SUPPLIMENTARY MATERIALS

COMPUTATIONAL COMPLEXITY ANALYSIS

In this paper, the computational complexity is defined as the number of complex addition ($CC^{[A]}$) and multiplication ($CC^{[M]}$) required to calculate the phase and channel information at each symbol. Moreover, the computational complexity of one complex division is defined as one complex multiplication due to their similar computational complexity. Furthermore, we assume that the complex exponential functions and the square root operations are implemented by means of a lookup table.

The channel amplitude estimation is calculated by (5), (7), and (8). Therefore, the computational complexity for a symbol can be calculated by

$$CC^{[\mathrm{M}]}_{|\hat{\mathbf{H}}|} = \underbrace{N_r N_t^2 / L_c}_{(5)} + \underbrace{N_r N_t / L_c}_{(7)}, \tag{36}$$

$$CC^{[\mathrm{A}]}_{|\hat{\mathbf{H}}|} = \underbrace{N_r N_t \left(N_t - 1\right)/L_c}_{(5)} + \underbrace{N_r N_t / L_c + N_r N_t / L_f}_{(7)}. \tag{37}$$

The WLLS one-shot estimation is calculated by (9), (10), (18)-(20). However, (9) can use the intermediate result of $\mathbf{Y}_i \mathbf{S}_i^H$ in (5). Therefore, the computational complexity for a symbol has the form of

$$\begin{aligned} CC^{[\mathrm{M}]}_{\mathrm{WLLS}} = &\underbrace{N_r N_t / L_c}_{(18)} + \underbrace{\left(N_r + N_t - 1\right) N_r N_t / L_f}_{(19)} \\ &+ \underbrace{N_r N_t (N_r + N_t - 1)^2 \Big/ L_f}_{(20)} \\ &+ \underbrace{N_r N_t \left(N_r + N_t - 1\right)/L_c}_{(20)}, \end{aligned} \tag{38}$$

$$\begin{aligned} CC^{[\mathrm{A}]}_{\mathrm{WLLS}} = &\underbrace{N_r N_t (N_r + N_t - 1)^2 \Big/ L_f}_{(20)} \\ &+ \underbrace{\left(N_r N_t - 1\right) \left(N_r + N_t - 1\right)/L_c}_{(20)}. \end{aligned} \tag{39}$$

The Wiener estimation is calculated by (23)-(29). However, (24) and (29) are constants. Therefore, the computational complexity for a symbol is

$$\begin{aligned} CC^{[\mathrm{M}]}_{\mathrm{Wiener}} = &\underbrace{\left(N_r + N_t - 1\right) N_r N_t / L_f}_{(23)} \\ &+ \underbrace{\left(2L_W + 1\right) \left(N_r + N_t - 1\right)/L_c}_{(25)} \\ &+ \underbrace{\left[\left(2L_W + 1\right)^3 + \left(2L_W + 1\right)\right] \Big/ L_f}_{(26)} + \underbrace{2L_W / L_f}_{(28)}, \end{aligned} \tag{40}$$

$$\begin{aligned} CC^{[\mathrm{A}]}_{\mathrm{Wiener}} = &\underbrace{\left(N_r + N_t - 1\right) \left(N_r N_t - 1\right)/L_f}_{(23)} \\ &+ \underbrace{2L_W \left(N_r + N_t - 1\right)/L_c}_{(25)} \\ &+ \underbrace{\left[\left(2L_W + 1\right) \cdot 2L_W + 2L_W\right]/L_f}_{(26)} + \underbrace{2L_W / L_f}_{(27)}. \end{aligned} \tag{41}$$

The phase recovery and channel estimation is calculated by (30)-(33). Again, $\mathbf{Y}_i \mathbf{S}_i^H$ in (30) can use the intermediate result of (5). Therefore, the computational complexity for a symbol can be represented as

$$\begin{aligned} CC^{[\mathrm{M}]}_{\hat{\mathbf{H}}} = &\underbrace{\left(2N_r N_t - N_r\right)/L_c}_{(30)} + \underbrace{\left(N_r + N_t - 1\right) L_d / L_c}_{(32)} \\ &+ \underbrace{\left(2N_r N_t - N_r\right) L_d / L_c}_{(33)}, \end{aligned} \tag{42}$$

$$CC^{[\mathrm{A}]}_{\hat{\mathbf{H}}} = \underbrace{N_r N_t / L_c}_{(31)} + \underbrace{\left(N_r + N_t - 1\right) \left(1 + 2L_d\right)/L_c}_{(32)}. \tag{43}$$

Combining (36)-(43), the overall computational complexity for each symbol is

$$\begin{aligned} CC = &C_M \left(CC^{[\mathrm{M}]}_{|\hat{\mathbf{H}}|} + CC^{[\mathrm{M}]}_{\mathrm{WLLS}} + CC^{[\mathrm{M}]}_{\mathrm{Wiener}} + CC^{[\mathrm{M}]}_{\hat{\mathbf{H}}} \right) \\ &+ \left(CC^{[\mathrm{A}]}_{|\hat{\mathbf{H}}|} + CC^{[\mathrm{A}]}_{\mathrm{WLLS}} + CC^{[\mathrm{A}]}_{\mathrm{Wiener}} + CC^{[\mathrm{A}]}_{\hat{\mathbf{H}}} \right), \end{aligned} \tag{44}$$

where C_M is the weighing coefficient for multiplication, indicating that the multiplication is much more complex than addition.

CRAMÉR–RAO LOWER BOUND

A. CRLB for One-shot Phase Estimation

An important question is how accurate the phase of different transmit and receive laser sources can be estimated in one pilot group (which is defined in assumption (A1)). Without loss of generality, the 1^{st} pilot group is considered in this subsection, and all the corresponding indices are restricted to the 1^{st} pilot group (i.e. $i = 1$, $1 \le m \le N_t$, $m_i = m_1 = \lceil N_t/2 \rceil$, $\mathbf{Y} = \mathbf{Y}_1$, $\mathbf{S} = \mathbf{S}_1$, and $\mathbf{N} = \mathbf{N}_1$).

As shown in Appendix A, the information of φ_{k,m_i} and ψ_{l,m_i} is only included in the angular terms of $\mathbf{Y}\mathbf{S}^H$, which are also the observed data of the CRLB in this subsection.

