the above complexity analysis only focuses on solving the optimization problem. The complexity of the channel estimation should be included as well to obtain the total complexity. As it is well known, the computational bottleneck of the EKF estimation method is (53), which requires $O(L^3)$ multiplications (the other formulas in (47)-(52) can be computed in $O(L^2)$). However, based on [22], (53) can be replaced with:

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{\Omega}(k)\,\mathbf{S}(k)\,\mathbf{\Omega}^H(k) \tag{56}$$

whose complexity is only $O(L^2)$. Similarly, the worst-case complexity of CKF estimation is on the same order, i.e. $O(L^2)$. Consequently, when compared to the solution of the optimization problem, with complexity order $O(M^2L^{3.5})$, the complexity of the channel estimation is negligible, even for moderate values of M and L. Finally, due to linear nature of the optimization problems in (20) or (41), the computational complexity of the randomization technique is also negligible compared to the optimization in (19) or (40).

B. Channel Estimation Accuracy

It is well known that the Kalman filter is a Best Linear Unbiased Estimator (BLUE) [22]. We shall assume that the unbiasedness property extends to the EKF and CKF, which is well supported by simulation experiments. Let us rewrite the observation vectors in (42) in concatenated form as

$$\mathbf{z}_m(k) = \mathbf{h}(\boldsymbol{\xi}(k)) + \mathbf{v}_m(k) \tag{57}$$

where $\boldsymbol{\xi}$ represents the unknown state vector at time k. Hence, for a given value of the unknown parameter vector $\boldsymbol{\xi} \equiv \boldsymbol{\xi}(k)$, we have $\mathbf{z}_m(k) \sim \mathcal{CN}(\mathbf{h}(\boldsymbol{\xi}), \mathbf{R})$, where $\mathbf{R}$ denote the covariance matrix of the measurement noise, assumed to be constant. Under these conditions, the Fisher information matrix for an unbiased estimator of $\boldsymbol{\xi}$ can be represented as [22]:

$$\mathbf{J}(\boldsymbol{\xi}) = \mathrm{E}\{[\frac{\partial \mathbf{h}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}]^H \mathbf{R}^{-1} [\frac{\partial \mathbf{h}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}]\} + \frac{1}{2}\mathrm{Tr}(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \boldsymbol{\xi}})^2) \tag{58}$$

For the problem proposed in Section IV, the Fisher information matrix is

$$\mathbf{J}(\boldsymbol{\xi}) = \mathbf{H}_{\boldsymbol{\xi}}^H \mathbf{R}^{-1} \mathbf{H}_{\boldsymbol{\xi}} \tag{59}$$

where $\mathbf{H}_{\boldsymbol{\xi}} = [\frac{\partial \mathbf{h}(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}]$ is obtained by stacking the matrices $\mathbf{H}_j(k)$ in (46) for $j \in \mathcal{J}$. Then, the covariance matrix of any unbiased estimator of the channel state vector$\boldsymbol{\xi}$, denoted as $\hat{\mathbf{C}}(\boldsymbol{\xi})$, satisfies

$$\hat{\mathbf{C}}(\boldsymbol{\xi}) \succeq \mathbf{J}^{-1}(\boldsymbol{\xi}) \tag{60}$$

which sets a lower bound on the state covariance matrix of our Bayesian estimator [22]. In this work, we employ the above simplified form of the CRLB based on a single KF iteration, where the calculations are made by averaging observed values of the $\mathbf{H}_{\boldsymbol{\xi}}$ during steady-state, i.e, after convergence of the CKF algorithm[8]. Beyond the above CRLB derivation, the general convergence analysis of the proposed algorithms remains very complex because of the intricate and coupled nature of the joint channel estimation and beamforming system under study. Some of the methods and theorems presented in [34] and [35] could be used to derive general conditions under which local convergence is possible. However, these theorems are based on assumptions that are difficult to demonstrate in practice. Nevertheless, this remains an interesting avenue for future work.

[8]For the dynamic state-space channel model under consideration, the CRLB should exhibit a time-varying behavior during the transient period following the onset of operation, i.e., gradually decreasing towards its steady-state value as additional observations become available. In fact, such recursive behavior of the CRLB has been studied in [33] for the special case of real-valued linear state-space models; however, the results are not directly applicable to the complex-valued EKF and CKF considered in this work.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

Table I: Parameters of computational complexity of different methods

parameters	n'	m	n_i^{soc}	N_{soc}	n_i^{sd}	N_{sd}
SD/SOCP [6]	$O\left(L^{0.5}\right)$	M	L^2+1	M	$\begin{cases} M & 1\le i\le M \\ 4LM & M+1\le i\le 2M \\ M & 2M+1\le i\le 3M \\ L & i=3M+1 \end{cases}$	$3M+1$
Non-Robust [8]	$O\left(L^{0.5}\right)$	M	0	0	$\begin{cases} M & i=1 \\ L & i=2 \end{cases}$	2
Worst case [16]	$O\left(L^{0.5}\right)$	$M+L$	L^2	$5M+4L$	$\begin{cases} M & i=1 \\ L & i=2 \end{cases}$	2
Adaptive approach [18]	$O\left(L^{0.5}\right)$	M	0	0	$\begin{cases} L^2+1 & 1\le i\le M \\ L & i=M+1 \end{cases}$	$M+1$
AMM in (19)	$O\left(L^{0.5}\right)$	M	0	0	$\begin{cases} M & i=1 \\ L & i=2 \end{cases}$	2
SM in (40)	$O\left(L^{0.5}\right)$	M	L^2+1	M	0	0

