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

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

To recover $\boldsymbol{S}$ from $\mathbf{Y}$, 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 $\mathbf{Y}_k = \boldsymbol{H}_k\boldsymbol{X} + \mathbf{W}_k$, 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
$$\mathbf{Y} := \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_K \end{bmatrix}, \quad \boldsymbol{H} := \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{H}_2 \\ \vdots \\ \boldsymbol{H}_K \end{bmatrix}, \quad \mathbf{W} := \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \vdots \\ \mathbf{W}_K \end{bmatrix},$$
the integrated base-band signal model $\mathbf{Y} = \boldsymbol{H}\boldsymbol{X} + \mathbf{W}$ 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 $\mathbf{Y}_r = \boldsymbol{H}_r\boldsymbol{X}_r + \mathbf{W}_r$, where $\mathbf{Y}_r \in \mathbb{C}^{K\times L}$, $\boldsymbol{H}_r \in \mathbb{C}^{K\times N}$, $\boldsymbol{X}_r \in \mathbb{C}^{N\times L}$, and $\mathbf{W}_r \in \mathbb{C}^{K\times L}$; 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. □

APPENDIX C
PROOF OF LEMMA 2

Proof: We have

$$\begin{aligned}
&\max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) \\
&\quad= \max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) + \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X}) \\
&\qquad - \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) \\
&\quad\le \left| \max_{\boldsymbol{H}} \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) \right| + \\
&\qquad \left| \max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) - \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X}) \right| \\
&\quad\le \max_{\boldsymbol{H}} \left| \phi(\boldsymbol{H}, \bar{\boldsymbol{X}}) - \phi(\bar{\boldsymbol{H}}, \bar{\boldsymbol{X}}) \right| + \\
&\qquad \max_{\boldsymbol{H}} \left| \min_{\boldsymbol{X}} \phi(\boldsymbol{H}, \boldsymbol{X}) - \min_{\boldsymbol{X}} \phi(\bar{\boldsymbol{H}}, \boldsymbol{X}) \right| \\
&\quad\le L_2 \cdot \max_{\boldsymbol{H}} \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| + L_1 \cdot \max_{\boldsymbol{H}} \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \\
&\quad= (L_1 + L_2) \cdot \theta.
\end{aligned}$$

This completes the proof. □

APPENDIX D
PROOF OF THEOREM 1

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

$$\begin{aligned}
\max_{\boldsymbol{H}} \min_{\boldsymbol{X}} \quad & \|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F^2 \\
\text{s.t.} \quad & \frac{1}{L}\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}} = \boldsymbol{R}, \\
& \|\boldsymbol{H} - \bar{\boldsymbol{H}}\| \le \theta.
\end{aligned} \tag{55}$$

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,

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

This completes the proof. □

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 $\boldsymbol{H}$, 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 $\|\boldsymbol{H}_1\boldsymbol{X} - \boldsymbol{S}\|_F + \|\boldsymbol{H}_2\boldsymbol{X} - \boldsymbol{S}\|_F \le B_1$. 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 $\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|_F \leq B\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|$; the existence of B is guaranteed due to the equivalence of norms on a finite-dimensional linear space. Just note that $|\|\boldsymbol{H}_1\|_F - \|\bar{\boldsymbol{H}}\|_F| \leq \|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\|_F \leq B\|\boldsymbol{H}_1 - \bar{\boldsymbol{H}}\| \leq B\theta$. The same argument also holds for $\boldsymbol{H}_2$. Hence, for every $\boldsymbol{H}_1, \boldsymbol{H}_2 \in \mathcal{H}$, we have

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

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

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. □

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 $f(0.5H_1 + 0.5H_2) \leq 0.5f(H_1) + 0.5f(H_2)$ for some realizations of H_1 and H_2, while $f(0.5H_1 + 0.5H_2) \geq 0.5f(H_1) + 0.5f(H_2)$ for other realizations. Therefore, f is neither concave nor convex. □

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., $\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{S}\|_F \geq |\|\boldsymbol{H}\boldsymbol{X}\|_F - \|\boldsymbol{S}\|_F|$; 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 $f(\bar{\boldsymbol{H}}) = \overline{f}(\bar{\boldsymbol{H}})$, 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

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

This completes the proof. □

APPENDIX H
PROOF OF PROPOSITION 4

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

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

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. □

APPENDIX I
INTUITIVE UNDERSTANDING OF SOLUTION METHOD 1

Motivated by Lemma 1 and Condition (27), we start with employing the upper bound $\overline{f}(\boldsymbol{H})$ in Proposition 3 because $\overline{f}(\boldsymbol{H})$ 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 $\overline{f}(\boldsymbol{H})$ is that $\bar{\boldsymbol{X}}$ exists for any specified uncertainty set $\mathcal{H}$ with any radius $\theta \geq 0$ because $\mathcal{H}$ is centered at $\bar{\boldsymbol{H}}$.

