Wireless AI: Challenges and Opportunities

Mérouane Debbah
Mathematical and Algorithmic Sciences Lab, Huawei, France

16th of November, 2018

Huawei [Wow Way]

180, 000

Employees

80, 000

R&D employees

170+

Countries

15

R&D Institutes & Centers

No. 70

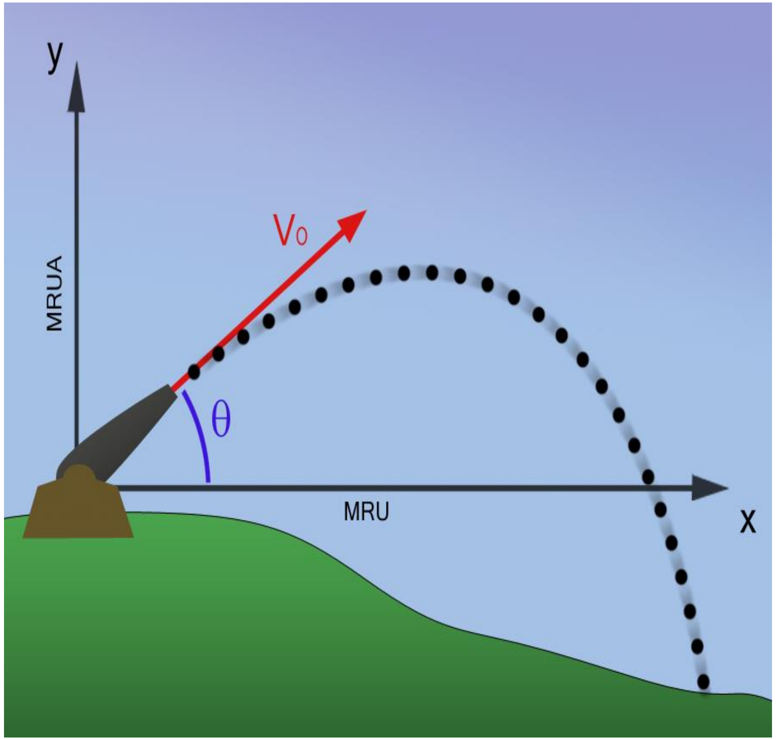
Interbrand’s Top 100 Best Global Brands

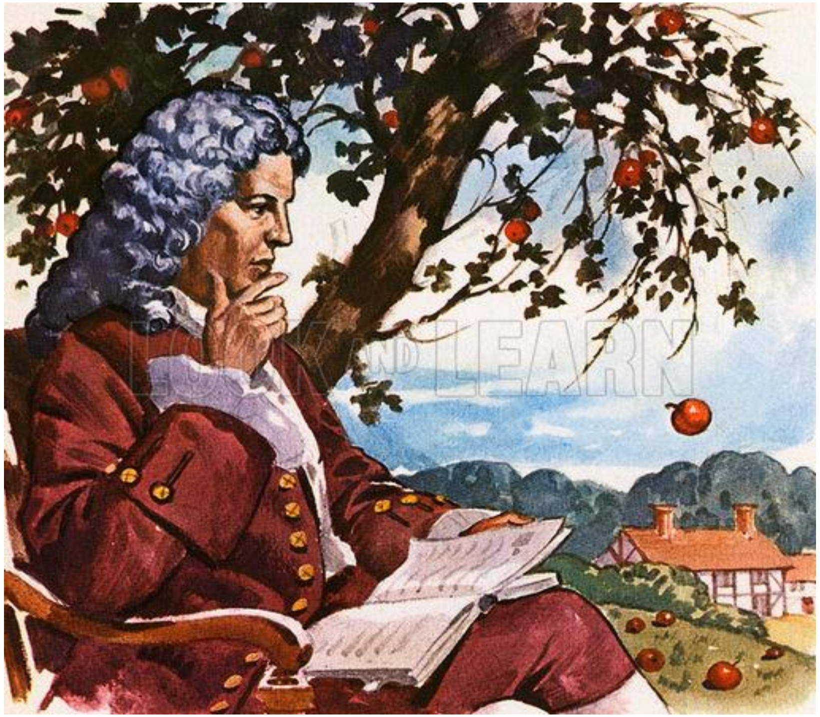
No. 83

Fortune Global 500

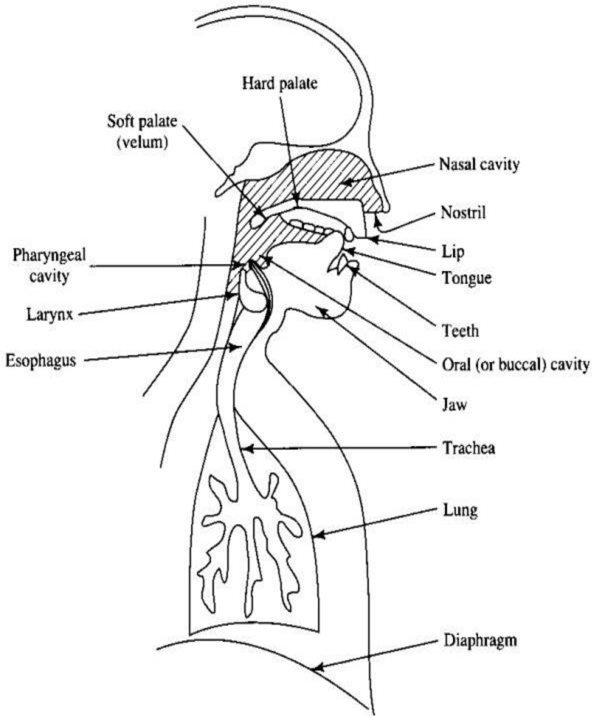
Contents

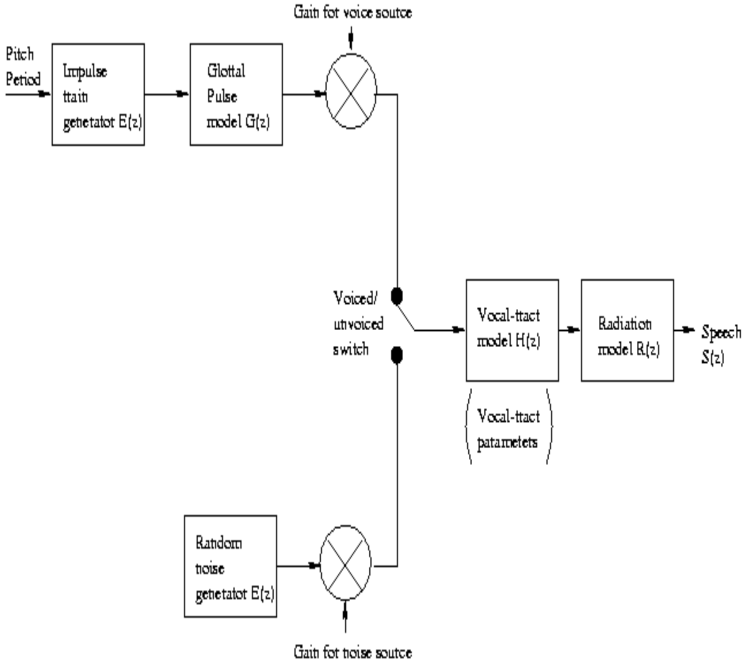
The cost of Understanding



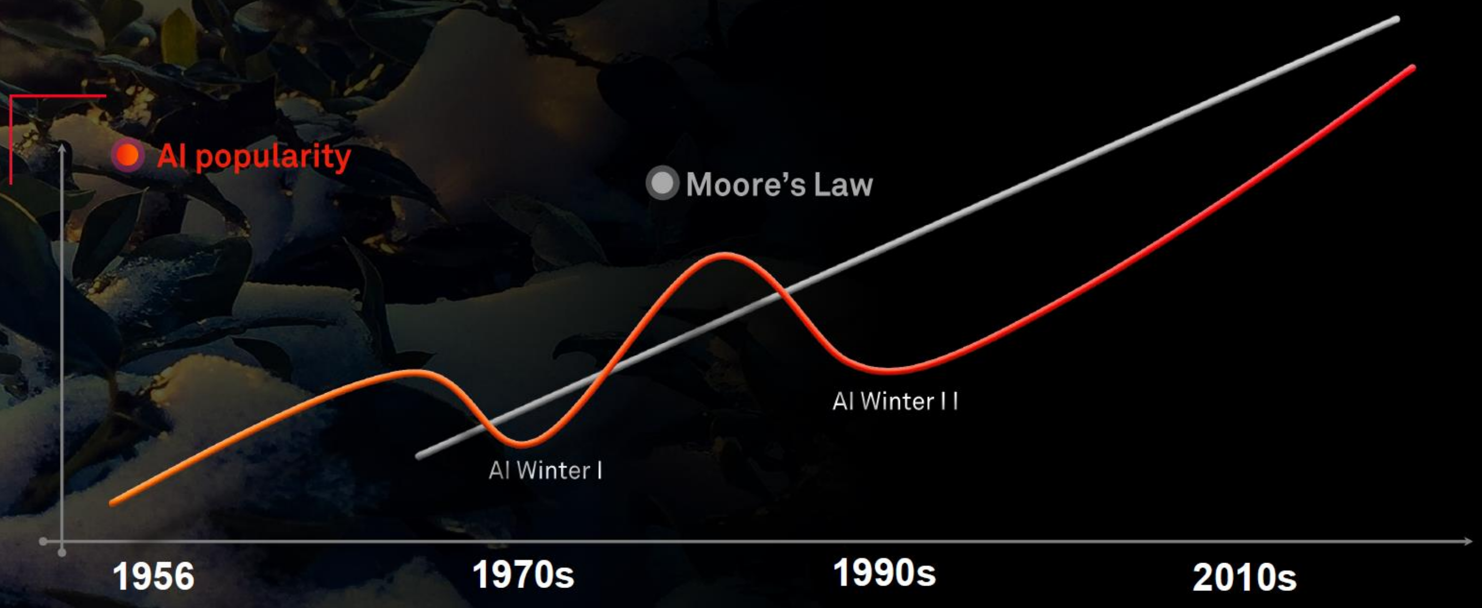


The limits of modelling





AI: Overall outcome of 60 years of development in ICT



Why it worked

•*Machine-learning algorithms* have progressed in recent years, especially through the development of deep learning and reinforcement-learning techniques based on neural networks.

