“©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Spatio-Temporal Power Optimization for MIMO Joint Communication and Radio Sensing Systems with Training Overhead

Xin Yuan, *Member, IEEE*, Zhiyong Feng, *Senior Member, IEEE*,
J. Andrew Zhang, *Senior Member, IEEE*, Wei Ni, *Senior Member, IEEE*,
Ren Ping Liu, *Senior Member, IEEE*, Zhiqing Wei, *Member, IEEE* and Changqiao Xu, *Senior Member, IEEE*

***Abstract*—In this paper, we study optimal spatio-temporal power mask design to maximize mutual information (MI) for a joint communication and (radio) sensing (JCAS, a.k.a., radar-communication) multi-input multi-output (MIMO) downlink system. We consider a typical packet-based signal structure which includes training and data symbols. We first derive the conditional MI for both sensing and communication under correlated channels by considering the training overhead and channel estimation error (CEE). Then, we derive a lower bound for the CEE and optimize the energy arrangement between the training and data signals to minimize the CEE. Based on the optimal energy arrangement, we provide optimal spatio-temporal power mask design for three scenarios, including maximizing MI for communication only and for sensing only, and maximizing a weighted sum MI for both communication and sensing. Extensive simulations validate the effectiveness of the proposed designs.**

***Index Terms*—Mutual information, joint communication and sensing, waveform design, training sequences.**

I. Introduction

A. Background and Motivation

A joint communication and (radio) sensing (JCAS, a.k.a., Radar-Communications) system that enables share of hardware and signal processing modules, can achieve efficient spectrum efficiency, enhanced security, and reduced cost, size, and weight [1]–[4]. JCAS systems can have many potential applications in intelligent transportation that require both communication links connecting vehicles and active environment sensing functions [5], [6]. For JCAS systems, it is crucial to use a waveform simultaneously performing both communication and sensing function, and help improve the availability of the limited spectrum resources. To this end, one of the main challenges in JCAS systems lies in designing optimal or adequate waveforms that serve both purposes of data transmission and radio sensing.

Mutual information (MI) is an important measure that can be used to study waveform designs for JCAS systems. To be specific, for communications the MI between wireless channels and the received communication signals can be employed as the waveform optimization criterion, while for sensing, the conditional MI between sensing channels and the reflected sensing signals can be measured [7], [8]. Despite a significant amount of research effort on waveform design in both communication and sensing systems, existing joint waveform designs for JCAS systems are still limited. It is known that the training sequences for channel estimation have a significant impact on communication capacity, particularly for multiple input multiple output (MIMO) systems [9], [10]. However, there has been no study on the waveform design for JCAS, which takes into consideration the typical signal packet structure containing the training sequences.

B. Related Work

Information theory has been used to design radar waveform [7], [8], [11]–[13]. Bell [7] was the first to apply information theory to optimize radar waveforms to improve target detection. In [12], the optimal radar waveform was proposed to maximize the detection performance of an extended target in a colored noise environment by using MI as waveform design criteria. Two criteria, namely, the maximization of the conditional MI and the minimization of the minimum mean-square error (MMSE), were studied in [8] to optimize the waveform design for MIMO radars by exploiting the covariance matrix of the extended target impulse response. In [11], the optimal waveform design for MIMO radars in colored noise was also investigated by considering two criteria: maximizing the MI and maximizing the relative entropy between two hypotheses that the target exists or does not exist in the echoes. In [13], a two-stage waveform optimization algorithm was proposed for an adaptive MIMO radar to unify the signal design and selection procedures. The algorithm is based on the constant learning of the radar environment at the receivers and the adaptation of the transmit waveform to dynamic radar scene. In [14], a robust waveform design based on the Cramér-Rao bound was proposed for co-located MIMO radars to improve the worst-case estimation accuracy in the presence of clutters.

For communication and radar co-existing systems that transmit and process respective signals, the MI has also been adopted for waveform design to minimize the interference to each other. In [15], [16], inner bounds on both the radar

X. Yuan and W. Ni are with the Data61, CSIRO, Marsfield, NSW2122, Australia (email: {Xin.Yuan, Wei.Ni}@data61.csiro.au).

Z. Feng and Z. Wei are with the Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, China (email: {fengzy, weizhiqing}@bupt.edu.cn).

J. A. Zhang and R. P. Liu are with the Global Big Data Technologies Center, University of Technology Sydney, Ultimo, NSW 2007, Australia (email: {Andrew.Zhang, RenPing.Liu}@uts.edu.au).

C. Xu is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China (email: cqxu@bupt.edu.cn).

estimation rate for sensing and the data rate for communication were derived for the co-existing systems. Liu *et al.* [17] studied transmit beamforming for spectrum sharing between MU-MIMO communication and co-located MIMO radar, to maximize the detection probability for sensing while guaranteeing the transmit power for downlink users. In [18], a minimum-estimation-error-variance waveform design method was proposed to optimize the spectral shape of a unimodular radar waveform and maximize the performance of both radar and communications. In [19], the waveform was designed based on a performance bound that is derived from jointly maximizing radar estimation rate and communication data rate.

Only several studies have investigated the MI for JCAS systems [20]–[22]. In [20], considering a JCAS MIMO setup, the expressions for radar MI and communication channel capacity were derived. In [21], [22], an integrated waveform design was proposed for OFDM JCAS systems to improve the MI for both communication and sensing by considering extended targets and frequency-selective fading channels.

C. Contributions

From an information-theoretic point-of-view, this paper studies the packet structure and spatio-temporal signal power mask for MIMO-based JCAS. MI is used as the performance metric of the study. (A power mask specifies the different transmit powers of different antennas to transmit different symbols within a frame.). MI is taken as the performance metric of our study. It is a critical (and arguably the universal) measure of how much information can be delivered over a known (or estimated) channel for communication, and how much reflections from unknown targets can be captured for sensing [7]. Different from other metrics (e.g., detection probability [23] and false-alarm probability [24]), the use of MI eliminates the need for (and the limitation to) a particular type of detector or detection algorithm, and provides understanding of JCAS.

With MI as the metric, the contributions of the paper include:

- As of sensing (around the transmitter), we derive the closed-form expression for the MI between the reflected signals $\mathbf{Y}_{\rm rad}$ at the transmitter and the sensing channel matrix $\mathbf{G}$ around the transmitter, given the knowledge of the transmitter on its transmitted signals $\mathbf{X} \in \mathbb{C}^{N\times L}$ (consisting of both the training and data signals).
- As of communication (between the transmitter and receiver), we derive the closed-form expression for the maximum of the MI between the transmitted data signals $\mathbf{X}_d \in \mathbb{C}^{N\times L_d}$ and the received signals $\mathbf{Y}_{\rm com}$. The maximum captures the channel estimation error (CEE) of the communication channel; and accounts for optimally allocated energy between the training and data, adapting to the statistics of the estimated channels.
- With the covariance matrices of the communication and sensing channels, i.e., $\mathbf{H}$ and $\mathbf{G}$, we derive the closed-form optimal spatio-temporal power mask of the data at the transmitter to separately maximize the sensing MI, the communication MI, or a weighted sum of both.

Our results reveal the trade-off of the MI between communication and sensing, as well as the non-negligible impact of the training length (and, in turn, the CEE) on JCAS.

A key difference of our work to [20]–[22] is that we consider both the training and data, where the training is used to estimate the communication channel and also contributes to the sensing MI. We also allocate energy between the training and data to maximize the best possible communication MI. In contrast, the existing related works [20]–[22] did not consider the training, solely focused on the data, and did not have to optimize the energy for data transmissions. Another difference of our work to [20]–[22] is that we consider the CEE of the communication channel for deriving the communication MI, allocating energy between the training and data, and optimizing the spatio-temporal power mask of the data. The spatio-temporal power mask is optimized by taking the contributions of both the training and data to the sensing MI into account. In contrast, [20]–[22] did not take the CEE into account, and all assumed that the channel was precisely estimated. Under the assumption, those studies [20]–[22] optimized the power mask of the data with only the contribution of the data to the sensing MI considered. More differences of our work to [21], [22] include that we consider a MIMO channel with all of the transmitter, receiver, and sensor/radar equipped with multiple antennas, as considered in [20]. In contrast, [21] and [22] considered a wideband single-input single-output (SISO) channel under perfect channel state information (CSI).

D. Organization and Notation

The rest of this paper is organized as follows. In Sec. II, the system model is introduced. In Sec. III, we derive the MI for both sensing and communication in the presence of non-negligible CEEs of the communication channel, and a tight lower bound for the CEE. In Sec. IV, we first design the optimal energy arrangement between the training and data signals, followed by the optimization of the spatio-temporal power mask for communication, sensing, and JCAS under the optimal energy arrangement. Sec. V presents simulation results, and Sec. VI concludes the paper.

Notation: Lower-case bold face $(\mathbf{x})$ indicates vector, and upper-case bold face $(\mathbf{X})$ indicates matrix. For a square matrix $\mathbf{X}$, $\mathbf{X}^a$ denotes the product matrix taking $\mathbf{X}$ and multiplying it by itself a-times. $\mathbf{I}$ denotes the identity matrix, $\mathbb{E}(\cdot)$ denotes expectation. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$, $(\cdot)^{-1}$ and $(\cdot)^\dagger$ denote transposition, conjugate transportation, conjugate, inverse and pseudo-inverse, respectively. $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ denote the determinant and trace of a matrix, respectively.

II. System Model

We consider a JCAS MIMO system where two nodes A and B perform point-to-point communications in time division duplex (TDD) mode, and simultaneously sense the environment to determine, e.g., the locations and speeds of nearby objects, as illustrated in Fig. 1. At the stage that node A is transmitting to node B, we consider downlink sensing where the reflection of the transmit signal is used

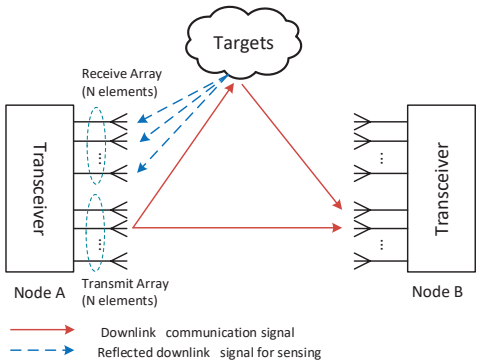


Fig. 1. A JCAS MIMO downlink system, where node A transmits data to node B, and simultaneously senses the environment to determine, e.g., the locations and speeds of nearby objects, by using the reflected signal. Each node is equipped with two spatially widely separated antenna arrays, i.e., N transmit antennas and N receive antennas, and each antenna array has dispersed elements to provide spatial diversity, and to suppress leakage signals from the transmitter and enable the reception of clear sensing signals.

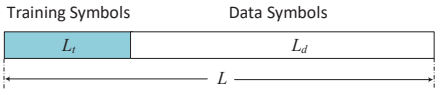


Fig. 2. Transmit symbols: including training and data symbols. For communications, the non-precoded training symbols are used for synchronization and channel estimation, and the data symbols are typically precoded data payload. While for sensing, both the training and data symbols are used for targets detection.

for sensing by node A. The transmitted symbols are known to node A. The channels of sensing and communications are correlated but different. To suppress leakage signals from the transmitter and enable the reception of clear sensing signals, each node is assumed to be equipped with two spatially widely separated antenna arrays, i.e., N transmit antennas and N receive antennas. Each antenna array has dispersed elements to provide spatial diversity. Detailed configurations of the transceiver for JCAS systems are beyond the scope of this paper, and readers can refer to [4] and [6] for more details.

In practice, a communication packet typically includes data payload, together with training signals for synchronization and channel estimation. The training signals can have various forms in different standards and systems. For example, it can be comb pilots or occupy whole resource blocks in 5G New Radio. Without loss of generality, we consider a general data structure consisting of a sequence of L_t training symbols and L_d data symbols for each spatial stream, as shown in Fig. 2. The total length of the transmit signals, $L = L_t + L_d$, is given in prior in the considered problem. Concatenating the symbols from N spatial streams into a matrix $\mathbf{X}$, we have $\mathbf{X} = [\mathbf{X}_t, \mathbf{X}_d]$, where $\mathbf{X}_t = [\boldsymbol{X}_t(1), \cdots, \boldsymbol{X}_t(N)]^T \in \mathbb{C}^{N\times L_t}$ and $\mathbf{X}_d = [\boldsymbol{X}_d(1), \cdots, \boldsymbol{X}_d(N)]^T \in \mathbb{C}^{N\times L_d}$, with $\boldsymbol{X}_t(n)$ and $\boldsymbol{X}_d(n)$ denoting the training and data symbols transmitted from the n-th antenna, respectively. We assume that the covariance matrix of $\mathbf{X}_d(n) \in \mathbb{C}^{L_d\times 1}$ is $\frac{1}{L_d}\mathbb{E}\left\{\mathbf{X}_d\mathbf{X}_d^H\right\} = \Sigma_{\mathbf{X}_d}$.

Let $\mathbf{X}_t(n) \in \mathbb{C}^{L_t\times 1}$ be typically designed to be orthogonal and $L_t \geq N$, and we have $\frac{1}{L_t}\mathbf{X}_t\mathbf{X}_t^H = \Sigma_{\mathbf{X}_t}$, where $\Sigma_{\mathbf{X}_t}$ is a scaled diagonal matrix. Existing designs of orthogonal codes, such as Walsh-Hadamard codes, can be used to encode the training sequences, and produce 2^n number of orthogonal sequences with 2^n bits per sequence [25]. By this means, the training sequences transmitted by different antennas remain orthogonal. The orthogonal design is a typical setting in MIMO communication systems, and also typically used in MIMO radar to exploit the degrees of freedom offered by multiple antennas [26].

