Solutions for low latency and massive connectivity in future wireless networks

Frédéric Lehmann

Télécom SudParis, Institut Mines-Télécom, IP Paris
Département CITI (Communications, Images et Traitement de l'Information)
Laboratoire SAMOVAR, UMR-CNRS 5157

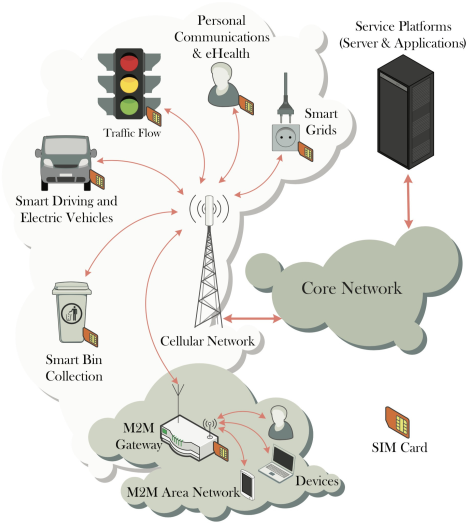
July, 2020

Outline

1 Prospects in wireless research

The rise of the era of IoT and big data...

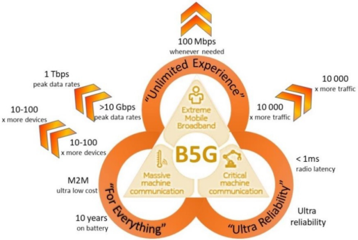
M2M communications



Beyond 5G requirements

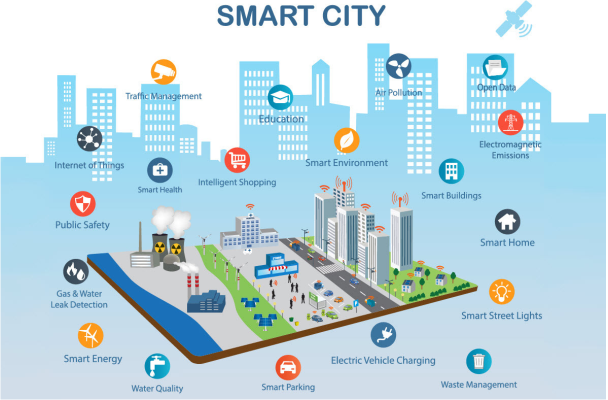
B5G, 5G+, 6G

- **Throughput**: peak data rate up to 1 Tbits/s ($\times 100$ vs. 5G)
- **Latency**: down to $10 - 100\mu$s (/ 10-100 vs. 5G)
- **Density**: up to 10^7 devices/km^2 ($\times 10$ vs. 5G)
- **Reliability**: packet failure rate down to 10^{-7} (/ 100 vs. 5G)



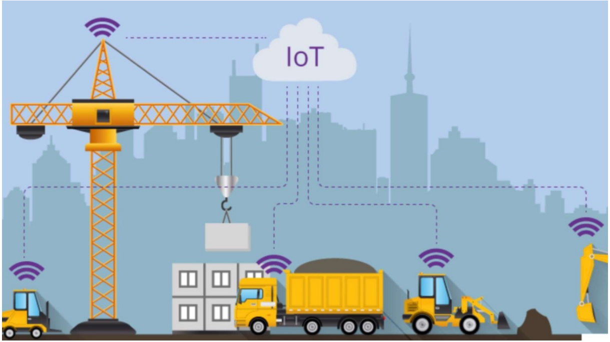
Use Cases: Smart Cities

Optimize traffic light schedule, energy consumption, security, etc



Use Cases: Autonomous Construction

Build roads or buildings according to plan, while respecting safety rules.

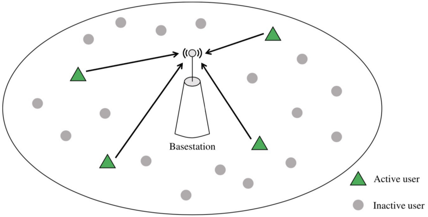


2 Challenges for massive access

Cellular IoT traffic

Assumptions in the uplink direction

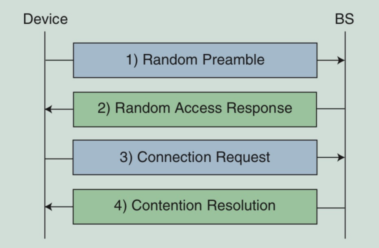
- Massive access: high density of devices per km^2
- Sporadic traffic: each device is active with low probability
- Small packets: active devices send only a few bytes
- Energy constraints: each device has scarce energy supply



