Munich Workshop on Shannon Coding Techniques

Unsourced Random Access in Cell-Free User-Centric Wireless Networks

Giuseppe Caire, Burak Çakmak, Eleni Gkiouzepi, Manfred Opper

Technische Universität Berlin

Technische Universität München, 3-5 April, 2024

CommIT
Communications and Information Theory Chair

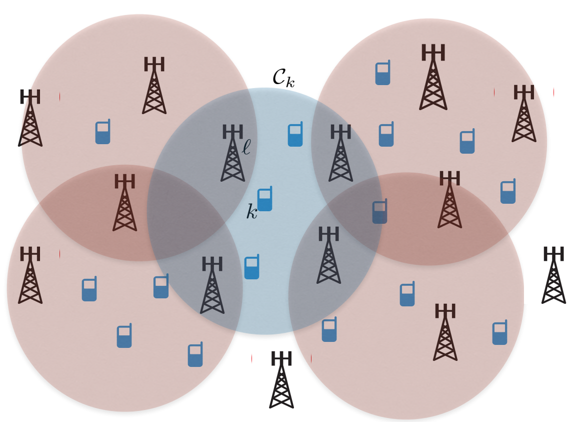
Cell-Free User-Centric Wireless Networks

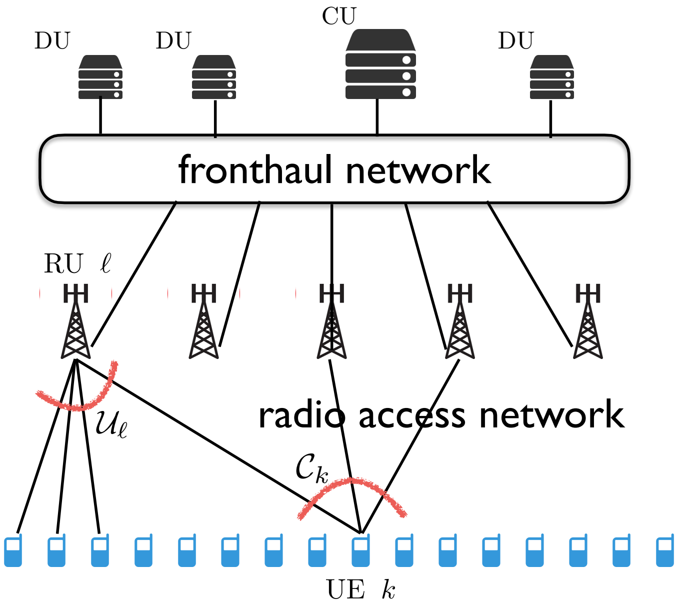
- Typically operating conventional (2 – 7 GHz) frequency bands (TDD reciprocity, UL/DL duality, pilot contamination/decontamination, linear precoding/detection).

- Scenarios: campus networks, ultra-dense deployments, super-high spectral efficiency ... Imagine a sport stadium with 10,000 users, on a 20-60 MHz bandwidth, served by 20 RUs with 10 antennas each, achieving ~ 50 bit/s/Hz per 10×10 m^2.

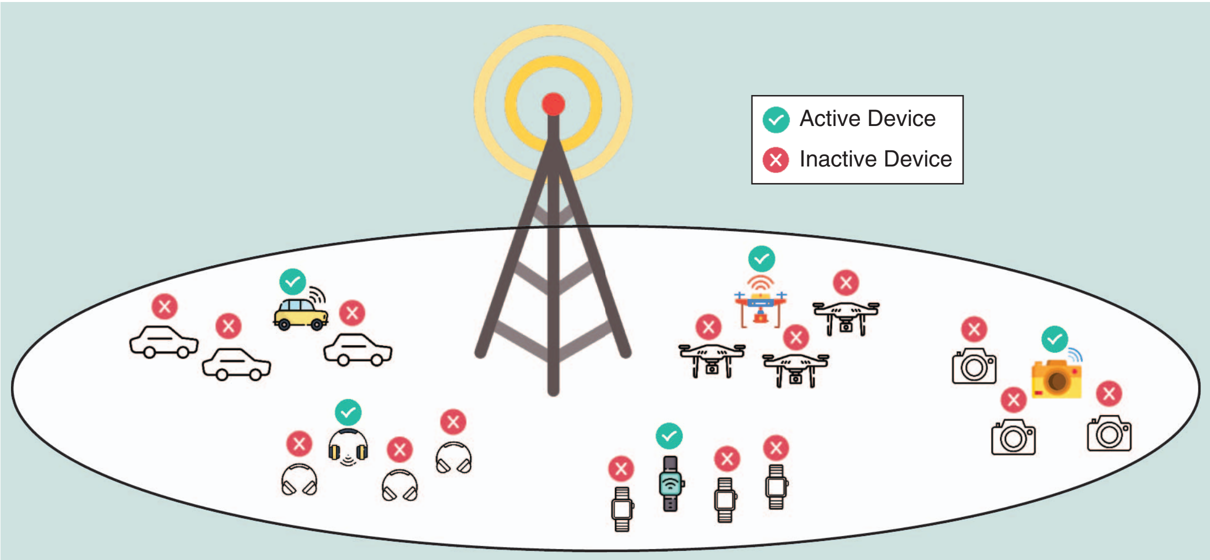
Cell-Free User-Centric Wireless Networks

- Each UE is served by a user-centric cluster of RUs; each RU participates in multiple user-centric clusters.
- The UE-RU association is described by a bipartite graph.
- RUs are connected with DUs via a flexible fronthaul network, and implement the user-centric cluster processors (PHY layer) as SDVNF.
- A CU implements higher level centralized functions.





Activity Detection and Unsourced Random Access



- Activity Detection (AD): each UE has a unique access code. Goal: determine who is active.

- Unsourced RA (uRA): when active, UEs transmit a randomly chosen codeword of the same codebook. Goal: determine the set of active messages.

- Similar but some fundamental differences ...

