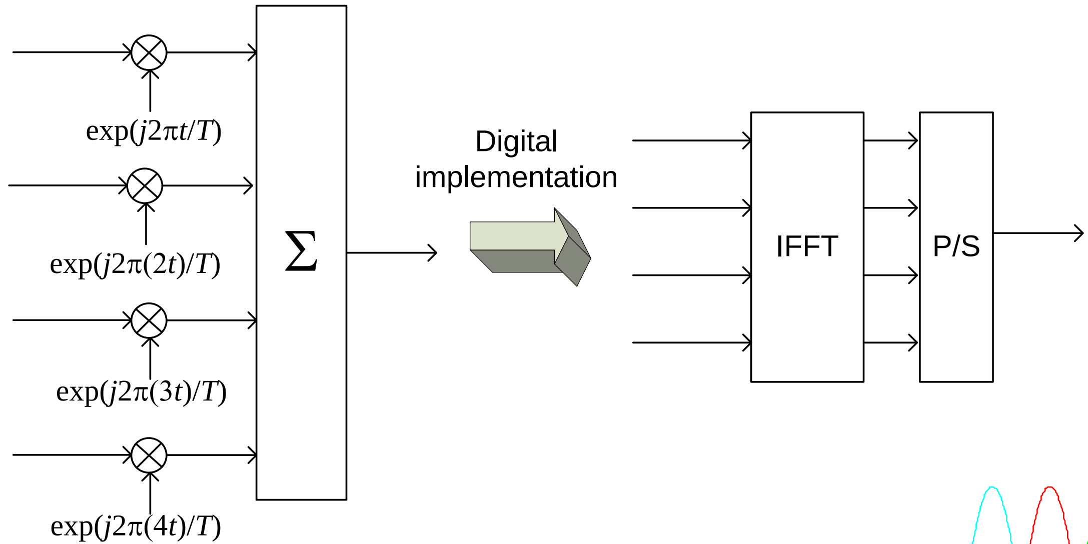
# On joint CFO and channel estimation in single-user and multi-user OFDM systems

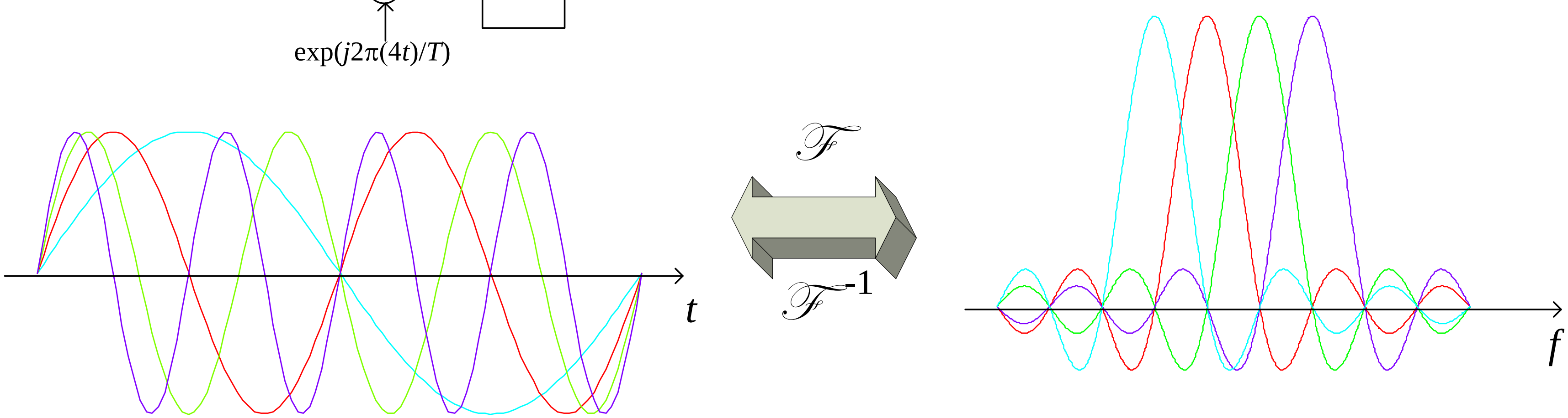
Yik-Chung Wu

The University of Hong Kong

Email: ycwu@eee.hku.hk, Webpage: www.eee.hku.hk/~ycwu

# OFDM basics



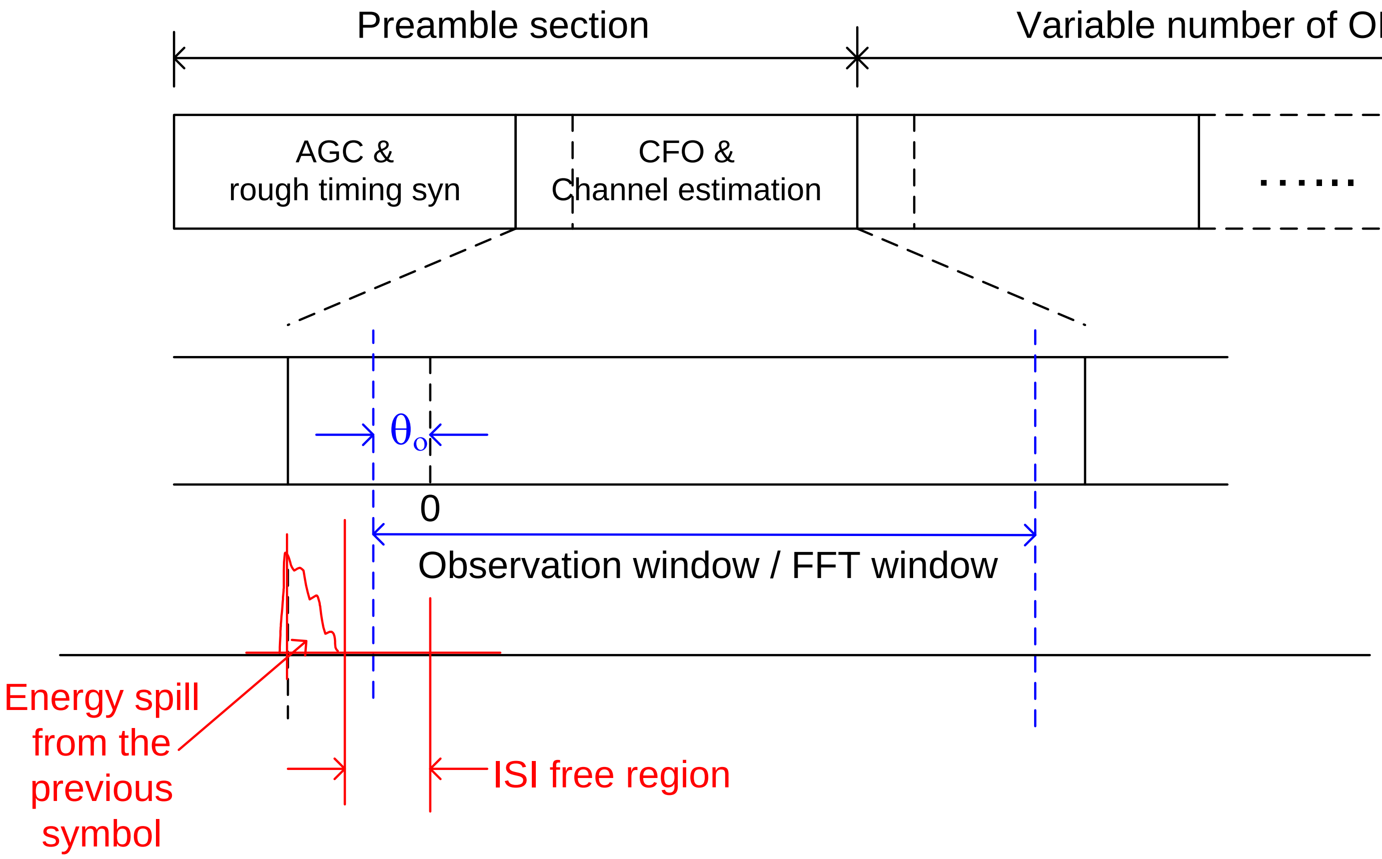

- Challenges: Time offset, frequency offset, unknown channel

# Outline

- Conventional OFDM [1]
  - ML estimator under timing uncertainty
  - Two ML estimators
  - Relationship and comparison
- OFDMA [2]
  - ML estimator
  - Optimization theorem
  - Importance sampling
- MIMO-OFDM [3]
  - CRB and asCRB
  - Optimal training
  - Performances of different kinds of training

# Conventional OFDM

- Packet structure



# Two equivalent signal models

- If the FFT window starts in the ISI-free region:

$$\mathbf{x} = \boldsymbol{\Gamma}(\omega_o)\mathbf{T}(\theta_o)\mathbf{F}^H\mathbf{D}\mathbf{F}_L\mathbf{h} + \mathbf{v} \qquad (1)$$

$\boldsymbol{\Gamma}(\omega_o)$ : Diagonal matrix with $e^{jn\omega_o}$ ($n$=0,1,...$N$-1) on the diagonal

$\omega_o \triangleq 2\pi\varepsilon_o / N$ : Normalized frequency offset

$\mathbf{T}(\theta_o)$ : Cyclic shift matrix

$\mathbf{F}$ : N×N FFT matrix

$\mathbf{D}$ : Diagonal matrix with training data on the diagonal

$\mathbf{F}_L$ : First L columns of $\mathbf{F}$ matrix

$\mathbf{h}$ : L×1 channel vector

- If we treat the delay as part of the channel:

$$\mathbf{x} = \boldsymbol{\Gamma}(\omega_o)\mathbf{F}^H\mathbf{D}\mathbf{F}_{L_{cp}}\boldsymbol{\xi} + \mathbf{v} \qquad (2)$$

where $\boldsymbol{\xi} \triangleq [0^T_{\theta_o\times 1} \;\; \mathbf{h}^T \;\; 0^T_{(L_{cp}-\theta_o-L)\times 1}]^T$

