DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Cognition-based networks: applying cognitive science to wireless networking

Michele Zorzi
zorzi@dei.unipd.it

University of Padova (ITALY)
and UC San Diego (USA)

Outline

- Cognition-based networks
  - Evolution of cognitive communications and networks
  - New holistic concept of cognition-based networks
  - Use of machine learning tools toward this vision
- Examples of application
  - QoE-driven video admission control
  - Context-aware handover optimization

Motivation

- Communication systems are becoming very complex, and call for intelligent solutions
- New technologies make such evolutions real
- For example:
  - White spaces in licensed spectrum (e.g., TV bands), spectrum re-used by a secondary user
  - Self-organizing networks in ad hoc (e.g., disaster) scenarios
  - Advanced paradigms in HetNets
- All of these are specific cases of a more general approach based on learning and context-awareness

Cognition applied to wireless

- Applying cognition is a way to deal with the complexity and challenges of future systems
- Cognition is already in use today in several cases
  - Cognitive radios, biologically inspired networks, node adaptation by learning, etc
- However, in order to draw the most benefit from this approach, one needs to
  - Consider all players in a more coherent manner
  - Apply cognition at all network layers and end-to-end
  - Apply the most advanced paradigms taken from cognitive science

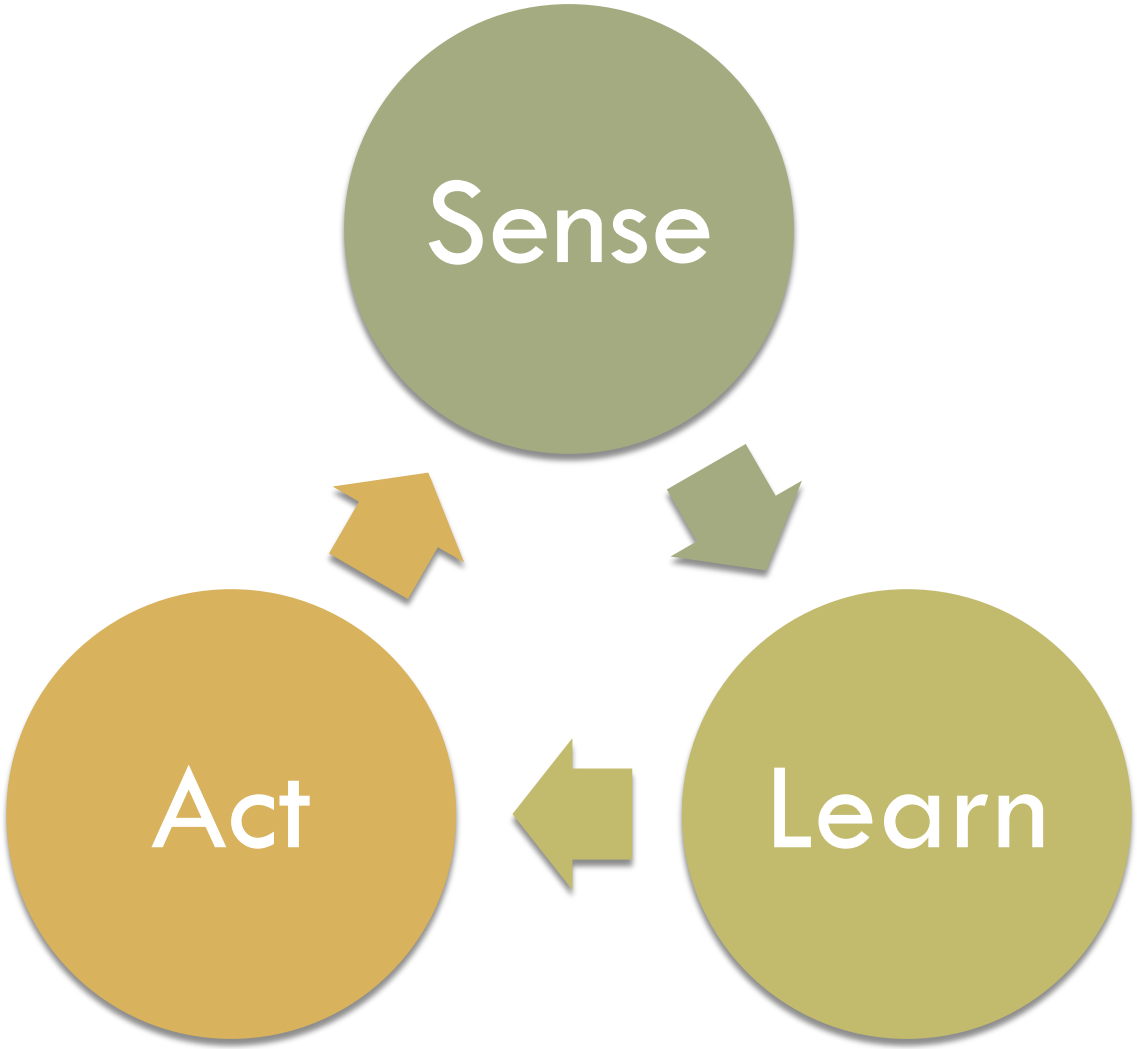
Cognition applied to wireless

- Mitola (2000) and Haykin (2005) actually gave a very general definition of the cognitive paradigm
- They spoke about the essence of cognition, including
  - Intelligent observation, learning, decision-making, emergent and collaborative behaviors
- It is clear that what has happened in this field since then has only scratched the surface
  - Cognitive networking in a broad sense remains an exciting and largely unexplored research field

Evolution of the cognitive paradigm

- Cognitive radios: frequency-agile devices for opportunistic access
- Cognitive radio networks: networks of CRs
- Cognitive networks: the cognitive approach applied to networking layers and end-to-end
  - Learning, decision-making, information sharing
- **Cognition-based networks:** drawing from the most recent results in cognitive science
  - Advanced unsupervised learning, emergent behaviors, evolutionary computation, model building, collective intelligence, combined with reconfigurable, software-defined communication techniques
  - Including out-stack information (related to the environment & the user)

The cognition cycle (simplified)



True for humans, true for networks

- **Sense**: nowadays devices are crammed with transducers / sensing apparatuses
  - needs efficient data handling
- **Learn**: optimization algorithms can be run at each node individually
  - needs (i) efficient algos (ii) harmonization
- **Act**: network modifies the environment
  - requires convergence of multiple devices

Supervised vs unsupervised learning

- Supervised learning requires a training set and/or explicit feedback about outcomes
  - Suitable when this is available, and when the goal is well-defined and known a priori
  - However, it has been shown to fail in some cases
- Unsupervised has no prior knowledge
  - No pre-existing model, more general
  - Reveals features that are not predefined, leading to the development of a data-driven worldview
  - Can be used in distributed optimization

Unsupervised learning

- The way we learn without prior information
- Cognitive stimuli are processed
  - We build a view of the world based on data
  - Generative model: probabilistic view of the world
  - Background for all our cognitive activities
- Has the potential to give an agent the ability to face and react to situations never seen before
- Can use the huge amount of data available
- The worldview provided will be an extremely valuable starting point for goal-specific learning

Learning in networks

- So far, we have discussed learning in nodes
- When dealing with populations of agents, we can refer to more complex phenomena, e.g.:
  - Evolutionary approaches mimic natural evolution
  - Emergent behaviors in agent based systems
  - Intelligent behaviors may emerge from simple entities (swarm intelligence)
- Network optimization as emergent property
- Nature has solved the scalability problem

Cognition-based Network

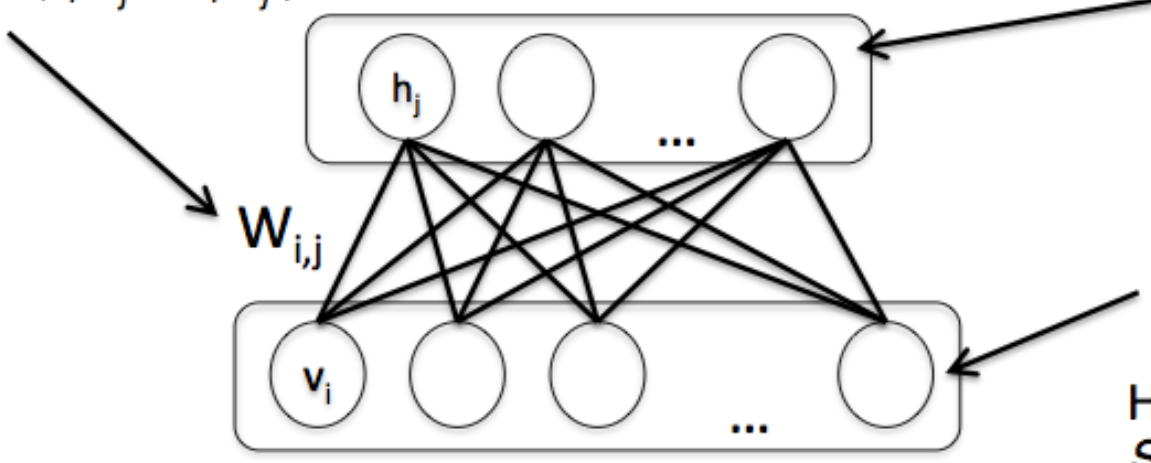
- Each node of the network:
  - exploits local information to achieve its goal
  - shares it with its neighbors

- Self-adaptation to the environment to achieve network wide goals
- Cognition applied to the entire network (not just at the PHY and MAC layers)
  - Both vertically and horizontally

Restricted Boltzmann Machine (RBM)