Table II: Computational complexity of different methods (in flops)

Different methods	SD/SOCP [6]	Non-Robust [8]	Worst case [16]	Adaptive approach [18]	AMM in (19)	SM in (40)
Computational complexity	1.6×10^9	3.7×10^5	1.7×10^8	6.96×10^{10}	3.7×10^5	4.2×10^5

VII. SIMULATION RESULTS

In this section, we present the results of Monte-Carlo simulations in order to illustrate the merits of our proposed approaches and compare their performance to various benchmarks.

A. Methodology

We consider a network with two transmitter-receiver pairs ($M = 2$) and $L = 20$ relays. We make the following assumptions about the data transmission[9]: As indicated in Section II.A, independent complex Rayleigh flat fading channel gains are employed to model the various radio links. There is no direct link between the transmitters and the receivers. The noise variance is -10 dBw, the source power is 0 dBw and the error variance of each channel is -20 dB. The true channel vectors are generated according to model (43)-(44) with parameters $\alpha = \beta = 0.99$, which corresponds to slowly-varying channels. BPSK modulation is assumed for data transmission.

Unless otherwise specified, the SINR threshold γ_j in (11b) is set to 5dB. To implement the IMM method, we consider a mixture of two models, referred to as Model-1 and Model-2 and respectively characterized with: $\mathbf{F}^{(1)} = \mathrm{BD}(\alpha^{(1)}\mathbf{I}_{LM}, \beta^{(1)}\mathbf{I}_{LM})$ and $\mathbf{F}^{(2)} = \mathrm{BD}(\alpha^{(2)}\mathbf{I}_{LM}, \beta^{(2)}\mathbf{I}_{LM})$ where $\alpha^{(1)} = \beta^{(1)} = 0.99$, $\alpha^{(2)} = \beta^{(2)} = 0.9$, and the measurements noise matrices are $\mathbf{R}_1 = \sigma_1^2\mathbf{I}_{L+1}$ and $\mathbf{R}_2 = \sigma_2^2\mathbf{I}_{L+1}$ where $\sigma_1^2 = -10$ dBw, and $\sigma_2^2 = 5$ dBw. Model-1 indeed corresponds to the true slowly-varying channel generation model, while Model-2 corresponds to an incorrect fast-varying channel model. The matrix of transition probabilities for the IMM is chosen as $[p_{ij}] = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}$, with initial model probabilities $[0.9 \;\; 0.1]$.

The following methods are compared:

- SDP/SOCP: The robust method in [6] which is a combination of SDP and SOCP problems.
- Full-CSI: The non-robust method in [8] which uses the true channel coefficients.
- Adaptive: The robust adaptive approach in [18].
- SM: Stochastic beamforming method in (40).
- AMM: Approximate mean method in (19).

In our experiment, we compare the performance of the EKF, CKF, IMM-EKF and IMM- CKF methods for channel estimation when used in connection with the proposed AMM and SM robust beamforming methods. For IMM EKF and IMM CKF, we use the parameter settings described above for Model-1 and Model-2. For EKF and CKF, we consider two different implementations of these algorithms respectively based on the parameters settings of (true) Model-1 and (incorrect) Model-2; we refer to the corresponding implementations as EKF/CKF Model-i, where $i = 1, 2$.

B. Evaluation

In Fig. 4, we plot the minimum transmission power required at the relays versus the target SINR. It can be seen that, as the measurement error covariance increases, more power is required. We first note that for both proposed beamforming methods, the use of CKF for channel tracking outperforms the EKF. Moreover, both proposed methods outperform the methods in [18] and [6] in terms of power consumption and feasibility region. This figure also shows that the performance of the proposed real-time channel estimation using CKF for both of beamforming methods is comparable with that of the Full-CSI method in [8]. As indicated earlier, both robust optimization problems formulated in (19) and (40) are NP hard and cannot be solved without resorting

[9]We do not mention the exact values of operating frequency and signal bandwidth, as they do not directly affect the conclusions of our study and may be adjusted in practice as per the specific needs of an application. While our simulation models target modern 4G and 5G wireless systems operating at higher frequencies, our simulation setup has been properly normalized so that it does not rely on specific choices of transmission frequency or signal bandwidth (e.g., signals are modeled in terms of their complex baseband analytic representations, spectral efficiency is normalized by the system bandwidth, etc.).

