Pilot Contamination: Is It Really A Stumbling Block For Massive MIMO?

Professor Sheng Chen

Southampton Wireless Group

Electronics and Computer Science

University of Southampton

Southampton SO17 1BJ, UK

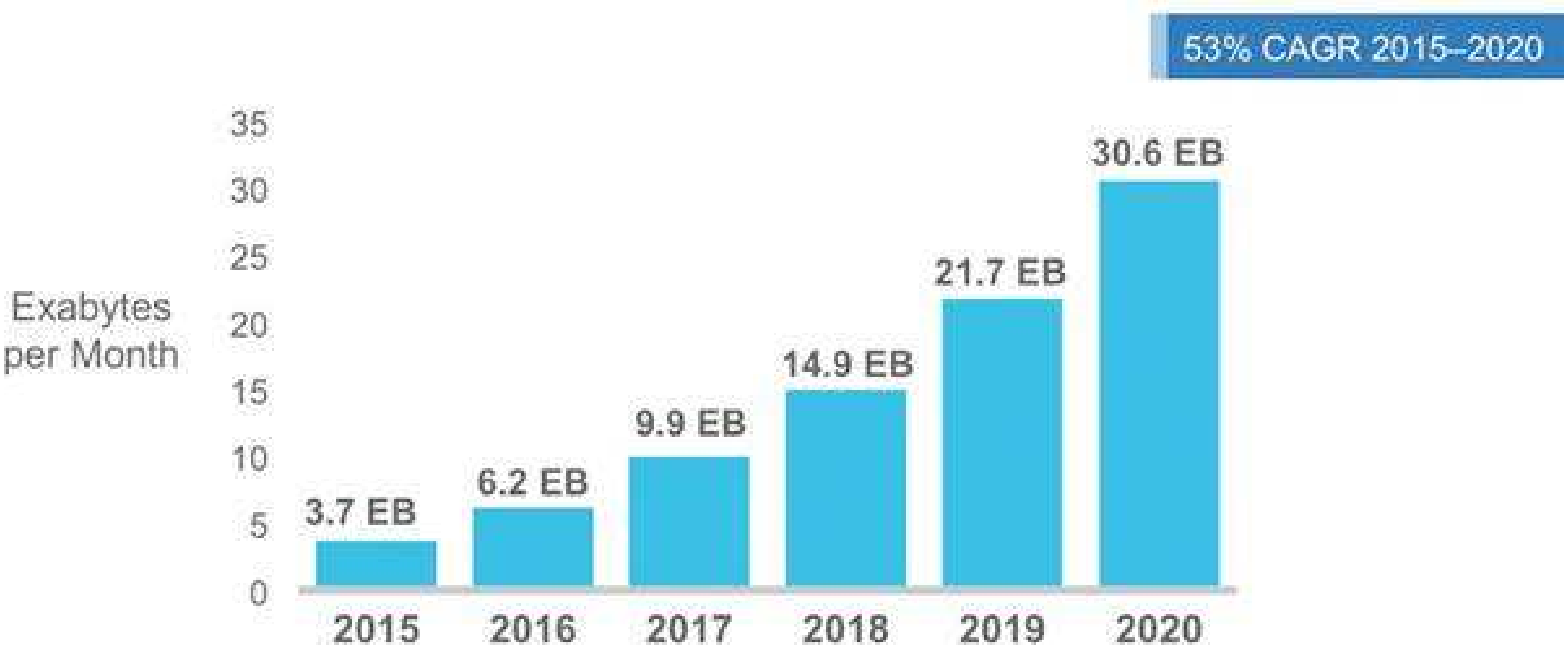
E-mail: sqc@ecs.soton.ac.uk

Joint work with: Miss Xinying Guo, Professor Lajos Hanzo

Talk at Toshiba Research Europe, Bristol, 26/09/2016

Where We Are

- Cisco Global Mobile Data Traffic Forecast Update, 2015–2016



- Several technologies, including MIMO and particularly **massive MIMO**, are promoted as enabling components for future mobile network to meet demand

How We Come to Where We Are

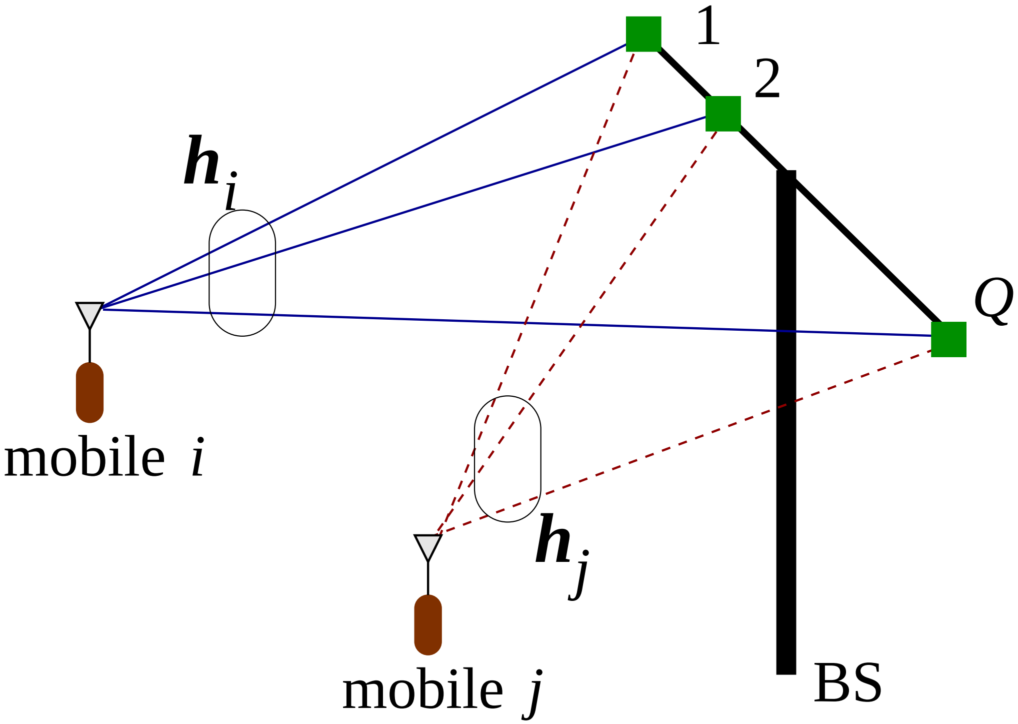
- 1G: mobile communication started – very very limited system capacity
- 2G: mobile communication spread – limited time/frequency resources, unable to meet increasing demand
- 3G: did not really created more ‘physical’ resources
  - but started fundamental paradigm shift: allow **non-orthogonal access** to support more users, but system becomes interference limited
- 4G: did not really created more ‘physical’ resources
  - To support high-rate applications, channel becomes extremely long, and hence OFDM and multi-carrier
- Meanwhile, at B3G, we started exploiting MIMO multiplexing gains, not for supporting higher rates, but for more users
  - Fundamental game change, **create** physically **new resources**

