Massive Random Access with Massive MIMO: Covariance Based Detection

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

University of Toronto
- 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)
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

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

(1)

where

- $\alpha_n \in \{1, 0\}$ activity indicator; $Z \in \mathbb{C}^{L \times M}$ Gaussian noise with variance $\sigma^2$
- $S \triangleq [s_1, \ldots, s_N] \in \mathbb{C}^{L \times N}$; $X \triangleq [\alpha_1 h_1, \ldots, \alpha_N h_N]^T \in \mathbb{C}^{N \times M}$
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 $\mathbf{Y} = \mathbf{S}\mathbf{X} + \mathbf{Z}$ [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 $\mu_{\mathbf{y} \rightarrow \mathbf{x}}$ as Gaussian in the large system limit $N, L \rightarrow \infty$
- Further simplify the messages such that only $N + L$ messages are tracked
Intuitive Interpretation of AMP

- Matched filtering \(\rightarrow\) Denoising \(\rightarrow\) Computing and correcting the residual

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

(2)

**Figure:** Soft thresholding function with \(\theta = 1\)
AMP Algorithm for MIMO

- The AMP algorithm expressed in matrix form:

\[ X^{t+1} = \eta_t(S^H R^t + X^t), \]

\[ R^{t+1} = Y - SX^{t+1} + \frac{N}{L} R^t \langle \eta'_t(S^H R^t + X^t) \rangle, \]

where

- \( X^{t+1} \), estimate at iteration \( t + 1 \);
- \( R^{t+1} \), 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: \( O(NLM) + \) 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 $\alpha_n$ and do not need $h_n$.

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

$$Y = \sum_{n=1}^{N} \alpha_n s_n h_n^T + Z \triangleq S \Gamma^{1/2} \tilde{H} + Z$$

- $S \triangleq [s_1, s_2, \cdots, s_N] \in \mathbb{C}^{L \times N}$, 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
- $\tilde{H} \triangleq [h_1/\sqrt{\beta_1}, h_2/\sqrt{\beta_2}, \cdots, h_N/\sqrt{\beta_N}]^T \in \mathbb{C}^{N \times M}$, normalized channel matrix
Statistics of the Received Signal

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

$$y_m = S\Gamma^{\frac{1}{2}}\tilde{h}_m + z_m$$  \hspace{1cm} (6)

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

Wei Yu (University of Toronto)

Covariance Based Detection  

9 / 64
Maximum Likelihood Estimation of $\Gamma$

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

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

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

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

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

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

$$= \log |\Sigma| + \text{tr} \left( \Sigma^{-1} \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}_m \mathbf{y}_m^H \right) + \text{const.} \quad (7)$$
Define the sample covariance matrix of the received signal as

\[ \hat{\Sigma} \triangleq \frac{1}{M} \sum_{m=1}^{M} y_m y_m^H = \frac{1}{M} YY^H. \] (8)

With the sample covariance matrix, the MLE of \( \Gamma \) can be expressed as

\[
\min_{\Gamma \succeq 0} f(\Gamma) := \log |\Sigma| + \text{tr} \left( \Sigma^{-1} \hat{\Sigma} \right) + \text{const.} \\
= \log |\mathbf{S}\Gamma\mathbf{S}^H + \sigma^2 \mathbf{I}| + \text{tr} \left( (\mathbf{S}\Gamma\mathbf{S}^H + \sigma^2 \mathbf{I})^{-1} \hat{\Sigma} \right) + \text{const.} \] (9)

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

Covariance Based Sparse Activity Detection

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

$$L \underbrace{\begin{pmatrix} \cdots \end{pmatrix}}_{M} + \begin{pmatrix} \cdots \end{pmatrix}$$

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

$$L \underbrace{\begin{pmatrix} \cdots \end{pmatrix}}_{M} \times \underbrace{\begin{pmatrix} \cdots \end{pmatrix}}_{N} = \underbrace{\begin{pmatrix} \cdots \end{pmatrix}}_{M} \times \underbrace{\begin{pmatrix} \cdots \end{pmatrix}}_{N} + \begin{pmatrix} \cdots \end{pmatrix}$$

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

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

**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 \( \Gamma \), need to solve the optimization problem

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

\[
= \log |S\Gamma S^H + \sigma^2 I| + \text{tr} \left( (S\Gamma S^H + \sigma^2 I)^{-1} \hat{\Sigma} \right) + \text{const.} \tag{10}
\]

