On learning CSI for 6G

Jean-Marie Gorce

INRIA-LYS / INSA-Lyon / CITI lab

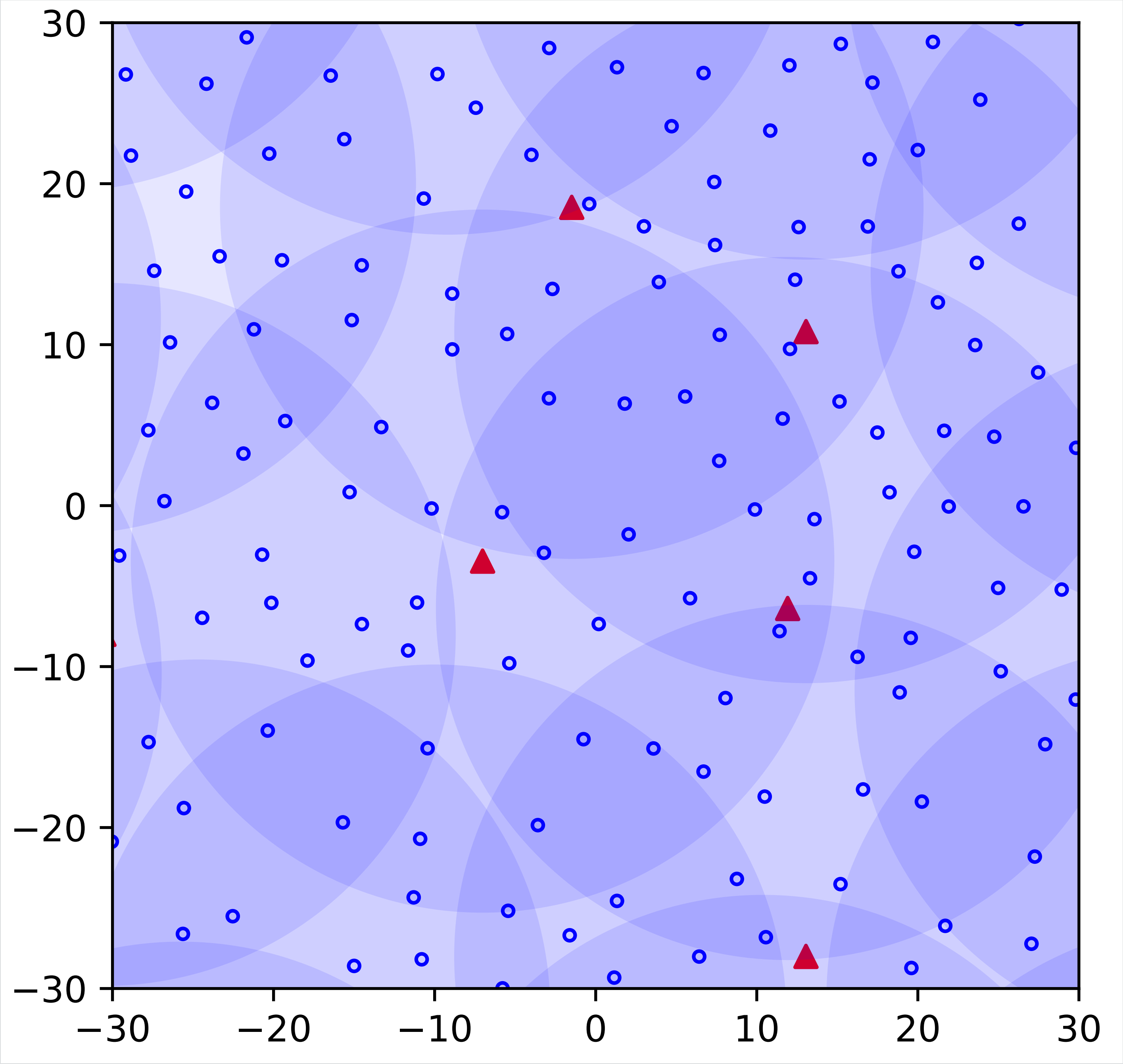
Outline

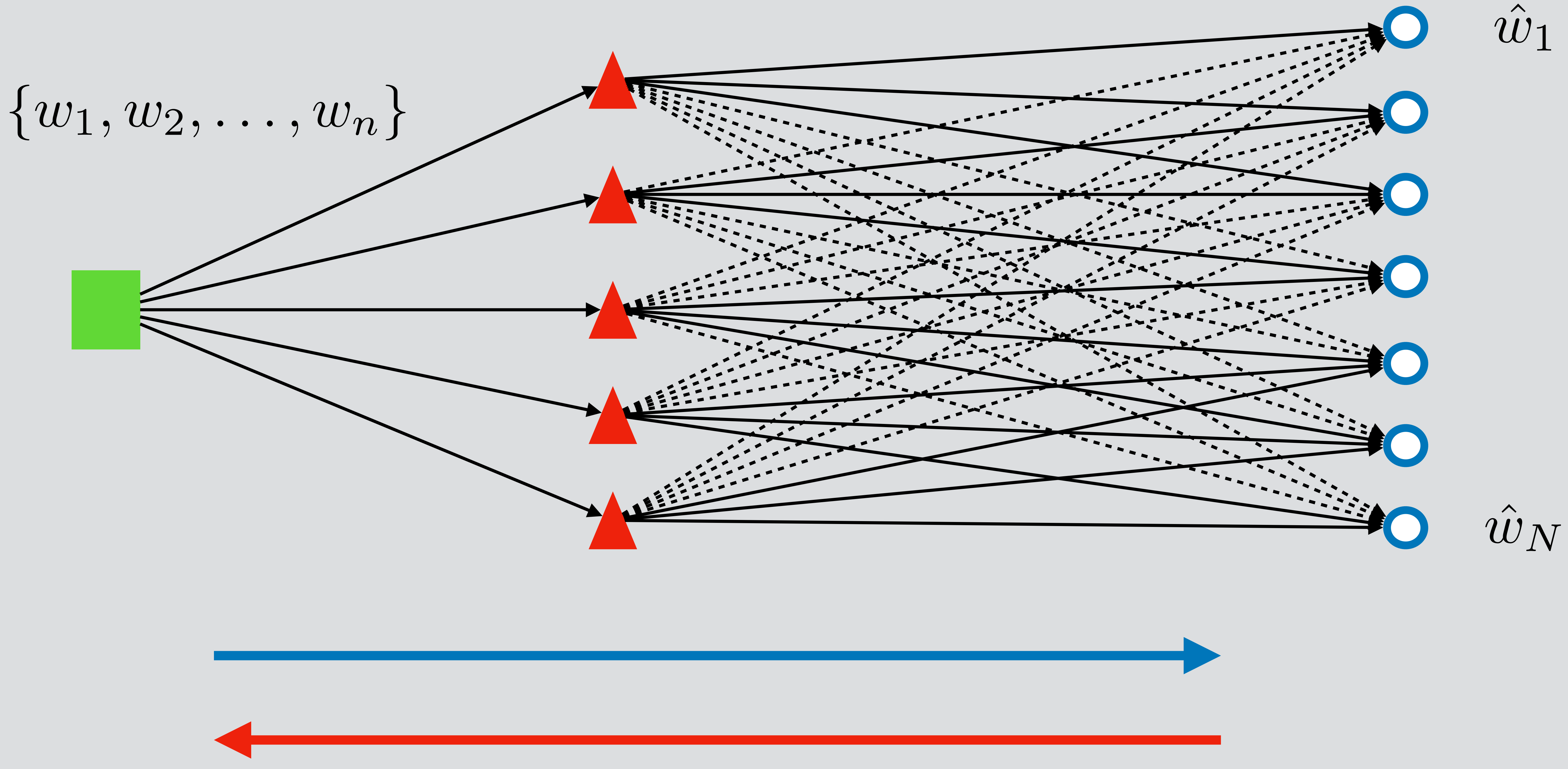
CSI=Channel State Information

- 1- Context and background: CSI needs are growing very fast in large scale cell-free systems.
- Case studies :
  - 2- CSI-R for massive random access - a compressed sensing problem.
  - 3- Non-explicit or ‘embedded’ CSI, learning based MU-MIMO systems.
  - 4- Federated learning for CSI-T estimation for the MU-MIMO BC : a preliminary study

1- Why are the needs for CSI increasing?

A cell-free massive access RAN can be defined as a large dimension linear system :





$$Y = H \cdot X(W) + N$$

$$\hat{W} = \phi_{dec}(Y)$$

Why is CSI going to be so important ?

$$Y = H \cdot X(W) + N$$

- **Exponential grows of CSI values :**
    - (Nb of Txs x Nb of antennas)x(Nb of Rxs x Nb of antennas)x(Nb of ressources)x(Nb of measures per time unit)
    - Higher frequencies, higher bandwidth
    - New paradigm with cell-free and IRS, joint communication and sensing : toward a distributed but controlled system.
- **Non asymptotic studies :**
    - Small information quantities (IoT, control) : Capacity/Rates cannot capture the problem well.
    - Delay/reliability with heterogeneous constraints —> static rate is not enough !
    - Transmission conditions change continuously : mobility, transmission policies, interference

Imperfect channel knowledge (at transmitter or at receiver)

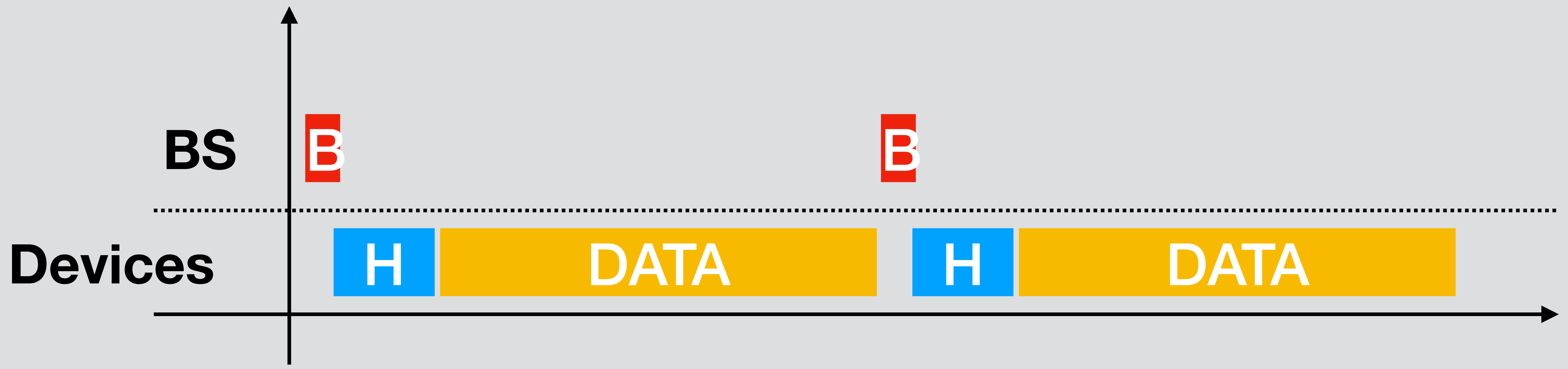
Trade-off : signaling overload versus accuracy

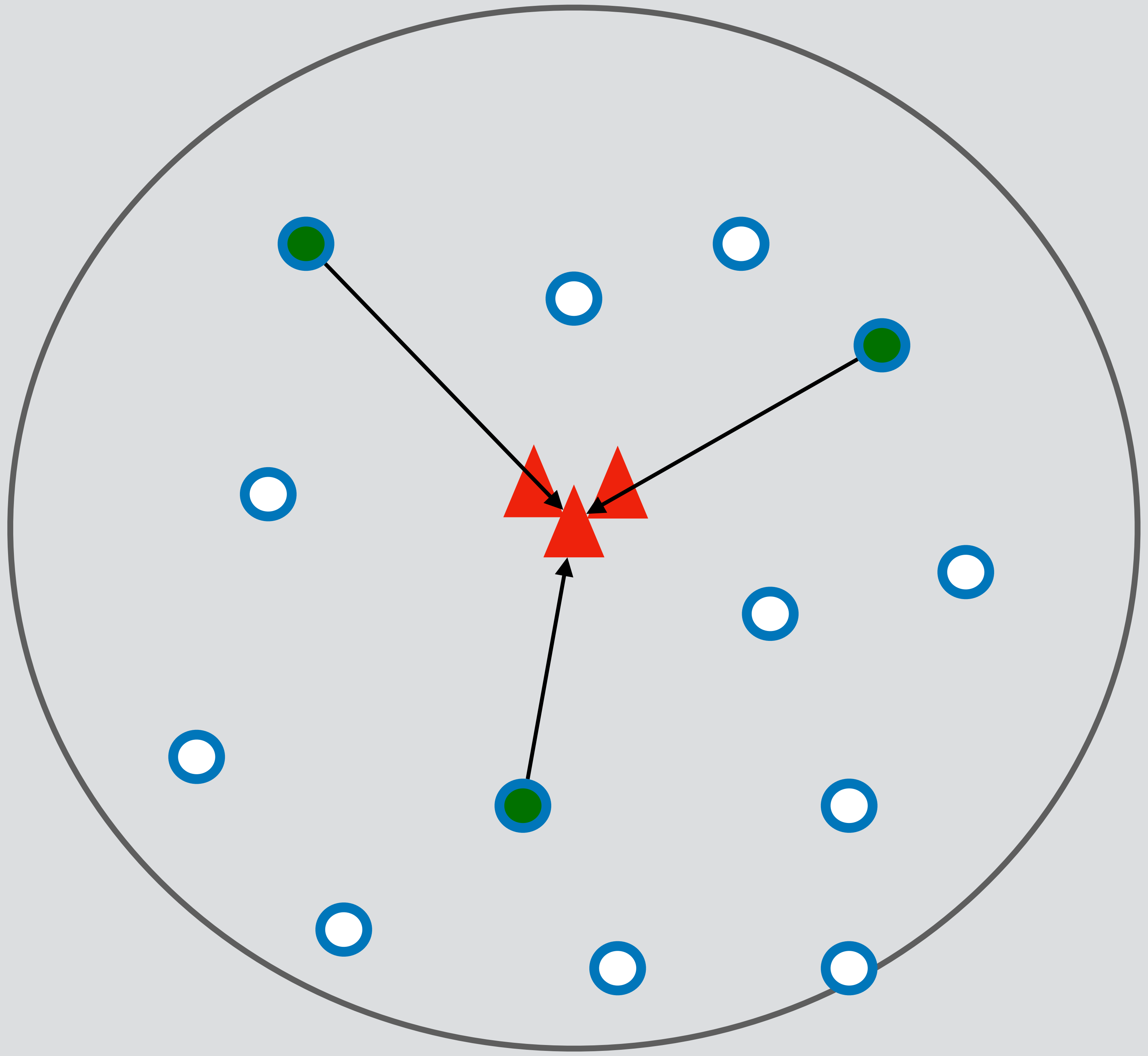
- **Tools/methods**
  - Theory : Combination of information theory, coding theory and queuing theory with control theory.
  - Algorithms : Leveraging on data based or machine learning to improve the exploitation of CSI
- **Some research directions**
  - Estimation of channels in Multi-user settings with a reduced number of pilots (ex.1)
  - Pilotless, embedded pilots —> blind techniques (ex. 2)
  - Compress and quantize CSI, « cooperative learning » (ex. 3)

2- Massive Grant Free Random Access

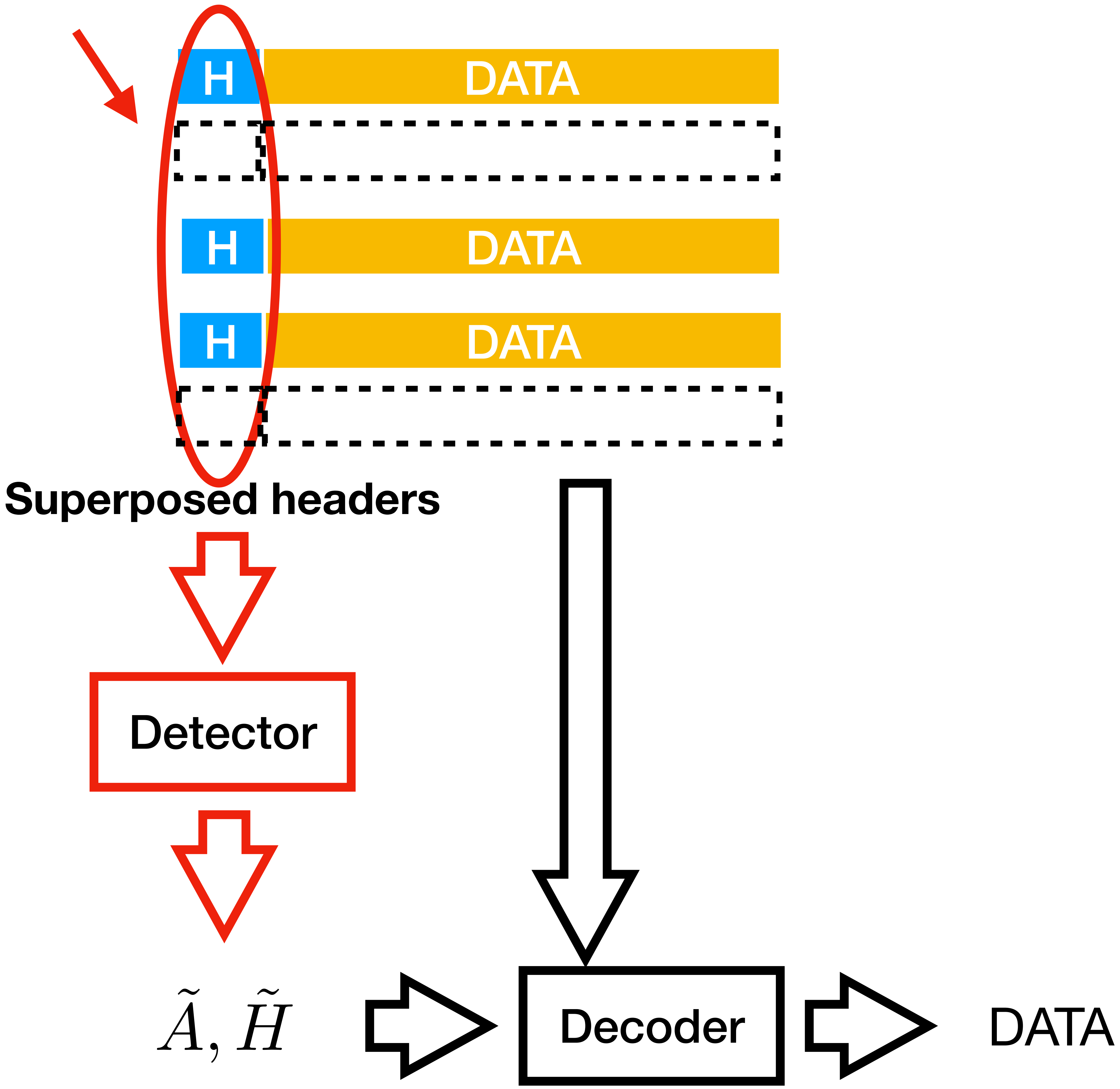
Joint work with <u>Lélio Chetot</u> and Malcolm Egan

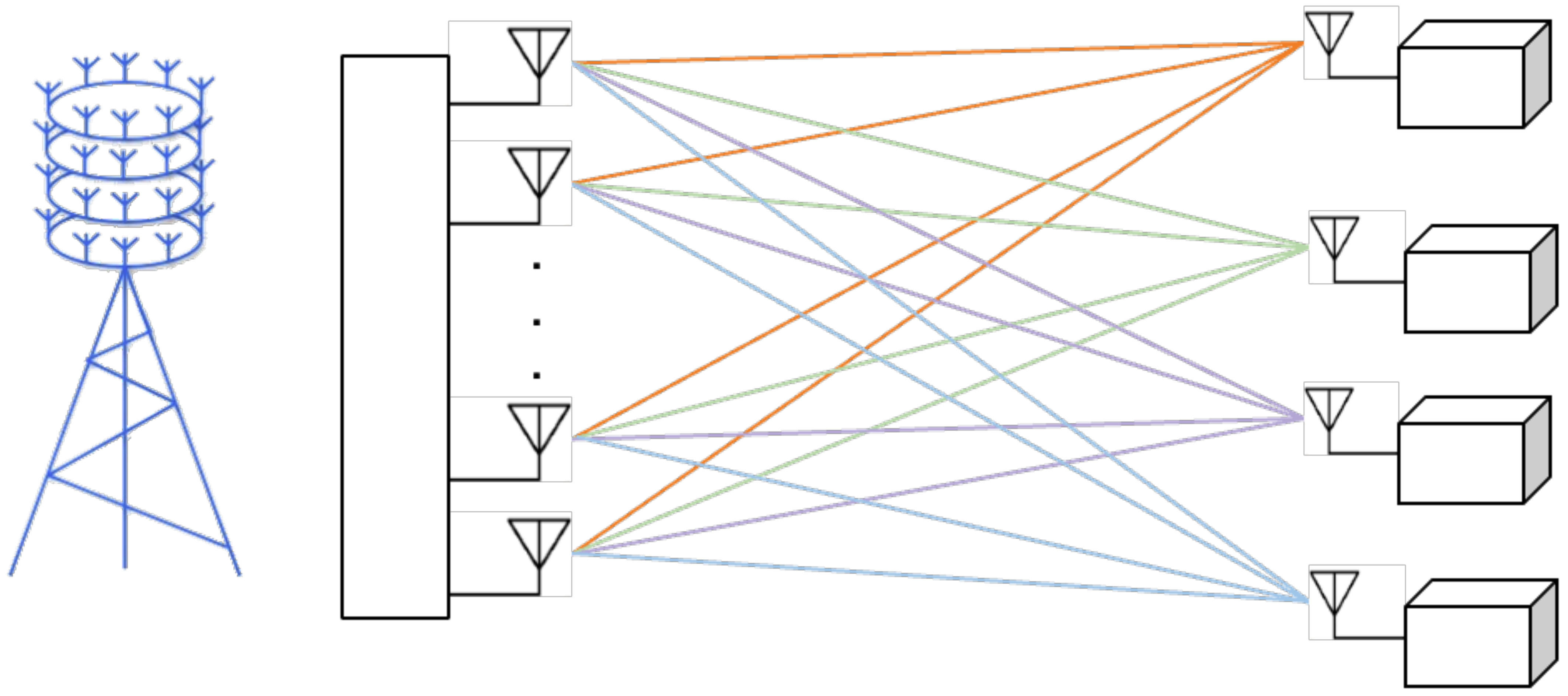
- Scenario : grant-free random access
    - 1 BS with N-antennas
    - Large number of M devices w. 1-antenna,
    - Low individual transmission probability
    - Slotted GF random access
    - Unknown channels, unknown activity

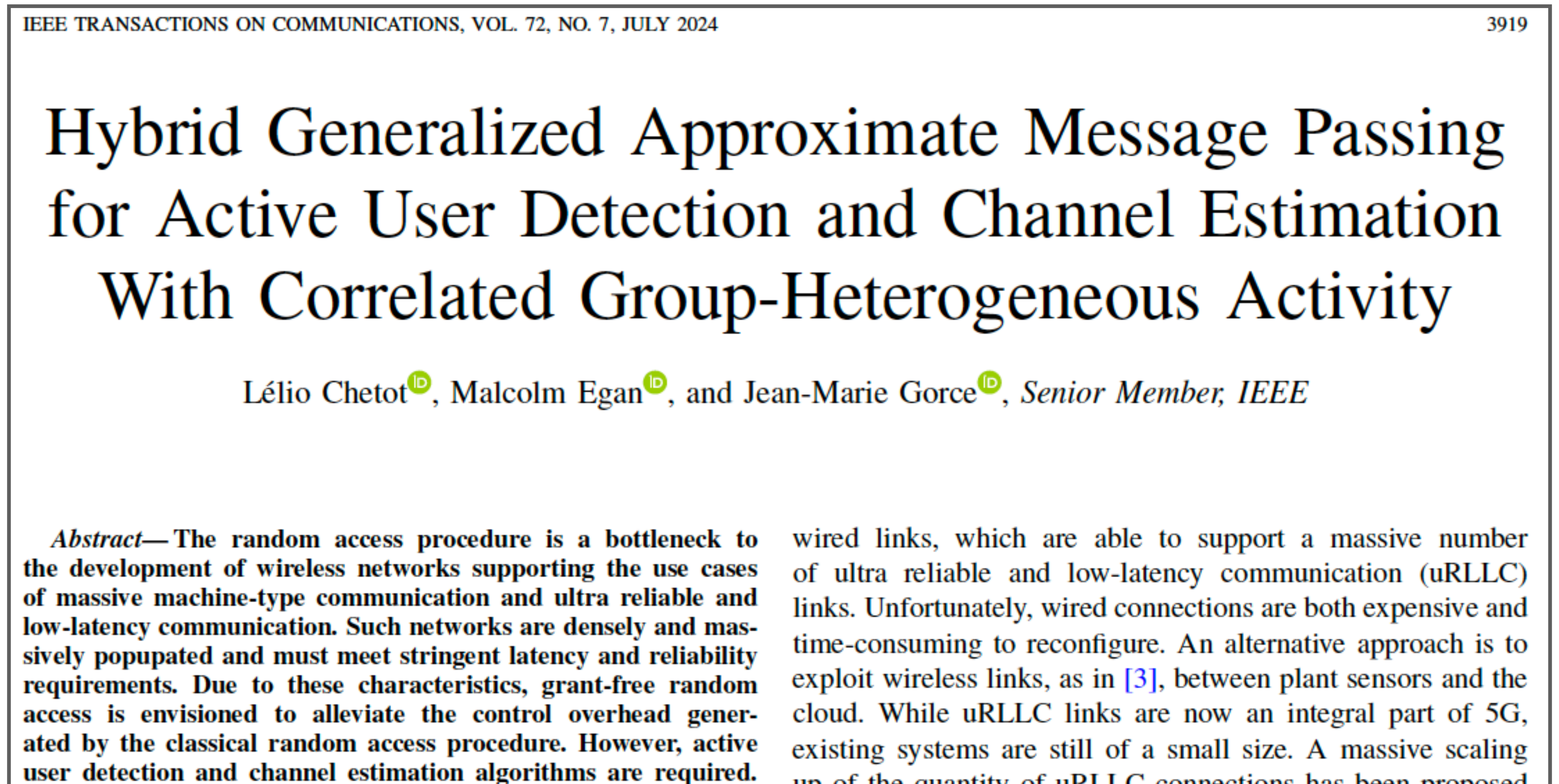




- AUDACE : Active User Detection and Channel Estimation






IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 72, NO. 7, JULY 2024 3919

Hybrid Generalized Approximate Message Passing for Active User Detection and Channel Estimation With Correlated Group-Heterogeneous Activity

Lélio Chetot, Malcolm Egan, and Jean-Marie Gorce, *Senior Member, IEEE*

***Abstract*—The random access procedure is a bottleneck to the development of wireless networks supporting the use cases of massive machine-type communication and ultra reliable and low-latency communication. Such networks are densely and massively popupated and must meet stringent latency and reliability requirements. Due to these characteristics, grant-free random access is envisioned to alleviate the control overhead generated by the classical random access procedure. However, active user detection and channel estimation algorithms are required.**

wired links, which are able to support a massive number of ultra reliable and low-latency communication (uRLLC) links. Unfortunately, wired connections are both expensive and time-consuming to reconfigure. An alternative approach is to exploit wireless links, as in [3], between plant sensors and the cloud. While uRLLC links are now an integral part of 5G, existing systems are still of a small size. A massive scaling

A compressed sensing (CS) problem

Given a BS with K antennas, its coverage area:
how many nodes (N) can be handled or how many resources (M) are required to solve AUDaCE ?

