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

Wei Yu
Joint Work with Zhilin Chen, Foad Sohrabi

University of Toronto

November 2021
A 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)

This talk: User activity detection problem for a multi-cell system.
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 $s_n$
- Channel $h_n$ of user $n$ includes both (i.i.d.) Rayleigh and large-scale fading
- For single-cell system, received signal $Y \in \mathbb{C}^{L \times M}$ at the BS is

$$Y = \sum_{n=1}^{N} \alpha_n s_n h_n^T + Z = SX + Z,$$  \hspace{1cm} (1)

where

- $\alpha_n \in \{1, 0\}$ activity indicator; $Z \in \mathbb{C}^{L \times M}$ Gaussian noise with variance $\sigma^2$
- $S \triangleq [s_1, \ldots, s_N] \in \mathbb{C}^{L \times N}$; $X \triangleq [\alpha_1 h_1, \cdots, \alpha_N h_N]^T \in \mathbb{C}^{N \times M}$
Aim to identify the $K$ non-zero rows of $X$ from $Y = SX + Z$.

- Multiple measurement vector (MMV) problem in compressed sensing
  - Columns of $X$ share the same sparsity pattern, i.e., row sparsity
  - Efficiently solved by the approximate message passing (AMP) algorithm
  - Detecting $KM$ variables based on $LM$ observations.
Joint Activity Detection and Large-Scale Fading Estimation

Key Assumption: We only need activity $\alpha_n$ and do not need $h_n$.

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

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

- $\mathbf{S} \triangleq [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N] \in \mathbb{C}^{L \times N}$, signature matrix
- $\Gamma \triangleq \text{diag}\{\alpha_1 \beta_1, \alpha_2 \beta_2, \cdots, \alpha_N \beta_N\} \in \mathbb{R}^{N \times N}$, where $\beta_n$ is large-scale fading
- $\tilde{\mathbf{H}} \triangleq [\mathbf{h}_1/\sqrt{\beta_1}, \mathbf{h}_2/\sqrt{\beta_2}, \cdots, \mathbf{h}_N/\sqrt{\beta_N}]^T \in \mathbb{C}^{N \times M}$, normalized channel matrix
The maximum likelihood estimation of $\Gamma$ can be expressed as

$$\min_{\Gamma \geq 0} - \log p(Y|\Gamma) \propto \min_{\Gamma \geq 0} - \frac{1}{M} \sum_{m=1}^{M} \left( \log \frac{1}{\pi \Sigma} \exp \left( -y_m^H \Sigma^{-1} y_m \right) \right)$$

$$= \min_{\Gamma \geq 0} \log |\Sigma| + \text{tr} \left( \Sigma^{-1} \hat{\Sigma} \right) + \text{const.}.$$  \hspace{1cm} (3)

$$= \min_{\Gamma \geq 0} \log |S \Gamma S^H + \sigma^2 I| + \text{tr} \left( (S \Gamma S^H + \sigma^2 I)^{-1} \hat{\Sigma} \right) + \text{const.}.$$ \hspace{1cm} (4)

Sample covariance matrix over the antennas is a sufficient statistics:

$$\hat{\Sigma} \triangleq \frac{1}{M} \sum_{m=1}^{M} y_m y_m^H = \frac{1}{M} YY^H.$$ \hspace{1cm} (5)

The above problem can be solved efficiently using coordinate descent.

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

$$\begin{align*}
\begin{bmatrix}
\mathbf{L} \\
\mathbf{M}
\end{bmatrix}
\begin{bmatrix}
\mathbf{L} \\
\mathbf{M} \\
\mathbf{N}
\end{bmatrix}
+ \\
\begin{bmatrix}
\mathbf{L} \\
\mathbf{N} \\
\mathbf{M}
\end{bmatrix}
\end{align*}$$

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

$$\begin{align*}
\begin{bmatrix}
\mathbf{L} \\
\mathbf{M}
\end{bmatrix}
\begin{bmatrix}
\mathbf{M} \\
\mathbf{L} \\
\mathbf{N}
\end{bmatrix}
= \\
\begin{bmatrix}
\mathbf{L} \\
\mathbf{N} \\
\mathbf{M}
\end{bmatrix}
\begin{bmatrix}
\mathbf{L} \\
\mathbf{N} \\
\mathbf{M}
\end{bmatrix}
\end{align*}$$

$$M \rightarrow \infty$$

$$\begin{align*}
\begin{bmatrix}
\mathbf{L} \\
\mathbf{L} \\
\mathbf{N}
\end{bmatrix}
+ \\
\begin{bmatrix}
\mathbf{N} \\
\mathbf{L}
\end{bmatrix}
\end{align*}$$