The transmitted signal $\mathbf{X}$ is used for both communication and radio sensing. Let E be the total energy of the transmit signal, E_t the energy of the training signals, and E_d the energy of the data signals. $E = E_t + E_d$. The average energy of the training and data symbols are $\sigma_t^2 = \frac{1}{NL_t}\sum_{n=1}^{N}\mathbf{X}_t(n)^H\mathbf{X}_t(n)$, and $\sigma_d^2 = \frac{1}{NL_d}\sum_{n=1}^{N}\mathbb{E}\left[\mathbf{X}_d(n)^H\mathbf{X}_d(n)\right]$, respectively. We also define a weighting value κ, $0 < \kappa < 1$, and have $E_d = \kappa E = NL_d\sigma_d^2$ and $E_t = (1-\kappa)E = \sigma_t^2 NL_t$.

A. Communication Model

For communication, the received training and data signals at node B can be respectively given by

$$\mathbf{Y}_{\text{com}}^t = \mathbf{H}\mathbf{X}_t + \mathbf{N}_{tc}; \tag{1}$$

$$\begin{aligned}\mathbf{Y}_{\text{com}}^d &= \mathbf{H}\mathbf{X}_d + \mathbf{N}_{dc} = \left(\hat{\mathbf{H}} + \Delta\mathbf{H}\right)\mathbf{X}_d + \mathbf{N}_{dc}\\ &= \hat{\mathbf{H}}\mathbf{X}_d + \underbrace{\Delta\mathbf{H}\mathbf{X}_d + \mathbf{N}_{dc}}_{\mathbf{N}_c'},\end{aligned} \tag{2}$$

where $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_j, \cdots, \mathbf{h}_N] \in \mathbb{C}^{N\times N}$ is the channel matrix with $\mathbf{h}_j = [h_{1,j}, h_{2,j}, \cdots, h_{N,j}]^T$ denoting the j-th row of $\mathbf{H}$; $\mathbf{N}_{tc} \in \mathbb{C}^{N\times L_t}$ and $\mathbf{N}_{dc} \in \mathbb{C}^{N\times L_d}$ are both additive white Gaussian noise (AWGN) with zero mean and element-wise variance σ_n^2. We assume that $\mathbf{N}_{tc}$, $\mathbf{N}_{dc}$ and $\mathbf{X}_d$ are mutually independent. $\mathbf{Y}_{\text{com}}^t$ is used for channel estimation. We assume that a linear channel estimation based on a MMSE criterion [27] is applied. In this case, the channel estimate $\hat{\mathbf{H}}$ and the estimation error $\Delta\mathbf{H}$ are uncorrelated [28]. Let $\Delta\mathbf{H} = [\Delta\mathbf{h}_1, \cdots, \Delta\mathbf{h}_j, \cdots, \Delta\mathbf{h}_N]$, where $\Delta\mathbf{h}_j = [\Delta h_{1,j}, \cdots, \Delta h_{N,j}]^T$ is the j-th row of $\Delta\mathbf{H}$. The coefficients Δh_{ij} are random variables following i.i.d. zero mean circularly symmetric complex Gaussian with variance σ_e^2, i.e., $\mathbb{E}\left[\Delta\mathbf{H}\Delta\mathbf{H}^H\right] = N\sigma_e^2\mathbf{I}_N$. We will evaluate σ_e^2 and link it to $\mathbf{X}_t$ and $\mathbf{N}_{tc}$ in Section III-C.

The matrix $\mathbf{N}_c'$ combines the CEE and noise, and can be viewed as an equivalent additive noise with zero mean and covariance. The variance $\sigma_{n'}^2$ can be obtained as

$$\mathbb{E}[\mathbf{N}_c'\mathbf{N}_c'^H] = \mathbb{E}[\Delta\mathbf{H}\mathbf{X}_d\mathbf{X}_d^H\Delta\mathbf{H}^H] + \mathbb{E}\left[\mathbf{N}_{dc}\mathbf{N}_{dc}^H\right] \tag{3a}$$

$$= \mathbb{E}\left\{[\Delta\mathbf{h}_1, \cdots, \Delta\mathbf{h}_N]^T\,\Sigma_{\mathbf{X}_d}\,[\Delta\mathbf{h}_1^*, \cdots, \Delta\mathbf{h}_N^*]\right\} + \mathbb{E}\left[\mathbf{N}_{dc}\mathbf{N}_{dc}^H\right] \tag{3b}$$

$$= \text{diag}\left\{\text{Tr}\left(\Sigma_{\mathbf{X}_d}\mathbb{E}[\Delta\mathbf{h}_1^*\Delta\mathbf{h}_1^T]\right), \cdots, \text{Tr}\left(\Sigma_{\mathbf{X}_d}\mathbb{E}[\Delta\mathbf{h}_N^*\Delta\mathbf{h}_N^T]\right)\right\} + L_d\sigma_n^2\mathbf{I}_N \tag{3c}$$

$$= NL_d\sigma_d^2\sigma_e^2\mathbf{I}_N + L_d\sigma_n^2\mathbf{I}_N \triangleq L_d\sigma_n'^2\mathbf{I}_N, \tag{3d}$$

where $\sigma_n'^2 = \frac{E_d}{L_d}\sigma_e^2 + \sigma_n^2$. Let $\mathbf{R}_H = \frac{1}{N}\mathbb{E}[\mathbf{H}^H\mathbf{H}]$ be the channel covariance matrix, $\text{Tr}(\mathbf{R}_H) = N\sigma_h^2$, and $\mathbf{R}_H$ is a

positive semi-definite matrix. We assume that $\mathbf{R}_H$ is known to Node A. We can write the random channel matrix as $\mathbf{H} = \mathbf{H}_0\mathbf{R}_H^{\frac{1}{2}}$, where the entries of $\mathbf{H}_0$ are i.i.d. zero mean circularly symmetric complex Gaussian with unit variance.

B. Sensing Model

Node A uses the reflection of the transmitted signal for sensing. The received signal, denoted by $\mathbf{Y}_{\rm rad}$, is given by

$$\mathbf{Y}_{\rm rad} = \mathbf{G}\mathbf{X} + \mathbf{N} = [\mathbf{G}\mathbf{X}_t + \mathbf{N}_{tr}, \mathbf{G}\mathbf{X}_d + \mathbf{N}_{tr}], \tag{4}$$

where $\mathbf{G} = [\mathbf{g}_1, \cdots, \mathbf{g}_N]$ is the channel matrix to be sensed with its j-th column being $\mathbf{g}_j = [g_{1,j}, \cdots, g_{N,j}]^T$, and elements of $\mathbf{g}_j$ are independent of each other; $\mathbf{N}_{tr} = [\mathbf{n}_{tr,1}, \cdots, \mathbf{n}_{tr,N}] \in \mathbb{C}^{L_t \times N}$ and $\mathbf{N}_{dr} = [\mathbf{n}_{dr,1}, \cdots, \mathbf{n}_{dr,N}] \in \mathbb{C}^{L_d \times N}$ are AWGN with zero mean and covariance matrix $\mathbb{E}\left\{\mathbf{N}_{tr}\mathbf{N}_{tr}^H\right\} = N\sigma_n^2\mathbf{I}_{L_t}$ and $\mathbb{E}\left\{\mathbf{N}_{dr}\mathbf{N}_{dr}^H\right\} = N\sigma_n^2\mathbf{I}_{L_d}$. Let $\Sigma_{\mathbf{G}} = \frac{1}{N}\mathbb{E}\{\mathbf{G}\mathbf{G}^H\}$ be the spatial correlation matrix. It is assumed to be full-rank and also known to Node A.

Note that for both communication and sensing, the channel matrices include large-scale path loss and small-scale fading. The path loss of sensing can vary for different multi-path components depending on the number of nearby objects and their locations, and therefore, we consider the mean path loss herein. Our optimization results only depend on the ratio between the mean path losses of communication and sensing.

III. Mutual Information

In this section, we first derive the expression for the sensing MI by using both the training and data symbols. Then, we present the MI for communications under CEEs, followed by the development of a lower bound for the CEE of the communication channel based on the training symbols.

A. MI for Sensing

The MI between the sensing channel matrix $\mathbf{G}$ (or the "target impulse response" matrix in radar) and reflected signals $\mathbf{Y}_{\rm rad}$ given the knowledge of $\mathbf{X}$ can be used to measure the sensing performance [11]. With our model (4), the MI is

$$\begin{aligned} I\left(\mathbf{G}; \mathbf{Y}_{\rm rad}|\mathbf{X}\right) &= h\left(\mathbf{Y}_{\rm rad}|\mathbf{X}\right) - h\left(\mathbf{Y}_{\rm rad}|\mathbf{X}, \mathbf{G}\right) \\ &= h\left(\mathbf{Y}_{\rm rad}|[\mathbf{X}_t, \mathbf{X}_d]^T\right) - h\left(\mathbf{N}_r\right), \end{aligned} \tag{5}$$

where $h(\cdot)$ denotes the entropy of a random variable. Provided the noise vector $\mathbf{N}_{r,j} = \begin{bmatrix}\mathbf{n}_{tr,j}\\ \mathbf{n}_{dr,j}\end{bmatrix}$, $j = 1, \cdots, N$ are independent of each other, the conditional probability density function (PDF) of $\mathbf{Y}_{\rm rad}$ conditioned on $\mathbf{X}$ is given by

$$p\left(\mathbf{Y}_{\rm rad}|\mathbf{X}\right) = \prod_{j=1}^{N} p\left(\mathbf{y}_{{\rm rad},j}|[\mathbf{X}_t, \mathbf{X}_d]^T\right) \tag{6a}$$

$$= \prod_{j=1}^{N} \frac{1}{\pi^L \det\left([\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L\right)} \tag{6b}$$

$$\times \exp\left(-\mathbf{y}_{{\rm rad},j}^H\left([\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{y}_{{\rm rad},j}\right)$$

$$= \frac{1}{\pi^{LN}\det^N\left([\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L\right)} \tag{6c}$$

$$\times \exp\left\{-{\rm Tr}\left[\left([\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{Y}_{\rm rad}\mathbf{Y}_{\rm rad}^H\right]\right\},$$

where (6b) is obtained based on the PDF of circularly symmetric complex Gaussian distribution, and

$$\begin{aligned} \mathbb{E}\{\mathbf{y}_{{\rm rad},i}\mathbf{y}_{{\rm rad},i}^H\} &= [\mathbf{X}_t, \mathbf{X}_d]^T\mathbb{E}\{\mathbf{g}_j\mathbf{g}_j^H\}[\mathbf{X}_t, \mathbf{X}_d]^* \\ &\quad + \mathbb{E}\{{\rm diag}\{\mathbf{n}_{tr,j}\mathbf{n}_{tr,j}^H, \mathbf{n}_{dr,j}\mathbf{n}_{dr,j}^H\}\} \quad (7a) \\ &= \frac{1}{N}[\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L, \quad (7b) \end{aligned}$$

where (7a) is conditioned on $\mathbf{X}$, and $\mathbb{E}\{\mathbf{g}_j\mathbf{g}_j^H\} = \frac{1}{N}\mathbb{E}\{\mathbf{G}\mathbf{G}^H\} = \Sigma_{\mathbf{G}}$ in (7b) since $\mathbf{g}_j$, $j = 1, \cdots, N$ are independent of each other.

Based on (6), the entropy of $\mathbf{Y}_{\rm rad}$ conditional on $\mathbf{X}$ can be obtained as

$$\begin{aligned} h\left(\mathbf{Y}_{\rm rad}|\mathbf{X}\right) &= LN\log_2(\pi) + LN \\ &\quad + N\log_2\left[\det\left([\mathbf{X}_t, \mathbf{X}_d]^T\Sigma_{\mathbf{G}}[\mathbf{X}_t, \mathbf{X}_d]^* + \sigma_n^2\mathbf{I}_L\right)\right] \quad (8a) \\ &= LN\log_2(\pi) + LN \\ &\quad + N\log_2\left[(\sigma_n^2)^{L-N}\det\left(\mathbf{X}_t^*\mathbf{X}_t^T\Sigma_{\mathbf{G}} + \mathbf{X}_d^*\mathbf{X}_d^T\Sigma_{\mathbf{G}} + \sigma_n^2\mathbf{I}_N\right)\right], \quad (8b) \end{aligned}$$

where (8b) is based on the *Sylvester's determinant* theorem [29], i.e., $\det\left(\mathbf{A}_{M\times N}\mathbf{B}_{N\times M} + \sigma_n^2\mathbf{I}_M\right) = (\sigma_n^2)^{M-N}\det\left(\mathbf{B}_{N\times M}\mathbf{A}_{M\times N} + \sigma_n^2\mathbf{I}_N\right)$. The columns of the noise matrix $\mathbf{N}_r$ follow the i.i.d. multivariate complex Gaussian distribution with zero mean and covariance matrix $\sigma_n^2\mathbf{I}_N$, and the entropy of $\mathbf{N}_r$ is given by

$$h\left(\mathbf{N}_r\right) = LN\log_2(\pi) + LN + N\log_2\left[\det\left(\sigma_n^2\mathbf{I}_N\right)\right]. \tag{9}$$

