**SCIENCE CHINA**
Information Sciences

• **Supplementary File** •

# Probabilistic Constrained Robust Secure Transmission for Wireless Powered Heterogeneous Networks

KaiZhi HUANG, Bo ZHANG*, Zhou ZHONG, YaJun CHEN & XiaoMing XU

*National Digital Switching System Engineering & Technological R&D Center, Information Engineering University, Zhenzhou* 450002*, china*

## Appendix A The received signal at $\mathrm{MU}_m$, $\mathrm{FU}_{nk}$ and $\mathrm{E}_b$

The received signal at $\mathrm{MU}_m$ is given by

$$\mathbf{y}_m = \mathbf{h}_m^H \mathbf{w}_m s_m + \mathbf{h}_m^H (\sum_{p\neq m}^{M} \mathbf{w}_p s_p + \mathbf{z}_0) + \sum_{n=1}^{N} \mathbf{h}_{n,m}^H (\sum_{k=1}^{K} \mathbf{w}_{nk} s_{nk} + \mathbf{z}_n) + n_m \tag{A1}$$

where $m \in [1, M]$, $\mathbf{h}_m \in C^{N_M \times 1}$ and $\mathbf{h}_{n,m} \in C^{N_F \times 1}$ denote the channel vectors from the MBS to $\mathrm{MU}_m$ and that from the $\mathrm{FBS}_n$ to $\mathrm{MU}_m$, respectively. $\mathbf{w}_f \in C^{N_M \times 1}(f = m, p)$ and $\mathbf{w}_{nk} \in C^{N_F \times 1}$ are the beamforming vectors from the MBS to $\mathrm{MU}_f$ and that from the $\mathrm{FBS}_n$ to $\mathrm{FU}_{nk}$, respectively. $\mathrm{s}_i(i = m, p)$ and $s_{nj}(j = k, t)$ stand for the message symbols from the MBS to $\mathrm{MU}_i$ and that from the $\mathrm{FBS}_n$ to $\mathrm{FU}_{nj}$, respectively, $\mathrm{E}\{\ |\mathrm{s}_i|^2\} = 1$, $\mathrm{E}\{\ |\mathrm{s}_{nj}|^2\} = 1$. $\mathbf{z}_0 \in C^{N_M \times 1}$ and $\mathbf{z}_n \in C^{N_F \times 1}$ represent the AN vectors at the MBS and $\mathrm{FBS}_n$, which follow $CN(0, \mathbf{Z}_0)$ and $CN(0, \mathbf{Z}_n)$, respectively. $\mathbf{Z}_0$ and $\mathbf{Z}_n$ indicate the corresponding covariance matrices of $\mathbf{z}_0$ and $\mathbf{z}_n$ [1], respectively. $n_m$ denotes the additive white Gaussian noise (AWGN) at $\mathrm{MU}_m$.

The received signal at $\mathrm{FU}_{nk}$ is expressed as

$$\mathbf{y}_{nk} = \mathbf{h}_{n,nk}^H \mathbf{w}_{nk} s_{nk} + \mathbf{h}_{n,nk}^H (\sum_{t\neq k}^{K} \mathbf{w}_{nt} s_{nt} + \mathbf{z}_n) + \sum_{a\neq n}^{N} \mathbf{h}_{a,nk}^H (\sum_{t=1}^{K} \mathbf{w}_{at} s_{at} + \mathbf{z}_a) + \mathbf{h}_{nk}^H (\sum_{m=1}^{M} \mathbf{w}_m s_m + \mathbf{z}_0) + n_m \tag{A2}$$

where $n \in [1, N]$, $k \in [1, K]$, $\mathbf{h}_{l,nk} \in C^{N_F \times 1}(l = n, a)$ and $\mathbf{h}_{nk} \in C^{N_M \times 1}$ denote the channel vectors from the $\mathrm{FBS}_l$ to $\mathrm{FU}_{nk}$ and that from the MBS to $\mathrm{FU}_{nk}$, respectively. $n_{nk}$ represents the AWGN at $\mathrm{FU}_{nk}$.

The received signal at $\mathrm{E}_b$ is given by

$$\mathbf{y}_{Eb} = \mathbf{h}_{Eb}^H \mathbf{w}_1 s_1 + \mathbf{h}_{Eb}^H (\sum_{m=1}^{M} \mathbf{w}_m s_m + \mathbf{z}_0) + \sum_{n=1}^{N} \mathbf{h}_{n,Eb}^H (\sum_{t=1}^{K} \mathbf{w}_{nk} s_{nk} + \mathbf{z}_n) + n_{Eb}, b \in [1, B] \tag{A3}$$

where $\mathbf{h}_{Eb} \in C^{N_M \times 1}$ and $\mathbf{h}_{n,Eb} \in C^{N_F \times 1}$ denote the channel vectors from the MBS to $\mathrm{E}_b$ and that from the $\mathrm{FBS}_n$ to $\mathrm{E}_b$, respectively. $n_{Eb}$ denotes the AWGN at $\mathrm{E}_b$. For the sake of simplification, we assume all the AWGN follow the independent and identical distribution, i.e., $CN(0, \sigma^2)$.

## Appendix B The SINR of $\mathrm{MU}_m$, $\mathrm{FU}_{nk}$ and $\mathrm{E}_b$

Based on (A1)∼(A3), the signal to interference plus noise ratio (SINR) of $\mathrm{MU}_m$, $\mathrm{FU}_{nk}$ and $\mathrm{E}_b$ can be respectively expressed as

$$\mathrm{SINR}_m = \frac{\left|\mathbf{h}_m^H \mathbf{w}_m\right|^2}{\sum\limits_{p\neq m}^{M} |\mathbf{h}_m^H \mathbf{w}_p|^2 + |\mathbf{h}_m^H \mathbf{z}_0|^2 + \sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} \left|\mathbf{h}_{n,m}^H \mathbf{w}_{nk}\right|^2 + \sum\limits_{n=1}^{N} (\left|\mathbf{h}_{n,m}^H \mathbf{z}_n\right|^2) + \sigma^2} \tag{B1}$$

$$\mathrm{SINR}_{nk} = \frac{\left|\mathbf{h}_{n,nk}^H \mathbf{w}_{nk}\right|^2}{A_{nk}} \tag{B2}$$

$$\mathrm{SINR}_{Eb} = \frac{\left|\mathbf{h}_{Eb}^H \mathbf{w}_1\right|^2}{\sum\limits_{m=2}^{M} \left|\mathbf{h}_{Eb}^H \mathbf{w}_m\right|^2 + \left|\mathbf{h}_{Eb}^H \mathbf{z}_0\right|^2 + \sum\limits_{n=1}^{N}\sum\limits_{k=1}^{K} \left|\mathbf{h}_{n,Eb}^H \mathbf{w}_{nk}\right|^2 + \sum\limits_{n=1}^{N} (\left|\mathbf{h}_{n,Eb}^H \mathbf{z}_n\right|^2) + \sigma^2} \tag{B3}$$

where

$$A_{nk} = \sum_{t\neq k}^{K} \left|\mathbf{h}_{n,nk}^H \mathbf{w}_{nt}\right|^2 + \left|\mathbf{h}_{n,nk}^H \mathbf{z}_n\right|^2 + \sum_{a\neq n}^{N}\sum_{t=1}^{K} \left|\mathbf{h}_{a,nk}^H \mathbf{w}_{at}\right|^2 + \sum_{a\neq n}^{N} (\left|\mathbf{h}_{a,nk}^H \mathbf{z}_a\right|^2) + \sum_{m=1}^{M} \left|\mathbf{h}_{nk}^H \mathbf{w}_m\right|^2 + \left|\mathbf{h}_{nk}^H \mathbf{z}_0\right|^2 + \sigma^2 \tag{B4}$$

* Corresponding author (email: zb100403@126.com)

Similar to [2], we assume that ERs can harvest energy from three parts: information beams, energy and AN signal of FBSs. Therefore, the harvested energy at $E_b$ is given by

$$E_{harvest}(b) = \kappa(\sum_{n=1}^{N}\sum_{k=1}^{K}\left|\mathbf{h}_{n,\mathrm{Eb}}^{H}\mathbf{w}_{nk}\right|^{2} + \sum_{n=1}^{N}(\left|\mathbf{h}_{n,\mathrm{Eb}}^{H}\mathbf{z}_{n}\right|^{2})) \tag{B5}$$

where $\kappa \in [0,1]$ is the coefficient of energy conversion.

