Robust Waveform Design for Integrated Sensing and Communication

Shixiong Wang, Wei Dai, Haowei Wang, and Geoffrey Ye Li, *Fellow, IEEE*

Abstract—Integrated sensing and communication (ISAC), which enables hardware, resources (e.g., spectra), and waveforms sharing, is becoming a key feature in future-generation communication systems. This paper investigates performance characterization and waveform design for ISAC systems when the underlying true communication channels are not accurately known. With uncertainty in a nominal communication channel, the nominal Pareto frontier of the sensing and communication performances cannot represent the true performance trade-off of a real-world operating ISAC system. Therefore, this paper portrays the robust (i.e., conservative) Pareto frontier considering the uncertainty in the communication channel. To be specific, the lower bound of the true (but unknown) Pareto frontier is investigated, technically by studying robust waveform design problems that find the best waveforms under the worst-case channels. The robust waveform design problems examined in this paper are shown to be non-convex and high-dimensional, which cannot be solved using existing optimization techniques. As such, we propose a computationally efficient solution framework to approximately solve them. Simulation results show that by solving the robust waveform design problems, the lower bound of the true but unknown Pareto frontier, which characterizes the sensing-communication performance trade-off under communication channel uncertainty, can be obtained.

Index Terms—ISAC, Performance Characterization, Robust Waveform Design, Non-Convex Non-Smooth Optimization.

I. INTRODUCTION

INTEGRATED sensing and communication (ISAC) is one of the enabling technologies for the sixth-generation (6G) communications. It uses one single hardware system to simultaneously realize the sensing and communication functions. This integration is able to improve spectral efficiency, reduce platform size, and control power consumption. From the signal processing perspective, one of the main features of ISAC is that the same transmitted waveform, called dual-functional waveform, is used for both sensing and communication functions [1]–[5]. This paper is concerned with performance characterization and waveform design for ISAC systems.

A. Performance Characterization and Waveform Design

Performance characterization for ISAC contains two connotations: 1) to evaluate the performances of an ISAC sys-

tem, in terms of sensing and communication, at given dual-functional waveforms; 2) to explore the performance boundary by depicting the Pareto frontier of sensing versus communication. Typical performance measures for sensing include the Cramér–Rao bounds for the estimated parameters [4], [6] and the sensing mutual information between received signals and sensing channels (i.e., sensing rate) [7]–[9], while those for communication include the distortion minimum mean-squared error [10] and the mutual information between the transmitted and the received signals (i.e., channel capacity) [4], [11]. Waveform design for ISAC, as a technical aspect of performance characterization, aims to depict the Pareto frontier (i.e., to achieve the performance boundary) through finding optimal dual-functional waveforms at different trade-off levels because an optimal waveform for sensing is not necessarily optimal for communication [4], [8], [12], and vice versa. Supposing that an ISAC system aims to minimize the sensing metric while maximizing the communication metric, e.g., the Cramér–Rao bound for sensing and the achievable sum-rate for communication, the performance Pareto frontier is shown in Fig. 1; cf., e.g., [12, Fig. 3]. Every point on the Pareto frontier is associated with at least one optimal dual-functional waveform that balances the performances between sensing and communication.
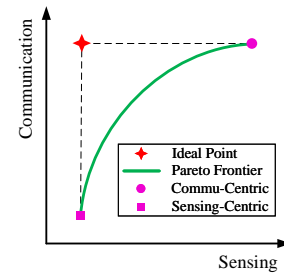

Fig. 1. The Pareto frontier represents the performance trade-off between sensing and communication. For a given ISAC system and specified sensing-communication performance measures, the Pareto frontier is fixed, which is obtained through waveform design. For communication-centric waveform design, the communication performance reaches its best level, while for sensing-centric waveform design, the sensing performance arrives at its limit.

According to the performance priority, as well as the underlying system and signal features, existing ISAC waveform design methods can be categorized into the following three classes [3], [13], [5, Fig. 1].

- *Communication-centric waveform design*: In this class, the communication performance is guaranteed to be (nearly) perfect, with a secondary focus on optimizing the sensing performance. Typically, communication

S. Wang, W. Dai, and G. Li are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, United Kingdom (E-mail: s.wang@u.nus.edu; wei.dai1@imperial.ac.uk; geoffrey.li@imperial.ac.uk).

H. Wang is with Rice-Rick Digitalization, Singapore 308900 (E-mail: haowei_wang@ricerick.com).

base stations and waveforms are used to realize the sensing function because the sensing information can be extracted from communication electromagnetic waves bounced back from environments and targets. While the communication performance can be virtually unaffected, the sensing performance may be limited and difficult to further enhance. Classical representatives of this category encompass the orthogonal frequency division multiplexing (OFDM) waveforms [14], [15], the orthogonal time frequency space (OTFS) waveforms [16], etc.

- *Sensing-centric waveform design*: In this class, the sensing performance is guaranteed to be (nearly) perfect, with a secondary focus on optimizing the communication performance. Typically, sensing base stations (e.g., radar) and waveforms are used to realize the communication function; to be specific, communication information is modulated onto sensing waveforms. Since these waveforms are (almost) optimal for sensing, the sensing performance can remain approximately uninfluenced, however, the communication performance may be restricted; for example, the communication data rate cannot be sufficiently high. Classical representatives of this category include the chirp waveforms [17], the index-modulation-based waveforms [18], etc; see also, e.g., [19].
- *Joint waveform design*: To fully balance the communication and sensing performances, i.e., to eliminate the inherent performance limits on either of them, ISAC systems can be independently designed and realized without relying on existing hardware and waveforms; see, e.g., [12], [20]. For an extensive review, see, e.g., [3, Sec. V], [5, Sec. IV].

Intuitively, joint waveform design portrays the Pareto frontier in Fig. 1, while communication-centric and sensing-centric designs delineate the two extreme points on the frontier. In terms of design methodology, the existing methods can also be categorized into two main streams. The first stream directly designs the dual-functional waveforms through optimizing and constraining various performance measures [12], [15], [21], while the second stream finds the balanced waveforms through similarity matching with known benchmark waveforms [20].

B. Problem Statement, Research Aims, and Related Works

The existing ISAC waveform design methods face the following major drawback: In practice, the **true** communication channel is unknown, and only the **nominal** (i.e., estimated or approximated) channel is available. This point can be understood from three aspects.

- First, the communication channel is time-selective (i.e., time-varying) in a frame.
- Second, the channel model used might be inexact.
- Third, any statistical method cannot give exact channel estimation when training samples are limited.

As a consequence, the nominal Pareto frontier obtained under the nominal communication channel cannot characterize the actual sensing-communication performance trade-off; see Fig. 2(a). To be specific, the true Pareto frontier associated with the actual communication channel may lie above, across, or

below the nominal Pareto frontier. When the actual channel varies over time, the true Pareto frontier may even not remain fixed. Since the true channel is unknown and even random (e.g., time-varying), characterizing the upper and lower bounds of the actual Pareto frontier(s) is of natural importance and significance; see Fig. 2(b). However, such uncertainty-aware performance characterization has not been discussed in the existing ISAC literature.

(a) Nominal and true frontiers   (b) Bounds of true frontiers
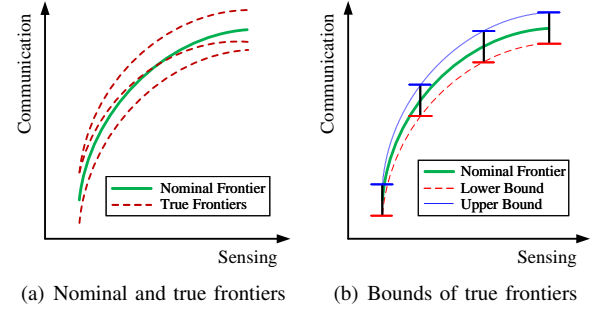
Fig. 2. Pareto frontiers of sensing-communication performances. The true Pareto frontier is unknown and even time-varying so the nominal frontier cannot characterize the actual performance trade-off; NB: every realization of the random communication channel corresponds to one true Pareto frontier. However, we can identify the lower and upper bounds of the true frontier(s).

As an initiative work on ISAC performance characterization under communication channel uncertainty, this paper focuses on exploring the lower bound of the actual Pareto frontier(s), that is, the **robust** (i.e., **conservative**) performance boundary. To achieve such conservativeness, robust waveform design methods that utilize the worst-case information of the communication channel are leveraged.

Literature on robust dual-functional waveform design for ISAC is rather lacking, and the highly related records include only [22] and [23]. In [22], a robust beamforming method for ISAC that handles the uncertainties in the communication channel and the target angles is studied, which, however, limits the ISAC waveforms to linear combinations of the communication symbols. Another ISAC waveform design work considering the uncertainty in target angles is [23], which, however, does not deal with the uncertainty in the communication channel. Other weakly-related works considering imperfect channel information encompass, e.g., robust beamforming for ISAC [24]–[27], which, as in [22], require the ISAC waveforms to be linear transforms of the communication symbols. For general ISAC waveform design, nevertheless, the technical methodology is beyond linear beamforming; see [20], [21]. The main issue of existing works on ISAC robust waveform design is that they do not reveal the relation between conservative performance characterization (i.e., lower bound discovery) and robust waveform design; cf. Fig. 2. Technically, this is because, facing the uncertainty in the communication channel, these works focus on guaranteeing the satisfaction of the worst-case communication performance rather than optimizing it; see, e.g., [22, Eq. (7)], [25, Eq. (17b)].

Remark 1 (Specificity of Robust Design): For any given existing ISAC waveform design method, in principle, we can propose a corresponding robust counterpart that is insensitive

to the potential uncertainties in the involved parameters and/or models. That is, the mathematical formulations of robust counterparts are specific to the adopted original (i.e., non-robust) ISAC waveform design frameworks because different waveform design philosophies give different waveform design formulations; cf. [3], [5], [20]. To develop this paper, as an illustration of uncertainty-aware performance characterization and robust design, we exemplify the waveform design framework proposed in [20], where balanced waveforms that are simultaneously close to the optimal waveforms for sensing and communication are studied. The future papers following this work can consider robustifying other existing ISAC waveform design methods. □

C. Contributions

In line with the basic system setups in [20], the contributions of this paper can be highlighted as follows:

1) A robust dual-functional waveform design framework that combats the uncertainty of the communication channel is proposed; see Subsection II-B. The framework is established leveraging the min-max optimization, where the minimization of the objective is to design the optimal waveforms and the maximization of the objective is to find the worst-case information of the communication channel; see Problems (9) and (10).

2) The concept of ISAC performance characterization under channel uncertainty is proposed; see Fig. 2. We show that finding the lower bound of the true Pareto frontier(s), i.e., the conservative performance boundary, is related to solving the robust waveform design problems; see Section III, especially Facts 1 and 2, and Remarks 3 and 4.

3) To tackle the two-layer min-max robust waveform design problems (9) and (10), an approximate solution framework is proposed; see Section IV, especially Lemmas 1 and 2, Insight 1, Theorems 1 and 2, Methods 1, 2, and 3, and Table I.

The first two contributions are new to the ISAC community, whereas the third contribution also enriches the general min-max optimization theory and practice (e.g., robustness theory, game theory, etc.).

D. Notations

Vectors (column by default) and matrices are denoted by bold symbols and are written in lowercase and uppercase letters, respectively; e.g., $\boldsymbol{x}$ is a vector while $\boldsymbol{X}$ is a matrix. Deterministic quantities are denoted by Italic symbols (e.g., $\boldsymbol{x}$, $\boldsymbol{X}$) while random quantities are denoted by normal symbols (e.g., $\mathbf{x}$, $\mathbf{X}$). Sets are denoted by calligraphic symbols, e.g., $\mathcal{X}$. Let $\mathbb{C}^d$, $\mathbb{R}^d$ stand for a d-dimensional complex, real space, respectively. Let $\boldsymbol{X}^{\mathsf{T}}$, $\boldsymbol{X}^{\mathsf{H}}$, $\mathrm{Tr}[\boldsymbol{X}]$, $\|\boldsymbol{X}\|$, $\mathrm{vec}(\boldsymbol{X})$, $\mathrm{diag}(\boldsymbol{X})$, and $\boldsymbol{X}^{-1}$ denote the transpose, the conjugate transpose, the trace, a norm (e.g., the Frobenius norm $\|\boldsymbol{X}\|_F$), the column vectorization, the diagonal entries, and the inverse of the matrix $\boldsymbol{X}$, respectively. When the inverse of $\boldsymbol{X}$ does not exist, we use $\boldsymbol{X}^{-1}$ to denote its Moore–Penrose pseudo-inverse. We use $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ to denote the circularly-symmetric complex normal distribution with zero mean, covariance matrix $\boldsymbol{\Sigma}$, and zero pseudo-covariance matrix. Let the running index set induced by an integer I be denoted by $[I] := \{1, 2, \ldots, I\}$ and the N-dimensional identity matrix by $\boldsymbol{I}_N$.

II. PROBLEM FORMULATION

A. Basic Setups and Traditional Waveform Design

Following the basic setups in [20], we consider an ISAC system with an N-antenna uniform linear array (ULA). The system simultaneously senses targets and communicates with K single-antenna downlink users using shared dual-functional waveform $\boldsymbol{X} \in \mathbb{C}^{N \times L}$, where L denotes the length of the communication frame (from the perspective of communication) and the number of snapshots (from the perspective of sensing). Letting $\boldsymbol{x}_l \in \mathbb{C}^N$ denote the transmitted N-dimensional signal vector at time stamp $l \in [L]$, we have $\boldsymbol{X} := [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_l, \ldots, \boldsymbol{x}_L]$.

On the communication side, $\boldsymbol{X}$ is the output of pre-coding and conveys the information of communication symbols. For downlink communication with K users, the base-band communication model is

$$\mathbf{Y} = \boldsymbol{H}\boldsymbol{X} + \mathbf{W}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{C}^{K \times L}$ denotes the received symbol matrix at K users, $\boldsymbol{H} := [\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_K]^{\mathsf{T}} \in \mathbb{C}^{K \times N}$ the communication channel matrix, and $\mathbf{W} := [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_L] \in \mathbb{C}^{K \times L}$ the channel noise; for every $k \in [K]$, $\boldsymbol{h}_k$ denotes the channel model of the kth user; for every $l \in [L]$, $\mathbf{w}_l \sim \mathcal{CN}(\mathbf{0}, N_0 \boldsymbol{I}_K)$ denotes the channel noise at the lth data unit and N_0 denotes the noise level. Note that given a waveform matrix $\boldsymbol{X}$ and a channel $\boldsymbol{H}$, the randomness of $\mathbf{Y}$ is from the randomness of $\mathbf{W}$. Suppose that the expected constellation-point matrix for the K downlink users is $\boldsymbol{S} \in \mathbb{C}^{K \times L}$. We have the multi-user interference (MUI) signals $\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}$ and the total MUI energy can be calculated as $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$; see [20], [28]. Therefore, from the perspective of communication, waveform design is to find a mapping $\mathcal{F}_{\boldsymbol{H}} : \boldsymbol{S} \to \boldsymbol{X}$ such that the total MUI energy $\|\boldsymbol{H} \cdot \mathcal{F}_{\boldsymbol{H}}(\boldsymbol{S}) - \boldsymbol{S}\|_F^2$ is minimized, where $\mathcal{F}_{\boldsymbol{H}}$ depends on the communication channel information $\boldsymbol{H}$. That is, the **perfect communication** problem can be stated as

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{X} \text{ satisfies some constraints.} \end{aligned} \tag{2}$$

The constraint can be, for example, to limit the total transmit power, i.e., $\|\boldsymbol{X}\|_F^2 / L \le P_{\mathsf{T}}$, where P_{T} denotes the maximum allowed total transmit power. Note that the smaller the MUI energy, the higher the communication performance. To be specific, for a communication system, the average achievable sum-rate (AASR; unit: bps/Hz/user) $R_{\boldsymbol{H},\boldsymbol{X}}$ is defined as [20]

$$R_{\boldsymbol{H},\boldsymbol{X}} := \frac{1}{K} \sum_{k=1}^{K} \log_2 \left(1 + \gamma_{\boldsymbol{H},\boldsymbol{X},k}\right) \tag{3}$$

and the average signal-to-interference-plus-noise ratio (SINR) $\gamma_{\boldsymbol{H},\boldsymbol{X},k}$ for the kth user is defined as

$$\gamma_{\boldsymbol{H},\boldsymbol{X},k} := \frac{\sum_{l=1}^{L} |s_{k,l}|^2 / L}{\sum_{l=1}^{L} |\boldsymbol{h}_k^{\mathsf{T}} \boldsymbol{x}_l - s_{k,l}|^2 / L + N_0}, \tag{4}$$

where $s_{k,l}$ is the (k,l)-entry of the constellation $\boldsymbol{S}$. As a result, minimizing MUI energy $\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2$ in (2) leads to the improvement of AASR in (3) because for a power-fixed constellation $\boldsymbol{S}$, the numerator in (4) is fixed.

Remark 2: The extensions of (2) to the case of multi-antenna users and the case of multi-carrier are technically trivial. For details, see Appendix A in supplementary materials. □

On the sensing side, the key is to design perfect beam patterns for target searching and/or tracking. According to [29], designing a beam pattern is equivalent to designing the cross-correlation matrix $\boldsymbol{R} \in \mathbb{C}^{N \times N}$ of the probing signals $\{\boldsymbol{x}_l\}_{l \in [L]}$. Therefore, from the perspective of sensing, waveform design is to find a mapping $\mathcal{F} : \boldsymbol{R} \to \boldsymbol{X}$ such that $\boldsymbol{XX}^{\mathsf{H}}/L = \mathcal{F}(\boldsymbol{R})\mathcal{F}^{\mathsf{H}}(\boldsymbol{R})/L = \boldsymbol{R}$. For an omnidirectional beam pattern, which is suitable for target searching, $\boldsymbol{R} := \frac{P_{\mathsf{T}}}{N}\boldsymbol{I}_N$ should be used. Therefore, the **perfect sensing** problem can be stated as

$$
\begin{aligned}
\text{find} \quad & \boldsymbol{X} \\
\text{s.t.} \quad & \boldsymbol{XX}^{\mathsf{H}} = L\boldsymbol{R} \\
& \boldsymbol{X} \text{ satisfies some constraints.}
\end{aligned} \tag{5}
$$

The constraint can be, for example, to limit the peak-to-average-power ratio (PAPR) [30, Eq. (18)].