# Two ML estimators

- For the first system model, with $\mathbf{A}(\theta)=\mathbf{T}(\theta)\mathbf{F}^H\mathbf{D}\mathbf{F}_L$ ,

$$\{\hat{\omega},\hat{\theta},\hat{\mathbf{h}}\} = \arg\min_{\tilde{\omega},\tilde{\theta},\tilde{\mathbf{h}}} \left\{ [\mathbf{x}-\mathbf{\Gamma}(\tilde{\omega})\mathbf{A}(\tilde{\theta})\tilde{\mathbf{h}}]^H [\mathbf{x}-\mathbf{\Gamma}(\tilde{\omega})\mathbf{A}(\tilde{\theta})\tilde{\mathbf{h}}]^H \right\}$$

- Since $\mathbf{h}$ is linear in the system model, ML for $\mathbf{h}$ is

$$\hat{\mathbf{h}} = (\mathbf{F}_L^H\mathbf{D}^H\mathbf{D}\mathbf{F}_L)^{-1}(\mathbf{\Gamma}(\tilde{\omega})\mathbf{A}(\tilde{\theta}))^H\mathbf{x} = \mathbf{A}^H(\tilde{\theta})\mathbf{\Gamma}^H(\tilde{\omega})\mathbf{x}$$

- Put 2nd eq. into the 1st eq. and dropping the irrelevant terms, we have

$$\{\hat{\omega},\hat{\theta}\} = \arg\max_{\tilde{\omega},\tilde{\theta}} \left\{ \left\| \mathbf{A}^H(\tilde{\theta})\mathbf{\Gamma}^H(\tilde{\omega})\mathbf{x} \right\|^2 \right\} \qquad (3)$$

- Eq. (3) requires a two-dimensional search

# Two ML estimators (cont.)

- With similar procedure applied to the second system model, we have

$$\hat{\boldsymbol{\xi}} = \mathbf{F}_{L_{cp}}^{H} \mathbf{D}^{H} \mathbf{F} \boldsymbol{\Gamma}^{H}(\tilde{\omega}) \mathbf{x}$$

$$\hat{\omega} = \arg\max_{\tilde{\omega}} \left\{ \left\| \mathbf{F}_{L_{cp}}^{H} \mathbf{D}^{H} \mathbf{F} \boldsymbol{\Gamma}^{H}(\tilde{\omega}) \mathbf{x} \right\|^{2} \right\} \qquad (4)$$

- Eq. (4) only requires a one-dimensional search
- Which estimator is better?
- Number of unknown parameters: L+2 vs $L_{cp}$+1
  - The first estimator will perform better
  - But with the price of higher complexity (due to 2-D search)

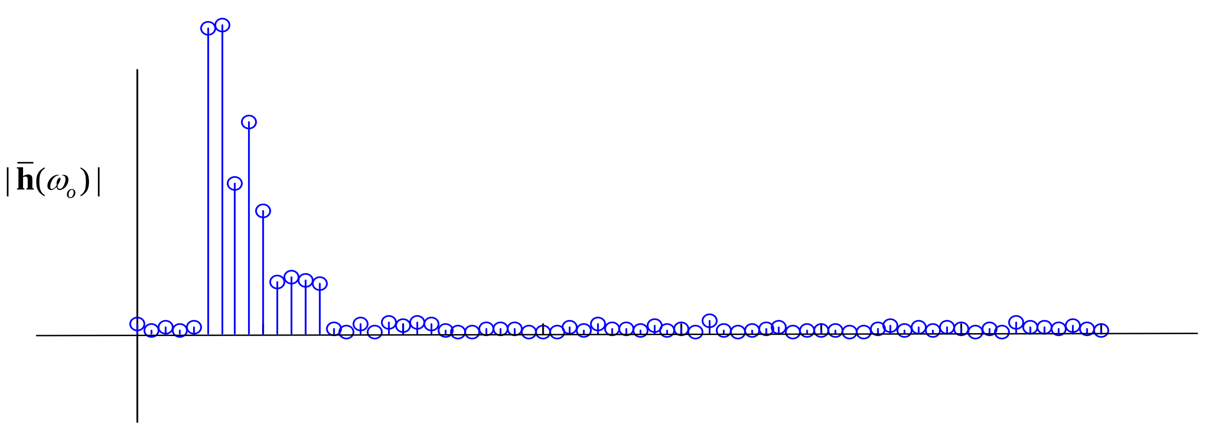
# Physical interpretation

- The cost functions in the two estimators are equivalent to

$$J_1(\tilde{\omega},\tilde{\theta}) = \left\| \mathbf{F}^H(:,\tilde{\theta}:\tilde{\theta}+L-1)\mathbf{D}^H\mathbf{F}\mathbf{\Gamma}^H(\tilde{\omega})\mathbf{x} \right\|^2$$

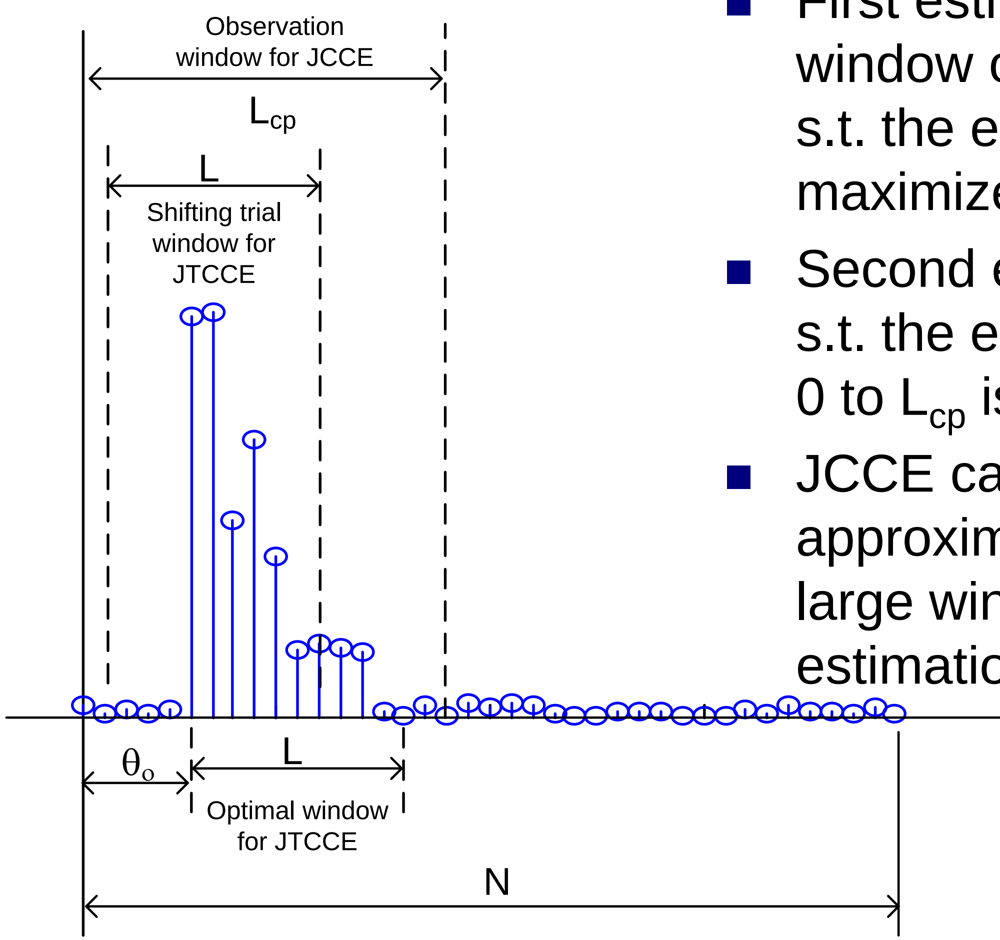
$$J_2(\tilde{\omega}) = \left\| \mathbf{F}^H(:,0:L_{cp}-1)\mathbf{D}^H\mathbf{F}\mathbf{\Gamma}^H(\tilde{\omega})\mathbf{x} \right\|^2$$

- They represent the energies of different sections of the vector $\bar{\mathbf{h}}(\tilde{\omega}) \triangleq \mathbf{F}^H\mathbf{D}^H\mathbf{F}\mathbf{\Gamma}^H(\tilde{\omega})\mathbf{x}$
- It can be shown that if $\tilde{\omega}=\omega_o$ , $\bar{\mathbf{h}}(\tilde{\omega})$ is a shifted time domain channel estimate



# Physical interpretation (cont.)

$$J_1(\tilde{\omega},\tilde{\theta}) = \left\| \overline{\mathbf{h}}(\tilde{\omega})_{\tilde{\theta}:\tilde{\theta}+L-1} \right\|^2, \qquad J_2(\tilde{\omega}) = \left\| \overline{\mathbf{h}}(\tilde{\omega})_{0:L_{cp}-1} \right\|^2$$




- First estimator (JTCCE): Locating a window of length L and finding an $\tilde{\omega}$ s.t. the energy within the window is maximized
- Second estimator (JCCE): Find an $\tilde{\omega}$ s.t. the energy within the window from 0 to $L_{cp}$ is maximized
- JCCE can be interpreted as an approximation to JTCCE by using a large window during frequency estimation

# Performance analysis

- The MSE and CRB expressions for the two estimators can be derived in closed form (eqs. not presented here)
- Analytical performance comparison:

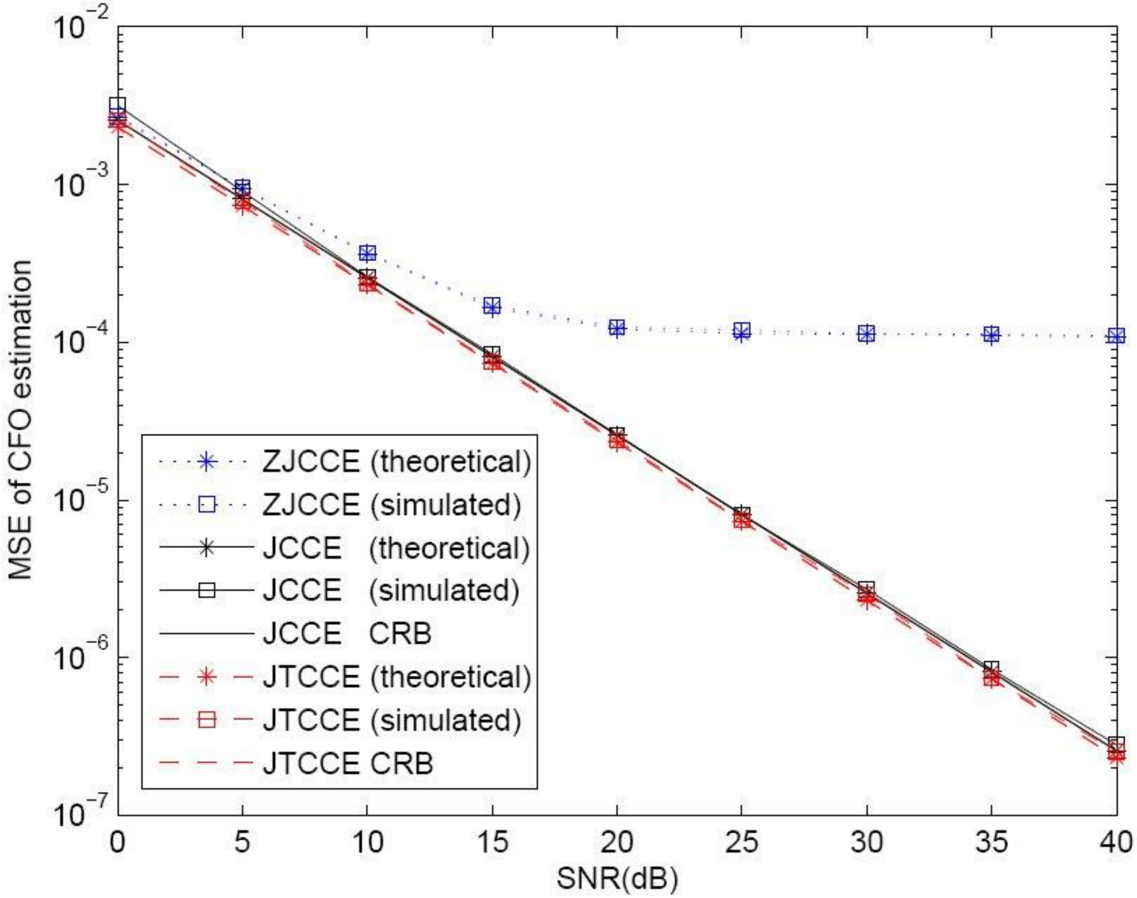
$$r \triangleq \frac{MSE_2}{MSE_1} = 1 + \frac{\left\| \mathbf{F}^H(:, L : L_{cp} - 1)\mathbf{D}^H \mathbf{F}\mathbf{M}\mathbf{F}^H \mathbf{D}\mathbf{F}_L \mathbf{h} \right\|^2}{\left\| \mathbf{F}^H(:, L_{cp} : N - 1)\mathbf{D}^H \mathbf{F}\mathbf{M}\mathbf{F}^H \mathbf{D}\mathbf{F}_L \mathbf{h} \right\|^2} > 1$$

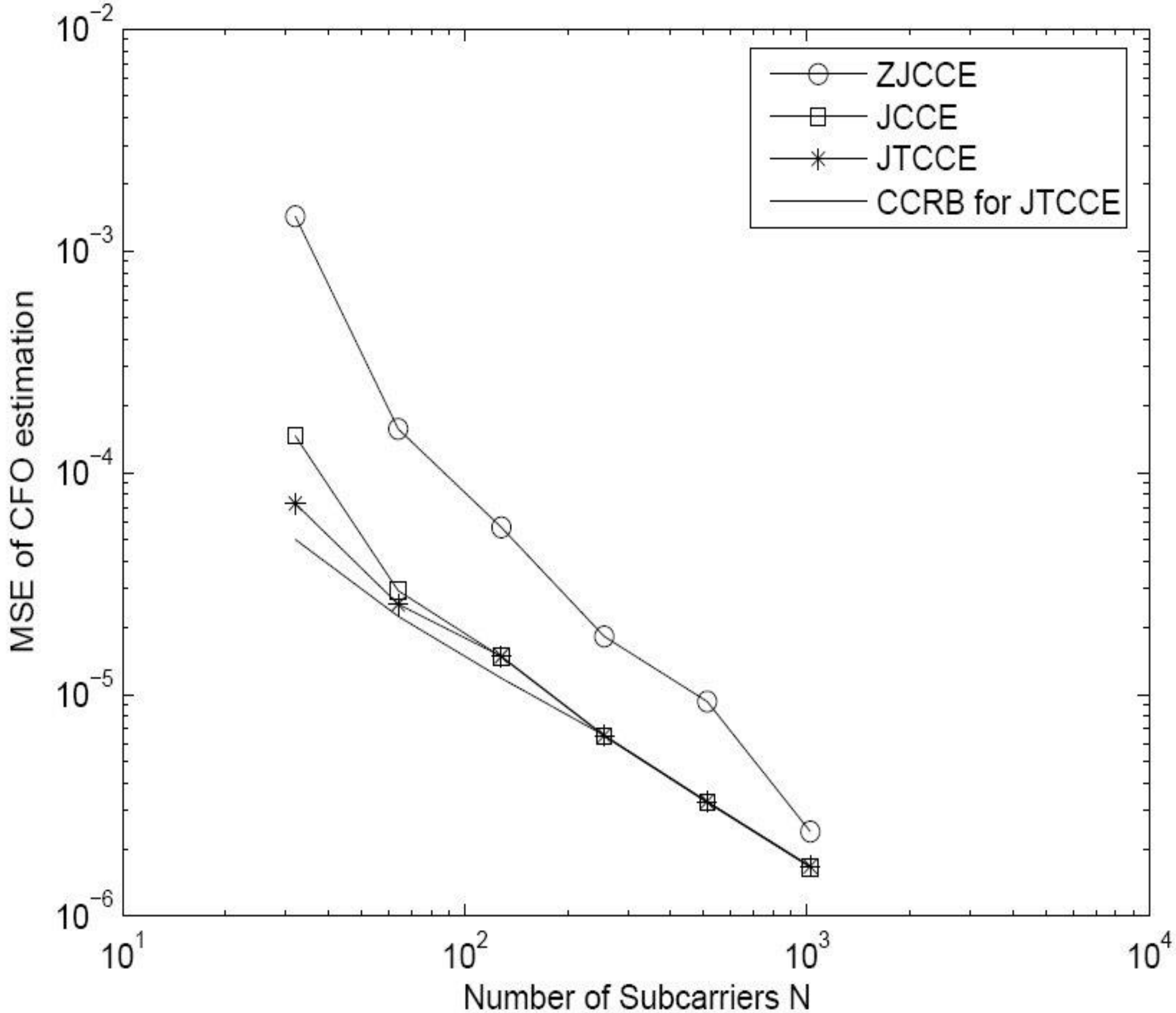
- The second term is the ratio between two projections of a common vector onto different subspaces of $\mathbf{F}$
- When N is large, dimension of $\mathbf{F}(:,L:L_{cp}\text{-}1)$ << that of $\mathbf{F}(:,L_{cp}:N\text{-}1)$

  $\Rightarrow$ Two estimators are asymptotically equivalent

# Numerical results

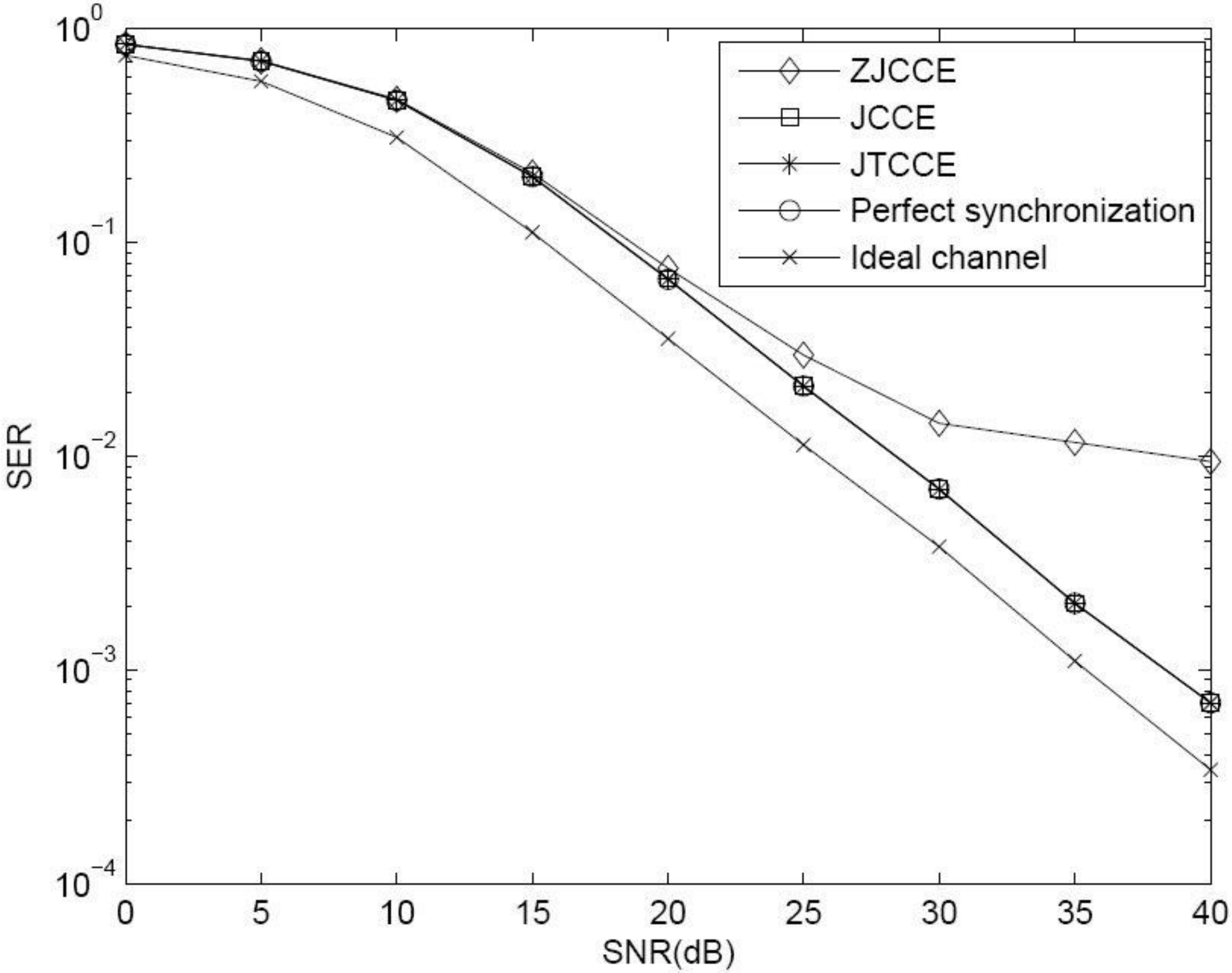
- N=64, $N_{cp}$=16, L=8
- Training: Chu sequence
- Rayleigh fading channel
- Exponential delay profile
- Normalized CFO uniformly distributed in [-0.5,0.5]
- $\theta_o$ uniformly distributed in the ISI-free region