A compressed sensing (CS) problem

Given a BS with K antennas, its coverage area:
how many nodes (N) can be handled or how many resources (M) are required to solve AUDaCE ?

- Unknown variables

$$\mathsf{s}_n = \begin{cases} 0 \Leftrightarrow \text{UE } n \text{ is inactive} \\ 1 \Leftrightarrow \text{UE } n \text{ is active} \end{cases}$$

$$\mathbb{P}_{\mathsf{s}}(s) = \prod_{n=1}^{N} q_n^{s_n}(1-q_n)^{1-s_n}$$

$$\mathbf{H} = [\mathbf{h}_k]_{k\in[\![1,K]\!]} \in \mathbb{C}^{N\times K}$$

$$f_{\mathsf{H}}(\boldsymbol{H}) = \prod_{n=1}^{N}\prod_{k=1}^{K}\Big((1-q_n)\delta(h_{nk}) + q_n\mathcal{CN}(h_{nk};\mu_{\mathsf{h}},\tau_{\mathsf{h}})\Big)$$

- Transmission system

Each device uses a pseudo-random code. Codebook $\mathbf{X} = [\mathbf{x}_n]_{n\in[\![1,N]\!]} \in \mathbb{C}^{M\times N}$

Then the observed signal is

$$Y = XH + W$$

Where H contains a small set of non null vectors

- A joint estimation-detection problem

Detection **Estimation**

$$\hat{\boldsymbol{s}} = \underset{\boldsymbol{s}\in\{0,1\}^N}{\arg\max}\, \mathbb{P}_{\mathsf{s}|\mathbf{Y}}(\boldsymbol{s} \mid \boldsymbol{Y}) \text{ and } \hat{\boldsymbol{H}} = \mathbb{E}[\boldsymbol{H} \mid \boldsymbol{Y}]$$

$$\mathbb{P}_{\mathsf{s}|\mathbf{Y}}(\boldsymbol{s} \mid \boldsymbol{Y}) = \frac{\int_{\mathbb{C}^{N\times K}} f_{\mathsf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s},\boldsymbol{H},\boldsymbol{Y})\mathrm{d}\boldsymbol{H}}{\mathcal{Z}(\boldsymbol{Y})}$$

$$f_{\mathbf{H}|\mathbf{Y}}(\boldsymbol{H} \mid \boldsymbol{Y}) = \frac{\sum_{\boldsymbol{s}\in\{0,1\}^N} f_{\mathsf{s},\mathbf{H},\mathbf{Y}}(\boldsymbol{s},\boldsymbol{H},\boldsymbol{Y})}{\mathcal{Z}(\boldsymbol{Y})}$$

Two families of algorithms :

1: Iterative algorithm with regularization : OMP, LASSO,…

2: Bayesian estimation with message passing (inc. BP, AMP, EP, GAMP)

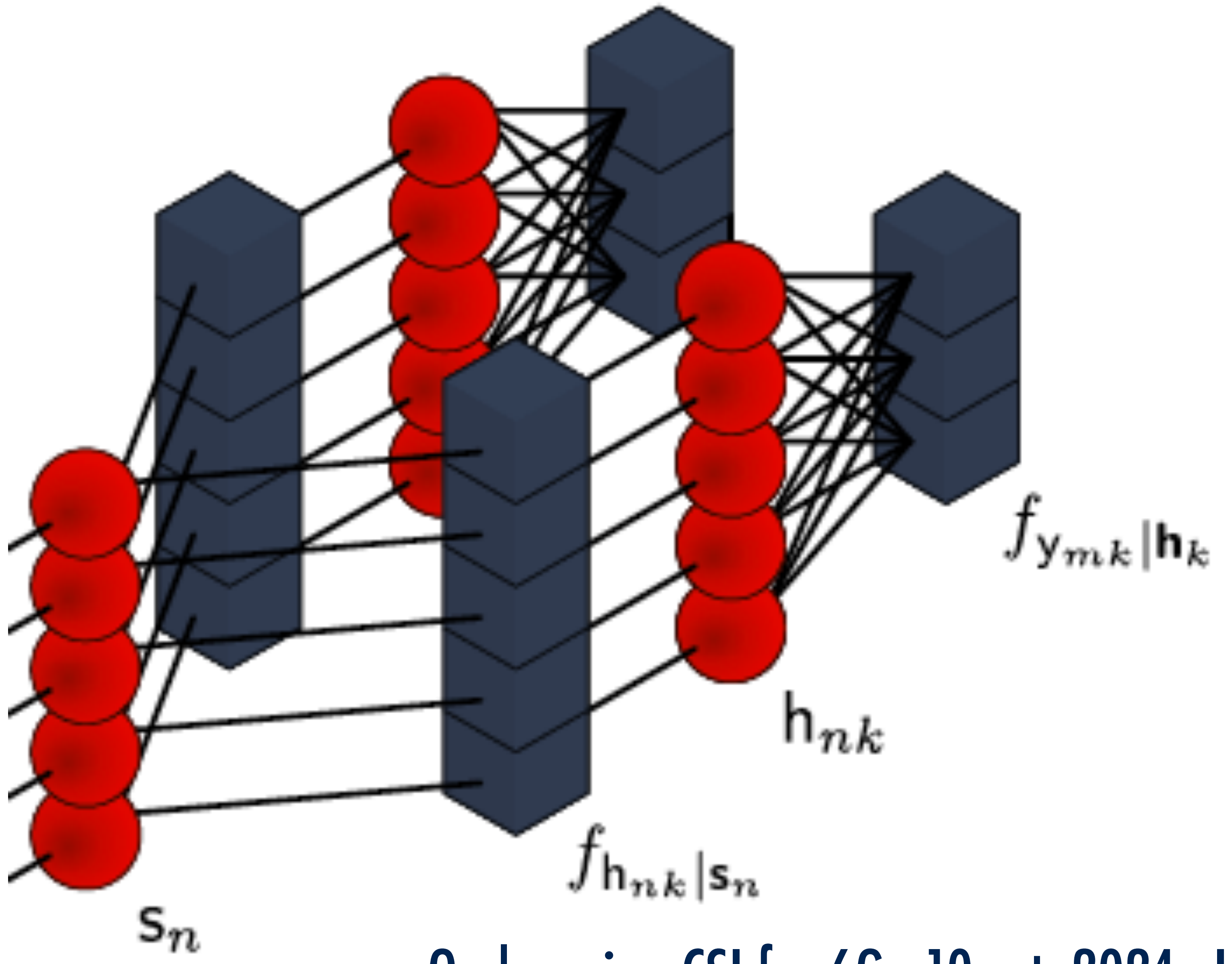
Markov chain $\mathsf{s} \rightarrow \mathbf{H} \rightarrow \mathbf{Y}$

$$f_{\mathbf{H},\mathsf{s},\mathsf{q}|\mathbf{Y}}(\boldsymbol{H},\boldsymbol{s} \quad \mid \boldsymbol{Y}) = f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y} \mid \boldsymbol{H}) f_{\mathbf{H}|\mathsf{s}}(\boldsymbol{H} \mid \boldsymbol{s})\, \mathbb{P}_{\mathsf{s}}(\boldsymbol{s})$$

$$f_{\mathbf{Y}|\mathbf{H}}(\boldsymbol{Y} \mid \boldsymbol{H}) = \prod_{k=1}^{K}\prod_{m=1}^{M} f_{\mathsf{y}_{mk}|\mathbf{h}_k}(y_{mk} \mid \boldsymbol{h}_k)$$

$$f_{\mathbf{H}|\mathsf{s}}(\boldsymbol{H} \mid \boldsymbol{s}) = \prod_{k=1}^{K}\prod_{n=1}^{N} f_{\mathsf{h}_{nk}|\mathsf{s}_n}(h_{nk} \mid s_n)$$

Graph used for message passing



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," IEEE Signal Process. Mag.,Sep. 2018
M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing based adaptive active user detection and channel estimation: Massive access meets massive MIMO," IEEE Trans. Signal Process. 2020
Q. Zou, H. Zhang, D. Cai, and H. Yang, "Message passing based joint channel and user activity estimation for uplink grant-free massive MIMO systems with low-precision ADCs," IEEE Signal Process. Lett. 2020

- Introducing dependencies in the data model

Correlated group-heterogenous activity with Copula

Something between independent and group

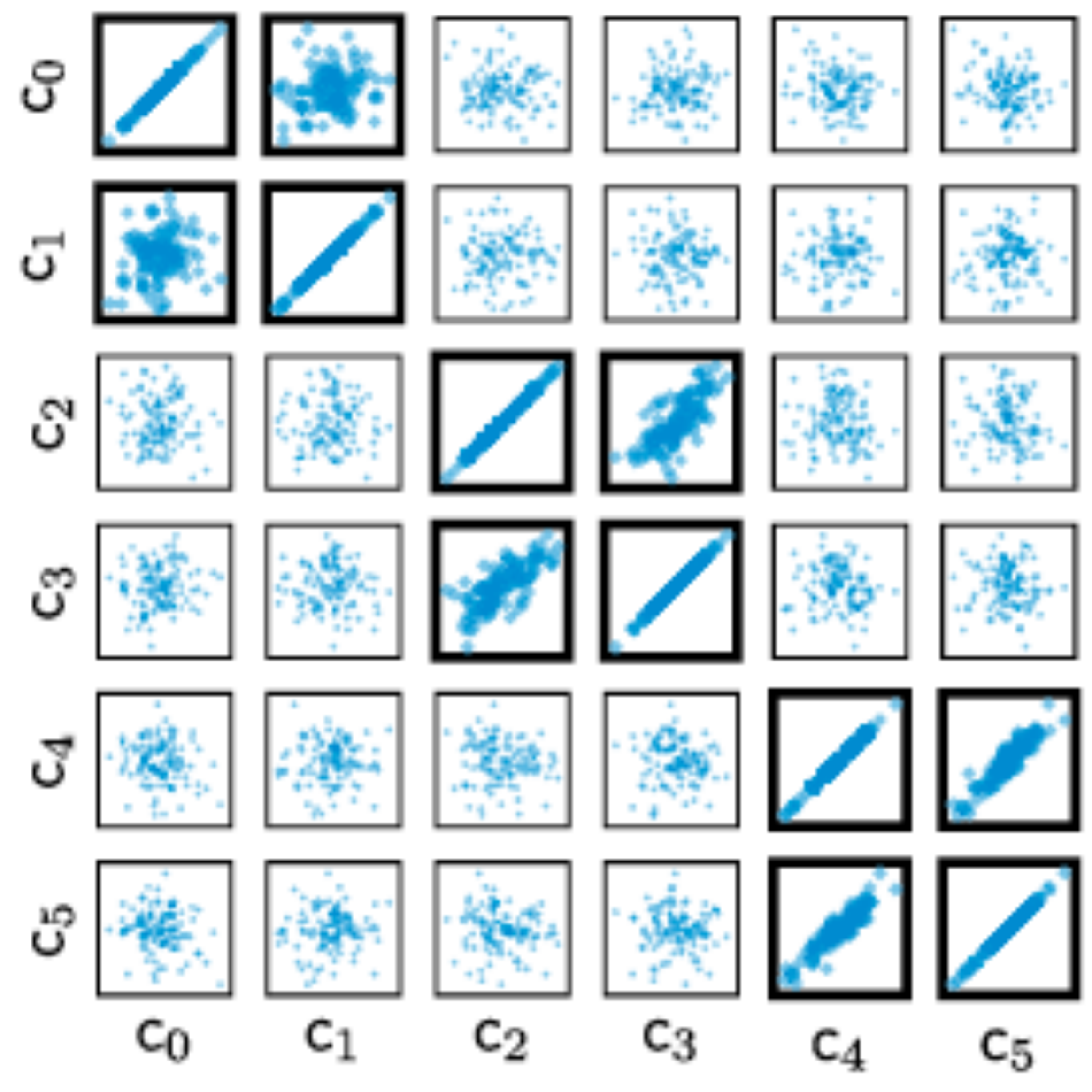
$$\mathbf{c} \rightarrow \mathbf{q} \rightarrow \mathbf{s}.$$

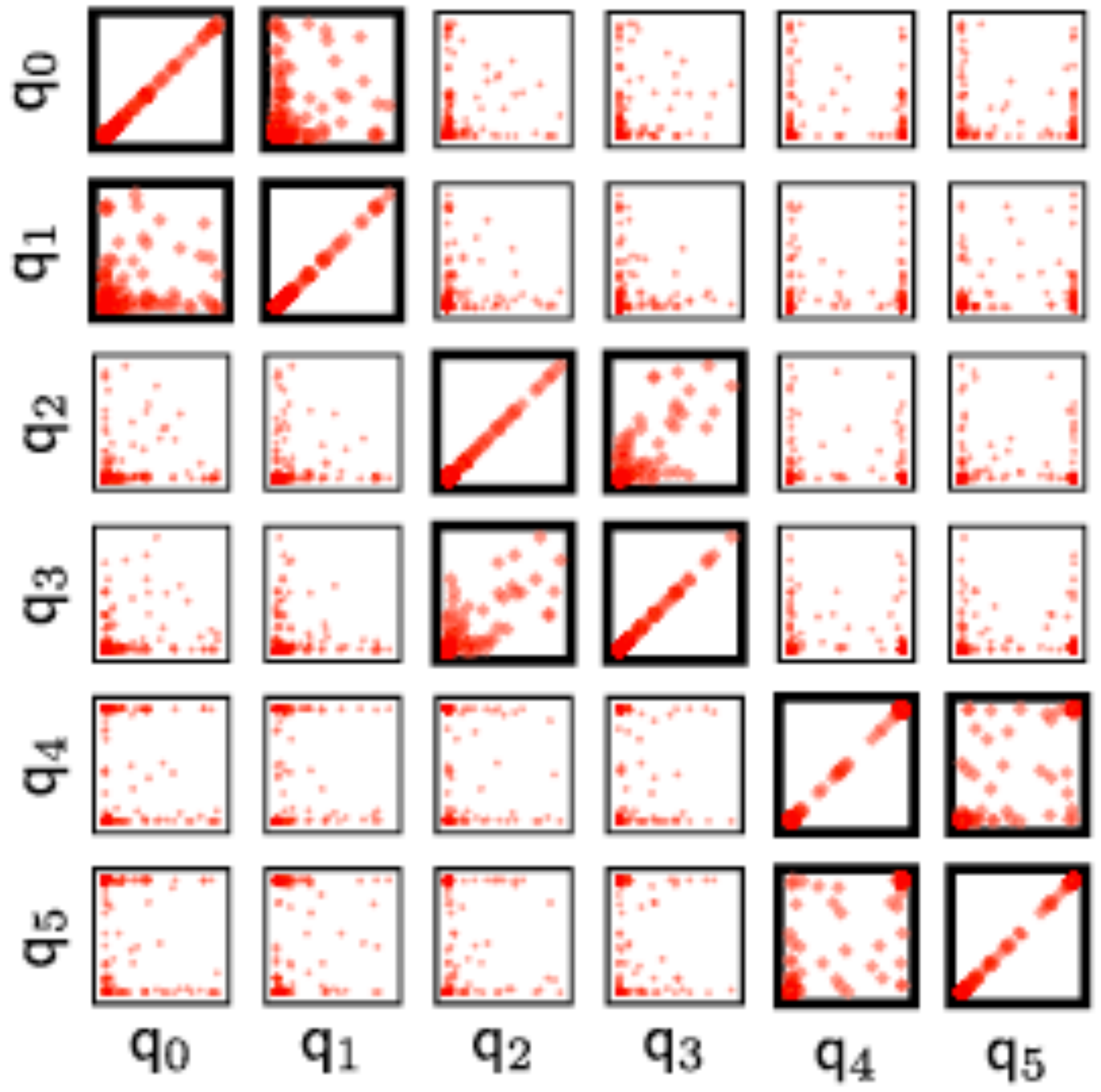
$$\mathbf{c} \sim \mathrm{Norm}(\mathbf{0}_{N,1}, \boldsymbol{K}_{\mathbf{c}})$$

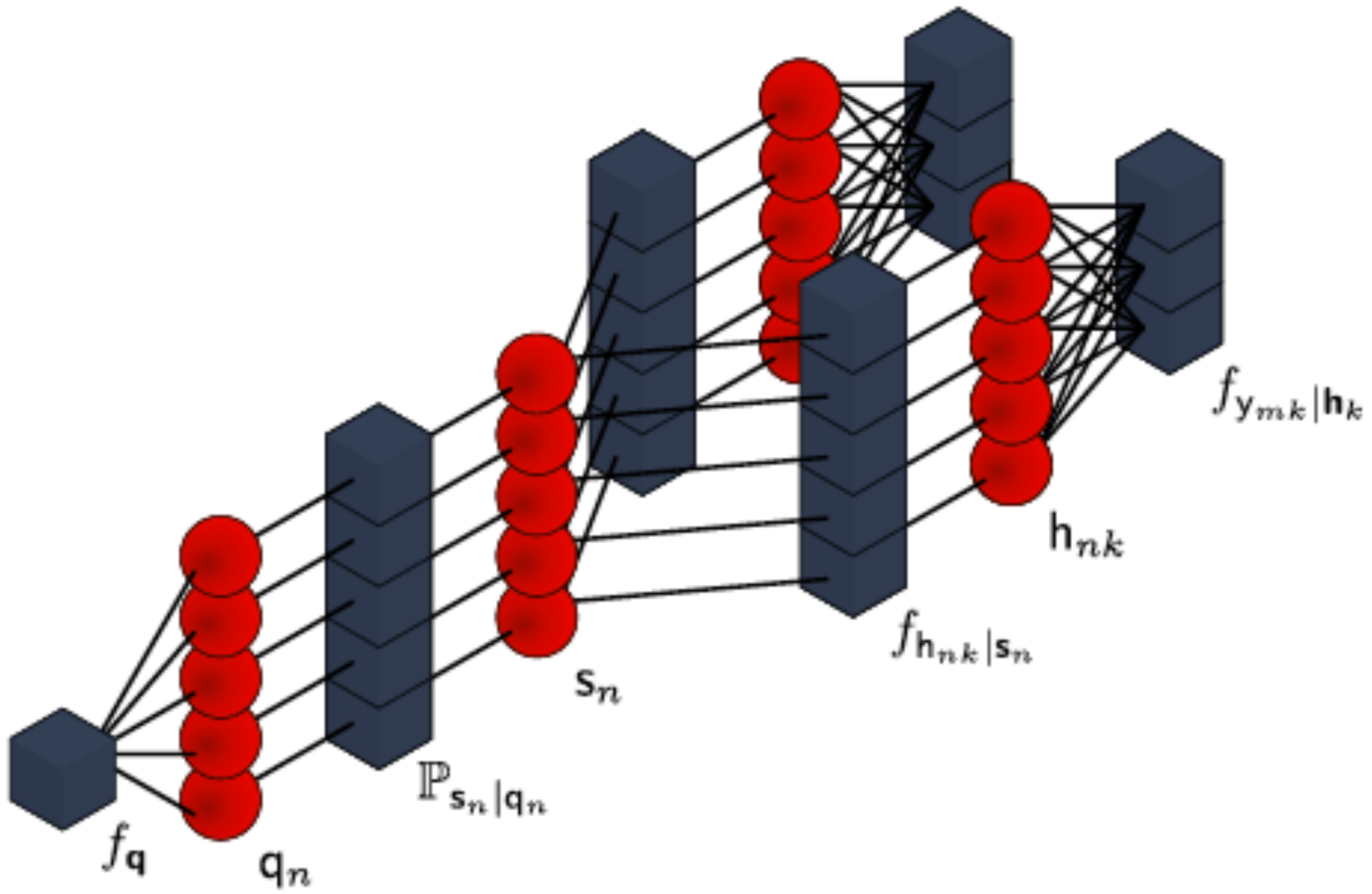
$$\boldsymbol{K}_{\mathbf{c}} = \begin{bmatrix} \boldsymbol{K}_{\mathbf{c},1} & & \\ & \ddots & \\ & & \boldsymbol{K}_{\mathbf{c},G} \end{bmatrix} \in \mathbb{R}^{N \times N}$$

$$\mathbf{q} = \left[(F_{\mathsf{q}_n}^{-1} \circ F_{\mathsf{c}_n})(\mathsf{c}_n) \right]^{\mathsf{T}}_{n \in [\![1,N]\!]}$$

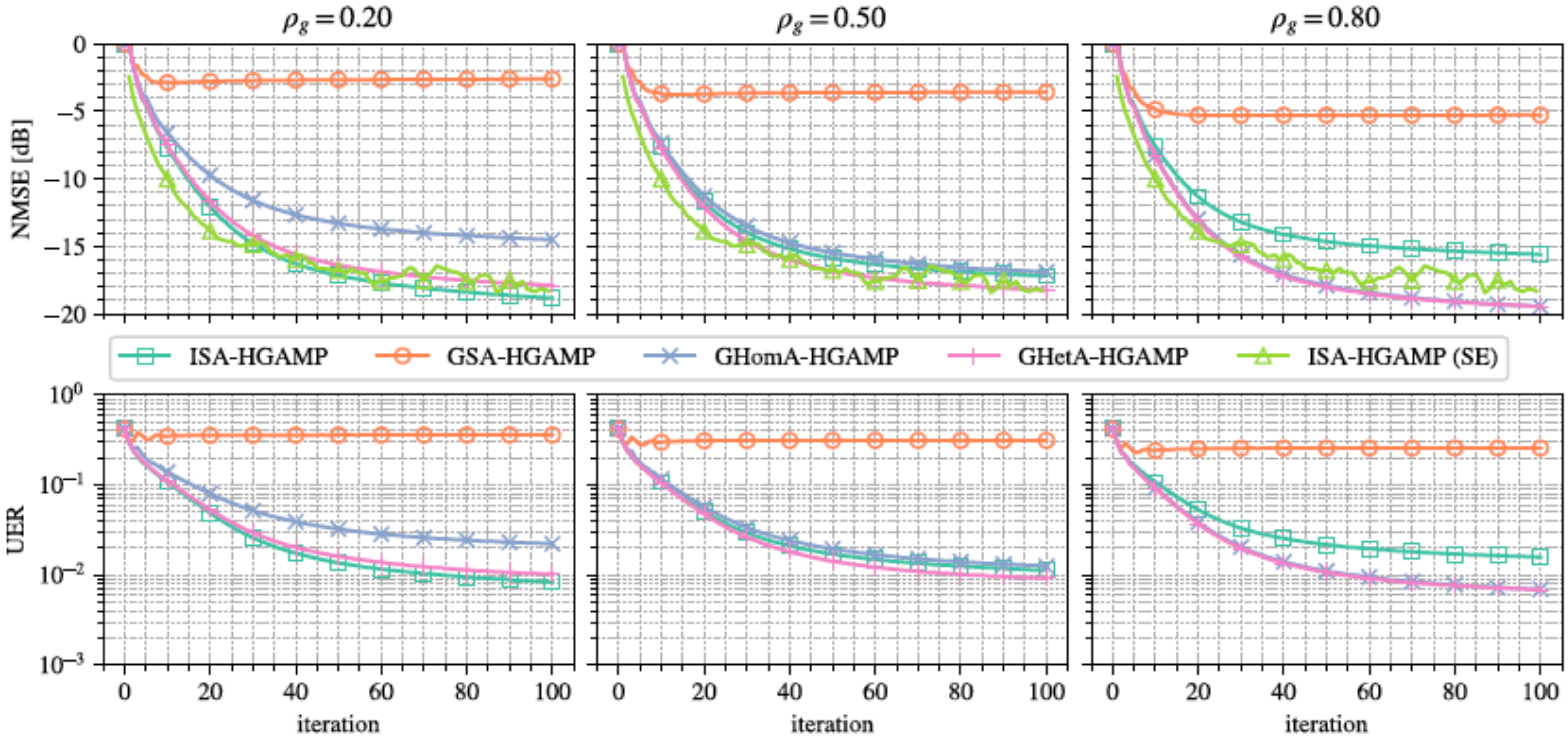
$$\mathsf{q}_n \sim \mathrm{Beta}(\alpha_n, \beta_n)$$







- Results with perfect model knowledge



SE = state evolution

Research trends

- Use of non orthogonal « pilots » to fasten the detection. Codebooks: optimal codes for detection and taking into account decoding complexity (semi-structured codes, block-codes, …): codebook size, minimal distance, decoding complexity : Grassmanian, theory of frames, …

- Taking data information (dependencies and prior knowledge on transmission probabilities, time and space) into account, may help to improve the decoding itself. Future: learning this information from the observations.