- \( f(\Gamma) \) 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, \( f(\Gamma) \) is minimized by the true value \( \Gamma^0 \):

\[
\hat{\Sigma} \triangleq \frac{1}{M} \sum_{m=1}^M y_m y_m^H \rightarrow \Sigma^0 \triangleq S\Gamma^0 S^H + \sigma^2 I, \quad \text{as } M \rightarrow \infty. \tag{11}
\]

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

\[
\min \log |\Sigma| + \text{tr}(\Sigma^{-1} \Sigma^0). \tag{12}
\]

By taking derivative, we see \( \Sigma^{\text{opt}} = \Sigma^0 \). For finite \( M \), we need to solve (10).
Coordinate Descent for Solving the MLE problem

Let $\gamma_n$ be the $n$-th diagonal entry of $\Gamma$. The MLE can be expressed as

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

(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

$$
\min_{d \geq -\hat{\gamma}_k} \log \left( 1 + ds_k^H \tilde{\Sigma}^{-1} s_k \right) - \frac{ds_k^H \tilde{\Sigma}^{-1} \hat{\Sigma} \tilde{\Sigma}^{-1} s_k}{1 + ds_k^H \tilde{\Sigma}^{-1} s_k}.
$$

(14)

- $\tilde{\Sigma} = \sum_{n=1}^{N} \hat{\gamma}_n s_n s_n^H + \sigma^2 I$ 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

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

(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

1: Initialize $\hat{\gamma} = 0$, $\tilde{\Sigma} = \sigma^2 I$, $\tilde{\Sigma}^{-1} = \sigma^{-2} I$.
2: for $i = 1, 2, \ldots, T$ do
3: Select a random permutation $i_1, i_2, \ldots, i_N$ of indices $\{1, 2, \ldots, N\}$.
4: for $n = 1$ to $N$ do
5: $d = \max \left\{ \frac{s_{in}^H \hat{\Sigma} \hat{\Sigma}^{-1} s_{in} - s_{in}^H \tilde{\Sigma}^{-1} s_{in}}{(s_{in}^H \hat{\Sigma}^{-1} s_{in})^2}, -\hat{\gamma}_{i_n} \right\}$
6: $\hat{\gamma}_{i_n} = \hat{\gamma}_{i_n} + d$ ← Coordinate descent update
7: $\tilde{\Sigma}^{-1} = \tilde{\Sigma}^{-1} - d \frac{\tilde{\Sigma}^{-1} s_{in} s_{in}^H \tilde{\Sigma}^{-1}}{1 + ds_{in}^H \tilde{\Sigma}^{-1} s_{in}}$ ← Rank-1 update of estimated covariance
8: end for
9: end for
10: Output $\hat{\gamma} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_N]^T$. Declare the activity with thresholding on $\hat{\gamma}$. 
Covariance Matching Approach

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

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

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

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

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

- 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.](image)

**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 \( \beta_n \) is known at the BS, only the activities \( \alpha_n \) need to be estimated.

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

\[
\min_{\alpha_1, \ldots, \alpha_N} f(\alpha_1, \ldots, \alpha_N) : = - \frac{1}{M} \log p(\mathbf{Y} | \alpha_1, \ldots, \alpha_N)
\]

\[
= - \frac{1}{M} \sum_{m=1}^{M} \log p(\mathbf{y}_m | \alpha_1, \ldots, \alpha_N)
\]

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

\[
= \log |\Sigma| + \text{tr} \left( \Sigma^{-1} \hat{\Sigma} \right) + \text{const.} \tag{18}
\]

Note that \( p(\mathbf{y}_m | \alpha_1, \ldots, \alpha_N) \) remains Gaussian with covariance \( \Sigma \).
Activity Detection with Known Large-Scale Fading

- The problem of detecting the binary activity indicator $\alpha_n$ is now:

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

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

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

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

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

\[
d = \min \left\{ \max \left\{ \frac{s_k^H \hat{\Sigma}^{-1} \hat{\Sigma}^{-1} s_k - s_k^H \hat{\Sigma}^{-1} s_k}{\beta_k (s_k^H \hat{\Sigma}^{-1} s_k)^2}, -\hat{\alpha}_k \right\}, 1 - \hat{\alpha}_k \right\} \tag{21}
\]

- 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 $[0, 1]$. 

