Machine learning for Wireless Networks: Challenges and opportunities

Mérouane Debbah
Mathematical and Algorithmic Sciences Lab, Huawei, France

8th of June, 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

Huawei [in France]

950+ employees, among with 150 in R&D

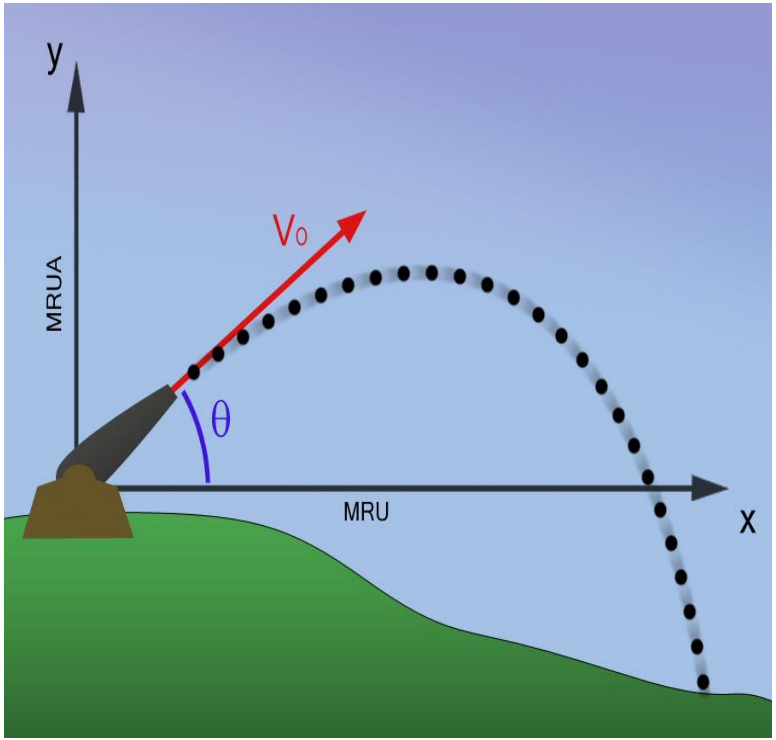
HQ in Boulogne-Billancourt

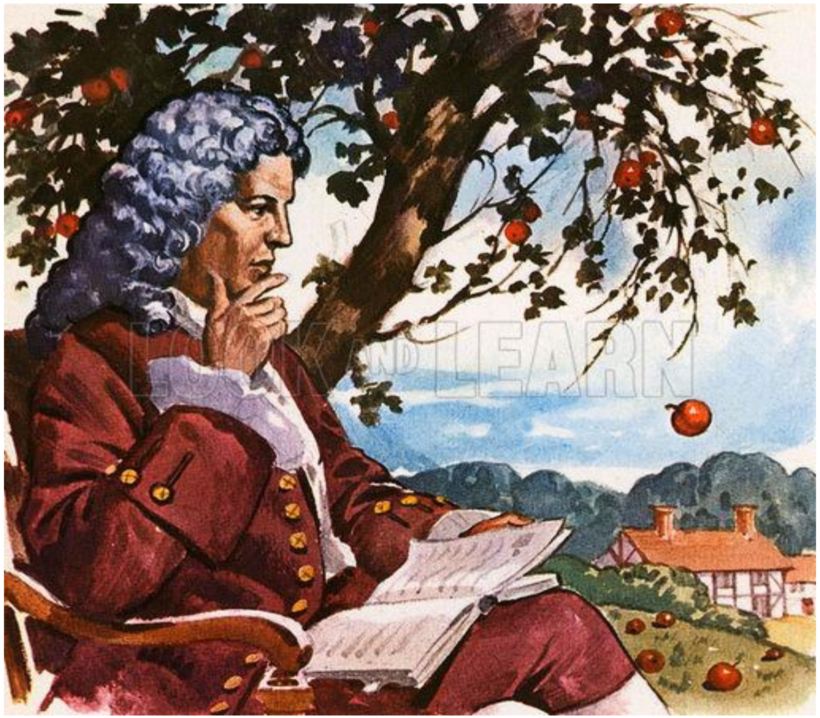
1 **Open Lab** (opening April 2018)

85 researchers in applied mathematics

4 R&D centers, Academic cooperation with Major Universities

The cost of Understanding





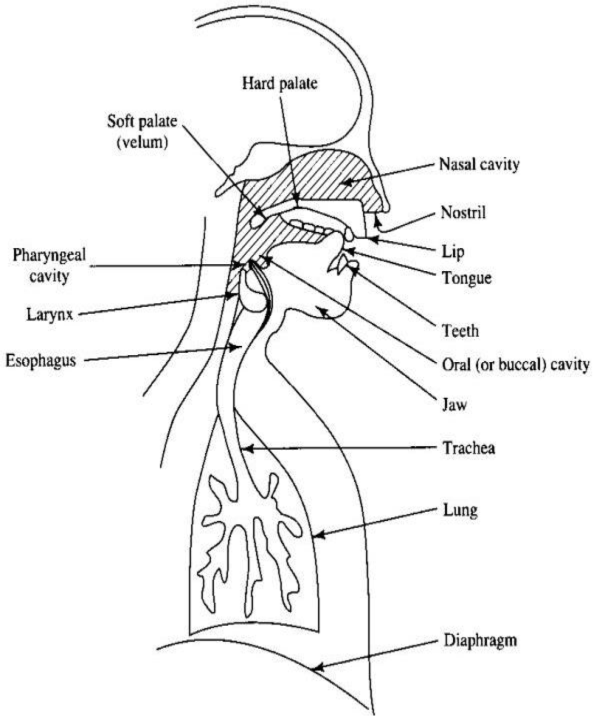
Why it works

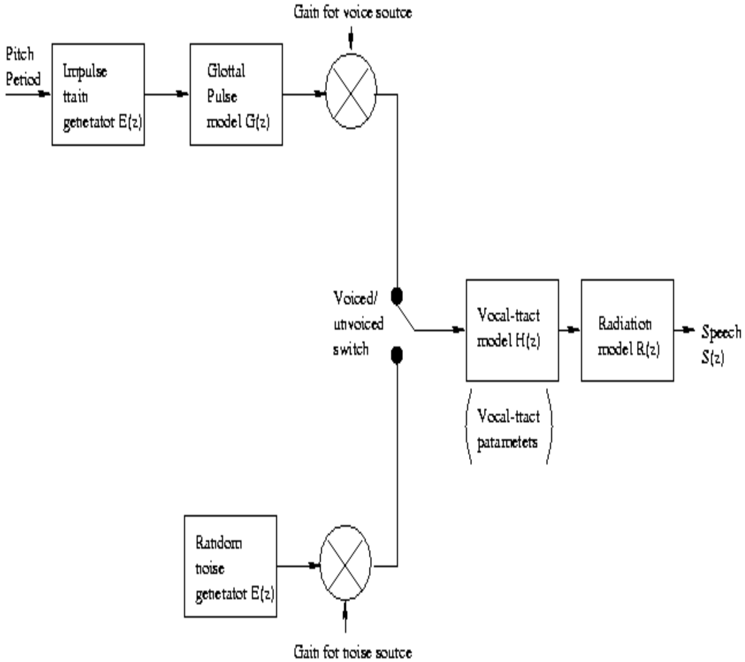
•***Machine-learning algorithms*** have progressed in recent years, especially through the development of deep learning and reinforcement-learning techniques based on neural networks.

