Massive Random Access with Massive MIMO: Sparse Activity Detection

Wei Yu

Joint Work with Zhilin Chen, Foad Sohrabi, Liang Liu

University of Toronto

2021

Introduction and Outline

Massive device connectivity is a key requirement for 5G cellular networks

- Machine-type (M2M) communications, Internet of Things (IoT), Sensors...
- Sporadic traffic with low latency requirement
- Large number of devices but only a few are active at a time

This talk is about how to design such a network:

- Sparsity device activity detection algorithms
- Massive connectivity with massive MIMO
- Scheduling and feedback in massive random access

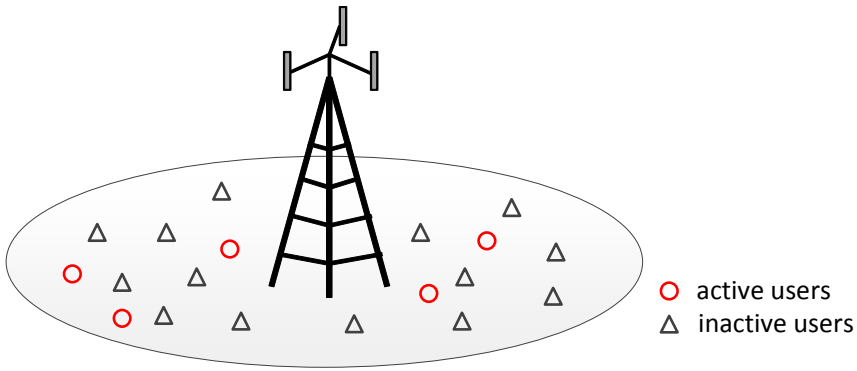
Main Messages

To support massive connectivity:

- The use of non-orthogonal pilots is inevitable.
- Compressed sensing techniques are indispensible for device detection.
- Massive MIMO can significantly enhance device activity detection.
- Channel estimation is the main bottleneck.
- Cooperative detection across multiple cells further improves performance.
- Scheduing and feedback are superior to uncoordinated random access.

Massive Random Access

- Cellular system with N users, but only K of which are active.



- BS needs to detect which users are active, then their messages.
- User activity pattern carries information. [Chen-Guo'14]

$$R + H(A) \leq I(X;Y) \tag{1}$$

- We also need to take the cost of channel estimation into account.

Fundamental Limit of Massive Random Access

For a massive device communications scenario $Y = HAX + Z$, the achievable sum rate of data transmission across all the users is approximately bounded by

$$R \lesssim I(X; Y|HA) - H(A) - I(HA; Y|X). \tag{2}$$

Interpretation:

- $I(X; Y|HA)$: Transmission rate with known channels and activity pattern;
- $H(A)$: Information content of device activity pattern;
- $I(HA; Y|X)$: Channel estimation and user activity detection.

Why? We see that $R + H(A) \leq I(X; Y)$.

$$\begin{aligned} I(HA, X; Y) &= I(X; Y) + I(HA; Y|X) &(3)\\ &= I(HA; Y) + I(X; Y|HA) &(4)\end{aligned}$$

Note that the $I(HA; Y)$ term is negligible.

Cost of User Activity Detection

Traditional MIMO system [Zheng-Tse'02, Lozano-Heath-Andrew'12]:

$$R \lesssim I(X; Y|H) - I(H; Y|X). \tag{5}$$

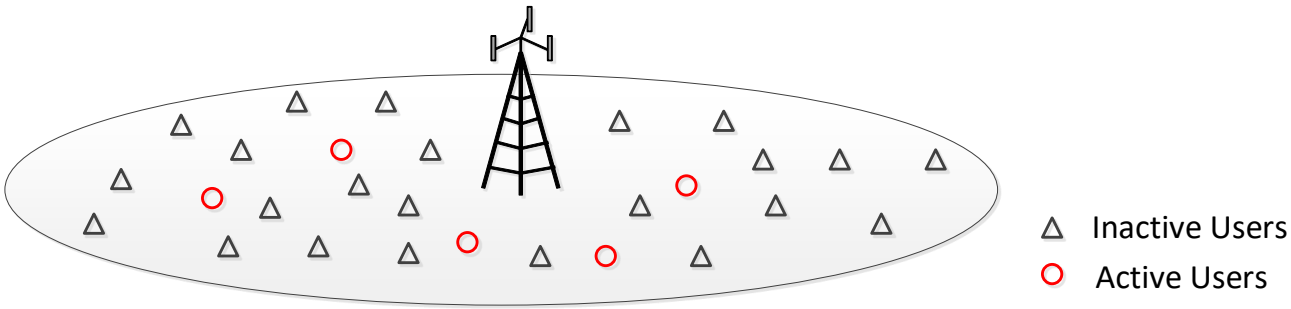
Massive connectivity system:

$$R \lesssim I(X; Y|HA) - H(A) - I(HA; Y|X). \tag{6}$$

where the cost of user activity detection is:

$$H(A) = Nh\left(\frac{K}{N}\right) \approx \log \binom{N}{K} \approx K \log(N/K). \tag{7}$$

User Activity Detection and Channel Estimation via Pilots



- BS equipped with M antennas
- N single-antenna devices, K of which are active at a time
- Each device is associated with a length-L unique signature sequence $\boldsymbol{s}_n$
- Channel $\boldsymbol{h}_n$ of user n is assumed to be fixed during the L symbols.
- For single-cell system, received signal $Y \in \mathbb{C}^{L\times M}$ at the BS is

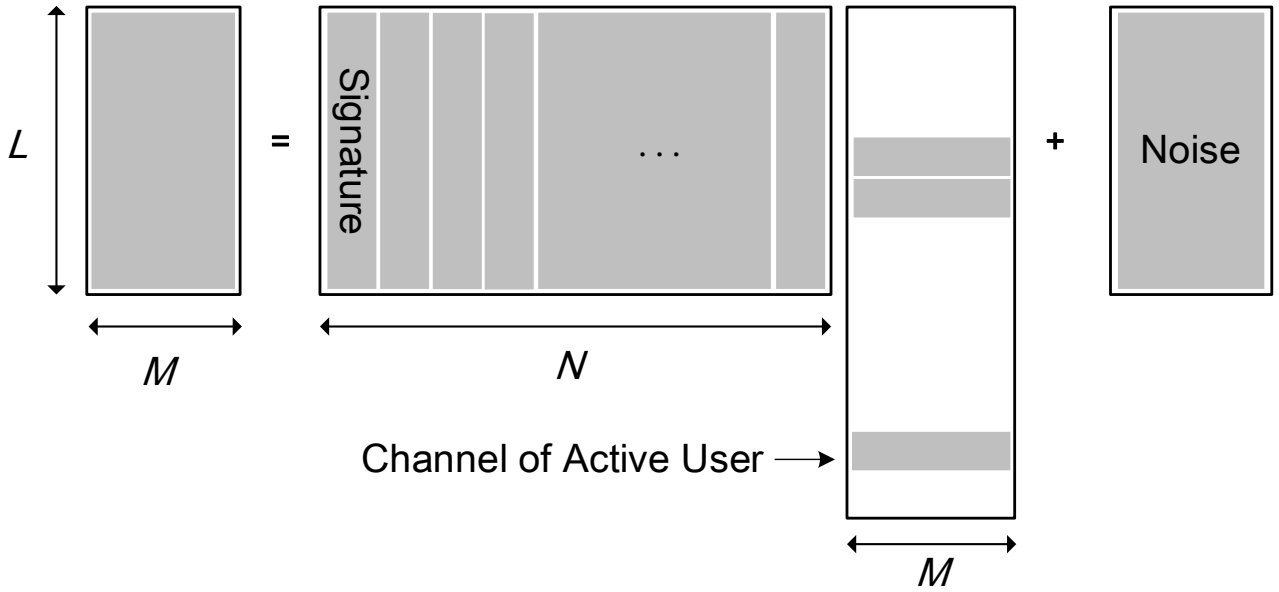
$$\boldsymbol{Y} = \sum_{n=1}^{N} \alpha_n \boldsymbol{s}_n \boldsymbol{h}_n^T + \boldsymbol{Z} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{Z}, \tag{8}$$

where

- $\alpha_n \in \{1, 0\}$ activity indicator; $\boldsymbol{Z} \in \mathbb{C}^{L\times M}$ Gaussian noise with variance σ^2
- $\boldsymbol{S} \triangleq [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N] \in \mathbb{C}^{L\times N}$; $\boldsymbol{X} \triangleq [\alpha_1 \boldsymbol{h}_1, \cdots, \alpha_N \boldsymbol{h}_N]^T \in \mathbb{C}^{N\times M}$

User Activity Detection via Compressed Sensing

Aim to identify the K non-zero rows of $\boldsymbol{X}$ from $\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{X} + \boldsymbol{Z}$.



- Multiple measurement vector (MMV) problem in compressed sensing
  - Columns of $\boldsymbol{X}$ share the same sparsity pattern, i.e., row sparsity
- Efficiently solved by the approximate message passing (AMP) algorithm

Practical Detector Design

Device Identification via Non-Orthogonal Pilots:

- Due to large number of potential devices N, orthogonal pilot is not feasible.
- Natural choice of pilot sequences: i.i.d. Gaussian signature
- Approximate Message Passing (AMP) [Donoho-Maleki-Montanari'09]

Prior work on compressed sensing for massive connectivity:

- Without channel estimation [Fletcher-Rangan-Goyal'09, Zhang-Luo-Guo'13]
- Joint user activity detection and channel estimation: Orthogonal matching pursuit [Schepker-Bockelmann-Dekorsy'13, Wunder-Jung-Ramadan'15, Wunder-Boche-Strohmer-Jung'15], Baysian [Xu-Rao-Lau'15]
- AMP is used for device detection in [Hannak-Mayer-Jung-Matz-Goertz'15].

Single-Antenna Case

Single-Antenna Case

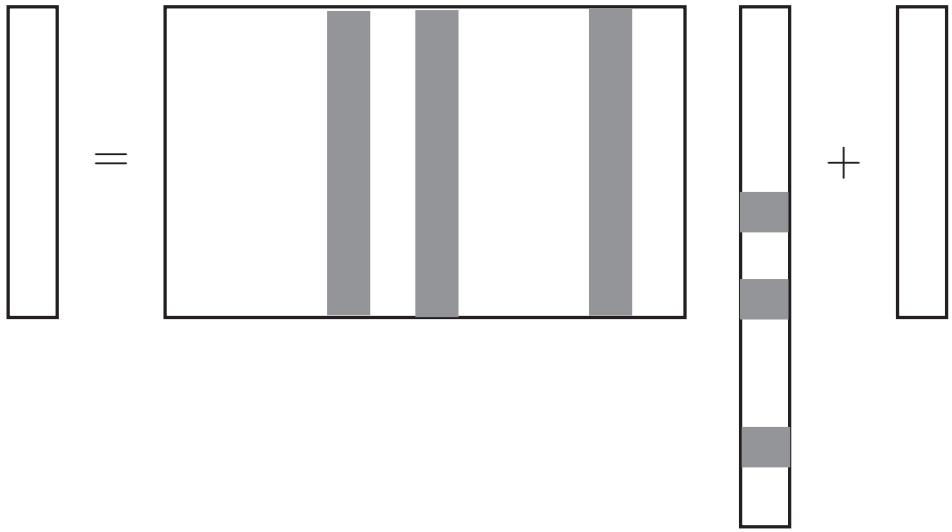
A single-antenna BS, N devices randomly located in a cell of radius R,

$$\boldsymbol{y} = \sum_{n=1}^{N} h_n \alpha_n \boldsymbol{s}_n + \boldsymbol{w} \triangleq \boldsymbol{S}\boldsymbol{x} + \boldsymbol{z} \tag{9}$$

- $h_n \in \mathbb{C}$: channel coefficient between user n and BS, including path-loss fading, shadowing and Rayleigh fading static within each block;
- $\alpha_n \in \{1, 0\}$: indicating whether user n is active
- $\boldsymbol{x} \triangleq [h_1\alpha_1, h_2\alpha_2, \cdots, h_N\alpha_N]^T \in \mathbb{C}^{N\times 1}$
- $\boldsymbol{s}_n \in \mathbb{C}^{L\times 1}$: signature sequence of user n generated as i.i.d. $\mathcal{CN}(0, 1/L)$
- $\boldsymbol{S} \triangleq [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_N]^T \in \mathbb{C}^{L\times N}$
- $\boldsymbol{z} \in \mathbb{C}^{L\times 1}$: effective noise following i.i.d. $\mathcal{CN}(0, \sigma^2)$

Sparse Recovery Problem

Identify the columns that correspond to non-zero elements in $\boldsymbol{x}$ via



LASSO formulation:

$$\hat{\boldsymbol{x}} = \arg\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \tag{10}$$

CoSaMP is computationally complex: Not scalable at $N = 10^5$.

Soft Thresholding Function

Consider a special case of a single measurement of a scalar, LASSO is

$$\hat{x} = \arg\min \frac{1}{2}|y - x|_2^2 + \lambda |x|_1 \tag{11}$$

The solution is explicitly given by

$$\hat{x} = \eta(y; \lambda), \tag{12}$$

where η is a soft thresholding function as

$$\eta(y; \theta) = \begin{cases} y - \theta, & y > \theta \\ 0, & -\theta \le y \le \theta \\ y + \theta, & y < -\theta \end{cases} \tag{13}$$

Soft Thresholding Function

This denoiser is nearly minimax optimal:

$$\eta(y;\theta) = \begin{cases} y-\theta, & y > \theta \\ 0, & -\theta \le y \le \theta \\ y+\theta, & y < -\theta \end{cases}$$

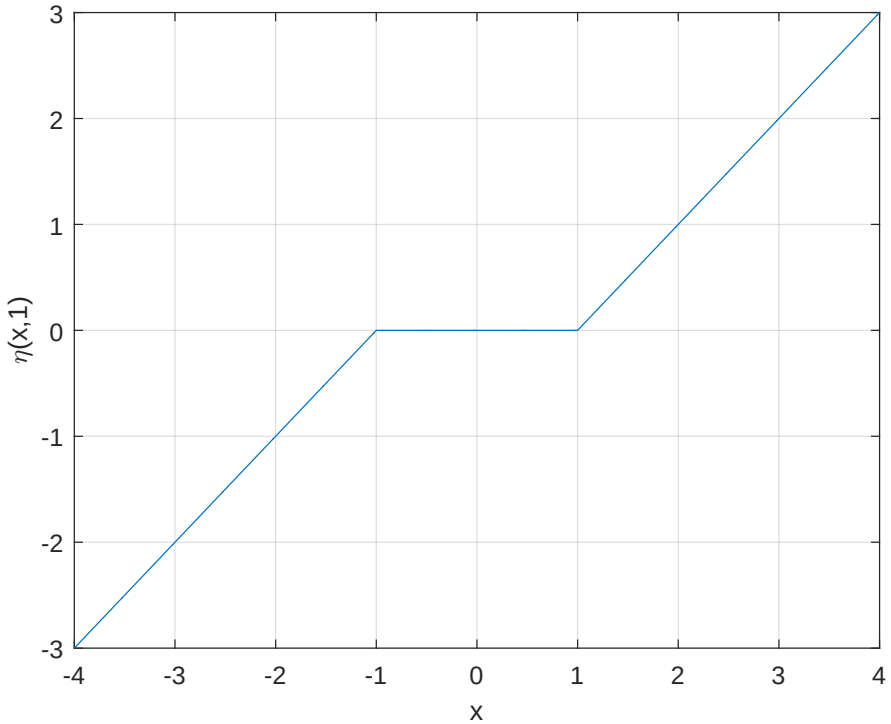
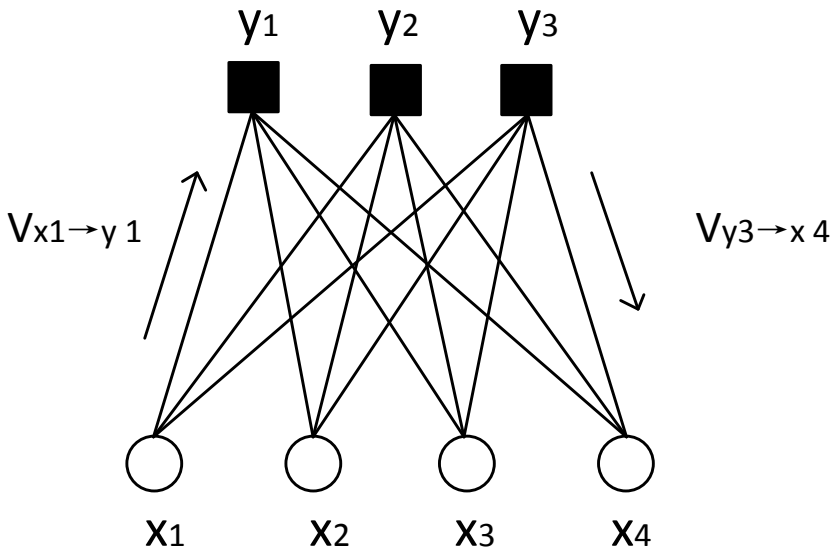


Figure: Soft thresholding function with $\theta = 1$

AMP via Graphical Model

Graphical model with message passing [Donoho-Maleki-Montanari'09]



Main features:

- Soft thresholding emerges in a minimax solution.
- State evolution describes the progress in iteration.
- Better denoiser design is possible by accounting for channel statistics.

AMP Algorithm

Algorithm: Correlate, denoise, then iterate with the residual

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{x}^t + \boldsymbol{S}^T \boldsymbol{r}^t; \lambda + \gamma^t) \tag{14}$$

$$\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}^t + \frac{1}{L}\boldsymbol{r}^{t-1}\|\boldsymbol{x}^t\|_0, \tag{15}$$

where the threshold satisfies

$$\gamma^{t+1} = \frac{\lambda + \gamma^t}{L}\|\boldsymbol{x}^{t+1}\|_0 \tag{16}$$

Note: the threshold γ^{t+1} is fixed by the recursion.

Without the last "Onsager term", this is the classical iterative soft thresholding.

AMP Algorithm – General Form

Recover $\boldsymbol{x}$ from $\boldsymbol{y}$ via AMP algorithm (complex case)

$$\boldsymbol{x}^{t+1} = \eta_t(\boldsymbol{S}^*\boldsymbol{r}^t + \boldsymbol{x}^t)$$

$$\boldsymbol{r}^{t+1} = \boldsymbol{y} - \boldsymbol{S}\boldsymbol{x}^{t+1} + \frac{\boldsymbol{r}^t}{\delta}\langle \eta_t'(\boldsymbol{S}^*\boldsymbol{r}^t + \boldsymbol{x}^t)\rangle$$

- $\boldsymbol{x}^t$: estimate of $\boldsymbol{x}$ at iteration t
- $\boldsymbol{r}^t$: residual at iteration t
- $\eta_t(\cdot)$: for soft thresholding, $\eta_t(\cdot) = \eta(\cdot, \theta\frac{1}{\sqrt{L}}\|\boldsymbol{r}^t\|_2)$, where θ is free parameter
- $\eta_t'(\cdot)$: first order derivative of $\eta_t(\cdot)$
- $\delta \triangleq \frac{L}{N}$: undersampling ratio
- $\langle\cdot\rangle$: averaging operation over all entries of a vector

State Evolution of AMP

The performance of AMP at each iteration can be predicted in the asymptotic regime where $L \to \infty, N \to \infty$ with fixed $\frac{L}{N}$

- $\boldsymbol{S}^* \boldsymbol{r}^t + \boldsymbol{x}^t$ can be modeled as signal plus noise, i.e., $\boldsymbol{x} + \boldsymbol{v}^t$
- $\boldsymbol{v}^t$ is i.i.d. Gaussian noise with variance τ_t tracked by state evolution equation

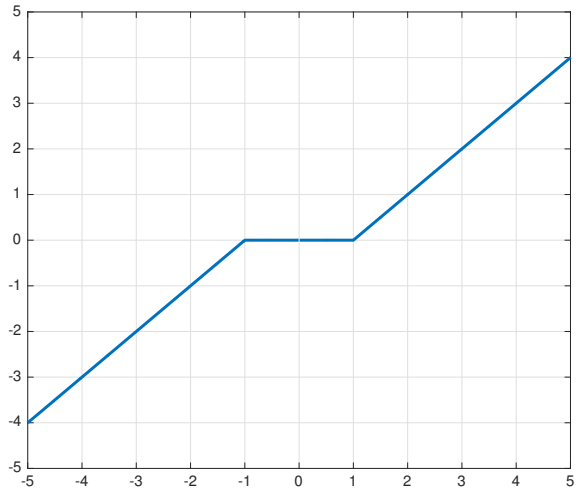
$$\tau_{t+1}^2 = \sigma_w^2 + \frac{1}{\delta}\mathbb{E}|\eta_t(X + \tau_t W) - X|^2 \tag{17}$$

  - X: random variable following the same distribution as $\boldsymbol{x}$
  - W: random variable following $\mathcal{CN}(0,1)$
  - initialization: $\tau_0 \triangleq \sigma_w^2 + \frac{1}{\delta}\mathbb{E}|X|^2$
- Interpretation of state evolution: vector estimation $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{w}$ is reduced to uncoupled scalar estimation $(\boldsymbol{x}^t + (\boldsymbol{S}^* \boldsymbol{r}^t)_i = x_i + v_i^t$

Denoiser for AMP

Complex soft thresholding denoiser:

$$\eta_t^{\text{soft}}(\tilde{x}^t) \triangleq \left(\tilde{x}^t - \theta\tau_t \frac{\tilde{x}^t}{|\tilde{x}^t|}\right) \mathbb{I}(|\tilde{x}^t| > \theta\tau_t) \qquad (18)$$



- θ: threshold control parameter
- τ_t: noise variance, estimated by $\hat{\tau}_t = \frac{1}{\sqrt{L}}\|\boldsymbol{r}^t\|_2$
- $\mathbb{I}(\cdot)$: indicator function

The above is the classical minimax denoiser based on soft thresholding.

Better MMSE denoiser can be designed while accounting for channel distribution:

$$\eta_t^{\text{mmse}}(\tilde{x}^t) \triangleq \mathbb{E}(X|\tilde{X}^t = \tilde{x}^t) \qquad (19)$$

where $\tilde{X}^t$ is the random variable defined as $\tilde{X}^t \triangleq X + \tau_t W$.

Comparison of Denoisers

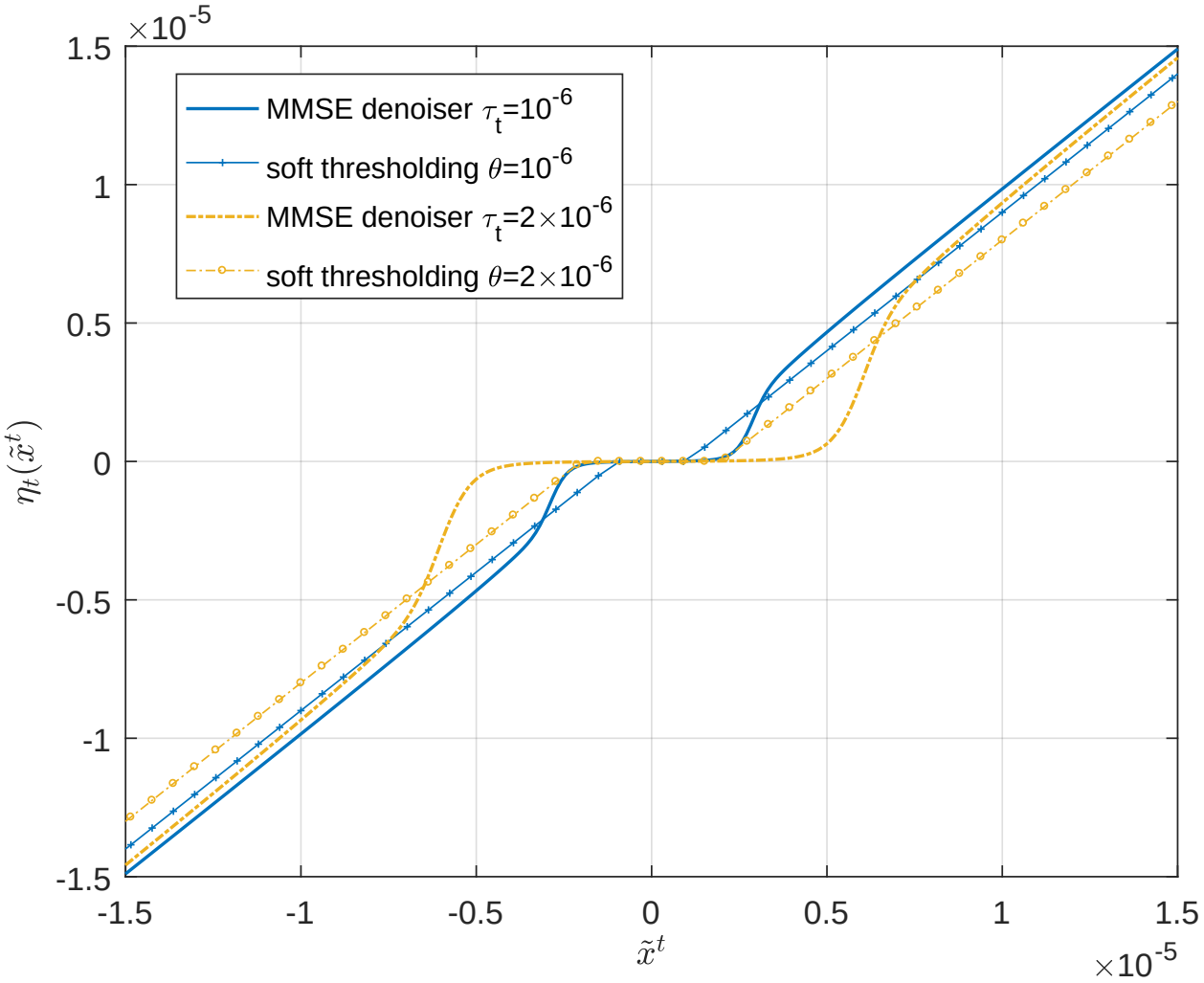


Figure: Soft thresholding denoiser and MMSE denoiser

User Activity Detection

Recall the signal plus noise model in AMP: $(\boldsymbol{S}^* \boldsymbol{r}^t + \boldsymbol{x}^t)_i = x_i + v_i^t$, which can be re-expressed as $\tilde{X}^t = X + \tau_t W$ via random variables $\tilde{X}^t, X, W$

Consider the hypothesis testing problem

$$\begin{cases} H_0 : X = 0, \text{ user is inactive} \\ H_1 : X \neq 0, \text{ user is active} \end{cases} \tag{20}$$

The optimal decision rule

$$LLR = \log \left(\frac{p_{\tilde{X}^t|X}(\tilde{x}^t | x \neq 0)}{p_{\tilde{X}^t|X}(\tilde{x}^t | x = 0)} \right) \underset{H_1}{\overset{H_0}{\lessgtr}} l_{th} \tag{21}$$

- LLR: log-likelihood ratio
- l_{th}: decision threshold determined by the detection criterion.

Analysis of Detection Error Probability

By state evolution, the likelihood distribution given X can be derived as:

$$p_{\tilde{X}^t|X}(\tilde{x}^t|x=0) = \frac{1}{\pi\tau_t^2}\exp\left(-\frac{|\tilde{x}^t|^2}{\tau_t^2}\right) \tag{22}$$

$$p_{\tilde{X}^t|X}(\tilde{x}^t|x\neq 0) = a\int_0^\infty \frac{\mathrm{erfc}(b\ln z + c)}{z^\gamma(z^2+\tau_t^2)}\exp\left(\frac{-|\tilde{x}^t|^2}{z^2+\tau_t^2}\right)dz \tag{23}$$

The log-likelihood ratio is computed as

$$LLR = \log\int_0^\infty \frac{a\pi\tau_t^2 z^{-\gamma}}{z^2+\tau_t^2}\mathrm{erfc}(b\ln z + c)\exp(|\tilde{x}^t|^2\Delta)dz \tag{24}$$

- $\Delta \triangleq \frac{1}{\tau_t^2} - \frac{1}{z^2+\tau_t^2}$
- LLR is monotonic in $|\tilde{x}^t|$

Missed Detection vs. False Alarm Probabilities

Based on the monotonicity, we simplify the decision rule as

$$|\tilde{x}^t| \underset{H_1}{\overset{H_0}{\lessgtr}} l'_{th} \tag{25}$$

The false alarm and missed detection probabilities:

$$P_F^t = \int_{|\tilde{x}^t|>l'_{th}} p_{\tilde{X}^t|X}(\tilde{x}^t|x=0)d\tilde{x}^t \tag{26}$$

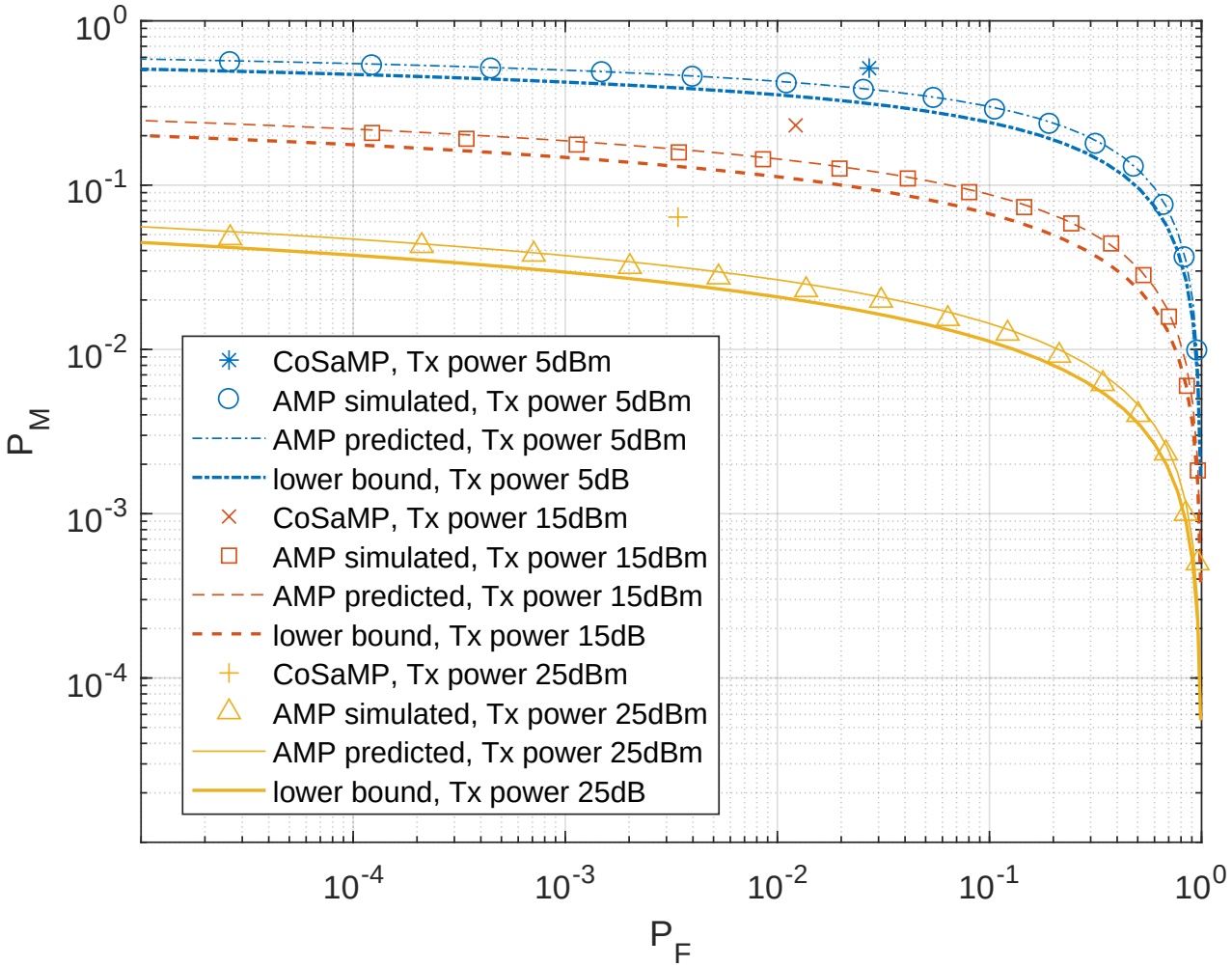
$$P_M^t = \int_{|\tilde{x}^t|<l'_{th}} p_{\tilde{X}^t|X}(\tilde{x}^t|x\neq 0)d\tilde{x}^t \tag{27}$$

- Decision is based on the amplitude of $\tilde{x}$
- Trade-off between P_F^t and P_M^t is achieved by adjusting l'_{th}
- P_F^t and P_M^t depend on noise variance τ_t (τ_∞ after converging), which can be tracked via the AMP state evolution

