Statistical Device Activity Detection for Massive Grant-free Access

Ying Cui

Department of Electrical Engineering
Shanghai Jiao Tong University
China

IEEE ICC Tutorial 2021

Outline

Outline

Internet of Things (IoT)

- IoT describes physical objects that connect and exchange data with other devices and systems over the Internet or other communications networks
  - Things include sensors, robots, smart meters, vehicles, etc.
- Typical IoT applications include smart health care, smart homes, smart manufacturing, smart transportation, smart surveillance, etc.
- IoT will impact the way we live and work in near future

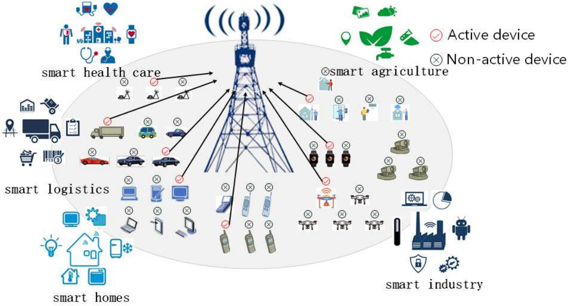


Figure: Typical IoT applications

Massive machine-type communication (mMTC)

- mMTC provides connections to a large number of devices that intermittently transmit small amount of traffic without the involvement of a human
  - The total number of connected devices in the world will be approximately 75.44 billion in 2025 and 125 billion in 2030
  - Very few devices from a large number of potential devices are active and send data at a time
  - Key performance indicators (KPIs) include number of connected devices, reliability, latency, etc.
- mMTC has been identified as one of the three main use cases for 5G, along with enhanced mobile broadband (eMBB) and ultra reliable, low-latency communications (URLLC)

Grant-free access for mMTC

- Grant-free access is proposed to eliminate the dynamic scheduling request and grant signaling overhead for uplink data transmission in mMTC
- Grant-free access relies on non-orthogonal pilot sequences (preambles) and operates in two phases
  - Each device is assigned a unique non-orthogonal pilot sequence, also serving as device ID
  - In phase I, active devices send pilot sequences, and the BS detects device activities and estimates active devices' channels
  - In phase II, active devices directly transmit data, and the BS detects transmitted data

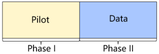


Figure: Grant-free access.

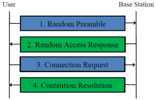


Figure: Grant-based random access.

Grant-free access	Grant-based random access
Pre-Assigned preambles	Random preambles
Unique preambles (ID)	Non-unique preambles
Non-orthogonal preambles	Orthogonal preambles
Access grant not needed	access grant needed
Accurate detection of colliding users	Accurate detection of non-colliding users
High access success probability	Low access success probability
High data transmission efficiency	Low data transmission efficiency
Low terminal energy consumption	High terminal energy consumption
High complexity	Low complexity

Table: Grant-free access vs. grant-based random access.

Grant-free access for mMTC

- Challenges of grant-free access
  - Activity detection and channel estimation for colliding devices with non-orthogonal pilot sequences
- Three application types
  - Devices just report their activities and do not send data
    - Device activity detection is sufficient
  - Active devices have very few data to send
    - Data can be embedded into pilots, and joint activity and data detection (extension of device activity detection) can be conducted
  - Active devices have more data to send
    - Separate activity detection and channel estimation (for detected devices with conventional methods) or joint activity detection and channel estimation can be conducted
- Focus on device activity detection (more fundamental)

Outline

Network model

- Consider a single-cell cellular network with one M-antenna BS and a large number N of single-antenna IoT devices
  - Denote $\mathcal{M} \triangleq \{1, \cdots, M\}$ and $\mathcal{N} \triangleq \{1, \cdots, N\}$
- Device activity patterns for IoT traffic are sporadic
  - Very few devices among all potential devices are active and access the BS at a time
- The device activity states, $\boldsymbol{\alpha} \triangleq (\alpha_n)_{n \in \mathcal{N}} \in \{0, 1\}^N$, are unknown to the BS and to be estimated
  - Can be modeled as unknown deterministic quantities or random variables with a known prior distribution
- Each device n is assigned a unique length-L pilot sequence $\mathbf{s}_n \in \mathbb{C}^L$, known to the BS
- The large-scale fading powers, $\mathbf{g} \triangleq (g_n)_{n \in \mathcal{N}} \in \mathbb{R}_{++}^N$, are assumed to be known to the BS
  - Can be jointly estimated with device activities if unknown
- Small-scale fading follows the block-fading channel model
- In each coherence block, all active devices synchronously send their pilots, and the BS detects the device activities

Flat Rayleigh fading model and receive signal

- Consider a narrow-band system
- Adopt the flat Rayleigh fading model for small-scale fading
  - $\mathrm{h}_n \in \mathbb{C}^M$ denotes the small-scale fading coefficients of device n
  - All elements of $\mathrm{h}_n, n \in \mathcal{N}$ are i.i.d. $\mathcal{CN}(0,1)$
- The receive signal over the L signal dimensions and M antennas, $\mathrm{Y} \in \mathbb{C}^{L\times M}$, is:

$$\mathrm{Y} = \sum_{n\in\mathcal{N}} \mathrm{s}_n \alpha_n \sqrt{g_n}\mathrm{h}_n^T + \mathrm{Z} = \mathrm{SAG}^{\frac{1}{2}}\mathrm{H} + \mathrm{Z}$$

  - $\mathrm{S} \triangleq [\mathrm{s}_1, \cdots, \mathrm{s}_N] \in \mathbb{C}^{L\times N}$ represents the pilot matrix
  - $\mathrm{A} \triangleq \mathrm{diag}(\boldsymbol{\alpha}) \in \{0,1\}^{N\times N}$ represents the device activities
  - $\mathrm{G} \triangleq \mathrm{diag}(\mathrm{g}) \in \mathbb{R}_{++}^{N\times N}$ represents the large-scale fading powers
  - $\mathrm{H} \triangleq [\mathrm{h}_1, \cdots, \mathrm{h}_N]^T \in \mathbb{C}^{N\times M}$ represents the small-scale fading coefficients, with all elements i.i.d. $\mathcal{CN}(0,1)$
  - $\mathrm{Z} \in \mathbb{C}^{L\times M}$ represents the additive white Gaussian noise (AWGN), with all elements i.i.d. $\mathcal{CN}(0,\sigma^2)$

Statistics of receive signal

- The receive signal at the m-th antenna, $\mathsf{Y}_{:,m} \in \mathbb{C}^{L}$, is:

$$\mathsf{Y}_{:,m} = \mathsf{SAG}^{\frac{1}{2}}\mathsf{H}_{:,m} + \mathsf{Z}_{:,m}$$

- Given device activities $\boldsymbol{\alpha}$, $\mathsf{Y}_{:,m}, m \in \mathcal{M}$ are i.i.d. $\mathcal{CN}(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \triangleq \mathsf{SAGS}^{H} + \sigma^2 \mathsf{I}_L \in \mathbb{C}^{L\times L}$
    - $\mathsf{H}_{:,m}, \mathsf{Z}_{:,m}, m \in \mathcal{M}$ are i.i.d. $\mathcal{CN}(0, \mathsf{I}_L)$
    - $\mathbb{E}\left[\mathsf{Y}_{:,m}\right] = \mathsf{SAG}^{\frac{1}{2}}\mathbb{E}\left[\mathsf{H}_{:,m}\right] + \mathbb{E}\left[\mathsf{Z}_{:,m}\right] = 0$
    - $\mathbb{E}\left[\mathsf{Y}_{:,m}\mathsf{Y}_{:,m}^{H}\right] = \mathsf{SAG}^{\frac{1}{2}}\mathbb{E}\left[\mathsf{H}_{:,m}\mathsf{H}_{:,m}^{H}\right]\mathsf{G}^{\frac{1}{2}}\mathsf{AS}^{H} + \mathsf{SAG}^{\frac{1}{2}}\mathbb{E}\left[\mathsf{H}_{:,m}\mathsf{Z}_{:,m}^{H}\right] + \mathbb{E}\left[\mathsf{H}_{:,m}^{H}\mathsf{G}^{\frac{1}{2}}\mathsf{AS}^{H}\mathsf{Z}_{:,m}\right] + \mathbb{E}\left[\mathsf{Z}_{:,m}\mathsf{Z}_{:,m}^{H}\right] = \mathsf{SAGAS}^{H} + \sigma^2\mathsf{I}_L = \mathsf{SAGS}^{H} + \sigma^2\mathsf{I}_L$

Statistics of receive signal

- The likelihood function of Y is:

$$f_{\boldsymbol{\alpha}}(\mathsf{Y}) \overset{(a)}{=} \prod_{m\in\mathcal{M}} \frac{\exp\left(-\mathsf{Y}_{:,m}^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m}\right)}{\pi^L|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|} = \frac{\exp\left(-\sum_{m\in\mathcal{M}} \mathsf{Y}_{:,m}^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m}\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|^M}$$

$$\overset{(b)}{=} \frac{\exp\left(-\sum_{m\in\mathcal{M}} \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m}\mathsf{Y}_{:,m}^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|^M} = \frac{\exp\left(-\operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\sum_{m\in\mathcal{M}} \mathsf{Y}_{:,m}\mathsf{Y}_{:,m}^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|^M}$$

$$= \frac{\exp\left(-\operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}\mathsf{Y}^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|^M}$$

  - (a) is due to that $\mathsf{Y}_{:,m}, m \in \mathcal{M}$ are i.i.d. $\mathcal{CN}(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$
  - (b) is due to

$$\mathsf{Y}_{:,m}^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m} = \operatorname{tr}\left(\mathsf{Y}_{:,m}^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m}\right) = \operatorname{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{Y}_{:,m}\mathsf{Y}_{:,m}^H\right)$$

Maximum likelihood (ML) estimation [Fengler et al. (2021)]

- Assume that $\boldsymbol{\alpha}$ are unknown deterministic quantities
- ML estimation of $\boldsymbol{\alpha}$:

$$\min_{\boldsymbol{\alpha}} \quad f_{\mathrm{ML}}(\boldsymbol{\alpha}) \triangleq \log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}| + \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}})$$

$$s.t. \quad \alpha_n \in \{0,1\} \text{ (relax to: } \alpha_n \in [0,1]),\ n \in \mathcal{N}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \triangleq \mathsf{SAGS}^H + \sigma^2\mathsf{I}_L \in \mathbb{C}^{L\times L}$, $\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}} \triangleq \frac{1}{M}\mathsf{YY}^H \in \mathbb{C}^{L\times L}$

  - $f_{\mathrm{ML}}(\boldsymbol{\alpha})$ is $-\frac{1}{M}\log f_{\boldsymbol{\alpha}}(\mathsf{Y})$ (omit the constant), where

$$-\log f_{\boldsymbol{\alpha}}(\mathsf{Y}) = M\log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}| + \mathrm{tr}((\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{YY}^H) + LM\log(\pi)$$

  - $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ represents the covariance matrix of $\mathsf{Y}_{:,m}, m \in \mathcal{M}$
  - $\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}}$ represents the sample covariance matrix of $\mathsf{Y}_{:,m}, m \in \mathcal{M}$
    - The average over M different antennas
    - $\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}} \to \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ as $M \to \infty$
    - Sufficient statistics: $f_{\mathrm{ML}}(\boldsymbol{\alpha})$ depends on Y only through $\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}}$
  - A binary solution can be conducted by performing thresholding
- Advantage in the massive MIMO regime: estimate N variables, $\boldsymbol{\alpha}$, from L^2 observations, $\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}}$, irrespective of M