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

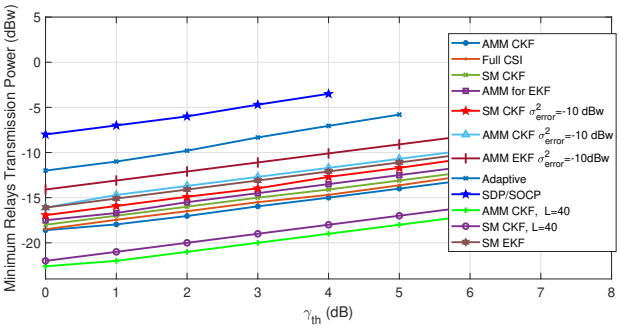

Figure 4: Minimum relays transmission power versus the SINR threshold

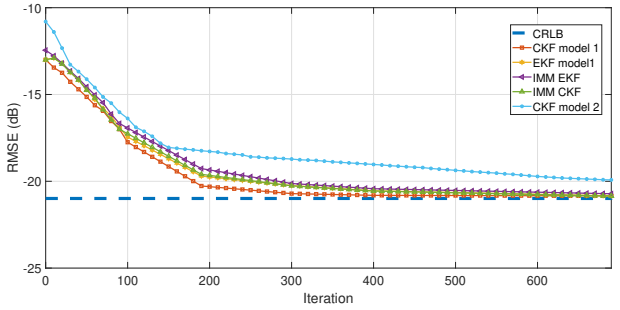

Figure 5: RMSE versus the number iterations

to approximations. Nevertheless, the above simulation results provide strong support for our proposed sub-optimal solution approaches, suggesting that they provide good approximations to the (unattainable) optimum solutions of the original NP problems.

The results of root mean square error (RMSE) for different channel estimation methods are compared in Fig. 5 and 6. We used the following definition for RMSE for channel estimation:

$$\text{RMSE} = \lim_{k\to\infty} \frac{1}{M}\sum_{j=1}^{M} \sqrt{\text{E}(\|\boldsymbol{\xi}_j(k) - \hat{\boldsymbol{\xi}}_j(k)\|_2^2)} \qquad (61)$$

where the ensemble average E is obtained by averaging over multiple channel realizations. We only use AMM for this comparison, but similar results are obtained with SM. In Fig. 5, we plot the RMSE for different approaches and channel models as a function of the KF iteration number. It is observed that he best results are obtained with CKF Model-1, closely followed by EKF Model 1. This is plausible since in this case, the parameters of the CKF and EKF trackers are perfectly matched to those of the true channel. Conversely, the worst results are obtained with CKF Model-2 and EKF Model-2 (in fact, the result for the latter algorithm are not shown as the filter diverges). In this case, the mismatch between the true and assumed model parameters results in significant deviations between the estimated and true channel state vectors. Interestingly, the RMSE for the proposed IMM-CKF and IMM-EKF methods achieves a useful compromise between the performance of the CKF Model-1 and CKF Model-2 filters (with IMM-CKF showing slightly better performance than IMM-EKF). This shows that by mixing multiple models for slowly and rapidly varying channels, IMM can effectively avoid severe performance degradations due to model mismatch. The steady-state CRLB calculated in [18] is also shown as a dashed horizontal line in this figure. We can see that the CKF Model-1, EKF Model-1, IMM-CKF and IMM-EKF all reaches the CRLB as the number of iterations increases, illustrating the efficiency of the proposed methods. channel model (Model-1), and IMM-CKF achieve the CRLB which shows the efficiency of the proposed method. However, the CKF Model-2 does not reach the CRLB due to model mismatch. It is indeed quite remarkable that the IMM allows convergence to the CRLB.

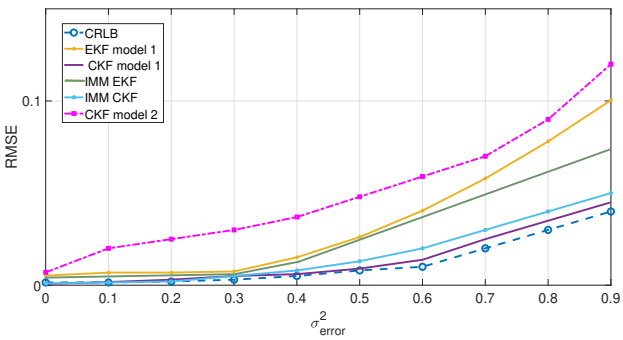

Figure 6: RMSE versus the measurement error level

Fig. 6 evaluates the performance of the different channel estimation methods versus measurement error level. As shown in this figure, CKF Model-1 whose parameters are matched to the true model achieves the best performance, closely followed by IMM-CKF. For larger values of measurement errors, the performance of EKF Model-1 and IMM EKF deviate in a more pronounced way from the CRLB. The deleterious effects of severe parameter mismatch are clearly exemplified by the overall poor performance of CKF Model-2. The results for EKF Model-2 are not shown since the algorithm diverges. These results further demonstrate the benefits of combining multiple state-space models through the IMM framework.

