CONNECT
COmmunicatioN-aware dyNamic Edge CompuTing

Sinem Coleri, Koc University
Deniz Gunduz, Imperial College London
Mehdi Bennis, University of Oulu

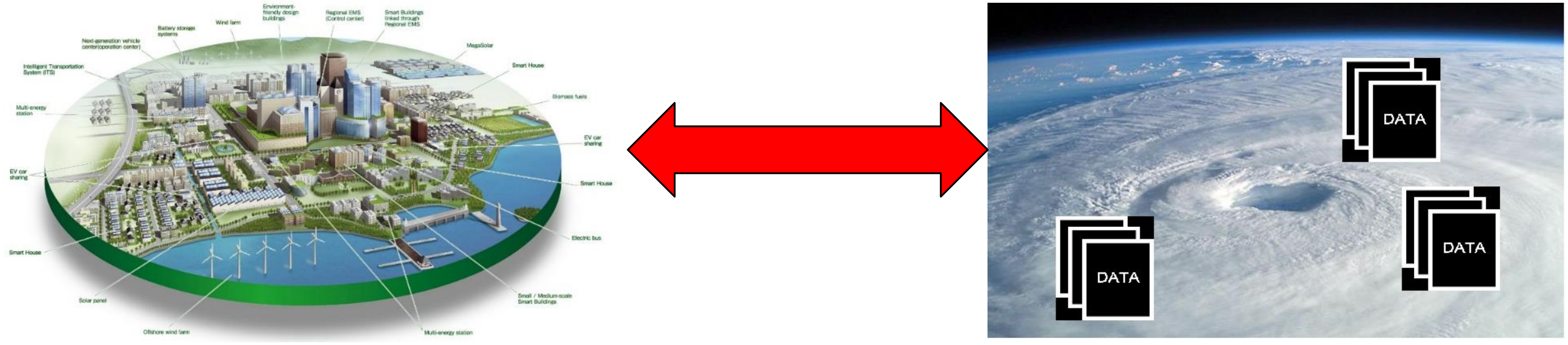
Why Machine Learning in B5G/6G?

- Communication and autonomy are **intertwined** – **ML/AI integral component of B5G/6G!**

- Communication must be ***self-sustaining*** – **system must operate itself by itself: No rigid and human-made protocols!**

- Communication should be ***proactive*** – **learn the user, system, machines dynamics/preferences, etc (beyond basic caching)!**

Massive Data


Environment-friendly design buildings
Regional EMS (Control center)
Smart Buildings linked through Regional EMS
Wind farm
Battery storage systems
Next-generation vehicle center(operation center)
MegaSolar
Smart House
Intelligent Transportation System (ITS)
Biomass fuels
Multi-energy station
EV car sharing
EV car sharing
Smart House
Electric bus
Smart House
Small / Medium-scale Smart Buildings
Solar panel
Offshore wind farm
Multi-energy station
DATA
DATA
DATA

- Massive volumes of data will be generated from sensors, wearables, surveillance cameras, etc.
- ***How to make sense of this hurricane of chaos?***
- Analyze, predict, and adapt to the wireless user behavior, user-centric wireless (e.g., virtual reality)

Heterogeneous Sources of Data

Devices

Smartphones

Applications

Wearables

Smart meters

Vehicles

Geo-location

Mobility patterns

GPS data

People/social

People links

Social networks

Demographics, preferences, etc.

Monitoring

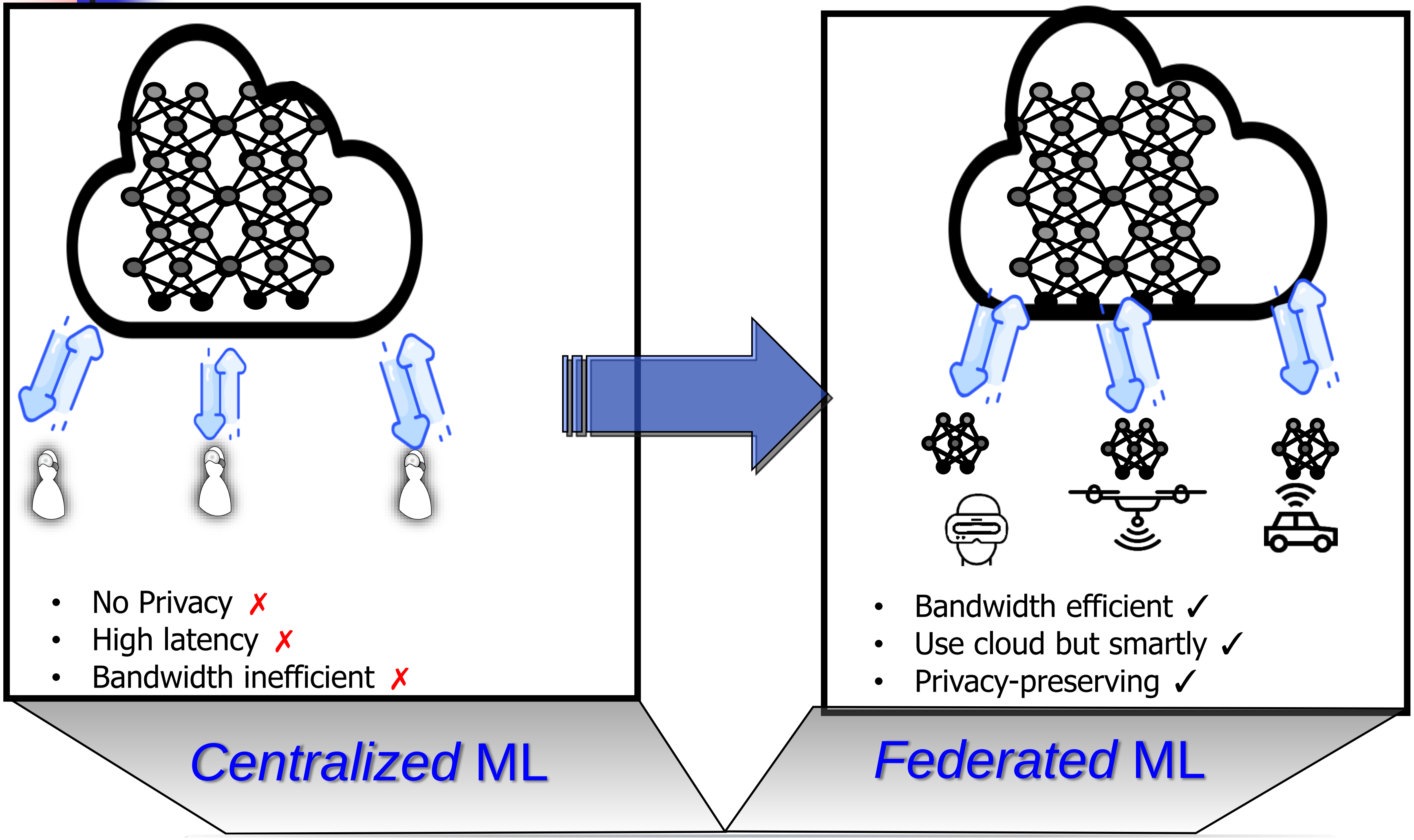
Drones

Surveillance camera

Air/water quality monitors

- **ML to combine to understand needs of wireless users, machines!**

Centralized → Federated ML



- No Privacy ✗
- High latency ✗
- Bandwidth inefficient ✗

- Bandwidth efficient ✓
- Use cloud but smartly ✓
- Privacy-preserving ✓

Centralized ML

Federated ML

Project Objectives

- Develop novel caching, distributed computing and networking methodologies to enable federated/distributed learning taking into account network dynamics

- Apply developed joint computing, caching and communication framework to a hierarchical heterogeneous architecture for vehicular ad-hoc networks
  - Large-scale simulations
  - Small scale implementation platform consisting of two cars and a roadside unit at Koc University

Work Package 2

- WP2. Distributed ML at the Edge: This WP focuses on distributed/ federated ML considering both training and inference at the network edge, targeting heterogeneous vehicular use cases.
  - *T2.1. Distributed learning at the wireless edge* (1-18): This task focuses on distributed training among moving edge devices taking into account per-device data sample size, network topology, connectivity, latency and reliability constraints.
  - *T2.2. Distributed and reliable computing over-the-air* (12-36): This task investigates the fundamental problem of stragglers which can adversely undermine the learning process.
  - *T2.3. Hierarchical heterogeneous networking architecture for distributed learning* (1-24): This task focuses on the design of smart clustering algorithms (for the network nodes) and efficient handover mechanisms considering data generation characteristics and ultra-high reliability and low latency constraints of distributed ML.

Work Package 3

- WP3. Energy and age-optimal IoT data caching across dynamic networks**:** This WP will focus on where, when and how much sensor data must be stored across a dynamic network under storage and communication constraints. Particular attention will be paid to the "age of information" as it will identify the relevance of data for various learning tasks.
  - *T3.1. Coded caching of data and computations under energy and bandwidth constraints* (6-18): This task focuses on the design of novel coded caching and delivery techniques focusing on high volume sensor data, mainly targeting computation tasks, which may benefit from caching computations.
  - *T3.2. Reinforcement learning (RL) for age-optimal caching and delivery of sensor data across a dynamic network* (12 – 36): This task focuses on the design of distributed collaborative reinforcement learning (RL) algorithms to optimize the caching of sensor data and offloading of computation tasks.
  - *T3.3. Age-optimal caching and delivery over hierarchical heterogeneous networks* (19– 24): This task will extend the caching and data access framework for distributed learning and computing developed in Tasks 3.1 and 3.2 to hierarchical and highly dynamic networks, such as VANETs.