Coordinate descent (CD) method for ML estimation

- The problem is non-convex, as $f_{\mathrm{ML}}(\boldsymbol{\alpha})$ is a difference of convex (DC) function of $\boldsymbol{\alpha}$
  - $\log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|$ is a concave function of $\boldsymbol{\alpha}$
  - $\mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}})$ is a convex function of $\boldsymbol{\alpha}$
- Standard methods for DC programming such as the convex-concave procedure are not computationally efficient
- The CD method is efficient as the coordinate optimization in each step can be solved analytically
  - Given $\boldsymbol{\alpha}$ obtained in the previous step, the optimization w.r.t. α_n equals to the optimization of the increment d in α_n:

$$\min_{d\in[-\alpha_n,1-\alpha_n]} f_{\mathrm{ML}}(\boldsymbol{\alpha}+d\mathsf{e}_n)=\log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|+\mathrm{tr}((\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}})$$

$$+\log(1+dg_n\mathsf{s}_n^H\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{s}_n^H)-\frac{dg_n\mathsf{s}_n^H\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{s}_n}{1+dg_n\mathsf{s}_n^H\boldsymbol{\Sigma}_{\mathsf{b}}\mathsf{s}_n}$$

  - The optimal solution is given by:

$$d^*_{\mathrm{ML,n}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1},\alpha_n\right)=\min\left\{\max\left\{\frac{\mathsf{s}_n^H\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{s}_n-\mathsf{s}_n^H\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{s}_n}{g_n(\mathsf{s}_n^H\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathsf{s}_n)^2},-\alpha_n\right\},1-\alpha_n\right\}$$

Prior distribution of device activities [SPAWC'20]

- Assume that $\boldsymbol{\alpha}$ is random, and its p.m.f., $p(\boldsymbol{\alpha})$, is known to the BS
- Adopt the Multivariate Bernoulli (MVB) model for $p(\boldsymbol{\alpha})$ [Ding et al. (2011)]:

$$p(\boldsymbol{\alpha}) = \exp\left(\sum_{\omega\in\Psi}\left(c_\omega \prod_{n\in\omega}\alpha_n\right) + b\right)$$

  - Ψ is the set of the nonempty subsets of $\mathcal{N}$
  - $b \triangleq -\log(\sum_{\boldsymbol{\alpha}\in\{0,1\}^N} \exp(\sum_{\omega\in\Psi}(c_\omega \prod_{n\in\omega}\alpha_n)))$ is the normalization factor
  - c_ω is the coefficient reflecting the correlation among α_n, $n \in \omega$
  - $c_\omega, \omega \in \Psi$ can be estimated based on the historical device activity data using existing methods [Ding et al. (2011)]
  - Given $p(\boldsymbol{\alpha})$ in any form, the coefficients $c_\omega, \omega \in \Psi$ can be calculated [Ding et al. (2011), Lem. 2.1]
- Two special cases of the MVB model:
  - Independent case: $c_\omega = 0$ for all $|\omega| > 1$
  - i.i.d. case: $c_\omega = 0$ for all $|\omega| > 1$ and $c_\omega = c$ for all $|\omega| = 1$

Two instances of MVB model

- The devices in $\mathcal{N}$ are divided into K groups, $\mathcal{N}_k \subseteq \mathcal{N}, k \in K$, where $\mathcal{K} \triangleq \{1, \cdots, K\}$
  - $\cup_{k\in\mathcal{K}} \mathcal{N}_k = \mathcal{N}$ and $\mathcal{N}_k \cap \mathcal{N}_{k'} = \varnothing$ for $k, k' \in \mathcal{K}$, $k \neq k'$
- The device activities in different groups are independent:

$$c_\omega = 0, \quad \omega \not\subseteq \mathcal{N}_k, k \in \mathcal{K}$$

- First instance:
  - Each group contains two devices, i.e., $|\mathcal{N}_k| = 2$, $k \in \mathcal{K}$
  - Every device is active with probability p_a
  - Every two devices in a group are correlated with correlation coefficient η
  - c_ω is given by [SPAWC'20, Lem. 1]:

$$c_\omega = \begin{cases} \frac{(\eta p_a + (1-\eta)p_a^2)(1+(\eta-2)p_a+(1-\eta)p_a^2)}{(1-\eta)^2(p_a-p_a^2)^2}, & |\omega| = 2 \\ \frac{(1-\eta)(p_a-p_a^2)}{1+(\eta-2)p_a+(1-\eta)p_a^2}, & |\omega| = 1 \end{cases}, \omega \subseteq \mathcal{N}_k, k \in \mathcal{K}$$

Two instances of MVB model

- Second instance:
  - The activity states of the devices in a group are the same
  - Each group $k \in \mathcal{K}$ is active with probability p_k
  - c_ω is given by [SPAWC'20, Lem. 2]:

$$c_\omega = \begin{cases} (-1)^{|\omega|} \log(\frac{1-p_k}{\epsilon}), & |\omega| < |\mathcal{N}_k| \\ \log(\frac{p_k}{1-p_k}), & |\omega| = |\mathcal{N}_k|, |\omega| \text{ is odd} \\ \log(\frac{p_k(1-p_k)}{\epsilon^2}), & |\omega| = |\mathcal{N}_k|, |\omega| \text{ is even} \end{cases}, \ \omega \subseteq \mathcal{N}_k, k \in \mathcal{K}$$

for arbitrarily small $\epsilon > 0$

Maximum posterior probability (MAP) estimation

- The conditional density of $\boldsymbol{\alpha}$, given Y, is given by:

$$f_{\boldsymbol{\alpha}|\mathrm{Y}}(\boldsymbol{\alpha}, \mathrm{Y}) = f_{\boldsymbol{\alpha}}(\mathrm{Y})p(\boldsymbol{\alpha}) = \frac{\exp\left(-\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathrm{Y}\mathrm{Y}^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}|^M} \exp\left(\sum_{\omega\in\Psi}\left(c_\omega \prod_{n\in\omega}\alpha_n\right) + b\right)$$

- MAP estimation of $\boldsymbol{\alpha}$:

$$\min_{\boldsymbol{\alpha}} \quad f_{\mathrm{MAP}}(\boldsymbol{\alpha}) \triangleq f_{\mathrm{ML}}(\boldsymbol{\alpha}) - \frac{1}{M}\sum_{\omega\in\Psi}\left(c_\omega \prod_{n\in\omega}\alpha_n\right)$$

$$s.t. \quad \alpha_n \in [0,1], \ n \in \mathcal{N}$$

  - $f_{\mathrm{MAP}}(\boldsymbol{\alpha})$ is $-\frac{1}{M}\log f_{\boldsymbol{\alpha}|\mathrm{Y}}(\boldsymbol{\alpha}, \mathrm{Y})$ (omit the constant), where

$$-\log f_{\boldsymbol{\alpha}|\mathrm{Y}}(\boldsymbol{\alpha}, \mathrm{Y}) = M\log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}| + \mathrm{tr}((\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1}\mathrm{Y}\mathrm{Y}^H) + LM\log(\pi)$$

$$-\sum_{\omega\in\Psi}\left(c_\omega \prod_{n\in\omega}\alpha_n\right) - b$$

  - The influence of $p(\boldsymbol{\alpha})$ decreases with M, as $|f_{\mathrm{ML}}(\boldsymbol{\alpha}) - f_{\mathrm{MAP}}(\boldsymbol{\alpha})|$ decreases with M

- The problem is non-convex

Coordinate descent (CD) method for MAP estimation

- The CD method is efficient as the coordinate optimization in each step can be solved analytically
  - Given $\boldsymbol{\alpha}$ obtained in the previous step, the CD optimization w.r.t. α_n equals to the optimization of the increment d in α_n:

$$\min_{d \in [-\alpha_n, 1-\alpha_n]} f_{\mathrm{MAP}}(\boldsymbol{\alpha} + d\mathbf{e}_n) = f_{\mathrm{MAP}}(\boldsymbol{\alpha}) + f_n(d, \boldsymbol{\alpha})$$

$$f_n(d, \boldsymbol{\alpha}) \triangleq \log(1 + d g_n \mathbf{s}_n^H \Sigma_{\boldsymbol{\alpha}}^{-1} \mathbf{s}_n) - \frac{d g_n \mathbf{s}_n^H \Sigma_{\boldsymbol{\alpha}}^{-1} \widehat{\Sigma}_{\mathsf{Y}} \Sigma_{\boldsymbol{\alpha}}^{-1} \mathbf{s}_n}{1 + d g_n \mathbf{s}_n^H \Sigma_{\boldsymbol{\alpha}}^{-1} \mathbf{s}_n} - d C_n$$

$$C_n \triangleq \frac{1}{M} \sum_{\omega \in \Psi : n \in \omega} \left(c_\omega \prod_{n' \in \omega, n' \neq n} \alpha_{n'} \right)$$

Coordinate descent (CD) method for MAP estimation

Theorem (Solution of Optimization w.r.t. α_n)

The optimal solution is given by:

$$d^*_{\mathrm{MAP,n}}\left(\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}, \alpha_n\right) = \begin{cases} \min\left\{\max\left\{s_n(\boldsymbol{\alpha}), -\alpha_n\right\}, 1-\alpha_n\right\}, & C_n \le 0 \\ \arg\min_{d \in \{s_n(\boldsymbol{\alpha}), 1-\alpha_n\}} f_n(d, \boldsymbol{\alpha}), & C_n > 0, \Delta_n > 0 \\ -\alpha_n + 1, & C_n > 0, \Delta_n \le 0 \end{cases}$$

where $s_n(\boldsymbol{\alpha}) \triangleq \frac{1-\sqrt{\Delta_n}}{2C_n} - \frac{1}{g_n \mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n}$ and $\Delta_n \triangleq 1 - \frac{4C_n \mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \widehat{\mathbf{\Sigma}}_{\mathrm{Y}} \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n}{g_n (\mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n)^2}$.

Corollary (Solution of Optimization w.r.t. α_n in i.i.d. case)

If α_n, $n \in \mathcal{N}$ are i.i.d. Bernoulli(p_a), the optimal solution is:

$$d^*_{\mathrm{MAP,n}}(\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}, \alpha_n) = \min\left\{\max\left\{\frac{M}{2\log(\frac{p_a}{1-p_a})} D_n - \frac{1}{g_n \mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n}, -\alpha_n\right\}, 1 - a_n\right\}$$

where $D_n \triangleq 1 - \sqrt{1 - \frac{\frac{4}{M}\log(\frac{p_a}{1-p_a}) \mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \widehat{\mathbf{\Sigma}}_{\mathrm{Y}} \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n}{g_n (\mathrm{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathrm{s}_n)^2}}$.

