Supplementary Materials for "Two-Dimensional Ice Filling Based Channel Estimation in Densifying MIMO Systems"

Zijian Zhang and Mingyao Cui

APPENDIX A
PROOF OF LEMMA 1

Given the channel model in (4) and $\mathbf{h} \equiv \mathrm{vec}(\mathbf{H})$, the vectorized channel can be rewritten as

$$\mathbf{h} = \sqrt{\frac{N_{\mathrm{T}} N_{\mathrm{R}}}{CR}} \sum_{c=1}^{C} \sum_{r=1}^{R} g_{c,r} \mathbf{b}^{*}(\varphi_{c,r}) \otimes \mathbf{a}(\theta_{c,r}). \tag{21}$$

Utilizing the commutative law of Kronecker product $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$, we have

$$\begin{aligned}&\left(\mathbf{b}^{*}(\varphi_{c,r}) \otimes \mathbf{a}(\theta_{c,r})\right)\left(\mathbf{b}^{T}(\varphi_{c,r}) \otimes \mathbf{a}^{H}(\theta_{c,r})\right) = \\ &\left(\mathbf{b}^{*}(\varphi_{c,r})\mathbf{b}^{T}(\varphi_{c,r})\right) \otimes \left(\mathbf{a}(\theta_{c,r})\mathbf{a}^{H}(\theta_{c,r})\right).\end{aligned} \tag{22}$$

Then, the covariance of channel $\mathbf{h}$ can be derived as (23), where (a) holds since the gains of different rays $\{g_{c,r}\}_{c=1,r=1}^{C,R}$ are i.i.d. with zero mean and normalized power. (b) holds according to (22). (c) holds since $\mathsf{E}\left(\mathbf{a}(\theta_{c,r})\mathbf{a}^{H}(\theta_{c,r})\right) = \mathsf{E}\left(\mathbf{a}(\theta_{c',r'})\mathbf{a}^{H}(\theta_{c',r'})\right)$ and $\mathsf{E}\left(\mathbf{b}^{*}(\varphi_{c,r})\mathbf{b}^{T}(\varphi_{c,r})\right) = \mathsf{E}\left(\mathbf{b}^{*}(\varphi_{c',r'})\mathbf{b}^{T}(\varphi_{c',r'})\right)$ hold for any $c, c' \in \{1, \cdots, C\}$ and $r, r' \in \{1, \cdots, R\}$. (d) holds by defining

$$\boldsymbol{\Sigma}_{\mathrm{T}} = N_{\mathrm{T}} \mathsf{E}\left(\mathbf{b}^{*}(\varphi_{c,r})\mathbf{b}^{T}(\varphi_{c,r})\right), \tag{25}$$

$$\boldsymbol{\Sigma}_{\mathrm{R}} = N_{\mathrm{R}} \mathsf{E}\left(\mathbf{a}(\theta_{c,r})\mathbf{a}^{H}(\theta_{c,r})\right), \tag{26}$$

wherein c and r can be arbitrarily selected from $\{1, \cdots, C\}$ and $\{1, \cdots, R\}$, respectively. One can find that, the matrix $\boldsymbol{\Sigma}_{\mathrm{T}}$ only depends on the steering vector $\mathbf{b}(\varphi)$ at the transmitter, while the matrix $\boldsymbol{\Sigma}_{\mathrm{R}}$ is only associated with the steering vector $\mathbf{a}(\theta)$ at the receiver. Thus, $\boldsymbol{\Sigma}_{\mathrm{T}}$ and $\boldsymbol{\Sigma}_{\mathrm{R}}$ can be viewed as the kernels that characterize the correlation among the transmitter antennas and that among the receiver antennas, respectively. This completes the proof.