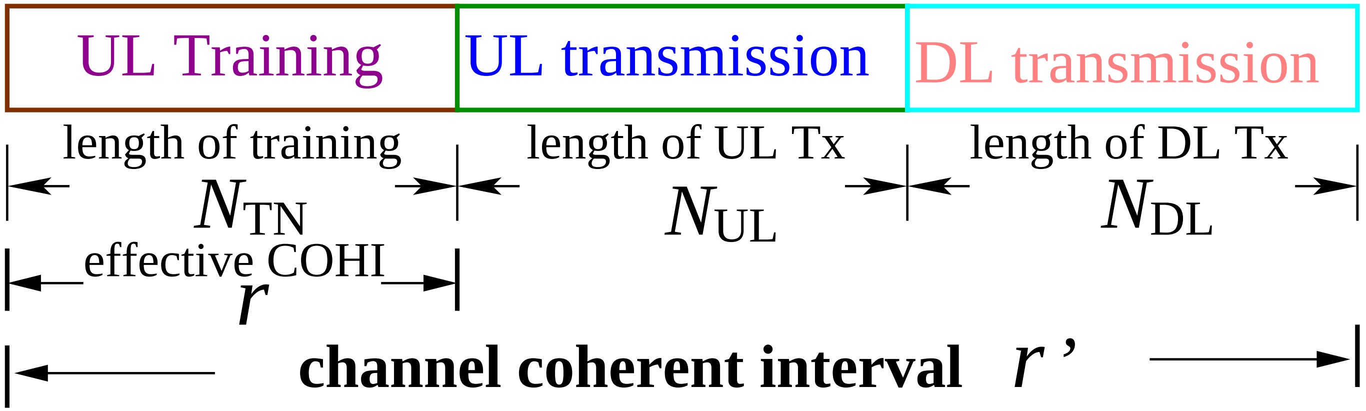
MIMO Wonderland

- MIMO for creating multiplexing gain and/or diversity gain is well understood
- At 2000, while everyone was busy on multi-carrier/OFDM for B3G, we were looking into MIMO for increasing user capacity
  - Exploit **user specific channel impulse responses**(CIRs) as **unique signatures** for distinguishing users
  - Unlike unique user spreading codes, we have no control on user specific CIRs, and base station with 4 antennas to support 4 or more users is tough
- This is a fundamental game change, potentially solve limited resource problem
  - To reach this MIMO wonderland requires accurate MIMO CSI, at Southampton, we carried out extensive research to offer effective solutions
  - Standard **linear** detection and precoding are offer **inadequate**, and we proposed nonlinear detection and precoding solutions
- Mobile providers do not like 'high complexity' **nonlinear** signal processing, want **have the cake and eat it**

Massive MIMO Wonderland

- Asymptotic spatial **orthogonality**
  - $Q \to \infty$: $\boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{h}_j = 0$
  - Potentially **infinite** spatial resource
  - Linear signal processing **sufficient**
- Reaching massive MIMO **promised** land
  - Need **accurate** MIMO CSI estimate
- Time division duplexing protocol
  - Uplink and downlink **reciprocal**





Pilot Contamination

- L cells, U users per cell, and BS antenna array having Q antenna elements
  - **Pilots** must be **orthogonal** so that least squares (LS) estimate has linear complexity
  - **Length** of pilot sequences must be **no longer** than **channel coherent time** (CCT) τ
  - With this maximum length, number of orthogonal pilots is τ
- Maximum number of users supported per cell is therefore $U = \tau$
  - With length of pilot sequences $\boldsymbol{\phi}_u \in \mathbb{C}^{\tau}$, $1 \le u \le U = \tau$, pilot set $\boldsymbol{\Phi} = \left[\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \cdots \boldsymbol{\phi}_U\right]^{\mathrm{T}} \in \mathbb{C}^{U\times\tau}$ with $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}} = \boldsymbol{I}_U$ must be **reused every cell**
- During UL training, received signal matrix of lth BS

$$\boldsymbol{Y}_l = \sum_{j=1}^{L} \boldsymbol{H}_{j,l}\boldsymbol{\Phi} + \boldsymbol{N}_l$$

  - $\boldsymbol{H}_{j,l} = \left[\boldsymbol{h}_{j,l,1} \ \boldsymbol{h}_{j,l,2} \cdots \boldsymbol{h}_{j,l,U}\right] \in \mathbb{C}^{Q\times U}$: channel matrix linking U users of jth cell to Q antennas of lth BS
- **Conventional** channel estimator (every BS estimates it channel matrix simultaneously)

$$\widehat{\boldsymbol{H}}_{l,l} = \boldsymbol{Y}_l\boldsymbol{\Phi}^{\mathrm{H}} = \boldsymbol{H}_{l,l} + \sum_{j\neq l} \boldsymbol{H}_{j,l} + \bar{\boldsymbol{N}}_l \quad \text{or} \quad \widehat{\boldsymbol{h}}_{l,l,u} = \boldsymbol{h}_{l,l,u} + \sum_{j\neq l} \boldsymbol{h}_{j,l,u} + \bar{\boldsymbol{n}}_{l,u}$$

Existing Solutions

- Pilot contamination becomes limiting factor, preventing us to reach massive MIMO promised land

- Extensive research leads to a range of existing state-of-the-arts in two categories, **none is effective and practical**

1. Schemes exploiting **user related features** with pilot assignment to combate pilot contamination
    - Acquisition of user related statistics is **costly** and requires considerable information exchange among cells
    - User related parameters are **time varying**, when they changes, the whole process has to be repeated

2. Schemes requiring no user related feature but at the expense of **sophisticated and long training** procedure to eliminate pilot contamination
    - Requiring **excessive long** CCT, unlikely to be met in practice
    - Achievable effective capacity is actually very **low**

Coordinated Channel Estimation

- H. Yin, D. Gesbert, M. Filippou, Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, 31(2), 264–273, 2013
  - Optimal Bayesian estimator, **do not suffer** from **pilot contamination**
  - So **problem solved? or is it?** until one examines what it requires
- This coordinated channel estimator requires the second-order statistics, i.e. channel covariance matrices, of all UL channels at every BS
  - Acquisition of such large amount of second-order statistics at BSs is extremely **time consuming**
  - Sharing them among BSs requires a **huge** amount of **back-haul** transmissions, too much coordinations needed among BSs
- This scheme is not practical, unless user related parameters are completely constant
  - Massive MIMO is primarily for increasing system capacity, but this estimator reduces effective throughput too much

Location-Aware Channel Estimation

- Z. Wang, C. Qian, L. Dai, J. Chen, C. Sun, S. Chen, "Location-based channel estimation and pilot assignment for massive MIMO systems," in *Proc. ICC 2015 Workshop*, June 8-12, 2015, 1264–1268
  - N-point DFT based **post-processing** on conventional channel estimate
  - For users with same pilot but non-overlapped AOAs, pilot contamination removed
  - Training duration is the **same** as conventional simultaneous training
  - Modest increase in complexity for N-point DFT ($N \geq Q$ and Q is array size)

- Use location-aware pilot assignment to ensure users with same pilot having **non-overlapped AOAs** as much as possible

- Requirement: **AOAs** of users
  - With aid of GPS or other positioning techniques, information of users' AOAs is much easier to obtain, compared with channel covariance matrices

- Location-aware channel estimation is currently most practical scheme available
  - Generally, **can only mitigate** pilot contamination

Pilot Contamination Elimination Schemes

1. J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, L. Hanzo, "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems," *IEEE J. Sel. Topics Signal Process.*, 8(5), 759–772, 2014
    - Consist of an amalgam of $(L+3)$ DL/UL training phases for L-cell system
    - Completely **eliminate pilot contamination**
    - Training duration is $(L+3)$ **times** of conventional simultaneous training