- Additional topics : improve transmission strategies (multi-carriers, power transmission) to reduce mis-detections, false alarms.

3- Learning aided MU MIMO

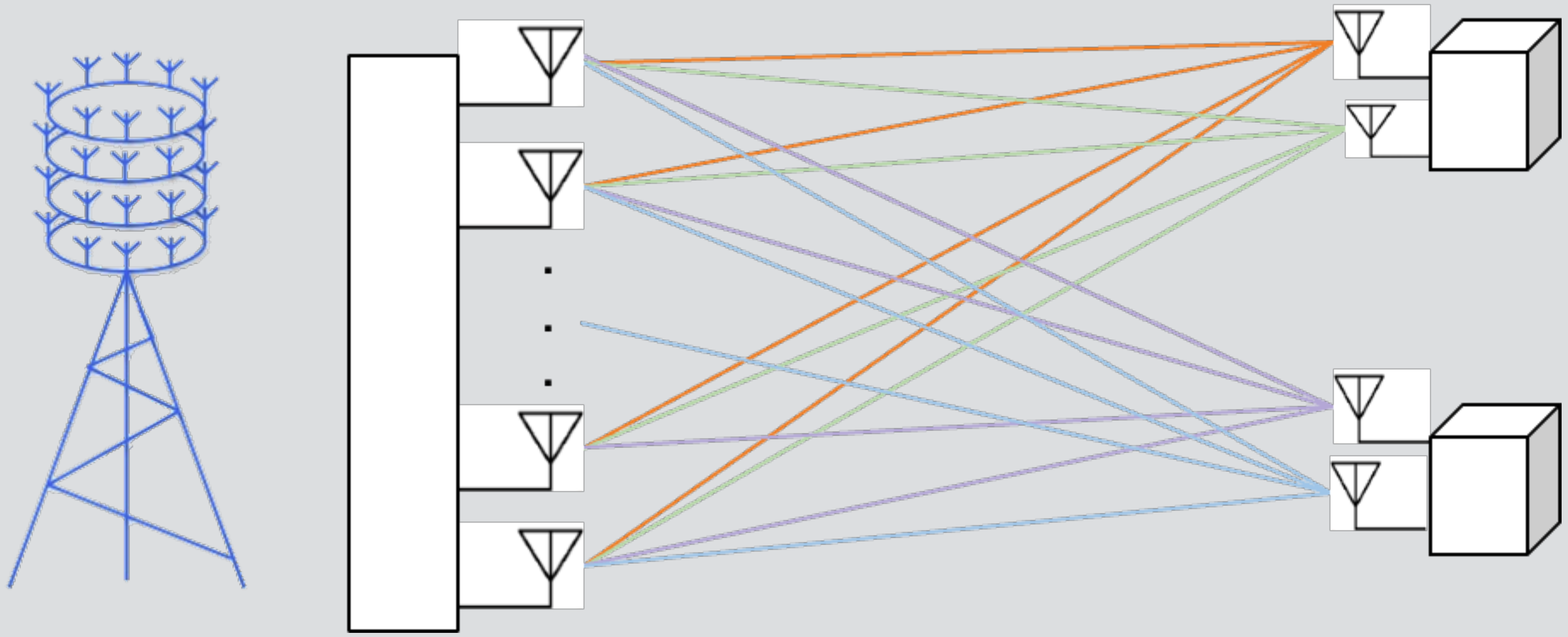
Joint work with Mathieu Goutay, Jakob Hoydis, Faycal Ait Aoudia

- Objectif : show learning based techniques allowing to reduce CSI impact on the transmission
    - Scenario 1 : Learning OFDM Waveforms with PAPR and ACLR Constraints

  Design an OFDM based transmission without explicit pilots

    - Scenario 2 : ML-enhanced Receive Processing for MU-MIMO

  Design a robust MU-MIMO receiver robust to CSI errors.

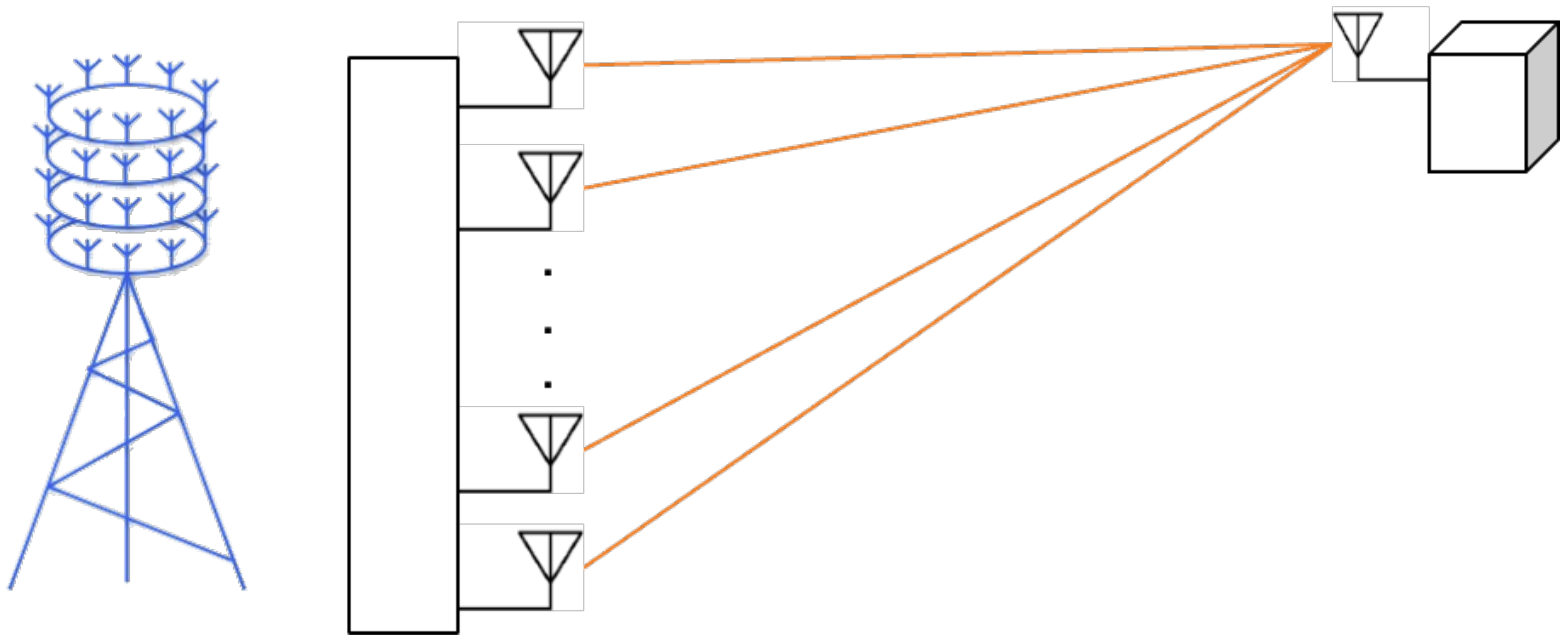


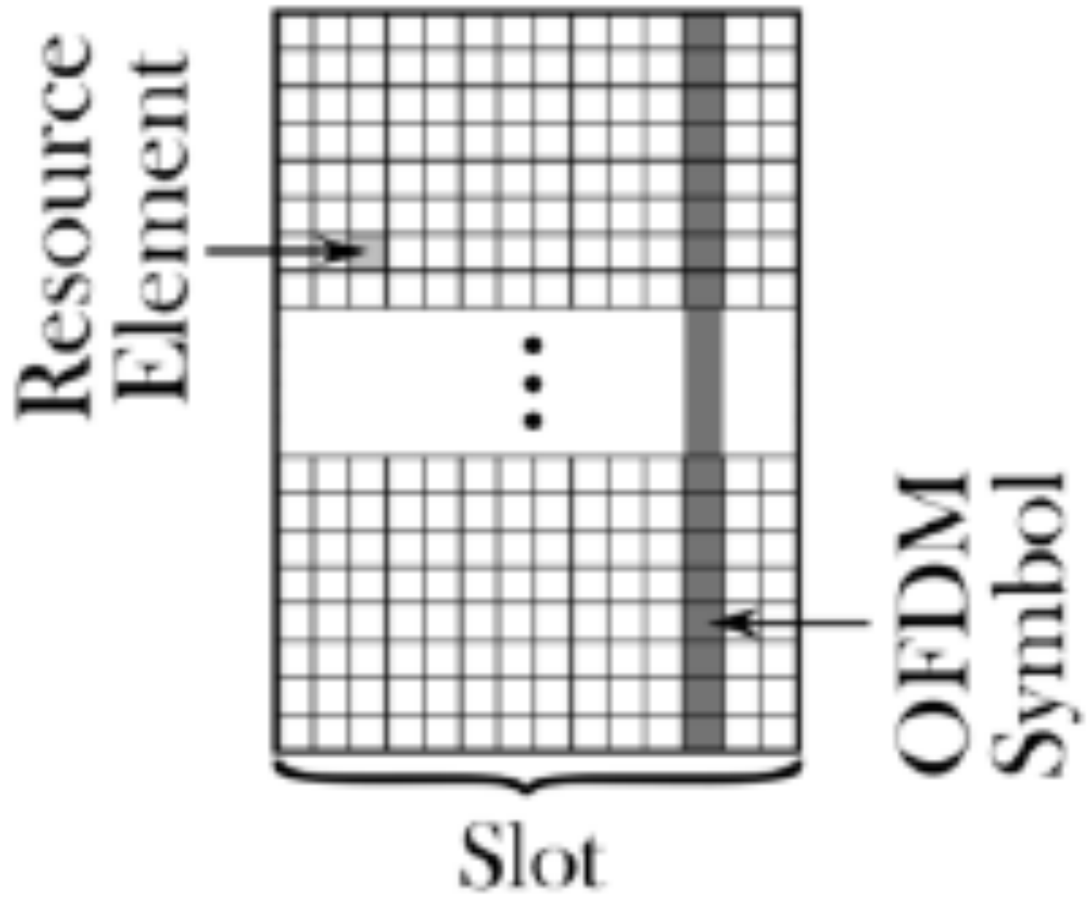
Scenario 1 : Learning OFDM Waveforms with PAPR and ACLR Constraints

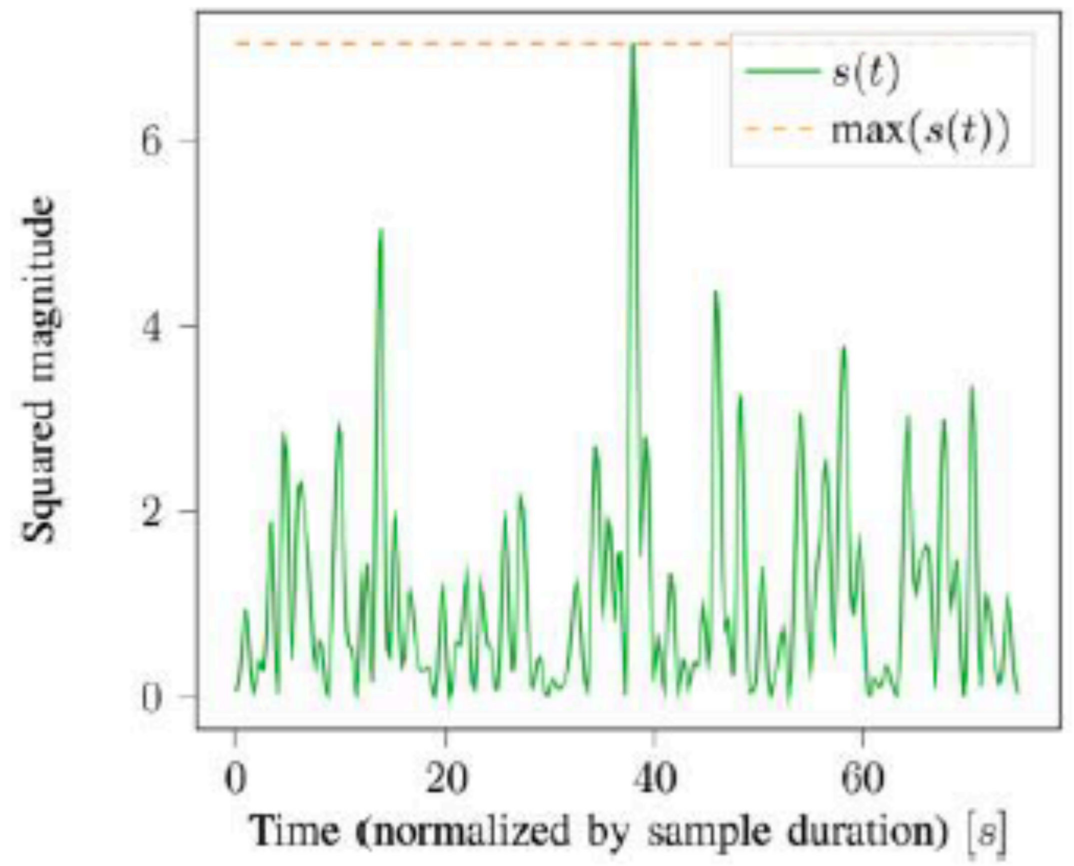
- Reference model
  - OFDM system with M symbols and N subcarriers
  - On each RE (m,n) a vector of bits is mapped :

$$\boldsymbol{b}_{m,n} \in \{0,1\}^Q \to x_{m,n} \in \mathcal{C}$$

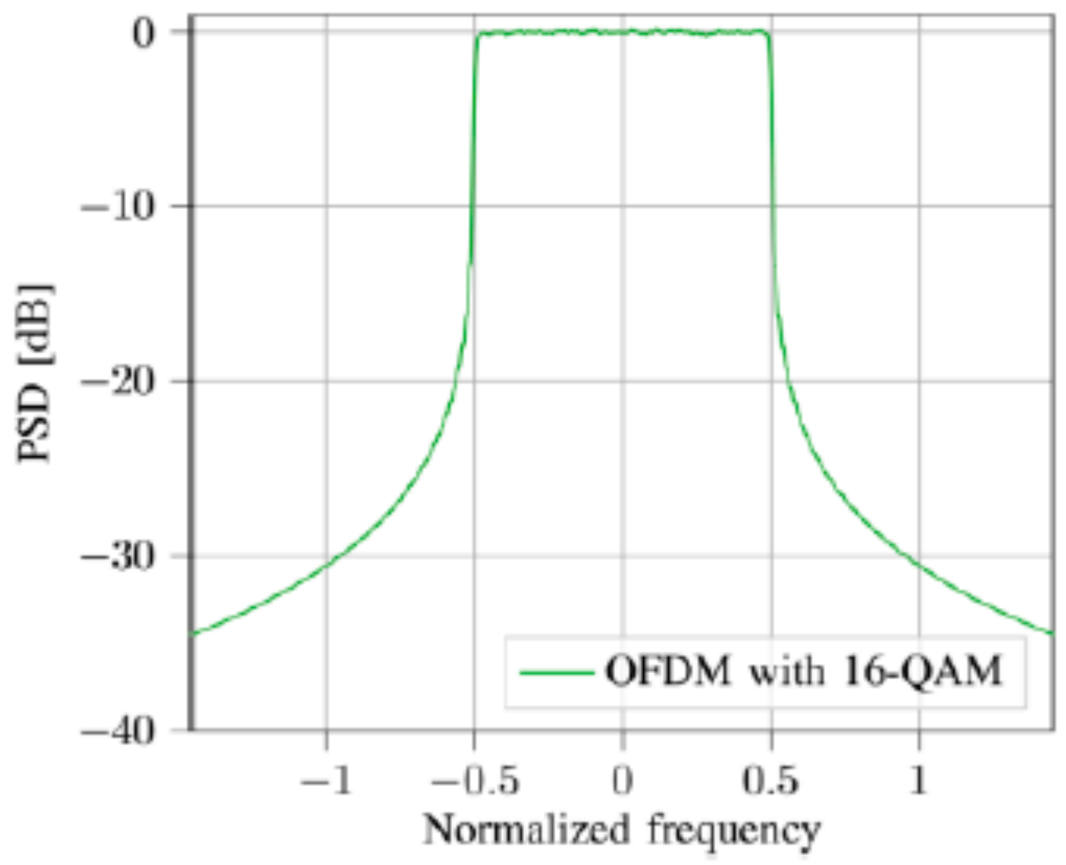
  - OFDM suffers from high peak to average power ratio (PAPR) and high adjacent channel leakage ratio (ACLR)







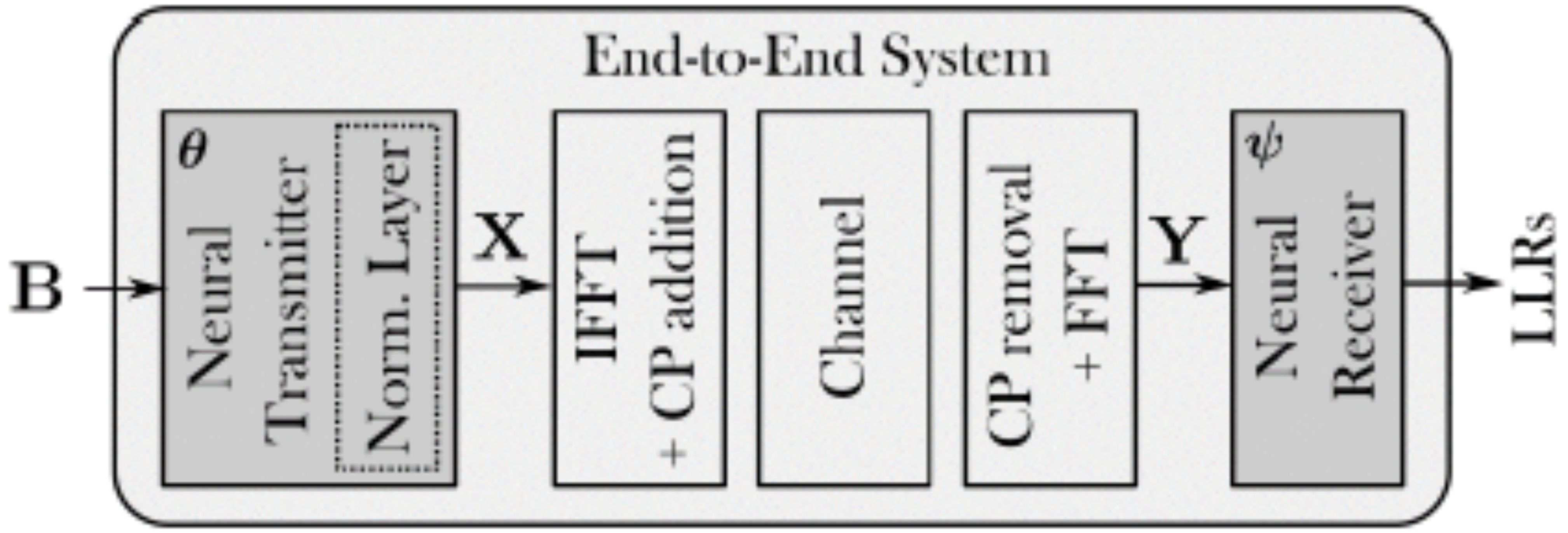
Example of time-domain OFDM signal

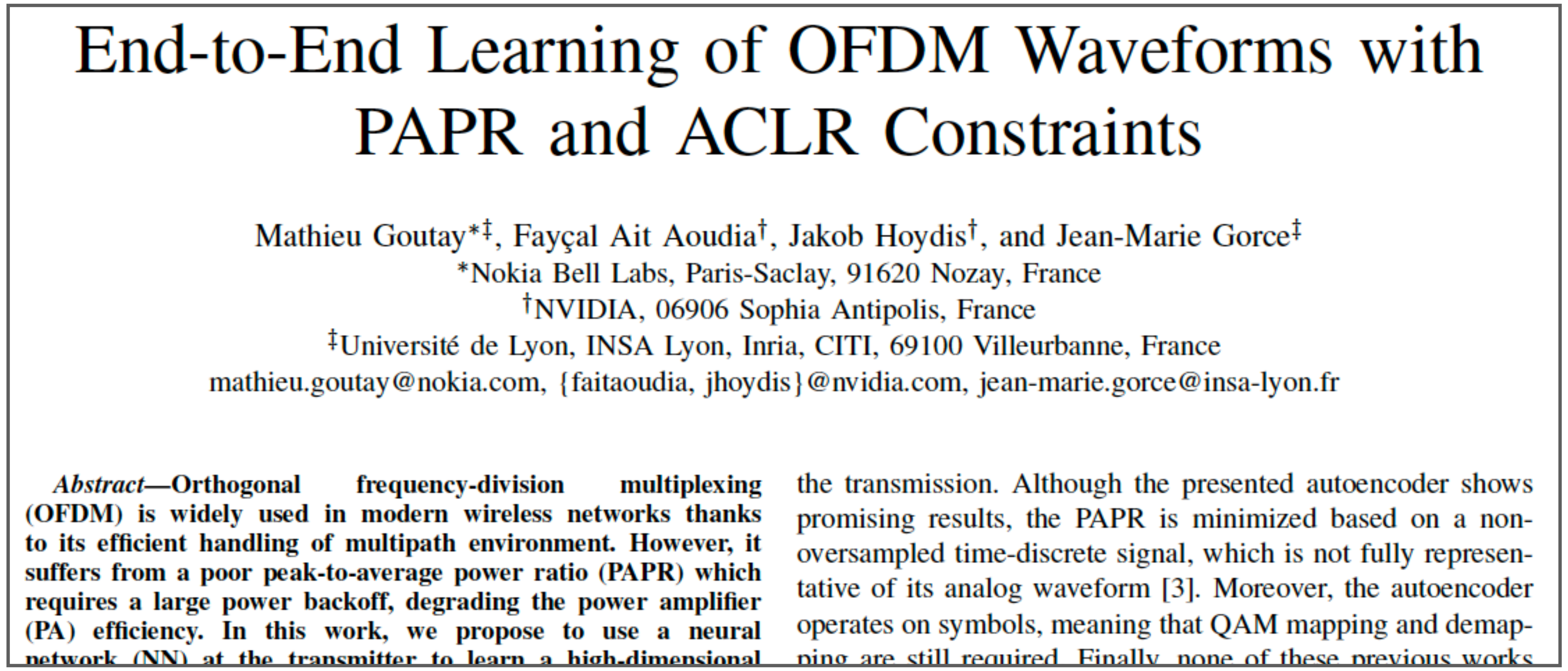


Power spectral density (PSD) of OFDM signals

- New system : the communication system (P2P) is modeled as an auto encoder, with parameter θ and Ψ, on an OFDM frame, maximizing a rate under side constraints:

$$\begin{array}{lll} \underset{\boldsymbol{\theta},\boldsymbol{\psi}}{\text{maximize}} & C(\boldsymbol{\theta},\boldsymbol{\psi}) & \longrightarrow \text{Achievable information rate} \\ \text{subject to} & \text{PAPR}\ (\boldsymbol{\theta}) = \gamma_{\text{peak}} & \longrightarrow \text{Target PAPR} \\ & \text{ACLR}(\boldsymbol{\theta}) \leq \beta_{\text{leak}} & \longrightarrow \text{Target ACLR} \end{array}$$