By substituting (8) and (9) into (5), the sensing MI is

$$I\left(\mathbf{G}; \mathbf{Y}_{\rm rad}|\mathbf{X}\right) = N\log_2\left[\det\left(\frac{\mathbf{X}_t^*\mathbf{X}_t^T\Sigma_{\mathbf{G}} + \mathbf{X}_d^*\mathbf{X}_d^T\Sigma_{\mathbf{G}}}{(\sigma_n^2)^{L-N}} + \mathbf{I}_N\right)\right]. \tag{10}$$

B. MI for Communication

The MI for communication is defined as the mutual dependence between the transmit signals of node A and the received signals at node B, conditional on $\hat{\mathbf{H}}$. With the Gaussian assumption of CEE, the conditional PDF of $\mathbf{Y}_{\rm com}^d$ on $\hat{\mathbf{H}}$ is

$$p\left(\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right) = \prod_{i=1}^{L_d} p\left(\mathbf{y}_{{\rm com},i}^d|\hat{\mathbf{H}}\right) \tag{11a}$$

$$= \frac{1}{\pi^{L_dN}\det^{L_d}\left(\hat{\mathbf{H}}\Sigma_{\mathbf{X}_d}\hat{\mathbf{H}}^H + \sigma_{n'}^2\mathbf{I}_N\right)} \times \exp\left\{-{\rm Tr}\left[\left(\hat{\mathbf{H}}\Sigma_{\mathbf{X}_d}\hat{\mathbf{H}}^H + \sigma_{n'}^2\mathbf{I}_N\right)^{-1}\mathbf{Y}_{\rm com}\mathbf{Y}_{\rm com}^H\right]\right\}. \tag{11b}$$

The columns of $\mathbf{N}_c'$ follow the i.i.d. multivariate complex Gaussian distribution with zero mean and covariance matrix $\sigma_n'^2\mathbf{I}_N$. By referring to (8) – (9), the entropy of $\mathbf{N}_c'$ can be given by

$$h\left(\mathbf{N}_c'\right) = L_dN\log_2(\pi) + L_dN + L_d\log_2\left[\det\left(\sigma_{n'}^2\mathbf{I}_N\right)\right]. \tag{12}$$

Therefore, the conditional MI between $\mathbf{X}_d$ and $\mathbf{Y}_{\rm com}^d$ is obtained as

$$I\left(\mathbf{X}_d;\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right)=h\left(\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right)-h\left(\mathbf{N}_c'\right) =L_d\log_2\left[\det\left(\frac{\hat{\mathbf{H}}\Sigma_{\mathbf{X}_d}\hat{\mathbf{H}}^H}{\sigma_{n'}^2}+\mathbf{I}_N\right)\right], \quad (13)$$

where $\sigma_n'^2=\frac{E_d}{L_d}\sigma_e^2+\sigma_n^2$.

C. Estimation Error (CEE) of Communication Channel

With an MMSE MIMO channel estimation, the estimated MIMO channel matrix is [10]

$$\hat{\mathbf{H}}=\mathbf{H}-\Delta\mathbf{H} =\mathbf{H}\mathbf{X}_t\mathbf{X}_t^H\left(\sigma_n^2\mathbf{I}_N+\mathbf{X}_t\mathbf{X}_t^H\right)^{-1}+\mathbf{N}_t\mathbf{X}_t^H\left(\sigma_n^2\mathbf{I}_N+\mathbf{X}_t\mathbf{X}_t^H\right)^{-1},$$

which is the actual channel matrix $\mathbf{H}$, contaminated by an additive estimation error $\Delta\mathbf{H}$. Here, $\mathbf{X}_t$ is the $N\times L_t$ training symbol matrix with the average energy σ_t^2 per entry. Take the singular value decomposition (SVD) of $\mathbf{R}_H$. $\mathbf{R}_H=\mathbf{U}_H\mathbf{\Lambda}_H\mathbf{U}_H^H$, where $\mathbf{\Lambda}_H=\mathrm{diag}(\delta_1,\delta_2,\cdots,\delta_N)$, and $\frac{1}{N}\sum_{i=1}^N\delta_i=\frac{1}{N}\mathrm{Tr}(\mathbf{R}_H)\triangleq\sigma_h^2$. Let $\mathbf{\Lambda}_{\rm CRLB}$ be the Cramér-Rao lower bound (CRLB) of the channel matrix estimation [30]. We have

$$\mathbb{E}\left[\Delta\mathbf{H}\Delta\mathbf{H}^H\right]=\mathbb{E}\left[\left(\mathbf{H}-\hat{\mathbf{H}}\right)\left(\mathbf{H}-\hat{\mathbf{H}}\right)^H\right]$$
$$\geq\mathbf{\Lambda}_{\rm CRLB}=\left(\frac{\Sigma_{\mathbf{X}_t}}{\sigma_n^2}+\mathbf{R}_H^{-1}\right)^{-1}=\left(\frac{\mathbf{U}_H^H\Sigma_{\mathbf{X}_t}\mathbf{U}_H}{\sigma_n^2}+\mathbf{\Lambda}_H{}^{-1}\right)^{-1}$$
$$=\mathrm{diag}\left(\frac{\sigma_n^2\delta_1}{\sigma_n^2+L_t\sigma_t^2\delta_1},\cdots,\frac{\sigma_n^2\delta_i}{\sigma_n^2+L_t\sigma_t^2\delta_i},\cdots,\frac{\sigma_n^2\delta_N}{\sigma_n^2+L_t\sigma_t^2\delta_N}\right),$$

which is due to the fact that $\Sigma_{\mathbf{X}_t}=\mathbf{X}_t\mathbf{X}_t^H=L_t\sigma_t^2\mathbf{I}_N$. A lower bound of the CEE is given by $\mathcal{C}_t=\mathrm{Tr}(\mathbf{\Lambda}_{\rm CRLB})=\sum_{i=1}^N\frac{\sigma_n^2\delta_i}{\sigma_n^2+L_t\sigma_t^2\delta_i}$. The lower bound depends on the channel covariance matrix $\mathbf{R}_H$. In other words, $\mathcal{C}_t$ is a function of δ_i with the constraint $\sum_{i=1}^N\delta_i=\mathrm{Tr}(\mathbf{R}_H)=N\sigma_h^2$.

We proceed to minimize the lower bound of the CEE $\mathcal{C}_t$, given the knowledge of average channel gains. We can obtain the lower bound of $\mathcal{C}_t$ by applying the Lagrange multiplier method. The Lagrangian function is given by

$$\mathcal{L}\left(\mathbf{\Lambda}_H\right)=\sum_{i=1}^N\frac{\sigma_n^2\delta_i}{\sigma_n^2+L_t\sigma_t^2\delta_i}+\tau\left(\sum_{i=1}^N\delta_i-\mathrm{Tr}(\mathbf{R}_H)\right),$$

where τ is the Lagrange multiplier. By solving $\frac{\partial\mathcal{L}(\mathbf{\Lambda}_H)}{\partial\delta_i}=0$, we have $\frac{\sigma_n^4}{(\sigma_n^2+\delta_iL_t\sigma_t^2)^2}+\tau=0$, which shows that the lower bound is achieved when $\delta_i=\frac{1}{N}\mathrm{Tr}(\mathbf{R}_H)=\frac{1}{N}\sum_{i=1}^N\delta_i=\sigma_h^2$, $i=1,\cdots,N$. The lower bound of $\mathcal{C}_t$ is $\mathcal{C}_t\geq\frac{N\sigma_n^2\sigma_h^2}{\sigma_n^2+L_t\sigma_t^2\sigma_h^2}=\mathcal{C}_t^l$. Therefore, for any diagonal element of $\mathbf{\Lambda}_{\rm CRLB}(i)$, we have

$$\mathbf{\Lambda}_{\rm CRLB}(i)\geq\frac{\sigma_n^2\sigma_h^2}{\sigma_n^2+L_t\sigma_t^2\sigma_h^2}\triangleq\mathcal{C}_e,\ i=1,\cdots,N. \quad (14)$$

It is noted that the channel characteristics are not controllable. Therefore, the equality in (14) can only be achieved when $\mathbf{R}_H$ is a scaled identity matrix, i.e., $\delta_1=\cdots=\delta_N$. Nevertheless, the minimum achievable lower bound, i.e., the right-hand side of (14), allows us evaluating the maximum possible MI of the communication from the transmitter to the receiver in Section IV-A.

IV. Optimal Spatio-temporal Power Mask of Data

In this section, we optimize the spatio-temporal structure of transmit signals, given an energy budget and a total number of symbols for the transmission. For mathematical tractability, the problem is decoupled into the following two stages.

- **Stage 1:** Given a transmission duration, we divide the total energy budget between training and data to maximize the upper bound of communication MI in the presence of the non-negligible CEEs of the communication channel.
- **Stage 2:** We optimize the spatio-temporal power mask of the data to maximize the communication MI, sensing MI and their weighted sum, given the allocated energy for the data and the a-priori knowledge of the correlation matrices of the communication and/or sensing channels.

A. Energy splitting between Training and Data

In general, there are some constraints on the maximum and average transmission powers of a transmitter. When such constraints are applied, there is a motivation for optimizing the energy arrangement between the training and data signals, especially for maximizing the MI for communications. Here, we optimize the energy arrangement only by referring to the communication MI, because its impact on communication MI is much stronger than on sensing MI. Larger CEE can cause substantially deteriorate communication performance while sensing MI can only be slightly affected since the training sequences are directly used for sensing.

Since $\hat{\mathbf{H}}=\mathbf{H}-\Delta\mathbf{H}$, we can obtain that $\hat{\mathbf{H}}$ is a random variable with zero mean and variance $\sigma_{\hat{\mathbf{H}}}^2=\frac{1}{N^2}\mathbb{E}\left[\mathrm{Tr}\{\hat{\mathbf{H}}\hat{\mathbf{H}}^H\}\right]$. According to the orthogonality principle for MMSE [9] and the obtained lower bound of CEE, we have $\sigma_{\hat{\mathbf{H}}}^2=\sigma_h^2-\sigma_e^2$. Therefore, $\hat{\mathbf{H}}$ can be normalized as $\tilde{\mathbf{H}}=\frac{1}{\sigma_{\hat{\mathbf{H}}}}\hat{\mathbf{H}}$, which has elements following i.i.d. Complex Gaussian distribution $\mathcal{CN}(0,1)$. The MI in (13) can be upper bounded by

$$I\left(\mathbf{X}_d;\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right)\leq L_d\log_2\left[\det\left(\frac{\sigma_h^2-\mathcal{C}_e}{\frac{E_d}{L_d}\mathcal{C}_e+\sigma_n^2}\tilde{\mathbf{H}}\Sigma_{\mathbf{X}_d}\tilde{\mathbf{H}}^H+\mathbf{I}_N\right)\right], \quad (15)$$

which is due to the lower bound of CEE, i.e., $\sigma_e^2\geq\mathcal{C}_t^l$.

By substituting $\mathcal{C}_e$ of (14) into (15) and exploiting Jensen's inequality, we can obtain the optimal energy arrangement that minimizes the CEE as summarized in the following theorem.

Theorem 1. *The optimal energy arrangement for maximizing the channel capacity under the training symbols is given by*

$$\kappa_{\rm op}=\begin{cases}\Gamma+\sqrt{\Gamma(\Gamma-1)}, & \text{if } L_d<N;\\ \frac{1}{2}, & \text{if } L_d=N;\\ \Gamma-\sqrt{\Gamma(\Gamma-1)}, & \text{if } L_d>N,\end{cases} \quad (16)$$

where $\Gamma=\frac{L_d}{L_d-N}\left(1+\frac{N\sigma_n^2}{E\sigma_h^2}\right)$.

The lower bound of the CEE can be given by

$$\mathcal{C}_t^l = \frac{N\sigma_n^2\sigma_h^2}{N\sigma_n^2 + (1-\kappa_{\rm op})\,E\sigma_h^2}. \tag{17}$$

Proof. The proof is provided in Appendix A. □

Corollary 1. *In a high SNR regime, Γ and $\kappa_{\rm op}$ can be approximated as $\Gamma \approx \frac{L_d}{L_d-N}$ and $\kappa_{\rm op} \approx \frac{\sqrt{L_d}}{\sqrt{L_d}+\sqrt{N}}$. In this case, $\rho \approx \rho_{\max} = \frac{L_d}{(\sqrt{L_d}+\sqrt{N})^2}\frac{E}{N\sigma_n^2}$. We can find that, in the high SNR regime, the energy arrangement between the data and training symbols depends on the number of the data symbols, L_d, and the number of antennas, N. Moreover, $\kappa_{\rm op}$ decreases with the growth of both L_d and N.*

In a low SNR regime, Γ and $\kappa_{\rm op}$ can be approximated as $\Gamma \approx \frac{L_d N\sigma_n^2}{(L_d-N)E\sigma_h^2}$ and $\kappa_{\rm op} \approx \frac{1}{2}$. In this case, $\rho \approx \rho_{\max} = \frac{E^2\sigma_h^4}{4N^2\sigma_n^4}$. We find that half of the total energy are allocated to the training symbols, and the maximum SNR in the low SNR regime quadratically increases with $\frac{P}{\sigma_n^2}$.

The optimal training length achieving the largest possible communication MI under imperfect channel estimation, i.e., (36), is provided in Corollary 2.