In Fig. 7, we plot the spectral efficiency (SE) of the proposed methods as a function of the minimum relay transmission power. Based on the definition in [36], the SE for a Gaussian channel can be expressed as[10]:

$$\text{SE} = \frac{1}{W}\sum_{j=1}^{M} D_j = \sum_{j=1}^{M} \log\left(1 + \text{SINR}_j\right) \qquad (62)$$

where D_j and W denote the data rate at the jth receiver and the system bandwidth, respectively. The results show

[10]For a channel with bandwidth W, the corresponding channel capacity C can be obtained as $C = W\text{SE}$.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

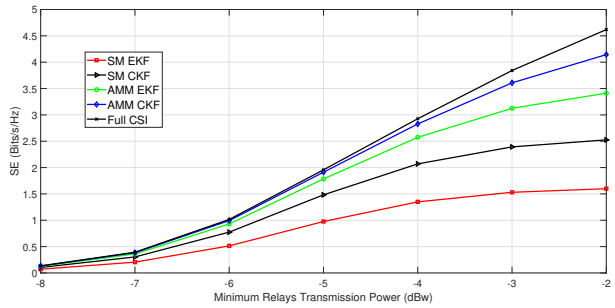

Figure 7: Comparison of SE versus minimum relay transmit power.

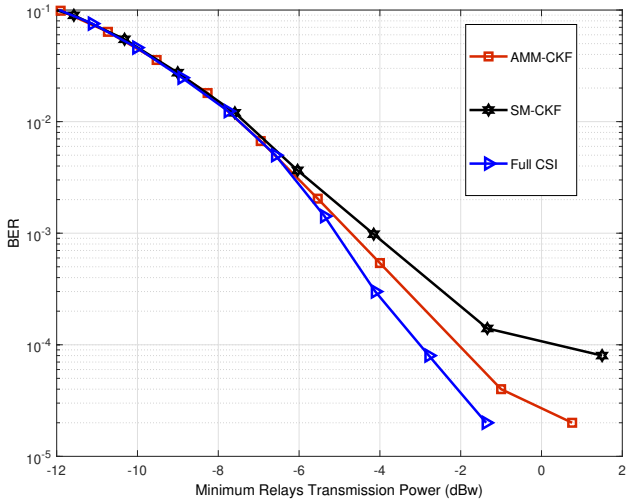

Figure 8: Comparison of BER versus the minimum relay transmit power.

that the SE for AMM CKF comes very close to that of the Full CSI approach over a wide range of relay transmission powers. For low to moderate levels of transmission power, the other approaches also exhibit similar SE levels. Moreover, Table III compares the ratio of system throughput to total power consumption for the different system configuration under study. In Fig. 8, we

Table III: Comparing the ratio of system throughput to total relay power consumption

$P_T(dBw)$	SM-EKF	SM-CKF	AMM-EKF	AMM-CKF	Full-CSI
-3	-0.50	-0.79	-1.05	-1.20	-1.29
-2	-0.81	-1.26	-1.72	-2.08	-2.32

compare AMM and SM methods by using CKF as a CSI estimator by using BER metric. As the result of this comparison shows, both of the proposed optimization approaches are almost identical with Full CSI method for low and moderate minimum relays transmission power.

VIII. CONCLUSION

In this paper, we proposed different methods for joint channel estimation and robust beamforming in a peer-to-peer communications system using a network of relays. Our methods were based on EKF and CKF for estimating the transmitter-relay and relay-receiver channels. The estimated CSI were used through a feedback mechanism for solving two robust beamforming problems, wherein the aim is to minimize the total transmission power of the relays subject to SINR constraint at each of the receiver nodes. Moreover, the IMM approach for mixing the non-stationary and stationary channel models was employed. Through numerical simulations, the proposed recursive CSI estimation methods were shown to be efficient, i.e., unbiased and converging to the steady-state. Furthermore, our results confirm the better performance of the proposed robust relay beamforming design algorithms compared to existing methods in terms of relevant transmission metrics, including relay power consumption and spectral efficiency.

While our approach assumes narrow-band flat fading channels, it can be extended to wide-band transmissions by applying it independently to individual sub-carriers of a multi-carrier system. However, this may require additional modeling assumptions (related to, e.g., relay power allocation or use of KF-based tracking across multiple subbands) and hence, further studies are needed to characterize the potential merits of this approach. Finally, we note that transmission delays within the feedback loop between the relays and the CPU will introduce additional network latency. In this regard, solving the problem in a distributed way instead of using a CPU would reduce network latency, at the cost of possible loss in performance. This remains an interesting avenue for future studies.