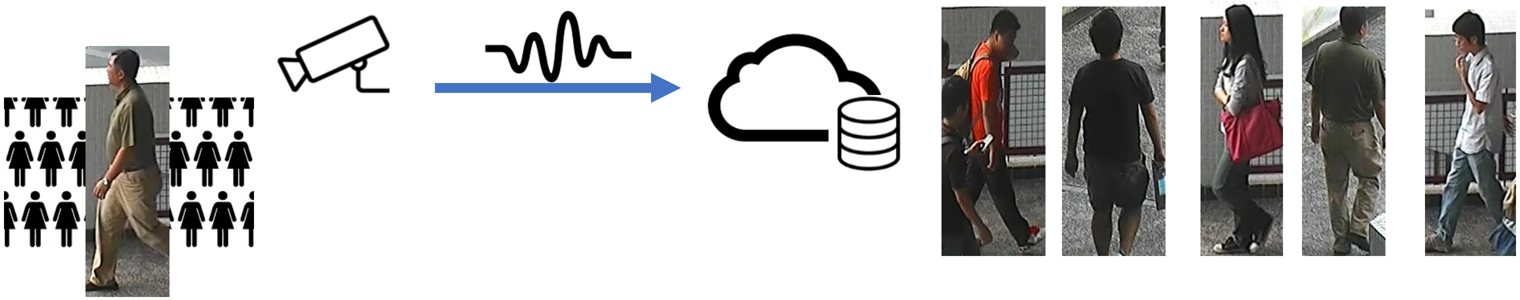
Work Package 4

- WP4. Software platform and testbed implementation**:** This WP will focus on building the collaborative software platforms for numerical simulations as well as the experimental testbed.
  - *T4.1. Software platform for numerical simulation of distributed learning, caching and networking algorithms* (1 – 36): This task will develop the software platform where all the distributed learning and caching algorithms will be tested.
  - *T4.2. Collection of requirements for caching and computing application* (1 – 6): This task focuses on the collection of communication and data generation characteristics of distributed computing and caching.
  - *T4.3. Wireless channel characterization for mobile VANETs* (7 – 18): This task focuses on the wireless channel modeling for VLC, IEEE 802.11p and cellular systems among moving vehicles and road side units under varying scenarios.
  - *T4.4. Designing a hybrid communication architecture for mobile VANETs* (19 – 30): This task focuses on the integration of VLC, IEEE 802.11p and cellular technologies on a hybrid transceiver architecture, targeting learning applications.
  - *T4.5. Validation of vehicular networking for caching and computing applications* (31 –36): This task focuses on the testing and validation of the performance of the developed algorithms on moving vehicles under various scenarios.

Imperial Team

- Extensive expertise in
    - Machine learning
    - Wireless communications
    - Information and coding theory
    - Optimization theory
    - Privacy / Security

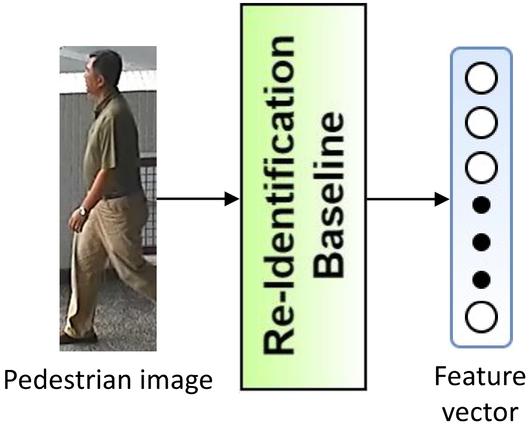
Image Retrieval at the Edge



Goal: match a pedestrian’s image from a wireless camera with another image in a large database

Standard approach:

- Transmit images to the cloud
- Determine features most relevant for re-identification over image database
- Re-ID baseline: Deep convolutional neural network, e.g., ResNet-50



Pedestrian image

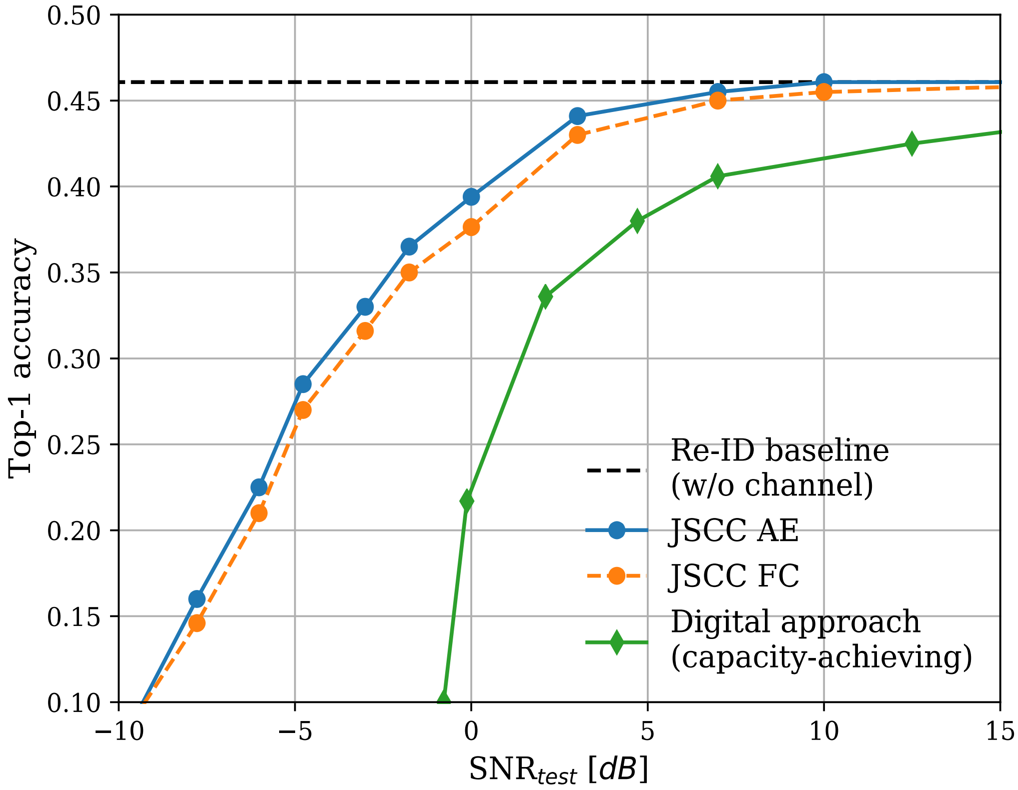
Re-Identification Baseline

Feature vector

Jankowski, Gunduz and Mikolajczyk, **Wireless image retrieval at the edge**, *IEEE Journal on Selected Areas in Communications (JSAC),* Jan. 2021.

Person Re-identification over Noisy Channels

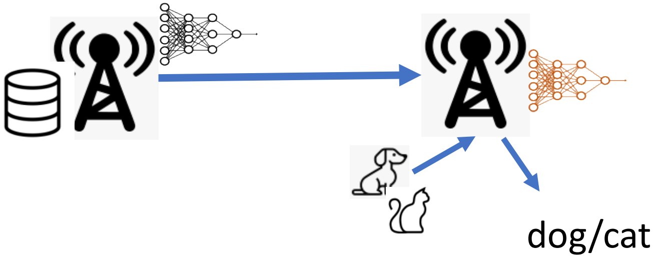
B = 128 channel uses



- CUHK03 dataset: 14096 images of 1467 identities taken from two camera views.
- 256x128 coloured images

Jankowski, Gunduz and Mikolajczyk, **Wireless image retrieval at the edge**, *IEEE Journal on Selected Areas in Communications (JSAC),* Jan. 2021.