Corollary 2. *Given $L_t\sigma_t^2 + L_d\sigma_d^2 = E$, the optimal length of the training sequence is $L_t = N$ for all L, and the upper bound of $\rho_{\max}$ is given by*

$$\rho_{\max} \le \begin{cases} \dfrac{(L-N)E}{(L-2N)N\sigma_n^2}\left(\sqrt{\Gamma_c} - \sqrt{\Gamma_c - 1}\right)^2, & L > 2N; \\ \dfrac{E^2\sigma_h^4}{4N\sigma_n^2(N\sigma_n^2 + E\sigma_h^2)}, & L = 2N; \\ \dfrac{(L-N)E}{(L-2N)N\sigma_n^2}\left(\sqrt{-\Gamma_c} - \sqrt{-\Gamma_c - 1}\right)^2, & L < 2N, \end{cases} \tag{18}$$

where $\Gamma_c = \frac{L-N}{L-2N}\left(1 + \frac{N\sigma_n^2}{E\sigma_h^2}\right)$.

Proof. In the case of $L_d > N$, the maximal SNR $\rho_{\max} = \frac{L_d E}{(L_d-N)N\sigma_n^2}\left(\sqrt{\Gamma} - \sqrt{\Gamma-1}\right)^2$ is a function of L_d. The first-order derivative of $\rho_{\max}$ with respect to L_d is given by

$$\frac{\partial \rho_{\max}}{\partial L_d} = \frac{\left(\sqrt{\Gamma} - \sqrt{\Gamma-1}\right)^2 E}{\left(L_d - N\right)^2 \sigma_n^2}\left(-1 + \frac{L_d}{N}\cdot\frac{\sqrt{\Gamma}}{\sqrt{\Gamma-1}}\right). \tag{19}$$

Clearly, $\frac{\partial \rho_{\max}}{\partial L_d} > 0$ for $L_d > N$. As $L_d = L - L_t$, we have $\frac{\partial \rho_{\max}}{\partial L_t} < 0$ for $L_t < L - N$, and $\rho_{\max}$ is a monotonically decreasing function of L_t. Since the shortest training length is N (i.e., $L_t \ge N$). Therefore, the optimal $\rho_{\max}$ is achieved when $L_t = N$, the optimal training length is N, and $\rho_{\max}^c = \frac{(L-N)E}{(L-2N)N\sigma_n^2}\left(\sqrt{\Gamma} - \sqrt{\Gamma-1}\right)^2$, where $\Gamma_c = \frac{L-N}{L-2N}\left(1 + \frac{N\sigma_n^2}{E\sigma_h^2}\right)$. Likewise, this conclusion can also be drawn in the cases of $L_d = N$ and $L_d < N$. The details are suppressed. □

Hereafter, we use the optimal energy allocation (16) and let $\mathcal{C}_e^l = \frac{\mathcal{C}_t^l}{N} = \frac{\sigma_n^2\sigma_h^2}{N\sigma_n^2 + (1-\kappa_{\rm op})P\sigma_h^2}$; unless otherwise stated. Our results are consistent with existing waveform designs based on the MSE or channel capacity [9], [10].

B. Optimal Waveform Design for Radio Sensing Only

To achieve the maximum sensing MI, or in other words, to make received signals $\mathbf{Y}_{\rm rad}$ containing rich information about $\mathbf{G}$, the transmit signals $\mathbf{X}$ (including $\mathbf{X}_t$ and $\mathbf{X}_d$) should be designed according to $\mathbf{G}$. Since $\mathbf{X}_t$ and $\mathbf{X}_d$ are independent and have different correlations, the optimization problem for maximizing the sensing MI can be decoupled into two separate optimization problems. As assumed, $\mathbf{X}_t$ contains deterministic orthogonal rows and $\mathbf{X}_t\mathbf{X}_t^H = L_t\sigma_t^2\mathbf{I}_N$. We only need to consider the optimization problem for the data signals.

For the data signals, the spatial correlation matrix can be diagonalized through SVD, i.e., $\Sigma_{\mathbf{G}} = \frac{1}{N}\mathbb{E}\{\mathbf{G}\mathbf{G}^H\} = \mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H$, where $\mathbf{U}_G$ is a unitary matrix and $\mathbf{\Lambda}_G = {\rm diag}\left\{\lambda_{11}, \cdots, \lambda_{ii}, \cdots, \lambda_{NN}\right\}$ is a diagonal matrix with λ_{ii} being the singular values. $\sigma_g^2 = \sum_{i=1}\lambda_{ii}$ is the mean channel gain for sensing channels. The MI in (10) can be rewritten as (20), where (20b) is based on *Sylvester's determinant* theorem. Define $\mathbf{Q}^{(t)} = \left(\mathbf{X}_t^T\mathbf{U}_G\right)^H\mathbf{X}_t^T\mathbf{U}_G$ and $\mathbf{Q}^{(d)} = \left(\mathbf{X}_d^T\mathbf{U}_G\right)^H\mathbf{X}_d^T\mathbf{U}_G$, and their (i,j)-th entries are $q_{ij}^{(t)}$ and $q_{ij}^{(d)}$.

According to Hadamard's inequality for the determinant and trace of an $N \times N$ positive semi-definite Hermitian matrix, we have the following inequalities: $\det\left(\mathbf{Q}_{N\times N}^{(t)}\right) \le \prod_{i=1}^N q_{ii}^{(t)}$, $\det\left(\mathbf{Q}_{N\times N}^{(d)}\right) \le \prod_{i=1}^N q_{ii}^{(d)}$, and ${\rm Tr}\left(\mathbf{Q}_{N\times N}^{(t)}{}^{-1}\right) \ge \prod_{i=1}^N \frac{1}{q_{ii}^{(t)}}$, ${\rm Tr}\left(\mathbf{Q}_{N\times N}^{(d)}{}^{-1}\right) \ge \prod_{i=1}^N \frac{1}{q_{ii}^{(d)}}$, where the equalities are achieved if and only if $\mathbf{Q}_{N\times N}$ is diagonal. As a result, the MI can be rewritten as

$$I\left(\mathbf{G};\mathbf{Y}_{\rm rad}|\mathbf{X}\right) \le N\log_2\left[\prod_{i=1}^N\left(\frac{\lambda_{ii}(q_{ii}^{(t)}+q_{ii}^{(d)})}{(\sigma_n^2)^{\frac{L}{N}}}+1\right)\right]. \tag{21}$$

Since $\mathbf{U}_G$ is unitary, we find ${\rm Tr}\left(\mathbf{Q}^{(d)}\right) = {\rm Tr}\left(\left(\mathbf{X}_d^T\mathbf{U}_G\right)^H\mathbf{X}_d^T\mathbf{U}_G\right) = {\rm Tr}\left(\mathbf{Z}^H\mathbf{Z}\right)$, where $\mathbf{Z} = \mathbf{X}_d^T\mathbf{U}_G$ is an $L_d \times N$ matrix. Under the constraint that E is finite, we have ${\rm Tr}\left(\mathbf{Q}^{(d)}\right) \le E_d = \kappa_{\rm op}E$. Since $\mathbf{X}_t$ satisfies the orthogonality condition, we have ${\rm Tr}\left(\mathbf{Q}^{(t)}\right) = {\rm Tr}\left(\left(\mathbf{X}_t^T\mathbf{U}_G\right)^H\mathbf{X}_t^T\mathbf{U}_G\right) = {\rm Tr}\left(\mathbf{X}_t^*\mathbf{X}_t^t\right) = {\rm Tr}\left(\mathbf{X}_t\mathbf{X}_t^H\right) = E_t = (1-\kappa_{\rm op})E$ and $q_{ii}^{(t)} = \frac{E_t}{N} = \frac{(1-\kappa_{\rm op})E}{N}$. Therefore, the maximum MI can be obtained by solving the following constrained problem:

$$\begin{aligned} F_r = &\max_{\mathbf{Q}^{(d)}} \sum_{i=1}^N\left\{\log_2\left(\frac{\lambda_{ii}\left(\frac{E_t}{N}+q_{ii}^{(d)}\right)}{(\sigma_n^2)^{\frac{L}{N}}}+1\right)\right\}, \\ &\text{subject to } \ {\rm Tr}\left(\mathbf{Q}^{(d)}\right) \le E_d; \\ &\quad q_{ii}^{(d)} \ge 0,\ 1 \le i \le N. \end{aligned} \tag{22}$$

We can apply the Lagrange multiplier method to solve (22). The Lagrangian can be written as

$$\mathcal{L}\left(\mathbf{Q}^{(d)}\right) = \sum_{i=1}^N\left\{\log_2\left(\frac{\lambda_{ii}\left(\frac{E_t}{N}+q_{ii}^{(d)}\right)}{(\sigma_n^2)^{\frac{L}{N}}}+1\right)\right\} + \alpha\sum_{i=1}^N q_{ii}^{(d)}, \tag{23}$$

$$I\left(\mathbf{G};\mathbf{Y}_{\rm rad}|\mathbf{X}\right)=N\log_2\left[\det\left(\frac{\mathbf{X}_t^*\mathbf{X}_t^T\mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H+\mathbf{X}_d^*\mathbf{X}_d^T\mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+\mathbf{I}_N\right)\right] \tag{20a}$$

$$=N\log_2\left[\det\left(\frac{\mathbf{\Lambda}_G\left(\mathbf{X}_t^T\mathbf{U}_G\right)^H\mathbf{X}_t^T\mathbf{U}_G+\mathbf{\Lambda}_G\left(\mathbf{X}_d^T\mathbf{U}_G\right)^H\mathbf{X}_d^T\mathbf{U}_G}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+\mathbf{I}_N\right)\right], \tag{20b}$$

where α is the Lagrange multiplier associated with $q_{ii}^{(d)}$, $i=1,\cdots,N$. Differentiating $\mathcal{L}(\mathbf{Q}^{(d)})$ with respect to $q_{ii}^{(d)}$, and setting the first-order derivative as 0, we can obtain $q_{ii}^{(d)}$ as

$$q_{ii}^{(d)}=\left(-\frac{\left(\sigma_n^2\right)^{\frac{L}{N}}}{\alpha\ln 2}-\frac{E_t}{N}-\frac{\left(\sigma_n^2\right)^{\frac{L}{N}}}{\lambda_{ii}}\right)^+,i=1,\cdots,N, \tag{24}$$

the optimality conditions are satisfied if $\sum_{i=1}^N\left(-\frac{\left(\sigma_n^2\right)^{\frac{L}{N}}}{\alpha\ln 2}-\frac{E_t}{N}-\frac{\left(\sigma_n^2\right)^{\frac{L}{N}}}{\lambda_{ii}}\right)^+=\kappa_{\rm op}E$ holds. Since the diagonal elements of $\mathbf{Q}^{(d)}$ are real and greater than 0, $\mathbf{Q}^{(d)\frac{1}{2}}$ exists. For any $L_d\times N$ matrix $\mathbf{\Psi}$ with orthogonal columns, if $\mathbf{\Psi}^H\mathbf{\Psi}=\mathbf{I}_N$, we have $\mathbf{Z}=\mathbf{\Psi}\mathbf{Q}^{(d)\frac{1}{2}}$. Since $\mathbf{Z}=\mathbf{X}_d^H\mathbf{U}_G$, we have $\mathbf{X}_d=\left(\mathbf{\Psi}\mathbf{Q}^{(d)\frac{1}{2}}\mathbf{U}_G^H\right)^H$.

With the optimal $\mathbf{X}_d$ for sensing, we can derive the corresponding communication MI which is not necessarily optimal, as given by

$$I\left(\mathbf{X}_d;\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right)\le L_d\log_2\left[\det\left(\frac{(\sigma_h^2-\mathcal{C}_e^l)\tilde{\mathbf{H}}\Sigma_{\mathbf{X}_d}\tilde{\mathbf{H}}^H}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+\mathbf{I}_N\right)\right] \tag{25a}$$

$$=L_d\log_2\left[\det\left(\frac{(\sigma_h^2-\mathcal{C}_e^l)\mathbf{U}_G\mathbf{Q}^{(d)}\mathbf{U}_G^H\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+\mathbf{I}_N\right)\right], \tag{25b}$$

where (25b) is obtained due to $\Sigma_{\mathbf{X}_d}=\frac{1}{L_d}\mathbb{E}\left\{\mathbf{X}_d\mathbf{X}_d^H\right\}=\left(\mathbf{\Psi}\mathbf{Q}^{(d)\frac{1}{2}}\mathbf{U}_G^H\right)^H\mathbf{\Psi}\mathbf{Q}^{(d)\frac{1}{2}}\mathbf{U}_G^H=\mathbf{U}_G\mathbf{Q}^{(d)}\mathbf{U}_G^H$.

Note that maximizing the sensing MI, i.e., (22), provides the maximal information extraction of the reflected signals at the transmitter. In other words, the higher the sensing MI is, the better estimation result can be expected to achieve.