APPENDIX I: DEFINITION OF MATRICES

Let $\bar{\mathbf{F}}_p = \text{diag}\left(\bar{\mathbf{f}}_p\right)$, $\bar{\mathbf{G}}_j = \text{diag}\left(\bar{\mathbf{g}}_j\right)$, $\bar{\mathbf{H}}_j^p = \text{diag}\left(\bar{\mathbf{h}}_j^p\right)$, $\mathbf{X} = \mathbf{w}\mathbf{w}^H$, $\mathcal{W} = \mathbf{W}\mathbf{W}^H$, $\mathcal{H}_j^p = \bar{\mathbf{h}}_j^p\bar{\mathbf{h}}_j^{pH}$, where $j,p \in \mathcal{J}$ and $j \neq p$. The matrices $\mathbf{Q_f}, \mathbf{Q_g}, \mathbf{Q_{f,g}}, \widehat{\mathbf{Q}}_{\mathbf{f,g}}, \mathbf{c_f}, \mathbf{c_g}$, and scalar d_j which have been derived in [6], are defined as:

$$\mathbf{Q_f} = \text{BD}([\mathbf{Q_f}]_{11}, [\mathbf{Q_f}]_{22}, \ldots, [\mathbf{Q_f}]_{jj}, \ldots, [\mathbf{Q_f}]_{MM}) \tag{63}$$

$$[\mathbf{Q_f}]_{jj} = P_j(\bar{\mathbf{G}}_j\mathbf{X}\bar{\mathbf{G}}_j^H) - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{G}}_p\mathbf{X}\bar{\mathbf{G}}_p^H$$

$$\mathbf{Q}_{\mathbf{g}_j} = \frac{1}{M}\text{BD}(\mathbf{Q'}_{\mathbf{g}_j}, \mathbf{Q'}_{\mathbf{g}_j}, \ldots, \mathbf{Q'}_{\mathbf{g}_j}) \tag{64}$$

$$\mathbf{Q'}_{\mathbf{g}_j} = P_j\bar{\mathbf{F}}_j^H\mathbf{X}\bar{\mathbf{F}}_j - \gamma_j\sigma_x^2\mathcal{W} - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{F}}_p^H\mathbf{X}\bar{\mathbf{F}}_p$$

$$\mathbf{Q_{f,g}}=\text{BD}\left([\mathbf{Q_{f,g}}]_{11}, [\mathbf{Q_{f,g}}]_{22}, \ldots, [\mathbf{Q_{f,g}}]_{MM}\right) \tag{65}$$

$$[\mathbf{Q}_{\mathbf{f},g}]_{jj} = P_j\bar{\mathbf{F}}_j^H\mathbf{X}\bar{\mathbf{G}}_j - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{F}}_p^H\mathbf{X}\bar{\mathbf{G}}_p$$

$$\widehat{\mathbf{Q}}_{\mathbf{f,g}}^H = \text{BD}([\widehat{\mathbf{Q}}_{\mathbf{f,g}}^H]_{11}, [\widehat{\mathbf{Q}}_{\mathbf{f,g}}^H]_{22}, \ldots, [\widehat{\mathbf{Q}}_{\mathbf{f,g}}^H]_{MM}) \tag{66}$$

$$[\widehat{\mathbf{Q}}_{\mathbf{f,g}}^H]_{jj} = P_j\text{diag}(\mathbf{w})^H\bar{\mathbf{H}}_j^j\text{diag}(\mathbf{w}) - \gamma_j\sum_{p\neq j}^{M} P_p\mathbf{W}^H\bar{\mathbf{H}}_p^p\mathbf{W}$$

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

$$\mathbf{c_f} = \mathrm{BD}[[\mathbf{c_f}]_{11}^{\mathrm{H}} \dots [\mathbf{c_f}]_{\mathrm{jj}}^{\mathrm{H}} \dots [\mathbf{c_f}]_{\mathrm{MM}}^{\mathrm{H}}] \tag{67}$$

$$[\mathbf{c_f}]_{jj}^{H} = P_j(\bar{\mathbf{h}}_j^j)^H \mathbf{X}\bar{\mathbf{G}}_j - \gamma_i \sum_{p\neq j}^{M} P_j \left(\bar{\mathbf{h}}_j^p\right)^H \mathbf{X}\bar{\mathbf{G}}_j$$

$$\mathbf{c_g} = \frac{1}{M}(P_j \mathbf{F}_j^H \mathbf{X}\bar{\mathbf{h}}_j^j - \gamma_j \sum_{p\neq j}^{M} P_p \mathbf{F}_p^H \mathbf{X}\bar{\mathbf{h}}_j^p - \gamma_j \sigma_x^2 \mathcal{W}\bar{\mathbf{g}}_j) \tag{68}$$

$$d_j = \mathbf{w}^H P_j \mathcal{H}_j^j \mathbf{w} - \gamma_j \sigma_x^2 \bar{\mathbf{g}}_j^H \mathcal{W}\bar{\mathbf{g}}_j - \gamma_j \mathbf{w}^H \sum_{p\neq j}^{M} P_p \mathcal{H}_j^p \mathbf{w} \tag{69}$$

APPENDIX II: DERIVATION OF MATRIX $\mathbf{U}_j$

In this section, we derive the expression of the matrix $\mathbf{U}_j$. The following Lemma is assumed.