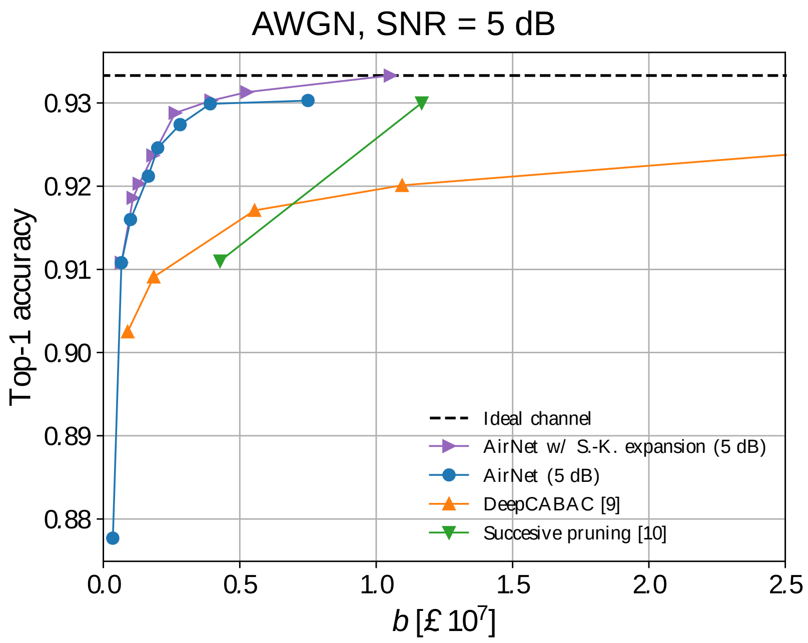
Neural Networks in the Air (AirNet)

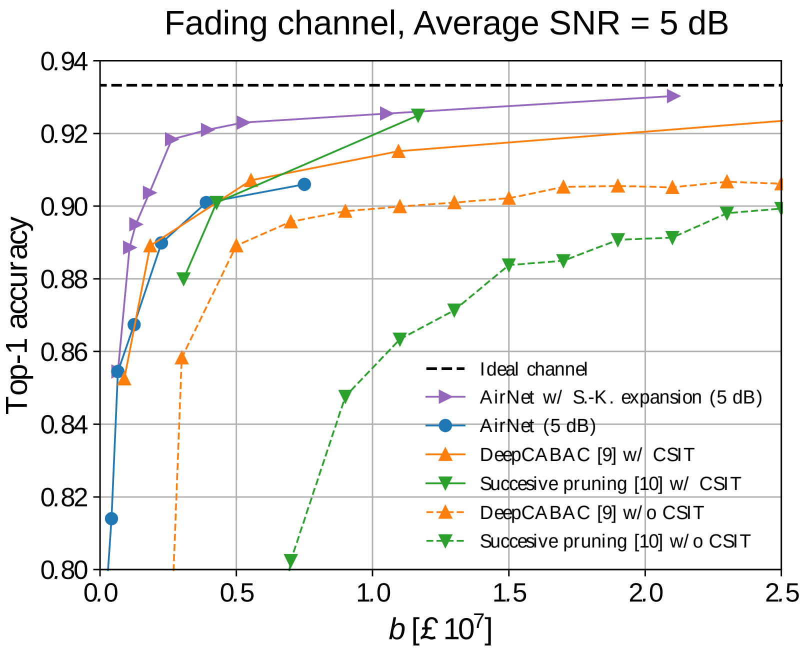


- Neural networks (NNs) may need to be transmitted, or stored over lossy 'analog' media (neuromorphic hardware)
- Classification will be done with the noisy NN
- Conventional approach: Compress NN weights, use channel coding against errors
- Proposed approach: Pruning (for bandwidth reduction) + noise injection during training + knowledge distillation

Jankowski, Gündüz, and Mikolajczyk, **AirNet: Neural network transmission over the air**, *arXiv*, 2021.

AirNet over AWGN Channels



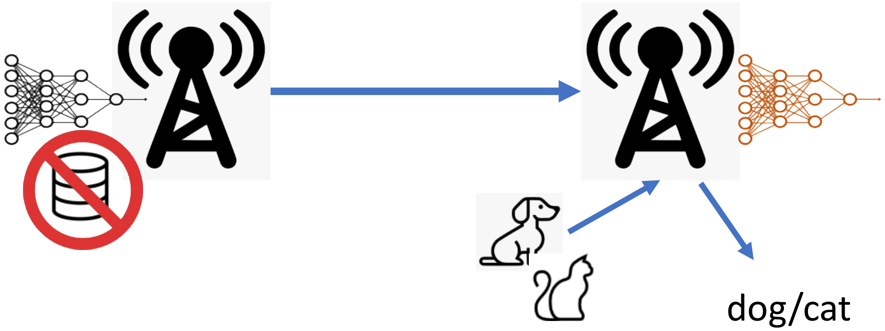


- Small-VGG16 for CIFAR-10 classification
- Observation: Better to prune more, then introduce redundancy through SK mapping

Jankowski, Gündüz, and Mikolajczyk, **AirNet: Neural network transmission over the air**, *arXiv*, 2021.

Denoising Noisy Neural Networks

$$\boldsymbol{r} = \boldsymbol{w} + \boldsymbol{z}$$



- w: neural network parameters
- z: independent Gaussian noise vector
- No data is available for training
- ML estimate: $\hat{\boldsymbol{w}}^{ML} = \boldsymbol{r}$
- Bayesian estimation
  - Assume iid Gaussian prior for network parameters: $W \sim \mathcal{N}(W; \mu_w, \sigma_w^2)$
    - Sample mean:

$$\mu_w = \frac{1}{d}\sum_{i=1}^{d} w[i]$$

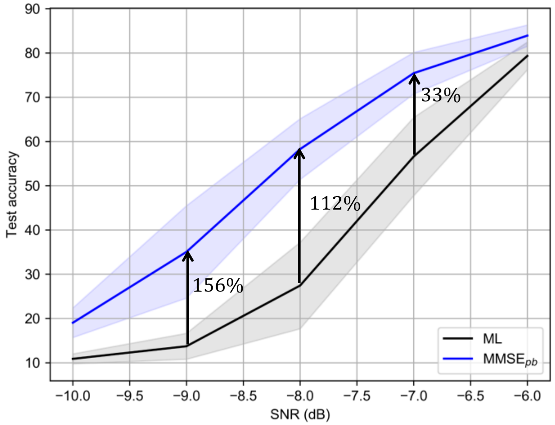
    - Sample variance:

$$\sigma_w^2 = \frac{1}{d}\sum_{i=1}^{d} (w[i] - \mu_w)^2$$

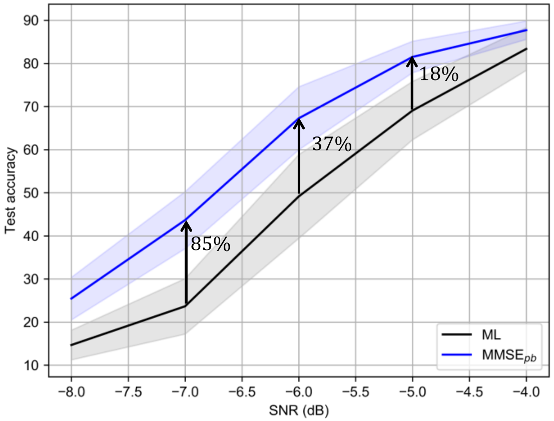
Shao, Liew and Gunduz, **Denoising Noisy Neural Networks: A Bayesian Approach with Compensation**, arXiv, May 2021.

Denoising Noisy Neural Networks

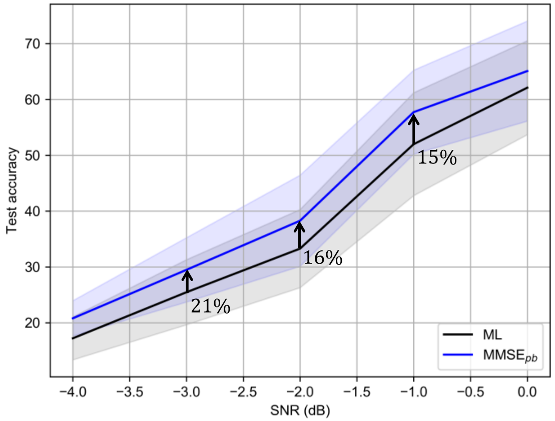
RestNet34 (CIFAR-10)



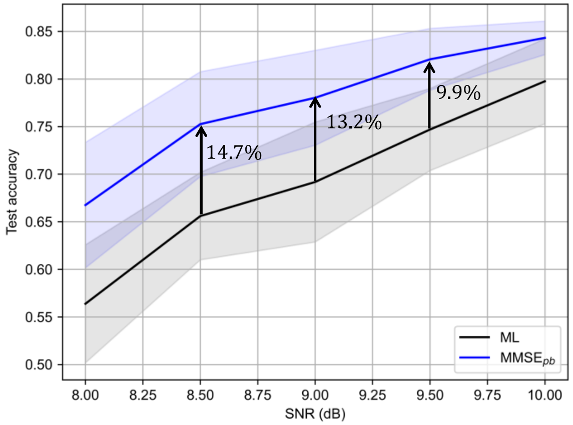
RestNet18 (CIFAR-10)



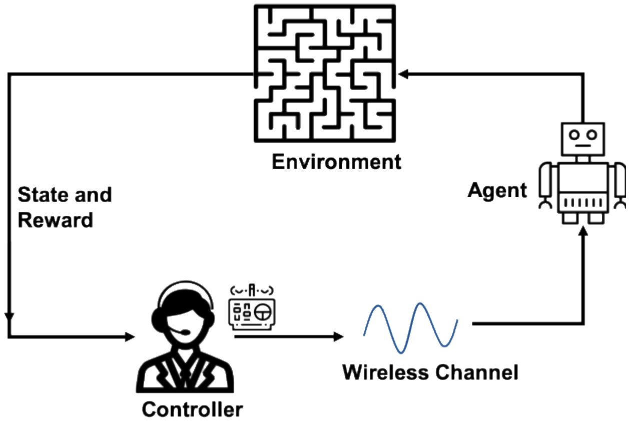
ShuffleNet V2 (CIFAR-10)