•*Computing capacity* has become available to train larger and more complex models much faster. Graphics processing units (GPUs), originally designed to render the computer graphics in video games, have been repurposed to execute the data and algorithm crunching required for machine learning at speeds many times faster than traditional processor chips. *Key Trend Emerging*: Specially design chips and Hardware for Machine Learning workloads (Tensor Units).

•**Massive amounts of data** that can be used to train Machine Learning models are being generated, for example through daily creation of billions of images, online click streams, voice and video, mobile locations, and sensors embedded in the Internet of Things devices.

Be careful: we will still need experts and not just data scientists. Ten Million uninformed opinions are not as good as one expert opinion!

Third Wave AI Technologies

DARPA Is Funding Research Into AI That Can Explain What It's "Thinking"

Krisztián Mátyás / Emily Cho

by **Kristin Houser** | **July 24, 2018** | **Artificial Intelligence**

LOOKING AHEAD. Researchers will hold the next wave of artificial intelligences (AI) to the same standard as high school math students everywhere: no credit if you don't show your work.

On Friday, Defense Advanced Research Projects Agency (DARPA), a Department of Defense (DoD) agency focused on breakthrough technologies, announced its Artificial Intelligence Exploration (AIE) program. This program will streamline the agency's process for funding AI research and development with a focus on third wave AI technologies — the kinds that can understand and explain *how* they arrived at an answer.

SHARE

#artificial intelligence #DARPA #military #the digest

ETSI Standards Initiative : New ETSI group on improving Operator Experience using Artificial Intelligence

- February 2017: ETSI announces the creation of the Industry Specification Group **'Experiential Network Intelligence' (ISG ENI).**

- The purpose of the group is to define a Cognitive Network Management architecture that is based on the **"observe-orient-decide-act"** control model. It uses AI (Artificial Intelligence) techniques and context-aware policies to ***adjust offered services*** based on changes ***in user needs, environmental conditions and business goals***.

- The system is ***experiential***, in that ***it learns from its operation and from decisions*** given to it by operators to improve its knowledge of how to act in the future. This will help operators automate their network configuration and monitoring processes, thereby reducing their operational expenditure and improving the use and maintenance of their networks.

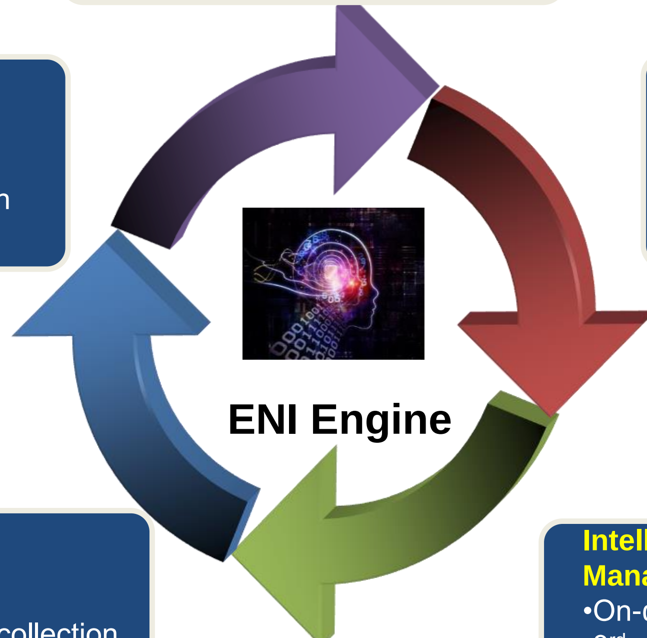
- "***The unique added value of the ETSI ISG ENI*** approach is to define new metrics to quantify the operator's experience; this enables the optimization and adjustment of the operator's experience over time, taking advantage of machine learning and reasoning." (ETSI ISG ENI Initiative Group Leader)

Source: ETSI ISG, Experiential Network Intelligence (ENI) Presentation, 2017

Experiential Networked Intelligence – Improving Experience Framework

Intelligent Service Deployment
- Intent based service management
- Service mapping
- Service atom

Intelligent Analyzing and Prediction
- Network analyzing
- Utilization/inventory Prediction
- Fault Prediction



Intelligent Policy Control
- Imperative policy
- Declarative policy
- Policy driven service/resource management

Intelligent Monitoring
- SDN Telemetry
- Network event & state collection
- Network performance collection

Intelligent Resource Management
- On-demand resource allocation
- 3rd party resource API
- Intent based resource management

Source: ETSI ISG, Experiential Network Intelligence (ENI) Presentation, 2017

New ETSI ISG ENI to Describe and Specify the Future Evolution of Network Intelligence to Enhance Experience

ETSI Standard
ENI – Initial
(Event-Condition-Action)

•ECA: ***Self-decision-making using Event –Condition-Action policy*** rule
•Service Model enables dynamic Service Automation

ETSI Standard
ENI – Mid
(Declarative)

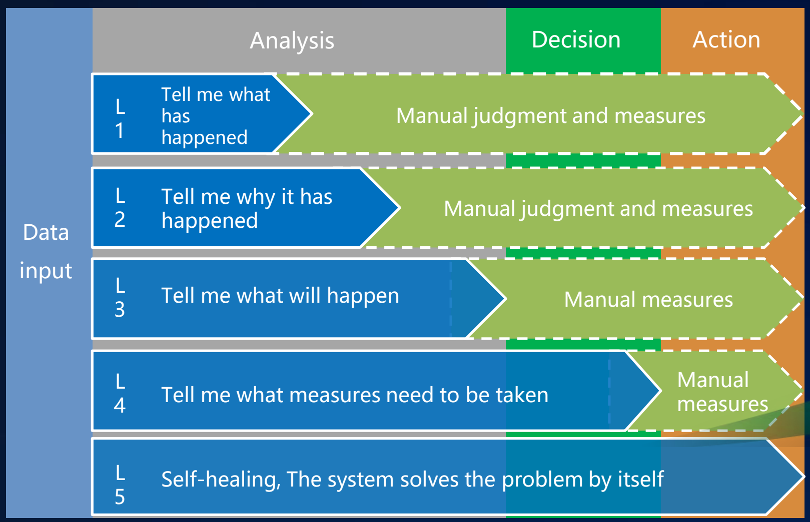
•Declarative: ***Self-decision-making with declarative policy*** statements
•Translation of user intent to service and resource management

ETSI Standard
ENI – Final
(Self-learning, AI)

•***Self-decision-making based on intelligence*** generated by Self-learning AI algorithms.
•Integrate all available resources to achieve high performance and improved Operator Experience

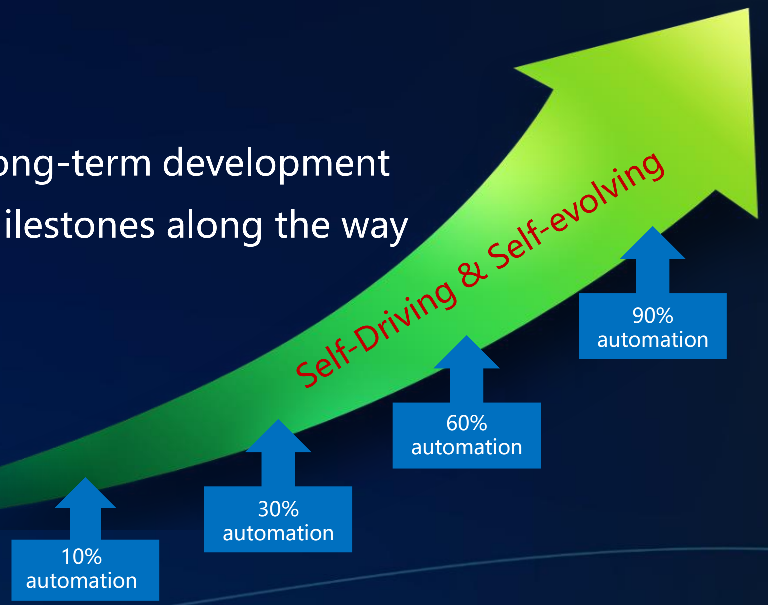
Source: ETSI ISG, Experiential Network Intelligence (ENI) Presentation, 2017

Five Stages for Self-Driving & Self-evolving Network ……



Long-term development

Milestones along the way



AI is not new in telecommunications

Philosophical Magazine, Ser.7, Vol. 41, No. 314 - March 1950.