Theorem 2. *Let* $\mathbf{A} \in \mathbb{R}^{p\times q}$, $\mathbf{B} \in \mathbb{R}^{q\times r}$, $\mathbf{C} \in \mathbb{R}^{r\times s}$, $\mathbf{D} \in \mathbb{R}^{s\times t}$ *be matrices which jointly have a multivariate Gaussian distribution. Then the following result holds:*

$$\begin{aligned}&\mathrm{E}\{\mathbf{ABCD}\} = \mathrm{E}\{\mathbf{AB}\}\mathrm{E}\{\mathbf{CD}\} \\ &+ \sum_{k=1}^{r} \mathrm{E}\{\mathbf{e}_k^T\mathbf{C}\otimes\mathbf{A}\}\mathrm{E}\{\mathbf{D}\otimes\mathbf{Be}_k\} \\ &+\mathrm{E}\{\mathbf{A}\mathrm{E}\{\mathbf{BC}\}\mathbf{D}\} - 2\mathrm{E}\{\mathbf{A}\}\mathrm{E}\{\mathbf{B}\}\mathrm{E}\{\mathbf{C}\}\mathrm{E}\{\mathbf{D}\}\end{aligned} \tag{70}$$

where $\mathbf{e}_k$ *denote a vector having element one at the kth position and zero elsewhere. For* $r = 1$ *the mentioned expression is simplified to:*

$$\begin{aligned}&\mathrm{E}\{\mathbf{ABCD}\}=\mathrm{E}\{\mathbf{AB}\}\mathrm{E}\{\mathbf{CD}\}+\mathrm{E}\{\mathbf{C}\otimes\mathbf{A}\}\mathrm{E}\{\mathbf{D}\otimes\mathbf{B}\} \\ &+\mathrm{E}\{\mathbf{A}\mathrm{E}\{\mathbf{BC}\}\mathbf{D}\}-2\mathrm{E}\{\mathbf{A}\}\mathrm{E}\{\mathbf{B}\}\mathrm{E}\{\mathbf{C}\}\mathrm{E}\{\mathbf{D}\}\end{aligned} \tag{71}$$

Proof. The proof of this theorem can be found in [37]. Note that $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ may consist of elements which are complex-valued variables. The assumption of Gaussianity is maintained. This means that the real and imaginary parts of the entries of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are assumed to be jointly Gaussian distributed. Under these conditions, Theorem 2 continues to hold for complex $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ matrices. This claim has also been proved in [37]. □

By applying (71) with $\mathbf{A} = \Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}^{\frac{1}{2}}$, $\mathbf{B} = \mathbf{A}^H$, $\mathbf{C} = \mathbf{A}$, and $\mathbf{D} = \mathbf{A}^H$, the first term of (34) can be written as:

$$\begin{aligned}&\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\} \\ &= \mathrm{Tr}(\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\}\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\}) \\ &+\mathrm{Tr}(\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}^{\frac{1}{2}}\otimes\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}^{\frac{1}{2}}\}\mathrm{E}\{\Delta\mathbf{H}_j\tilde{\mathbf{Q}}^{\frac{1}{2}}\otimes\Delta\mathbf{H}_j\tilde{\mathbf{Q}}^{\frac{1}{2}}\}) \\ &+\mathrm{Tr}(\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\mathbf{\Sigma}\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\})\end{aligned} \tag{72}$$

Let us now focus on simplifying the term $\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\}$; other terms can be simplified by following a similar procedure.

$$\begin{aligned}\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\} &= \mathrm{Tr}(\mathrm{E}\{\Delta\mathbf{H}_j^H\tilde{\mathbf{Q}}\Delta\mathbf{H}_j\}) \\ &= \mathrm{Tr}(\mathrm{E}\{\Delta\mathbf{H}_j\Delta\mathbf{H}_j^H\}\tilde{\mathbf{Q}}) \\ &= \mathrm{Tr}(\Sigma\tilde{\mathbf{Q}}) \\ &= \mathrm{Tr}\left(\sigma_f^2\mathbf{R_I} + 2\sigma_{fg}^2\mathbf{R_{IQ}} + \sigma_g^2\mathbf{R_Q}\right) \\ &= \mathrm{Tr}\left(\sigma_f^2\mathbf{Q_f} + \sigma_g^2\mathbf{Q_g}\right) \\ &= \sigma_f^2\mathrm{vec}^H(\mathbf{X})\,\mathrm{vec}(\phi^1) \\ &+ \frac{\sigma_g^2}{M}\mathrm{vec}^H(\mathbf{X})\,\mathrm{vec}(\phi^2)\end{aligned} \tag{73}$$

The matrix $\mathbf{U}_j$ is then calculated as

$$\mathbf{U}_j \triangleq \delta^1 + \delta^2 + \delta^3 + \delta^4 + \delta^5 \tag{74}$$

where

$$\delta^1 \triangleq 2\sigma_f^4\mathrm{vec}(\phi^1)\mathrm{vec}^H(\phi^1) \tag{75}$$

$$\delta^2 \triangleq 2\sigma_g^4\mathrm{vec}(\phi^2)\mathrm{vec}^H(\phi^2) \tag{76}$$

$$\delta^3 \triangleq 4\sigma_f^2\sigma_g^2\mathrm{vec}(\phi^1)\mathrm{vec}^H(\phi^2) \tag{77}$$