For an ISAC system, **dual-functional waveform design** can be stated as finding a mapping $\mathcal{F}_{\boldsymbol{H}} : (\boldsymbol{S}, \boldsymbol{R}) \to \boldsymbol{X}$ such that the total MUI energy $\|\boldsymbol{H} \cdot \mathcal{F}_{\boldsymbol{H}}(\boldsymbol{S}, \boldsymbol{R}) - \boldsymbol{S}\|_F^2$ is minimized and the cross-correlation matrix $\mathcal{F}_{\boldsymbol{H}}(\boldsymbol{S}, \boldsymbol{R})\mathcal{F}_{\boldsymbol{H}}^{\mathsf{H}}(\boldsymbol{S}, \boldsymbol{R})/L$ is as close as possible to the desired $\boldsymbol{R}$. In [20, Eq. (10)], the following **sensing-centric** ISAC waveform design problem is defined:

$$
\begin{aligned}
\min_{\boldsymbol{X}} \quad & \|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 \\
\text{s.t.} \quad & \frac{1}{L}\boldsymbol{XX}^{\mathsf{H}} = \boldsymbol{R},
\end{aligned} \tag{6}
$$

which aims to minimize the communication MUI energy while guaranteeing the perfect sensing performance. If we desire a balanced performance between communication and sensing, we can consider a **joint design** problem:

$$
\begin{aligned}
\min_{\boldsymbol{X}} \quad & \rho\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{XX}^{\mathsf{H}} - L\boldsymbol{R}\|_F^2 \\
\text{s.t.} \quad & \boldsymbol{X} \text{ satisfies some constraints,}
\end{aligned} \tag{7}
$$

where $\rho \in [0, 1]$ is the trade-off coefficient. Problem (7) is technically difficult to solve because the objective function is quartic in terms of the waveform matrix $\boldsymbol{X}$. Therefore, we need to formulate simplifications. In [20], the alternative objective function $\rho\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2$ is intensively considered, which relaxes (7) to

$$
\begin{aligned}
\min_{\boldsymbol{X}} \quad & \rho\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2 \\
\text{s.t.} \quad & \boldsymbol{X} \text{ satisfies some constraints,}
\end{aligned} \tag{8}
$$

where $\boldsymbol{X}_s$ is a desired sensing waveform, for example, the waveform in (5) for perfect sensing regardless of communications. Compared with (6), the hard constraint $\boldsymbol{XX}^{\mathsf{H}} = L\boldsymbol{R}$ for perfect sensing is relaxed to penalizing the mismatch $\boldsymbol{XX}^{\mathsf{H}} - L\boldsymbol{R}$ in (7) and is further relaxed to penalizing the discrepancy $\boldsymbol{X} - \boldsymbol{X}_s$ in (8) due to the technical tractability in developing the solution method, where $\boldsymbol{X}_s\boldsymbol{X}_s^{\mathsf{H}} = L\boldsymbol{R}$. Note that the objective of (8) is quadratic (not quartic) in $\boldsymbol{X}$.

B. Robust Waveform Design

1) Robust Counterpart of Sensing-Centric Design (6)*:* Considering the modeling errors in the communication channel $\boldsymbol{H}$, the robust counterpart of (6) can be formulated as

$$
\begin{aligned}
\min_{\boldsymbol{X}} \max_{\boldsymbol{H}} \quad & \|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 \\
\text{s.t.} \quad & \frac{1}{L}\boldsymbol{XX}^{\mathsf{H}} = \boldsymbol{R} \\
& \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \leq \theta,
\end{aligned} \tag{9}
$$

where $\bar{\boldsymbol{H}}$ is a nominal value (e.g., an estimate) of the true channel matrix, $\|\cdot\|$ denotes any suitable matrix norm (e.g., the Frobenius norm $\|\cdot\|_F$), and $\theta \geq 0$ is our trust level of the nominal $\bar{\boldsymbol{H}}$; the smaller the θ, the more we trust $\bar{\boldsymbol{H}}$. Problem (9) is non-convex in $\boldsymbol{X}$ and non-concave in $\boldsymbol{H}$.

2) Robust Counterpart of Joint Design (8)*:* Similarly, the robust counterpart of (8) is

$$
\begin{aligned}
\min_{\boldsymbol{X}} \max_{\boldsymbol{H}} \quad & \rho\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2 \\
\text{s.t.} \quad & \boldsymbol{X} \text{ satisfies some constraints} \\
& \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \leq \theta,
\end{aligned} \tag{10}
$$

where the constraints of $\boldsymbol{X}$ can specifically be

$$
\frac{1}{L}\|\boldsymbol{X}\|_F^2 = P_{\mathsf{T}}, \tag{11}
$$

for the total power constraint (TPC), or

$$
\text{diag}(\boldsymbol{XX}^{\mathsf{H}}) = \frac{L \cdot P_{\mathsf{T}}}{N}\boldsymbol{I}_N, \tag{12}
$$

for the per-antenna power constraint (PAPC). Problem (10), particularized by either (11) or (12), is non-convex in $\boldsymbol{X}$ and non-concave in $\boldsymbol{H}$.

3) Remarks on Robust Counterparts: From the viewpoint of optimization theory, the min-max formulations (9) and (10) are new and technically challenging in their own right. In the authors' knowledge, no literature working on similar optimization problems can be found. The only related research is [31], which somehow has a subtle connection with (9), however, did not solve the same problem.

Suppose that the communication performance of the investigated ISAC system is measured by AASR. To obtain the lower bound of AASR in Fig. 2(b), we need to find the upper bound of the MUI energy, which justifies why we maximize the MUI energy $\|\boldsymbol{HX} - \boldsymbol{S}\|_F^2$ in (9) and (10) over the admissible channels $\boldsymbol{H}$. For detailed analyses on min-max modeling, see Section III.

III. ROBUST WAVEFORM DESIGN: PROPERTY ANALYSES

This section studies the rationale behind the min-max robust formulations (9) and (10). Without loss of generality, we exemplify using (9); the analysis on (10) is similar, and therefore, omitted.

A. Interpretation of Min-Max Formulation (9)

The min-max robust formulation (9) can be seen as a game between a waveform designer selecting the waveform $\boldsymbol{X}$ and a fictitious adversary selecting the channel state $\boldsymbol{H}$; the adversary tries to deteriorate the communication performance

by lifting the MUI energy $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$, while the waveform designer tries to improve the communication performance by diminishing the MUI energy. Therefore, min-max robust waveforms perform best under worst-possible channel states.

Let the true communication channel be $\boldsymbol{H}_0$. According to (6), the problem that we really want to solve should be

$$\min_{\boldsymbol{X}:\ \boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} = L\boldsymbol{R}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \tag{13}$$

which, however, cannot be conducted in practice because $\boldsymbol{H}_0$ is unknown. Nevertheless, we can assume that $\boldsymbol{H}_0$ is included in an **uncertainty set**

$$\mathcal{H} := \{\boldsymbol{H}:\ \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \le \theta\} \tag{14}$$

for some matrix norm $\|\cdot\|$ where $\bar{\boldsymbol{H}}$ is an estimate of $\boldsymbol{H}_0$ and θ denotes the trust level; we suppress the dependence of $\mathcal{H}$ on $(\bar{\boldsymbol{H}}, \theta, \|\cdot\|)$ to avoid notational clutter. This assumption is practically reasonable because the estimate $\bar{\boldsymbol{H}}$ of the ground truth $\boldsymbol{H}_0$ can be close to $\boldsymbol{H}_0$. In addition, we denote $\mathcal{X}$ the collection of all waveforms for perfect sensing, i.e.,

$$\mathcal{X} := \{\boldsymbol{X}: \boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} = L\boldsymbol{R}\}. \tag{15}$$

Facing the unavailability of $\boldsymbol{H}_0$, the pragmatic strategy is to find a practically available upper bound $\mathcal{UB}(\boldsymbol{X})$ for the true cost function $\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$ for every $\boldsymbol{X} \in \mathcal{X}$, i.e.,

$$\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \le \mathcal{UB}(\boldsymbol{X}), \quad \forall \boldsymbol{X} \in \mathcal{X}. \tag{16}$$

As a result, when the upper bound $\mathcal{UB}(\boldsymbol{X})$ is minimized by some $\boldsymbol{X}^*$, the **true cost** of the waveform $\boldsymbol{X}^*$ evaluated at the true channel state $\boldsymbol{H}_0$, i.e.,

$$\|\boldsymbol{H}_0\boldsymbol{X}^* - \boldsymbol{S}\|_F^2,$$

which defines the true communication performance in practical operation, is also reduced. In consideration of the uncertainty set $\mathcal{H}$, $\mathcal{UB}(\boldsymbol{X})$ can be constructed as

$$\mathcal{UB}(\boldsymbol{X}) := \max_{\boldsymbol{H} \in \mathcal{H}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2, \quad \forall \boldsymbol{X} \in \mathcal{X}. \tag{17}$$

The upper bound $\mathcal{UB}(\boldsymbol{X})$ above is tight in the sense that when the uncertainty set $\mathcal{H}$ contains only $\boldsymbol{H}_0$ (i.e., when $\theta = 0$), the equality in (16) holds. In other words, provided that $\boldsymbol{H}_0 \in \mathcal{H}$, the more accurate the $\bar{\boldsymbol{H}}$ (i.e., the smaller the θ), the tighter the upper bound (17).

The fact below explains the rationale of the robust waveform design model (9).

Fact 1 (Achievability of Robust Estimate): By minimizing the practically-available upper bound $\max_{\boldsymbol{H} \in \mathcal{H}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$, we can also upper bound the **truly optimal cost**

$$\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \tag{18}$$

and reduce the **true cost** of the **robust waveform** $\boldsymbol{X}^*$ evaluated at the true channel $\boldsymbol{H}_0$, i.e.,

$$\|\boldsymbol{H}_0\boldsymbol{X}^* - \boldsymbol{S}\|_F^2 \tag{19}$$

because

$$\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}_0\boldsymbol{X}^* - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2,$$

where

$$(\boldsymbol{X}^*, \boldsymbol{H}^*) \in \operatorname*{argmin}_{\boldsymbol{X} \in \mathcal{X}} \operatorname*{argmax}_{\boldsymbol{H} \in \mathcal{H}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$$

and $\boldsymbol{H}^*$ is the **worst-case channel state** associated with $\boldsymbol{X}^*$. Note that in real-world operation, the true communication performance of a specified waveform $\boldsymbol{X}$ is defined by the true cost $\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$ evaluated at $\boldsymbol{H}_0$; cf. (19). $\square$

In contrast, the nominally optimal waveform(s) cannot provide such a performance guarantee.

Fact 2 (Unachievability of Nominal Estimate): The **nominally optimal cost**

$$\min_{\boldsymbol{X} \in \mathcal{X}} \|\bar{\boldsymbol{H}}\boldsymbol{X} - \boldsymbol{S}\|_F^2 \tag{20}$$

cannot upper bound the **truly optimal cost**

$$\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$$

and reduce the true cost of the **nominally optimal waveform** $\bar{\boldsymbol{X}}$ evaluated at the true channel $\boldsymbol{H}_0$, i.e.,

$$\|\boldsymbol{H}_0\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \tag{21}$$

because

$$\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}_0\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \overset{?}{\le} \|\bar{\boldsymbol{H}}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2,$$

where $\bar{\boldsymbol{X}} \in \operatorname{argmin}_{\boldsymbol{X} \in \mathcal{X}} \|\bar{\boldsymbol{H}}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ and the symbol $\overset{?}{\le}$ means that the relation is not guaranteed. $\square$

Let $\boldsymbol{X}_0 \in \operatorname{argmin}_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$ denote a **truly optimal waveform**. The motivation and the necessity of the robust waveform design are summarized in the remark below.

Remark 3 (Benefit of Robust Waveform Design): In the practice of communications, finding the tight upper bound of the truly optimal cost $\|\boldsymbol{H}_0\boldsymbol{X}_0 - \boldsymbol{S}\|_F^2$ and the true cost $\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$ of the specified waveform $\boldsymbol{X}$ is critical. Suppose that an ISAC system can tolerate at most C MUI energies for communications: i.e., for an employed waveform $\boldsymbol{X}'$, we must have $\|\boldsymbol{H}_0\boldsymbol{X}' - \boldsymbol{S}\|_F^2 \le C$. However, if we use the nominally optimal waveform $\bar{\boldsymbol{X}}$ designed at $\bar{\boldsymbol{H}}$, the designed nominally optimal cost $\|\bar{\boldsymbol{H}}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \le C$ does not necessarily imply the true cost $\|\boldsymbol{H}_0\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \le C$. This may lead to significant dissatisfaction (i.e., bad user experiences) among communication users because the announced system performance C cannot be guaranteed, and therefore, a serious reliability issue in communications arises. In contrast, if we employ the robust waveform design in (9), the dissatisfaction issue can be fixed because the robust cost $\|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2 \le C$ can imply the true cost $\|\boldsymbol{H}_0\boldsymbol{X}^* - \boldsymbol{S}\|_F^2 \le C$; recall Fact 1. $\square$

To further clarify the benefit of robust design, in addition to Remark 3, the remark below exemplifies the average achievable sum-rate defined in (3).

Remark 4: Given a communication system (1), the performance is typically characterized using the AASR $R_{\boldsymbol{H},\boldsymbol{X}}$, which is specific to the channel state $\boldsymbol{H}$ and the waveform $\boldsymbol{X}$. When the true channel state $\boldsymbol{H}_0$ is exactly known, the communication system design is to find the optimal waveform $\boldsymbol{X}_0$ such that

$$R_{\boldsymbol{H}_0,\boldsymbol{X}} \le R_{\boldsymbol{H}_0,\boldsymbol{X}_0}, \quad \forall \boldsymbol{X} \in \mathcal{X}.$$

As a result, the truly optimal AASR R_{H_0, X_0} is the performance frontier of this communication system; cf. Fig. 1. In practice, H_0 is unknown, and the nominal channel $\bar{H}$ acts as a surrogate of H_0. However, the nominally optimal AASR $R_{\bar{H}, \bar{X}}$ cannot serve as a performance indicator of this communication system because

$$R_{\bar{H}, \bar{X}} \overset{?}{\leq} R_{H_0, \bar{X}} \leq R_{H_0, X_0};$$

cf. Fig. 2(a). In contrast, when the robust waveform X^* is utilized, we have

$$R_{H^*, X^*} \leq R_{H_0, X^*} \leq R_{H_0, X_0},$$

which means that R_{H^*, X^*} defines the worst-case performance bound of this communication system: that is, in real-world operations, the system's true running performance is no worse than R_{H^*, X^*}. Hence, R_{H^*, X^*} characterizes the robust or conservative performance frontier in terms of communication; cf. Fig. 2(b). $\square$

B. Price of Robustness

As we can see from Fact 1, the robust design specifies an upper bound for the truly optimal cost and the true cost. In practice, however, this upper bound may be overly loose, i.e.,

$$\min_{X \in \mathcal{X}} \|H_0 X - S\|_F^2 \leq \|H_0 X^* - S\|_F^2 \ll \|H^* X^* - S\|_F^2.$$

To be specific, the robust waveform X^* that solves the min-max robust problem may induce a loose estimate of the truly optimal cost and the true cost, and therefore, the optimality of the communication performance might be compromised. This is because the worst-case channel(s) does not necessarily frequently occur in real-world operations. As a result, the robustness-optimality trade-off is raised: to obtain robustness under uncertain conditions (i.e., when H_0 is not exactly known), the price to pay is to sacrifice optimality under perfect conditions (i.e., when H_0 is perfectly known).

C. Budget of Uncertainty Set

To reduce the conservativeness of the robust design, a practical trick is to limit the size of the uncertainty set $\mathcal{H}$. To be specific, we employ a **shrunken uncertainty set** $\mathcal{H}_\beta := \{H : \|H - \bar{H}\| \leq \beta\theta\}$ where $\beta \in [0, 1]$ is called the **budget of the uncertainty set**. Since

$$\min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}_\beta} \|HX - S\|_F^2 \leq \min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}} \|HX - S\|_F^2,$$

a practically tighter upper bound (i.e., the left-hand side of the above display) can be suggested, which however **cannot** theoretically serve as an upper bound for the truly optimal cost and the true cost. The motivation is that in practice, H_0 is still included in the shrunken set $\mathcal{H}_\beta$ with high probability. Nevertheless, the price is to meet the worst-case: H_0 might be outside of $\mathcal{H}_\beta$. The fact below practically lifts Fact 1.

Fact 3 (Achievability in Probability): Suppose that the true channel H_0 is included in $\mathcal{H}_\beta$ with probability η (e.g., $\eta = 95\%$). Then the robust cost $\min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}_\beta} \|HX - S\|_F^2$ upper bounds the truly optimal cost and the true cost with probability η; cf. Fact 1. $\square$

IV. Robust Waveform Design: Solution Methods

This section studies the solution methods to the robust counterparts (9) and (10). We start with the following preparatory and motivational lemma.

Lemma 1: Let $\mathcal{X}$ define a generic feasible region of waveforms. Suppose that $\phi : \mathcal{H} \times \mathcal{X} \to \mathbb{R}$ is the performance objective function of an ISAC system, e.g., as in (9) and (10). For every fixed point $X^* \in \mathcal{X}$, we have

$$\begin{aligned} \max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X) & \leq \min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}} \phi(H, X) \\ & \leq \max_{H \in \mathcal{H}} \phi(H, X^*). \end{aligned} \quad (22)$$

In addition, if there exists an $X^* \in \mathcal{X}$ such that

$$\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X) = \max_{H \in \mathcal{H}} \phi(H, X^*), \quad (23)$$

then the strong min-max property holds for the min-max game on $(\phi, \mathcal{H}, \mathcal{X})$: i.e.,

$$\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X) = \min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}} \phi(H, X). \quad (24)$$

Proof: See Appendix B in supplementary materials. $\square$

In the context of robustness analysis, the max-min optimization $\max_H \min_X \phi(H, X)$ can be easier to solve than the original min-max robust counterpart $\min_X \max_H \phi(H, X)$ because for every fixed channel H, the solution to the waveform design problem $\min_X \phi(H, X)$ is usually available in literature. Therefore, if the strong mix-max property (24) holds, the robustification problem $\min_X \max_H \phi(H, X)$ can be simplified to the problem of finding the worst-case channel through $\max_{H \in \mathcal{H}} \phi(H, X_H^*)$ where X_H^* solves $\min_X \phi(H, X)$ for every H.