Wei Yu (University of Toronto)
Recap of Problem Formulations

- Sparse user activity detection with channel $\alpha_n h_n \sim \alpha_n \sqrt{\beta_n} \mathcal{CN}(0, I)$:

- If channel estimate is needed for subsequent data transmission:
  - We can use AMP, which gives a rough estimate of the instantaneous $h_n$.
- If only user activities ($\alpha_n$) 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 ($\beta_n$) is known:
  - We can modify the covariance method to estimate $\alpha_n$. 

- Active Users
- Inactive Users
# Comparison of AMP vs Covariance Approaches

<table>
<thead>
<tr>
<th>Derived from</th>
<th>Compressed Sensing (AMP)</th>
<th>Covariance Based Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approx. marginals of $p(X, Y)$</td>
<td>Maximization of $p(Y</td>
<td>\Gamma)$</td>
</tr>
<tr>
<td>Prior needed</td>
<td>Sparsity level for design of $\eta_t(\cdot)$</td>
<td>None (deterministic $\Gamma$)</td>
</tr>
<tr>
<td>Estimate</td>
<td>Activities $\alpha_n$ and channels $h_n$</td>
<td>Activities $\alpha_n$ and large-scale fading $\beta_n$</td>
</tr>
<tr>
<td>Preferred regime</td>
<td>Fix $\epsilon \triangleq \frac{K}{N}$, $\delta \triangleq \frac{L}{N}$, and $M$ Let $N, L, K \to \infty$</td>
<td>Fix $N, K, L$ Let $M \to \infty$</td>
</tr>
<tr>
<td>Complexity</td>
<td>Roughly $O(NLM)$ per iteration</td>
<td>Roughly $O(NL^2)$ via CD per iteration</td>
</tr>
</tbody>
</table>
Remark on Complexity

Figure: Numerical runtime for $\hat{\gamma}$ to converge to an $\epsilon$ norm ball around $\gamma$ 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 $L$. 
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 $13$dBm; 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 $M$.

  \[ \mathbf{X}^{t+1} = (1 - \alpha) \eta_t (\mathbf{S}^H \mathbf{R}^t + \mathbf{X}^t) + \alpha \mathbf{X}^t, \]

  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 $M = 8$

![Graph 1](image1.png)

**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 $L = 100$

![Graph showing performance and complexity of AMP vs covariance based estimation](image)

**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 $L = 100$

![Graph showing error probability and running time comparison between AMP and Covariance based methods.](image)

**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 $K \gtrsim L$.

- **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 \( \hat{\Sigma} (M \to \infty) \), we plot the estimation error of \( \Gamma \) under different \((K, L)\) with \( N = 2000 \)
Analysis
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 \( \tau_t \) tracked by state evolution equation

\[
\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E} \left| \eta_t(X + \tau_t Z) - X \right|^2
\]

(22)

for the \( M = 1 \) case.

- Interpretation of state evolution: Vector estimation \( \mathbf{y} = \mathbf{Sx} + \mathbf{z} \) is reduced to uncoupled scalar estimation \( (\mathbf{x}_t + \mathbf{S}^H \mathbf{r}_t)_i = x_i + v_i^t \)
Recall the MLE formulation, and let $\gamma$ denote the diagonal entries of $\Gamma$

$$f(\gamma) : = -\frac{1}{M} \log p(Y|\gamma) = -\frac{1}{M} \sum_{m=1}^{M} \log p(y_{m}|\gamma)$$

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

- 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 $J(\gamma)$, plays a critical role in the asymptotic analysis. The $(i,j)$-th entry of $J(\gamma)$ is defined as

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

- Key assumption for the analysis: $M \to \infty$. 
The Fisher Information matrix can be also written as the negative expected second derivative of the log-likelihood function

\[
[J(\gamma)]_{ij} = E \left[ \frac{\partial \log p(Y|\gamma)}{\partial \gamma_i} \frac{\partial \log p(Y|\gamma)}{\partial \gamma_j} \right] = -E \left[ \frac{\partial^2 \log p(Y|\gamma)}{\partial \gamma_i \partial \gamma_j} \right]
\] (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 $\gamma$ be a parameter, and let $\hat{\gamma}$ be an unbiased estimator of $\gamma$. Then the covariance of estimation error is lower bounded by

$$\mathbb{E} [(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)^T] \geq J^{-1}(\gamma) \quad (26)$$

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

  Consistency: $\hat{\gamma}^{ML} \xrightarrow{P} \gamma \quad (27)$

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

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 $\gamma^0$ is **identifiable**, i.e., there exists no other $\gamma' \neq \gamma^0$ with
    \[ p(Y|\gamma^0) = p(Y|\gamma'). \] (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
    \[ N \gg L^2, \] (30)
    i.e., when the dimension of $\gamma^0$ is larger than the dimensions of the sample covariance $\hat{\Sigma}$, there are too many parameters to estimate.
  - The true parameter $\gamma^0$ in fact always lies on the **boundary** of its parameter space $[0, \infty)^N$, because most of the entries of $\gamma^0$ are zero.