## Appendix C The actual channels in (A1)∼(A3) of appendix A

In practical network, it is difficult for the MBS and FBSs to obtain the perfect CSI of corresponding users due to the channel estimation and feedback errors. In this paper, we consider the stochastic CSI error model, where CSI error is stochastic and follows a certain distribution [3]-[5]. Then, the actual channels in (A1)∼(A3) can be respectively expressed in (C1), where $\mathbf{h}$, $\hat{\mathbf{h}}$ and $\mathbf{e}$ denote actual channel vector, corresponding estimated channel vector and corresponding channel error vector, respectively. From (C1), we can observe that each channel error vector follows a complex-valued Gaussian distribution. Thus, each channel error vector can be respectively expressed as the following [24]

$$\mathbf{h}_m = \hat{\mathbf{h}}_m + \mathbf{e}_m, \mathbf{e}_m \sim CN(0, \mathbf{Q}_m) \tag{C1a}$$

$$\mathbf{h}_{n,m} = \hat{\mathbf{h}}_{n,m} + \mathbf{e}_{n,m}, \mathbf{e}_{n,m} \sim CN(0, \mathbf{Q}_{n,m}) \tag{C1b}$$

$$\mathbf{h}_{l,nk} = \hat{\mathbf{h}}_{l,nk} + \mathbf{e}_{l,nk}, \mathbf{e}_{l,nk} \sim CN(0, \mathbf{Q}_{l,nk}) \tag{C1c}$$

$$\mathbf{h}_{nk} = \hat{\mathbf{h}}_{nk} + \mathbf{e}_{nk}, \mathbf{e}_{nk} \sim CN(0, \mathbf{Q}_{nk}) \tag{C1d}$$

$$\mathbf{h}_{\mathrm{Eb}} = \hat{\mathbf{h}}_{\mathrm{Eb}} + \mathbf{e}_{\mathrm{Eb}}, \mathbf{e}_{Eb} \sim CN(0, \mathbf{Q}_{Eb}) \tag{C1e}$$

$$\mathbf{h}_{n,\mathrm{Eb}} = \hat{\mathbf{h}}_{n,\mathrm{Eb}} + \mathbf{e}_{n,\mathrm{Eb}}, \mathbf{e}_{n,Eb} \sim CN(0, \mathbf{Q}_{n,Eb}) \tag{C1f}$$

where $\mathbf{Q}$ denotes corresponding covariance matrix of channel error vector, $\mathbf{I}_N$ is an unit matrix of $N \times N$, $\mathbf{Q}_m \in C^{N_M \times N_M}$, $\mathbf{Q}_{n,m} \in C^{N_F \times N_F}$, $\mathbf{Q}_{l,nk} \in C^{N_F \times N_F}$, $\mathbf{Q}_{nk} \in C^{N_M \times N_M}$, $\mathbf{Q}_{Eb} \in C^{N_M \times N_M}$, $\mathbf{Q}_{n,Eb} \in C^{N_F \times N_F}$. Meanwhile, the following expressions

$$\mathbf{e}_m = \mathbf{Q}_m{}^{1/2}\mathbf{r}_m, \mathbf{r}_m \sim CN(0, \mathbf{I}_{N_M}) \tag{C2a}$$

$$\mathbf{e}_{n,m} = \mathbf{Q}_{n,m}{}^{1/2}\mathbf{r}_{n,m}, \mathbf{r}_{n,m} \sim CN(0, \mathbf{I}_{N_F}) \tag{C2b}$$

$$\mathbf{e}_{l,nk} = \mathbf{Q}_{l,nk}{}^{1/2}\mathbf{r}_{l,nk}, \mathbf{r}_{l,nk} \sim CN(0, \mathbf{I}_{N_F}) \tag{C2c}$$

$$\mathbf{e}_{nk} = \mathbf{Q}_{nk}{}^{1/2}\mathbf{r}_{nk}, \mathbf{r}_{nk} \sim CN(0, \mathbf{I}_{N_M}) \tag{C2d}$$

$$\mathbf{e}_{Eb} = \mathbf{Q}_{Eb}{}^{1/2}\mathbf{r}_{Eb}, \mathbf{r}_{Eb} \sim CN(0, \mathbf{I}_{N_M}) \tag{C2e}$$

hold on.

## Appendix D The formulation and solution of optimization problem

In this paper, our interest is to study the joint information beamforming, energy transfer and AN design under the stochastic CSI error case, such that the total transmit power of system is minimized while satisfying the probabilistic QoS constraints on MUs, FUs and Eves, and the probabilistic EH constraints on ERs. Hence, the problem of minimizing the total transmit power can be formulated as

$$\min_{\mathbf{W}_m,\mathbf{Z}_0\mathbf{W}_{nk},\mathbf{Z}_n,\mathbf{U}_n} \left\{\sum_{m=1}^{M}\mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N}\left\{\left\{\sum_{k=1}^{K}\mathrm{Tr}(\mathbf{W}_{nk})\right\} + \mathrm{Tr}(\mathbf{Z}_n)\right\}\right\} \tag{D1a}$$

$$s.t. \quad \Pr\{\mathrm{SINR}_m \leqslant \gamma_m\} \leqslant \rho_{\mathrm{m}}, m \in [1, M] \tag{D1b}$$

$$\Pr\{\mathrm{SINR}_{nk} \leqslant \gamma_{nk}\} \leqslant \rho_{nk}, n \in [1, N], k \in [1, K] \tag{D1c}$$

$$\Pr\{\mathrm{SINR}_{Eb} \geqslant \gamma_{Eb}\} \leqslant \rho_{Eb}, b \in [1, B] \tag{D1d}$$

$$\Pr\{\mathrm{E}_{harvest}(b) \leqslant \phi_b\} \leqslant q_{Eb}, b \in [1, B] \tag{D1e}$$

$$\mathrm{Rank}(\mathbf{W}_m) = \mathrm{Rank}(\mathbf{W}_{nk}) = 1, m \in [1, M], n \in [1, N], k \in [1, K] \tag{D1f}$$

where $\gamma_m$, $\gamma_{nk}$, $\gamma_{Eb}$, $\phi_b$ denote the predefined SINR thresholds of $\mathrm{MU}_m$, $\mathrm{FU}_{nk}$, $\mathrm{E}_b$ and EH threshold of $\mathrm{E}_b$, respectively. $\rho_{\mathrm{m}} \in (0,1]$, $\rho_{nk} \in (0,1]$, $\rho_{Eb} \in (0,1]$ and $q_{Eb} \in (0,1]$ denote the QoS probability threshold at $\mathrm{MU}_m$, $\mathrm{FU}_{nk}$, $\mathrm{E}_b$ and the EH probability threshold at $\mathrm{E}_b$, respectively. The secrecy rate of system can be expressed as

$$R_s = \{\log_2(1 + \mathrm{SINR}_1) - \max_{b\in[1,B]} \log_2(1 + \mathrm{SINR}_{\mathrm{E}b})\}^{+} \tag{D2}$$

where $\{a\}^{+}$ denotes $\max\{a, 0\}$. From (D2), we can observe that the secrecy rate is increasing with $\mathrm{SINR}_1$, however, decreasing with $\mathrm{SINR}_{\mathrm{E}b}$ monotonically. Therefore, the constraints (D1b) and (D1c) together can guarantee the required lower bound on secrecy rate by

$$R_s \geqslant \{\log_2(1 + \gamma_1) - \max_{b\in[1,B]} \log_2(1 + \gamma_{\mathrm{E}b})\}^{+} \tag{D3}$$