- $d^*_{\mathrm{MAP,n}}(\cdot)$ reduces to $d^*_{\mathrm{ML,n}}(\cdot)$, as $M \to \infty$ or $p_a \to 0.5$

Algorithm for statistical device activity detection

Algorithm (CD for statistical device activity detection)

1: **Initialization** choose $\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_L$ and $\boldsymbol{\alpha} = 0$.
2: **repeat**
3: **for** $n \in \mathcal{N}$ **do**
4: Calculate $d_n = d^*_{\mathrm{ML,n}}\left(\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}, \alpha_n\right)$ (ML) or $d_n = d^*_{\mathrm{MAP,n}}\left(\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}, \alpha_n\right)$ (MAP).
5: Update $\alpha_n = \alpha_n + d_n$ (CD update).
6: Update $\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} = \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} - \frac{d_n g_n \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathbf{s}_n \mathbf{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}}{1 + d g_n \mathbf{s}_n^H \mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1} \mathbf{s}_n}$ (estimated covariance matrix update).
7: **end for**
8: **until** $\boldsymbol{\alpha}$ satisfies some stopping criterion.

- Update $\mathbf{\Sigma}_{\boldsymbol{\alpha}}^{-1}$ instead of $\mathbf{\Sigma}_{\boldsymbol{\alpha}}$ to avoid the calculation of matrix inversion and improve the computation efficiency

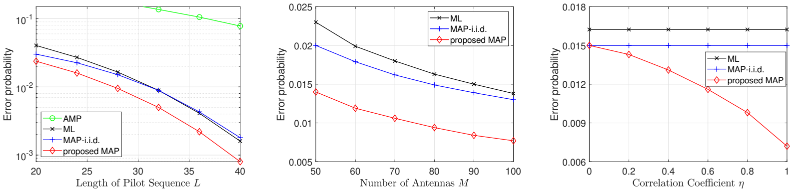
Algorithm for statistical device activity detection

- The algorithm converges to a stationary point of the corresponding statistical estimation problem, as the number of iterations goes to infinity [Bertsekas (1999), Prop. 2.7.1]
  - Different initial points usually correspond to different stationary points
  - Numerical results show that the stationary point corresponding to the initial point $\boldsymbol{\alpha} = 0$ usually provides good detection performance
- The computational complexities of each iteration of ML and MAP are $\mathcal{O}(NL^2)$ and $\mathcal{O}(N2^N + NL^2)$, respectively
  - The actual computational complexity of each iteration of MAP is much lower as $\boldsymbol{\alpha}$ is a sparse vector

Simulation setup

- N devices are uniformly distributed in a disk with radius R
- Each device is active with probability p_a (marginal p.m.f.)
- Generalize the symbols of each pilot according to i.i.d. $\mathcal{CN}(0, 1)$ and then normalize its norm to $\sqrt{L}$
- Independently generate 2000 realizations for the locations of devices and s_n, α_n, h_n, $n \in \mathcal{N}$, and evaluate the average error probability over the 2000 realizations
- Choose $R = 200$, $\gamma = 3$, $L = 28$, $M = 80$, and $\sigma^2 = \frac{R^{-\gamma}}{10}$ ($\mathrm{SNR} = \frac{R^{-\gamma}}{10}$), unless otherwise stated
- Consider three baseline schemes: AMP [Liu & Yu (2018)], ML [Fengler et al. (2021)], and MAP-i.i.d (assuming that $\alpha_n, n \in \mathcal{N}$ are i.i.d. Bernoulli(p_a))
  - The thresholds are numerically optimized

Group device activities in first instance

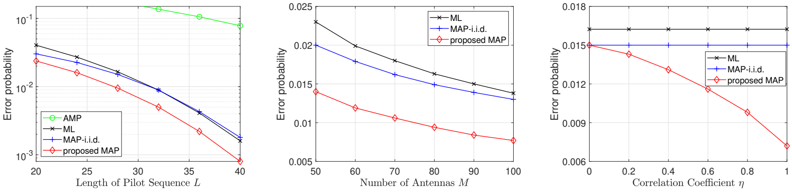


(a) L (b) M (c) η

Figure: Error probability versus pilot length L, number of antennas M, and correlation coefficient η. $N = 1000$, $p_a = 0.05$.

- The statistical estimation schemes significantly outperform AMP
- MAP-i.i.d. outperforms ML, especially at small L and M
  - The gain comes from the incorporation of the marginal p.m.f. of $\alpha_n, n \in \mathcal{N}$ and becomes large at small L and M
- Proposed MAP outperforms MAP-i.i.d., especially at large η (stronger correlation)
  - The gain derives from the explicit consideration of the correlation among $\alpha_n, n \in \mathcal{N}$ and becomes large at large η

Group device activities in first instance



(a) L (b) M (c) η

Figure: Error probability versus pilot length L, number of antennas M, and correlation coefficient η. $N = 1000$, $p_a = 0.05$.

- The error probability of each scheme decreases with L and M
- The error probability of proposed MAP significantly decreases with η, while the error probabilities of MAP-i.i.d. and ML do not change with η
  - Demonstrate the value of utilizing the correlation among $\alpha_n, n \in \mathcal{N}$

Group device activities in second instance

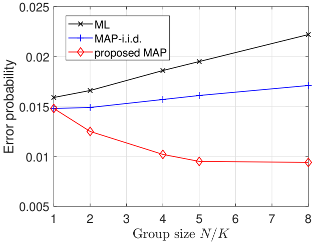


Figure: Error probability versus group size N/K. $N = 1000$, $p_k = 0.05$, $k \in \mathcal{K}$.

- MAP-i.i.d. outperforms ML, especially at large N/K
  - The gain comes from the incorporation of the marginal p.m.f. of $\alpha_n, n \in \mathcal{N}$ and becomes large at large N/K
- Proposed MAP outperforms MAP-i.i.d., especially at large N/K
  - The gain derives from the explicit consideration of the correlation among $\alpha_n, n \in \mathcal{N}$ and becomes large at large N/K

Group device activities in second instance

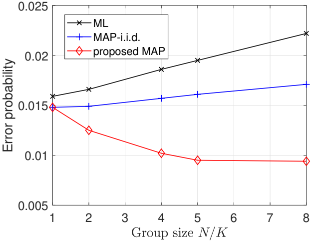


Figure: Error probability versus group size N/K. $N = 1000$, $p_k = 0.05$, $k \in \mathcal{K}$.

- When N/K increases, the variance of the number of active devices increases and the sample space of device activities reduces
- The error probabilities of MAP-i.i.d. and ML increase with N/K
  - The error probability significantly increases when the number of active devices is large if the correlation is not utilized
- The error probability of proposed MAP decreases with N/K
  - The exploitation of the correlation among $\alpha_n, n \in \mathcal{N}$ narrows down the set of possible activity states

Conclusion

- We consider device activity detection in a single-cell network
- We formulate the problem for the MAP estimation of device activities based on the tractable MVB model, explicitly specifying the general correlation among device activities
- We propose an efficient iterative algorithm to obtain a stationary point of the MAP estimation problem using the coordinate descent method
- The proposed MAP estimation enhances the existing ML estimation by exploiting the prior distribution of device activities, at the cost of increasing computational complexity

Outline

Network model [WCNC'20, TWC'21]

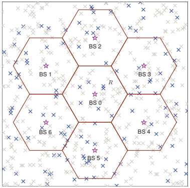


Figure: System model.

- Consider a multi-cell network which consists of M-antenna BSs and single-antenna IoT devices
- The locations of BSs are distributed according to the hexagonal grid model with the side length of each hexagonal cell R
  - Can be extended [TWC'21]
- The BSs and their cells are indexed by $j \in \mathcal{J}$, $\mathcal{J} \triangleq \{0, 1, \cdots\}$
- The devices are indexed by $n \in \mathcal{N}$, $\mathcal{N} \triangleq \{1, 2, \cdots\}$
- $\mathcal{N}_j$ denotes the set of N_j devices in cell j
- $\boldsymbol{\alpha}_j \triangleq (\alpha_n)_{n \in \mathcal{N}_j} \in \{0, 1\}^{N_j}$ denotes the activities of the devices in cell j

Channel model

- Adopt the power-law path loss model for interfering devices
  - $d_{n,j}$ denotes the distance between device n and BS j
  - $\gamma > 2$ denotes the path loss exponent
  - $g_{n,j} = d_{n,j}^{-\gamma}$ denotes the path loss between device n and BS j
  - Commonly used for large-scale random networks
- Adopt the block-fading channel model for small-scale fading
- Consider a narrow-band system and adopt the flat Rayleigh fading model
  - $\mathrm{h}_{n,j} \in \mathbb{C}^M$ denotes the channel vector between device n and BS j
  - $\mathrm{h}_{n,j}$, $n \in \mathcal{N}$, $j \in \mathcal{J}$ are i.i.d. $\mathcal{CN}(0, \mathrm{I}_M)$

Massive grant-free access in a multi-cell network

- Adopt a massive grant-free access scheme
  - Each device n is assigned a length-L pilot $\mathbf{s}_n \triangleq (s_{n,\ell})_{\ell \in \mathcal{L}}$, where $\mathcal{L} \triangleq \{1, 2, \cdots, L\}$ and $L \ll N$
  - $\mathbf{s}_n, n \in \mathcal{N}$ are i.i.d. $\mathcal{CN}(0, \mathbf{I}_L)$
  - $\mathbf{S}_j \triangleq (\mathbf{s}_n)_{n \in \mathcal{N}_j} \in \mathbb{C}^{L \times N_j}$ denotes the pilot matrix for the devices in cell j
- In each coherence block, all active devices synchronously send their pilots, and each BS detects the activities of its associated devices
- The receive signal over L signal dimensions and M antennas at BS j, $\mathbf{Y}_j \in \mathbb{C}^{L \times M}$, is:

$$\mathbf{Y}_j = \sum_{n \in \mathcal{N}} \mathbf{s}_n \alpha_n g_{n,j}^{\frac{1}{2}} \mathbf{h}_{n,j}^T + \mathbf{Z}_j, \quad j \in \mathcal{J}$$

  - $\mathbf{Z}_j \in \mathbb{C}^{L \times M}$ represents the AWGN, with all elements i.i.d. $\mathcal{CN}(0, \sigma^2)$

Two device activity detection mechanisms

- Consider device activity detection at a typical BS at the origin
  - The typical BS is denoted as BS 0
  - The six neighbor BSs of BS 0 are indexed with $1, 2, \cdots, 6$
- Non-cooperative device activity detection:
  - BS 0 knows the pilot matrix S_0 for the devices in cell 0 and the path losses $\mathsf{g}_0 \triangleq (g_{n,0})_{i \in \mathcal{N}_0}$ between devices in cell 0 and BS 0
  - BS 0 detects the activities of the devices in cell 0 from (the sufficient statistics of) Y_0
- Cooperative device activity detection:
  - $\overline{\mathcal{N}}_0 \triangleq \cup_{j=0}^{6} \mathcal{N}_j$ denotes the $\overline{N}_0 \triangleq \sum_{j=0}^{6} N_j$ devices in cell 0 and its six neighbor cells $1, \cdots, 6$
  - BS 0 knows the pilot matrices $\overline{\mathsf{S}}_0 \triangleq (\mathsf{S}_j)_{j \in \{0,1,\cdots,6\}}$ for the devices in the 7 cells and the path losses $\overline{\mathsf{g}}_j \triangleq (g_{n,j})_{n \in \overline{\mathcal{N}}_0}$, $j \in \{0, 1, \cdots, 6\}$ between the devices in the 7 cells and 7 BSs
  - Each BS $j \in \{1, 2, \cdots, 6\}$ transmits (the sufficient statistics of) Y_j to BS 0 via an error-free backhaul link
  - BS 0 detects the activities of the devices in the 7 cells from (the sufficient statistics of) $\overline{\mathsf{Y}}_0 \triangleq [\mathsf{Y}_0, \mathsf{Y}_1, \cdots, \mathsf{Y}_6]$ and utilizes the detection results for the devices in cell 0