4-step Random access procedure

Handshake protocol example: #10 ms in 4G, #1 ms in 5G

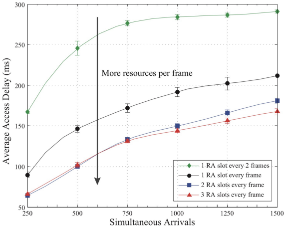
- **Msg 1**: Device sends a randomly chosen preamble
- **Msg 2**: BS sends timing advance and resources for Msg3
- **Msg 3**: Devices sends access request with device identifier
- **Msg 4**: BS sends contention resolution message



4-step Random access procedure

Limitations with high arrival rate

- Low probability of access success
- High preamble collision rate
- High access delay
- High energy consumption



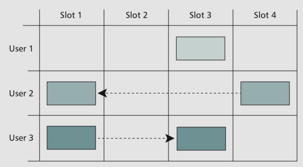
Enhanced random access procedures

Coded random access

- Example: each packet is repeated $\bar{d}$ times on average
- IC/backtracking cancellation to recover collided packets

Limitations

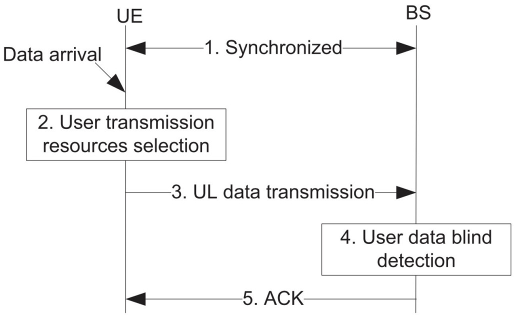
- Keep track of UE ID of each packet (overhead)
- What if synch. and CSI are imperfect?



Grant-free access

2-step procedure

- Active UE sends data over grant-free resources
- BS responds with ACK



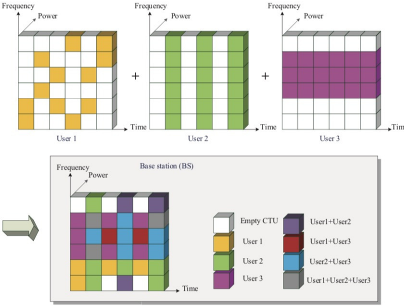
Grant-free access

More complicated communication problem

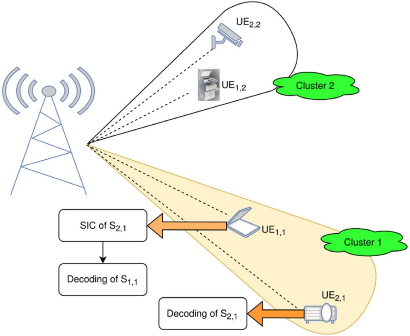
- Active UEs: perform uncoordinated transmission
- Channel: collisions are allowed by design
- BS: resolve collisions
  1. **User Activity Detection** (UAD)
  2. **Multi-User Detection** (MUD)

UE uplink transmission using NOMA

Code domain (CD-NOMA)



Power domain (PD-NOMA)



Multi-access system model

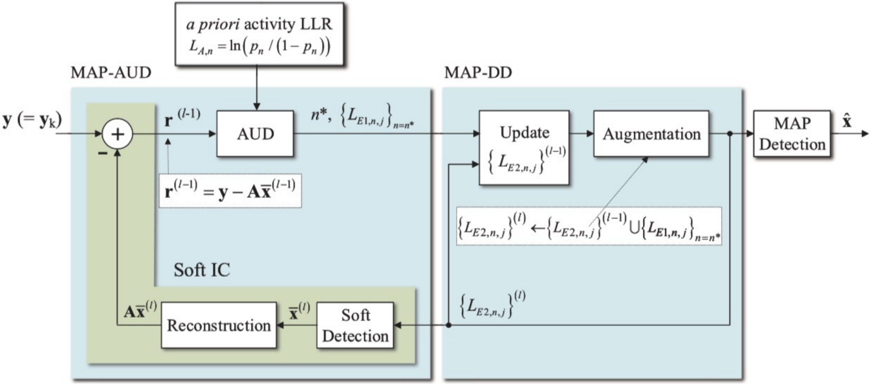
$n \in \{1, 2, \ldots, N\}$: user index - $\mathcal{A}$: set of active users, $|\mathcal{A}| = K$
$m \in \{1, 2, \ldots, M\}$: resource element (RE) index:
frequency (CD-NOMA) / Rx antenna element (PD-NOMA)

x_n: complex symbol sent by user n
s_{mn}: m-th component of the spreading sequence (CD-NOMA) or square-root of power (PD-NOMA) for user n
g_{mn}: user n flat-fading channel gain over RE m