- Stochastic neural network with input layer symmetrically connected with a hidden layer of feature-detectors
  - (probabilistic graphical model: undirected graph)
- Unsupervised learning of an internal model of the data (features or latent causes).
- Objective function: minimize contrastive divergence (Kullback-Liebler) between input data and (top-down) reconstruction of the data

$$W_{i,j} = (v_i^+ h_j^+ \quad v_i h_j)$$



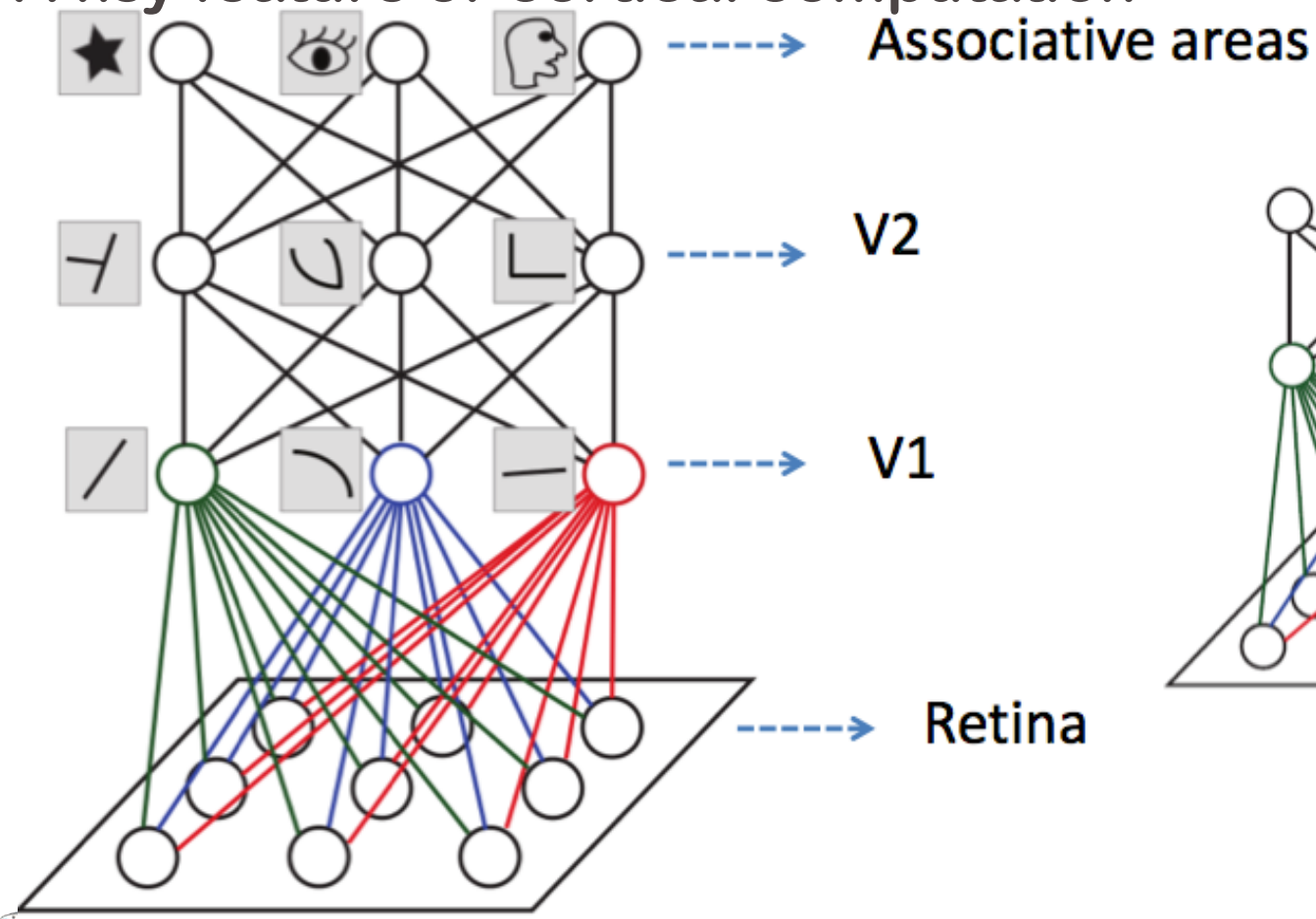
Hidden neurons
(feature detectors)

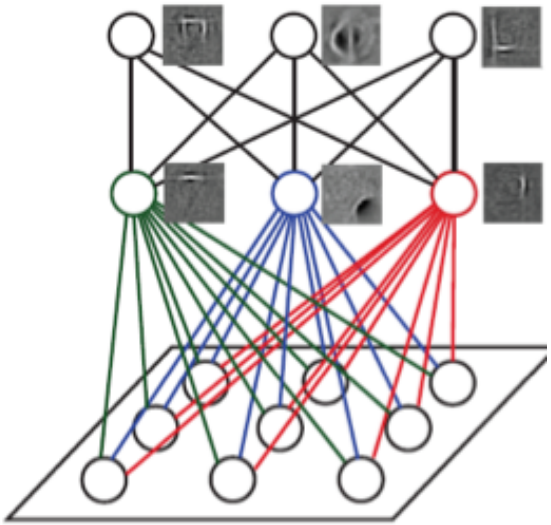
Visible neurons

Hinton & Salakhutdinov, 2006, *Science*

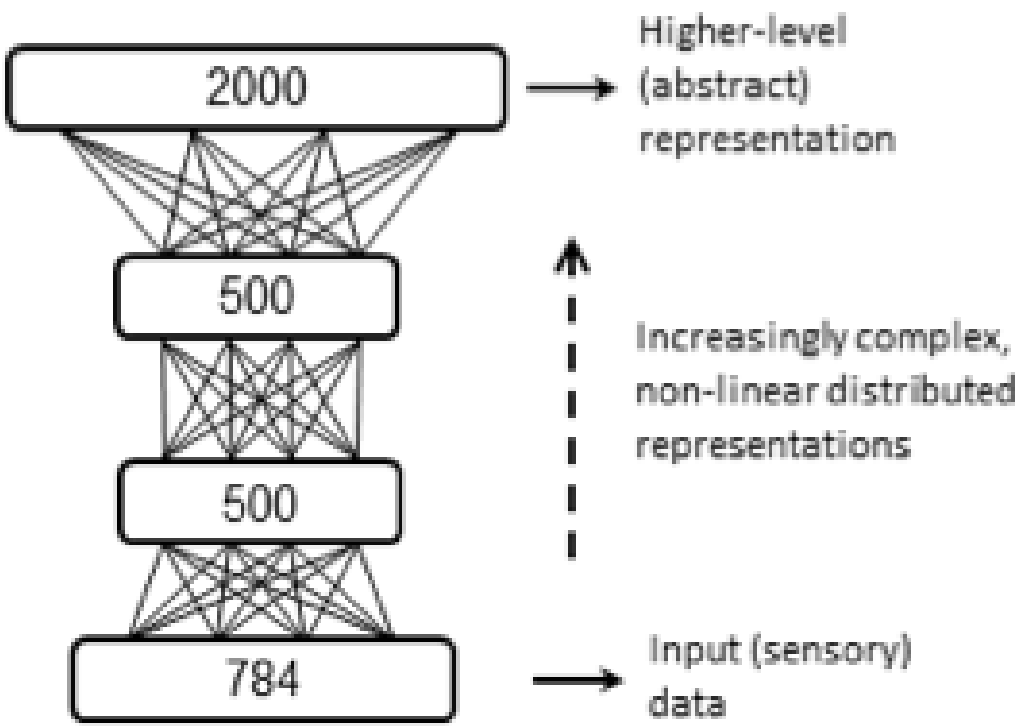
Hierarchical processing

A key feature of cortical computation





Deep learning



- Bottom-up:
  - Higher layer neurons "activate" when input presents some specific features
  - Higher layer provides an abstract representation of input features
- Top-down:
  - Activating hidden layer neurons according to their weight and propagating back toward inputs we can generate signals with similar features of training signal

Example of learning/generation

Large scale simulation
Deep network 9 layers (1 billion connections)
10 million (200x200) unlabeled images

(Le et al., 2012)