Non-cooperative device activity detection

- The receive signal $\mathsf{Y}_0 \in \mathbb{C}^{L\times M}$ at BS 0 can be rewritten as:

$$\mathsf{Y}_0 = \mathsf{S}_0\mathsf{A}_0\mathsf{G}_0^{\frac{1}{2}}\mathsf{H}_0^T + \sum_{n\in\mathcal{N}\setminus\mathcal{N}_0} \mathsf{s}_n\alpha_n g_{n,0}^{\frac{1}{2}}\mathsf{h}_{n,0}^T + \mathsf{Z}_0$$

  - $\mathsf{A}_0 \triangleq \mathrm{diag}(\boldsymbol{\alpha}_0)$, $\mathsf{G}_0 \triangleq \mathrm{diag}(\mathsf{g}_0)$, $\mathsf{H}_0 \triangleq (\mathsf{h}_{n,0})_{n\in\mathcal{N}_0}$
- Given $\alpha_n, \mathsf{g}_{n,0}, \mathsf{s}_n, n \in \mathcal{N}$, all M columns of Y_0 are i.i.d. $\mathcal{CN}(0, \mathsf{S}_0\mathsf{A}_0\mathsf{G}_0\mathsf{S}_0^H + \widetilde{\mathsf{X}} + \delta^2\mathsf{I}_L)$ with $\widetilde{\mathsf{X}} \triangleq \sum_{n\in\mathcal{N}\setminus\mathcal{N}_0} \alpha_n\ g_{n,0}\mathsf{s}_n\mathsf{s}_n^H$
  - $\widetilde{\mathsf{X}} \in \mathbb{C}^{L\times L}$ represents the covariance matrix of inter-cell interference and has also to be estimated
  - $\widetilde{\mathsf{X}}$ is diagonally dominant, as $\mathsf{s}_n, n \in \mathcal{N}$ are i.i.d. $\mathcal{CN}(0, \mathsf{I}_L)$ [Chen & Yu (2019)]
- Approximate $\widetilde{\mathsf{X}}$ with $\mathsf{X} \triangleq \mathrm{diag}(\mathsf{x})$, $\mathsf{x} \triangleq (x_\ell)_{\ell\in\mathcal{L}} \in \mathbb{R}_+^L$, $x_\ell \triangleq \sum_{n\in\mathcal{N}\setminus\mathcal{N}_0} \alpha_n\ g_{n,0}|\mathsf{s}_{n,\ell}|^2$ to reduce the estimation complexity [Chen et al. (2019); Andrews et al. (2007); Choi (2019)]
  - Diagonal elements $\mathsf{x} \in \mathbb{R}_+^L$ of $\mathsf{X} \in \mathbb{R}_+^{L\times L}$ can be interpreted as the inter-cell interference powers over the L signal dimensions

Non-cooperative device activity detection

- Given α_0 and x, all M columns of Y_0 are approximated as i.i.d. $\mathcal{CN}(0, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}} \triangleq \mathrm{S}_0\mathrm{A}_0\mathrm{G}_0\mathrm{S}_0^H + \mathrm{X} + \delta^2\mathrm{I}_L \in \mathbb{C}^{L\times L}$
- The likelihood function of Y_0 is:

$$f_{\boldsymbol{\alpha}_0,\mathrm{x}}(\mathrm{Y}_0) = \frac{\exp\left(-\mathrm{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{Y}_0\mathrm{Y}_0^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}|^M}$$

- Consider the joint ML estimation and the joint MAP estimation of N_0 device activities $\boldsymbol{\alpha}_0$ and L interference powers x without BS cooperation, respectively

Joint ML estimation without BS cooperation

- Assume that $\boldsymbol{\alpha}_0$ and $\mathbf{x}$ are unknown deterministic quantities
- Joint ML estimation of $\boldsymbol{\alpha}_0$ and $\mathbf{x}$ without BS cooperation:

$$\begin{aligned}\min_{\boldsymbol{\alpha}_0,\mathbf{x}} \quad & f_{\mathrm{ML}}(\boldsymbol{\alpha}_0,\mathbf{x}) \triangleq \log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathbf{x}}| + \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathbf{x}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_0}) \\ s.t. \quad & \alpha_n \in [0,1], \quad n \in \mathcal{N}_0 \\ & x_\ell \geq 0, \quad \ell \in \mathcal{L}\end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathbf{x}} \triangleq \mathbf{S}_0\mathbf{A}_0\mathbf{G}_0\mathbf{S}_0^H + \mathbf{X} + \delta^2\mathbf{I}_L \in \mathbb{C}^{L\times L}$ and
$\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_0} \triangleq \frac{1}{M}\mathbf{Y}_0\mathbf{Y}_0^H \in \mathbb{C}^{L\times L}$

  - Jointly estimate $N_0 + L$ variables, $\boldsymbol{\alpha}_0$ and $\mathbf{x}$, from L^2 observations, $\widehat{\boldsymbol{\Sigma}}_{\mathbf{Y}_0}$, in the multi-cell network with inter-cell interference
  - $f_{\mathrm{ML}}(\boldsymbol{\alpha}_0,\mathbf{x})$ is $-\frac{1}{M}\log f_{\boldsymbol{\alpha}_0,\mathbf{x}}(\mathbf{Y}_0)$ (omit the constant), where

$$-\log f_{\boldsymbol{\alpha}_0,\mathbf{x}}(\mathbf{Y}_0) = M\log|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathbf{x}}| + \mathrm{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathbf{x}}^{-1}\mathbf{Y}_0\mathbf{Y}_0^H) + LM\log(\pi)$$

- The problem is non-convex, as $f_{\mathrm{ML}}(\boldsymbol{\alpha}_0,\mathbf{x})$ is a DC function

Coordinate descent (CD) method for joint ML estimation

- At each step of one iteration, optimize $f_{\mathrm{ML}}(\boldsymbol{\alpha}_0, \mathsf{x})$ w.r.t. one coordinate in $\{\alpha_n \colon n \in \mathcal{N}_0\} \cup \{x_\ell : \ell \in \mathcal{L}\}$
- Given $\boldsymbol{\alpha}_0$ and x obtained in the previous step, the CD optimization w.r.t. α_n equals to the optimization of the increment d in α_n:

$$\min_{d \in [-\alpha_n, 1-\alpha_n]} f_{\mathrm{ML}}(\boldsymbol{\alpha}_0 + d\mathsf{e}_n, \mathsf{x}) = f_{\mathrm{ML}}(\boldsymbol{\alpha}_0, \mathsf{x}) + \log |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0, \mathsf{x}} + d g_{n,0} \mathsf{s}_n \mathsf{s}_n^H| + \mathrm{tr}((\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0, \mathsf{x}} + d g_{n,0} \mathsf{s}_n \mathsf{s}_n^H)^{-1} \widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}_0})$$

  and the CD optimization w.r.t. x_ℓ equals to the optimization of the increment d in x_ℓ:

$$\min_{d \in [-x_\ell, +\infty)} f_{\mathrm{ML}}(\boldsymbol{\alpha}_0, \mathsf{x} + d\mathsf{e}_\ell) = f_{\mathrm{ML}}(\boldsymbol{\alpha}_0, \mathsf{x}) + \log |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0, \mathsf{x}} + d \mathsf{e}_\ell \mathsf{e}_\ell^T| + \mathrm{tr}((\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0, \mathsf{x}} + d \mathsf{e}_\ell \mathsf{e}_\ell^T)^{-1} \widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}_0})$$

Coordinate descent (CD) method for joint ML estimation

Theorem (Solutions of Optimizations w.r.t. α_n and x_ℓ)

Given $\boldsymbol{\alpha}_0$ and x obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. α_n is

$$d^*_{\mathrm{ML,1,n}}(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}}, \alpha_n) \triangleq \min\left\{\max\left\{\frac{\mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_0} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n - \mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n}{g_{n,0}(\mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n)^2}, -\alpha_n\right\}, 1-\alpha_n\right\}$$

and the optimal solution of the coordinate optimization w.r.t. x_ℓ is

$$d^*_{\mathrm{ML},2,\ell}(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}}, x_\ell) \triangleq \max\left\{\frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_0} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell - \mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell}{(\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell)^2}, -x_\ell\right\}$$

Prior distribution of device activities

- Assume that $\boldsymbol{\alpha}_0$ is random, and its p.m.f., $p_0(\boldsymbol{\alpha}_0)$, is known to BS 0
- Adopt the MVB model for $p_0(\boldsymbol{\alpha}_0)$ [Ding et al. (2011)]:

$$p_0\left(\boldsymbol{\alpha}_0\right) = \exp\left(\sum_{\omega\in\Psi_0}\left(c_\omega \prod_{n\in\omega}\alpha_n\right) + b_0\right),$$

  - Ψ_0 is the set of the nonempty subsets of $\mathcal{N}_0$
  - $b_0 \triangleq \log(\sum_{\boldsymbol{\alpha}_0\in\{0,1\}^{N_0}} \exp(\sum_{\omega\in\Psi_0}(c_\omega \prod_{n\in\omega}\alpha_n)))$ is the normalization factor
  - c_ω is the coefficient reflecting the correlation among α_n, $n\in\omega$
  - $c_\omega, \omega\in\Psi_0$ can be estimated based on the historical device activity data using existing methods [Ding et al. (2011)]
  - Given $p_0\left(\boldsymbol{\alpha}_0\right)$ in any form, the coefficients $c_\omega, \omega\in\Psi_0$ can be calculated [Ding et al. (2011), Lem. 2.1]

Prior distribution of interference powers

- Assume that x is random, and its p.d.f., $g(\mathsf{x})$, is known to BS 0
- Assume that the locations of the active interfering devices in $\mathcal{N} \setminus \mathcal{N}_0$ follow a homogeneous Poisson point process (PPP) with density λ
- Approximate the p.d.f. of x with a Gaussian distribution with the same mean and variance

$$g(\mathsf{x}) \approx \frac{1}{(\sqrt{2\pi}\delta)^L} \exp\left(-\frac{\sum_{\ell \in \mathcal{L}} (x_\ell - \mu)^2}{2\delta^2}\right)$$

  - $x_\ell = \sum_{n \in \mathcal{N} \setminus \mathcal{N}_0} \alpha_n \, g_{n,0} |\mathsf{s}_{n,\ell}|^2$
  - $\mathsf{s}_{n,\ell}, n \in \mathcal{N}$ are i.i.d. $\mathcal{CN}(0,1)$

Prior distribution of interference powers

Lemma (Mean and variance for hexagon model)

If Cell 0 is modeled as a hexagon with side length R, then we have:

$$\mu = 12\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_{0}^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\frac{\alpha}{2}} \mathrm{d}y\mathrm{d}x$$

$$\delta^2 = 12\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_{0}^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\alpha} \mathrm{d}y\mathrm{d}x$$

Lemma (Mean and variance for disk model)

If Cell 0 is modeled as a disk with radius R, then we have:

$$\mu = \frac{2\pi\lambda R^{2-\alpha}}{\alpha - 2}, \quad \delta^2 = \frac{\pi\lambda R^{2-2\alpha}}{\alpha - 1}$$

Prior distribution of interference powers

- The Gaussian distribution with the same mean and variance is a good approximation of the exact p.d.f. of x

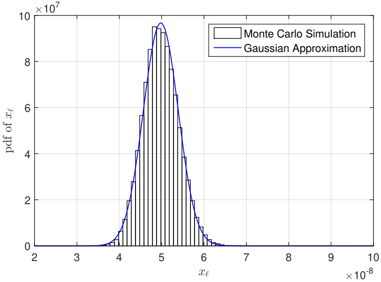


Figure: Comparison between the histogram of x_ℓ (reflecting the p.d.f. of x_ℓ) and its corresponding Gaussian approximation. $R = 200$, $\lambda = 0.0005$, and $\gamma = 4$.