In other words, the required lower bound on secrecy rate can be guaranteed through adjusting $\gamma_1$ and $\gamma_{\mathrm{E}b}$ to achieve secure communication for the system. Meanwhile, (D1b), (D1c) and (D1e) can guarantee the reliability of system together. In summary, the security and reliability of system are guaranteed when the constraints (D1b)∼(D1e) are satisfied. However, due to the non-convexity and the fact that probabilistic functions have no closed-form expressions, the problem (D1) is hard to solve. Meanwhile, due to the existence of cross-tier interference, the former study on the scenario of single-BS with CSI uncertainty can not be applied to solve it. In next section, we will propose two different methods to cope with this problem.

**D.1 Equivalent transformation for original problem**

According to the above analysis, we can see that the probabilistic constraints (D1b)∼(D1e) have a similar structure. Therefore, we will adopt the similar way to deal with them. In the next, we first deal with (D1b), which can be instructive for the processing of (D1c)∼(D1e). For the sake of simplifying the transformation of (D1b), we define the following expressions

$$\begin{cases} \mathbf{h}_{Mm} = [(\hat{\mathbf{h}}_m + \mathbf{e}_m)^H, (\hat{\mathbf{h}}_{1,m} + \mathbf{e}_{1,m})^H, \cdots, (\hat{\mathbf{h}}_{N,m} + \mathbf{e}_{N,m})^H]^H \\ \hat{\mathbf{h}}_{Mm} = [(\hat{\mathbf{h}}_m)^H, (\hat{\mathbf{h}}_{1,m})^H, \cdots, (\hat{\mathbf{h}}_{N,m})^H]^H \\ \mathbf{e}_{Mm} = [(\mathbf{e}_m)^H, (\mathbf{e}_{1,m})^H, \cdots, (\mathbf{e}_{N,m})^H]^H \\ \mathbf{A}_{Mm} = diag[(\mathbf{W}_m - \gamma_m \sum_{p \neq m}^{M} (\mathbf{W}_p + \mathbf{Z}_0)), -\gamma_m((\mathbf{W}_{1k} + \mathbf{Z}_1), \cdots, (\mathbf{W}_{Nk} + \mathbf{Z}_N))] \end{cases} \tag{D4}$$

where $diag[A, B] = \begin{bmatrix} A & 0_{N\times M} \\ 0_{M\times N} & B \end{bmatrix}$, $A \in C^{N\times N}$, $B \in C^{M\times M}$. Then, (D1b) can be transformed as

$$\Pr\{\mathbf{r}_{Mm}^H \mathbf{A}_{Mm} \mathbf{r}_{Mm} + 2\,\mathrm{Re}\{\mathbf{e}_{Mm}^H \mathbf{A}_{Mm} \hat{\mathbf{h}}_{Mm} + \mathbf{e}_{Mm}^H \mathbf{A}_{Mm} \mathbf{e}_{Mm} - \gamma_m \leqslant 0\} \leqslant \rho_m, m \in [1, M] \tag{D5}$$

Combining (C1), we further define the following expressions

$$\begin{cases} \mathbf{e}_{Mm} = \mathbf{Q}_{Mm} \mathbf{r}_{Mm} \\ \mathbf{Q}_{Mm} = diag[{\mathbf{Q}_m}^{1/2}, {\mathbf{Q}_{1,m}}^{1/2}, \cdots, {\mathbf{Q}_{N,m}}^{1/2}] \\ \mathbf{r}_{Mm} = [(\mathbf{r}_m)^H, (\mathbf{r}_{1,m})^H, \cdots, (\mathbf{r}_{N,m})^H]^H \end{cases} \tag{D6}$$

Due to the independence between different channels, we can acquire $\mathbf{r}_{Mm} \sim CN(0, \mathbf{I}_{(N_M+N\times N_F)})$ from $\mathbf{r}_m \sim CN(0, \mathbf{I}_{N_M})$ and $\mathbf{r}_{n,m} \sim CN(0, \mathbf{I}_{N_F})$. Then, (D5) can be equivalently transformed as

$$\Pr\{\mathbf{r}_{Mm}^H \mathbf{B}_{Mm} \mathbf{r}_{Mm} + 2\,\mathrm{Re}\{\mathbf{r}_{Mm}^H \mathbf{c}_{Mm}\} \leqslant d_{Mm}\} \leqslant \rho_m, m \in [1, M] \tag{D7}$$

where

$$\begin{cases} \mathbf{B}_{Mm} = \mathbf{Q}_{Mm} \mathbf{A}_{Mm} \mathbf{Q}_{Mm} \\ \mathbf{c}_{Mm} = \mathbf{Q}_{Mm} \mathbf{A}_{Mm} \hat{\mathbf{h}}_{Mm} \\ d_{Mm} = \gamma_m * \sigma^2 - \hat{\mathbf{h}}_{Mm}^H \mathbf{A}_{Mm} \hat{\mathbf{h}}_{Mm} \end{cases} \tag{D8}$$

Similarly, (D1c)∼(D1e) can be respectively transformed as

$$\Pr\{\mathbf{r}_{Fnk}^H \mathbf{B}_{Fnk} \mathbf{r}_{Fnk} + 2\,\mathrm{Re}\{\mathbf{r}_{Fnk}^H \mathbf{c}_{Fnk}\} \leqslant d_{Fnk}\} \leqslant \rho_{nk}, n \in [1, N], k \in [1, K] \tag{D9}$$

$$\Pr\{\mathbf{r}_{EEb}^H \mathbf{B}_{EEb} \mathbf{r}_{EEb} + \mathrm{Re}\{\mathbf{r}_{EEb}^H \mathbf{c}_{EEb}\} \geqslant d_{EEb}\} \leqslant \rho_{EEb}, b \in [1, B] \tag{D10}$$

$$\Pr\{\mathbf{r}_{CEb}^H \mathbf{B}_{CEb} \mathbf{r}_{CEb} + 2\,\mathrm{Re}\{\mathbf{r}_{CEb}^H \mathbf{c}_{CEb}\} \leqslant d_{CEb}\} \leqslant q_{Eb}, b \in [1, B] \tag{D11}$$