In practice, the condition (23) might be harsh to satisfy. A compromise, however, can be made if the gap between $\max_H \phi(H, X^*)$ and $\max_H \min_X \phi(H, X)$ in (22) is small. The lemma below characterizes this gap in a special case.

Lemma 2: Suppose that $\bar{X}$ solves the nominal problem $\min_{X \in \mathcal{X}} \phi(\bar{H}, X)$. If the function

$$H \mapsto \min_{X \in \mathcal{X}} \phi(H, X), \quad \forall H \in \mathcal{H} \quad (25)$$

is L_1-Lipschitz continuous in H on $\mathcal{H}$ and the function

$$H \mapsto \phi(H, \bar{X}), \quad \forall H \in \mathcal{H} \quad (26)$$

is L_2-Lipschitz continuous in H on $\mathcal{H}$, both with respect to the norm $\|\cdot\|$ used in defining $\mathcal{H}$ in (14), then

$$\max_{H \in \mathcal{H}} \phi(H, \bar{X}) - \max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X) \leq (L_1 + L_2) \cdot \theta.$$

Proof: See Appendix C in supplementary materials. $\square$

Lemma 2 suggests that, if the functions in (25) and (26) are Lipschitz continuous, and the radius θ of $\mathcal{H}$ is small, then the gap between $\max_H \phi(H, \bar{X})$ and $\max_H \min_X \phi(H, X)$ would be naturally small. Lemma 2 further implies that if we let $X^* := \bar{X}$ in (22), then the extent of breaching the equality in (23) can be inherently slight. The following insight is useful for designing approximate solution methods to original min-max problems.

Insight 1 (Solve Max-Min Counterpart): When employing the max-min counterpart $\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X)$ for designing the approximate solution method to the original

min-max problem $\min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}} \phi(H, X)$, the key is to choose an appropriate X^* such that the discrepancy between $\max_{H \in \mathcal{H}} \phi(H, X^*)$ and $\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X)$ can be minimized to its utmost extent. □

In what follows, based on Lemmas 1 and 2 and Insight 1, we discuss the solution methods to the robust counterparts (9) and (10), respectively in Subsections IV-A and IV-B. In Subsection IV-C, the proposed methods and their computational complexities are summarized; see Tables I and II.

A. Solution Method to Robust Counterpart (9)

This subsection designs the solution method to the robust counterpart (9) of the sensing-centric waveform design problem (6).

1) Model Reformulation: Since Problem (9) is non-convex in X and non-concave in H, conventional solution methods (e.g., alternating descent, Lagrangian duality) to min-max problems are not applicable.

Equipped with Lemma 1, the theorem below transforms the robust counterpart (9) into a tractable equivalent.

Theorem 1: Let $\mathcal{X} := \{X : XX^{\mathsf{H}} = LR\}$. Suppose that the beampattern-inducing matrix R is positive definite and F is invertible such that $R = FF^{\mathsf{H}}$ (e.g., Cholesky decomposition).[1] If there exists $X^* \in \mathcal{X}$ such that

$$\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \|HX - S\|_F^2 = \max_{H \in \mathcal{H}} \|HX^* - S\|_F^2, \quad (27)$$

then Problem (9) is equivalent to

$$\begin{aligned} \max_{H} \quad & \|\sqrt{L} \cdot H \cdot F \cdot U I_{N \times L} V^{\mathsf{H}} - S\|_F^2 \\ \text{s.t.} \quad & \|H - \bar{H}\| \leq \theta, \\ & U \Sigma V^{\mathsf{H}} \overset{\text{SVD}}{=} F^{\mathsf{H}} H^{\mathsf{H}} S, \end{aligned} \quad (28)$$

where $U \Sigma V^{\mathsf{H}} \overset{\text{SVD}}{=} F^{\mathsf{H}} H^{\mathsf{H}} S$ means the singular value decomposition (SVD) of $F^{\mathsf{H}} H^{\mathsf{H}} S$ and $I_{N \times L} := [I_N, \mathbf{0}_{N \times (L-N)}]$; the $N \times (L-N)$ zero matrix is denoted by $\mathbf{0}_{N \times (L-N)}$.

Proof: See Appendix D in supplementary materials. □

Theorem 1 implies that the original two-layer min-max problem (9) can be conditionally equivalently solved by the single-layer maximization problem (28). The equivalence stems from the condition (27) and the existence of the closed-form solution X_H^* for every given H [20, Eq. (15)]:

$$X_H^* = \sqrt{L} F U I_{N \times L} V^{\mathsf{H}}, \quad \forall H \in \mathcal{H}. \quad (29)$$

Condition (27), which is motivated by Lemma 1, enforces the strong min-max property between $\min_{X \in \mathcal{X}} \max_{H \in \mathcal{H}} \|HX - S\|_F^2$ and $\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \|HX - S\|_F^2$. Since directly attacking the original problem (9) is technically challenging, this paper solves the reformulated problem (28), wherein the condition (27) needs to be stringently satisfied to ensure the equivalence between (9) and (28).

Although the condition (27) is particularized from the general case (23), it is still technically difficult to verify the existence of X^* in (27). Therefore, approximate solutions to the robust counterpart (9) can be obtained if the strict equality in the condition (27) can be compromised; cf. Insight

1. To be specific, we aim to seek a solution X^* to (28) such that $\max_{H \in \mathcal{H}} \|HX^* - S\|_F^2$ can approach or equal its lower bound $\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \|HX - S\|_F^2$ from above; note that the closer the right-hand side of (27) is to the left-hand side, the smaller the gap between (9) and (28); cf. Lemma 1.

We start with examining the properties of the maximization problem (28).

2) Property Analysis of Maximization Problem (28): Denote the objective function in Problem (28) by

$$f(H) := \|\sqrt{L} \cdot H \cdot F \cdot U I_{N \times L} V^{\mathsf{H}} - S\|_F^2, \quad (30)$$

which is a particularization of (25). The proposition below establishes the Lipschitz continuity of the function f on $\mathcal{H}$.

Proposition 1: The function f is L_f-Lipschitz continuous in H on $\mathcal{H}$ with respect to the norm $\|\cdot\|$ used in defining $\mathcal{H}$, where

$$L_f = 2BLP_{\mathsf{T}} \cdot (\|\bar{H}\|_F + B\theta) + 2B\sqrt{LP_{\mathsf{T}}} \cdot \|S\|_F \quad (31)$$

and B is a finite positive real number such that $\|H\|_F \leq B\|H\|$ for every $H \in \mathcal{H}$.

Proof: See Appendix E in supplementary materials. □

It is well believed that a Lipschitz continuous objective function is much easier to be globally maximized [32, p. 7]. An extreme case is that the Lipschitz constant is zero on $\mathcal{H}$ so f is constant on $\mathcal{H}$. In this case, we just need to evaluate only one point on $\mathcal{H}$ to globally maximize f. In addition, the Lipschitz continuity of f is also important to control the extent of breaching the equality in (27); recall Lemma 2. Since f is Lipschitz continuous, it is almost everywhere differentiable. However, f is non-convex and non-concave in H on $\mathcal{H}$.

Proposition 2: The function f is neither convex nor concave in H on $\mathcal{H}$.

Proof: See Appendix F in supplementary materials. □

Nevertheless, the objective function $f(H)$ has tight upper bound $\overline{f}(H)$ and lower bound $\underline{f}(H)$. In addition, $\overline{f}(H)$ is positive-definite quadratic (thus convex).

Proposition 3: The function $f(H)$ is upper bounded by

$$\overline{f}(H) := \|H\bar{X} - S\|_F^2$$

and lower bounded by

$$\underline{f}(H) := \left[\sqrt{L \operatorname{Tr}[H^{\mathsf{H}} H R]} - \|S\|_F \right]^2$$

for all possible $H \in \mathbb{C}^{K \times N}$, not necessarily on $\mathcal{H}$, where

$$\bar{X} := \sqrt{L} F \bar{U} I_{N \times L} \bar{V}^{\mathsf{H}},$$

$\bar{U} \bar{\Sigma} \bar{V}^{\mathsf{H}} \overset{\text{SVD}}{=} F^{\mathsf{H}} \bar{H}^{\mathsf{H}} S$, and $\bar{H}$ is the center of $\mathcal{H}$. In addition, the following are true.

1) The upper bound $\overline{f}(H)$ is positive-definite quadratic (thus convex) in $\operatorname{vec}(H)$.
2) The upper bound $\overline{f}(H)$ and the lower bound $\underline{f}(H)$ are both tight in the sense that the two bounds can be reached for some H.
3) The upper bound $\overline{f}(H)$ is L_f-Lipschitz continuous.
4) The difference between $\overline{f}(H)$ and $\underline{f}(H)$ is uniformly bounded by $2L_f \cdot \theta$ on $\mathcal{H}$, i.e.,

$$\overline{f}(H) - \underline{f}(H) \leq 2L_f \cdot \theta.$$

---

[1]Note that given R, there may exist multiple F.

As a consequence, the positive-definite quadratic function $\overline{f}(\boldsymbol{H}) - 2L_f \cdot \theta$ is also a convex lower bound of $f(\boldsymbol{H})$ on $\mathcal{H}$: i.e., $\underline{f}(\boldsymbol{H}) := \overline{f}(\boldsymbol{H}) - 2L_f \cdot \theta$ is also a possible choice.

Proof: See Appendix G in supplementary materials. □

Note that $\overline{f}(\boldsymbol{H})$ is a particularization of (26), due to which the Lipschitz continuity is crucial. Since the Lipschitz continuity constant in (31) is rather loose, the upper bound $2L_f \cdot \theta$ of the difference between $\overline{f}(\boldsymbol{H})$ and $f(\boldsymbol{H})$ can be much smaller in practice.

Propositions 2 and 3 suggest that the landscape of the objective function $f(\boldsymbol{H})$ is "globally positive-definite quadratic" but "locally non-convex and non-concave". An illustration is shown in Fig. 3, in whose caption the implications from Propositions 2 and 3 are clarified.
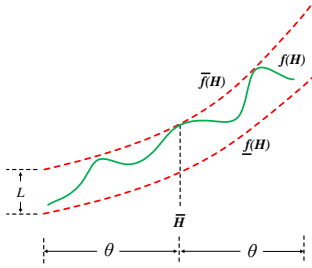

Fig. 3. An one-dimensional illustration of the landscape of the objective function $f(\boldsymbol{H})$ on $\mathcal{H}$. The upper bound and the lower bound of f are $\overline{f}$ and $\underline{f}$, respectively; $\underline{f} := \overline{f} - L$, where L is the uniform gap between $\overline{f}$ and $\underline{f}$. Therefore, the upper bound $\overline{f}$ and the lower bound $\underline{f}$ are both positive-definite quadratic and thus convex. Note that $L \leq 2L_f \cdot \theta$. The center and the scale of the feasible region $\mathcal{H}$ are $\overline{\boldsymbol{H}}$ and θ, respectively. At the center $\overline{\boldsymbol{H}}$, $f(\overline{\boldsymbol{H}})$ reaches its upper bound $\overline{f}(\overline{\boldsymbol{H}})$. Since the three functions f, $\overline{f}$, and $\underline{f}$ share the same Lipschitz continuity constant, the objective function f cannot be overly rugged (i.e., it is relatively flat) in between the two bounds $\overline{f}$ and $\underline{f}$. Note that f is almost everywhere differentiable on $\mathcal{H}$.

In consideration of the landscape and other properties (i.e., non-convexity, non-concavity, Lipschitz continuity, etc.) of the objective function $f(\boldsymbol{H})$ in Problem (28), the zero-order optimization methods such as heuristic optimization [33]–[36], Bayesian optimization [37]–[39], etc., are the last possible choices. However, a zero-order method is globally optimal for continuous objective functions if and only if the evaluation points governed by the searching algorithm are dense in the feasible region [40]. Nevertheless, the searching space of Problem (28) is extremely large; when the number K of communication users and the number N of transmit antennas are large, the dimension of $\boldsymbol{H}$ will be large because $\boldsymbol{H} \in \mathbb{C}^{K \times N}$. Recall that for large-scale multiple-input multiple-output (MIMO) communication systems, the antenna number N is expected to be large [41]. To be specific, there are in total $2KN$ real numbers involved. Even if we only query I points in each dimension, we must evaluate the function $f(\boldsymbol{H})$ at least I^{2KN} times to ensure global optimality for Problem (28). This is a well-identified dilemma known as "the curse of dimensionality" in global optimization; for rigorous complexity results in numerical accuracy and numerical computation, see, e.g., [32, p. 10], [42], [43, Table 1].

In summary, Problem (28) has challenging properties as a global optimization problem: large dimensionality, non-

convexity, and non-concavity. Therefore, an ad-hoc and efficient algorithm for approximately solving Problem (28) is to be designed.

3) Approximate Solution Method to (9)*:* This subsection studies an approximate solution method to the robust counterpart (9) by leveraging the maximization problem (28). The approximation arises from sacrificing the strict equality in the condition (27).

By introducing auxiliary variables $\boldsymbol{A}$, $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$, we consider an optimization equivalent of Problem (28):

$$
\begin{aligned}
\max_{\boldsymbol{H},\boldsymbol{A},\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}} \quad & \|\sqrt{L}\boldsymbol{H}\boldsymbol{F}\boldsymbol{A} - \boldsymbol{S}\|_F^2 \\
\text{s.t.} \quad & \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \leq \theta, \\
& \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{S}, \\
& \boldsymbol{A} = \boldsymbol{U}\boldsymbol{I}_{N \times L}\boldsymbol{V}^{\mathsf{H}}, \ \boldsymbol{U}\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{I}_N, \ \boldsymbol{V}\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{I}_L, \\
& \boldsymbol{\Sigma} \text{ is diagonal, non-negative, and real;}
\end{aligned}
$$
(32)

note from Theorem 1 that for every given $\boldsymbol{H}$, the corresponding optimal waveform is

$$
\boldsymbol{X}_{\boldsymbol{H}}^* = \sqrt{L}\boldsymbol{F}\boldsymbol{A}. \tag{33}
$$

In terms of the variables $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$, the transformed problem (32) is just a feasibility problem because the three variables are not involved in the objective function. Since the rank of the matrix $\boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{S}$ is no larger than K where $K \leq N \leq L$, the SVD of $\boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{S}$ is not unique for every feasible $\boldsymbol{H}$; to be specific, the feasible values of $\boldsymbol{U}$ and $\boldsymbol{V}$ are not unique. As a result, the feasible value of $\boldsymbol{A}$ is also not unique given $\boldsymbol{H}$.

In terms of the variable $\boldsymbol{H}$, the vectorized optimization equivalent of (32) is

$$
\begin{aligned}
\max_{\boldsymbol{H}} \max_{\boldsymbol{A},\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}} \quad & \|\sqrt{L}((\boldsymbol{F}\boldsymbol{A})^{\mathsf{T}} \otimes \boldsymbol{I}_K)\operatorname{vec}(\boldsymbol{H}) - \operatorname{vec}(\boldsymbol{S})\|_2^2 \\
\text{s.t.} \quad & (\boldsymbol{F}^{\mathsf{T}} \otimes \boldsymbol{S}^{\mathsf{H}})\operatorname{vec}(\boldsymbol{H}) = \operatorname{vec}(\boldsymbol{V}\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{U}^{\mathsf{H}}), \\
& \|\operatorname{vec}(\boldsymbol{H}) - \operatorname{vec}(\bar{\boldsymbol{H}})\| \leq \theta, \\
& \boldsymbol{A} = \boldsymbol{U}\boldsymbol{I}_{N \times L}\boldsymbol{V}^{\mathsf{H}}, \\
& \boldsymbol{U}\boldsymbol{U}^{\mathsf{H}} = \boldsymbol{I}_N, \ \boldsymbol{V}\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{I}_L, \\
& \boldsymbol{\Sigma} \text{ is diagonal, non-negative, and real.}
\end{aligned}
$$
(34)

The key properties of the problem (34) are given below.

Proposition 4: Consider the reformulated problem (34). The following is true.

1) The objective function is positive-definite quadratic in $\operatorname{vec}(\boldsymbol{H})$.
2) In terms of both $\boldsymbol{H}$ and $\boldsymbol{A}$, the objective function is convex, and the constraints are also convex.[2]

Proof: See Appendix H in supplementary materials. □

To limit the technical complexity of, and obtain an elegant solution to, Problem (34), we impose an additional constraint

$$
\boldsymbol{U}\boldsymbol{I}_{N \times L}\boldsymbol{V}^{\mathsf{H}} = \bar{\boldsymbol{U}}\boldsymbol{I}_{N \times L}\bar{\boldsymbol{V}}^{\mathsf{H}} \tag{35}
$$

to limit the size and shape of the feasible region of $\boldsymbol{H}$, and thus, simplify the problem. Intuitively speaking, this proposal

---

[2]However, note that the optimization problem (34) is not convex in $\boldsymbol{H}$ and $\boldsymbol{A}$ because it is a maximization problem.

is to let the nominally optimal solution $\bar{X}$ **simultaneously** solve the robust problem, that is,

$$\max_{H} \min_{X} \|HX - S\|_F^2 = \max_{H} \|H\bar{X} - S\|_F^2, \qquad (36)$$

which adheres to the condition (27); cf. Lemma 2. Suppose that H^* solves (28) (or equivalently (32) and (34)) and X^* is associated with H^* through (29); that is

$$(H^*, X^*) \in \operatorname*{argmax}_{H \in \mathcal{H}} \operatorname*{argmin}_{X \in \mathcal{X}} \|HX - S\|_F^2.$$

By employing the strategies (35) and (36), the relation of

$$X^* = \bar{X} \qquad (37)$$

can be guaranteed where $\bar{X} = \sqrt{L}F\bar{A} = \sqrt{L}F\bar{U}I_{N \times L}\bar{V}^{\mathsf{H}}$. We do not directly consider the constraint $\sqrt{L}FUI_{N \times L}V^{\mathsf{H}} = \bar{X}$ (i.e., $X = \bar{X}$) because F is invertible, and therefore, the two ways are equivalent.

The method below formalizes the above reasoning, which approximately solves (28) and compromises the exact equality constraint in (27).