•***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.

From Learning to recognize





Why would we learn to optimize in Communications?

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

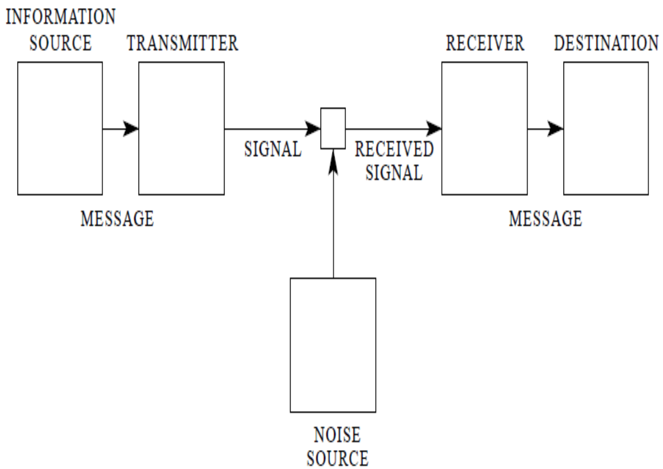
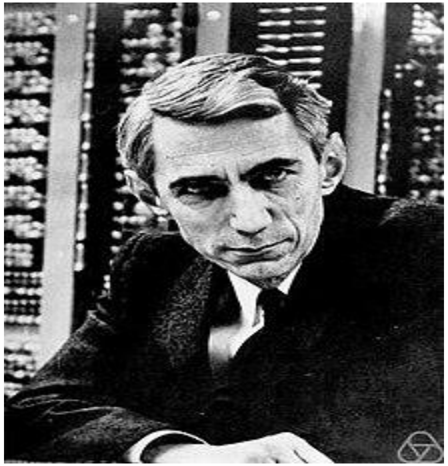
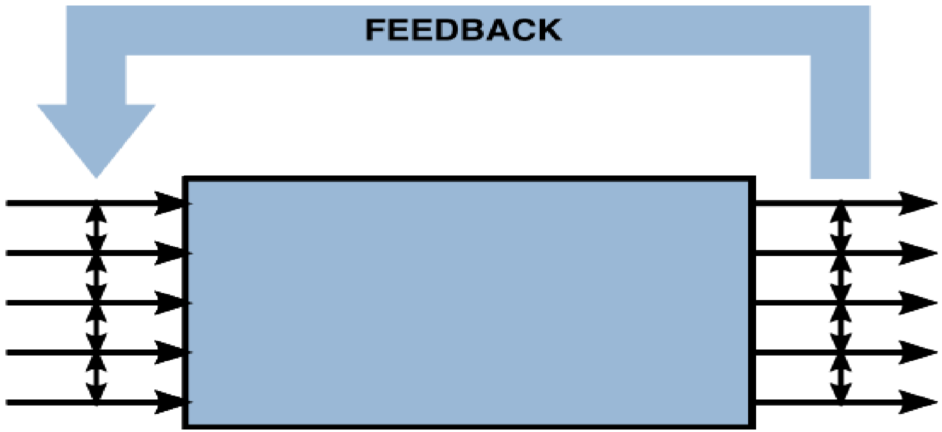


Fig. 1—Schematic diagram of a general communication system.

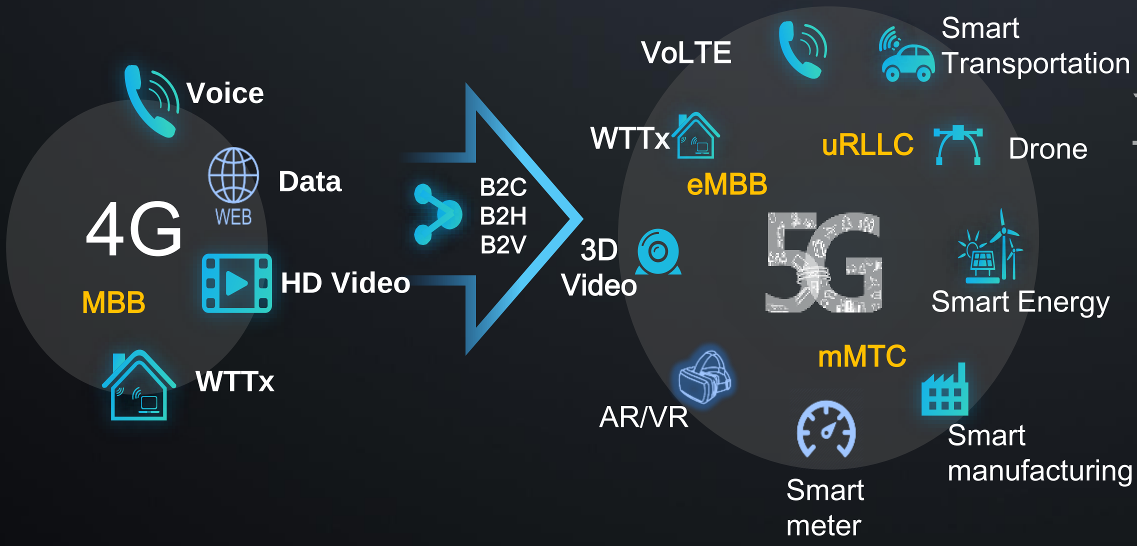


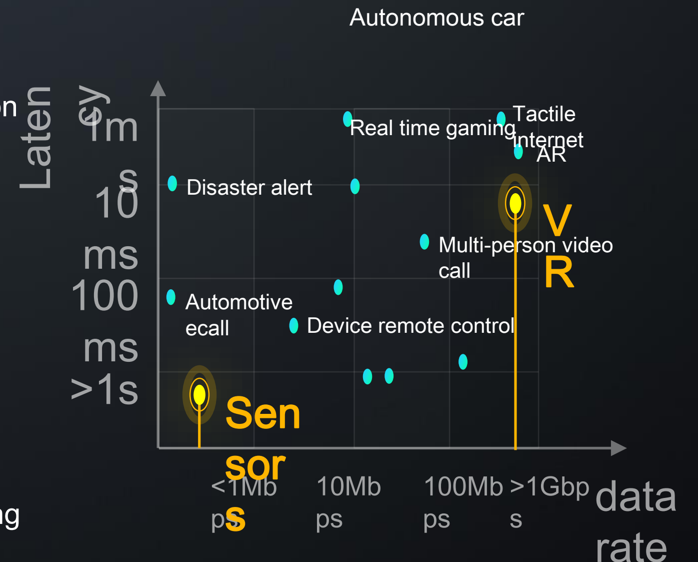
Deep Communications now?