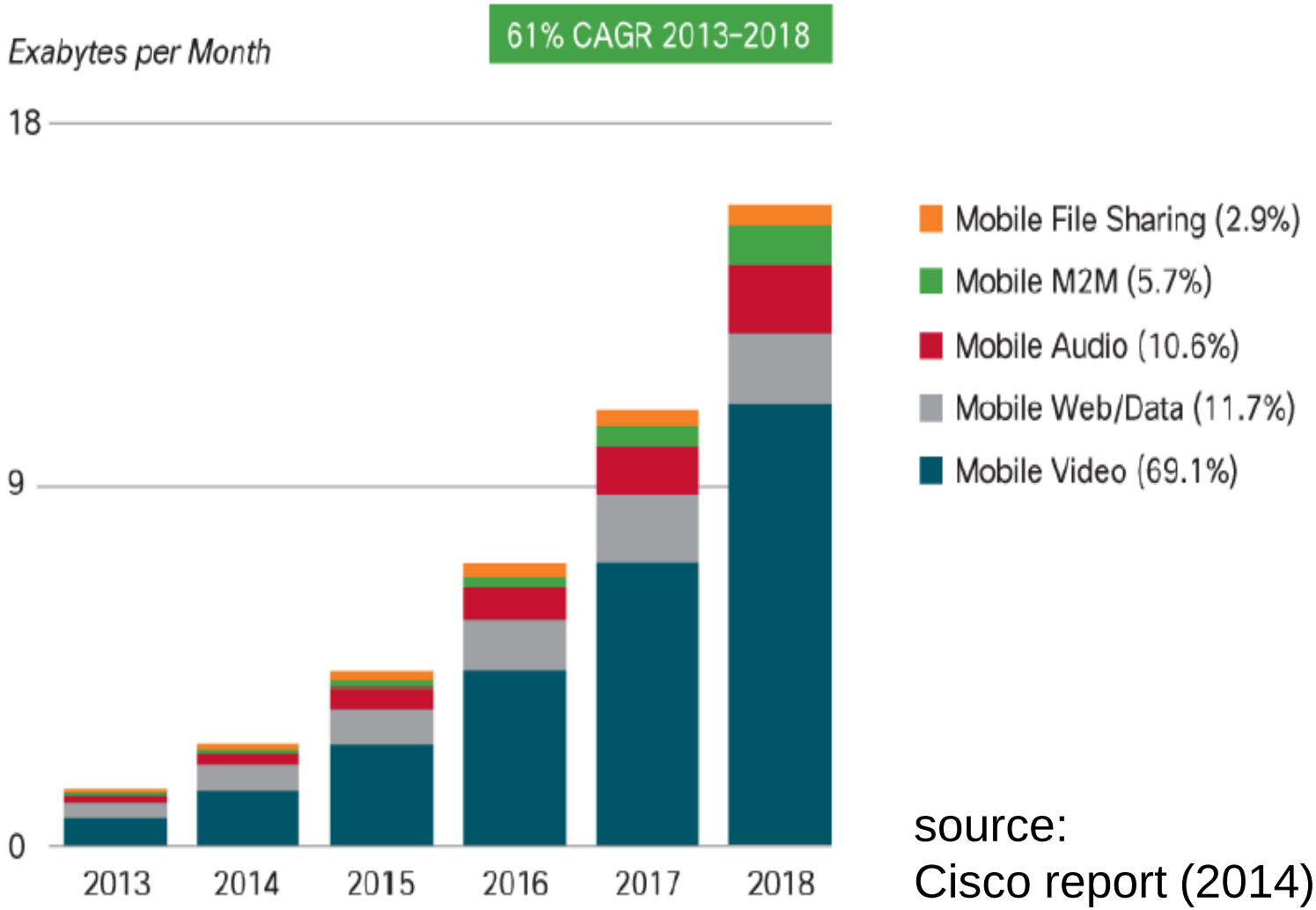
Advantages of deep learning

- Classification objective: divide inputs according to certain criteria
- Standard approach: supervised training of classifier
  - e.g., linear classifier, or neural network
  - representative set of input signals with associated classes
  - Apply classifier to new signals and look at outputs
- Classification is (often) better if classifier is trained by using higher-layer deep network neurons as inputs in place of original signal

Examples of application

- There are many instances where learning on a network-wide basis can be used
- Here we consider two of them, i.e.,
- Content-based video management
  - Learn video features and apply this knowledge to some useful networking task
- Context-dependent handover in HetNets
  - Learn environmental features and make context-based handover decisions accordingly

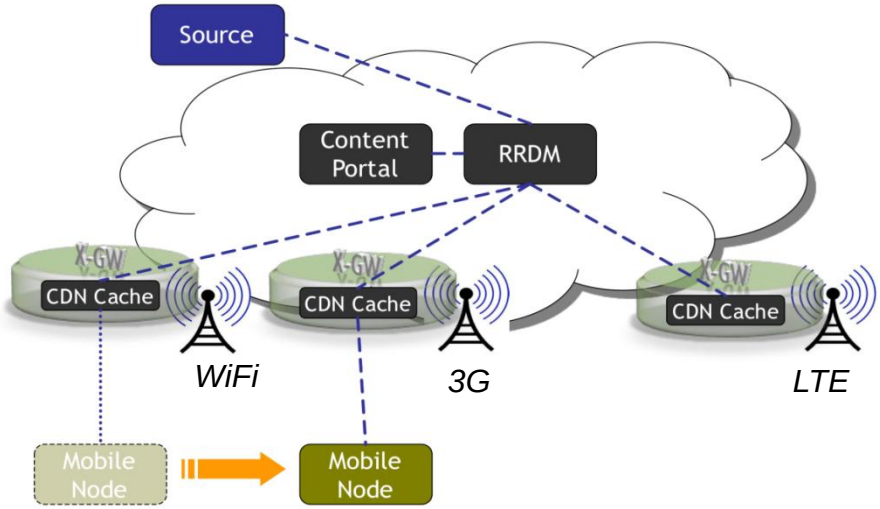
Multimedia traffic growth



source:
Cisco report (2014)

Challenges for video systems

- Heavy data
- Hierarchical system
- Backhaul network capacity



- Must handle different access techniques
- Needs to account for video popularity, heterogeneous user terminals
- Quality-of-Experience of video is hard to capture

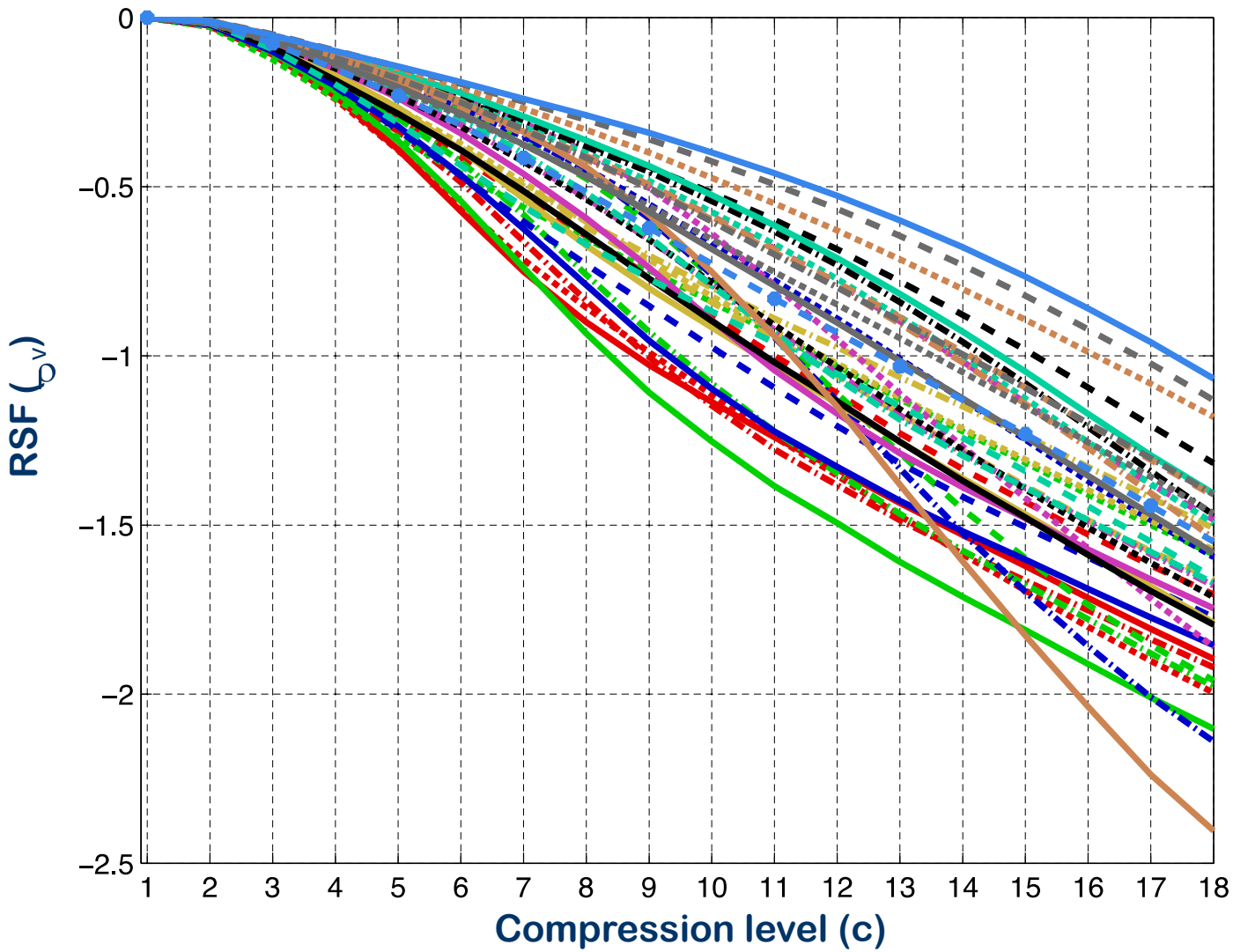
Our approach

- We propose a way to facilitate video handling
- Proposed approach:
  - Represent video characteristic in terms of rate
  - Capture the relationship between rate and QoE
  - Use this to determine resources needed
  - Make admission decisions based on QoE
- We consider a set of video clips and apply machine learning techniques

Analysis

- We consider a test set of 38 video clips, all encoded in an H.264-AVC format
- All the videos are encoded with a 16-frame structure (1 I-frame, 15 P-frames) and compressed with 18 different quantization strategies
  - Transmit rate [bit/s] of video v at compression level c: $r_v(c)$
  - Rate Scaling Factor (RSF): $\rho_v(c)=\log(r_v(c)/r_v(1))$

Rate Scaling Factor *vs* compression



QoE characterization

- Depending on the content, the perceived quality of a given compression level changes
- There are several metrics to measure quality of a video signal
- Here, video quality is expressed in terms of Structural Similarity (SSIM)