**Crucial Advantage:** We now detect $K$ variables based on $L^2$ observations!
Let $\mathcal{I}$ be an index set corresponding to zero entries of $\gamma^0$, i.e., $\mathcal{I} \triangleq \{i \mid \gamma^0_i = 0\}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

\begin{align*}
\mathcal{N} & \triangleq \{ x \mid x^T J(\gamma^0) x = 0, x \in \mathbb{R}^N \}, \\
\mathcal{C} & \triangleq \{ x \mid x_i \geq 0, i \in \mathcal{I}, x \in \mathbb{R}^N \},
\end{align*}

where $x_i$ is the $i$-th entry of $x$. 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 $\mathcal{N} \cap \mathcal{C} = \{0\}$.

$\mathcal{N}$ is the “null space” of the Fisher Information Matrix $J(\gamma^0)$; $\mathcal{C}$ is a cone with non-negative entries indexed by the inactive users $\mathcal{I}$.

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 $\gamma^0$ even in the large $M$ limit.
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.
Suppose that large-scale fading is known, activity detection now becomes over the binary indicators $\alpha_n$:

$$\min_{\{\alpha_n\}} \log |S \Sigma S^H + \sigma^2 I| + \text{tr} \left((S \Sigma S^H + \sigma^2 I)^{-1} \Sigma\right)$$  \hspace{1cm} (8a)

subject to $\alpha_n \in \{0, 1\}, \ n = 1, 2, \ldots, N$ \hspace{1cm} (8b)

Binary $\alpha_n$ is challenging to deal with. We relax the constraint such that

$$\alpha_n \in [0, 1], \quad n = 1, 2, \ldots, N$$  \hspace{1cm} (9)

The relaxed problem can be solved by coordinate descent:

$$d = \min \left\{ \max \left\{ \frac{s_k^H \tilde{\Sigma}^{-1} \Sigma \tilde{\Sigma}^{-1} s_k - s_k^H \hat{\Sigma}^{-1} s_k}{\beta_k (s_k^H \tilde{\Sigma}^{-1} s_k)^2}, -\hat{\alpha}_k \right\}, 1 - \hat{\alpha}_k \right\}. \hspace{1cm} (10)$$

With unknown large-scale fading $\beta_n$, we estimate $\gamma_n = \alpha_n \beta_n$ in $[0, \infty]$. With known large-scale fading $\beta_n$, we estimate $\alpha_n$ in $[0, 1]$. 
The covariance method directly estimates the activity indicator $\alpha_n$ in $[0, 1]$ instead of $\gamma_n = \alpha_n \beta_n$ in $[0, \infty)$. Let $\alpha \triangleq [\alpha_1, \ldots, \alpha_N]^T$ and let the true value of $\alpha$ be $\alpha^0$.

**Theorem**

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

$$\mathcal{N}' \triangleq \{x | x^T J(\gamma^0) x = 0, x \in \mathbb{R}^N\}, \tag{11}$$

$$\mathcal{C}' \triangleq \{x | x_i \geq 0, i \in \mathcal{I}, x_i \leq 0, i \notin \mathcal{I}, x \in \mathbb{R}^N\}, \tag{12}$$

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

The extra constraint in the definition of $\mathcal{C}$ is due to that $\alpha_n$ is upper bounded.
Figure: Phase transition comparison of the cases with and without knowing large-scale fading. $N = 1000$. With known large-scale fading, $\alpha_n$ is both lower and upper bounded.

When $\frac{K}{N} \approx 1$, then inactive users are sparse!
User Activity Detection in Multicell Systems

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

- How to overcome the inter-cell interference?
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 $s_{bn}$;
- Received signal $Y_b \in \mathbb{C}^{L \times M}$ at BS $b$ is

$$Y_b = \sum_{n=1}^{N} \alpha_{bn} s_{bn} h_{bbn}^T + \sum_{j=1}^{B} \sum_{n=1}^{N} \alpha_{jn} s_{jn} h_{bjn}^T + Z_b$$

$$= S_b A_b G_{bb}^{1/2} \tilde{H}_{bb} + \sum_{j=1}^{B} S_j A_j G_{bj}^{1/2} \tilde{H}_{bj} + Z_b$$

$$= S_b \Gamma_{bb}^{1/2} \tilde{H}_{bb} + \sum_{j=1}^{B} S_j \Gamma_{bj}^{1/2} \tilde{H}_{bj} + Z_b$$