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

Networks are becoming very complex





Source: GSMA/Huawei X Labs

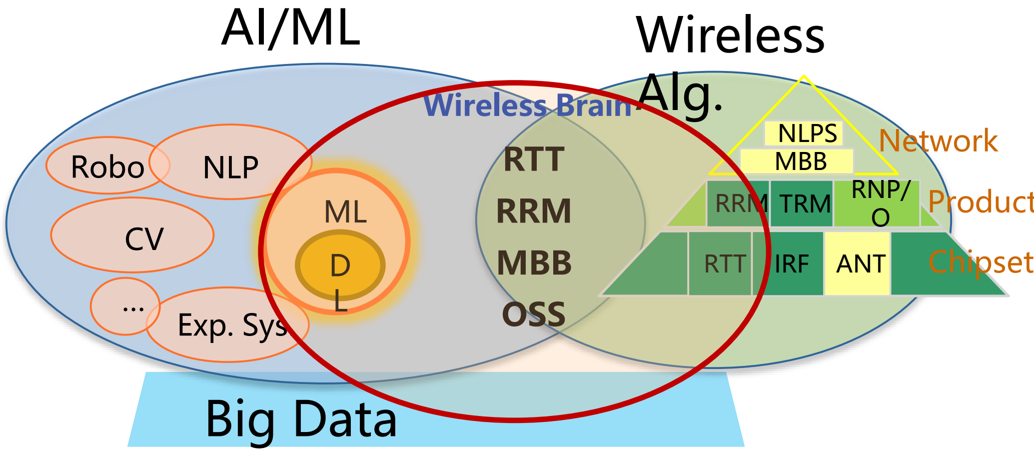
One Network 4 Services One Network Hundreds of Services

Diversified Services SLA Requirement

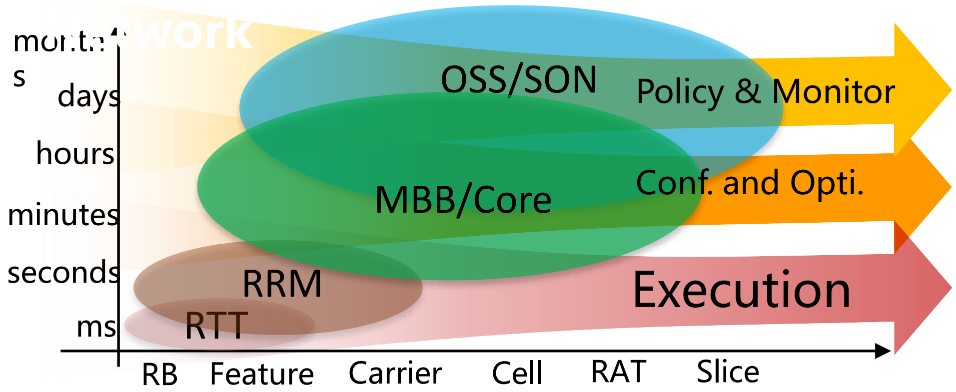
Wireless AI: Key Technology

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

AI in wireless network

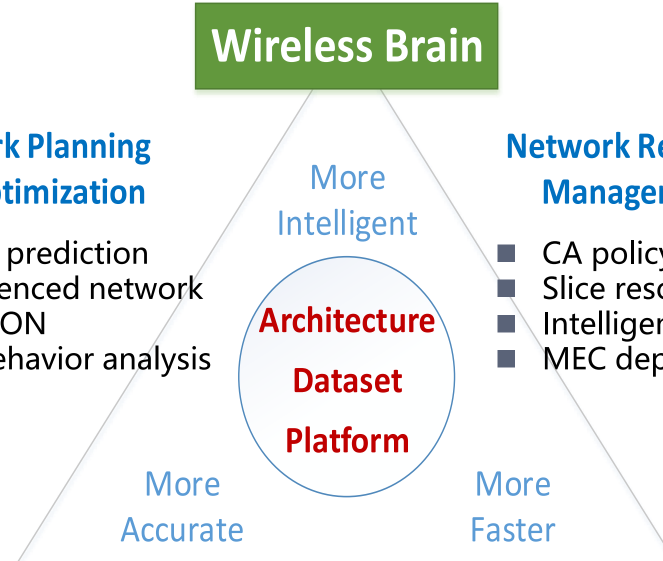
Reconstruct Wireless network using AI technique

Technique trend in AI/ML

- New deep learning network architecture has been proposed
- Rapid development of unsupervised learning
- Development of AI chipset /TPU

Trend when AI/ML used in wireless network

- AI/ML technique is designed into the network pipe, to enable wireless network autonomic
- Improve the network operations
- Reduce the complexity of network fault diagnosis