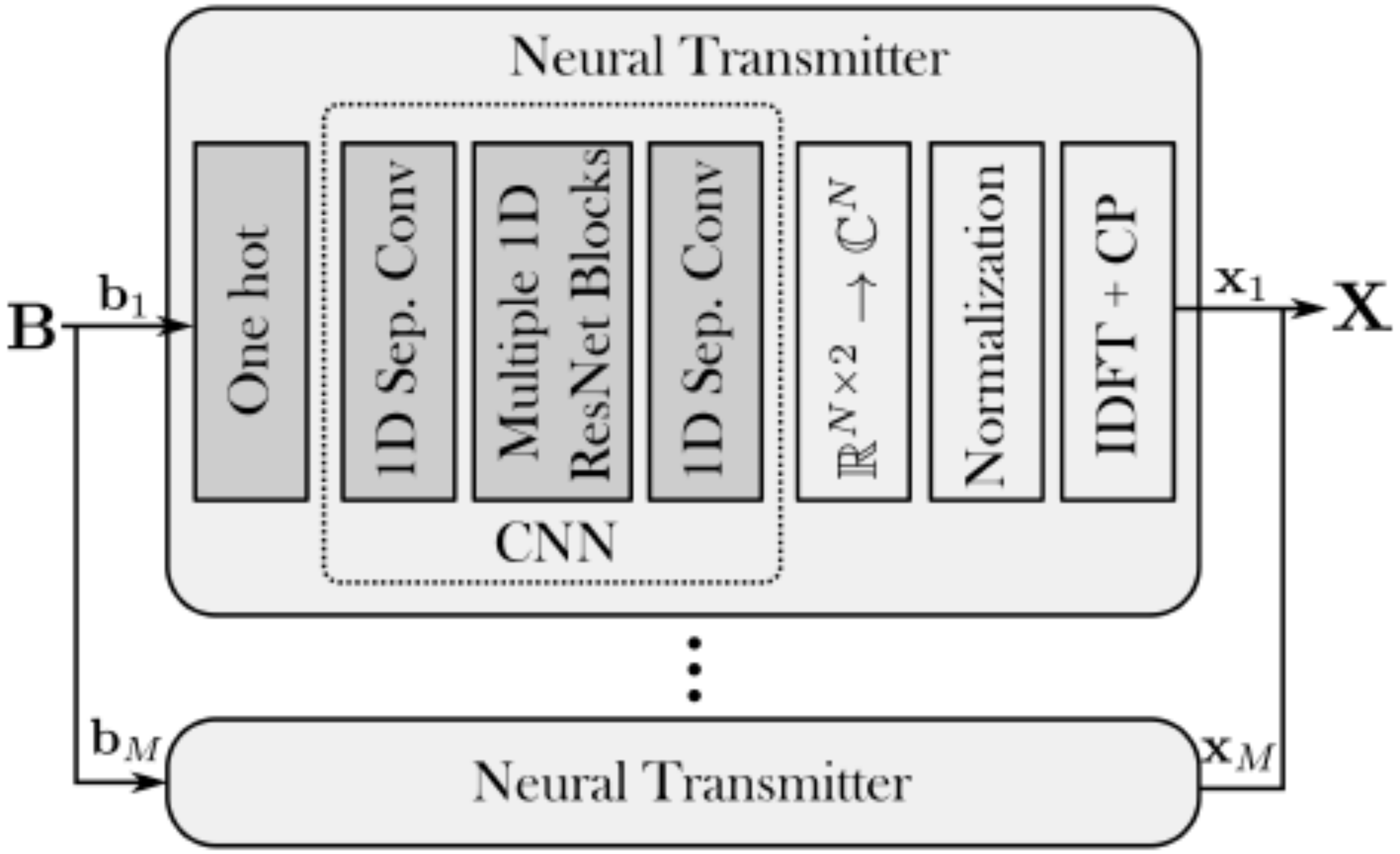

End-to-End Learning of OFDM Waveforms with PAPR and ACLR Constraints

Mathieu Goutay*‡, Fayçal Ait Aoudia†, Jakob Hoydis†, and Jean-Marie Gorce‡
*Nokia Bell Labs, Paris-Saclay, 91620 Nozay, France
†NVIDIA, 06906 Sophia Antipolis, France
‡Université de Lyon, INSA Lyon, Inria, CITI, 69100 Villeurbanne, France
mathieu.goutay@nokia.com, {faitaoudia, jhoydis}@nvidia.com, jean-marie.gorce@insa-lyon.fr

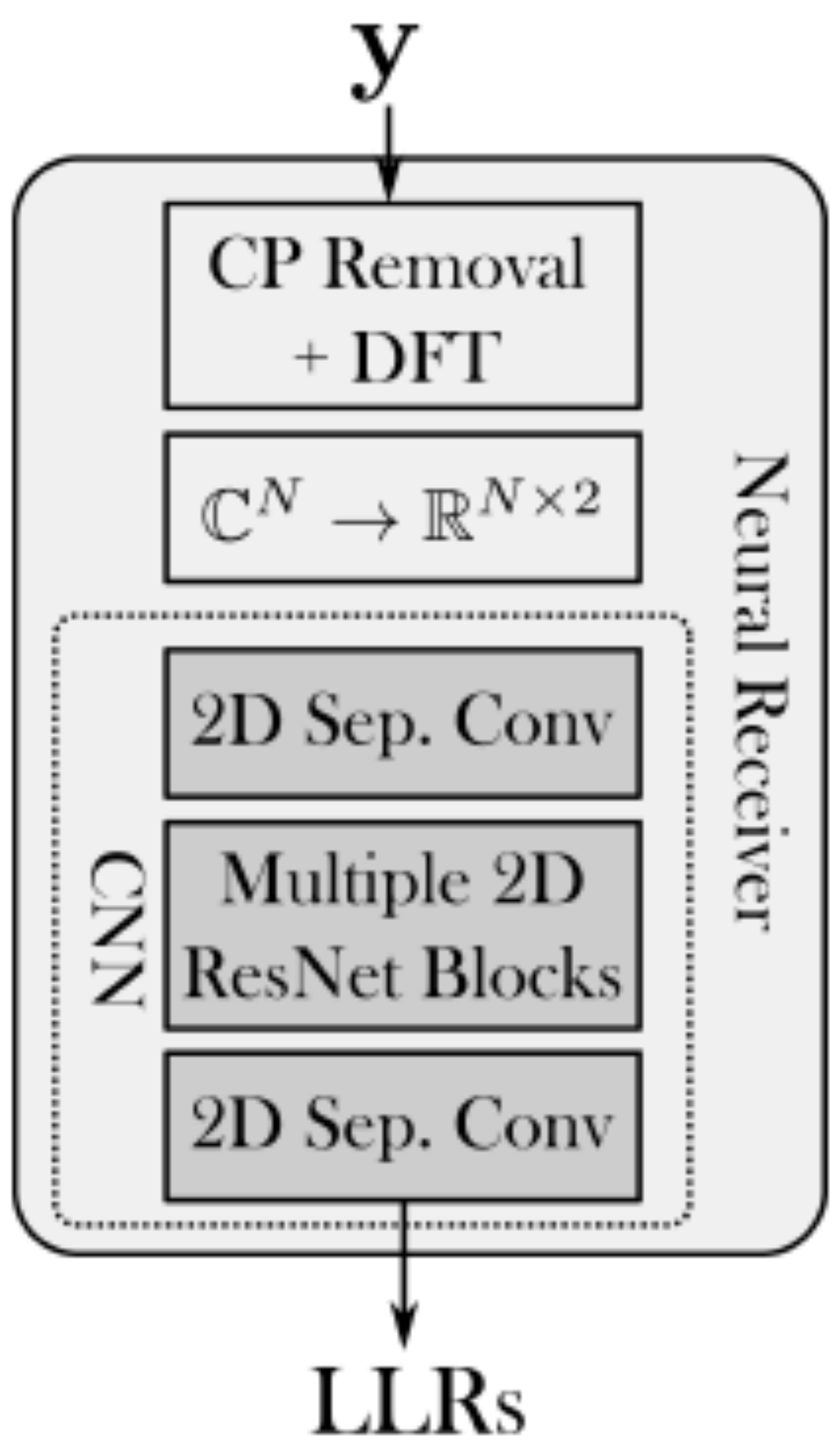
***Abstract*—Orthogonal frequency-division multiplexing (OFDM) is widely used in modern wireless networks thanks to its efficient handling of multipath environment. However, it suffers from a poor peak-to-average power ratio (PAPR) which requires a large power backoff, degrading the power amplifier (PA) efficiency. In this work, we propose to use a neural**

the transmission. Although the presented autoencoder shows promising results, the PAPR is minimized based on a non-oversampled time-discrete signal, which is not fully representative of its analog waveform [3]. Moreover, the autoencoder operates on symbols, meaning that QAM mapping and demapping are still required.

NN based architecture, end-to-end learned on simulated channels



(a) Neural Transmitter.



(b) Neural Receiver.

- Assumptions :
  - Average channel is known at the transmitter and power control is applied to achieve an average SNR.
  - The system is trained on 3GPP-compliant urban microcell (UMi) LOS and non-LOS models

Sketch-up of the algorithmic solution with an end-to-end loss (Augmented Lagrangian based) :

$$
\begin{aligned}
\overline{L}(\boldsymbol{\theta}, \boldsymbol{\psi}, \lambda_p, \lambda_l, \mu_p, \mu_l) = {} & L_C(\boldsymbol{\theta}, \boldsymbol{\psi}) \\
& + \lambda_p L_{\gamma_{\text{peak}}}(\boldsymbol{\theta}) + \frac{1}{2}\mu_p |L_{\gamma_{\text{peak}}}(\boldsymbol{\theta})|^2 \\
& + \frac{1}{2\mu_l}\left(\max(0, \lambda_l + \mu_l L_{\beta_{\text{leak}}}(\boldsymbol{\theta}))^2 - \lambda_l^2\right).
\end{aligned}
$$

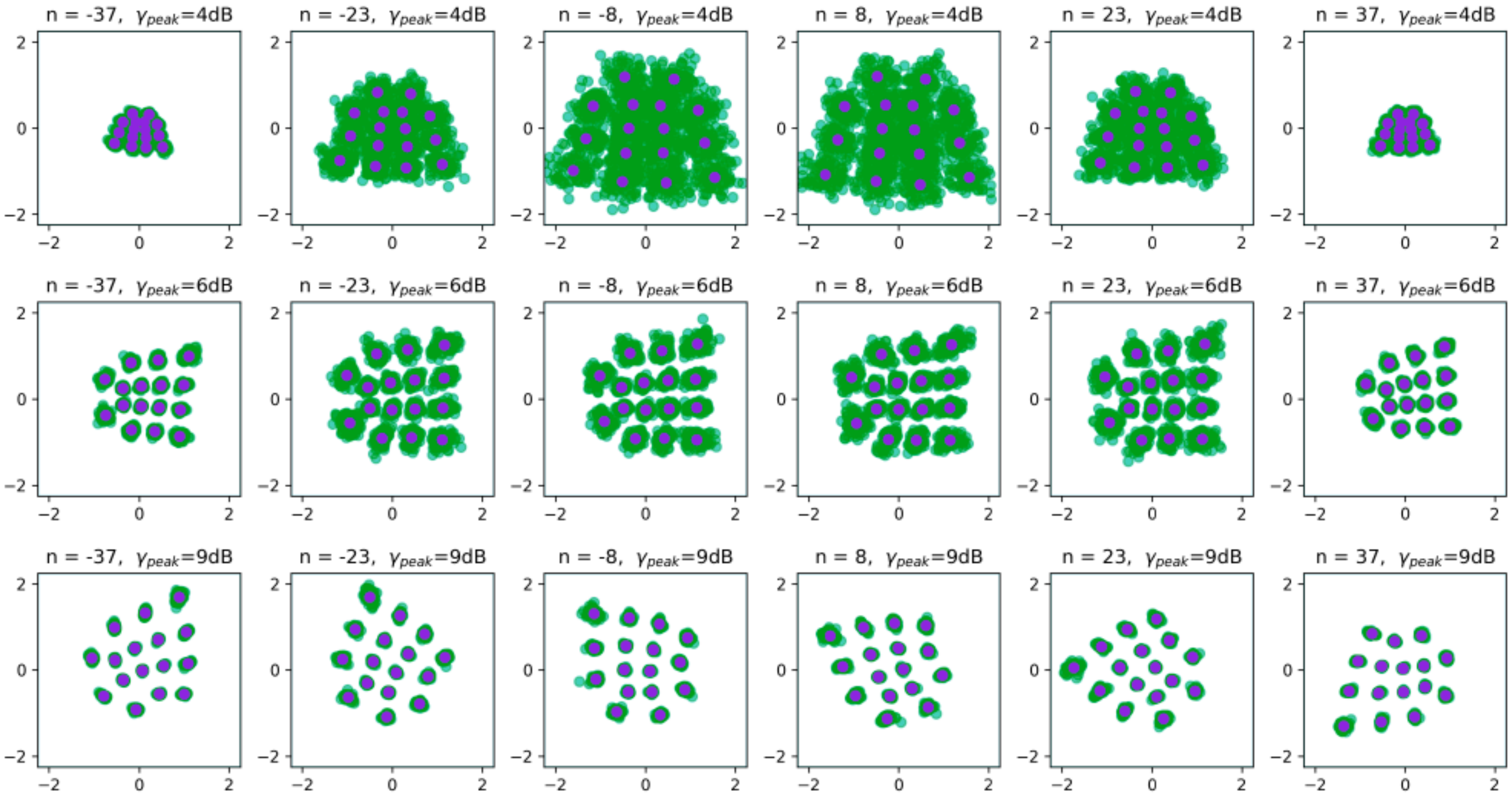
where $\lambda_p, \lambda_l, \mu_p, \mu_l$ are optimization parameters

Algorithm 1: Training procedure

Initialize $\boldsymbol{\theta}, \boldsymbol{\psi}, \lambda_p^{(0)}, \lambda_l^{(0)}, \mu_p^{(0)}, \mu_l^{(0)}$
for $u = 0, \ldots$ **do**
  ▷ Perform multiple steps of SGD on $\overline{L}(\boldsymbol{\theta}, \boldsymbol{\psi}, \lambda, \lambda_l, \mu_p, \mu_l)$ w.r.t. $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$
  ▷ Update optimization hyperparameters :
  $\lambda_p^{(u+1)} = \lambda_p^{(u)} + \mu_p^{(u)} L_{\gamma_{\text{peak}}}(\boldsymbol{\theta})$
  $\lambda_l^{(u+1)} = \max\left(0, \lambda_l^{(u)} + \mu_l^{(u)} L_{\beta_{\text{leak}}}(\boldsymbol{\theta})\right)$
  $\mu_p^{(u+1)} = (1+\tau)\mu_p^{(u)}$
  $\mu_l^{(u+1)} = (1+\tau)\mu_l^{(u)}$
end

Illustration of the obtained constellations

- Here, the learned constellations comply with a lax ACLR and variable PAPR constraints
- The green points are effective constellations transmitted, showing variability to cope with PAPR constraint.
- The channel is learned from the non symmetric distribution, at the receiver

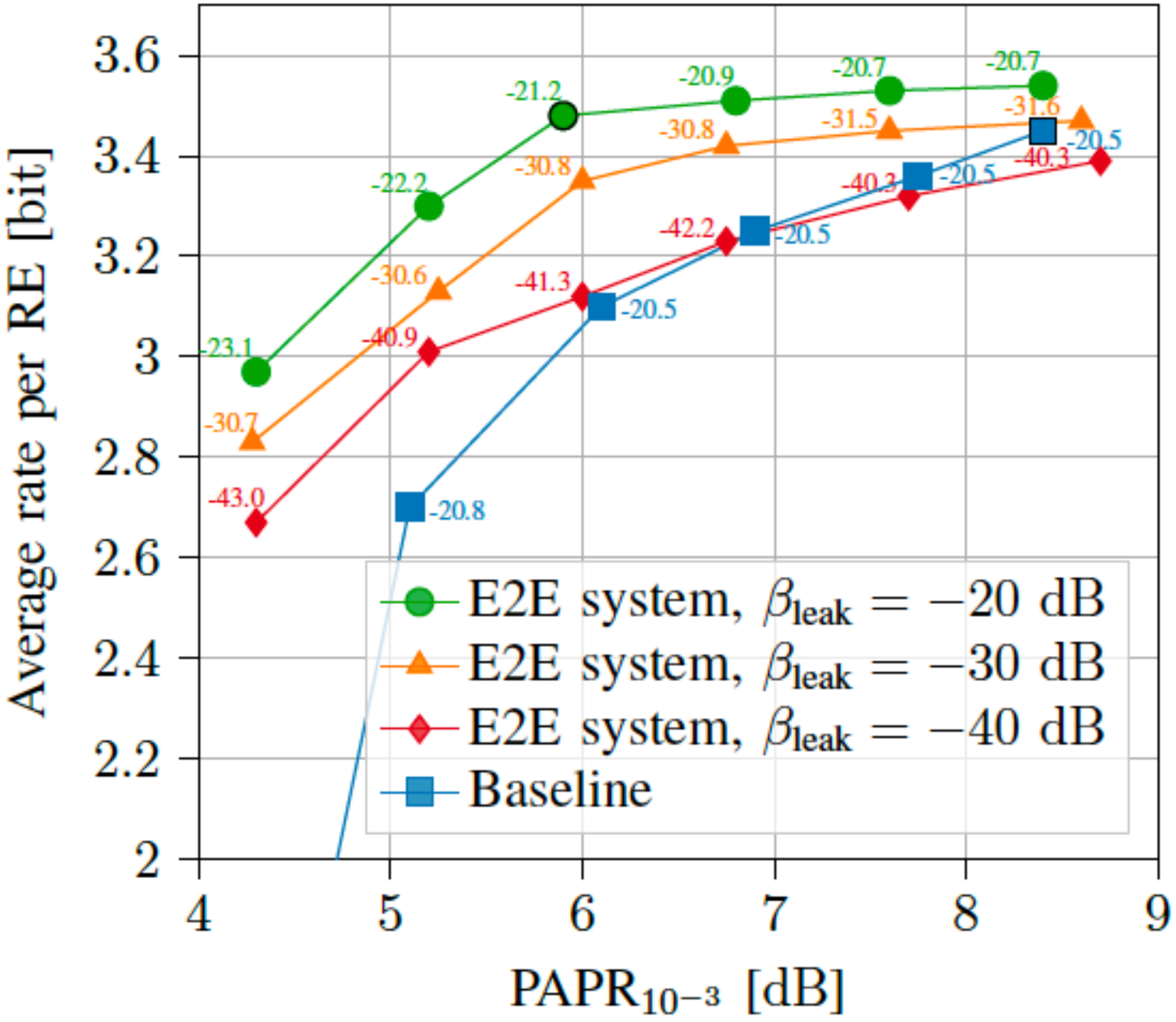


Low PAPR required

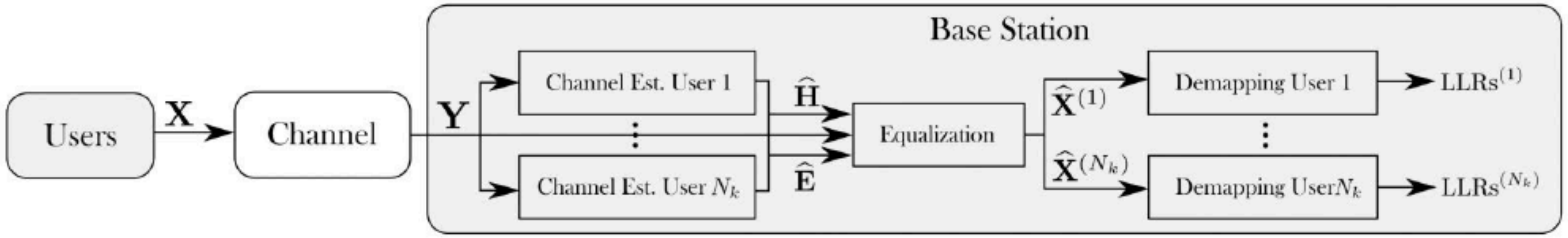
Higher PAPR allowed

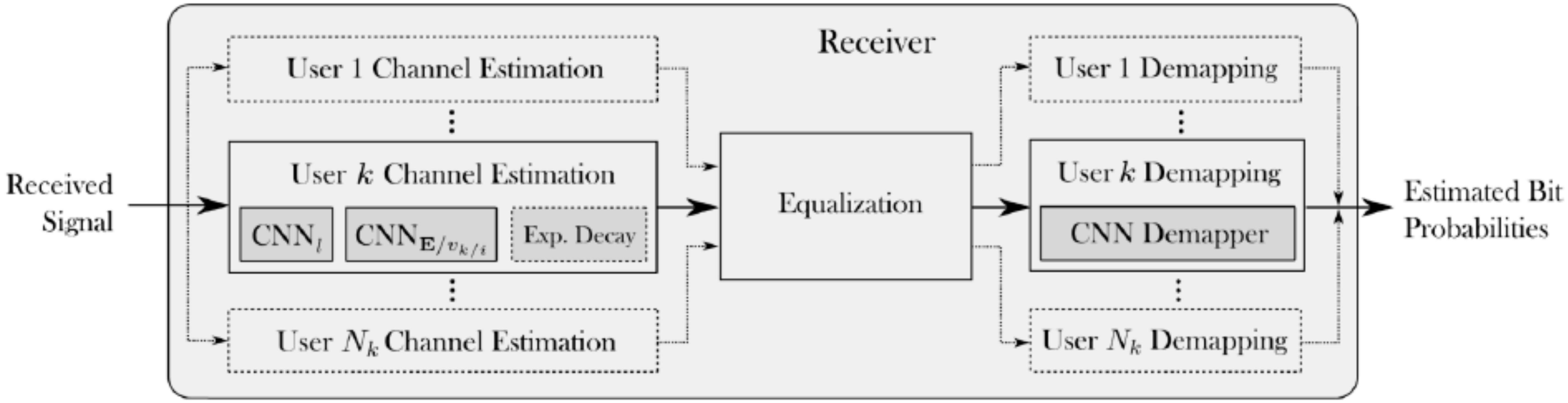
Learning OFDM Waveforms with PAPR and ACLR Constraints

- Results
  - The proposed approach outperforms the trade-offs (rate, PAPR, ACLR)
  - Note : leakage is optimized in the baseline with PRTs (peak reduction tones)