XXII. Programming a Computer for Playing Chess[1]

By CLAUDE E. SHANNON

Bell Telephone Laboratories, Inc., Murray Hill, N.J.[2]

[Received November 8, 1949]

1. INTRODUCTION

This paper is concerned with the problem of constructing a computing routine or "program" for a modern general purpose computer which will enable it to play chess. Although perhaps of no practical importance, the question is of theoretical interest, and it is hoped that a satisfactory solution of this problem will act as a wedge in attacking other problems of a similar nature and of greater significance. Some possibilities in this direction are: -

AI is not new in Telecommunications

(1)Machines for designing filters, equalizers, etc.

(2)Machines for designing relay and switching circuits.

(3)Machines which will handle routing of telephone calls based on the individual circumstances rather than by fixed patterns.

(4)Machines for performing symbolic (non-numerical) mathematical operations.

(5)Machines capable of translating from one language to another.

(6)Machines for making strategic decisions in simplified military operations.

(7)Machines capable of orchestrating a melody.

(8)Machines capable of logical deduction.

AI is not new in Telecommunications

Unfortunately a machine operating according to the type A strategy would be both slow and a weak player. It would be slow since even if each position were evaluated in one microsecond (very optimistic) there are about 10^9 evaluations to be made after three moves (for each side). Thus, more than 16 minutes would be required for a move, or 10 hours for its half of a 40-move game.

It would be weak in playing skill because it is only seeing three moves deep and because we have not included any condition about quiescent positions for evaluation. The machine is operating in an extremely inefficient fashion - it computes all variations to exactly three moves and then stops (even though it or the opponent be in check). A good human player examines only a few selected variations and carries these out to a reasonable stopping point. A world champion can construct (at best) combinations say, 15 or 20 moves deep.

Contents

We designed Communication schemes without AI

- « A Mathematical Theory of Communication », Bell System Technical Journal, 1948, C. E. Shannon

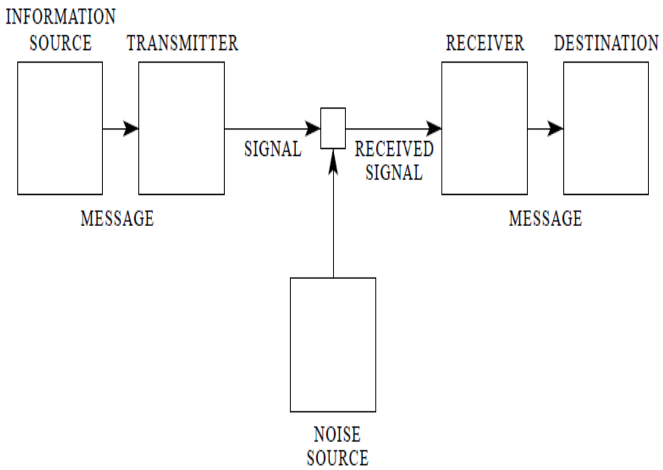
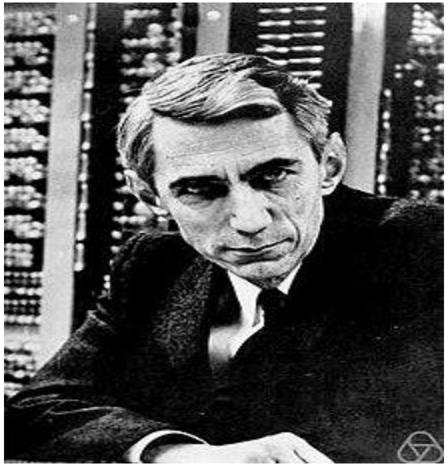
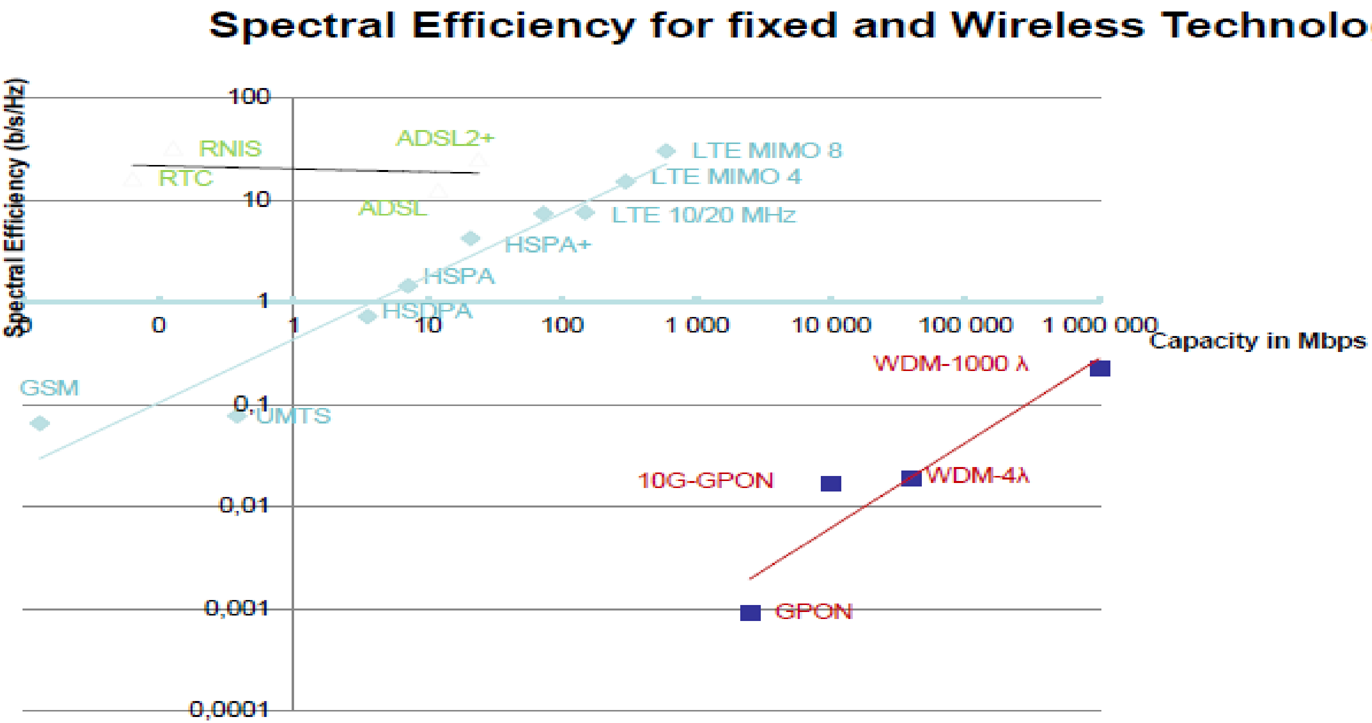


Fig. 1—Schematic diagram of a general communication system.



And we have been quite successful!



We learned without AI

- ” Cybernetics, or Control and Communication in the Animal and the Machine” , Herman et Cie/The Technology Press, 1948, N. Wiener

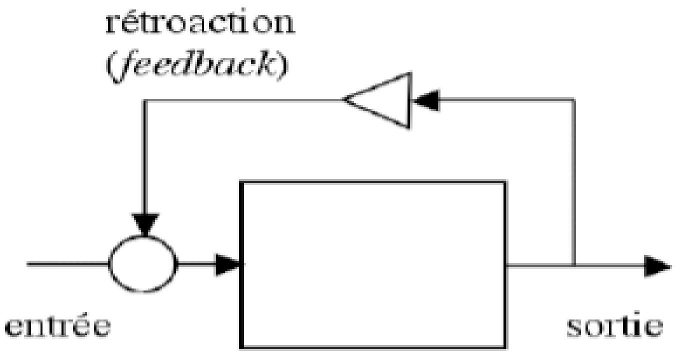
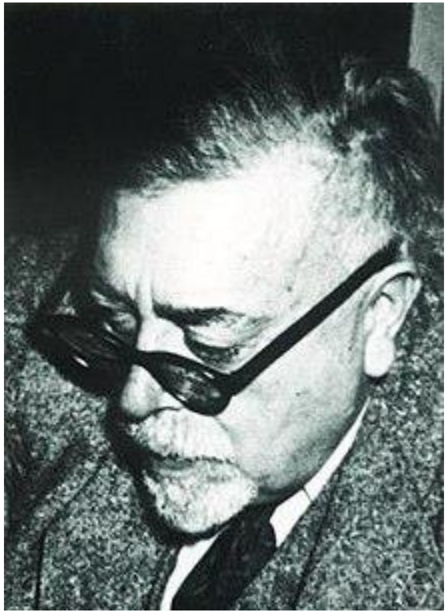


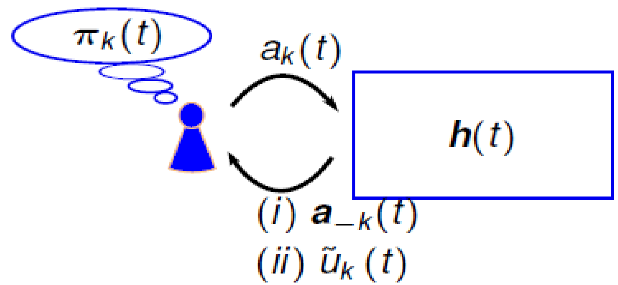
FIG. 1 – Boucle de rétroaction



and we have been quite successful: Control theory can rigorously prove that your design will perform as intended!