Simulation Parameters

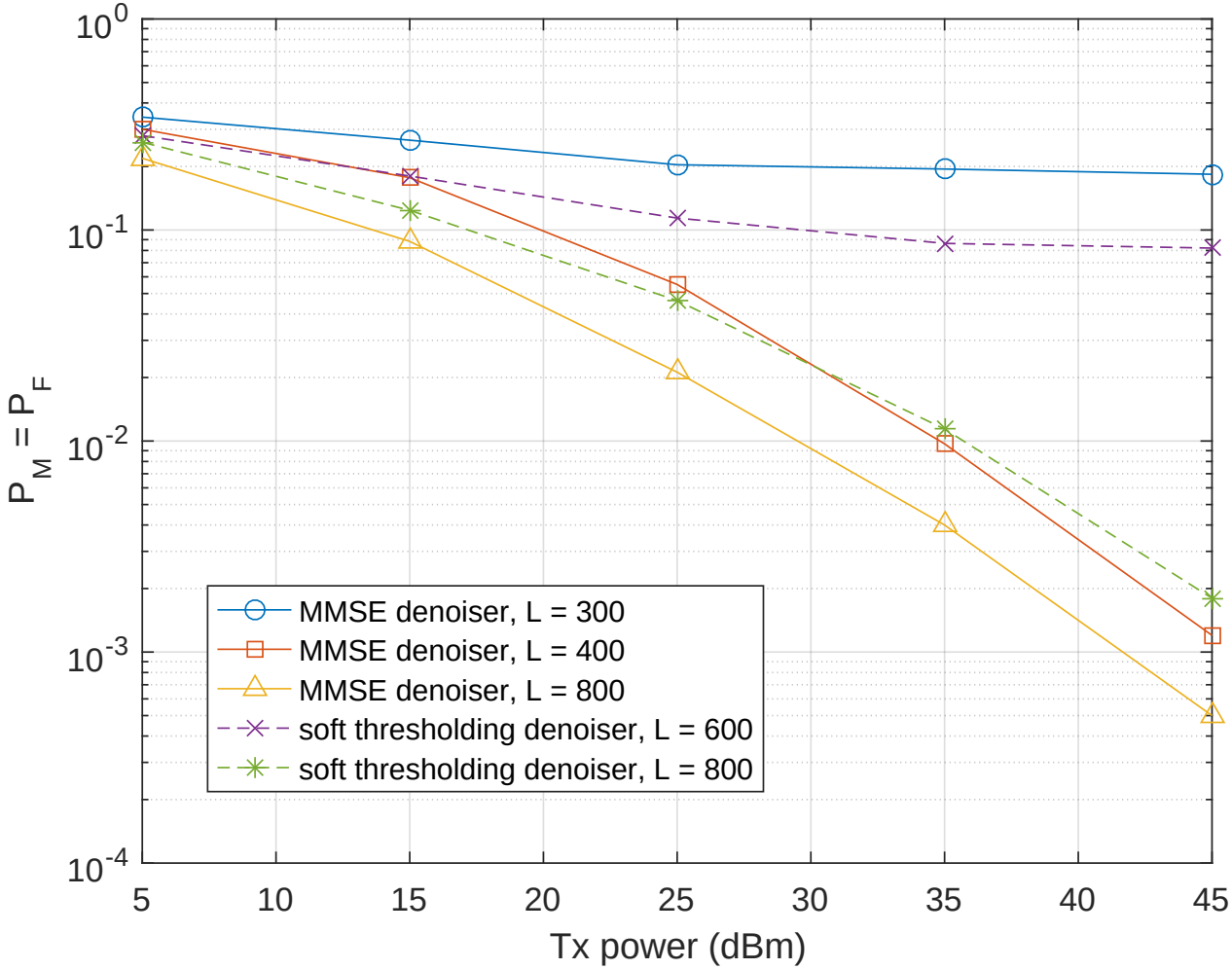
User number N	4000
Cell radius R	1000m
Activity probability ϵ	0.05
Signature sequence length L	800
Pathloss parameter α	15.3
Pathloss parameter β	37.6
Shadowing parameter σ_{SF}	8 dB
Background noise power	-99 dBm
Transmission power	5, 15, 25 dBm

Missed Detection vs. False Alarm



Small mismatch of predicted vs simulated curves due to neglecting shadowing

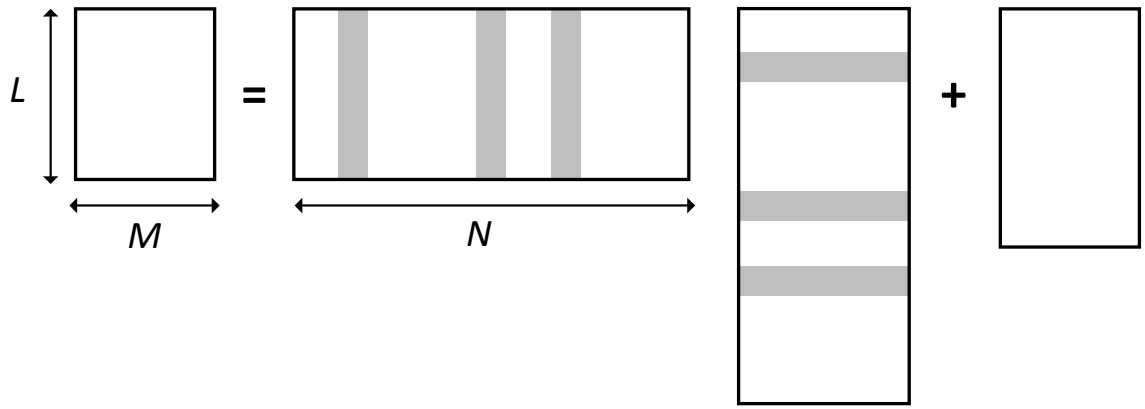
AMP Performance vs. SNR



Threshold for MMSE denoiser is better than soft thresholding denoiser.

Multi-Antenna Case

Multiple Antennas at the BS



- Multiple measurement vector (MMV) problem
  - Better performance than single measurement vector (SMV)
- Asymptotic analysis: Fix M, let $N, K, L \to \infty$, $\epsilon = \frac{K}{N}$, $\delta = \frac{L}{N}$,
- Main insight: Perfect user detection is possible when $M \to \infty$!
- But, the multi-antenna case is also more challenging:
  (i) convergence is slower;
  (ii) channel estimation error.

Two-Phase Transmission

- Pilot Transmission Phase
  - Due to large number of devices, non-orthogonal pilots are inevitable.
  - The same pilots can be used for both activity detection and channel estimation.
- Data Transmission Phase
  - The achievable rates are limited by the channel estimation error.

Signal Model in Pilot Phase

Received signal in pilot phase:

$$\boldsymbol{Y} = \sqrt{\xi}\sum_{n=1}^{N}\alpha_n \boldsymbol{s}_n \boldsymbol{h}_n^T + \boldsymbol{Z} \triangleq \sqrt{\xi}\boldsymbol{S}\boldsymbol{X} + \boldsymbol{Z} \tag{28}$$

- $\xi = \rho^{\text{pilot}} L$: total transmit energy in pilot phase
- $\boldsymbol{h}_n \in \mathbb{C}^{M\times 1}$: channel coefficient between user n and BS, including path-loss fading, shadowing and Rayleigh fading static within each block;
- $\alpha_n \in \{1, 0\}$: indicating whether user n is active
- $\boldsymbol{X} \triangleq [\boldsymbol{h}_1\alpha_1, \boldsymbol{h}_2\alpha_2, \cdots, \boldsymbol{h}_N\alpha_N]^T \in \mathbb{C}^{N\times M}$
- $\boldsymbol{s}_n \in \mathbb{C}^{L\times 1}$: signature sequence of user n following i.i.d. $\mathcal{CN}(0, 1/L)$
- $\boldsymbol{S} \triangleq [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_N]^T \in \mathbb{C}^{L\times N}$
- $\boldsymbol{Z} \in \mathbb{C}^{L\times M}$: effective noise following i.i.d. $\mathcal{CN}(0, \sigma^2\boldsymbol{I})$

Vector Approximate Message Passing

- Vector generalization of AMP works iteratively as follows:

$$\boldsymbol{x}_n^{t+1} = \eta_{t,n}((\boldsymbol{R}^t)^H \boldsymbol{s}_n + \boldsymbol{x}_n^t)$$

$$\boldsymbol{R}^{t+1} = \boldsymbol{Y} - \boldsymbol{S}\boldsymbol{X}^{t+1} + \frac{1}{\delta}\boldsymbol{R}^t \sum_{n=1}^{N} \frac{\eta'_{t,n}((\boldsymbol{R}^t)^H \boldsymbol{s}_n + \boldsymbol{x}_n^t)}{N}$$

  where $\boldsymbol{R}^t = [\boldsymbol{r}_1^t, \cdots, \boldsymbol{r}_L^t]^T \in \mathbb{C}^{L \times M}$ is the residual

- Use MMSE denoise that accounts for channel distribution: (Here $\delta = \frac{L}{N}$)
  - $\eta_{t,n}(\cdot)$: denoiser that depends on β_n
  - $\eta'_{t,n}(\cdot)$: first-order derivative of $\eta_{t,n}(\cdot)$

State Evolution

- Performance analysis by state evolution for $K, N, L \to \infty$ with $\frac{L}{N} = \delta$ [Bayati-Montanari'11], [Kim-Chang-Jung-Baron-Ye'11], [Rangan'11]:

$$\Sigma_{t+1} = \frac{\sigma^2}{\xi}\boldsymbol{I} + \frac{1}{\delta}\mathbb{E}\left[(\eta_{t,\beta}(\boldsymbol{X}_\beta + \Sigma_t^{\frac{1}{2}}\boldsymbol{V}) - \boldsymbol{X}_\beta)(\eta_{t,\beta}(\boldsymbol{X}_\beta + \Sigma_t^{\frac{1}{2}}\boldsymbol{V}) - \boldsymbol{X}_\beta)^H\right] \tag{29}$$

- AMP is statistically equivalent to applying the denoiser to

$$\hat{\boldsymbol{x}}_{t,n} = \boldsymbol{x}_n + \Sigma_t^{\frac{1}{2}}\boldsymbol{v}_n = \alpha_n \boldsymbol{h}_n + \Sigma_t^{\frac{1}{2}}\boldsymbol{v}_n \tag{30}$$

- MMSE denoiser:

$$\eta_{t,n}(\hat{\boldsymbol{x}}_{t,n}) = \phi_{t,n}\beta_n(\beta_n\boldsymbol{I} + \Sigma_t)^{-1}\hat{\boldsymbol{x}}_{t,n} \tag{31}$$

$$\phi_{t,n} = \frac{1}{1 + \frac{1-\varepsilon}{\varepsilon}\exp\left(-\frac{M}{2}\left(\pi_{t,n} - \psi_{t,n}\right)\right)} \tag{32}$$

$$\pi_{t,n} = \frac{\hat{\boldsymbol{x}}_{t,n}^H(\Sigma_t^{-1} - (\Sigma_t + \beta_n\boldsymbol{I})^{-1})\hat{\boldsymbol{x}}_{t,n}}{M} \tag{33}$$

$$\psi_{t,n} = \frac{\log\det(\boldsymbol{I} + \beta_n\Sigma_t^{-1})}{M} \tag{34}$$

Simplified MMSE Denoiser

- With i.i.d. fading, Σ_{t+1} is a diagonal matrix with identical diagonal entries

$$\Sigma_t = \tau_t^2 \boldsymbol{I}$$

- MMSE denoiser reduces to

$$\eta_{t,n}(\hat{\boldsymbol{x}}_{t,n}) = \phi_{t,n} \frac{\beta_n}{\beta_n + \tau_t^2} \hat{\boldsymbol{x}}_{t,n} \tag{35}$$

$$\phi_{t,n} = \frac{1}{1 + \frac{1-\varepsilon}{\varepsilon} \exp\left(-\frac{M}{2}\left(\pi_{t,n} - \psi_{t,n}\right)\right)} \tag{36}$$

$$\pi_{t,n} = \frac{\left(\frac{1}{\tau_t^2} - \frac{1}{\tau_t^2 + \beta_n}\right) \hat{\boldsymbol{x}}_{t,n}^H \hat{\boldsymbol{x}}_{t,n}}{M} \tag{37}$$

$$\psi_{t,n} = \log\left(1 + \frac{\beta_n}{\tau_t^2}\right) \tag{38}$$

- Asymptotically as $M \to \infty$, $\phi_{t,n}$ is either 0 or 1 depending on whether device n is active or not.

Massive MIMO Guarantees Perfect Activity Detection

Theorem: A massive MIMO system can detect device activities perfectly, i.e.,

$$\lim_{M\to\infty} P_M^{t,n}(M) = \lim_{M\to\infty} P_F^{t,n}(M) = 0$$

Proof: By strong law of large numbers:

$$\pi_{t,n} \to \begin{cases} \beta_n/\tau_t^2, & \text{if } \alpha_n = 1 \\ \beta_n/(\beta_n + \tau_t^2), & \text{if } \alpha_n = 0 \end{cases}$$

The proof follows as $a > \log(1+a) > \frac{a}{1+a}$ for all $a > 0$.

What is the cost of massive connectivity?

Channel Estimation Error

- Covariance of estimated channel $\hat{\boldsymbol{h}}_{t,k}$: $\mathrm{Cov}(\hat{\boldsymbol{h}}_{t,k}, \hat{\boldsymbol{h}}_{t,k}) = \upsilon_{t,k}(M)\boldsymbol{I}$
- Covariance of channel estimation error $\Delta\boldsymbol{h}_{t,k} = \boldsymbol{h}_{t,k} - \hat{\boldsymbol{h}}_{t,k}$

$$\mathrm{Cov}(\Delta\boldsymbol{h}_{t,k}, \Delta\boldsymbol{h}_{t,k}) = \Delta\upsilon_{t,k}(M)\boldsymbol{I} \tag{39}$$

- As $M \to \infty$

$$\lim_{M\to\infty} \upsilon_k(M) = \frac{\beta_k^2}{\beta_k + \tau_\infty^2} \tag{40}$$

$$\lim_{M\to\infty} \Delta\upsilon_k(M) = \frac{\beta_k \tau_\infty^2}{\beta_k + \tau_\infty^2} \tag{41}$$

where τ_∞^2 is the fixed-point solution to state evolution: $(\epsilon = \frac{K}{N},\ \delta = \frac{L}{N})$

$$\tau_{t+1}^2 = \frac{\sigma^2}{\xi} + \frac{\epsilon}{\delta}\mathbb{E}_\beta\left[\frac{\beta\tau_t^2}{\beta + \tau_t^2}\right] \tag{42}$$

Assuming $L > K$ and high SNR, then $\tau_\infty^2 \to \frac{\sigma^2}{\xi(1-\frac{\epsilon}{\delta})}$.

Data Transmission Phase

- Consider a system with K users transmitting to BS with M antennas.
- Received signal at the BS with estimated channel $\tilde{\boldsymbol{h}}_k$'s:

$$\boldsymbol{y} = \sum_{k\in\mathcal{K}} \boldsymbol{h}_k \sqrt{\rho^{\text{data}}} u_k + \boldsymbol{z} = \sum_{k\in\mathcal{K}} \tilde{\boldsymbol{h}}_k \sqrt{\rho^{\text{data}}} u_k + \sum_{k\in\mathcal{K}} \Delta\boldsymbol{h}_k \sqrt{\rho^{\text{data}}} u_k + \boldsymbol{z}$$

- Maximum ratio combining:

$$\hat{u}_k = \boldsymbol{w}_k^H \tilde{\boldsymbol{h}}_k \sqrt{\rho^{\text{data}}} u_k + \boldsymbol{w}_k^H \left(\sum_{n\in\mathcal{K}, n\neq k} \tilde{\boldsymbol{h}}_n \sqrt{\rho^{\text{data}}} u_n + \sum_{n\in\mathcal{K}} \Delta\boldsymbol{h}_n \sqrt{\rho^{\text{data}}} u_n + \boldsymbol{z} \right)$$

- The achievable rate of user k is [Hassibi-Hochwald'03]

$$R_k = \frac{T-L}{T} \mathbb{E}[\log_2(1+\gamma_k)], \quad \forall k \in \mathcal{K},$$

where

$$\gamma_k = \frac{\rho^{\text{data}} |\boldsymbol{w}_k^H \hat{\boldsymbol{h}}_k|^2}{\rho^{\text{data}} \sum\limits_{n\in\mathcal{K}, n\neq k} |\boldsymbol{w}_k^H \hat{\boldsymbol{h}}_n|^2 + \rho^{\text{data}} \|\boldsymbol{w}_k\|^2 \sum\limits_{n\in\mathcal{K}} \frac{\beta_n \tau_\infty^2}{\beta_n + \tau_\infty^2} + \sigma^2 \|\boldsymbol{w}_k\|^2}$$

Achievable Rate with MMSE Beamforming

- With MMSE: $\boldsymbol{w}_k^{\mathrm{MMSE}} = \left(\sum_{n\in\mathcal{K}} \rho^{\mathrm{data}} \hat{\boldsymbol{h}}_n \hat{\boldsymbol{h}}_n^H + \sum_{n\in\mathcal{K}} \frac{\rho^{\mathrm{data}} \beta_n \tau_\infty^2}{\beta_n + \tau_\infty^2} \boldsymbol{I} + \sigma^2 \boldsymbol{I} \right)^{-1} \hat{\boldsymbol{h}}_k$

$$\lim_{M\to\infty} \gamma_k^{\mathrm{MMSE}} \to \frac{\beta_k^2}{\beta_k + \tau_\infty^2} \Gamma$$

where Γ is fixed-point solution to ($\mu = \frac{K}{M}$):

$$\Gamma = \frac{1}{\mu \mathbb{E}\left[\frac{\beta^2}{\beta + \tau_\infty^2 + \beta^2 \Gamma}\right] + \mu \mathbb{E}\left[\frac{\beta \tau_\infty^2}{\beta + \tau_\infty^2}\right]}$$

In the special case of perfect CSI, the above result reduces to [Tse-Hanly'99]

Cost of Massive Uncoordinated Access

- Fixed number of K users:

$$\gamma_k \to \frac{\beta_k^2}{\beta_k + \frac{\sigma^2}{\rho^{\text{pilot}} L}} \Gamma \tag{43}$$

$$\frac{1}{\Gamma} = \frac{1}{M} \sum_{n \in \mathcal{K}} \frac{\beta_n^2}{\beta_n + \frac{\sigma^2}{\rho^{\text{pilot}} L} + \beta_n^2 \Gamma} + \frac{1}{M} \sum_{n \in \mathcal{K}} \frac{\frac{\beta_n \sigma^2}{\rho^{\text{pilot}} L}}{\beta_n + \frac{\sigma^2}{\rho^{\text{pilot}} L}} \tag{44}$$

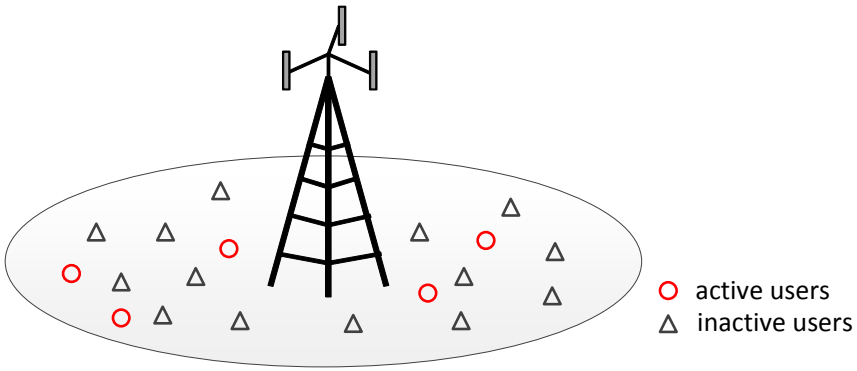
- Massive number of N potential users with K active user:

$$\gamma_k \to \frac{\beta_k^2}{\beta_k + \tau_\infty^2} \Gamma \tag{45}$$

$$\frac{1}{\Gamma} = \frac{1}{M} \sum_{n \in \mathcal{K}} \frac{\beta_n^2}{\beta_n + \tau_\infty^2 + \beta_n^2 \Gamma} + \frac{1}{M} \sum_{n \in \mathcal{K}} \frac{\beta_n \tau_\infty^2}{\beta_n + \tau_\infty^2} \tag{46}$$

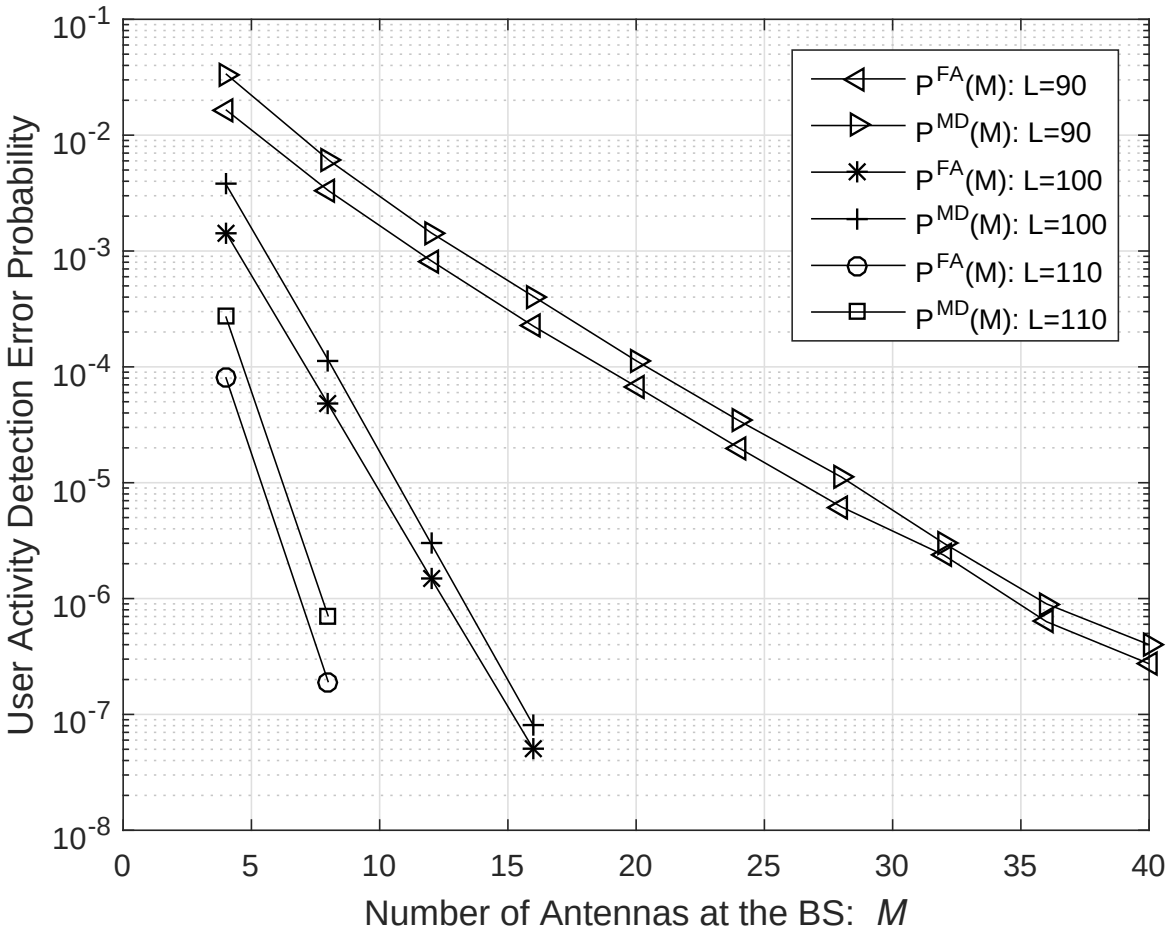
- At high SNR: $\tau_\infty^2 \approx \frac{\sigma^2}{\rho^{\text{pilot}}(L-K)}$
- Channel estimation error is increased due to the non-orthogonal pilots

Numerical Example



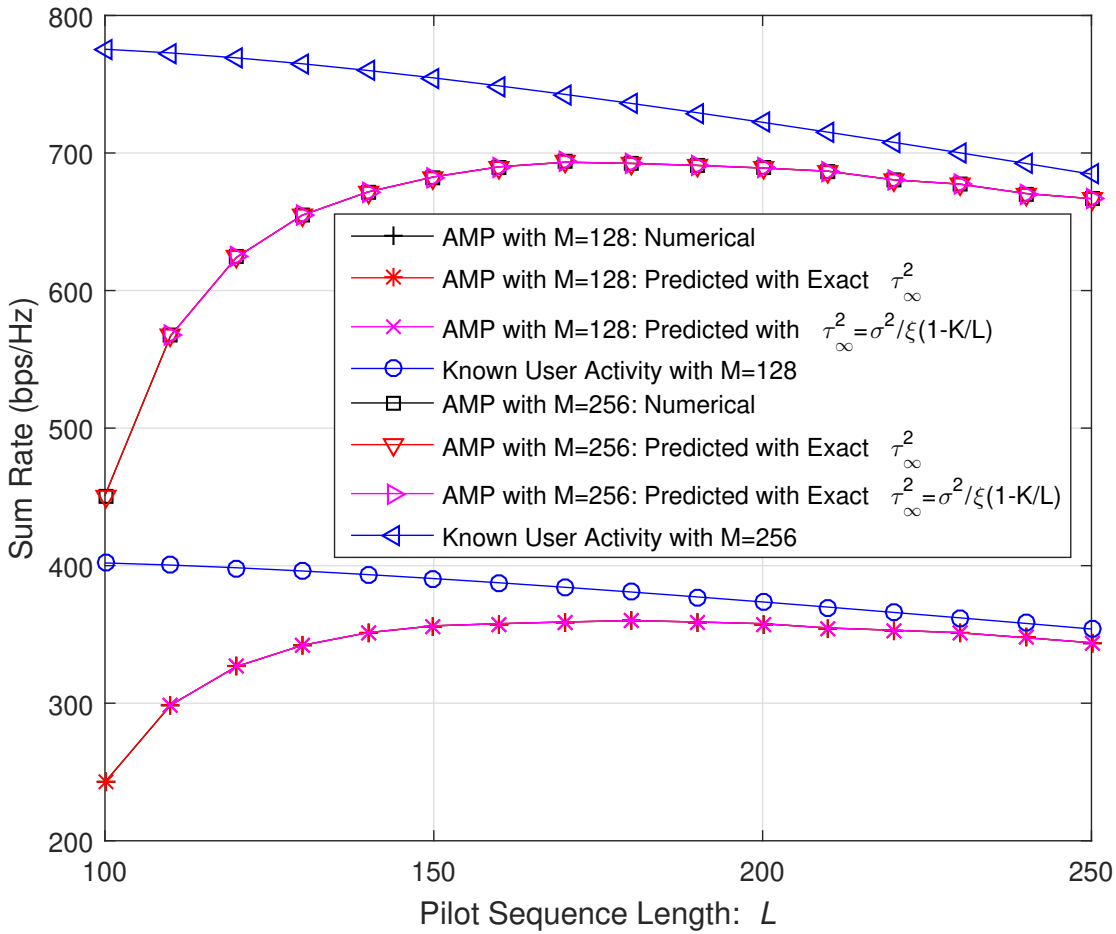
- $N = 2000$ users
- $K = 100$ active users
- Transmit power: $\rho^{\text{pilot}} = \rho^{\text{data}} = 23\text{dBm}$
- User distance to BS: $[0.5\text{km}, 1\text{km}]$
- Path loss: $\beta_n = -128.1 - 36.7 \log_{10}(d_n)$, $\forall n$
- 100kHz bandwidth, 10ms coherence time
- $T = 1000$ symbols per coherence time

User Activity Detection



- Probabilities of missed detection and false alarm reduce as L increases
- Probabilities of missed detection and false alarm go to zero as M increases

Achievable Rate with Massive MIMO

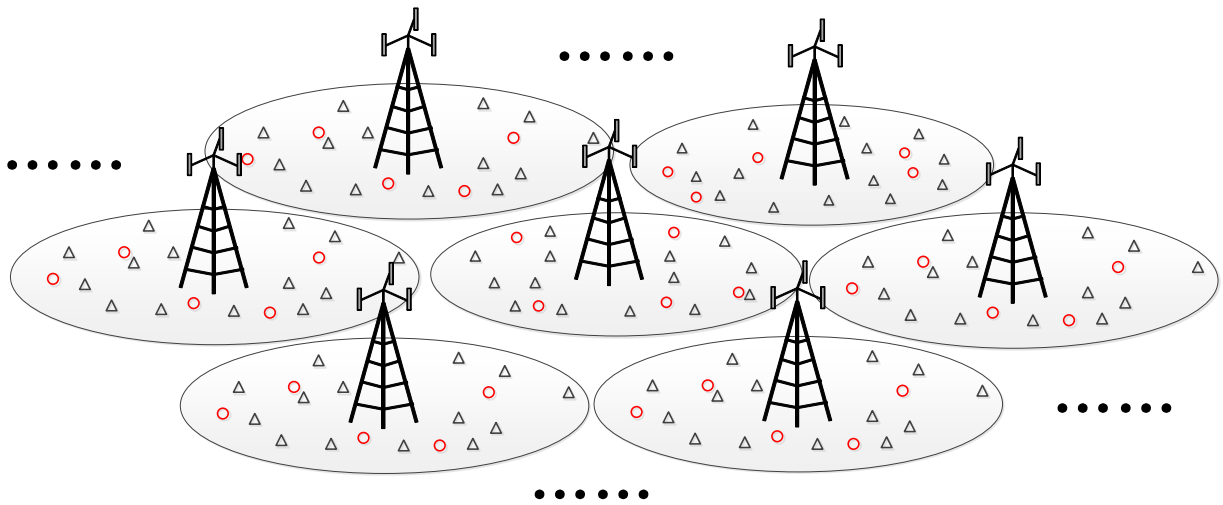


- The optimal L when user activity is unknown needs to be longer
- There is a loss in sum-rate due to the need for longer pilot

Multi-Cell Systems

User Activity Detection in Multicell Systems

- What is the impact of the inter-cell interference?



- How to overcome the inter-cell interference?

Activity Detection in Multicell Systems

- Multi-cell system with B BSs each equipped with M antennas;
- N single-antenna devices per cell, K of which are active;
- Device n in cell b is assigned a length-L unique signature sequence $\boldsymbol{s}_{bn}$;
- Received signal $\boldsymbol{Y}_b \in \mathbb{C}^{L\times M}$ at BS b is

$$\begin{aligned}\boldsymbol{Y}_b &= \sum_{n=1}^{N} \alpha_{bn}\boldsymbol{s}_{bn}\boldsymbol{h}_{bbn}^T + \sum_{j=1,j\neq b}^{j=B}\sum_{n=1}^{N} \alpha_{jn}\boldsymbol{s}_{jn}\boldsymbol{h}_{bjn}^T + \boldsymbol{Z}_b \\ &= \boldsymbol{S}_b\boldsymbol{X}_{bb} + \sum_{j=1,j\neq b}^{j=B} \boldsymbol{S}_j\boldsymbol{X}_{bj} + \boldsymbol{Z}_b, \end{aligned} \tag{47}$$

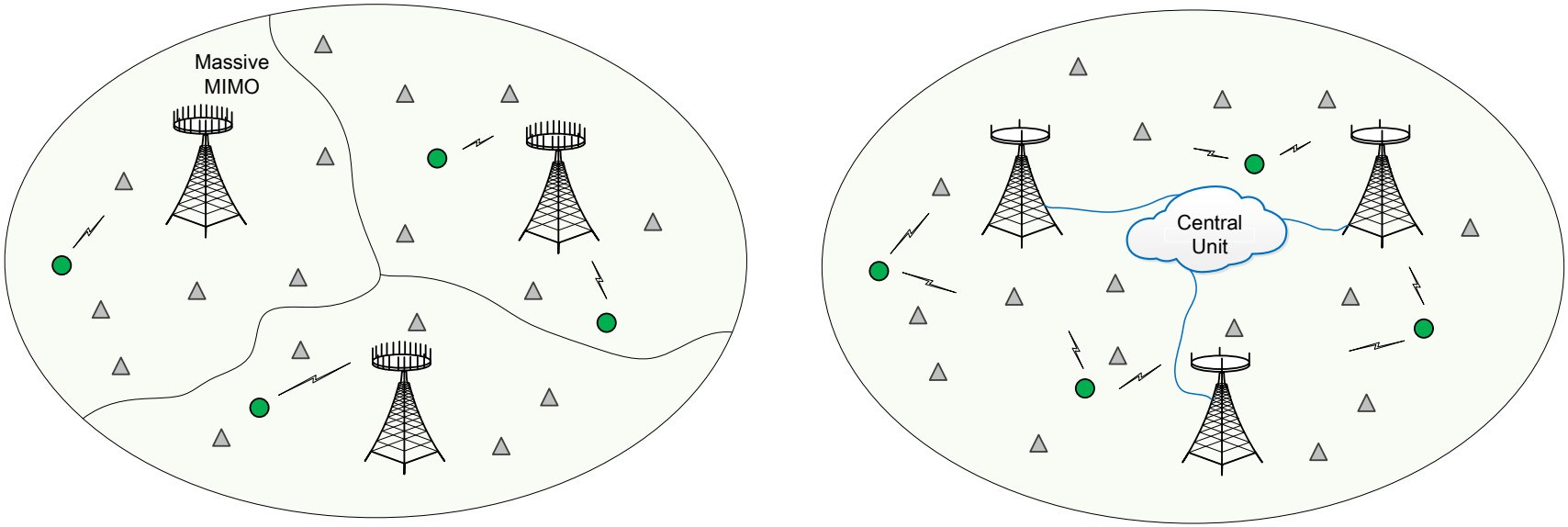
where

- $\alpha_{bn} \in \{1, 0\}$ activity indicator; $\boldsymbol{Z}_b \in \mathbb{C}^{L\times M}$ Gaussian noise with variance σ^2.
- $\boldsymbol{h}_{bjn} \in \mathbb{C}^{M\times 1}$ is the channel from user n in cell j to BS b
- $\boldsymbol{S}_j \triangleq [\boldsymbol{s}_{j1}, \cdots, \boldsymbol{s}_{jN}] \in \mathbb{C}^{L\times N}$; $\boldsymbol{X}_{bj} \triangleq [\alpha_{j1}\boldsymbol{h}_{bj1}, \cdots, \alpha_{jN}\boldsymbol{h}_{bjN}]^T \in \mathbb{C}^{N\times M}$

The inter-cell interference brings performance degradation for activity detection.