ZJCCE
JCCE
JTCCE
CCRB for JTCCE
MSE of CFO estimation
Number of Subcarriers N

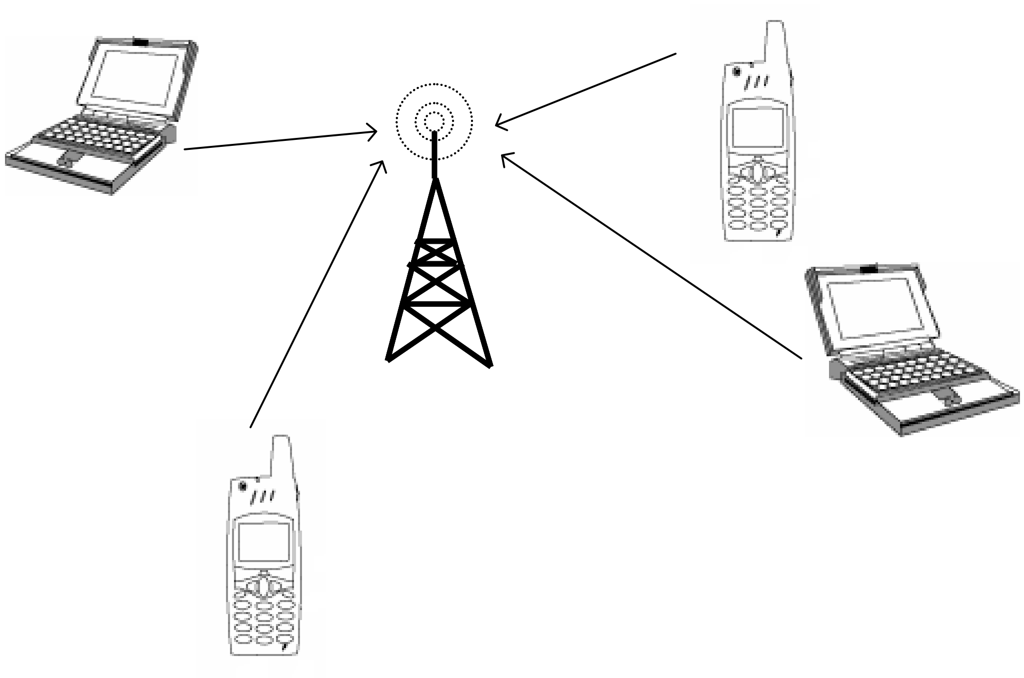
- 16QAM
- 5 OFDM symbols after the preamble

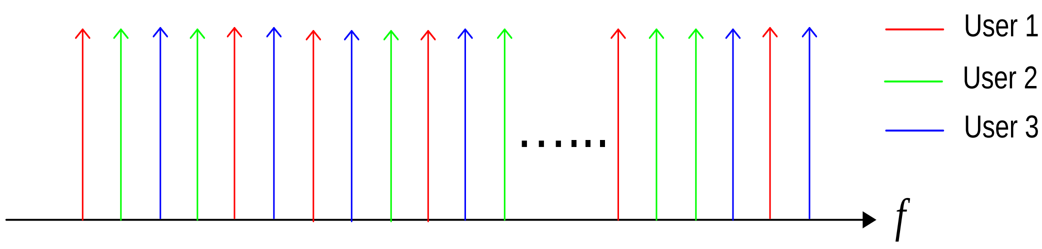


# Summary

- Based on two different signal models, two ML estimators for joint CFO and channel estimation with timing ambiguity have been derived
- The first estimator needs a 2-D search
- The second estimator only requires 1-D search
- The first estimator perform slightly better than the second one, with the price of higher complexity
- Asymptotically, the two estimators are equivalent

# OFDMA uplink

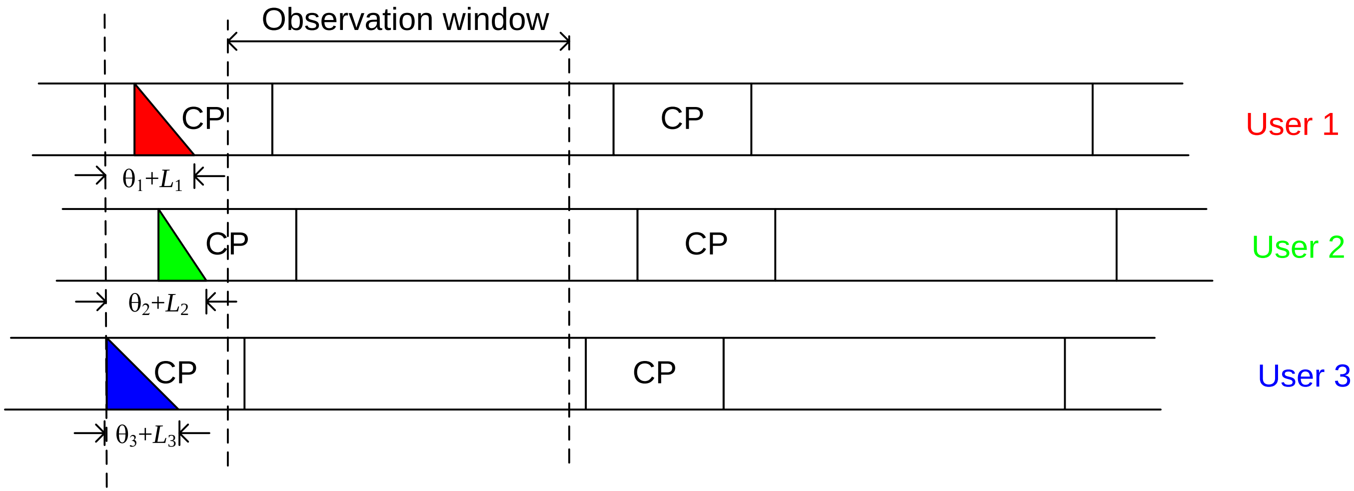





- Each user modulates an exclusive set of subcarrier
- Subcarrier can be allocated blockwise, interleaved, or arbitrarily
- Challenges: Different user have different timing delays, CFOs and channels

# Timing offset problem

- The user's timing is roughly synchronized using the downlink synchronization channel
- Timing offsets in the uplink are mainly due to the propagation delay from different users => Limited to a few samples
- Can be treated as part of the channel
- Called this quasi-synchronous system
- As long as $\max(L_k+\theta_k) < L_{cp}$, we will have an observation window free of ISI



# System model

- Signal of user k

$$\mathbf{x}_k = \mathbf{\Gamma}(\omega_k)\mathbf{F}^H \mathbf{D}_k \mathbf{F}_L \mathbf{h}_k$$

$\mathbf{D}_k$ : Diagonal matrix with training data for user k

$\mathbf{h}_k$ : User's k channel, including the time delay $\theta_k$

- Received signal $\mathbf{x} = \sum_{k=1}^{K} \mathbf{\Gamma}(\omega_k)\mathbf{F}^H \mathbf{D}_k \mathbf{F}_L \mathbf{h}_k + \mathbf{n}$

- Or in matrix form $\mathbf{x} = \mathbf{Q}(\boldsymbol{\omega})\mathbf{h} + \mathbf{n}$

$$\mathbf{Q}(\boldsymbol{\omega}) = [\mathbf{\Gamma}(\omega_1)\mathbf{F}^H \mathbf{D}_1 \mathbf{F}_L \quad \mathbf{\Gamma}(\omega_2)\mathbf{F}^H \mathbf{D}_2 \mathbf{F}_L \quad \ldots \quad \mathbf{\Gamma}(\omega_K)\mathbf{F}^H \mathbf{D}_K \mathbf{F}_L]$$

$$\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \ldots \ \mathbf{h}_K^T]^T$$

- Goal: To estimate $\boldsymbol{\omega}$ and $\mathbf{h}$, based on a single OFDM training symbol

# ML joint CFO and channel estimator

- Based on the standard procedure of deriving ML estimator, we have

$$\hat{\boldsymbol{\omega}} = \arg\max_{\tilde{\boldsymbol{\omega}}} \{\mathbf{x}^H \mathbf{Q}(\tilde{\boldsymbol{\omega}})(\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})\mathbf{Q}(\tilde{\boldsymbol{\omega}}))^{-1}\mathbf{Q}^H(\tilde{\boldsymbol{\omega}})\mathbf{x}\}$$

$$\hat{\mathbf{h}} = (\mathbf{Q}^H(\hat{\boldsymbol{\omega}})\mathbf{Q}(\hat{\boldsymbol{\omega}}))^{-1}\mathbf{Q}^H(\hat{\boldsymbol{\omega}})\mathbf{x}$$