Learning Iterative Steps:

- **Choose** action $a_k(t) \sim \pi_k(t)$.
- **Observe** game outcome, e.g.,
  $\boldsymbol{a}_{-k}(t)$
  $u_k(a_k(t), \boldsymbol{a}_{-k}(t))$.
- **Improve** $\pi_k(t+1)$.



Thus, we can expect that: $\forall k \in \mathcal{K}$,

$$\pi_k(t) \xrightarrow{t\to\infty} \pi_k^* \tag{1}$$

$$\bar{u}_k\,(\pi_k(t), \pi_{-k}(t)) \xrightarrow{t\to\infty} \bar{u}_k\,(\pi_k^*, \pi_{-k}^*) \tag{2}$$

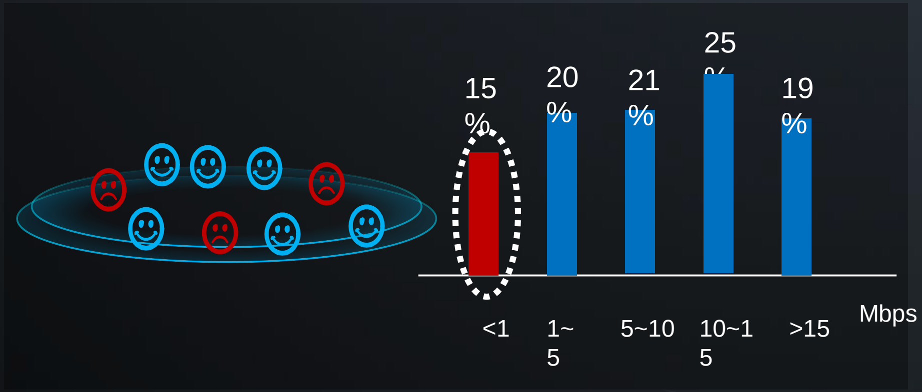
where, $\pi^* = (\pi_1^*, \ldots, \pi_K^*)$ is a NE strategy profile.

Most Common Learning Algorithms

- Best Response Dynamics (BRD)
- Fictitious Paly (FP)
- Reinforcement Learning (RL)
- Joint Utility Strategy Learning (JUSTE)
- Trial and Error Learning (TE)
- Regret Matching Learning
- Q-Learning
- Multi-Arm Bandits
- Imitation Learning

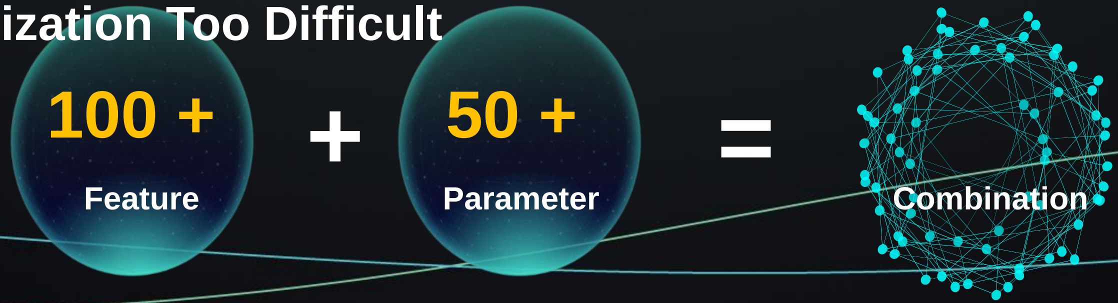
Edge User Throughput and VoLTE Background

Edge User THP Affect User Experience Than Average THP



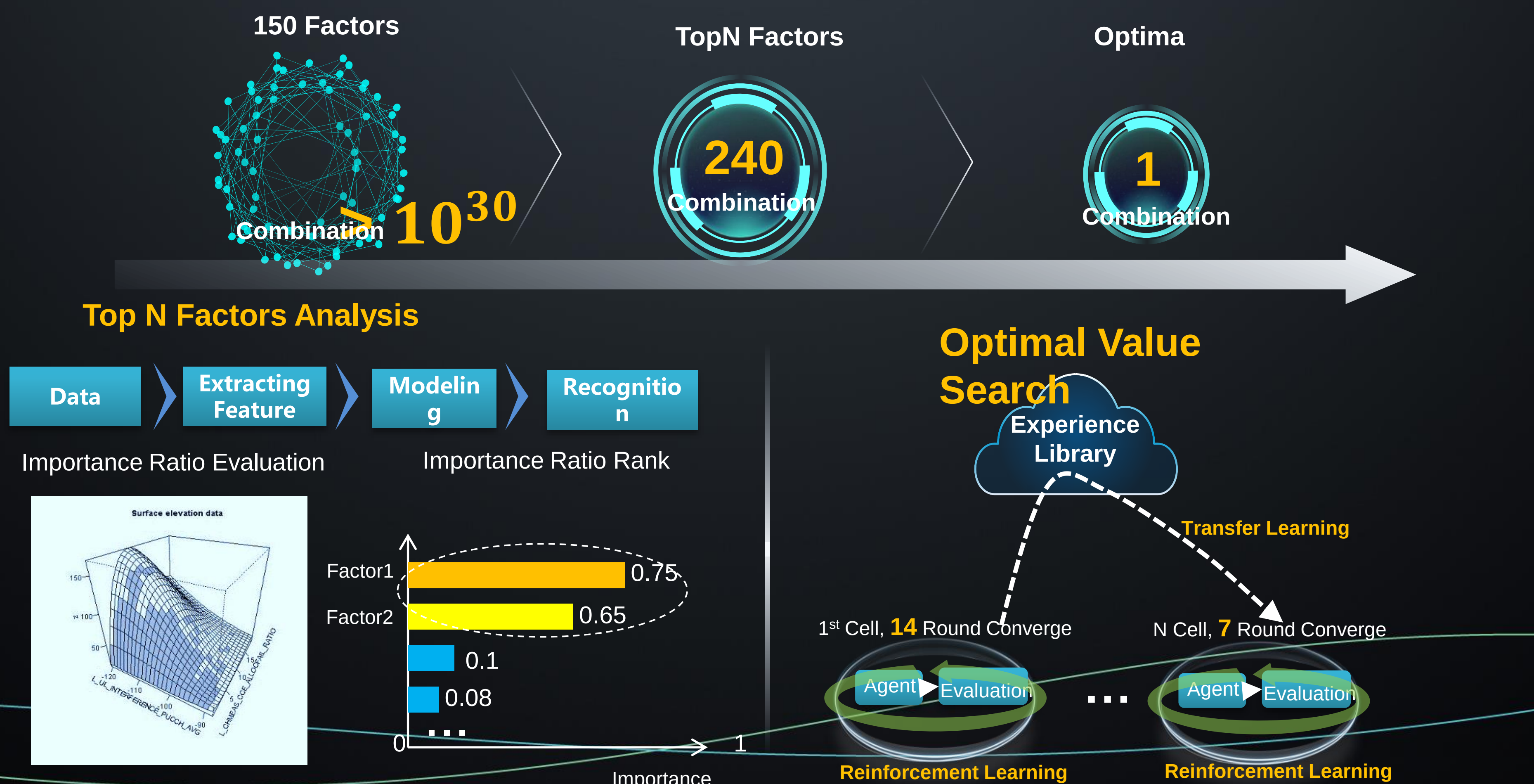
VoLTE PLR(Packet Loss Rate) Correlated with MOS

Too Many Factors Affect Edge Throughput and VoLTE, Make Optimization Too Difficult