For the 1^{st} pilot group, (2) can be rewritten as

$$\begin{aligned} y_{k,m} &= \sum_{l=1}^{N_t} h_{k,l} e^{j\left(\varphi_{k,m} + \psi_{l,m}\right)} s_{l,m} + n_{k,m} \\ &= \sum_{l=1}^{N_t} h_{k,l} e^{j\left(\varphi_{k,m_i} + \psi_{l,m_i} + \gamma_{k,l,m}\right)} s_{l,m} + n_{k,m}, \end{aligned} \tag{45}$$

where

$$\gamma_{k,l,m} = (\varphi_{k,m} - \varphi_{k,m_i}) + (\psi_{l,m} - \psi_{l,m_i}) = \begin{cases} -\sum_{m'=m+1}^{m_i} (\Delta\varphi_{k,m'} + \Delta\psi_{l,m'}), & (m < m_i) \\ \sum_{m'=m_i+1}^{m} (\Delta\varphi_{k,m'} + \Delta\psi_{l,m'}), & (m > m_i) \\ 0, & (m = m_i). \end{cases} \tag{46}$$

For practical laser sources, the phase noise innovations are small [15], and the approximation below holds

$$e^{j\gamma_{k,l,m}} \approx 1 + j\gamma_{k,l,m}. \tag{47}$$

Therefore, (45) can be approximated by

$$\begin{aligned} \mathbf{y}_{k,:} \approx & \sum_{l=1}^{N_t} h_{k,l} e^{j\left(\varphi_{k,m_i} + \psi_{l,m_i}\right)} \mathbf{s}_{l,:} \\ & + j \sum_{l=1}^{N_t} h_{k,l} e^{j\left(\varphi_{k,m_i} + \psi_{l,m_i}\right)} \left(\mathbf{s}_{l,:} \odot \boldsymbol{\gamma}_{k,l,:}\right) + \mathbf{n}_{k,:}. \end{aligned} \tag{48}$$

Note the fact that $\mathbf{YS}^H$ is an orthogonal transformation of the observed data $\mathbf{Y}$, it does not change any information if the observed data is given by $\mathbf{YS}^H$. Moreover, as shown in Appendix A, the information of φ_{k,m_i} and ψ_{l,m_i} is only included in the angular terms of $\mathbf{YS}^H$. When the perfect channel estimation is assumed, which is the optimal case of the phase estimation, the element at the k^{th} row and l^{th} column of the angular terms of the observed data matrix $\mathbf{O}$ can be given by Appendix A as

$$\begin{aligned} o_{k,l} = & \varphi_{k,m_i} + \psi_{l,m_i} + \frac{n_{k,l}^{(2)}}{|h_{k,l}|} \\ & - \sum_{l'=1}^{N_t} \sum_{m'=2}^{m_i} \sum_{m=1}^{m'-1} \xi_{k,l,l',m} \left(\Delta\varphi_{k,m'} + \Delta\psi_{l',m'}\right) \\ & + \sum_{l'=1}^{N_t} \sum_{m'=m_i+1}^{N_t} \sum_{m=m'}^{N_t} \xi_{k,l,l',m} \left(\Delta\varphi_{k,m'} + \Delta\psi_{l',m'}\right), \end{aligned} \tag{49}$$

where $n_{k,l}^{(2)} \sim \mathcal{N}\left(0, \sigma_n^2/(2N_t)\right)$ is i.i.d. real AWGN, and

$$\xi_{k,l,l',m} \triangleq \Re\left(\frac{h_{k,l'}}{N_t h_{k,l}} e^{j\left(\psi_{l',m_i} - \psi_{l,m_i}\right)} s_{l',m} s_{l,m}^*\right). \tag{50}$$

where $\Re\{\cdot\}$ is the real part of a complex number, and $(\cdot)^*$ is the complex conjugate operator. It is also worth noting that the phase noise terms in (49) are from the symbols before and after the reference index m_i, respectively.

In order to calculate the Fisher information matrix, it is necessary to rearrange the angular terms of the observed data to vector form as

$$\boldsymbol{\upsilon} = \left[\mathbf{o}_{1,:}, \mathbf{o}_{2,:}, \cdots, \mathbf{o}_{N_r,:}\right]^T. \tag{51}$$

The expected value of $\boldsymbol{\upsilon}$ is denoted by $\boldsymbol{\mu}_{\boldsymbol{\upsilon}}$. Note the fact that all the noise terms in (49) are zero mean i.i.d. AWGN, the elements of $\boldsymbol{\mu}_{\boldsymbol{\upsilon}}$ can therefore be represented as

$$\mu_{\upsilon(N_t(k-1)+l)} = \varphi_{k,m_i} + \psi_{l,m_i}. \tag{52}$$

Considering that the variables to be estimated are given by (13), (52) can be rewritten as

$$\mu_{\upsilon(N_t(k-1)+l)} = \begin{cases} \beta_{k,m_i} + \beta_{N_r+l,m_i}, & (l \neq N_t) \\ \beta_{k,m_i}, & (l = N_t), \end{cases} \tag{53}$$

and the first order derivative of $\boldsymbol{\mu}_{\boldsymbol{\upsilon}}$ can be calculated as

$$\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\upsilon}}}{\partial \beta_{q,m_i}} = \begin{cases} \left[\underbrace{0, \cdots, 0}_{(q-1)N_t}, \underbrace{1, \cdots, 1}_{N_t}, \underbrace{0, \cdots, 0}_{(N_r-q)N_t}\right]^T, & (q \leq N_r) \\ \left[\overbrace{\underbrace{0, \cdots, 0}_{q-N_r-1}, 1, \underbrace{0, \cdots, 0}_{N_r+N_t-q}, \cdots}^{\text{repeat } N_r \text{ times}}\right]^T, & (q > N_r). \end{cases} \tag{54}$$

On the other hand, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}$ of $\boldsymbol{\upsilon}$ is defined as

$$\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}} \triangleq E\left[\left(\boldsymbol{\upsilon} - \boldsymbol{\mu}_{\boldsymbol{\upsilon}}\right)\left(\boldsymbol{\upsilon} - \boldsymbol{\mu}_{\boldsymbol{\upsilon}}\right)^T\right]. \tag{55}$$

where $\mathrm{E}(\cdot)$ denotes the expected value of a variable.