APPENDIX B
PROOF OF LEMMA 2

Using some matrix techniques, the MI $I(\mathbf{y};\mathbf{h})$ can be rewritten as equation (24), where (a) holds since $\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA})$ and $\boldsymbol{\Xi} = \sigma^2 \mathrm{blkdiag}\left(\mathbf{W}_1^H \mathbf{W}_1, \cdots, \mathbf{W}_Q^H \mathbf{W}_Q\right)$; (b) holds according to the property that $(\mathbf{a} \otimes \mathbf{B})\mathbf{C}(\mathbf{a}^H \otimes \mathbf{D}) = (\mathbf{a}\mathbf{a}^H) \otimes (\mathbf{BCD})$ if all dimensions meet the requirements of matrix multiplications. To find more insights, we perform singular value decomposition (SVD) on all $\{\mathbf{W}_q\}_{q=1}^{Q}$ and then substitute all decomposition formulas $\mathbf{W}_q = \boldsymbol{\Pi}_q \boldsymbol{\Omega}_q \boldsymbol{\Upsilon}_q^H$ into (24). It is evident that $\mathbf{W}_q(\mathbf{W}_q^H \mathbf{W}_q)^{-1}\mathbf{W}_q^H = \boldsymbol{\Pi}_q \boldsymbol{\Pi}_q^H$, thus the MI $I(\mathbf{y};\mathbf{h})$ can be rewritten as

$$I(\mathbf{y};\mathbf{h}) = \log_2 \det\left(\mathbf{I}_{N_{\mathrm{R}}N_{\mathrm{T}}} + \frac{1}{\sigma^2}\sum_{q=1}^{Q}\left((\mathbf{v}_q^{*}\mathbf{v}_q^{T}) \otimes (\boldsymbol{\Pi}_q \boldsymbol{\Pi}_q^H)\right)\boldsymbol{\Sigma}_{\mathbf{h}}\right). \tag{27}$$

Observing (27), one can find that the MI $I(\mathbf{y};\mathbf{h})$ in (7) only relies on the orthogonal matrix $\boldsymbol{\Pi}_q \in \mathbb{C}^{N \times N_{\mathrm{RF}}}$ decomposed from $\mathbf{W}_q$ for all $q \in \{1, \cdots, Q\}$, while it does not depend on any $\boldsymbol{\Omega}_q$ or $\boldsymbol{\Upsilon}_q$. It indicates that imposing $\mathbf{W}_q = \boldsymbol{\Pi}_q$ does

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{h}} = \mathsf{E}\left(\mathbf{h}\mathbf{h}^H\right) &\overset{(a)}{=} \frac{N_{\mathrm{T}}N_{\mathrm{R}}}{CR}\sum_{c=1}^{C}\sum_{r=1}^{R}\mathsf{E}\left(\left(\mathbf{b}^{*}(\varphi_{c,r}) \otimes \mathbf{a}(\theta_{c,r})\right)\left(\mathbf{b}^{T}(\varphi_{c,r}) \otimes \mathbf{a}^{H}(\theta_{c,r})\right)\right) \\ &\overset{(b)}{=} \frac{N_{\mathrm{T}}N_{\mathrm{R}}}{CR}\sum_{c=1}^{C}\sum_{r=1}^{R}\mathsf{E}\left(\mathbf{b}^{*}(\varphi_{c,r})\mathbf{b}^{T}(\varphi_{c,r})\right) \otimes \mathsf{E}\left(\mathbf{a}(\theta_{c,r})\mathbf{a}^{H}(\theta_{c,r})\right) \\ &\overset{(c)}{=} N_{\mathrm{T}}N_{\mathrm{R}}\mathsf{E}\left(\mathbf{b}^{*}(\varphi_{c,r})\mathbf{b}^{T}(\varphi_{c,r})\right) \otimes \mathsf{E}\left(\mathbf{a}(\theta_{c,r})\mathbf{a}^{H}(\theta_{c,r})\right) \overset{(d)}{=} \boldsymbol{\Sigma}_{\mathrm{T}} \otimes \boldsymbol{\Sigma}_{\mathrm{R}}.\end{aligned} \tag{23}$$

$$\begin{aligned}I(\mathbf{y};\mathbf{h}) &\overset{(a)}{=} \log_2\det\Bigg(\mathbf{I}_{N_{\mathrm{R}}N_{\mathrm{T}}} + \\ &\quad \frac{1}{\sigma^2}\left[\mathbf{v}_1^{*} \otimes \mathbf{W}_1, \cdots, \mathbf{v}_Q^{*} \otimes \mathbf{W}_Q\right]\mathrm{blkdiag}\left(\left(\mathbf{W}_1^H\mathbf{W}_1\right)^{-1}, \cdots, \left(\mathbf{W}_Q^H\mathbf{W}_Q\right)^{-1}\right)\left[\mathbf{v}_1^{*} \otimes \mathbf{W}_1, \cdots, \mathbf{v}_Q^{*} \otimes \mathbf{W}_Q\right]^H\boldsymbol{\Sigma}_{\mathbf{h}}\Bigg) \\ &\overset{(b)}{=} \log_2\det\left(\mathbf{I}_{N_{\mathrm{R}}N_{\mathrm{T}}} + \frac{1}{\sigma^2}\sum_{q=1}^{Q}\left((\mathbf{v}_q^{*}\mathbf{v}_q^{T}) \otimes \left(\mathbf{W}_q(\mathbf{W}_q^H\mathbf{W}_q)^{-1}\mathbf{W}_q^H\right)\right)\boldsymbol{\Sigma}_{\mathbf{h}}\right).\end{aligned} \tag{24}$$