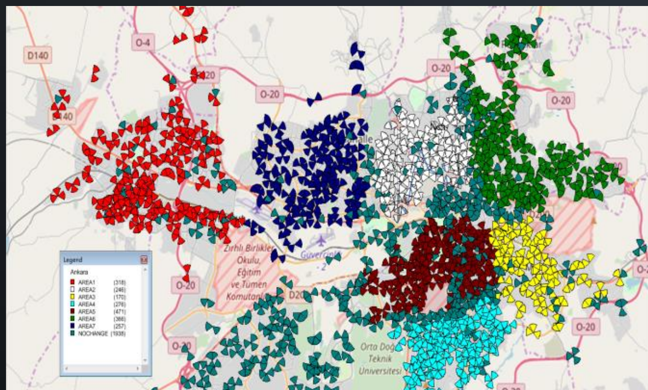
$>10^{30}$

AI Assisted VoLTE and Edge User THP Optimization

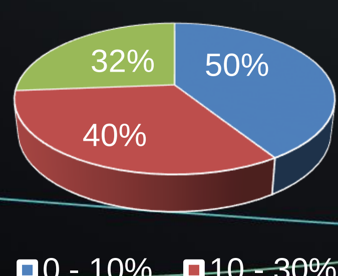


Innovation in Turkey – Edge THP Optimization

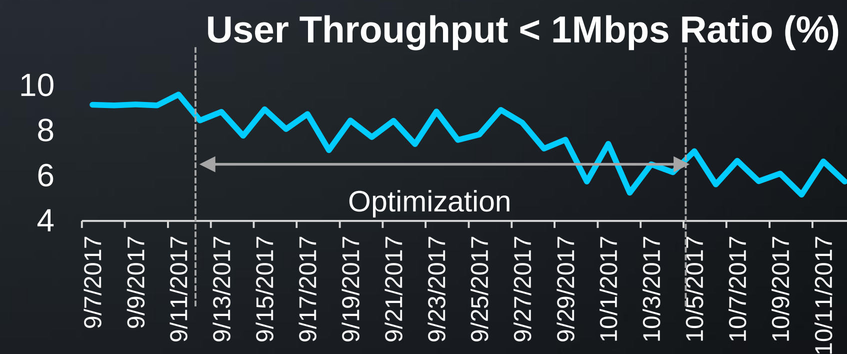
Test Area: Ankara **Scope**: 699 Site, 2281LCell **<1 Mbps Use Ratio Decrease while User Num Increase.**

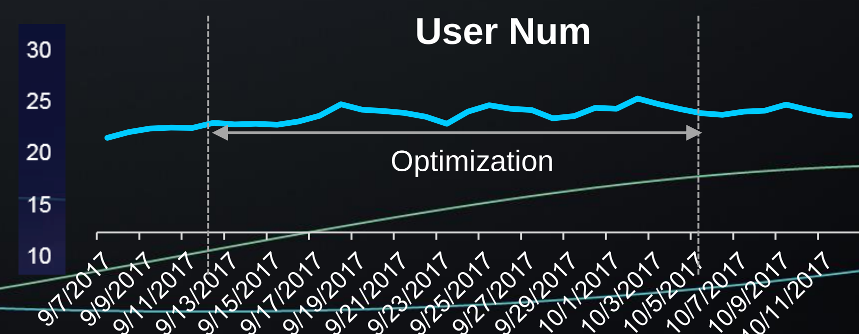


Whole Network Gain:



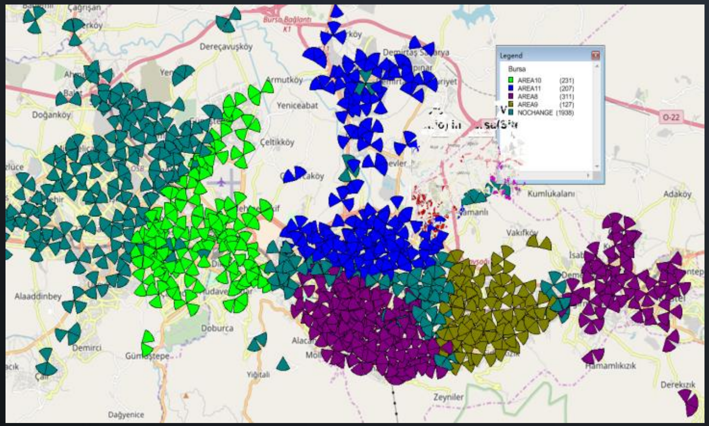
Ankara Top 10% Gain Cells





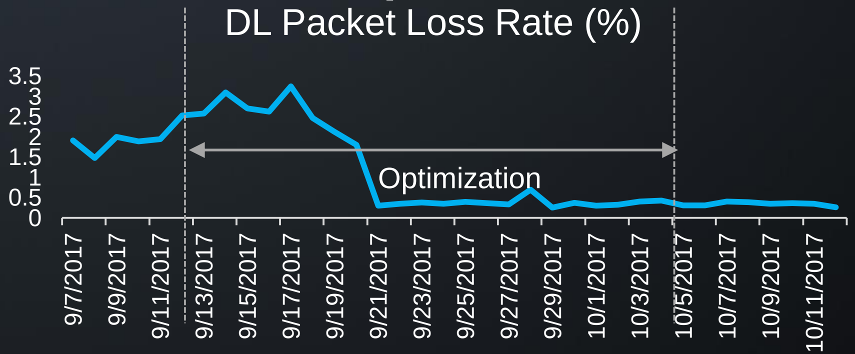
Innovation in Turkey – VoLTE Optimization

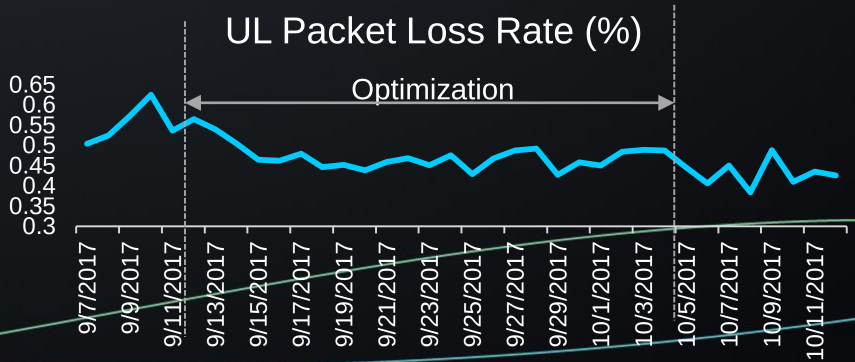
Test Area: Bursa **Scope: 188 Site, 877 LCell**



DL/UL PLR Decrease Dramatically

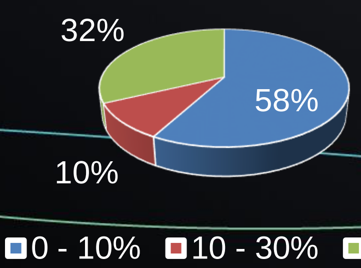
Bursa Top 10% Gain Cells



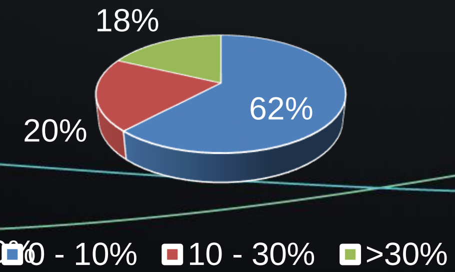


Whole Network Gain:

DL PLR Gain Distribution

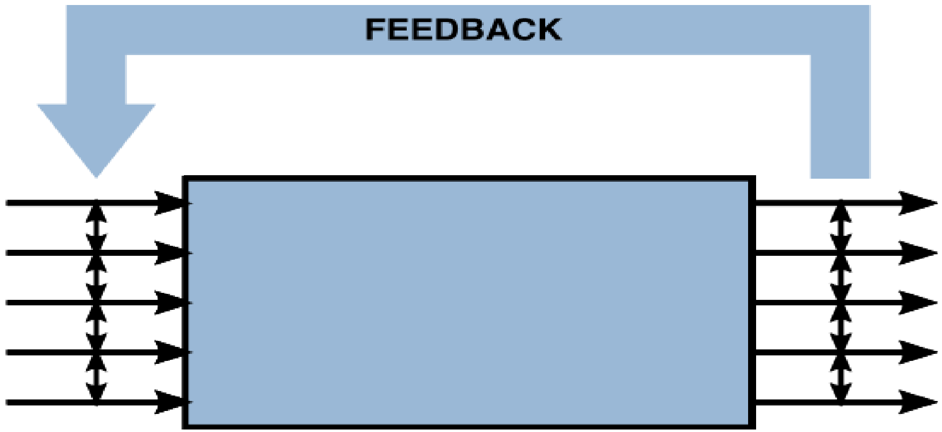


UL PLR Gain Distribution



PLR: Packet Loss Rate

Why Deep Communications now?