Moreover, as shown in Appendix B, the element at the $(N_t(k_1-1)+l_1)^{th}$ row and $(N_t(k_2-1)+l_2)^{th}$ column of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}$ has the form of

$$\begin{aligned} & \Sigma_{\upsilon(N_t(k_1-1)+l_1, N_t(k_2-1)+l_2)} \\ & = \frac{\delta(k_1-k_2)\,\delta(l_1-l_2)}{2N_t |h_{k_1,l_1}| |h_{k_2,l_2}|} \sigma_n^2 \\ & + \sum_{m'=2}^{\lceil N_t/2 \rceil} \left[\left(\sum_{m_1=1}^{m'-1} \sum_{l_1'=1}^{N_t} \xi_{k_1,l_1,l_1',m_1} \right) \right. \\ & \left. \times \left(\sum_{m_2=1}^{m'-1} \sum_{l_2'=1}^{N_t} \xi_{k_2,l_2,l_2',m_2} \right) \sigma_{\Delta\varphi}^2 \delta(k_1-k_2) \right] \\ & + \sum_{l'=1}^{N_t} \sum_{m'=2}^{\lceil N_t/2 \rceil} \sum_{m_1=1}^{m'-1} \sum_{m_2=1}^{m'-1} \xi_{k_1,l_1,l',m_1} \xi_{k_2,l_2,l',m_2} \sigma_{\Delta\psi}^2 \\ & + \sum_{m'=\lceil N_t/2 \rceil+1}^{N_t} \left[\left(\sum_{m_1=m'}^{N_t} \sum_{l_1'=1}^{N_t} \xi_{k_1,l_1,l_1',m_1} \right) \right. \\ & \left. \times \left(\sum_{m_2=m'}^{N_t} \sum_{l_2'=1}^{N_t} \xi_{k_2,l_2,l_2',m_2} \right) \sigma_{\Delta\varphi}^2 \delta(k_1-k_2) \right] \\ & + \sum_{l'=1}^{N_t} \sum_{m'=\lceil N_t/2 \rceil+1}^{N_t} \sum_{m_1=m'}^{N_t} \sum_{m_2=m'}^{N_t} \xi_{k_1,l_1,l',m_1} \xi_{k_2,l_2,l',m_2} \sigma_{\Delta\psi}^2, \end{aligned} \tag{56}$$

where $\delta(\cdot)$ is the unit sampling function which is defined as

$$\delta(x) \triangleq \begin{cases} 1, (x = 0) \\ 0, (x \neq 0), \end{cases} \tag{57}$$

and the first order derivative of the covariance matrix can be calculated directly from (56) as (58), where the first derivative

of (50), which is calculated in Appendix C, has the form of

$$\frac{\partial \xi_{k,l,l',m}}{\partial \beta_{q,m_i}} = \begin{cases} -\Im\left(\frac{h_{k,l'}}{N_t h_{k,l}} e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)} s_{l',m} s_{l,m}^*\right), \\ \qquad\qquad (l \neq l' = q - N_r) \\ \Im\left(\frac{h_{k,l'}}{N_t h_{k,l}} e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)} s_{l',m} s_{l,m}^*\right), \\ \qquad\qquad (l' \neq l = q - N_r) \\ 0, \qquad\qquad \text{otherwise}, \end{cases} \tag{59}$$

where $\Im\{\cdot\}$ is the imaginary part of a complex number.

Using (54), (56), and (58), the element at the ${q_1}^{th}$ row and ${q_2}^{th}$ column of the Fisher information matrix $\mathbf{FIM}$ can be represented by the classic form as [28, p. 47, (3.31)]

$$\mathrm{FIM}_{q_1,q_2} = \frac{\partial \boldsymbol{\mu}_{\boldsymbol{v}}^T}{\partial \beta_{q_1}} \boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{v}}}{\partial \beta_{q_2}} + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{v}}}{\partial \beta_{q_1}} \boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{v}}}{\partial \beta_{q_2}}\right), \tag{60}$$

and the CRLB for one-shot phase estimation can be calculated as [28, p. 44, (3.24)]

$$\mathbf{CRLB}(\boldsymbol{\beta}) = \mathrm{diag}\left(\mathbf{FIM}^{-1}\right), \tag{61}$$

where $\mathrm{diag}(\mathbf{X})$ denotes a vector which contains the diagonal elements of matrix $\mathbf{X}$.

Remark 6: The CRLB derived in this subsection depends only on the angular information of the observed data. Considering that most known practical phase estimation algorithms only use phase information of the observed data for the phase estimation, this CRLB gives a practical lower bound for the phase estimation in MIMO systems.

B. CRLB for Element-wise Wiener Estimation

Although the ultimate bound of the phase estimation performance in the MIMO system is very difficult to obtain, a very useful CRLB for element-wise Wiener estimation can be obtained to evaluate the performance of the proposed algorithm.

Note the fact that Wiener estimator is the optimal filter for the element-wise phase estimation problem [15], [27], the CRLB is equal to the MSE performance of a two-sided infinite impulse response (IIR) Wiener filter.

Without loss of generality, we only consider the MSE performance of the Wiener estimation for the first pilot group. For the q^{th} IIR Wiener filter, $\sigma_{pW(q)}^2$ is given by (24), $\sigma_{\beta(q)}^2 = \mathbf{CRLB}(\beta_q)$ is given by (61), and the coefficient of the i^{th} pilot group can be rewritten from (26) as [15]

$$\omega_{i,q} = \begin{cases} \frac{\kappa\tau}{1-\kappa^2}\kappa^{i-1}, (i \geq 1) \\ \frac{\kappa\tau}{1-\kappa^2}\kappa^{1-i}, (i < 1), \end{cases} \tag{62}$$

where

$$\kappa = (1+\tau/2) - \sqrt{(1+\tau/2)^2 - 1}, \tag{63}$$

$$\tau = \sigma_{pW(q)}^2 / \sigma_{\beta(q)}^2. \tag{64}$$

By setting $L_W \to +\infty$ in (25), and noting that the phase noise is a stationary process with independent increments, the lower bound of the MSE performance of the Wiener estimator can be written as

$$\sigma_{W(q)}^2 = \sum_{i=-\infty}^{+\infty} \omega_{i,q}^2 \sigma_{\beta(q)}^2 + \sum_{i=2}^{\infty}\left(\sum_{i'=i}^{\infty}\omega_{i',q}\right)^2 \sigma_{pW(q)}^2 + \sum_{i=-\infty}^{0}\left(\sum_{i'=-\infty}^{i}\omega_{i',q}\right)^2 \sigma_{pW(q)}^2. \tag{65}$$