not change the value of $I(\mathbf{y};\mathbf{h})$. As a result, the orthogonality constraint $\mathbf{W}_q^H\mathbf{W}_q=\mathbf{\Pi}_q^H\mathbf{\Pi}_q=\mathbf{I}_{\mathrm{RF}}$ can be safely introduced into the problem formulation regarding $I(\mathbf{y};\mathbf{h})$, which completes the proof.

Appendix C
Proof of MI increment $I(\bar{\mathbf{y}}_{t+1};\mathbf{h})-I(\bar{\mathbf{y}}_t;\mathbf{h})$

Using some matrix partition operations, the MI $I(\bar{\mathbf{y}}_{t+1};\mathbf{h})$ can be rewritten as

$$\begin{aligned}I(\bar{\mathbf{y}}_{t+1};\mathbf{h}) &\overset{(a)}{=} \log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}Q}+\frac{1}{\sigma^2}\bar{\mathbf{X}}_{t+1}^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_{t+1}\right)\\ &=\log_2\det\begin{bmatrix}\mathbf{I}_{N_{\mathrm{RF}}t}+\frac{1}{\sigma^2}\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t & \frac{1}{\sigma^2}\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_{t+1}\\ \frac{1}{\sigma^2}\mathbf{X}_{t+1}^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t & \mathbf{I}_{N_{\mathrm{RF}}}+\frac{1}{\sigma^2}\mathbf{X}_{t+1}^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_{t+1}\end{bmatrix}\\ &\overset{(b)}{=}\log_2\det\begin{bmatrix}\mathbf{I}_{N_{\mathrm{RF}}t}+\frac{1}{\sigma^2}\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t & \frac{1}{\sigma^2}\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_{t+1}\\ \mathbf{0}_{N_{\mathrm{RF}}\times N_{\mathrm{RF}}t} & \mathbf{I}_{N_{\mathrm{RF}}}+\frac{1}{\sigma^2}\mathbf{X}_{t+1}^H\mathbf{\Sigma}_{t}\mathbf{X}_{t+1}\end{bmatrix}\\ &=I(\bar{\mathbf{y}}_t;\mathbf{h})+\log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}}+\frac{1}{\sigma^2}\mathbf{X}_{t+1}^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_{t+1}\right),\end{aligned}\quad(28)$$

where (a) holds since according to **Lemma 2** and (b) holds by performing matrix triangularization. In particular, $\mathbf{\Sigma}_t$ is given by $\mathbf{\Sigma}_t=\mathbf{\Sigma}_{\mathbf{h}}-\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t\left(\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t+\sigma^2\mathbf{I}_{N_{\mathrm{RF}}t}\right)^{-1}\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}$, which completes the proof.

Appendix D
Proof of Lemma 3

The key idea of the proof is to rewrite the $\bar{\mathbf{X}}_t$-related terms in (10) as $\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t=\mathbf{\Sigma}_{\mathbf{h}}\left[\bar{\mathbf{X}}_{t-1},\mathbf{X}_t\right]$ and

$$\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t=\begin{bmatrix}\bar{\mathbf{X}}_{t-1}^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_{t-1} & \bar{\mathbf{X}}_{t-1}^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_t\\ \mathbf{X}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_{t-1} & \mathbf{X}_t^H\mathbf{\Sigma}_{\mathbf{h}}\mathbf{X}_t\end{bmatrix}.\quad(30)$$