C. Optimal Waveform Design for Communication Only

Based on the optimal energy allocation (i.e., Theorem 1) and correlated channel matrix $\tilde{\mathbf{H}}$, we analyze the maximum MI of the communication by optimizing the spatio-temporal waveforms of the data signals, i.e., $\mathbf{X}_d$. Let the SVD of the spatial correlation matrix of $\tilde{\mathbf{H}}$ be $\Sigma_{\tilde{\mathbf{H}}}=\frac{1}{N}\mathbb{E}\{\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\}=\mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Lambda}_{\tilde{\mathbf{H}}}\mathbf{U}_{\tilde{\mathbf{H}}}^H$, where $\mathbf{U}_{\tilde{\mathbf{H}}}$ is a unitary matrix and $\mathbf{\Lambda}_{\tilde{\mathbf{H}}}=\mathrm{diag}\left\{\mu_{11},\cdots,\mu_{ii},\cdots,\mu_{NN}\right\}$ is a diagonal matrix with μ_{ii} being the singular values. Also let $\mathbf{\Xi}=\left(\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}\right)^H\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}=\left(\mathbf{Y}_{\rm com}^d\right)^H\mathbf{Y}_{\rm com}^d$ with its (i,j)-th entry being ξ_{ij}. Based on Hadamard's inequality for the determinant and trace of a positive semi-definite Hermitian matrix, we have $\det\left(\mathbf{\Xi}_{N\times N}\right)\le\prod_{i=1}^N\xi_{ii}$. As derived in the proof of Theorem 1 in Appendix A, the maximum possible MI can be obtained by further maximizing the upper bound of the MI, as given by

$$\begin{aligned}F_c=\max\ & L_d\sum_{i=1}^N\left\{\log_2\left[\frac{\left(\sigma_h^2-\mathcal{C}_e^l\right)\mu_{ii}\xi_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+1\right]\right\}\\ \text{subject to}\ & \mathrm{Tr}\left(\mathbf{\Xi}\right)\le\kappa_{\rm op}E;\\ & \xi_{ii}\ge 0,\ 1\le i\le N.\end{aligned} \tag{26}$$

The optimal solution for (26) satisfies $\sum_{i=1}^N\left[-\frac{1}{\beta'\ln 2}-\frac{\sigma_n^2+\frac{E_d}{L_d}\mathcal{C}_e^l}{\mu_{ii}\left(\sigma_h^2-\mathcal{C}_e^l\right)}\right]^+=\kappa_{\rm op}E$, where β' is the Lagrange multiplier associated with ξ_{ii}. Therefore, ξ_{ii} can be obtained as

$$\xi_{ii}=\left[-\frac{1}{\beta'\ln 2}-\frac{\sigma_n^2+\frac{E_d}{L_d}\mathcal{C}_e^l}{\mu_{ii}\left(\sigma_h^2-\mathcal{C}_e^l\right)}\right]^+,\ i=1,\cdots,N. \tag{27}$$

Since the diagonal elements of $\mathbf{\Xi}$ are real and positive, $\mathbf{\Xi}^{\frac{1}{2}}$ exists. For any $L_d\times N$ matrix $\mathbf{\Theta}$ with orthonormal columns, if $\mathbf{\Theta}^H\mathbf{\Theta}=\mathbf{I}_N$, we have $\mathbf{Y}_{\rm com}^d=\mathbf{\Theta}\mathbf{\Xi}^{\frac{1}{2}}$. Since $\mathbf{Y}_{\rm com}^d=\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}$, we have $\mathbf{X}_d=\left(\mathbf{\Theta}\mathbf{\Xi}^{\frac{1}{2}}\mathbf{U}_{\tilde{\mathbf{H}}}^H\right)^H$. Based on the optimal singular values of the covariance matrix for the data signals in (27), we can obtain the sensing MI under the maximum possible communication MI, as given by

$$I\left(\mathbf{G};\mathbf{Y}_{\rm rad}|\mathbf{X}\right)$$
$$=N\log_2\left[\det\left(\frac{\left(\mathbf{X}_t^T\mathbf{U}_G\right)^H\mathbf{X}_t^T\mathbf{U}_G+\mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Xi}\mathbf{U}_{\tilde{\mathbf{H}}}^H\Sigma_{\mathbf{G}}}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+\mathbf{I}_N\right)\right],$$

which is obtained due to $\Sigma_{\mathbf{X}_d}=\frac{1}{L_d}\mathbb{E}\left\{\mathbf{X}_d^*\mathbf{X}_d^T\right\}=\frac{1}{L_d}\mathbb{E}\left\{\mathbf{X}_d\mathbf{X}_d^H\right\}=\left(\mathbf{\Theta}\mathbf{\Xi}^{\frac{1}{2}}\mathbf{U}_{\tilde{\mathbf{H}}}^H\right)^H\mathbf{\Theta}\mathbf{\Xi}^{\frac{1}{2}}\mathbf{U}_{\tilde{\mathbf{H}}}^H=\mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Xi}\mathbf{U}_{\tilde{\mathbf{H}}}^H$.

D. Joint Maximization of a Weighted Sum of MI

In this section, we conduct the waveform optimization for jointly considering the MI for both communication and radio sensing. Since there is generally no solution that can simultaneously maximize the MI for communication and radio sensing, a weighted sum of them is exploited and given by

$$F_w=\frac{w_r}{F_r}I\left(\mathbf{G};\mathbf{Y}_{\rm rad}|\mathbf{X}\right)+\frac{1-w_r}{F_c}I\left(\mathbf{X}_d;\mathbf{Y}_{\rm com}^d|\hat{\mathbf{H}}\right). \tag{29}$$

To maximize the weighted sum, $\mathbf{X}_d$ is designed according to the correlation matrices of both $\mathbf{H}$ and $\mathbf{G}$. Based on the SVD, $\Sigma_{\mathbf{H}}=\frac{1}{N}\mathbb{E}\{\mathbf{H}\mathbf{H}^H\}=\mathbf{U}_H\mathbf{\Lambda}_H\mathbf{U}_H^H$ and $\Sigma_{\mathbf{G}}=\frac{1}{N}\mathbb{E}\{\mathbf{G}\mathbf{G}^H\}=\mathbf{U}_G\mathbf{\Lambda}_G\mathbf{U}_G^H$, (29) can be rewritten as (27).

Define $\mathbf{\Pi}=(\mathbf{X}_d^T\mathbf{U}_H)^H\mathbf{X}_d^T\mathbf{U}_H=(\mathbf{X}_d^T\mathbf{U}_G)^H\mathbf{X}_d^T\mathbf{U}_G$ with the (i,j)-th entry, ϖ_{ij}, and $\mathrm{tr}\left(\mathbf{\Pi}\right)=\mathrm{tr}\left(\Sigma_{\mathbf{X}_{\rm d}}\right)$. According to Hadamard's inequality for the determinant and

$$F_w = \frac{w_r N}{F_r}\log_2\left[\det\left(\frac{\mathbf{\Lambda}_G\left(\mathbf{X}_t^T\mathbf{U}_G\right)^H\mathbf{X}_t^T\mathbf{U}_G+\mathbf{\Lambda}_G\left(\mathbf{X}_d^T\mathbf{U}_G\right)^H\mathbf{X}_d^T\mathbf{U}_G}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+\mathbf{I}_N\right)\right] + \frac{(1-w_r)L_d}{F_c}\log_2\left[\det\left(\frac{(\sigma_h^2-\mathcal{C}_e^l)\mathbf{\Lambda}_{\tilde{\mathbf{H}}}\left(\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}\right)^H\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}}{\frac{P_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+\mathbf{I}_N\right)\right]. \quad (27a)$$

trace of a positive semi-definite Hermitian matrix, we have $\det\left(\mathbf{\Pi}_{N\times N}\right)\le\prod_{i=1}^N\varpi_{ii}$. The MI maximization problem is formulated as

$$F_w \le \max_{\mathbf{\Pi}}\sum_{i=1}^N\left\{\frac{w_r}{F_r}N\log_2\left(\frac{\lambda_{ii}\left(\frac{E_t}{N}+\varpi_{ii}\right)}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+1\right) + \frac{1-w_r}{F_c}L_d\log_2\left(\frac{\left(\sigma_h^2-\mathcal{C}_e^l\right)\mu_{ii}\varpi_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+1\right)\right\} \quad (30a)$$

$$\text{subject to}\ \ \mathrm{Tr}\left(\mathbf{\Pi}\right)\le P_d;\ \varpi_{ii}\ge 0,\ 1\le i\le N, \quad (30b)$$

where ω_r is the weighting coefficient of sensing, F_r and F_c are the maximum MI (22) in Section IV-B and the communication capacity (26) in Section IV-C, respectively.

The objective function in (30a) is concave, since it is a non-negative weighted sum of two concave functions of ϖ_{ii}, i.e.,

$$N\log_2\left(\frac{\lambda_{ii}\left(\frac{E_t}{N}+\varpi_{ii}\right)}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+1\right);\ L_d\log_2\left(\frac{\left(\sigma_h^2-\mathcal{C}_e^l\right)\mu_{ii}\varpi_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+1\right).$$

Besides, $\mathrm{Tr}\left(\mathbf{\Pi}\right)\le E_d$ and $\varpi_{ii}\ge 0$ are affine. Therefore, the maximization of the concave problem in (30) can be reformulated equivalently to minimize the convex objective. The optimization problem can be solved by using Karush-Kuhn-Tucker (KKT) conditions. Let $\nu_i=\frac{\lambda_{ii}}{\left(\sigma_n^2\right)^{\frac{L}{N}}}$, $\varphi_i=\frac{\left(\sigma_h^2-\mathcal{C}_e^l\right)\mu_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}$, $\epsilon=\frac{w_rN}{\ln 2F_r}$, and $\eta=\frac{(1-w_r)L_d}{\ln 2F_c}$, we have

$$\zeta-\zeta_i=\frac{\epsilon\nu_i}{1+\nu_i(\frac{E_t}{N}+\varpi_{ii})}+\frac{\eta\varphi_i}{1+\varphi_i\varpi_{ii}}; \quad (31a)$$

$$\zeta\left(\sum_{i=1}^N\varpi_{ii}-E_d\right)=0;\ \zeta_i\varpi_{ii}=0; \quad (31b)$$

$$\zeta\ge 0;\ \zeta_i\ge 0, i=1,\cdots,N, \quad (31c)$$

where ζ and $\zeta_i,\ i=1,\cdots,N$, are the Lagrange multipliers. Since $\hat{\varpi}_{ii}\ne 0$, we have $\zeta_i=0,\ i=1,\cdots,N$. By solving (31a), the optimal solution for (31) is given by

$$\hat{\varpi}_{ii}=\frac{1}{2}\left[\frac{1}{\zeta}(\epsilon+\eta)-\left(\frac{E_t}{N}+\frac{1}{\nu_i}+\frac{1}{\varphi_i}\right) +\sqrt{\left[\left(\frac{E_t}{N}+\frac{1}{\nu_i}-\frac{1}{\varphi_i}\right)+\frac{1}{\zeta}(\eta-\epsilon)\right]^2+\frac{4\epsilon\eta}{\zeta^2}}\right]^+, \quad (32)$$

where $[x]^+=\max\{x,0\}$, and $\hat{\varpi}_{ii}$ satisfy the equality: $\sum_{i=1}^N\hat{\varpi}_{ii}-E_d=0$. The positive ζ can be obtained by the bisection search over the interval: $0<\frac{1}{\zeta}<1/\min_i\left\{\frac{\epsilon\nu_i}{(\frac{E_t}{N}+E_d)\nu_i+1}+\frac{\eta\varphi_i}{\varphi_iE_d+1}\right\}$. Once ζ is obtained, the optimal covariance matrix of the data signals in (30) can be obtained. In the case of $w_r=0$, $\hat{\varpi}_{ii}$ is consistent with (27) in Section IV-C. In the case of $w_r=1$, $\hat{\varpi}_{ii}$ coincides with (24), as described in Section IV-B.

The maximum relative MI, i.e., the sum of the relative communication and sensing MI, is given by

$$R_{\text{total}}=\sum_{i=1}^N\left\{\frac{N}{F_r}\log_2\left(\frac{\lambda_{ii}(\frac{E_t}{N}+\hat{\varpi}_{ii})}{\left(\sigma_n^2\right)^{\frac{L}{N}}}+1\right) +\frac{L_d}{F_c}\log_2\left[\frac{(\sigma_h^2-\mathcal{C}_e^l)\mu_{ii}\hat{\varpi}_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e^l+\sigma_n^2}+1\right]\right\}. \quad (33)$$

E. Complexity Analysis

The computational complexities of JCAS, OPTC and OPTS are in the same order of magnitude. In particular, OPTC and OPTS are solved by using the water-filling technique (a special class of the Lagrange multiplier method in coupling with the KKT conditions) [31]. At each iteration with a given Lagrange multiplier (or in other words, the water level, i.e., α and β' in (24) and (27), respectively), each of OPTC and OPTS has a computational complexity of $\mathcal{O}(N)$ to evaluate (24) and (27) for $i=1,\cdots,N$, respectively. The water level is recursively adjusted until convergence. With a targeted accuracy of the convergent Lagrange multiplier, ϵ_0, the overall complexity of OPTC and OPTS is $\mathcal{O}(N\log(\frac{1}{\epsilon_0}))$. JCAS is also solved by using the Lagrange multiplier method in coupling with the KKT conditions. At each iteration with the Lagrange multiplier ζ (since $\zeta_i=0$ for $i=1,\cdots,N$), JCAS has the computational complexity of $\mathcal{O}(N)$ to evaluate (32) for $i=1,\cdots,N$. The Lagrange multiplier ζ is recursively adjusted until convergence. With the targeted accuracy of the convergent Lagrange multiplier, ϵ_0, the overall complexity of JCAS is also $\mathcal{O}(N\log(\frac{1}{\epsilon_0}))$. As a result, all of OPTC, OPTS and JCAS have a linear complexity with the number of antennas, N.

V. Simulation Results

In this section, we conduct extensive simulations to evaluate the proposed methods. A system with 2 nodes is considered, and each node is equipped with 8 antennas. We consider (correlated) MIMO Rayleigh fading for both the communications and sensing channels, and the channels are independent of each other. The channels remain unchanged during the transmission of L symbols. Correlated channels are generated based on the Kronecker model, where the normalized correlation matrix has identical diagonal elements and random off-diagonal elements following uniform distributions between 0 and a maximal correlation coefficient ϵ_c. The values of ϵ_c are 0.1 and 0.8 for communication and sensing, respectively. We

assume that the communication and sensing channels have the same mean path losses, i.e., $\sigma_h^2 = \sigma_g^2 = 1$. This corresponds to the case where the mean sensing distance is approximately the square root of the communication distance. In all the simulations, the noise is a complex AWGN. The length of the training sequences L_t is equal to the number of antennas N. Without loss of generality, we set $N = L_t = 8$, and $L = 160$. The value of SNR is $(E/L)/\sigma_n^2$. For any given SNR and L, we compute the value of E with $\sigma_n^2 = 1$, and then decide the value for E_t and E_d based on κ_{op} in Theorem 1.