Received signal over the m-th RE:

$$y_m = \sum_{n \in \mathcal{A}} g_{mn} s_{mn} x_n + v_m, \;\; v_m \sim \mathcal{N}(0, \sigma^2) \quad (1)$$

Serial user activity and data detection[1]

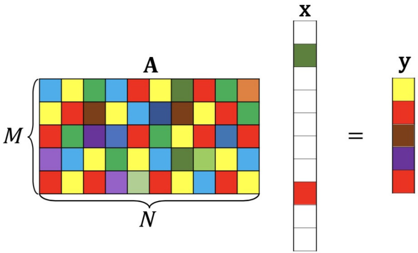


[1] B. K. Jeong, B. Shim and K. B. Lee, "MAP-Based Active User and Data Detection for Massive Machine-Type Communications," IEEE Trans. on Veh. Technol., vol. 67, no. 9, pp. 8481-8494, Sept. 2018.

Compressed-sensing user activity and data detection[2]

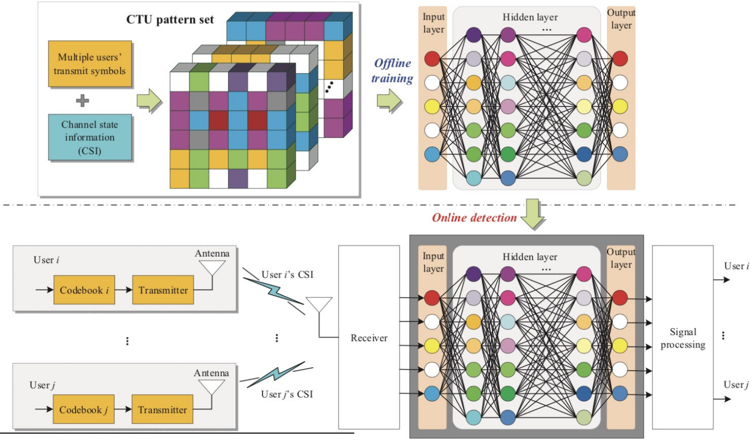
We can rewrite (1) in matrix-vector form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$,

where $[\mathbf{A}]_{mn} = g_{mn}s_{mn}$, $\mathbf{x} = [x_1, \ldots, x_N]^T$, $\mathbf{v} = [v_1, \ldots, v_M]^T$.



[2]L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic and E. de Carvalho, "Sparse Signal Processing for Grant-Free Massive Connectivity: A Future Paradigm for Random Access Protocols in the Internet of Things," in IEEE Sig. Proc. Mag., vol. 35, no. 5, pp. 88-99, Sept. 2018.

AI-based user activity and data detection[3]



[3]W. Kim, Y. Ahn and B. Shim, "Deep Neural Network-Based Active User Detection for Grant-Free NOMA Systems," in IEEE Trans. on Comm., vol. 68, no. 4, pp. 2143-2155, April 2020.

3 Multi-antenna grant-free access

Compressed-sensing grant-free access

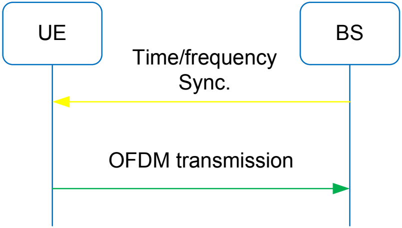
Limitations:

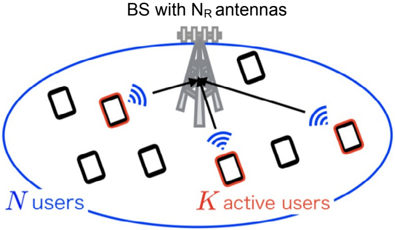
- Channel/spreading sensing matrix $\mathbf{A}$: known constant at the BS
- Needs uplink channel sounding even for inactive UE
- Not robust to fast channel variations

Multi-antenna OFDM grant-free access

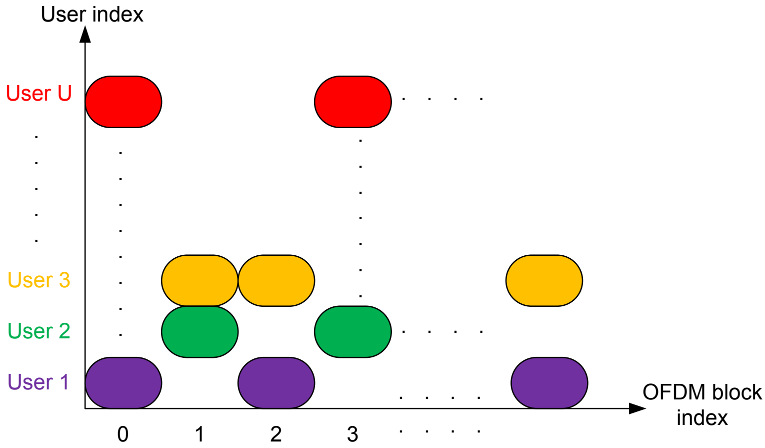
Characteristics:

- Time/frequency synchronization through downlink signal
- Pilot subcarriers embedded in each uplink OFDM blocks
- UEs are distinguished by their random data interleavers





Multi-antenna OFDM grant-free access: sparse traffic flows



Multi-antenna OFDM grant-free access

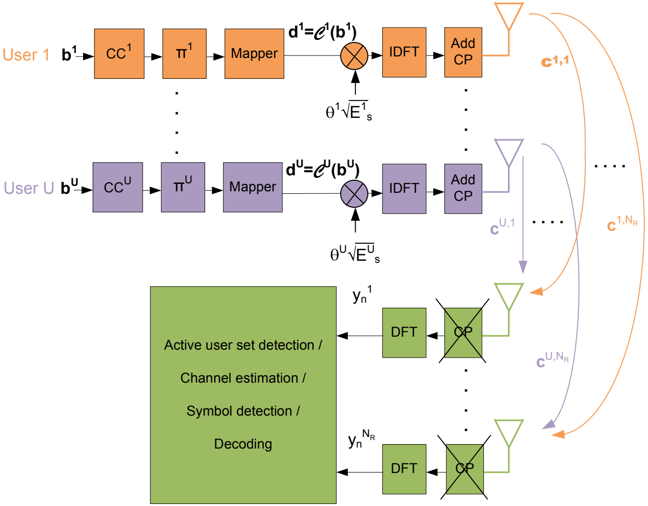
Tx parameters:

- $1 \leq u \leq U$: user index
- $\mathbf{b}^u$: uniformly distributed information bits for user u
- CC^u: convolutional encoder for user u
- π^u: user-specific interleaver for user u
- $\mathbf{d}^u = \mathcal{C}^u(\mathbf{b}^u)$: length-$N$ complex symbol vector for user u
- $\theta^u \in \{0, 1\}$: activity variable for user u
- E_s^u: energy per complex symbol for user u

Channel parameters: N_R Rx antennas

$\mathbf{c}^{(u,m)}$: channel impulse response from user u to Rx antenna m
$\{C_n^{u,m}\}_{n=0}^{N-1} = \mathcal{F}_N(\mathbf{c}^{(u,m)})$: corr. channel frequency response (CFR)

Multi-antenna OFDM grant-free access: System model



Multi-antenna OFDM grant-free access: Modeling

User-u CFR over subcarrier-n: $\mathbf{x}_n^u = \left[C_n^{u,1}, C_n^{u,2} \ldots, C_n^{u,N_R}\right]^T$

Gauss-Markov state-space model: $\mathbf{x}_n^u = \mathbf{x}_{n-1}^u + \mathbf{\Delta}_n^u$
i.i.d. Gaussian driving noise $\mathbf{\Delta}_n^u \sim \mathcal{N}_C(\mathbf{\Delta}_n^u : \mathbf{0}_{N_R \times 1}, \mathbf{Q}^u)$

Observation at Rx antenna m over subcarrier-n:

$$y_n^m = \sum_{u=1}^{U} \theta^u \sqrt{E_s^u} d_n^u C_n^{u,m} + w_n^m,\ w_n^m \sim \mathcal{N}_C(w_n^m : 0, N_0) \text{ i.i.d.}$$

In vector-matrix form: $\mathbf{y}_n = \sum_{u=1}^{U} \theta^u \mathbf{H}_n^u(d_n^u)\mathbf{x}_n^u + \mathbf{w}_n$, where

$\mathbf{y}_n = [y_n^1, \ldots, y_n^{N_R}]^T$, $\mathbf{H}_n^u(d_n^u) = \operatorname{diag}\left\{\left[\sqrt{E_s^u} d_n^u, \ldots, \sqrt{E_s^u} d_n^u\right]\right\}$,
$\mathbf{w}_n = [w_n^1, \ldots, w_n^{N_R}]^T$

Joint posterior distribution and its marginals

Notations

- Collection of hidden variables: $\{x, y_1, y_2, \dots\}$ (circles)
- Consider a factorization of the joint posterior distribution:

$$p(x, y_1, y_2, \dots | \mathbf{o}) = f(x, y_1, y_2 \dots) \prod_j h_j(x, \dots),$$