BERT (SST-2)



Remote-Controlled Markov Decision Process (MDP)

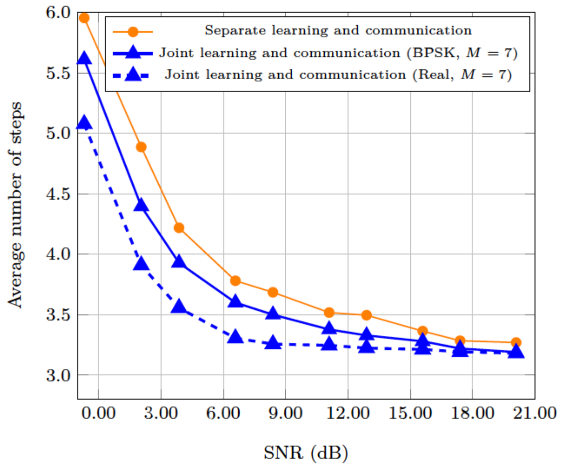


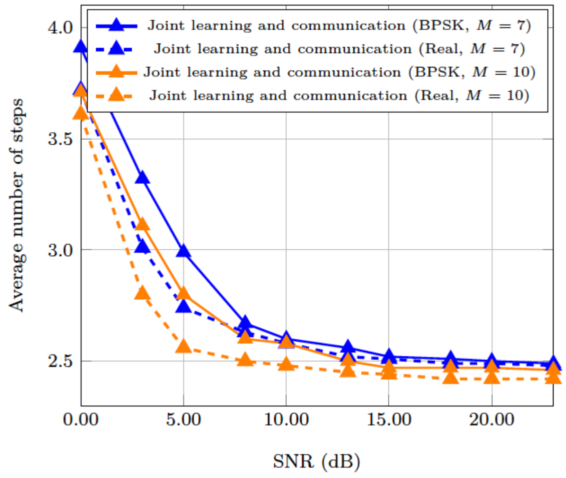
- Create a two-agent version of an MDP
- Controller observes the state and reward, but agent takes the action
- A noisy communication channel in between
- Agent can depend solely on the received signal, or may have some limited observation of the system state

Tung, Kobus, Roig Pujol, Gunduz, **Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning over Noisy Channels**, IEEE JSAC, 2021.

Example: Grid World

- L x L grid world
- Agent can take 16 actions (1 or 2 steps in every direction)
- Arrives at a random neighbouring cell w.p. $\boldsymbol{\delta}$
- Find treasure at a random location as fast as possible
- Channels:
  - Binary input AWGN channel
  - AWGN channel with average power constraint



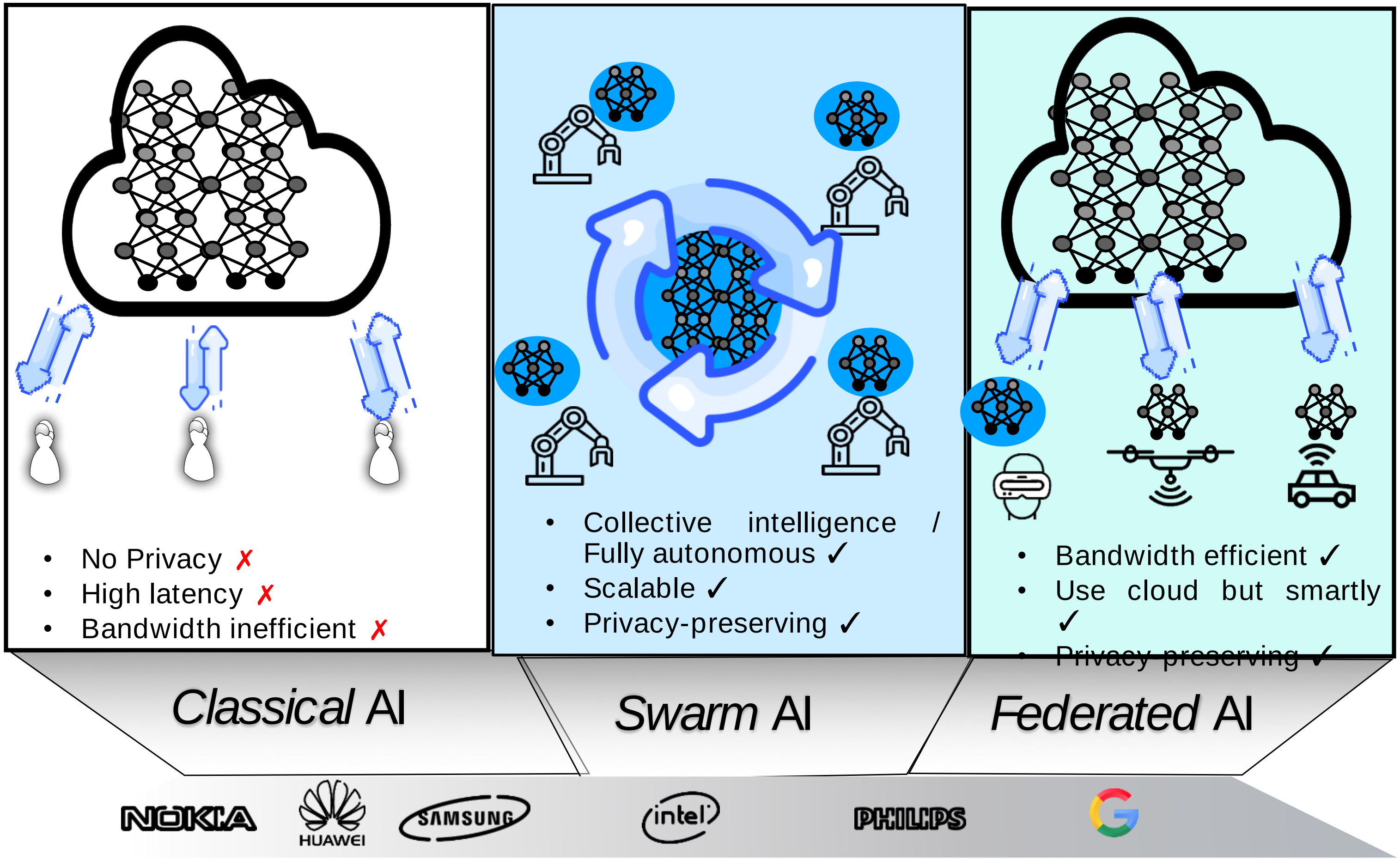


Tung, Kobus, Roig Pujol, Gunduz, **Effective Communications: A Joint Learning and Communication Framework for Multi-Agent Reinforcement Learning over Noisy Channels**, IEEE JSAC, 2021.

Oulu Team

- Extensive expertise in
  - Distributed ML
  - Mobile edge/fog computing
  - URLLC
  - Vehicular communications