- Multi-dimensional search in $\boldsymbol{\omega} \Rightarrow$ Computational expensive
- Exhaustive search impossible for K $\geq$ 3
- Possible method: alternative projection
- However, no guarantee of global maximum solution
- We employ an optimization theorem to solve this problem

# Simple solution in asymptotic case

- If the number of subcarrier $N \to \infty$, it can be shown that

$$(\mathbf{Q}^H(\boldsymbol{\omega})\mathbf{Q}(\boldsymbol{\omega}))^{-1} = \begin{bmatrix} (\mathbf{A}_1^H\mathbf{A}_1)^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & (\mathbf{A}_K^H\mathbf{A}_K)^{-1} \end{bmatrix}$$

  where $\mathbf{A}_k = \mathbf{F}^H\mathbf{D}_k\mathbf{F}_L$
- Therefore, the optimal CFO estimator can be decoupled as

$$\hat{\omega}_k = \arg\max_{\tilde{\omega}_k}\{\mathbf{x}^H\boldsymbol{\Gamma}(\tilde{\omega}_k)\mathbf{A}_k(\mathbf{A}_k^H\mathbf{A}_k)^{-1}\mathbf{A}_k^H\boldsymbol{\Gamma}^H(\tilde{\omega}_k)\mathbf{x}\}$$

- For large enough but finite $N$, it can be viewed as an approximate solution
- For small $N$, it suffer great performance loss

# Optimization theorem

- Optimization theorem by Pincus [R1]:

The global optimal solution (if it is unique) for a multi-dimensional optimization problem maximizing $L'(\tilde{\boldsymbol{\omega}})$ is given by

$$\hat{\omega}_k = \lim_{\rho_1 \to \infty} \frac{\int \ldots \int \omega_k \exp(\rho_1 L'(\boldsymbol{\omega})) d\boldsymbol{\omega}}{\int \ldots \int \exp(\rho_1 L'(\boldsymbol{\omega})) d\boldsymbol{\omega}} \qquad k = 1, \ldots, K \tag{5}$$

- Good news: If we are smart enough to perform the integration, we can get the optimal solution analytically!!
- Bad news: The integration usually is too complex to be computed analytically
- Physical meaning of the theorem:
  - Taking exponential make the largest peak in $L'(\omega)$ 'peaker' and other smaller peaks lower
  - When $\rho_1$ is large enough, $\exp(\rho_1 L'(\omega))$ will have only a single peak
  - The maximum point is the mean of the resultant function

[R1] M. Pincus, "A closed form solution for certain programming problems," Oper. Res., pp. 690-694, 1962

# Approximation by sample mean

- The optimization theorem can rewritten as (for large $\rho_1$)

$$\hat{\omega}_k = \int \ldots \int \omega_k \bar{L}(\boldsymbol{\omega}) d\boldsymbol{\omega} \qquad k = 1, \ldots, K$$

  where $\bar{L}(\boldsymbol{\omega}) = \dfrac{\exp(\rho_1 L'(\boldsymbol{\omega}))}{\int \ldots \int \exp(\rho_1 L'(\boldsymbol{\omega})) d\boldsymbol{\omega}}$ is termed pseudo-PDF

- This is just the statistical mean of $\omega_k$ w.r.t. PDF $\bar{L}(\boldsymbol{\omega})$
- If we can generate a large number of realization of $\omega_k$ according to the PDF $\bar{L}(\boldsymbol{\omega})$, the integration can be approximated by the sample mean

$$\hat{\omega}_k \approx \frac{1}{T} \sum_{i=1}^{T} \omega_k^i \qquad k = 1, \ldots, K$$

- Then the question becomes how to generate samples from $\bar{L}(\boldsymbol{\omega})$

# Importance sampling

- In general, generating samples directly from $\bar{L}(\boldsymbol{\omega})$ is difficult since it is a multi-dimensional PDF
- Generating realization from an arbitrary but fixed PDF is a well-studied problem in statistics
- Here, we use a technique called importance sampling
- It is based on the observation that

$$\hat{\omega}_k = \int \ldots \int \omega_k \bar{L}(\boldsymbol{\omega}) d\boldsymbol{\omega} = \int \ldots \int \omega_k \frac{\bar{L}(\boldsymbol{\omega})}{\bar{g}(\boldsymbol{\omega})} \bar{g}(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

  where $\bar{g}(\boldsymbol{\omega}) = \dfrac{g(\boldsymbol{\omega})}{\int \ldots \int g(\boldsymbol{\omega}) d\boldsymbol{\omega}}$ is called normalized importance function

- If $\bar{g}(\boldsymbol{\omega})$ is chosen s.t. realization of $\omega_k$ can be easily generated

$$\hat{\omega}_k \approx \frac{1}{T} \sum_{i=1}^{T} \omega_k^i \frac{\bar{L}(\boldsymbol{\omega}^i)}{\bar{g}(\boldsymbol{\omega}^i)}$$

# Further simplification

- In our problem, $\omega_k$ is the CFO, so it is a circular R.V.
- The estimator can be further rewritten as

$$\hat{\omega}_k = \frac{1}{T}\sum_{i=1}^{T}\omega_k^i \frac{\bar{L}(\boldsymbol{\omega}^i)}{\bar{g}(\boldsymbol{\omega}^i)} = \frac{1}{2\pi}\angle\frac{1}{T}\sum_{i=1}^{T}\exp(j2\pi\omega_k^i)\frac{L(\boldsymbol{\omega}^i)}{g(\boldsymbol{\omega}^i)}$$

  where $L(\boldsymbol{\omega})=\exp(\rho_1\mathbf{x}^H\mathbf{Q}(\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{x})$ is the non-normalized version of $\bar{L}(\boldsymbol{\omega})$ and $g(\boldsymbol{\omega})$ is the non-normalized version of $\bar{g}(\boldsymbol{\omega})$

- Advantages:
  - Eliminates potential bias
  - Computation of normalization constants for $L(\boldsymbol{\omega})$ and $g(\boldsymbol{\omega})$ can be avoided

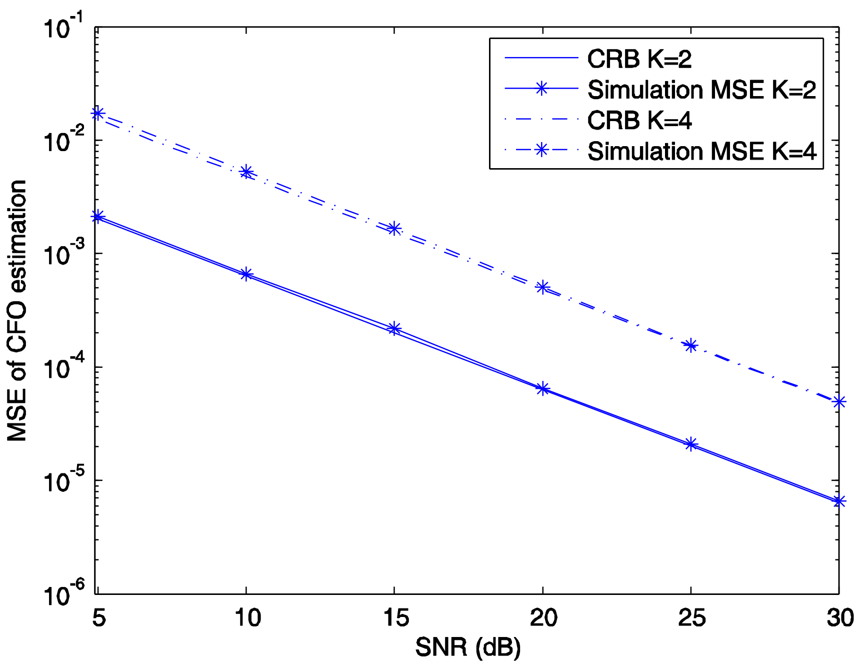
# Choosing $g(\boldsymbol{\omega})$

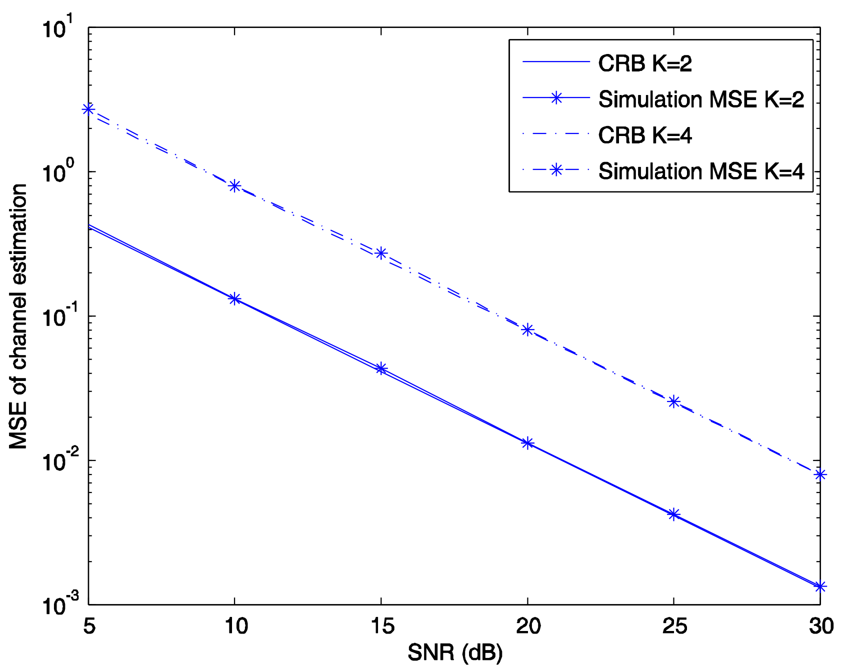
- By the strong laws of large number, the estimate $\hat{\omega}_k$ will converge to optimal value, regardless of choice of $g(\boldsymbol{\omega})$
- Choice of $g(\boldsymbol{\omega})$ only affects
  - computational complexity
  - How fast the estimate converge to the true value
- General guidelines for choosing $g(\boldsymbol{\omega})$ :
  - Easy sample generation
  - Close to $L(\boldsymbol{\omega})$ in order to reduce variance of the estimate
- From the discussion in asymptotic case, propose to choose $g(\boldsymbol{\omega})$ as (with $\rho_2 < \rho_1$ )