Step 1. Maximize the upper bound $\overline{f}(\boldsymbol{H})$ 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:

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

b) Bound of the Nominally Optimal Cost:

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

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.,

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

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:

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

b) Bound of the Nominally Optimal Cost:

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

(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. □

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. □

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 $\langle \nabla p(\boldsymbol{y}),\ \boldsymbol{h} - \boldsymbol{y}\rangle \le 0$, for every $\boldsymbol{h}$ such that $\|\boldsymbol{h} - \bar{\boldsymbol{h}}\| \le \theta$, where $\nabla p(\boldsymbol{y})$ denotes the gradient of p evaluated at $\boldsymbol{y}$. Therefore, if we have $\max_{\boldsymbol{y}\in\mathcal{Y}} \max_{\boldsymbol{h}:\|\boldsymbol{h}-\bar{\boldsymbol{h}}\|\le\theta} \langle \nabla p(\boldsymbol{y}),\ \boldsymbol{h} - \boldsymbol{y}\rangle \le 0$, 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. □

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

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

The above display is further equivalent to

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

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. □

Appendix N
Proof of Proposition 7

Proof: Problem (43) is equivalent to

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

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 $\boldsymbol{h}^* = \theta\boldsymbol{z}^* + \bar{\boldsymbol{h}}$. □

Appendix O
Proof of Proposition 8

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

$$\min_{\boldsymbol{U}} \|\boldsymbol{U}\boldsymbol{A}_1 - \boldsymbol{B}_1\|_F^2, \quad \text{s.t.}\ \ \boldsymbol{U}\boldsymbol{U}^\mathsf{H} = \boldsymbol{I}_N, \tag{57}$$

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

$$\min_{\boldsymbol{V}} \|\boldsymbol{V}\boldsymbol{A}_2 - \boldsymbol{B}_2\|_F^2, \quad \text{s.t.}\ \ \boldsymbol{V}\boldsymbol{V}^\mathsf{H} = \boldsymbol{I}_L, \tag{58}$$

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 r^{th} iteration. Following the defined iteration process in Proposition 8, we have

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

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 r^{th} iteration, the values of $(\boldsymbol{U}^r,\boldsymbol{\Sigma}^r,\boldsymbol{V}^r)$ may not be unique. This completes the proof. □

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

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

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

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: $I = 1000$

- // ***Stage 1: Engine Initialization***
- $N \leftarrow 16$, $K \leftarrow 4$, $L \leftarrow 30$, $f \leftarrow 5.9 \times 10^9$, $P_{\mathsf{T}} \leftarrow 2.5$, $\theta \leftarrow 0 : 0.01 : 0.2$
- Generate $\boldsymbol{H}_{\text{ref}}$
- Design Perfect-Sensing Waveform $\boldsymbol{X}_s$
- // ***Stage 2: Offline Design Using Practically Available Nominal Channel***
- Generate Constellation $\boldsymbol{S}$
- Generate Nominal Channel $\bar{\boldsymbol{H}}$ Using (54)
- Design Nominally Optimal Waveform $\bar{\boldsymbol{X}}$ Using $\bar{\boldsymbol{H}}$
- Calculate Nominally Estimated AASR $R_{\bar{\boldsymbol{H}},\bar{\boldsymbol{X}}}$
- Design Robust Waveform $\boldsymbol{X}^*$ Using $\bar{\boldsymbol{H}}$ for Every θ (NB: $\boldsymbol{X}^*$ depends on θ)
- Calculate Robustly Estimated AASR $R_{\boldsymbol{H}^*,\boldsymbol{X}^*}$
- // ***Stage 3: Online Test Using Random And Practically Unknown True Channel***
- $i \leftarrow 0$
- **while** true **do**
  - // ***Calculate True AASRs***
  - Uniformly Generate $\boldsymbol{H}_0$ Using (54)
  - Calculate True AASR $R_{\boldsymbol{H}_0,\bar{\boldsymbol{X}}}$ at Nominally Optimal Waveform $\bar{\boldsymbol{X}}$
  - Calculate True AASR $R_{\boldsymbol{H}_0,\boldsymbol{X}^*}$ at Robust Waveform $\boldsymbol{X}^*$ for Every θ (NB: $\boldsymbol{X}^*$ depends on θ)
  - // ***Next Episode***
  - $i \leftarrow i + 1$
  - // ***End of Simulation***
  - **if** $i > I$ **then**
    - **break while**
  - **end if**
- **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.