Method 1: Suppose a SVD of $F^{\mathsf{H}}\bar{H}^{\mathsf{H}}S$ is $\bar{U}\bar{\Sigma}\bar{V}^{\mathsf{H}} \stackrel{\text{SVD}}{=} F^{\mathsf{H}}\bar{H}^{\mathsf{H}}S$. Let $\bar{A} := \bar{U}I_{N \times L}\bar{V}^{\mathsf{H}}$. If we consider U and V such that $U\Sigma V^{\mathsf{H}} \stackrel{\text{SVD}}{=} F^{\mathsf{H}}H^{\mathsf{H}}S$ and $UI_{N \times L}V^{\mathsf{H}} = \bar{U}I_{N \times L}\bar{V}^{\mathsf{H}}$, then Problem (34) is approximately solved by $(H^*, \bar{A}, U^*, \Sigma^*, V^*)$ where H^* is a global maximum of the upper bound function $\overline{f}(H)$, i.e.,

$$\begin{aligned} H^* &\in \operatorname*{argmax}_{H} \quad \|\sqrt{L}((F\bar{A})^{\mathsf{T}} \otimes I_K)\operatorname{vec}(H) - \operatorname{vec}(S)\|_2^2 \\ &\text{s.t.} \quad \|\operatorname{vec}(H) - \operatorname{vec}(\bar{H})\| \leq \theta, \end{aligned} \tag{38}$$

and (U^*, Σ^*, V^*) is a global minimum of

$$\begin{aligned} \min_{U, \Sigma, V} \quad & \alpha\|U\Sigma V^{\mathsf{H}} - F^{\mathsf{H}}H^{*\mathsf{H}}S\|_F^2 + \|UI_{N \times L}V^{\mathsf{H}} - \bar{A}\|_F^2 \\ \text{s.t.} \quad & UU^{\mathsf{H}} = I_N, \ VV^{\mathsf{H}} = I_L, \\ & \Sigma \text{ is diagonal, non-negative, and real,} \end{aligned} \tag{39}$$

where $\alpha \geq 0$ is a large real number to numerically ensure that $U\Sigma V^{\mathsf{H}}$ is a SVD of $F^{\mathsf{H}}H^{*\mathsf{H}}S$. □

In Method 1, the exact equality constraint in (35), and therefore in (36), is sacrificed; cf. Condition (27) and Insight 1. Specifically, we just require $UI_{N \times L}V^{\mathsf{H}}$ to be as close as possible to $\bar{A}$. This relaxation brings two benefits:

1) First, the relaxed problem (39) can be technically readily solved over U, Σ, and V (see Proposition 8);
2) Second, the values of U and V that strictly satisfy $UI_{N \times L}V^{\mathsf{H}} = \bar{A}$ might not exist for every specified H; some technical regularization conditions are needed to ensure the existence. However, if we use the soft-constraint counterpart as in (39), we do not need to explicitly derive the regularization conditions.

As a result of Method 1, compared to the strict equality in (37), an approximation is achieved:

$$X^* \approx \bar{X}. \qquad (40)$$

For an intuitive interpretation of Method 1, see Appendix I in supplementary materials.

As indicated by Method 1, the key to approximately solving the robust counterpart (9) is to solve the two sub-problems (38) and (39).

Solution to Problem (38): Since Problem (38) is not convex, it is not straightforward to solve. Nevertheless, Problem (38) has nice properties to benefit the design of a globally optimal algorithm: it is positive-definite quadratic in $\operatorname{vec}(H)$ and has convex constraints for $\operatorname{vec}(H)$. We rewrite (38) in real spaces in a compact form:

$$\begin{aligned} \max_{h \in \mathbb{R}^{2KN}} \quad & \|Ch - s\|_2^2, \\ \text{s.t.} \quad & \|h - \bar{h}\| \leq \theta, \end{aligned} \tag{41}$$

where

$$h := \begin{bmatrix} \operatorname{real}(\operatorname{vec}(H)) \\ \operatorname{imag}(\operatorname{vec}(H)) \end{bmatrix} \in \mathbb{R}^{2KN}$$

is a real-valued vector constructed by stacking the real and imaginary components of $\operatorname{vec}(H)$; the quantities

$$C \in \mathbb{R}^{2KL \times 2KN}, \quad s \in \mathbb{R}^{2KL}, \quad \text{and } \bar{h} \in \mathbb{R}^{2KN}$$

are constructed from (38) in a similar way, during which the lemma below is useful.

Lemma 3: Suppose $\Xi := \Gamma + \Theta j$ and $\xi := a + bj$ where j denotes the imaginary unit; Γ and Θ are real matrices and a and b are real vectors. We have

$$\begin{bmatrix} \operatorname{real}(\Xi\xi) \\ \operatorname{imag}(\Xi\xi) \end{bmatrix} = \begin{bmatrix} \Gamma & -\Theta \\ \Theta & \Gamma \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Proof: See Appendix J in supplementary materials. □

Let the objective function of (41) be $p(h)$. We can further show that $p(h)$ is positive-definite and strongly convex.

Lemma 4: The objective function $p(h)$ of (41) is positive definite and strongly convex in h.

Proof: See Appendix K in supplementary materials. □

Equipped with Lemma 4, Proposition 5 below provides the globally optimal solution(s) to Problem (41), and hence, to Problem (38).

Proposition 5 (Solution to Upper-Bound Problem (41)*):* The globally optimal solution to the upper-bound problem (41), and therefore to (38), is given by the following algorithmic iteration process.

Step a) Initialization: Let the iteration count $k = 0$. Choose h_0 from the domain $\{h : \|h - \bar{h}\| \leq \theta\}$ but $h_0 \neq (C^{\mathsf{T}}C)^{-1}C^{\mathsf{T}}s$.

Step b) Solve

$$\begin{aligned} y_k = \operatorname*{argmax}_{y \in \mathbb{R}^{2KN}} \quad & (h_{k-1}^{\mathsf{T}}C^{\mathsf{T}}C - s^{\mathsf{T}}C) \cdot y \\ \text{s.t.} \quad & \|Cy - s\|_2^2 = \|Ch_{k-1} - s\|_2^2. \end{aligned} \tag{42}$$

The constraint can be changed to $\|Cy - s\|_2^2 \leq \|Ch_{k-1} - s\|_2^2$ without losing the optimality.

Step c) Solve

$$\begin{aligned} h_k = \operatorname*{argmax}_{h \in \mathbb{R}^{2KN}} \quad & (y_k^{\mathsf{T}}C^{\mathsf{T}}C - s^{\mathsf{T}}C) \cdot h \\ \text{s.t.} \quad & \|h - \bar{h}\| \leq \theta. \end{aligned} \tag{43}$$

Step d) Repeat Step b) and Step c) until

$$\langle C^{\mathsf{T}}Cy_k - C^{\mathsf{T}}s, \ h_k - y_k \rangle \leq 0 \qquad (44)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in real spaces.

When the iteration process terminates, h_k is a globally optimal solution to Problem (41). In addition, for every k, it holds that $p(h_k) > p(h_{k-1})$; i.e., in every iteration round, the solution h_k is improved.

Proof: See Appendix L in supplementary materials. □

In addition to Proposition 5, [44, Algo. 2] is an alternative method that has been empirically shown to be computationally efficient and almost globally optimal for convex-quadratic maximization. Problem (42) can be analytically solved.

Proposition 6 (Solution to (42)*):* Suppose that $M^\mathsf{T} M$ denotes a square-root decomposition (e.g., Cholesky) of $C^\mathsf{T} C$ and M is invertible. Let

$$\gamma := \sqrt{\|Ch_{k-1} - s\|_2^2 + s^\mathsf{T} C(C^\mathsf{T} C)^{-1} C^\mathsf{T} s - s^\mathsf{T} s}.$$

Then, a maximum of (42) is

$$M^{-1} \cdot \left[\gamma \frac{M^{-\mathsf{T}}(C^\mathsf{T} Ch_{k-1} - C^\mathsf{T} s)}{\|M^{-\mathsf{T}}(C^\mathsf{T} Ch_{k-1} - C^\mathsf{T} s)\|_2} + M^{-\mathsf{T}} C^\mathsf{T} s \right].$$

Proof: See Appendix M in supplementary materials. □

As for (43), when the norm constraint is defined using the infinity norm, it is particularized into a linear program, which can be efficiently solved using the simplex method. In addition, we prove that a closed-form solution exists for (43) if the norm constraint is defined using the 2-norm.

Proposition 7 (Solution to (43)*):* If the norm constraint is defined using the 2-norm, then a maximum of (43) is

$$\theta \frac{C^\mathsf{T} Cy_k - C^\mathsf{T} s}{\|C^\mathsf{T} Cy_k - C^\mathsf{T} s\|_2} + \bar{h}.$$

Proof: See Appendix N in supplementary materials. □

After addressing Problem (38), we solve Problem (39), which is termed a *remedy* problem.

Solution to Problem (39): The solution to Problem (39) is summarized in the proposition below.

Proposition 8 (Solution to Remedy Problem (39)*):* The solution to Problem (39) is given by the following iteration.

Step a) Fix Σ and V, and minimize over U. The optimal solution is $U^* = U_1 V_1^\mathsf{H}$ where $U_1 \Sigma_1 V_1^\mathsf{H} \overset{\text{SVD}}{=} B_1 A_1^\mathsf{H}$, $A_1 := [I_{N \times L} V^\mathsf{H}, \sqrt{\alpha} \Sigma V^\mathsf{H}]$, and $B_1 := [\bar{A}, \sqrt{\alpha} F^\mathsf{H} H^{*\mathsf{H}} S]$.

Step b) Fix Σ and U, and minimize over V. The optimal solution is $V^* = U_2 V_2^\mathsf{H}$ where $U_2 \Sigma_2 V_2^\mathsf{H} \overset{\text{SVD}}{=} B_2 A_2^\mathsf{H}$, $A_2 := [I_{N \times L}^\mathsf{H} U^\mathsf{H}, \sqrt{\alpha} \Sigma^\mathsf{H} U^\mathsf{H}]$, and $B_2 := [\bar{A}^\mathsf{H}, \sqrt{\alpha} S^\mathsf{H} H^* F]$.

Step c) Fix U and V, and minimize over Σ. The optimal solution is $\Sigma^* = \text{Proj}_{\Omega_{N \times L}} [U^\mathsf{H} F^\mathsf{H} H^{*\mathsf{H}} SV]$ where Proj denotes the projection operator and $\Omega_{N \times L}$ denotes the diagonal, non-negative, and real space with the dimension of $N \times L$.

Step d) Repeat Steps a)-c) until convergence.

Regarding the objective value, the iteration process is guaranteed to converge for any initial values of (U, Σ, V).

Proof: See Appendix O in supplementary materials. □

Remark 5 (Initialization in Proposition 8): The iteration in Proposition 8 can be initialized using the SVD of $F^\mathsf{H} H^{*\mathsf{H}} S$, that is, $U \Sigma V^\mathsf{H} \overset{\text{SVD}}{:=} F^\mathsf{H} H^{*\mathsf{H}} S$. □

Remark 6 (Projection in Proposition 8): For $\text{Proj}_{\Omega_{N \times L}}[\cdot]$, one strategy is to use the singular value matrix of the argument. Another strategy is to only keep the real non-negative components in the diagonal entries and zero the rest. □

4) Another Approximate Solution Method to (9)*:* This subsection proposes another approximate solution method to the robust counterpart (9). Based on the design philosophy of Method 1, the following alternative method can be motivated.

Method 2: In Method 1, Problem (39) can be replaced by

$$\min_{X \in \mathcal{X}} \quad \alpha \|H^* X - S\|_F^2 + \|X - \bar{X}\|_F^2. \tag{45}$$

Therefore, Problem (34) can also be approximately solved by (38) and (45). □

Note that in the objective of Problem (39), the first term is to guarantee that X solves $\min_{X \in \mathcal{X}} \|H^* X - S\|_F^2$, while the second term is to reduce $\|X - \bar{X}\|_F^2$. This is why Problem (45) can directly be an alternative to Problem (39). However, the two problems are not equivalent, although both perform effectively and efficiently in experiments; see Section V. The closed-form solution to (45) is given in the proposition below.

Proposition 9: Problem (45) is analytically solved by

$$X^* = \sqrt{L} F \tilde{U} I_{N \times L} \tilde{V}^\mathsf{H}$$

where

$$\tilde{U} \tilde{\Sigma} \tilde{V}^\mathsf{H} \overset{\text{SVD}}{=} F^\mathsf{H} \tilde{H}^\mathsf{H} \tilde{S},$$

$$\tilde{H} := \begin{bmatrix} \sqrt{\alpha} H^* \\ I_N \end{bmatrix}, \quad \text{and} \quad \tilde{S} := \begin{bmatrix} \sqrt{\alpha} S \\ \bar{X} \end{bmatrix}.$$

Proof: By defining $\tilde{H}$ and $\tilde{S}$, Problem (45) becomes $\min_{X \in \mathcal{X}} \|\tilde{H} X - \tilde{S}\|_F^2$, which is solved by X^* defined in the statement; cf. [20, Eq. (15)] and (29). □

B. Solution Method to Robust Counterpart (10)

This subsection designs the solution method to the robust counterpart (10) of the joint waveform design problem (8).

1) Model Reformulation: Depending on the design preference, let the feasible region $\mathcal{X}$ of waveforms X be defined as one of the follows:

- TPC: $\mathcal{X} := \{X : \|X\|_F^2 = LP_\mathsf{T}\}$, for the total power constraint in (11);
- PAPC: $\mathcal{X} := \{X : \text{diag}(X X^\mathsf{H}) = \frac{L \cdot P_\mathsf{T}}{N} I_N\}$, for the per-antenna power constraint in (12).

In analogy with Theorem 1 for Problem (9), we give a reformulation of Problem (10).

Theorem 2: Let $\phi(H, X) := \rho \|HX - S\|_F^2 + (1-\rho)\|X - X_s\|_F^2$. If there exists $X^* \in \mathcal{X}$ such that

$$\max_{H \in \mathcal{H}} \min_{X \in \mathcal{X}} \phi(H, X) = \max_{H \in \mathcal{H}} \phi(H, X^*), \tag{46}$$

then Problem (10) is equivalent to

$$\begin{aligned} \max_{H} \quad & \rho \|H X_H^* - S\|_F^2 + (1-\rho)\|X_H^* - X_s\|_F^2 \\ \text{s.t.} \quad & \|H - \bar{H}\| \le \theta, \end{aligned} \tag{47}$$

where X_H^*, which depends on H, is given by

- [20, Algo. 1], for the total power constraint (11);
- [20, Algo. 2], for the per-antenna power constraint (12).

Proof: This is obvious from Lemma 1 and [20]. □

Denote the objective function in Problem (47) by

$$g(\boldsymbol{H}) := \rho\|\boldsymbol{H}\boldsymbol{X}_{\boldsymbol{H}}^* - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X}_{\boldsymbol{H}}^* - \boldsymbol{X}_s\|_F^2, \quad (48)$$

which is a particularization of (25). Suppose that $\bar{\boldsymbol{X}}$ solves the nominal joint waveform design problem (8) under the nominal communication channel $\bar{\boldsymbol{H}}$; for technical details, see [20, Algos. 1, 2].

Motivated by Lemma 2, the proposition below establishes the Lipschitz continuity of the function g on $\mathcal{H}$.

Proposition 10: The function $g(\boldsymbol{H})$ defined in (48) is upper bounded by the function $\bar{g}(\boldsymbol{H})$, for every $\boldsymbol{H} \in \mathcal{H}$, where

$$\bar{g}(\boldsymbol{H}) := \rho\|\boldsymbol{H}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\bar{\boldsymbol{X}} - \boldsymbol{X}_s\|_F^2. \quad (49)$$

In addition, g and $\bar{g}$ are L_g-Lipschitz continuous in $\boldsymbol{H}$ on $\mathcal{H}$ with respect to the norm $\|\cdot\|$ used in defining $\mathcal{H}$, where

$$\begin{aligned}L_g &= \rho \cdot L_f \\ &= 2\rho B L P_\mathsf{T} \cdot (\|\bar{\boldsymbol{H}}\|_F + B\theta) + 2\rho B\sqrt{LP_\mathsf{T}} \cdot \|\boldsymbol{S}\|_F\end{aligned} \quad (50)$$

and L_f is defined in (31); B is a finite positive real number such that $\|\boldsymbol{H}\|_F \le B\|\boldsymbol{H}\|$ for every $\boldsymbol{H} \in \mathbb{C}^{K \times N}$.

Proof: The proof is routine given the techniques in the proof of Proposition 1. For details, see Appendix P in supplementary materials. □

Note that $\bar{g}(\boldsymbol{H})$ is a particularization of (26), due to which the Lipschitz continuity is crucial. Proposition 10 implies that the function g in (48) is more "flat" than f in (30) since $\rho \le 1$, which can further reduce the gap studied in Lemma 2. In addition, from the perspective of global optimization, such flatness would benefit the design of solution methods, specifically, improving the efficiency of the search process of iterative algorithms; see [43], [45].

2) Approximate Solution Method to (10)*:* Motivated by Method 2, an approximate solution method to (10) can be summarized in the method below.

Method 3: Problem (10) is approximately solved by $(\boldsymbol{H}^*, \boldsymbol{X}^*)$ where $\boldsymbol{H}^*$ is a global maximum of the upper bound function $\bar{g}(\boldsymbol{H})$, that is,

$$\begin{aligned}\boldsymbol{H}^* \in \underset{\boldsymbol{H}}{\arg\max} \quad &\|(\bar{\boldsymbol{X}}^\mathsf{T} \otimes \boldsymbol{I}_K)\operatorname{vec}(\boldsymbol{H}) - \operatorname{vec}(\boldsymbol{S})\|_2^2 \\ \text{s.t.} \quad &\|\operatorname{vec}(\boldsymbol{H}) - \operatorname{vec}(\bar{\boldsymbol{H}})\| \le \theta,\end{aligned} \quad (51)$$

and $\boldsymbol{X}^*$ is a global minimum of

$$\begin{aligned}\min_{\boldsymbol{X} \in \mathcal{X}} \quad &\alpha \cdot [\rho\|\boldsymbol{H}^*\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2] + \\ &\|\boldsymbol{X} - \bar{\boldsymbol{X}}\|_F^2\end{aligned} \quad (52)$$

where $\alpha \ge 0$ is a large real number to numerically ensure that $\boldsymbol{X}$ solves

$$\min_{\boldsymbol{X} \in \mathcal{X}} \rho\|\boldsymbol{H}^*\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2. \quad □$$

Problem (51) can be transformed into (41) and solved by Proposition 5. Problem (52) can be transformed into

$$\min_{\boldsymbol{X} \in \mathcal{X}} \frac{\alpha}{\alpha+1}\|\tilde{\boldsymbol{H}}\boldsymbol{X} - \tilde{\boldsymbol{S}}\|_F^2 + \frac{1}{\alpha+1}\|\boldsymbol{X} - \bar{\boldsymbol{X}}\|_F^2, \quad (53)$$

where

$$\tilde{\boldsymbol{H}} := \begin{bmatrix} \sqrt{\rho}\boldsymbol{H}^* \\ \sqrt{1-\rho}\boldsymbol{I}_N \end{bmatrix}, \quad \text{and} \quad \tilde{\boldsymbol{S}} := \begin{bmatrix} \sqrt{\rho}\boldsymbol{S} \\ \sqrt{1-\rho}\boldsymbol{X}_s \end{bmatrix}.$$

The transformed problem (53) can be solved by

- [20, Algo. 1], for the total power constraint (11);
- [20, Algo. 2], for the per-antenna power constraint (12).