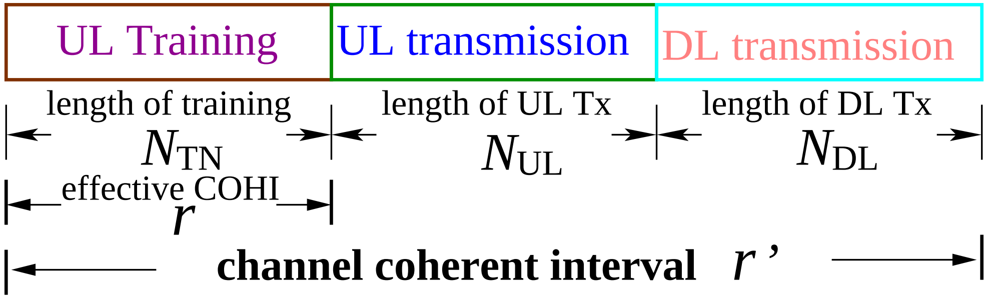
2. T.X. Vu, T.A. Vu, T.Q.S. Quek, "Successive pilot contamination elimination in multiantenna multicell networks," *IEEE Wireless Commun. Let.*, 3(6), 617–620, 2014
    - Consist of $(L+1)$ training phases with signal cancellation operations
    - Completely **eliminate pilot contamination**
    - Training duration is $(L+1)$ **times** of conventional simultaneous training
    - Signal cancellations **amplify noise** and reduce estimation accuracy

- Both require **excessively long** channel coherent time, unlikely to be met in practice

Implications of Training Duration

- For OFDM, define COHI as ratio of CCT t_{CCT} over OFDM symbol duration T_{OFDM}

$$r' = \left\lfloor \frac{t_{\mathrm{CCT}}}{T_{\mathrm{OFDM}}} \right\rfloor ,$$



- Training duration must satisfy $N_{\mathrm{TN}} \le r$, with effective COHI for performing channel estimation

$$r = r' - \left(N_{\mathrm{UL}} + N_{\mathrm{DL}}\right),$$

N_{UL} and N_{DL}: numbers of OFDM symbols transmitted during UL and DL data transmissions

- Let C_{UL} and C_{DL} be ideal UL and DL sum-rates, without taking into account training overhead
- **Effective** UL and DL **sum-rates** C_{UL}^{ef} and C_{DL}^{ef} are obtained respectively as

$$C_{\mathrm{UL}}^{ef} = \frac{N_{\mathrm{UL}}}{\frac{1}{2}N_{\mathrm{TN}} + N_{\mathrm{UL}}} C_{\mathrm{UL}}$$

$$C_{\mathrm{DL}}^{ef} = \frac{N_{\mathrm{DL}}}{\frac{1}{2}N_{\mathrm{TN}} + N_{\mathrm{DL}}} C_{\mathrm{DL}}$$

What We/You Want

Given network with: number of cells L, number of antennas at each BS Q, maximum number of users supported per cell U, number of subcarriers N, maximum delay spread or length of CIRs K, and effective COHI r

- Design an optimal scheme at network planning stage capable of eliminating or significantly reducing pilot contamination
  - **With the minimum training duration N_{TN}**
  - **Depend only on above network parameters**
  - **Does not depend on any user related features**
- The design remains unchanged during entire network operational life time
- We have designed such an optimal scheme, and it is extremely simple

X. Guo, S. Chen, J. Zhang, X. Mu, L. Hanzo, "Optimal pilot design for pilot contamination elimination/reduction in large-scale multiple-antenna aided OFDM systems," ***IEEE Trans. Wireless Commun.*****, to appear, 2016**

Our Time-Domain Channel Estimation

- CIR linking uth user of cell l to qth antenna of cell l':

$$\underline{\mathbf{G}}_{l,l',q}^{u} = \left[G_{l,l',q}^{u}[1]\ G_{l,l',q}^{u}[2] \cdots G_{l,l',q}^{u}[K]\right]^{\mathrm{T}} \in \mathbb{C}^{K}$$

  - FDCHTF vector $\underline{\mathbf{H}}_{l,l',q}^{u} = \mathbf{F}\underline{\mathbf{G}}_{l,l',q}^{u} \in \mathbb{C}^{N}$ with FFT matrix $\mathbf{F} \in \mathbb{C}^{N\times K}$
- Signal vector $\underline{\mathbf{Y}}_{l',q} \in \mathbb{C}^{N}$ received by qth antenna of l'th BS and collected over N subcarriers:

$$\underline{\mathbf{Y}}_{l',q} = \sqrt{p_r}\sum_{u'=1}^{U} \mathbf{X}_{l'}^{u'}\mathbf{F}\underline{\mathbf{G}}_{l',l',q}^{u'} + \sqrt{p_r}\sum_{l=1,l\neq l'}^{L}\sum_{u=1}^{U} \mathbf{X}_{l}^{u}\mathbf{F}\underline{\mathbf{G}}_{l,l',q}^{u} + \underline{\mathbf{W}}_{l',q}$$

  - $\mathbf{X}_{l}^{u} = \mathrm{diag}\{X_{l}^{u}[1], X_{l}^{u}[2], \cdots, X_{l}^{u}[N]\}$: frequency domain **pilot symbol** of user u in lth cell, with unity power; p_r: average user power
- A **difference** between our approach and existing schemes
  - We consider signal collected over **all N OFDM subcarriers for an individual BS antenna**
  - All existing works consider signal **over all Q target BS's antennas for an individual subcarrier**
- Our approach to UL training has a significant **advantage**
  - Our approach for simultaneous UL training requires effective COHI $r \geq 1$
  - Conventional simultaneous UL training requires effective COHI $r \geq U$
  - Pilot contamination elimination scheme 1 requires effective COHI $r \geq (L+3)U$
  - Pilot contamination elimination scheme 2 requires effective COHI $r \geq (L+1)U$

Optimal Frequency-Domain Pilot Design

- Design a FD PS matrix set for all LU users in all cells according to (Li, 2002)

$$\begin{aligned} \mathbf{P} &= \left\{\mathbf{X}_l^u, 1 \le u \le U, 1 \le l \le L\right\} = \left\{\mathbf{P}[i], 1 \le i \le LU\right\} \\ &= \left\{\mathbf{X}_1^1, \mathbf{X}_2^1, \cdots, \mathbf{X}_L^1; \mathbf{X}_1^2, \mathbf{X}_2^2, \cdots, \mathbf{X}_L^2; \cdots; \mathbf{X}_1^U, \mathbf{X}_2^U, \cdots, \mathbf{X}_L^U\right\} \end{aligned}$$

  - which contains LU diagonal PS matrices of

$$\mathbf{P}[i] = \mathbf{P}[(u-1)L+l] = \mathbf{X}_l^u, \; i = (u-1)L+l, \; 1 \le u \le U, 1 \le l \le L$$

ith element of $\mathbf{P}$ is generated from reference $\mathbf{P}[1] = \mathbf{X}_1^1$ according to

$$\mathbf{P}[i] = \mathbf{\Phi}[i]\mathbf{P}[1], \; 1 \le i \le LU$$

with

$$\mathbf{\Phi}[i] = \text{diag}\left\{e^{\text{j}2\pi\frac{(i-1)\zeta 0}{N}}, e^{\text{j}2\pi\frac{(i-1)\zeta 1}{N}}, \cdots, e^{\text{j}2\pi\frac{(i-1)\zeta(N-1)}{N}}\right\}, \; 1 \le i \le LU$$

  - If $\zeta = \lfloor\frac{N}{LU}\rfloor \ge K$, then all PS matrices $\mathbf{P}[i]$, $1 \le i \le LU$, are **orthogonal**

- **No pilot contamination** in simultaneous UL training, and MSE of channel estimate attains

$$\text{CRLB} = \frac{K\sigma_w^2}{Np_r}, \;\; \text{with } \sigma_w^2 \text{ channel noise power}$$