By substituting (62) into (65) and calculating the summation, the following equation holds

$$\sigma_{W(q)}^2 = \sigma_{\beta(q)}^2 \frac{\kappa^2\tau^2\left(1+\kappa^2\right)}{(1+\kappa)^3(1-\kappa)^3} + 2\sigma_{pW(q)}^2 \frac{\kappa^4\tau^2}{(1+\kappa)^3(1-\kappa)^5}. \tag{66}$$

By substituting (63) and (64) into (66), $\sigma_{W(q)}^2$ can be represented by $\sigma_{\beta(q)}^2$ and $\sigma_{pW(q)}^2$. Moreover, the equation can be simplified by Mathematica™ 11.2.0.0 as

$$\sigma_{W(q)}^2 = \left[\frac{4}{\sigma_{pW(q)}^2\sigma_{\beta(q)}^2} + \frac{1}{\sigma_{\beta(q)}^4}\right]^{-\frac{1}{2}}. \tag{67}$$

Wiener filter is the optimal filter for the element-wise phase estimation problem [28]. Noting that the noise MSE in (67) is given by (24) and (61), (67) is also the CRLB for the element-wise Wiener estimators.

$$\begin{aligned} &\frac{\partial \Sigma_{v(N_t(k_1-1)+l_1,N_t(k_2-1)+l_2)}}{\partial \beta_{q,m_i}} \\ &= \sum_{m'=2}^{\lceil N_t/2\rceil}\sum_{m_1=1}^{m'-1}\sum_{l_1'=1}^{N_t}\sum_{m_2=1}^{m'-1}\sum_{l_2'=1}^{N_t}\left(\frac{\partial \xi_{k_1,l_1,l_1',m_1}}{\partial \beta_{q,m_i}}\xi_{k_2,l_2,l_2',m_2} + \xi_{k_1,l_1,l_1',m_1}\frac{\partial \xi_{k_2,l_2,l_2',m_2}}{\partial \beta_{q,m_i}}\right)\sigma_{\Delta\varphi}^2\delta(k_1-k_2) \\ &+ \sum_{l'=1}^{N_t}\sum_{m'=2}^{\lceil N_t/2\rceil}\sum_{m_1=1}^{m'-1}\sum_{m_2=1}^{m'-1}\left(\frac{\partial \xi_{k_1,l_1,l',m_1}}{\partial \beta_{q,m_i}}\xi_{k_2,l_2,l',m_2} + \xi_{k_1,l_1,l',m_1}\frac{\partial \xi_{k_2,l_2,l',m_2}}{\partial \beta_{q,m_i}}\right)\sigma_{\Delta\psi}^2 \\ &+ \sum_{m'=\lceil N_t/2\rceil+1}^{N_t}\sum_{m_1=m'}^{N_t}\sum_{l_1'=1}^{N_t}\sum_{m_2=m'}^{N_t}\sum_{l_2'=1}^{N_t}\left(\frac{\partial \xi_{k_1,l_1,l_1',m_1}}{\partial \beta_{q,m_i}}\xi_{k_2,l_2,l_2',m_2} + \xi_{k_1,l_1,l_1',m_1}\frac{\partial \xi_{k_2,l_2,l_2',m_2}}{\partial \beta_{q,m_i}}\right)\sigma_{\Delta\varphi}^2\delta(k_1-k_2) \\ &+ \sum_{l'=1}^{N_t}\sum_{m'=\lceil N_t/2\rceil+1}^{N_t}\sum_{m_1=m'}^{N_t}\sum_{m_2=m'}^{N_t}\left(\frac{\partial \xi_{k_1,l_1,l',m_1}}{\partial \beta_{q,m_i}}\xi_{k_2,l_2,l',m_2} + \xi_{k_1,l_1,l',m_1}\frac{\partial \xi_{k_2,l_2,l',m_2}}{\partial \beta_{q,m_i}}\right)\sigma_{\Delta\psi}^2. \end{aligned} \tag{58}$$

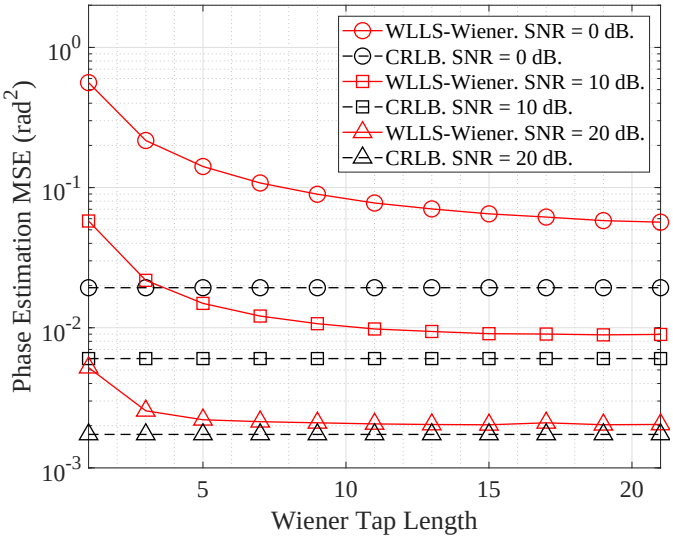

Fig. 15. MSE performance of WLLS-Wiener phase estimation (red lines) and corresponding CRLB (black dashed lines) in a 2×2 MIMO system. The influence of Wiener filter tap length ($L_{tap} = 1, 3, 5, ..., 21$) at SNR $= 0$ (circles), 10 (squares), and 20 dB (triangles).

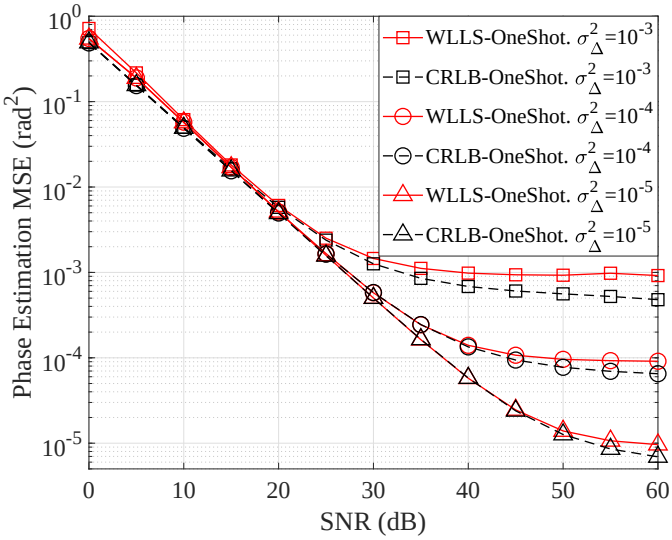

Fig. 16. MSE performance of WLLS one-shot phase estimation (red lines) and corresponding CRLB (black dashed lines) in a 2×2 MIMO system. $\sigma^2_\Delta = 10^{-3}$ (squares), 10^{-4} (circles), 10^{-5} (triangles).