Centralized to Federated & Swarm/Distributed ML

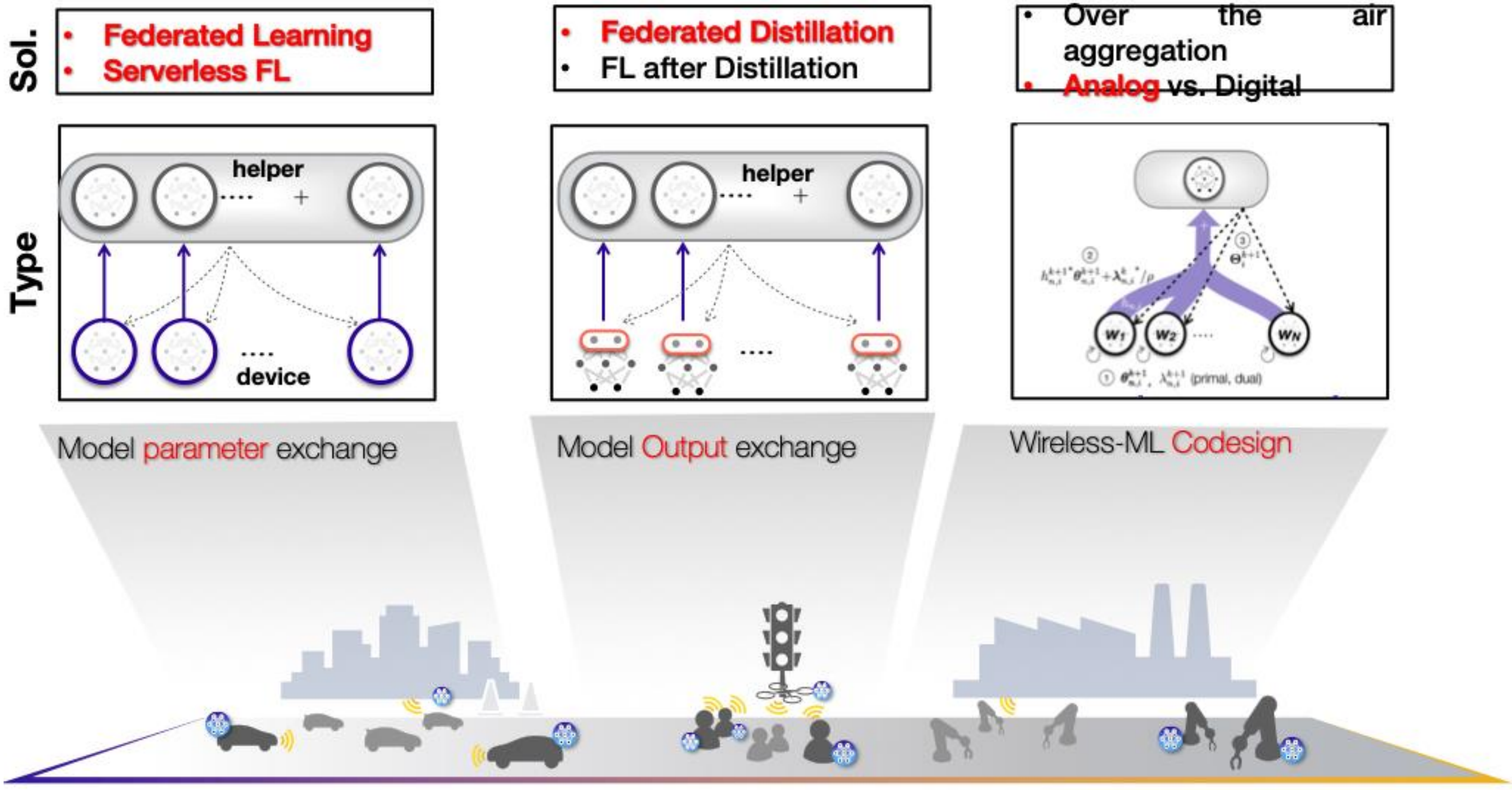


- No Privacy ✗
- High latency ✗
- Bandwidth inefficient ✗

- Collective intelligence / Fully autonomous ✓
- Scalable ✓
- Privacy-preserving ✓

- Bandwidth efficient ✓
- Use cloud but smartly ✓
- Privacy-preserving ✓

Federated Learning



Sol.

- Federated Learning
- Serverless FL

- Federated Distillation
- FL after Distillation

- Over the air aggregation
- Analog vs. Digital

Type

helper
+
....
device
....
helper
+
....
②
$h_{m,i}^{k+1*}\theta_{m,i}^{k+1}+\lambda_{m,i}^{k*}/\rho$
③
Θ_i^{k+1}
w_1
w_2
w_N
① $\theta_{m,i}^{k+1}$, $\lambda_{m,i}^{k+1}$ (primal, dual)

Model parameter exchange

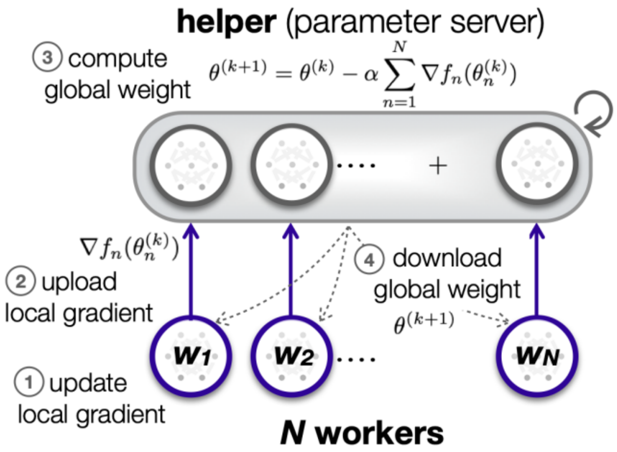
Model Output exchange

Wireless-ML Codesign

Model Training Beyond Parameter Server

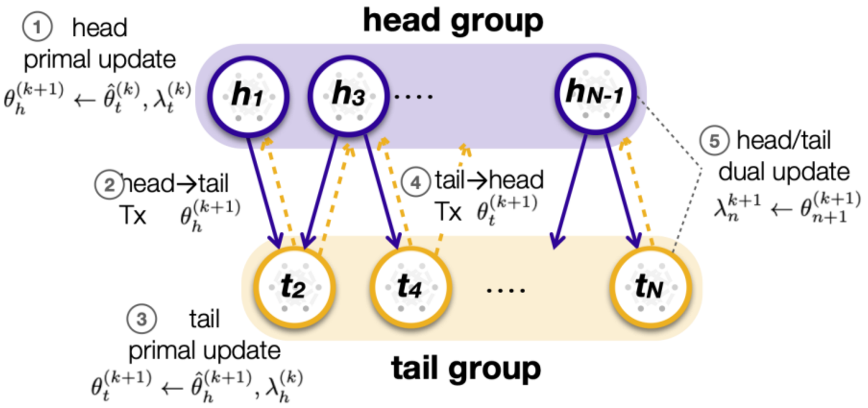
FL (GD based)

$$\underset{\{\theta_n\}}{\text{Minimize}} \quad \sum_{n=1}^{N} f_n(\theta_n)$$
$$\text{s.t.} \quad \theta_n = \theta \quad \forall n$$



GADMM

$$\underset{\{\theta_n\}}{\text{Minimize}} \quad \sum_{n=1}^{N} f_n(\theta_n)$$
$$\text{s.t.} \quad \theta_n = \theta_{n+1} \quad \forall n$$

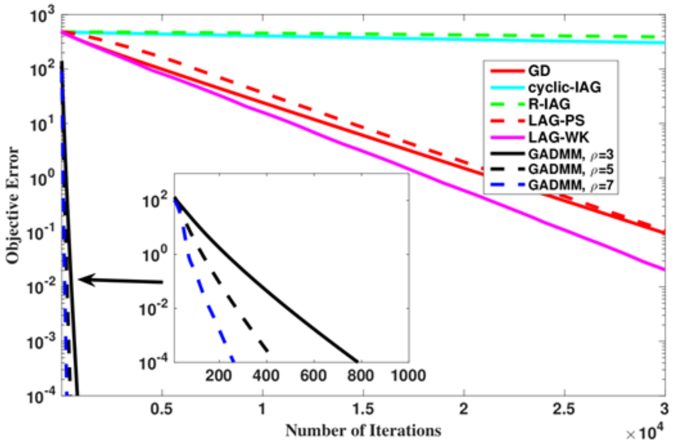


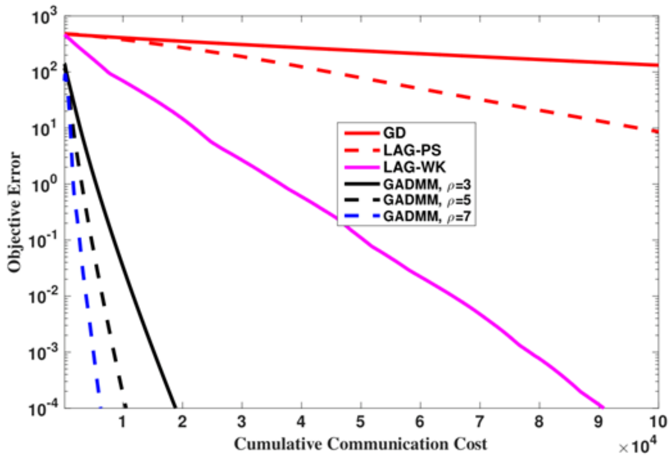
① head primal update
$\theta_h^{(k+1)} \leftarrow \hat{\theta}_t^{(k)}, \lambda_t^{(k)}$
head group
h_1
h_3
....
h_{N-1}
② head→tail Tx $\theta_h^{(k+1)}$
④ tail→head Tx $\theta_t^{(k+1)}$
⑤ head/tail dual update
$\lambda_n^{k+1} \leftarrow \theta_{n+1}^{(k+1)}$
t_2
t_4
....
t_N
③ tail primal update
$\theta_t^{(k+1)} \leftarrow \hat{\theta}_h^{(k+1)}, \lambda_h^{(k)}$
tail group

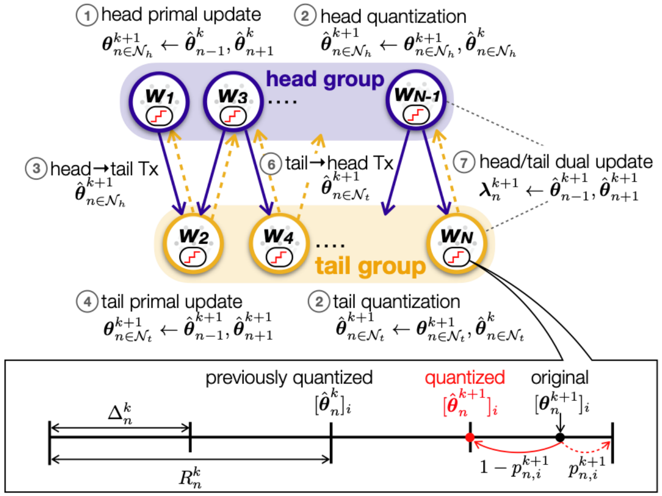
A. Elgabli, J. Park, A. S. Bedi, V. Aggarwal, and M. Bennis, "**GADMM: Fast and Communication Efficient Distributed Machine Learning Framework**," *JMLR20*