Structural Similarity (SSIM)

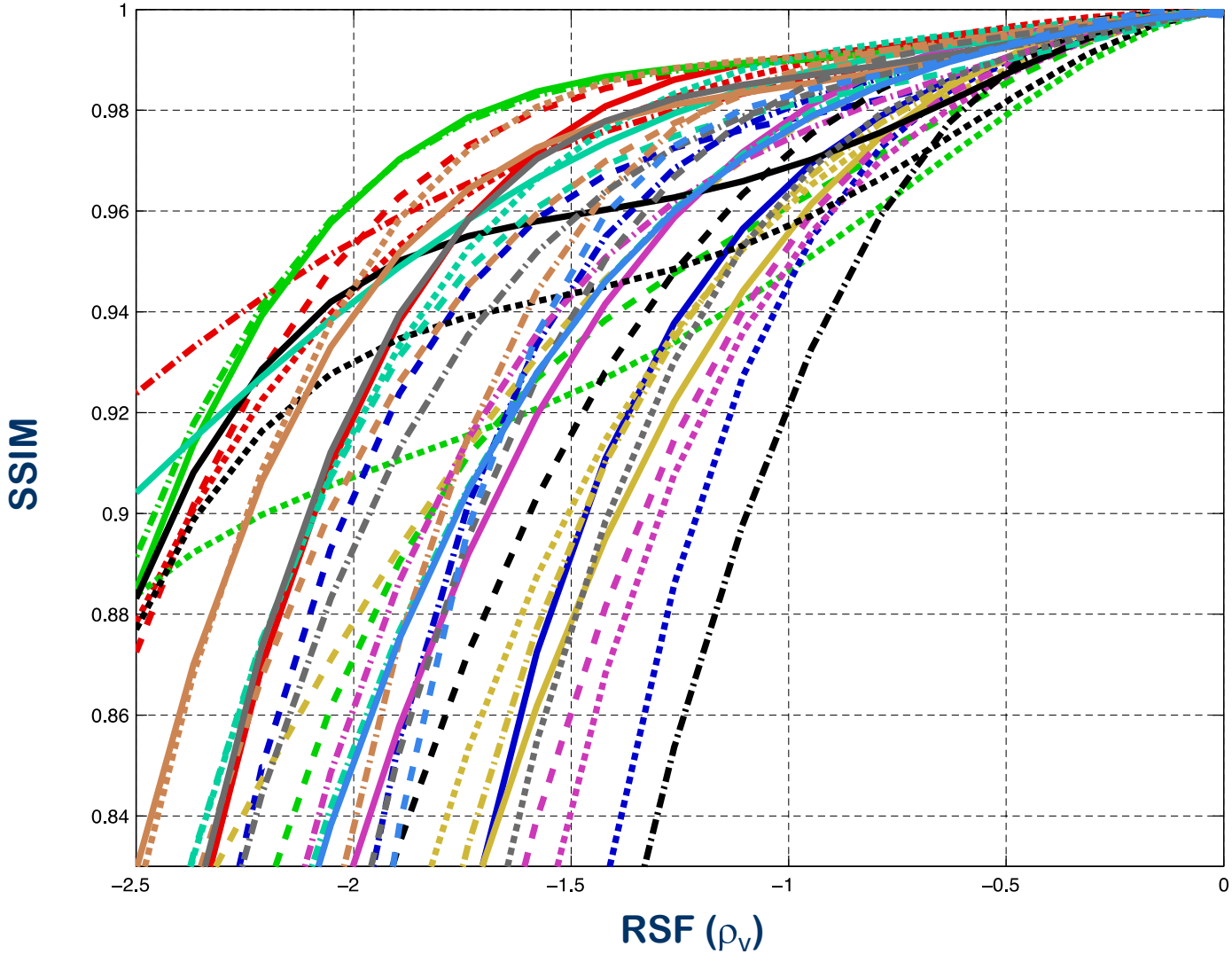
- SSIM measures the closeness of square sets of pixels, and is computed as

$$SSIM(X,Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)} \qquad (1)$$

with μ and σ^2 denoting the mean and variance of the luminance value in the corresponding window, and c_1 and c_2 being variables to stabilize the division with weak denominator

  - Measures image degradation in terms of perceived structural information change
  - represents quality as seen by the human eye

SSIM versus RSF



Remarks

- All the videos exhibit similar trends
  - monotonic descent
  - a steep "fall" after a threshold
- However, there are quantitative differences
  - different perceived end quality
  - different resource requirements
- These characteristics are roughly consistent within the same (homogeneous) video

SSIM polynomial approximation

- We introduce a polynomial approximation to express SSIM behavior
  - This provides a compact representation for use in VAC and RM

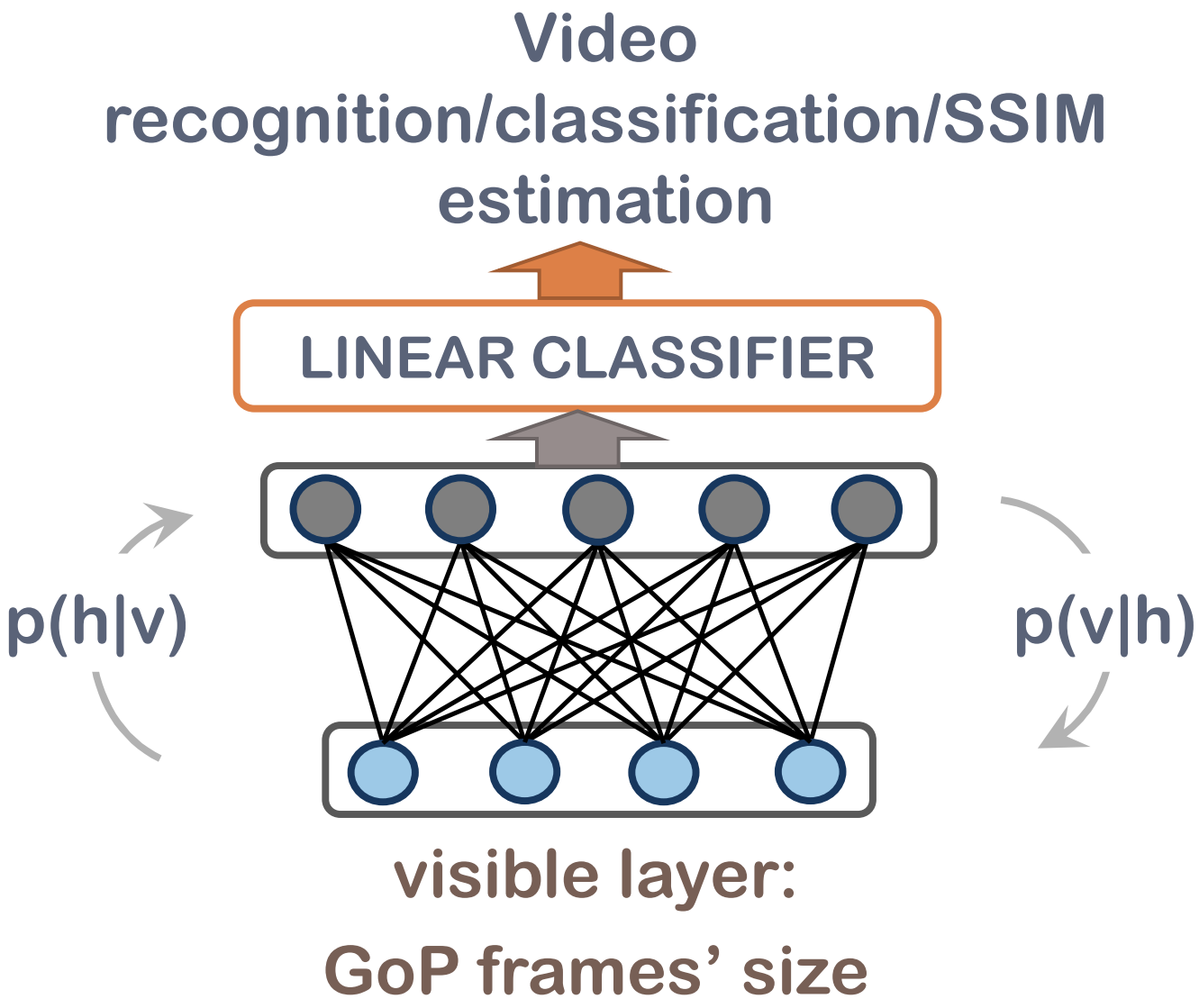
$$F_v^{(n)}(\rho) \simeq 1 + a_{v,1}\rho + a_{v,2}\rho^2 + a_{v,3}\rho^3 + \ldots + a_{v,n}\rho^n$$

- A 4-degree polynomial provides a quite accurate approximation of the SSIM vs RSF curve

Possible video applications

- Knowledge of these characteristics of the video may be useful for
  - (i) QoE-aware admission/congestion control
  - (ii) determining popular content
  - (iii) inferring user behavioral characteristics
- Two different tasks can be envisioned:
  - video recognition (exactly identify it)
  - video classification (just categorize it)