Detailed Simulation Results

C. Further Details on Phase Estimation

In order to quantify the influence of the Wiener filter tap length, Fig. 15 compares the phase estimation MSE performance of WLLS-Wiener phase estimator for different Wiener filter tap lengths ($L_{tap} = 1, 3, 5, ..., 21$). As shown in Fig. 15, when the SNR is lower, the effective filter length will be larger and the curve converges slower. When the filter tap length is small, we can observe more MSE performance degradation due to the inadequate tap length. When the filter length is large and the curve converges, an MSE performance degradation is observed, and the performance degradation is larger when the SNR is lower. We believe this is mainly due to the small angle assumption in the derivation.

Fig. 5 indicates that the MSE has a minimum value, or floor in the High SNR region. In order to verify the origin of this floor, Fig. 16 compares the phase estimation MSE performance of the WLLS one-shot phase estimator for different phase noise variances ($\sigma^2_\Delta = 10^{-3}, 10^{-4}, 10^{-5}$). In the low SNR region, the MSE is similar as it is dominated by AWGN. In the high SNR region, the MSE floor is dominated by intra-pilot-group phase noise, which is proportional to the phase noise variance. Therefore, the MSE floor is also proportional to the phase noise variance. Moreover, similar to Fig. 5, the MSE penalty in the very high SNR region is due to neglecting the intra-pilot-group phase noise in the WLLS estimator.

Fig. 17 compares the phase estimation MSE performance of the WLLS-Wiener phase estimator for different phase noise variances ($\sigma^2_\Delta = 10^{-3}, 10^{-4}, 10^{-5}$). Unlike Fig. 16, where similar performance is observed for all phase noise variances in the low SNR region, a phase noise dependent MSE performance is observed. This is because the effective length of the Wiener filter is shorter when the phase noise is larger, and so there is less noise averaging. Again, the MSE

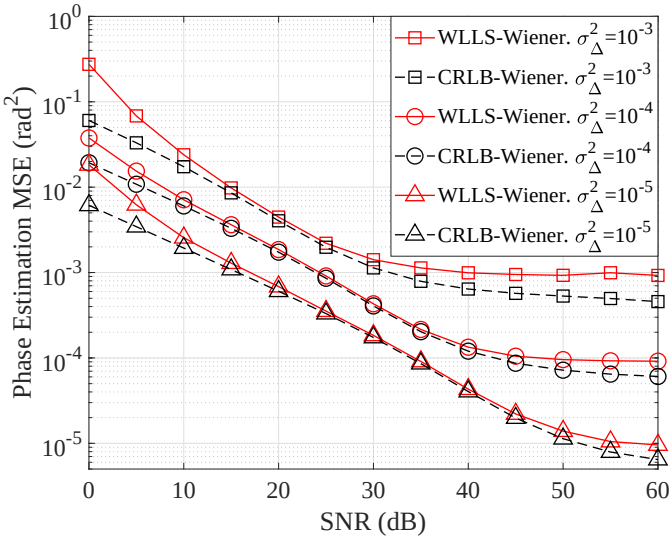

Fig. 17. MSE performance of WLLS-Wiener phase estimation (red lines) and corresponding CRLB (black dashed lines) in a 2×2 MIMO system. $\sigma^2_\Delta = 10^{-3}$ (squares), 10^{-4} (circles), 10^{-5} (triangles).

degradation in the low SNR region is mainly due to the small angle approximation.

D. Further Details on BER performance

Fig. 18 compares the BER performance of the proposed phase and channel estimation algorithm for different phase noise variances ($\sigma^2_\Delta = 10^{-3}, 10^{-4}, 10^{-5}$). In the low SNR region, although the phase noise dependent MSE performance is observed in Fig. 17, we only observe very small BER penalty in Fig. 18. This is because the dominant factor in this region is AWGN rather than phase estimation error. In the high SNR region, the BER floor is lower when the phase noise is smaller, which is a direct consequence of the MSE floor in Fig. 17

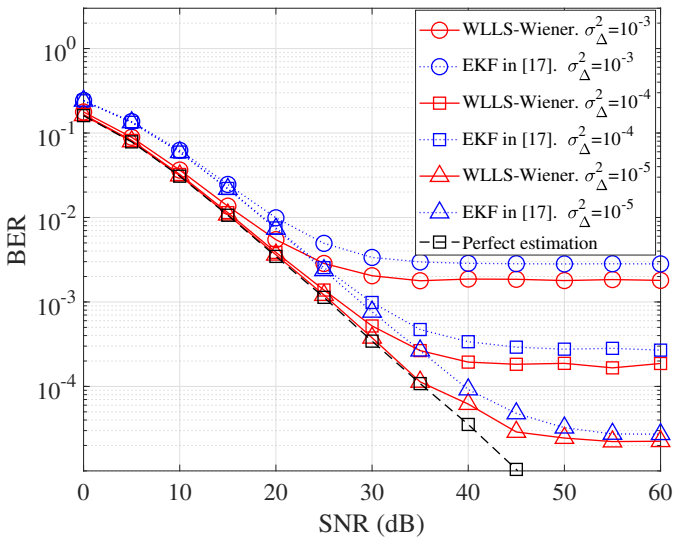

Fig. 18. BER of a 2×2 MIMO system. $\sigma^2_\Delta = 10^{-3}$ (circles), 10^{-4} (squares), 10^{-5} (triangles). Red lines: proposed algorithm; blue dotted lines: EKF in [17]; black dashed lines: perfect phase and channel estimation.