Network Planning and Optimization

- Traffic prediction
- Experienced network
- AI in SON
- Use behavior analysis

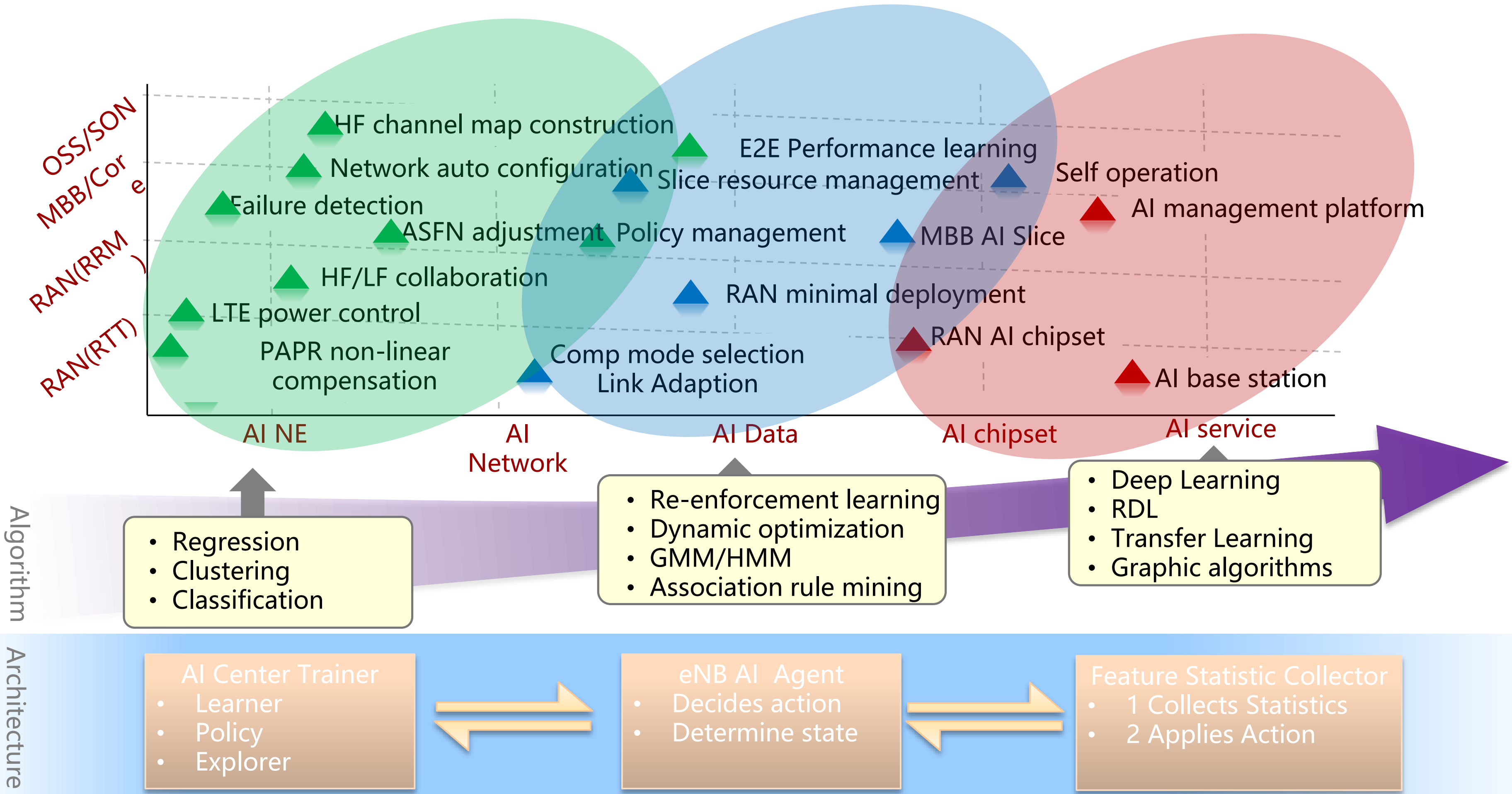
Network Resource Management

- CA policy selection
- Slice resource managemen
- Intelligent base station
- MEC deployment

Failure Detection and Analysis

- Physical Sub-health detection
- VoLTE root cause analysis
- Failure prediction
- Network security risk analysis

AI in Wireless



Example

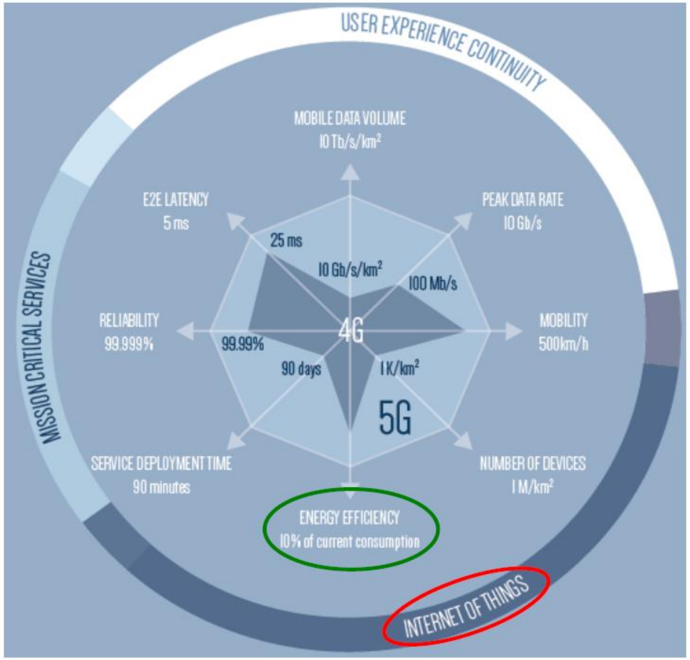
Motivation and Definition of EE

Why do we need energy efficiency in communications?

"Energy efficiency is defined as the number of bits which can be transmitted per Joule of energy. 5G should support a 1000 times traffic increase in the next 10 years timeframe, with an energy consumption by the whole network of only half that is typically consumed by today's networks. This leads to the requirement of an energy efficiency increase of x2000 in the next 10 years timeframe."[1]

"5G will bring drastic energy efficiency improvement and develop energy harvesting everywhere. This energy chase will cover terminal devices, network elements, and the network as a whole including data centers."[2]

Why do we need energy efficiency in communications?



5G Disruptive Capabilities

The *energy efficiency* is defined as the system benefit-cost ratio in terms of amount of data reliably transmitted over the energy that is required to do so.

- Transmit power p in the time slot T, over a bandwidth W, and channel h. We can reliably transmit $TW \log_2(1 + p|h|^2)$ bit of information.
- But we consume $T(\mu p + P_c)$ Joule of energy.
- $\mu \geq 1$ (amplifier non-idealities); $P_c > 0$ (static power consumption, e.g. DAC/ADC, filters, signal processing operation, ...).
- If we set $\mu = 0$ and $P_c = 1$, we fall back to channel capacity.

Summing up, the energy efficiency is

$$\mathrm{EE} = \frac{W \log_2(1 + p|h|^2)}{\mu p + P_c} \qquad [\mathrm{bit/Joule}]$$

- It is called the EARTH power model[3]. More sophisticated models exist (non-linearities in the amplifier, power-dependent static power).