Proposed approach



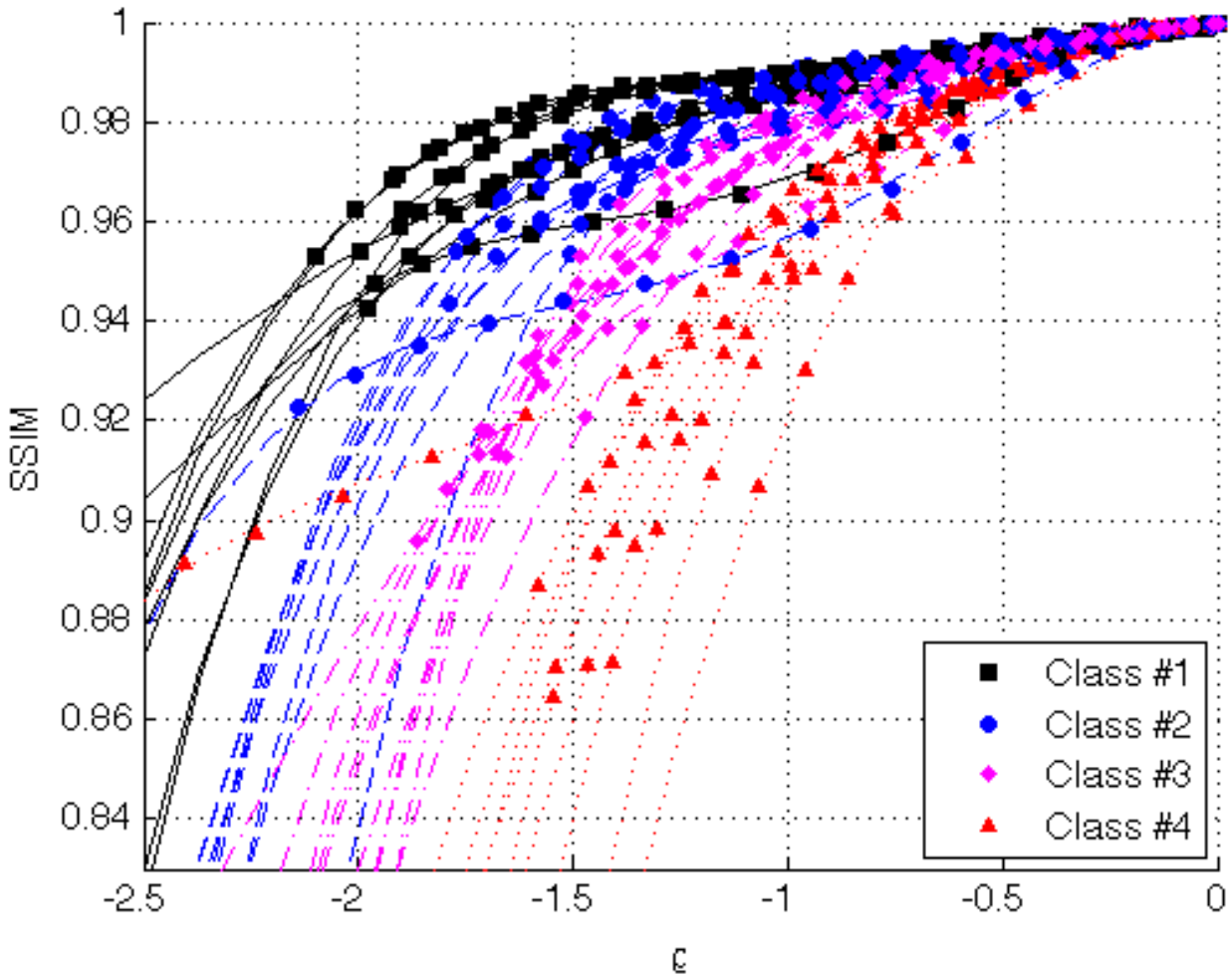
Proposed approach

- Input to the RBM: **frame size only**
  - this is done for a whole GoP (isolated)
  - The RBM "learns" by creating certain patterns in the hidden layer
  - This enables a sparser representation of the input in the hidden layer
- After that, we apply a linear classifier for recognition / classification / SSIM estimation
  - Note: we must train a different linear classifier for each case, while the RBM is trained just once for all

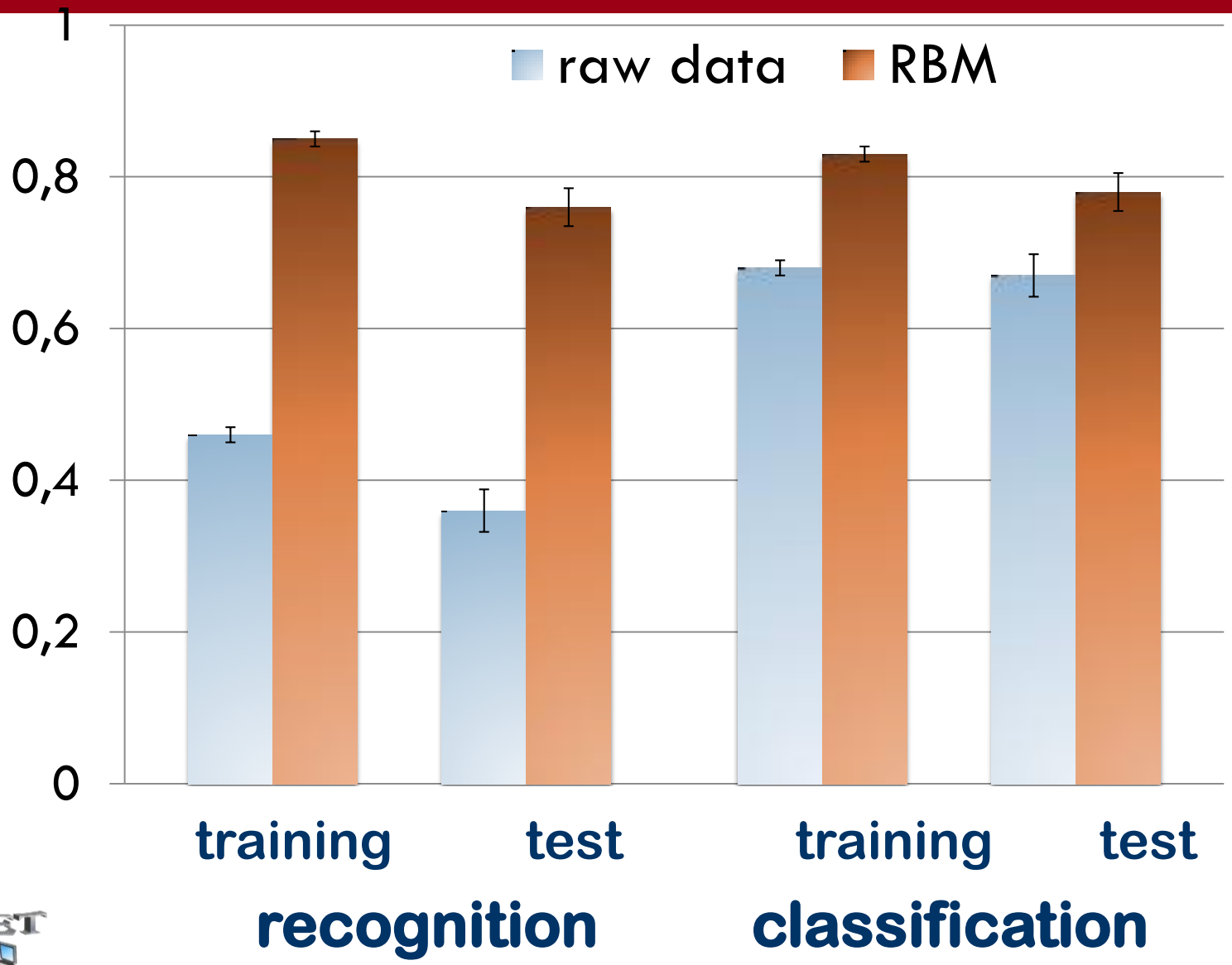
Setting the learning machine

- Dataset is split: Training Set - Test Set
- Input of the RBM (visible layer): 32 units
  - for each of the 16 frames, size of the uncompressed version and of the version compressed at intermediate rate 9
- Hidden layer set (empirically) to 70 units
- The result is compared to just using the raw data as the input of the classifier