The corresponding variables in (D9)∼(D11) are respectively recast as

$$\begin{cases} \mathbf{r}_{Fnk} = \left[(\mathbf{r}_{n,nk})^H, (\mathbf{r}_{nk})^H, \cdots, (\mathbf{r}_{n-1,nk})^H, (\mathbf{r}_{n+1,nk})^H, \cdots\right]^H \\ \mathbf{B}_{Fnk} = \mathbf{Q}_{Fnk} \mathbf{A}_{Fnk} \mathbf{Q}_{Fnk} \\ \mathbf{c}_{Fnk} = \mathbf{Q}_{Fnk} \mathbf{A}_{Fnk} \hat{\mathbf{h}}_{Fnk} \\ d_{Fnk} = \gamma_{nk} * \sigma^2 - (\hat{\mathbf{h}}_{Fnk})^H \mathbf{A}_{Fnk} \hat{\mathbf{h}}_{Fnk} \\ \mathbf{A}_{Fnk} = diag((\mathbf{W}_{nk} - \gamma_{nk}(\sum_{t=1,t\neq k}^{K} \mathbf{W}_{nt} + \mathbf{Z}_n)), (-\gamma_{nk}(\sum_{m=1}^{M} \mathbf{W}_m + \mathbf{Z}_0)), \\ -\gamma_{nk}(\cdots, (\mathbf{W}_{n-1t} + \mathbf{Z}_{n-1}), (\mathbf{W}_{n+1t} + \mathbf{Z}_{n+1}), \cdots) \\ \mathbf{Q}_{Fnk} = diag[{\mathbf{Q}_{n,nk}}^{1/2}, {\mathbf{Q}_{nk}}^{1/2} \cdots, {\mathbf{Q}_{n-1,nk}}^{1/2}, {\mathbf{Q}_{n+1,nk}}^{1/2}, \cdots] \\ \hat{\mathbf{h}}_{Fnk} = \left[(\hat{\mathbf{h}}_{n,nk})^H, (\hat{\mathbf{h}}_{nk})^H, \cdots, (\hat{\mathbf{h}}_{n-1,nk})^H, (\hat{\mathbf{h}}_{n+1,nk})^H, \cdots\right]^H \end{cases} \tag{D12}$$

$$\begin{cases} \mathbf{r}_{EEb} = [(\mathbf{r}_{Eb})^H, (\mathbf{r}_{1,Eb})^H, \cdots, (\mathbf{r}_{N,Eb})^H]^H \\ \mathbf{B}_{EEb} = \mathbf{Q}_{EEb}\mathbf{A}_{EEb}\mathbf{Q}_{EEb} \\ \mathbf{c}_{EEb} = \mathbf{Q}_{EEb}\mathbf{A}_{EEb}\underset{EEb}{\hat{\mathbf{h}}} \\ d_{EEb} = \gamma_{Eb} * \sigma^2 - (\hat{\mathbf{h}}_{EEb})^H \mathbf{A}_{EEb}\hat{\mathbf{h}}_{EEb} \\ \mathbf{Q}_{EEb} = diag[\mathbf{Q}_{Eb}{}^{1/2}, \mathbf{Q}_{1,Eb}{}^{1/2} \cdots, \mathbf{Q}_{N,Eb}{}^{1/2}] \\ \mathbf{A}_{EEb} = diag((\mathbf{W}_1 - \gamma_{Eb}(\sum_{m=2}^{M}\mathbf{W}_m + \mathbf{Z}_0)), -\gamma_{Eb}(\mathbf{W}_{1k} + \mathbf{Z}_1), \\ \cdots, -\gamma_{Eb}(\mathbf{W}_{Nk} + \mathbf{Z}_N)) \\ \hat{\mathbf{h}}_{EEb} = \left[(\hat{\mathbf{h}}_{Eb})^H, (\hat{\mathbf{h}}_{1,Eb})^H, \cdots, (\hat{\mathbf{h}}_{N,Eb})^H\right]^H \end{cases} \tag{D13}$$

$$\begin{cases} \mathbf{r}_{CEb} = [(\mathbf{r}_{1,Eb})^H, \cdots, (\mathbf{r}_{N,Eb})^H]^H \\ \mathbf{B}_{CEb} = \mathbf{Q}_{CEb}\mathbf{A}_{CEb}\mathbf{Q}_{CEb} \\ \mathbf{c}_{CEb} = \mathbf{Q}_{CEb}\mathbf{A}_{CEb}\hat{\mathbf{h}}_{CEb} \\ d_{CEb} = \phi_b/\kappa * \sigma^2 - (\hat{\mathbf{h}}_{CEb})^H \mathbf{A}_{CEb}\hat{\mathbf{h}}_{CEb} \\ \mathbf{Q}_{CEb} = diag(\mathbf{Q}_{1,Eb}{}^{1/2}, \cdots, \mathbf{Q}_{N,Eb}{}^{1/2}) \\ \mathbf{A}_{CEb} = diag((\mathbf{W}_{1k} + \mathbf{Z}_1), \cdots, (\mathbf{W}_{Nk} + \mathbf{Z}_N)) \\ \hat{\mathbf{h}}_{CEb} = \left[(\hat{\mathbf{h}}_{Eb})^H, (\hat{\mathbf{h}}_{1,Eb})^H, \cdots, (\hat{\mathbf{h}}_{N,Eb})^H\right]^H \end{cases} \tag{D14}$$

where $\mathbf{r}_{Fnk} \sim CN(0, \mathbf{I}_{(N_M+N\times N_F)})$, $\mathbf{r}_{EEb} \sim CN(0, \mathbf{I}_{(N_M+N\times N_F)})$, $\mathbf{r}_{CEb} \sim CN(0, \mathbf{I}_{(N\times N_F)})$. Up to now, the problem (D1) can be equivalently transformed as

$$\min_{\mathbf{W}_m,\mathbf{Z}_0,\mathbf{W}_{nk},\mathbf{Z}_n} \left\{ \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N} \left\{ \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_{nk}) + \mathrm{Tr}(\mathbf{Z}_n) \right\} \right\} \tag{D15a}$$

$$s.t.(D7), (D9) \sim (D11), (D1f) \tag{D15b}$$

However, the problem (D15) is still hard to solve due to the probabilistic constraints (D7), (D9)∼(D11). In next section, we will propose two different methods to deal with them.

**D.2 Conservative Reformulation by Bernstein-type inequality**

Due to the probabilistic form of probabilistic constraints (D7), (D9)∼(D11), they can not be utilized to solve the problem (D15) directly. Observing their form, we find that the framework of Bernstein-type inequality can be applied to acquire the solvable form of them. In this regard, we introduce the following Lemma 1 [6].

**Lemma 1 (Bernstein-type inequality):** For any $\mathbf{D} \in \mathrm{H}^N$, $\mathbf{d} \in C^N$, $\mathbf{r} \sim CN(0, \mathbf{I}_N)$, and $\alpha \geqslant 0$, we have

$$\Pr\{\mathbf{r}^H\mathbf{D}\mathbf{r} + 2\,\mathrm{Re}\{\mathbf{r}^H\mathbf{d}\} \leqslant \mathrm{Tr}(\mathbf{D}) - \sqrt{2\alpha(\|\mathrm{vec}(\mathbf{D})\|^2 + 2\|\mathbf{d}\|^2)} - \alpha s^-(\mathbf{D})\} \leqslant \exp(-\alpha) \tag{D16a}$$

$$\Pr\{\mathbf{r}^H\mathbf{D}\mathbf{r} + 2\,\mathrm{Re}\{\mathbf{r}^H\mathbf{d}\} \geqslant \mathrm{Tr}(\mathbf{D}) + \sqrt{2\alpha(\|\mathrm{vec}(\mathbf{D})\|^2 + 2\|\mathbf{d}\|^2)} + \alpha s^+(\mathbf{D})\} \leqslant \exp(-\alpha) \tag{D16b}$$

where $s^-(\mathbf{D}) = \max\{\ \lambda_{\max}(\ -\mathbf{D}), 0\}$ , $s^+(\mathbf{D}) = \max\{\ \lambda_{\max}(\mathbf{D}), 0\}$ , and $\lambda_{\max}(\mathbf{D})$ represents the maximum eigenvalue of the matrix $\mathbf{D}$.