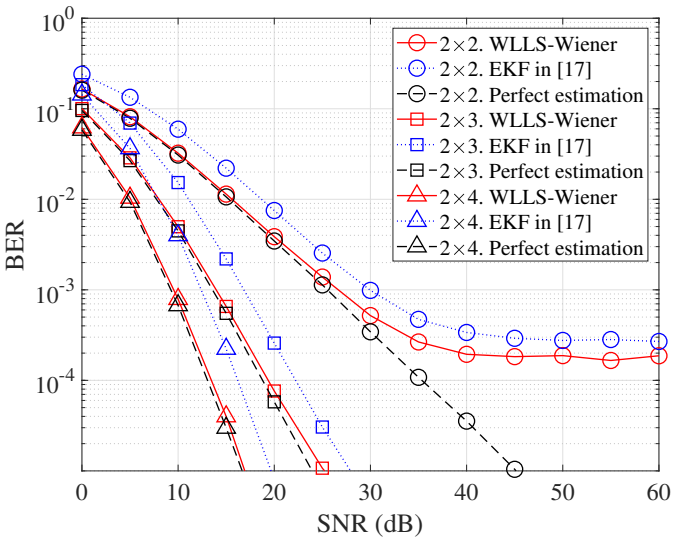

Fig. 20. BER of different MIMO systems. 2×2 (circles), 2×3 (squares), and 2×4 (triangles). Red lines: proposed algorithm; blue dotted lines: EKF in [17]; black dashed lines: perfect phase and channel estimation.

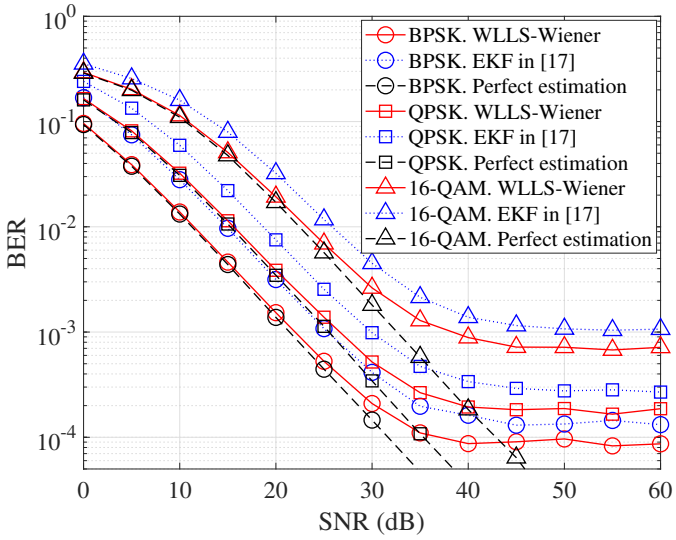

Fig. 19. BER of a 2×2 MIMO system. Different modulation formats (BPSK (circles), QPSK (squares), and 16-QAM (triangles)). Red lines: proposed algorithm; blue dotted lines: EKF in [17], black dashed lines: perfect phase and channel estimation.

In order to verify the compatibility with higher modulation formats such as quadrature amplitude modulation (QAM), Fig. 19 compares the BER performance of the proposed phase and channel estimation algorithm for BPSK, quadrature phase shift keying (QPSK), and 16-QAM. A higher modulation format is more vulnerable to the phase error, leading to a higher BER floor. Compared with the perfect phase and channel estimation scenario at the HD-FEC limit, the proposed algorithm has an SNR penalty of approximately 0.3 dB, 0.5 dB, and 1.2 dB, while the conventional algorithm has an SNR penalty of approximately 3.5 dB, 3.5 dB, and 4.0 dB for BPSK, QPSK, and 16-QAM, respectively.

Fig. 20 compares the BER performance of the proposed phase and channel estimation algorithm for different MIMO systems (2×2, 2×3, and 2×4. The results indicate that for a fixed transmit antenna number (N_t), a larger receive antenna number (N_r) can lead to a better BER performance. This is because the diversity order of MMSE MIMO decoder is $N_r - N_t + 1$, and a larger diversity order leads to a better performance for the reference system with perfect estimation [11]. Moreover, a lower BER floor is also observed when N_r is larger. We believe redundant degrees of freedom in MIMO systems can lead to better resistance to the imperfect phase and channel estimations. Compared with the perfect phase and channel estimation scenario at the HD-FEC limit, the proposed algorithm has an SNR penalty of approximately 0.5 dB, 0.3 dB, and 0.2 dB, while the conventional algorithm has an SNR penalty of approximately 3.6 dB, 3.2 dB, and 3.3 dB for 2×2, 2×3, and 2×4 MIMO systems, respectively.

APPENDIX A
THE ANGULAR TERMS OF THE OBSERVED DATA

Consider the observed data below

$$\mathbf{O}' = \frac{1}{N_t}\mathbf{Y}\mathbf{S}^H. \tag{68}$$

The element at the k^{th} row and l^{th} column of $\mathbf{O}'$ can be calculated by (48) as

$$\begin{aligned} o'_{k,l} =& \frac{1}{N_t}\mathbf{y}_{k,:}\mathbf{s}^H_{l,:} \\ \approx& \frac{1}{N_t}\sum_{l'=1}^{N_t} h_{k,l'}e^{j\left(\varphi_{k,m_i}+\psi_{l',m_i}\right)}\mathbf{s}_{l',:}\mathbf{s}^H_{l,:} + \frac{1}{N_t}\mathbf{n}_{k,:}\mathbf{s}^H_{l,:} \\ &+ j\frac{1}{N_t}\sum_{l'=1}^{N_t} h_{k,l'}e^{j\left(\varphi_{k,m_i}+\psi_{l',m_i}\right)}\left(\mathbf{s}_{l',:}\odot\boldsymbol{\gamma}_{k,l',:}\right)\mathbf{s}^H_{l,:}. \end{aligned} \tag{69}$$

Note the fact that $\mathbf{S}$ is an orthogonal matrix with normalized elements, (69) can be simplified as

$$o'_{k,l} \approx h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)} + n^{(1)}_{k,l} + j \sum_{l'=1}^{N_t} \left[\frac{h_{k,l'} e^{j\left(\varphi_{k,m_i}+\psi_{l',m_i}\right)}}{N_t} \sum_{m=1}^{N_t} \left(s_{l',m} s^*_{l,m} \gamma_{k,l',m} \right) \right], \tag{70}$$

where $n^{(1)}_{k,l} \sim \mathcal{CN}\left(0, \sigma_n^2/N_t\right)$ is i.i.d. circularly-symmetric complex AWGN.