AMP Based Activity Detection for Multi-cell

With AMP, we consider two strategies to deal with the inter-cell interference



- **Massive MIMO**: Each BS has a large-scale antenna array, and operates independently, while treating the inter-cell interference as noise.
- **Cooperative MIMO**: Each BS has a moderate number of antennas, and is connected to a central unit (CU), where cooperative detection is performed.

Activity Detection in Massive MIMO System

- Each BS is equipped with a large-scale antenna array, i.e., M is large.
- Each BS aims to detect the active devices within its own cell, and the inter-cell interference is treated as noise:

$$\begin{aligned} \boldsymbol{Y}_b &= \boldsymbol{S}_b \boldsymbol{X}_{bb} + \sum_{j \neq b} \boldsymbol{S}_j \boldsymbol{X}_{bj} + \boldsymbol{Z}_b \\ &\triangleq \boldsymbol{S}_b \boldsymbol{X}_{bb} + \boldsymbol{Z'}_b \end{aligned} \tag{48}$$

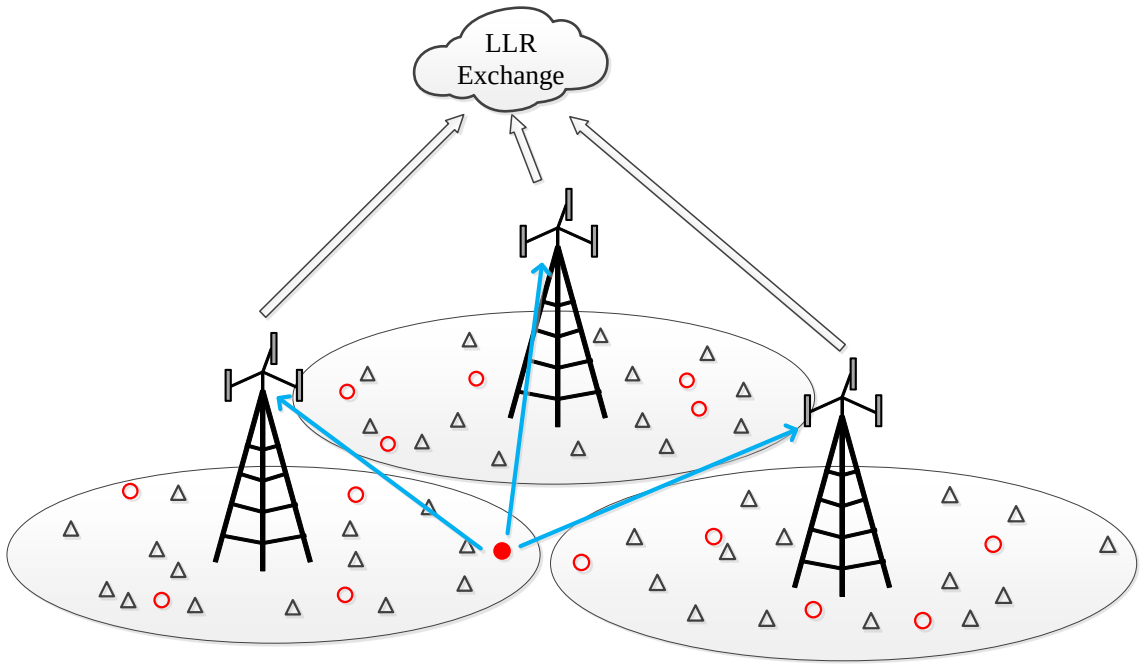
- By approximating $\boldsymbol{Z'}_b$ as a Gaussian noise, the resulting system model in multicell case is similar to that in the single-cell case.
- AMP can be used to detect the active devices in cell b by recovering the non-zero rows of $\boldsymbol{X}_{bb}$ based on $\boldsymbol{Y}_b$.

Activity Detection in Cooperative MIMO System

Potential ways to perform the cooperative detection with BSs connected to CU:

- **Centralized detection**: The received signals at the BSs $\boldsymbol{Y}_b$'s are forwarded to the CU, where a large-scale AMP is used for activity detection.
  Interference is completely avoided. However, need high-bandwidth BS-CU links.

- **Distributed detection**: Each BS performs a preliminary activity detection, and forwards the results to the CU, where an aggregation is performed.
  Forwarding the detection LLRs can save bandwidth of the BS-CU links.



Cooperative Activity Detection

- Each BS detects the active devices in all B cells using the knowledge of all signature sequences. This can be achieved by recovering the interference as

$$\begin{aligned} \boldsymbol{Y}_b &= \boldsymbol{S}_b \boldsymbol{X}_{bb} + \sum_{j \neq b} \boldsymbol{S}_j \boldsymbol{X}_{bj} + \boldsymbol{Z}_b \\ &= \begin{bmatrix} \boldsymbol{S}_1 & \cdots & \boldsymbol{S}_B \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{1b} \\ \vdots \\ \boldsymbol{X}_{Bb} \end{bmatrix} + \boldsymbol{Z}_b \\ &\triangleq \boldsymbol{S} \boldsymbol{X}_b + \boldsymbol{Z}_b \end{aligned} \tag{49}$$

- **Preliminary detection**: The BS detects the active devices by estimating the non-zero row of $\boldsymbol{X}_b$ from $\boldsymbol{Y}_b$ using AMP. This is similar to the single-cell case.
- **Quantization and forwarding**: The detection results by AMP at each BS are quantized and sent to the CU in the form of LLRs (e.g., 3-4 bits per LLR.)
- **Aggregation**: CU aggregates the independent LLRs and declares activities.

Comparison of Massive MIMO and Cooperative MIMO

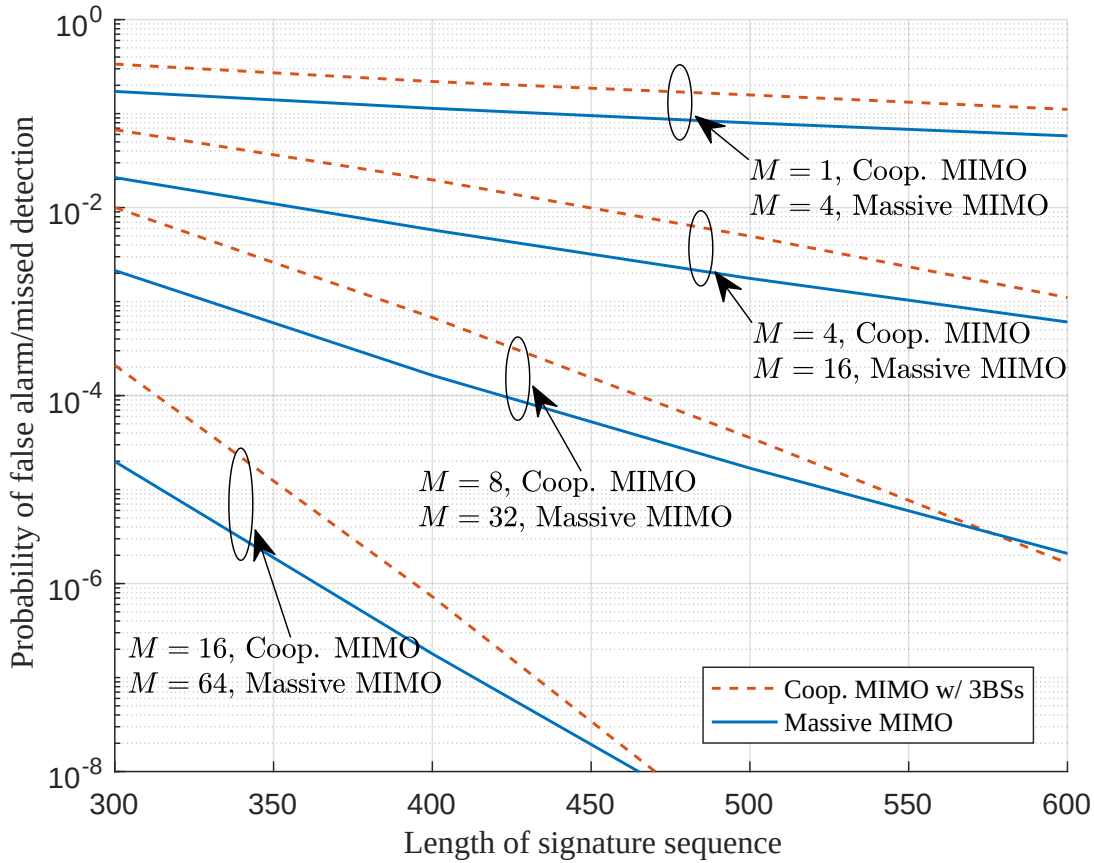


Figure: Cell-edge user performance in a network with 19 cells and 2000 devices per cell, among which 100 devices are active. To achieve comparable performance as cooperative MIMO, four times as many as antennas are required in the massive MIMO case.

Summary

AMP is a practical sparse user activities detection algorithm:

- State evolution provides accurate detector performance analysis.
- Denoiser should be designed to match channel characteristics.
- Detection becomes accurate with massive MIMO but convergence is slower.
- Cooperation can improve the cell-edge performance.

Implications for network design:

- The use of non-orthogonal pilots is inevitable.
- Massive MIMO needs to be deployed for good detection performance.
- Multi-cell cooperation can further help.
- Channel estimation is the main bottleneck.

Further Information

Zhilin Chen, Foad Sohrabi, and Wei Yu,

"Sparse Activity Detection for Massive Connectivity",

IEEE Transactions on Signal Processing, vol. 66, no. 7, pp. 1890-1904, April 2018.

Liang Liu and Wei Yu,

"Massive Connectivity with Massive MIMO – Part I: Device Activity Detection and Channel Estimation and Part II: Achievable Rate Characterization",

IEEE Transactions on Signal Processing, vol. 66, no. 11, pp. 2933-2959, June 2018.

Zhilin Chen, Foad Sohrabi, and Wei Yu,

"Multi-Cell Sparse Activity Detection for Massive Connectivity: Massive MIMO versus Cooperative MIMO",

IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4060-4074, August 2019.

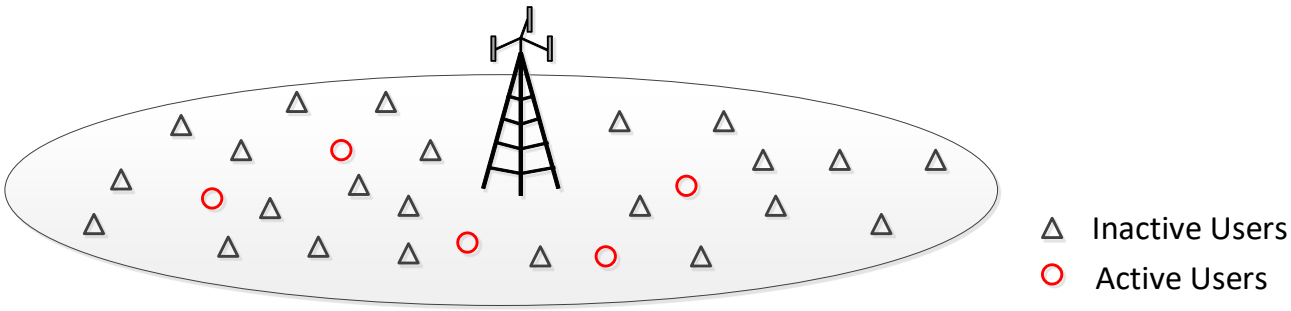
Massive Random Access with Massive MIMO: Covariance Based Detection

Wei Yu

Joint Work with Zhilin Chen, Foad Sohrabi, Ya-Feng Liu

University of Toronto

Massive Random Access for Internet-of-Things (IoT)



- Large number of devices with sporadic activity
- Low latency random access scheme for massive users is required
- Non-orthogonal signature sequences need to be used
- User activity detection (user identification) performed at base station (BS)

System Model

- BS equipped with M antennas
- N single-antenna devices, K of which are active at a time
- Each device is associated with a length-L unique signature sequence $\mathbf{s}_n$
- Channel $\mathbf{h}_n$ of user n includes both (i.i.d.) Rayleigh and large-scale fading
- For single-cell system, received signal $\mathbf{Y} \in \mathbb{C}^{L\times M}$ at the BS is

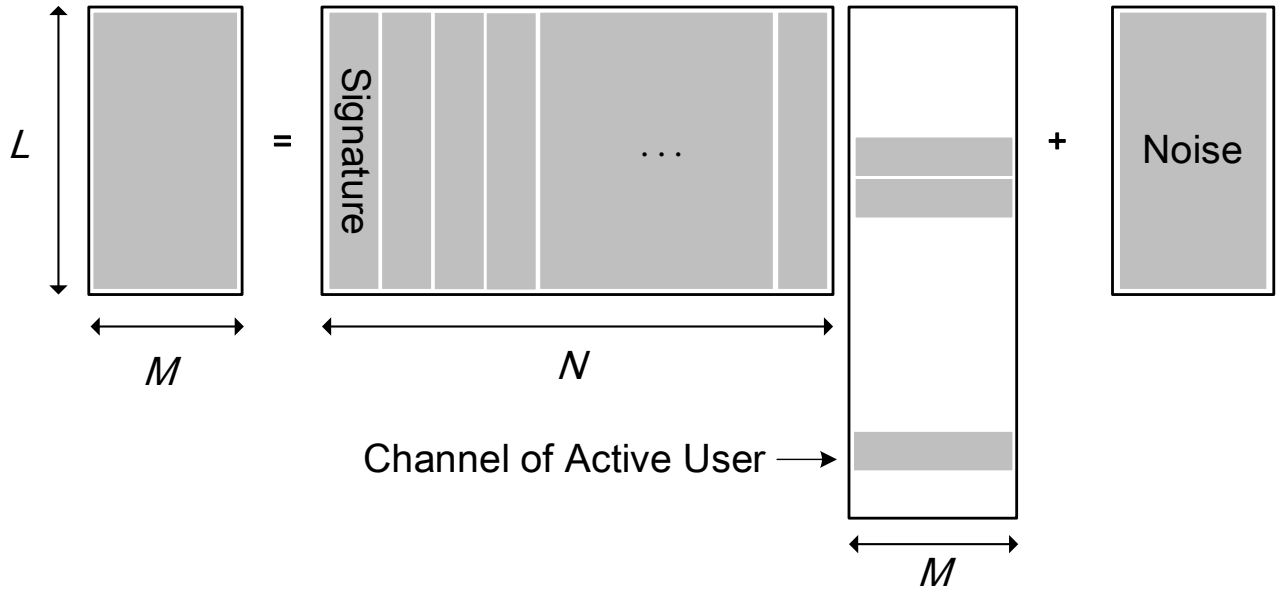
$$\mathbf{Y} = \sum_{n=1}^{N} \alpha_n \mathbf{s}_n \mathbf{h}_n^T + \mathbf{Z} = \mathbf{S}\mathbf{X} + \mathbf{Z}, \tag{1}$$

where

- $\alpha_n \in \{1, 0\}$ activity indicator; $\mathbf{Z} \in \mathbb{C}^{L\times M}$ Gaussian noise with variance σ^2
- $\mathbf{S} \triangleq [\mathbf{s}_1, \ldots, \mathbf{s}_N] \in \mathbb{C}^{L\times N}$; $\mathbf{X} \triangleq [\alpha_1 \mathbf{h}_1, \cdots, \alpha_N \mathbf{h}_N]^T \in \mathbb{C}^{N\times M}$

Joint Sparse Activity Detection and Channel Estimation

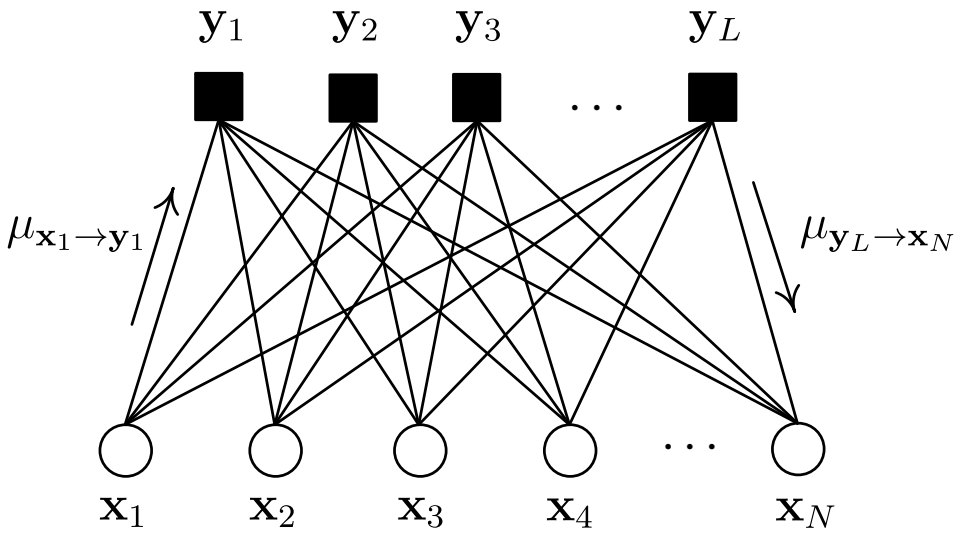
Aim to identify the K non-zero rows of $\mathbf{X}$ from $\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{Z}$.



- Multiple measurement vector (MMV) problem in compressed sensing
  - Columns of $\mathbf{X}$ share the same sparsity pattern, i.e., row sparsity
- Efficiently solved by the approximate message passing (AMP) algorithm

Approximate Message Passing (AMP)

Derived from graphical model of $\mathbf{Y} = \mathbf{SX} + \mathbf{Z}$ [Donoho-Maleki-Montanari'09]



- Suppose entries of $\mathbf{S}$ are i.i.d. random
- Aim to compute the marginals of the joint distribution $p(\mathbf{X}, \mathbf{Y})$
- Approximate $\mu_{\mathbf{y} \to \mathbf{x}}$ as Gaussian in the large system limit $N, L \to \infty$
- Further simplify the messages such that only $N + L$ messages are tracked

Intuitive Interpretation of AMP

- Matched filtering $\longrightarrow$ Denoising $\longrightarrow$ Computing and correcting the residual

$$\eta(y;\theta) = \begin{cases} y-\theta, & y > \theta \\ 0, & -\theta \le y \le \theta \\ y+\theta, & y < -\theta \end{cases} \tag{2}$$

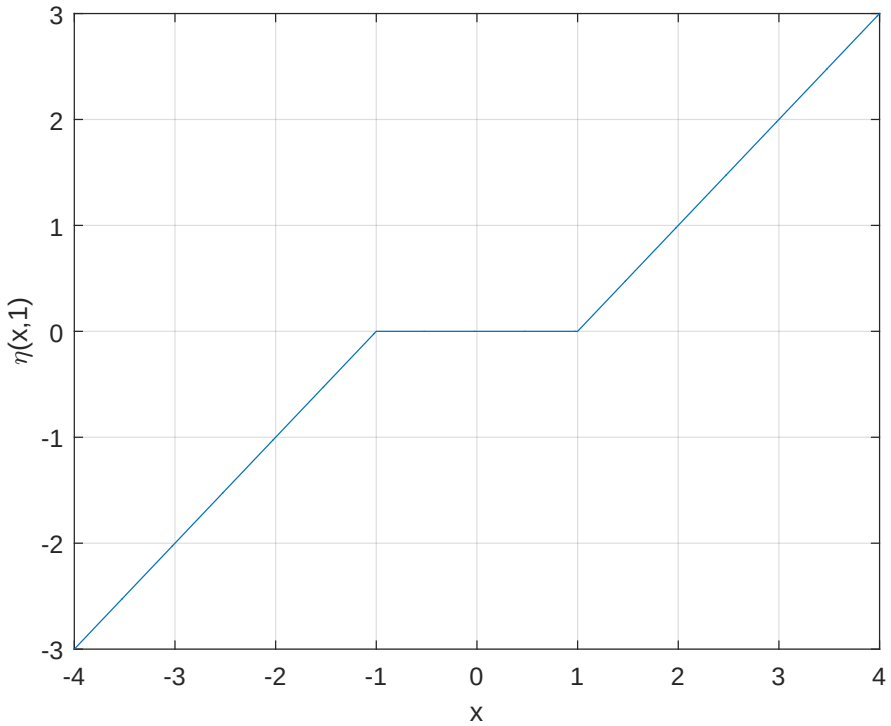


Figure: Soft thresholding function with $\theta = 1$

AMP Algorithm for MIMO

- The AMP algorithm expressed in matrix form:

$$\mathbf{X}^{t+1} = \eta_t(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t), \tag{3}$$

$$\mathbf{R}^{t+1} = \mathbf{Y} - \mathbf{S}\mathbf{X}^{t+1} + \frac{N}{L}\mathbf{R}^t\langle\eta_t'(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t)\rangle, \tag{4}$$

  where
  - $\mathbf{X}^{t+1}$, estimate at iteration $t+1$;
  - $\mathbf{R}^{t+1}$, residual at iteration $t+1$;
  - $\eta_t(\cdot)$, a non-linear function known as denoiser that performs on each row
  - $\langle\cdot\rangle$, sample averaging operation
- Works well if M is fixed, and L, N, $K \to \infty$.
- Complexity: $\mathcal{O}(NLM)$ + complexity of $\eta_t(\cdot)$ per iteration

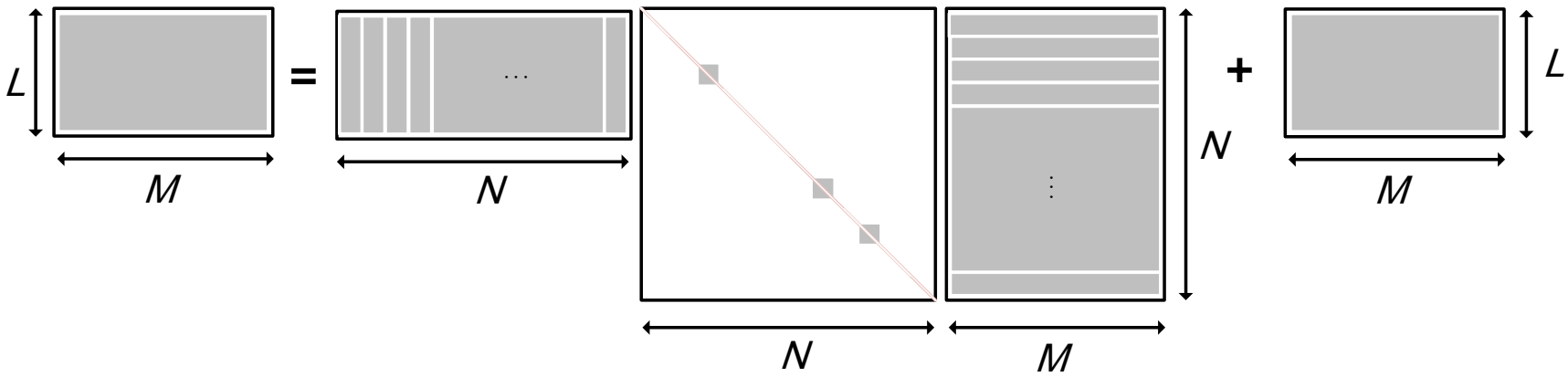
  But what if M is large? AMP becomes increasingly difficult to converge.

Joint Activity Detection and Large-Scale Fading Estimation

Key Assumption: We only need activity α_n and do not need $\mathbf{h}_n$.

Reformulate sparse activity detection as a large-scale-fading estimation problem:

$$\mathbf{Y} = \sum_{n=1}^{N} \alpha_n \mathbf{s}_n \mathbf{h}_n^T + \mathbf{Z} \triangleq \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{H}} + \mathbf{Z} \tag{5}$$

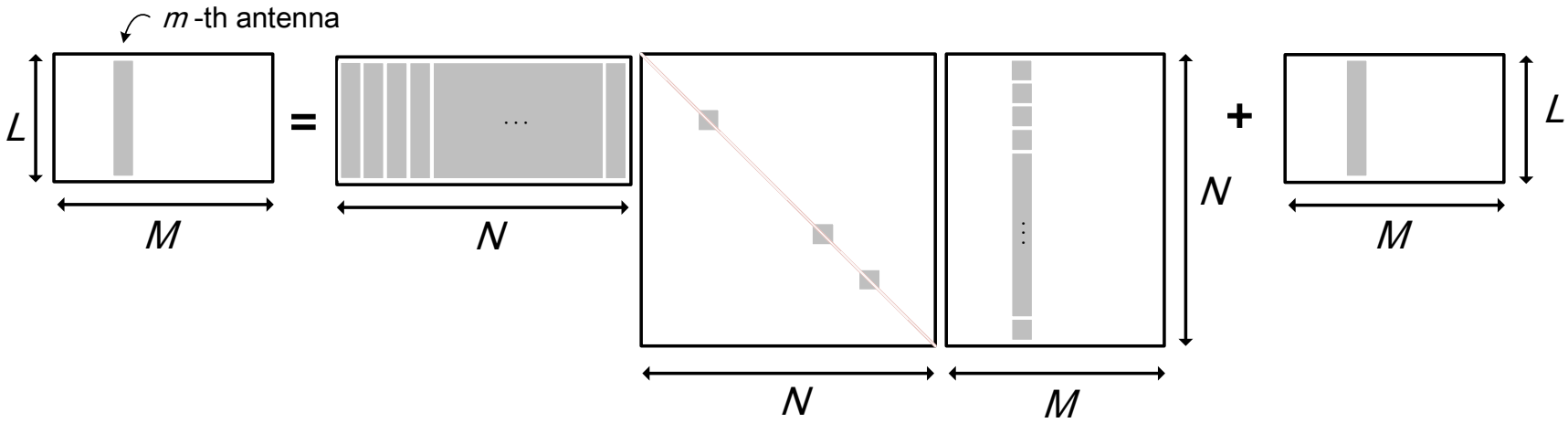


- $\mathbf{S} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N] \in \mathbb{C}^{L\times N}$, signature matrix
- $\boldsymbol{\Gamma} \triangleq \text{diag}\{\alpha_1\beta_1, \alpha_2\beta_2, \cdots, \alpha_N\beta_N\} \in \mathbb{R}^{N\times N}$, where β_n is large-scale fading
- $\tilde{\mathbf{H}} \triangleq \left[\mathbf{h}_1/\sqrt{\beta_1}, \mathbf{h}_2/\sqrt{\beta_2}, \cdots, \mathbf{h}_N/\sqrt{\beta_N}\right]^T \in \mathbb{C}^{N\times M}$, normalized channel matrix

Statistics of the Received Signal

Let $\mathbf{y}_m$ be the received signal at the m-th antenna, and let $\tilde{\mathbf{h}}_m$ be the normalized channel and $\mathbf{z}_m$ be the noise. Then, $\mathbf{y}_m$ can be expressed as

$$\mathbf{y}_m = \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}}\tilde{\mathbf{h}}_m + \mathbf{z}_m \tag{6}$$



- Model: Small-scale fading is i.i.d. Rayleigh across M received antennas.
- Then, $\tilde{\mathbf{h}}_m$ follows $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. Also, $\mathbf{z}_m$ follows $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$.
- Therefore, given $\boldsymbol{\Gamma}$, $\mathbf{y}_m$ is i.i.d. across m as $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}$.

Maximum Likelihood Estimation of $\mathbf{\Gamma}$

The sparse device activity is included in the diagonal matrix $\mathbf{\Gamma}$, which can be estimated using the maximum likelihood estimation (MLE) as:

$$\min_{\mathbf{\Gamma}\geq\mathbf{0}} \quad f(\mathbf{\Gamma}) := -\frac{1}{M}\log p(\mathbf{Y}|\mathbf{\Gamma}) \qquad \longleftarrow \text{ minimization of negative log-likelihood}$$

$$= -\frac{1}{M}\sum_{m=1}^{M}\log p(\mathbf{y}_m|\mathbf{\Gamma}) \qquad \longleftarrow \text{ i.i.d. over antennas}$$

$$= -\frac{1}{M}\sum_{m=1}^{M}\log\left(\frac{1}{|\pi\mathbf{\Sigma}|}\exp\left(-\mathbf{y}_m^H\mathbf{\Sigma}^{-1}\mathbf{y}_m\right)\right) \qquad \longleftarrow \text{ Gaussian distribution}$$

$$= -\frac{1}{M}\sum_{m=1}^{M}\log\left(\frac{1}{|\pi\mathbf{\Sigma}|}\right) - \frac{1}{M}\sum_{m=1}^{M}\log\left(\exp\left(-\mathbf{y}_m^H\mathbf{\Sigma}^{-1}\mathbf{y}_m\right)\right)$$

$$= \log|\mathbf{\Sigma}| + \frac{1}{M}\sum_{m=1}^{M}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{y}_m\mathbf{y}_m^H\right) + const. \qquad \longleftarrow \mathbf{x}^H\mathbf{A}\mathbf{x} = \mathrm{tr}\left(\mathbf{A}\mathbf{x}\mathbf{x}^H\right)$$

$$= \log|\mathbf{\Sigma}| + \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m^H\right) + const. \qquad (7)$$

Sample Covariance as a Sufficient Statistic

Define the sample covariance matrix of the received signal as

$$\hat{\mathbf{\Sigma}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}_m\mathbf{y}_m^H = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H. \tag{8}$$

With the sample covariance matrix, the MLE of $\mathbf{\Gamma}$ can be expressed as

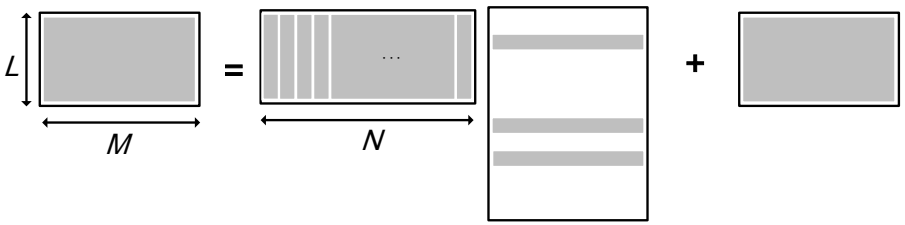
$$\begin{aligned}\min_{\mathbf{\Gamma}\succeq\mathbf{0}} \quad f(\mathbf{\Gamma}) &:= \log|\mathbf{\Sigma}| + \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right) + const. \\ &= \log|\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}| + \mathrm{tr}\left((\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}\right) + const. \end{aligned} \tag{9}$$

- $\hat{\mathbf{\Sigma}}$ is computed by averaging over different antennas, not time slots
- $\hat{\mathbf{\Sigma}}$ is a sufficient statistics since $f(\mathbf{\Gamma})$ depends on $\mathbf{Y}$ only through $\hat{\mathbf{\Sigma}}$
- The size of the MLE problem depends on N, L only, not M.