C. Remarks on Solution Methods

This section studies three solution methods to robust counterparts: Methods 1 and 2 are designed for the robust counterpart (9), whereas Method 3 is for the robust counterpart (10). As base-band digital signal processing methods, they are realized in digital signal processors of transmitters, for example, in distributed units (DUs) of wireless communication systems. The three methods are summarized in Table I.

TABLE I
SUMMARY OF SOLUTION METHODS

Problem	Method	Component	Reformulation	Solution
(9)	1	(38)	(41)	Prop. 5
		(39)	N/A	Prop. 8
	2	(38)	(41)	Prop. 5
		(45)	N/A	Prop. 9
(10)	3	(51)	(41)	Prop. 5
		(52)	(53)	[20, Algo. 1, 2]

All three methods involve a convex quadratic maximization problem (41), specifically, (38) in Methods 1 and 2, and (51) in Method 3, which is solved using Proposition 5. At each iteration in Proposition 5, two dominating operations are included: (42) and (43). However, both (42) and (43) can be analytically solved; see Propositions 6 and 7 where one Cholesky decomposition and several matrix inversion and multiplication operations are included. Problem (39) is solved by the iteration process in Proposition 8; at each iteration, three SVD and several matrix multiplication operations are contained. Problem (45) has the closed-form solution in Proposition 9, where one SVD and several matrix multiplication operations are encompassed.

Table II summarizes the asymptotically computational complexities of the involved solution methods, in terms of floating-point operation (FLOP). Note that in real-world operation, we have $K \le N \le L$. Therefore, the time complexity of the full SVD of an $N \times L$ matrix, which computes all SVD components, is $\mathcal{O}(L^2N + N^2L + N^3)$ using the Golub–Reinsch algorithm [46, p. 493]. For the Cholesky decomposition and the inversion of an $N \times N$ matrix, the time complexity is $\mathcal{O}(N^3)$. Since the variables in (41) are constructed from matrices, they are inherently large-dimensional. As a result, computing $\boldsymbol{C}^\mathsf{T}\boldsymbol{C}$ requires $\mathcal{O}(K^3N^2L)$ FLOPs and inverting $\boldsymbol{C}^\mathsf{T}\boldsymbol{C}$ requires $\mathcal{O}(K^3N^3)$ FLOPs because $\boldsymbol{C} \in \mathbb{R}^{2KL \times 2KN}$.

Table II suggests that all the methods in this paper are solvable in polynomial times, and in this sense, they are computationally efficient. However, note that the time complexities of the algorithms in Propositions 5, 8, and [20, Algo. 2] also depend on the number of numerical iterations. On the other hand, as the closed-form solution exists at every dominating operation, the actual running speeds of the proposed methods (i.e., Propositions 5, 6, 7, 8, and 9) are empirically very fast; see the experimental illustrations in Subsection V-B5.

TABLE II
ASYMPTOTICALLY COMPUTATIONAL COMPLEXITY ANALYSES; CF. TABLE I

Solution	Dominating Operation	Time Complexity	Solution Type	Remark
Prop. 5	(42), Prop. 6, Chol & MatInv & MatMul	$\mathcal{O}(K^3N^3 + K^3N^2L + K^3L^2N)$	Closed-Form	At Each Iteration
	(43), Prop. 7, MatMul	$\mathcal{O}(K^3N^2L)$	Closed-Form	
Prop. 8	Step a, SVD & MatMul	$\mathcal{O}(N^3 + N^2L + N^2K + L^2N + NKL)$	Closed-Form	At Each Iteration
	Step b, SVD & MatMul	$\mathcal{O}(L^3 + N^2L + N^2K + L^2N + NKL)$	Closed-Form	
	Step c, SVD & MatMul	$\mathcal{O}(N^3 + N^2L + L^2K + L^2N + NKL)$	Closed-Form	
Prop. 9	SVD & MatMul	$\mathcal{O}(N^3 + N^2L + N^2K + L^2N + NKL)$	Closed-Form	Overall
[20, Algo. 1]	See [20, Sec. IV-E]	$\mathcal{O}(N^3 + N^2L + N^2K + NKL)$	Iterative	Overall
[20, Algo. 2]	See [20, Sec. IV-E]	$\mathcal{O}(N^2L + NKL)$	Iterative	At Each Iteration

Chol: Cholesky Decomposition;   MatInv: Matrix Inversion;   MatMul: Matrix Multiplication

V. EXPERIMENTS

The experiments are conducted using MATLAB 2023a on an HP® Desktop PC equipped with 12th Gen Intel® Core™ i7-12700K processor (3.60 GHz), 64.0 GB RAM, and 64-bit operating system. All the source data and codes are available online at GitHub: https://github.com/Spratm-Asleaf/Robust-Waveform. One may use them to reproduce the experimental results and verify the empirical claims in this paper. For each simulation case, we conduct 1000 independent Monte–Carlo episodes, over which the reported results are averaged; for details, see Appendix Q in supplementary materials.

A. Experimental Setups

Following the real-world experimental setups in Beijing, China, in [47, Fig. 3], we imagine a MIMO ISAC system placed at an intersection where two perpendicular streets cross each other, serving four downlink communication users while sensing two targets; see Fig. 4. The key system parameters are configured as follows.

- The carrier frequency is 5.9 GHz; see [47, Tab. 1];
- The total transmitting power is $P_\mathsf{T} = 34$ dBm $= 2.5$ watts; see [47, Tab. 1];
- In [47, p. 136], the signal-to-noise ratio (SNR) is 24 dB, that is, the noise power is $N_0 = P_\mathsf{T}/10^{(24/10)} = 0.01$ (watts). In our experiments, we use a lower SNR = 10 dB, i.e., $N_0 = 0.25$ watts, which is a more complex case;
- The number of radio paths for each communication channel is in-between 60 and 100; the average is 80; see [47, Fig. 7].

For sufficient generality and practicality, the mature MATLAB Phased Array System Toolbox and Communications Toolbox are employed to simulate the transmitters and the scattering-based communication channels.

For other system parameters, we follow the setups in [20] for the convenience of comparison.

- The ISAC base station is located at the origin $(0, 0)$;
- $K = 4$, $N = 16$, $L = 30$; cf. (1);
- The constellation S is constructed using 4-point quadrature phase-shift keying (QPSK);
- The two targets are located with azimuths of $(-45, 45)$ degrees relative to the vertical line of sight of the base station, and with randomly generated ranges of $(1.51, 1.39)$ kilometers (Km) from the base station, respectively;
- The expected beamwidth at each target direction is 10 degrees;

(a) Satellite View   (b) Map View
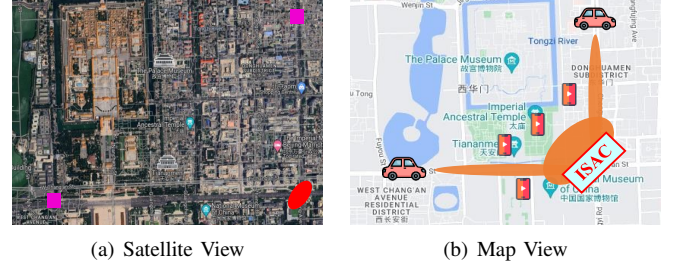
Fig. 4. The ISAC system is placed at a crossing of two perpendicular streets in Beijing, China. It senses two targets (cars) and serves four downlink-communication users (phones). For adequate sensing qualities, two high-directional beams point to the two targets, respectively. In Fig. (a), the red ellipse represents the ISAC station, and magenta squares denote cars. In Fig. (b), the orange elongated ellipses illustrate beams. (Icon Credit: FLATICON.com; Photo Credit: Google Map.)

- The locations of the four downlink communication users are randomly generated according to the uniform distribution on $[0, 1] \times [-1, 1]$ Km².

We suppose that the nominal channel $\bar{H}$ and the true channel H_0 are generated from the small perturbations of a reference channel H_ref, that is,

$$\begin{aligned} H_0 &= H_\mathrm{ref} + \epsilon \cdot \Delta_1, \\ \bar{H} &= H_\mathrm{ref} + \epsilon \cdot \Delta_2, \end{aligned} \tag{54}$$

where H_ref is simulated using the MATLAB toolboxes and fixed throughout the experiments; $\bar{H}$ is also fixed after channel estimation; $\epsilon := 0.05$ is a small number and the perturbation matrices Δ_1 and Δ_2 are distributed according to entry-wise standard complex Gaussian. Note that the true channel H_0 randomly varies from one Monte–Carlo test to another, however, it is close to the fixed nominal channel $\bar{H}$. Note also that waveform designs are based on the nominal channel $\bar{H}$, while performance tests are on the true channels H_0 because, in simulations, H_0 is known. In practice, the channel's uncertainty quantification parameter θ such that $\|H_0 - \bar{H}\| \le \theta$ is unknown for a specified norm $\|\cdot\|$ because we only know $\bar{H}$ and all other information about the true channel H_0 is missing.[3] Hence, in real-world operation, θ can be a tuning parameter unless it is specified in the channel estimation stage. In the experiments, $\|\cdot\|$ is particularized to the matrix F-norm or the vector 2-norm; cf. Proposition 7.

Note that the parameter configurations described above do not essentially impact the experimental results in this paper.

---

[3]The relation $\|H_0 - \bar{H}\| \le \theta$ holds in probability due to (54); cf. Fact 3.

Therefore, these setups are merely employed as demonstrative examples; one may use the shared source codes to change the configurations to verify the claims in this section.

B. Experimental Results

1) Perfect Waveforms for Sensing and Communication: The beam pattern for perfect sensing is shown in Fig. 5(a), where two high-directional beams point to the two targets located with azimuths of $(-45, 45)$ degrees, respectively. For each target, the beam width is 10 degrees. The theoretically perfect beam pattern for sensing is not practically achievable, and therefore, the mean-squared-error minimization method [29] is used to design the practically perfect beam pattern (i.e., the beampattern-inducing matrix $\boldsymbol{R}$). When $\boldsymbol{R}$ is available, the cyclic algorithm [30] is used to design the perfect sensing waveform $\boldsymbol{X}_s$ with constrained peak-to-average power ratio (PAPR); recall (5).

(a) Perfect Sensing   (b) Perfect Communication (for $\bar{\boldsymbol{H}}$)
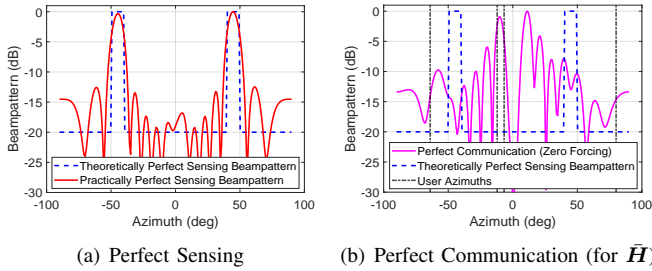
Fig. 5. The beam patterns for perfect sensing and perfect communication, respectively. The perfect communication, obtained using zero-forcing precoding, is specific to the nominal channel $\bar{\boldsymbol{H}}$.

As for perfect communication, in the sense of maximum AASR $R_{\boldsymbol{H}_0, \boldsymbol{X}_0}$ defined in (3), it cannot be specified because $R_{\boldsymbol{H}_0, \boldsymbol{X}_0}$ depends on the true but unknown channel $\boldsymbol{H}_0$ and the associated truly optimal waveform $\boldsymbol{X}_0$. For the purpose of illustration, we show the perfect communication waveform under the nominal channel, that is, the zero-forcing waveform $\boldsymbol{X}_c := \bar{\boldsymbol{H}}^{\mathsf{H}}(\bar{\boldsymbol{H}}\bar{\boldsymbol{H}}^{\mathsf{H}})^{-1}\boldsymbol{S}$; see Fig. 5(b). The zero-forcing waveform is perfect for communication in the sense of interference mitigation (i.e., MUI cancellation). We see that to reduce MUI energy, the communication beams do not necessarily point to users. In addition, the optimal waveform for communication is indeed not necessarily the same as the optimal waveform for sensing; see Table III. The detection probability of the target at the azimuth of -45 degrees is calculated using [48, Eq. (69)].

TABLE III
SYSTEM PERFORMANCES FOR SEPARATE WAVEFORM DESIGN

Performance Metric	Waveform	Value
Detection Probability	$\boldsymbol{X}_s$	1
(for Sensing)	$\boldsymbol{X}_c$	0.38
AASR	$\boldsymbol{X}_s$	0.30 (bps/Hz/user)
(for Communication)	$\boldsymbol{X}_c$	3.07 (bps/Hz/user)
Azimuth of Target to Detect: -45 Degrees		

2) Nominal Waveform Design: In this subsection, we show the practical issues of the nominal waveform design methods (6) and (8) using $\bar{\boldsymbol{H}}$. To be specific, we demonstrate the issues of the nominal Pareto frontiers (specified by nominal designs)

and the necessity of uncertainty-aware performance characterization (i.e., robust/conservative Pareto frontier). Experimental results are shown in Fig. 6; cf. Figs. 1 and 2.

(a) Sensing-Centric Design

(b) Joint Design (TPC; $\rho \le 0.1$)   (c) Joint Design (TPC; $\rho \ge 0.1$)

(d) Joint Design (PAPC; $\rho \le 0.1$)   (e) Joint Design (PAPC; $\rho \ge 0.1$)
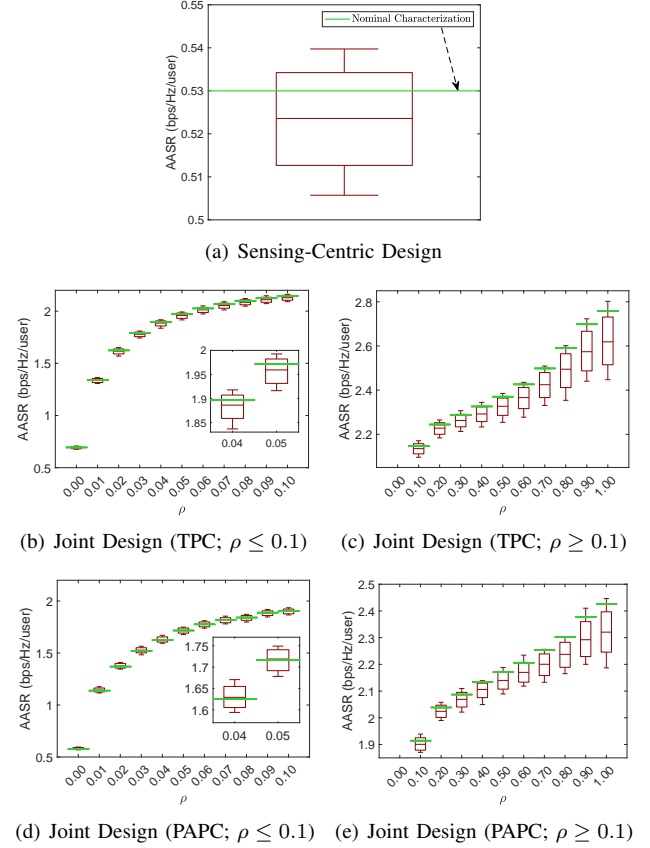
Fig. 6. Nominal Pareto frontiers (i.e., nominal AASR performances; green) and true AASR performances (box plots). In (b)-(e), sensing-centric and communication-centric designs are obtained for $\rho = 0$ and $\rho = 1$, respectively; cf. (8). (Cf. Fig. 2(a).)

In Fig. 6, the box plots of the true communication performances $R_{\boldsymbol{H}_0, \bar{\boldsymbol{X}}}$ using the nominally optimal waveform $\bar{\boldsymbol{X}}$ under 1000 independent Monte–Carlo simulations (i.e., 1000 different realizations of $\boldsymbol{H}_0$) are shown. The box plots display the minimum value, 5% percentile, median, 95% percentile, and the maximum value among all AASRs $R_{\boldsymbol{H}_0, \bar{\boldsymbol{X}}}$. The green lines indicate the nominal characterization of sensing-communication performances, which, however, cannot fully depict the actual performance trade-off; cf. Fig 2(a).

In addition, the box plots in Fig. 6 support that the joint design methods can improve communication performances by sacrificing sensing performances via changing the trade-off parameter ρ. Moreover, the joint design under the total power constraint (TPC) brings higher AASR than that under the per-antenna power constraint (PAPC).

3) Robust Sensing-Centric Waveform Design (9)*:* The experimental results of sensing-centric robust waveform designs are shown in Fig. 7, where the box plots of the true communication performances $R_{\boldsymbol{H}_0, \boldsymbol{X}^*}$ using the robust waveforms $\boldsymbol{X}^*$ with different θs are shown. As we can see, in robust designs, the radius θ of the uncertainty set $\mathcal{H}$ should be elegantly specified: if θ is overly large, the associated performance bound would be extremely conservative (i.e., loose), while if

θ is overly small, the performance lower bound cannot be correctly identified; recall Subsections III-B and III-C.

(a) Method 1 against θ

(b) Method 1 ($\theta = 0.127$)

(c) Method 2 against θ

(d) Method 2 ($\theta = 0.127$)
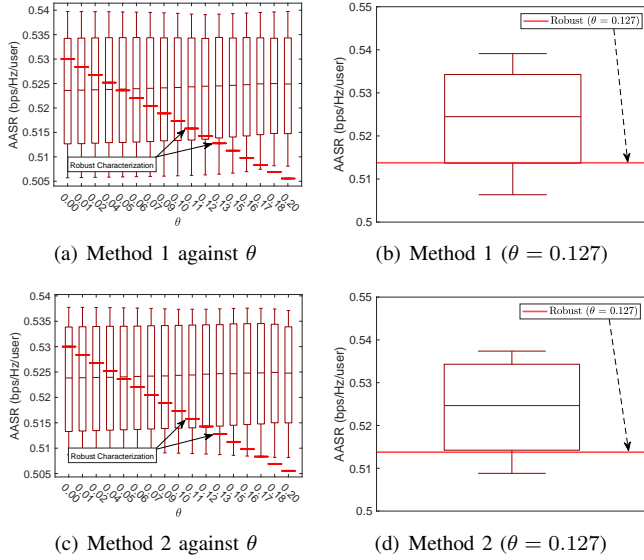
Fig. 7. Conservative characterization under robust sensing-centric waveform design. Red step values show the robust performance characterizations R_{H^*, X^*} under different θ. When $\theta \approx 0.127$, R_{H^*, X^*} correctly identifies the tightest conservative performance bounds of true AASRs, with probability 95%. NB: $\theta = 0$ gives nominal characterizations. (Cf. Fig. 2(b).)