Joint MAP estimation without BS cooperation

- The conditional density of $\boldsymbol{\alpha}_0$ and x, given Y_0, is given by:

$$f_{\boldsymbol{\alpha}_0,\mathrm{x}|\mathrm{Y}_0}(\boldsymbol{\alpha}_0,\mathrm{x},\mathrm{Y}_0) = f_{\boldsymbol{\alpha}_0,\mathrm{x}}(\mathrm{Y}_0)p_0(\boldsymbol{\alpha}_0)g(\mathrm{x})$$
$$= \frac{\exp\left(-\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}}\mathrm{Y}_0\mathrm{Y}_0^H\right)\right)}{\pi^{LM}|\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}|^M(\sqrt{2\pi}\delta)^L}\exp\left(\sum_{\omega\in\Psi_0}\left(c_\omega\prod_{n\in\omega}\alpha_n\right)+b_0-\sum_{\ell\in\mathcal{L}}\frac{(x_\ell-\mu)^2}{2\delta^2}\right)$$

- Joint MAP estimation of $\boldsymbol{\alpha}_0$ and x without BS cooperation:

$$\min_{\boldsymbol{\alpha}_0,\mathrm{x}} \quad f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0,\mathrm{x}) \triangleq f_{\mathrm{ML}}(\boldsymbol{\alpha}_0,\mathrm{x})-\frac{1}{M}\sum_{\omega\in\Psi_0}\left(c_\omega\prod_{n\in\omega}\alpha_n\right)+\frac{1}{M}\sum_{\ell\in\mathcal{L}}\frac{(x_\ell-\mu)^2}{2\delta^2}$$
$$s.t. \quad \alpha_n\in[0,1], \quad n\in\mathcal{N}_0$$
$$x_\ell\geq 0, \quad \ell\in\mathcal{L}$$

  - The impacts of prior distributions of $\boldsymbol{\alpha}_0$ and x decrease with M, as $|f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0,\mathrm{x})-f_{\mathrm{ML}}(\boldsymbol{\alpha}_0,\mathrm{x})|$ decreases with M
- The problem is non-convex

Coordinate descent (CD) method for joint MAP estimation

- At each step of one iteration, optimize $f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0, \mathrm{x})$ w.r.t. one coordinate in $\{\alpha_n\colon n \in \mathcal{N}_0\} \cup \{x_\ell : \ell \in \mathcal{L}\}$
- Given $\boldsymbol{\alpha}_0$ and x obtained in the previous step, the CD optimization w.r.t. α_n equals to the optimization of the increment d in α_n:

$$\min_{d\in[-\alpha_n,1-\alpha_n]} f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0 + d\mathrm{e}_i, \mathrm{x}) = f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0, \mathrm{x}) + f_{\boldsymbol{\alpha}_0,n}(d, \boldsymbol{\alpha}_0, \mathrm{x})$$

$$f_{\boldsymbol{\alpha}_0,n}(d, \boldsymbol{\alpha}_0, \mathrm{x}) \triangleq \log(1 + dg_{n,0}\mathrm{s}_n^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{s}_n) - \frac{dg_{n,0}\mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_0}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{s}_n}{1 + dg_{n,0}\mathrm{s}_n^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{s}_n}$$
$$- \frac{d}{M}\sum_{\omega\subseteq\mathcal{N}_0:n\in\omega}\left(c_\omega \prod_{n'\in\omega,n'\neq n}\alpha_{n'}\right)$$

and the CD optimization w.r.t. x_ℓ equals to the optimization of the increment d in x_ℓ:

$$\min_{d\in[-x_\ell,+\infty)} f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0, \mathrm{x} + d\mathrm{e}_\ell) = f_{\mathrm{MAP}}(\boldsymbol{\alpha}_0, \mathrm{x}) + f_{\mathrm{x},\ell}(d, \boldsymbol{\alpha}_0, \mathrm{x})$$

$$f_{\mathrm{x},\ell}(d, \boldsymbol{\alpha}_0, \mathrm{x}) \triangleq \frac{(x_\ell - \mu + d)^2}{2M\sigma^2} - \frac{d\mathrm{e}_\ell^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_0}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{e}_\ell}{1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{e}_\ell} + \log(1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathrm{x}}^{-1}\mathrm{e}_\ell)$$

Coordinate descent (CD) method for joint MAP estimation

Theorem (Solution of Optimization w.r.t. α_n)

Given $\boldsymbol{\alpha}_0$ and x obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. $\boldsymbol{\alpha}_0$ is:

$$d^*_{\mathrm{MAP,1,n}}(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}, \alpha_n)$$

$$= \begin{cases} \min\left\{\max\left\{s_n(\boldsymbol{\alpha}_0,\mathsf{x}), -\boldsymbol{\alpha}_n\right\}, 1-\boldsymbol{\alpha}_n\right\}, & C_n \leq 0 \\ \arg\min_{d\in\{s_n(\boldsymbol{\alpha}_0,\mathsf{x}),1-\alpha_n\}} f_{\boldsymbol{\alpha}_0,n}(d,\boldsymbol{\alpha}_0,\mathsf{x}), & 0 < C_n < \frac{g_{n,0}(\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n)^2}{4\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n} \\ 1-\alpha_n, & C_n \geq \frac{g_{n,0}(\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n)^2}{4\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n} \end{cases}$$

where

$$s_n(\boldsymbol{\alpha}_0,\mathsf{x}) \triangleq \frac{1}{2C_n}\left(1-\sqrt{1-\frac{4C_n\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\widehat{\boldsymbol{\Sigma}}_{\mathsf{Y}_0}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n}{g_{n,0}(\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{s}_n)^2}}\right) - \frac{1}{g_{n,0}\mathsf{s}_n^H\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathsf{x}}\mathsf{g}_{n,0}}$$

$$C_n \triangleq \frac{1}{M}\sum_{\omega\in\Psi_0:n\in\omega} c_\omega \prod_{n'\in\omega,n'\neq n} \alpha_{n'}$$

Coordinate descent (CD) method for joint MAP estimation

Theorem (Solution of Optimization w.r.t. x_ℓ)

Given $\boldsymbol{\alpha}_0$ and x obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. x is:

$$d^*_{\mathrm{MAP},2,\ell}(\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}}, x_\ell) = \underset{d \in \mathcal{X}_\ell(\boldsymbol{\alpha}_0,\mathrm{x}) \cup \{-x_\ell\}}{\arg\min} f_{\mathrm{x},\ell}(d, \boldsymbol{\alpha}_0, \mathrm{x})$$

where

$$\mathcal{X}_\ell(\boldsymbol{\alpha}_0,\mathrm{x}) \triangleq \{d \in [-x_\ell, +\infty) : h_{\mathrm{x},\ell}(d, \boldsymbol{\alpha}_0, \mathrm{x}) = 0\}$$

$$h_{\mathrm{x},\ell}(d, \boldsymbol{\alpha}_0, \mathrm{x}) \triangleq \frac{d + x_\ell - \mu}{M\delta^2} - \frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_0} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell}{(1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell)^2} + \frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell}{1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell}$$

Algorithm for statistical device activity detection

Algorithm (Statistical device activity detection without BS cooperation)

Initialization: choose $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0}^{-1} = \frac{1}{\sigma^2}\mathsf{I}_L$, $\boldsymbol{\alpha}_0 = 0$, $\mathsf{x} = 0$.
repeat
for $n \in \mathcal{N}_0$ **do**
  Calculate $d_n = d^*_{\mathrm{ML},1,\mathrm{n}}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}, \alpha_n)$ (ML) or $d_n = d^*_{\mathrm{MAP},1,\mathrm{n}}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}, \alpha_n)$ (MAP).
  Update $\alpha_n = \alpha_n + d_n$ (CD update).
  Update $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} - \frac{d_n g_n \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} \mathsf{s}_n \mathsf{s}_n^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}}{1 + d_n g_n \mathsf{s}_n^H \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} \mathsf{s}_n}$ (estimated covariance matrix update).
end for
for $\ell \in \mathcal{L}$ **do**
  Calculate $d_\ell = d^*_{\mathrm{ML},2,\ell}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}, x_\ell)$ (ML) or $d_\ell = d^*_{\mathrm{MAP},2,\ell}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}, x_\ell)$ (MAP).
  Update $x_\ell = x_\ell + d_\ell$ (CD update).
  Update $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} - \frac{d_\ell \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} \mathsf{e}_\ell \mathsf{e}_\ell^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1}}{1 + d_\ell \mathsf{e}_\ell^T \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_0,\mathsf{x}}^{-1} \mathsf{e}_\ell}$ (estimated covariance matrix update)
end for
until $\boldsymbol{\alpha}_0$ and x satisfy some stopping criterion.

Algorithm for statistical device activity detection without BS cooperation

- The algorithm converges to a stationary point of the corresponding statistical estimation problem, as the number of iterations goes to infinity [Bertsekas (1999),Prop. 2.7.1]
  - Different initial points usually correspond to different stationary points
  - Numerical results show that the stationary point corresponding to the initial point $\boldsymbol{\alpha}_0 = 0$, $\mathrm{x} = 0$ usually provides good detection performance
- The computational complexities of each iteration of the joint ML estimation and the joint MAP estimation without BS cooperation are $\mathcal{O}(N_0 L^2 + L^3)$ and $\mathcal{O}(N_0 2^{N_0} + N_0 L^2 + L^3)$, respectively
  - The actual computational complexity for the joint MAP estimation is much lower as $\boldsymbol{\alpha}_0$ is a sparse vector

Cooperative device activity detection

- The receive signal at BS j, $\mathsf{Y}_j \in \mathbb{C}^{L\times M}$, can be rewritten as:

$$\mathsf{Y}_j = \overline{\mathsf{S}}_0 \overline{\mathsf{A}}_0 \overline{\mathsf{G}}_j^{\frac{1}{2}} \overline{\mathsf{H}}_j^T + \sum_{n\in\mathcal{N}\backslash\overline{\mathcal{N}}_0} \mathsf{s}_n \alpha_n g_{n,j}^{\frac{1}{2}} \mathsf{h}_{n,j}^T + \mathsf{Z}_j,\ j\in\{0,1,\cdots,6\}$$