To provide comparable results to the *communication rate*, we introduce the *sensing rate* as the sensing MI per unit time, and assume each symbol lasts 1 unit time. Hence, the sensing rate and communication rate are equal to the ratios between their respective MI and the total number of symbols transmitted. Note that the term "sensing rate" is not widely used in the literature as the sensing channel is typically assumed to remain unchanged during the period of interest. Hence the sensing rate depends on how fast sensing channel changes.

For convenience, the abbreviations of the spatio-temporal signal design schemes proposed in this paper and other comparison schemes for the legends in figures are listed as follows, where all the schemes include optimal energy arrangement (EA), unless otherwise specified.

- *OPTC with CEE (OPTC w/ CEE or OPTC)*: produces the spatio-temporal power mask, where the singular values of the covariance matrix of the power masks are optimized to maximize the communication MI with CEE at the receiver, given the covariance matrix of the communication channel.
- *OPTC without (w/o) CEE [20]*: produces the spatio-temporal power mask, where the singular values of the covariance matrix of the power masks are optimized to maximize the communication MI without CEE at the receiver, given the covariance matrix of the communication channel.
- *OPTS*: produces the spatio-temporal power mask, where the singular values of the covariance matrix of the power masks are optimized to maximize the sensing MI, given the covariance matrix of the sensing channel.
- *Equal*: produces the spatio-temporal power mask, where the singular values of the covariance matrix of the power mask are configured to be invariably equal (as opposed to being optimized in response to the statistics of the communication or sensing channel).
- *Random*: produces the spatio-temporal power mask, where the singular values of the covariance matrix of the power mask are randomly configured.

For each result, Monte-Carlo simulations with 5,000 independent trials are conducted and the average results are provided.

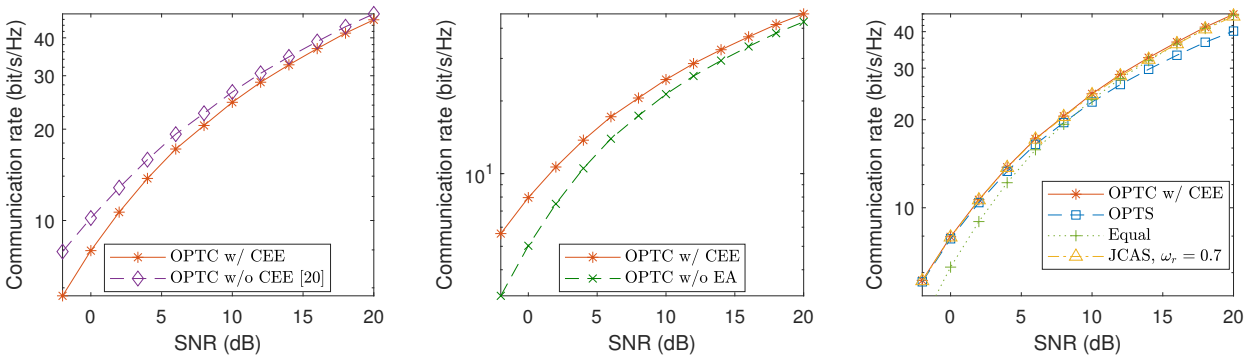
Fig. 3 plots the communication rates of the schemes under correlated channels with CEE. Fig. 4 plots the corresponding sensing rates. In Fig. 3, we also see that OPTC w/o CEE [20] is better than OPTC w/ CEE, and the gap decreases with the growth of the SNR. The impact of the CEE is non-negligible on the communication MI, as shown in (13). JCAS lies between OPTC and OPTS, and the gap to OPTC decreases with the growth of the SNR. Further, we see that OPTC w/o EA is lower than OPTC w/ CEE, especially in low SNR regimes, and the gap decreases with the increase of the SNR. In Fig. 4, we see that the sensing rate of JCAS approaches that of OPTS with the increasing SNR. Equal provides the lowest sensing rate, but its gaps to the other schemes decrease with the increase of the SNR.

Fig. 5 plots the communication rates of the considered schemes against the ratio of the training length to the total signal length under $\text{SNR} = 1$ dB and 10 dB. Fig. 6 provides the corresponding sensing rates. To plot the curves, we change the ratio of the training while keeping the total length L unchanged. We see in Fig. 5 that the communication rates of the schemes decrease significantly with the growing ratio of the training, because the number of data symbols reduces and consequently the communication rate decreases. We also see in Fig. 5 that OPTC w/ CEE is consistently better than the other schemes in terms of the communication rate, as it is optimized to maximize the communication MI. OPTS is optimized on a different purpose to maximize the sensing MI. As shown in Fig. 6, OPTS is invariably better than the other schemes in terms of the sensing rate, as it is optimized purposefully.

We see in Fig. 6 that the sensing rates of OPTC, OPTS, and JCAS decline, as the ratio of the training increases. One reason is because the feasible solution regions of problems (22), (26), and (30) are bounded by the energy for the data. With the increasing training energy, the energy for the data decreases and consequently the feasible solution regions of problems (22), (26), and (30) reduce. As a result, the objectives of problems (22), (26), and (30) decrease. In the case of OPTS, another reason for the decreasing sensing rate with the increasing training energy is that we notice that the maximization of the sensing MI requires the sum of the corresponding singular values of the training and data, i.e., $\frac{E_t}{N} + q_{ii}^{(d)}$ in (22), to be large for large λ_{ii}'s; see (24). This follows the optimal water-filling principle [31], since (22) exhibits the standard structure of a water-filling problem if $\frac{E_t}{N} + q_{ii}^{(d)}$ is treated in whole as an optimization variable. To this end, the more energy is allocated for the training (and evenly distributed between the singular values λ_{ii}, $\forall i$), the less energy could be spared to increase $q_{ii}^{(d)}$ (or $\frac{E_t}{N} + q_{ii}^{(d)}$) for the large λ_{ii}'s, and the differences between $\frac{E_t}{N} + q_{ii}^{(d)}$, $\forall i$ are smaller. As a result, the benefit of water-filling in terms of improving the objective of problem (22) decreases, and so does the sensing MI.

In Fig. 6, we also see a big sensing rate drop, as the ratio of the training symbols approaches 1. The reason is that all the methods become training-only, as the ratio of the training approaches 1. The training consists of orthogonal sequences and is the same across the different methods. Therefore, all the methods converge to the same sensing MI value offered by the orthogonal training. On the other hand, Equal has equal singular values of the covariance matrix of the data, as the orthogonal training sequences do. As a result, the sensing rate of Equal is unaffected by the ratio of the training, as shown in Fig. 6.

Fig. 7 plots the total relative MI, i.e., the sum of the relative communication MI and sensing MI, against the weighting co-



(a) Impact of CEE. (b) Impact of energy allocation. (c) Impact of signal structure.

Fig. 3. Communication rate vs. SNR, where $L = 160$, and $L_t = 8$.

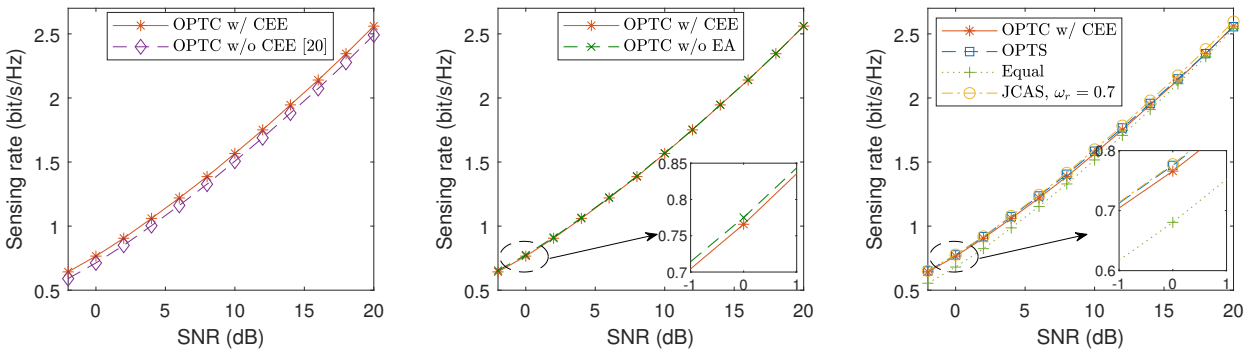
efficient of sensing, ω_r, under different signal design schemes and the maximal correlation coefficient ϵ_c. We see that the total relative MI of JCAS is the highest when $\epsilon_c = 0.1$ for communication and $\epsilon_c = 0.8$ for sensing. The total relative MI of JCAS increases first and then decreases with w_r, while the total relative MI of the other schemes does not change with w_r. The total relative MI of JCAS under $w_r = 0$ and $w_r = 1$ is equal to the total relative MI of OPTC and OPTS. Random has the lowest total relative MI. In the case of $\epsilon_c = 0.5$ for both communication and sensing, we see the weighted sum of the communication MI and sensing MI is nearly the same between OPTC, OPTS and JCAS. In other words, the transmit signal can be optimized to maximize both the communication and sensing MI, when the communication and sensing channels exhibit the same correlation characteristics.

Fig. 8 plots the total relative MI against the ϵ_c of the sensing channel $\mathbf{G}$ under both SNR $= 1$ dB and 10 dB. We can see that the total relative MI decreases with the growth of ϵ_c, especially when ϵ_c is large. JCAS is less affected by the channel correlation and outperforms the other schemes. In contrast, the increase of ϵ_c has a significant impact on the total relative MI for random and Equal, and the MI is drastically reduced with the increasing correlation.

Fig. 9 plots the sensing rate versus the communication rate, where OPTC, OPTS and JCAS are considered. The curves are plotted by varying the SNR values. As shown in the figure, OPTS and OPTC have a consistent gap with the increase of the SNR. We note that OPTC and OPTS are two special cases of JCAS when only the communication MI or the sensing MI is maximized. In contrast, JCAS maximizes a weighted sum of the communication and sensing MI. JCAS offers the flexibility to accommodate both the communication and sensing and bridges the gap between OPTC and OPTS by adjusting the weights of the two aspects in the optimization of the spatio-temporal power mask of the data (as compared to the sole focus of OPTC and OPTS on one of the aspects).

The benefit of JCAS is to provide the flexibility to trade off between the communication and sensing capabilities in an effective way. As shown in Fig. 9, OPTS maximizes the sensing rate at the cost of communication rate. As also shown in Fig. 9, the use of JCAS bridges the gap between OPTS and OPTC, and strives for high communication capacity without substantially penalizing the sensing capability (or the other way around).

By maximizing the sensing MI, the considered system maximizes its capacity to detect and resolve subtle changes in the sensing channel (e.g., caused by small changes in the environment, e.g., intruders or trespassers). Given the maximized sensing MI or capacity, the detection accuracy is expected to improve. In Fig. 10, we take the MMSE-based



(a) Impact of CEE. (b) Impact of energy allocation. (c) Impact of signal structure.

Fig. 4. Sensing rate vs. SNR, where $L = 160$, and $L_t = 8$.

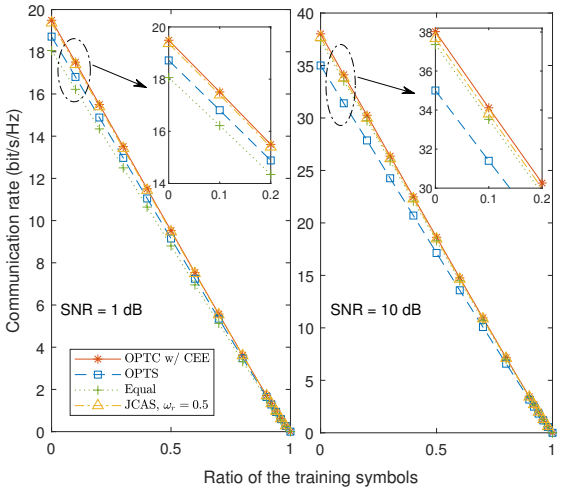


Fig. 5. Communication rate vs. the ratio of training symbols where the total number of the transmit symbols $L = 160$, SNR $= 1, 10$ dB.

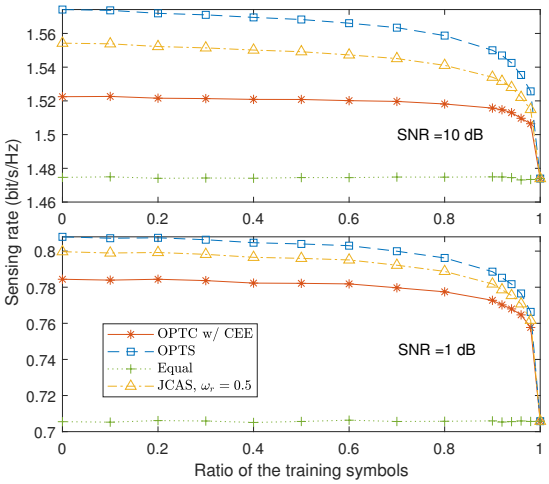


Fig. 6. Sensing rate vs. the ratio of training symbols where the total number of the transmit symbols $L = 160$, SNR $= 1$ dB.