SSIM-based Video Classification

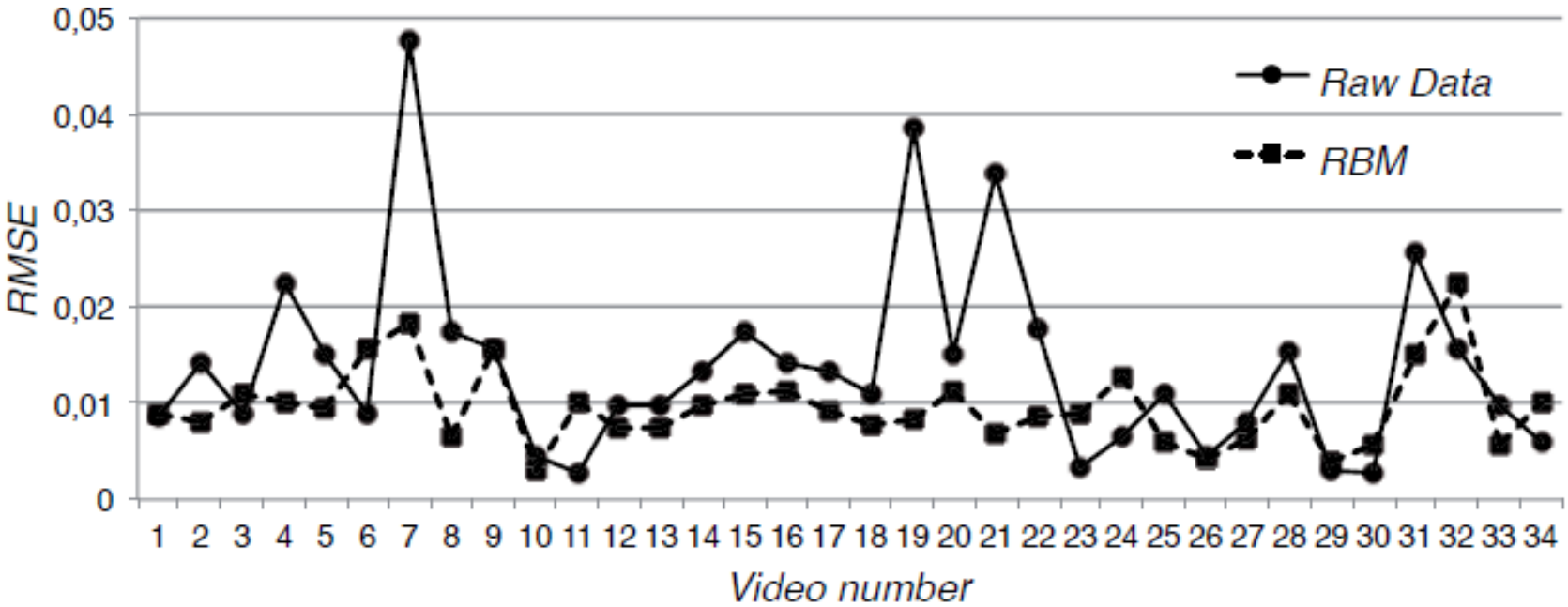


Preliminary results



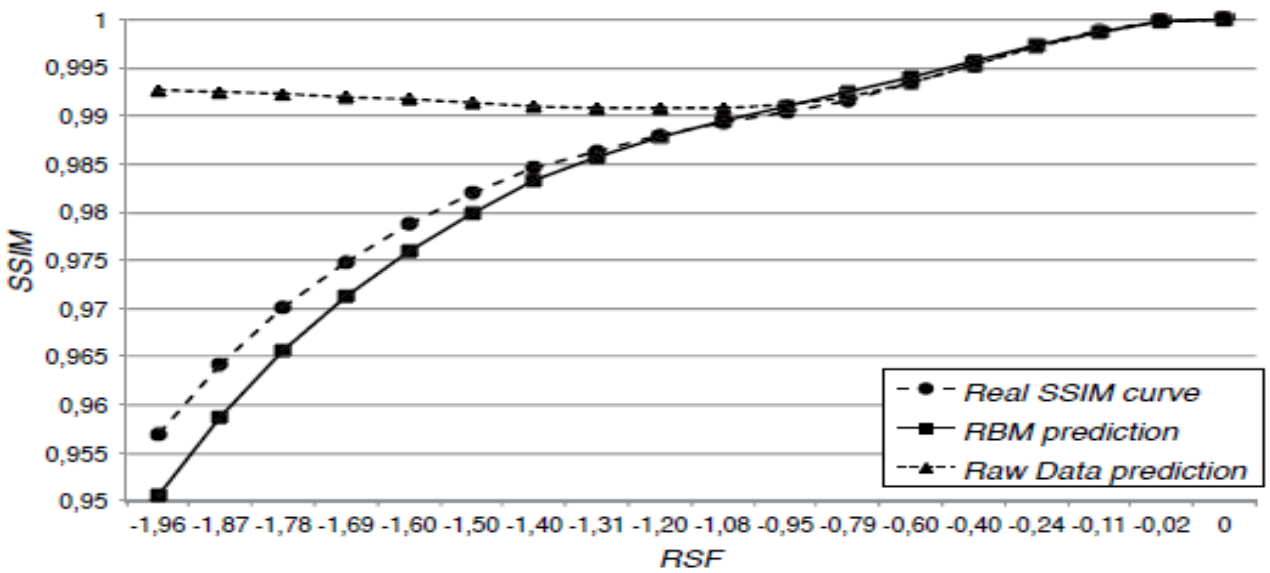
Learning via RBM: accuracy

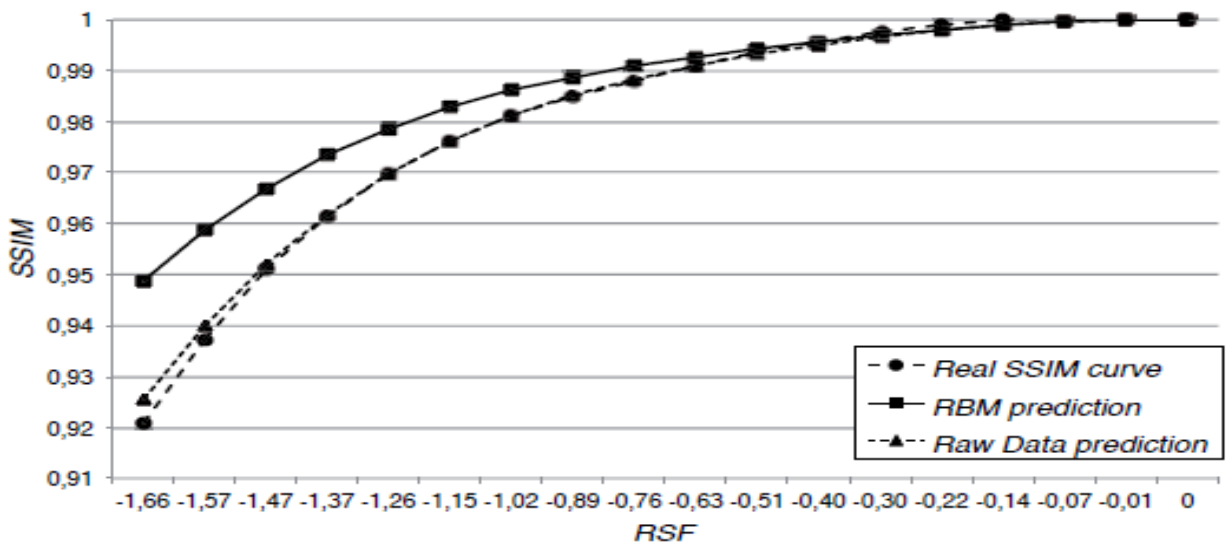
- Root Mean Square Error (RMSE) between exact SSIM and polynomial approximation with estimated coefficients



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Exact vs. estimated SSIM curves for two random videos





:orzi@dei.unipd.it

CognitiveVideo Admission Control & Resource Management

- Videos multiplexed into a shared link of capacity R
- **Resource Manager (RM)**: detects changes and triggers optimization to adapt video rates to maximize QoE utility function
- **Video Admission Controller (VAC)**: determines whether a new video request can be accepted without decreasing QoE of any video below a threshold F*

Cognitive VAC

- At each new video flow request, VAC invokes RM to get the "best" resource allocation according to a specified policy
- RM returns the resource that can be assigned to each video
- VAC computes the SSIM of each video with the best compression level, compatibly with the allotted resources
  - If estimated SSIM of all active videos is above the quality threshold the video request is accepted

Cognitive RM

- The optimization problem addressed by RM is as follows

Resource share allotted to video "v"

Utility function

$$\Gamma_{\text{opt}} = \arg\max_{\Gamma} U(\Gamma, R, \{F_v\}) \quad \text{s.t.} \quad \sum_v \gamma_v \leq 1$$

Rate allocation vector

Channel rate

SSIM videos' characteristics

zorzi@dei.unipd.it

Cognitive RM algorithms

Possible utility functions

Rate fairness (RF) $$U(\Gamma, R, \{F_v\}) = \left[\sum_v \left| \log \frac{\gamma_v R}{r_v(1)} \right| \right]^{-1}$$

SSIM fairness (SF) $$U(\Gamma, R, \{F_v\}) = \min_v F_v(\log(\gamma_v R / r_v(1)))$$

Resource share to be allotted:

RF: $$\gamma_v = \frac{r_v(1)}{\sum_j r_j(1)}$$

SF: $$\gamma_v = r_v(1) 10^{F_v^{-1}(\phi)}$$

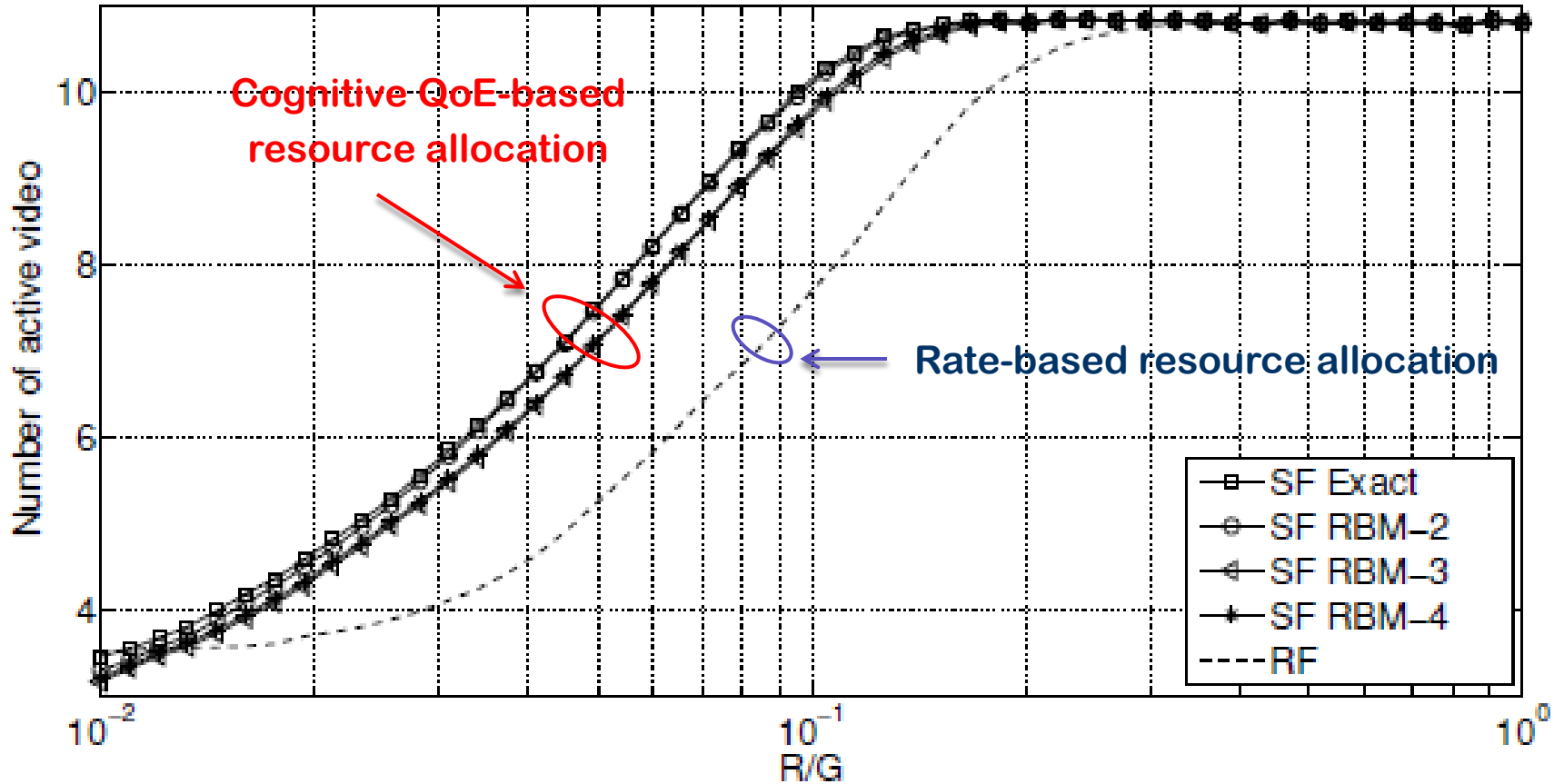
where $$\phi = \arg\max_\phi \left\{ \sum_v r_v(1) 10^{F_v^{-1}(\phi)} \leq R \right\}$$

Simulation setup

- Poisson video requests (0.66 req./s)
- average offered load of 11 videos
- Aggregate rate of G = 161 Mbit/s
- F* = 0.95 SSIM value to reach MOS 4

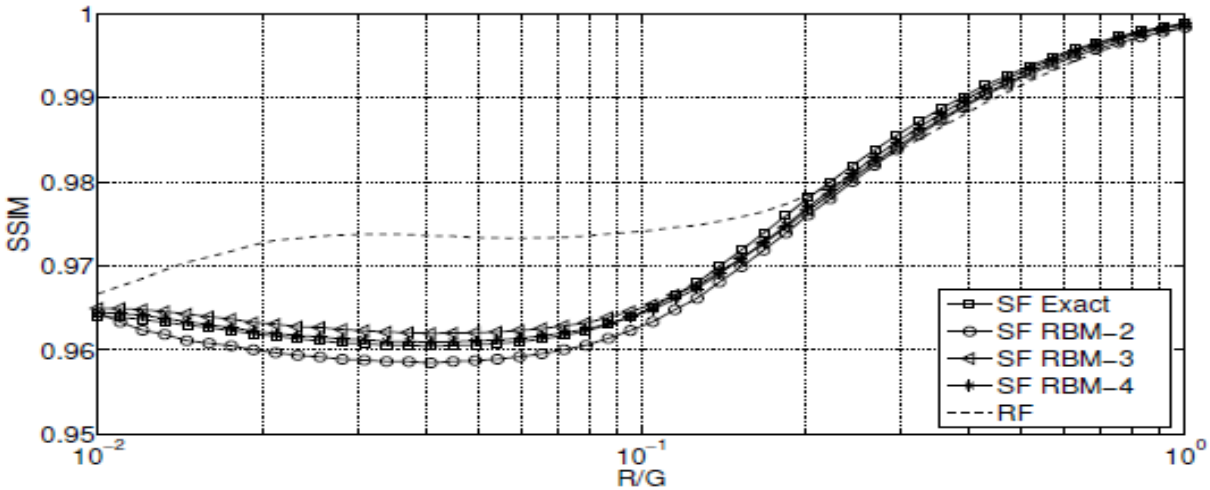
SSIM	MOS	Quality	Impairment
≥ 0.99	5	Excellent	Imperceptible
$[0.95, 0.99)$	4	Good	Perceptible but not annoying
$[0.88, 0.95)$	3	Fair	Slightly annoying
$[0.5, 0.88)$	2	Poor	Annoying
< 0.5	1	Bad	Very annoying