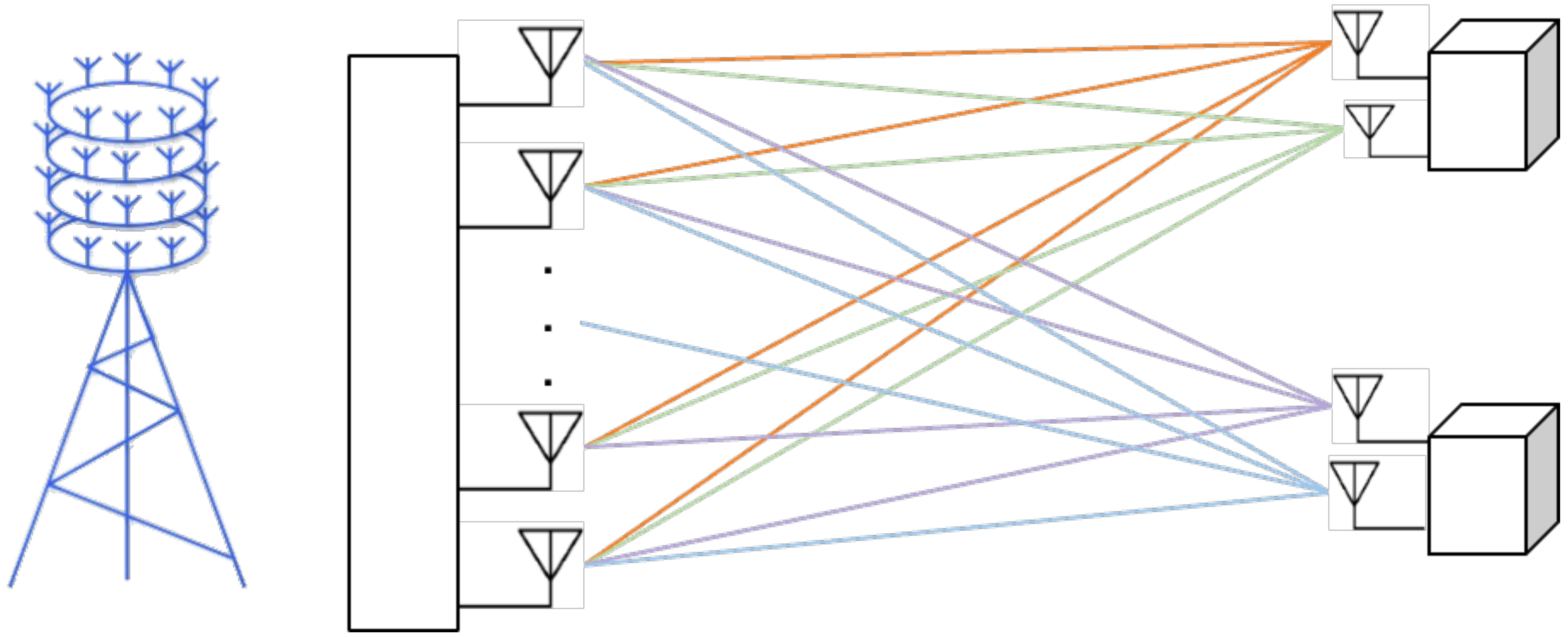


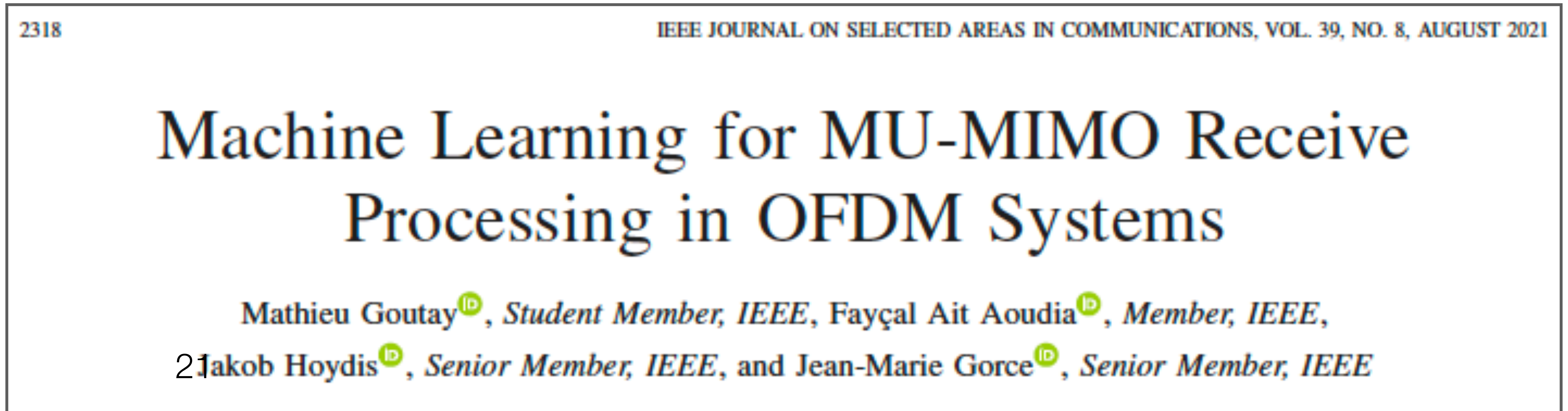
Scénario 2 : ML-enhanced Receive Processing for MU-MIMO





Enhance multiple blocks with ML components



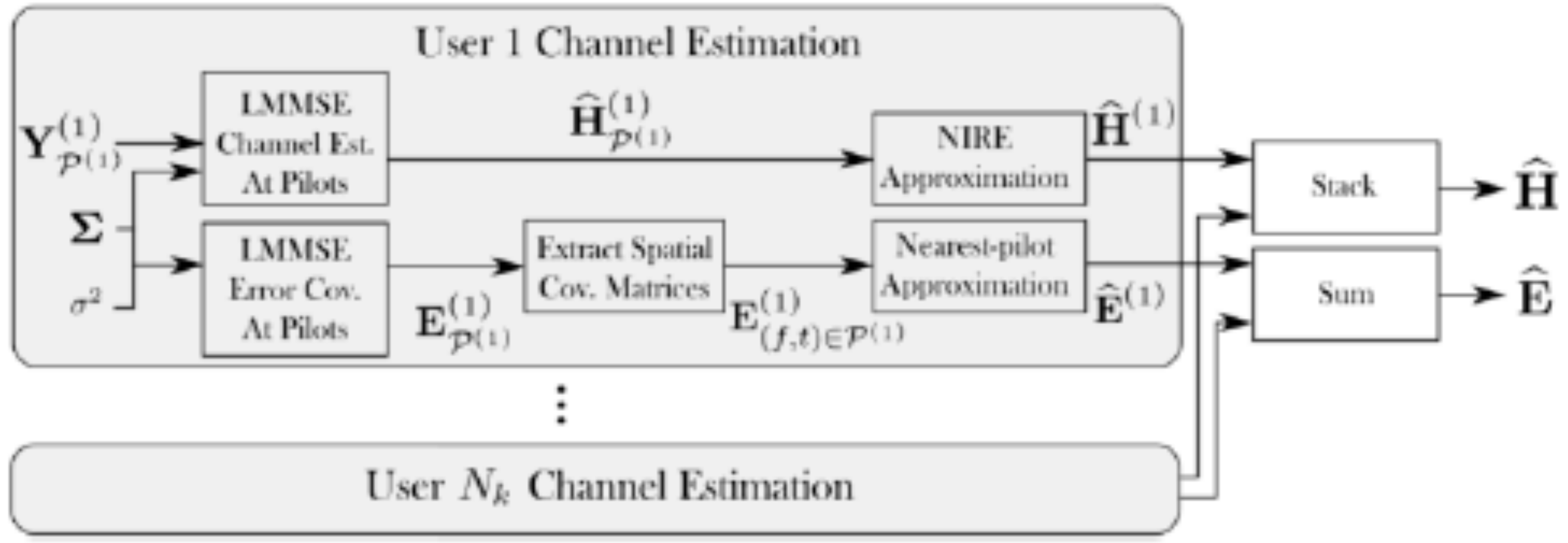

2318 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 39, NO. 8, AUGUST 2021

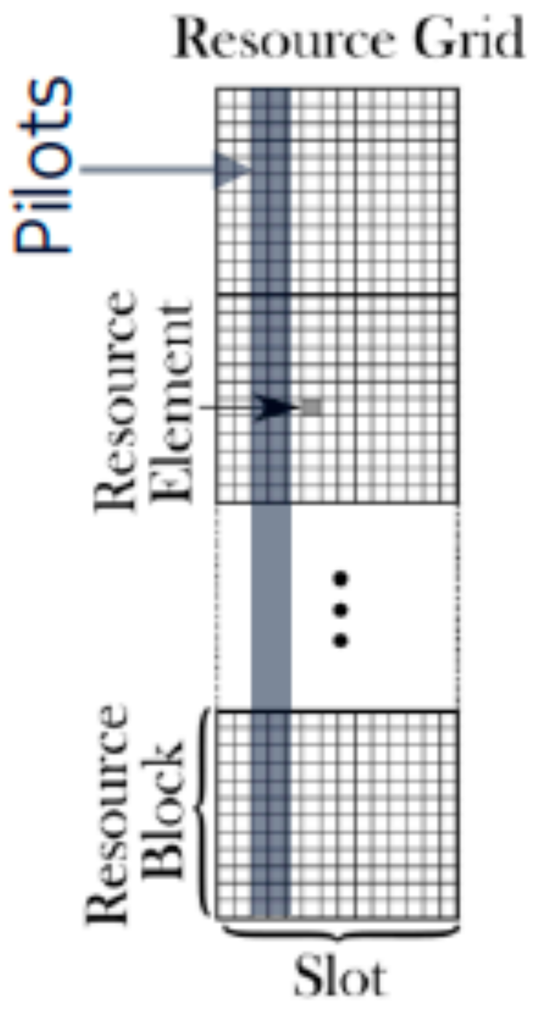
Machine Learning for MU-MIMO Receive Processing in OFDM Systems

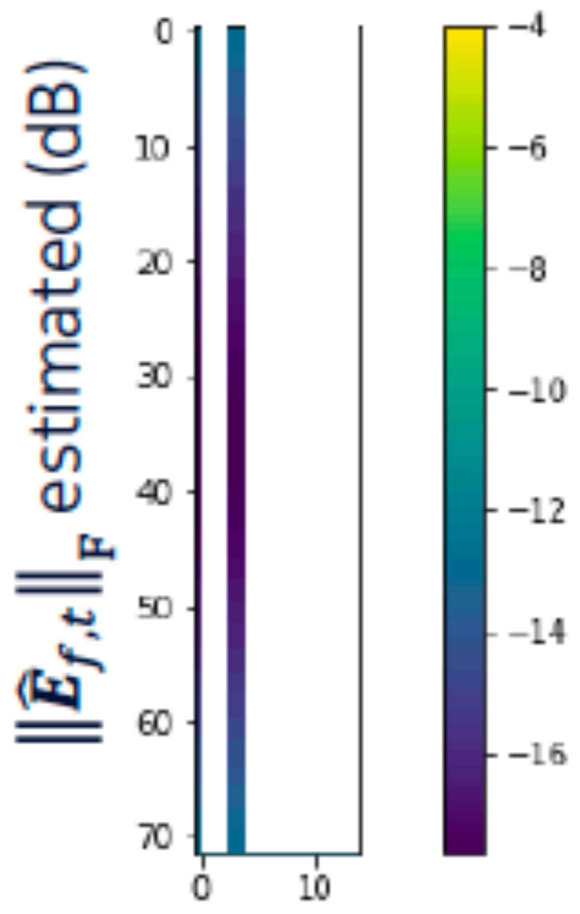
Mathieu Goutay, *Student Member, IEEE*, Fayçal Ait Aoudia, *Member, IEEE*, Jakob Hoydis, *Senior Member, IEEE*, and Jean-Marie Gorce, *Senior Member, IEEE*

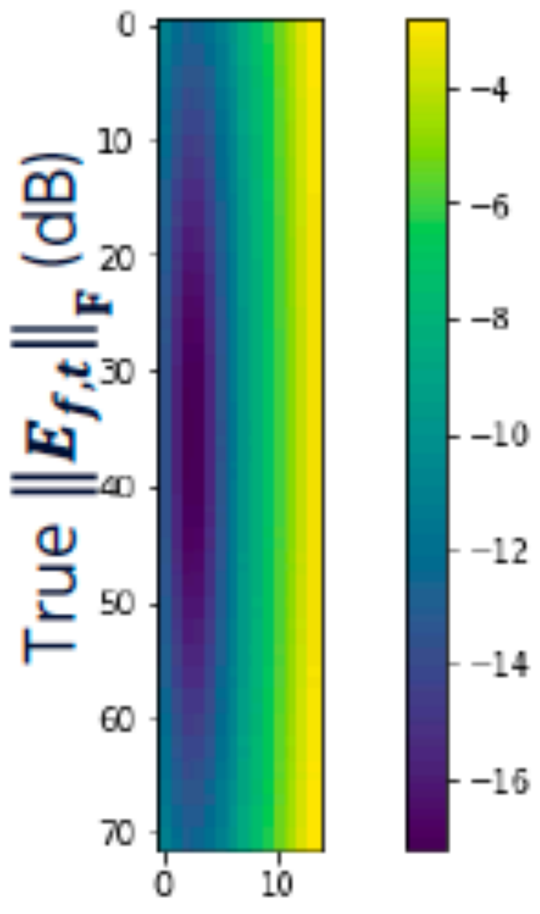
ML-enhanced receiver : channel estimate errors

Problem: The estimates of $\widehat{\boldsymbol{E}}$ are sub-optimal

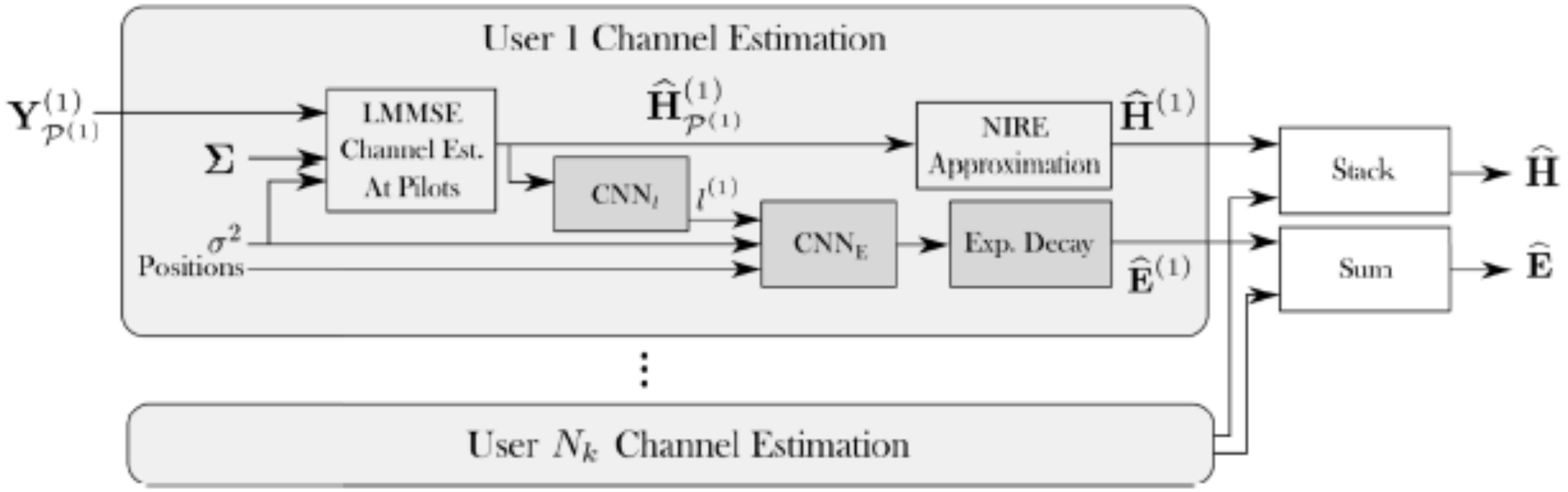








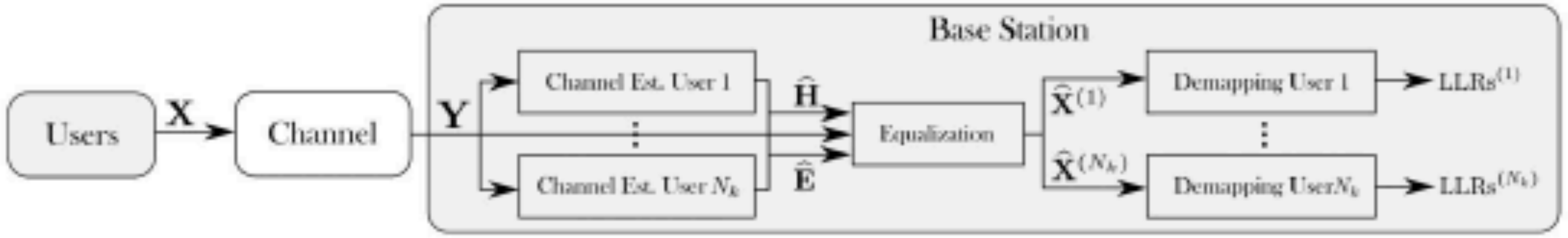
Solution: Use a CNN to estimate $\boldsymbol{E}$



- The CNN will learn the error statistics during training
- But $\boldsymbol{E}$ is a tensor of dimension $\boldsymbol{N} \times \boldsymbol{M} \times \boldsymbol{L} \times \boldsymbol{L}$!

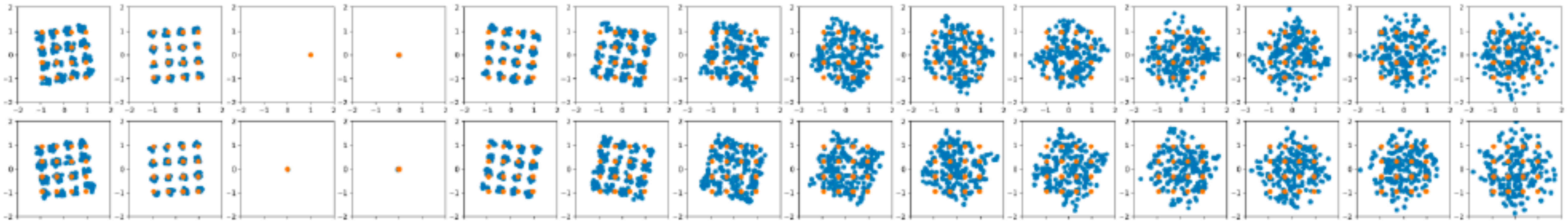
  → The CNN outputs 2 parameters $\alpha_{f,t}, \beta_{f,t}$ per RE and use a complex exp decay model

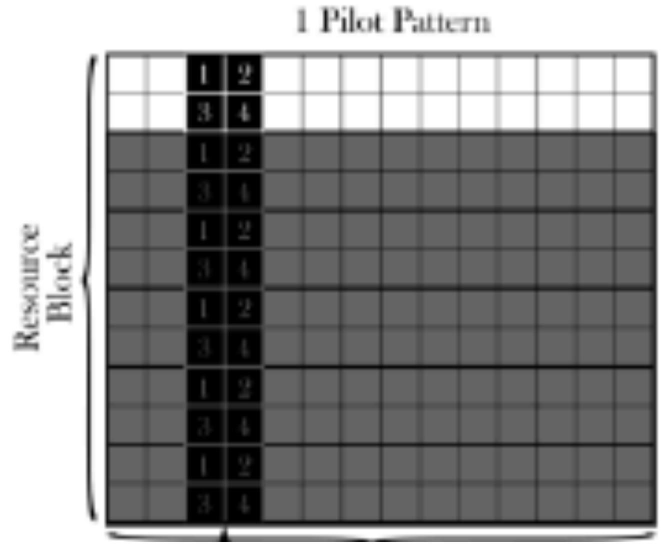
ML-enhanced receiver : unmapping



Problem

Equalized symbols suffer from channel aging

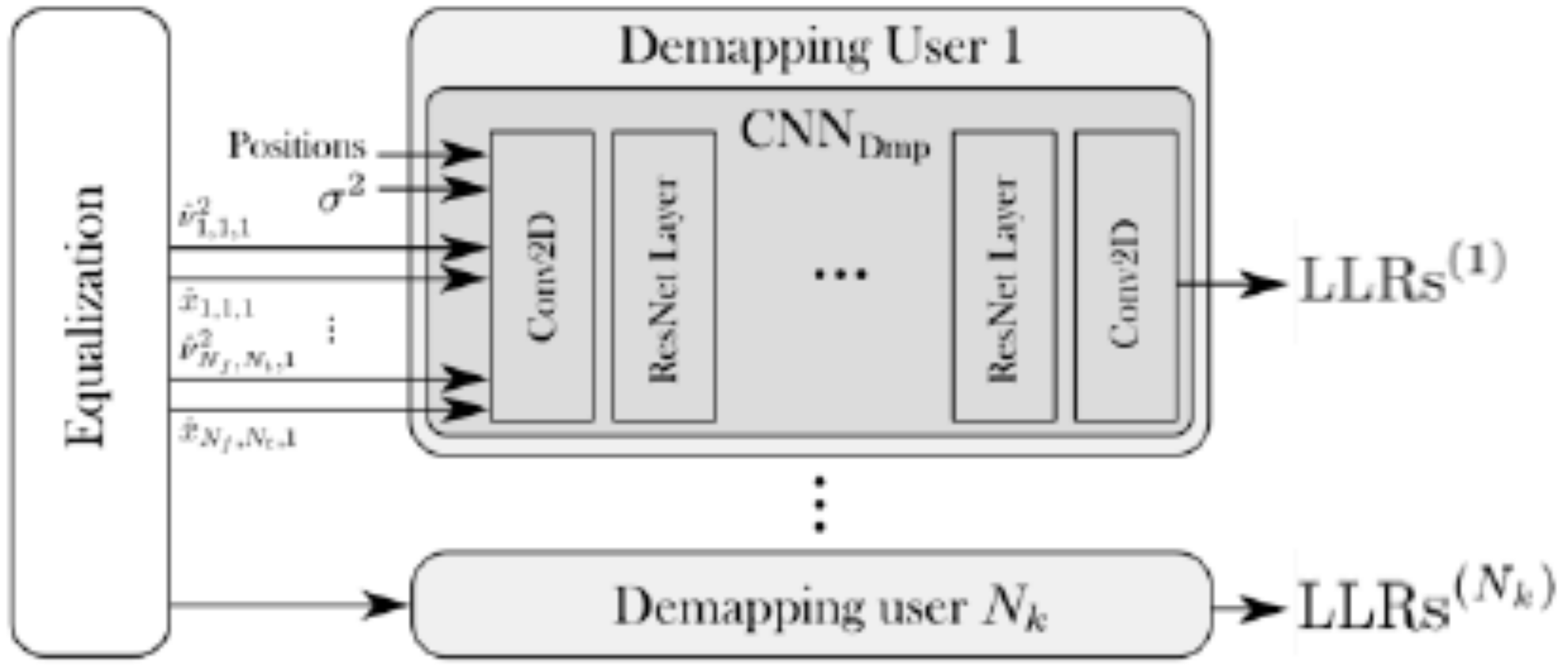




Solution

Use a CNN to demap the symbols jointly on all REs

- The CNN can estimate the channel aging by looking at the entire OFDM Resource Grid

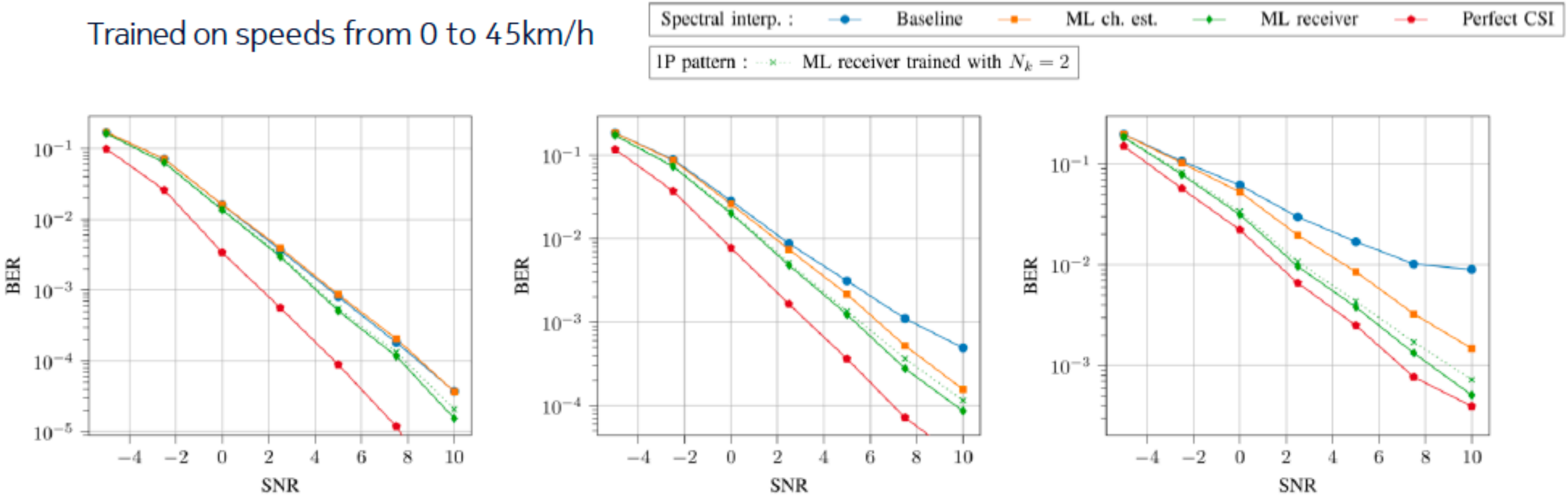


Results, uplink, @slow speeds

- 16-QAM, code rate $\eta = 1/2$
- $K = 4$ users
- $L = 16$ Rx antennas
- 3GPP UMi NLOS (Generated with Quadriga)
- 3.5Ghz, 15Khz subcarrier spacing