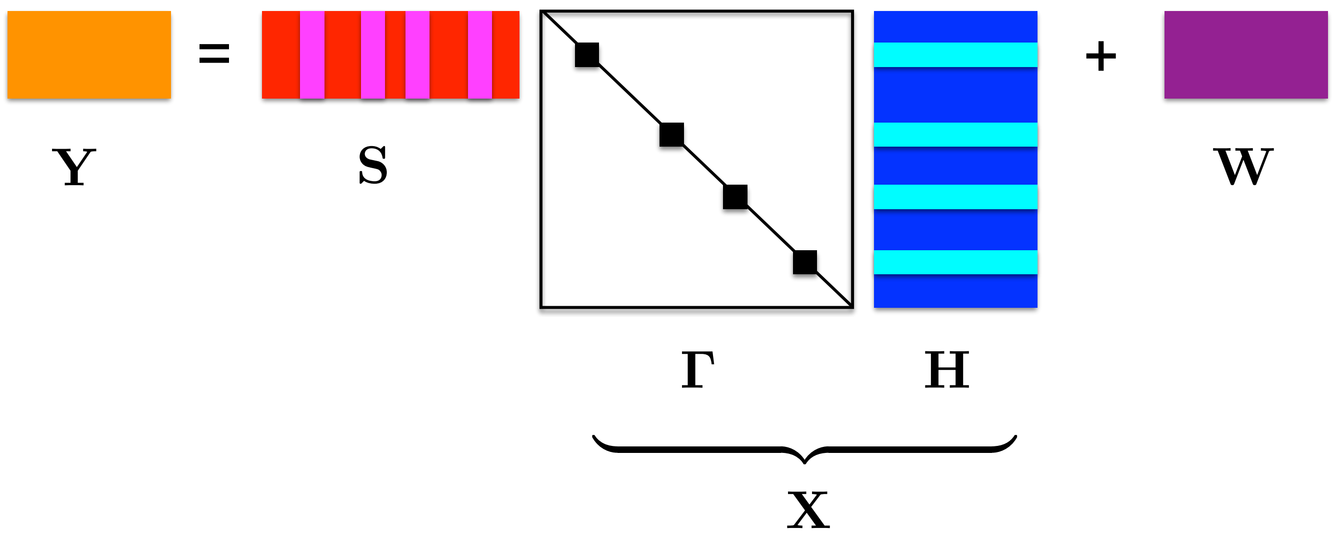
The multiple antenna case: activity detection

- The received baseband signal over the L channel uses and the M antennas:

$$\mathbf{Y} = \mathbf{S}\mathbf{A}\mathbf{G}^{1/2}\mathbf{H} + \mathbf{W}$$

  where $\mathbf{W}$ is Gaussian i.i.d. noise.

- $\mathbf{\Gamma} = \mathbf{A}\mathbf{G} = \text{diag}(\gamma)$, where $\gamma_k = a_k g_k$ for user k with LSFC g_k.
- $\mathbf{X} = \mathbf{\Gamma}\mathbf{H}$ contains Bernoulli-Gaussian rows. Each k-th row has i.i.d. $\mathcal{CN}(0, g_k)$the elements, given $a_k = 1$.



Technische Universität Berlin

Scaling regimes for $K_{\text{tot}}, K_a, L, M$

- Compressed Sensing (CS) regime: in order to obtain a "stable" estimate the (sparse) rows of $\mathbf{\Gamma}^{1/2}\mathbf{H}$ we need $L \geq K_a \log(K_{\text{tot}}/K_a)$ (more measurements than unknowns).

- Identifiability regime (quadratic): two K_a-sparse vectors $\gamma, \gamma' \in \mathbb{R}_+^{K_{\text{tot}}}$ can be distinguished based on $\mathbf{Y}$ if $\mathbf{A}(\mathbf{\Gamma} - \mathbf{\Gamma}')\mathbf{A}^{\mathsf{H}} \neq \mathbf{0}$. This yields $K_a \leq L^2$ (quadratic in the signature dimension).

- Achievability of the quadratic regime: AMP fails (as well as any CS algorithm) but a relaxed ML algorithm, practically implemented by componentwise rank-1 update minimization, achieves the quadratic regimes [Fengler, Haghighatshoar, Jung, and GC, TIT 2021].

- Note: Beyond the linear CS regime, it is "impossible" to obtain "good" channel estimates (activity can be still detected for large M, but the estimated channels has large MSE). Therefore, the quadratic regime is intrinsically "non-coherent".

Key Difference Between AD and uRA

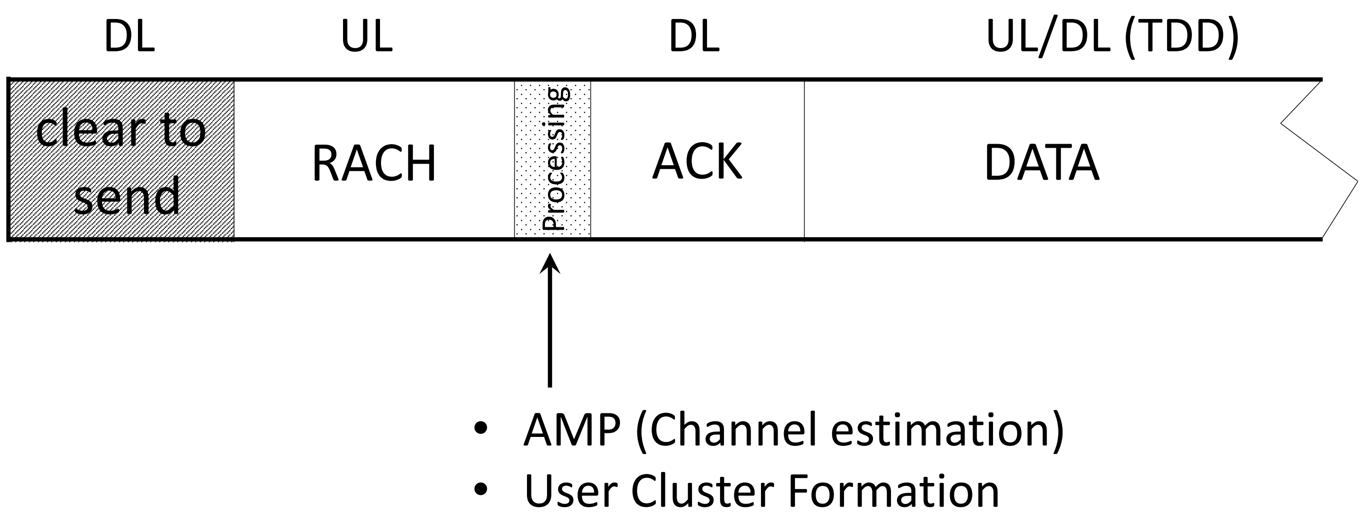
- In AD, provided that the users' LSFCs are known, we can treat $\mathbf{X} = \mathbf{\Gamma H}$ as row-wise Bernoulli-Gaussian: Bayesian estimation formulation is "relatively simple", in particular the Posterior Mean Estimate (PME) denoising function in AMP is tractable.