GADMM (full precision) and Quantization

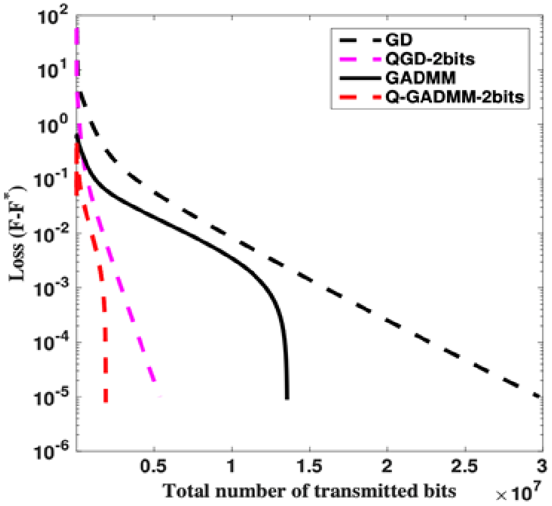
GADMM, Linear Regression



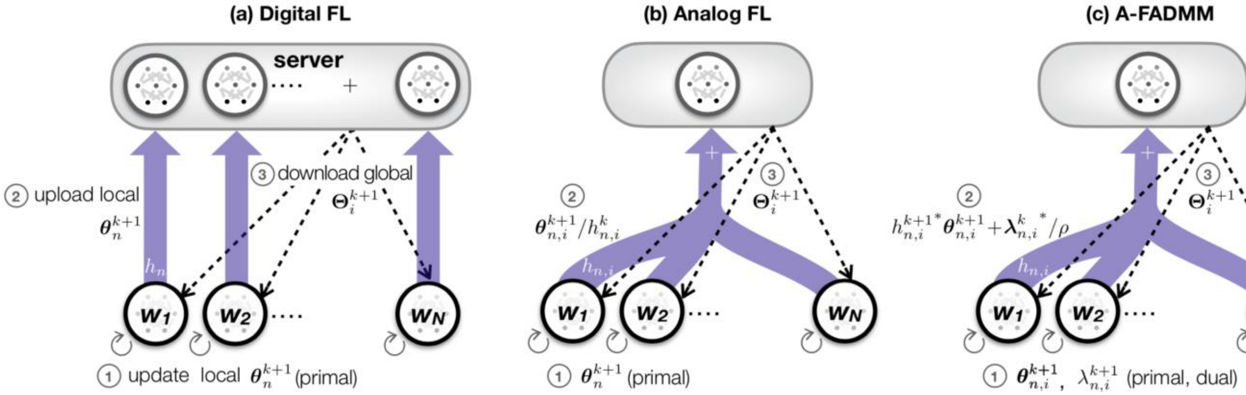


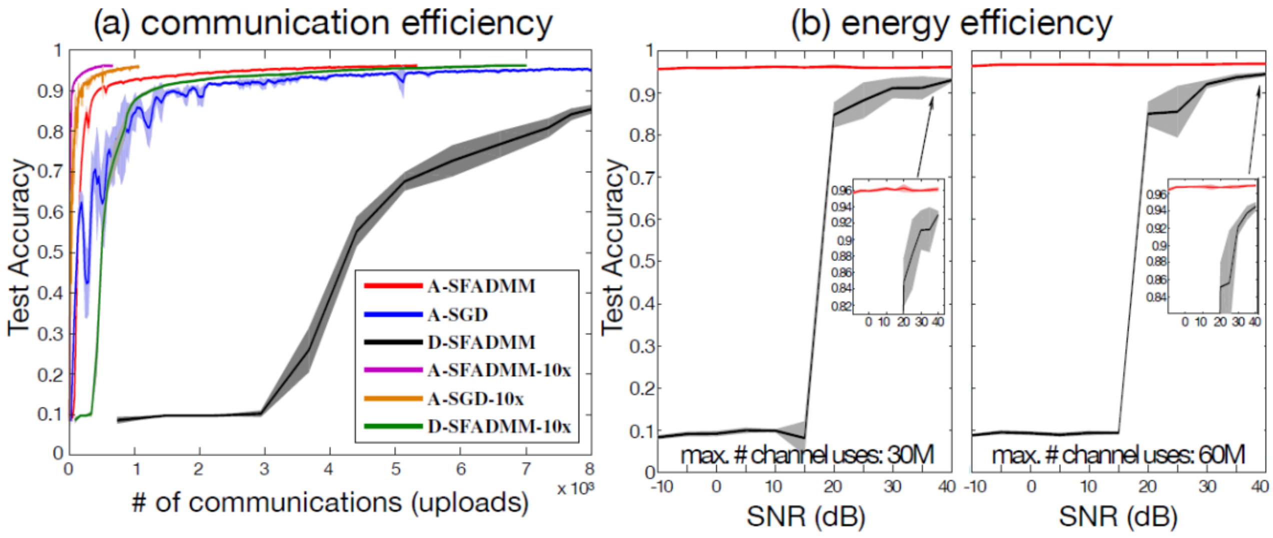


① head primal update
$\theta^{k+1}_{n\in\mathcal{N}_h} \leftarrow \hat{\theta}^k_{n-1}, \hat{\theta}^k_{n+1}$
② head quantization
$\hat{\theta}^{k+1}_{n\in\mathcal{N}_h} \leftarrow \theta^{k+1}_{n\in\mathcal{N}_h}, \hat{\theta}^k_{n\in\mathcal{N}_h}$
W_1
W_3
head group
....
W_{N-1}
③ head→tail Tx
$\hat{\theta}^{k+1}_{n\in\mathcal{N}_h}$
⑥ tail→head Tx
$\hat{\theta}^{k+1}_{n\in\mathcal{N}_t}$
⑦ head/tail dual update
$\lambda^{k+1}_n \leftarrow \hat{\theta}^{k+1}_{n-1}, \hat{\theta}^{k+1}_{n+1}$
W_2
W_4
....
W_N
tail group
④ tail primal update
$\theta^{k+1}_{n\in\mathcal{N}_t} \leftarrow \hat{\theta}^{k+1}_{n-1}, \hat{\theta}^{k+1}_{n+1}$
⑤ tail quantization
$\hat{\theta}^{k+1}_{n\in\mathcal{N}_t} \leftarrow \theta^{k+1}_{n\in\mathcal{N}_t}, \hat{\theta}^k_{n\in\mathcal{N}_t}$
previously quantized
$[\hat{\theta}^k_n]_i$
quantized
$[\hat{\theta}^{k+1}_n]_i$
original
$[\theta^{k+1}_n]_i$
Δ^k_n
R^k_n
$1-p^{k+1}_{n,i}$
$p^{k+1}_{n,i}$



Analog Federated ADMM (A-FADMM)





A. Elgabli, J. Park, C. Issaid, and M. Bennis, "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning," *IEEE TCOM 2021*

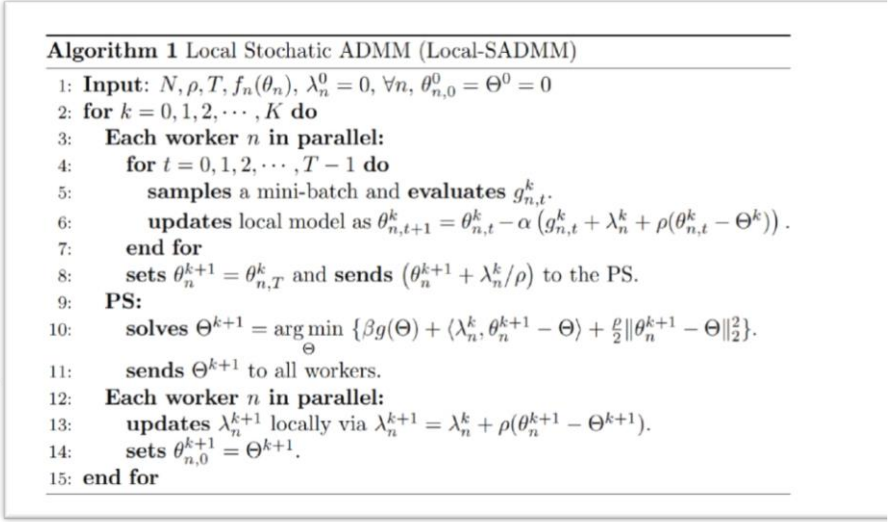
Extension to Non-convex Settings

Goal: To design communication-efficient primal-dual distributed learning algorithm in the **non-convex** setting.