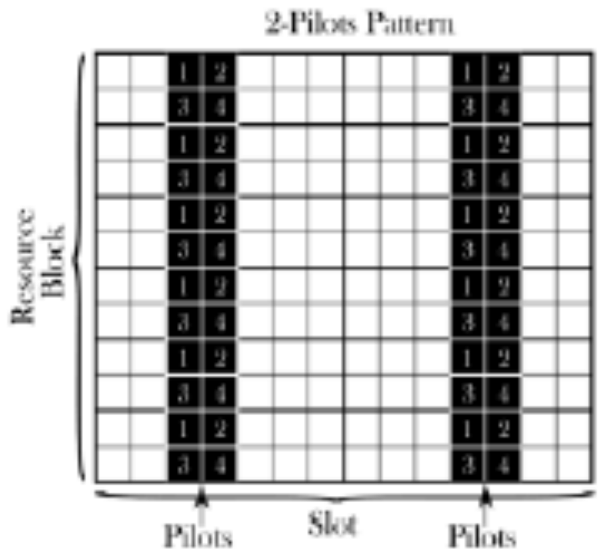
Trained on speeds from 0 to 45km/h

- Baseline : traditional ch. est. and demapper
- ML ch. est : ML channel estimator only
- ML receiver : ML ch. est. + ML demapper
- Perfect CSI : perfect $\hat{H}$ at pilot, perfect $\hat{E}$ everywhere



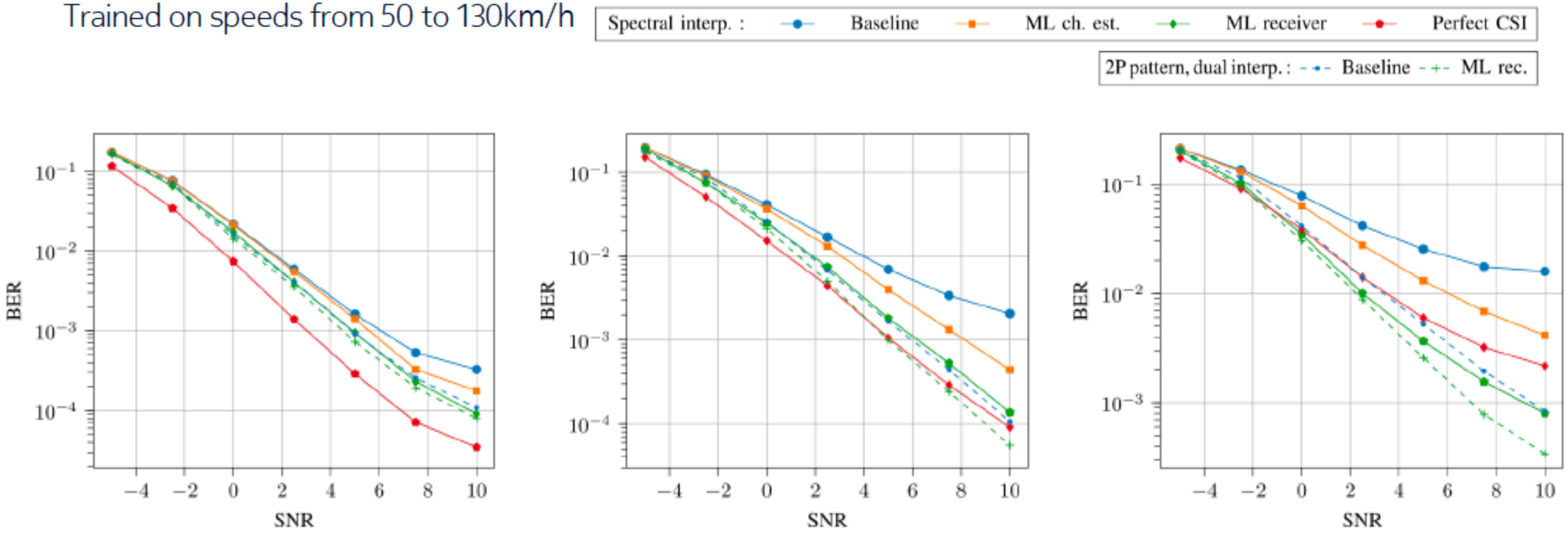
(a) 1P pilot pattern at 0 to 15 km h^{-1}.
(b) 1P pilot pattern at 15 to 30 km h^{-1}.
(c) 1P pilot pattern at 30 to 45 km h^{-1}.

Results, uplink, @high speeds



Trained on speeds from 50 to 130km/h

- Baseline : traditional ch. est. and demapper
- ML ch. est : ML channel estimator only
- ML receiver : ML ch. est. + ML demapper
- Perfect CSI : perfect $\hat{H}$ at pilot, perfect $\hat{E}$ everywhere



(d) 2P pilot pattern at 50 to 70 km h^{-1}.

(e) 2P pilot pattern at 80 to 100 km h^{-1}.

(f) 2P pilot pattern at 110 to 130 km h^{-1}.

Research trends

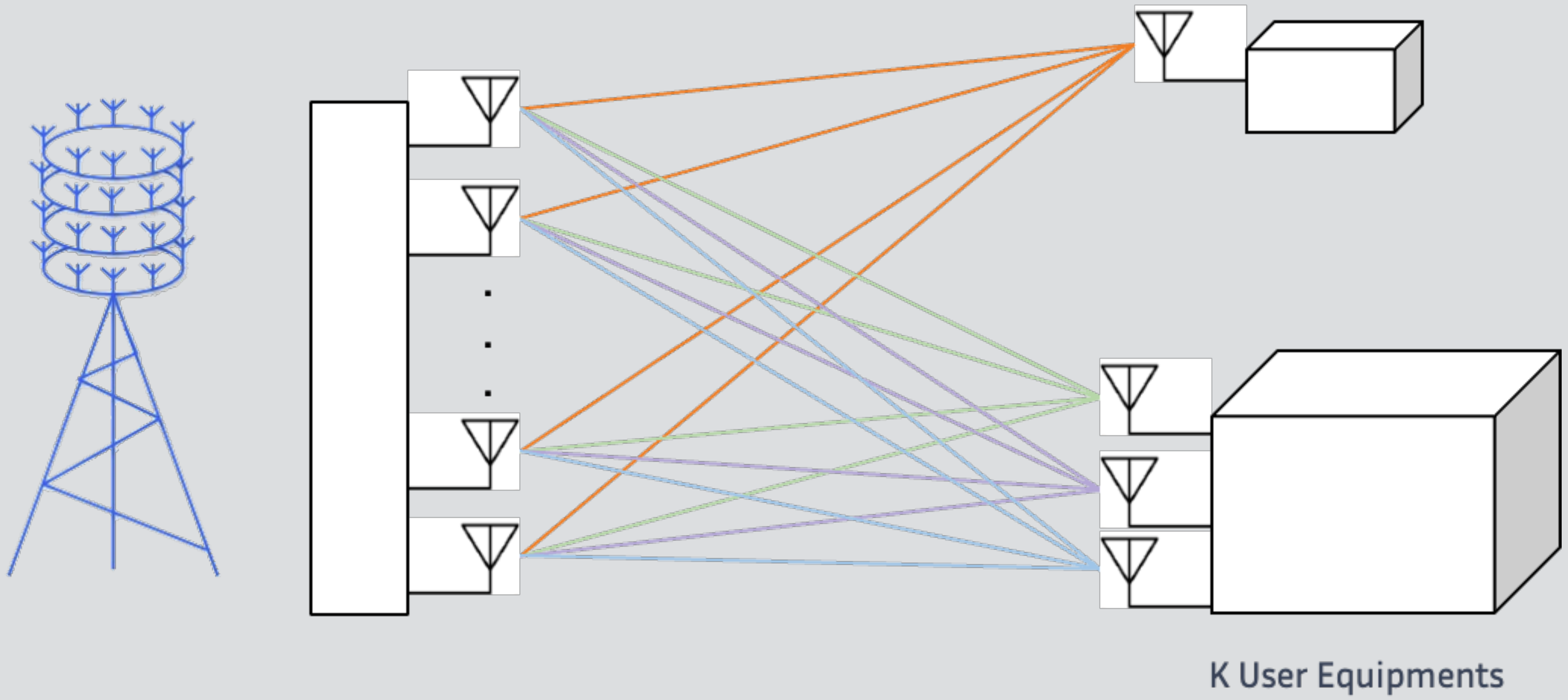
- Learning waveforms over multiple symbols allows to determine more robust detectors, with less (or no) pilots —> **Embedded pilots, optimized receiver**
- Other End-to-end approaches including constellation shaping, encoders, … are proposed in the literature, e.g. :

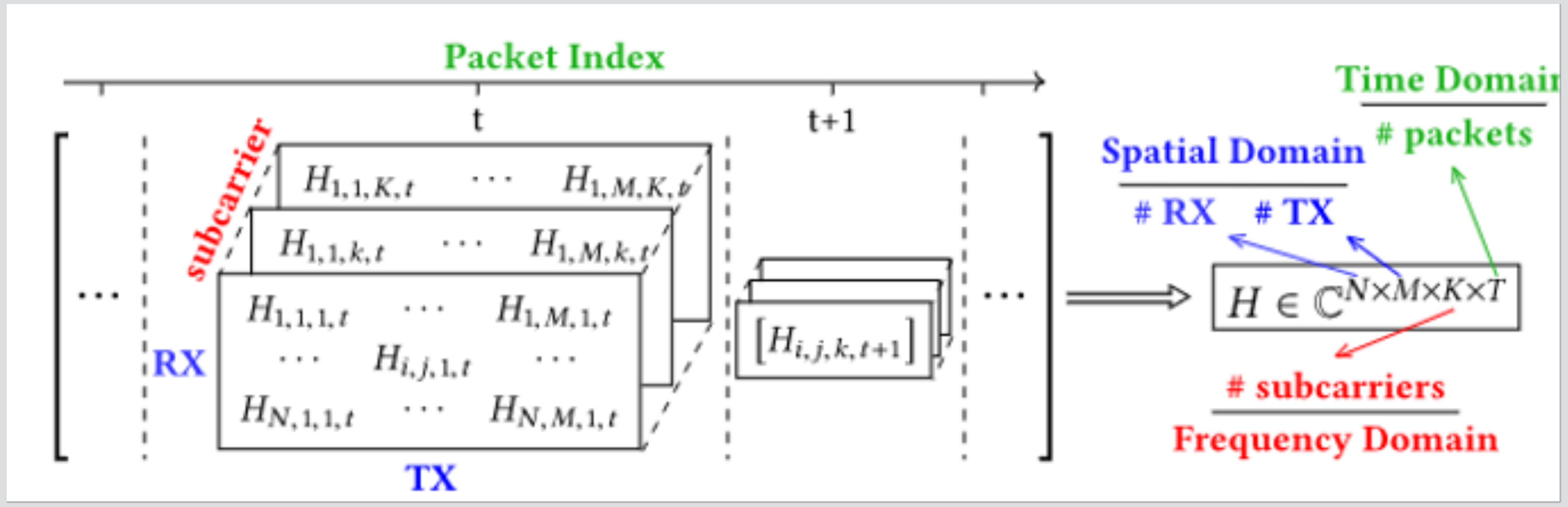
  Aoudia, F. A., & Hoydis, J. Waveform learning for next-generation wireless communication systems. IEEE Transactions on Communications (2022)

- **The performance should be balanced with the computational complexity.**

4- Federated learning for CSI estimation

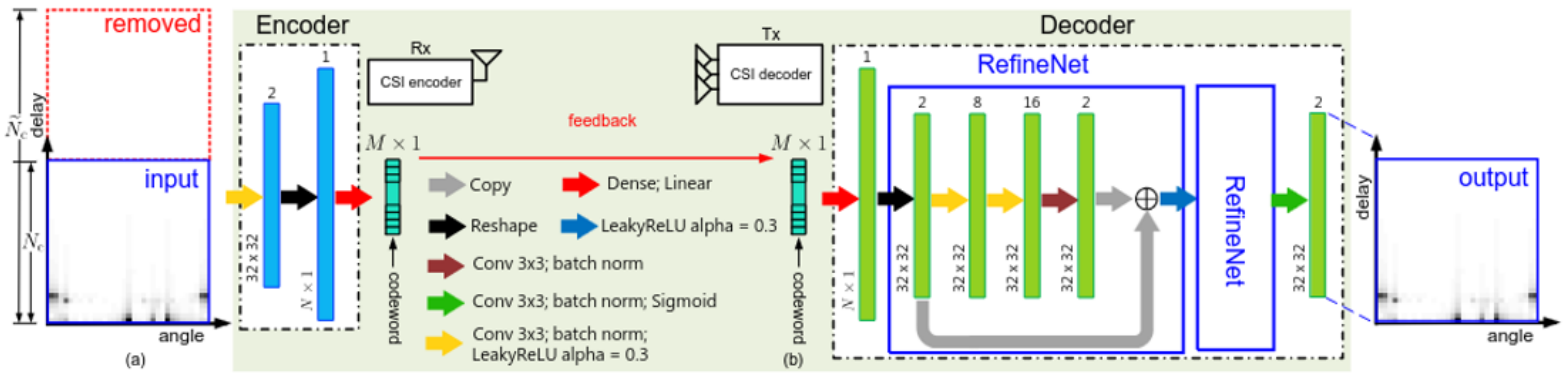
Joint work with Loukas Duque, Tan Khiem Huyn, Hadrien Hendricks, Florence Forbes, Malcolm Egan





Z. Du, H. Li, L. Li, B. Zhang, Z. Liu and X. Gu, "Training CSI Feedback Model with Federated Learning in Massive Mimo Systems," 2023 8th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC), Beijing, China, 2023

- Reference solution : CSInet



1- Offline learning (e.g. on extensive simulations)

2- Sharing the encoder with UEs

3- Replace standard CQI with compressed feedback

$$L(\theta_{enc}, \theta_{dec}) = \mathbb{E}\left[NMSE(H, \theta_{enc}, \theta_{dec})\right]$$
$$= \frac{1}{T}\sum_t \frac{||f_{dec}\left(f_{enc}\left(H, \theta_{enc}\right), \theta_{dec}\right) - H_t||^2}{||H_t||^2}$$

C. -K. Wen, W. -T. Shih and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," in IEEE Wireless Communications Letters, 2018

- **Our (on going) work :**
  - Evaluate « Distributed learning solution » for this CSI compression problem to
    - Adapt the model « encoder/decoder » to the actual BS/UEs environment (propagation conditions)
    - One model per BS : each BS learns its own model
    - One model per UE ? Learn the encoder only or encoder/decoder ?
    - **Trade-off generalization-personalization**
- **Scenario under investigation**
  - Each UE = an agent trying to learn a model on-line.
  - Exploit observations among UEs under similar channel conditions but avoid too much generalization

- ## What is Federated learning ?

Algorithm 1: Local SGD

Input: Local datasets, number of communication rounds T, number of local steps K, local learning rate η_l, global learning rate η_g
Output: $\hat{w}_T$
Initialize w_0 ;
for $t = 0, \ldots, T-1$ **do**
  server selects a set $\mathcal{N}_t$ of N available clients ;
  server broadcast w_t to every selected client ;
  for *each client* $m \in \mathcal{N}_t$ **do**
    $w_{t+1}^m \leftarrow$ `LocalUpdate` (w_t, η_l, K) ;
    client sends w_{t+1}^m to the server ;
  end
  server update:
  $$w_{t+1} \leftarrow w_t - \frac{\eta_g}{N} \sum_{m \in \mathcal{N}_t} \left(w_{t+1}^* - w_t \right)$$
end
Function `LocalUpdate` (w_0, η_l, K):
  for $e = 1, \ldots, K$ **do**
    Receive sample x_e from the stream ;
    $w_e \leftarrow w_{e-1} - \eta_l \nabla f_m \left(w_{e-1}; x_e \right)$
  end
  Result: w_K
end

- Each client performs K SGD step on their local model between each communication round
- Communication step to prevent divergence
- Communication cost dominates!

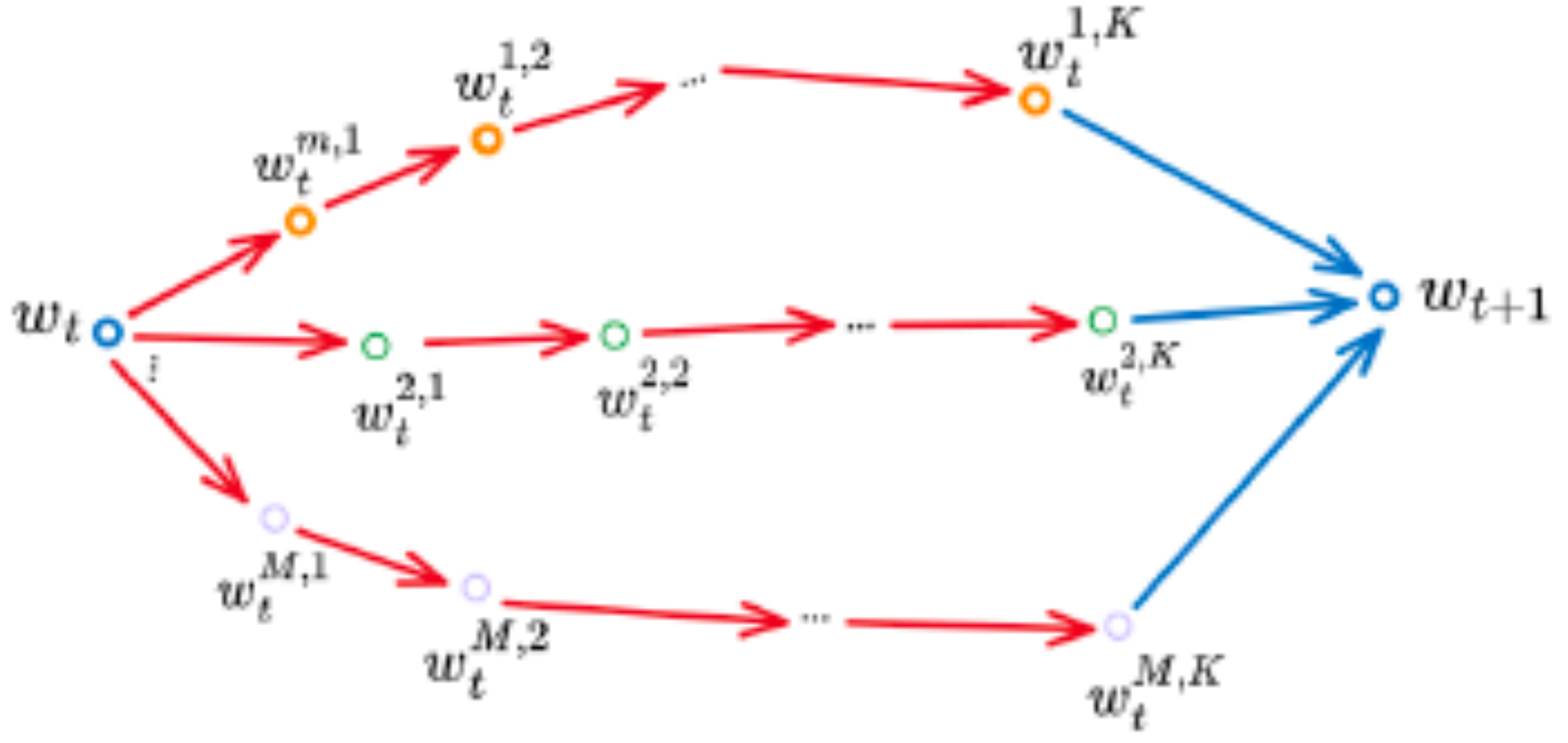
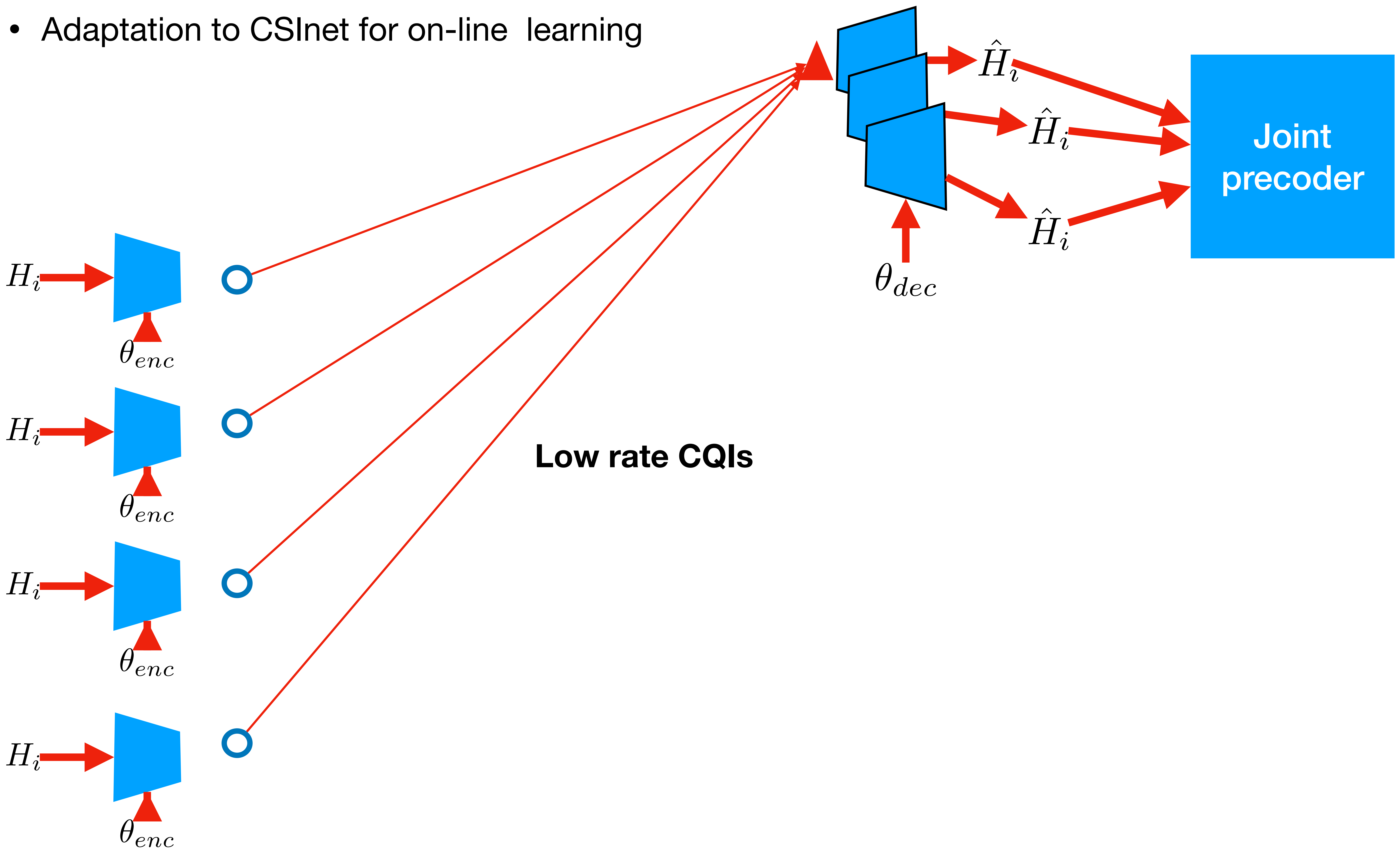