- In uRA, K_{tot} is irrelevant (it could be arbitrarily large): what counts is the number of codewords N (number of columns of $\mathbf{S}$).

- Any active user can pick any codeword: a fixed correspondence between the columns of $\mathbf{S}$ and the LSFCs is not possible.

- As a consequence, $\mathbf{X} = \mathbf{\Gamma H}$ is not row-wise Bernoulli-Gaussian: Bayesian estimation formulation is "much more complicated" (and heavily relies on assumptions on the LSFC distribution for randomly placed users, .. risk of model mismatch, etc ...).

uRA for Cell-Free User-Centric Networks



- Interestingly, this arrangement is conceptually very similar to the 2-step RACH currently studied in 3GPP.

uRA for Cell-Free User-Centric Networks

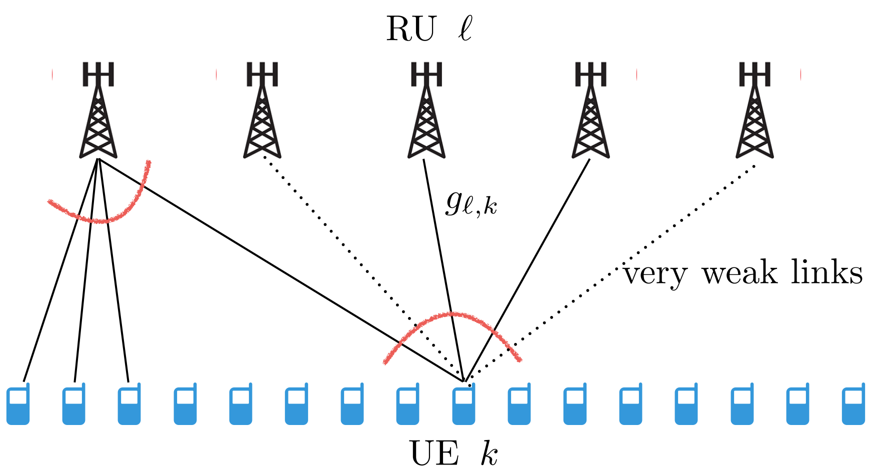
- In uRA for cell-free systems the statistics of the aggregate channel vector over all RUs is more complicated. Conditioning on the position from which codeword k may be transmitted, we have

$$\mathbf{x}_k = a_k[\mathbf{h}_{k,1}, \mathbf{h}_{k,2}, \ldots, \mathbf{h}_{k,B}]$$

from which

$$\mathbf{x}_k \sim (1-\lambda)\delta(\mathbf{x}_k) + \lambda \mathfrak{g}(\mathbf{x}_k|\mathbf{0}, \mathbf{\Sigma}_k)$$

where $\mathbf{\Sigma}_k = \text{diag}(g_{k,b} : b \in [B]) \otimes \mathbf{I}_M$.



uRA for Cell-Free User-Centric Networks

- However, message k may be transmitted by any user in any random position. Hence, $\{g_{k,b} : b \in [B]\}$ are a set of B jointly distributed LSFCs that depend on the (random) transmitter position.

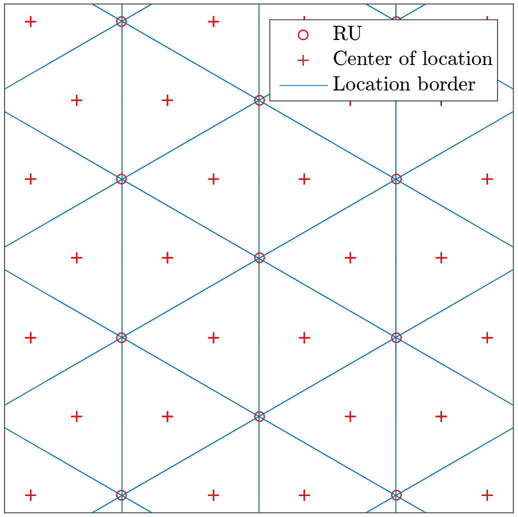
- Removing the conditioning involves a very complicated B-dimensional integral (or, equivalently, integrating over the random transmitter position).

- In 3G-4G-5G the RACH codebook is cell-dependent.

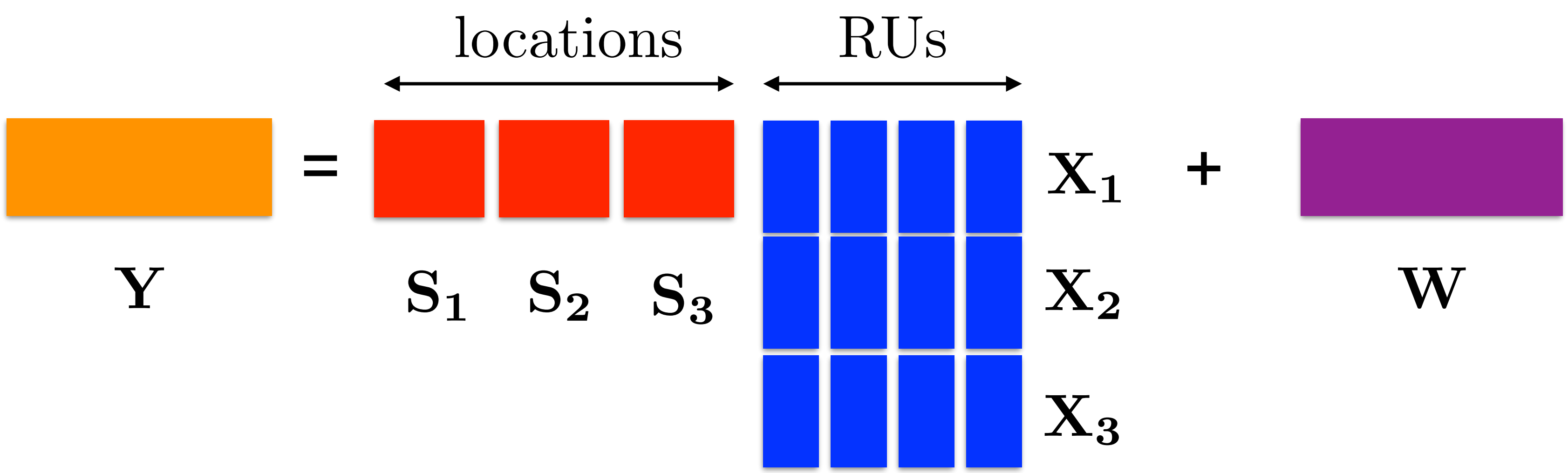
- In a cell-free user-centric network, we propose location-dependent access codebooks.

