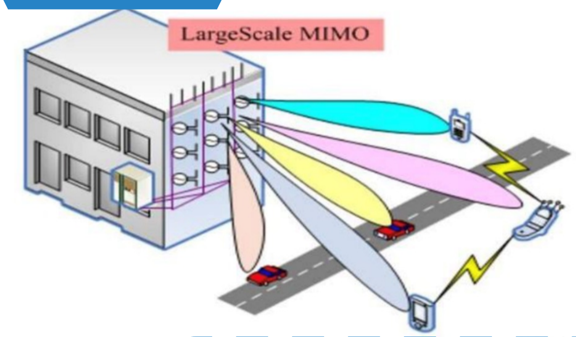
University of Virginia
SCHOOL of ENGINEERING & APPLIED SCIENCE
Charles L. Brown Department of Electrical and Computer Engineering

Calibration of Phase Shifter Network

for Hybrid Beamforming in mmWave Massive MIMO Systems

Xizixiang (Victor) Wei, Ph.D. Candidate, U.Va.

1 RESEARCH BACKGROUND



mmWave and massive MIMO together is key to the 5/6th generation (5G, 6G) wireless communications

mmWave

Tens of GigaHertz of spectrum bands are available in the range of 30–300 GHz and have the potential for future eMBB service.

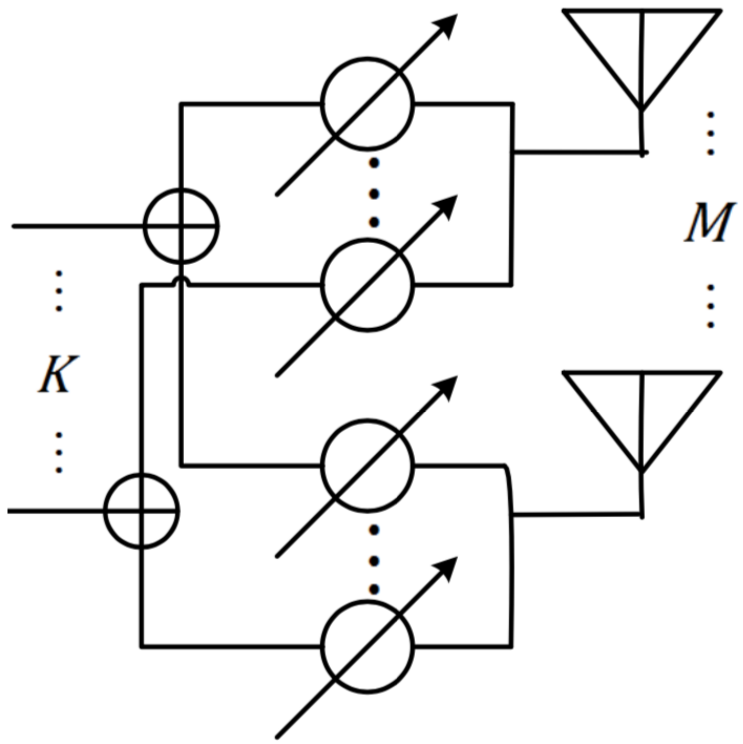
Massive MIMO

The large array gain of massive MIMO can **compensate the severe path loss** of the mmWave through directional beamforming.

Hybrid Beamforming

- Fully-digital beamforming schemes would require too many RF chains, resulting in high cost and high power consumption.
- Hybrid beamforming has been proposed to reap the great gain of massive MIMO with a moderate number of RF chains

1 RESEARCH BACKGROUND



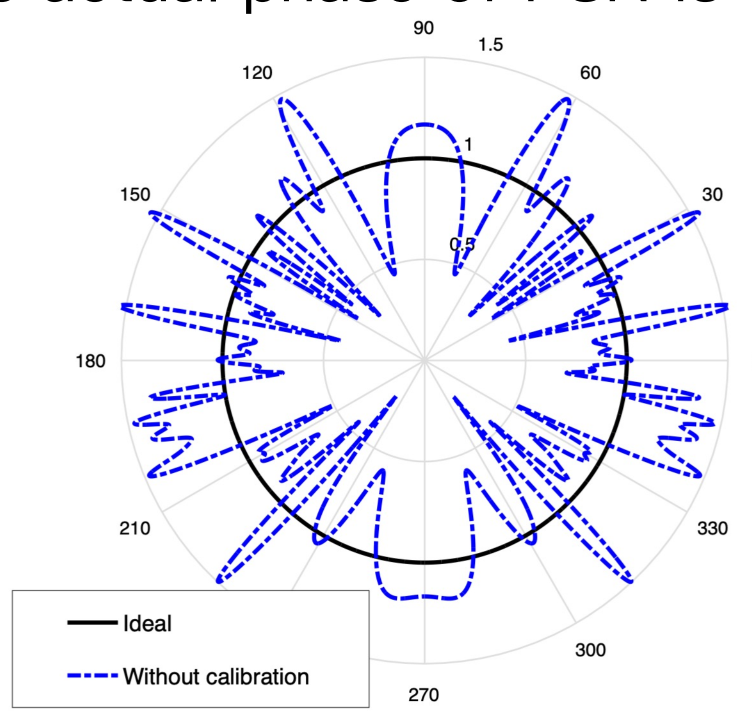
- For hybrid beamforming, BS deploys a **Phase Shifter Network (PSN) $\boldsymbol{\Phi} \in \mathbb{C}^{K\times M}$** between the **K RF chains** and an antenna array **(e.g. a M-element ULA)**

$$\left|\boldsymbol{\phi}^{(km)}\right| = 1 \quad \angle\boldsymbol{\phi}^{(km)} \in [0,2\pi]$$

- There would be distinct **unknown phase deviation** of each shifter due to the time or temperature change

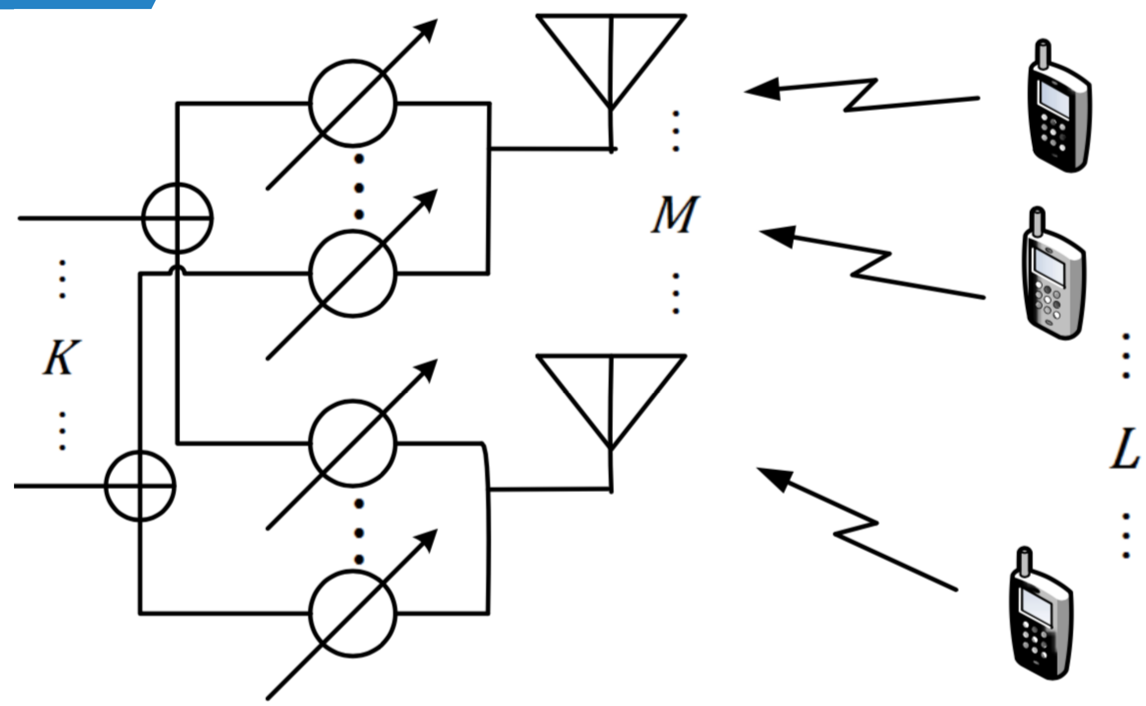
$$w_{km} = e^{j\omega_{km}} \quad \angle\omega_{km} \in [0,2\pi]$$

- The actual phase of PSN is $\boldsymbol{\Phi} \odot \mathbf{W}(\boldsymbol{\Omega})$



- Calibration of PSN is **necessary** for the effective hybrid beamforming
- The core of PSN calibration is the **estimation of the phase deviations**
- Calibration should be done **online** and **over the air**

2 SYSTEM MODEL: Calibration Steps



- K : Number of RF chains
- M: Number of antenna elements
- L: Number of UEs involved for calibration
- N : Number of PSN beamformer

Involved UEs **send training sequence** via different orthogonal channels (e.g. sub–carriers)

↓

BS use **multiple PSN (analog) beamformers** to **receive the training signals** from all involved UEs