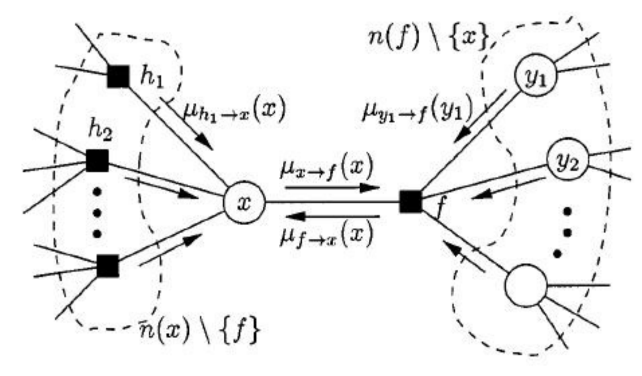
- Local function: depends on subset of hidden variables (square)

Compute the marginal posterior distribution of x

$$p(x|\mathbf{o}) = \sum_{y_1, y_2, \dots} p(x, y_1, y_2, \dots | \mathbf{o})$$

General factor graph representation

Propagate messages along each edge in both directions



Distributed inference on the factor graph

Sum-product algorithm (SPA)

- Variable node to local function node messages:

$$\mu_{x \to f}(x) \propto \prod_{h \in n(x) \setminus \{f\}} \mu_{h \to x}(x)$$

- Local function node to variable node messages:

$$\mu_{f \to x}(x) \propto \sum_{n(f) \setminus \{x\}} \left(f(n(f)) \prod_{y \in n(f) \setminus \{x\}} \mu_{y \to f}(y) \right)$$

- Compute the marginal posterior of x:
$$p(x|\mathbf{o}) \propto \prod_{h \in n(x)} \mu_{h \to x}(x)$$

Multi-antenna OFDM grant-free access: Posterior

Posterior distribution of all hidden variables:

$$p\left(\theta^1,\ldots,\theta^U,\{\mathbf{x}_n^1\}_{n=0}^{N-1},\ldots,\{\mathbf{x}_n^U\}_{n=0}^{N-1},\mathbf{d}^1,\ldots,\mathbf{d}^U,\mathbf{b}^1,\ldots,\mathbf{b}^U|\{\mathbf{y}_n\}_{n=0}^{N-1}\right)$$

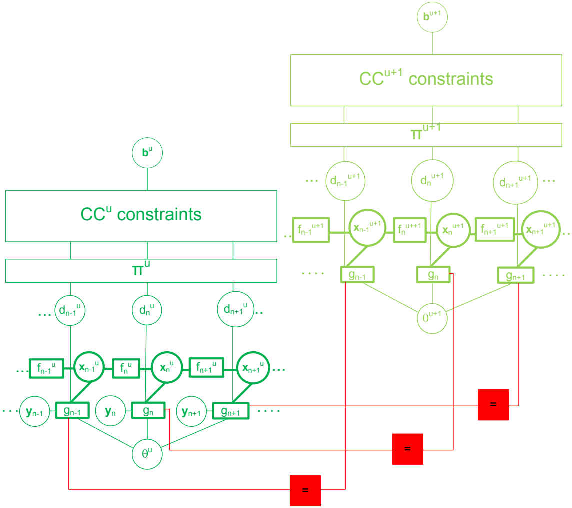
$$\propto p\left(\{\mathbf{y}_n\}_{n=0}^{N-1}|\theta^1,\ldots,\theta^U,\{\mathbf{x}_n^1\}_{n=0}^{N-1},\ldots,\{\mathbf{x}_n^U\}_{n=0}^{N-1},\mathbf{d}^1,\ldots,\mathbf{d}^U\right)$$

$$\times\prod_{u=1}^{U}\left\{P(\theta^u)p\left(\{\mathbf{x}_n^u\}_{n=0}^{N-1}\right)P(\mathbf{d}^u|\mathbf{b}^u)P(\mathbf{b}^u)\right\}$$

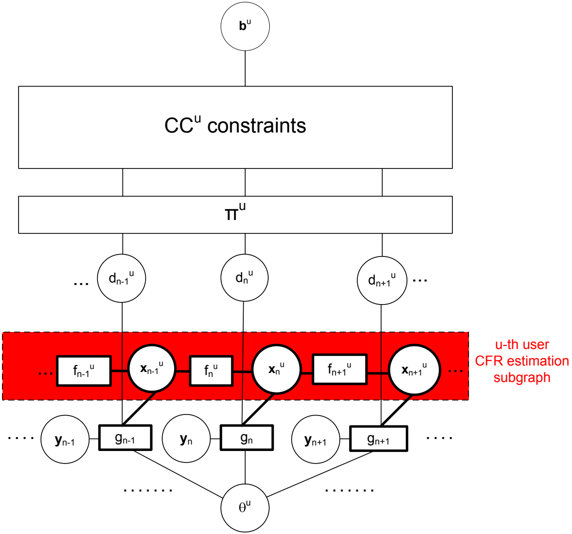
$$\propto\prod_{n=0}^{N-1}p(\mathbf{y}_n|\theta^1,\ldots,\theta^U,\mathbf{x}_n^1,\ldots,\mathbf{x}_n^U,d_n^1,\ldots,d_n^U)$$