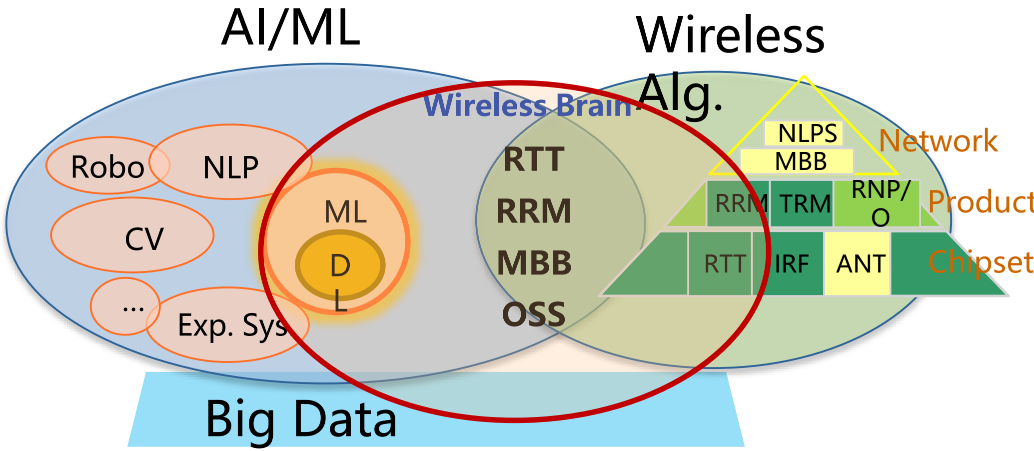


- Models are expensive to obtain
- The E2E objective function is not defined mathematically
- High Dimensional space with many parameters

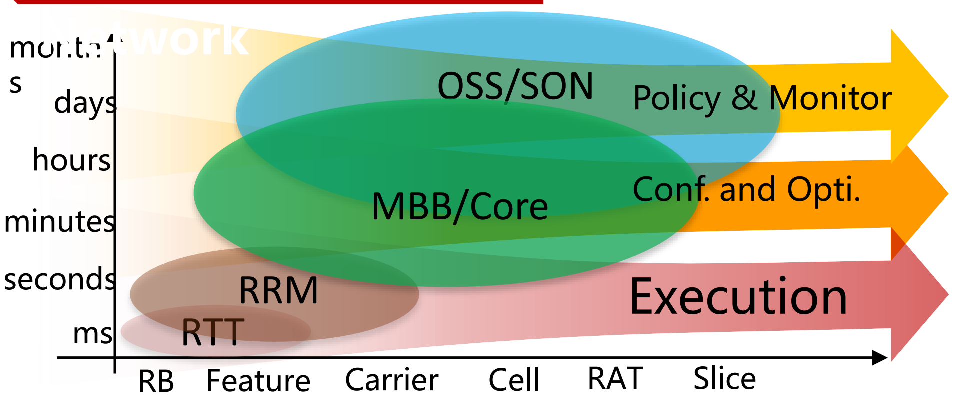
Wireless AI: Key Technology

What is Wireless AI

Goal oriented and self control in network management and optimization solution, can overcome the problem when the network cannot be accurately expressed with formula based on big data and machine learning technology.



AI in Wireless Network

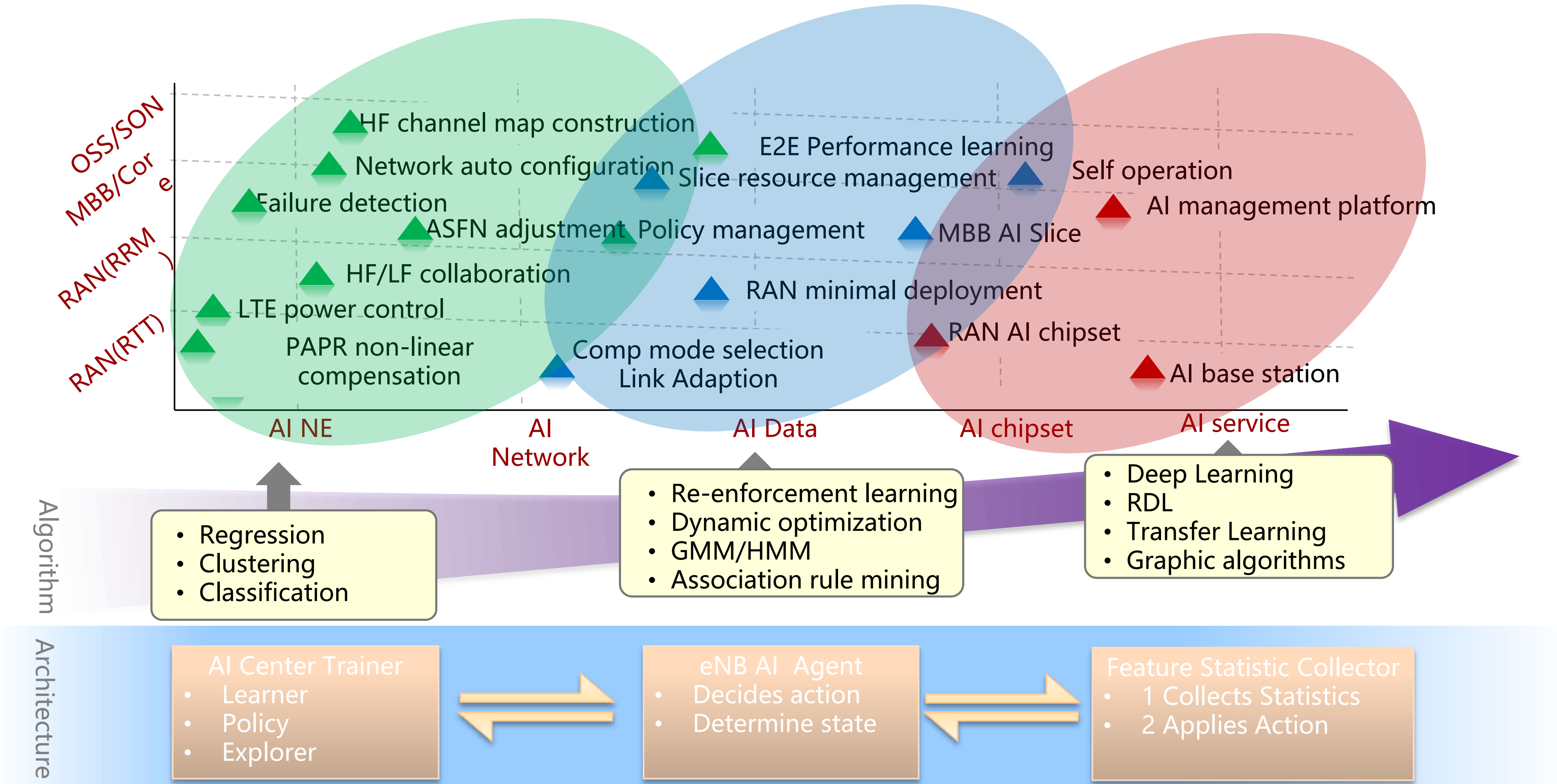


Comparison

	AI Algorithm	Wireless Algorithm
Value	Data	Link
Scenario	Automatically	Manually
Target	Global probability optimization	Local determined optimization
Scope	E2E network	Locally Modelling
method	Big data, learning	Formula , optimization
Usage	Set the target goal	Tune parameters manually

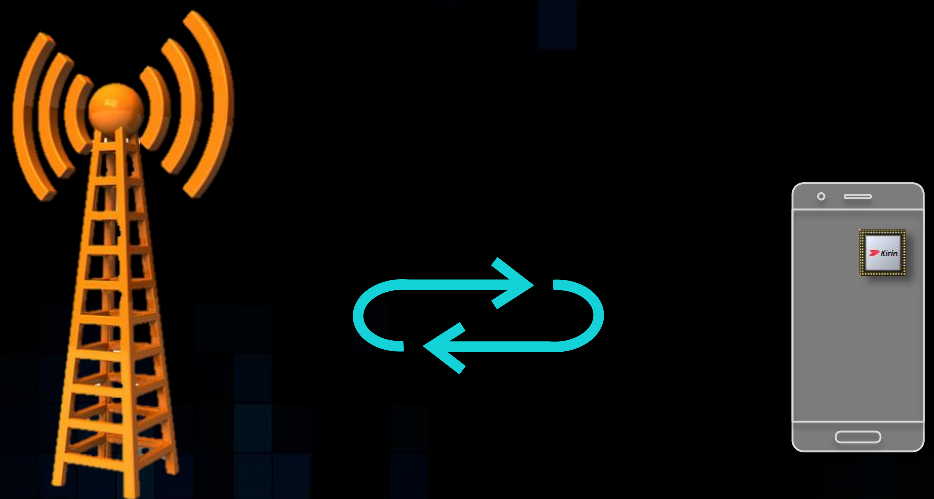
Feature

AI in Wireless



Mobile AI: Where to compute AI?