- Y. Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, 1(1), 67–75, 2002

Sufficient/Insufficient Subcarrier Resource

- With **sufficient** subcarrier resource, namely, $N \geq KLU$, we can always design **orthogonal** PS matrices for all LU users
  - **Simultaneous** UL training **does not suffer** from pilot contamination, and only requires **minimum** effective COHI $r = 1$
- In practice for CIR having large path K and/or large number of users per cell U and/or large number of cells L, the available subcarrier resource becomes **insufficient**, i.e. $N < KLU$
  - **Not all** $\mathbf{P}[i]$, $1 \leq i \leq LU$, are orthogonal, and simultaneous UL training suffers from **some pilot contamination**
- Depending on system parameters

$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor, \; f = \left\lceil \frac{K}{\zeta} \right\rceil, \; n_u = \left\lfloor \frac{LU}{f} \right\rfloor \quad R = \text{Rem}\left\{ \frac{LU}{f} \right\}$$

  we can always divide LU users into f or $f+1$ **groups**
  - Each group contains no more than n_u users
  - PS matrices associated with users of every group are **orthogonal**
  - $2 \leq f < L$ and $n_u > f$
- Hence we can always implement f or $f+1$ phases of UL training, which **completely eliminates** pilot contamination
  - Only require effective channel coherent interval $r \geq f$ or $f+1$

Optimal Grouping

1. Optimally grouping LU users into f groups given $R = 0$

$LU = n_u f + R$, $R = 0$, $i = (u-1)L + l$, $1 \leq i \leq LU, 1 \leq u \leq U, 1 \leq l \leq L$.				
Group	User indexes i in each group			
1	1	$f+1$	$\cdots$	$LU-(f-1)$
2	2	$f+2$	$\cdots$	$LU-(f-2)$
$\vdots$	$\vdots$			
$f-1$	$f-1$	$2f-1$	$\cdots$	$LU-1$
f	f	$2f$	$\cdots$	LU

2. Optimally grouping LU users into f groups given $R \neq 0$ and $n_u f = LU - R \leq \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$

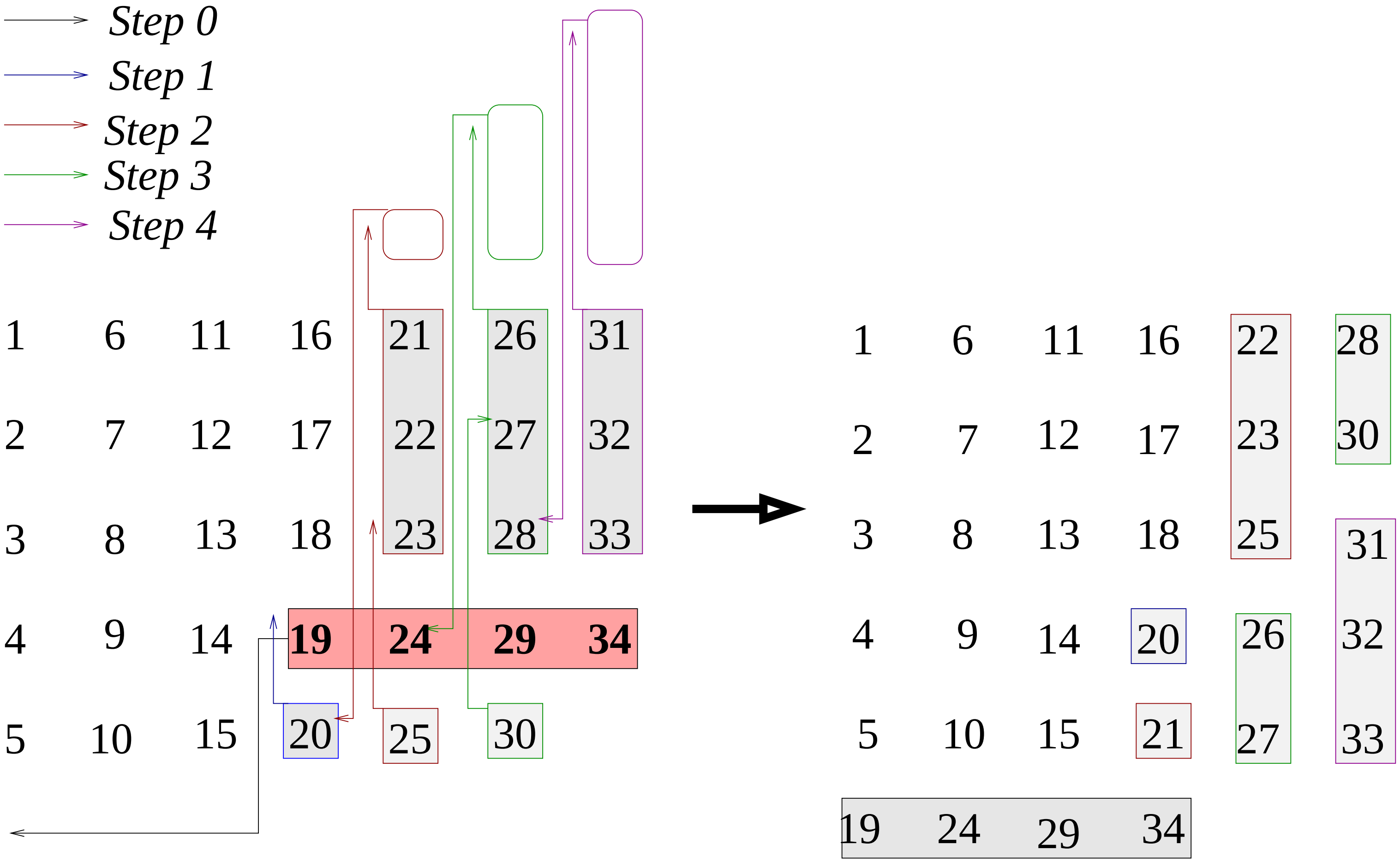
$LU = n_u f + R$, $R \in \{1, 2, \cdots, f-1\}$, $i = (u-1)L + l$, $1 \leq i \leq LU, 1 \leq u \leq U, 1 \leq l \leq L$.							
Group	User indexes i in each group						
1	1	$f+1$	$\cdots$	$(n_u-(R-1))f+1$	$\cdots$	$(n_u-1)f+1$	$n_u f+1$
2	2	$f+2$	$\cdots$	$(n_u-(R-1))f+2$	$\cdots$	$(n_u-1)f+2$	$n_u f+2$
$\vdots$	$\vdots$						
$R-1$	$R-1$	$f+R-1$	$\cdots$	$(n_u-(R-1))f+R-1$	$\cdots$	$(n_u-1)f+R-1$	$n_u f+R-1$
R	R	$f+R$	$\cdots$	$\boldsymbol{(n_u-(R-1))f+R}$	$\boldsymbol{\cdots}$	$\boldsymbol{(n_u-1)f+R}$	$\boldsymbol{n_u f+R}$
$R+1$	$R+1$	$f+R+1$	$\cdots$	$(n_u-(R-1))f+R+1$	$\cdots$	$(n_u-1)f+R+1$	
$\vdots$	$\vdots$						
f	f	$2f$	$\cdots$	$(n_u-(R-2))f$	$\cdots$	$n_u f$	

3. Optimally grouping LU users into $f+1$ groups given $R \neq 0$ and $n_u f = LU - R > \left\lfloor \frac{N-K}{\lfloor \frac{N}{LU} \rfloor} \right\rfloor$
   - A simple procedure to rearrange 2. into $f+1$ groups

Example of $f+1$ Optimal Grouping

- $N = 206$, $LU = 34$ and $K = 29$. Clearly, $N < KLU$

$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor = 6,\ f = \left\lceil \frac{K}{\zeta} \right\rceil = 5,\ n_u = \left\lfloor \frac{LU}{f} \right\rfloor = 6,\ R = \text{Rem}\left\{ \frac{LU}{f} \right\} = 4$$