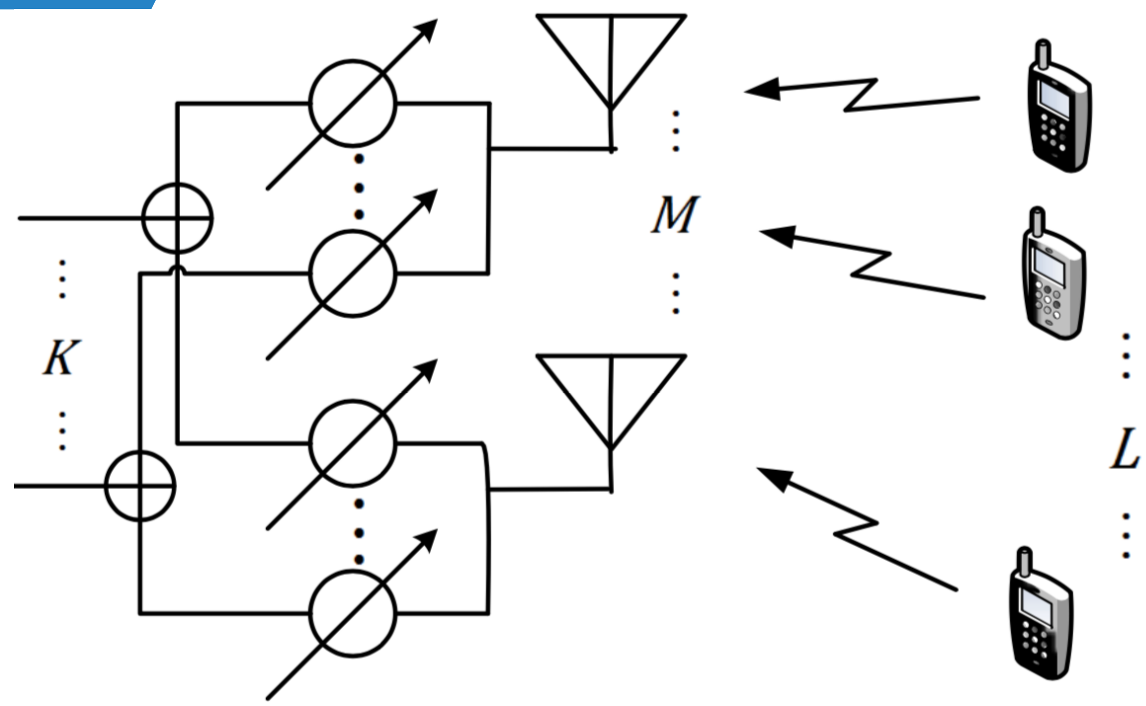
↓

BS **measure** the received signals and use it to **estimate** the PSN phase deviations

↓

BS use the phase deviation estimates to **calibrate** the PSN

2 SYSTEM MODEL: Signal Notations



- K : Number of RF chains
- M: Number of antenna elements
- L: Number of UEs involved for calibration
- N : Number of PSN (analog) beamformer

- Steering vector of a M–element ULA

$$a(\theta) = \left[1, e^{-\frac{j2\pi d \sin(\theta)}{\lambda}}, e^{-\frac{j4\pi d \sin(\theta)}{\lambda}}, \ldots, e^{-\frac{j2(M-1) d \sin(\theta)}{\lambda}}\right]$$

- LOS channel of the l–th user at θ_l:

$$h_l = \gamma_l a(\theta_l)$$

- PSN beamformer $\mathbf{\Phi_n} \in \mathbb{C}^{K \times M}$

$$\left|\boldsymbol{\phi}_{\boldsymbol{n}}^{(km)}\right| = 1 \quad \angle\boldsymbol{\phi}_{\boldsymbol{n}}^{(km)} \in [0, 2\pi]$$

- Phase deviation matrix $\mathbf{W}(\mathbf{\Omega}) \in \mathbb{C}^{K \times M}$

$$w_{km} = e^{j\omega_{km}}$$

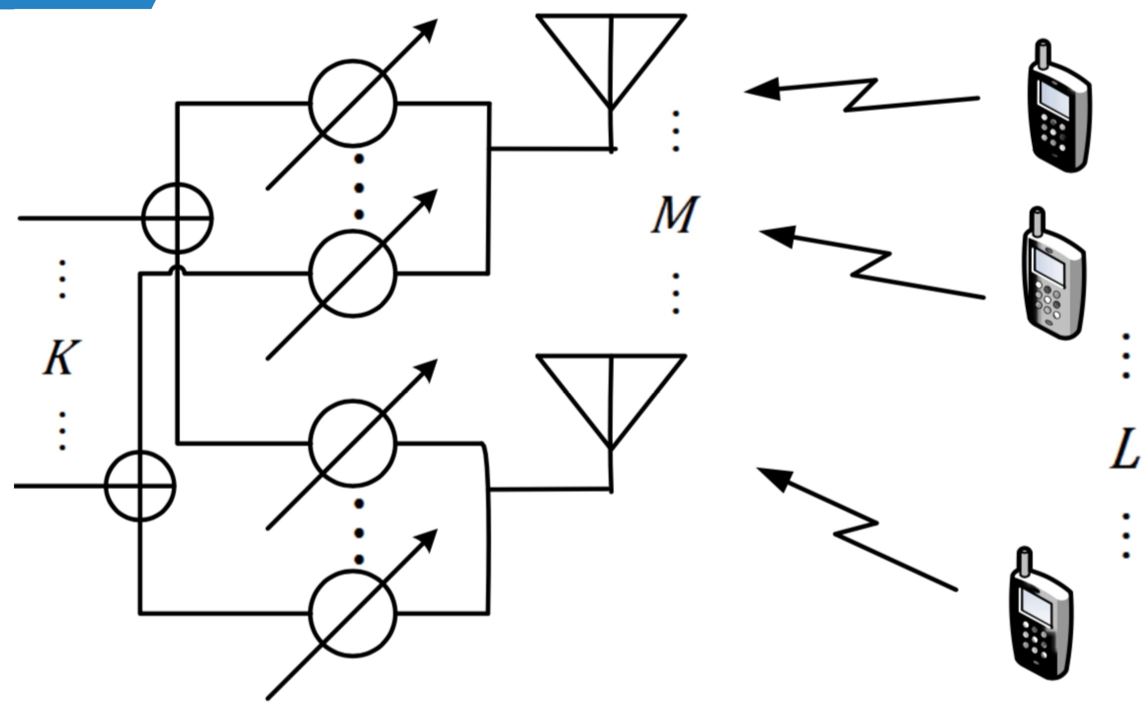
- Effective channel of the l–th user:

$$h_{n,eff}^{(l)} = [\mathbf{W} \odot \mathbf{\Phi}_n] a(\theta_l) \gamma_l \in \mathbb{C}^{K \times 1}$$

- Measure the Effective channel:

$$y_n^{(l)} = h_{n,eff}^{(l)} + z_l \in \mathbb{C}^{K \times 1}$$

2 SYSTEM MODEL: Matrix Form



Define

$$\mathbf{A}(\boldsymbol{\Theta}) = [a(\theta_1), a(\theta_2), \ldots a(\theta_l)] \in \mathbb{C}^{M\times L}$$

$$\boldsymbol{\Gamma} = diag(\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_l) \in \mathbb{C}^{L\times L}$$

Measurements of all L user when BS applying PSN training beamformer $\boldsymbol{\Phi}_{\mathbf{n}}$

$$\mathbf{Y}_{\mathbf{n}} = [\mathbf{W} \odot \boldsymbol{\Phi}_{\mathbf{n}}]\mathbf{A}(\boldsymbol{\Theta})\boldsymbol{\Gamma} + \mathbf{Z}_{\mathbf{n}}\mathbb{C}^{K\times L}$$

Stack

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \\ \boldsymbol{\Phi}_3 \\ \vdots \\ \boldsymbol{\Phi}_n \end{bmatrix} \mathbb{C}^{NK\times M} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \mathbf{Z}_3 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} \mathbb{C}^{NK\times L}$$

All Measurements is given by:

$$\mathbf{Y} = \left[\left(1_{\mathbf{N}} \otimes \mathbf{W}(\boldsymbol{\Omega})\right) \odot \boldsymbol{\Phi}\right]\mathbf{A}(\boldsymbol{\Theta})\boldsymbol{\Gamma} + \mathbf{Z} \in \mathbb{C}^{NK\times L}$$

2 SYSTEM MODEL: Estimation Program

Phase Deviation Estimation Program

$$\hat{\boldsymbol{\Omega}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

$$\text{s.t.} \quad |w_{km}| = 1 \text{ for } k = 1 \cdots K; m = 1, \cdots M$$

A **non-convex tri-variables** program with **unimodular constraints**

We can optimize the three matrices in **an alternating manner** to minimize the objective function

3 ALGORITHM

$$\widehat{\mathbf{\Omega}}, \widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Gamma}} = \arg \min_{\mathbf{\Omega},\mathbf{\Theta},\mathbf{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_{\boldsymbol{N}} \otimes \mathbf{W}(\mathbf{\Omega}) \right) \odot \mathbf{\Phi} \right] \mathbf{A}(\mathbf{\Theta}) \mathbf{\Gamma} \right\|_{\mathrm{F}}^{2}$$

Step1: Estimation of channel gain matrix $\mathbf{\Gamma}$

Given $\widehat{\mathbf{\Theta}}, \widehat{\mathbf{\Omega}}$, diagonal matrix $\mathbf{\Gamma}$ can be estimated as by the least-square root

$$\widehat{\gamma_l} = \frac{\boldsymbol{b}_l^H \boldsymbol{y}_l}{\boldsymbol{b}_l^H \boldsymbol{b}_l},$$

where

$$\boldsymbol{b}_l = [(1_{\boldsymbol{N}} \otimes \mathbf{W}) \odot \mathbf{\Phi}] \boldsymbol{a}(\boldsymbol{\theta}_l) \in \mathbb{C}^{NK \times 1}$$

Complexity: $\mathbf{O}(\boldsymbol{L})$

3 ALGORITHM

$$\hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Gamma}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_{\boldsymbol{N}} \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_{\mathrm{F}}^{2}$$