Improving communication experience



Improving coverage
Optimizing resources

Field test
Algorithm improvement
Performance optimization

HUAWEI

Huawei Mobile AI strategy

On-Device AI

+

Cloud AI

HUAWEI

Huawei Mobile AI strategy

Intelligent Phones

+

Intelligent Network

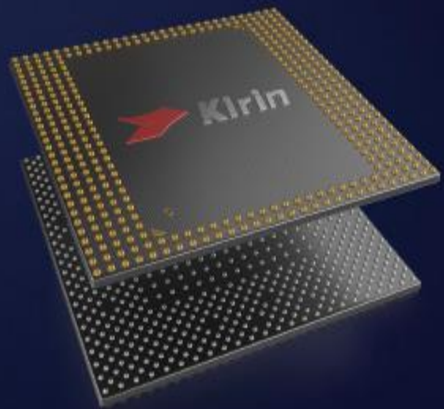
HUAWEI Kirin 970

The World's First Smartphone AI Computing Platform with HiAI Architecture

Leading Process Technology
10nm Process Technology

High Performance NPU
Up to 25x performance
Up to 50x power efficiency

High Performance 8-Core CPU
4xA73 @2.4GHz
4xA53 @1.8GHz



High Efficiency 12-Core GPU
First-to-Market
Mali G72MP12

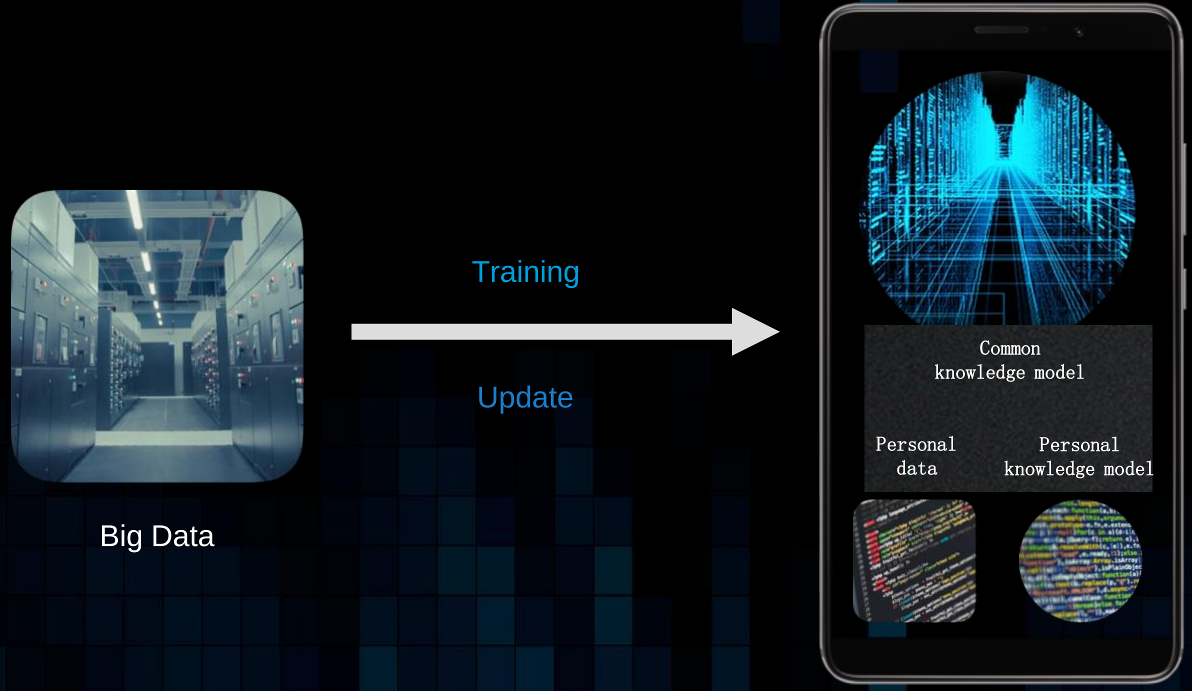
Advanced Dual ISP
4-Hybrid Focus
Low-light & Motion Shooting

Ultra-Fast 4.5G LTE Modem
4.5G LTE Cat.18 up to
1.2Gbps Download speeds

4.5G

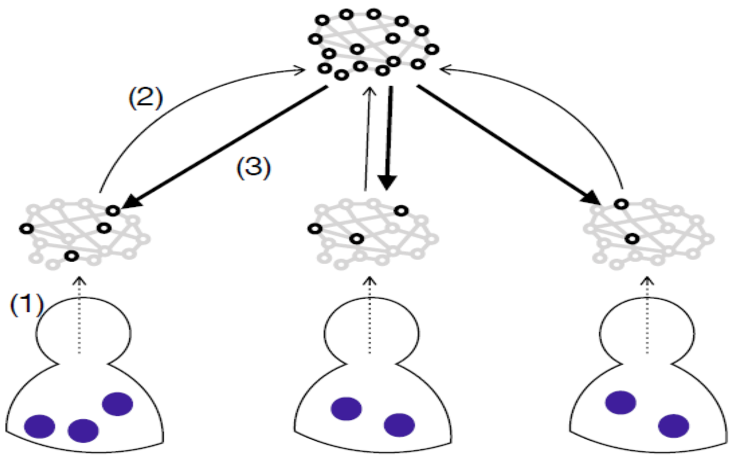
NPU: Neural Network Processing Unit

Mobile AI: What is the right architecture?



New Open Possibilities in Wireless to cope with 5G Requirements

Federated Learning (FL)



Federated Learning for Ultra-Reliable Low-Latency V2V Communications

Sumudu Samarakoon*, Mehdi Bennis*, Walid Saad†, and Mérouane Debbah‡

* Centre for Wireless Communication, University of Oulu, Finland, email: {sumudu.samarakoon,mehdi.bennis}@oulu.fi

†Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, email: walids@vt.edu

‡ Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France, (email: merouane.debbah@huawei.com)

***Abstract*—In this paper, a novel joint transmit power and resource allocation approach for enabling ultra-reliable low-latency communication (URLLC) in vehicular networks is proposed. The objective is to minimize the network-wide power consumption of vehicular users (VUEs) while ensuring high reliability in terms of**

a probabilistic constraint to maintain small queue lengths.

Although a probabilistic constraint on the queue length improves network reliability, it fails to control rare events in which large queue lengths occur with low probability.

HUAWEI

Contents

What has been done so far with AI for Wireless

AI Inside

- AI based Transceiver design
- AI based Self-Optimization
- AI based scheduling
- AI based Prediction

AI based Transceiver Design

- CSI acquisition in FDD Massive MIMO
- Auto-encoder for end to end communication systems
- AI based Localization
- AI based Distributed Calibration
- AI based Waveform Design
- AI based Multiple Access
- AI based PAPR reduction

AI based Self-Optimization

- Energy Efficient Power Control by Deep Neural Networks
- Collaborative AI for User-Cell Association
- Network Traffic Prediction with AI
- AI for traffic Association in C-RAN
- AI for the deployment of UAV’ s
- AI for multiple access in IoT
- AI for scheduling
- AI for rate adaptation in WiFi
- AI for interference management in SON

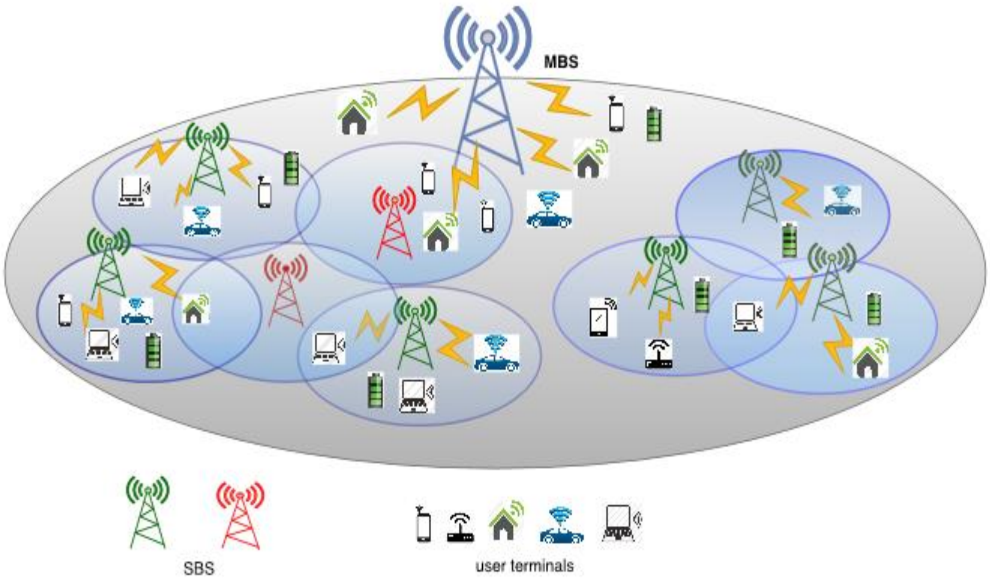
Collaborative Artificial Intelligence (AI) for User-Cell Association in Ultra-Dense Cellular Systems

Abstract—In this paper, the problem of cell association between small base stations (SBSs) and users in dense wireless networks is studied using artificial intelligence (AI) techniques. The problem is formulated as a mean-field game in which the users' goal is to maximize their data rate by exploiting local data and the data available at neighboring users via an imitation process. Such a learning process prevents the users from exchanging their data directly via the cellular network's limited backhaul links and, thus, allows them to improve their cell association policy collaboratively with minimum computing. To solve this problem, a deep reinforcement learning algorithm is proposed that enables the users to predict their reward function using a neural network whose input is the SBSs selected by neighboring users and the local data of the considered user. Simulation results show that the proposed imitation-based mechanism for cell association converges faster than the cell association without imitation to the optimal solution.