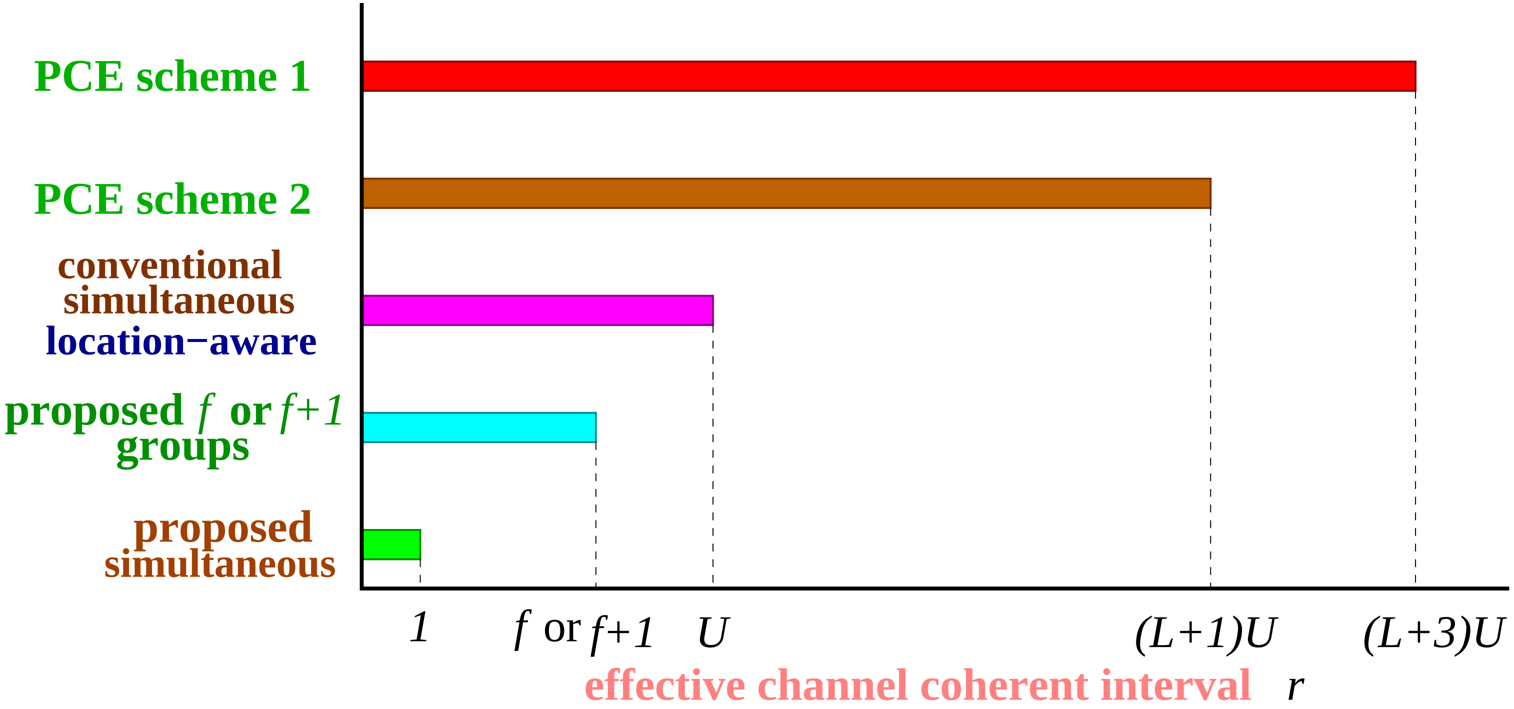
Step 0
Step 1
Step 2
Step 3
Step 4

1 6 11 16 21 26 31
2 7 12 17 22 27 32
3 8 13 18 23 28 33
4 9 14 **19** **24** **29** **34**
5 10 15 20 25 30

1 6 11 16 22 28
2 7 12 17 23 30
3 8 13 18 25 31
4 9 14 20 26 32
5 10 15 21 27 33
19 24 29 34

Comparison

- Implementation requirement in terms of effective channel coherent interval r:



- If subcarrier resource is sufficient, proposed simultaneous completely **eliminate** pilot contamination with minimum required $r = 1$
- If r meets their individual requirements
  - Proposed (f or $f + 1$ groups), PCE schemes 1 and 2: completely **eliminate** PC
  - Location-aware: significantly **reduces** pilot contamination
- PCE schemes 1 and 2 can no longer be implemented for $r < (L + 3)U$ or $(L + 1)U$, but proposed scheme can still be implemented for any $1 \leq r < f$ with significantly reduced PC

Simulation System Setup

Number of cells L	7
Radius of each cell	$1000\,\text{m}$
Number of MSs per-cell U	8
Number of antennas at each BS Q	100
Average transmit power at each MS p_r	$0\,\text{dB}$
Average transmit power at each BS p_f	$10\,\text{dB}$
Path loss exponent	3
Mean of path AoAs θ	90°
Standard deviation of path AoAs σ_{AoA}	90°
Antenna spacing D	$\frac{\lambda}{2}$
Length of CIRs K	54
Number of subcarriers N	1024

- **Insufficient** subcarrier resources as $N < KLU$, and optimal grouping is $f + 1 = 4$ groups as

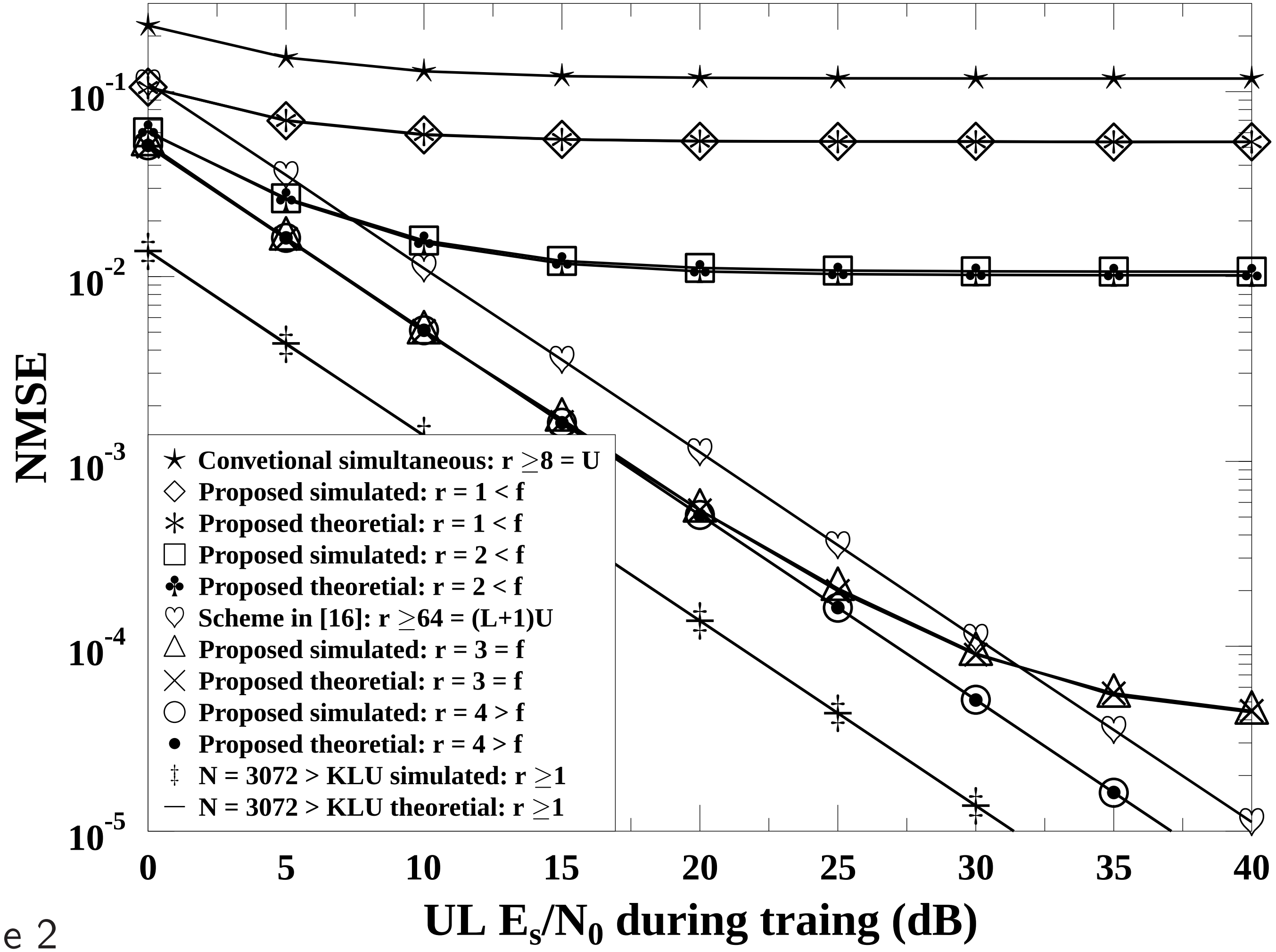
$$\zeta = \left\lfloor \frac{N}{LU} \right\rfloor = 18, \; f = \left\lceil \frac{K}{\zeta} \right\rceil = 3, \; n_u = \left\lfloor \frac{LU}{f} \right\rfloor = 18,$$