Our Idea: Location-Based Codebooks

- All users in "location" $u \in \mathcal{U}$ make use of codebook $\mathbf{S}_u$.
- Assuming perfectly co-located users, the channel statistics of all users $k \in \mathcal{L}_u$ are defined by nominal LSFCs $\{\tilde{g}_{u,b} : b \in [B]\}$.
- This establishes a fixed relation between codewords and LSFCs.



Channel/Observation Model for uRA in Cell-Free



Channel/Observation Model for uRA in Cell-Free

- The aggregate signal received during the RACH slot by all the RUs is given by

$$\mathbf{Y} = \sum_{u=1}^{U} \mathbf{S}_u \mathbf{X}_u + \mathbf{W}$$

- With $F = BM$, $\mathbf{W} \sim_{\text{i.i.d.}} \mathcal{CN}(\mathbf{0}, \sigma_w^2)$, $\mathbf{X}_u \sim_{\text{i.i.d.}} \mathbf{x}_u$ for an *arbitrary* and independent (for each u) random vector $\mathbf{x}_u \in \mathbb{C}^F$ with fulfilling the "finite-moments" condition $\mathbb{E}[\|\mathbf{x}_u\|^p] < \infty, \;\; \forall p \in \mathbb{N}_+$.

- Moreover, $\mathbf{S}_u \sim_{\text{i.i.d.}} \mathcal{CN}(\mathbf{0}, 1/L)$ has dimension $L \times N_u$, and we assume $N_u/L = \alpha_u$ as $L \to \infty$ (system scaling parameter).

New Multisource AMP

For the system at hand, we consider the following AMP algorithm:

- Initialize: $\mathbf{X}_u^{(1)} = \mathbf{0}$ for $u \in [U]$ and $\mathbf{Z}^{(0)} = \mathbf{0}$.
- For iteration steps $t = 1, 2, \ldots, T$, repeat:

$$\mathbf{V}_u^{(t)} = \mathbf{S}_u \mathbf{X}_u^{(t)} - \frac{1}{\alpha_u} \mathbf{Z}^{(t-1)} \mathbf{Q}_u^{(t)} \tag{1a}$$

$$\mathbf{Z}^{(t)} = \mathbf{Y} - \sum_{u=1}^{U} \mathbf{V}_u^{(t)} \tag{1b}$$

$$\mathbf{R}_u^{(t)} = \mathbf{S}_u^{\mathsf{H}} \mathbf{Z}^{(t)} + \mathbf{X}_u^{(t)} \tag{1c}$$

$$\mathbf{X}_u^{(t+1)} = \eta_{u,t}(\mathbf{R}_u^{(t)}) \tag{1d}$$

Composite/Multisource MMV-AMP

- where $\eta_{u,t}(\cdot) : \mathbb{C}^F \to \mathbb{C}^F$ is an appropriately defined deterministic and (u,t)-dependent “denoiser” function.

- $\eta_{u,t}(\cdot)$ applied to an $N_u \times F$ matrix $\mathbf{R}_u$ denotes the $N \times F$ matrix with its nth row given by $\eta_{u,t}(\mathbf{r}_n)$, i.e.,

$$\eta_{u,t}(\mathbf{R}) = [\eta_{u,t}(\mathbf{r}_1)^\top, \eta_{u,t}(\mathbf{r}_2)^\top, \cdots, \eta_{u,t}(\mathbf{r}_{N_u})^\top]^\top .$$

- where we define

$$\mathbf{Q}_u^{(t+1)} \doteq \mathbb{E}[\eta'_{u,t}(\mathbf{x}_u + \boldsymbol{\phi}^{(t)})],$$

with $\mathbf{Q}_u^{(1)} = \mathbf{0}$ and $\{\boldsymbol{\phi}^{(t)}\}_{t\in[T]}$ is a Gaussian process (defined in the following) independent of the random vector $\mathbf{x}_u$.

Composite/Multisource MMV-AMP

- For a differentiable vector-valued function $\eta(\mathbf{r}) : \mathbb{C}^F \to \mathbb{C}^F$ we denote by $\eta'(\mathbf{r})$ its $F \times F$ Jacobian matrix with the entries

$$[\eta'(\mathbf{r})]_{ij} = \frac{\partial[\eta(\mathbf{r})]_j}{\partial r_i} \quad \forall i, j \in [F].$$

(here, for a complex number $r = x + \mathrm{i}y$, the complex (Wirtinger) derivative is defined $\frac{\partial}{\partial r} = \frac{1}{2}(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y})$. For $\mathbf{R} \in \mathbb{C}^{N\times F}$).

Main Result: a Rigorous State Evolution

Definition 1. (State Evolution) *Let* $\{\boldsymbol{\phi}^{(t)} \in \mathbb{C}^{1\times F}\}_{t\in[T]}$ *be a zero-mean (discrete-time) Gaussian process with its two-time covariances* $\mathbf{C}^{(t,s)} \doteq \mathbb{E}[(\boldsymbol{\phi}^{(t)})^{\mathsf{H}}\boldsymbol{\phi}^{(s)}]$ *for all* $t, s \in [T]$ *constructed recursively according to*

$$\mathbf{C}^{(t,s)} = \sigma^2\mathbf{I} + \sum_{u=1}^{U} \alpha_u \mathbb{E}[(\mathbf{x}_u^{(t)} - \mathbf{x}_u)^{\mathsf{H}}(\mathbf{x}_u^{(s)} - \mathbf{x}_u)] \ , \tag{2}$$

where we define the random vectors for $t \in [T]$ *and* $u \in [U]$

$$\mathbf{x}_u^{(t+1)} := \eta_{u,t}(\mathbf{x}_u + \boldsymbol{\phi}^{(t)}), \tag{3}$$

independent of $\mathbf{x}_u^{(1)}$. $\Diamond$

Main Result: a Rigorous State Evolution

Theorem 1. *Let the matrices in* $\{\{\mathbf{S}_u, \mathbf{X}_u\}, \mathbf{W}\}$ *be defined as before. Then, as* $L \to \infty$, *for all* $t \in [T]$ *and* $u \in [U]$ *there exists a constant* C_p *for all* $p \in \mathbb{N}_+$ *such that*

$$\mathbb{E}\left[\left\|\mathbf{R}_u^{(t)} - (\mathbf{X}_u + \mathbf{\Phi}_u^{(t)})\right\|_{\mathrm{F}}^p\right]^{\frac{1}{p}} \leq C_p \qquad (4)$$

where $\mathbf{\Phi}_u^{(t)} \sim_{\text{i.i.d.}} \phi^{(t)}$ *with the Gaussian process* $\phi^{(t)}$ *as in Definition 1 and* $\{\mathbf{\Phi}_u^{(t)}\}_{u\in[U]}$ *are mutually independent.* □