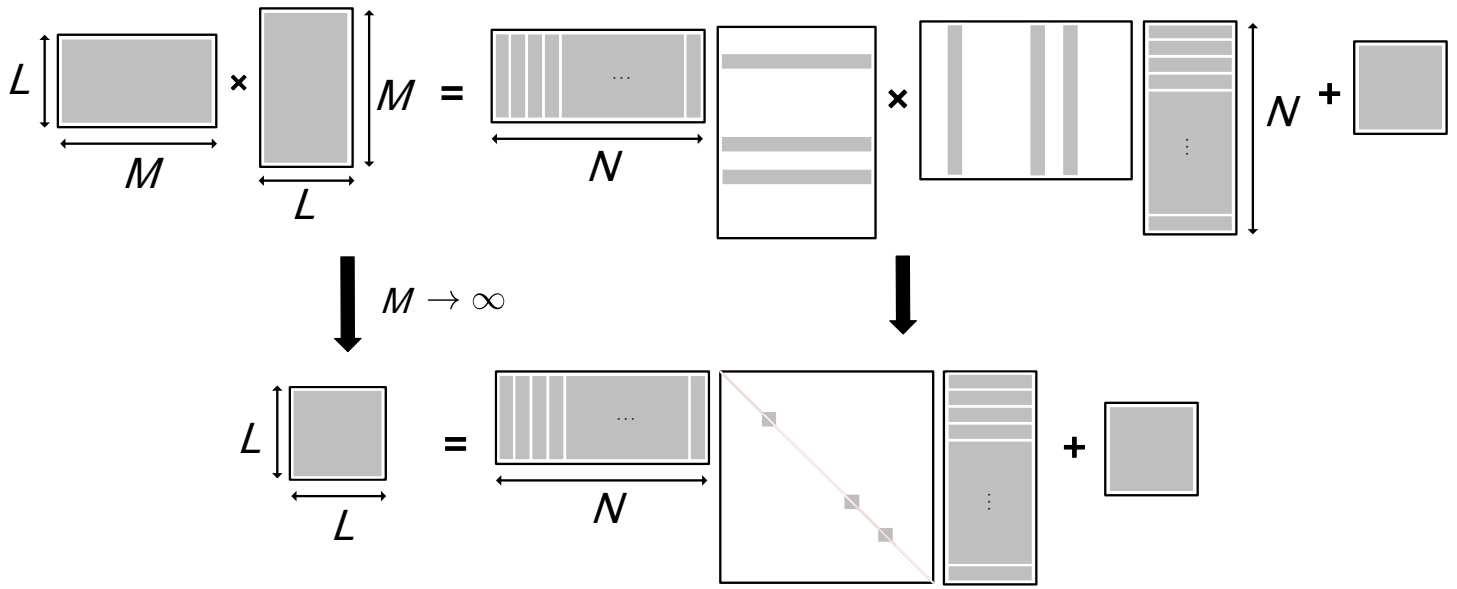
A. Fengler, S. Haghighatshoar, P. Jung, and G. Caire: "Non-Bayesian Activity Detection, Large-Scale Fading Coefficient Estimation, and Unsourced Random Access with a Massive MIMO Receiver", *IEEE Trans. Inf. Theory*, May 2021. http://arxiv.org/abs/1910.11266

Covariance Based Sparse Activity Detection

Instead of jointly estimating the channel, i.e., the non-zero rows in $\mathbf{X}$ based on $\mathbf{Y}$:



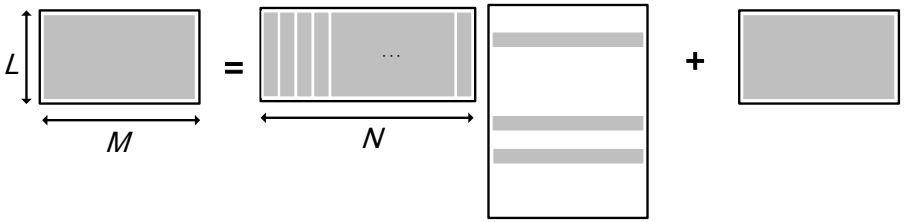
We now estimate large-scale fading $\mathbf{\Gamma}$ based on $\hat{\mathbf{\Sigma}} = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H$:



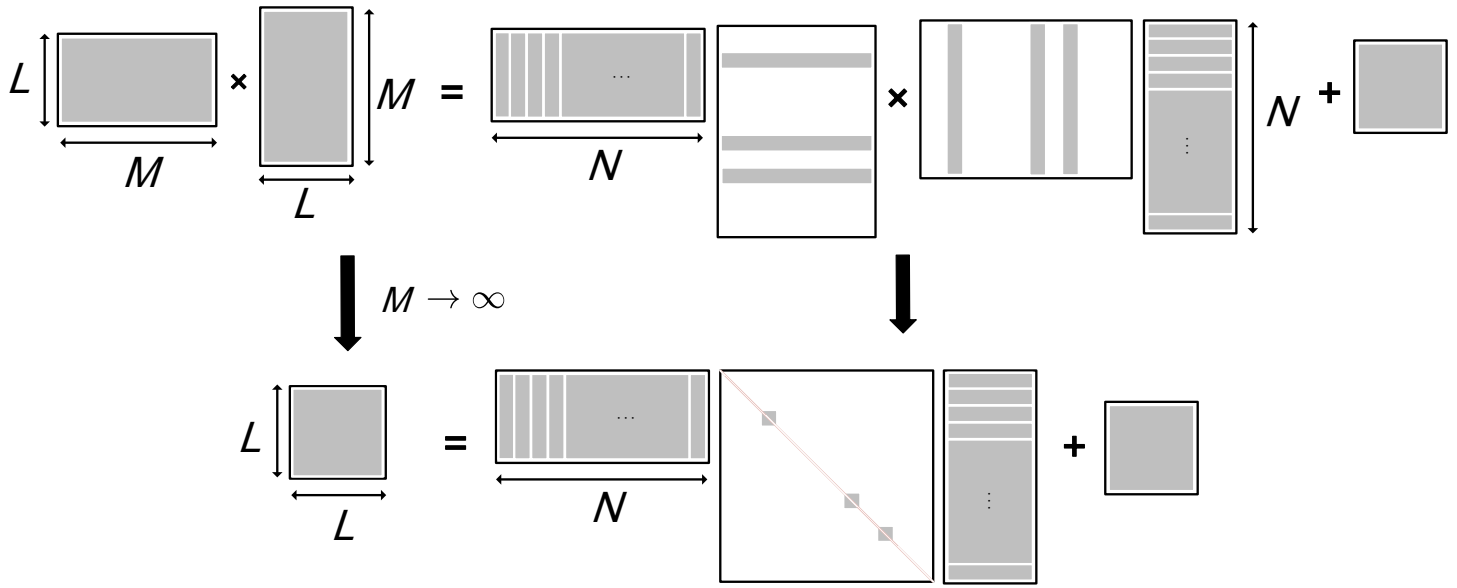
In the massive MIMO regime, i.e., if we let $M \to \infty$, this can be thought of detecting a diagonal sparse matrix from the sample covariance.

Covariance Based Sparse Activity Detection

Instead of jointly estimating the channel, i.e., the non-zero rows in $\mathbf{X}$ based on $\mathbf{Y}$:



We now estimate large-scale fading $\mathbf{\Gamma}$ based on $\hat{\mathbf{\Sigma}} = \frac{1}{M}\mathbf{Y}\mathbf{Y}^H$:



Crucial Advantage: Instead of detecting KM variables based on LM observations, we now detect K variables based on L^2 observations!

Covariance Based Sparse Activity Detection

To estimate $\mathbf{\Gamma}$, need to solve the optimization problem

$$\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad f(\mathbf{\Gamma}) := -\frac{1}{M} \log p(\mathbf{Y}|\mathbf{\Gamma})$$

$$= \log |\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I}| + \mathrm{tr}\left((\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\mathbf{\Sigma}} \right) + const. \qquad (10)$$

- $f(\mathbf{\Gamma})$ is non-convex (since it is concave function + convex function)
  - Expectation-maximization [Wipf-Rao '07] (Sparse Bayesian Learning)
  - Coordinate descent [Haghighatshoar-Jung-Caire '18]
- Observe: In the large M limit, $f(\mathbf{\Gamma})$ is minimized by the true value $\mathbf{\Gamma}^0$:

$$\hat{\mathbf{\Sigma}} \triangleq \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}_m \mathbf{y}_m^H \quad \to \quad \mathbf{\Sigma}^0 \triangleq \mathbf{S}\mathbf{\Gamma}^0\mathbf{S}^H + \sigma^2 \mathbf{I}, \qquad \text{as} \quad M \to \infty. \qquad (11)$$

Now consider the optimization (10) with $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}^0$, optimizing over $\mathbf{\Sigma}$ as in:

$$\min_{\mathbf{\Sigma}} \log |\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^0). \qquad (12)$$

By taking derivative, we see $\mathbf{\Sigma}^{\mathrm{opt}} = \mathbf{\Sigma}^0$. For finite M, we need to solve (10).

Coordinate Descent for Solving the MLE problem

Let γ_n be the n-th diagonal entry of $\mathbf{\Gamma}$. The MLE can be expressed as

$$\min_{\gamma_1,\ldots,\gamma_N \geq 0} \log \left| \sum_{n=1}^{N} \gamma_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I} \right| + \mathrm{tr} \left(\left(\sum_{n=1}^{N} \gamma_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I} \right)^{-1} \hat{\mathbf{\Sigma}} \right). \tag{13}$$

- **Basic Idea**: Update the coordinates $\gamma_1, \ldots, \gamma_N$ alternatively
- **Coordinate update**: Let $\hat{\gamma}_n, \forall n$ be the current estimates. Update $\hat{\gamma}_k$ with other $\hat{\gamma}_n, n \neq k$ fixed at a time. Let $\hat{\gamma}_k + d$ be the update. Determine d by

$$\min_{d \geq -\hat{\gamma}_k} \quad \log \left(1 + d \mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k \right) - \frac{d \mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{\Sigma}} \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}{1 + d \mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}. \tag{14}$$

  - $\tilde{\mathbf{\Sigma}} = \sum_{n=1}^{N} \hat{\gamma}_n \mathbf{s}_n \mathbf{s}_n^H + \sigma^2 \mathbf{I}$ is the current value of the covariance based on $\hat{\gamma}_n$.
  - The constraint $d \geq -\hat{\gamma}_k$ ensures the new $\hat{\gamma}_k + d$ is always non-negative.
  - By taking the derivative of the objective in (14), a closed-form solution is

$$d = \max \left\{ \frac{\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \hat{\mathbf{\Sigma}} \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k - \mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k}{(\mathbf{s}_k^H \tilde{\mathbf{\Sigma}}^{-1} \mathbf{s}_k)^2}, -\hat{\gamma}_k \right\}. \tag{15}$$

- **Advantages**: Efficient due to closed-form solution; empirically performs well.

Coordinate Descent for Device Activity Detection

We use two loops in the coordinate decent. The inner loop update all $\gamma_1, \ldots, \gamma_N$ in a random permuted order to ensure that each coordinate will be visited once.

Algorithm 1 Coordinate descent for device activity detection

Initialize $\hat{\boldsymbol{\gamma}} = \mathbf{0}, \tilde{\boldsymbol{\Sigma}} = \sigma^2 \mathbf{I}, \tilde{\boldsymbol{\Sigma}}^{-1} = \sigma^{-2}\mathbf{I}$.
for $i = 1, 2, \ldots, T$ **do**
  Select a random permutation $i_1, i_2, \ldots, i_N$ of indices $\{1, 2, \ldots, N\}$.
  for $n = 1$ to N **do**
    $$d = \max\left\{ \frac{\mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n} - \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n}}{(\mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n})^2}, -\hat{\gamma}_{i_n} \right\}$$
    $\hat{\gamma}_{i_n} = \hat{\gamma}_{i_n} + d$ ← Coordinate descent update
    $$\tilde{\boldsymbol{\Sigma}}^{-1} = \tilde{\boldsymbol{\Sigma}}^{-1} - d \frac{\tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n} \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1}}{1 + d \mathbf{s}_{i_n}^H \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{s}_{i_n}}$$ ← Rank-1 update of estimated covariance
  end for
end for
Output $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_N]^T$. Declare the activity with thresholding on $\hat{\boldsymbol{\gamma}}$.

Covariance Matching Approach

Recall that the MLE aims to recover $\mathbf{\Gamma}$ by solving the following problem

$$\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad f(\mathbf{\Gamma}) := -\frac{1}{M} \log p(\mathbf{Y}|\mathbf{\Gamma}) = \log |\mathbf{\Sigma}| + \text{tr}\left(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}\right) + const. \tag{16}$$

- The objective can be seen as the distance between $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma} = \mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}$ measured in the log-det Bregman matrix divergence.
- The MLE aims to match the sample covariance $\hat{\mathbf{\Sigma}}$ to the true covariance $\mathbf{\Sigma}$.

We can also use other distance metrics. With Frobenius norm as metric, we get

$$\min_{\mathbf{\Gamma} \geq \mathbf{0}} \quad \|\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} - \hat{\mathbf{\Sigma}}\|_F^2 \tag{17}$$

- This method is also known as non-negative least square (NNLS).
- The objective is convex. Coordinate descent can also be used to solve NNLS.
- A scaling law on N, L, K, and M has been established under NNLS.

MLE versus NNLS for Device Activity Detection

We compare the detection performance of MLE and NNLS via simulations.

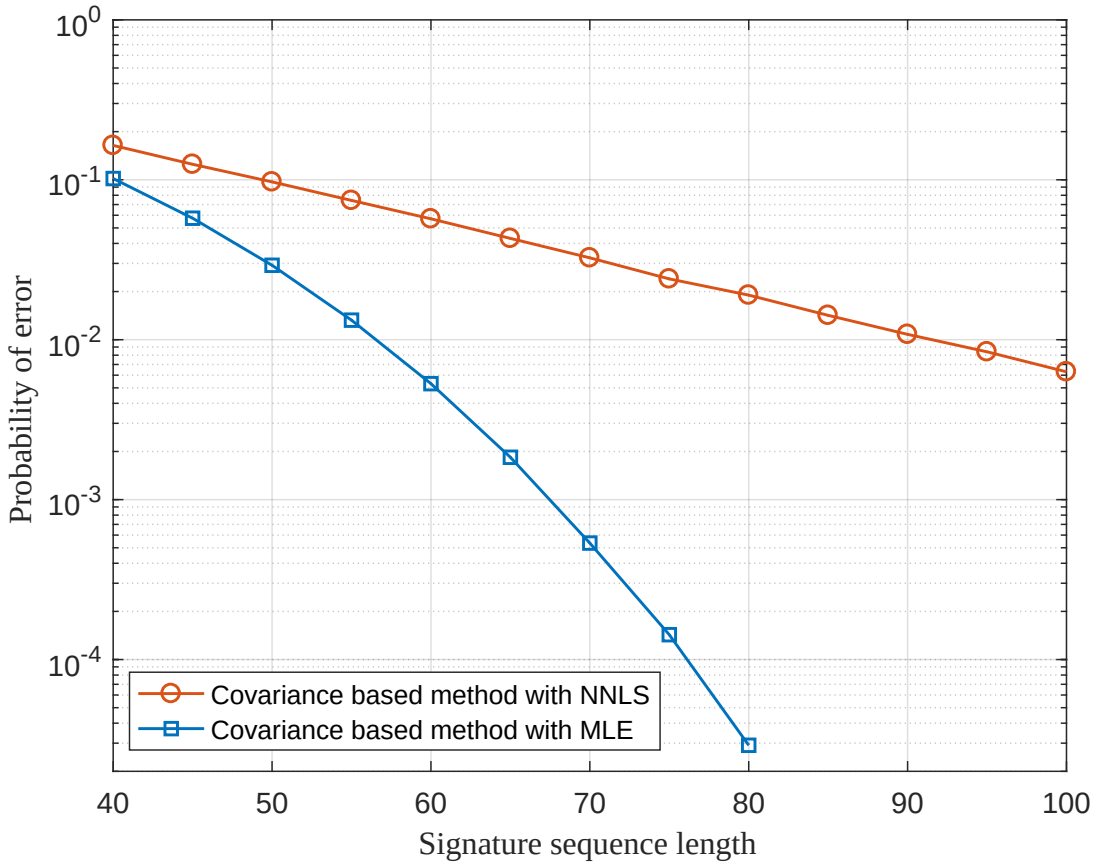


Figure: Performance comparison of MLE and NNLS. $N = 2000$, $K = 100$, and $M = 64$. MLE outperforms NNLS. The performance gap becomes more substantial as L increases.

Activity Detection with Known Large-Scale Fading

The covariance approach detects the device activity by estimating $\gamma_n \triangleq \alpha_n \beta_n$. There are scenarios in which the large-scale fading β_n is known at the BS, only the activities α_n need to be estimated.

- Maximizing the log-likelihood function of $\mathbf{Y}$ given $\alpha_1, \ldots, \alpha_N$ can be cast as

$$\begin{aligned}
\min_{\alpha_1,\ldots,\alpha_N} \quad f(\alpha_1, \ldots, \alpha_N) :&= -\frac{1}{M} \log p(\mathbf{Y}|\alpha_1, \ldots, \alpha_N) \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m|\alpha_1, \ldots, \alpha_N) \\
&= -\frac{1}{M} \sum_{m=1}^{M} \log \left(\frac{1}{|\pi \mathbf{\Sigma}|} \exp\left(-\mathbf{y}_m^H \mathbf{\Sigma}^{-1} \mathbf{y}_m\right) \right) \\
&= \log |\mathbf{\Sigma}| + \mathrm{tr}\left(\mathbf{\Sigma}^{-1} \hat{\mathbf{\Sigma}}\right) + const. \qquad (18)
\end{aligned}$$

Note that $p(\mathbf{y}_m|\alpha_1, \ldots, \alpha_N)$ remains Gaussian with covariance $\mathbf{\Sigma}$.

Activity Detection with Known Large-Scale Fading

- The problem of detecting the binary activity indicator α_n is now:

$$\min_{\{\alpha_n\}} \quad \log\left|\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right| + \mathrm{tr}\left((\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I})^{-1}\hat{\boldsymbol{\Sigma}}\right) \tag{19a}$$

$$\text{s.t.} \quad \alpha_n \in \{0,1\},\ n = 1,2,\ldots,N \tag{19b}$$

- Binary α_n is challenging to deal with. We relax the constraint such that

$$\alpha_n \in [0,1], \qquad n = 1,2,\ldots,N \tag{20}$$

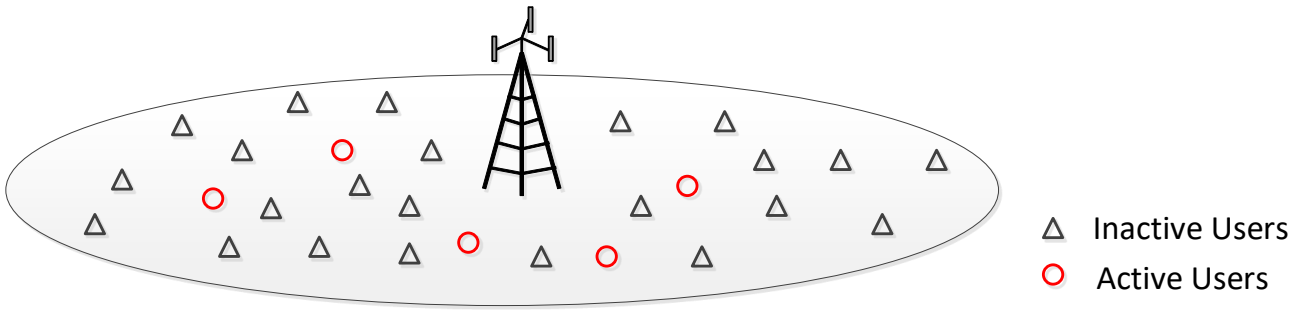
- The relaxed problem can be solved by coordinated descent with minor modifications:

$$d = \min\left\{\max\left\{\frac{\mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k - \mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k}{\beta_k(\mathbf{s}_k^H\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{s}_k)^2}, -\hat{\alpha}_k\right\}, 1-\hat{\alpha}_k\right\}. \tag{21}$$

- With unknown large-scale fading β_n, we estimate $\gamma_n = \alpha_n\beta_n$ in $[0,\infty]$. With known large-scale fading β_n, we estimate α_n in $[0,1]$.

Recap of Problem Formulations

- Sparse user activity detection with channel $\alpha_n \mathbf{h}_n \sim \alpha_n \sqrt{\beta}_n \mathcal{CN}(\mathbf{0}, \mathbf{I})$:



- If channel estimate is needed for subsequent data transmission:
  - We can use AMP, which gives a rough estimate of the instantaneous $\mathbf{h}_n$.
- If only user activities (α_n) are needed and large-scale fading is not known:
  - We can estimate large-scale fading ($\alpha_n \beta_n$) using the covariance method.
- If the users are not mobile and large-scale fading (β_n) is known:
  - We can modify the covariance method to estimate α_n.

Comparison of AMP vs Covariance Approaches

	Compressed Sensing (AMP)	Covariance Based Estimation
Derived from	Approx. marginals of $p(\mathbf{X}, \mathbf{Y})$	Maximization of $p(\mathbf{Y}\|\mathbf{\Gamma})$
Prior needed	Sparsity level for design of $\eta_t(\cdot)$	None (deterministic $\mathbf{\Gamma}$)
Estimate	Activities α_n and channels $\mathbf{h}_n$	Activities α_n and large-scale fading β_n
Preferred regime	Fix $\epsilon \triangleq \frac{K}{N}$, $\delta \triangleq \frac{L}{N}$, and M Let $N, L, K \to \infty$	Fix N, K, L Let $M \to \infty$
Complexity	Roughly $\mathcal{O}(NLM)$ per iteration	Roughly $\mathcal{O}(NL^2)$ via CD per iteration

Remark on Complexity

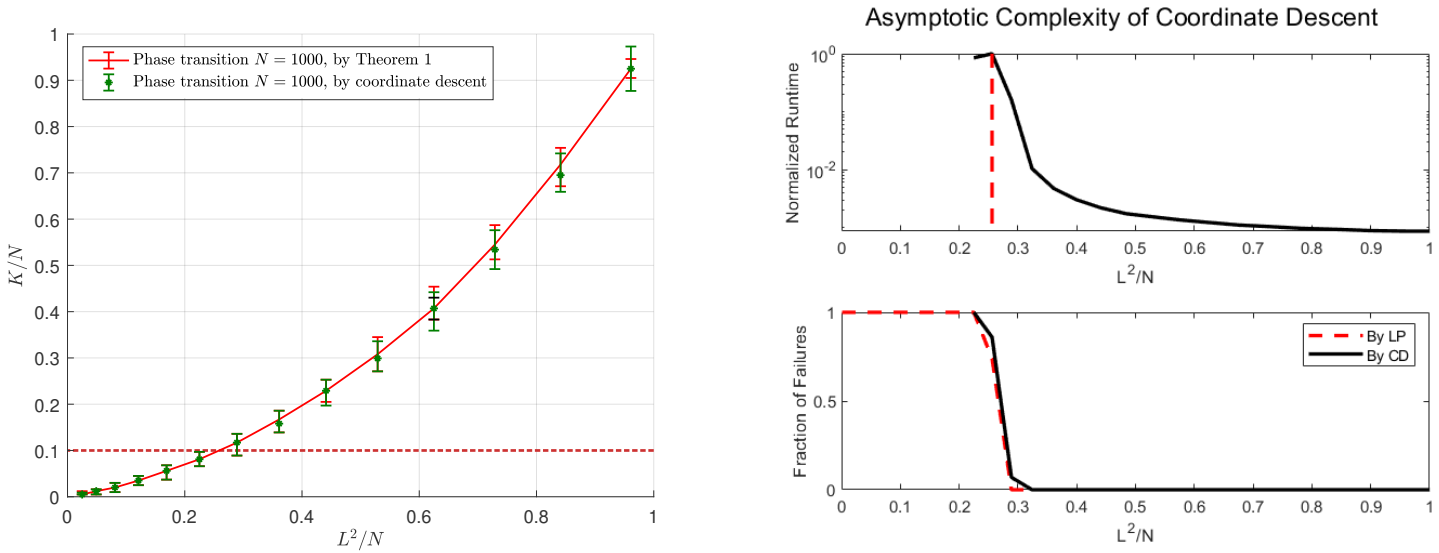
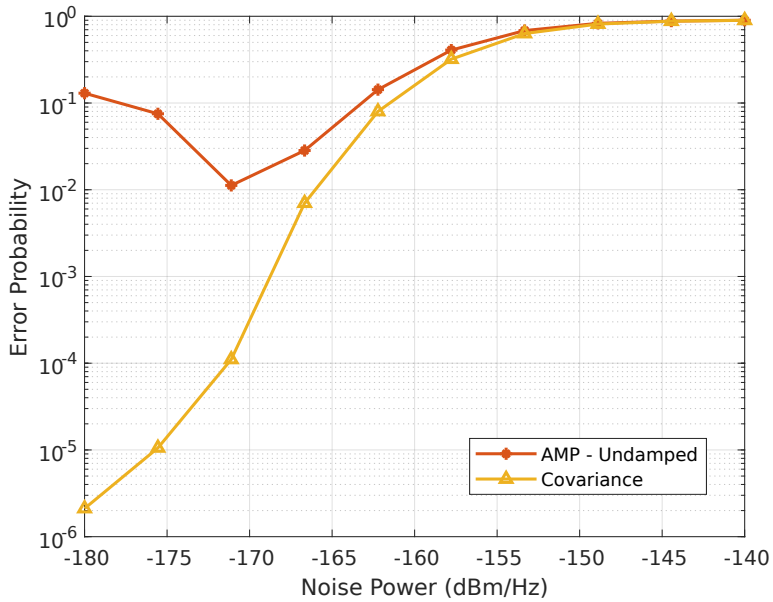


Figure: Numerical runtime for $\hat{\gamma}$ to converge to an ϵ norm ball around γ for fixed $\frac{K}{N} = 0.1$

- Each coordinate descent step requires $O(L^2)$ operations so updating all N coordinates resulting in a complexity of $O(L^2N)$ per iteration.
- The average number of iterations required to converge to solution increases as the operating point approaches the phase transition boundary.
- The complexity of each iteration grows as $O(L^2)$, but the complexity of the overall algorithm decreases with L.

Instability of AMP at High SNR

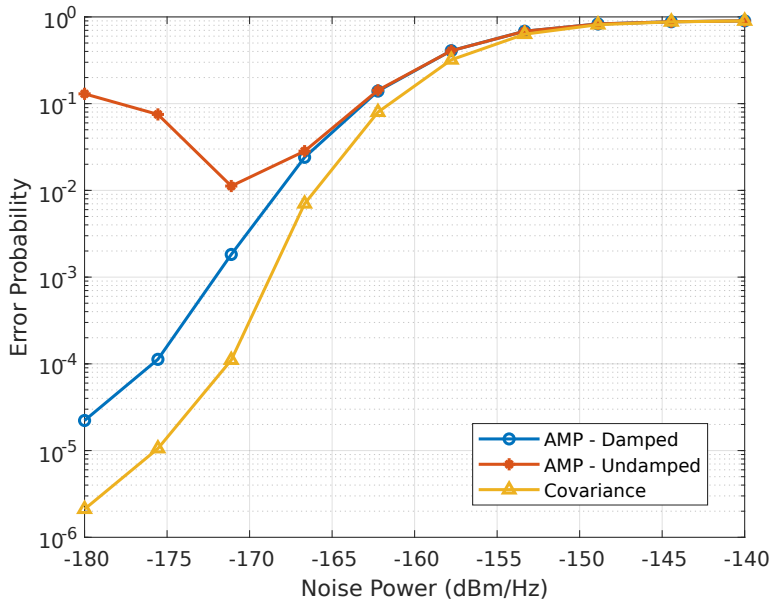
- Total users $N = 1000$, active users $K = 100$, BS antennas $M = 64$, $L = 110$.



- Users are uniformly distributed between 0.8 and 1km from the BS.
- User transmit power is 13dBm; Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$
- Error probability is the probability chosen such that $N \cdot P_{FA} \approx K \cdot P_{MD}$

Damping for AMP

- Total users $N = 1000$, active users $K = 100$, BS antennas $M = 64$, $L = 110$.



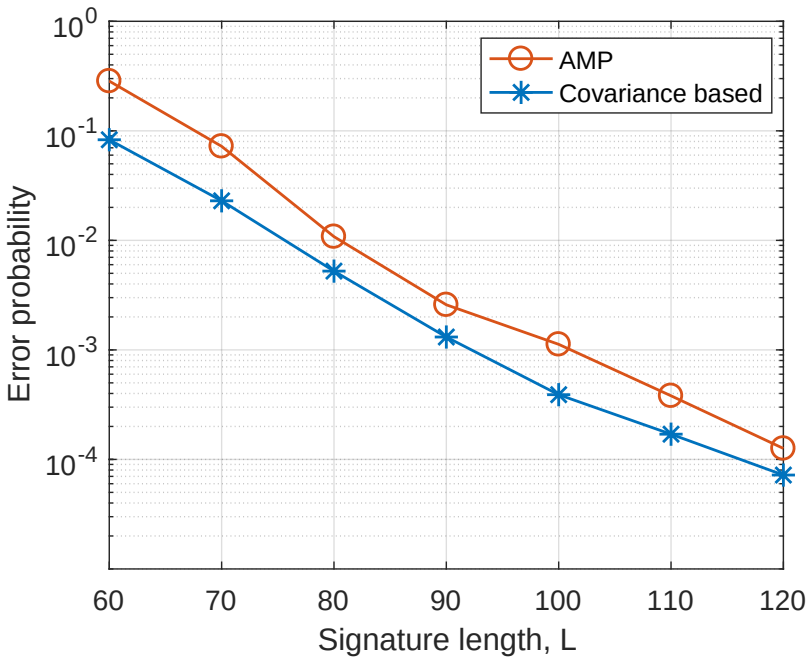
- Consider a *damping term* in the AMP update, improving stability and convergence, making AMP more effective with large M.

$$\mathbf{X}^{t+1} = (1-\alpha)\eta_t\left(\mathbf{S}^H\mathbf{R}^t + \mathbf{X}^t\right) + \alpha\mathbf{X}^t,$$

where $\alpha \in [0, 1]$ is a damping factor ($\alpha = 0$ for standard AMP).