  - $\overline{\mathsf{A}}_0 \triangleq \mathrm{diag}(\overline{\boldsymbol{\alpha}}_0)$, $\overline{\boldsymbol{\alpha}}_0 \triangleq (\alpha_n)_{n\in\overline{\mathcal{N}}_0}$, $\overline{\mathsf{G}}_j \triangleq \mathrm{diag}(\overline{\mathsf{g}}_j)$, $\overline{\mathsf{H}}_j \triangleq (\mathsf{h}_{n,j})_{n\in\overline{\mathcal{N}}_0}$
- Given $\alpha_n, \mathsf{g}_{n,j}, \mathsf{s}_n, n\in\mathcal{N}$, Y_j, $j\in\{0,1,\cdots,6\}$ are independent and for all $j\in\{0,1,\cdots,6\}$, the M columns of Y_j are i.i.d. $\mathcal{CN}(0, \overline{\mathsf{S}}_0\overline{\mathsf{A}}_0\overline{\mathsf{G}}_j\overline{\mathsf{S}}_0^H + \sum_{n\in\mathcal{N}\backslash\overline{\mathcal{N}}_0} \alpha_n g_{n,j} \mathsf{s}_n \mathsf{s}_n^H + \sigma^2 \mathsf{I}_L)$
- For all $j\in\{0,1,\cdots,6\}$, approximate $\sum_{n\in\mathcal{N}\backslash\overline{\mathcal{N}}_0} \alpha_n g_{n,j}^{\frac{1}{2}} \mathsf{s}_n \mathsf{s}_n^H$ with $\mathsf{X}_j \triangleq \mathrm{diag}(\mathsf{x}_j)$, $\mathsf{x}_j \triangleq (x_{j,\ell})_{\ell\in\mathcal{L}}$, $x_{j,\ell} \triangleq \sum_{n\in\mathcal{N}\backslash\overline{\mathcal{N}}_0} \alpha_n g_{n,j} |\mathsf{s}_{n,\ell}|^2$
  - x_j can be interpreted as the inter-cell interference powers over the L signal dimensions in Y_j

Cooperative device activity detection

- Given $\overline{\boldsymbol{\alpha}}_0$ and x_j, all M columns of Y_j are approximated as i.i.d. $\mathcal{CN}(0, \overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j})$ with $\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j} \triangleq \overline{\mathsf{S}}_0\overline{\mathsf{A}}_0\overline{\mathsf{G}}_j\overline{\mathsf{S}}_0^H + \mathsf{X}_j + \sigma^2\mathsf{I}_L$
- The likelihood function of Y_j is:

$$\bar{f}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}(\mathsf{Y}_j) = \frac{\exp\left(-\mathrm{tr}\left(\overline{\boldsymbol{\Sigma}}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}\mathsf{Y}_j\mathsf{Y}_j^H\right)\right)}{\pi^{LM}|\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}|^M}, \quad j \in \{0, 1, \cdots, 6\}$$

- The likelihood function of $\overline{\mathsf{Y}}_0$ is:

$$\bar{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathsf{x}}_0}(\overline{\mathsf{Y}}_0) \overset{(a)}{=} \prod_{j=0}^{6} \bar{f}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}(\mathsf{Y}_j) = \frac{\exp\left(-\sum_{j=0}^{6}\mathrm{tr}\left(\overline{\boldsymbol{\Sigma}}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}\mathsf{Y}_j\mathsf{Y}_j^H\right)\right)}{\pi^{7LM}\prod_{j=0}^{6}|\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathsf{x}_j}|^M}$$

  where $\overline{\mathsf{x}}_0 \triangleq [\mathsf{x}_0^T, \cdots, \mathsf{x}_6^T]^T$
  - (a) is due to that Y_j, $j \in \{0, 1, \cdots, 6\}$ are independent
- Consider the joint ML estimation and the joint MAP estimation of $\overline{N}_0$ device activities $\overline{\mathsf{a}}_0$ and $7L$ interference powers $\overline{\mathsf{x}}_0$ with BS cooperation, respectively

Joint ML estimation with BS cooperation

- Joint ML estimation of $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ with BS cooperation:

$$\min_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0} \quad \overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0) \triangleq \sum_{j=0}^{6} \Bigg(\underbrace{\log|\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}| + \mathrm{tr}\Big(\overline{\boldsymbol{\Sigma}}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}\widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j}\Big)}_{\overline{f}_{ML,j}(\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j)} \Bigg)$$

$$s.t. \quad \alpha_n \in [0,1], \quad n \in \overline{\mathcal{N}}_0$$

$$x_{j,\ell} \geq 0, \quad j \in \{0,1,\cdots,6\}, \ \ell \in \mathcal{L}$$

where $\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j} \triangleq \overline{\mathrm{S}}_0\overline{\mathrm{A}}_0\overline{\mathrm{G}}_j\overline{\mathrm{S}}_0^H + \mathrm{X}_j + \sigma^2\mathrm{I}_L \in \mathbb{C}^{L\times L}$ and
$\widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j} \triangleq \frac{1}{M}\mathrm{Y}_j\mathrm{Y}_j^H \in \mathbb{C}^{L\times L}$

  - Jointly estimate $\overline{N}_0 + 7L$ variables, $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$, from $7L^2$ observations, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j}, j \in \{0,1,\cdots,6\}$, in the multi-cell network with inter-cell interference
  - $\overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)$ is $-\frac{1}{M}\log \overline{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0}(\overline{\mathrm{Y}}_0)$ (omit the constant), where

$$-\log \overline{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0}(\overline{\mathrm{Y}}_0) = \sum_{j=0}^{6}\left(M\log|\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}| + \mathrm{tr}\left(\overline{\boldsymbol{\Sigma}}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}\mathrm{Y}_j\mathrm{Y}_j^H\right)\right) + 7LM\log(\pi)$$

  - The problem is non-convex, as $\overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)$ is a DC function

Coordinate descent (CD) method for joint ML estimation

- At each step of one iteration, optimize $\overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$ w.r.t. one coordinate in $\{\alpha_n : n \in \overline{\mathcal{N}}_0\} \cup \{x_{j,\ell} : j \in \{0, \cdots, 6\}, \ell \in \mathcal{L}\}$
- Given $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ obtained in the previous step, the coordinate optimization w.r.t. α_n equals to the optimization of the increment of d in α_n:

$$\min_{d \in [-\alpha_n, 1-\alpha_n]} \overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0 + d\mathrm{e}_i, \overline{\mathrm{x}}_0) = \overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) + \overline{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$$

$$\overline{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \sum_{j=0}^{6} \left(\log(1 + d g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha}_0,\mathrm{x}}^{-1} \mathrm{s}_n) - \frac{d g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha}_0,\mathrm{x}}^{-1} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j} \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha}_0,\mathrm{x}}^{-1} \mathrm{s}_n}{1 + d g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha}_0,\mathrm{x}}^{-1} \mathrm{s}_n} \right)$$

and coordinate optimization w.r.t. $x_{j,\ell}$ equals to the optimization of the increment of d in $x_{j,\ell}$:

$$\min_{d \in [-x_{j,\ell}, \infty)} \overline{f}_{\mathrm{ML,j}}(\overline{\boldsymbol{\alpha}}_0, \mathrm{x}_j + d\mathrm{e}_\ell) = \overline{f}_{\mathrm{ML,j}}(\overline{\boldsymbol{\alpha}}_0, \mathrm{x}_j) + \overline{f}_{x,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$$

$$\overline{f}_{x,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \frac{\mathrm{e}_\ell^T \Sigma_j^{-1} \mathrm{e}_\ell}{1 + d \mathrm{e}_\ell^T \Sigma_j^{-1} \mathrm{e}_\ell} - \frac{\mathrm{e}_\ell^T \Sigma_j^{-1} \widehat{\Sigma}_{\mathrm{Y}_j} \Sigma_j^{-1} \mathrm{e}_\ell}{(1 + d \mathrm{e}_\ell^T \Sigma_j^{-1} \mathrm{e}_\ell)^2}$$

Coordinate descent (CD) method for joint ML estimation

Theorem (Solutions of Optimizations w.r.t. α_n and $x_{j,\ell}$)

Given $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. α_n is

$$\overline{d}^*_{\mathrm{ML},1,\mathrm{n}}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}}) \triangleq \underset{d \in \overline{\mathcal{A}}_n(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0) \cup \{-\alpha_n, 1-\alpha_n\}}{\arg\min} \overline{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$$

and the optimal solution of the coordinate optimization w.r.t. $x_{j,\ell}$ is

$$\overline{d}^*_{\mathrm{ML},2,\ell}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}}) \triangleq \max\left\{ \frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j} \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell - \mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell}{(\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{e}_\ell)^2}, -x_{j,\ell} \right\}$$

where

$$\overline{\mathcal{A}}_n(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \{d \in [-\alpha_n, 1-\alpha_n] : \overline{h}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) = 0\}$$

$$\overline{h}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \sum_{j=0}^{6} \left(\frac{g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n}{1 + d g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n} - \frac{g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j} \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n}{(1 + d g_{n,j} \mathrm{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha}_0,\mathrm{x}} \mathrm{s}_n)^2} \right)$$

Prior distribution of device activities

- Assume that $\boldsymbol{\alpha}_j$, $j \in \{0, 1, \cdots, 6\}$ are random, and their p.m.f.s $p_j(\boldsymbol{\alpha}_j), j \in \{0, 1, \cdots, 6\}$ are known to BS 0
- Adopt the MVB model for $p_j(\boldsymbol{\alpha}_j), j \in \{0, 1, ..., 6\}$ [Ding et al. (2011)]:

$$p_j(\boldsymbol{\alpha}_j) = \exp\left(\sum_{\omega \in \Psi_j}\left(c_\omega \prod_{n \in \omega} \alpha_n\right) + b_j\right)$$

  - Ψ_j is the set of the nonempty subsets of $\mathcal{N}_j$
  - $b_j \triangleq \log(\sum_{\boldsymbol{\alpha}_j \in \{0,1\}^{N_j}} \exp(\sum_{\omega \in \Psi_j}(c_\omega \prod_{n \in \omega} \alpha_n)))$ is the normalization factor
  - c_ω is the coefficient reflecting the correlation among α_n, $n \in \omega$
  - $c_\omega, \omega \in \Psi_j$ can be estimated based on the historical device activity data using existing methods [Ding et al. (2011)]
  - Given $p_j(\boldsymbol{\alpha}_j)$ in any form, the coefficients $c_\omega, \omega \in \Psi_j$ can be calculated [Ding et al. (2011), Lem. 2.1]

Prior distribution of interference powers

- Assume that x_j, $j \in \{0, 1, \cdots, 6\}$ are random, and their p.d.f.s, $g(x_j)$, are known to BS 0
- Assume that the locations of the active interfering devices in $\mathcal{N} \setminus \overline{\mathcal{N}}_0$ follow a homogeneous PPP with density λ
- Approximate the p.d.f. of x_j with a Gaussian distribution with the same mean and variance

$$g_j(\mathrm{x}_j) = \frac{1}{(\sqrt{2\pi}\delta_j)^L} \exp\left(- \frac{\sum_{\ell \in \mathcal{L}} (x_{j,\ell} - \mu_j)^2}{2\delta_j^2} \right), j \in \{0, 1, ..., 6\}$$

  - $x_{j,\ell} = \sum_{n \in \mathcal{N} \setminus \overline{\mathcal{N}}_0} \alpha_n g_{n,j} |s_{n,\ell}|^2$
  - $s_{n,\ell}, n \in \mathcal{N}$ are i.i.d. $\mathcal{CN}(0, 1)$
- Define:

$$U_0(x) = \begin{cases} \frac{\sqrt{3}}{3}x, & \frac{\sqrt{3}}{2}R \le x < \sqrt{3}R \\ -\frac{\sqrt{3}}{3}x + 2R, & \sqrt{3}R \le x \le \frac{3\sqrt{3}}{2}R \\ 0, & \frac{3\sqrt{3}}{2}R \le x \end{cases} \quad U_1(x) = \begin{cases} \frac{\sqrt{3}}{3}x + R, & \sqrt{3}R \le x < \frac{3\sqrt{3}}{2}R \\ -\frac{\sqrt{3}}{3}x + 4R, & \frac{3\sqrt{3}}{2}R \le x \le 2\sqrt{3}R \\ -\frac{\sqrt{3}}{3}x + 3R, & 2\sqrt{3}R \le x < \frac{5\sqrt{3}}{2}R \end{cases}$$