$$\times\prod_{u=1}^{U}\left\{P(\theta^u)\quad p(\mathbf{x}_0^u)\prod_{n=1}^{N-1}p(\mathbf{x}_n^u|\mathbf{x}_{n-1}^u)\quad I(\mathbf{d}^u=\mathcal{C}^u(\mathbf{b}^u))\right\}$$

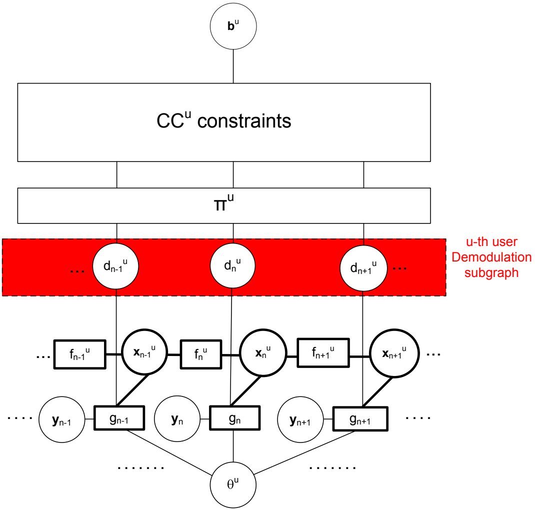
Complete factor graph



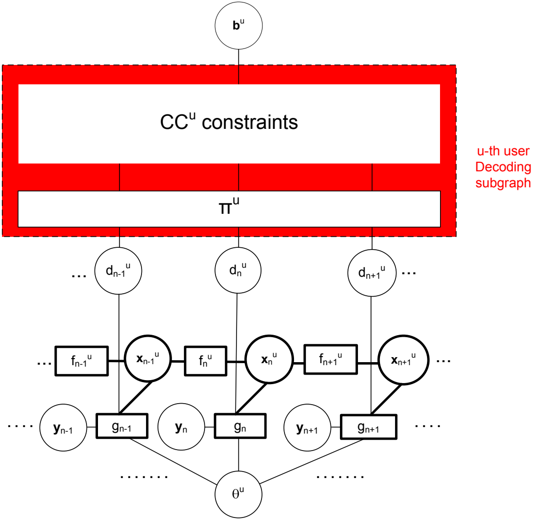
CFR estimation subgraph for user u



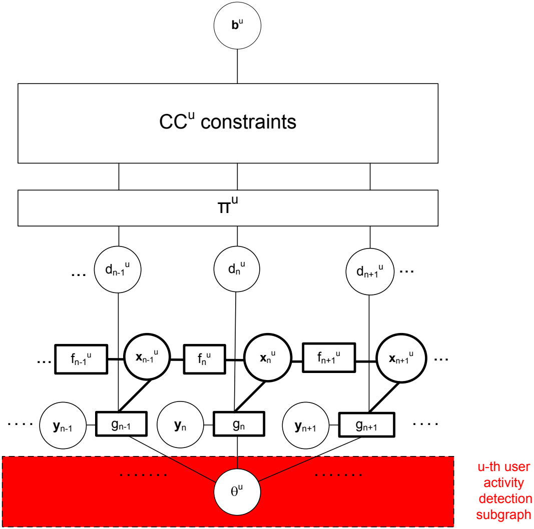
Demodulation subgraph for user u



Decoding subgraph for user u



User activity detection subgraph for user u



Complexity issues

Ordinary message passing:

Complexity order per-subcarrier: exponential in U

Complexity reduction:

Collapse all Gaussian mixture messages to a single Gaussian.
Complexity order becomes

- Linear in U,
- Linear in size of the modulation alphabet
- Linear in the number of receiver iterations
- Cubic in N_R

Simulation example

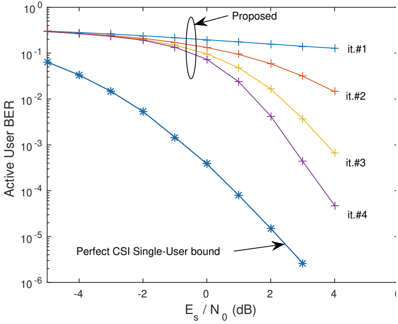
Setup:

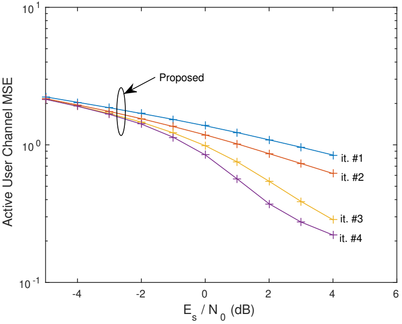
- $(1, 5/7)$ convolutional code, QPSK
- OFDM modulation: $N = 1024$ subcarriers, length-$N/8$ CP
- $(U, N_R) = (16, 4)$
- 4 out of 16 active users
- Channel: block Rayleigh fading with delay spread $=$ 3 samples
- Receiver with 4 message passing iterations

Simulation example:

Missed-detection probability:
below 10^{-3} at $\frac{E_s}{N_0} > -3$ dB

False alarm probability:
below 10^{-3} at $\frac{E_s}{N_0} > 2$ dB





4 Perspectives

Perspectives

Improvements:

- Robustness to higher order modulations?
- Other NOMA schemes: SCMA, IDMA, etc?
- Other detection algorithms: EPA, variational message passing?
- Other applications needing UAD: as cognitive radio, decode-and-forward relaying, etc

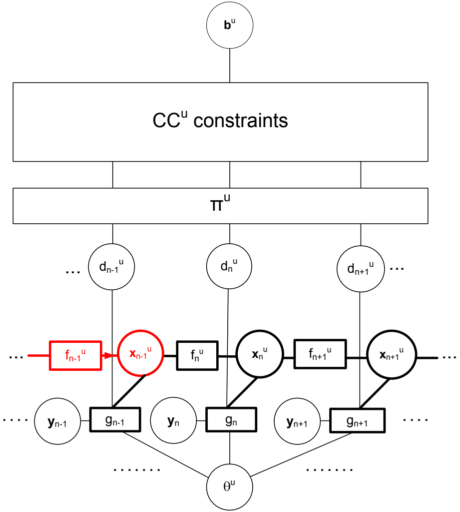
Future work...

- Reduced-order modeling/filtering
- Mixed discrete-continuous filtering

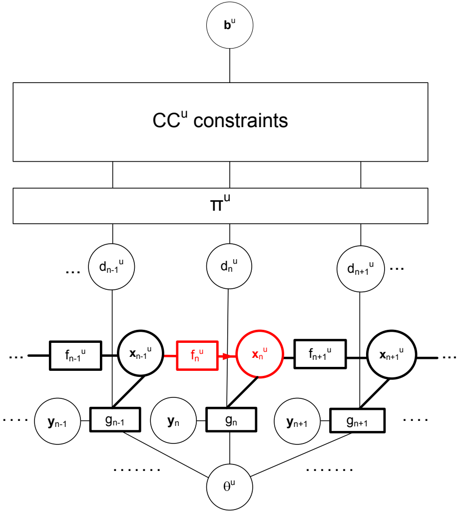
End

Thank you for your attention !

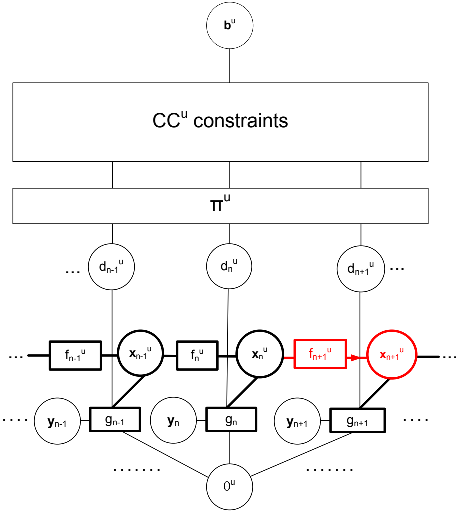
CFR estimation subgraph for user u: Forward pass (1)



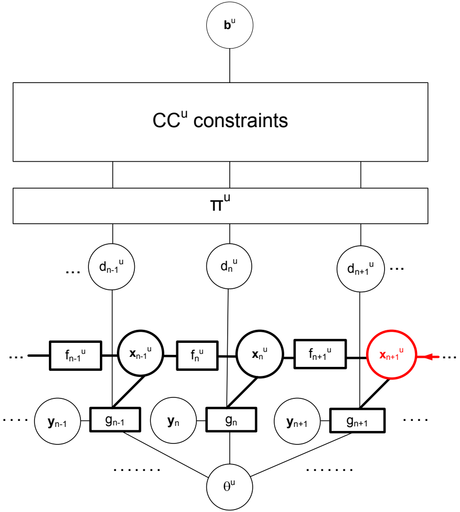
CFR estimation subgraph for user u: Forward pass (2)