Need new analysis!
We first derive the Fisher information matrix for the activity detection problem:

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

(31)

• \(p(y_m|\gamma)\) is Gaussian, the second derivative of \(\log p(Y|\gamma) = \sum_m \log p(y_m|\gamma)\) is

\[
\frac{\partial^2 \log p(Y|\gamma)}{\partial \gamma_i \partial \gamma_j} = M \text{tr}(\Sigma^{-1} s_j s_j^H \Sigma^{-1} s_i s_i^H) - M \text{tr}(\Sigma^{-1} s_j s_j^H \Sigma^{-1} s_i s_i^H \Sigma^{-1} \hat{\Sigma}) \\
- M \text{tr}(\Sigma^{-1} s_i s_i^H \Sigma^{-1} s_j s_j^H \Sigma^{-1} \hat{\Sigma}).
\]

(32)

• Taking the expectation of \(\hat{\Sigma}\) using the fact that \(\mathbb{E}[\hat{\Sigma}] = \Sigma\) gives

\[
-\mathbb{E} \left[ \frac{\partial^2 \log p(Y|\gamma)}{\partial \gamma_i \partial \gamma_j} \right] = M(s_i^H \Sigma^{-1} s_j)(s_j^H \Sigma^{-1} s_i).
\]

(33)
The Fisher information matrix $J(\gamma)$ can be further written in a matrix form as

$$J(\gamma) = M (P \odot P^*) ,$$

(34)

where $P \triangleq S^H (S\Sigma S^H + \sigma^2 I)^{-1} S$; $\odot$ element-wise product; $(\cdot)^*$ conjugate

$J(\gamma)$ is a real symmetric matrix of dimensions $N \times N$, whose rank satisfies:

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

(35)

where

(a) is due to $\text{Rank}(U \odot V) \leq \text{Rank}(U) \text{Rank}(V)$;

(b) is due to $\text{Rank}(P) \leq \text{Rank}(S) \leq \min\{N, L\}$.

Thus $J(\gamma)$ is rank-deficient if $N > L^2$ since $P \odot P^*$ is of size $N \times N$.

Our new analysis takes rank-deficiency of $J(\gamma)$ into consideration
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 $\hat{\gamma}^{ML}$ can approach the true parameter $\gamma^0$ as $M \to \infty$?
- This helps identify the desired operating regime of the system parameters for getting an accurate estimate $\hat{\gamma}^{ML}$ via MLE with massive MIMO.

- If $M$ is finite, how is the estimation error $\hat{\gamma} - \gamma^0$ 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 $\gamma^0$, 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

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

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

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 $\gamma^0$ 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 $\mathbb{R}_+^N$.
- $\mathcal{N} \cap \mathcal{C} = \{0\}$ ensures that the true parameter $\gamma^0$ 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 $\mathcal{N} \cap \mathcal{C} = \{0\}$ is that $\dim(\mathcal{N}) < |I|$.
- Since $\dim(\mathcal{N})$ is roughly $N - L^2$ and $|I| = N - K$, 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 $J(\gamma^0) \in \mathbb{R}^{N \times N}$ as follows.

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

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

\[
\begin{align*}
\text{find} & \quad x & \quad \text{(38a)} \\
\text{subject to} & \quad (A - BC^{-1}B^T)x > 0, & \quad \text{(38b)}
\end{align*}
\]

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

Note that matrix $M$ satisfying $M^T x > 0$ for some $x$, 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 $J(\gamma^0)$ and that $\forall M$: (i) $Mx = 0$ has no solution for $x \geq 0$ and $x \neq 0$, is equivalent to (ii) $M^T x > 0$ has solutions.
Covariance Matching Perspective

Analyzing the optimization problem:

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

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

**Intuition**

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

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

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

**Proposition**

Let \( \hat{S} \in \mathbb{C}^{L^2 \times N} \) be the column-wise Kronecker product (Khatri-Rao product) of \( S^* \) and \( S \), i.e., \( \hat{S} \triangleq [s_1^* \otimes s_1, \ldots, s_N^* \otimes s_N] \). We define a set \( \tilde{N} \) in the space \( \mathbb{R}^N \) as

\[
\tilde{N} \triangleq \{ x \mid \hat{S}x = 0, x \in \mathbb{R}^N \}.
\] (40)