Combining (D16a) and introducing an auxiliary variable $\lambda_{Mm} = -\ln(\rho_m)$ into (D17), we can acquire the following expression

$$\Pr\{\mathbf{r}_{Mm}^H\mathbf{B}_{Mm}\mathbf{r}_{Mm} + 2\,\mathrm{Re}\{\mathbf{r}_{Mm}^H\mathbf{c}_{Mm}\} \leqslant \mathrm{Tr}(\mathbf{B}_{Mm}) - \sqrt{2\lambda_{Mm}(\|\mathrm{vec}(\mathbf{B}_{Mm})\|^2 + 2\|\mathbf{c}_{Mm}\|^2)} - \lambda_{Mm}s^-(\mathbf{B}_{Mm})\} \leqslant \rho_m \tag{D17}$$

Comparing (D17) with (D7), we can see that the constraint (D7) must be satisfied when the following constraint

$$d_{Mm} \leqslant \mathrm{Tr}(\mathbf{B}_{Mm}) - \sqrt{2\alpha_{Mm}(\|\mathrm{vec}(\mathbf{B}_{Mm})\|^2 + 2\|\mathbf{c}_{Mm}\|^2)} - \lambda_{Mm}s^-(\mathbf{B}_{Mm}), m \in [1, M] \tag{D18}$$

holds true.

Similarly, we respectively apply the Bernstein-type inequality to (D9)∼(D11), and obtain the corresponding conservative reformulations as

$$d_{Fnk} \leqslant \mathrm{Tr}(\mathbf{B}_{Fnk}) - \sqrt{2\lambda_{Fnk}(\|\mathrm{vec}(\mathbf{B}_{Fnk})\|^2 + 2\|\mathbf{c}_{Fnk}\|^2)} - \lambda_{Fnk}s^-(\mathbf{B}_{Fnk})\} \tag{D19}$$

$$d_{EEb} \geqslant \mathrm{Tr}(\mathbf{B}_{EEb}) + \sqrt{2\lambda_{EEb}(\|\mathrm{vec}(\mathbf{B}_{EEb})\|^2 + 2\|\mathbf{c}_{EEb}\|^2)} + \lambda_{EEb}s^+(\mathbf{B}_{EEb}) \tag{D20}$$

$$d_{CEb} \leqslant \mathrm{Tr}(\mathbf{B}_{CEb}) - \sqrt{2\lambda_{CEb}(\|\mathrm{vec}(\mathbf{B}_{CEb})\|^2 + 2\|\mathbf{c}_{CEb}\|^2)} - \lambda_{CEb}s^-(\mathbf{B}_{CEb}) \tag{D21}$$

where $\lambda_{Fnk} = -\ln(\rho_{nk})$, $\lambda_{EEb} = -\ln(\rho_{Eb})$, $\lambda_{CEb} = -\ln(q_{Eb})$.

Furthermore, the constraints (D19)∼(D21) can be equivalently reformulated as

$$\begin{cases} \mathrm{Tr}(\mathbf{B}_{Mm}) - u_{Mm}\sqrt{2\lambda_{Mm}} - \lambda_{Mm}v_{Mm} \geqslant d_{Mm} \\ \left\| \begin{matrix} \mathrm{vec}(\mathbf{B}_{Mm}) \\ \sqrt{2}\mathbf{c}_{Mm} \end{matrix} \right\| \leqslant u_{Mm} \\ v_{Mm}\mathbf{I}_{N_M+NN_F} + \mathbf{B}_{Mm} \geqslant 0 \\ v_{Mm} \geqslant 0 \end{cases} \tag{D22}$$

$$\begin{cases} \mathrm{Tr}(\mathbf{B}_{Fnk}) - u_{Fnk}\sqrt{2\lambda_{Fnk}} - \lambda_{Fnk}v_{Fnk} \geqslant d_{Fnk} \\ \left\| \begin{matrix} \mathrm{vec}(\mathbf{B}_{Fnk}) \\ \sqrt{2}\mathbf{c}_{Fnk} \end{matrix} \right\| \leqslant u_{Fnk} \\ v_{Fnk}\mathbf{I}_{N_M+NN_F} + \mathbf{B}_{Fnk} \geqslant 0 \\ v_{Fnk} \geqslant 0 \end{cases} \tag{D23}$$

$$\begin{cases} \mathrm{Tr}(\mathbf{B}_{EEb}) + u_{EEb}\sqrt{2\lambda_{EEb}} + \lambda_{EEb}v_{EEb} \leqslant d_{EEb} \\ \left\| \begin{matrix} \mathrm{vec}(\mathbf{B}_{EEb}) \\ \sqrt{2}\mathbf{c}_{EEb} \end{matrix} \right\| \leqslant u_{EEb} \\ v_{EEb}\mathbf{I}_{N_M+NN_F} - \mathbf{B}_{EEb} \geqslant 0 \\ v_{EEb} \geqslant 0 \end{cases} \tag{D24}$$

$$\begin{cases} \mathrm{Tr}(\mathbf{B}_{CEb}) - u_{CEb}\sqrt{2\lambda_{CEb}} - \lambda_{CEb}v_{CEb} \geqslant d_{CEb} \\ \left\| \begin{matrix} \mathrm{vec}(\mathbf{B}_{CEb}) \\ \sqrt{2}\mathbf{c}_{CEb} \end{matrix} \right\| \leqslant u_{CEb} \\ v_{CEb}\mathbf{I}_{NN_F} + \mathbf{B}_{CEb} \geqslant 0 \\ v_{CEb} \geqslant 0 \end{cases} \tag{D25}$$

where $u_{Mm}$, $v_{Mm}$, $u_{Fnk}$, $v_{Fnk}$, $u_{EEb}$, $v_{EEb}$, $u_{CEb}$, $v_{CEb}$ are auxiliary variables.

Up to now, the problem (D15) can be reformulated as

$$\min_{\mathbf{W}_m,\mathbf{Z}_0,\mathbf{W}_{nk},\mathbf{Z}_n} \left\{ \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N} \left\{ \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_{nk}) + \mathrm{Tr}(\mathbf{Z}_n) \right\} \right\} \tag{D26a}$$

$$s.t. (D22) \sim (D25), (D1f) \tag{D26b}$$

Due to the rank one constraint (D1f), the problem (D26) is still non-convex. Employing the semi-definite relaxation (SDR) technique, the problem (D26) becomes a convex semi-definite program (SDP) problem by abandoning the rank one constraint, which can be solved efficiently by the interior point method. It should be noted that the problem (D26) can be relaxed when the rank of $\mathbf{W}_m$ and $\mathbf{W}_{nk}$ are one. However, the optimal solutions of the problem (D26) may be high-dimension. To obtain the rank-one solutions of the problem (D26), we reformulate it as

$$\min_{\mathbf{W}_m,\mathbf{Z}_0,\mathbf{W}_{nk},\mathbf{Z}_n} \left\{ \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N} \left\{ \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_{nk}) + \mathrm{Tr}(\mathbf{Z}_n) \right\} \right\} \tag{D27a}$$

$$s.t. (D22) \sim (D25) \tag{D27b}$$

$$\mathbf{W}_m = \mathbf{w}_m\mathbf{w}_m^H, \mathbf{W}_{nk} = \mathbf{w}_{nk}\mathbf{w}_{nk}^H \tag{D27c}$$

Constraint (D27c) is a quadratic equality constraint. In this regard, we introduce the following Lemma 2 [7].