Corollary 1. *Under the premises of Th. 1, for any small constant* $c > 0$

$$\frac{1}{L^c}\left\|\mathbf{R}_u^{(t)} - (\mathbf{X}_u + \mathbf{\Phi}_u^{(t)})\right\|_{\mathrm{F}} \to 0$$

where convergence is both a.s. and in the p*-th mean.* □

Main Result: a Rigorous State Evolution

- Decoupling principle:

$$\mathbf{r}_{u,n}^{(t)} \sim \mathbf{x}_{u,n} + \boldsymbol{\phi}_{u,n}^{(t)}$$

where $\phi_{u,n}^{(t)}$ is the n-th row of $\boldsymbol{\Phi}_u^{(t)}$ in Th. 1.

- By the Lipschiz-continuity of $\eta_{u,t}$ Th. 1 implies that there exists C_p such that

$$\mathbb{E}\left[\left\|\mathbf{X}_u^{(t+1)} - \eta_{u,t}(\mathbf{X}_u + \boldsymbol{\Phi}_u^{(t)})\right\|_{\mathrm{F}}^p\right]^{\frac{1}{p}} \leq C_p$$

- Convergence of the empirical squared error:

$$\frac{1}{N_u}(\mathbf{X}_u - \mathbf{X}_u^{(t)})^{\mathsf{H}}(\mathbf{X}_u - \mathbf{X}_u^{(t)}) \overset{a.s.}{\to} \mathbb{E}\left[(\mathbf{x}_u - \mathbf{x}_u^{(t)})^{\mathsf{H}}(\mathbf{x}_u - \mathbf{x}_u^{(t)})\right].$$

MSE and PME Bayesian Denoiser

- The decoupling principle and the MSE suggest to choose $\eta_{u,t}(\cdot)$ in order to minimize $\mathbb{E}\left[(\mathbf{x}_u - \mathbf{x}_u^{(t)})^{\mathsf{H}}(\mathbf{x}_u - \mathbf{x}_u^{(t)})\right]$.

- This yields the Posterior Mean Estimator (PME) for $\mathbf{x}_u$ form the observation $\mathbf{r}_u^{(t)} = \mathbf{x}_u + \boldsymbol{\phi}^{(t)}$, i.e.,

$$\eta_{u,t}(\mathbf{r}) = \mathbb{E}[\mathbf{x}_u|\mathbf{r}_u^{(t)}]$$

  that can be easily calculated in closed form given the nominal LFSCs $\{\tilde{g}_{u,b}\}$ and the Bernoulli-Gaussian prior distribution of $\mathbf{x}_u$.

- Interestingly, in this case we obtain also the Jacobian matrix $\eta'_{u,t}(\mathbf{r})$ in closed form (see paper).

Message Detection

- Let $\mathbf{C} := \mathbf{C}^{(T,T)}$ and $\mathbf{R}_u := \mathbf{R}_u^{(T)} \sim \mathbf{X}_u + \mathbf{\Phi}_u^{(T)}$. The decoupled channel model suggests the binary hypothesis test for the detection (active/inactive) of message (u, n) with the two hypotheses:

$$\mathbf{r}_{n,u} \sim \begin{cases} \mathcal{CN}(\mathbf{0}, \mathbf{C}) & a_{n,u} = 0 \ \text{(Hypothesis } \mathcal{H}_0\text{)} \\ \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_u + \mathbf{C}) & a_{n,u} = 1 \ \text{(Hypothesis } \mathcal{H}_1\text{)} \end{cases}$$

- Although we use the prior activity probability λ for the PME denoiser, we prefer to use a Neyman-Pearson test for message activity detection since MD and FA probabilities have a different impact on the system performance.

Message Detection

- The LLR test takes on the form

$$\Delta_{n,u} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \nu_u$$

where ν_u is a suitable threshold and where $\Delta_{n,u}$ is a Hermitian Quadratic form of Gaussian Circularly Symmetric Random Variables (HQF-GRV) under both $\mathcal{H}_0$ and $\mathcal{H}_1$.

- As a consequence, the MD and FA probabilities can be computed in closed form!!! (or approximated with any desired degree of accuracy using the method of Laplace Inversion with Gauss-Chebyshev Quadrature Rules (see oldie [J. Ventura-Traveset, GC, E. Biglieri, and G. Taricco TCOM 1997]).

Achievable rate in the ACK/DL transmission

- Based on the active message channel estimates, for each UE (detected active message) we allocate a user-centric cluster corresponding by the Q RUs $b \in [B]$ with largest $\widetilde{g}_{u,b}$.
- We use MRT to transmit a beamformed coded ACK to the users (and possibly allocate resource for further data communication).
- Using the asymptotic analysis, we obtain the semi-closed form ergodic rate expression:

$$R_{u,n}^{\mathrm{UatF}} = \log\left(1 + \frac{\left|\sum_{b\in\mathcal{C}_u} \mathcal{M}_{u,b}\right|^2}{\sigma_w^2/\rho_{\mathrm{DL}} + \sum_{b\in\mathcal{C}_u} \mathcal{V}_{u,b} + L\sum_{u'\in[U]}\sum_{b\in\mathcal{C}_{u'}} \lambda_{u'}\alpha_{u'}\widetilde{g}_{u,b}\mathcal{Z}_{u',b}}\right),$$

where, for all $(u,b) \in [U]\times[B]$, we define

$$\mathcal{M}_{u,b} \triangleq \mathbb{E}\left[\mathbf{h}_{u,b}\,\eta_{u,T}(\mathbf{h}_{u,b} + \mathbf{z}_b\mathbf{C}_b^{\frac{1}{2}})^{\mathsf{H}}|\mathbf{h}_u, \mathbf{z}\in\mathcal{D}_u\right]$$