User Cell Association

- Each user has to choose the best SBS among its options.
- After selection it will receive a reward based on congestion level of the SBS and the channel gain between itself and the SBS.
- Each user try to solve an optimal control problem for selecting the SBS over the time.
- Network is ultra-dense so we model it using mean field games.

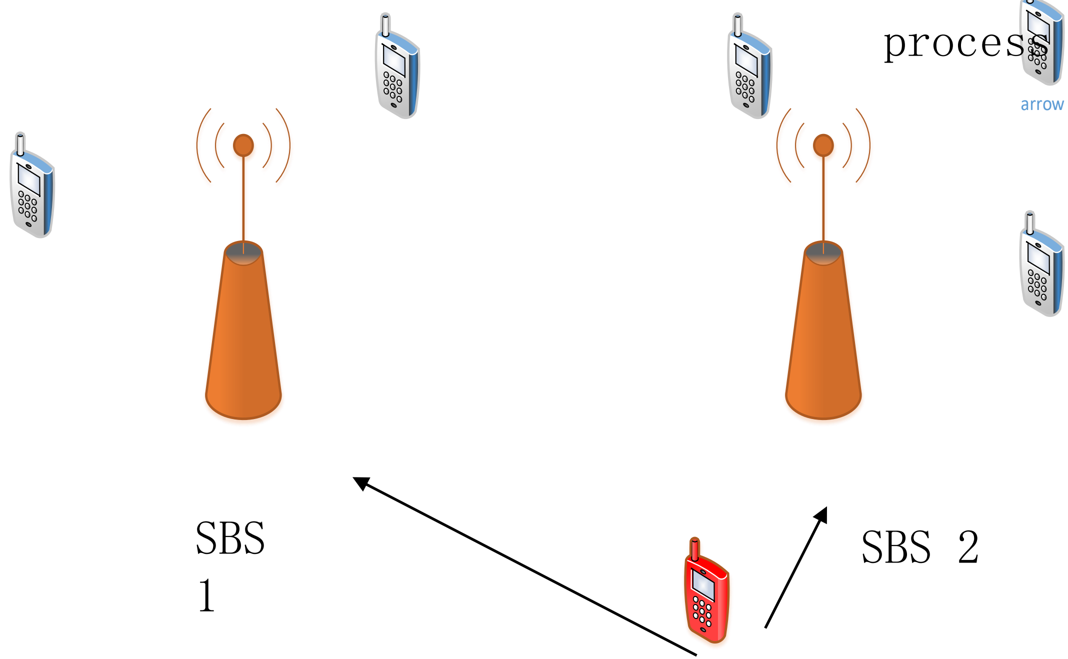
User Cell Association



Neighboring devices are in the same conditions.

User Cell Association

how to exploit similarity between users to reduce the communication load between users and make the decision process faster?



Users' Reward Model

- Users' reward function

Effect of congestion on SBS: a function of the total number of users connected to SBS s

$$r_{u,s} = B \log_2 \left(1 + \frac{p_{us} h_{us}}{\sigma^2 + I(\sum_{u'} a_{su'})}\right) - c(\sum_{u'=1}^{U} a_{su'}), \qquad (1)$$

Reward for user u when connected to SBS s

Interference: a function of the total number of users connected to SBS s

Ergodic Control Problem

- Optimization problem for user u located at x:

SBS selected by user u' located at y at time t

Distance between user u located at x and user u' located at y_u

$$\sup_{s(.,\boldsymbol{x})} \lim_{T\to\infty} \inf \frac{1}{T} \sum_{t=0}^{T} \frac{-1}{\epsilon^2} \sum_{u' \in \mathcal{C}_u} \left(s(t,\boldsymbol{x}) - s(t,\boldsymbol{y}_i)\right)^2 \frac{1}{\epsilon} g\left(\frac{\|\boldsymbol{x} - \boldsymbol{y}_{u'}\|}{\epsilon}\right)) + r_{u,s(t,x)} \quad (1)$$

Gaussian kernel

Mexican Wave as a Mean Field Equilibrium

$$\inf_{z(t,x)} \liminf_{T\to+\infty} \int_0^T \left(\left[\frac{1}{\epsilon^2} \int_{\mathbb{R}} (z(t,x) - z(t,x-y))^2 \frac{1}{\epsilon} g\left(\frac{y}{\epsilon}\right) dy \right] + F(z(t,x)) + \frac{\dot{z}(t,x)^2}{2} \right) dt$$

where for any $t > 0$ and for any $x \in [0, L]$, $z(t, x) \in [0, 1]$ denotes the position at time t of any individual which is sitting into the stadium at the position x ($z = 0$ means that the individual is sitting, $z = 1$ that is standing). Moreover, g is a Gaussian kernel, and the term in square brackets models, roughly speaking, the taste of mimicry of each individual. Finally, F is a given function (is the cost payed if one decides to stay neither sitting nor standing) and the quadratic term is, intuitively, the cost due to the effort of standing up or sitting.

Mean-Field Game (MFG) Formulation

- Hamilton-Jacobi-Bellman equation:

$$V_s^t = \max_{P_s^t \in \mathbb{S}_s(\mathcal{G})} \left\{ r_s(\boldsymbol{\pi}^t, P_s^t) + \sum_{j \in \mathcal{V}_s} P_{sj}^t V_j^{t+1} \right\}$$

vector that represents the transition probability from SBS s to all other SBSs in $\mathbb{S}$

- Fokker-Planck equation (the evolution of users distribution over time):

$$\pi_s^{t+1} = \sum_{j \in \mathcal{V}_s} P_{js}^t \pi_j^t$$

probability that a user connected to SBS j switches to SBS s at time t

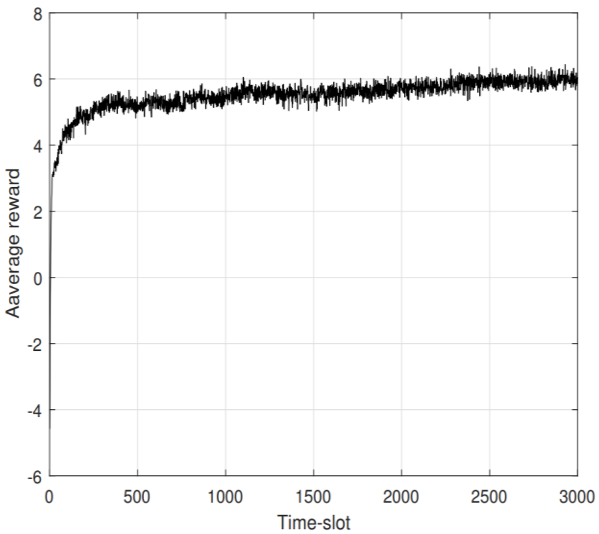
Transforming MFG to MDP

- Solving the mean-field game depends on the form of the utility function which restricts the application domain of MFG
- Given that the MFG problem is difficult to solve, we transform MFG problem into a Markov decision process (MDP).
- For solving MDP, we use deep reinforcement learning.
- Users approximate the value function using adaptive linear neuron (ADALINE) neural networks. The user then trains its network using Widrow-Hoff algorithm (exploration) with the probability ϵ and chooses the best BS with probability $1-\epsilon$.

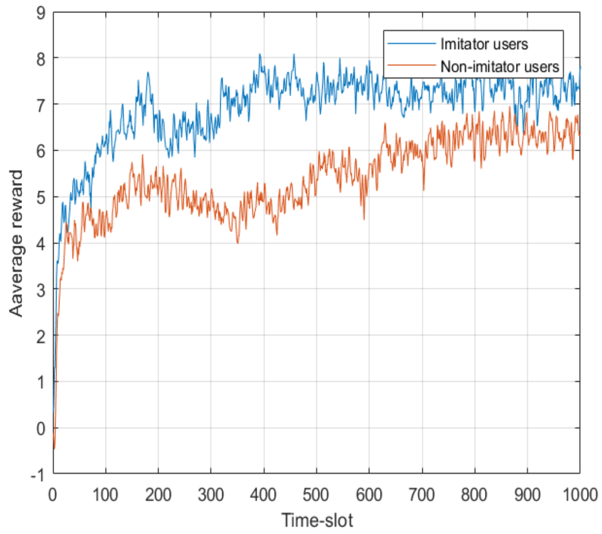
Simulation Results

- We have a system with 10 SBSs in equilibrium. (part1)
- 20 imitator users and 20 non-imitator users enter the system and start to learn the environment. (part2)
- Imitator users can learn the environment faster than non-imitator users.
- Imitator users use experience of the existing users in the system.

Simulation Results



Equilibrium for the users in the first part.



average reward for 20 imitator users and 20 non-imitator users in the second part.

Conclusion

- AI is at the heart of the Computation and Communication paradigm
- As engineers, we should Open the black Box!
- What are the adequate architectures to implement Wireless AI?
- Collaborative AI: Learn from one-self or learn from others?