Define

$$\eta_{k,l,l',m} \triangleq \frac{h_{k,l'}}{N_t h_{k,l}} e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)} s_{l',m} s^*_{l,m}. \tag{71}$$

Equation (70) can be represented as

$$\begin{aligned} o'_{k,l} \approx\ & h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)} \\ & + h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)} \cdot \frac{n^{(1)}_{k,l}}{h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)}} \\ & + h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)} j \sum_{l'=1}^{N_t} \sum_{m=1}^{N_t} \eta_{k,l,l',m} \gamma_{k,l',m}. \end{aligned} \tag{72}$$

When the complex term X is assumed to be small, below approximation holds

$$1 + X \approx (1 + \Re(X)) e^{j\Im(X)}. \tag{73}$$

By extracting the term $h_{k,l} e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i}\right)}$ in (72) and using (73), (72) can be rewritten as

$$\begin{aligned} o'_{k,l} \approx\ & h_{k,l} \left(1 + \frac{n^{(3)}_{k,l}}{|h_{k,l}|} + \sum_{l'=1}^{N_t} \sum_{m=1}^{N_t} \Im\left(\eta_{k,l,l',m}\right) \gamma_{k,l',m} \right) \\ & \times e^{j\left(\varphi_{k,m_i}+\psi_{l,m_i} + \frac{n^{(2)}_{k,l}}{|h_{k,l}|} + \sum\limits_{l'=1}^{N_t} \sum\limits_{m=1}^{N_t} \Re\left(\eta_{k,l,l',m}\right) \gamma_{k,l',m} \right)}, \end{aligned} \tag{74}$$

where

$$\begin{cases} n^{(2)}_{k,l} = \Im\left(n^{(1)}_{k,l} e^{-j\left(\angle h_{k,l}+\varphi_{k,m_i}+\psi_{l,m_i}\right)} \right) \\ n^{(3)}_{k,l} = \Re\left(n^{(1)}_{k,l} e^{-j\left(\angle h_{k,l}+\varphi_{k,m_i}+\psi_{l,m_i}\right)} \right). \end{cases} \tag{75}$$

are i.i.d. real AWGN with zero mean and variance $\sigma_n^2/2N_t$.

Noting the fact that $n^{(1)}_{k,l}$ is i.i.d. circularly-symmetric complex AWGN, the exponential term in (75) does not change the distribution of $n^{(2)}_{k,l}$ and $n^{(3)}_{k,l}$. Therefore, the information of φ_{k,m_i} and ψ_{l,m_i} is only included in the angular term of (74). If perfect channel estimation is assumed, which is the optimal case of the phase estimation, the effective observed data for the phase estimation is the angular term of (74), which can be written as

$$o_{k,l} = \varphi_{k,m_i} + \psi_{l,m_i} + \frac{n^{(2)}_{k,l}}{|h_{k,l}|} + \sum_{l'=1}^{N_t} \sum_{m=1}^{N_t} \xi_{k,l,l',m} \gamma_{k,l',m}, \tag{76}$$

where $\xi_{k,l,l',m} = \Re\left(\eta_{k,l,l',m}\right)$ is given by (50).

By substituting (46) into (76), the equation can be rewritten as

$$\begin{aligned} o_{k,l} =\ & \varphi_{k,m_i} + \psi_{l,m_i} + \frac{n^{(2)}_{k,l}}{|h_{k,l}|} \\ & - \sum_{l'=1}^{N_t} \sum_{m=1}^{m_i-1} \sum_{m'=m+1}^{m_i} \xi_{k,l,l',m} \left(\Delta\varphi_{k,m'} + \Delta\psi_{l',m'} \right) \\ & + \sum_{l'=1}^{N_t} \sum_{m=m_i+1}^{N_t} \sum_{m'=m_i+1}^{m} \xi_{k,l,l',m} \left(\Delta\varphi_{k,m'} + \Delta\psi_{l',m'} \right), \end{aligned} \tag{77}$$

and (49) can be directly obtained by changing the order of summation in (77).

APPENDIX B
DERIVATION OF (56)

Equation (55) can be rewritten in element-wise form as

$$\begin{aligned} & \Sigma_{\upsilon(N_t(k_1-1)+l_1, N_t(k_2-1)+l_2)} \\ & = \mathrm{E}\left[\left(o_{k_1,l_1} - \mu_{\upsilon(N_t(k_1-1)+l_1)} \right) \left(o_{k_2,l_2} - \mu_{\upsilon(N_t(k_2-1)+l_2)} \right) \right], \end{aligned} \tag{78}$$

Noting the fact that all the phase noise increment and AWGN are mutually independent random variables, the cross-covariance between any two different noise sources is zero. By substituting (49) and (53) into (78), the 1^{st} non-zero term of (78) can be written as

$$\begin{aligned} \mathrm{E}\left(\frac{n^{(2)}_{k_1,l_1}}{|h_{k_1,l_1}|} \frac{n^{(2)}_{k_2,l_2}}{|h_{k_2,l_2}|} \right) &= \frac{1}{|h_{k_1,l_1}|\,|h_{k_2,l_2}|} \mathrm{E}\left(n^{(2)}_{k_1,l_1} n^{(2)}_{k_2,l_2} \right) \\ &= \frac{\sigma_n^2 \delta(k_1-k_2)\,\delta(l_1-l_2)}{2N_t\,|h_{k_1,l_1}|\,|h_{k_2,l_2}|}. \end{aligned} \tag{79}$$

The 2^{nd} non-zero term of (78) can be written as

$$\begin{aligned} & \mathrm{E}\left(- \sum_{m'_1=2}^{m_i} \left(\sum_{m_1=1}^{m'_1-1} \sum_{l'_1=1}^{N_t} \xi_{k_1,l_1,l'_1,m_1} \right) \Delta\varphi_{k_1,m'_1} \right. \\ & \left. \times \left[- \sum_{m'_2=2}^{m_i} \left(\sum_{m_2=1}^{m'_2-1} \sum_{l'_2=1}^{N_t} \xi_{k_2,l_2,l'_2,m_2} \right) \Delta\varphi_{k_2,m'_2} \right] \right) \\ & = \sum_{m'=2}^{m_i} \left[\left(\sum_{m_1=1}^{m'-1} \sum_{l'_1=1}^{N_t} \xi_{k_1,l_1,l'_1,m_1} \right) \right. \\ & \left. \times \left(\sum_{m_2=1}^{m'-1} \sum_{l'_2=1}^{N_t} \xi_{k_2,l_2,l'_2,m_2} \right) \sigma^2_{\Delta\varphi} \delta(k_1-k_2) \right]. \end{aligned} \tag{80}$$