Then, using the Schur's matrix inversion formula to expand the term $\left(\bar{\mathbf{X}}_t^H\mathbf{\Sigma}_{\mathbf{h}}\bar{\mathbf{X}}_t+\sigma^2\mathbf{I}_{N_{\mathrm{RF}}t}\right)^{-1}$ in (10), the following recursion formula of can be obtained:

$$\mathbf{\Sigma}_{t+1}=\mathbf{\Sigma}_t-\mathbf{\Sigma}_t\mathbf{X}_{t+1}\left(\mathbf{X}_{t+1}^H\mathbf{\Sigma}_t\mathbf{X}_{t+1}+\sigma^2\mathbf{I}_{N_{\mathrm{RF}}}\right)^{-1}\mathbf{X}_{t+1}^H\mathbf{\Sigma}_t,\quad(31)$$

When $\mathbf{X}_{t+1}=\sqrt{P}\mathbf{U}_t\left(:,[1,\cdots,N_{\mathrm{RF}}]\right)$, we have $\mathbf{\Sigma}_t\mathbf{X}_{t+1}=\mathbf{X}_{t+1}\mathrm{diag}\left(\lambda_1\left(\mathbf{\Sigma}_t\right),\cdots,\lambda_{N_{\mathrm{RF}}}\left(\mathbf{\Sigma}_t\right)\right)$ and $\mathbf{X}_{t+1}^H\mathbf{\Sigma}_t\mathbf{X}_{t+1}=P\mathrm{diag}\left(\lambda_1\left(\mathbf{\Sigma}_t\right),\cdots,\lambda_{N_{\mathrm{RF}}}\left(\mathbf{\Sigma}_t\right)\right)$. Thus, the following equality holds:

$$\begin{aligned}\mathbf{\Sigma}_{t+1}=&\mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H-\mathbf{X}_{t+1}\mathrm{diag}\Bigg(\frac{\lambda_1^2\left(\mathbf{\Sigma}_t\right)}{P\lambda_1\left(\mathbf{\Sigma}_t\right)+\sigma^2},\cdots,\\ &\frac{\lambda_{N_{\mathrm{RF}}}^2\left(\mathbf{\Sigma}_t\right)}{P\lambda_{N_{\mathrm{RF}}}\left(\mathbf{\Sigma}_t\right)+\sigma^2}\Bigg)\mathbf{X}_{t+1}^H.\end{aligned}\quad(32)$$

Given that $\mathbf{X}_{t+1}\mathrm{diag}(\frac{\lambda_1^2(\mathbf{\Sigma}_t)}{P\lambda_1(\mathbf{\Sigma}_t)+\sigma^2},\cdots,\frac{\lambda_{N_{\mathrm{RF}}}^2(\mathbf{\Sigma}_t)}{P\lambda_{N_{\mathrm{RF}}}(\mathbf{\Sigma}_t)+\sigma^2})\mathbf{X}_{t+1}^H=\mathbf{U}_t\mathrm{diag}(\frac{P\lambda_1^2(\mathbf{\Sigma}_t)}{P\lambda_1(\mathbf{\Sigma}_t)+\sigma^2},\cdots,\frac{P\lambda_{N_{\mathrm{RF}}}^2(\mathbf{\Sigma}_t)}{P\lambda_{N_{\mathrm{RF}}}(\mathbf{\Sigma}_t)+\sigma^2},\underbrace{0,\cdots,0}_{N_{\mathrm{R}}N_{\mathrm{T}}-N_{\mathrm{RF}}})\mathbf{U}_t^H$ and $\mathbf{\Sigma}_t=\mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^H$, the equality in (13) can be derived from (32), which completes the proof.

Appendix E
Proof of Corollary 1

According to **Lemma 1** and equality $\left(\mathbf{ABA}^H\right)\otimes\left(\mathbf{CDC}^H\right)=\left(\mathbf{A}\otimes\mathbf{C}\right)\left(\mathbf{B}\otimes\mathbf{D}\right)\left(\mathbf{A}^H\otimes\mathbf{C}^H\right)$, the kernel $\mathbf{\Sigma}_{\mathbf{h}}$ can be decomposed as