Step2: Estimation of user direction $\boldsymbol{\Theta}$

Given $\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Omega}}$, $\boldsymbol{\Theta}$ can be estimated separately by 1–D search since each direction is independent:

$$\theta_l = \arg\min \|\boldsymbol{y_n} - \mathbf{Q}a(\theta_l)\|_2^2$$

where

$$\mathbf{Q} = \boldsymbol{\gamma_l}[(1_{\boldsymbol{N}} \otimes \mathbf{W})] \odot \boldsymbol{\Phi} \in \mathbb{C}^{NK \times M}$$

We can adapt FFT to estimate θ_l, due to the special structure of $a(\theta_l)$

2 ways to obtain more accurate estimates:

- Use large–length FFTs (by padding zeros)
- Use moderate–length FFTs and some local search algorithms such as the back–tracking line search algorithm (more efficient)

Complexity: $\mathbf{O}(\boldsymbol{LM} \ \log(\boldsymbol{M}))$

3 ALGORITHM

$$\hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Gamma}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega})\right) \odot \boldsymbol{\Phi}\right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

- **Step3: Estimation of phase deviation $\mathbf{W}(\boldsymbol{\Omega})$**

Given $\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Theta}}$, each row of $\mathbf{W}$ can be estimate separately.
Take the k–th row of each $\mathbf{Y_n}$ and $\boldsymbol{\Phi}_\mathbf{n}$ and stack:

$$\tilde{\mathbf{Y}}_k \triangleq \begin{bmatrix} \mathbf{Y}_1(\boldsymbol{k},:) \\ \mathbf{Y}_2(\boldsymbol{k},:) \\ \vdots \\ \mathbf{Y}_N(\boldsymbol{k},:) \end{bmatrix} \mathbb{C}^{N\times M} \text{ and } \tilde{\boldsymbol{\Phi}}_k \triangleq \begin{bmatrix} \boldsymbol{\Phi}_1(\boldsymbol{k},:) \\ \boldsymbol{\Phi}_2(\boldsymbol{k},:) \\ \vdots \\ \boldsymbol{\Phi}_N(\boldsymbol{k},:) \end{bmatrix} \mathbb{C}^{N\times M}$$

Then, we have K quadratic sub–programs:

$$\min_{\mathrm{W(k,:)}} \left\| \tilde{\mathbf{Y}}_k - \left[\left(1_N \otimes \hat{\mathbf{W}}(k,:)\right) \odot \tilde{\boldsymbol{\Phi}}_k\right] \mathbf{A}(\hat{\boldsymbol{\Theta}}) \boldsymbol{\Gamma} \right\|_F^2 = \min_{\mathrm{W(k,:)}} \left\| \mathcal{Y}_k - \mathbf{R}_k \left[\hat{\mathbf{W}}(k,:)\right]^T \right\|_2^2$$

where $\mathcal{Y}_k = \mathrm{vec}(\tilde{\mathbf{Y}}_k) \in \mathbb{C}^{NL\times 1}, \mathbf{R}_k = \left(\mathbf{A}(\hat{\boldsymbol{\Theta}})\boldsymbol{\Gamma}\right)^T * \tilde{\boldsymbol{\Phi}}_k \in \mathbb{C}^{NL\times M}$

These programs are **hard** due to **the unimodular constraints: $|w_{km}| = 1$**

3 ALGORITHM

$$\min_{\mathbf{W}(k,:)} \left\| \tilde{\mathbf{Y}}_k - \left[\left(1_N \otimes \widehat{\mathbf{W}}(k,:) \right) \odot \tilde{\boldsymbol{\Phi}}_k \right] \mathbf{A}(\widehat{\boldsymbol{\Theta}}) \boldsymbol{\Gamma} \right\|_F^2 = \min_{\mathbf{W}(k,:)} \left\| \mathcal{Y}_k - \mathbf{R}_k \left[\widehat{\mathbf{W}}(k,:) \right]^T \right\|_2^2$$

◆ Rewrite into a UQP

Rewrite into a standard Unimodular Quadratic Program (UQP) :

$$\max_{\mathbf{v}^{(k)}} \mathbf{v}^{(k)^H} \mathbf{U}_k \, \mathbf{v}^{(k)}$$

$$\text{s.t.} \left| v_i^{(k)} \right| = 1$$

$$\text{for } i = 1,2,\cdots, M+1$$

where

$$\mathbf{U}_k = \begin{bmatrix} -\mathbf{R}_k^H \mathbf{R}_k & \mathbf{R}_k \mathcal{Y}_k \\ \mathcal{Y}_k^H \mathbf{R}_k & 0 \end{bmatrix} \text{ and}$$

$$\mathbf{v}^{(k)} = \begin{bmatrix} e^{j\xi} \left[\widehat{\mathbf{W}}(k,:) \right]^T \\ e^{j\xi} \end{bmatrix} \quad \xi \in [0, 2\pi)$$

◆ Solve the UQP: Power Method

Power method is an **efficient convergent** descent algorithm to find **local** optimum for UQP:

Iterations:

$$\mathbf{v}_{i+1}^{(k)} = e^{j \arg\left(\boldsymbol{U}_k \mathbf{v}_i^{(k)} \right)}$$

Ref: M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization,"IEEE Transactions on Signal Processing, vol. 62, No. 5,pp. 1221–1234, March 2014.

3 ALGORITHM: Summary

Algorithm 1 Calibration of Phase Shifter Network

Input: number of antennas M; number of RF chains K; number of precoders N; number of users L; measured effective channel vectors $\mathbf{Y}$;

Output: Phase deviation matrix $\hat{\mathbf{W}} \in \mathbb{C}^{K \times M}$;

```
Random initialization of Ŵ and Θ̂
do
    for l = 1 : L do                                   Step1
        Estimate γ̂_l in (12);
    end for
    for l = 1 : L do                                   Step2
        Estimate θ̂_l in (14) by FFT and backtracking line search;
    end for
    for k = 1 : K do                                   Step3
        Estimate Ŵ(k, :) in (20) via the power method;
    end for
    Ŷ = [(1_N ⊗ Ŵ) ⊙ Φ] A(Θ̂)Γ̂
while change in ||Ŷ||_F from the previous iteration is less than ε
return Ŵ;
Perform calibration according to Ŵ
```

$$\hat{\mathbf{Y}} = \left[(\mathbf{1_N} \otimes \hat{\mathbf{W}}) \odot \mathbf{\Phi}\right] \mathbf{A}(\hat{\mathbf{\Theta}})\hat{\mathbf{\Gamma}}$$

Step4: Repeat 1–3 until convergence

◆ Low complexity

- Estimation of $\mathbf{\Gamma}$: closed form
- Estimation of $\mathbf{\Theta}$: FFTs
- Estimation of $\mathbf{W}$: Efficient Power Method
- Fast convergence: about 20 iterations (Numerical results)

◆ Convergence

- Each step is monotonically decreasing
- The problem is bounded

4 PERFORMANCE ANALYSIS

Cramer–Rao Bound (CRLB)

Define:

$$\boldsymbol{\eta} \triangleq [\boldsymbol{\Omega}^T, \boldsymbol{\Theta}^T, \mathrm{Re}\{\boldsymbol{\gamma}^T\}, \mathrm{Im}\{\boldsymbol{\gamma}^T\}]^T \in \mathbb{R}^{MK+3L-2}$$

where

$$\boldsymbol{\Omega} = [\omega_{21}, \cdots, \omega_{K1}, \omega_{12}, \cdots, \omega_{K2}, \cdots, \omega_{1M}, \cdots, \omega_{KM}] \in \mathbb{R}^{MK-1}$$

$$\boldsymbol{\Theta} = [\theta_2, \theta_3, \cdots, \theta_L] \in \mathbb{R}^{L-1}$$

$$\boldsymbol{\gamma} = [\gamma_2, \gamma_3, \cdots, \gamma_L] \in \mathbb{R}^{L}$$

Fisher information matrix (FIM):

$$\mathbf{F} = \frac{2}{\sigma_z^2} \sum_{n=1}^{N} \mathrm{Re}\left[\frac{\partial \boldsymbol{\mu}_n^H(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]$$

where

$$\boldsymbol{\mu}_n = \left[\mathbf{I}_K \otimes \left(\boldsymbol{\Gamma}^T \mathbf{A}^T(\boldsymbol{\Theta})\right)\right] \mathcal{D}\left(\mathbf{w}_{\mathrm{vec}} \boldsymbol{\Phi}_{\mathrm{vec}}^{(n)}\right) \in \mathbb{C}^{KL \times 1}$$

Compute different parts of $\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$:

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathrm{Re}\{\boldsymbol{\gamma}\}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathrm{Im}\{\boldsymbol{\gamma}\}}\right] \in \mathbb{C}^{KL \times (MK+3L-2)}$$

4 PERFORMANCE ANALYSIS

Fisher information

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}} = j[\mathbf{I}_K \otimes (\boldsymbol{\Gamma}^T \mathbf{A}^T(\boldsymbol{\Theta}))]\mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right) \times [\mathcal{D}(\mathbf{w}_{\text{vec}})(:,2:KM)] \in \mathbb{C}^{KL\times(MK-1)}$$

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}} = j\pi\left\{\begin{array}{c}\mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ \gamma_2 \boldsymbol{a}^T(\theta_2)\mathbf{D}_{\theta_2} \\ 0_M^T \\ \vdots \\ 0_M^T \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}}, \mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ 0_M^T \\ \gamma_3 \boldsymbol{a}^T(\theta_3)\mathbf{D}_{\theta_3} \\ \vdots \\ 0_M^T \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}}, \\ \cdots, \mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ 0_M^T \\ \vdots \\ 0_M^T \\ \gamma_L \boldsymbol{a}^T(\theta_L)\mathbf{D}_{\theta_L} \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}} \end{array}\right\} \in \mathbb{C}^{KL\times(L-1)}$$