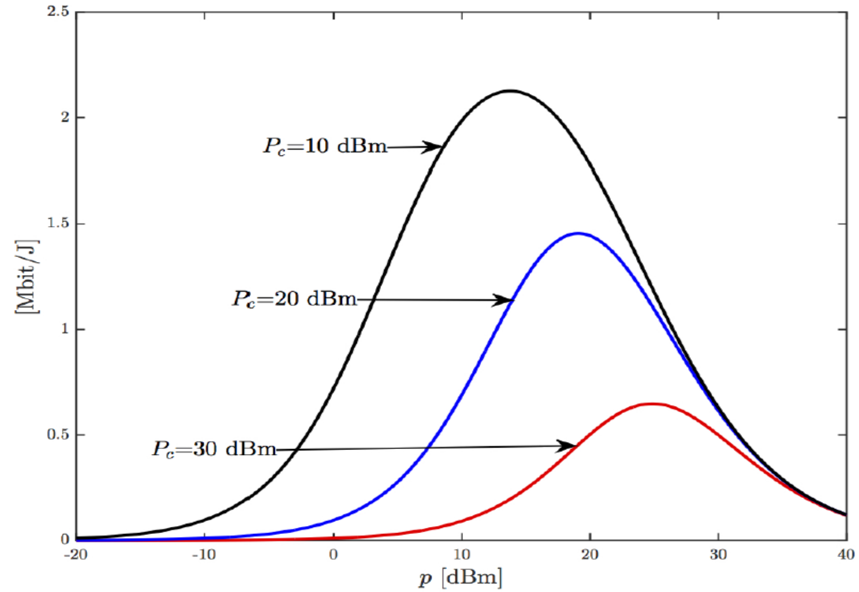


Figure: EE vs p, for fixed transmit directions.

The EE is not concave and does not always increase with the transmit power

What if we have K communication links? How to combine the EEs to define a network EE?

Several choices are possible:

1. Global Energy Efficiency

$$\mathrm{GEE} = \frac{W \sum_{k=1}^{K} \log_2(1 + p_k |h_k|^2)}{\sum_{k=1}^{K} \mu_k p_k + P_{c,k}} .$$

   It is the ratio between the global benefit and global cost of the network, but does not enable to tune the EE of the individual links.

2. Weighted arithmetic mean of the EEs

$$\text{Sum-EE} = \sum_{k=1}^{K} w_k \frac{\log_2(1 + p_k |h_k|^2)}{\mu_k p_k + P_{c,k}} .$$

   It allows tuning the EE of the individual links by a suitable choice of the weights (very useful in heterogeneous networks).

EE maximization is usually written as

$$\arg\max_{\boldsymbol{p}} \ \mathrm{EE}(\boldsymbol{p}, \boldsymbol{H})$$

wherein

- EE could be GEE, or WSEE, or WPEE, or WMEE...
- $\boldsymbol{p}$ is the transmit power vector, $\boldsymbol{H}$ collects all network channels.

But we can also write the problem as

$$F(\boldsymbol{H}) = \arg\max_{\boldsymbol{p}} \mathrm{EE}(\boldsymbol{p}, \boldsymbol{H})$$

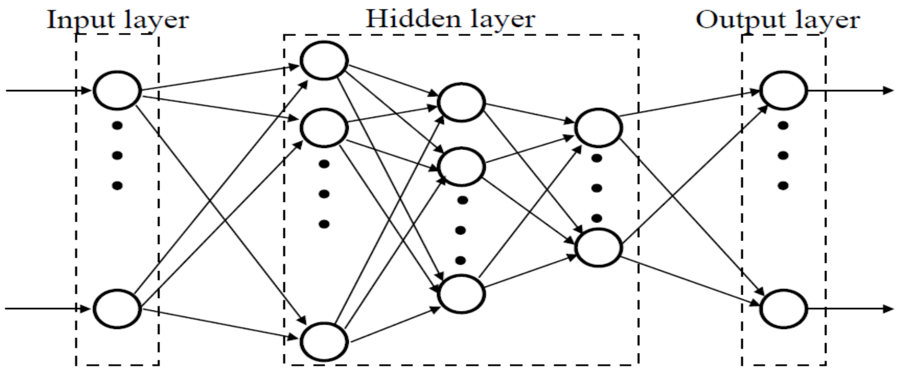
Proposed approach

Feedforward neural networks are universal function approximators [2].
A neural network can *learn* F.

References

[2] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989

Deep Forward Neural Network

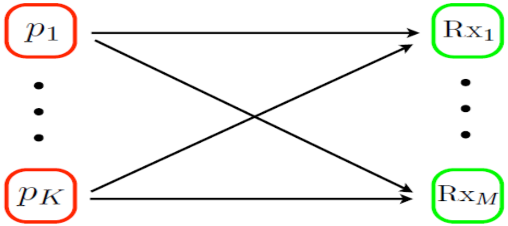


- L fully-connected layers, each having N_ℓ processing units called neurons.
- Data propagates only forward (from input to output)
- Each neuron performs an affine combination of its inputs, applies a non-linear function, and propagates the result.

$$x_{\ell+1}(n) = f\left(\sum_{i=1}^{N_\ell} x_\ell(i) w_{\ell+1,n}(i) + b_{\ell+1,n}\right)$$

GEE Maximization in Interference Limited Networks

Interference network with K transmitters and M receivers.



- The receive filter used by receiver m to decode the data from user k is $\boldsymbol{c}_{m,k}$.
- All receivers have N receive antennas and thermal noise power σ^2.
- Each transmitter k has a single antenna and transmits with power p_k.
- The channel between transmitter k and receiver m is $\boldsymbol{h}_{k,m}$.

The SINR obtained by transmitter k at receiver m is:

$$\gamma_{k,m} = \frac{|\boldsymbol{c}_{m,k}^H \boldsymbol{h}_{k,m}|^2 p_k}{\sigma^2 + \sum_{j \neq k} p_j |\boldsymbol{c}_{m,k}^H \boldsymbol{h}_{j,m}|^2}$$

GEE Maximization Problem

The goal is to maximize the system global energy efficiency (GEE), i.e.

$$\mathrm{GEE}(\boldsymbol{p}) = W \frac{\sum_{k=1}^{K} \log_2(1+\gamma_k)}{P_c + \sum_{k=1}^{K} \mu_k p_k} \ ,$$

subject to the power constraints $p_k \in [0, P_{max,k}]$ for all k, and wherein

- W is the transmission bandwidth.
- $\mu_k \geq 1$ is the inverse of the amplifier efficiency of transmitter k.
- P_c is the total hardware power dissipated in all network nodes.

We use a deep neural network to learn the map:

$$F(\boldsymbol{H}) = \arg \max_{\boldsymbol{p} \in \mathcal{S}} \mathrm{GEE}(\boldsymbol{p}, \boldsymbol{H})$$

with $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]$.