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{h}}&=\left(\mathbf{U}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}\mathbf{U}_{\mathrm{T}}{}^H\right)\otimes\left(\mathbf{U}_{\mathrm{T}}\mathbf{\Lambda}_{\mathrm{T}}\mathbf{U}_{\mathrm{T}}{}^H\right)\\ &=\underbrace{\left(\mathbf{U}_{\mathrm{T}}\otimes\mathbf{U}_{\mathrm{R}}\right)}_{\mathbf{U}_0}\underbrace{\left(\mathbf{\Lambda}_{\mathrm{T}}\otimes\mathbf{\Lambda}_{\mathrm{R}}\right)}_{\text{Eigenvalue matrix}}\left(\mathbf{U}_{\mathrm{T}}^H\otimes\mathbf{U}_{\mathrm{R}}^H\right)\\ &=\sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}\sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}\alpha_{n_{\mathrm{T}}}\beta_{n_{\mathrm{R}}}\left(\mathbf{a}_{n_{\mathrm{T}}}\otimes\mathbf{b}_{n_{\mathrm{R}}}\right)\left(\mathbf{a}_{n_{\mathrm{T}}}^H\otimes\mathbf{b}_{n_{\mathrm{R}}}^H\right),\end{aligned}\quad(33)$$

Based on (33), one can verify without difficulty that (14) is exactly the eigenvalue decomposition of $\mathbf{\Sigma}_{\mathbf{h}}$, which completes the proof.

Appendix F
Proof of Lemma 4

Given the new constraints $\mathbf{v}_{t+1}\in\left\{\sqrt{P}\mathbf{a}_{n_{\mathrm{T}}}^*\right\}_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}$ and $\mathbf{w}_{t+1,k}\in\left\{\mathbf{b}_{n_{\mathrm{R}}}\right\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}$ for all $k\in\{1,\cdots,N_{\mathrm{RF}}\}$, problem (12) can be reorganized as

$$\begin{aligned}\max_{\mathbf{v}_{t+1},\mathbf{W}_{t+1}}\ & f\left(\mathbf{v}_{t+1},\mathbf{W}_{t+1}\right)\\ \text{s.t. }& \mathbf{v}_{t+1}\in\left\{\sqrt{P}\mathbf{a}_{n_{\mathrm{T}}}^*\right\}_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}},\\ & \mathbf{w}_{t+1,k}\in\left\{\mathbf{b}_{n_{\mathrm{R}}}\right\}_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}},\forall k\in\{1,\cdots,N_{\mathrm{RF}}\},\\ & \mathbf{w}_{t+1,k}\neq\mathbf{w}_{t+1,k'},\forall k\neq k',\end{aligned}\quad(34)$$

where the objective function is given in (29), in which (a) holds according to the definition in (15) and (b) holds by utilizing the property to the property that $\left(\mathbf{A}\otimes\mathbf{B}\right)\left(\mathbf{C}\otimes\mathbf{D}\right)=\left(\mathbf{AC}\right)\otimes\left(\mathbf{BD}\right)$. Note that, the constraint $\mathbf{w}_{t+1,k}\neq\mathbf{w}_{t+1,k'}$ for all $k\neq k'$ in (34) ensures the orthogonality of $\mathbf{W}_{t+1}$. Observing (34), one can find that our goal becomes finding optimal indexes n_{T} and $\{n_{\mathrm{R},k}\}_{k=1}^{N_{\mathrm{RF}}}$ that maximize the MI increment $f\left(\mathbf{v}_{t+1},\mathbf{W}_{t+1}\right)$. Assuming that the optimal indexes are expressed by $n_{\mathrm{T}}^{\mathrm{opt}}$ and $\{n_{\mathrm{R},k}^{\mathrm{opt}}\}_{k=1}^{N_{\mathrm{RF}}}$, the optimal precoder and the optimal combiner are

$$\mathbf{v}_{t+1}^{\mathrm{opt}}=\sqrt{P}\mathbf{a}_{n_{\mathrm{T}}^{\mathrm{opt}}}^*\text{ and }\mathbf{W}_{t+1}^{\mathrm{opt}}=\left[\mathbf{b}_{n_{\mathrm{R},1}^{\mathrm{opt}}},\cdots,\mathbf{b}_{n_{\mathrm{R},N_{\mathrm{RF}}}^{\mathrm{opt}}}\right],\quad(35)$$