$$R = \text{Rem}\left\{ \frac{LU}{f} \right\} = 2, \; n_u f = 54 > \left\lfloor \frac{N-K}{\zeta} \right\rfloor = 53$$

- We also show the system with **sufficient** subcarrier resources of $N = 3072 > KLU$

Estimation Result Comparison

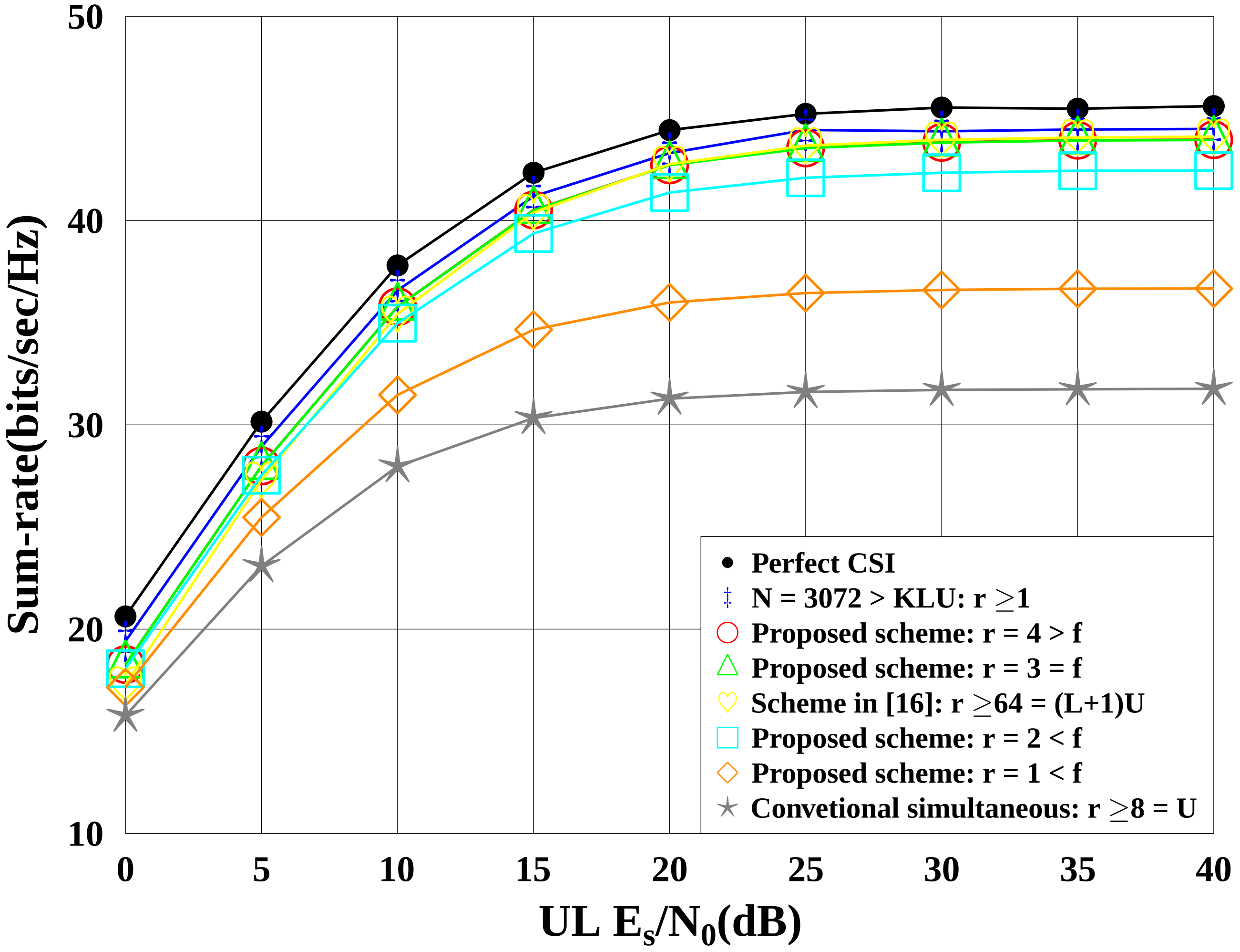
Normalized MSE of channel estimate (averaged over 100 channel realizations) $\text{NMSE} = \dfrac{\sum_{l=1}^{L}\sum_{u=1}^{U}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left|\widehat{H}_{l,l,q}^{u}[n]-H_{l,l,q}^{u}[n]\right|^2}{\sum_{l=1}^{L}\sum_{u=1}^{U}\sum_{q=1}^{Q}\sum_{n=1}^{N}\left|H_{l,l,q}^{u}[n]\right|^2}$