ANN Training & Validation

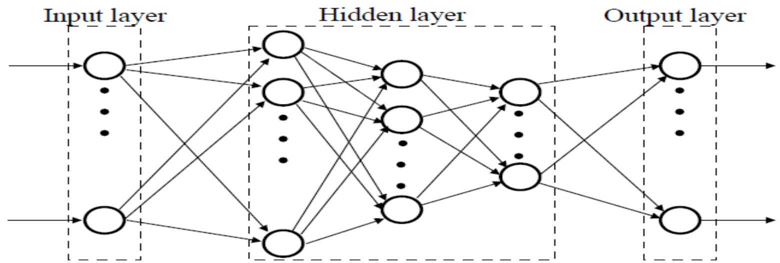
Training and Validation Procedure

- Optimize the weights and bias terms of each layer.
- The training set consists of I pairs $\{(\boldsymbol{H}_i, \boldsymbol{p}_i)\}_{i=1}^{I}$, with $\boldsymbol{p}_i$ the optimal power allocation when the system channels are $\boldsymbol{H}_i$.
- The training set is generated *offline* solving the EE maximization problem.
- 10% of the training set is used as validation set, to set the network hyper-parameters (number of layers, number of neurons, etc...).
- Mini-batch stochastic gradient descent is used as training algorithm.
- No regularization and no batch normalization (very simple setup).

Training and validation
need to be performed only sporadically

Online Implementation

Very low complexity, namely closed-form "optimal" power allocation for yet unobserved channels.



$$x_{\ell+1}(n) = f\left(\sum_{i=1}^{N_\ell} x_\ell(i) w_{\ell+1,n}(i) + b_{\ell+1,n}\right)$$

Network Testing

- Trained network is tested over a test set of J new channel realizations $\{\boldsymbol{H}_j\}_{j=1}^{J}$.
- For each channel $\boldsymbol{H}_j$ the output of the network is compared to the solution of the GEE maximization problem.

Building the training set

If multi-user interference is not present (e.g. ZF receivers are used), building the training set is simple:

- The EE has a concave numerator and an affine denominator.
- Fractional programming (e.g. Dinkelbach's algorithm) can be used to find the EE-maximizing $\boldsymbol{p}$, given any $\boldsymbol{H}$, [1].

If multi-user interference is present, building the training set is harder (but is done offline):

- The numerator of the EE is not concave and EE maximization becomes in general NP-hard.
- For any given $\boldsymbol{H}$, a (near-)optimal $\boldsymbol{p}$ is found by the sequential fractional programming framework [3].
- Alternatively, for any given $\boldsymbol{H}$, the globally optimal solution is found by monotonic fractional programming [4].

References

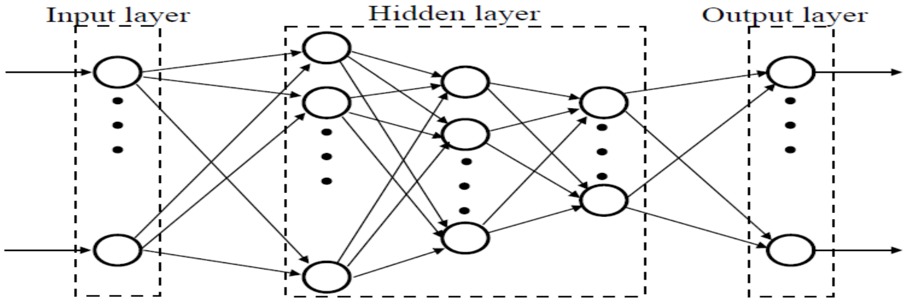
[1] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3-4, pp. 185–396, 2015

[3] A. Zappone, L. Sanguinetti, G. Bacci, E. A. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1668–1683, April 2016

[4] A. Zappone, E. Björnson, L. Sanguinetti, and E. A. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2844–2859, June 2017

Neural Network Architecture

- $N_L = 5$ fully connected layers (no convolutional layers), with $18, 16, 14, 12, 10$ nodes in Layer $1, 2, 3, 4, 5$, respectively.
- Training plus validation of $I = 10^4$ points (split with ratio 0.9). This is small for typical deep networks applications, but low complexity.
- The neural network is tested over a test set of dimension $J = 10^4$.



- $K = 10$ users, $M = 3$ BSs with $N = 10$ antennas each, Rayleigh fading
- $\mu = 10$, P_c generated according to the model from [5]

References

[5] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, 2017

GEE Comparison. LMMSE reception is used.

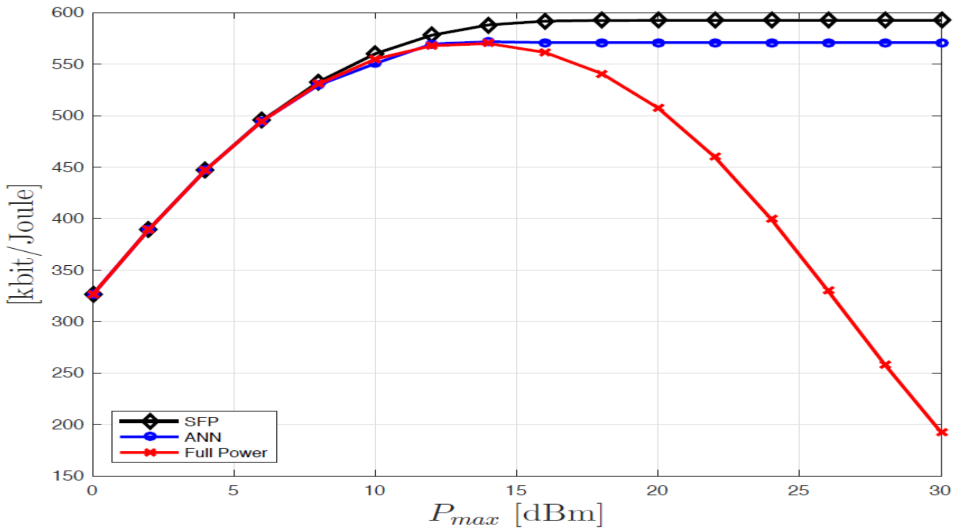


Figure: GEE versus SNR= P_{max}/σ^2 for: GEE maximization by fractional programming (black line); GEE Maximization by ANN (blue line); Full Power transmission (red line)

The maximum of the ratio between the values of the two curves is 1.0224, obtained for SNR=0 dB.