$$\delta^4 \triangleq \sigma_f^4\mathrm{vec}(\phi^3)\mathrm{vec}^H(\phi^3) \tag{78}$$

$$\delta^5 \triangleq \frac{\sigma_g^4}{M^2}\mathrm{vec}(\phi^4)\mathrm{vec}^H(\phi^4) \tag{79}$$

and

$$\phi^1 \triangleq \sum_{j=1}^{M} P_j\bar{\mathbf{G}}_j\bar{\mathbf{G}}_j^H - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{G}}_p\bar{\mathbf{G}}_p^H \tag{80}$$

$$\phi^2 \triangleq \sum_{j=1}^{M} P_j\bar{\mathbf{F}}_j\bar{\mathbf{F}}_j^H - \gamma_j\sigma_x^2\mathrm{diag}\,(\mathbf{I}) - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{F}}_p\bar{\mathbf{F}}_p^H \tag{81}$$

$$\phi^3 \triangleq \sum_{j=1}^{M} P_j\bar{\mathbf{G}}_j(\bar{\mathbf{h}}_j^j)^H - \gamma_i\sum_{p\neq j}^{M} P_j\bar{\mathbf{G}}_j\left(\bar{\mathbf{h}}_j^p\right)^H \tag{82}$$

$$\phi^4 \triangleq \sum_{j=1}^{M} P_j\bar{\mathbf{h}}_j^j\mathbf{F}_j^H - \gamma_j\sum_{p\neq j}^{M} P_p\bar{\mathbf{h}}_j^p\mathbf{F}_p^H - \gamma_j\sigma_x^2\mathrm{diag}(\mathbf{I})\bar{\mathbf{g}}_j \tag{83}$$

REFERENCES

[1] Y. Xiao, X. Jin, Y. Shen, and Q. Guan, "Joint relay selection and adaptive modulation and coding for wireless cooperative communications," *IEEE Sens. J.*, vol. 21, no. 22, pp. 25 508–25 516, Nov. 2021.

[2] A. K. Shukla, V. Singh, P. K. Upadhyay, A. Kumar, and J. M. Moualeu, "Performance analysis of energy harvesting-assisted overlay cognitive NOMA systems with incremental relaying," *IEEE Open J. of Comm.*, vol. 2, pp. 1558–1576, June 2021.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. Part I. System description," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[4] P. Xu, Z. Yang, Z. Ding, I. Krikidis, and Q. Chen, "A novel probabilistic buffer-aided relay selection scheme in cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4548–4552, Feb. 2020.

[5] M. A. Maleki Sadr, M. Ahmadian-Attari and B. Mahboobi, "Low-complexity robust relay optimisation for multiple peer-to-peer beamforming: A safe tractable approximation approach," *IET Commun.*, vol. 9, no. 16, pp. 1968–1979, Dec. 2015.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

[6] M. A. Maleki Sadr, B. Mahboobi, S. Mehrizi, M. Ahmadian-Attari, and M. Ardebilipour, "Stochastic robust collaborative beamforming: Non-regenerative relay," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 947–958, Mar. 2016.

[7] M. A. Maleki Sadr and M. Ahmadian-Attari, "Secure robust relay beamforming: a convex conic approximation approach," *IET Commun.*, vol. 10, no. 10, pp. 1138–1150, Nov. 2016.

[8] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor, "Multiple peer-to-peer communications using a network of relays," *IEEE Trans. Signal. Process.*, vol. 57, no. 8, pp. 3053–3062, Mar. 2009.

[9] F. Gao and Y. Chen, "Performance analysis for an opportunistic DF based WPC system with interferers over log-normal fading channels," in *13th Int. Conf. Commun. Softw. Netw.*, June 2021, pp. 6–10.

[10] A. R. Heidarpour, M. Ardakani, and C. Tellambura, "Network-coded cooperation with outdated CSI," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1720–1723, June 2018.

[11] S. Srivastava, M. S. Kumar, A. Mishra, S. Chopra, A. K. Jagannatham, and L. Hanzo, "Sparse doubly-selective channel estimation techniques for OSTBC MIMO-OFDM systems: A hierarchical bayesian Kalman filter based approach," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4844–4858, May 2020.

[12] S. Srivastava, J. Nath, and A. K. Jagannatham, "Data aided quasistatic and doubly-selective CSI estimation using affine-precoded superimposed pilots in millimeter wave MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6983–6998, June 2021.

[13] U. Uyoata and M. Dlodlo, "Relay assisted device-to-device communication: Approaches and issues," *arXiv preprint arXiv:1810.07799*, 2018.

[14] O. A. Amodu, M. Othman, N. K. Noordin, and I. Ahmad, "Relay-assisted D2D underlay cellular network analysis using stochastic geometry: Overview and future directions," *IEEE Access*, vol. 7, pp. 115 023–115 051, Aug. 2019.

[15] O. Bello and S. Zeadally, "Intelligent device-to-device communication in the internet of things," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1172–1182, Sept. 2016.