Prior distribution of interference powers

Lemma (Mean and variance for hexagon model)

If cell $j \in \{0, 1, \cdots, 6\}$ is modeled as a hexagon with side length R, then we have:

$$\mu_0 = 12\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_0^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\frac{\alpha}{2}} \mathrm{d}y\mathrm{d}x - 12\lambda \int_{\frac{\sqrt{3}R}{2}}^{\frac{3\sqrt{3}R}{2}} \int_0^{U_0(x)} (x^2 + y^2)^{-\frac{\alpha}{2}} \mathrm{d}y\mathrm{d}x$$

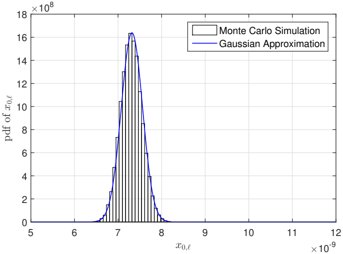
$$\mu_j = \frac{\mu_0}{2} + 6\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_0^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\frac{\alpha}{2}} \mathrm{d}y\mathrm{d}x - 2\lambda \int_{\sqrt{3}R}^{\frac{5\sqrt{3}R}{2}} \int_{U_0(x)}^{U_1(x)} (x^2 + y^2)^{-\frac{\alpha}{2}} \mathrm{d}y\mathrm{d}x \quad j \in \{1, \cdots, 6\}$$

$$\delta_0^2 = 12\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_0^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\alpha} \mathrm{d}y\mathrm{d}x - 12\lambda \int_{\frac{\sqrt{3}R}{2}}^{\frac{3\sqrt{3}R}{2}} \int_0^{U_0(x)} (x^2 + y^2)^{-\alpha} \mathrm{d}y\mathrm{d}x$$

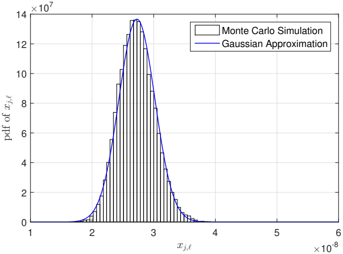
$$\delta_j^2 = \frac{\sigma_0^2}{2} + 6\lambda \int_{\frac{\sqrt{3}}{2}R}^{\infty} \int_0^{\frac{\sqrt{3}}{3}x} (x^2 + y^2)^{-\alpha} \mathrm{d}y\mathrm{d}x - 2\lambda \int_{\sqrt{3}R}^{\frac{5\sqrt{3}R}{2}} \int_{U_0(x)}^{U_1(x)} (x^2 + y^2)^{-\alpha} \mathrm{d}y\mathrm{d}x \quad j \in \{1, \cdots, 6\}$$

Prior distribution of interference powers

- The Gaussian distribution with the same mean and variance is a good approximation of the exact p.d.f. of x_j



(a) BS 0



(b) BS 1

Figure: Comparison between the histogram of $x_{j,\ell}$ (reflecting the p.d.f. of $x_{j,\ell}$) and its corresponding Gaussian approximation. $R = 200$, $\lambda = 0.0005$ and $\alpha = 4$.

Joint MAP estimation for cooperative activity detection

- The conditional density of $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$, given $\overline{\mathrm{Y}}_0$, is given by:

$$\overline{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0|\overline{\mathrm{Y}}_0}\left(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0,\overline{\mathrm{Y}}_0\right)=\overline{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0}(\overline{\mathrm{Y}}_0)\left(\prod_{j=0}^{6}p_j(\boldsymbol{\alpha}_j)\right)\left(\prod_{j=0}^{6}g_j(\mathrm{x}_j)\right)$$

$$=\frac{\exp\left(-\sum_{j=0}^{6}\mathrm{tr}\left(\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}^{-1}\mathrm{Y}_j\mathrm{Y}_j^H\right)\right)}{\pi^{7LM}\prod_{j=0}^{6}|\overline{\boldsymbol{\Sigma}}_{j,\overline{\boldsymbol{\alpha}}_0,\mathrm{x}_j}|^M(\sqrt{2\pi}\delta_j)^L}\exp\left(\sum_{j=0}^{6}\sum_{\omega\in\Psi_j}\left(c_\omega\prod_{i\in\omega}a_i\right)+b_j\right)\exp\left(-\sum_{j=0}^{6}\sum_{\ell=1}^{L}\frac{(x_{j,\ell}-\mu_j)^2}{2\delta_j^2}\right)$$

- Joint MAP estimation of $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ with BS cooperation:

$$\min_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0}\quad \overline{f}_{\mathrm{MAP}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)\triangleq\overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)-\frac{1}{M}\sum_{j=0}^{6}\sum_{\omega\in\Psi_j}\left(c_\omega\prod_{n\in\omega}\alpha_n\right)+\frac{1}{M}\sum_{j=0}^{6}\sum_{\ell=1}^{L}\frac{(x_{j,\ell}-\mu_j)^2}{2\delta_j^2}$$

$$s.t.\quad \alpha_n\in[0,1],\quad n\in\overline{\mathcal{N}}_0$$

$$x_{j,\ell}\geq 0,\quad j\in\{0,1,\cdots,6\},\ \ell\in\mathcal{L}$$

  - $\overline{f}_{\mathrm{MAP}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)$ is $-\frac{1}{M}\overline{f}_{\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0|\overline{\mathrm{Y}}_0}\left(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0,\overline{\mathrm{Y}}_0\right)$ (omit the constant)
  - The impacts of the prior distributions of $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ decrease with M, as $|\overline{f}_{\mathrm{MAP}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)-\overline{f}_{\mathrm{ML}}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)|$ decreases with M

- The problem is non-convex

Coordinate descent (CD) method for joint MAP estimation

- At each step of an iteration, optimize $\overline{f}_{\rm MAP}(\overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0)$ w.r.t. one coordinate in $\{\alpha_n : n \in \overline{\mathcal{N}}_0\} \cup \{x_{j,\ell} : j \in \{0, \cdots, 6\}, \ell \in \mathcal{L}\}$
- Given $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\rm x}_0$ obtained in the previous step, the coordinate optimization w.r.t. α_n equals to the optimization of the increment d in α_n:

$$\min_{d\in[-\alpha_n, 1-\alpha_n]} \overline{f}_{\rm MAP}(\overline{\boldsymbol{\alpha}}_0 + d{\rm e}_i, \overline{\rm x}_0) = \overline{f}_{\rm MAP}(\overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0) + \widetilde{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0)$$

$$\widetilde{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0) \triangleq \overline{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0) - \frac{d}{M} \sum_{j=0}^{6} \sum_{\omega\in\Psi_j : n\in\omega} \left(c_\omega \prod_{n'\in\omega, n'\neq n} \alpha_{n'} \right)$$

and the coordinate optimization w.r.t. $x_{j,\ell}$ equals to the optimization of the increment d in $x_{j,\ell}$:

$$\min_{d\in[-x_{j,\ell}, +\infty)} \overline{f}_{\rm ML,j}(\overline{\boldsymbol{\alpha}}_0, {\rm x}_j + d{\rm e}_\ell) + \frac{(x_{j,\ell} - \mu_j + d)^2}{2M\sigma_j^2} = \overline{f}_{\rm ML,j}(\overline{\boldsymbol{\alpha}}_0, {\rm x}_j) + \widetilde{f}_{x,j,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0)$$

$$\widetilde{f}_{x,j,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\rm x}_0) \triangleq \log(1 + d{\rm e}_\ell^T \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha},{\rm x}}^{-1} {\rm e}_\ell) - \frac{d{\rm e}_\ell^T \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha},{\rm x}}^{-1} \widehat{\boldsymbol{\Sigma}}_{{\rm Y}_j} \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha},{\rm x}}^{-1} {\rm e}_\ell}{1 + d{\rm e}_\ell^T \boldsymbol{\Sigma}_{j,\boldsymbol{\alpha},{\rm x}}^{-1} {\rm e}_\ell} + \frac{(x_{j,\ell} - \mu_j + d)^2}{2M\sigma_j^2}$$

Coordinate descent (CD) method for joint MAP estimation

Theorem (Solution of Optimization w.r.t. α_n)

Given $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. α_n is:

$$\overline{d}^*_{\mathrm{MAP,1,n}}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}}) = \underset{d \in \widetilde{\mathcal{A}}_n(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \cup \{-\alpha_n, 1-\alpha_n\}}{\arg\min} \widetilde{f}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$$

where

$$\widetilde{\mathcal{A}}_n(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \{d \in [-\alpha_n, 1-\alpha_n] : \widetilde{h}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) = 0\}$$

$$\widetilde{h}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \overline{h}_{\alpha,n}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) - \frac{1}{M} \sum_{j=0}^{6} \sum_{\omega \in \Psi_j : n \in \omega} \left(c_\omega \prod_{n' \in \omega, n' \neq n} \alpha_{n'} \right)$$

Coordinate descent (CD) method for joint MAP estimation

Theorem (Optimal Solution of Optimization w.r.t. $x_{j,\ell}$)

Given $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathrm{x}}_0$ obtained in the previous step, the optimal solution of the coordinate optimization w.r.t. $x_{j,\ell}$ is:

$$\overline{d}^*_{\mathrm{MAP},2,\ell}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}}) = \underset{d \in \widetilde{\mathcal{X}}_{j,\ell}(\overline{\boldsymbol{\alpha}}_0,\overline{\mathrm{x}}_0)\cup\{-x_{j,\ell}\}}{\arg\min} \widetilde{f}_{x,j,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0)$$

where

$$\widetilde{\mathcal{X}}_{j,\ell}(\overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \{d \geq -x_{j,\ell} : \widetilde{h}_{x,j,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) = 0\}$$

$$\widetilde{h}_{x,j,\ell}(d, \overline{\boldsymbol{\alpha}}_0, \overline{\mathrm{x}}_0) \triangleq \frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha},\mathrm{x}} \mathrm{e}_\ell}{1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha},\mathrm{x}} \mathrm{e}_\ell} - \frac{\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha},\mathrm{x}} \widehat{\boldsymbol{\Sigma}}_{\mathrm{Y}_j} \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha},\mathrm{x}} \mathrm{e}_\ell}{(1 + d\mathrm{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\boldsymbol{\alpha},\mathrm{x}} \mathrm{e}_\ell)^2} + \frac{d + x_{j,\ell} - \mu_j}{M\delta_j^2}$$

Algorithm for statistical device activity detection

Algorithm 4 (Statistical device activity detection with BS cooperation)

Initialization: choose $\boldsymbol{\Sigma}_{\overline{\boldsymbol{\alpha}}_0}^{-1} = \frac{1}{\sigma^2}\mathbf{I}_L$, $\overline{\boldsymbol{\alpha}}_0 = 0$, $\overline{\mathbf{x}} = 0$.
repeat
for $n \in \overline{\mathcal{N}}_0$ **do**
  Calculate $d_n = \overline{d}^*_{\mathrm{ML,1,n}}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}})$ (ML) or $d_n = \overline{d}^*_{\mathrm{MAP,1,n}}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}})$ (MAP).
  Update $\alpha_n = \alpha_n + d_n$ (CD update).
  Update $\boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} = \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} - \frac{d_n g_n \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} \mathbf{s}_n \mathbf{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}}}{1 + d_n g_n \mathbf{s}_n^H \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} \mathbf{s}_n}, j \in \{0, 1, \cdots, 6\}$ (estimated covariance matrix update).
end for
for $j = 0$ to 6 **do**
  for $\ell \in \mathcal{L}$ **do**
    Calculate $d_\ell = \overline{d}^*_{\mathrm{ML},2,\ell}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}})$ (ML) or $d_\ell = \overline{d}^*_{\mathrm{MAP},2,\ell}(\overline{\boldsymbol{\alpha}}_0, \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}})$ (MAP).
    Update $x_\ell = x_\ell + d_\ell$ (CD update).
    Update $\boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} = \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} - \frac{d_\ell \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} \mathbf{e}_\ell \mathbf{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}}}{1 + d_\ell \mathbf{e}_\ell^T \boldsymbol{\Sigma}^{-1}_{j,\overline{\boldsymbol{\alpha}}_0,\overline{\mathbf{x}}} \mathbf{e}_\ell}$ (estimated covariance matrix update).
  end for
end for
until $\overline{\boldsymbol{\alpha}}_0$ and $\overline{\mathbf{x}}$ satisfy some stopping criterion.