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}} = j\pi\left\{\begin{array}{c}\mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ \gamma_2 \boldsymbol{a}^T(\theta_2)\mathbf{D}_{\theta_2} \\ 0_M^T \\ \vdots \\ 0_M^T \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}}, \mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ 0_M^T \\ \gamma_3 \boldsymbol{a}^T(\theta_3)\mathbf{D}_{\theta_3} \\ \vdots \\ 0_M^T \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}}, \\ \cdots, \mathbf{I}_K \otimes \begin{bmatrix} 0_M^T \\ 0_M^T \\ \vdots \\ 0_M^T \\ \gamma_L \boldsymbol{a}^T(\theta_L)\mathbf{D}_{\theta_L} \end{bmatrix} \mathcal{D}\left(\boldsymbol{\Phi}_{\text{vec}}^{(n)}\right)\mathbf{w}_{\text{vec}} \end{array}\right\} \in \mathbb{C}^{KL\times(L-1)}$$

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \text{Im}\{\boldsymbol{\gamma}\}} = j\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \text{Re}\{\boldsymbol{\gamma}\}}$$

4 PERFORMANCE ANALYSIS

Fisher information

The Fisher information of $\{\omega_{km}\}_{k=1,m=1}^{K,M}$ based on the observation Y is the first KM−1 diagonal elements of the FIM, which is

$$\mathcal{J}(\omega_{km}) = \frac{2N}{\sigma_z^2}|\gamma_l|^2$$

A more **accurate** estimates may be obtained via a larger number of measurements (PSN beamformer)

Cramer–Rao Lower Bound

The CRLB matrix can be obtained by the inverse of the FIM:

$$\mathrm{C}_\eta = \mathbf{F}^{-1}$$

The first KM–1 diagonal elements of C_η can be then used as the lower bound to the variances of our estimations $\mathbf{\Omega}$.

4 PERFORMANCE ANALYSIS

Minimum of required number of measurements

Counting the degrees of freedom (DOF) of the variables :

$$\mathrm{DOF}(\mathbf{\Omega}) = MK - 1,\ \mathrm{DOF}(\mathbf{\Theta}) = L - 1,\ \mathrm{DOF}(\mathbf{\Gamma}) = 2L,$$

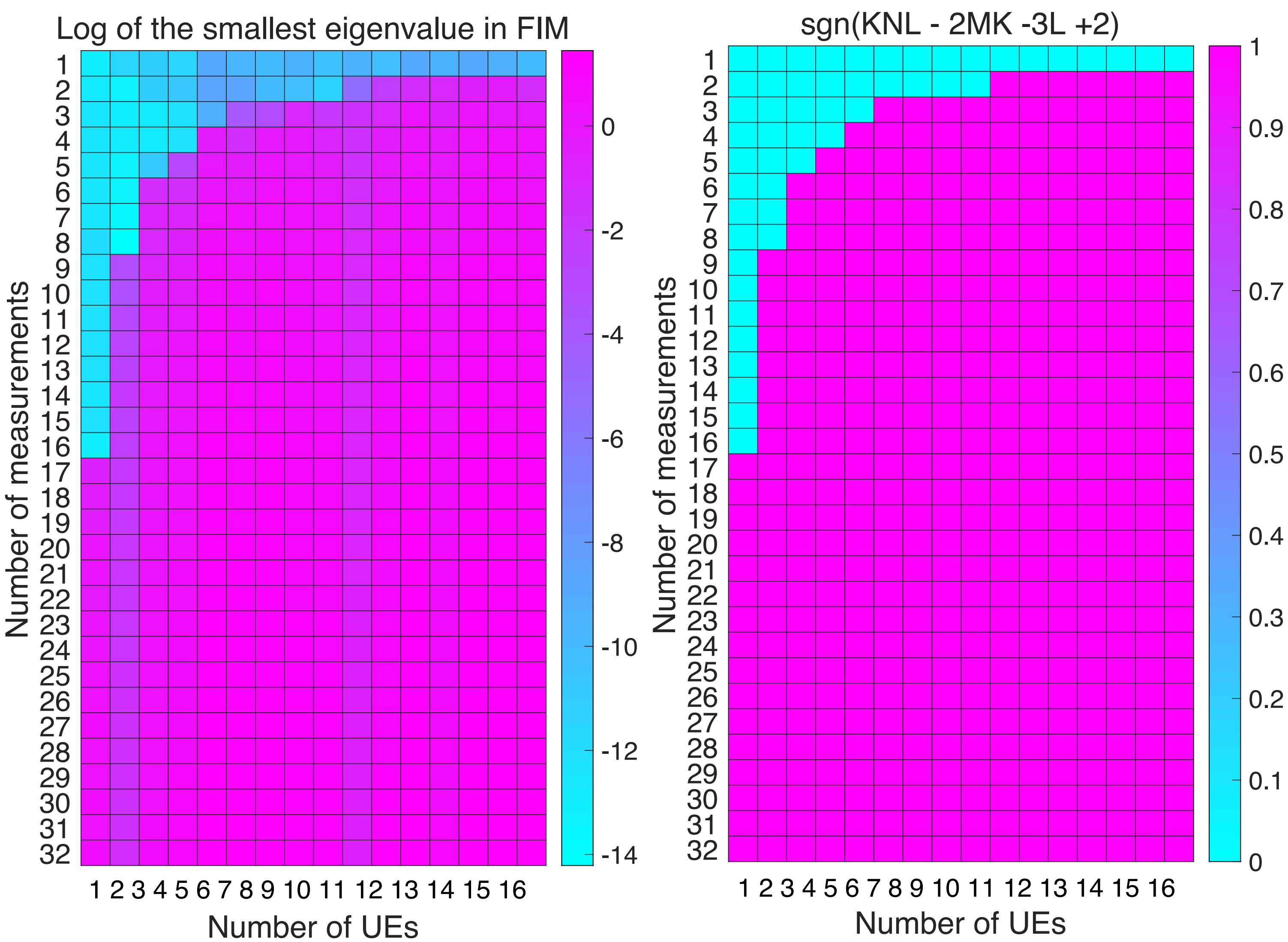
The number of observations in the matrix Y is 2KNL. To ensure that the phase deviation is estimable we need to have

$$2KNL \geq MK + 3L - 2.$$

Hence, the lower bound of N is

$$N \geq \left\lceil \frac{M}{2L} + \frac{3}{2K} - \frac{2}{KL} \right\rceil.$$

4 PERFORMANCE ANALYSIS

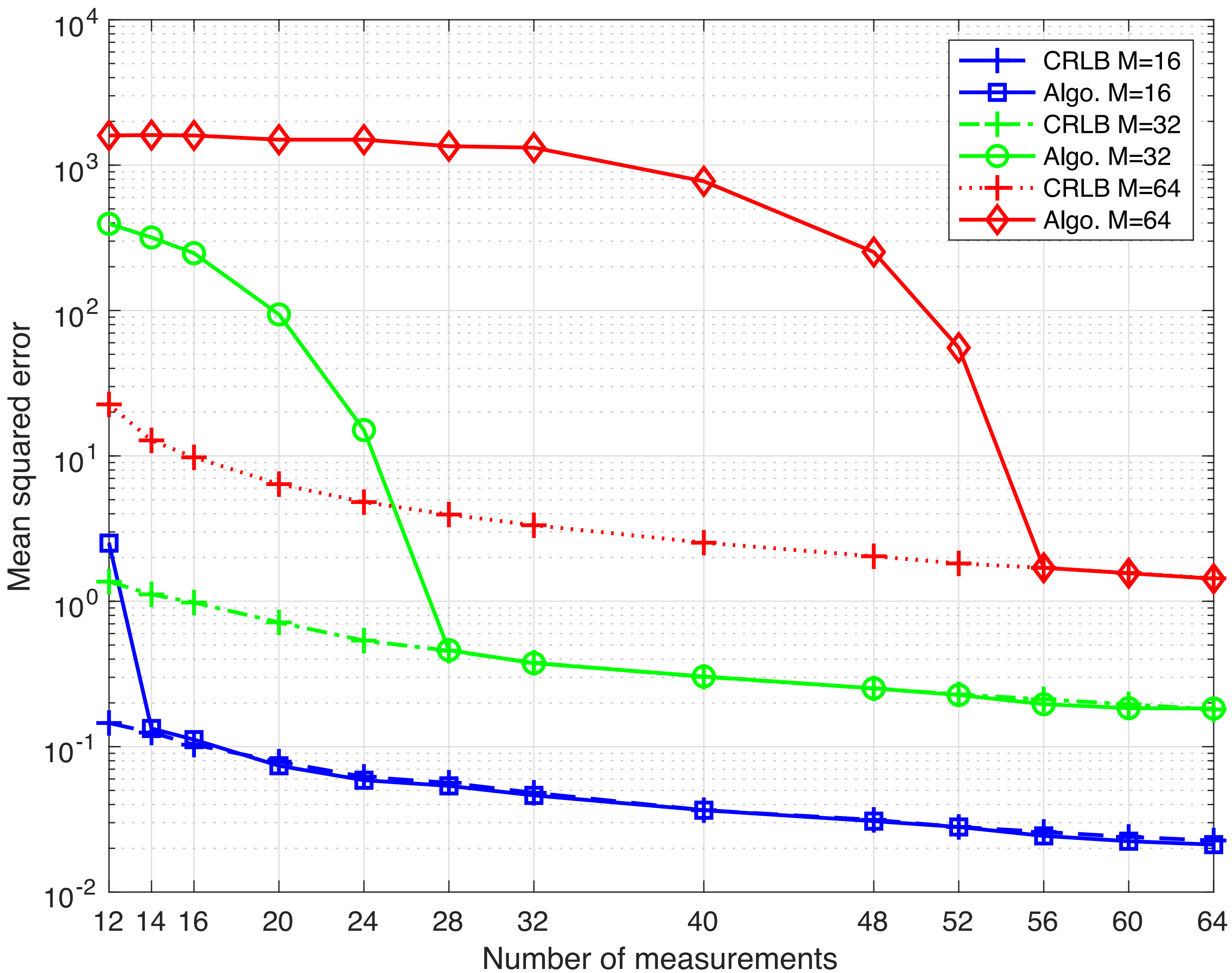


Blue pixels:
$2KNL - MK - 3L + 2 < 0$

Red pixels :
$2KNL - MK - 3L + 2 \geq 0$

Number of antennas (M)	Number of RF chains (K)	Number of PSN beamformer (N)	Number of UEs (L)	Channel gain	UE distribution
32	4	1-32	1-16	Random Phase	Uniform between (0,pi)