**Lemma 2:** The equality constraint (i.e., $\mathbf{W} = \mathbf{w}\mathbf{w}^H$) is equivalent to the following constraint

$$\begin{cases} \left\| \begin{bmatrix} \mathbf{A}_1 & \mathbf{W} & \mathbf{w} \\ \mathbf{W}^H & \mathbf{A}_2 & \mathbf{w} \\ \mathbf{w}^H & \mathbf{w}^H & 1 \end{bmatrix} \right\| \geqslant 0 \\ \mathrm{Tr}(\mathbf{A}_1 - \mathbf{w}\mathbf{w}^H) \leqslant 0 \end{cases} \tag{D28}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the newly introduced slack matrices.

Applying Lemma 2, we can approximate the problem (D27) as

$$\min_{\mathbf{w}_m,\mathbf{Z}_0,\mathbf{w}_{nk},\mathbf{Z}_n} \left\{ \sum_{m=1}^{M} \mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N} \left\{ \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_{nk}) + \mathrm{Tr}(\mathbf{Z}_n) \right\} \right\} \tag{D29a}$$

$$s.t. (D22) \sim (D25) \tag{D29b}$$

$$\left\| \begin{bmatrix} \mathbf{A}_{BS1m} & \mathbf{W}_m & \mathbf{w}_m \\ \mathbf{W}_m^H & \mathbf{A}_{BS2m} & \mathbf{w}_m \\ \mathbf{w}_m^H & \mathbf{w}_m^H & 1 \end{bmatrix} \right\| \geqslant 0 \tag{D29c}$$

$$\mathrm{Tr}(\tilde{\mathbf{w}}_m(x)\tilde{\mathbf{w}}_m(x)^H) + 2\mathrm{Re}\{\mathrm{Tr}(\mathbf{w}_m - \tilde{\mathbf{w}}_m(x))\tilde{\mathbf{w}}_m(x)^H\} \geqslant \mathrm{Tr}(\mathbf{A}_{BS1m}) \tag{D29d}$$

$$\left\|\begin{bmatrix}\mathbf{A}_{BS1nk} & \mathbf{W}_{nk} & \mathbf{w}_{nk}\\ \mathbf{W}_{nk}^H & \mathbf{A}_{BS2nk} & \mathbf{w}_{nk}\\ \mathbf{w}_{nk}^H & \mathbf{w}_{nk}^H & 1\end{bmatrix}\right\| \geqslant 0 \tag{D29e}$$

$$\mathrm{Tr}(\tilde{\mathbf{w}}_{nk}(x)\tilde{\mathbf{w}}_{nk}(x)^H) + 2\mathrm{Re}\{\mathrm{Tr}(\mathbf{w}_{nk} - \tilde{\mathbf{w}}_{nk}(x))\tilde{\mathbf{w}}_{nk}(x)^H\} \geqslant \mathrm{Tr}(\mathbf{A}_{BS1nk}) \tag{D29f}$$

where $\tilde{\mathbf{w}}_m(x)$ and $\tilde{\mathbf{w}}_{nk}(x)$ is the variable at the $n$ - th iteration by using first-order Taylor series approximation. An initial values $\tilde{\mathbf{w}}_m(0)$ and $\tilde{\mathbf{w}}_{nk}(0)$ can be randomly generated until satisfying the feasible conditions for the problem (D29).

**D.2 Conservative Reformulation by S-procedure**

Due to the probabilistic form of (D7), (D9)∼(D11), they can not be utilized to solve the problem (D15) directly. Observing their form, we find that the framework of S-procedure can be applied to acquire the solvable form of them. Therefore, in this subsection, we introduce the following Lemma 3 [8] to convert them into the worst-case deterministic constraints.

**Lemma 3:** Let $\mathbf{r} \sim C^{N\times 1}$ be a continuous random vector, which follows a certain statistical distribution. And $G(\mathbf{r}) : C^{N\times 1} \sim \Re$ is a function of $\mathbf{r}$. Then, there exists $R > 0$ and the following expression

$$\begin{cases} G(\mathbf{r}) \geqslant 0, \forall\|\mathbf{r}\|^2 \leqslant R^2\\ \Pr\{\|\mathbf{r}\|^2 \leqslant R^2\} \geqslant 1-\rho\end{cases} \Rightarrow \Pr\{G(\mathbf{r}) \leqslant 0\} \leqslant \rho \tag{D30}$$

where $\rho \in (0, 1]$.

According to Lemma 3, we can see (D7) must be satisfied if the following expressions

$$\begin{cases} \mathbf{r}_{Mm}^H\mathbf{B}_{Mm}\mathbf{r}_{Mm} + 2\,\mathrm{Re}\{\mathbf{r}_{Mm}^H\mathbf{c}_{Mm}\} - d_{Mm} \geqslant 0, \forall\|\mathbf{r}_{Mm}\|^2 \leqslant (R_{Mm})^2\\ \Pr\{\|\mathbf{r}_{Mm}\|^2 \leqslant (R_{Mm})^2\} \geqslant 1-\rho_m\end{cases} \tag{D31}$$

hold true. From (D30), we know the value of $R_{Mm}$ needs to be determined. $2\|\mathbf{r}_{Mm}\|^2$ is a Chi-square random variable with $2(N_M + N \times N_F)$ degrees of freedom because $\mathbf{r}_{Mm} \sim CN(0, \mathbf{I}_{(N_M+N\times N_F)})$. Thus,

$$R_{Mm} = \sqrt{\mathrm{Inv}_{2(N_M+N\times N_F)}(1-\rho_m)/2} \tag{D32}$$

where Inv() indicates the inverse cumulative distribution function of Chi-square random variable. Similarly, applying Lemma 3 to (D9)∼(D11), we can obtain the following expressions

$$\mathbf{r}_{Fnk}^H\mathbf{B}_{Fnk}\mathbf{r}_{Fnk} + 2\,\mathrm{Re}\{\mathbf{r}_{Fnk}^H\mathbf{c}_{Fnk}\} - d_{Fnk} \geqslant 0, \forall\|\mathbf{r}_{Fnk}\|^2 \leqslant (R_{Fnk})^2, n\in[1,N], k\in[1,K] \tag{D33}$$

$$-\mathbf{r}_{EEb}^H\mathbf{B}_{EEb}\mathbf{r}_{EEb} - 2\,\mathrm{Re}\{\mathbf{r}_{EEb}^H\mathbf{c}_{EEb}\} + d_{EEb} \geqslant 0, \forall\|\mathbf{r}_{\mathrm{EEb}}\|^2 \leqslant (R_{\mathrm{EEb}})^2, b\in[1,B] \tag{D34}$$

$$\mathbf{r}_{CEb}^H\mathbf{B}_{CEb}\mathbf{r}_{CEb} + 2\,\mathrm{Re}\{\mathbf{r}_{CEb}^H\mathbf{c}_{CEb}\} - d_{CEb} \geqslant 0, \forall\|\mathbf{r}_{CEb}\|^2 \leqslant (R_{CEb})^2, b\in[1,B] \tag{D35}$$

where

$$\begin{cases} R_{Fnk} = \sqrt{\mathrm{Inv}_{2(N_M+N\times N_F)}(1-\rho_{nk})/2}\\ R_{EEb} = \sqrt{\mathrm{Inv}_{2(N_M+N\times N_F)}(1-\rho_{Eb})/2}\,, n\in[1,N], k\in[1,K], b\in[1,B]\\ R_{CEb} = \sqrt{\mathrm{Inv}_{2N\times N_F}(1-q_{Eb})/2}\end{cases} \tag{D36}$$

Obviously, (D31) and (D33)∼(D35) are semi-infinite constraints. To make them more tractable, we further introduce the following Lemma 4 [9].