$$\mathcal{V}_{u,b} \triangleq \mathrm{Var}\left(\mathbf{h}_{u,b}\,\eta_{u,T}(\mathbf{h}_{u,b} + \mathbf{z}_b\mathbf{C}_b^{\frac{1}{2}})^{\mathsf{H}}|\mathbf{h}_u, \mathbf{z}\in\mathcal{D}_u\right),$$

and the Tx power normalization coefficient

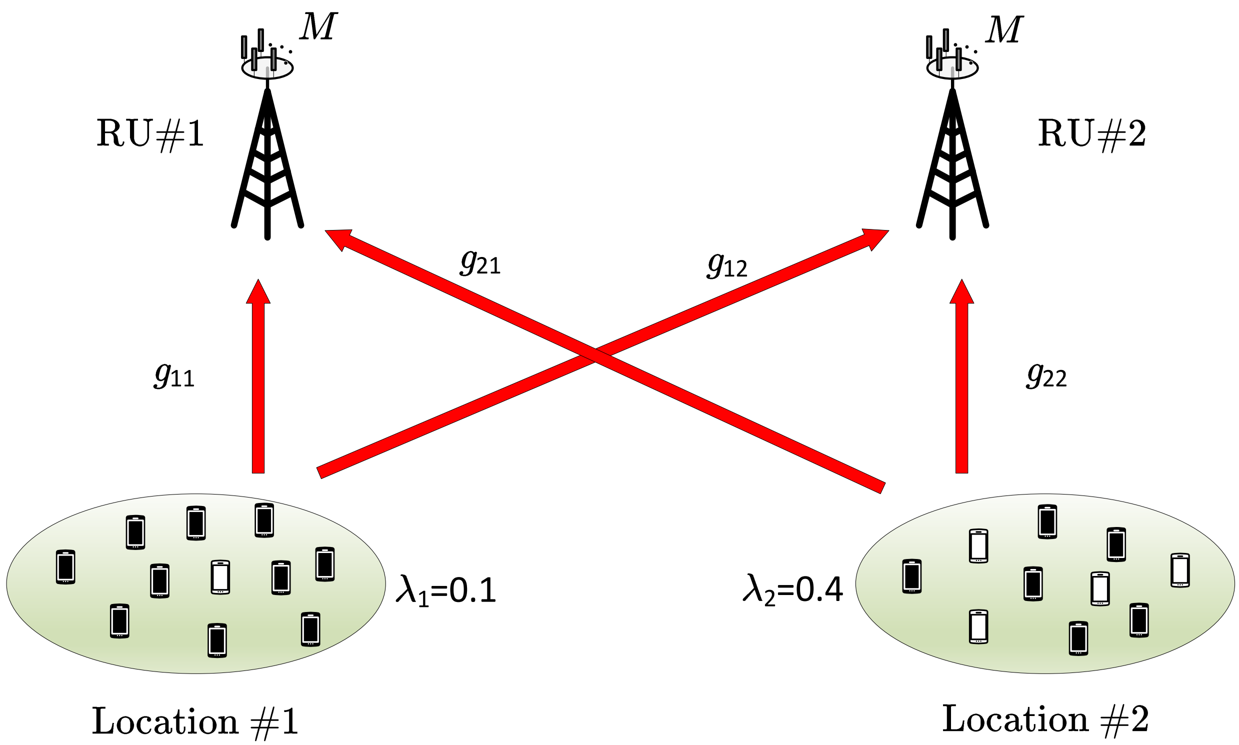
$$\rho_{\mathrm{DL}} = \frac{1}{L} \frac{\sum_{u=1}^{U} \lambda_u \alpha_u}{\sum_{b=1}^{B} \sum_{u \in \mathcal{S}_b} \lambda_u \alpha_u \mathcal{Z}_{u,b}}$$

with

$$\begin{aligned} \mathcal{Z}_{u,b} \quad &\triangleq \quad (1 - \mathrm{P}_u^{\mathrm{md}}) \mathbb{E}\left[\|\eta_{u,T}(\mathbf{h}_{u,b} + \mathbf{z}_b \mathbf{C}_b^{\frac{1}{2}})\|^2 \mid \mathbf{z}, \mathbf{h}_u \in \mathcal{D}_u \right] \\ &\quad + (\lambda_u^{-1} - 1) \mathrm{P}_u^{\mathrm{fa}} \mathbb{E}\left[\|\eta_{u,T}(\mathbf{z}_b \mathbf{C}_b^{\frac{1}{2}})\|^2 \mid \mathbf{z} \in \mathcal{F}_u \right]. \end{aligned}$$

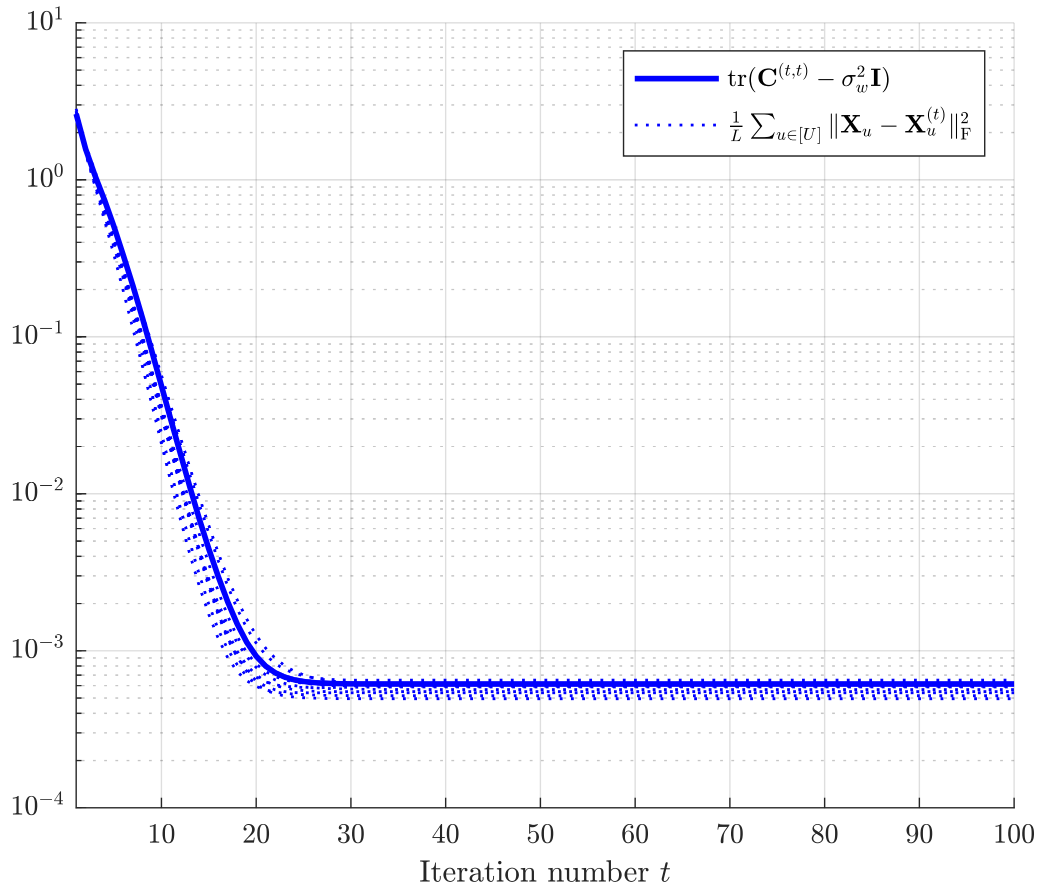
where $\mathbf{h}_u \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_u)$, $\mathbf{z} \sim_{\mathrm{i.i.d.}} \mathcal{CN}(0, 1)$ are mutually independent, $\mathbf{h}_{u,b}$ and $\mathbf{z}_b$ denote the b-th segment of size $1 \times M$ of $\mathbf{u}_u$ and $\mathbf{z}$, respectively, and the events $\mathcal{D}_u$ and $\mathcal{F}_u$ are defined before.