- A set of N workers communicating with a PS to learn a global model
- The learning problem is given by

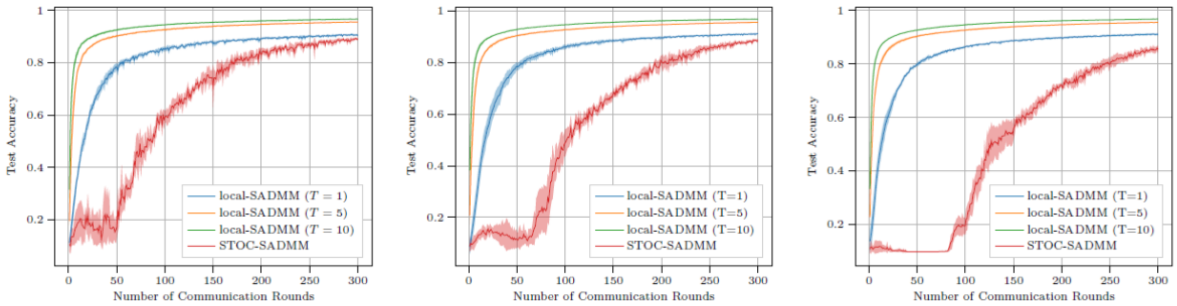
$$\textbf{(P1)} \quad \min_{\Theta} \sum_{n=1}^{N} f_n(\Theta) + \beta g(\Theta)$$

- $\Theta \in \mathbb{R}^d$: global model
- f_n : local *differentiable* and *non-convex* functions
- g : *non-smooth* and *convex* regularization function
- $\beta > 0$: regularization parameter



Algorithm 1 Local Stochatic ADMM (Local-SADMM)

```
Input: $N, \rho, T, f_n(\theta_n)$, $\lambda_n^0 = 0$, $\forall n$, $\theta_{n,0}^0 = \Theta^0 = 0$
for $k = 0, 1, 2, \cdots, K$ do
  Each worker $n$ in parallel:
    for $t = 0, 1, 2, \cdots, T-1$ do
      samples a mini-batch and evaluates $g_{n,t}^k$.
      updates local model as $\theta_{n,t+1}^k = \theta_{n,t}^k - \alpha\left(g_{n,t}^k + \lambda_n^k + \rho(\theta_{n,t}^k - \Theta^k)\right)$.
    end for
    sets $\theta_n^{k+1} = \theta_{n,T}^k$ and sends $\left(\theta_n^{k+1} + \lambda_n^k/\rho\right)$ to the PS.
  PS:
    solves $\Theta^{k+1} = \arg\min_{\Theta} \{\beta g(\Theta) + \langle \lambda_n^k, \theta_n^{k+1} - \Theta \rangle + \frac{\rho}{2}\|\theta_n^{k+1} - \Theta\|_2^2\}$.
    sends $\Theta^{k+1}$ to all workers.
  Each worker $n$ in parallel:
    updates $\lambda_n^{k+1}$ locally via $\lambda_n^{k+1} = \lambda_n^k + \rho(\theta_n^{k+1} - \Theta^{k+1})$.
    sets $\theta_{n,0}^{k+1} = \Theta^{k+1}$.
end for
```



C. Ben Issaid,et al, "Local Stochastic ADMM for Communication-Efficient Distributed Learning." IEEE WCNC 2022

Koc Team

- Extensive expertise in
  - Wireless networking
  - Machine-to-machine communications
  - Sensor networks
  - Vehicular communication networks

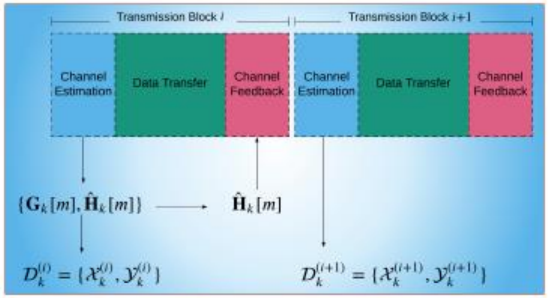
FL for Physical Layer Design: Channel Estimation

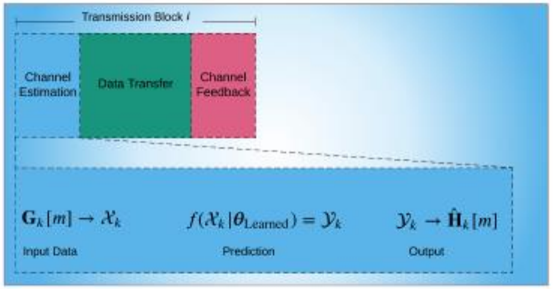
Training Stage

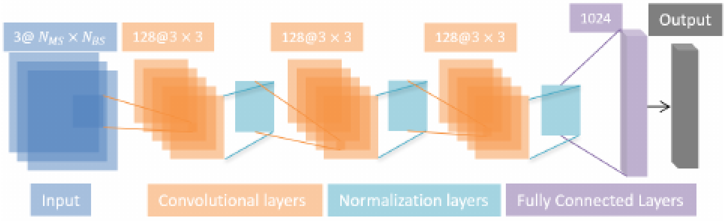
- We need input-label tuples for various channel realizations.
- Each user collects the received pilot data and perform channel estimation via a model-based method.
- Users perform FL with their local datasets.

Prediction Stage

- The trained model is available at the users
- Each user feed the received pilot data to predict its own channel.
- The users can then feedback the channel info to the BS


Transmission Block i
Transmission Block $i+1$
Channel Estimation
Data Transfer
Channel Feedback
Channel Estimation
Data Transfer
Channel Feedback
$\{\mathbf{G}_k[m], \hat{\mathbf{H}}_k[m]\}$
$\hat{\mathbf{H}}_k[m]$
$\mathcal{D}_k^{(i)} = \{\mathcal{X}_k^{(i)}, \mathcal{Y}_k^{(i)}\}$
$\mathcal{D}_k^{(i+1)} = \{\mathcal{X}_k^{(i+1)}, \mathcal{Y}_k^{(i+1)}\}$

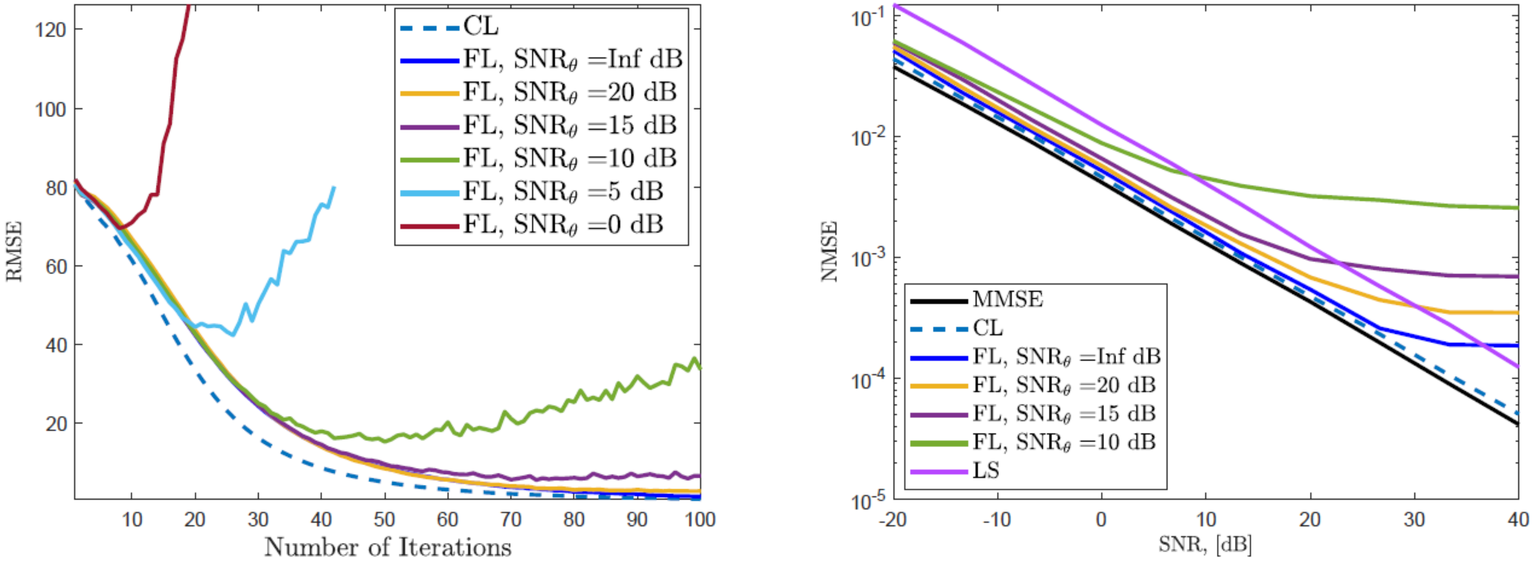

Transmission Block i
Channel Estimation
Data Transfer
Channel Feedback
$\mathbf{G}_k[m] \rightarrow \mathcal{X}_k$
Input Data
$f(\mathcal{X}_k|\boldsymbol{\theta}_{\text{Learned}}) = \mathcal{Y}_k$
Prediction
$\mathcal{Y}_k \rightarrow \hat{\mathbf{H}}_k[m]$
Output



A. M. Elbir and S. Coleri, "**Federated Learning for Channel Estimation in Conventional and IRS-Assisted Massive MIMO**", accepted to IEEE Transactions on Wireless Communications.

FL for Physical Layer Design: Channel Estimation

Training RMSE and Prediction NMSE
when model parameters are corrupted during FL



(a)

(b)

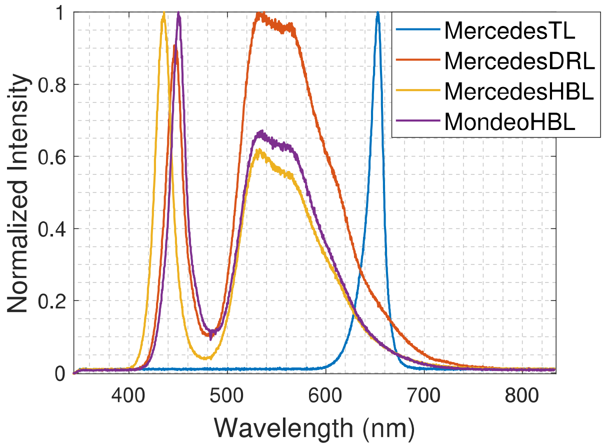
Fig. 5. Validation RMSE (a) and channel estimation NMSE (b) with respect to SNR_θ in massive MIMO scenario, respectively.