[16] B. Chalise and L. Vandendorpe, "MIMO relay design for multipoint-to-multipoint communications with imperfect channel state information," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2785–2796, July 2009.

[17] J. Yang, B. Champagne, Y. Zou, and L. Hanzo, "Joint optimization of transceiver matrices for MIMO-Aided multiuser AF relay networks: Improving the QoS in the presence of CSI errors," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1434–1451, Mar. 2016.

[18] M. A. Maleki Sadr, M. Ahmadian-Attari, and R. Amiri, "Real-time cooperative adaptive robust relay beamforming based on Kalman filtering channel estimation," *IEEE Trans. Wireless. Commun.*, vol. 18, no. 12, pp. 5600–5612, Dec. 2019.

[19] Proakis, *Digital Communications 5th Edition*. McGraw Hill, 2007.

[20] A. I. Dvirnyi and V. I. Slyn'ko, "On stability of solutions of nonlinear nonstationary systems of impulsive differential equations in a critical case," *Nonlinear Oscillations*, vol. 14, no. 4, Mar. 2012.

[21] I. Arasaratnam, "Cubature Kalman filtering theory & applications," Ph.D. dissertation, 2009.

[22] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory algorithms and software*. John Wiley & Sons, 2004.

[23] R. D. Gitlin, J. F. Hayes, and S. B. Weinstein, *Data communications principles*. Springer Science & Business Media, Dec. 2012.

[24] G. Liu, F. R. Yu, H. Ji, V. C. Leung, and X. Li, "In-band full-duplex relaying: A survey, research issues and challenges," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 2, pp. 500–524, Jan. 2015.

[25] D. Simon, "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms," *IET Control. Theory Appl.*, vol. 4, no. 8, pp. 1303–1318, Aug. 2010.

[26] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.

[27] A. Dalir and H. Aghaeinia, "Maximizing first-order approximate mean of SINR under imperfect channel state information for throughput enhancement of MIMO interference networks," *IJST-T Electr. Eng.*, vol. 43, no. 1, pp. 121–132, July 2019.

[28] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[29] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[30] S. Janson, *Gaussian Hilbert Spaces*. Cambridge university press, 1997, no. 129.

[31] W. Wu, P. Cao, and Z. D. Pan, *Maneuver Target Tracking Based on Kalman IMM Algorithm*. Springer, 2012, pp. 479–483.

[32] J. F. Sturm, *Implementation of Interior Point Methods for Mixed Semidefinite and Second Order Cone Optimization Problems*, 2002.

[33] G. Hendeby, "Fundamental estimation and detection limits in linear non-gaussian systems," Ph.D. dissertation, Institutionen för systemteknik, 2005.

[34] H. Leung, Z. Zhu, and Z. Ding, "An aperiodic phenomenon of the extended Kalman filter in filtering noisy chaotic signals," *IEEE Trans. Sig. Process.*, vol. 48, no. 6, pp. 1807–1810, June 2000.

[35] J. Zarei and E. Shokri, "Convergence analysis of non-linear filtering based on cubature Kalman filter," *IET Sci. Meas. Technol.*, vol. 9, no. 3, pp. 294–305, May 2015.

[36] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Aug. 2002.

[37] P. H. M. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Automat. Contr.*, vol. 33, no. 9, pp. 867–870, Sept. 1988.

Mohammad Amin Maleki Sadr received his B.Sc. degree from University of Isfahan, Iran, and the M.Sc., and Ph.D. degrees from K.N. Toosi University of Technology, Tehran, Iran in 2012, 2014, and 2018, respectively, all in Electrical Engineering. During 2020-2021, he was a postdoc research fellow at Department of ECE department in McGill university, Canada. He is now a He is now a postdoc research fellow at department of statistics & actuarial science in university of Waterloo, Canada. His research interests include Deep Learning, Anomaly detection, localization, signal processing, and Beyond 5G wireless communication.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2022.3207410

Prof. Benoit Champagne received the B.Ing. in Engineering Physics from École Polytechnique de Montréal (1983), the M.Sc. in Physics from Université de Montréal (1985), and the Ph.D. in Electrical Engineering from University of Toronto (1990). From 1990 to 1999, he was Assistant and then Associate Professor at INRS-Telecommunications, Université du Quebec, Montreal. In1999, he joined McGill University, Montreal, where he is now a Professor within the Dept. of Electrical and Computer Engineering; he served as Associate Chairman of Graduate Studies in the Department from 2004 to 2007. His research focuses on the study of advanced methods and algorithms for the processing of information bearing signals by digital means. His interests span many areas of statistical signal processing and learning, including detection and estimation, sensor array processing, adaptive filtering, and applications thereof to broadband wireless communications and speech processing; he has co-authored more than 350 referred publications in these areas. He has been an Associate Editor for the IEEE Signal Processing Letters, the IEEE Trans. on Signal Processing and the EURASIP Journal on Applied Signal Processing. He has also served on the Technical Committees of several international conferences in the fields of communications and signal processing.

© 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: McGill University. Downloaded on November 21,2022 at 05:22:18 UTC from IEEE Xplore. Restrictions apply.