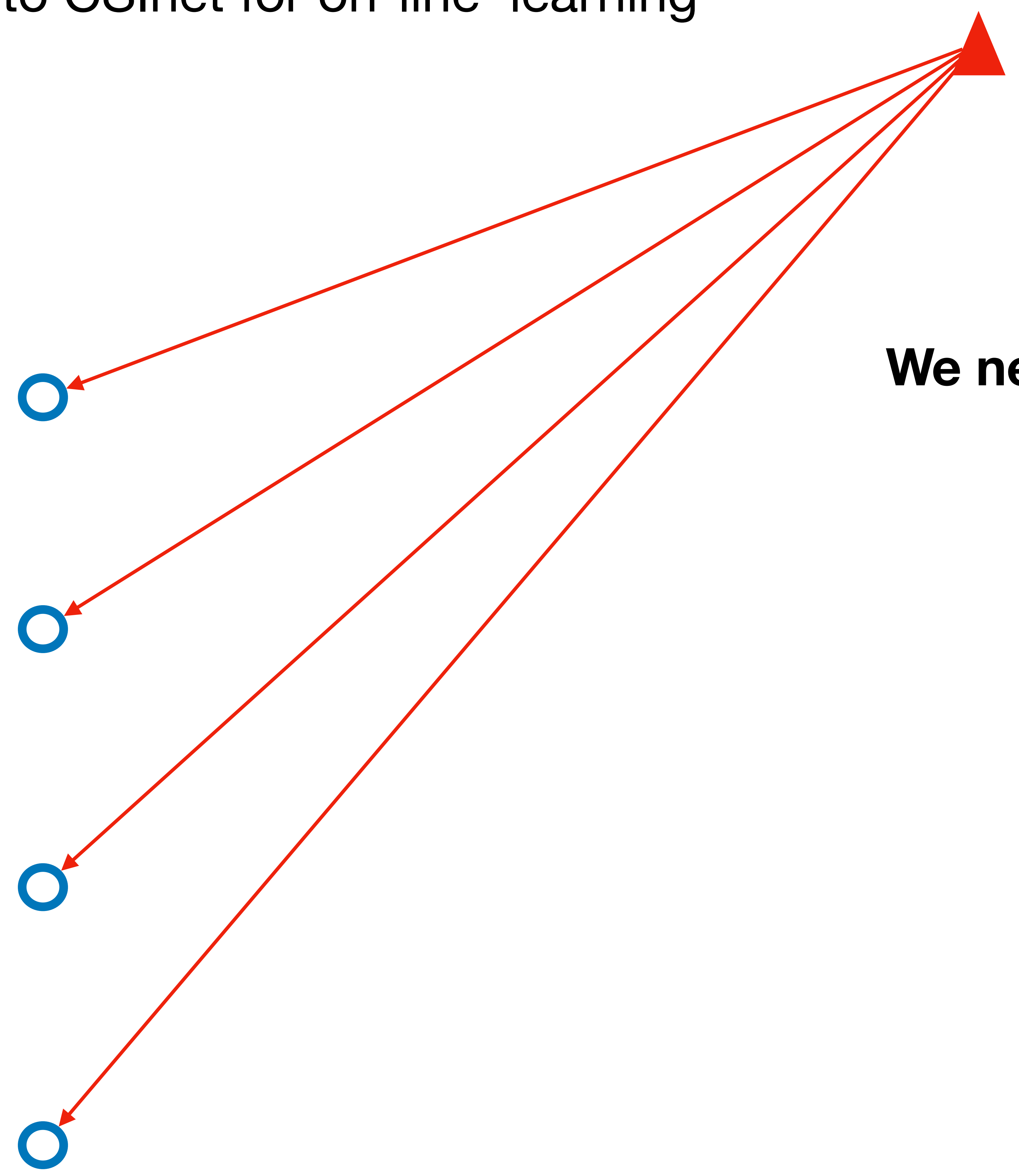
Figure: Local drift

- Adaptation to CSInet for on-line learning



- Adaptation to CSInet for on-line learning

$$\left\{\theta_{enc}^{(0)}, \theta_{dec}^{(0)}\right\}$$



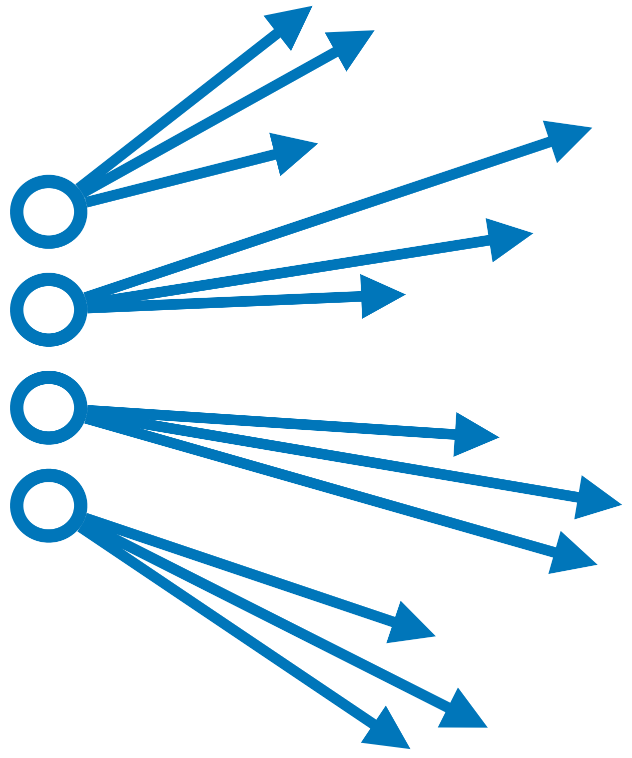
$$\left\{\theta_{enc}^{(0)}, \theta_{dec}^{(0)}\right\}$$

We need an additional control channels to update the system parameters

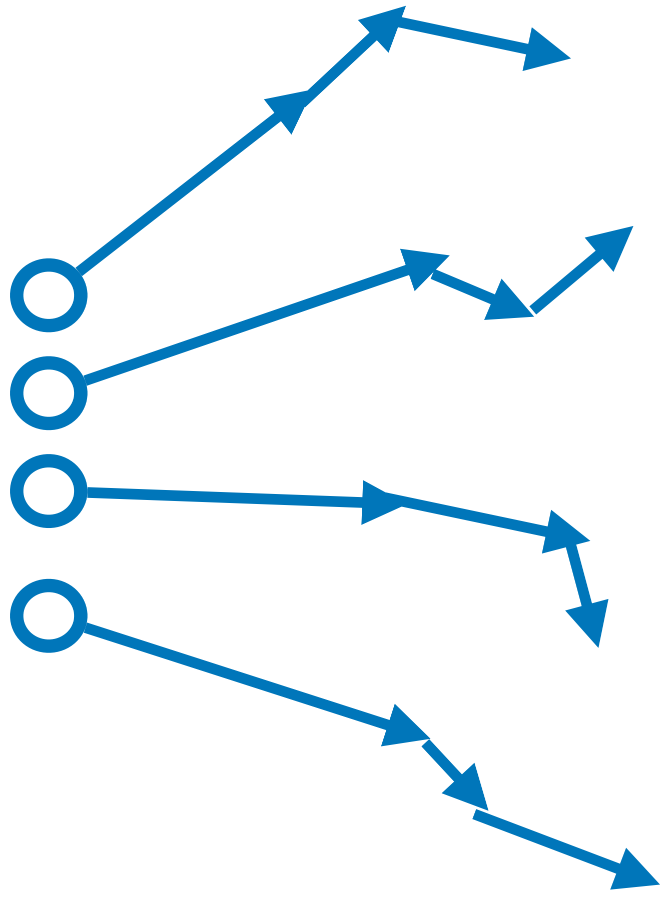
- Adaptation to CSInet for on-line learning

$$\left\{\theta_{enc}^{(0)}, \theta_{dec}^{(0)}\right\}$$

$$\left\{\theta_{enc}^{(0)}, \theta_{dec}^{(0)}\right\}$$



$$\left\{\theta_{enc}^{(0)}, \theta_{dec}^{(0)}\right\}$$

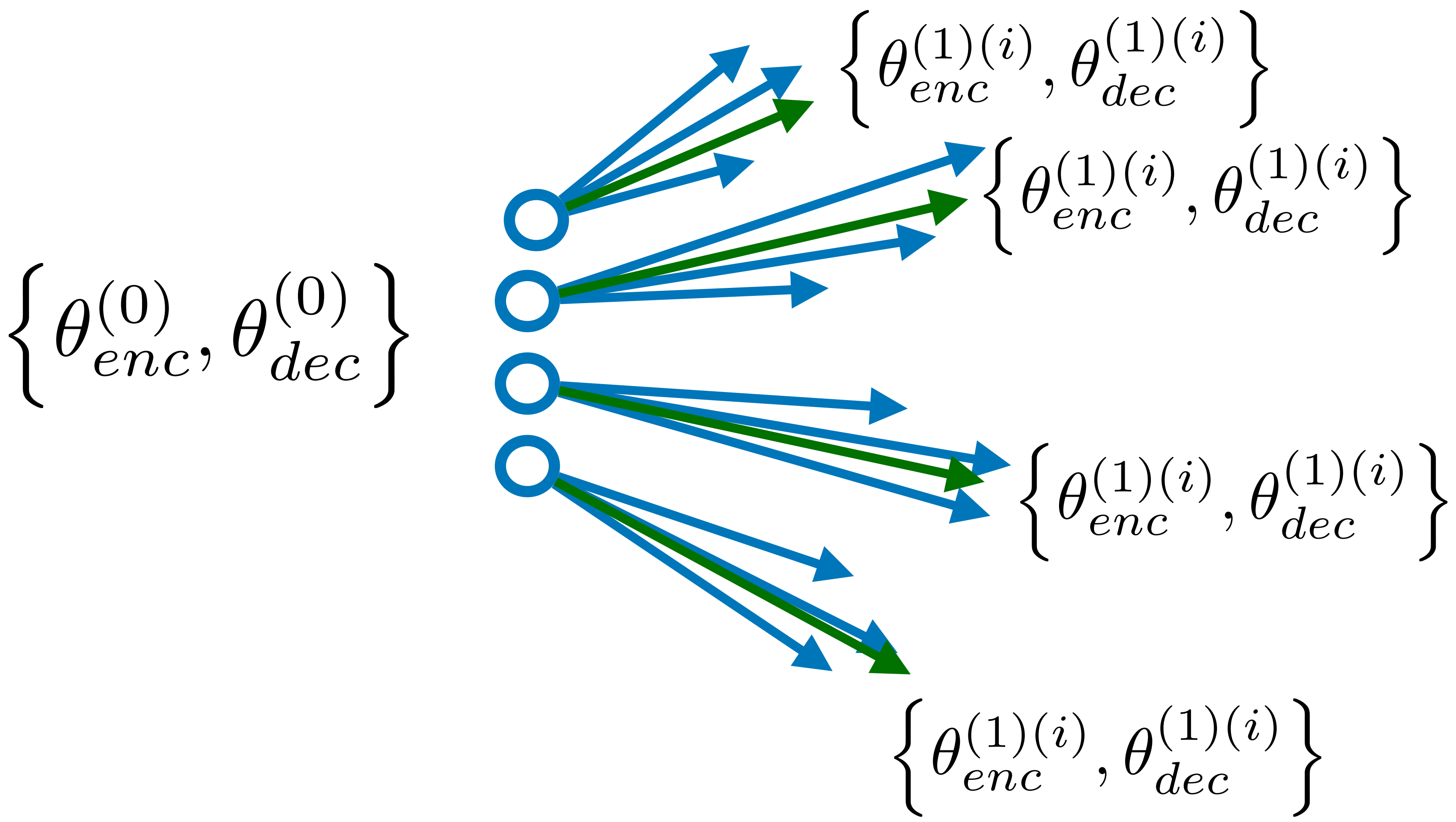


Learning strategies :

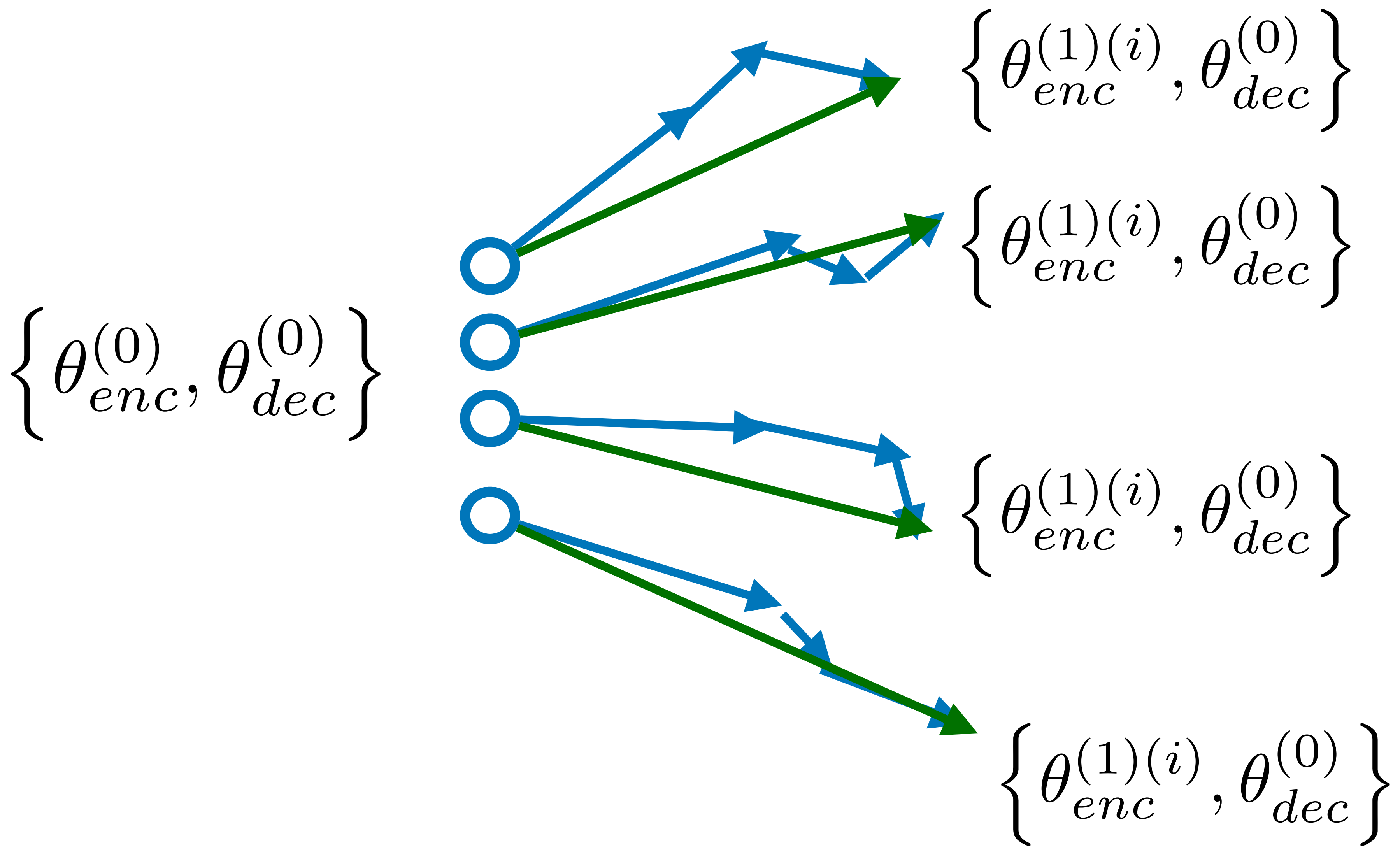
- **learn both parameter sets, without immediate use**
- **Learn encoder parameters, under fixed decoder**

- Adaptation to CSInet for on-line learning

$$\left\{\theta^{(0)}_{enc}, \theta^{(0)}_{dec}\right\}$$



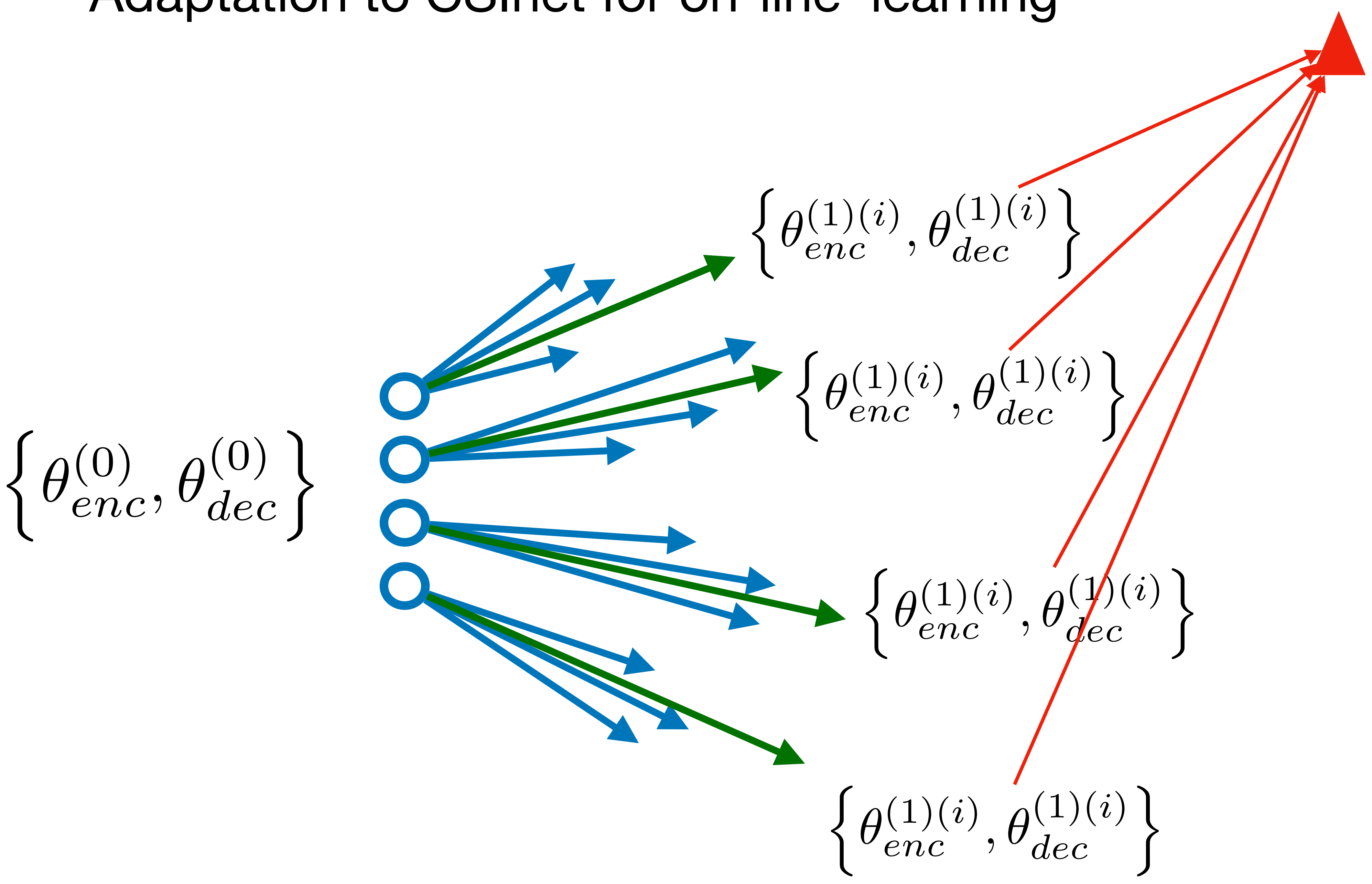
Averaged model or gradient

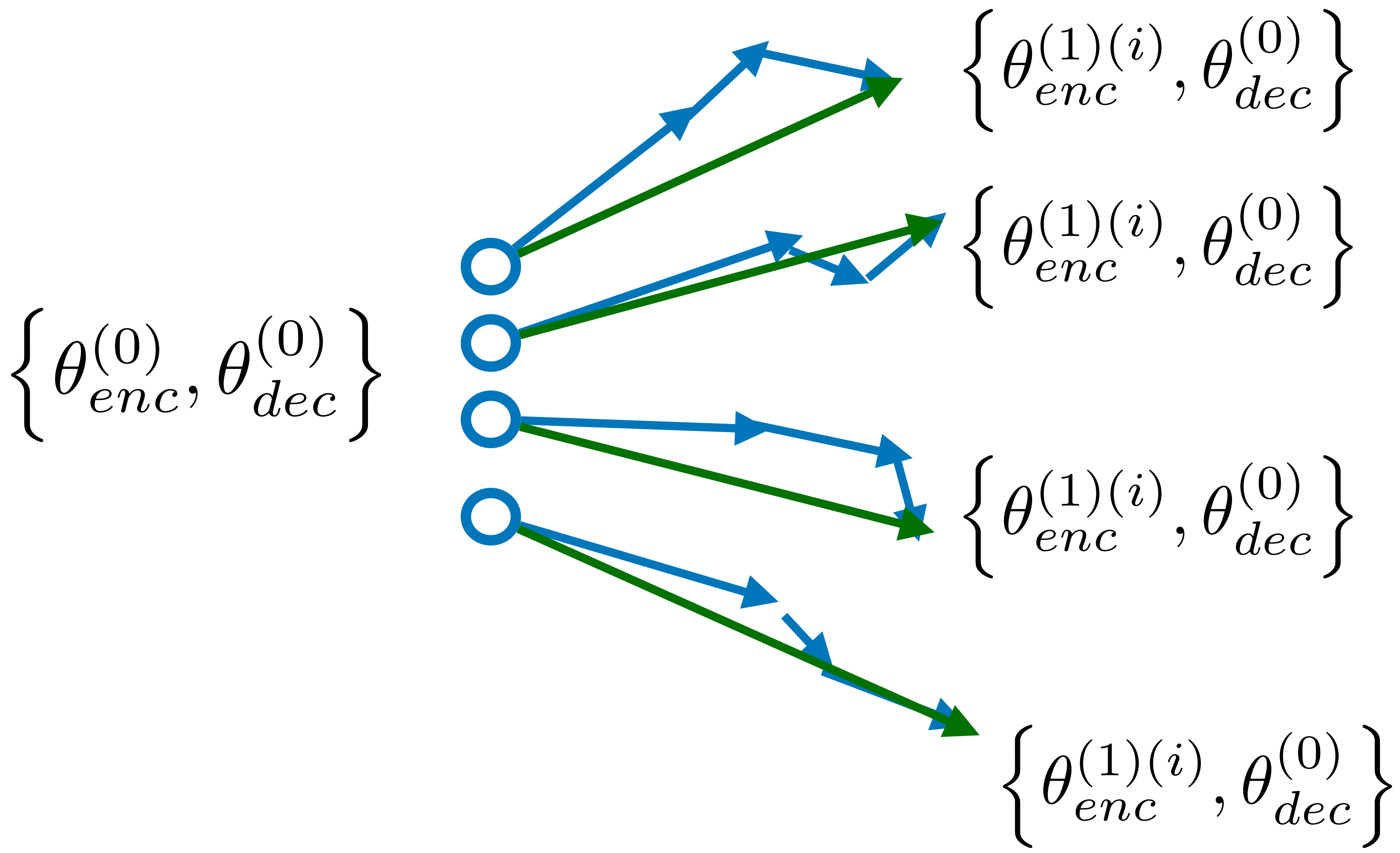


sum-gradient

- Adaptation to CSInet for on-line learning

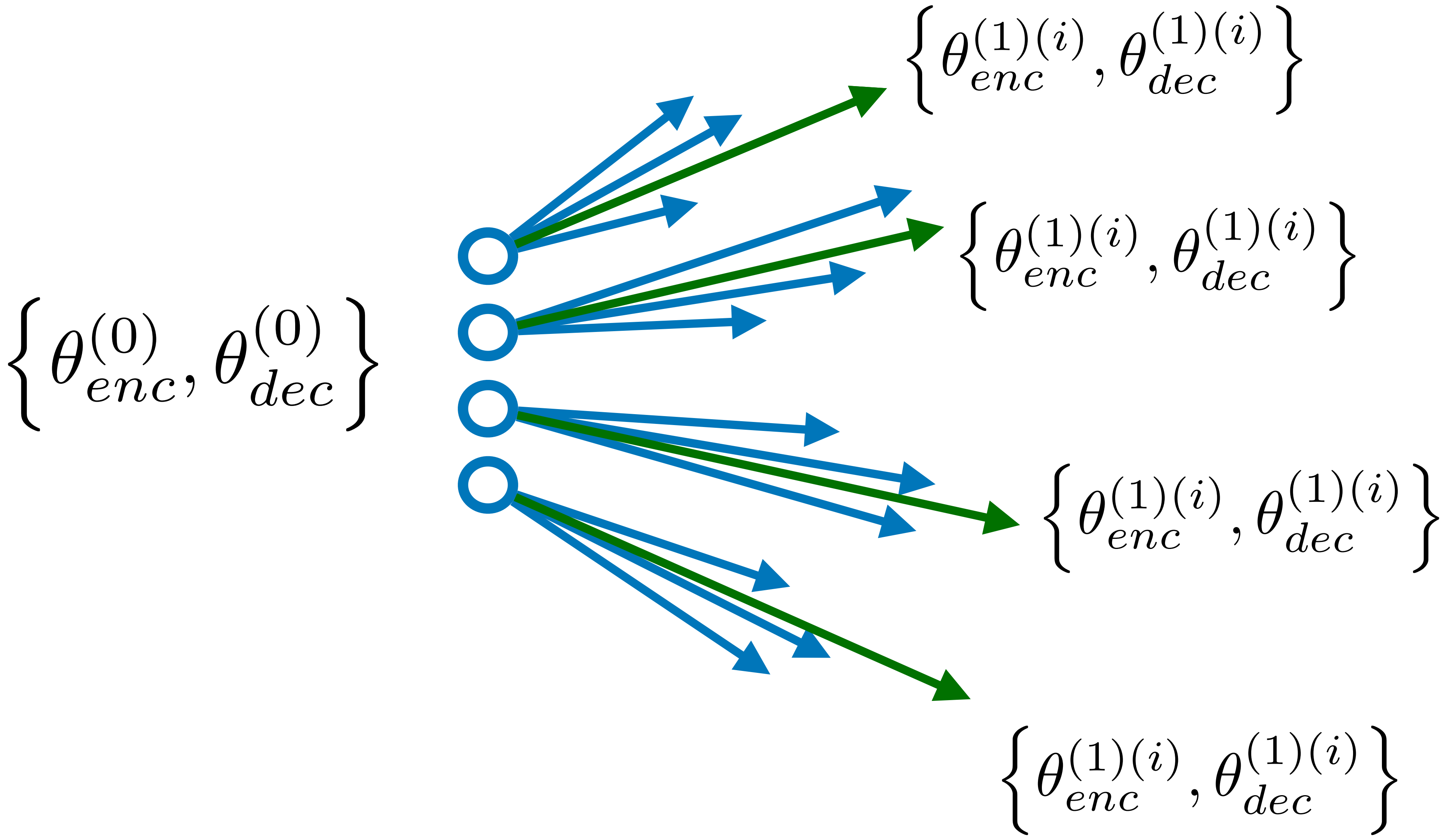
$$\left\{\theta_{enc}^{(1)}, \theta_{dec}^{(1)}\right\} = \frac{1}{N}\sum_i \left\{\theta_{enc}^{(1)}(i), \theta_{dec}^{(1)}(i)\right\}$$