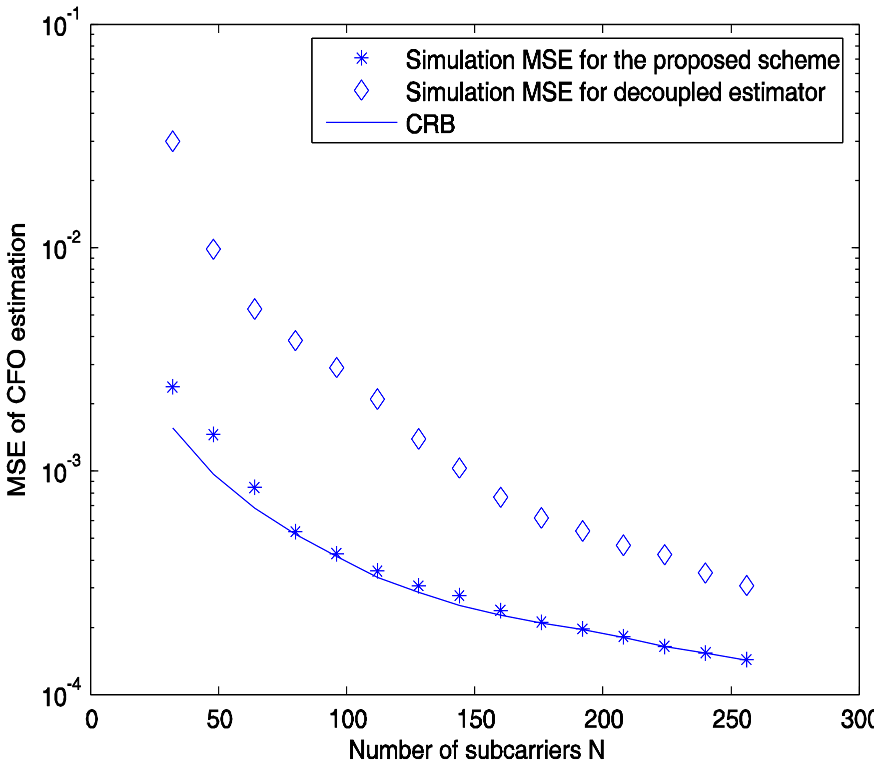
$$
\begin{aligned}
g(\boldsymbol{\omega}) &= \exp(\rho_2 \sum_{k=1}^{K} \mathbf{x}^H \boldsymbol{\Gamma}(\omega_k) \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H \boldsymbol{\Gamma}^H(\omega_k) \mathbf{x}) \\
&= \prod_{k=1}^{K} \underbrace{\exp(\rho_2 \mathbf{x}^H \boldsymbol{\Gamma}(\omega_k) \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H \boldsymbol{\Gamma}^H(\omega_k) \mathbf{x})}_{\triangleq g_k(\omega_k)}
\end{aligned}
$$

# Numerical results

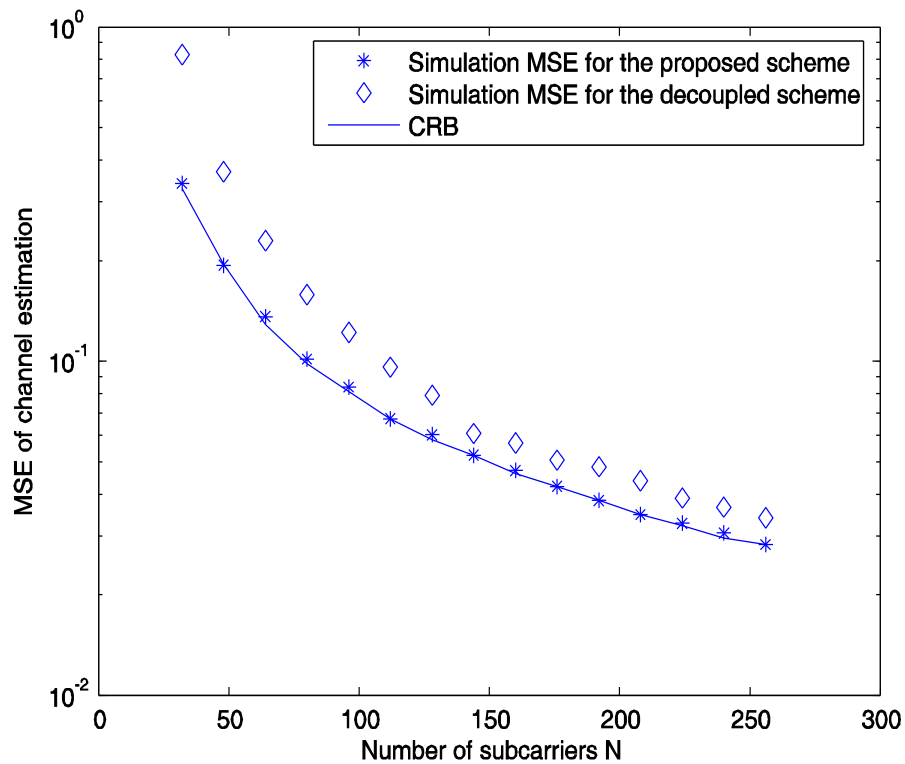
- $N$=64, $L_{cp}$=16, $L$=8, no time delay, random subcarrier allocation, $\varepsilon_k$ uniformly distributed [-N/2,N/2]
- $\rho_1$=2/K, $\rho_2$=1/K, $T$=2000
- Training: constant modulus white seq in frequency domain
- $MSE_{CFO} \triangleq \sum_{k=1}^{K} (\hat{\omega}_k - \omega_k)^2, \quad MSE_{ch} \triangleq \| \hat{\mathbf{h}} - \mathbf{h} \|^2$





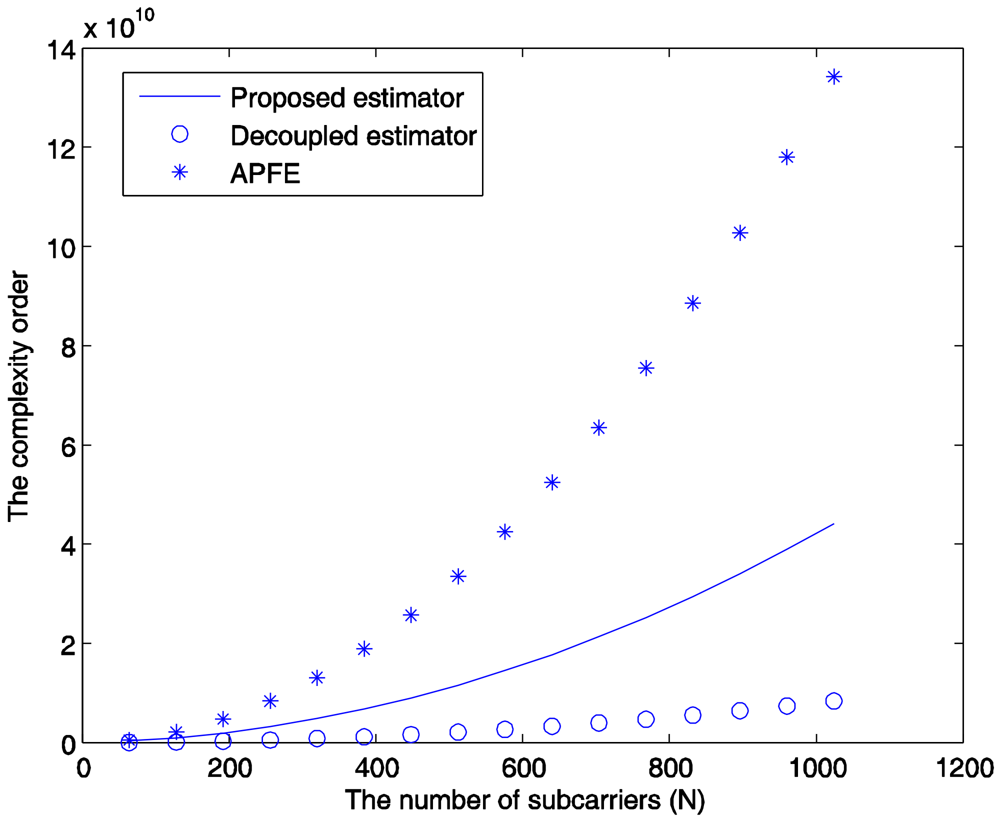
Simulation MSE for the proposed scheme
Simulation MSE for decoupled estimator
CRB
MSE of CFO estimation
Number of subcarriers N

Simulation MSE for the proposed scheme
Simulation MSE for the decoupled scheme
CRB
MSE of channel estimation
Number of subcarriers N