$$\begin{aligned}f\left(\mathbf{v}_{t+1},\mathbf{W}_{t+1}\right)&\overset{(a)}{=}\log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}}+\frac{1}{\sigma^2}\sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}\sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}\lambda_{t,n_{\mathrm{T}},n_{\mathrm{R}}}\left(\mathbf{v}_{t+1}^T\otimes\mathbf{W}_{t+1}^H\right)\left(\mathbf{a}_{n_{\mathrm{T}}}\otimes\mathbf{b}_{n_{\mathrm{R}}}\right)\left(\mathbf{a}_{n_{\mathrm{T}}}^H\otimes\mathbf{b}_{n_{\mathrm{R}}}^H\right)\left(\mathbf{v}_{t+1}^*\otimes\mathbf{W}_{t+1}\right)\right)\\ &\overset{(b)}{=}\log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}}+\frac{1}{\sigma^2}\sum_{n_{\mathrm{T}}=1}^{N_{\mathrm{T}}}\sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}\lambda_{t,n_{\mathrm{T}},n_{\mathrm{R}}}\left|\mathbf{a}_{n_{\mathrm{T}}}^H\mathbf{v}_{t+1}^*\right|^2\mathbf{W}_{t+1}^H\mathbf{b}_{n_{\mathrm{R}}}\mathbf{b}_{n_{\mathrm{R}}}^H\mathbf{W}_{t+1}\right)\end{aligned}\quad(29)$$

respectively. Then, we have

$$\mathbf{a}_{n_{\mathrm{T}}}^{H}(\mathbf{v}_{t+1}^{\mathrm{opt}})^{*}=\begin{cases}\sqrt{P}, & n_{\mathrm{T}}=n_{\mathrm{T}}^{\mathrm{opt}}\\ 0, & \text{else}\end{cases}, \tag{36a}$$

$$\mathbf{b}_{n_{\mathrm{R}}}^{H}\mathbf{W}_{t+1}^{\mathrm{opt}}=\begin{cases}\mathbf{e}_{n_R}^{T}, & n_{\mathrm{R}}\in\{n_{\mathrm{R},k}^{\mathrm{opt}}\}_{k=1}^{N_{\mathrm{RF}}}\\ \mathbf{0}_{N_{\mathrm{RF}}}^{T}, & \text{else}\end{cases}, \tag{36b}$$

where $\mathbf{e}_{n_R}$ denotes an N_{RF}-dimensional vector whose n_R-th entry is one and the other entries are zero. By substituting (36) into (34), the optimal MI increment $f\left(\mathbf{v}_{t+1}^{\mathrm{opt}},\mathbf{W}_{t+1}^{\mathrm{opt}}\right)$ can be expressed by

$$\begin{aligned}
&f\left(\mathbf{v}_{t+1}^{\mathrm{opt}},\mathbf{W}_{t+1}^{\mathrm{opt}}\right)\\
&=\log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}}+\frac{P}{\sigma^2}\sum_{n_{\mathrm{R}}=1}^{N_{\mathrm{R}}}\lambda_{t,n_{\mathrm{T}}^{\mathrm{opt}},n_{\mathrm{R}}}(\mathbf{W}_{t+1}^{\mathrm{opt}})^{H}\mathbf{b}_{n_{\mathrm{R}}}\mathbf{b}_{n_{\mathrm{R}}}^{H}\mathbf{W}_{t+1}^{\mathrm{opt}}\right)\\
&=\log_2\det\left(\mathbf{I}_{N_{\mathrm{RF}}}+\frac{P}{\sigma^2}\mathrm{diag}\left(\lambda_{t,n_{\mathrm{T}}^{\mathrm{opt}},n_{\mathrm{R},1}^{\mathrm{opt}}},\cdots,\lambda_{t,n_{\mathrm{T}}^{\mathrm{opt}},n_{\mathrm{R},N_{\mathrm{RF}}}^{\mathrm{opt}}}\right)\right)\\
&=\sum_{k=1}^{N_{\mathrm{RF}}}\log_2\left(1+\frac{P\lambda_{t,n_{\mathrm{T}}^{\mathrm{opt}},n_{\mathrm{R},k}^{\mathrm{opt}}}}{\sigma^2}\right),
\end{aligned} \tag{37}$$

which only relies on the eigenvalues of $\mathbf{\Sigma}_t$. In this context, the problem becomes finding n_{T} and $\{n_{\mathrm{R},k}\}_{k=1}^{N_{\mathrm{RF}}}$ that maximize $f\left(\mathbf{v}_{t+1},\mathbf{W}_{t+1}\right)$, as formulated in (17). This completes the proof.