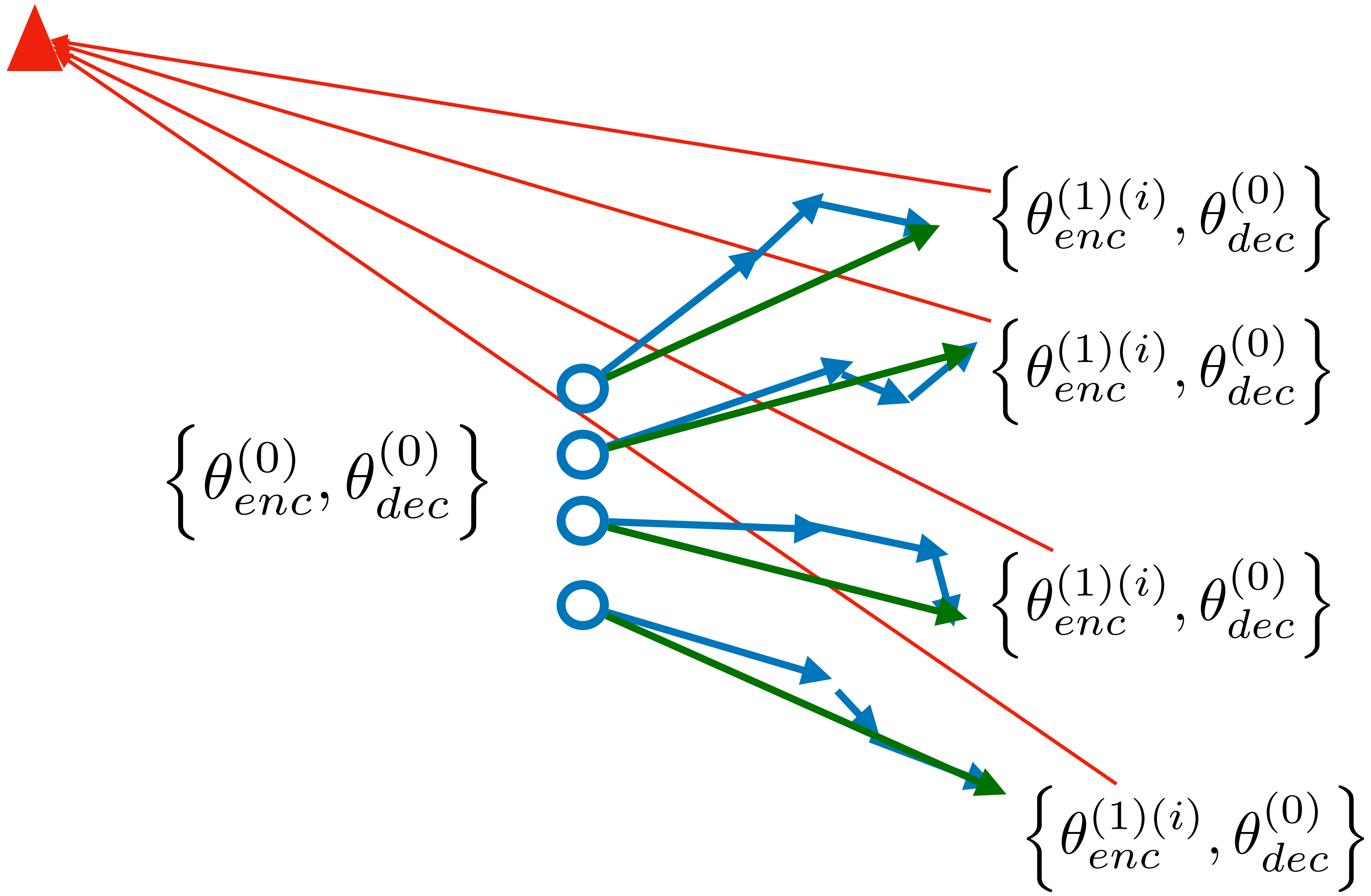




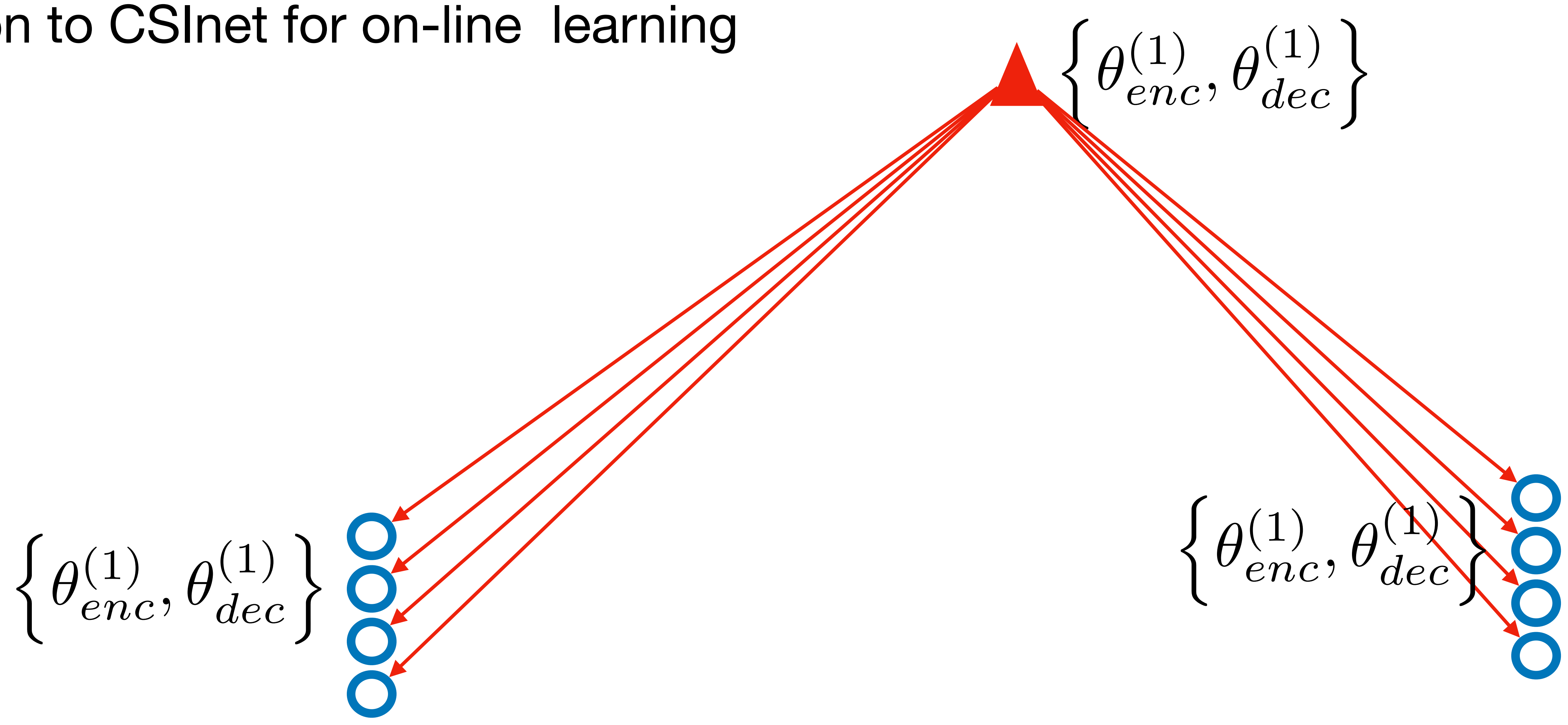
$$\left\{\theta^{(1)}_{enc}, \theta^{(1)}_{dec}\right\} = \left\{\frac{1}{N}\sum_i \theta^{(1)}_{enc}(i), \theta^{(1)}_{dec}(i) = \theta^{(0)}_{dec}(i) + \frac{1}{N}\sum_i \Delta\theta_i\right\}$$

- Adaptation to CSInet for on-line learning





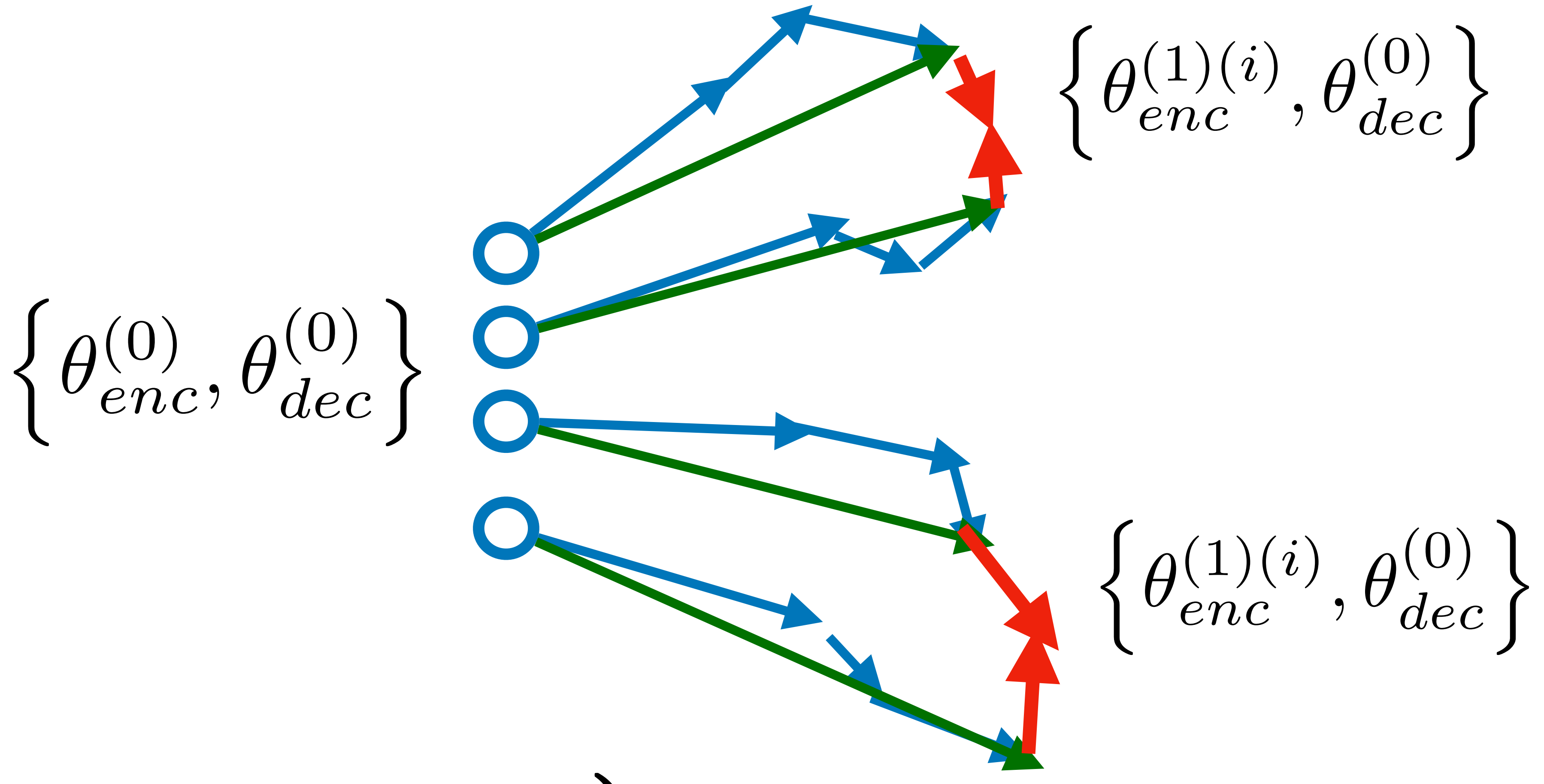
- Adaptation to CSInet for on-line learning



- Adaptation to CSInet for on-line learning

  with personalization

  - Determine clusters

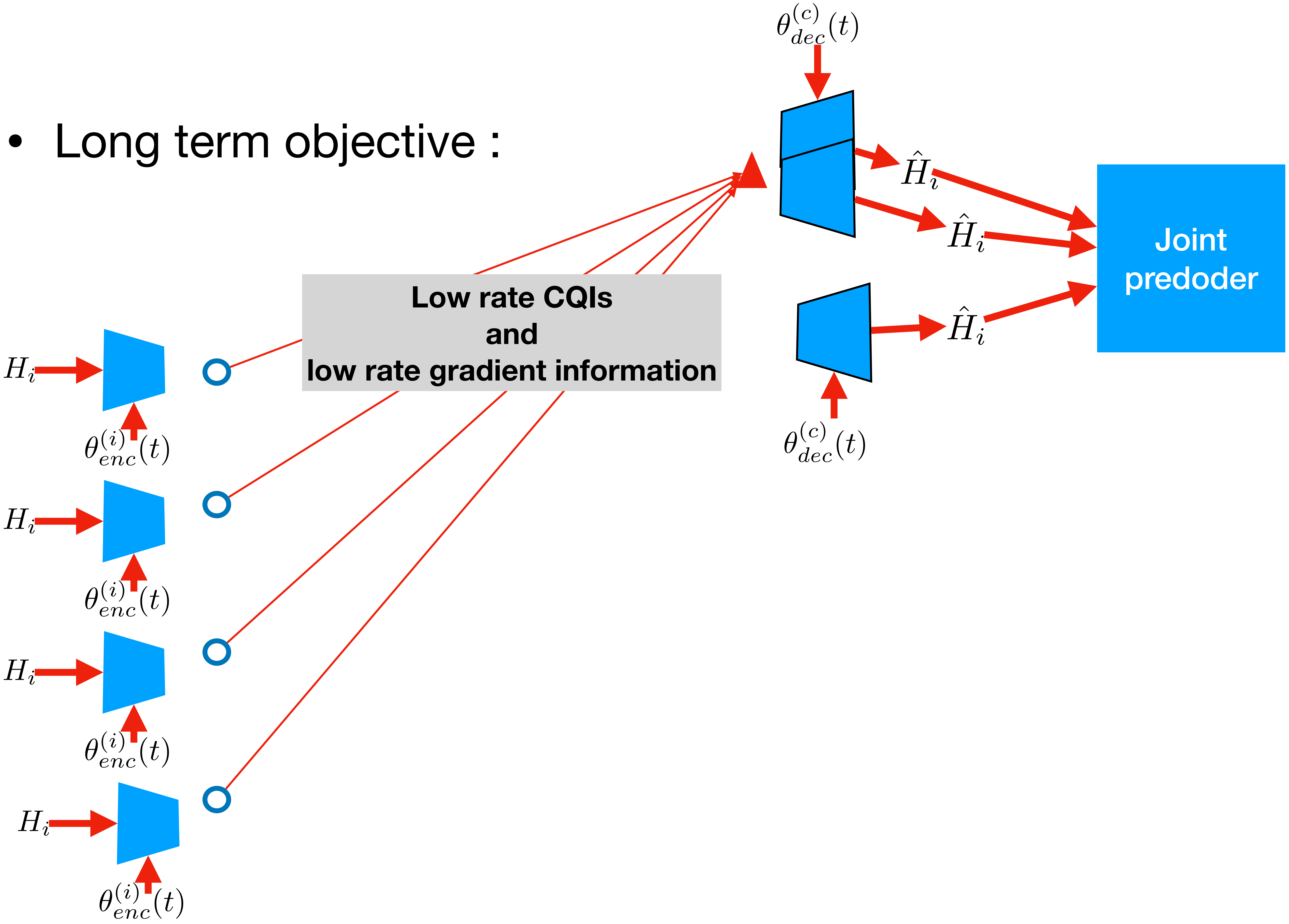


  - Détermine personalized models :

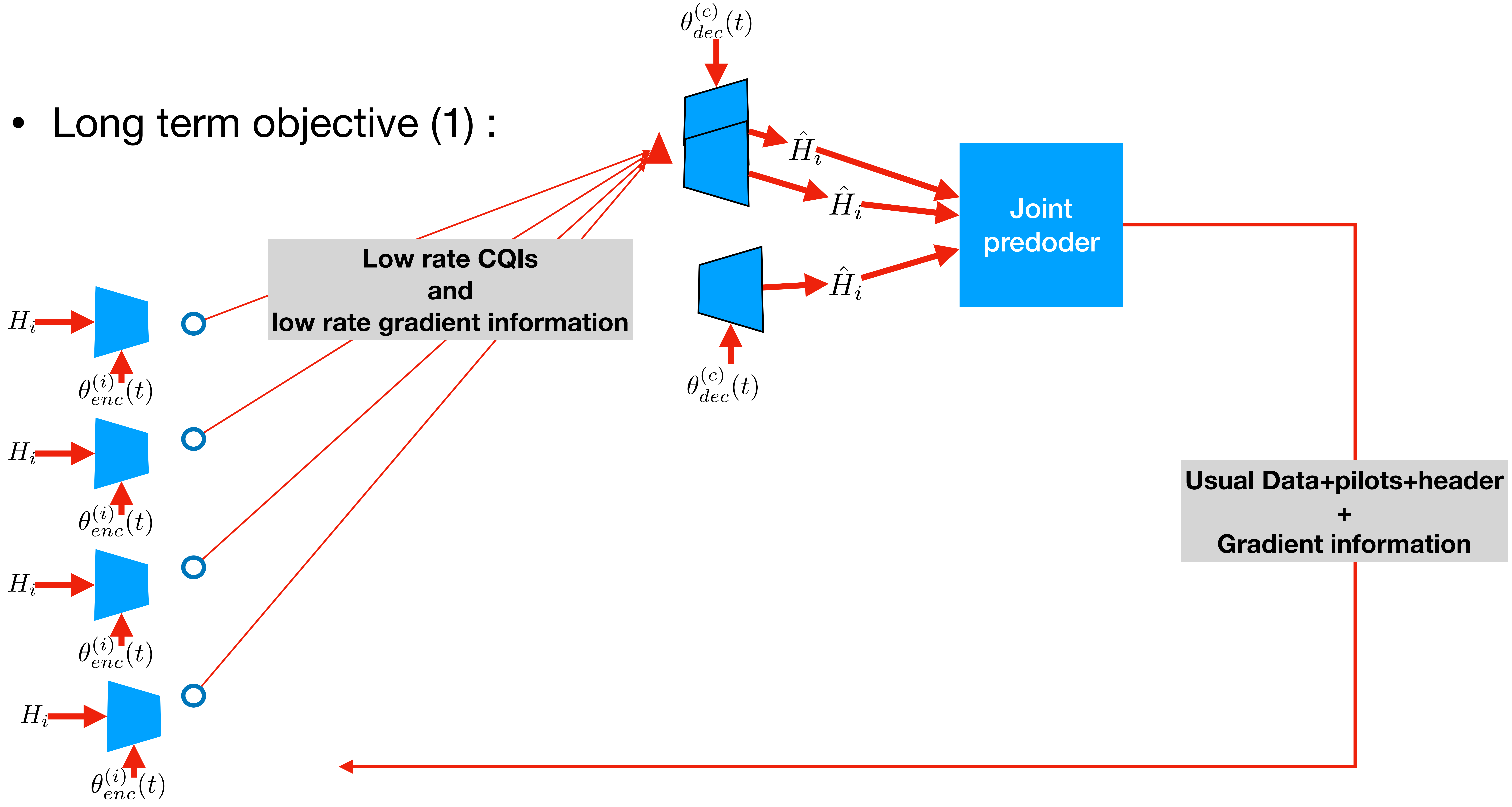
$$\left\{\theta_{enc}^{(1)}, \theta_{dec}^{(1)}\right\}_{c\in\mathcal{C}} = \left\{\frac{1}{N_c}\sum_{i\in\mathcal{I}(c)} \theta_{enc}^{(1)}(i), \theta_{dec}^{(1)}(i) = \theta_{dec}^{(0)}(i) + \Delta\theta\right\}$$

  - Update models in the cell

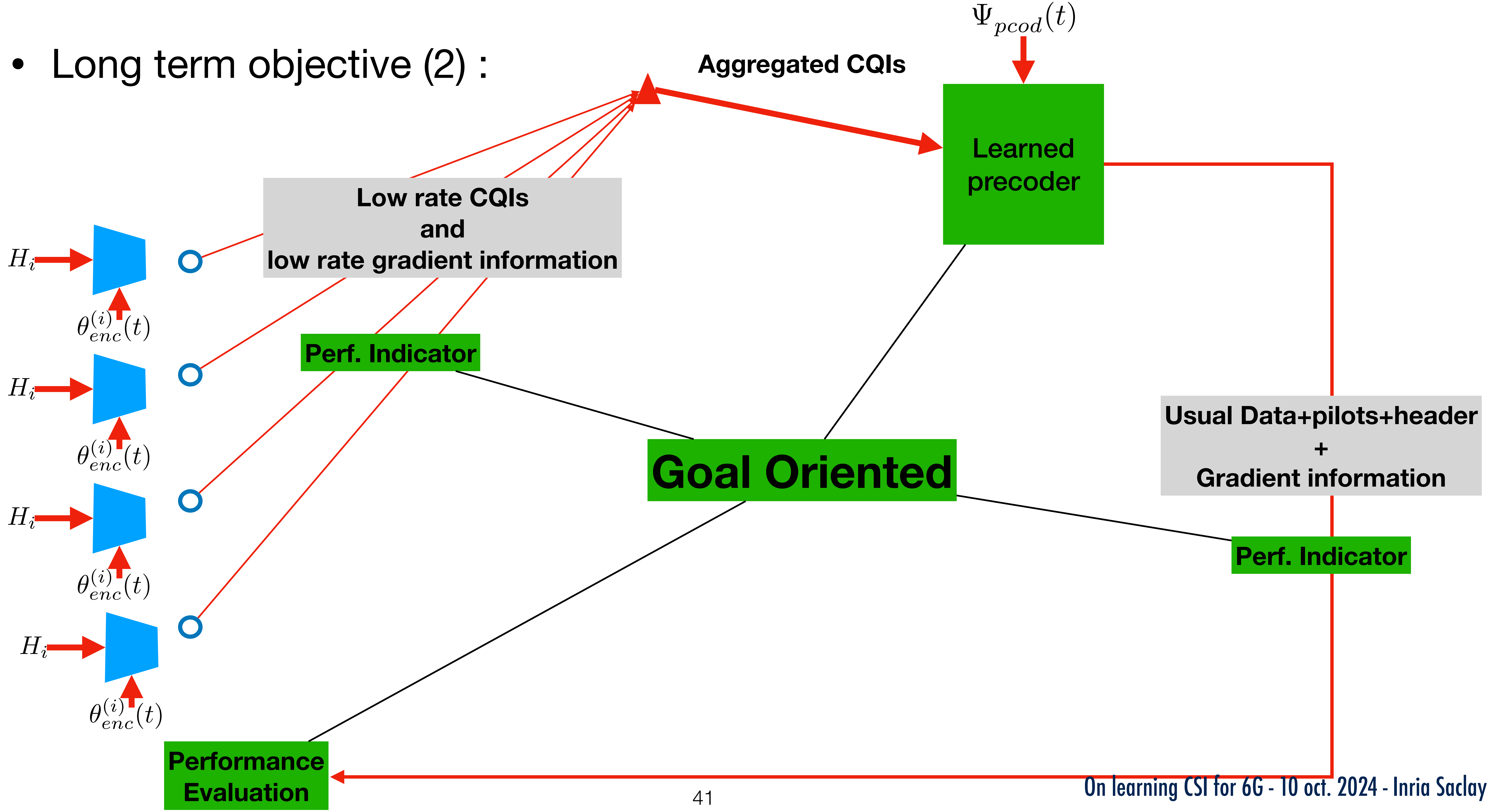
- Long term objective :



- Long term objective (1) :



- Long term objective (2) :



Research trends

- The problem is not just about using distributed learning, but ho to combine and optimize learning, communication and feedbacks.

- Perf. metrics are multiple :

  - Service (rate, delay, …) versus costs (bandwidth resources, energy, computational load)

- Finding tradeoff between generalization (global optimal algorithms) versus personalization

  - Is personalization interesting to reduce model size, communication costs and complexity ?

5- (few) take away messages

- CSI is a critical point to operate future wireless networks
  - Avoid to track a full CSI everywhere : not necessary, resource consuming.
  - Non-coherent approaches may embed channel information, more efficiently than pilots but the decoding is more complex.
  - Joint multi-dimensional processing (time/frequency/antennas) is theoretically more efficient but introduce higher processing complexity.
  - Extensions in cell-free and with IRS is fundamental to achieve expected high performance.