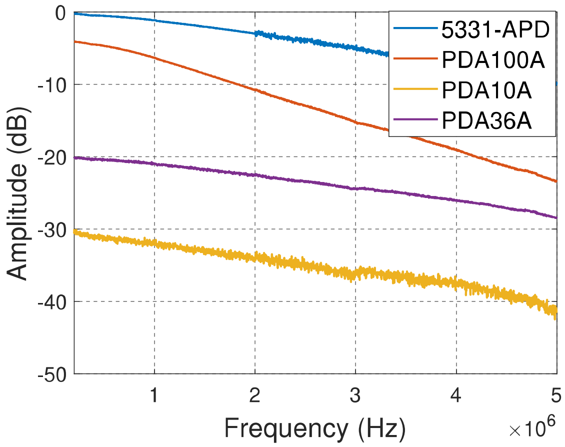
Heterogeneous architecture

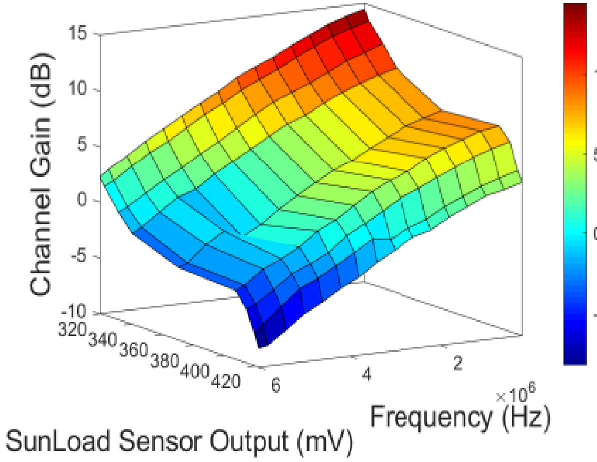
- Communication technologies
  - IEEE 802.11p
  - C-V2X
  - VLC

- Network selection
  - Different network capabilities at different vehicles
  - Changing network topology
  - Sensor readings for network selection (proximity, fog sensors)

Channel Modeling







- Machine learning based V2V V-VLC channel modeling
  - Incorporates multi-dimensional channel physical parameters as inputs, free from assumptions and analytical expression limitations
  - Predict channel parameters by learning robust patterns in V-VLC channel data

B. Turan and S. Coleri, "**Machine Learning based Channel Modeling for Vehicular Visible Light Communication**", IEEE Transactions on Vehicular Technology, 2021.

Path Loss Prediction

Method	RMSE (dB)	R-Square
Piecewise Lambertian	7.457	0.8539
Exponential Fit	7.459	0.8538
Linear Fit	10.220	0.7254
Two Term Exponential	7.002	0.8712

Algorithm	Optimal Hyperparameters	RMSE (dB)	MAE (dB)
Random Forest	Number of Estimators (S) 253 , Maximum Depth (T) 710	3.8107	2.4541
MLP-NN	35-10 2 layer network, tansig activation function	3.9502	2.1856
RBF-NN	Spread Factor 0.4, NN Size 551	3.5305	1.8854

- RBF-NN predicts path loss with the highest accuracy, 3.47 dB better prediction performance than the best fitting two-term exponential model

Heterogeneous Architecture Modeling

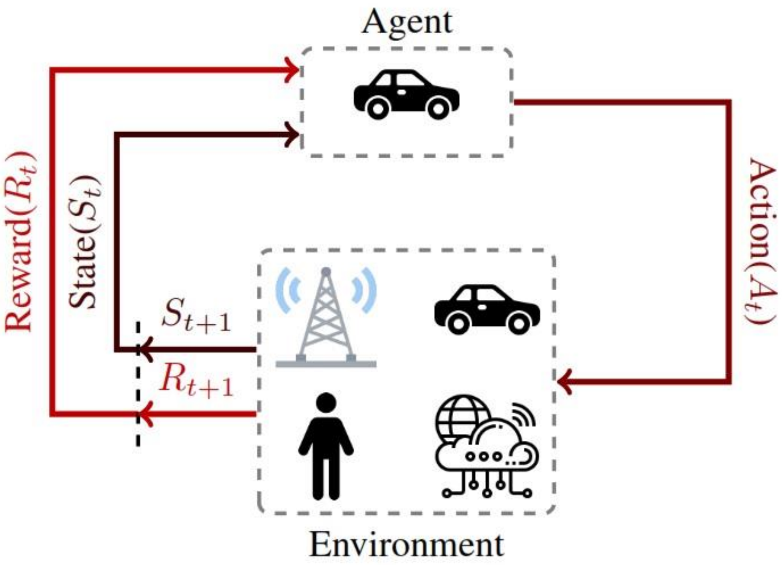
- Propose link quality estimation and jamming detection scheme for IEEE 802.11p and V-VLC
  - Use random forest regression and classifier based algorithms
- Test framework on real-world measurement data
  - 2.34dB and 0.56dB mean absolute error (MAE) improvement for V-VLC and IEEE 802.11p, respectively
  - 88.3% accuracy to detect noise interference injection for IEEE 802.11p links

B. Turan, A. Uyrus, O. N. Koc, E. Kar and S. Coleri, "**Machine Learning Aided Path Loss Estimation and Radio Jamming Detection for Heterogeneous Vehicular Communications**", IEEE GLOBECOM 2021.

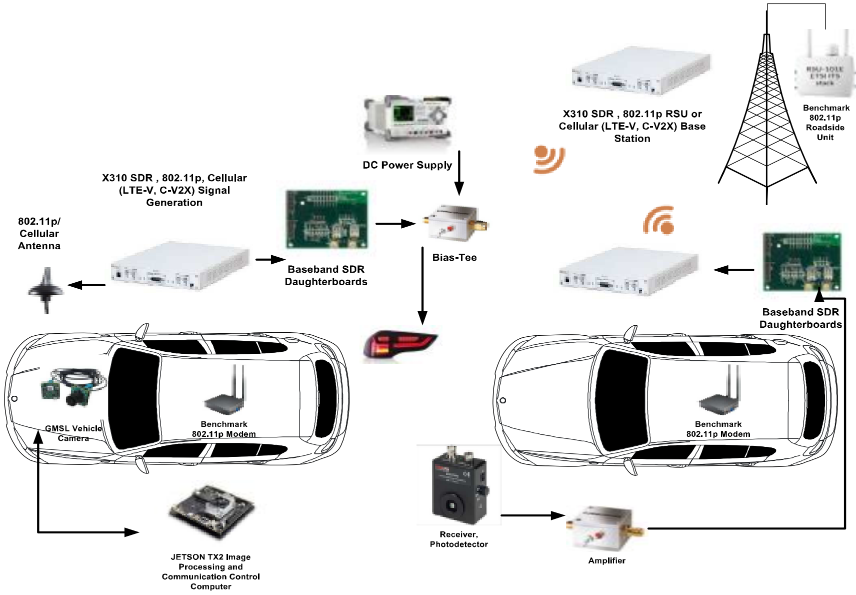
Network Selection in Heterogeneous Architecture

- Reinforcement learning algorithm

- States: Magnitude of vehicle’s local update, network availability parameters and SNR
- Actions: Discard, IEEE 802.11p, LTE-V2X, 5G NR, VLC
- Reward: Reliability, delay



Collaborative Work


X310 SDR , 802.11p RSU or Cellular (LTE-V, C-V2X) Base Station
Benchmark 802.11p Roadside Unit
DC Power Supply
X310 SDR , 802.11p, Cellular (LTE-V, C-V2X) Signal Generation
802.11p/ Cellular Antenna
Baseband SDR Daughterboards
Bias-Tee
Baseband SDR Daughterboards
GMSL Vehicle Camera
Benchmark 802.11p Modem
Benchmark 802.11p Modem
Receiver, Photodetector
Amplifier
JETSON TX2 Image Processing and Communication Control Computer

- Validation of algorithms on moving vehicles
- Joint paper preparation on vehicular cooperative sensing and communication

- Visit of Oulu team at Koc university for a period of 3-months in 2022
- Hybrid transceiver architecture ready
  - VVLC, IEEE 802.11p and C-V2X modems controlled by the same edge computer (EC)

Dissemination

- Published/submitted many papers in high-impact journals
- Gave many invited talks and tutorials
- Provide software codes on project webpage
- Imperial design for novel beam selection algorithm using LIDAR data collected from vehicles won third place in AI/ML in 5G challenge organized by International Telecommunications Union
- Oulu team disseminated results to 6GENESIS project

Contact

- Project webpage: https://connect.ku.edu.tr

- Sinem Coleri: scoleri@ku.edu.tr
- Deniz Gunduz: d.gunduz@imperial.ac.uk
- Mehdi Bennis: mehdi.bennis@oulu.fi