In Fig. 8, we show the beam patterns of the three waveforms: the perfect-sensing waveform X_s, the nominal waveform $\bar{X}$, and the robust waveform X^* when $\theta = 0.127$. Since this is a sensing-centric design scheme, all waveforms are guaranteed to have the same beam pattern.

(a) Method 1

(b) Method 2
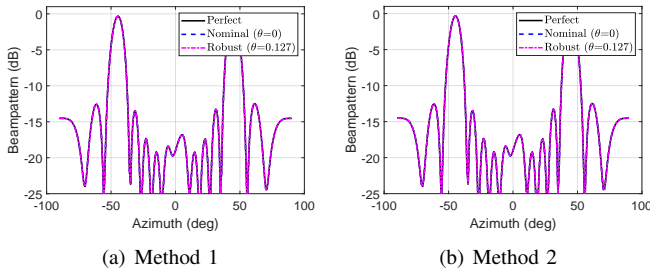
Fig. 8. Beam patterns of sensing-centric ISAC waveforms (including perfect-sensing, nominal, and robust waveforms); the two sub-figures are identical.

4) Robust Joint Waveform Design (10)*:* The experimental results of robust joint waveform design, subject to TPC and PAPC, are shown in Figs. 9 and 10, respectively. In both cases, conservative (i.e., robust) performance boundaries are successfully identified. As joint design schemes, nominally optimal waveforms and robust waveforms are not guaranteed to have the same beam pattern as the perfect waveform for sensing; see Fig. 11.

5) Results on Algorithmic Convergence and Computation Burdens: In Propositions 5 and 8, iterative processes are involved. However, experiments show that the two processes converge extremely fast on average, almost within six or seven iterations. This is attributed to the existence of closed-form solutions at each iteration. One illustration is given in Fig. 12.

(a) TPC against θ ($\rho = 0.25$)

(b) TPC ($\theta = 0.16$, $\rho = 0.25$))

(c) TPC ($\theta = 0.16$, $\rho \leq 0.1$)

(d) TPC ($\theta = 0.16$, $\rho \geq 0.1$)
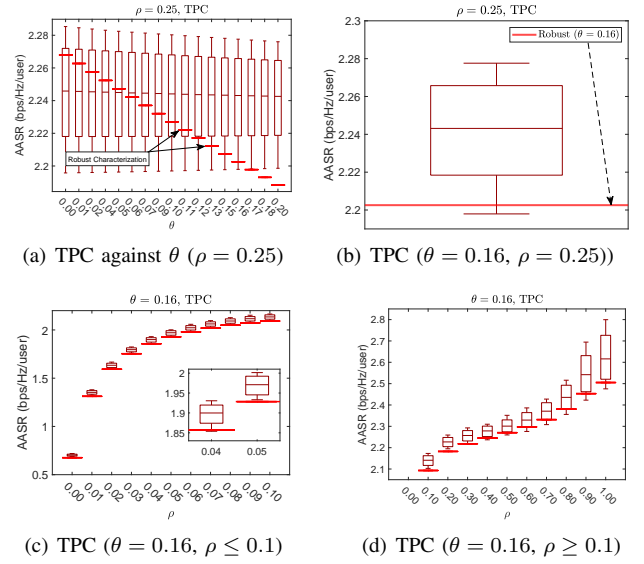
Fig. 9. Conservative characterization under robust joint waveform design subject to total power constraint (TPC). When $\theta = 0.16$, the tight conservative performance boundary is identified, with probability 95%. (Cf. Fig. 2(b).)

(a) PAPC against θ ($\rho = 0.25$)

(b) PAPC ($\theta = 0.16$, $\rho = 0.25$)

(c) PAPC ($\theta = 0.16$, $\rho \leq 0.1$)

(d) PAPC ($\theta = 0.16$, $\rho \geq 0.1$)
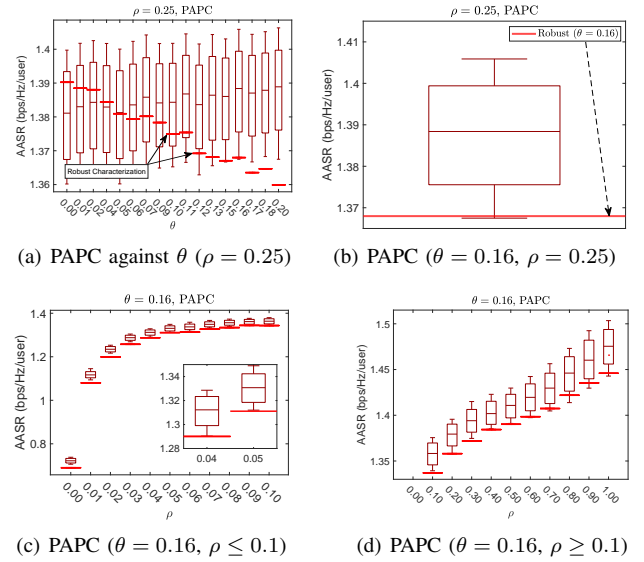
Fig. 10. Conservative characterization under robust joint waveform design subject to per-antenna power constraint (PAPC). When $\theta = 0.16$, the tight conservative performance boundary is identified, with probability 95%. (Cf. Fig. 2(b).) Since high computational burdens exist at each iteration (cf. Table IV), the solution algorithm is forced to terminate after limited iterations. As a result, the performances are not globally optimal and stable; see, e.g., (a).

The average running times of involved methods are shown in Table IV; see Table I for the components of each method. Table IV suggests the following: 1) non-robust nominal designs consume significantly fewer running times than their robust counterparts; 2) the joint design method under PAPC has significantly more computational burdens than the sensing-centric design and the joint design under TPC; 3) the sensing-centric design and the joint design under TPC are computationally comparable; 4) Methods 1, 2, and 3 under TPC are computationally comparable.

(a) TPC ($\rho = 0.25$)

(b) TPC ($\rho = 0.75$)

(c) PAPC ($\rho = 0.25$)

(d) PAPC ($\rho = 0.75$)
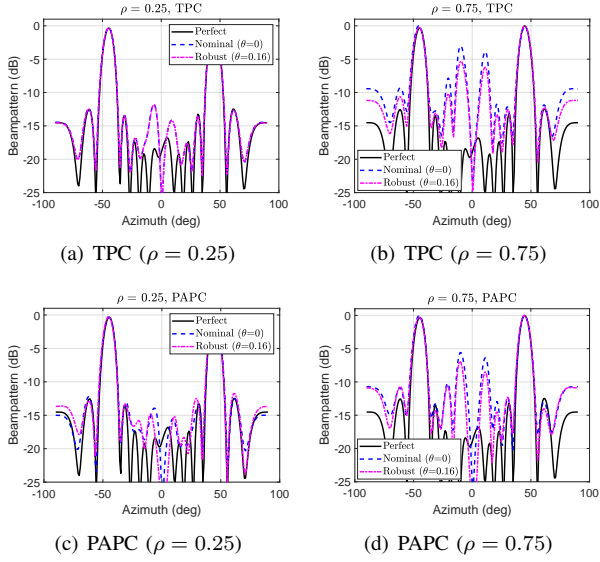
Fig. 11. Beam patterns of joint-design ISAC waveforms (including perfect-sensing, nominal, and robust waveforms).

(a) Iteration in Proposition 5

(b) Iteration in Proposition 8
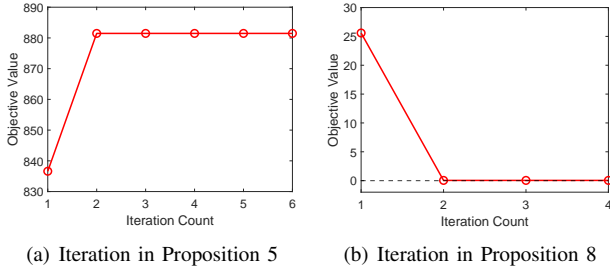
Fig. 12. Empirical convergence illustrations of the iteration processes in Propositions 5 and 8, terminating within six steps on average. Note that at each iteration step, closed-form solutions exist.

TABLE IV
AVERAGE RUNNING TIMES OF DESIGN METHODS; CF. TABLE I

	Sensing Centric			Joint (TPC)		Joint (PAPC)	
	N.	M. 1	M. 2	N.	M. 3	N.	M. 3
Times	0.36	5.83	4.60	0.69	5.76	132.35	180.17

N.: <u>N</u>ominal Designs;   M.: <u>M</u>ethod (Robust Designs)
Time Unit: milliseconds (ms)

(a) Nominal ($\epsilon = 0.001$)

(b) Conservative ($\epsilon = 0.001$)

(c) Nominal ($\epsilon = 0.01$)

(d) Conservative ($\epsilon = 0.01$)

(e) Nominal ($\epsilon = 0.1$)

(f) Conservative ($\epsilon = 0.1$)

(g) Nominal ($\epsilon = 0.5$)

(h) Conservative ($\epsilon = 0.5$)

Fig. 13. Nominal and conservative characterizations against different values of ϵ under the sensing-centric design; cf. Figs. 6(a) and 7. To obtain tight performance lower bounds, when $\epsilon = 0.001$, $\theta = 0.12 \times 10^{-2}$; when $\epsilon = 0.01$, $\theta = 0.12 \times 10^{-1}$; when $\epsilon = 0.1$, $\theta = 0.12$; when $\epsilon = 0.5$, $\theta = 0.60$.

6) Results on Different ϵ: In previous experiments, we use $\epsilon = 0.05$; cf. (54). In this subsection, we offer more empirical results against different values of ϵ to show the applicability of the proposed robust method for different uncertainty levels. As demonstrations, we adopt the sensing-centric design scheme; see Fig. 13. From Fig. 13, the following two observations can be outlined: First, the larger the uncertainty level ϵ, the more dispersive the AASRs; second, for different uncertainty levels ϵ, the proposed robust design method can provide conservative characterizations under appropriate θs (i.e., identify the tight performance lower bounds with probability 95%). For more discussions, see Appendix R in supplementary materials.

VI. CONCLUSIONS

This paper investigates uncertainty-aware performance characterization and robust waveform design for ISAC. With uncer-

tainties in nominal communication channels, the phenomenon of unreliable Pareto frontier is noticed and the concept of robust (i.e., conservative) performance characterization is proposed; see Fig. 2. We show that min-max robust waveform design formulations can obtain the conservative performance boundary; see Section III. To solve these min-max formulations, an approximate solution framework is presented; see Section IV. The experiments validate that the proposed methods are effective in robust performance characterization for ISAC and are also computationally efficient; see Section V.

However, in real-world operations, the true radius θ of the uncertainty set is unknown, which can therefore be left as an empirically tunable parameter for investigated real-world ISAC systems unless it can be elegantly specified in the channel estimation stage; cf. Figs. 7, 9, and 10.

REFERENCES

[1] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.

[2] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. on Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.

[3] J. A. Zhang, F. Liu, C. Masouros, R. W. Heath, Z. Feng, L. Zheng, and A. Petropulu, "An overview of signal processing techniques for joint communication and radar sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1295–1315, 2021.

[4] A. Liu, Z. Huang, M. Li, Y. Wan, W. Li, T. X. Han, C. Liu, R. Du, D. K. P. Tan, J. Lu *et al.*, "A survey on fundamental limits of integrated sensing and communication," *IEEE Commun. Surveys & Tutorials*, vol. 24, no. 2, pp. 994–1034, 2022.

[5] W. Zhou, R. Zhang, G. Chen, and W. Wu, "Integrated sensing and communication waveform design: A survey," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1930–1949, 2022.

[6] T. S. Ferguson, *A Course in Large Sample Theory*. Chapman & Hall, 1996.

[7] C. Ouyang, Y. Liu, and H. Yang, "Performance of downlink and uplink integrated sensing and communications (ISAC) systems," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1850–1854, 2022.

[8] C. Ouyang, Y. Liu, H. Yang, and N. Al-Dhahir, "Integrated sensing and communications: A mutual information-based framework," *IEEE Communications Magazine*, vol. 61, no. 5, pp. 26–32, 2023.

[9] L. Xie, F. Liu, J. Luo, and S. Song, "Sensing mutual information with random signals in gaussian channels: Bridging sensing and communication metrics," *arXiv preprint arXiv:2402.03919*, 2024.

[10] P. Kumari, S. A. Vorobyov, and R. W. Heath, "Adaptive virtual waveform design for millimeter-wave joint communication–radar," *IEEE Trans. on Signal Processing*, vol. 68, pp. 715–730, 2019.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[12] Y. Xiong, F. Liu, Y. Cui, W. Yuan, T. X. Han, and G. Caire, "On the fundamental tradeoff of integrated sensing and communications under gaussian channels," *IEEE Trans. on Inform. Theory*, 2023.

[13] K. V. Mishra, M. B. Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 100–114, 2019.

[14] S. Aditya, O. Dizdar, B. Clerckx, and X. Li, "Sensing using coded communications signals," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 134–152, 2022.

[15] Z. Wei, J. Piao, X. Yuan, H. Wu, J. A. Zhang, Z. Feng, L. Wang, and P. Zhang, "Waveform design for MIMO-OFDM integrated sensing and communication system: An information theoretical approach," *IEEE Trans. on Commun.*, 2023.

[16] L. Gaudio, M. Kobayashi, G. Caire, and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 9, pp. 5951–5965, 2020.

[17] X. Ouyang and J. Zhao, "Orthogonal chirp division multiplexing," *IEEE Trans. on Commun.*, vol. 64, no. 9, pp. 3946–3957, 2016.

[18] T. Huang, N. Shlezinger, X. Xu, Y. Liu, and Y. C. Eldar, "Majorcom: A dual-function radar communication system using index modulation," *IEEE Trans. on Signal Processing*, vol. 68, pp. 3423–3438, 2020.

[19] A. Ahmed, E. Aboutanios, and Y. D. Zhang, "Sensing-Centric ISAC Signal Processing," in *Integrated Sensing and Communications*. Springer, 2023, pp. 179–209.

[20] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Toward dual-functional radar-communication systems: Optimal waveform design," *IEEE Trans. on Signal Processing*, vol. 66, no. 16, pp. 4264–4279, 2018.

[21] B. Guo, J. Liang, G. Wang, B. Tang, and H. So, "Bistatic MIMO DFRC system waveform design via fractional programming," *IEEE Trans. on Signal Processing*, 2023.

[22] B. Liao, X. Xiong, and Z. Quan, "Robust beamforming design for dual-function radar-communication system," *IEEE Transactions on Vehicular Technology*, 2023.

[23] X. Li, V. C. Andrei, U. J. Mönich, and H. Boche, "Optimal and robust waveform design for MIMO-OFDM channel sensing: A cramer-rao bound perspective," in *2023 IEEE International Conference on Communications*, 2023.

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

[25] N. Zhao, Y. Wang, Z. Zhang, Q. Chang, and Y. Shen, "Joint transmit and receive beamforming design for integrated sensing and communication," *IEEE Commun. Letters*, vol. 26, no. 3, pp. 662–666, 2022.

[26] Z. Ren, L. Qiu, J. Xu, and D. W. K. Ng, "Robust transmit beamforming for secure integrated sensing and communication," *IEEE Trans. on Commun.*, 2023.

[27] M. Luan, B. Wang, Z. Chang, T. Hämäläinen, and F. Hu, "Robust beamforming design for ris-aided integrated sensing and communication system," *IEEE Trans. on Intell. Transport. Syst.*, 2023.

[28] A. Bazzi and M. Chafii, "On integrated sensing and communication waveforms with tunable PAPR," *IEEE Transactions on Wireless Communications*, 2023.

[29] D. R. Fuhrmann and G. San Antonio, "Transmit beamforming for MIMO radar systems using signal cross-correlation," *IEEE Trans. on Aerosp. Electron. Syst.*, vol. 44, no. 1, pp. 171–186, 2008.

[30] P. Stoica, J. Li, and X. Zhu, "Waveform synthesis for diversity-based transmit beampattern design," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2593–2598, 2008.

[31] S. Ahmed and I. M. Jaimoukha, "A relaxation-based approach for the orthogonal procrustes problem with data uncertainties," in *Proceedings of 2012 UKACC International Conference on Control*. IEEE, 2012, pp. 906–911.

[32] Y. D. Sergeyev, R. G. Strongin, and D. Lera, *Introduction to global optimization exploiting space-filling curves*. Springer Science & Business Media, 2013.

[33] S. Desale, A. Rasool, S. Andhale, and P. Rane, "Heuristic and meta-heuristic algorithms and their relevance to the real world: a survey," *International Journal of Computer Engineering in Research Trends*, vol. 351, no. 5, pp. 2349–7084, 2015.

[34] R. Mart, P. M. Pardalos, and M. G. Resende, *Handbook of Heuristics*. Springer Publishing, 2018.

[35] A. Kaveh and T. Bakhshpoori, *Metaheuristics: Outlines, MATLAB Codes, and Examples*. Springer, 2019.

[36] E. D. Taillard, *Design of Heuristic Algorithms for Hard Optimization*. Springer Publishing, 2023.

[37] E. Vazquez and J. Bect, "Convergence properties of the expected improvement algorithm with fixed mean and covariance functions," *Journal of Statistical Planning and Inference*, vol. 140, no. 11, pp. 3088–3095, 2010.

[38] A. D. Bull, "Convergence rates of efficient global optimization algorithms." *Journal of Machine Learning Research*, vol. 12, no. 10, 2011.

[39] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2015.

[40] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, pp. 455–492, 1998.

[41] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—what is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, 2019.

[42] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek, "Scalable global optimization via local bayesian optimization," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[43] C. Malherbe and N. Vayatis, "Global optimization of lipschitz functions," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2314–2323.