**Lemma 4 (S-procedure):** Let $g_k(x) = x^H F_k x + 2\,\mathrm{Re}\{g_k^H x\} + q_k$, $k = 1, 2$, where $F_k \in H^L$, $g_k \in C^L$, $q_k \in R$. When $g_1(\tilde{x}) < 0$, then $g_1(\tilde{x}) \leqslant 0 \Rightarrow g_2(\tilde{x}) \leqslant 0$ holds if and only if there is $\alpha \geqslant 0$ and

$$\alpha\begin{bmatrix}F_1 & g_1\\ g_1^H & q_1\end{bmatrix} - \begin{bmatrix}F_2 & g_2\\ g_2^H & q_2\end{bmatrix} \geqslant 0 \tag{D37}$$

Applying Lemma 4, we can respectively transform (D31) and (D33)∼(D35) into

$$\begin{bmatrix}\alpha_{Mm}\mathbf{I}_{N_M+NN_F} + \mathbf{B}_{Mm} & \mathbf{c}_{Mm}\\ \mathbf{c}_{Mm}^H & -d_{Mm} - \alpha_{Mm}(R_{Mm})^2\end{bmatrix} \geqslant 0, m\in[1,M] \tag{D38}$$

$$\begin{bmatrix}\alpha_{Fnk}\mathbf{I}_{N_M+NN_F} + \mathbf{B}_{Fnk} & \mathbf{c}_{Fnk}\\ \mathbf{c}_{Fnk}^H & -d_{Fnk} - \alpha_{Fnk}(R_{Fnk})^2\end{bmatrix} \geqslant 0, n\in[1,N], k\in[1,K] \tag{D39}$$

$$\begin{bmatrix}\alpha_{EEb}\mathbf{I}_{N_M+NN_F} - \mathbf{B}_{\mathrm{EEb}} & -\mathbf{c}_{EEb}\\ -\mathbf{c}_{EEb}^H & d_{EEb} - \alpha_{EEb}(R_{EEb})^2\end{bmatrix} \geqslant 0, b\in[1,B] \tag{D40}$$

$$\begin{bmatrix}\alpha_{CEb}\mathbf{I}_{NN_F} + \mathbf{B}_{CEb} & \mathbf{c}_{CEb}\\ \mathbf{c}_{CEb}^H & -d_{CEb} - \alpha_{CEb}(R_{CEb})^2\end{bmatrix} \geqslant 0, b\in[1,B] \tag{D41}$$

where $\alpha_{Mm} \geqslant 0$, $\alpha_{Fnk} \geqslant 0$, $\alpha_{EEb} \geqslant 0$, $\alpha_{CEb} \geqslant 0$ are auxiliary variables.

Similarly, we can apply Lemma 2 to obtain the rank-one solutions. Up to now, the problem (D15) can be reformulated as

$$\min_{\mathbf{w}_m,\mathbf{Z}_0,\mathbf{w}_{nk},\mathbf{Z}_n} \left\{\sum_{m=1}^{M}\mathrm{Tr}(\mathbf{W}_m) + \mathrm{Tr}(\mathbf{Z}_0) + \sum_{n=1}^{N}\left\{\left\{\sum_{k=1}^{K}\mathrm{Tr}(\mathbf{W}_{nk})\right\} + \mathrm{Tr}(\mathbf{Z}_n)\right\}\right\} \tag{D42a}$$

$$s.t.(D38) \sim (D41), (D29c) \sim (D29f) \tag{D42b}$$

## Appendix E Analysis of signaling overhead and complexity

In this paper, for the sake of simplification, the methods in subsection 3.2.1 and 3.2.2 are named the Bernstein method and the Worst-case method, respectively. The two compared robust methods are the Bernstein with isotropic AN method and the Worst-case with isotropic AN method, respectively, where the isotropic AN is imposed in the null space of estimated main channels in two compared methods. In this subsection, we will analyze the signaling overhead of BS cooperation and the complexity of various robust methods.

**E.1 The Signaling Overhead of Base Stations Cooperation Analysis**

The signaling overhead of BSs cooperation is determined by the information exchange between the MBS and FBSs [10]. Firstly, the major information need to be changed between them is the local CSI of FUs including that of ERs and that between MUs and FBSs. Then, the MBS can obtain the global CSIs by information exchange. After obtaining the optimized vectors, the MBS sends them to FBSs. It should be noted that the structure of AN is fixed and only the power of AN needs to be optimized in the two compared robust methods. Therefore, signaling overhead of BSs cooperation in various robust methods are given in Table 1. From Table 1, we can see that the former two methods and the latter ones have the same overhead, respectively, and the former two methods has a little higher signaling overhead than the latter ones.

**Table E1** Signaling overhead of various robust methods

| Method | Corresponding information exchanged |
|---|---|
| Bernstein | $(NN_F(K+B)+NMN_F+NKN_F+2NN_F{}^2)$ |
| Worst-case | $(NN_F(K+B)+NMN_F+NKN_F+2NN_F{}^2)$ |
| Bernstein with isotropic AN | $(NN_F(K+B)+NMN_F+NKN_F+NN_F^2+N)$ |
| Worst-case with isotropic AN | $(NN_F(K+B)+NMN_F+NKN_F+NN_F^2+N)$ |

**E.2 The complexity analysis**

The complexity of solving an optimization problem through the method of interior point in each iteration is $\mathrm{O}(\sqrt{\theta}L\log(1/\omega))$[11]. Among them, $\omega$ is the precision of solution, $\mathrm{O}(\sqrt{\theta}\log(1/\omega))$ is the number of iteration used to find the optimal solution while $\mathrm{O}(L)$ is the expenditure of each iteration. When an optimization problem has SOC constraints and PSD constraints, its barrier parameter is

$$\theta=\sum_{i=1}^{m_{\mathrm{sdp}}}k_{i,\mathrm{sdp}}+2m_{\mathrm{socp}} \tag{E1}$$

where $m_{\mathrm{sdp}}$, $k_{i,\mathrm{sdp}}$ and $m_{\mathrm{socp}}$ are the number of PSD constraint, dimension in $i$ - th PSD constraint and SOC constraint, respectively. The expenditure of each computation, i.e.,$\mathrm{O}(L)$, is the order of

$$L=n_{op}\sum_{i=1}^{m_{\mathrm{sdp}}}k_{i,\mathrm{sdp}}^3+n_{op}^2\sum_{i=1}^{m_{\mathrm{sdp}}}k_{i,\mathrm{sdp}}^2+n_{op}\sum_{i=1}^{m_{\mathrm{socp}}}k_{i,\mathrm{socp}}^2+n_{op}^3 \tag{E2}$$

where $n_{op}$ is the number of optimization variables while $k_{i,\mathrm{socp}}$ is the dimensions of $i$ - th SOC constraint. The number and dimension of variables and constraints in four robust methods are summarized in Table 2, where the linear constraint is seen as an one-dimension PSD constraint.