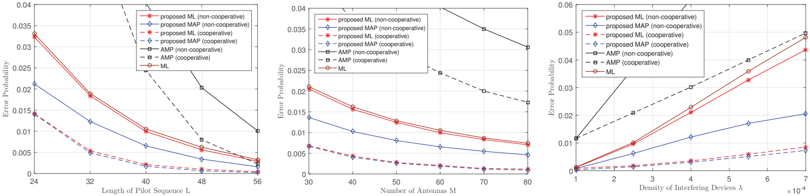
Algorithm for statistical device activity detection with BS cooperation

- Under the mild condition that each coordinate optimization has a unique optimal solution, the algorithm converges to a stationary point of the corresponding statistical estimation problem, as the number of iterations goes to infinity [Bertsekas (1999), Prop. 2.7.1]
  - Different initial points usually correspond to different stationary points
  - Numerical results show that the stationary point corresponding to the initial point $\overline{\boldsymbol{\alpha}}_0 = 0$, $\overline{\mathrm{x}} = 0$ usually provides good detection performance
- The computational complexities of each iteration of the joint ML estimation and the joint MAP estimation with BS cooperation are $\mathcal{O}(\overline{N}_0 L^2 + L^3)$ and $\mathcal{O}(\sum_{j=0}^{6} N_j 2^{N_j} + \overline{N}_0 L^2 + L^3)$, respectively
  - The actual computational complexity for the joint MAP estimation is much lower as $\overline{\boldsymbol{\alpha}}_0$ is a sparse vector

Simulation setup

- N_0 devices are uniformly distributed in cell 0, and each device in cell 0 is active with probability p_a (marginal p.m.f.)
- The locations of the devices out of cell 0 are distributed according to a homogeneous PPP with λ
  - The number of active devices in any other cell is random and has average $\frac{3\sqrt{3}}{2}R^2\lambda$
- Treat the devices in cell 0 and the other cells differently to separate the impacts of N_0 and inter-cell interference intensity for non-cooperative detection
- Independently generate 2000 realizations for the locations of devices, s_n, α_n, $n \in \mathcal{N}$ and $\mathrm{h}_{n,j}$, $n \in \mathcal{N}$, $j \in \{0, 1, \cdots, 6\}$, and evaluate the average error probability over all 2000 realizations
- Choose $R = 200$, $\lambda = 0.005$, $p_a = 0.05$, $N_0 = 500$, $\gamma = 3$, $L = 40$, $M = 60$, and $\sigma^2 = \frac{R^{-\gamma}}{10}$, unless otherwise stated
- Consider three baseline schemes: AMP (non-cooperative) [Liu & Yu (2018)], AMP (cooperative) [Chen et al. (2020)], ML [Fengler et al. (2021)] with numerically optimized thresholds

i.i.d. device activities

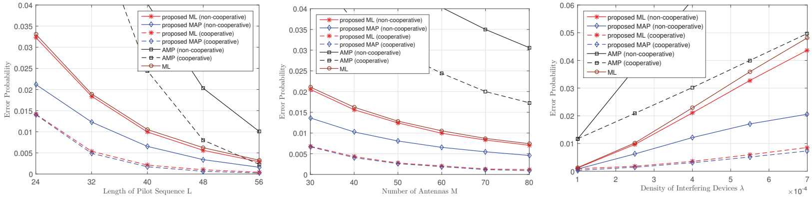


(a) L (b) M (c) λ

Figure: Error probability versus pilot length L, number of antennas M, and density of active interfering devices λ.

- Proposed ML (non-cooperative) significantly outperforms ML, especially in the high interference regime
  - The gain comes from the explicit consideration of interference
- Proposed MAP (non-cooperative) significantly outperforms proposed ML (non-cooperative), and the gain decreases with L and M and increases with λ
  - The gain derives from the incorporation of the prior distributions of the device activities and interference powers

i.i.d. device activities

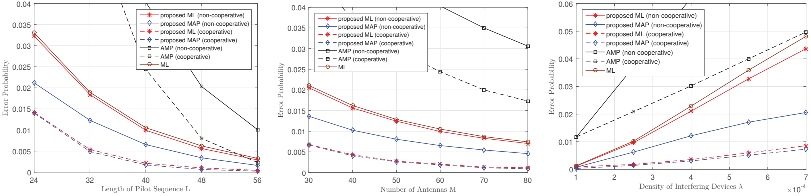


(a) L (b) M (c) λ

Figure: Error probability versus pilot length L, number of antennas M, and density of active interfering devices λ.

- Each proposed cooperative scheme significantly outperforms its non-cooperative counterpart
  - The gain is due to the exploitation of more observations from neighbor BSs and the utilization of more network parameters
- Performance of proposed MAP (cooperative) is similar to that of proposed ML (cooperative)
  - Prior knowledge of the device activities and interference powers brings a relatively smaller gain under BS cooperation

i.i.d. device activities



(a) L (b) M (c) λ

Figure: Error probability versus pilot length L, number of antennas M, and density of active interfering devices λ.

- The statistical estimation schemes (i.e., proposed joint MLs, proposed joint MAPs, and ML) significantly outperform AMPs
- The error probability of each scheme decreases with L and M and increases with λ

Group device activities in first instance

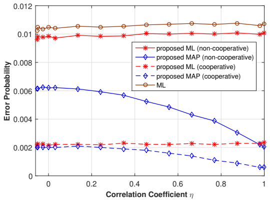


Figure: Error probability versus correlation coefficient η.

- The error probabilities of proposed MAP (non-cooperative) and the proposed MAP (cooperative) significantly decrease with η, while the error probabilities of the other schemes nearly do not change with η
    - Demonstrate the value of exploiting the correlation among device activities

Group device activities in second instance

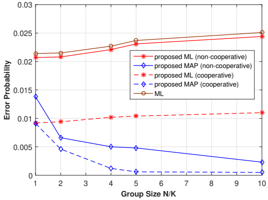


Figure: Error probability versus group size $\frac{N_0}{K}$. $L = 30$.

- When N/K increases, the variance of the number of active devices increases and the sample space of device activities reduces
- The error probabilities of proposed ML (non-cooperative), proposed ML (cooperative), and ML increase with $\frac{N_0}{K}$
  - The error probability significantly increases when the number of active devices is large if correlation is not utilized
- The error probabilities of proposed MAP (non-cooperative) and proposed MAP (cooperative) decrease with $\frac{N_0}{K}$
  - The exploitation of correlation narrows down the set of possible activity states

Conclusion

- We consider non-cooperative device activity detection and cooperative device activity detection in a multi-cell network
- Under each detection mechanism, we formulate the problems for the joint ML estimation and the joint MAP estimation of both device activities and interference powers
- We propose an iterative algorithm to obtain a stationary point of each problem using the coordinate descent method
- Each proposed joint ML estimation extends the existing ML estimation by additionally estimating interference powers
- Each proposed joint MAP estimation further enhances the corresponding joint ML estimation by exploiting prior distributions of device activities and interference powers
- The proposed cooperative joint ML and MAP estimations outperform their non-cooperative counterparts, at the costs of increasing backhaul burden, knowledge of network parameters and computational complexities

Publications

- D. Jiang and Y. Cui*, "ML and MAP Device Activity Detections for Grant-Free Massive Access in Multi-Cell Networks," to appear in IEEE Trans. Wireless Commun., 2021.
- Y. Jia, W. Jiang, and Y. Cui*, "Statistical Device Activity Detection for Massive Grant-Free Access under Frequency-Selective Rayleigh Fading," submitted to IEEE Trans. Wireless Commun., 2021.
- Y. Jia, W. Jiang, and Y. Cui*, "Device Activity Detection for Grant-Free Massive Access Under Frequency-Selective Rayleigh Fading," in Proc. of GLOBECOM, Dec. 2021.
- D. Jiang and Y. Cui*, "MAP-based pilot state detection in grant-free random access for mMTC," in Proc. of IEEE SPAWC, May 2020.
- D. Jiang and Y. Cui*, "ML estimation and MAP estimation for device activities in grant-free random access with interference," in Proc. of IEEE WCNC, Apr. 2020.

Reference I

Andrews, J. G., Choi, W., & Heath, R. W. (2007). Overcoming interference in spatial multiplexing mimo cellular networks. *IEEE Wireless Commun.*, *14*, 95–104.

Bertsekas, D. (1999). *Nonlinear Programming*. Athena Scientific.

Chen, Z., Sohrabi, F., Liu, Y.-F., & Yu, W. (2020). Phase Transition Analysis for Covariance Based Massive Random Access with Massive MIMO. *arXiv e-prints*, (p. arXiv:2003.04175). `arXiv:2003.04175`.

Chen, Z., Sohrabi, F., & Yu, W. (2019). Multi-cell sparse activity detection for massive random access: Massive mimo versus cooperative mimo. *IEEE Trans. Wireless Commun.*, *18*, 4060–4074.

Chen, Z., & Yu, W. (2019). Phase transition analysis for covariance based massive random access with massive mimo. In *Proc. ASILOMAR* (pp. 1–5).

Choi, J. (2019). Noma-based compressive random access using gaussian spreading. *IEEE Trans. Commun.*, *67*, 5167–5177.

Reference II

Ding, S., Wahba, G., & Zhu, J. (2011). Learning higher-order graph structure with features by structure penalty. In *Advances in Neural Information Processing Systems 24* (pp. 253–261). Curran Associates, Inc.

Fengler, A., Haghighatshoar, S., Jung, P., & Caire, G. (2021). Non-bayesian activity detection, large-scale fading coefficient estimation, and unsourced random access with a massive mimo receiver. *IEEE Trans. Inf. Theory*, *67*, 2925–2951.

Liu, L., & Yu, W. (2018). Massive connectivity with massive mimo—part i: Device activity detection and channel estimation. *IEEE Trans. Signal Process.*, *66*, 2933–2946.