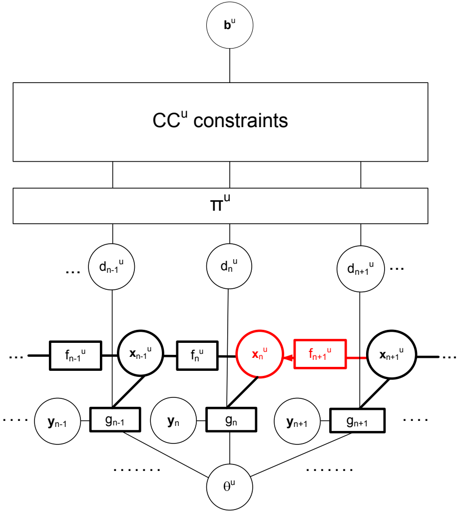
CFR estimation subgraph for user u: Forward pass (3)



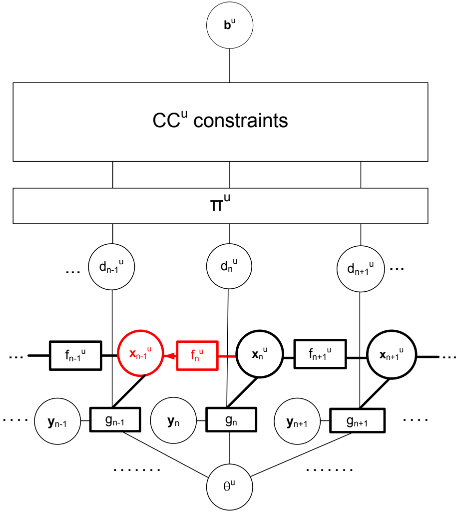
CFR estimation subgraph for user u: Backward pass (1)



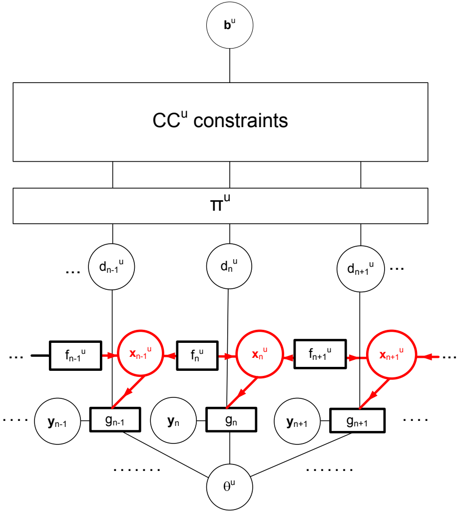
CFR estimation subgraph for user u: Backward pass (2)



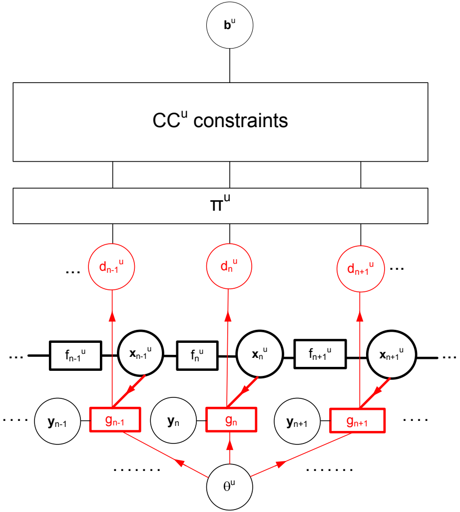
CFR estimation subgraph for user u: Backward pass (3)



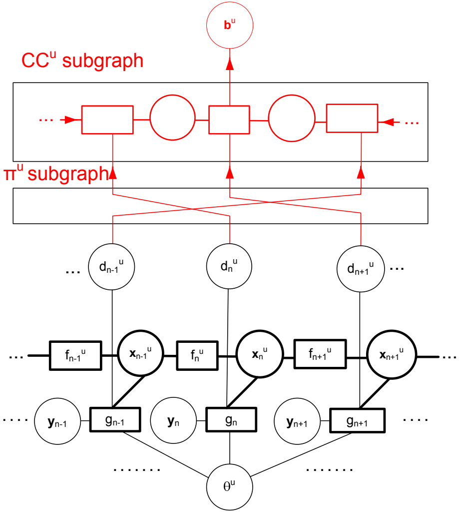
CFR estimation subgraph for user u: Smoothing pass



Demodulation subgraph for user u: Soft Demodulation



Decoding subgraph for user u: Soft Decoding



User activity detection subgraph for user u: Soft UAD