**Table 2** Related variables and constraints of various robust methods.

| Methods | Variables (dimension, number) | PSD constraints (dimension, number) | SOC constraints (dimension, number) |
|---|---|---|---|
| Bernstein | $(N_M,3M+1)$<br>$(N_F,3NK+2N)$<br>$(1,2M+2NK+4B)$ | $(NN_F,B)$<br>$(N_M,M+1)$<br>$(N_F,NK+2N)$<br>$(1,3M+3NK+4B)$<br>$(N_M+NN_F,M+NK+B)$ | $(2N_M+2,M)$<br>$(2N_F+2,NK)$<br>$((NN_F)^2+NN_F+1,B)$<br>$((N_M+NN_F)^2+N_M+NN_F+1,M+NK+B)$ |
| Worst-case | $(N_M,3M+1)$<br>$(N_F,3NK+2N)$<br>$(1,M+NK+2B)$ | $(NN_F,B)$<br>$(N_M,M+1)$<br>$(N_F,NK+2N)$<br>$(1,3M+3NK+4B)$<br>$(N_M+NN_F,M+NK+B)$ | $(2N_M+1,M)$<br>$(2N_F,NK)$ |
| Bernstein with isotropic AN | $(N_M,3M)$<br>$(N_F,3NK+N)$<br>$(1,2M+2NK+4B+N+1)$ | $(NN_F,B)$<br>$(N_M,M)$<br>$(N_F,NK+N)$<br>$(1,3M+3NK+4B+N+1)$<br>$(N_M+NN_F,M+NK+B)$ | $(2N_M+2,M)$<br>$(2N_F+2,NK)$<br>$((NN_F)^2+NN_F+1,B)$<br>$((N_M+NN_F)^2+N_M+NN_F+1,M+NK+B)$ |
| Worst-case with isotropic AN | $(N_M,3M)$<br>$(N_F,3NK+N)$<br>$(1,2M+2NK+2B+N+1)$ | $(NN_F+1,B)$<br>$(N_M,M)$<br>$(N_F,NK+N)$<br>$(1,3M+3NK+4B+N+1)$<br>$(N_M+NN_F+1,M+NK+B)$ | $(2N_M+1,M)$<br>$(2N_F,NK)$ |

The complexity of two algorithms under the same precision is determined by $\theta$ and $L$. From the second column to the fourth column and combining (E2), we can observe the four robust methods have the same determined parameters. Therefore, we only need to compare their barrier parameters. According to (E1), we can acquire their corresponding barrier parameters as

$$\begin{cases} \theta_{\text{Bernstein}} = M(2N_M + NN_F + 7) + NK(N_M + NN_F + N_F + 7) + B(N_M + 2NN_F + 8) + N_M + 2NN_F \\ \theta_{\text{Worst - case}} = M(2N_M + NN_F + 5) + NK(N_M + NN_F + N_F + 5) + B(N_M + NN_F + 3) + N_M + 2NN_F \\ \theta_{\text{Bernstein with isotropic AN}} = M(A+2) + NK(N_M + NN_F + N_F + 7) + B(N_M + 2NN_F + 8) + NN_F + N + 1 \\ \theta_{\text{Worst - case with isotropic AN}} = MA + NK(N_M + NN_F + N_F + 5) + B(N_M + NN_F + 3) + NN_F + N + 1 \end{cases} \tag{E3}$$

where $A = 2N_M + NN_F + 5$. Obviously, $\theta_{\text{Bernstein}} > \theta_{\text{Bernstein with isotropic AN}} \geqslant \theta_{\text{Worst - case}} > \theta_{\text{Worst - case with isotropic AN}}$. So the Bernstein method has the highest complexity, followed by the Bernstein with isotropic AN method, Worst-case method and Worst-case with isotropic AN. In following simulation analysis, we can see that the achieved performance of the four robust methods have the same order.

## Appendix F The specific simulation parameters setting

In this section, we provide the simulation results to evaluate the performance of the proposed methods. Similar to [12], we consider the simplified large-scale fading model and the small scale fading model. In specific, the large-scale fading model is expressed as $(d/d_0)^{-\alpha}$, where $d$, $d_0 = 5$ and $\alpha = 3$ denote the distance between BS and user, the reference distance and the path loss exponent, respectively. The small scale fading obeys Rayleigh fading. Meanwhile, unless otherwise specified, the distance between the MBS and all users are 60 meters, and that between FBSs and MUs, FUs, ERs are 30, 20 and 10 meters, respectively; the number of antennas at the MBS and each FBS are $N_M = 5$, $N_F = 4$, respectively; the number of FBSs, MUs, ERs and FUs are $N = 1$, $M = 2$, $B = 2$, and $K = 1$, respectively; the coefficient of energy conversion $\kappa$ is 1. In addition, for the sake of simplifying the analysis, we assume that the CSI error of MUs and FUs has the same value, i.e., $\mathbf{Q}_m = \mathbf{Q}_{nk} = \varepsilon_b \mathbf{I}_{N_M}$, $\mathbf{Q}_{n,m} = \mathbf{Q}_{l,nk} = \varepsilon_b \mathbf{I}_{N_F}$, $\varepsilon_b = 0.002$; meanwhile, the CSI error of ERs has the same value $\varepsilon_e$, i.e., $\mathbf{Q}_{Eb} = \varepsilon_e \mathbf{I}_{N_M}$, $\mathbf{Q}_{n,Eb} = \varepsilon_e \mathbf{I}_{N_F}$; furthermore, the SINR threshold of MUs, FUs, Eves and the EH threshold are set to be the same, respectively, i.e., $\gamma_m = \gamma_{nk} = \gamma_u$, $\gamma_{Eb} = \gamma_e$, $\phi_b = \phi$, $\forall m, n, k, b$, while all the probability threshold are set to be the same, i.e., $\rho_{\mathrm{m}} = \rho_{nk} = \rho_{Eb} = q_{Eb} = 0.1$, $\forall m, n, k, b$. In the simulations, the Perfect CSI method is presented as a benchmark to compare with the four robust methods, which does not consider the CSI error and regards the estimated channel vectors as actual channel vectors. All simulation results are averaged over 1000 randomly generated channel realizations.

**References**

1. Wang H, Zheng T, Yuan J. Physical layer security in heterogeneous cellular networks[J]. IEEE Transactions on Communications, 2016, 64(3):1204-1219.
2. Y Ren, T Lv and H Gao. Secure Wireless Information and Power Transfer in Heterogeneous Networks[J]. IEEE Access, 2017, 5(99): 4967-4979.
3. Z. Zhu, Z. Chu, Z. Wang, and I. Lee. Outage constrained robust beamforming for secure broadcasting systems with energy harvesting[J]. IEEE Transactions on Wireless Communications, 2016, 15(11): 7610-7620.
4. Le T A, Vien Q T, Nguyen H X et al. Robust Chance-Constrained Optimization for Power-Efficient and Secure SWIPT Systems[J]. IEEE Transactions on Green Communications and Networking, 2017, 1(3): 333-346.
5. Xu J, Xu S, Xu C. Probabilistic robust secure beamforming in MISO channels with imperfect LCSI and statistical ECSI[J]. IEEE Access, 2017, 5: 10277-10284.
6. F. Zhou, Z. Li, J. Cheng et al. Robust AN-Aided Beamforming and Power Splitting Design for Secure MISO Cognitive Radio with SWIPT[J]. IEEE Transactions on Wireless Communications, 2017, 16(4): 2450-2464.
7. Ma S, Sun D. Chance Constrained Robust Beamforming in Cognitive Radio Networks[J]. IEEE Communications Letters, 2013, 17(1): 67-70.
8. Rashid U, Tuan H D, Kha H H et al. Joint Optimization of Source Precoding and Relay Beamforming in Wireless MIMO Relay Networks[J]. IEEE Transactions on Communications, 2014, 62(2): 488-499.
9. Wang K Y, Chang T H, Ma W K et al. A semidefinite relaxation based conservative approach to robust transmit beamforming with probabilistic SINR constraints[C]. Signal Processing Conference, 2010, European. IEEE, 2010: 407-411.
10. Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press, 2004.
11. Zhang L, Jin L, Luo W. Robust secure transmission for multiuser MISO systems with probabilistic QoS constraints[J]. Science China Information Sciences, 2016, 59(2): 1-13.
12. Z. Chu, Z. Zhu, M. Johnston, and S. Y. L. Goff. Simultaneous wireless information power transfer for MISO secrecy channel[J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 6913-6925.