(13)

where
- $\alpha_{bn} \in \{1, 0\}$ activity indicator; $A_j \triangleq \text{diag}\{\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jN}\} \in \{0, 1\}^{N \times N}$;
- $h_{bjn} \in \mathbb{C}^{M \times 1}$ is the channel from user $n$ in cell $j$ to BS $b$
- $\tilde{H}_{bj} \triangleq \left[ h_{bj1}/\sqrt{\beta_{bj1}}, \cdots, h_{bjN}/\sqrt{\beta_{bjN}} \right]^T \in \mathbb{C}^{N \times M}$, normalized channel
- $S_j \triangleq [s_{j1}, s_{j2}, \cdots, s_{jN}] \in \mathbb{C}^{L \times N}$; $G_{bj} \triangleq \text{diag}\{\beta_{bj1}, \beta_{bj2}, \cdots, \beta_{bjN}\} \in \mathbb{R}^{N \times N}$
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 $G_{bj}, \forall b, j$ are known, the device activity detection problem can be formulated differently.

When $G_{bj}$ are not known, we need to estimate $\Gamma_{bj} = A_j G_{bj}, \forall b, j$, which has $B^2 N$ unknown parameters.

When $G_{bj}$ are known, we only need to estimate $A_b, \forall b$, which contains $BN$ 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 \( \Gamma_{bj} = A_j G_{bj}, \forall b, j \) from the received signals \( Y_b, \forall b \).

The likelihood function of \( Y_b \)'s given \( \Gamma_{bj} \)'s can be expressed as

\[
p(Y_1, \ldots, Y_B | \Gamma_{11}, \Gamma_{12}, \ldots, \Gamma_{BB}) = \prod_{b=1}^{B} p(Y_b | \Gamma_{11}, \Gamma_{12}, \ldots, \Gamma_{BB}) = \prod_{b=1}^{B} \frac{1}{\pi \sum_b |M|} \exp \left( - \text{tr} \left( M \Sigma_b^{-1} \hat{\Sigma}_b \right) \right).
\]

(14)

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

\[
\min_{\{\Gamma_{bj}\}} \sum_{b=1}^{B} \left( \log |\Sigma_b| + \text{tr} \left( \Sigma_b^{-1} \hat{\Sigma}_b \right) \right)
\]

(15a)

s. t. \( \gamma_{bjn} \in [0, \infty), \forall b, j, n \)

(15b)

This is a challenging problem to solve.
Cooperative Detection with Known Large-scale Fading

Assuming that all large-scale fading matrices $G_{bj}$'s, we directly estimate the device activity $A_b$'s using the MLE. The likelihood function of $Y_b$'s can be expressed as

$$p(Y_1, \ldots, Y_B | A_1, \ldots, A_B) = \prod_{b=1}^{B} p(Y_b | A_1, \ldots, A_B)$$

$$= \prod_{b=1}^{B} \frac{1}{|\pi \Sigma_b|^M} \exp \left( - \text{tr} \left( M \Sigma_b^{-1} \hat{\Sigma}_b \right) \right). \quad (16)$$

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

$$\min_{\{A_b\}} \sum_{b=1}^{B} \left( \log |\Sigma_b| + \text{tr} \left( \Sigma_b^{-1} \hat{\Sigma}_b \right) \right) \quad (17a)$$

s. t. \hspace{1em} $\alpha_{bn} \in \{0, 1\}, \forall b, n \quad (17b)$

**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.
Let $\alpha \triangleq [\alpha_1, \ldots, \alpha_{BN}]^T$ be the activity indicators, and let true value of $\alpha$ be $\alpha^0$.

**Theorem**

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

\[ \mathcal{N}'' \triangleq \{ x \mid x^T J(\gamma^0) x = 0, x \in \mathbb{R}^{BN} \} , \]  

\[ C'' \triangleq \{ x \mid x_i \geq 0, i \in I, x_i \leq 0, i \notin I, x \in \mathbb{R}^{BN} \} , \]  

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

All the dimensions scale by $B$, so we expect the same phase transition curve.
Phase Transition of Multicell vs. Single-Cell

**Figure:** Phase transition comparison for multicell system with $B = 7$, $N = 200$. 
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.
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.

- Covariance based MLE can detect activities without estimating the channels.

- Analysis and algorithm can be extended from single-cell to multicell systems.

- Multicell cooperative detection has a similar phase transition as single-cell system, \textit{but only if the large-scale fading is known}.

- In practice, there is a performance degradation due to intercell interference.
Zhilin Chen, Foad Sohrabi, Ya-Feng Liu, Wei Yu,
“Phase Transition Analysis for Covariance Based Massive Random Access with Massive MIMO”,

Zhilin Chen, Foad Sohrabi, and Wei Yu,
“Sparse Activity Detection in Multi-Cell Massive MIMO Exploiting Channel Large-Scale Fading”,