[16]: PCE scheme 2

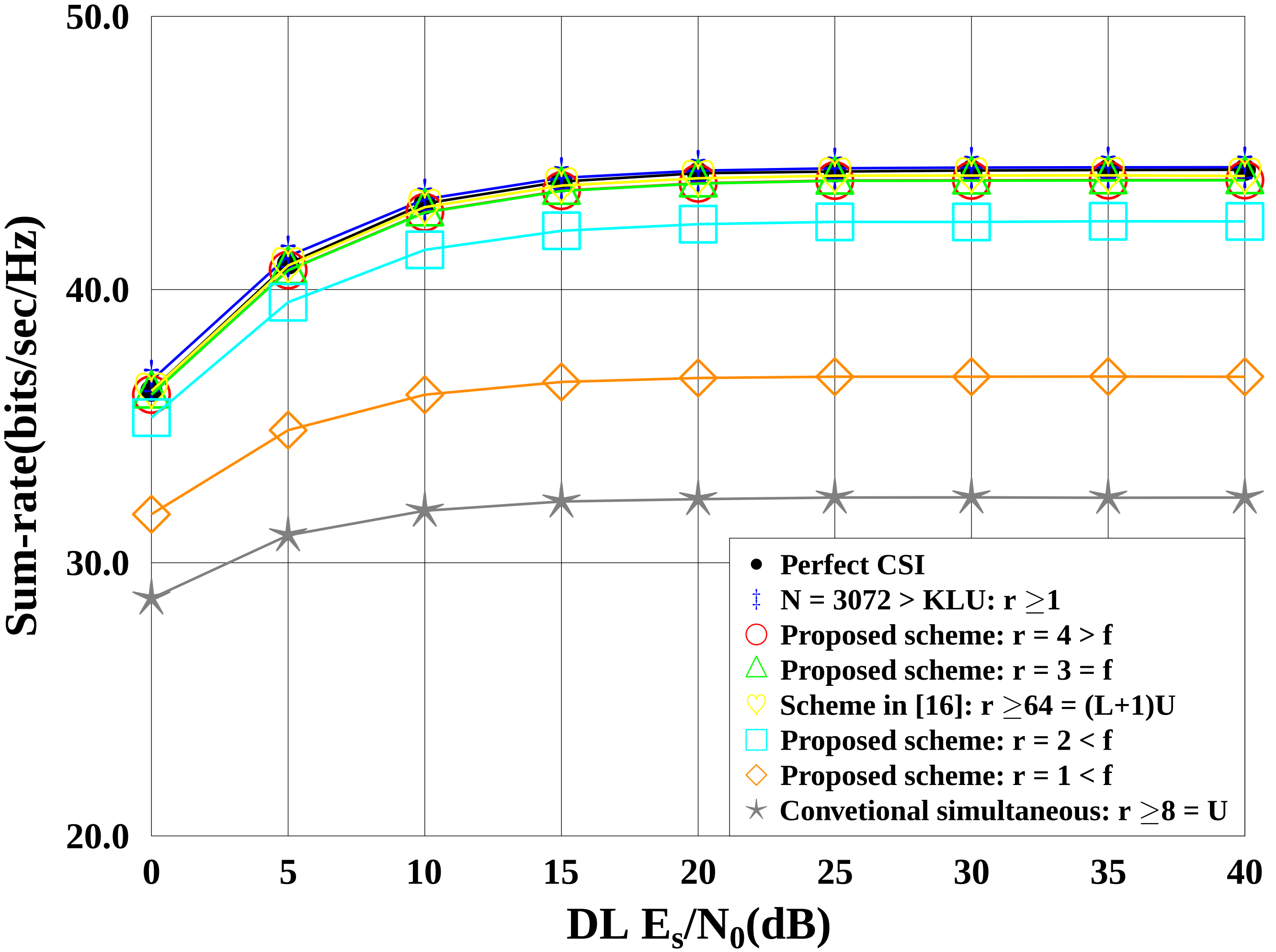
Ideal Uplink Beamforming Performance

Ideal per-cell UL sum-rate performance (without considering impact of training duration) as functions of UL system's SNR with UL training SNR equal to UL system's SNR, using maximum-ratio combining



Ideal Downlink Precoding Performance

Ideal per-cell DL sum-rate performance (without considering impact of training duration) as functions of DL system's SNR where UL training SNR is fixed to $E_s/N_0 = 20\,\text{dB}$, using zero-forcing precoding

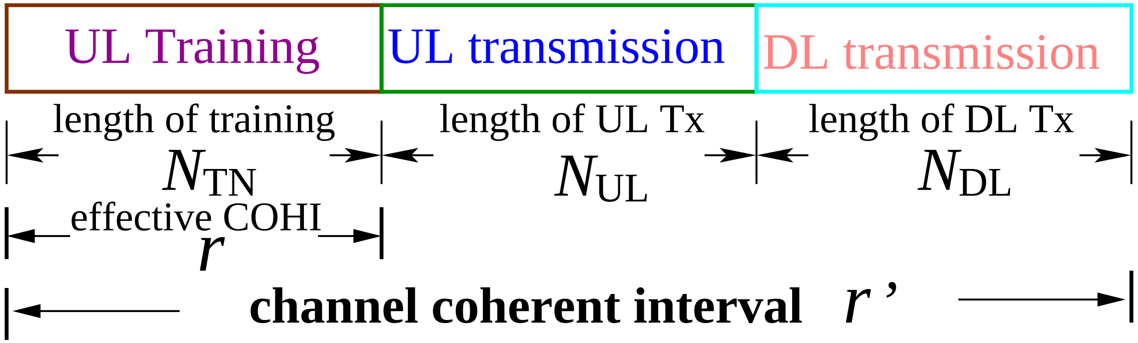


Effective Sum-Rate Performance

- We have to consider very slow fading system with COHI $r' = 84$ OFDM symbols so that PCS scheme in [16] can be implemented with $N_{\text{TN}} = 64$ and $N_{\text{UL}} = N_{\text{DL}} = 10$

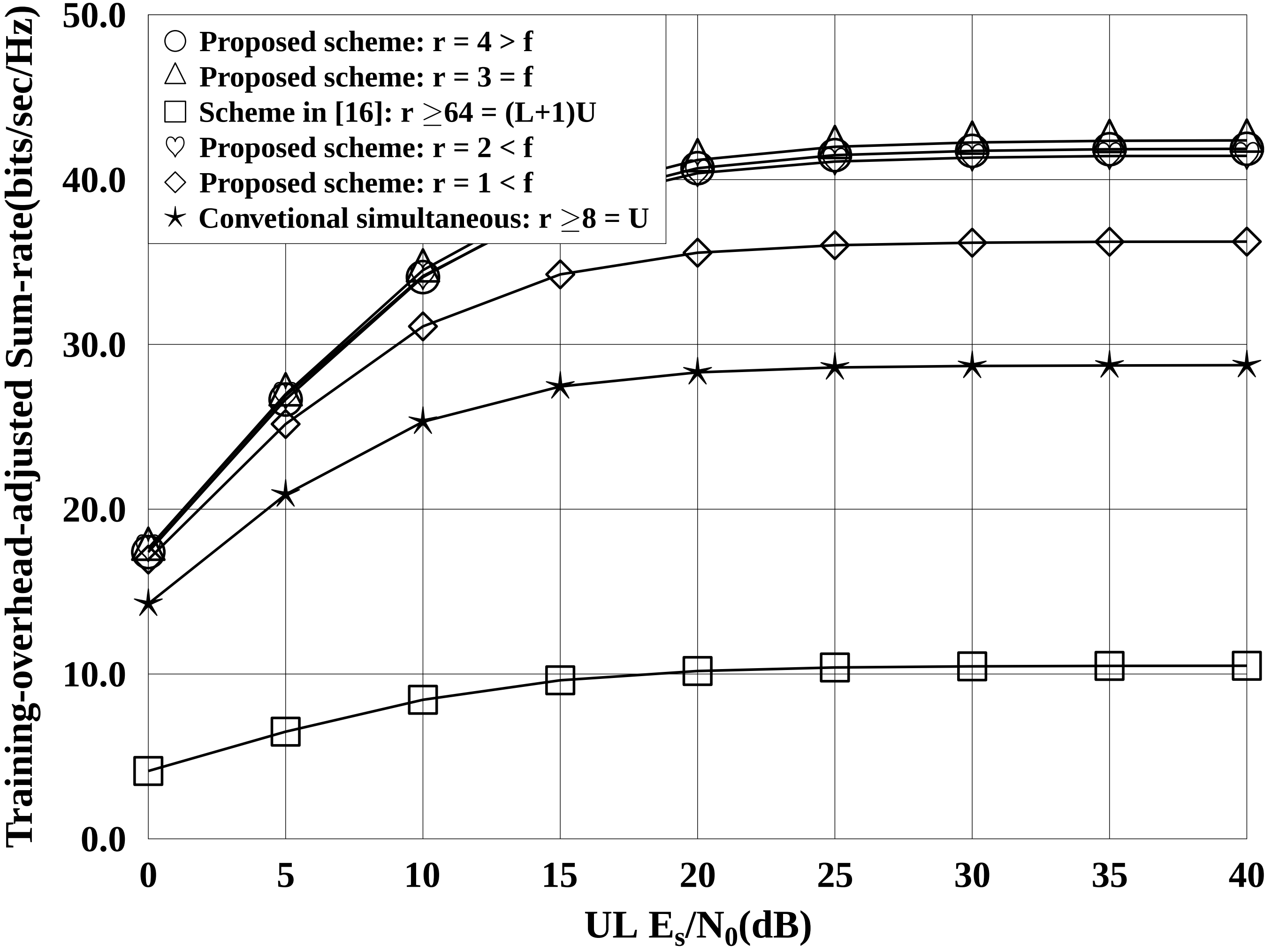
- Conventional simultaneous UL training scheme requires $N_{\text{TN}} = 8$, and can support UL and DL transmissions with $N_{\text{UL}} = N_{\text{DL}} = 38$