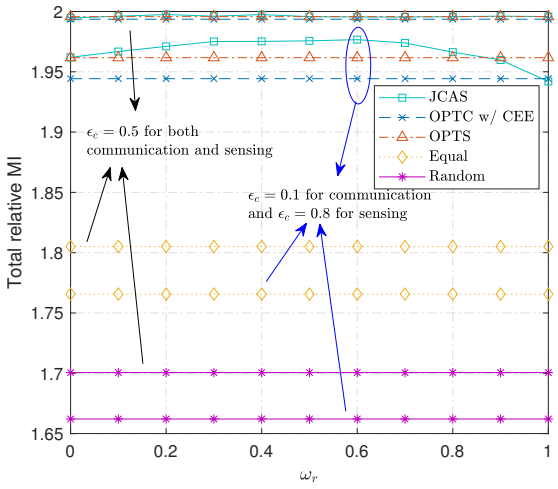


Fig. 7. Total relative MI vs. ω_r for different signal design methods with SNR $= 1$ dB, and ϵ_c for communication and sensing are 0.8 and 0.1, respectively.

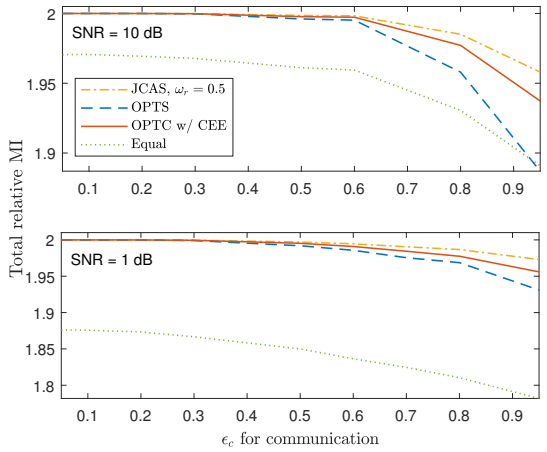


Fig. 8. Total relative MI vs. ϵ_c for communication channels, where SNR $= 1$ dB, and ϵ_c for sensing is 0.3.

estimation of the sensing channel for example. Fig. 10(a) compares the estimation errors based on the differently optimized spatio-temporal power masks of the data signals, including OPTS, OPTC, JCAS (and Equal). We see that the lowest root mean square error (RMSE) is achieved based on the spatio-temporal power mask optimized by OPTS to maximize the sensing MI, followed by the one optimized by JCAS to jointly consider both the communication and sensing MI, and the one optimized by OPTC to maximize the communication MI. As also shown in Fig. 10(a), the gain of OPTS over OPTC is considerable, especially when the SNR is low. OPTS and OPTC are two special cases of JCAS, and provide the tight performance bounds for JCAS. In particular, JCAS can turn into OPTS or OPTC by configuring the weighting coefficients of sensing and communication MI.

In Figs. 10(b) and 10(c), we pre-populate the 5,000 realizations of the sensing channel used to plot Fig. 10(a), and nominate the one with the smallest total distance to the rest of the realizations to be the target-free sensing channel. We run the MMSE technique to estimate the sensing channel and measure the Euclidean distance between the estimated channel and the target-free sensing channel. The distance is normalized by the number of pairwise channels. We can compare the distance against a detection threshold, δ_{th}, and alert significant changes exceeding the threshold in the sensing channel. For each SNR and detection threshold value, the proposed algorithms are tested under all the realizations with independently generated receiver noises. The results in Figs. 10(b) and 10(c) are consistent with Fig. 10(a).

VI. Conclusion

We presented the optimal signal designs of MIMO JCAS systems by considering both training and data symbols. We

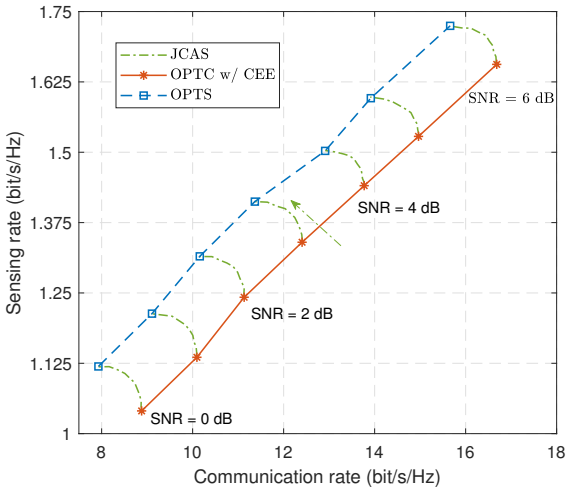


Fig. 9. Trade-off curve between communication and radio sensing, where ϵ_c for communication and sensing are 0.8 and 0.3, respectively.

proposed an optimal energy arrangement scheme under MMSE estimators for MIMO communication channels. Among the three optimization strategies we studied, the design that maximizes a weighted sum of relative MI is shown to achieve the best overall performance for JCAS. The design is also less affected by the varying channel correlation than the other two design methods. Extensive simulations corroborated the merit of the proposed techniques.

Appendix A
Proof of Theorem 1

As defined earlier, the SVD of spatial correlation matrix of $\tilde{\mathbf{H}}$ is $\Sigma_{\tilde{\mathbf{H}}} = \frac{1}{N}\mathbb{E}\{\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H\} = \mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Lambda}_{\tilde{\mathbf{H}}}\mathbf{U}_{\tilde{\mathbf{H}}}^H$, where $\mathbf{U}_{\tilde{\mathbf{H}}}$ is a unitary matrix and $\mathbf{\Lambda}_{\tilde{\mathbf{H}}} = \mathrm{diag}\{\mu_{11}, \cdots, \mu_{NN}\}$ is a diagonal matrix. Based on (13) and the lower bound of CEE, we can get an upper bound for the MI between $\mathbf{X}_d$ and $\mathbf{Y}_{\mathrm{com}}^d$ as

$$I\left(\mathbf{X}_d; \mathbf{Y}_{\mathrm{com}}^d|\hat{\mathbf{H}}\right) \leq L_d \log_2\left[\det\left(\frac{\left(\sigma_h^2 - \mathcal{C}_e\right)\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}\mathbf{\Lambda}_{\tilde{\mathbf{H}}}\mathbf{U}_{\tilde{\mathbf{H}}}^H\mathbf{X}_d^H}{\frac{E_d}{L_d}\mathcal{C}_e + \sigma_n^2} + \mathbf{I}_N\right)\right] \tag{34a}$$

$$= L_d \log_2\left[\det\left(\frac{\left(\sigma_h^2 - \mathcal{C}_e\right)\mathbf{\Lambda}_{\tilde{\mathbf{H}}}\left(\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}\right)^H\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}}{\frac{E_d}{L_d}\mathcal{C}_e + \sigma_n^2} + \mathbf{I}_N\right)\right], \tag{34b}$$

where the equality in (34a) is achieved when the lower bound CEE $\mathcal{C}_e$ is achieved; (34b) is based on the *Sylvester's determinant* theorem [29].

Because $\mathbf{\Xi} = \left(\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}}\right)^H\mathbf{X}_d\mathbf{U}_{\tilde{\mathbf{H}}} = \left(\mathbf{Y}_{\mathrm{com}}^d\right)^H\mathbf{Y}_{\mathrm{com}}^d$ and its (i,j)-th entry is ξ_{ij}, based on Hadamard's inequality for the determinant and trace of a positive semi-definite Hermitian matrix, we have $\det\left(\mathbf{\Xi}_{N\times N}\right) \leq \prod_{i=1}^N \xi_{ii}$. The upper bound of the MI between $\mathbf{X}_d$ and $\mathbf{Y}_{\mathrm{com}}^d$ can be obtained as

$$I\left(\mathbf{X}_d; \mathbf{Y}_{\mathrm{com}}^d|\hat{\mathbf{H}}\right) \leq \sum_{i=1}^N \underbrace{\left\{L_d \log_2\left[\frac{\left(\sigma_h^2 - \mathcal{C}_e\right)\mu_{ii}\xi_{ii}}{\frac{E_d}{L_d}\mathcal{C}_e + \sigma_n^2} + 1\right]\right\}}_{f(\xi_{ii})}, \tag{35}$$

where the equality is achieved if and only if $\mathbf{\Xi}$ is a diagonal matrix.

As $f\left(\xi_{ii}\right)$ is a monotonically decreasing and concave function of ξ_{ii}, based on Jensen's inequality, we can obtain the expectation of $I\left(\mathbf{X}_d; \mathbf{Y}_{\mathrm{com}}^d|\hat{\mathbf{H}}\right)$, i.e., $\mathbb{E}\left[I\left(\mathbf{X}_d; \mathbf{Y}_{\mathrm{com}}^d|\hat{\mathbf{H}}\right)\right]$, as given in (36), where (36c) is obtained by substituting $\mathbb{E}\left[\xi_{ii}\right] = \frac{1}{N}\mathrm{Tr}\left(\Sigma_{\mathbf{X_d}}\right) = \frac{1}{N}\mathrm{Tr}\left(\mathbf{\Xi}\right) = L_d\sigma_d^2 = \frac{\kappa E}{N}$ and $(1-\kappa)E = NL_t\sigma_t^2$ into (36b).

$$\mathbb{E}\left[I\left(\mathbf{X}_d; \mathbf{Y}_{\mathrm{com}}^d|\hat{\mathbf{H}}\right)\right] \leq \mathbb{E}\left[\sum_{i=1}^N f\left(\xi_{ii}\right)\right] \tag{36a}$$

$$= \sum_{i=1}^N \mathbb{E}\left[f\left(\xi_{ii}\right)\right] \leq \sum_{i=1}^N f\left(\mathbb{E}\left[\xi_{ii}\right]\right) \tag{36b}$$

$$= L_d \sum_{i=1}^N \left\{\log_2\left[\frac{EL_d}{(L_d - N)N\sigma_n^2}\frac{\kappa\left(1-\kappa\right)}{-\kappa + \frac{L_d}{L_d - N}\left(1 + \frac{N\sigma_n^2}{E\sigma_h^2}\right)} + 1\right]\right\}. \tag{36c}$$

From (36c), we can see that different values of κ can lead to different mean MI given E and N, denoted as $\rho = \frac{L_dE}{(L_d-N)N\sigma_n^2} \cdot \frac{\kappa(1-\kappa)}{-\kappa + \frac{L_d}{L_d-N}\left(1+\frac{N\sigma_n^2}{E\sigma_h^2}\right)}$. Referring to the cases considered in [9], to maximize ρ over $0 \leq \kappa \leq 1$, we can consider the following three cases:

1) $L_d = N$: The maximal ρ, denoted by $\rho_{\max}$, is $\rho_{\max} = \frac{E^2\sigma_h^4}{4N\sigma_n^2(N\sigma_n^2 + E\sigma_h^2)}$, from which it follows that $\kappa_{\mathrm{op}} = \frac{1}{2}$.
2) $L_d > N$: We rewrite ρ as $\rho = \frac{L_dE}{(L_d-N)N\sigma_n^2} \cdot \frac{\kappa(1-\kappa)}{-\kappa+\Gamma}$, where $\Gamma = \frac{L_d}{L_d-N}\left(1 + \frac{N\sigma_n^2}{E\sigma_h^2}\right) > 1$. The maximal SNR $\rho_{\max}$ can be obtained as $\rho_{\max} = \frac{L_dE}{(L_d-N)N\sigma_n^2}\left(\sqrt{\Gamma} - \sqrt{\Gamma-1}\right)^2$, and it follows that $\kappa_{\mathrm{op}} = \Gamma - \sqrt{\Gamma(\Gamma-1)}$.
3) $L_d < N$: We rewrite ρ as $\rho = \frac{L_dE}{(N-L_d)N\sigma_n^2} \cdot \frac{\kappa(1-\kappa)}{\kappa-\Gamma}$, where $\Gamma = \frac{L_d}{L_d-N}\left(1 + \frac{N\sigma_n^2}{E\sigma_h^2}\right) < 0$. The maximal SNR $\rho_{\max}$ can be obtained as $\rho_{\max} = \frac{L_dE}{(L_d-N)N\sigma_n^2}\left(\sqrt{-\Gamma} - \sqrt{-\Gamma-1}\right)^2$, and it follows that $\kappa_{\mathrm{op}} = \Gamma + \sqrt{\Gamma(\Gamma-1)}$.

Therefore, we can obtain the lower bound of the CEE as presented in (17).

References

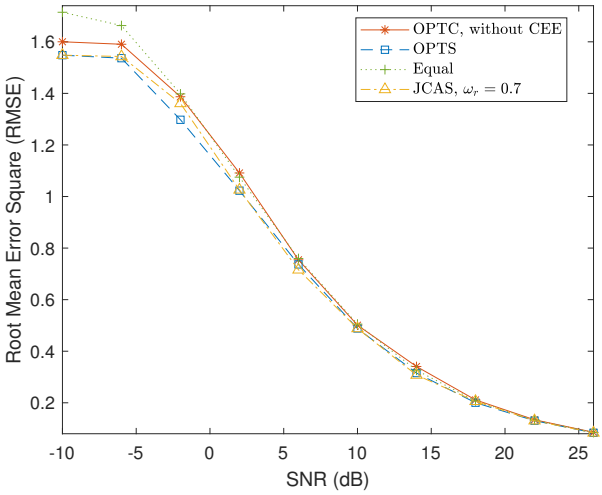
[1] P. Kumari, M. E. Eltayeb, and R. W. Heath, "Sparsity-aware adaptive beamforming design for ieee 802.11ad-based joint communication-radar," in *2018 IEEE Radar Conference (RadarConf18)*, Apr. 2018, pp. 0923–0928.