5 NUMERICAL RESULTS

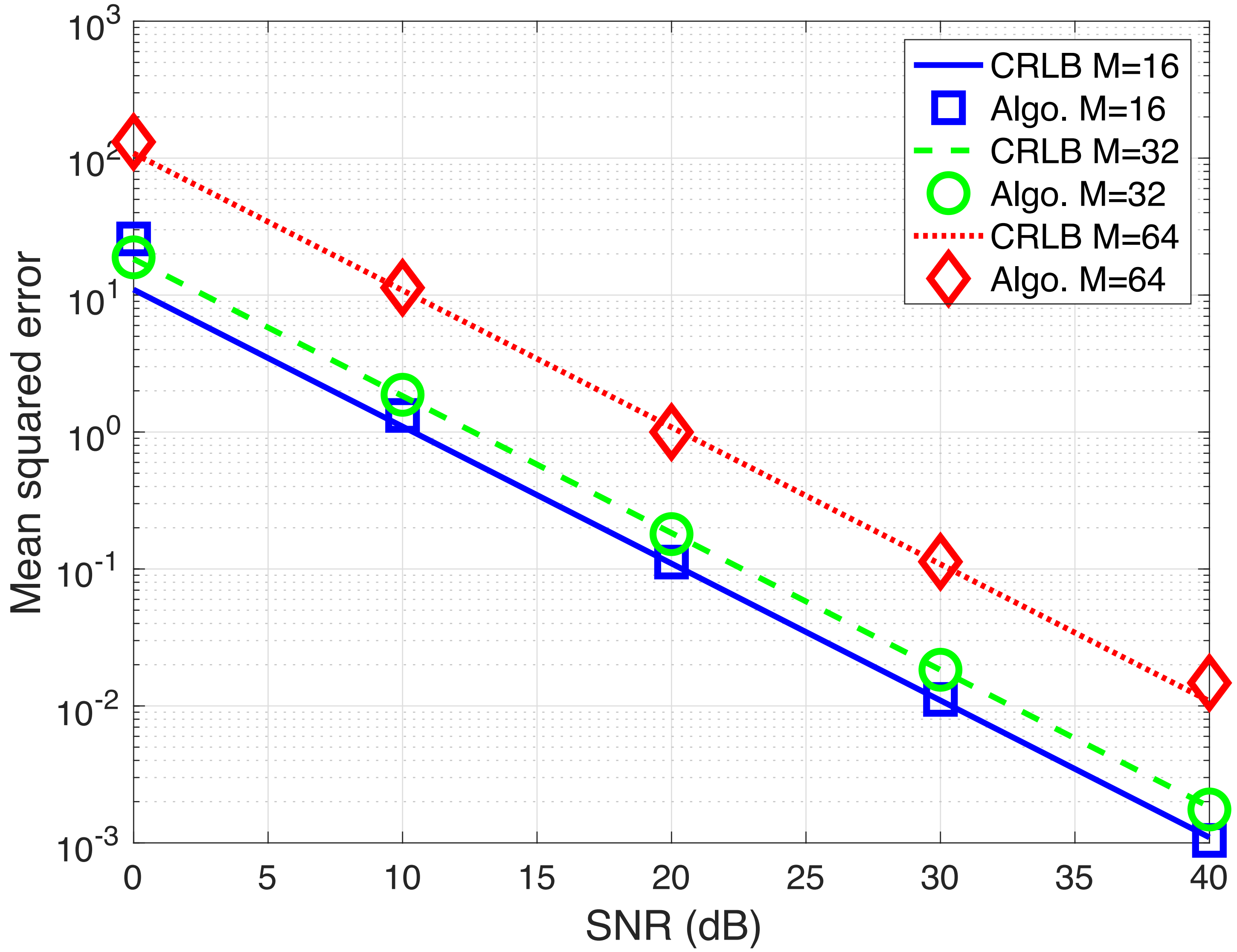


Threshold effect

- The lower bound of number of PSN beamformer guarantee the problem to be estimable, but do not ensure it reach the CRLB. **(local optimum)**
- When the number of beamformer large enough, (e.g. slightly less than M), the estimates will close to CRLB. **(global optimum)**

Number of antennas (M)	Number of RF chains (K)	Number of PSN beamformer (N)	Number of UEs (L)	Channel gain	UE distribution	Measurement SNR (dB)
16,32,64	M/8	12-64	4	Random Phase	Uniform between (0,pi)	20

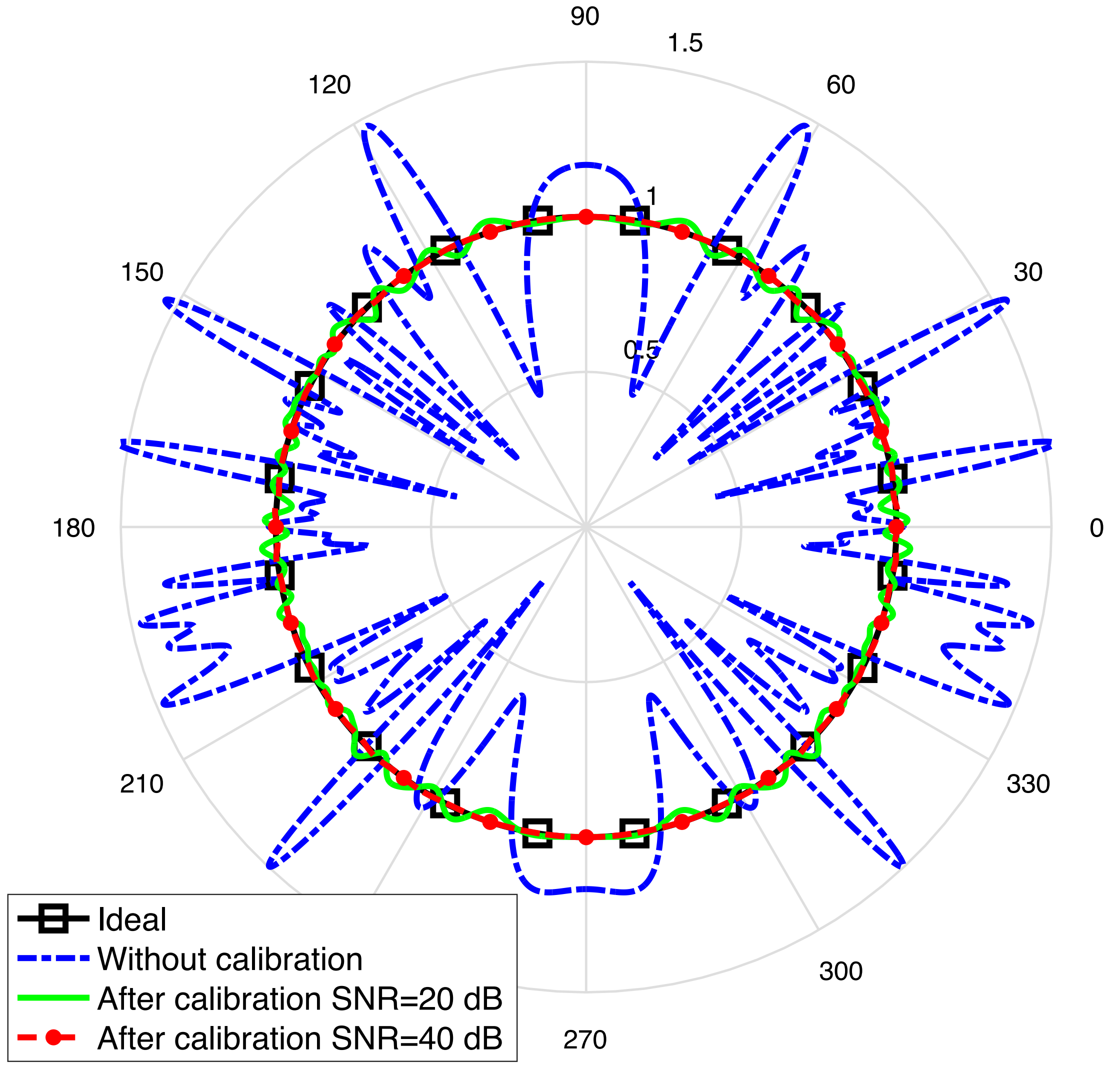
5 NUMERICAL RESULTS



- The phase deviation estimates are close to the CRLB even when measurement **SNR is very low**.