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 50$, Number of antennas $M = 8$



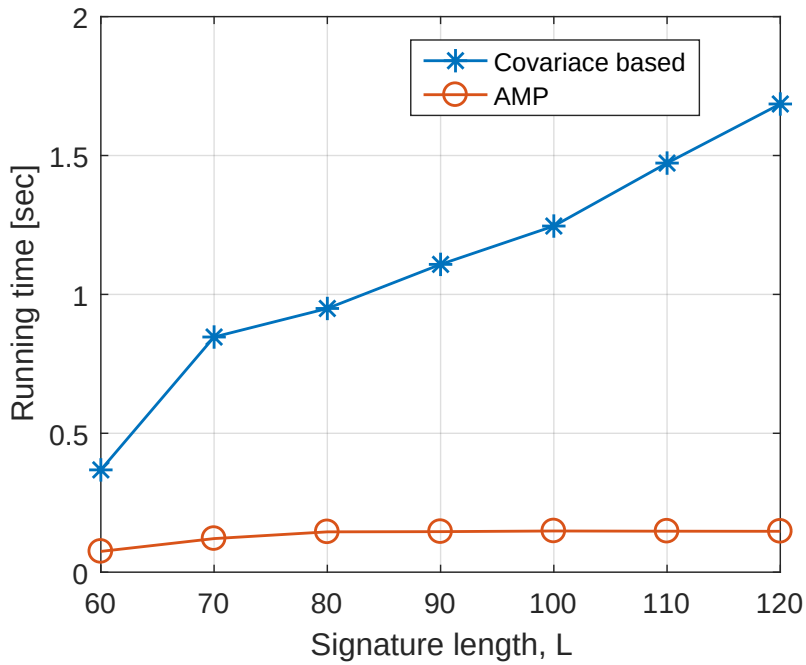
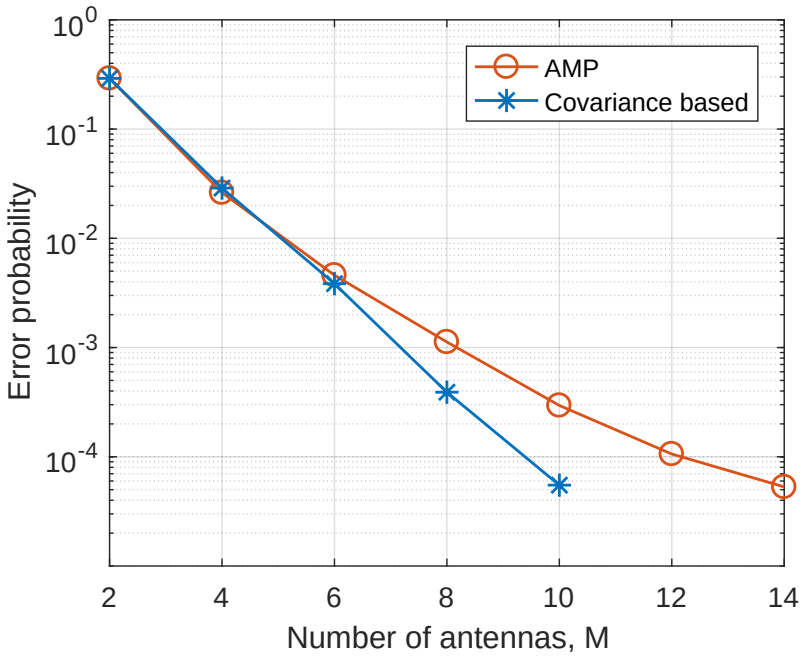


Figure: Performance and complexity of AMP vs covariance based estimation

- All users are located in the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 50$, Signature length $L = 100$



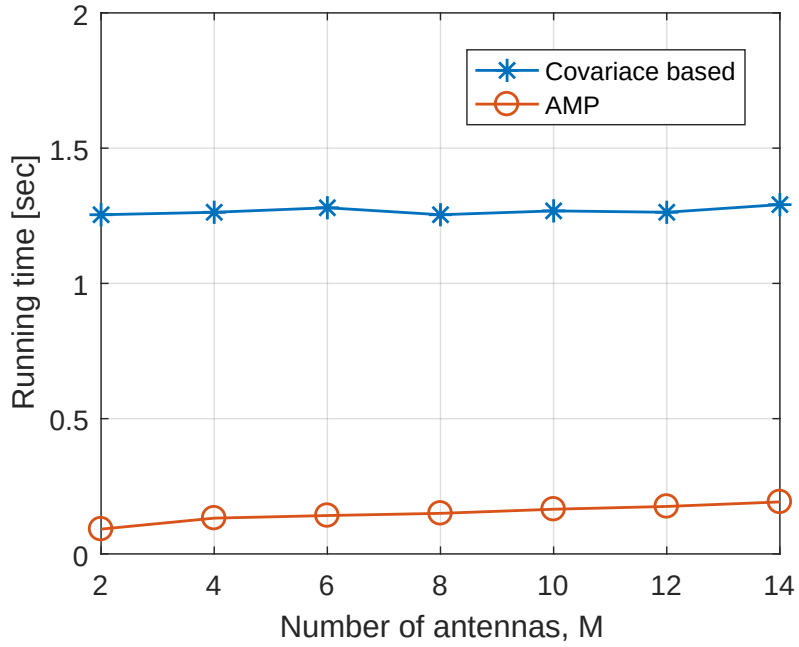
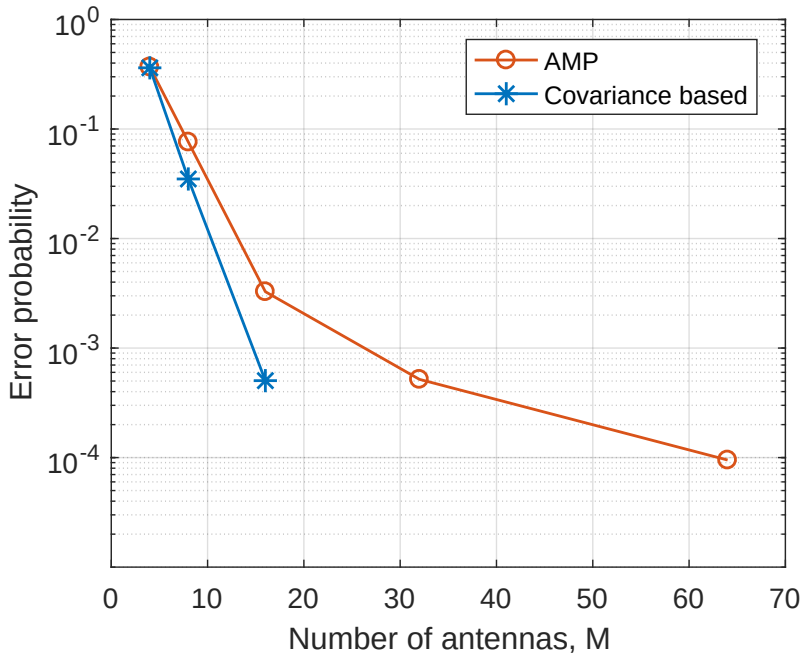


Figure: Performance and complexity of AMP vs covariance based estimation

- All users are located at the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

Numerical Comparison of AMP vs. Covariance Approach

- Total users $N = 1000$, Active users $K = 90$, Signature length $L = 100$



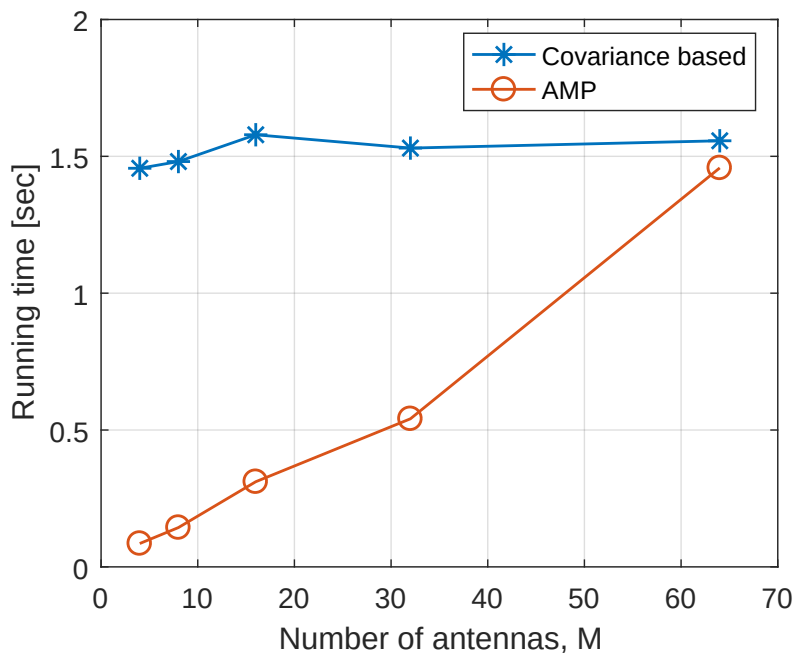


Figure: Performance and complexity of AMP vs covariance based estimation

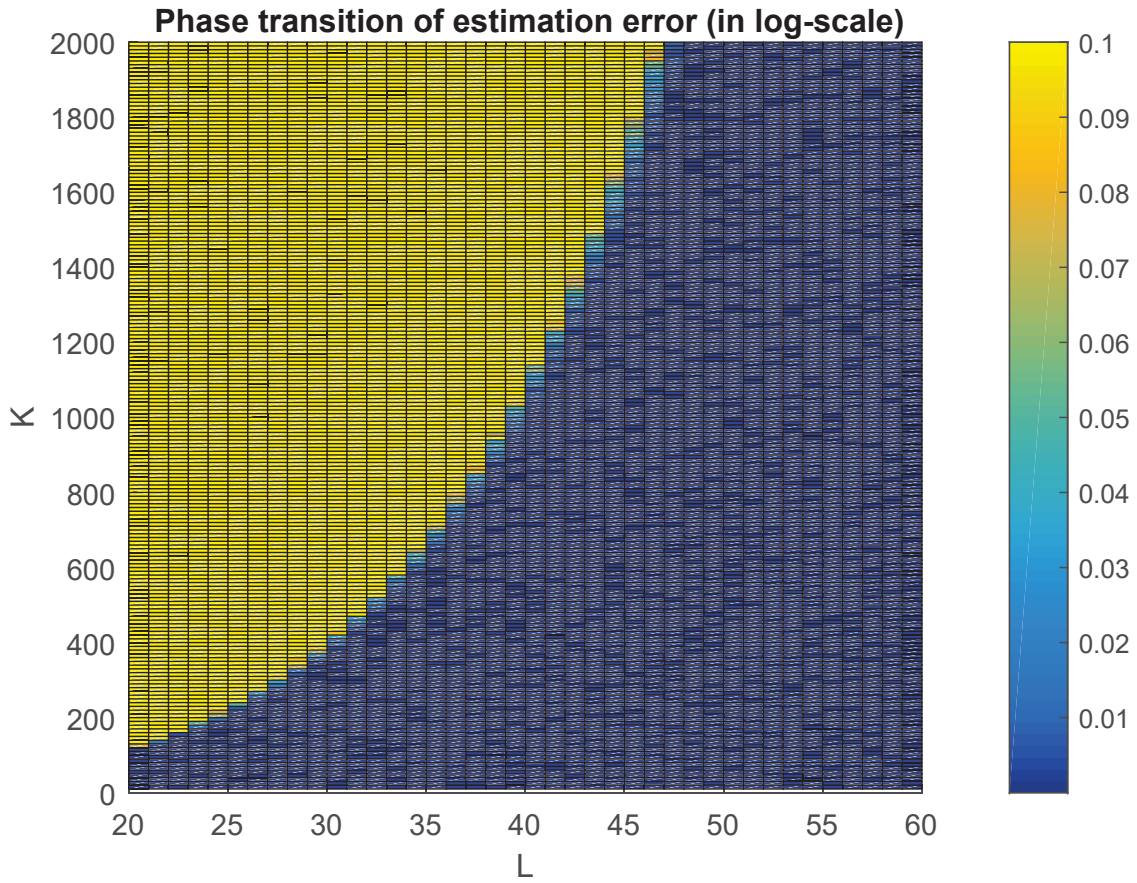
- All users are located at the cell-edge (1000m) with transmit power 23dBm.
- Path-loss model $128.1 + 37.6 \log(d[\text{in km}])$.
- Error probability is defined as the average of $\frac{\#(\text{Incorrectly detected users})}{K}$

AMP vs Covariance Approach

- Objectives:
  - Both algorithms perform sparse activity detection for massive random access.
  - AMP aims to recover the channels as well.
- Performance:
  - AMP and covariance approach have similar performance if $K \ll L$ and M small
  - Covariance approach is more effective in exploiting large M and when $K \gtrapprox L$.
- Complexity:
  - AMP is more computationally efficient when $K \ll L$ and M small.
- Crucial advantage of covariance method:
  - Being able to accommodate $K \gg L$ (!)

Scaling Law of the Covariance Approach

Suppose high SNR, perfect sampled covariance matrix $\hat{\mathbf{\Sigma}}$ ($M \to \infty$), we plot the estimation error of $\mathbf{\Gamma}$ under different (K, L) with $N = 2000$



Analysis

Analysis of AMP via State Evolution

The performance of AMP at each iteration can be predicted in the asymptotic regime where $L \to \infty, N \to \infty$ with fixed $\frac{L}{N}$

- $\mathbf{S}^H \mathbf{r}^t + \mathbf{x}^t$ can be modeled as signal plus noise, i.e., $\mathbf{x} + \mathbf{v}^t$
- $\mathbf{v}^t$ is i.i.d. Gaussian noise with variance τ_t tracked by state evolution equation

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta}\mathbb{E}|\eta_t(X + \tau_t Z) - X|^2 \tag{22}$$

  for the $M = 1$ case.

- Interpretation of state evolution: Vector estimation $\mathbf{y} = \mathbf{S}\mathbf{x} + \mathbf{z}$ is reduced to uncoupled scalar estimation $(\mathbf{x}^t + \mathbf{S}^H \mathbf{r}^t)_i = x_i + v_i^t$

Analysis of Covariance Approach via Fisher Info Matrix

Recall the MLE formulation, and let γ denote the diagonal entries of $\mathbf{\Gamma}$

$$\min_{\gamma \geq \mathbf{0}} \quad f(\gamma) := -\frac{1}{M} \log p(\mathbf{Y}|\gamma) = -\frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m|\gamma)$$

$$= \log |\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I}| + \text{tr}\left((\mathbf{S}\mathbf{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\mathbf{\Sigma}} \right) + const. \qquad (23)$$

- Analyzing the solution to (23) under coordinate descent is hard.
- Instead, let's analyze the true optimum of (23), i.e., MLE solution $\hat{\gamma}^{ML}$.
- Investigate asymptotic property of $\hat{\gamma}^{ML}$ in the massive MIMO regime.
- The Fisher information matrix, denoted by $\mathbf{J}(\gamma)$, plays a critical role in the asymptotic analysis. The (i,j)-th entry of $\mathbf{J}(\gamma)$ is defined as

$$[\mathbf{J}(\gamma)]_{ij} = \mathbb{E}\left[\frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_i} \frac{\partial \log p(\mathbf{Y}|\gamma)}{\partial \gamma_j} \right]. \qquad (24)$$

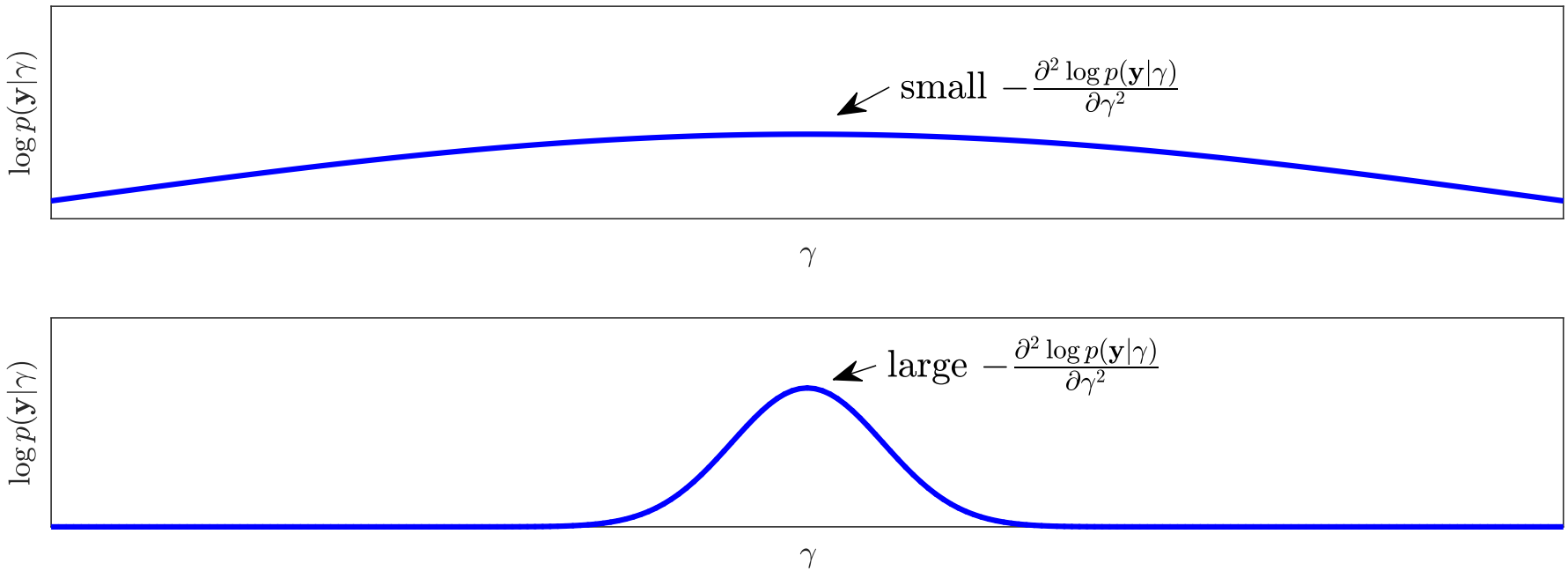
- Key assumption for the analysis: $M \to \infty$.

Fisher Information Matrix

- The Fisher Information matrix can be also written as the negative expected second derivative of the log-likelihood function

$$[\mathbf{J}(\boldsymbol{\gamma})]_{ij} = \mathbb{E}\left[\frac{\partial \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i}\frac{\partial \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_j}\right] = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\right] \tag{25}$$

- **Intuitive interpretation**: Fisher information matrix measures how informative the likelihood function is, and how effective the MLE can be



Cramer-Rao Bound and Asymptotic Property of MLE

Fisher information matrix plays a critical role in classic estimation theory.

- **Cramer-Rao bound**: Let γ be a parameter, and let $\hat{\gamma}$ be an unbiased estimator of γ. Then the covariance of estimation error is lower bounded by

$$\mathbb{E}\left[(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)^T\right] \geq \mathbf{J}^{-1}(\gamma) \tag{26}$$

- **Asymptotic properties of the MLE**: Let $\hat{\gamma}^{ML}$ be the maximum likelihood estimator of γ. Then, under certain regularity conditions, as $M \to \infty$

$$\text{Consistency:} \quad \hat{\gamma}^{ML} \xrightarrow{P} \gamma \tag{27}$$

$$\text{Asymptotic normality:} \quad \sqrt{M}(\hat{\gamma}^{ML}-\gamma) \xrightarrow{D} \mathcal{N}(\mathbf{0}, M\mathbf{J}^{-1}(\gamma)) \tag{28}$$

It means that the maximum likelihood estimator $\hat{\gamma}^{ML}$ is asymptotically unbiased and asymptotically attains the Cramer-Rao bound, i.e., asymptotically efficient.

Regularity Conditions

- The regularity conditions for consistency and asymptotic normality include
  - The true parameter γ^0 is identifiable, i.e.,, there exists no other $\gamma' \neq \gamma^0$ with

$$p(\mathbf{Y}|\gamma^0) = p(\mathbf{Y}|\gamma'). \tag{29}$$

  - The true parameter should be in the interior of the feasible region, as otherwise $\hat{\gamma}^{ML} - \gamma^0$ cannot be Gaussian distributed.
- These conditions are usually reasonable and mild.
- But, these conditions are NOT always satisfied for sparse activity detection.
  - The identifiability may not be guaranteed when

$$N \gg L^2, \tag{30}$$

    i.e., when the dimension of γ^0 is larger than the dimensions of the sample covariance $\hat{\mathbf{\Sigma}}$, there are too many parameters to estimate.
  - The true parameter γ^0 in fact always lies on the boundary of its parameter space $[0, \infty)^N$, because most of the entries of γ^0 are zero.

Need new analysis!

Fisher Information Matrix in Sparse Activity Detection

We first derive the Fisher information matrix for the activity detection problem:

$$[\mathbf{J}(\boldsymbol{\gamma})]_{ij} = -\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\right]. \tag{31}$$

- $p(\mathbf{y}_m|\boldsymbol{\gamma})$ is Gaussian, the second derivative of $\log p(\mathbf{Y}|\boldsymbol{\gamma}) = \sum_m \log p(\mathbf{y}_m|\boldsymbol{\gamma})$ is

$$\begin{aligned}\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j} &= M \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\boldsymbol{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H) - M \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\boldsymbol{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}) \\ &\quad - M \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{s}_i\mathbf{s}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{s}_j\mathbf{s}_j^H\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}).\end{aligned} \tag{32}$$

- Taking the expectation of $\hat{\boldsymbol{\Sigma}}$ using the fact that $\mathbb{E}[\hat{\boldsymbol{\Sigma}}] = \boldsymbol{\Sigma}$ gives

$$-\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\gamma})}{\partial \gamma_i \partial \gamma_j}\right] = M(\mathbf{s}_i^H\boldsymbol{\Sigma}^{-1}\mathbf{s}_j)(\mathbf{s}_j^H\boldsymbol{\Sigma}^{-1}\mathbf{s}_i). \tag{33}$$

Fisher Information Matrix in Sparse Activity Detection

- The Fisher information matrix $\mathbf{J}(\boldsymbol{\gamma})$ can be further written in a matrix form as

$$\mathbf{J}(\boldsymbol{\gamma}) = M\left(\mathbf{P} \odot \mathbf{P}^*\right), \tag{34}$$

  where $\mathbf{P} \triangleq \mathbf{S}^H \left(\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I}\right)^{-1}\mathbf{S}$; $\odot$ element-wise product; $(\cdot)^*$ conjugate
- $\mathbf{J}(\boldsymbol{\gamma})$ is a real symmetric matrix of dimensions $N \times N$, whose rank satisfies:

$$\text{Rank}(\mathbf{P} \odot \mathbf{P}^*) \overset{(a)}{\leq} \text{Rank}(\mathbf{P})^2 \overset{(b)}{\leq} L^2, \tag{35}$$

  where
  - (a) is due to $\text{Rank}(\mathbf{U} \odot \mathbf{V}) \leq \text{Rank}(\mathbf{U})\,\text{Rank}(\mathbf{V})$;
  - (b) is due to $\text{Rank}(\mathbf{P}) \leq \text{Rank}(\mathbf{S}) \leq \min\{N, L\}$.
- Thus $\mathbf{J}(\boldsymbol{\gamma})$ is rank-deficient if $N > L^2$ since $\mathbf{P} \odot \mathbf{P}^*$ is of size $N \times N$.

Our new analysis takes rank-deficiency of $\mathbf{J}(\boldsymbol{\gamma})$ into consideration

Performance Analysis of Activity Detection

Since the regularity conditions may not hold in the sparse activity detection problem, we need to ask:

- What are the conditions on the system parameters such that $\hat{\boldsymbol{\gamma}}^{ML}$ can approach the true parameter $\boldsymbol{\gamma}^0$ as $M \to \infty$?
- This helps identify the desired operating regime of the system parameters for getting an accurate estimate $\hat{\boldsymbol{\gamma}}^{ML}$ via MLE with massive MIMO

- If M is finite, how is the estimation error $\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0$ distributed?
- This helps characterize the error probabilities in device activity detection.

We answer these questions by examining the "null space" of the Fisher information matrix.

Necessary and Sufficient Condition for $\hat{\gamma}^{ML} \to \gamma^0$

Theorem

Let $\mathcal{I}$ be an index set corresponding to zero entries of γ^0, i.e., $\mathcal{I} \triangleq \{i \mid \gamma_i^0 = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

$$\mathcal{N} \triangleq \{\mathbf{x} \mid \mathbf{x}^T \mathbf{J}(\gamma^0)\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N\}, \tag{36}$$

$$\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}, \tag{37}$$

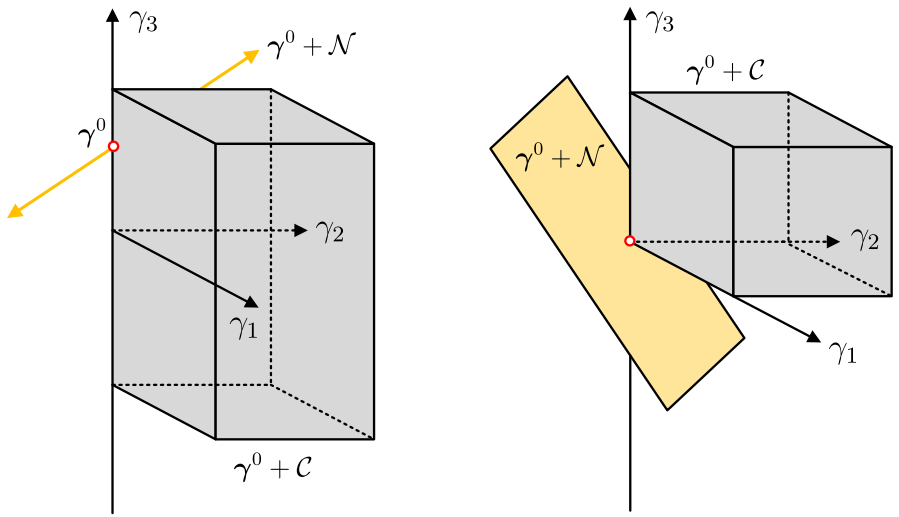
where x_i is the i-th entry of $\mathbf{x}$. Then a necessary and sufficient condition for the consistency of $\hat{\gamma}^{ML}$, i.e., $\hat{\gamma}^{ML} \to \gamma^0$ as $M \to \infty$, is $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

$\mathcal{N}$ is the "null space" of $\mathbf{J}(\gamma^0)$; $\mathcal{C}$ is a cone with non-negative entries indexed by $\mathcal{I}$.

This condition leads to a phase analysis for the covariance based method, i.e., set of (N, L, K) outside of which $\hat{\gamma}^{ML}$ cannot approach γ^0 even in the large M limit.

Interpretation of the Condition

- $\mathcal{N}$ corresponds to all directions in $\mathbb{R}^N$ along which likelihood stays constant. In these directions, the true parameter cannot be identified via the likelihood.
- $\mathcal{C}$ is the directions along which parameters remain within constraint $\mathbb{R}_+^N$
- $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ ensures that the true parameter $\boldsymbol{\gamma}^0$ is uniquely identifiable via the likelihood in its feasible neighborhood, also termed as local identifiability



- Local identifiability is clearly necessary.
- Sufficiency due to equivalence of local and global identifiability in this case.
- A necessary condition for $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ is that $dim(\mathcal{N}) < |\mathcal{I}|$.
- Since $dim(\mathcal{N})$ is roughly $N - L^2$ and $|\mathcal{I}| = N - K$, we have $K < L^2$.

Numerically Verify the Condition via $\mathcal{M}^+$ Criterion

Proposition

Let $\mathcal{I} \triangleq \{i \mid \gamma_i^0 = 0\}$ *and* $\mathcal{I}^c \triangleq \{i \mid \gamma_i^0 > 0\}$ *be two index sets with* $|\mathcal{I}| = N - K$ *and* $|\mathcal{I}^c| = K$*. We define three submatrices of* $\mathbf{J}(\boldsymbol{\gamma}^0) \in \mathbb{R}^{N \times N}$ *as follows.*

- $\mathbf{A} \in \mathbb{R}^{(N-K)\times(N-K)}$*, row indices and column indices from* $\mathcal{I}$
- $\mathbf{B} \in \mathbb{R}^{(N-K)\times K}$*, row indices from* $\mathcal{I}$ *and column indices from* $\mathcal{I}^c$
- $\mathbf{C} \in \mathbb{R}^{K\times K}$*, row indices and column indices from* $\mathcal{I}^c$

If $\mathbf{C}$ *is invertible, then the condition* $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to the feasibility of*

$$\text{find} \quad \mathbf{x} \tag{38a}$$

$$\text{subject to} \quad (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)\mathbf{x} > \mathbf{0}, \tag{38b}$$

where vector $\mathbf{x} \in \mathbb{R}^{N-K}$*.*

Note that matrix $\mathbf{M}$ satisfying $\mathbf{M}^T\mathbf{x} > \mathbf{0}$ for some $\mathbf{x}$, i.e., row span intersecting the positive orthant, is referred to as $\mathcal{M}^+$ [Bruckstein-Elad-Zibulevsky'08].

Proof based on analyzing the null space of $\mathbf{J}(\boldsymbol{\gamma}^0)$ and that $\forall \mathbf{M}$: (i) $\mathbf{M}\mathbf{x} = \mathbf{0}$ has no solution for $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{x} \neq \mathbf{0}$, is equivalent to (ii) $\mathbf{M}^T\mathbf{x} > \mathbf{0}$ has solutions.

Covariance Matching Perspective

Analyzing the optimization problem:

$$\min_{\boldsymbol{\gamma} \succeq \mathbf{0}} \quad f(\boldsymbol{\gamma}) = \log |\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I}| + \mathrm{tr}\left((\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\boldsymbol{\Sigma}} \right) \tag{39}$$

- By taking the derivative, we see that ideally we need: $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I} = \hat{\boldsymbol{\Sigma}}$
- For finite M, usually $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I} \neq \hat{\boldsymbol{\Sigma}}$ since $\hat{\boldsymbol{\Sigma}}$ is the *sample* covariance
- For $M \to \infty$, $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I} = \hat{\boldsymbol{\Sigma}}$ holds at true $\boldsymbol{\gamma}^0$, i.e., $\boldsymbol{\gamma}^0$ minimizes $f(\boldsymbol{\gamma})$.

Intuition

$\hat{\boldsymbol{\gamma}}^{ML} \to \boldsymbol{\gamma}^0$ *as* $M \to \infty$ $\iff$ $\boldsymbol{\gamma}^0$ *uniquely minimizes* $f(\boldsymbol{\gamma})$ *in the limit* $M \to \infty$

$\iff$ $\boldsymbol{\gamma}^0$ *is the unique solution to* $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I} = \hat{\boldsymbol{\Sigma}}$ *in the limit* $M \to \infty$.

A necessary and sufficient condition for the consistency of $\hat{\boldsymbol{\gamma}}^{ML}$ can be derived by studying the uniqueness of $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2 \mathbf{I} = \hat{\boldsymbol{\Sigma}}$ in the limit $M \to \infty$.

Equivalent Necessary and Sufficient Condition

Proposition

Let $\widehat{\mathbf{S}} \in \mathbb{C}^{L^2 \times N}$ *be the column-wise Kronecker product (Khatri-Rao product) of* $\mathbf{S}^*$ *and* $\mathbf{S}$*, i.e.,* $\widehat{\mathbf{S}} \triangleq [\mathbf{s}_1^* \otimes \mathbf{s}_1, \ldots, \mathbf{s}_N^* \otimes \mathbf{s}_N]$*. We define a set* $\widetilde{\mathcal{N}}$ *in the space* $\mathbb{R}^N$ *as*

$$\widetilde{\mathcal{N}} \triangleq \{\mathbf{x} \mid \widehat{\mathbf{S}}\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbb{R}^N\}. \tag{40}$$

Then a necessary and sufficient condition for $\boldsymbol{\gamma}^0$ *being the unique solution to* $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} = \hat{\boldsymbol{\Sigma}}$ *in the limit* $M \to \infty$*, is* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$*, where* $\mathcal{C}$ *is as in* (37)*.*

The proof is obtained by vectorizing $\mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^H + \sigma^2\mathbf{I} = \hat{\boldsymbol{\Sigma}}$ in the limit $M \to \infty$, and studying the resulting linear equation.

Proposition

We have that $\widetilde{\mathcal{N}}$ *defined in* (40) *and* $\mathcal{N}$ *defined in* (36) *are identical. Thus, the condition* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to* $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$

Key advantage of using $\widetilde{\mathcal{N}}$ is that it depends only on S and is independent of SNR.

Alternative Way to Numerically Verify the Condition

Theorem

Let $\mathbf{r}_i^T = [s_{i1}, s_{i2}, \ldots, s_{iN}]$ *be the* i*-th row of* $\mathbf{S}$*. Based on* $\mathbf{r}_i^T$*, we construct two sets of row vectors to represent the real and imaginary parts of rows of* $\widehat{\mathbf{S}}$*:*

$$\left\{\mathrm{Re}(\mathbf{r}_i^T) \odot \mathrm{Re}(\mathbf{r}_j^T) + \mathrm{Im}(\mathbf{r}_i^T) \odot \mathrm{Im}(\mathbf{r}_j^T),\ 1 \le i \le j \le L\right\}, \tag{41}$$

$$\left\{\mathrm{Re}(\mathbf{r}_i^T) \odot \mathrm{Im}(\mathbf{r}_j^T) - \mathrm{Im}(\mathbf{r}_i^T) \odot \mathrm{Re}(\mathbf{r}_j^T),\ 1 \le i < j \le L\right\}. \tag{42}$$

Let $\mathbf{D} \in \mathbb{R}^{L^2 \times N}$ *be the matrix formed by all* L^2 *rows from these two sets. The condition* $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ *is equivalent to the infeasibility of the following problem*

$$\begin{aligned} \text{find} \quad & \mathbf{x} && \text{(43a)} \\ \text{subject to} \quad & \mathbf{D}\mathbf{x} = \mathbf{0}, && \text{(43b)} \\ & \|\mathbf{x}\|_1 = 1, && \text{(43c)} \\ & x_i \ge 0, i \in \mathcal{I}, && \text{(43d)} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^N$ *and the constraint* (43c) *guarantees* $\mathbf{x} \neq \mathbf{0}$*.*

Scaling Law for NNLS Formulation

Theorem (Fengler-Haghighatshoar-Jung-Caire '19)

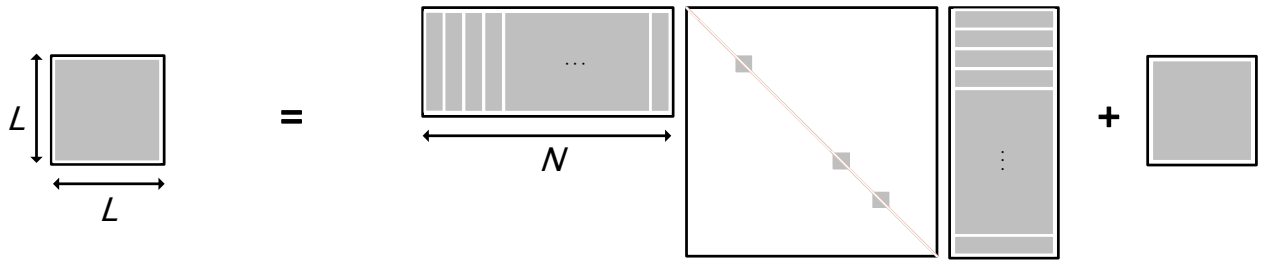
Let $\mathbf{S} \in \mathbb{C}^{L \times N}$ *be the signature sequence matrix whose columns are uniformly drawn from the sphere of radius* $\sqrt{L}$ *in an i.i.d. fashion. There exist some constants* c_1, c_2, c_3, *and* c_4 *whose values do not depend on* K, L, *and* N *such that if* $K \leq c_1 L^2 / \log^2(eN/L^2)$, *then with probability at least* $1 - \exp(-c_2 L)$, *the solution of the NNLS problem,* $\hat{\gamma}^{NNLS}$, *satisfies*

$$\|\gamma^0 - \hat{\gamma}^{NNLS}\|_2 \leq c_3 \left(\sqrt{\frac{L}{K}} + c_4 \right) \frac{\|\mathbf{S}\boldsymbol{\Gamma}^0\mathbf{S}^H + \sigma^2\mathbf{I} - \hat{\boldsymbol{\Sigma}}\|_F}{L}. \tag{44}$$

- The derivation is based on restricted isometry property in compressed sensing.
- It implies that the error vanishes as $M \to \infty$, because $\hat{\boldsymbol{\Sigma}} \to \mathbf{S}\boldsymbol{\Gamma}^0\mathbf{S}^H + \sigma^2\mathbf{I}$.
- The result is for specific sequence $\mathbf{S}$.
- Since $K < L^2$, we get a simpler form of scaling law: $L^2 \approx K \log^2(N/K)$.

Scaling Law for MLE Formulation

- Scaling laws in compressed sensing:
  - For $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A}$ satisfying restricted isometry property, the number of measurements needed to recover a K-sparse vector $\mathbf{x}$ of length-N is
  $$L = O(K \log(N/K)).$$
  - For $\hat{\mathbf{\Sigma}} = \mathbf{S\Gamma S}^H + \sigma^2\mathbf{I}$ with $\widehat{\mathbf{S}}$ satisfying robust null space property, the number of measurements needed to receover a K-sparse diagonal matrix $\mathbf{\Gamma}$ of size N^2 is
  $$L^2 = O(K \log^2(N/K)).$$
- Based on the same robust NSP of $\widehat{\mathbf{S}}$, we can derive the scaling law of MLE:



Theorem

Under the same scaling law for K, L, N and for the same randomly chosen $\mathbf{S}$, $\widetilde{\mathcal{N}} \cap \mathcal{C} = \{\mathbf{0}\}$ holds with probability at least $1 - \exp(-c_2 L)$.