[2] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Radar-communications convergence: Coexistence, cooperation, and co-design," *IEEE Trans. on Cogn. Commun. Netw.*, vol. 3, no. 1, pp. 1–12, Mar. 2017.

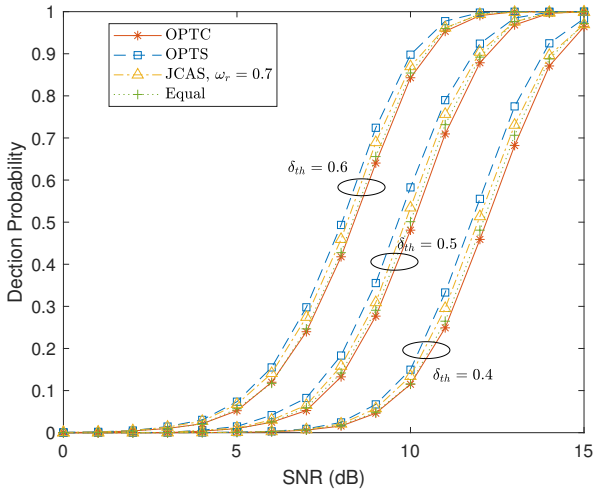
[3] B. Paul, A. R. Chiriyath, and D. W. Bliss, "Survey of RF communications and sensing convergence research," *IEEE Access*, vol. 5, pp. 252–270, Dec. 2017.

[4] M. L. Rahman, J. A. Zhang, X. Huang, Y. J. Guo, and R. W. Heath Jr, "Framework for a perceptive mobile network using joint communication and radar sensing," *IEEE Trans. Aerosp. Electron. Syst.*, 2019.

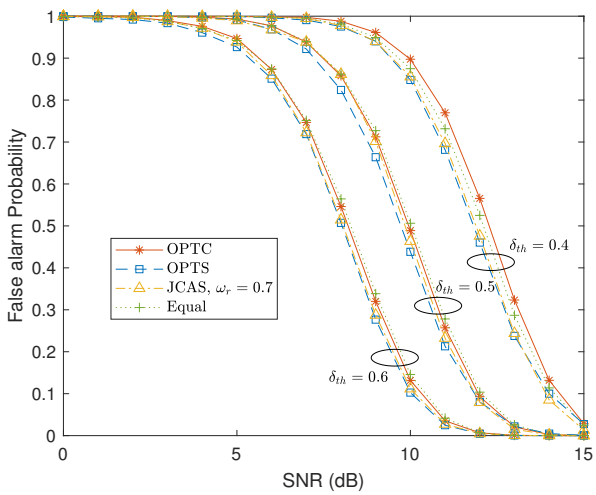
[5] R. C. Daniels, E. R. Yeh, and R. W. Heath, "Forward collision vehicular radar with IEEE 802.11: Feasibility demonstration through measurements," *IEEE Trans. Veh. Tech.*, vol. 67, no. 2, pp. 1404–1416, Feb 2018.



(a) RMSE vs. SNR



(b) Detection probability vs. SNR



(c) False alarm probability vs. SNR

Fig. 10. Sensing performance vs. SNR, where $L = 160$ and $L_t = 8$.

[6] J. A. Zhang, X. Huang, Y. J. Guo, J. Yuan, and R. W. Heath, "Multibeam for joint communication and radar sensing using steerable analog antenna arrays," *IEEE Trans. Veh. Tech.*, vol. 68, no. 1, pp. 671–685, 2019.

[7] M. R. Bell, "Information theory and radar waveform design," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1578–1597, Sept. 1993.

[8] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 1, pp. 330–343, Jan. 2007.

[9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[10] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[11] B. Tang, J. Tang, and Y. Peng, "MIMO radar waveform design in colored noise based on information theory," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4684–4697, Sept. 2010.

[12] Z. Zhu, S. Kay, and R. S. Raghavan, "Information-theoretic optimal radar waveform design," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 274–278, Mar. 2017.

[13] Y. Chen *et al.*, "Adaptive distributed MIMO radar waveform optimization based on mutual information," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1374–1385, Apr. 2013.

[14] Y. Liu, H. Wang, and J. Wang, "Robust multiple-input multiple-output radar waveform design in the presence of clutter," *IET Radar, Sonar Navigation*, vol. 10, no. 7, pp. 1249–1259, 2016.

[15] A. R. Chiriyath, B. Paul, and D. W. Bliss, "Joint radar-communications information bounds with clutter: The phase noise menace," in *2016 IEEE Radar Conference (RadarConf)*, May 2016, pp. 1–6.

[16] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, "Inner bounds on performance of radar and communications co-existence," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 464–474, Jan. 2016.

[17] F. Liu, C. Masouros, A. Li, and T. Ratnarajah, "Robust MIMO beamforming for cellular and radar coexistence," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 374–377, Jun. 2017.

[18] H. D. M. A. R. Chiriyath, S. Ragi and D. W. Bliss, "Novel radar waveform optimization for a cooperative radar-communications system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 3, p. 1160–1173, Jun. 2019.

[19] B. Paul, A. R. Chiriyath, and D. W. Bliss, "Joint communications and radar performance bounds under continuous waveform optimization: The waveform awakens," in *2016 IEEE Radar Conference (RadarConf)*, May 2016, pp. 1–6.

[20] R. Xu, L. Peng, W. Zhao, and Z. Mi, "Radar mutual information and communication channel capacity of integrated radar-communication system using MIMO," *ICT Express*, vol. 1, no. 3, pp. 102 – 105, 2015.

[21] Y. Liu, G. Liao, J. Xu, Z. Yang, and Y. Zhang, "Adaptive OFDM integrated radar and communications waveform design based on information theory," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2174–2177, Oct. 2017.

[22] Y. Liu, G. Liao, R. S. Blum, and Z. Yang, "Robust OFDM integrated radar and communications waveform design," *arXiv preprint arXiv:1805.01511*, 2018.

[23] D. A. Shnidman, "Radar detection probabilities and their calculation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 3, pp. 928–950, 1995.

[24] B. K. Chalise, M. G. Amin, and B. Himed, "Performance tradeoff in a unified passive radar and communications system," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1275–1279, 2017.

[25] C. Park, "Recursive algorithm for sliding walsh hadamard transform," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2827–2836, 2014.

[26] G. Galati and G. Pavan, "Waveforms design for modern and MIMO radar," in *Eurocon 2013*, Jul. 2013, pp. 508–513.

[27] C. Artigue and P. Loubaton, "On the precoder design of flat fading MIMO systems equipped with MMSE receivers: A large-system approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4138–4155, Jul. 2011.

[28] J. Wang, O. Y. Wen, and S. Li, "Near-optimum pilot and data symbols power allocation for MIMO spatial multiplexing system with zero-forcing receiver," *IEEE Signal Process. Lett.*, vol. 16, no. 5, pp. 358–361, May 2009.

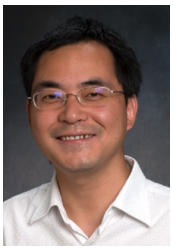
[29] G. T. Gilbert, "Positive definite matrices and sylvester's criterion," *The American Mathematical Monthly*, vol. 98, no. 1, pp. 44–46, 1991.

[30] L. Berriche, K. Abed-Meraim, and J. C. Belfiore, "Cramer-rao bounds for MIMO channel estimation," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, May 2004, pp. iv–397–iv–400 vol.4.

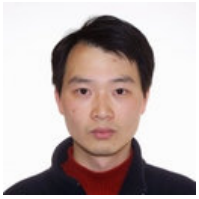
[31] L. Lai and H. El Gamal, "The water-filling game in fading multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2110–2122, 2008.

Xin Yuan (S'16-M'19) received the B.E. degree from Taiyuan University of Technology, Shanxi, China, in 2013, and the dual Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, and University of Technology Sydney (UTS), Sydney, Australia, in 2019 and 2020, respectively. Currently she is a Research Scientist at CSIRO, Australia. Her research interests include unmanned aerial vehicle communication and networking, modeling and performance analysis of UAV networks.

Zhiyong Feng (M'08-SM'15) received her B.E., M.E., and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China. She is a professor at BUPT, and the director of the Key Laboratory of the Universal Wireless Communications, Ministry of Education, P.R.China. She is a senior member of IEEE, vice chair of the Information and Communication Test Committee of the Chinese Institute of Communications (CIC). Currently, she is serving as Associate Editors-in-Chief for China Communications, and she is a technological advisor for international forum on NGMN. Her main research interests include wireless network architecture design and radio resource management in 5th generation mobile networks (5G), spectrum sensing and dynamic spectrum management in cognitive wireless networks, and universal signal detection and identification.



J. Andrew Zhang (M'04-SM'11) received the B.Sc. degree from Xi'an JiaoTong University, China, in 1996, the M.Sc. degree from Nanjing University of Posts and Telecommunications, China, in 1999, and the Ph.D. degree from the Australian National University, in 2004. Currently, Dr. Zhang is an Associate Professor in the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He was a researcher with Data61, CSIRO, Australia from 2010 to 2016, the Networked Systems, NICTA, Australia from 2004 to 2010, and ZTE Corp., Nanjing, China from 1999 to 2001. Dr. Zhang's research interests are in the area of signal processing for wireless communications and sensing, and autonomous vehicular networks. He has published more than 140 papers in leading international Journals and conference proceedings, and has won 4 best paper awards for his work. He is a recipient of CSIRO Chairman's Medal and the Australian Engineering Innovation Award in 2012 for exceptional research achievements in multi-gigabit wireless communications.



Wei Ni (M'09-SM'15) received the B.E. and Ph.D. degrees in Electronic Engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. Currently he is a Senior Scientist, and Group and Project Leaders at CSIRO, Australia, since 2009. He also holds honorary positions at the University of New South Wales (UNSW), Macquarie University (MQ) and the University of Technology Sydney (UTS). Prior to this he was a postdoctoral research fellow at Shanghai Jiaotong University (2005-2008), Research Scientist and Deputy Project Manager at the Bell Labs R&I Center, Alcatel/Alcatel-Lucent (2005-2008), and Senior Researcher at Devices R&D, Nokia (2008-2009). His research interests include optimization, game theory, graph theory, as well as their applications to network and security.

Dr Ni serves as Vice Chair of IEEE NSW VTS Chapter since 2018, served as Editor for Hindawi Journal of Engineering from 2012-2015, secretary of IEEE NSW VTS Chapter from 2015-2018, Track Chair for VTC-Spring 2017, Track Co-chair for IEEE VTC-Spring 2016, and Publication Chair for BodyNet 2015. He also served as Student Travel Grant Chair for WPMC 2014, a Program Committee Member of CHINACOM 2014, a TPC member of IEEE ICC14, ICCC15, EICE14, and WCNC10.

Ren Ping Liu (M'09-SM'14) is a Professor and Head of Discipline of Network and Cybersecurity in the School of Electrical and Data Engineering at University of Technology Sydney. Prior to that he was a Principal Scientist and Research Leader at CSIRO, where he led wireless networking research activities. Professor Liu was the winner of Australian Engineering Innovation Award and CSIRO Chairman medal. He specialises in protocol design and modelling, and has delivered networking solutions to a number of government agencies and industry customers. His research interests include 5G, VANET, IoT, cybersecurity, and Blockchain.

Professor Liu was the founding chair of IEEE NSW VTS Chapter and a Senior Member of IEEE. He served as Technical Program Committee chairs and Organising Committee chairs in a number of IEEE Conferences. Ren Ping Liu received his B.E.(Hon) and M.E. degrees from Beijing University of Posts and Telecommunications, China, and the Ph.D. degree from the University of Newcastle, Australia.

Zhiqing Wei (S'12-M'15) received his B.S. and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT) in 2010 and 2015. Now he is an associate professor at BUPT. He has published 1 books, 3 book chapters and over 50 papers. He was granted the Exemplary Reviewer of IEEE Wireless Communications Letters in 2017, the Best Paper Award of International Conference on Wireless Communications and Signal Processing (WCSP) 2018. He was the Registration Co-Chair of IEEE/CIC International Conference on Communications in China (ICCC) 2018 and the publication Co-Chair of IEEE/CIC ICCC 2019. His research interest is the performance analysis and optimization of mobile ad hoc networks.

Changqiao Xu (S'12-M'15) received the Ph.D. degree from the Institute of Software, Chinese Academy of Sciences (ISCAS) in Jan. 2009. He was an Assistant Research Fellow and R&D Project Manager in ISCAS from 2002 to 2007. He was a researcher at Athlone Institute of Technology and Joint Training PhD at Dublin City University, Ireland during 2007-2009. He joined Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in Dec. 2009. Currently, he is a full Professor with the State Key Laboratory of Networking and Switching Technology, and Director of the Next Generation Internet Technology Research Center at BUPT. His research interests include Future Internet Technology, Mobile Networking, Multimedia Communications, and Network Security. He has published over 160 technical papers in prestigious international journals and conferences, including IEEE Comm. Surveys & Tutorials, IEEE Wireless Comm., IEEE Comm. Magazine, IEEE/ACM ToN, IEEE TMC etc. He has served a number of international conferences and workshops as a Co-Chair and Technical Program Committee member. He is currently serving as the Editor-in-Chief of Transactions on Emerging Telecommunications Technologies (Wiley). He is Senior member of IEEE.