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

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

**Proposition**

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

Key advantage of using \( \tilde{N} \) is that it depends only on \( S \) and is independent of SNR.
Alternative Way to Numerically Verify the Condition

**Theorem**

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

\[
\begin{align*}
\{ & \text{Re}(r_i^T) \odot \text{Re}(r_j^T) + \text{Im}(r_i^T) \odot \text{Im}(r_j^T), \ 1 \leq i \leq j \leq L \}, \\
\{ & \text{Re}(r_i^T) \odot \text{Im}(r_j^T) - \text{Im}(r_i^T) \odot \text{Re}(r_j^T), \ 1 \leq i < j \leq L \}. 
\end{align*}
\]

(41) \hspace{1cm} (42)

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

\[
\begin{align*}
\text{find} \quad & x \\
\text{subject to} \quad & Dx = 0, \\
& \|x\|_1 = 1, \\
& x_i \geq 0, \ i \in \mathcal{I},
\end{align*}
\]

(43a) \hspace{1cm} (43b) \hspace{1cm} (43c) \hspace{1cm} (43d)

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

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

(44)

- The derivation is based on restricted isometry property in compressed sensing.
- It implies that the error vanishes as $M \to \infty$, because $\hat{\Sigma} \to S\Gamma^0 S^H + \sigma^2 I$.
- The result is for specific sequence $S$.
- 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 $Ax = b$ with $A$ satisfying restricted isometry property, the number of measurements needed to recover a $K$-sparse vector $x$ of length-$N$ is
  \[ L = O(K \log(N/K)). \]
- For $\hat{\Sigma} = S\Gamma S^H + \sigma^2 I$ with $\hat{S}$ satisfying robust null space property, the number of measurements needed to recover a $K$-sparse diagonal matrix $\Gamma$ of size $N^2$ is
  \[ L^2 = O(K \log^2(N/K)). \]
- Based on the same robust NSP of $\hat{S}$, we can derive the scaling law of MLE:

\[
\begin{align*}
L & = \text{Diagram with dimensions}
\end{align*}
\]

Theorem

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

Wei Yu (University of Toronto)  Covariance Based Detection  47 / 64
**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 $\hat{\Sigma}$ ($M \rightarrow \infty$), we plot the estimation error of $\Gamma$ under different $(K/N, L^2/N)$ with $N = 2000$

Performance of coordinate descent algorithm is very close to the optimal MLE!
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., $\mathcal{I} \triangleq \{ i \mid \alpha_i^0 = 0 \}$. We define two sets $\mathcal{N}, \mathcal{C}$ in the space $\mathbb{R}^N$, respectively, as follows

\[
\mathcal{N} \triangleq \{ \mathbf{x} \mid \mathbf{x}^T J(\gamma^0) \mathbf{x} = 0, \mathbf{x} \in \mathbb{R}^N \}, \quad (45)
\]
\[
\mathcal{C} \triangleq \{ \mathbf{x} \mid x_i \geq 0, i \in \mathcal{I}, x_i \leq 0, i \notin \mathcal{I}, \mathbf{x} \in \mathbb{R}^N \}, \quad (46)
\]

where $x_i$ is the $i$-th entry of $\mathbf{x}$. Then a necessary and sufficient condition for the consistency of $\hat{\alpha}^{ML}$, i.e., $\hat{\alpha}^{ML} \to \alpha^0$ 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 $\alpha_n$ 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!
Asymptotic Distribution of the ML Estimation Error

For MLE solutions not on boundary, we have \( \sqrt{M}(\hat{\gamma}_{ML} - \gamma) \xrightarrow{D} N(0, MJ^{-1}(\gamma)) \).

For MLE with boundary constraint: \( C \triangleq \{ x \mid x_i \geq 0, i \in I, x \in \mathbb{R}^N \} \):

**Theorem**

\( \text{Let } x \in \mathbb{R}^{N \times 1} \sim N(0, MJ^\dagger(\gamma^0)) \), where \( \dagger \) denotes Moore-Penrose inverse. Let \( \mu \in \mathbb{R}^{N \times 1} \) be a solution to the constrained quadratic programming problem

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{M} (x - \mu)^T J(\gamma^0)(x - \mu) \\
\text{subject to} & \quad \mu \in C,
\end{align*}
\]

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

Note that \( \mu \) is random due to the randomness of \( x \).