# Complexity comparison

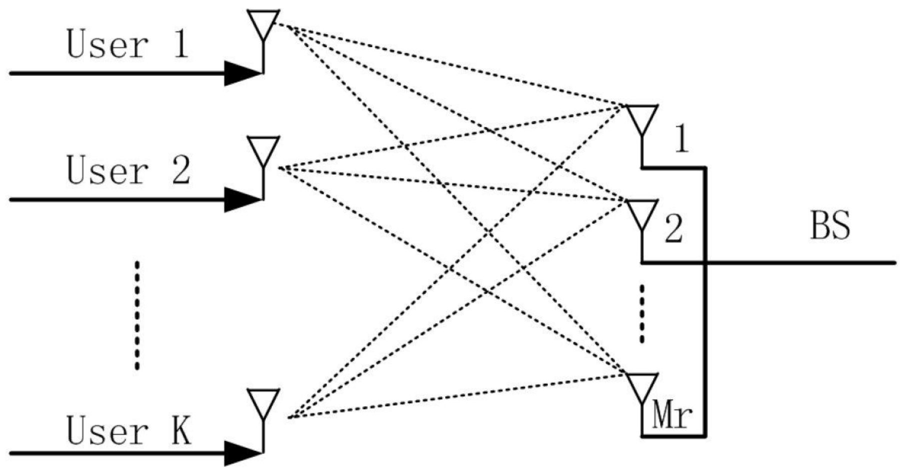
- Complexity order has been derived for the proposed estimator, the decoupled estimator and the alternative projection method (APFE)



# Summary

- Direct implementation of ML joint CFO and channel estimation is impractical due to the multi-dimensional search
- An optimization theorem, together with the importance sampling technique is used to solve this problem
- The proposed method can guarantee global optimality without the need of providing a ‘good’ initial estimate

# Multi-user MIMO-OFDM system




- Assume quasi-synchronous s.t. the small time delays can be lumped into the channels
- With spatial multiplexing (e.g., BLAST), each user is using all the subcarriers at the same time
- The receive antennas at BS share the same oscillator
- All users are driven by different oscillators
- Can be easily generalized to the case where user equips with more than one antenna

# System model

- Received signal (one OFDM symbol) at the jth receive antenna $$\mathbf{x}_j = \sum_{i=1}^{K} \mathbf{\Gamma}(\omega_i)\mathbf{A}_i\mathbf{h}_{ij} + \mathbf{n}_j$$
- Stack all the received vector from different receive antenna

$$\mathbf{x} = \mathbf{Q}(\boldsymbol{\omega})\mathbf{h} + \mathbf{n}$$

where $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \ldots \ \mathbf{x}_{M_r}^T]^T, \ \mathbf{n} = [\mathbf{n}_1^T \ \mathbf{n}_2^T \ \ldots \ \mathbf{n}_{M_r}^T]^T$

$$\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \ldots \ \mathbf{h}_{M_r}^T]^T \text{ with } \mathbf{h}_j = [\mathbf{h}_{1j}^T \ \mathbf{h}_{2j}^T \ \ldots \ \mathbf{h}_{Kj}^T]^T$$

$$\mathbf{Q}(\boldsymbol{\omega}) = \mathbf{I}_{M_r} \otimes [\mathbf{\Gamma}(\omega_1)\mathbf{A}_1 \ \ldots \ \mathbf{\Gamma}(\omega_K)\mathbf{A}_K] \text{ with } \mathbf{A}_i = \mathbf{F}^H\mathbf{D}_i\mathbf{F}_L$$

- This linear model is in the same form as OFDMA case
- For estimation of $\boldsymbol{\omega}$ and $\mathbf{h}$, we can use a similar procedure as in OFDMA case
- What is the optimal training sequences?

# CRB

- It can be shown that the CRB for the joint CFO and channel estimation problem is

$$CRB(\boldsymbol{\omega}) = \frac{\sigma^2}{2}(\Re\{\mathbf{Z}^H \mathbf{\Pi}_Q^{\perp} \mathbf{Z}\})^{-1}$$

$$CRB(\mathbf{h}) = \frac{\sigma^2}{2}\left[2(\mathbf{Q}^H\mathbf{Q})^{-1} + (\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{Z}(\Re\{\mathbf{Z}^H\mathbf{\Pi}_Q^{\perp}\mathbf{Z}\})^{-1}\mathbf{Z}^H\mathbf{Q}(\mathbf{Q}^H\mathbf{Q})^{-H}\right]$$

where $\mathbf{\Pi}_Q^{\perp} = \mathbf{I}_{M_r N} - \mathbf{Q}(\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{21} \cdots & \mathbf{Z}_{K1} \\ \mathbf{Z}_{12} & \mathbf{Z}_{22} \quad \cdots & \mathbf{Z}_{K2} \\ \vdots & \vdots \quad \ddots & \vdots \\ \mathbf{Z}_{1M_r} & \mathbf{Z}_{1M_r} \cdots & \mathbf{Z}_{KM_r} \end{bmatrix}, \text{ with } \mathbf{Z}_{ij} = \underbrace{diag(0,1,\ldots,N-1)}_{\triangleq \mathbf{M}} \cdot \mathbf{\Gamma}(\omega_k)\mathbf{A}_i\mathbf{h}_{ij}$$

# Special case: CFO-free

- If there is no CFO, the CRB reduces to

$$CRB(\mathbf{h})_{CFO-free} = \sigma^2 \mathbf{I}_{M_r} \otimes (\mathbf{A}^H \mathbf{A})^{-1}$$

$\mathbf{A} \triangleq [\mathbf{A}_1 \ \mathbf{A}_2 \ \ldots \ \mathbf{A}_K]$

- It is shown [R2] that the condition of minimizing $Tr\{CRB(\mathbf{h})_{CFO-free}\}$ is $\mathbf{A}^H \mathbf{A} \propto \mathbf{I}_{KL}$
- It is further shown in [2] that two types of training satisfy the condition
  - FDM pilot: $$d_k(n) = \sum_{l=0}^{L_1-1} b_k^{(l)} \delta[n - lN / L_1 - k]$$ $n = 0,1,\ldots,N-1$, $k = 0,1,\ldots,K-1$
    - * $\{b_k^{(l)}\}$ are constant-modulus symbols
    - * $L_1$ is any integer s.t. $N / L_1$ is an integer while $L \le L_1 \le N / K$
  - CDM(F) pilot: $$d_k(n) = \exp(j\zeta_n)\exp(j\phi_k)\exp(-j2\pi(k-1)n / N)$$
    - * $\zeta_n$ and $\phi_k$ are R.V.s in $[0, 2\pi]$ w.r.t. $n$ and $k$

[R2] H. Minn and N. Al-Dhahir, “Optimal training signals for MIMO OFDM channel estimation,” *IEEE Trans. Commun.*, May 2006

# Asymptotic CRB

- If the CFOs are not zero, the conditions for the optimal training cannot be obtained in closed-form
- We turn to asymptotic CRB:

$$asCRB(\boldsymbol{\omega}) = \frac{6\sigma^2}{N^3}(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\})^{-1}$$

$$asCRB(\mathbf{h}) = \frac{\sigma^2}{N}\left[\mathbf{R}^{-1} + \frac{3}{2}\mathbf{H}(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\})^{-1}\mathbf{H}^H\right]$$

Much Simpler!!!

where $\mathbf{H} = [\mathbf{H}_1^T \; \mathbf{H}_2^T \; \ldots \; \mathbf{H}_{M_r}^T]^T$ with $\mathbf{H}_j = diag([\mathbf{h}_{1j} \; \mathbf{h}_{2j} \; \ldots \; \mathbf{h}_{Kj}])$

$$\mathbf{R} = \mathbf{I}_{M_r} \otimes \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1K} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{K1} & \mathbf{R}_{K2} & \cdots & \mathbf{R}_{KK} \end{bmatrix}, \text{ with } \mathbf{R}_{ij} = \frac{\mathbf{A}_i^H \mathbf{A}_j}{N}\delta(\omega_i - \omega_j)$$

# Minimizing the asCRB

- Using the fact that $((\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\})^{-1})_{k,k} \geq \frac{1}{(\Re\{\mathbf{H}^H\mathbf{R}\mathbf{H}\})_{k,k}}$

  with equality holds iff $\mathbf{H}^H\mathbf{R}\mathbf{H}$ is diagonal
- It can be shown that

$$\mathrm{Tr}(asCRB(\boldsymbol{\omega})) \geq \frac{6\sigma^2}{N^3}\sum_{k=1}^{K}\frac{1}{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{R}_{k,k}\mathbf{h}_{ki}}$$

$$\mathrm{Tr}(asCRB(\mathbf{h})) \geq \frac{\sigma^2}{N}\left[\mathrm{Tr}(\mathbf{R}^{-1}) + \frac{3}{2}\sum_{k=1}^{K}\frac{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{h}_{ki}}{\sum_{i=1}^{M_r}\mathbf{h}_{ki}^H\mathbf{R}_{k,k}\mathbf{h}_{ki}}\right]$$

  with equality hold iff $\mathbf{R} \propto \mathbf{I}_{M_r KL}$
- It can be further shown that this is equivalent to