Number of antennas (M)	Number of RF chains (K)	Number of PSN beamformer (N)	Number of UEs (L)	Channel gain	UE distribution	Measurement SNR (dB)
16,32,64	M/8	M	4	Random Phase	Uniform between (0,pi)	0,10,20,30,40

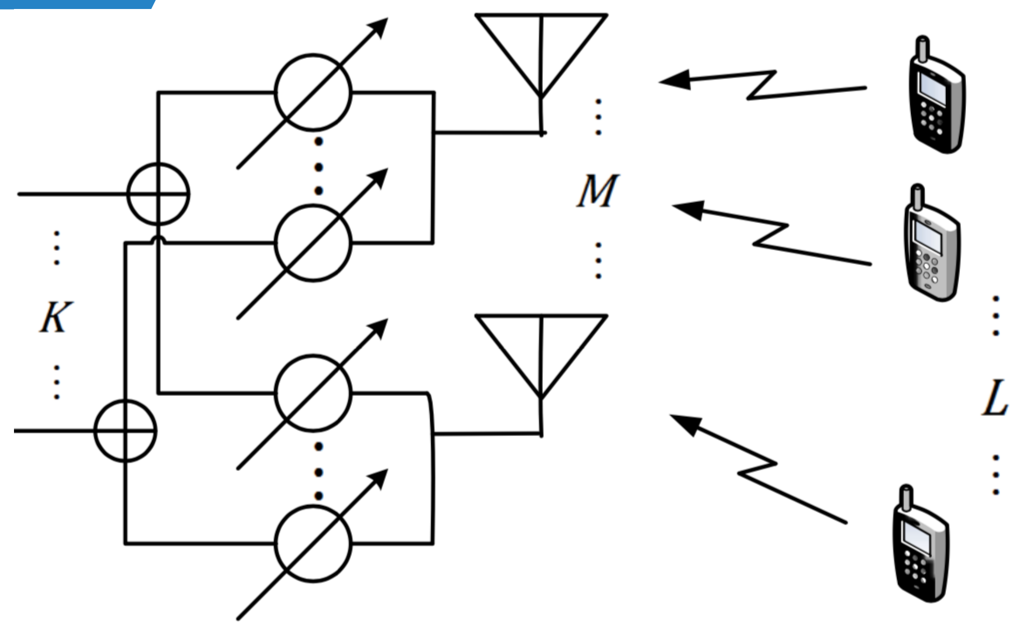
5 NUMERICAL RESULTS



- The beam pattern is **significantly improved**.

Number of antennas (M)	Number of RF chains (K)	Number of PSN beamformer (N)	Number of UEs (L)	Channel gain	UE distribution	Measurement SNR (dB)
32	2	32	4	Random Phase	Uniform between (0,pi)	20,40

6 Extend to non-LOS Channel Model



- K : Number of RF chains
- M: Number of antenna elements
- L: Number of UEs involved for calibration
- N : Number of PSN (analog) beamformer

- Steering vector of a M–element ULA

$$a(\theta) = \left[1, e^{-\frac{j2\pi d\sin(\theta)}{\lambda}}, e^{-\frac{j4\pi d\sin(\theta)}{\lambda}}, \dots, e^{-\frac{j2(M-1)d\sin(\theta)}{\lambda}}\right]$$

- Phase deviation matrix $\mathbf{W}(\boldsymbol{\Omega}) \in \mathbb{C}^{K\times M}$

$$w_{km} = e^{j\omega_{km}}$$

- PSN beamformer $\boldsymbol{\Phi}_\mathbf{n} \in \mathbb{C}^{K\times M}$

$$\left|\boldsymbol{\phi}_{\boldsymbol{n}}^{(km)}\right| = 1 \quad \angle\boldsymbol{\phi}_{\boldsymbol{n}}^{(km)} \in [0,2\pi]$$

- Non–LOS channel of the l–th user

$$h_l = \sum_{i}^{d_l} \gamma_l^{(i)} a(\theta_l^{(i)})$$

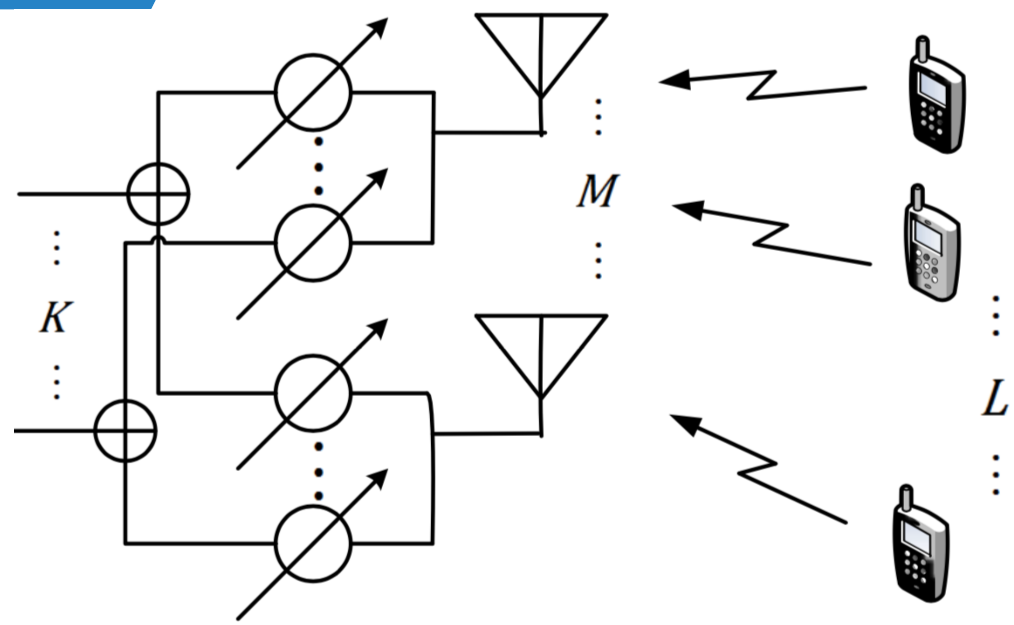
- Effective channel of the l–th user:

$$h_{n,eff}^{(l)} = [\mathbf{W} \odot \boldsymbol{\Phi}_n] \sum_{i}^{d_l} \gamma_l^{(i)} a(\theta_l^{(i)})$$

- Measure the Effective channel:

$$y_n^{(l)} = h_{n,eff}^{(l)} + z_l$$

6 Extend to non-LOS Channel Model



Define

$$\mathbf{A}(\mathbf{\Theta}) = \Big[\boldsymbol{a}(\theta_1^{(1)}), \cdots, \boldsymbol{a}(\theta_1^{(d_1)}), \boldsymbol{a}(\theta_2^{(1)}), \cdots, \boldsymbol{a}(\theta_2^{(d_2)}), \cdots, \boldsymbol{a}(\theta_L^{(1)}), \cdots, \boldsymbol{a}(\theta_L^{(d_L)})\Big] \in \mathbb{C}^{M\times(d_1+d_2+\cdots+d_L)},$$

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma_L \end{bmatrix} \in \mathbb{C}^{(d_1+d_2+\cdots+d_L)\times L}$$

$$\gamma_l^T \triangleq [\gamma_l^{(1)}, \gamma_l^{(2)}, \cdots, \gamma_l^{(d_l)}] \ \ \forall l = 1, 2, \cdots, L.$$

Measurements of all L users when BS applying PSN training beamformer $\mathbf{\Phi_n}$

$$\mathbf{Y_n} = [\mathbf{W} \otimes \mathbf{\Phi_n}]\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma} + \mathbf{Z_n} \mathbb{C}^{K\times L}$$

Stack

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_1 \\ \mathbf{\Phi}_2 \\ \mathbf{\Phi}_3 \\ \vdots \\ \mathbf{\Phi}_n \end{bmatrix} \mathbb{C}^{NK\times M} \ \text{ and } \ \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \mathbf{Z}_3 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix} \mathbb{C}^{NK\times L}$$

All Measurements of L users when applying all N PSN training beamformers are given by:

$$\mathbf{Y} = \big[\big(1_\mathbf{N} \otimes \mathbf{W}(\mathbf{\Omega})\big) \odot \mathbf{\Phi}\big]\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma} + \mathbf{Z} \in \mathbb{C}^{NK\times L}$$

6 Extend to non-LOS Channel Model

Phase deviation Estimation Program

$$\hat{\boldsymbol{\Omega}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

$$\text{s.t.} \quad |w_{km}| = 1 \text{ for } k = 1 \cdots K; m = 1, \cdots M$$

A **non–convex tri–variable** program with **unimodular constraints**

We can optimize the three matrices in **an alternating manner** to minimize the objective function

6 Extend to non-LOS Channel Model

$$\hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\Gamma}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

- **Joint Estimation of Multipath Channel Parameters $\boldsymbol{\Theta}, \boldsymbol{\Gamma}$**

Given $\widehat{\mathbf{W}}$, channel parameters for the L users can be estimated separately.