Detection error can be characterized based on the distribution of \( \hat{\gamma}_{ML} - \gamma^0 \).
**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 $M = 256$. 
**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 $s_{bn}$;
- Received signal $Y_b \in \mathbb{C}^{L \times M}$ at BS $b$ is

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

$$= S_b X_{bb} + \sum_{j=1,j\neq b}^{B} S_j X_{bj} + Z_b, \quad (48)$$

where

- $\alpha_{bn} \in \{1, 0\}$ activity indicator; $Z_b \in \mathbb{C}^{L \times M}$ Gaussian noise with variance $\sigma^2$.
- $h_{bjn} \in \mathbb{C}^{M \times 1}$ is the channel from user $n$ in cell $j$ to BS $b$.
- $S_j \triangleq [s_{j1}, \cdots, s_{jN}] \in \mathbb{C}^{L \times N}$; $X_{bj} \triangleq [\alpha_{j1} h_{bj1}, \cdots, \alpha_{jN} h_{bjN}]^{T} \in \mathbb{C}^{N \times M}$

The inter-cell interference brings performance degradation for activity detection.
To use the covariance approach, the signal at BS $b$ is re-written as

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

$$
= S_b A_b G_{bb}^{\frac{1}{2}} \tilde{H}_{bb} + \sum_{j=1, j \neq b}^{N} S_j A_j G_{bj}^{\frac{1}{2}} \tilde{H}_{bj} + Z_b
$$

$$
= S_b \Gamma_{bb}^{\frac{1}{2}} \tilde{H}_{bb} + \sum_{j=1, j \neq b}^{N} S_j \Gamma_{bj}^{\frac{1}{2}} \tilde{H}_{bj} + Z_b
$$

(49)

- $S_j \triangleq [s_{j1}, s_{j2}, \cdots, s_{jN}] \in \mathbb{C}^{L \times N}$; $A_j \triangleq \text{diag}\{\alpha_{j1}, \alpha_{j2}, \cdots, \alpha_{jN}\} \in \{0, 1\}^{N \times N}$
- $G_{bj} \triangleq \text{diag}\{\beta_{bj1}, \beta_{bj2}, \cdots, \beta_{bjN}\} \in \mathbb{R}^{N \times N}$ large-scale fading matrix
- $\Gamma_{bj} \triangleq \text{diag}\{\alpha_{j1}\beta_{bj1}, \alpha_{j2}\beta_{bj2}, \cdots, \alpha_{jN}\beta_{bjN}\} \in \mathbb{R}^{N \times N}$
- $\tilde{H}_{bj} \triangleq \left[ h_{bj1}/\sqrt{\beta_{bj1}}, \cdots, h_{bjN}/\sqrt{\beta_{bjN}} \right]^T \in \mathbb{C}^{N \times M}$, normalized channel

Similar to single-cell case, all $\Gamma_{bj}$ are treated as deterministic unknown parameters and all $\tilde{H}_{bj}$ are treated as random samples.
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 $B^2N$ 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

$$p(Y_1, \ldots, Y_B | \Gamma_{11}, \Gamma_{12}, \ldots, \Gamma_{BB}) = \prod_{b=1}^{B} p(Y_b | \Gamma_{11}, \Gamma_{12}, \ldots, \Gamma_{BB})$$

$$= \prod_{b=1}^{B} \frac{1}{|\pi \Sigma_b|^M} \exp \left( - \text{tr} \left( M \Sigma_b^{-1} \hat{\Sigma}_b \right) \right). \quad (50)$$

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

$$\min_{\{\Gamma_{bj}\}} \sum_{b=1}^{B} \left( \log |\Sigma_b| + \text{tr} \left( \Sigma_b^{-1} \hat{\Sigma}_b \right) \right) \quad (51a)$$

s. t. $\gamma_{bjn} \in [0, \infty), \forall b, j, n \quad (51b)$

- The problem can be solved using coordinate descent.
Cooperative Detection with Known Large-scale Fading

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

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

$$= \prod_{b=1}^{B} \frac{1}{\pi |\Sigma_b|^M} \exp \left( - \operatorname{tr} \left( M \Sigma_b^{-1} \hat{\Sigma}_b \right) \right). \quad (52)$$

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

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

$$\text{s. t.} \quad \alpha_{bn} \in \{0, 1\}, \forall b, n \quad (53a)$$

$$\text{Known large-scale fading:} \text{ Single-cell and multicell have same phase transition}$$

- 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.
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 $K = O(L)$. 
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”,