- Our proposed scheme (insufficient subcarrier resources of $N < KLU$)
  - 4-group implementation (optimal and no PC): $N_{\text{TN}} = 4$ and $N_{\text{UL}} = N_{\text{DL}} = 40$
  - 3-group implementation: $N_{\text{TN}} = 3$ and $N_{\text{UL}} = N_{\text{DL}} = 40.5$
  - 2-group implementation: $N_{\text{TN}} = 2$ and $N_{\text{UL}} = N_{\text{DL}} = 41$
  - 1-group implementation (simultaneous): $N_{\text{TN}} = 1$ and $N_{\text{UL}} = N_{\text{DL}} = 41.5$



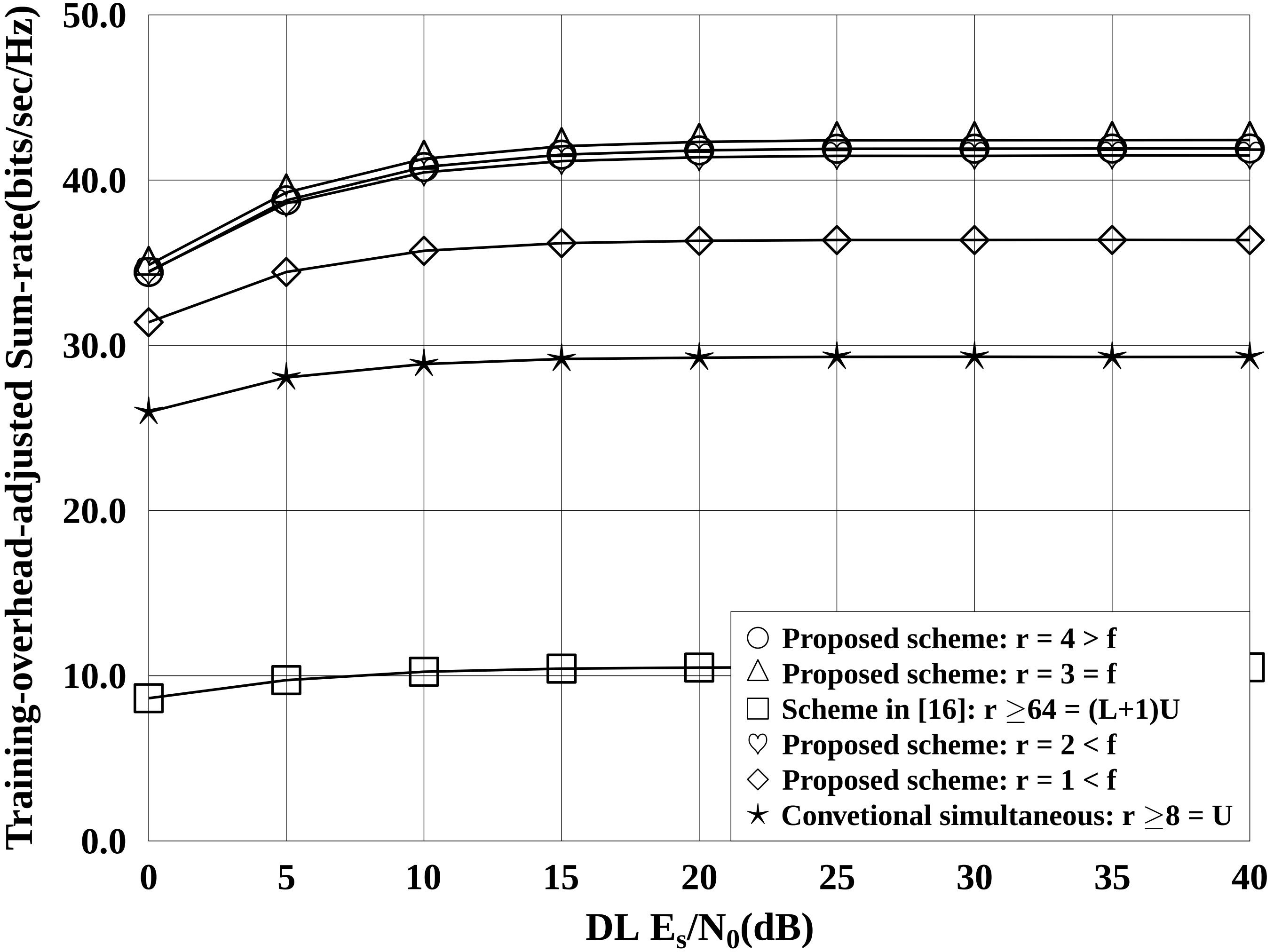
Effective Uplink Performance

Effective per-cell UL sum-rate performance (considering impact of training duration) as functions of UL system's SNR with UL training SNR equal to UL system's SNR, using maximum-ratio combining



Effective Downlink Precoding Performance

Effective per-cell DL sum-rate performance (considering impact of training duration) as functions of DL system's SNR where UL training SNR is fixed to $E_s/N_0 = 20\,$dB, using zero-forcing precoding



Summary

- Given number of subcarriers N, maximum length of CIRs K, maximum number of users supported per cell U, number of cells L:
  - Optimal set of pilot symbol matrices for all LU users are obtained straightforwardly, and they remain **fixed**
  - Optimal grouping for pilot contamination-free UL training is **determined**
  - Design remains valid for entire network operating life time, and can be implemented even effective CHOI only lasts one OFDM symbol duration
- Our scheme achieves **PC-free** UL training with **lowest** training overhead
  - **No user related features or statistics needed**
  - **No information exchange between cells needed**
  - **Nothing needs changed**
- Sound 'too good to be true'? – **It is true**
  - Just to prove the **best** solution is also the **simplest** one

Conclusions

- Massive MIMO has been recognized as a promising and key component for future mobile network

- However, pilot contamination has been a **stumbling block** preventing us from reaching massive MIMO **promised** land
  - Existing PC elimination/reduction solutions either require too much or demanding excessively long training duration, which cannot be met in practice
  - These so-called state-of-the-arts may actually be less effective than conventional simultaneous UL training

- With our proposed simple yet effective approach
  - Pilot contamination problem **is solved** for OFDM based massive MIMO systems

- Much works remain to be done in order to bring massive MIMO from concept to protocol