Power Consumption. LMMSE is used

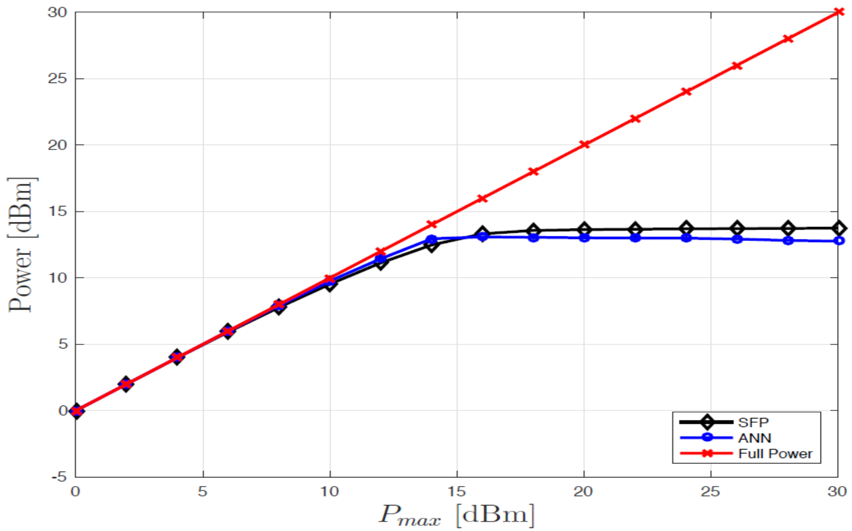
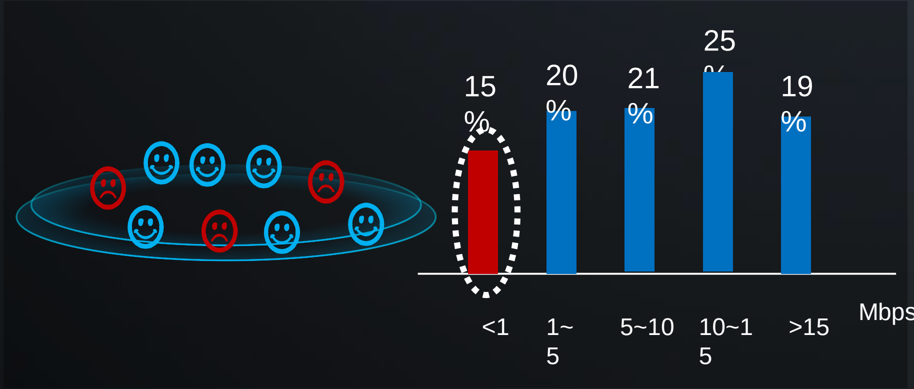


Figure: GEE versus SNR= P_{max}/σ^2 for: GEE maximization by fractional programming (black line); GEE Maximization by ANN (blue line); Full Power transmission (red line)

Example

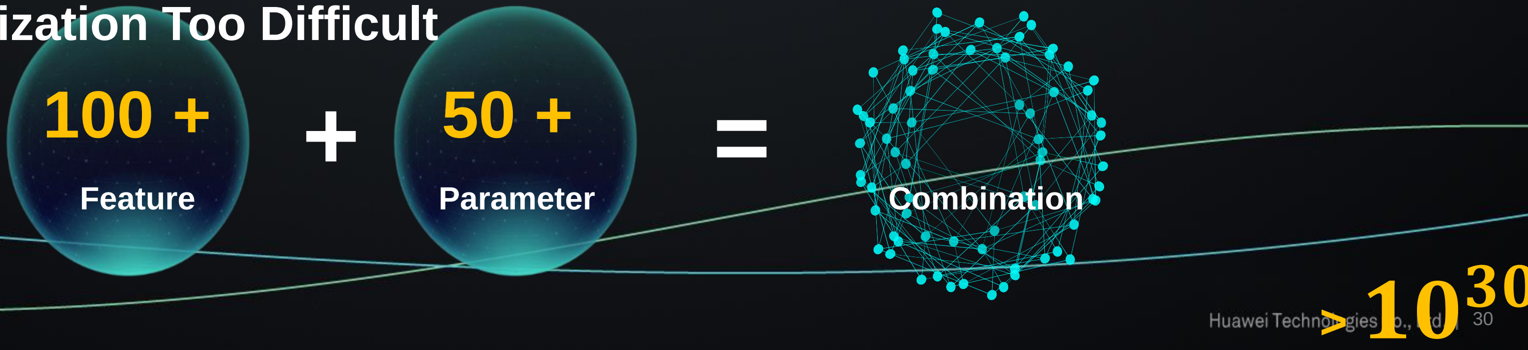
Edge User Throughput and VoLTE Background

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



AI Assisted VoLTE and Edge User THP Optimization

150 Factors

$>10^{30}$ Combination

TopN Factors

240 Combination

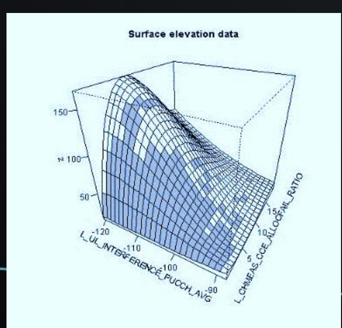
Optima

1 Combination

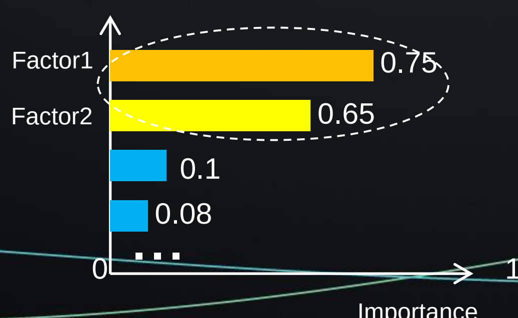
Top N Factors Analysis

Data → Extracting Feature → Modeling → Recognition

Importance Ratio Evaluation



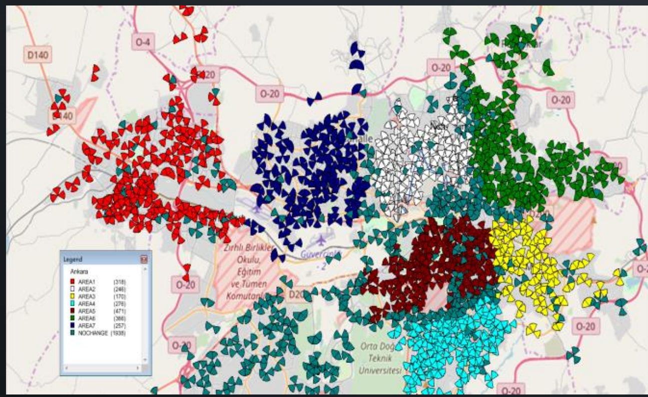
Importance Ratio Rank



Optimal Value Search

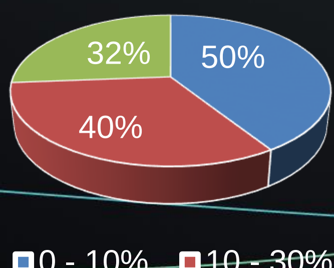
Innovation in Turkey – Edge THP Optimization