[44] A. Ben-Tal and E. Roos, "An algorithm for maximizing a convex function based on its minimum," *INFORMS Journal on Computing*, vol. 34, no. 6, pp. 3200–3214, 2022.

[45] P. Hansen and B. Jaumard, *Lipschitz Optimization*. Springer, 1995.

[46] G. H. Golub and C. F. Van Loan, *Matrix Computations*. JHU press, 2013.

[47] H. Zhang, R. He, B. Ai, S. Cui, and H. Zhang, "Measuring sparsity of wireless channels," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 133–144, 2020.

[48] A. Khawar, A. Abdelhadi, and C. Clancy, "Target detection performance of spectrum sharing MIMO radars," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 4928–4940, 2015.

Online Supplementary Materials

This document contains appendices to the paper, including:

- Extensive discussions on Model (2); see Appendix A;
- Proofs of all lemmas, theorems, and propositions;
- An intuitive explanation of Method 1; see Appendix I;
- The specifics of how the simulations in Section V are conducted; see Appendix Q.

APPENDIX A
EXTENSIONS OF MODEL (2)

For the signal model $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}$ in (1), it can be written as

$$\boldsymbol{Y} = \boldsymbol{S} + (\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}) + \boldsymbol{W}.$$

To recover $\boldsymbol{S}$ from $\boldsymbol{Y}$, we aim to minimize the interference signals $\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}$ over waveforms $\boldsymbol{X}$. This is another reason to minimize MUI energy $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ as in (2). Below we discuss the case of multi-antenna users and the case of multi-carrier.

- **Multi-Antenna Case**. Suppose that we have K downlink users and each user is equipment with R receive antennas. Then, the channel matrix of each user is $\boldsymbol{H}_k \in \mathbb{C}^{R \times N}$ for every $k \in [K]$, where N is the number of transmit antennas at the base station. As a result, for each user k, the base-band signal model is $\boldsymbol{Y}_k = \boldsymbol{H}_k\boldsymbol{X} + \boldsymbol{W}_k$, where $\boldsymbol{X} \in \mathbb{C}^{N \times L}$ is the transmitted waveform and L is the frame length. By constructing $\boldsymbol{Y}$, $\boldsymbol{H}$, and $\boldsymbol{W}$ as

$$\boldsymbol{Y} := \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_K \end{bmatrix}, \quad \boldsymbol{H} := \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_K \end{bmatrix}, \quad \boldsymbol{W} := \begin{bmatrix} \boldsymbol{W}_1 \\ \boldsymbol{W}_2 \\ \vdots \\ \boldsymbol{W}_K \end{bmatrix},$$

the integrated base-band signal model $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W}$ can be obtained. Note that in this case, $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ no longer physically means the multi-user interference (MUI) energy. However, as in (2), minimizing $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ with respect to $\boldsymbol{X}$ is still the technical focus to improve communication performance (i.e., to reduce the restoration error of the information matrix $\boldsymbol{S}$).

- **Multi-Carrier Case**. Suppose that we have R sub-carriers and for each sub-carrier $r \in [R]$, the base-band signal model is $\boldsymbol{Y}_r = \boldsymbol{H}_r\boldsymbol{X}_r + \boldsymbol{W}_r$, where $\boldsymbol{Y}_r \in \mathbb{C}^{K \times L}$, $\boldsymbol{H}_r \in \mathbb{C}^{K \times N}$, $\boldsymbol{X}_r \in \mathbb{C}^{N \times L}$, and $\boldsymbol{W}_r \in \mathbb{C}^{K \times L}$; K is the number of downlink single-antenna users, L is the frame length, and N is the number of transmit antennas at the base station. Since every sub-carrier has an independent base-band signal model, all signal processing operations can be applied separately for every base-band model indexed by $r \in [R]$.

APPENDIX B
PROOF OF LEMMA 1

Proof: The first inequality in the lemma is due to the weak min-max property (also known as the min-max inequality), which is unconditionally true for any ϕ, $\mathcal{H}$, and $\mathcal{X}$. The second inequality is due to the feasibility of the solution $\boldsymbol{X}^*$ in $\mathcal{X}$. This completes the proof. $\square$

APPENDIX C
PROOF OF LEMMA 2

Proof: We have

$$\max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X})$$
$$= \max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) + \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X})$$
$$\qquad - \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X})$$
$$\leq \big| \max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) \big| +$$
$$\qquad \big| \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) - \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X}) \big|$$
$$\leq \max_{\boldsymbol{H}} \big| \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) \big| +$$
$$\qquad \max_{\boldsymbol{H}} \big| \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) - \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X}) \big|$$
$$\leq L_2 \cdot \max_{\boldsymbol{H}} \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| + L_1 \cdot \max_{\boldsymbol{H}} \|\boldsymbol{H} - \bar{\boldsymbol{H}}\|$$
$$\leq (L_1 + L_2) \cdot \theta.$$

This completes the proof. $\square$

APPENDIX D
PROOF OF THEOREM 1

Proof: First, we consider the max-min counterpart of (9):

$$\max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \quad \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$$
$$\text{s.t.} \quad \frac{1}{L} \boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} = \boldsymbol{R}, \qquad (55)$$
$$\|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \leq \theta.$$

For every feasible $\boldsymbol{H}$, the inner sub-problem $\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ is solved by

$$\boldsymbol{X}_{\boldsymbol{H}}^* = \sqrt{L} \cdot \boldsymbol{F} \cdot \boldsymbol{U}\boldsymbol{I}_{N \times L}\boldsymbol{V}^{\mathsf{H}},$$

where $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}} \overset{\text{SVD}}{=} \boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{S}$ and $\boldsymbol{I}_{N \times L} := [\boldsymbol{I}_N, \boldsymbol{0}_{N \times (L-N)}]$; the $N \times (L - N)$ zero matrix is denoted by $\boldsymbol{0}_{N \times (L-N)}$; see [20, Eq. (15)]. Note that the optimal solution $\boldsymbol{X}_{\boldsymbol{H}}^*$ depends on $\boldsymbol{H}$, and $\boldsymbol{X}_{\boldsymbol{H}}^*$ may not be unique given $\boldsymbol{H}$. Plugging in $\boldsymbol{X}_{\boldsymbol{H}}^*$ back to (55) yields (28).

Second, according to Lemma 1 and Condition (27), the strong min-max property holds, that is,

$$\min_{\boldsymbol{X} \in \mathcal{X}} \max_{\boldsymbol{H} \in \mathcal{H}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2 = \max_{\boldsymbol{H} \in \mathcal{H}} \min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$$
$$= \max_{\boldsymbol{H} \in \mathcal{H}} \|\sqrt{L} \cdot \boldsymbol{H} \cdot \boldsymbol{F} \cdot \boldsymbol{U}\boldsymbol{I}_{N \times L}\boldsymbol{V}^{\mathsf{H}} - \boldsymbol{S}\|_F^2.$$
$$(56)$$

This completes the proof. $\square$

APPENDIX E
PROOF OF PROPOSITION 1

One may verify that it is difficult to prove the continuity of f directly using the definition in (30) because the SVD of a matrix might not be unique; specifically, given $\bar{\boldsymbol{H}}$, there may exist multiple $\boldsymbol{U}$s and $\boldsymbol{V}$s such that $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}} = \boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{S}$. The complication arises when $\boldsymbol{U}$ and $\boldsymbol{V}$ correspond to zero singular value(s) in $\boldsymbol{\Sigma}$. We, therefore, investigate the continuity of f from its original definition.

Proof: Recall that $f(\boldsymbol{H}) := \min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ where $\mathcal{X} := \{\boldsymbol{X} : \boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} = L\boldsymbol{R}\}$. First, note that for every $\boldsymbol{X} \in \mathcal{X}$ and every $\boldsymbol{H}_1, \boldsymbol{H}_2 \in \mathcal{H}$, there exists an upper bound $0 < B_1 < \infty$ such that $\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F + \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \leq B_1$. A loose

choice can be $B_1 := 2\sqrt{LP_\mathsf{T}} \cdot (\|\bar{\boldsymbol{H}}\|_F + B\theta) + 2\|\boldsymbol{S}\|_F$ where $0 < B < \infty$ is a real-valued constant such that $\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|_F \le B\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|$; the existence of B is guaranteed due to the equivalence of norms on a finite-dimensional linear space. Just note that $|\|\boldsymbol{H}_1\|_F - \|\bar{\boldsymbol{H}}\|_F| \le \|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|_F \le B\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\| \le B\theta$. The same argument also holds for $\boldsymbol{H}_2$. Hence, for every $\boldsymbol{H}_1, \boldsymbol{H}_2 \in \mathcal{H}$, we have

$$
\begin{aligned}
&|f(\boldsymbol{H}_1) - f(\boldsymbol{H}_2)| \\
&= \left| \min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F^2 - \min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F^2 \right| \\
&\le \max_{\boldsymbol{X} \in \mathcal{X}} \left| \|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F^2 - \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F^2 \right| \\
&= \max_{\boldsymbol{X} \in \mathcal{X}} \left| \left(\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F + \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \right) \cdot \right. \\
&\qquad\qquad \left. \left(\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F - \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \right) \right| \\
&\le \max_{\boldsymbol{X} \in \mathcal{X}} \left| \|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F + \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \right| \cdot \\
&\qquad\qquad \left| \|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F - \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \right| \\
&\le B_1 \cdot \max_{\boldsymbol{X} \in \mathcal{X}} \left\| \boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{H}_2\boldsymbol{X} \right\|_F \\
&\le B_1 \cdot \left\| \boldsymbol{H}_1 - \boldsymbol{H}_2 \right\|_F \cdot \max_{\boldsymbol{X} \in \mathcal{X}} \left\| \boldsymbol{X} \right\|_F \\
&= B_1 \sqrt{LP_\mathsf{T}} \cdot \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|_F \\
&\le B_1 \sqrt{LP_\mathsf{T}} \cdot B\|\boldsymbol{H}_1 - \boldsymbol{H}_2\|.
\end{aligned}
$$

Note that $\|\boldsymbol{X}\|_F = \sqrt{\operatorname{Tr} \boldsymbol{X}^\mathsf{H}\boldsymbol{X}} = \sqrt{L \cdot \operatorname{Tr} \boldsymbol{R}} = \sqrt{LP_\mathsf{T}}$. $\quad\square$

APPENDIX F
PROOF OF PROPOSITION 2

Proof: The non-convexity and non-concavity of the objective function $f(\boldsymbol{H}) = \min_{\boldsymbol{X}: \ \boldsymbol{X}\boldsymbol{X}^\mathsf{H} = L\boldsymbol{R}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ can be verified through the definitions of convexity and concavity by constructing counterexamples, completing the proof. $\quad\square$

The example below specifically justifies the claim above.

Example 1: Let the nominal channel be $\bar{H} := 0.9 + 0.5j$, $\epsilon := 0.05$, $R := 1$, and $S := 1 + 1j$. Suppose that H_1 and H_2 are generated according to the following formulas:

$$
\begin{aligned}
H_1 &= \bar{H} + \epsilon \cdot \Delta_1, \\
H_2 &= \bar{H} + \epsilon \cdot \Delta_2,
\end{aligned}
$$

where Δ_1 and Δ_2 are mutually-independent standard complex Gaussian variables. One may verify that $f(0.5H_1 + 0.5H_2) \le 0.5f(H_1) + 0.5f(H_2)$ for some realizations of H_1 and H_2, while $f(0.5H_1 + 0.5H_2) \ge 0.5f(H_1) + 0.5f(H_2)$ for other realizations. Therefore, f is neither concave nor convex. $\quad\square$

APPENDIX G
PROOF OF PROPOSITION 3

Proof: The upper bound is obtained by plugging $\bar{\boldsymbol{X}}$ into the optimization problem $f(\boldsymbol{H}) := \min_{\boldsymbol{X}: \ \boldsymbol{X}\boldsymbol{X}^\mathsf{H} = L\boldsymbol{R}} \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ because $\bar{\boldsymbol{X}}$ is a feasible solution. The lower bound is obtained by the reverse triangle inequality, i.e., $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F \ge |\|\boldsymbol{H}\boldsymbol{X}\|_F - \|\boldsymbol{S}\|_F|$; note that $\|\boldsymbol{H}\boldsymbol{X}\|_F = \sqrt{L \operatorname{Tr}[\boldsymbol{H}^\mathsf{H}\boldsymbol{H}\boldsymbol{R}]}$ because $\boldsymbol{X}\boldsymbol{X}^\mathsf{H} = L\boldsymbol{R}$. The upper bound is positive-definite quadratic (thus convex) in $\operatorname{vec}(\boldsymbol{H})$ because $\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}^\mathsf{H} = L\boldsymbol{R}$ and $\boldsymbol{R}$ is positive definite. Since at the center $\bar{\boldsymbol{H}}$ of $\mathcal{H}$ we have $f(\bar{\boldsymbol{H}}) = \bar{f}(\bar{\boldsymbol{H}})$, the upper bound is tight. Also, the reverse triangle inequality is tight, and therefore, the lower bound is

tight. The third claim is obvious. We show the uniform bound in the fourth claim below. We have

$$
\begin{aligned}
\bar{f}(\boldsymbol{H}) - f(\boldsymbol{H}) &= \bar{f}(\boldsymbol{H}) - \bar{f}(\bar{\boldsymbol{H}}) + f(\bar{\boldsymbol{H}}) - f(\boldsymbol{H}) \\
&\le |\bar{f}(\boldsymbol{H}) - \bar{f}(\bar{\boldsymbol{H}})| + |f(\bar{\boldsymbol{H}}) - f(\boldsymbol{H})| \\
&\le L_f \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| + L_f \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \\
&\le 2L_f \cdot \theta.
\end{aligned}
$$

This completes the proof. $\quad\square$

APPENDIX H
PROOF OF PROPOSITION 4

Proof: The Hessian of the objective function in terms of $\operatorname{vec}(\boldsymbol{H})$ is

$$
\begin{aligned}
&L((\boldsymbol{F}\boldsymbol{A})^\mathsf{T} \otimes \boldsymbol{I}_K)^\mathsf{H}((\boldsymbol{F}\boldsymbol{A})^\mathsf{T} \otimes \boldsymbol{I}_K) \\
&= L\{[(\boldsymbol{F}\boldsymbol{A})^\mathsf{T}]^\mathsf{H} \otimes \boldsymbol{I}_K^\mathsf{H}\}\{(\boldsymbol{F}\boldsymbol{A})^\mathsf{T} \otimes \boldsymbol{I}_K\} \\
&= L[(\boldsymbol{F}\boldsymbol{A})^\mathsf{T}]^\mathsf{H}(\boldsymbol{F}\boldsymbol{A})^\mathsf{T} \otimes \boldsymbol{I}_K^\mathsf{H}\boldsymbol{I}_K \\
&= L[(\boldsymbol{F}\boldsymbol{A})(\boldsymbol{F}\boldsymbol{A})^\mathsf{H}]^\mathsf{T} \otimes \boldsymbol{I}_K^\mathsf{H}\boldsymbol{I}_K \\
&= L\boldsymbol{R}^\mathsf{T} \otimes \boldsymbol{I}_K \succ \boldsymbol{0},
\end{aligned}
$$

because the beampattern-inducing matrix $\boldsymbol{R}$ is positive definite. Hence, the objective function of (34) is positive definite in $\operatorname{vec}(\boldsymbol{H})$, and therefore, convex in $\operatorname{vec}(\boldsymbol{H})$. The convexity of the objective function in terms of $\operatorname{vec}(\boldsymbol{A})$ can be shown in a similar way: just note that the objective function can only be shown to be positive semi-definite because the Hessian $L(\boldsymbol{I}_L \otimes \boldsymbol{H}\boldsymbol{F})^\mathsf{H}(\boldsymbol{I}_L \otimes \boldsymbol{H}\boldsymbol{F})$ is not necessarily positive definite.

The feasible region of $\boldsymbol{H}$ is convex because the norm constraint is convex and the equality constraint is linear. The feasible region of $\boldsymbol{A}$ is convex because it is built by a linear equality constraint. This completes the proof. $\quad\square$

APPENDIX I
INTUITIVE UNDERSTANDING OF SOLUTION METHOD 1

Motivated by Lemma 1 and Condition (27), we start with employing the upper bound $\bar{f}(\boldsymbol{H})$ in Proposition 3 because $\bar{f}(\boldsymbol{H})$ is close to $f(\boldsymbol{H})$ if the radius θ of the uncertainty set $\mathcal{H}$ is small; recall from Subsection III-C that the radius of uncertainty set should be controlled to small. The other reason to employ $\bar{\boldsymbol{X}}$ and $\bar{f}(\boldsymbol{H})$ is that $\bar{\boldsymbol{X}}$ exists for any specified uncertainty set $\mathcal{H}$ with any radius $\theta \ge 0$ because $\mathcal{H}$ is centered at $\bar{\boldsymbol{H}}$.

Step 1. Maximize the upper bound $\bar{f}(\boldsymbol{H})$ to obtain $\boldsymbol{H}^*$:

$$
\boldsymbol{H}^* = \operatorname*{argmax}_{\boldsymbol{H} \in \mathcal{H}} \|\boldsymbol{H}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2.
$$

Interpretation. Step 1 gives a feasible approximation solution $(\bar{\boldsymbol{X}}, \boldsymbol{H}^*)$ to Problem (28), and therefore Problem (9), in the sense of Fact 1. To be specific, we have

a) Bound of the Truly Optimal Cost and the True Cost:

$$
\min_{\boldsymbol{X} \in \mathcal{X}} \|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}_0\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}^*\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2;
$$

b) Bound of the Nominally Optimal Cost:

$$
\min_{\boldsymbol{X} \in \mathcal{X}} \|\bar{\boldsymbol{H}}\boldsymbol{X} - \boldsymbol{S}\|_F^2 = \|\bar{\boldsymbol{H}}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \le \|\boldsymbol{H}^*\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2;
$$

where the nominally optimal solution $\bar{\boldsymbol{X}} := \sqrt{L}\boldsymbol{F}\bar{\boldsymbol{U}}\boldsymbol{I}_{N \times L}\bar{\boldsymbol{V}}^\mathsf{H}$ solves the nominal waveform design problem $\min_{\boldsymbol{X} \in \mathcal{X}} \|\bar{\boldsymbol{H}}\boldsymbol{X} - \boldsymbol{S}\|_F^2$ and $\mathcal{X} = \{\boldsymbol{X} : \boldsymbol{X}\boldsymbol{X}^\mathsf{H} = L\boldsymbol{R}\}$.