Numerical Results – Scaling Law of Covariance Approach

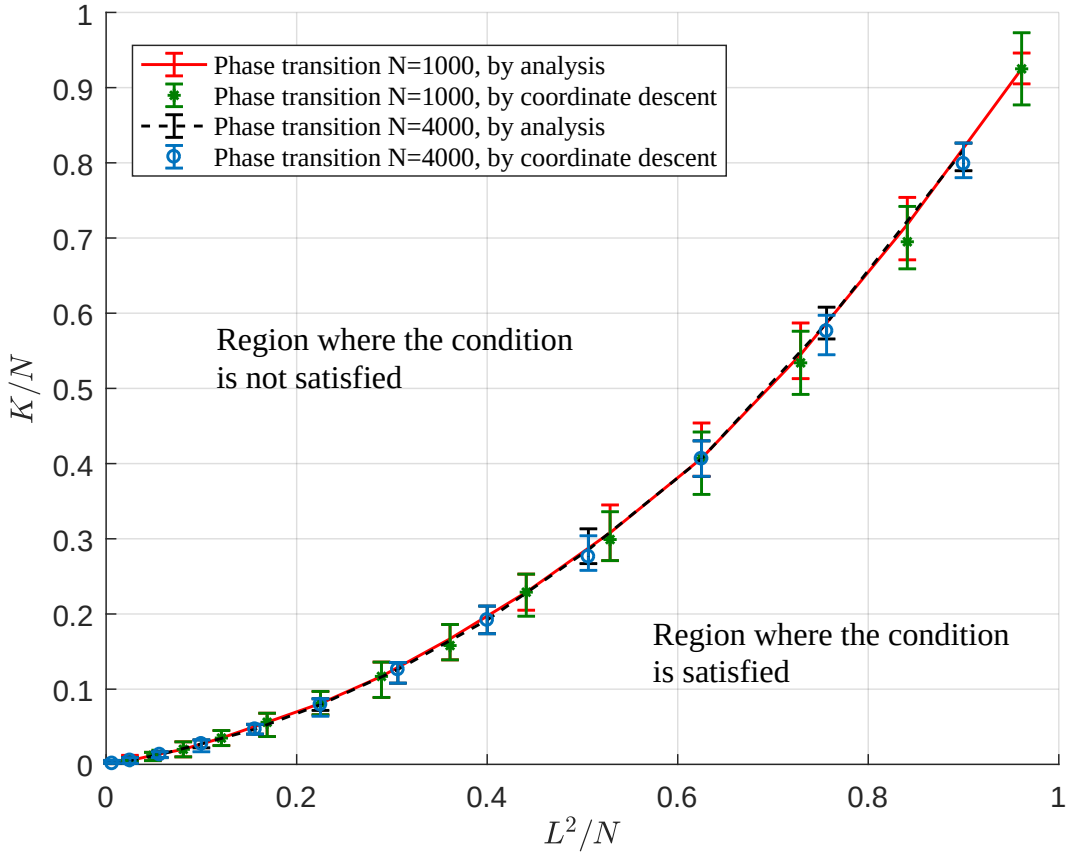
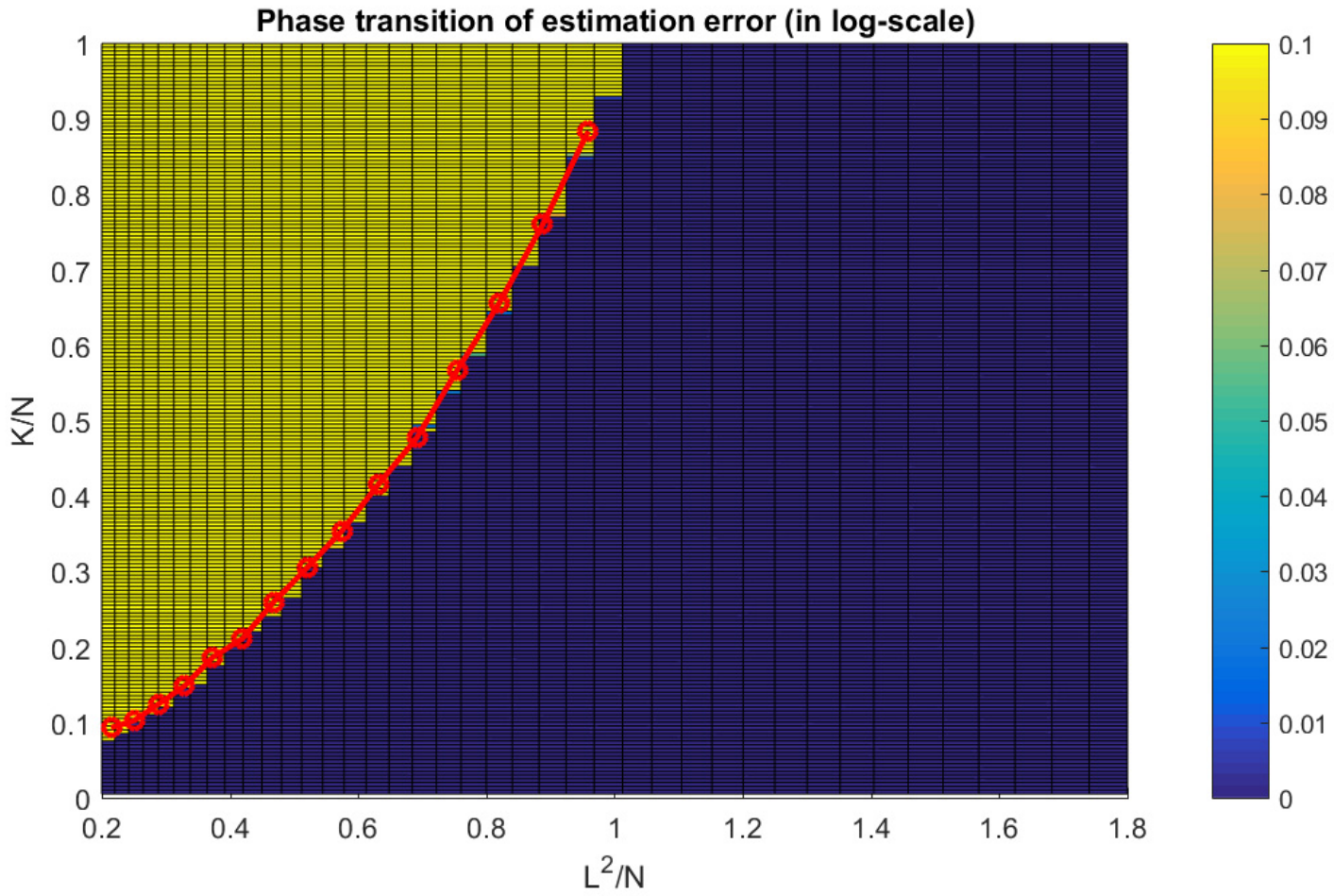


Figure: Phase transition in the space of N, L, K. All users are located at the cell-edge (1000m) with transmit power 23dBm. Path-loss is $128.1 + 37.6 \log(d[\text{km}])$. Generated by 100 Monte Carlo simulations. Error bars indicate the range below which all 100 realizations satisfy the condition and above which none satisfies the condition.

Phase Transition of the Covariance Approach

Suppose high SNR, perfect sampled covariance matrix $\hat{\mathbf{\Sigma}}$ ($M \to \infty$), we plot the estimation error of $\mathbf{\Gamma}$ under different (K/N, L^2/N) with $N = 2000$



Performance of coordinate descent algorithm is very close to the optimal MLE!

Value of Knowing Large-Scale Fading

The covariance method directly estimates the activity indicator α_n in $[0, 1]$ instead of $\gamma_n = \alpha_n \beta_n$ in $[0, \infty)$. Let $\boldsymbol{\alpha} \triangleq [\alpha_1, \ldots, \alpha_N]^T$ and let the true value of $\boldsymbol{\alpha}$ be $\boldsymbol{\alpha}^0$.

Theorem

Let $\mathcal{I}$ be an index set corresponding to zero entries of $\boldsymbol{\alpha}^0$, i.e., $\mathcal{I} \triangleq \{i \mid \alpha_i^0 = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

$$\mathcal{N} \triangleq \{\mathbf{x} \mid \mathbf{x}^T \mathbf{J}(\boldsymbol{\gamma}^0)\mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N\}, \tag{45}$$

$$\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, x_i \leq 0, i \notin \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}, \tag{46}$$

where x_i is the i-th entry of $\mathbf{x}$. Then a necessary and sufficient condition for the consistency of $\hat{\boldsymbol{\alpha}}^{ML}$, i.e., $\hat{\boldsymbol{\alpha}}^{ML} \to \boldsymbol{\alpha}^0$ as $M \to \infty$, is $\mathcal{N} \cap \mathcal{C} = \{\mathbf{0}\}$.

The extra constraint in defining $\mathcal{C}$ is due to the fact that α_n is upper bounded.

Value of Knowing Large-Scale Fading

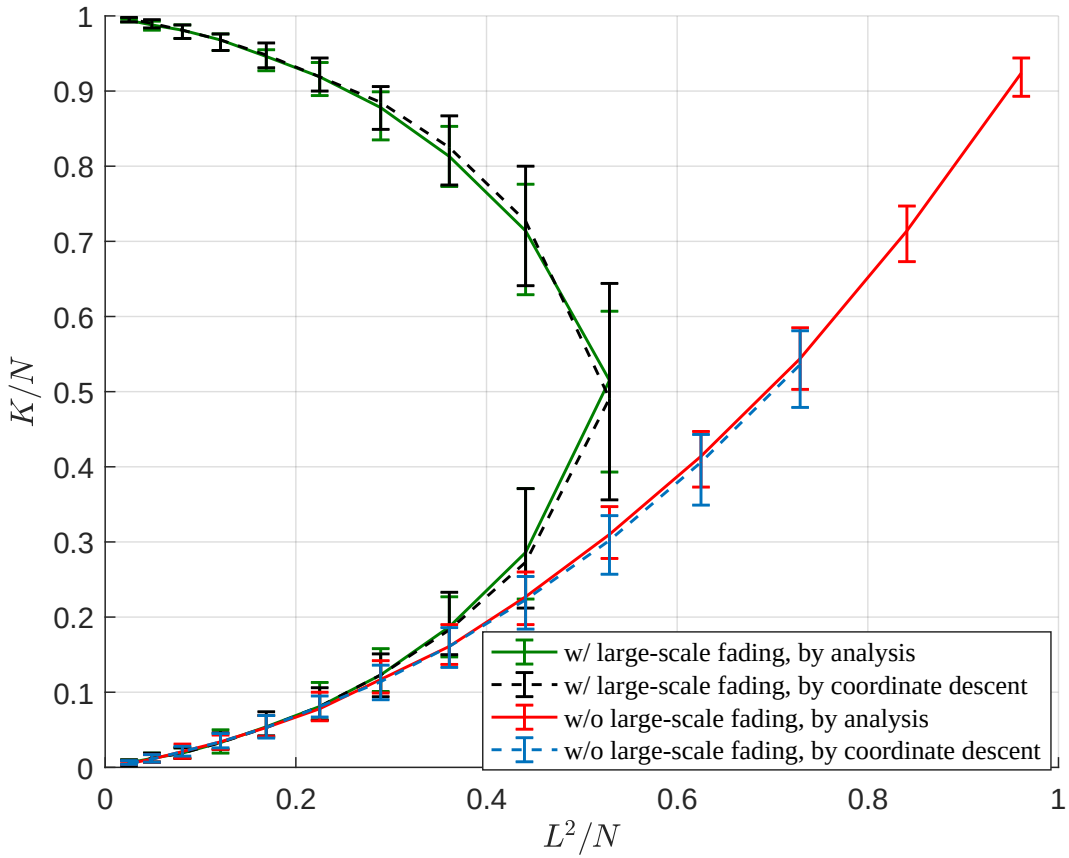


Figure: Phase transition comparison of the cases with and without knowing large-scale fading. $N = 1000$. With known large-scale fading, α_n is both lower and upper bounded.

When $\frac{K}{N} \approx 1$, then inactive users are sparse!

Asymptotic Distribution of the ML Estimation Error

For MLE solutions not on boundary, we have $\sqrt{M}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}) \xrightarrow{D} \mathcal{N}(\mathbf{0}, M\mathbf{J}^{-1}(\boldsymbol{\gamma}))$.
For MLE with boundary constraint: $\mathcal{C} \triangleq \{\mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, \mathbf{x} \in \mathbb{R}^N\}$:

Theorem

Let $\mathbf{x} \in \mathbb{R}^{N\times 1} \sim \mathcal{N}\left(\mathbf{0}, M\mathbf{J}^{\dagger}(\boldsymbol{\gamma}^0)\right)$, where $\dagger$ denotes Moore-Penrose inverse. Let $\boldsymbol{\mu} \in \mathbb{R}^{N\times 1}$ be a solution to the constrained quadratic programming problem

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \frac{1}{M}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{J}(\boldsymbol{\gamma}^0)(\mathbf{x}-\boldsymbol{\mu}) \tag{47a}$$

$$\text{subject to} \quad \boldsymbol{\mu} \in \mathcal{C}, \tag{47b}$$

where $\mathcal{C}$ is defined in (37). For the case without knowing large-scale fading, assume that $\hat{\boldsymbol{\gamma}}^{ML} \to \boldsymbol{\gamma}^0$, then there exists a sequence of $\boldsymbol{\mu}$ such that $M^{\frac{1}{2}}(\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma})$ has asymptotically the same distribution as $\boldsymbol{\mu}$ as $M \to \infty$.

Note that $\boldsymbol{\mu}$ is random due to the randomness of $\mathbf{x}$.

Detection error can be characterized based on the distribution of $\hat{\boldsymbol{\gamma}}^{ML} - \boldsymbol{\gamma}^0$.

Distribution of Estimation Error

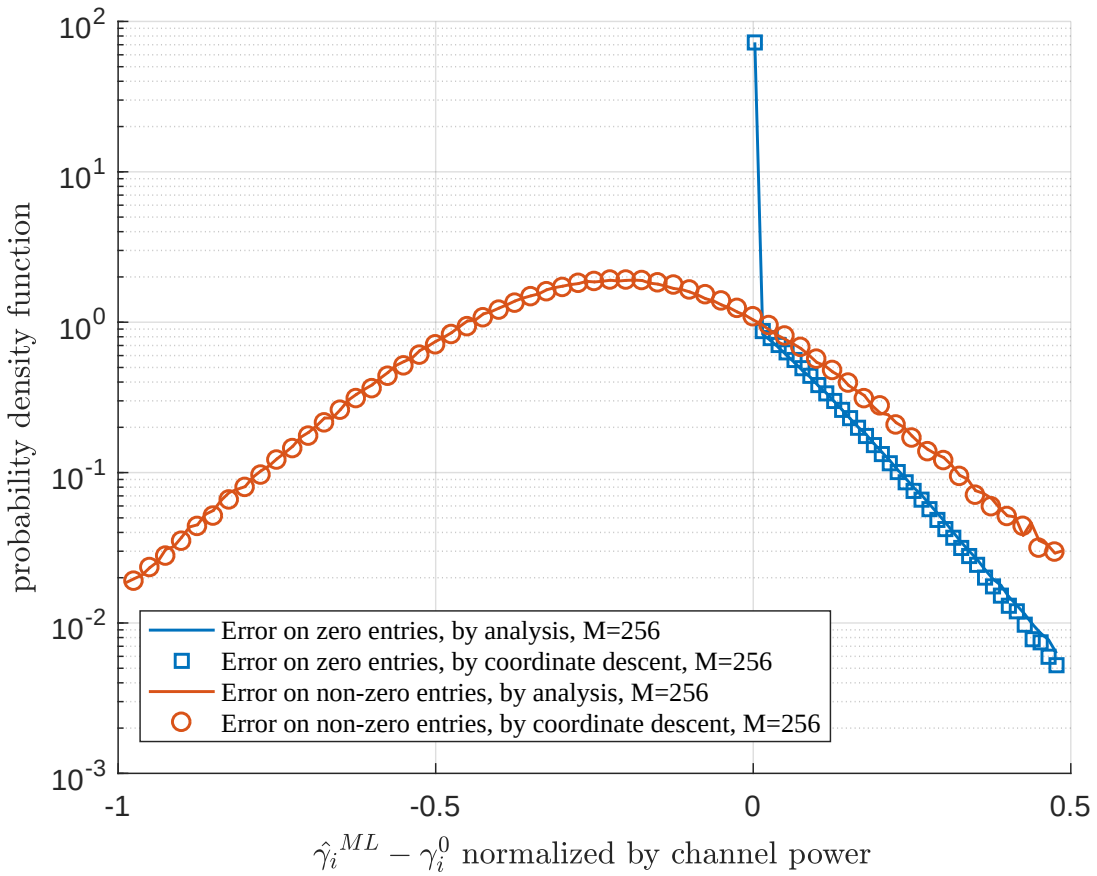


Figure: Probability density functions (PDFs) of the error $\hat{\gamma}_i^{ML} - \gamma_i^0$ (normalized). The parameters are $N = 1000$, $K = 50$, and $L = 20$ ($L^2/N = 0.4, K/N = 0.05$). Note that there is a point mass in the distribution of the error for the zero entries. This is the probability that the inactive devices are correctly identified at finite $M = 256$.

Detection Error Probabilities

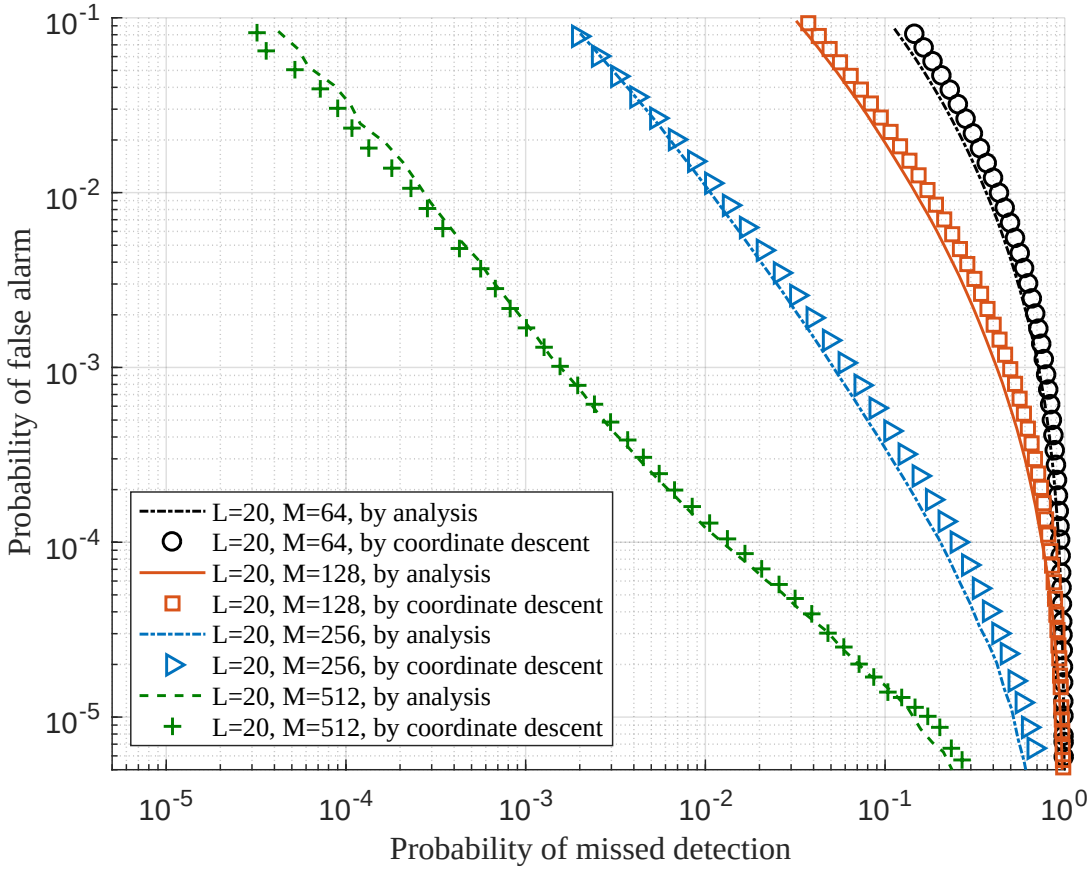
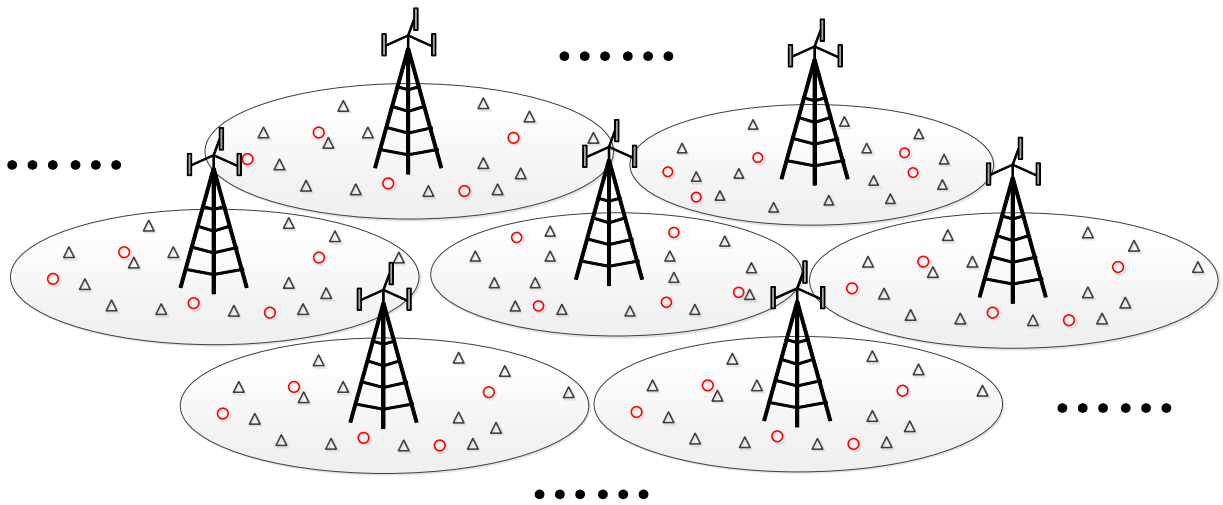


Figure: Probability of missed detection vs. probability of false alarm. The parameters are $N = 1000$, $K = 50$, and $L = 20$ ($L^2/N = 0.4, K/N = 0.05$). All users are located at the cell-edge (1000m) with transmit power 23dBm. Path-loss is $128.1 + 37.6 \log(d[\text{km}])$.

User Activity Detection in Multicell Systems

- What is the impact of the inter-cell interference?



- How to overcome the inter-cell interference?

Activity Detection in Multicell Systems

- Multi-cell system with B BSs each equipped with M antennas;
- N single-antenna devices per cell, K of which are active;
- Device n in cell b is assigned a length-L unique signature sequence $\mathbf{s}_{bn}$;
- Received signal $\mathbf{Y}_b \in \mathbb{C}^{L\times M}$ at BS b is

$$\begin{aligned}\mathbf{Y}_b &= \sum_{n=1}^{N} \alpha_{bn}\mathbf{s}_{bn}\mathbf{h}_{bbn}^T + \sum_{j=1,j\neq b}^{j=B}\sum_{n=1}^{N} \alpha_{jn}\mathbf{s}_{jn}\mathbf{h}_{bjn}^T + \mathbf{Z}_b \\ &= \mathbf{S}_b\mathbf{X}_{bb} + \sum_{j=1,j\neq b}^{j=B} \mathbf{S}_j\mathbf{X}_{bj} + \mathbf{Z}_b, \end{aligned} \tag{48}$$

where

- $\alpha_{bn} \in \{1, 0\}$ activity indicator; $\mathbf{Z}_b \in \mathbb{C}^{L\times M}$ Gaussian noise with variance σ^2.
- $\mathbf{h}_{bjn} \in \mathbb{C}^{M\times 1}$ is the channel from user n in cell j to BS b
- $\mathbf{S}_j \triangleq [\mathbf{s}_{j1}, \cdots, \mathbf{s}_{jN}] \in \mathbb{C}^{L\times N}$; $\mathbf{X}_{bj} \triangleq [\alpha_{j1}\mathbf{h}_{bj1}, \cdots, \alpha_{jN}\mathbf{h}_{bjN}]^T \in \mathbb{C}^{N\times M}$

The inter-cell interference brings performance degradation for activity detection.

Covariance Based Activity Detection for Multi-cell

To use the covariance approach, the signal at BS b is re-written as

$$\begin{aligned}\mathbf{Y}_b &= \sum_{n=1}^{N} \alpha_{bn}\mathbf{s}_{bn}\mathbf{h}_{bbn}^T + \sum_{j=1,j\neq b}^{j=B}\sum_{n=1}^{N} \alpha_{jn}\mathbf{s}_{jn}\mathbf{h}_{bjn}^T + \mathbf{Z}_b \\ &= \mathbf{S}_b\mathbf{A}_b\mathbf{G}_{bb}^{\frac{1}{2}}\tilde{\mathbf{H}}_{bb} + \sum_{j=1,j\neq b}^{j=B} \mathbf{S}_j\mathbf{A}_j\mathbf{G}_{bj}^{\frac{1}{2}}\tilde{\mathbf{H}}_{bj} + \mathbf{Z}_b \\ &= \mathbf{S}_b\mathbf{\Gamma}_{bb}^{\frac{1}{2}}\tilde{\mathbf{H}}_{bb} + \sum_{j=1,j\neq b}^{j=B} \mathbf{S}_j\mathbf{\Gamma}_{bj}^{\frac{1}{2}}\tilde{\mathbf{H}}_{bj} + \mathbf{Z}_b \end{aligned} \tag{49}$$

- $\mathbf{S}_j \triangleq [\mathbf{s}_{j1}, \mathbf{s}_{j2}, \cdots, \mathbf{s}_{jN}] \in \mathbb{C}^{L\times N}$; $\mathbf{A}_j \triangleq \text{diag}\{\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jN}\} \in \{0,1\}^{N\times N}$
- $\mathbf{G}_{bj} \triangleq \text{diag}\{\beta_{bj1}, \beta_{bj2}, \cdots, \beta_{bjN}\} \in \mathbb{R}^{N\times N}$ large-scale fading matrix
- $\mathbf{\Gamma}_{bj} \triangleq \text{diag}\{\alpha_{j1}\beta_{bj1}, \alpha_{j2}\beta_{bj2}, \cdots, \alpha_{jN}\beta_{bjN}\} \in \mathbb{R}^{N\times N}$
- $\tilde{\mathbf{H}}_{bj} \triangleq \left[\mathbf{h}_{bj1}/\sqrt{\beta_{bj1}}, \cdots, \mathbf{h}_{bjN}/\sqrt{\beta_{bjN}}\right]^T \in \mathbb{C}^{N\times M}$, normalized channel

Similar to single-cell case, all $\mathbf{\Gamma}_{bj}$ are treated as deterministic unknown parameters and all $\tilde{\mathbf{H}}_{bj}$ are treated as random samples.

Cooperative Activity Detection via Covariance Approach

- Assume that each BS is equipped with a large-scale antenna array.
- **Cooperative detection**: To alleviate the impact of inter-cell interference, we further consider BS cooperation by assuming all BSs are connected to a CU.
- Depending on whether the large-scale fading matrices $\mathbf{G}_{bj}, \forall b, j$ are known, the device activity detection problem can be formulated differently.
- When $\mathbf{G}_{bj}$ are not known, we need to estimate $\mathbf{\Gamma}_{bj} = \mathbf{A}_j \mathbf{G}_{bj}$, $\forall b, j$, which has

$$\boxed{B^2 N \text{ unknown parameters}}$$

- When $\mathbf{G}_{bj}$ are known, we only need to estimate $\mathbf{A}_b, \forall b$, which contains

$$\boxed{BN \text{ unknown parameters}}$$

Device activity detection is much easier if large-scale fading is known!

Cooperative Detection with Unknown Large-scale Fading

We aim to estimate $\mathbf{\Gamma}_{bj} = \mathbf{A}_j \mathbf{G}_{bj}, \forall b, j$ from the received signals $\mathbf{Y}_b, \forall b$.
The likelihood function of $\mathbf{Y}_b$'s given $\mathbf{\Gamma}_{bj}$'s can be expressed as

$$p(\mathbf{Y}_1, \ldots, \mathbf{Y}_B | \mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}, \ldots, \mathbf{\Gamma}_{BB}) = \prod_{b=1}^{B} p(\mathbf{Y}_b | \mathbf{\Gamma}_{11}, \mathbf{\Gamma}_{12}, \ldots, \mathbf{\Gamma}_{BB})$$
$$= \prod_{b=1}^{B} \frac{1}{|\pi \mathbf{\Sigma}_b|^M} \exp\left(- \operatorname{tr}\left(M \mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b \right) \right). \quad (50)$$

The MLE problem can be cast as minimization of negative log-likelihood:

$$\min_{\{\mathbf{\Gamma}_{bj}\}} \quad \sum_{b=1}^{B} \left(\log |\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b \right) \right) \quad (51a)$$
$$\text{s. t.} \quad \gamma_{bjn} \in [0, \infty), \forall b, j, n \quad (51b)$$

- The problem can be solved using coordinate descent.

Cooperative Detection with Known Large-scale Fading

Assuming that all large-scale fading matrices $\mathbf{G}_{bj}$'s, we directly estimate the device activity $\mathbf{A}_b$'s using the MLE. The likelihood function of $\mathbf{Y}_b$'s can be expressed as

$$p(\mathbf{Y}_1, \ldots, \mathbf{Y}_B | \mathbf{A}_1, \ldots, \mathbf{A}_B) = \prod_{b=1}^{B} p(\mathbf{Y}_b | \mathbf{A}_1, \ldots, \mathbf{A}_B)$$
$$= \prod_{b=1}^{B} \frac{1}{|\pi \mathbf{\Sigma}_b|^M} \exp\left(- \operatorname{tr}\left(M \mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b \right) \right). \quad (52)$$

Since the activity α_{bn} is binary, the maximization of likelihood can be cast as

$$\min_{\{\mathbf{A}_b\}} \quad \sum_{b=1}^{B} \left(\log |\mathbf{\Sigma}_b| + \operatorname{tr}\left(\mathbf{\Sigma}_b^{-1} \hat{\mathbf{\Sigma}}_b \right) \right) \quad (53a)$$
$$\text{s.t.} \quad \alpha_{bn} \in \{0, 1\}, \forall b, n \quad (53b)$$

Known large-scale fading: Single-cell and multicell have same phase transition

- Multicell problem: Find a BN-dim sparse vector in $(BN - BL^2)$-dim subspace.
- Single-cell problem: Find a N-dim sparse vector in $(N - L^2)$-dim subspace.

Performance of Covariance Based Detection for Multi-cell

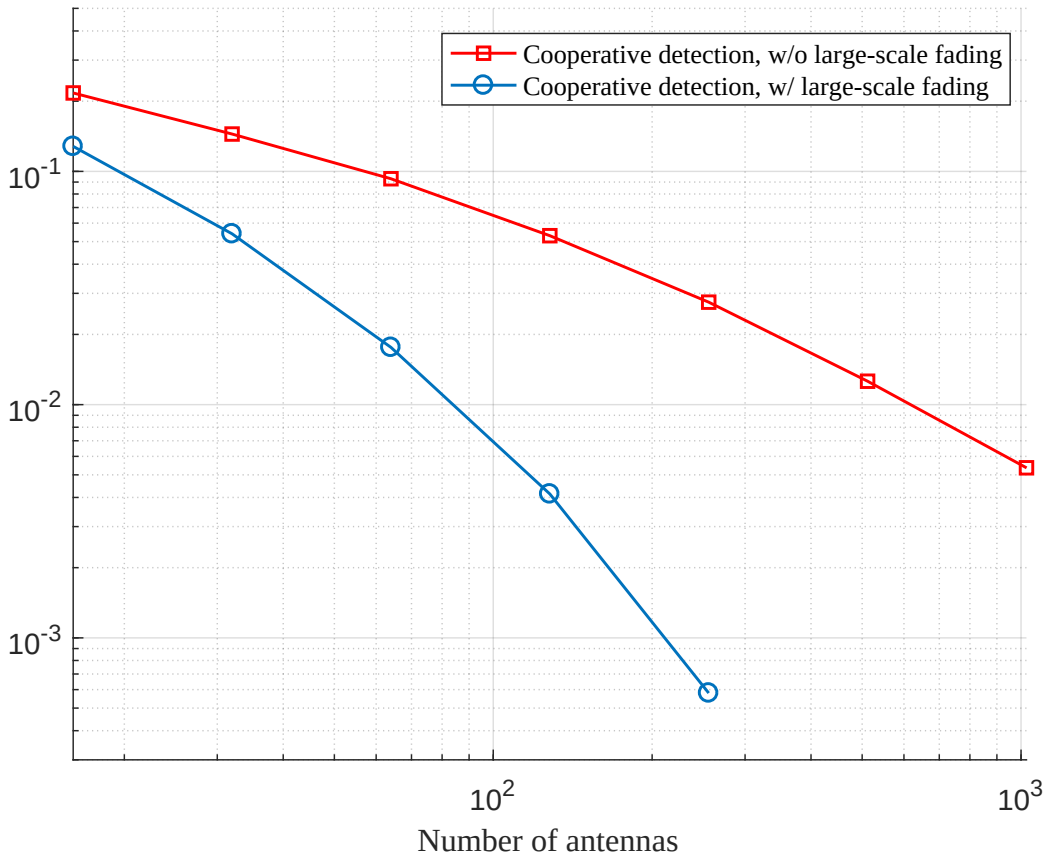


Figure: Performance comparison of the multicell covariance approach with and without knowing large-scale fading. $B = 7$, $N = 200$, $K = 20$, and $L = 20$. We observe that knowing the large-scale fading brings substantial improvement.