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

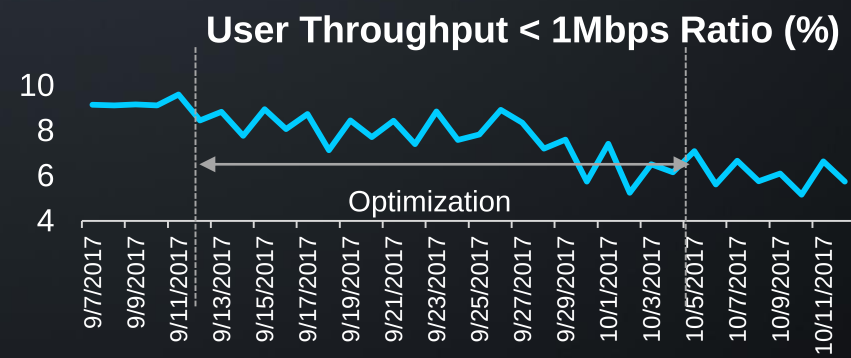


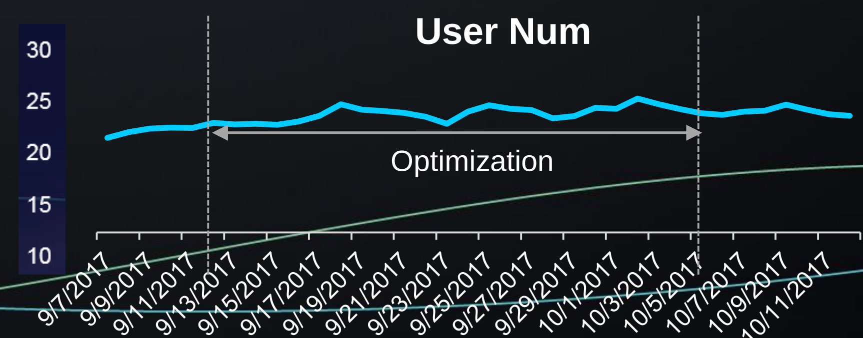
Whole Network Gain:

Throughput Gain Distribution



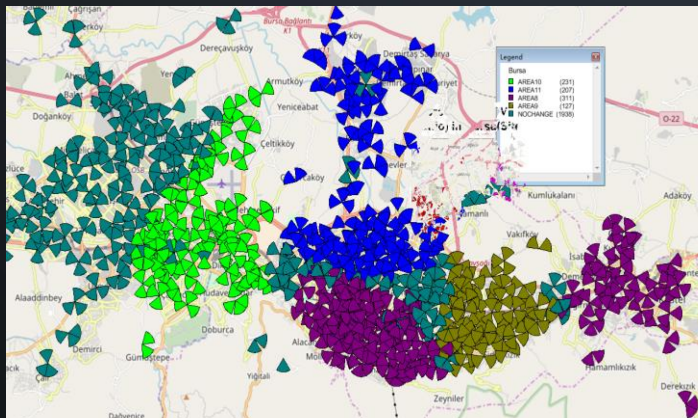
Ankara Top 10% Gain Cells





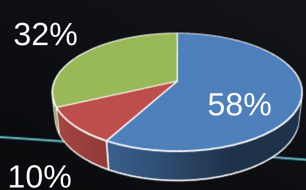
Innovation in Turkey – VoLTE Optimization

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

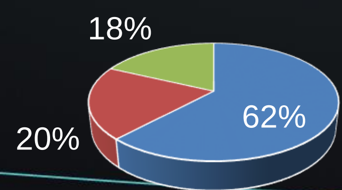


Whole Network Gain:

DL PLR Gain Distribution



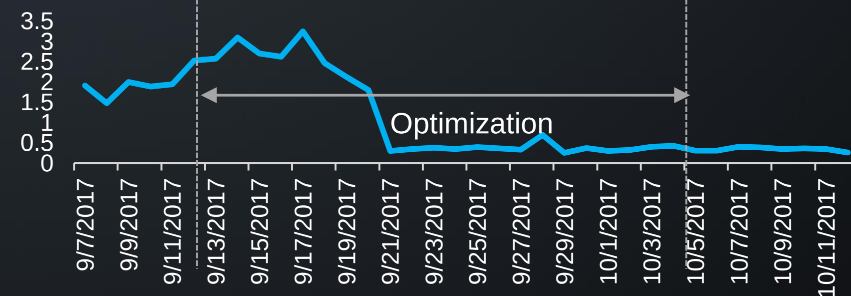
UL PLR Gain Distribution



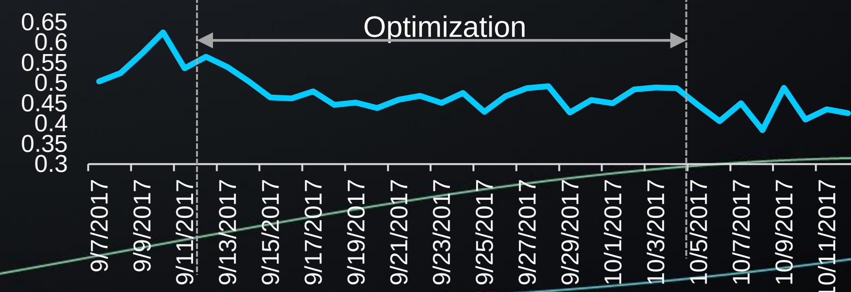
DL/UL PLR Decrease Dramatically

Bursa Top 10% Gain Cells

DL Packet Loss Rate (%)



UL Packet Loss Rate (%)



PLR: Packet Loss Rate

Model versus Data

Wireless AI: From Model-Based to Data-Driven Communication Networks Design

Alessio Zappone[1], Marco Di Renzo[2], Merouane Debbah[1,3]
[1] LANEAS Group, Laboratory of Signals and Systems (CentraleSupelec - CNRS - Univ. Paris-Sud), Paris, France
[2] Laboratory of Signals and Systems (CNRS - CentraleSupelec - Univ. Paris-Sud), Paris, France
[3] Mathematical and Algorithmic Sciences Laboratory, France Research Center, Huawei Technologies, Paris, France

...In writing...
(Invited paper @ IEEE Transactions on Communications)