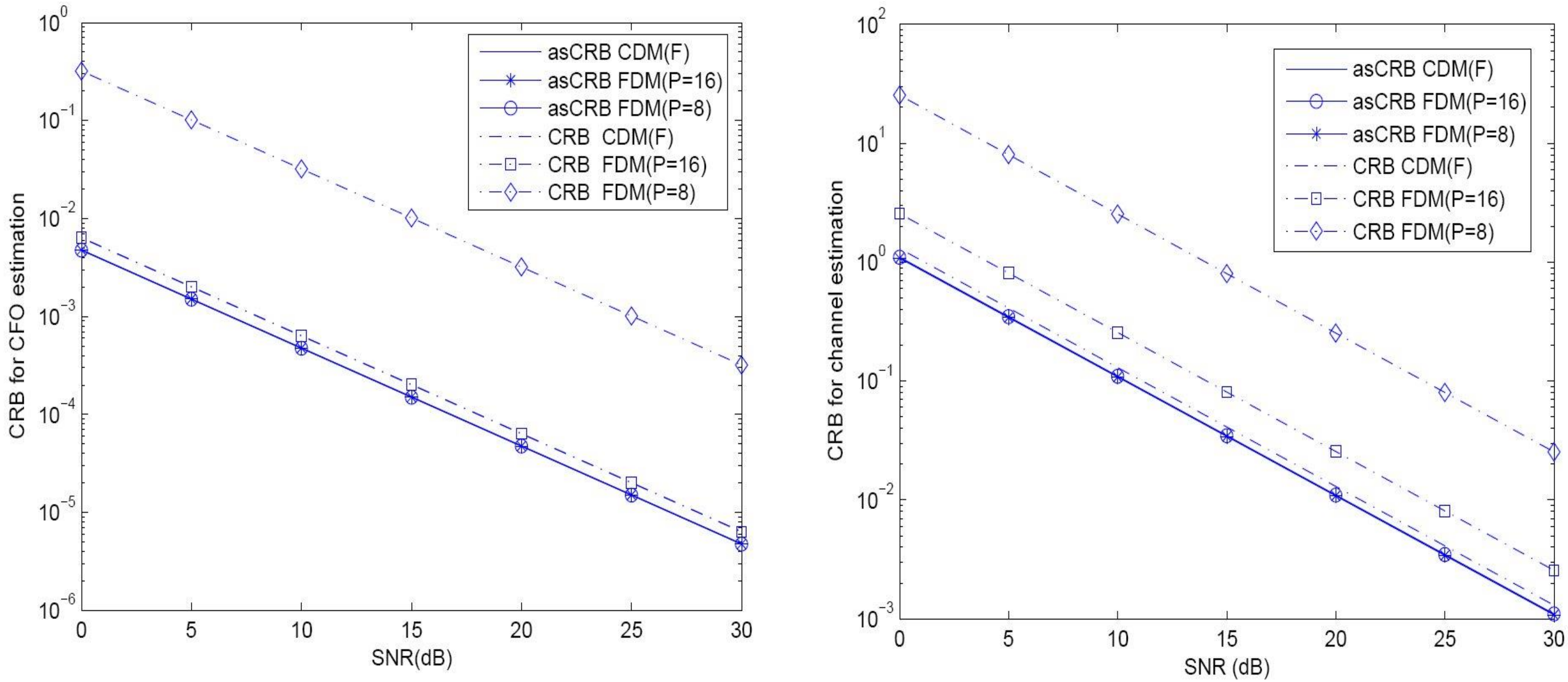
$$\mathbf{A}^H\mathbf{A} \propto \mathbf{I}_{KL}$$

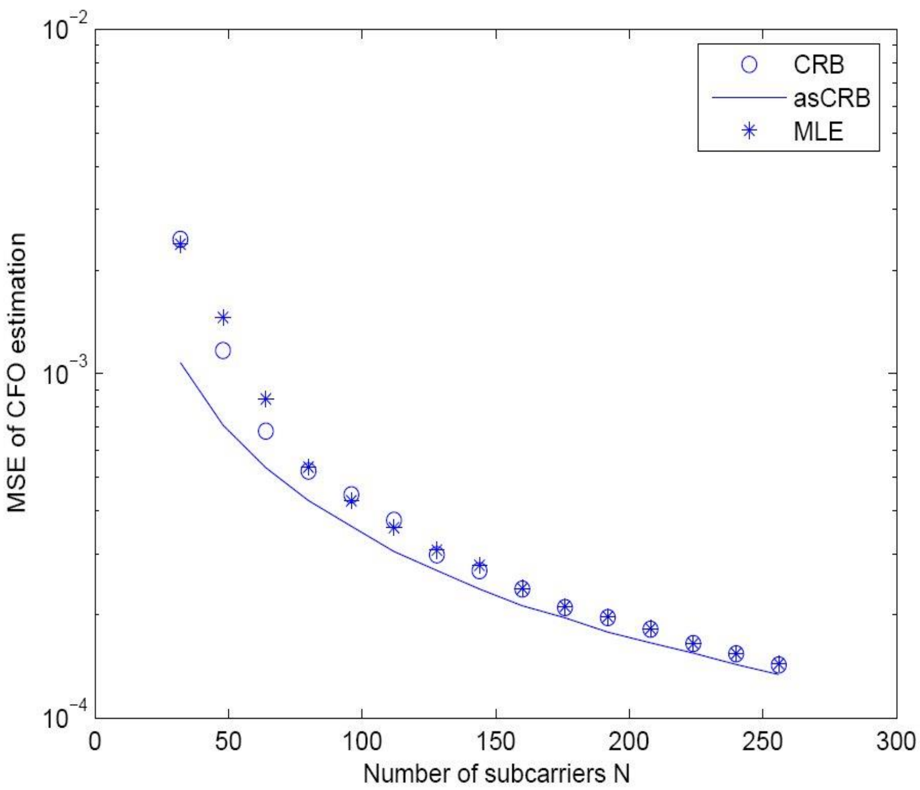
# Optimal training

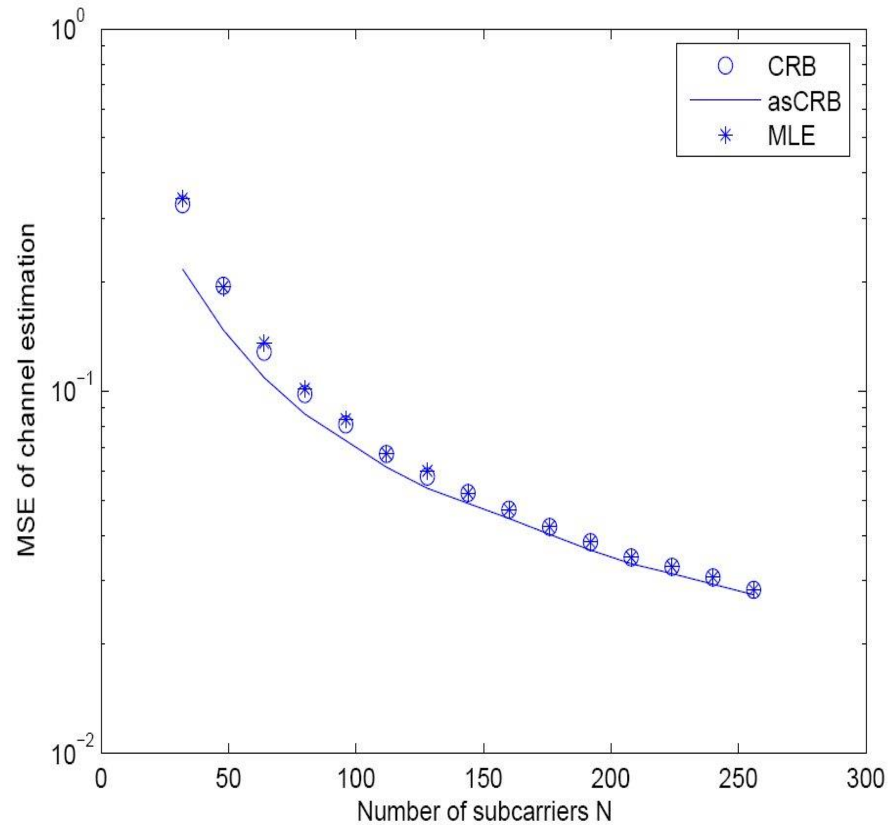
- The optimal training for CFO-free case (FDM and CDM(F) sequences) are also asymptotically optimal for joint CFO and channel estimation
- Question: for finite number of subcarriers, will they perform differently?
- Fact 1: CRB $\geq$ asCRB
- Fact 2: Training with correlation in time domain would have the CRB depart from the asCRB
- FDM sequence is repetitive in time domain $\Rightarrow$ correlated
- CDM(F) sequence has relatively a long correlation
- Prediction: CDM(F) performs better than FDM sequence

# Numerical results

- $N$=64, $L_{cp}$=16, $L$=8, $M_r$=2, $K$=2
- Exponential delay profile




CRB
asCRB
MLE
$10^{-2}$
$10^{-3}$
$10^{-4}$
MSE of CFO estimation
0
50
100
150
200
250
300
Number of subcarriers N


CRB
asCRB
MLE
$10^{0}$
$10^{-1}$
$10^{-2}$
MSE of channel estimation
0
50
100
150
200
250
300
Number of subcarriers N

# Summary

- Condition for optimal training has been derived by minimizing the asCRB
- Both CDM(F) and FDM sequences are asymptotically the best
- For finite number of subcarriers, CDM(F) seq perform better than FDM seq

# References

[1] Jianwu Chen, Yik-Chung Wu, Shaodan Ma and Tung-Sang Ng, “ML Joint CFO and Channel Estimation in OFDM systems with Timing Ambiguity,” *IEEE Trans. on Wireless Communications*, vol. 7, no. 7, pp. 2436-2440, Jul 08.

[2] Jianwu Chen, Yik-Chung Wu, S. C. Chan and Tung-Sang Ng, “Joint maximum-likelihood CFO and channel estimation for OFDMA uplink using importance sampling,” *IEEE Trans. on Vehicular Technology*, vol. 57, no. 6, pp.3462-3470, Nov. 2008.

[3] Jianwu Chen, Yik-Chung Wu, Shaodan Ma and Tung-Sang Ng, “Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences,” *IEEE Trans. on Signal Processing*, vol. 56, no. 8, pp. 4008-4019, Aug 08.

Further related readings:

- Kun Cai, Xiao Li, Jian Du, Yik-Chung Wu and Feifei Gao, “CFO Estimation in OFDM Systems under Timing and Channel Length Uncertainties with Model Averaging,” *IEEE Trans. on Wireless Communications,* Vol. 9, no. 3, pp. 970-974, Mar 2010.
- Xun Cai, Yik-Chung Wu, Hai Lin and Katsumi Yamashita, “Estimation and Compensation of CFO and I/Q Imbalance in OFDM Systems under Timing Ambiguity,” *IEEE Trans. on Vehicular Technology*, Vol.60, no.3, pp.1200-1205, Mar 2011.
- Gongpu Wang, Feifei Gao, Yik-Chung Wu and Chintha Tellambura, “Joint CFO and Channel Estimation for OFDM-Based Two-Way Relay Network,” *IEEE Trans. on Wireless Communications*, Vol.10, no.2, pp.456-465, Feb 2011.