The 3^{rd} non-zero term of (78) can be written as

$$\begin{aligned} & \mathrm{E}\left(\left[- \sum_{l_1{}'=1}^{N_t} \sum_{m_1{}'=2}^{m_i} \left(\sum_{m_1=1}^{m_1{}'-1} \xi_{k_1,l_1,l_1{}',m_1} \right) \Delta\psi_{l_1{}',m_1{}'} \right] \right. \\ & \left. \times \left[- \sum_{l_2{}'=1}^{N_t} \sum_{m_2{}'=2}^{m_i} \left(\sum_{m_2=1}^{m_2{}'-1} \xi_{k_2,l_2,l_2{}',m_2} \right) \Delta\psi_{l_2{}',m_2{}'} \right] \right) \\ & = \sum_{l'=1}^{N_t} \sum_{m'=2}^{m_i} \sum_{m_1=1}^{m'-1} \sum_{m_2=1}^{m'-1} \xi_{k_1,l_1,l_1{}',m_1} \xi_{k_2,l_2,l_2{}',m_2} \sigma^2_{\Delta\psi}. \end{aligned} \tag{81}$$

Similar to (80), the 4^{th} non-zero term of (78) can be written as

$$\begin{aligned}
&\mathrm{E}\left[\sum_{m_1'=m_i+1}^{N_t}\left(\sum_{m_1=m_1'}^{N_t}\sum_{l_1'=1}^{N_t}\xi_{k_1,l_1,l_1',m_1}\right)\Delta\varphi_{k_1,m_1'}\right.\\
&\left.\times\sum_{m_2'=m_i+1}^{N_t}\left(\sum_{m_2=m_2'}^{N_t}\sum_{l_2'=1}^{N_t}\xi_{k_2,l_2,l_2',m_2}\right)\Delta\varphi_{k_2,m_2'}\right]\\
&=\sum_{m'=m_i+1}^{N_t}\left[\left(\sum_{m_1=m'}^{N_t}\sum_{l_1'=1}^{N_t}\xi_{k_1,l_1,l_1',m_1}\right)\right.\\
&\left.\times\left(\sum_{m_2=m'}^{N_t}\sum_{l_2'=1}^{N_t}\xi_{k_2,l_2,l_2',m_2}\right)\sigma_{\Delta\varphi}^2\delta\left(k_1-k_2\right)\right].
\end{aligned}\tag{82}$$

Similar to (81), the 5^{th} non-zero term of (78) can be written as

$$\begin{aligned}
&\mathrm{E}\left(\left[\sum_{l_1'=1}^{N_t}\sum_{m_1'=m_i+1}^{N_t}\left(\sum_{m_1=m_1'}^{N_t}\xi_{k_1,l_1,l_1',m_1}\right)\Delta\psi_{l_1',m_1'}\right]\right.\\
&\left.\times\left[\sum_{l_2'=1}^{N_t}\sum_{m_2'=m_i+1}^{N_t}\left(\sum_{m_2=m_2'}^{N_t}\xi_{k_2,l_2,l_2',m_2}\right)\Delta\psi_{l_2',m_2'}\right]\right)\\
&=\sum_{l'=1}^{N_t}\sum_{m'=m_i+1}^{N_t}\sum_{m_1=m'}^{N_t}\sum_{m_2=m'}^{N_t}\xi_{k_1,l_1,l',m_1}\xi_{k_2,l_2,l_2',m_2}\sigma_{\Delta\psi}^2.
\end{aligned}\tag{83}$$

Moreover, all the other terms in (78) are equal to 0, and (56) can be obtained by substituting (79), (80), (81), (82), (83) into (78).

Appendix C
Derivation of (59)

By substituting (13) into (50), the first derivative under different cases can be calculated as below:

When $l'\neq l$, $q=l'+N_r$, and $l\neq N_t$, the first derivative of (50) has the form of

$$\begin{aligned}
\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}&=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\beta_{q,m_i}-\beta_{l+N_r,m_i}\right)}s_{l',m}s_{l,m}^*\right)\\
&=-\Im\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)}s_{l',m}s_{l,m}^*\right).
\end{aligned}\tag{84}$$

When $l'\neq l$, $q=l'+N_r$, and $l=N_t$, the first derivative of (50) has the form of

$$\begin{aligned}
\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}&=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\beta_{q,m_i}}s_{l',m}s_{l,m}^*\right)\\
&=-\Im\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)}s_{l',m}s_{l,m}^*\right).
\end{aligned}\tag{85}$$

When $l'\neq l$, $q=l+N_r$, and $l'\neq N_t$, the first derivative of (50) has the form of

$$\begin{aligned}
\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}&=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\beta_{l'+N_r,m_i}-\beta_{q,m_i}\right)}s_{l',m}s_{l,m}^*\right)\\
&=\Im\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)}s_{l',m}s_{l,m}^*\right).
\end{aligned}\tag{86}$$

When $l'\neq l$, $q=l+N_r$, and $l'=N_t$, the first derivative of (50) has the form of

$$\begin{aligned}
\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}&=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{-j\beta_{q,m_i}}s_{l',m}s_{l,m}^*\right)\\
&=\Im\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)}s_{l',m}s_{l,m}^*\right).
\end{aligned}\tag{87}$$

When $l'\neq l$, $q\neq l+N_r$, and $q\neq l'+N_r$, the first derivative of (50) has the form of

$$\begin{aligned}
\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}&=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}e^{j\left(\psi_{l',m_i}-\psi_{l,m_i}\right)}s_{l',m}s_{l,m}^*\right)\\
&=0.
\end{aligned}\tag{88}$$

When $l'=l$, the first derivative of (50) has the form of

$$\frac{\partial\xi_{k,l,l',m}}{\partial\beta_{q,m_i}}=\frac{\partial}{\partial\beta_{q,m_i}}\Re\left(\frac{h_{k,l'}}{N_th_{k,l}}s_{l',m}s_{l,m}^*\right)=0,\tag{89}$$

and (59) is obtained by directly combining (84)-(89).