Performance of Covariance Based Detection for Multi-cell

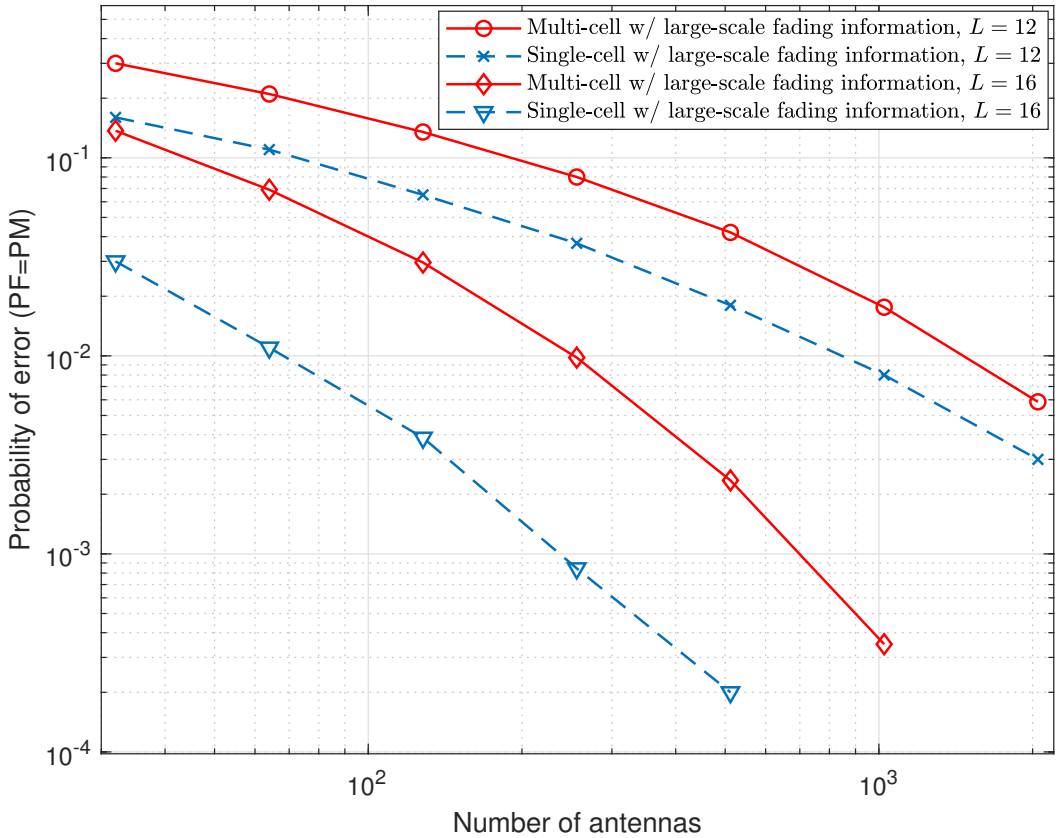


Figure: Performance comparison of the multicell covariance approach with single-cell system, with knowing large-scale fading. $B = 7$, $N = 200$, $K = 20$.

Conclusions

- Device activity detection for massive random access in machine-type and IoT communications is a sparse recovery problem.
- Two detection algorithms for user activity detection:
  - Signal-based AMP for estimating the user activity and the exact channel.
  - Covariance-based MLE for estimating the user activity only.
- Analyses for AMP and the covariance approach:
  - State evolution for AMP: Low complexity, works best for small M.
  - Fisher information matrix for covariance approach: Suited for massive MIMO.
- Advantage of covariance-based approach is that it can handle $K = O(L^2)$, as compared to AMP which can only handle $K = O(L)$.

Further Information

Zhilin Chen, Foad Sohrabi, Ya-Feng Liu, Wei Yu,

"Phase Transition Analysis for Covariance Based Massive Random Access with Massive MIMO",

[Online] available: https://arxiv.org/abs/2003.04175, March 2020.

Zhilin Chen, Foad Sohrabi, and Wei Yu,

"Sparse Activity Detection in Multi-Cell Massive MIMO Exploiting Channel Large-Scale Fading",

To appear in *IEEE Transactions on Signal Processing*, 2021.

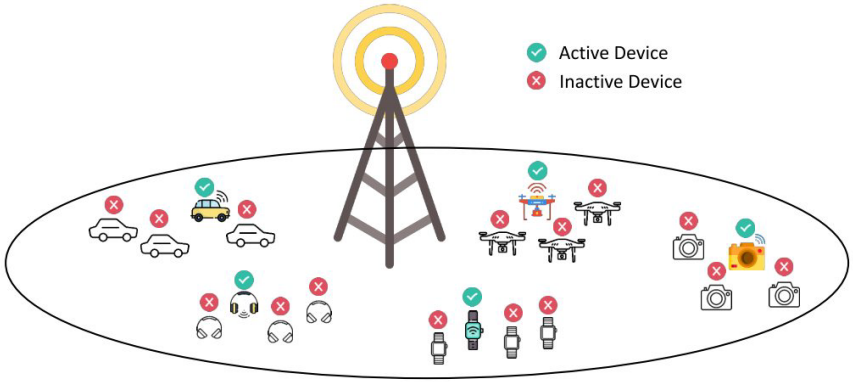
Massive Random Access with Massive MIMO: Contention versus Scheduling

Wei Yu

Joint Work with Justin Kang

University of Toronto

Massive Connectivity



- Massive connectivity is a crucial requirement for Internet-of-Things (IoT)
- Requires up to $10^5 \sim 10^6$ devices connected per base station (BS).
  - Sporadic traffic, making device identification & scheduling challenging.
  - Assigning each user an orthogonal resource requires coordination.
- Activity Detection is a first step toward coordination.
- Equally importantly, we need to schedule users to transmission slots.

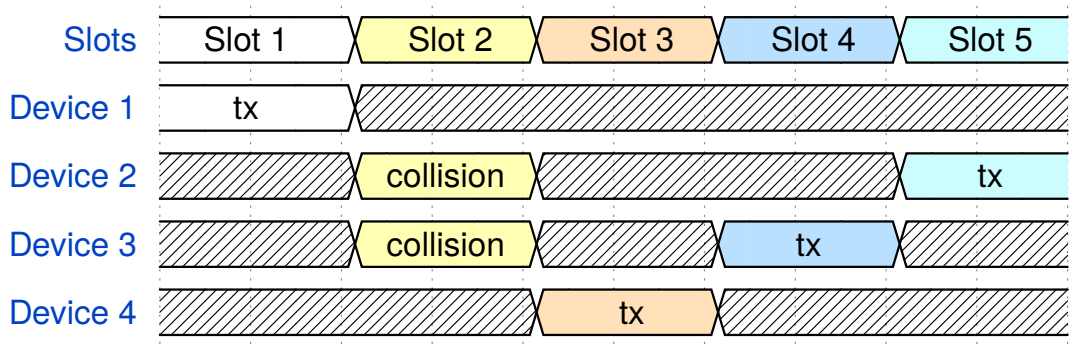
What is the cost of coordinated scheduling?

Contention-Based vs Coordinated Scheduling

- **Uncoordinated Random Access:**
  - Classic Slotted ALOHA: Contention-based uncoordinated scheduling.
  - Coded ALOHA can alleviate some of the inefficiencies of classic ALOHA.
- **Coordinated Random Access:**
  - Coordinated scheduling requires feedback from the BS to the users.
  - What is the minimum feedback rate for scheduling?
- **Massive Random Access with Massive MIMO:**
  - Coded Pilot Access vs. Scheduled Random Access

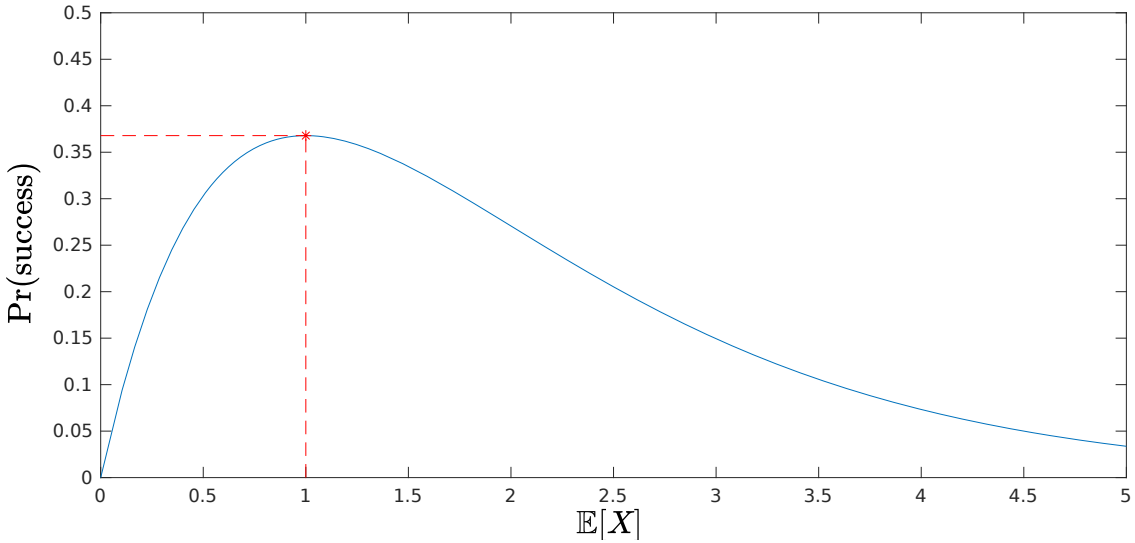
Classic Solution: Slotted ALOHA

Slotted ALOHA involves contention and is uncoordinated involving no communication between BS and users.



- Users become active and transmit at random with probability p.
- Transmission is successful only if a single user transmits in a slot.
- If there is a collision, users must re-transmit their payload.

Slotted ALOHA: Analysis

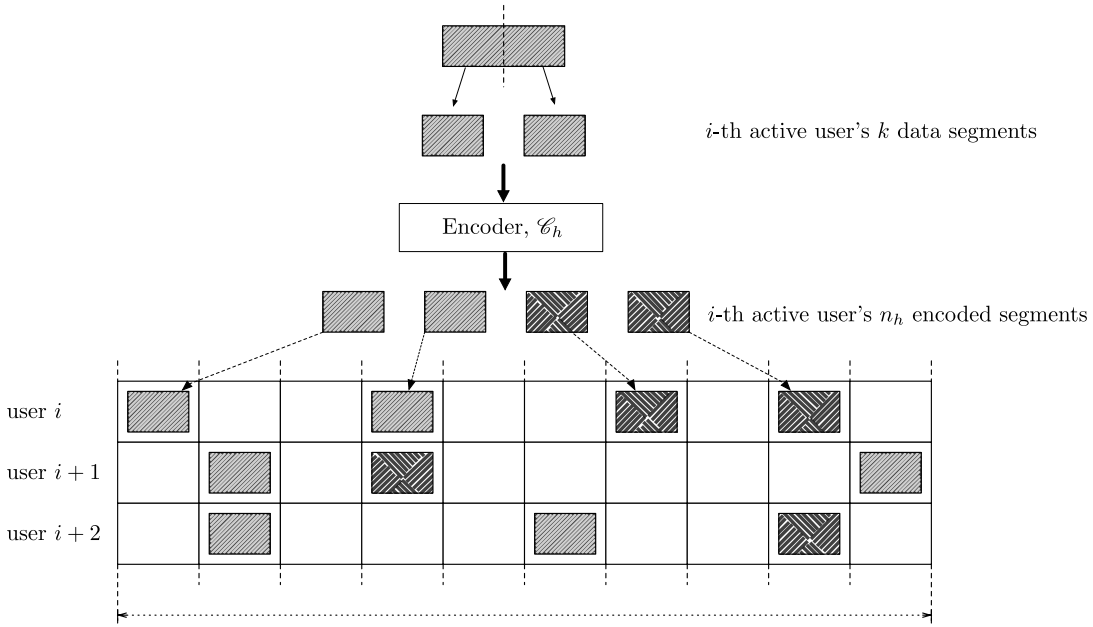


- Let X be the number of users that transmit in a slot.
- Since X is sum of independent Bernoulli trials, it follows Poisson distribution

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda k}}{k!}, \quad \text{where} \;\; \mathbb{E}[X] = \lambda. \tag{1}$$

- Successful transmission only when $k = 1$, with probability $\lambda e^{-\lambda}$.
- Optimize over λ. Throughput is maximized when $\lambda = 1$ with P(success) $= \frac{1}{e}$.
- Slots with collision or slots with no transmission (i.e., 63% slots) are wasted.

Coded Slotted ALOHA



- Coded Slotted ALOHA: Use packet-level erasure codes and successive interference cancellation (SIC) to extract information from collisions.
- Each user chooses an (n_h, k) erasure code $\mathscr{C}_h$ to encode their k segments.
- Code is chosen from a finite set $\{\mathscr{C}_h\}_{h=1}^{\theta}$ according to some p.m.f., and the n_h packets are transmitted randomly over a fixed frame.

E. Paolini, G. Liva, and M. Chiani, "Coded Slotted ALOHA: A Graph-Based Method for Uncoordinated Multiple Access," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6815–6832, 2015.

Coded Slotted ALOHA: Graph Representation

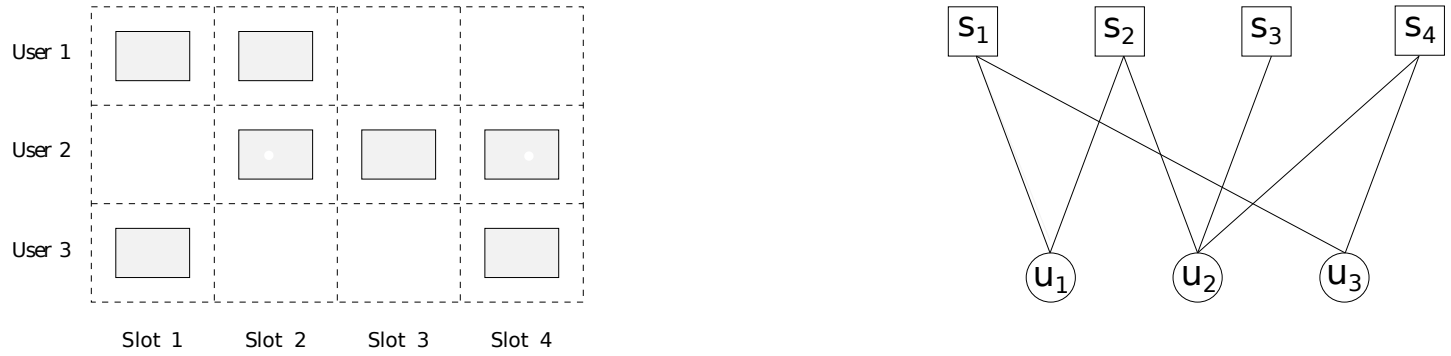


Figure: Bipartite graph model for contention resolution

- Users are represented by variable nodes, slots by check nodes.
- A user node u_i is connected to slot node s_j if user i transmits in slot j.
- Decoding process is identical to the peeling decoder for erasure channel.
- If users select repetition codes, this is known as Contention Resolution Diversity Slotted ALOHA (CRDSA).

Coded Slotted ALOHA: Decoding Example

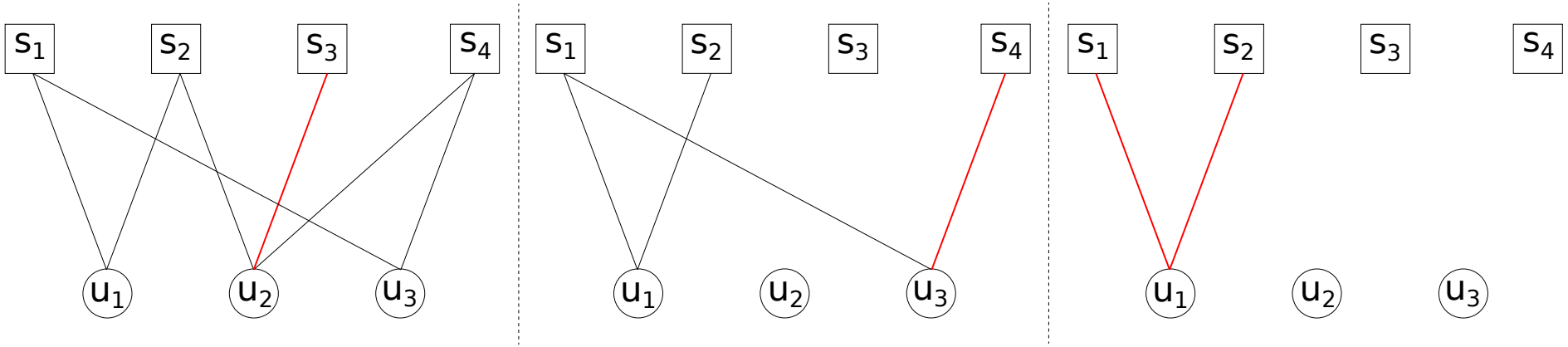
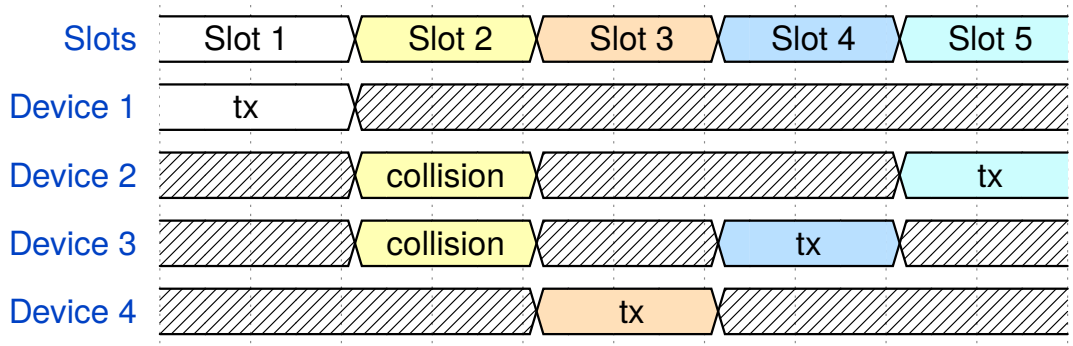


Figure: Peeling decoding for CRDSA on a bipartite graph.

- Decoding procedure for CRDSA is similar to Fountain code or LT code.
- This connection allows us to show that the optimal user-node degree distribution is the soliton distribution [Narayanan-Pfister'12].
- With this degree distribution, the throughput $\triangleq \frac{\#\text{ of decoded users}}{\#\text{ of slots}} \to 1$ asymptotically as the number of users and slots go to infinity.

Contention vs. Scheduling

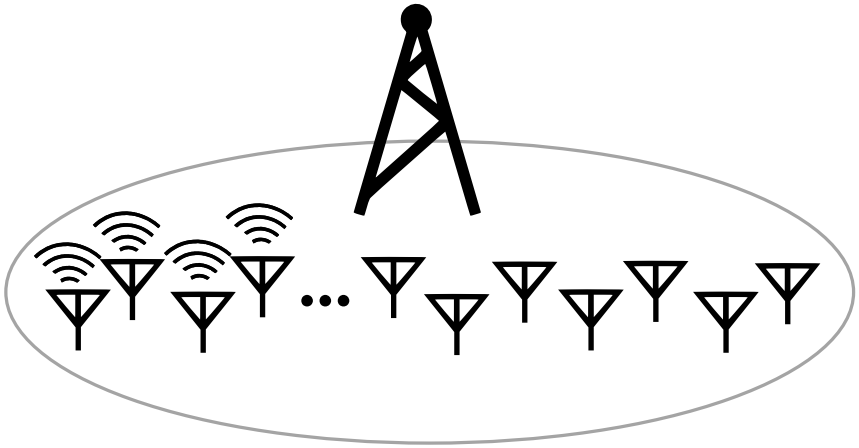


- Slotted ALOHA based schemes all involve contention and collision resolution
  - Multiple transmissions increases power consumption.
  - Collision resolution increases delay.
  - Practical schemes cannot operate at optimal throughput.
- Scheduling is an alternative approach to contention.
- Contention-based schemes are often justified based on the assumption that the cost of coordination is too great.

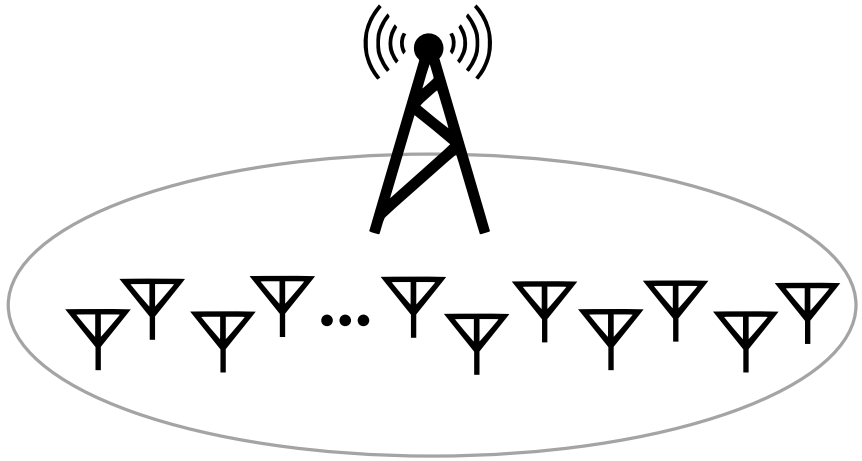
What is the cost of scheduling?

Feedback-Based Scheduling for Random Access

Each of n potential users is assigned a unique non-orthogonal pilot.

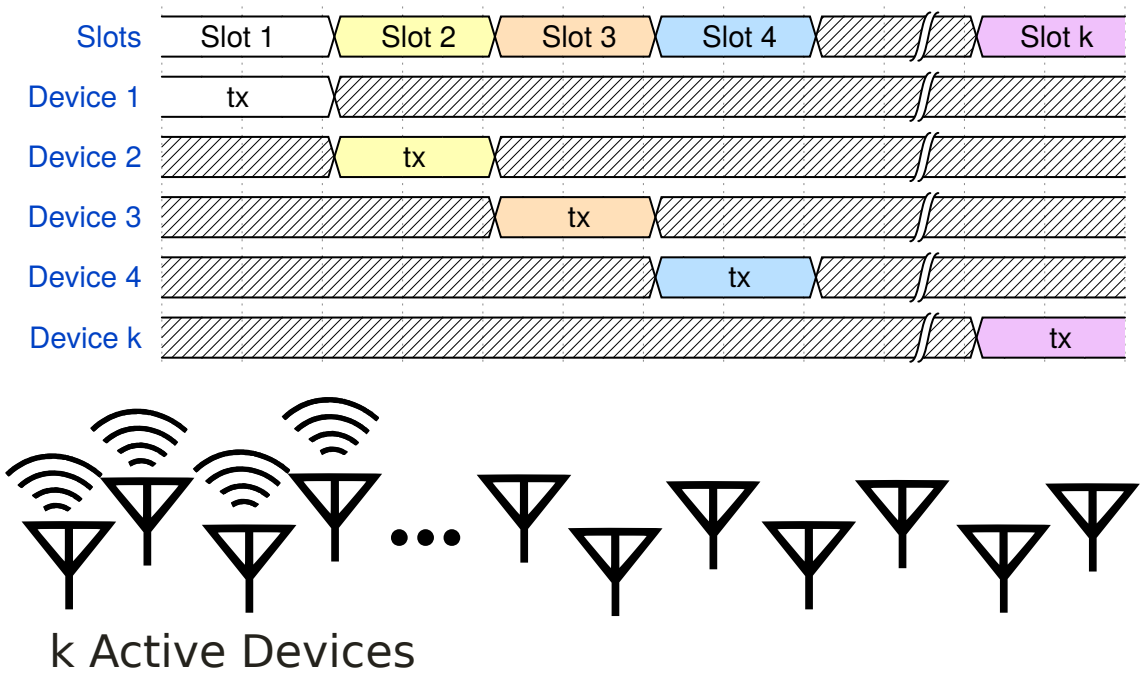


Phase 1 (Activity Detection):
The k active users ($k \ll n$) send their pilots synchronuously to the BS.



Phase 2 (Downlink Feedback): BS sends a *common* feedback message to schedule the data transmissions of k active users.

Feedback-Based Scheduling for Random Access



Phase 3 (Uplink Payload Transmission): The k active users transmit their payload in the k slots based on the schedule provided by the BS, while avoiding collision.

What is the minimum feedback needed to ensure collision-free scheduling?

Straightforward Feedback Scheme

- A naive scheme to schedule k out of n users:
  - Assign a unique index to each of the n users;
  - The BS detects the k active users based on the pilots;
  - The BS lists the k users in the order in which they should transmit;
  - Each active user finds its index in the list, waits for its turn to transmit.

- The feedback overhead of this scheme is $k \log(n)$ bits.
  - When $n = 10^6$, the cost of identification is $\log(n) = 20$ bits per user.

Can we do better?

Why Can We Do Better?

- The naive $k \log(n)$ feedback scheme is not optimal.
- There is flexibility in the order that users are scheduled.

  Example: Users $1, \ldots, k$ are to be scheduled. The BS can schedule according to any of the $k!$ permutations of these users, e.g. $\{1, \ldots, k\}$ or $\{k, \ldots, 1\}$.

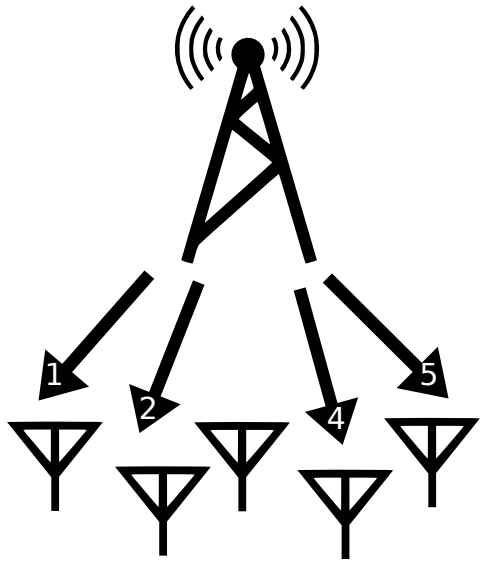
  We can remove this extraneous cost via *enumerative source coding*.

  This still requires $\log \binom{n}{k}$ bits feedback, which scales as $O(\log(n))$ for fixed k.
- Each user only needs to know its own slot, and NOT the other users' slots. Removing this extraneous information is the key to further reducing feedback.

G. K. Facenda and D. Silva, "Efficient Scheduling for the Massive Random Access Gaussian Channel," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7598–7609, Aug. 2020.

Identification Capacity

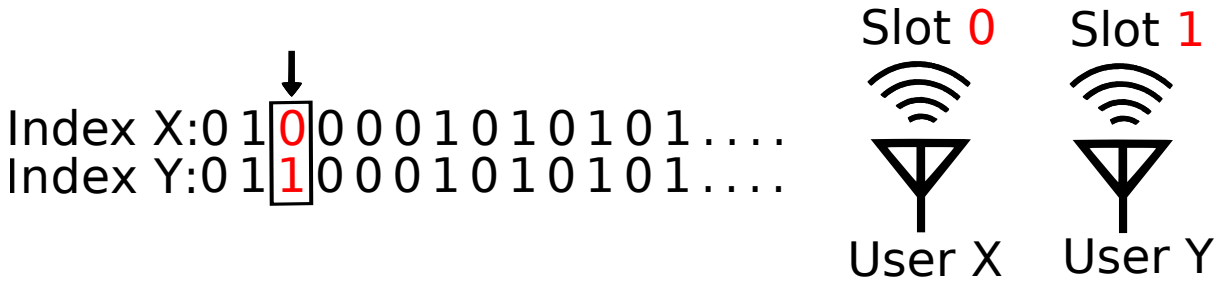
- *Identification via channels* [Ahlswede-Dueck, 1992] says that identifying one out of n users only requires $O(\log\log(n))$ bits! — This eliminates the extraneous information as users no longer know which other users are active.
- Identification codes lead to a feedback rate of $O\left(k\log\log(n)\right)$.



- A careful construction can beat even this scheme!

Two-User Case

- Consider the case of two active users $(k = 2)$, out of a total of n users:
- Any two distinct binary vectors differ in at least one index:



- BS simply transmits the location in which the user indices differ.
- The user with 0 transmits first, and the user with 1 transmits second.
- This requires only $R = \lceil \log \lceil \log (n) \rceil \rceil$ feedback with a fixed-length encoding.

Optimal for $k = 2$!

Feedback Scheduling Code for Arbitrary (n, k)

- Notation: $[n] = \{1, \cdots, n\}$. $\binom{[n]}{k} \triangleq$ set of all k-element subsets of $[n]$.
- The BS encodes the "activity pattern" into an index t

$$f : \binom{[n]}{k} \to \{1, 2, \ldots, T\} \triangleq [T].$$

- Each user "decodes" its scheduled slot using

$$g_i : [T] \to [k], \quad i \in [n].$$

(We consider k slots here, but having more slots can decrease feedback.)