“Wyner model” with $U = B = 2$



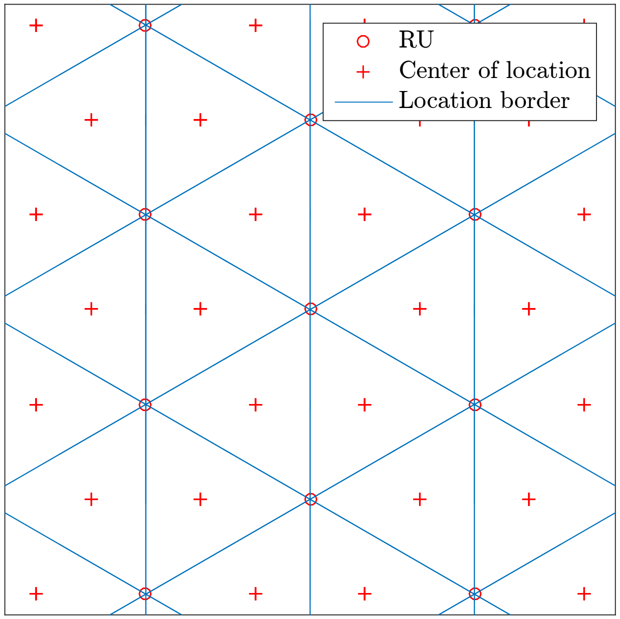
Toy Setup with $U = B = 2$

- $N_1 = N_2 = 2048$, $L = 1024$, SNR$= 10$ dB, $\lambda_1 = 0.1, \lambda_2 = 0.2$, $M = 2$.



More Realistic Setup $U = 16$, $B = 12$

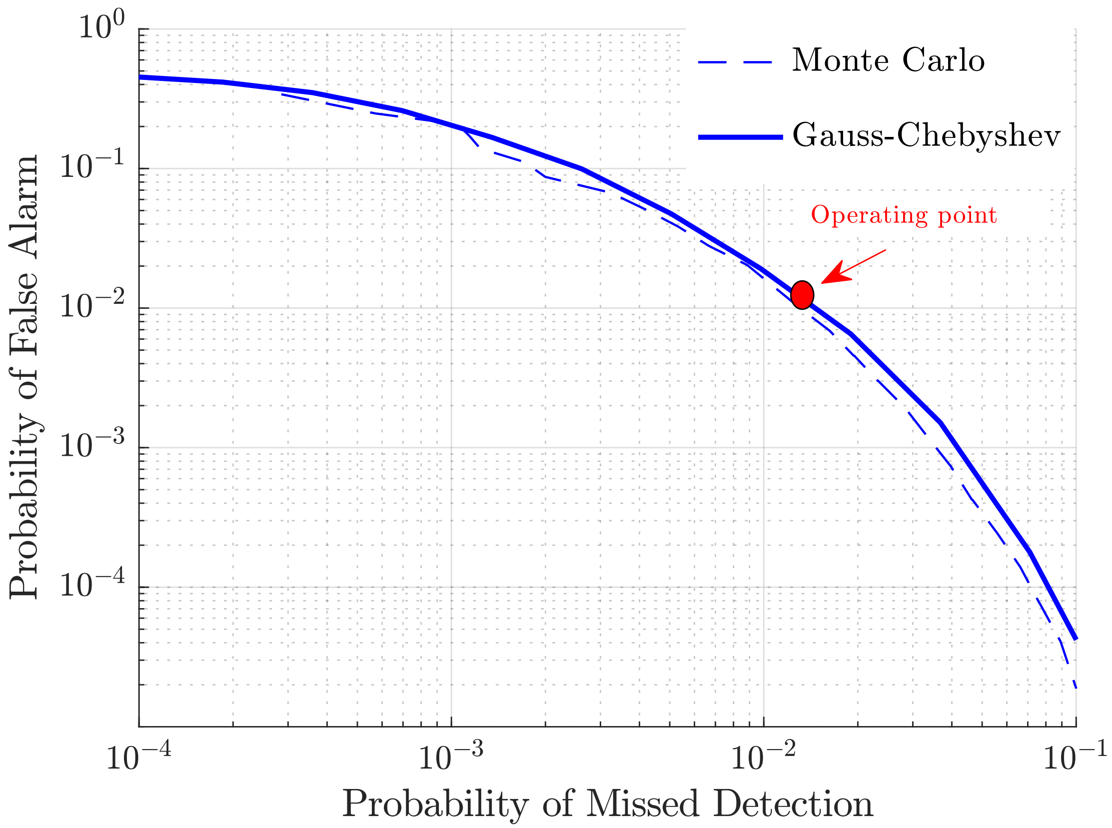
- $N_u = 2048$, $L = 1024$, realistic SNR and distance dependent pathloss model, $\lambda_u \in \{0.003, 0.002\}$, repeated in a periodic pattern.