However, maximizing the upper-bound $\overline{f}(\boldsymbol{H})$ gives an extremely conservative solution. Specifically, the robust cost $\|\boldsymbol{H}^*\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2$ would be overly large than the true cost $\|\boldsymbol{H}_0\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2$ and the truly optimal cost $\min_{\boldsymbol{X}\in\mathcal{X}}\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2$. Hence, a refinement is needed.

Step 2. Refine the robust cost $\|\boldsymbol{H}^*\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2$: i.e.,

$$\boldsymbol{X}^* = \underset{\boldsymbol{X}\in\mathcal{X}}{\operatorname{argmin}}\|\boldsymbol{H}^*\boldsymbol{X} - \boldsymbol{S}\|_F^2.$$

Consequence. However, the resulting cost $\|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2$ cannot be guaranteed to upper bound the truly optimal cost, the true cost, and the nominally optimal cost. To be specific, it is unnecessary to have

a) Bound of the Truly Optimal Cost and the True Cost:

$$\min_{\boldsymbol{X}\in\mathcal{X}}\|\boldsymbol{H}_0\boldsymbol{X} - \boldsymbol{S}\|_F^2 \leq \|\boldsymbol{H}_0\boldsymbol{X}^* - \boldsymbol{S}\|_F^2 \overset{?}{\leq} \|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2.$$

b) Bound of the Nominally Optimal Cost:

$$\min_{\boldsymbol{X}\in\mathcal{X}}\|\bar{\boldsymbol{H}}\boldsymbol{X} - \boldsymbol{S}\|_F^2 = \|\bar{\boldsymbol{H}}\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 \overset{?}{\leq} \|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2.$$

(Note that $\boldsymbol{H}^*$ is obtained by the maximization at $\bar{\boldsymbol{X}}$.) Thus, we propose a remedy strategy.

Step 3. Design a mechanism to let $\boldsymbol{X}^* = \bar{\boldsymbol{X}}$.
Interpretation. If it technically holds that $\boldsymbol{X}^* = \bar{\boldsymbol{X}}$ (or $\boldsymbol{X}^* \approx \bar{\boldsymbol{X}}$), then the conservativeness of the solution $(\boldsymbol{H}^*, \bar{\boldsymbol{X}})$ will be controlled, and equivalently, the feasibility of the solution $(\boldsymbol{H}^*, \boldsymbol{X}^*)$ in the sense of Fact 1 will be guaranteed.

The three algorithmic steps above provide an intuitive understanding of Method 1.

APPENDIX J
PROOF OF LEMMA 3

Proof: We have $(\boldsymbol{\Gamma} + \boldsymbol{\Theta}j)(\boldsymbol{a} + \boldsymbol{b}j) = (\boldsymbol{\Gamma}\boldsymbol{a} - \boldsymbol{\Theta}\boldsymbol{b}) + (\boldsymbol{\Theta}\boldsymbol{a} + \boldsymbol{\Gamma}\boldsymbol{b})j$. Hence, the stacking scheme to construct real quantities from complex quantities is given in the statement of the lemma. This completes the proof. $\qquad\square$

APPENDIX K
PROOF OF LEMMA 4

Proof: According to Proposition 4, we immediately have $\boldsymbol{C}^\mathsf{T}\boldsymbol{C} \succ \boldsymbol{0}$, and therefore, $p(\boldsymbol{h})$ is positive definite in $\boldsymbol{h}$. In addition, we have $p(\boldsymbol{h}) - \frac{\mu}{2}\boldsymbol{h}^\mathsf{T}\boldsymbol{h} = \boldsymbol{h}^\mathsf{T}(\boldsymbol{C}^\mathsf{T}\boldsymbol{C} - \frac{\mu}{2}\boldsymbol{I}_{2KN})\boldsymbol{h} - 2\boldsymbol{s}^\mathsf{T}\boldsymbol{C}\boldsymbol{h} + \boldsymbol{s}^\mathsf{T}\boldsymbol{s}$, where $\mu > 0$ is a positive number. Since $\boldsymbol{C}^\mathsf{T}\boldsymbol{C} \succ \boldsymbol{0}$, there exist $\mu > 0$ such that $\boldsymbol{C}^\mathsf{T}\boldsymbol{C} - \frac{\mu}{2}\boldsymbol{I}_{2KN} \succeq \boldsymbol{0}$. As a result, the function $p(\boldsymbol{h}) - \frac{\mu}{2}\boldsymbol{h}^\mathsf{T}\boldsymbol{h}$ can be a convex function for some $\mu > 0$, which means that the objective function $p(\boldsymbol{h})$ of (41) is strongly convex. $\qquad\square$

APPENDIX L
PROOF OF PROPOSITION 5

Proof: According to [49, Thm. 1], a point $\boldsymbol{y}$ is a globally optimal solution to (41) if and only if $\langle \nabla p(\boldsymbol{y}), \boldsymbol{h} - \boldsymbol{y} \rangle \leq 0$, for every $\boldsymbol{h}$ such that $\|\boldsymbol{h} - \bar{\boldsymbol{h}}\| \leq \theta$, where $\nabla p(\boldsymbol{y})$ denotes the gradient of p evaluated at $\boldsymbol{y}$. Therefore, if we have $\max_{\boldsymbol{y}\in\mathcal{Y}}\max_{\boldsymbol{h}:\|\boldsymbol{h}-\bar{\boldsymbol{h}}\|\leq\theta} \langle \nabla p(\boldsymbol{y}), \boldsymbol{h} - \boldsymbol{y} \rangle \leq 0$, for some dedicated $\mathcal{Y}$, then every $\boldsymbol{y}$ that solves the above optimization is a global maximum of (41). This proposition, which is

adapted from [49, Algo. 1] for Problem (41), formalizes the above intuition. The global optimality and convergence are therefore guaranteed by [49, Thm. 4]. For rigorous and complete technical proof, see [49]; just note that in (42), the equality constraint can be changed to its convex inequality counterpart because the optima of linear objective functions lie on the boundary of feasible regions. $\qquad\square$

APPENDIX M
PROOF OF PROPOSITION 6

Proof: Since $\boldsymbol{C}^\mathsf{T}\boldsymbol{C}$ is positive definite, the invertible $\boldsymbol{M}$ exists; see Appendix K. Problem (42) is equivalent to

$$\boldsymbol{y}_k = \underset{\boldsymbol{y}\in\mathbb{R}^{2KN}}{\operatorname{argmax}} \quad (\boldsymbol{h}_{k-1}^\mathsf{T}\boldsymbol{C}^\mathsf{T}\boldsymbol{C} - \boldsymbol{s}^\mathsf{T}\boldsymbol{C}) \cdot \boldsymbol{y}$$
$$\text{s.t.} \qquad \|\boldsymbol{M}\boldsymbol{y} - \boldsymbol{M}^{-\mathsf{T}}\boldsymbol{C}^\mathsf{T}\boldsymbol{s}\|_2 = \gamma.$$

The above display is further equivalent to

$$\underset{\boldsymbol{z}\in\mathbb{R}^{2KN}}{\max} \quad (\boldsymbol{h}_{k-1}^\mathsf{T}\boldsymbol{C}^\mathsf{T}\boldsymbol{C} - \boldsymbol{s}^\mathsf{T}\boldsymbol{C})\boldsymbol{M}^{-1} \cdot (\gamma\boldsymbol{z} + \boldsymbol{M}^{-\mathsf{T}}\boldsymbol{C}^\mathsf{T}\boldsymbol{s}),$$
$$\text{s.t.} \qquad \|\boldsymbol{z}\|_2 = 1.$$

Due to Cauchy–Schwarz inequality, the maximum is

$$\boldsymbol{z}^* = \frac{\boldsymbol{M}^{-\mathsf{T}}(\boldsymbol{C}^\mathsf{T}\boldsymbol{C}\boldsymbol{h}_{k-1} - \boldsymbol{C}^\mathsf{T}\boldsymbol{s})}{\|\boldsymbol{M}^{-\mathsf{T}}(\boldsymbol{C}^\mathsf{T}\boldsymbol{C}\boldsymbol{h}_{k-1} - \boldsymbol{C}^\mathsf{T}\boldsymbol{s})\|_2}.$$

Then, a maximum of (42) is $\boldsymbol{y}^* = \boldsymbol{M}^{-1} \cdot (\gamma\boldsymbol{z}^* + \boldsymbol{M}^{-\mathsf{T}}\boldsymbol{C}^\mathsf{T}\boldsymbol{s})$. This completes the proof. $\qquad\square$

APPENDIX N
PROOF OF PROPOSITION 7

Proof: Problem (43) is equivalent to

$$\underset{\boldsymbol{z}\in\mathbb{R}^{2KN}}{\max} \quad (\boldsymbol{y}_k^\mathsf{T}\boldsymbol{C}^\mathsf{T}\boldsymbol{C} - \boldsymbol{s}^\mathsf{T}\boldsymbol{C}) \cdot (\theta\boldsymbol{z} + \bar{\boldsymbol{h}}), \text{ s.t. } \|\boldsymbol{z}\|_2 = 1.$$

Due to Cauchy–Schwarz inequality, the maximum is

$$\boldsymbol{z}^* = \frac{\boldsymbol{C}^\mathsf{T}\boldsymbol{C}\boldsymbol{y}_k - \boldsymbol{C}^\mathsf{T}\boldsymbol{s}}{\|\boldsymbol{C}^\mathsf{T}\boldsymbol{C}\boldsymbol{y}_k - \boldsymbol{C}^\mathsf{T}\boldsymbol{s}\|_2}.$$

As a result, a maximum of (43) is $\boldsymbol{h}^* = \theta\boldsymbol{z}^* + \bar{\boldsymbol{h}}$. $\qquad\square$

APPENDIX O
PROOF OF PROPOSITION 8

Proof: In terms of $\boldsymbol{U}$, we can rewrite (39) as

$$\underset{\boldsymbol{U}}{\min}\|\boldsymbol{U}\boldsymbol{A}_1 - \boldsymbol{B}_1\|_F^2, \quad \text{s.t.} \quad \boldsymbol{U}\boldsymbol{U}^\mathsf{H} = \boldsymbol{I}_N, \qquad (57)$$

where $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ are defined in Proposition 8. Problem (57) is a standard orthogonal Procrustes problem whose closed-form solution is given in Proposition 8; see technical details in [50].

In terms of $\boldsymbol{V}$, we can rewrite (39) as

$$\underset{\boldsymbol{V}}{\min}\|\boldsymbol{V}\boldsymbol{A}_2 - \boldsymbol{B}_2\|_F^2, \quad \text{s.t.} \quad \boldsymbol{V}\boldsymbol{V}^\mathsf{H} = \boldsymbol{I}_L, \qquad (58)$$

where $\boldsymbol{A}_2$ and $\boldsymbol{B}_2$ are defined in Proposition 8. Problem (58) is a standard orthogonal Procrustes problem whose closed-form solution is given in Proposition 8; see technical details in [50].

In terms of $\boldsymbol{\Sigma}$, Problem (39) is a positive-definite-quadratic convex program. Note that the space $\Omega_{N\times L}$ of the matrices

with diagonal, non-negative, and real entries is convex. The gradient of the objective function of (39) with respect to $\boldsymbol{\Sigma}$ is

$$2\boldsymbol{U}^{\mathsf{H}}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{H}} - \boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{*\mathsf{H}}\boldsymbol{S})\boldsymbol{V} = 2\boldsymbol{\Sigma} - 2\boldsymbol{U}^{\mathsf{H}}\boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{*\mathsf{H}}\boldsymbol{S}\boldsymbol{V}.$$

Therefore, the optimal solution is given by the projected point of $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{F}^{\mathsf{H}}\boldsymbol{H}^{*\mathsf{H}}\boldsymbol{S}\boldsymbol{V}$ onto $\Omega_{N \times L}$.

The convergence proof of the iteration process is straightforward. We rewrite (39) in shorthand as

$$\min_{\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}} \psi(\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}),$$

where the constraints of the variables $(\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V})$ are implicitly defined by (39). Let $(\boldsymbol{U}^0,\boldsymbol{\Sigma}^0,\boldsymbol{V}^0)$ denote the initial values of the variables and $(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)$ the values at the rth iteration. Following the defined iteration process in Proposition 8, we have

$$\begin{aligned}
\psi(\boldsymbol{U}^0,\boldsymbol{\Sigma}^0,\boldsymbol{V}^0) &\geq \psi(\boldsymbol{U}^1,\boldsymbol{\Sigma}^0,\boldsymbol{V}^0) \\
&\geq \psi(\boldsymbol{U}^1,\boldsymbol{\Sigma}^0,\boldsymbol{V}^1) \\
&\geq \psi(\boldsymbol{U}^1,\boldsymbol{\Sigma}^1,\boldsymbol{V}^1) \\
&\vdots \\
&\geq \psi(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r),
\end{aligned}$$

for every $r \geq 1$. Therefore, the sequence $\{\psi(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)\}$, which is indexed by r, is decreasing as r increases. Since $\psi(\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V}) \geq 0$ for every feasible $(\boldsymbol{U},\boldsymbol{\Sigma},\boldsymbol{V})$, according to the monotone convergence theorem, $\psi(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)$ monotonically converges to a non-negative value as r goes to infinity.

Note that $(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)$ is not guaranteed to converge because at the rth iteration, the values of $(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)$ may not be unique. This completes the proof. $\square$

APPENDIX P
PROOF OF PROPOSITION 10

Proof: According to Lemma 1, the upper bound function $\bar{g}(\boldsymbol{H})$ is straightforward to obtain because $\bar{\boldsymbol{X}}$ is a feasible solution in $\mathcal{X}$.

The Lipschitz continuity of g is shown as follows. For every $\boldsymbol{H}_1, \boldsymbol{H}_2 \in \mathcal{H}$, we have

$$\begin{aligned}
&|g(\boldsymbol{H}_1) - g(\boldsymbol{H}_2)| \\
&= \Big| \min_{\boldsymbol{X}\in\mathcal{X}} \rho\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2 - \\
&\qquad \min_{\boldsymbol{X}\in\mathcal{X}} \rho\|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2 \Big| \\
&\leq \max_{\boldsymbol{X}\in\mathcal{X}} \Big| \big[\rho\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2\big] - \\
&\qquad \big[\rho\|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F^2 + (1-\rho)\|\boldsymbol{X} - \boldsymbol{X}_s\|_F^2\big] \Big| \\
&= \rho \cdot \max_{\boldsymbol{X}\in\mathcal{X}} \Big| \|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F^2 - \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F^2 \Big| \\
&\leq \rho \cdot L_f \cdot \|\boldsymbol{H}_1 - \boldsymbol{H}_2\|.
\end{aligned}$$

Using the same manipulations, $\bar{g}(\boldsymbol{H})$ can also be shown to be ρL_f-Lipschitz continuous. This completes the proof. $\square$

APPENDIX Q
DETAILS ON EXPERIMENTS

In this appendix, we detail the logic flow upon which the shared source codes are written.

Algorithm 1 Simulation Engine

Definition: Let I denote the number of Monte–Carlo episodes.
Remark: The notation $0 : 0.01 : 0.2$ means a discrete vector starting at 0, ending with 0.2, and uniformly spaced with 0.01.

Input: $I = 1000$
1: *// Stage 1: Engine Initialization*
2: $N \leftarrow 16$, $K \leftarrow 4$, $L \leftarrow 30$, $f \leftarrow 5.9 \times 10^9$, $P_{\mathsf{T}} \leftarrow 2.5$, $\theta \leftarrow 0 : 0.01 : 0.2$
3: Generate $\boldsymbol{H}_{\text{ref}}$
4: Design Perfect-Sensing Waveform $\boldsymbol{X}_s$
5: *// Stage 2: Offline Design Using Practically Available Nominal Channel*
6: Generate Constellation $\boldsymbol{S}$
7: Generate Nominal Channel $\bar{\boldsymbol{H}}$ Using (54)
8: Design Nominally Optimal Waveform $\bar{\boldsymbol{X}}$ Using $\bar{\boldsymbol{H}}$
9: Calculate Nominally Estimated AASR $R_{\bar{\boldsymbol{H}},\bar{\boldsymbol{X}}}$
10: Design Robust Waveform $\boldsymbol{X}^*$ Using $\bar{\boldsymbol{H}}$ for Every θ (NB: $\boldsymbol{X}^*$ depends on θ)
11: Calculate Robustly Estimated AASR $R_{\boldsymbol{H}^*,\boldsymbol{X}^*}$
12: *// Stage 3: Online Test Using Random And Practically Unknown True Channel*
13: $i \leftarrow 0$
14: **while** true **do**
15:     *// Calculate True AASRs*
16:     Uniformly Generate $\boldsymbol{H}_0$ Using (54)
17:     Calculate True AASR $R_{\boldsymbol{H}_0,\bar{\boldsymbol{X}}}$ at Nominally Optimal Waveform $\bar{\boldsymbol{X}}$
18:     Calculate True AASR $R_{\boldsymbol{H}_0,\boldsymbol{X}^*}$ at Robust Waveform $\boldsymbol{X}^*$ for Every θ (NB: $\boldsymbol{X}^*$ depends on θ)
19:     *// Next Episode*
20:     $i \leftarrow i + 1$
21:     *// End of Simulation*
22:     **if** $i > I$ **then**
23:         **break while**
24:     **end if**
25: **end while**
Output: AASRs for all I episodes (used for box plots)

APPENDIX R
ADDITIONAL RESULTS ON DIFFERENT ϵ

The gap defined in Lemma 2, i.e., the value of $\|\boldsymbol{H}^*\bar{\boldsymbol{X}} - \boldsymbol{S}\|_F^2 - \|\boldsymbol{H}^*\boldsymbol{X}^* - \boldsymbol{S}\|_F^2$ (recall Methods 1 and 2), is shown in Table V. The gap is statistically small even for large ϵ.

TABLE V
GAP IN LEMMA 2

ϵ	0.1	0.25	0.5
Gap	0.055 ± 0.031	0.234 ± 0.089	0.706 ± 0.188

Note: Format: mean $\pm$ std; when $\epsilon < 0.1$, the gap is almost zero.

REFERENCES

[49] R. Enhbat, "An algorithm for maximizing a convex function over a simple set," Journal of Global Optimization, vol. 8, pp. 379–391, 1996.
[50] R. Everson, "Orthogonal, but not orthonormal, procrustes problems," Advances in Computational Mathematics, vol. 3, no. 4, 1998.