- In order for no collisions between active users, we must have:

$$\forall \mathbf{A} \in \binom{[n]}{k}, \ \exists t \in [T] \ \text{ s.t. } \ \forall i \neq j \in \mathbf{A} \ \ g_i(t) \neq g_j(t).$$

Scheduling via Set-Partitioning

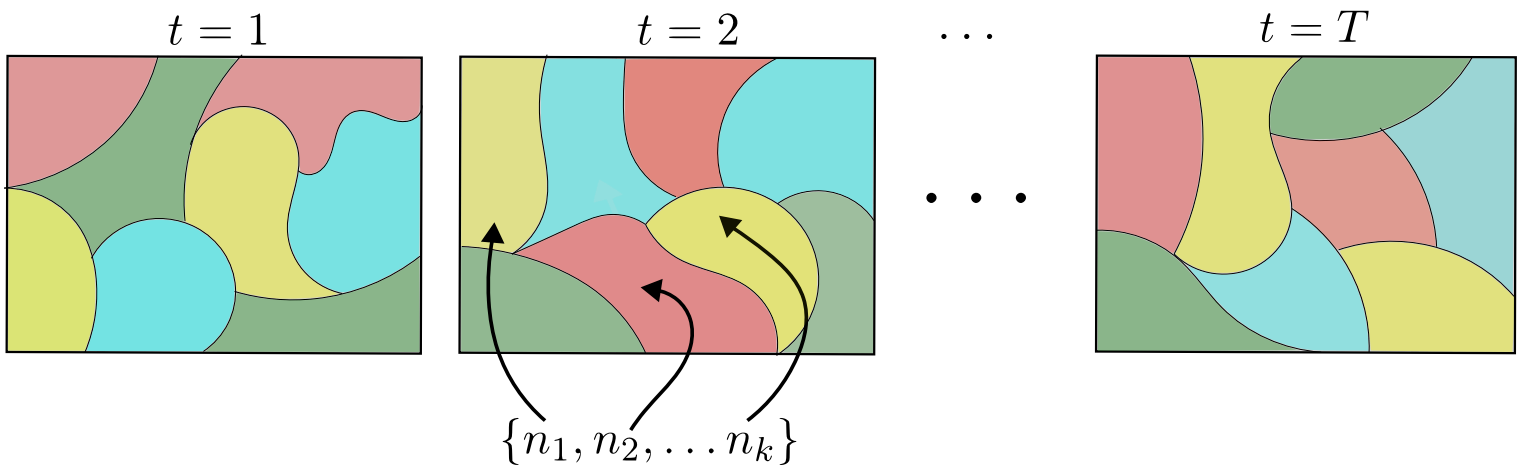
- Define a k-partition of a set $[n]$ to be a tuple of subsets $\bar{\mathbf{X}} = (\mathbf{X}_1, \ldots, \mathbf{X}_k)$ such that $\mathbf{X}_i \bigcap \mathbf{X}_j = \emptyset$, $\forall i, j$, and $\bigcup_{i=1}^{k} \mathbf{X}_i = [n]$.
- Define the set of activity patterns that can be covered by $\bar{\mathbf{X}}$ as

$$\mathbf{C}\left(\bar{\mathbf{X}}\right) = \left\{\{x_1, \ldots, x_k\} \mid x_i \in \mathbf{X}_i, i = 1, \ldots, k\right\}.$$

  i.e., there is exactly one active user in each distinct subset of the partition $\bar{\mathbf{X}}$.
- Example: For the set $[4]$, if $\bar{\mathbf{X}} = (\{1, 2\}, \{3, 4\})$, then

$$\mathbf{C}\left(\bar{\mathbf{X}}\right) = \left\{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\right\}.$$

Set-Partition Encoding



- To cover all activity patterns, we construct T partitions $\bar{\mathbf{X}}^{(1)}, \ldots, \bar{\mathbf{X}}^{(T)}$ s.t.

$$\bigcup_{t=1}^{T} \mathbf{C}\left(\bar{\mathbf{X}}^{(t)}\right) = \binom{[n]}{k}.$$

- For activity pattern $\mathbf{A}$, the following encoder/decoders ensure no collision:

$$f(\mathbf{A}) = t \quad \text{s.t.} \quad \mathbf{A} \in \mathbf{C}\left(\bar{\mathbf{X}}^{(t)}\right);$$

$$g_i(t) = j \quad \text{if} \quad i \in \mathbf{X}_j^{(t)}.$$

Tetra Code: An Example for $(n, k) = (9, 3)$

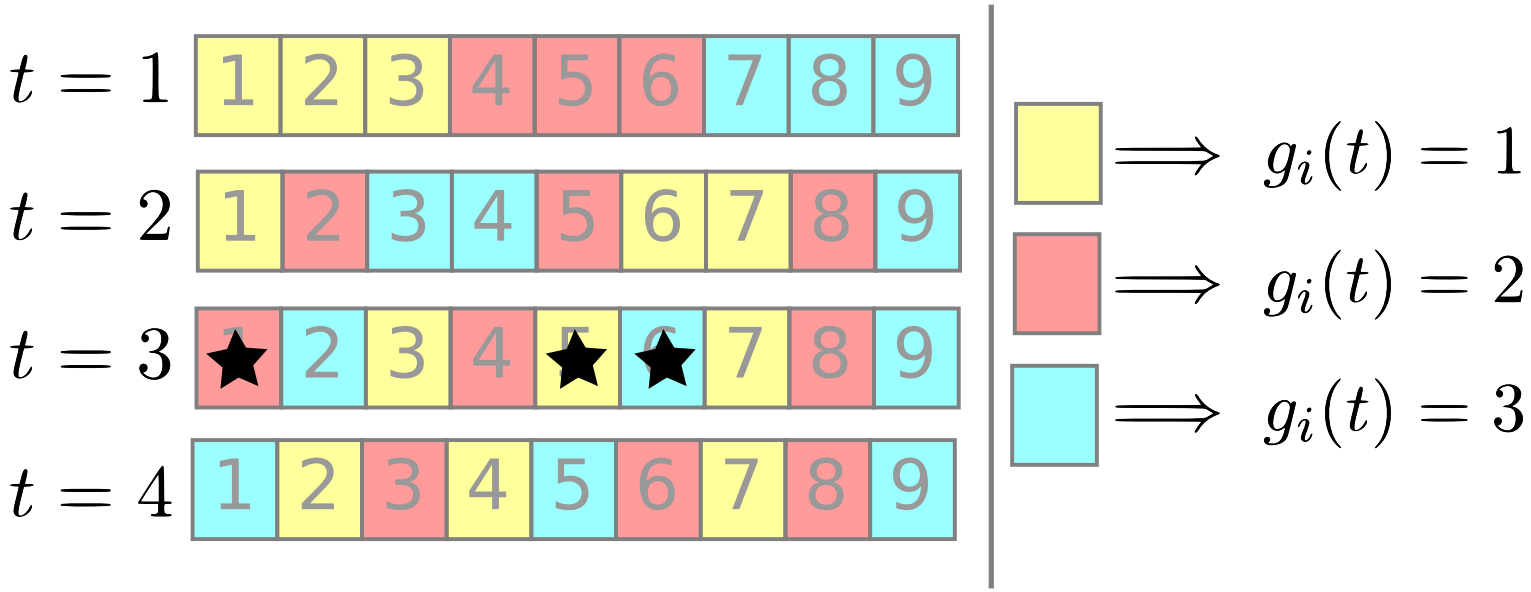


Figure: The tetra code can be used to define 4 partitions.

- Example: For the activity pattern $\mathbf{A} = \{1, 5, 6\}$, the $t = 3$ partition has all three active users in separate subsets, thus $f(\mathbf{A}) = 3$ ensures no collision.
- Only 2 bits of feedback as required! Optimal [Körner and Marton, 1988].

Set-Partition Encoding

- Any set of collision-free encoding and decoding function can be described with the set-partition framework.
- Given the decoding functions $g_i' : [T] \to [k]$, we can define T partitions $\bar{\mathbf{X}}^{(t)} = \left(\mathbf{X}_1^{(t)}, \ldots, \mathbf{X}_k^{(t)}\right)$, $t \in [T]$, where

$$\mathbf{X}_j^{(t)} = \{i \mid g_i'(t) = j, \quad i \in [n]\}.$$

- For a fixed-length feedback code, we define the feedback rate as

$$R_f^*(n,k) \triangleq \log(T^*)$$

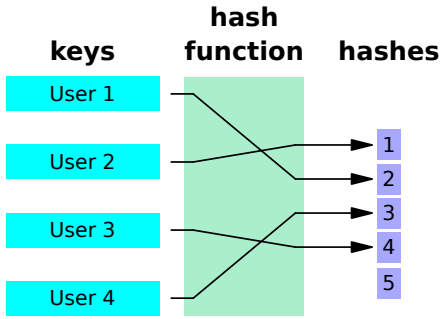
where T^* is the minimum number of partitions needed to cover all activity patterns.

Finding the minimum-rate zero-collision feedback code
now reduces to finding T^*.

Perfect Hashing Families

Finding T^* is equivalent to the *perfect hashing families* problem.

- An (n, b, k)-family of perfect hash functions is a family of functions from $[n] \to [b]$ for $n \geq b \geq k$ such that for every $\mathbf{A} \subset [n], |\mathbf{A}| = k$, there exists a function in the family that is injective on $\mathbf{A}$.



- We can view our decoding functions as a (n, k, k)-family perfect hash functions from $[n] \to [k]$ if we swap the argument and the subscript.

Theorem (Fredman and Komlós, 1984, Körner and Marton, 1988)

The minimum size T^* of an (n, b, k) perfect hash family is bounded as:

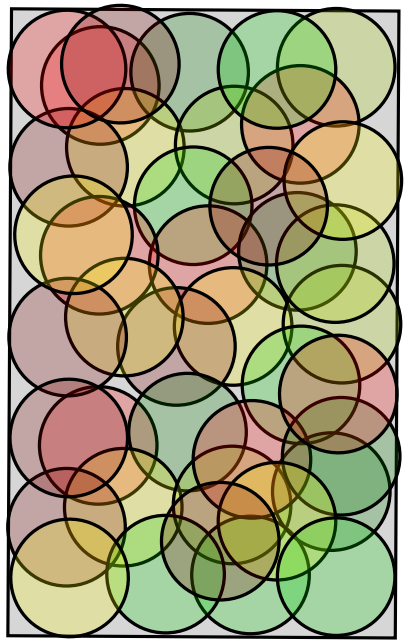
$$\frac{\log n}{\min_{1 \leq s \leq k-1} \frac{b^{\underline{s}}}{b^s} \log \frac{b-s+1}{k-s}} \lesssim T^* \lesssim \frac{(k-1)\log n}{\log \frac{1}{1-\frac{b^{\underline{k}}}{b^k}}}.$$

- The proof uses a notion of hypergraph entropy, but we can derive simpler, but still instructive bounds. Here, $b^{\underline{k}} \triangleq \frac{b!}{(b-k)!}$ is the falling factorial.

Random Partition Construction

- Take T random partitions $\bar{\mathbf{X}}^{(1)}, \ldots, \bar{\mathbf{X}}^{(T)}$, then the probability an activity pattern $\mathbf{A}$ is not covered is

$$\Pr\left(\mathbf{A} \notin \bigcup_{t=1}^{T} \mathbf{C}\left(\bar{\mathbf{X}}^{(t)}\right)\right) = \left(1 - \frac{k!}{k^k}\right)^T.$$



- By the union bound we have:

$$\Pr\left(\bigcup_{t=1}^{T} \mathbf{C}\left(\bar{\mathbf{X}}^{(t)}\right) \neq \binom{[n]}{k}\right) \leq \binom{n}{k}\left(1 - \frac{k!}{k^k}\right)^T.$$

- If the RHS of the above falls below 1, it means that there exists a family of partitions that cover all activity patterns.

Achievability Bound on Minimum Feedback Rate

- Using the fact $1 - x < e^{-x}$, we can show that the RHS falls below 1 for:

$$T \geq \left(\ln \binom{n}{k}\right)\left(\frac{k^k}{k!}\right).$$

Proposition

The minimum rate for a fixed-length collision-free feedback code must be upper bounded as:

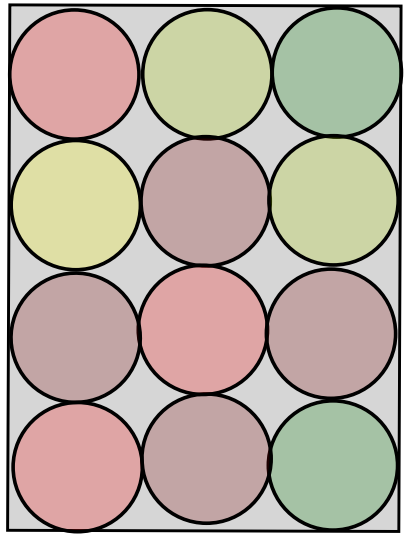
$$R_f^*(n,k) \triangleq \log(T^*) \leq k\log(e) + \log\left(\ln\left(\frac{n}{k}\right)+1\right) + \frac{1}{2}\log\left(\frac{k}{2\pi}\right).$$

Key observation: $R_f^*(n,k) \leq O(\log\log(n))$, plus a linear term in $k\log(e)$.

Converse: Volume Bound

- Since each partition can cover at most only a small fraction of the activity patterns, we can also place a volume bound on the covering:

$$T^* \geq \frac{\binom{n}{k}}{\left\lceil \frac{n}{k} \right\rceil^{n \bmod k} \left\lfloor \frac{n}{k} \right\rfloor^{k-n \bmod k}}.$$



Proposition

The minimum rate for a fixed-length collision-free feedback code must be lower bounded as:

$$R_f^*(k,n) \geq k \log(e) - \log\left(\frac{n^k}{n(n-1)...(n-k+1)}\right) - \frac{1}{2}\log(2\pi k) - \frac{\log(e)}{12k}.$$

Thus, $R_f^*(n,k) \geq O(k)$.

Converse: Exclusion Bound

- A partition $\bar{\mathbf{X}}^{(1)}$ cannot have covered any activity pattern which has all its elements drawn from $\mathbf{S_1} = [n] - \mathbf{X}_j^{(1)}$, as

$$\mathbf{C}\left(\bar{\mathbf{X}}^{(1)}\right) \cap \binom{[n] - \mathbf{X}_j^{(1)}}{k} = \emptyset, \quad j = 1, \ldots, k.$$

  i.e., activity patterns with indices exclusively drawn from $\mathbf{S_1}$ are *excluded*.

- Since one of the partitions $\mathbf{X}_j^{(i)}$ is at most size $\left\lfloor \frac{n}{k} \right\rfloor$, we have:

$$|\mathbf{S_1}| = m_1(n,k) \geq n\left(1 - \frac{1}{k}\right).$$

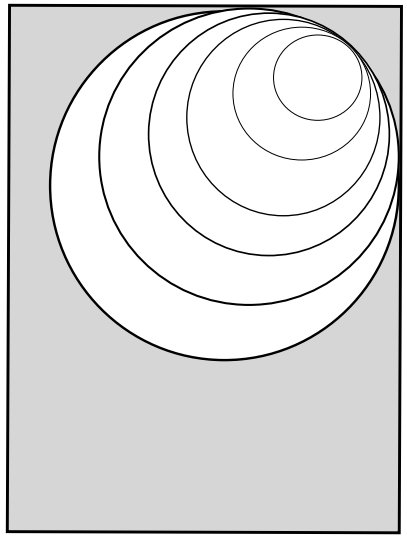
Exclusion Bound

- By repeated application of the exclusion argument:

$$m_t(n,k) \geq n\left(1-\frac{1}{k}\right)^t.$$

- For this exclusion set to shrink down to the null set (not containing any activity pattern), we need

$$n\left(1-\frac{1}{k}\right)^T \leq k-1.$$



With each partition, the exclusion region shrinks.

Proposition

The minimum rate for a fixed-length collision-free feedback code must be lower bounded as:

$$R_f^*(n,k) \geq \log\log\left(\frac{n}{k-1}\right) + \log(k) - 1.$$

From Fixed to Variable Length Feedback Code

Fixed-length collision-free feedback code:

- Random Partition: $R_f^*(n,k)$ scales at most as $k\log(e)$ plus $O(\log\log(n))$.
- Volume Bound: $R_f^*(n,k)$ scales at least as $k\log(e)$ for large n.
- Exclusion Bound: $R_f^*(n,k)$ scales at least as $\Omega(\log\log(n))$ for fixed k.

Thus, rate of fixed-length code scales linearly as $k\log(e)$ and as $\Theta(\log\log(n))$.

Can we do better?

Variable-length collision-free feedback code:

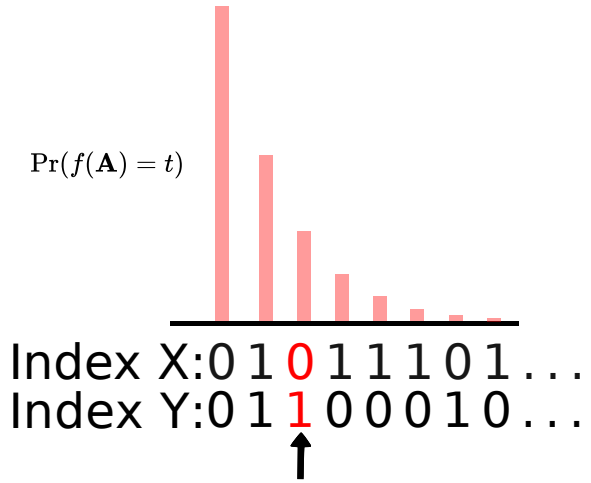
- Treat $\mathbf{A}$ as a random variable with distribution $Q(\mathbf{A})$ and define $R_v(n,k) \triangleq \mathrm{H}(f(\mathbf{A}))$, corresponding to optimal entropy coding.
- Focusing on the worst-case activity distribution, define the optimal rate as:

$$R_v^*(n,k) \triangleq \sup_{Q(\cdot)} \mathrm{H}(f(\mathbf{A})).$$

- It turns out we can remove even the $\Theta(\log\log(n))$ growth in n.

Greedy Encoding for $k = 2$

- Consider the index based feedback strategy for $k = 2$, but now greedily choose the first position where user indices differ.
- If the user activity is the worst-case uniform distribution, $f(\mathbf{A})$ follows a truncated geometric distribution.



- A direct application of Huffman Coding results in a code of rate:

$$R_v(n, 2) = 2 - \frac{\log(n) + 1}{n - 1}.$$

- This implies $\lim_{n\to\infty} R_v^*(n, 2) \leq 2$, thus the achievable feedback rate remains bounded as n tends to infinity.

Greedy Encoding for $k > 2$

- We again use the concept of greedy encoding strategy. Given a family of T k-partitions $\underline{\mathbf{B}} = (\bar{\mathbf{X}}^{(1)}, \ldots, \bar{\mathbf{X}}^{(T)})$, define the greedy encoder $f_{\underline{\mathbf{B}}}$:

$$f_{\underline{\mathbf{B}}}(\mathbf{A}) = \min_{t \in [T]} t, \quad \text{s.t.} \quad \mathbf{A} \in \mathbf{C}\left(\bar{\mathbf{X}}^{(t)}\right), \text{ else } \quad T+1,$$

  and the resulting distribution $p_{\underline{\mathbf{B}}}(t) \triangleq \Pr(f_{\underline{\mathbf{B}}}(\mathbf{A}) = t)$.
- Denote the set of all families of k-partitions of size T as $\mathcal{B}$, regardless of whether each of them covers all activity patterns, or not.
- Consider an encoder that chooses $\underline{\mathbf{B}}$ uniformly at random from $\mathcal{B}$. Define $p_{\mathcal{B}}(t) \triangleq \mathbb{E}_{\mathcal{B}}\left[p_{\underline{\mathbf{B}}}(t)\right]$. The first T terms in this distribution are:

$$p_{\mathcal{B}}(t) = \frac{k!}{k^k}\left(1 - \frac{k!}{k^k}\right)^{t-1}, \quad t = 1, \ldots, T,$$

  with the remainder of the mass at $T+1$, regardless of the distribution of $\mathbf{A}$.

Variable-Length Feedback Bounds

- With Jensen's inequality, this implies the following bound independent of T:

$$\mathbb{E}_{\mathcal{B}}\left[\mathrm{H}\left(p_{\underline{\mathbf{B}}}(t)\right)\right] \leq \mathrm{H}\left(p_{\mathcal{B}}(t)\right) \leq (k+1)\log(e).$$

- For families of partitions of size T, let $1-\epsilon$ be the fraction of collision-free families in $\mathcal{B}$, then the rate for collision-free feedback can be bounded as:

$$R_v^*(n,k) \leq \frac{1}{1-\epsilon}(k+1)\log(e)$$

- Now, we can let $T \to \infty$, so $\epsilon \to 0$, implying $R_v^*(n,k) \leq (k+1)\log(e)$.
- The volume bound converse can also be extended to variabe-length codes.

Theorem

The minimum rate for variable-length collision-free feedback code is bounded as

$$(k+1)\log(e) \;\geq\; R_v^*(n,k) \;\geq\; k\log(e) - \log\left(\frac{n^k}{n^{\underline{k}}}\right) - \frac{1}{2}\log(2\pi k) - \frac{\log(e)}{12k}.$$

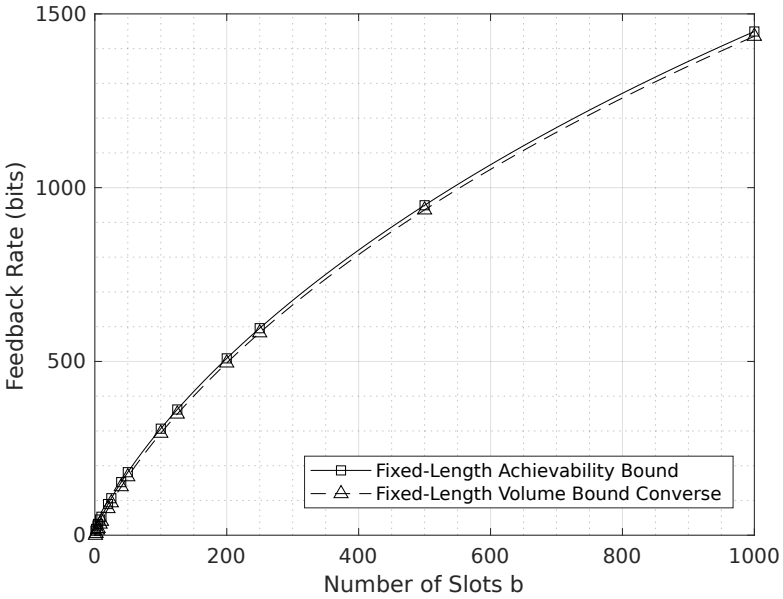
Practical Implementations

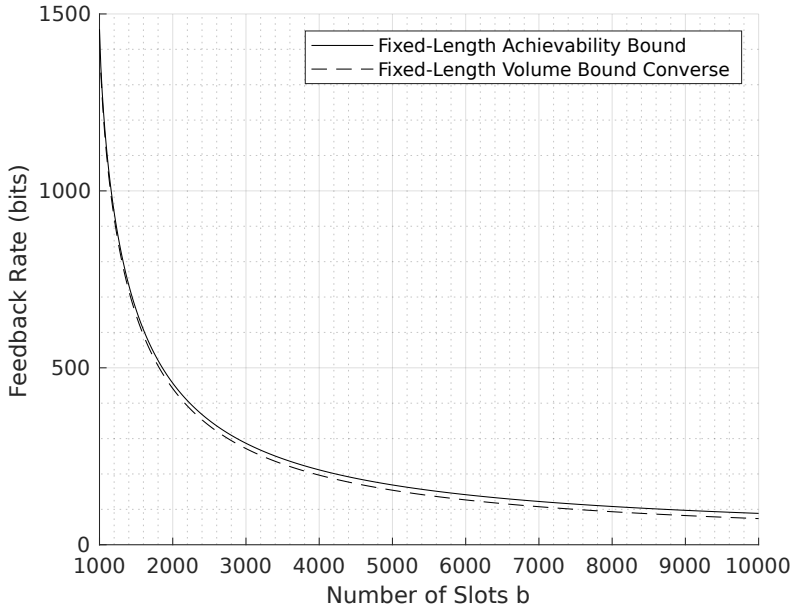
- Consider a system with $n = 10^6$ potential users and $k = 10^3$ active users:
  - Naive scheme would require **20 kbits**
  - Enumerative source coding requires **11.5 kbits**.
  - Optimal feedback only needs approximately **1.5 kbits**.
- Some practical schemes come close to achieving the $k \log(e)$ linear scaling:

Table: Practical Hashing/Feedback Algorithms

Method	Bits Per User
Random Coding	1.44
Boolean SAT	1.83
Compress-Hash-Displace	2.07

More Slots and Multiple Users per Slot





- These bounds can be extended to the case of:
  - $b \geq k$ slots (over-provisioned system), and
  - $b \leq k$ slots for systems where the BS can decode multiple users per slot.

Summary

What is the cost of coordinating collision-free scheduling?

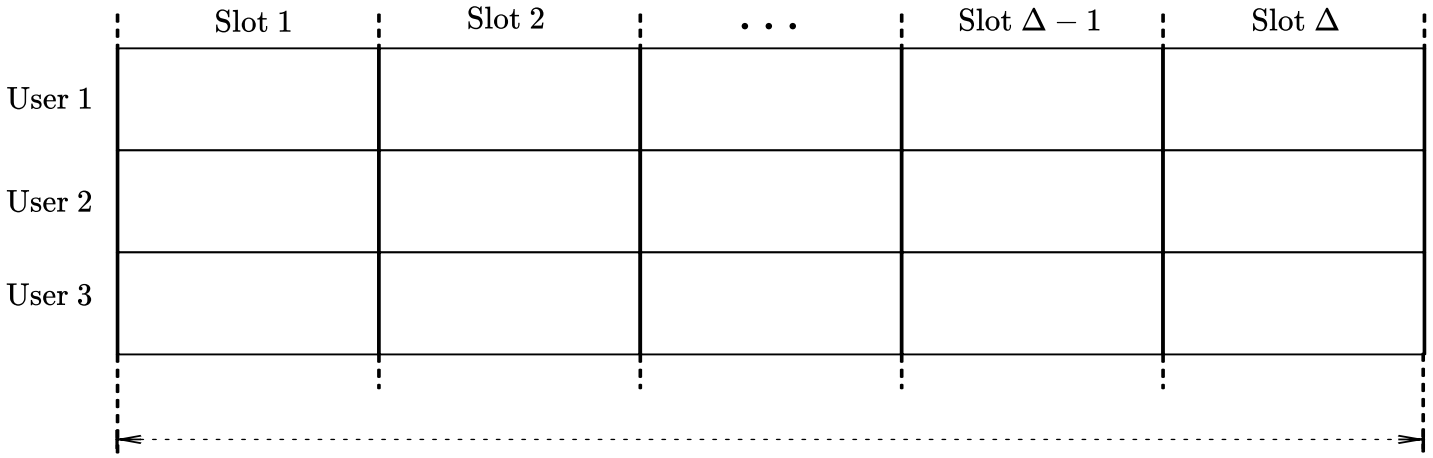
- Fixed-length feedback codes for collision-free scheduling of k active users among n potential users into k slots requires a rate of approximately $k \log(e)$ bits, plus a $\Theta(\log \log(n))$ term.
- Using variable-length feedback codes can reduce the required feedback rate for collision-free scheduling to $(k+1) \log(e)$ bits, independent of n.
- If $b \geq k$ slots are available, or more than one user can be decoded per slot, feedback can be further reduced.

Random Access for Massive MIMO Systems

1. Uncoordinated Random Access for Massive MIMO
   - Channel estimation and data transmission must both be without coordination.
   - Coded ALOHA can be adapted to Massive MIMO systems to enable uncoordinated communication.
   - We will consider a variant of coded ALOHA known as *Coded Pilot Access*.

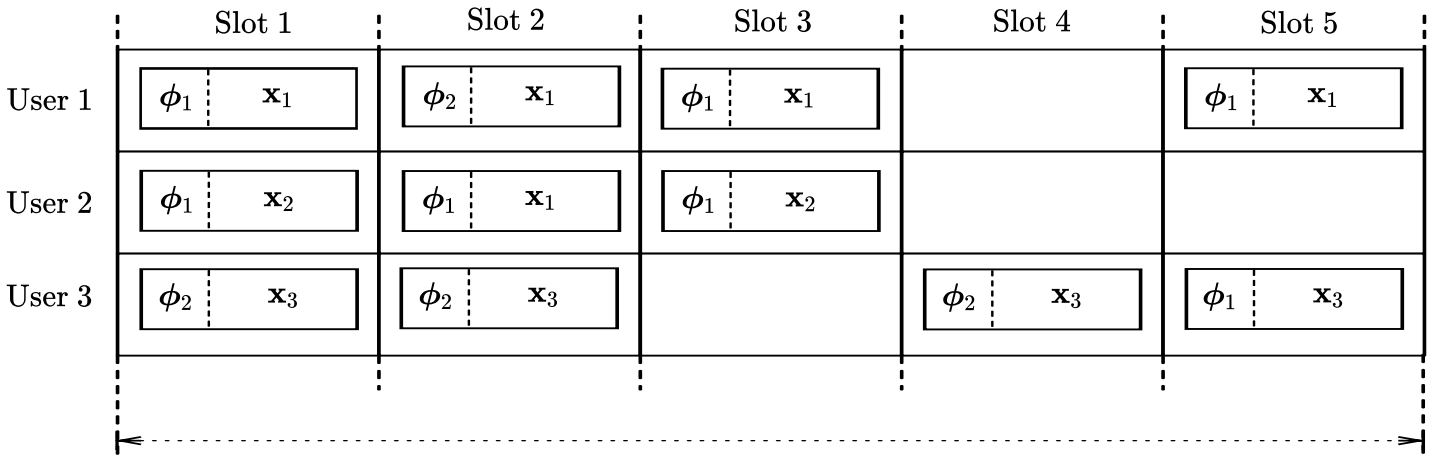
2. Scheduled Random Access for Massive MIMO
   - Activity detection can serve as an initial step for scheduled random access.
   - A relatively small amount of feedback can be used to ensure collision-free scheduling for the users.
   - Users are assigned orthogonal pilots for channel estimation.

Slotted Random Access



- The BS is equipped with M antennas.
- There are n single-antenna devices k of which are active.
- Active users transmit across Δ temporal slots each containing L symbols.
- The channels $\mathbf{h}_{d,i} \sim \mathcal{CN}(0,1)$ is i.i.d for each user i in the d^{th} slot. We assume users apply inverse power control to compensate for large scale fading.
- The BS uses the received signal $\mathbf{Y}_d$ over Δ slots to decode the messages of k active users.

Coded Pilot Access

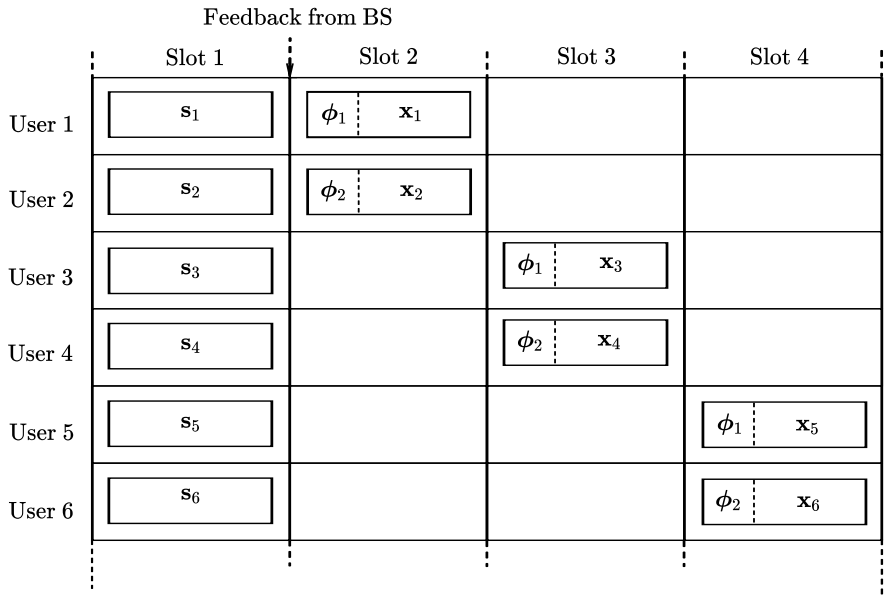


- Users transmit their payload $\mathbf{x}_i$ multiple times, each time preceded by a pilot randomly selected from a set of orthogonal pilots $\{\boldsymbol{\phi}_t\}_{t=1}^{\tau}$.
- In cases with no collision, the BS can perform channel estimation and data decoding for that user.
- The data contains the location of the other slots where the user has transmitted, allowing the BS to perform SIC.

J. H. Sørensen, E. De Carvalho, Č. Stefanović, and P. Popovski, "Coded Pilot Random Access for Massive MIMO Systems", *IEEE Trans. Wireless Commun.*, vol.17, no.12, pp.8035–8046, 2018.

Scheduled Random Access for Massive MIMO

- Users first transmit non-orthogonal pilots $\mathbf{s}_i \in \mathbb{C}^L$ for activity detection.
- BS sends scheduling message.
- Each user is assigned a unique (slot, orthogonal pilot) pair based on common feedback from the BS.



- The BS performs channel estimation using the orthogonal pilots, and then maximum ratio combining to reconstruct the payload.
- Each user is only required to transmit twice, in contrast to Coded ALOHA.

Scheduled Random Access vs. Coded Pilot Access

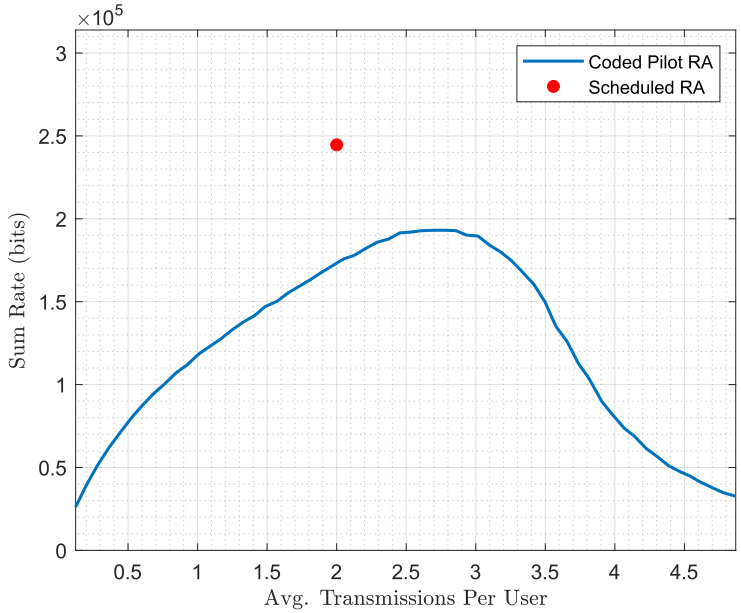


Figure: Throughput of Coded Pilot Access and Scheduled Random Access with $k = 1000$, $n = 10000$, SNR $= 10$dB, $M = 400$ BS antennas, $\tau = 64$ orthogonal pilots

- Each slot consists of $L = 300$ symbols.
  - Number of slots $\Delta = 20$ for coded pilot access;
  - Number of slots $\Delta = 17$ for scheduled random access.
- Activity detection is done via covariance method over one slot for SRA.
- Sum rate calculation assumes MRC beamforming and perfect SIC for CPA.
- Sum rate gain of 50 kbits at moderate cost of 1.44 kbits of feedback.

Conclusions

- Classic random access is contention based.
- Coded random access can alleviate some of the loss due to collision.
- If feedback is available from BS to the users:
  - BS can first detect the active users using sparse recovery methods;
  - BS can then schedule orthogonal pilots to users for channel estimation;
  - Finally, the users transmit their data to the BS.
- Significant performance improvement can be obtained at moderate feedback of 1.44 bits/user for scheduling.

Further Information

Justin Kang and Wei Yu,

"Minimum Feedback for Collision-Free Scheduling in Massive Random Access",

Submitted to IEEE Transactions on Information Theory, 2020.

[Online] available: https://arxiv.org/abs/2007.15497.