CommIT
Communications and Information Theory Chair

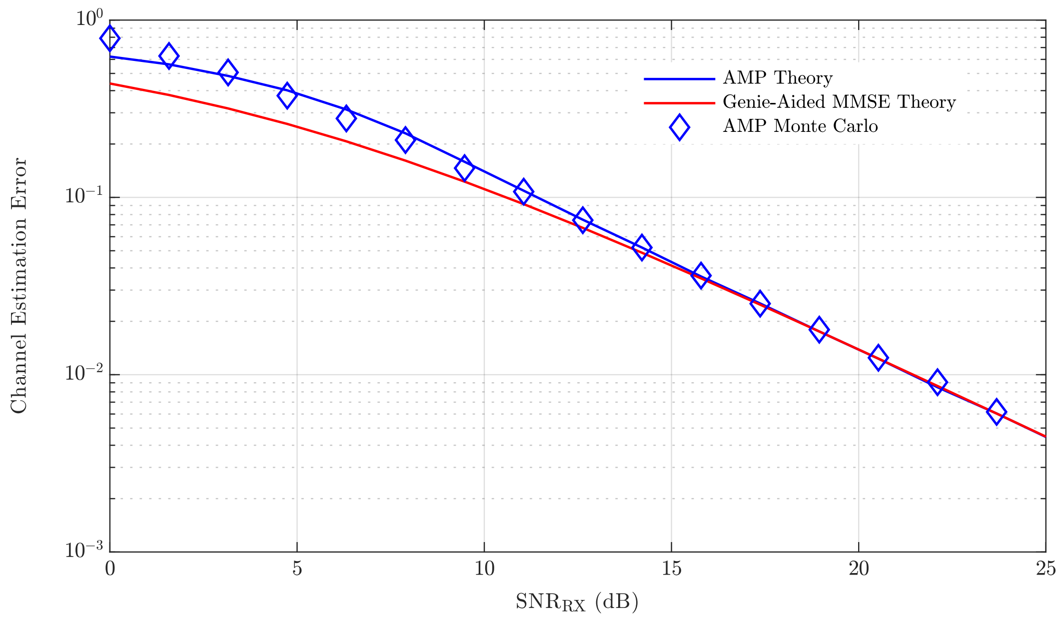
More Realistic Setup $U = 16$, $B = 12$

- We choose to work at the point where $P_{\text{fa}} = P_{\text{md}}$.



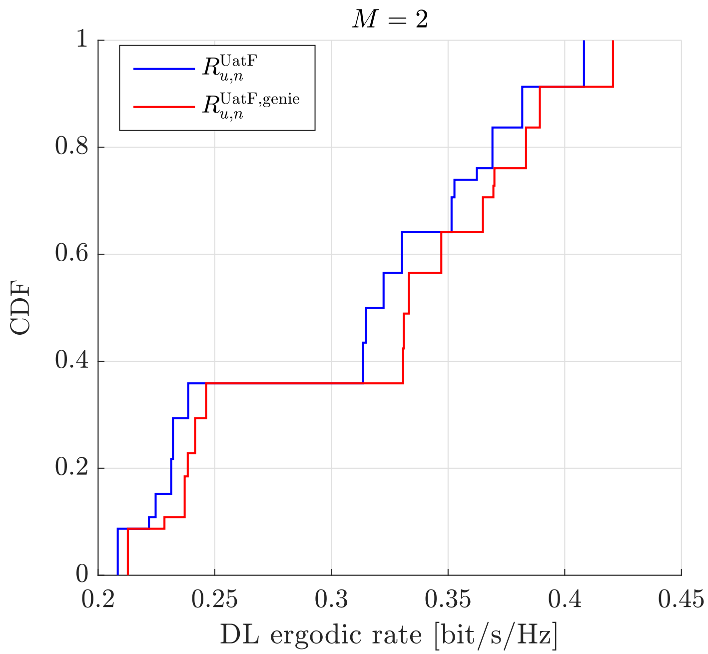
More Realistic Setup $U = 16$, $B = 12$

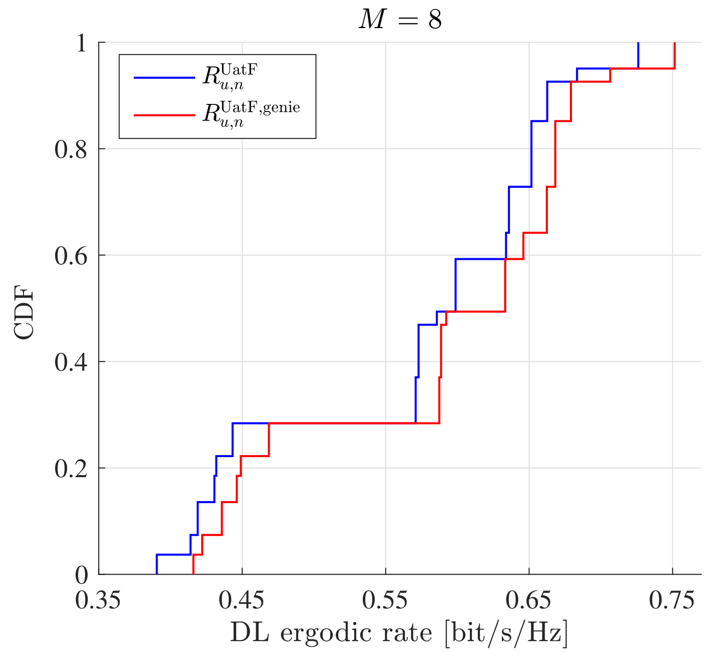
- In these conditions, the channel estimation for the messages in $\mathcal{A}_{\mathrm{d}}$ is excellent:



More Realistic Setup $U = 16$, $B = 12$

- Ergodic rate CDF (over the user population) for the MRT downlink transmission.





Conclusion

- Key Ideas: 1) location-based access codebooks; 2) a novel "multisource" AMP.
- The multisource AMP generalizes conventional AMP and MMV-AMP and can be rigorously analyzed.
- Not shown (see long paper): the Replica-Analysis yields results that coincide with the SE.
- The asymptotic output statistics of the AMP allows (almost) closed-form evaluation of very large systems.
- Work in progress: a) assess mismatch for non-colocated users (done!); b) Joint message detection and RSS-based position estimation (done!); c) Thorough comparison between uRA-based "seamless connectivity" and conventional pilot-allocation/cluster formation (in progress); d) Extension to the multipath/freq.selective case (in progress).

Technische Universität Berlin

Thank You

CommIT
Communications and Information Theory Chair