Recall:

$$\boldsymbol{\gamma}_l = \left[\gamma_l^{(1)}, \gamma_l^{(2)}, \cdots, \gamma_l^{(d_l)} \right]^T, \boldsymbol{\theta}_l = \left[\theta_l^{(1)}, \theta_l^{(2)}, \cdots, \theta_l^{(d_l)} \right]^T \quad h_l = \sum_i^{d_l} \gamma_l^{(i)} a(\theta_l^{(i)})$$

$$\hat{\boldsymbol{\gamma}}_l, \hat{\boldsymbol{\Theta}}_l = \arg \min_{\boldsymbol{\gamma}_l, \boldsymbol{\Theta}_l} \left\| \mathbf{y}_l - \left[(\mathbf{1}_N \otimes \mathbf{W}) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}_l) \boldsymbol{\gamma}_l \right\|_2^2$$

Define

$$\mathbf{B} \triangleq \left[(\mathbf{1}_N \otimes \hat{\mathbf{W}}) \odot \boldsymbol{\Phi} \right] \in \mathbb{C}^{KN \times M},$$

Concentrate out $\widehat{\gamma}_l$

$$\hat{\boldsymbol{\gamma}}_l = \left(\mathbf{A}^H(\hat{\boldsymbol{\Theta}}_l) \mathbf{B}^H \mathbf{B} \mathbf{A}(\hat{\boldsymbol{\Theta}}_l) \right)^{-1} \mathbf{A}(\hat{\boldsymbol{\Theta}}_l)^H \mathbf{B}^H \mathbf{y}_l.$$

6 Extend to non-LOS Channel Model

$$\widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Gamma}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

- **Joint Estimation of Multipath Channel Parameters $\boldsymbol{\Theta}, \boldsymbol{\Gamma}$**

Estimation of $\widehat{\Theta}_l$

$$\hat{\boldsymbol{\Theta}}_l = \arg \max_{\boldsymbol{\Theta}_l} \mathbf{y}_l^H \mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l) \left(\mathbf{A}^H(\boldsymbol{\Theta}_l) \mathbf{B}^H \mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l) \right)^{-1} \mathbf{A}^H(\boldsymbol{\Theta}_l) \mathbf{B}^H \mathbf{y}_l$$

$$= \arg \max_{\boldsymbol{\Theta}_l} \quad \mathbf{y}_l^H \boldsymbol{\Lambda}_l \mathbf{y}_l .$$

where $\boldsymbol{\Lambda}_l$ is the projection matrix of $\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l)$.

If $d_l = 1$, Θ_l becomes a scalar $\theta_l^{(1)}$, we can use an 1–D search to estimate it

$$\hat{\theta}_l^{(1)} = \arg \max_{\theta_l^{(1)}} \frac{\left| \mathbf{y}_l^H \mathbf{B} \boldsymbol{a}(\theta_l^{(1)}) \right|^2}{\boldsymbol{a}^H(\theta_l^{(1)}) \mathbf{B}^H \mathbf{B} \boldsymbol{a}(\theta_l^{(1)})}$$

FFT to compute

6 Extend to non-LOS Channel Model

$$\widehat{\boldsymbol{\Omega}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Gamma}} = \arg \min_{\boldsymbol{\Omega},\boldsymbol{\Theta},\boldsymbol{\Gamma}} \left\| \boldsymbol{Y} - \left[\left(1_N \otimes \mathbf{W}(\boldsymbol{\Omega}) \right) \odot \boldsymbol{\Phi} \right] \mathbf{A}(\boldsymbol{\Theta}) \boldsymbol{\Gamma} \right\|_F^2$$

◆ Joint Estimation of Multipath Channel Parameters $\boldsymbol{\Theta}, \boldsymbol{\Gamma}$

Lemma 1. *Given a j-element subset of $\boldsymbol{\Theta}_l$:*

$$\boldsymbol{\Theta}_l^{(j)} \triangleq \{\theta_l^{(1)}, \theta_l^{(2)} \cdots \theta_l^{(j)}\} \quad 1 \leq j < d_l, \tag{27}$$

$\theta_l^{(j+1)}$ *can be obtained by a 1-D search:*

$$\hat{\theta}_l^{(j+1)} = arg \max_{\theta_l^{(j+1)}} \frac{\left| \mathbf{y}_l^H \boldsymbol{\Lambda}_{l,j}^{\perp} \mathbf{B} \boldsymbol{a}(\theta_l^{(j+1)}) \right|^2}{\boldsymbol{a}^H(\theta_l^{j+1}) \mathbf{B}^H \boldsymbol{\Lambda}_{l,j}^{\perp} \mathbf{B} \boldsymbol{a}(\theta_l^{j+1})}, \tag{28}$$

where $\boldsymbol{\Lambda}_{l,j}^{\perp} \triangleq \mathcal{P}^{\perp}\{\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l^{(j)})\}$ *is the orthogonal projection matrix of* $\mathbf{B}\mathbf{A}(\boldsymbol{\Theta}_l^{(j)})$.

FFT to compute

Proof. The proof can be found in our extended journal version.

Algorithm 1 Repeated orthogonal projection algorithm

Input: B in (22); number of multipath d_l
Output: Angle vector $\hat{\boldsymbol{\Theta}}_l$ and gain vector $\hat{\gamma}_l$ of multipath channel;

Estimate $\hat{\theta}_l^{(1)}$ in (25);
for $j = 2 : d_l$ **do**
  Set $\boldsymbol{\Theta}_l^{(j-1)} \leftarrow \{\hat{\theta}_l^{(1)} \cdots \hat{\theta}_l^{(j-1)}\}$ and use (28) to estimate $\theta_l^{(j)}$
  Use (28) to update all $\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}, \cdots \hat{\theta}_l^{(j)}$ iteratively until "practical convergence";
end for
Estimate $\hat{\gamma}_l$ with (23);
return $\hat{\boldsymbol{\Theta}}_l$ and $\hat{\gamma}_l$

Assume $d_l = 1$
Obtain $\hat{\theta}_l^{(1)}$

Set
$\boldsymbol{\Theta}_l^{(1)} = \{\hat{\theta}_l^{(1)}\}$

Estimate $\hat{\theta}_l^{(2)}$ via (28)

Assume $d_l = 2$, Set
$\boldsymbol{\Theta}_l^{(2)} = \{\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}\}$

Estimate $\hat{\theta}_l^{(3)}$ via (28)

Assume $d_l = 3$, Set
$\boldsymbol{\Theta}_l^{(3)} = \{\hat{\theta}_l^{(1)}, \hat{\theta}_l^{(2)}, \hat{\theta}_l^{(3)}\}$
…

6 Extend to non-LOS Channel Model

Cramer–Rao Bound (CRLB)

Define:

$$\boldsymbol{\eta} \triangleq [\boldsymbol{\Omega}^T, \boldsymbol{\Theta}^T, \mathrm{Re}\{\boldsymbol{\gamma}^T\}, \mathrm{Im}\{\boldsymbol{\gamma}^T\}]^T \in \mathbb{R}^{MK+3L-2}$$

where

$$\boldsymbol{\Omega} = [\omega_{21}, \cdots, \omega_{K1}, \omega_{12}, \cdots, \omega_{K2}, \cdots, \omega_{1M}, \cdots, \omega_{KM}] \in \mathbb{R}^{MK-1}$$

$$\boldsymbol{\Theta} = \left[\theta_1^{(2)} \cdots \theta_1^{(d_1)}, \theta_2^{(1)}, \cdots \theta_2^{(d_2)}, \theta_3^{(1)}, \cdots, \theta_L^{(d_L)}\right] \in \mathbb{R}^{d_1+d_2+\cdots+\mathrm{d}_L-1}$$

$$\boldsymbol{\gamma} = \left[\gamma_1^{(1)}, \cdots, \gamma_1^{(d_1)}, \gamma_2^{(1)}, \cdots, \gamma_2^{(d_2)}, \gamma_3^{(1)} \cdots, \gamma_L^{(d_L)}\right] \in \mathbb{R}^{d_1+d_2+\cdots d_L}$$

Fisher information matrix (FIM):

$$\mathbf{F} = \frac{2}{\sigma_z^2} \sum_{n=1}^{N} \mathrm{Re}\left[\frac{\partial \boldsymbol{\mu}_n^H(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right]$$

where

$$\boldsymbol{\mu}_n = [\mathbf{I}_K \otimes (\boldsymbol{\Gamma}^T \mathbf{A}^T(\boldsymbol{\Theta}))] \mathcal{D}\left(\mathbf{w}_{\mathrm{vec}} \boldsymbol{\Phi}_{\mathrm{vec}}^{(n)}\right) \in \mathbb{C}^{KL\times 1}$$

Compute different parts of $\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$:

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = \left[\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathrm{Re}\{\boldsymbol{\gamma}\}}, \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \mathrm{Im}\{\boldsymbol{\gamma}\}}\right] \in \mathbb{C}^{KL\times(MK+3(d_1+\cdots+d_L)-2)}$$

6 Extend to non-LOS Channel Model

Fisher information

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Omega}} = j\left[\mathbf{I}_K \otimes \left(\boldsymbol{\Gamma}^T \mathbf{A}^T(\boldsymbol{\Theta})\right)\right] \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n) \times [\mathcal{D}(\mathbf{w}_{\text{vec}})(:, 2:KM)] \in \mathbb{C}^{KL\times(MK-1)},$$

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \boldsymbol{\Theta}} = j\pi \times \left[\mathbf{q}_1^{(2)}, \cdots, \mathbf{q}_1^{(d_1)}, \mathbf{q}_2^{(1)}, \cdots, \mathbf{q}_2^{(d_2)}, \cdots, \mathbf{q}_L^{(1)}, \cdots, \mathbf{q}_L^{(d_L)}\right] \in \mathbb{C}^{KL\times(d_1+\cdots+d_L-1)},$$

and each column of which is

$$\mathbf{q}_i^{(j)} = \mathbf{I}_K \otimes \begin{bmatrix} \left.\begin{matrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{matrix}\right\} i-1 \\ \gamma_i^{(j)} \boldsymbol{a}^T(\theta_i^{(j)}) \mathbf{D}_{\theta_i^{(j)}} \\ \left.\begin{matrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{matrix}\right\} L-i \end{bmatrix} \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n)\mathbf{w}_{\text{vec}}$$

for $i = 1, 2, \cdots L$ and $j = 1, 2, \cdots, d_i$.

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Re}\{\boldsymbol{\gamma}\}} = \left[\mathbf{r}_1^{(1)}, \cdots, \mathbf{r}_1^{(d_1)}, \mathbf{r}_2^{(1)}, \cdots, \mathbf{r}_2^{(d_2)}, \cdots, \mathbf{r}_L^{(1)}, \cdots, \mathbf{r}_L^{(d_L)}\right] \in \mathbb{C}^{KL\times(d_1+\cdots+d_L)},$$

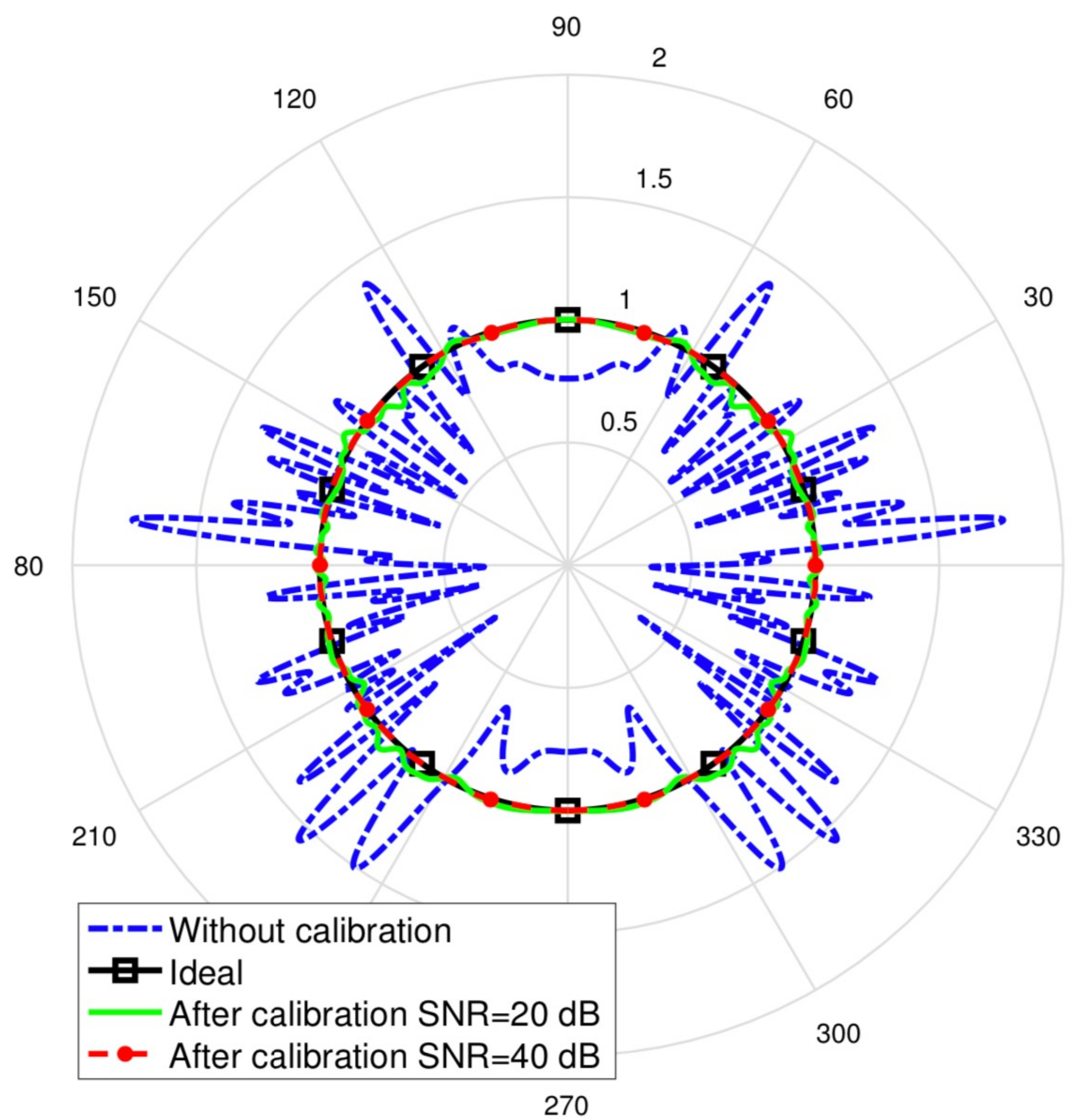
where

$$\mathbf{r}_i^{(j)} = \mathbf{I}_K \otimes \begin{bmatrix} \left.\begin{matrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{matrix}\right\} i-1 \\ \boldsymbol{a}^T(\theta_i^{(j)}) \\ \left.\begin{matrix} \mathbf{0}_M^T \\ \vdots \\ \mathbf{0}_M^T \end{matrix}\right\} L-i \end{bmatrix} \mathcal{D}(\boldsymbol{\Phi}_{\text{vec}}^n)\mathbf{w}_{\text{vec}}$$

for $i = 1, 2, \cdots L$ and $j = 1, 2, \cdots, d_i$.

$$\frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\eta})}{\partial \operatorname{Im}\{\boldsymbol{\gamma}\}} = j\frac{\partial \boldsymbol{\mu}(\boldsymbol{\eta})}{\partial \operatorname{Re}\{\boldsymbol{\gamma}\}}.$$

6 Extend to non-LOS Channel Model



- The beam pattern is **significantly improved**.

Number of antennas (M)	Number of RF chains (K)	Number of PSN beamformer (N)	Number of UEs (L)	Channel model	UE distribution	Measurement SNR (dB)
32	2	32	4	geometric-based channel model	Uniform between (0,pi)	20,40

7 Reference

- **Xizixiang Wei**, Yi Jiang, Qingwen Liu and Xin Wang, *Calibration of Phase Shifter Network for Hybrid Beamforming in mmWave Massive MIMO Systems*, in IEEE Transactions on Signal Processing, vol. 68, pp. 2302–2315, 2020.

- **Xizixiang Wei**, Yi Jiang, Xin Wang and Cong Shen, *Tx–Rx Calibration of Massive MIMO Systems with Analog Phase Shifter Network*, IEEE Wireless Communications Letters, vol. 11, no. 2, pp. 431–435, Feb. 2022.

- **Xizixiang Wei**, Yi Jiang, and Xin Wang, *Calibration of Phase Shifter Network for Hybrid Beamforming in mmWave Massive MIMO Systems*, in Proc. IEEE International Conference on Communications (ICC), May 2019.

- **Xizixiang Wei**, Yi Jiang, and Xin Wang, *Online Calibration of Phase Shifter Network for mmWave Massive MIMO Systems in Multipath Channels*, in Proc. 2019 International Conference on Wireless Communications and Signal Processing (WCSP), Oct. 2019.

8 Other Background

Different Topics of Wireless Federated Learning

–Resource allocation, low–complexity design, convergence guarantee, differential privacy guarantee…

Publication:

– **Xizixiang Wei**, et al., Random Orthogonalization for Federated Learning in Massive MIMO Systems, IEEE Transactions on Communications, submitted.

– **Xizixiang Wei**, et al., FLORAS: Differentially Private Wireless Federated Learning Using Orthogonal Sequences, in Proc. IEEE ICC, submitted.

– **Xizixiang Wei**, et al., Random Orthogonalization for Federated Learning in Massive MIMO Systems, in Proc. IEEE ICC, May 2022.

– **Xizixiang Wei** and Cong Shen, Federated Learning over Noisy Channels: Convergence Analysis and Design Examples, IEEE Transactions on Cognitive Communications and Networking, vol. 8, no. 2, pp. 1253–1268, June 2022.

– **Xizixiang Wei** and Cong Shen, Federated Learning over Noisy Channels, in Proc. IEEE ICC, June 2021.

UNIVERSITY of VIRGINIA | SCHOOL of ENGINEERING & APPLIED SCIENCE
Charles L. Brown Department of Electrical and Computer Engineering

Thank you!

Q&A

9 APPENDIX

Ambiguity

$$\mathbf{Y} = \left[\left(1_\mathbf{N} \otimes \mathbf{W}(\mathbf{\Omega})\right) \odot \mathbf{\Phi}\right]\mathbf{A}(\mathbf{\Theta})\mathbf{\Gamma} + \mathbf{Z} \in \mathbb{C}^{NK\times L}$$

$\mathbf{\Theta}, \mathbf{\Gamma}, \mathbf{W}$ can not be be uniquely determined

- If $\widehat{\mathbf{\Gamma}}, \widehat{\mathbf{W}}$ are solutions, so are $\mathbf{e}^{\mathbf{j}\beta}\widehat{\mathbf{\Gamma}}, \mathbf{e}^{-\mathbf{j}\beta}\widehat{\mathbf{W}}$, β is a random phase
- If $\mathbf{A}(\widehat{\mathbf{\Theta}}), \widehat{\mathbf{W}}$ are solutions, so are, $\widehat{\mathbf{W}}\mathbf{T}$, $\mathbf{T}^{-1}\mathbf{A}(\widehat{\mathbf{\Theta}})$,
  where $\mathbf{T}=\text{diag}(1, \mathbf{e}^{\mathbf{j}\alpha}, \mathbf{e}^{\mathbf{j}2\alpha}, \dots, \mathbf{e}^{(\mathbf{M}-1)\alpha})$, α is a random phase

Ambiguity do not affect hybrid beamforming

$$||(\mathbf{W} \odot \mathbf{\Phi})\mathbf{A}(\mathbf{\Theta}_l)\boldsymbol{\gamma}_l||^2 = ||(e^{-j\beta}\mathbf{W}\mathbf{T} \odot \mathbf{\Phi})\mathbf{T}^{-1}\mathbf{A}(\mathbf{\Theta}_l)\boldsymbol{\gamma}_l||^2$$

W.L.O.G, we can set $\omega_{11} = 0$ and $\theta_1 = 0$.
This constraints can be ignored when running the Algorithm.
After obtaining a solution, we can simply remove the ambiguity by $\omega_{km} = (\omega_{km} - \omega_{11})$ and $\theta_l = (\theta_l - \theta_1)$.