Cognitive video admission control

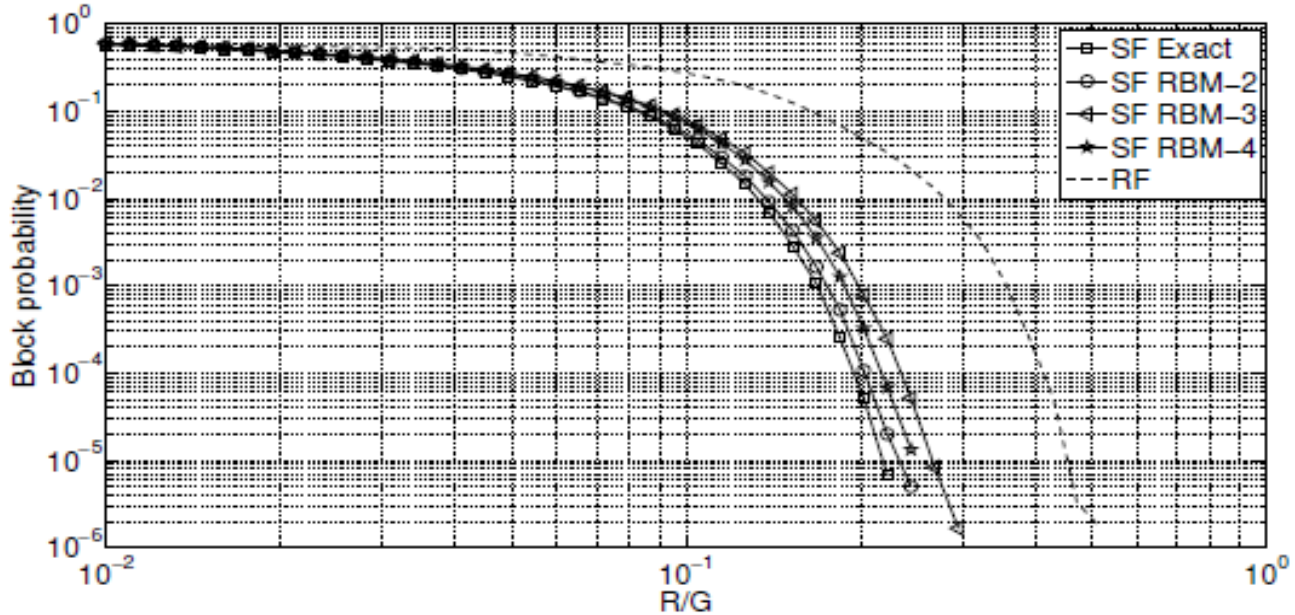


Link rate over aggregate full-quality video rate

Effect of SSIM approximation



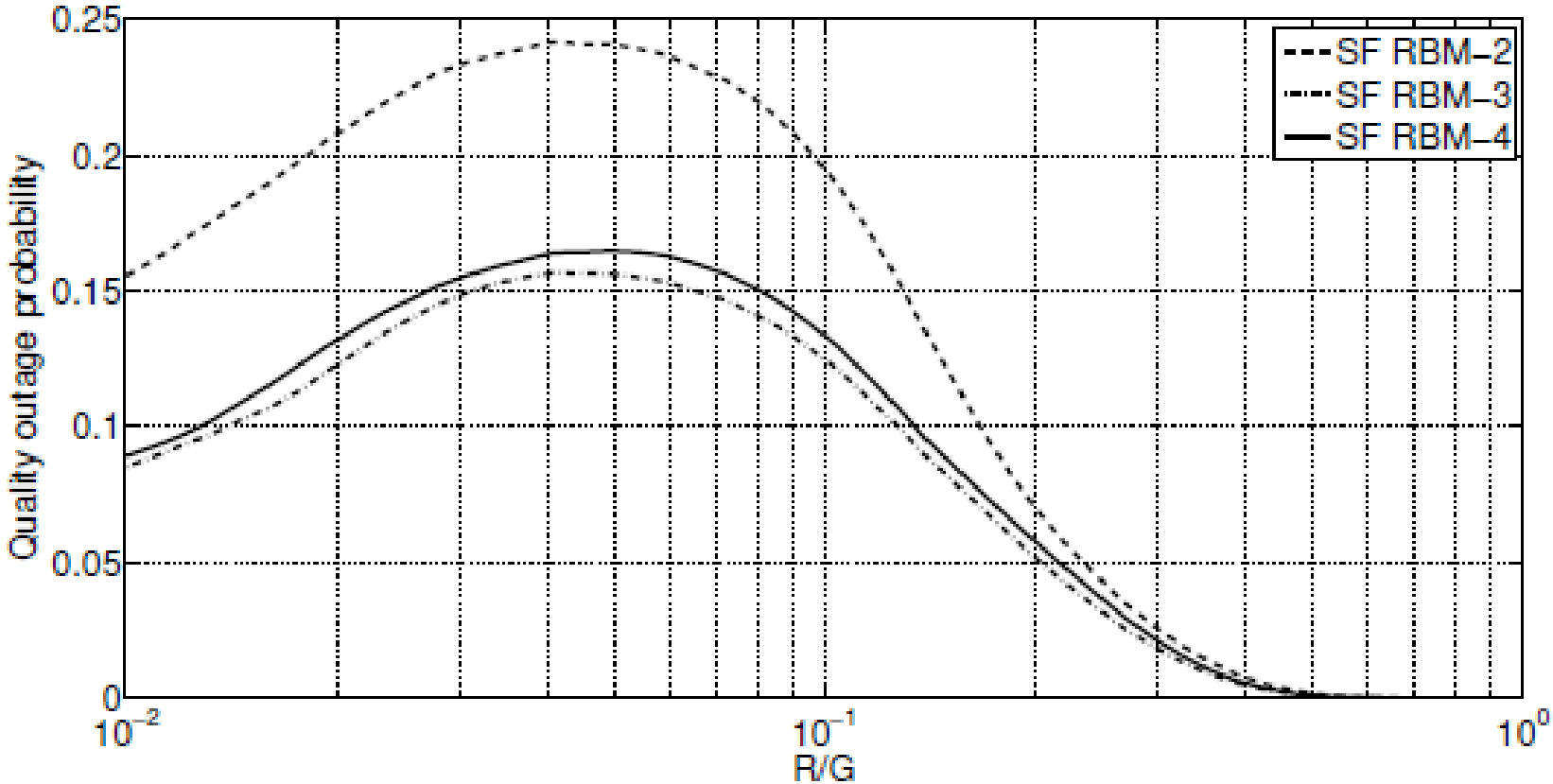
SF RBM-n: SF algorithm, with n-degree polyn. approx. of SSIM curve obtained by RBM approach



Quality outage probability

- Because of SSIM approximation errors, videos may occasionally be accepted even if quality threshold is not met
- Tradeoff on SSIM polynomial approximation: the fewer the coefficients, the coarser the approximation, but the better the RBM coefficient estimate
- Question: is it better to have a well-estimated low degree or a coarsely estimated high-degree polynomial?

Quality outage probability



- n=2 is too small, n=3 or 4 give very similar results: n=3 is the best choice in this case

Conclusions

- Optimizing resource allocation for video transmission is challenging
  - many numerical parameters involved
  - subjective QoE issues
  - high signaling exchange
- Learning-based approaches are useful to
  - obtain a compact representation
  - extrapolate the most significant data
  - Provide a framework with no need for prior models

Publications

- Daniele Munaretto, Andrea Zanella, Daniel Zucchetto, Michele Zorzi, "Data-driven QoE optimization techniques for multi-user wireless networks" in the Proceedings of the 2015 International Conference on Computing, Networking and Communications, February 16-19, 2015, Anaheim, California, USA

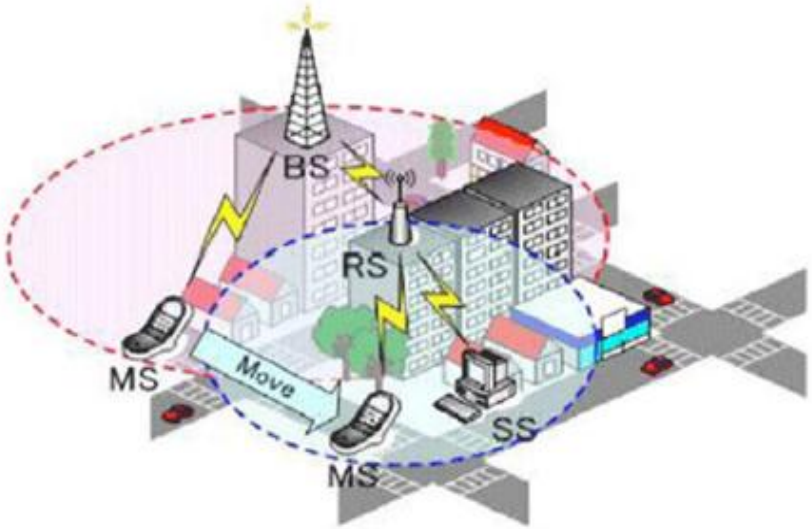
- Alberto Testolin, Marco Zanforlin, Michele De Filippo De Grazia, Daniele Munaretto, Andrea Zanella, Marco Zorzi, Michele Zorzi, "A Machine Learning Approach to QoE-based Video Admission Control and Resource Allocation in Wireless Systems" in the Proceedings of IEEE IFIP Annual Mediterranean Ad Hoc Networking Workshop, Med-Hoc-Net 2014, June 2-4, 2014, Piran, Slovenia.

- Leonardo Badia, Daniele Munaretto, Alberto Testolin, Andrea Zanella, Marco Zorzi, Michele Zorzi , "Cognition-based networks: applying cognitive science to multimedia wireless networking" in the Proceedings of Video Everywhere (VidEv) Workshop of IEEE WoWMoM'14, 16 June, 2014, Sydney, Australia.

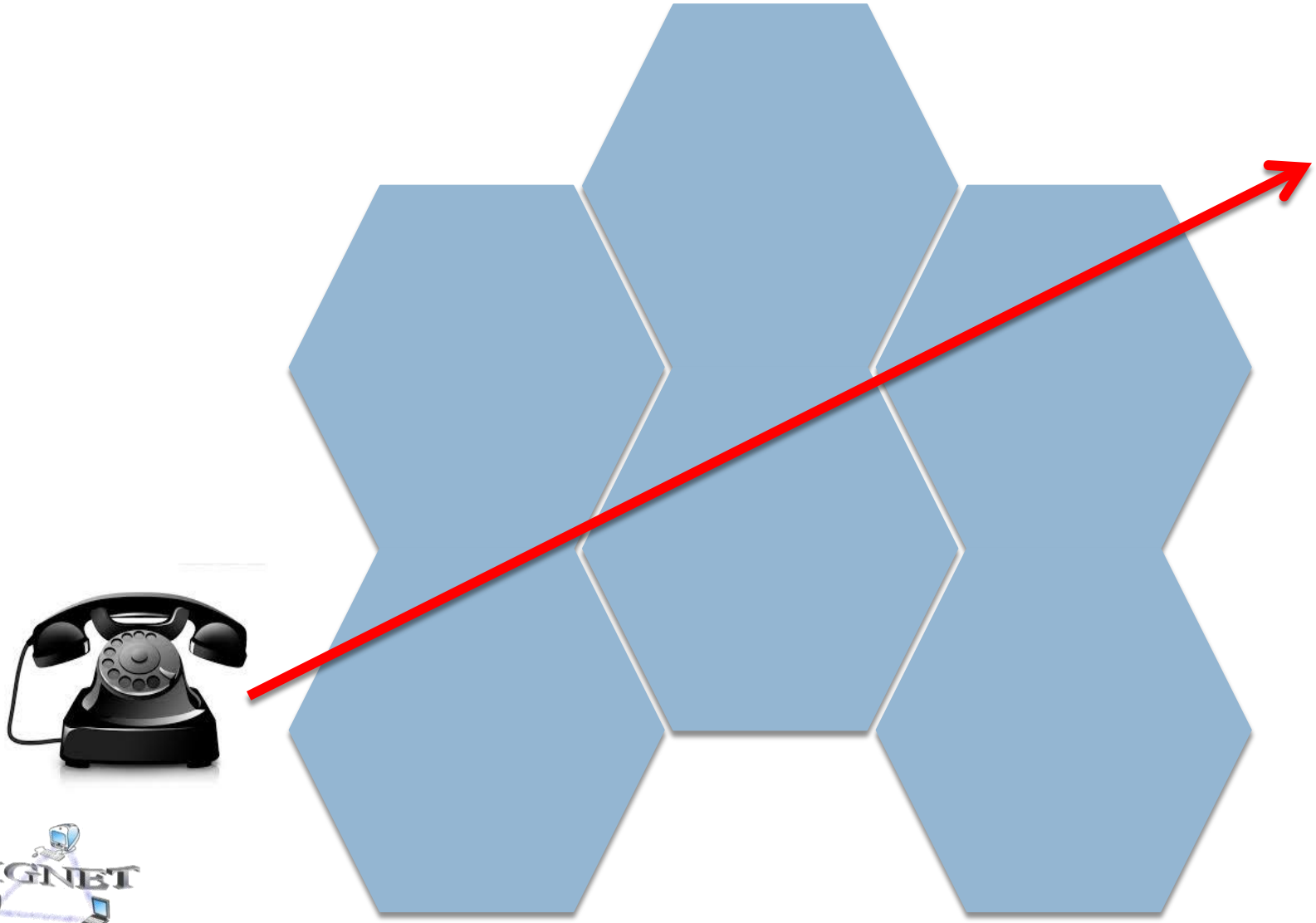
- Marco Zanforlin, Daniele Munaretto, Andrea Zanella, Michele Zorzi, "SSIM-based video admission control and resource allocation algorithms" in the Proceedings of WiOpt workshop WiVid'14, May 12-16, 2014, Hammamet, Tunisia.

Handover process

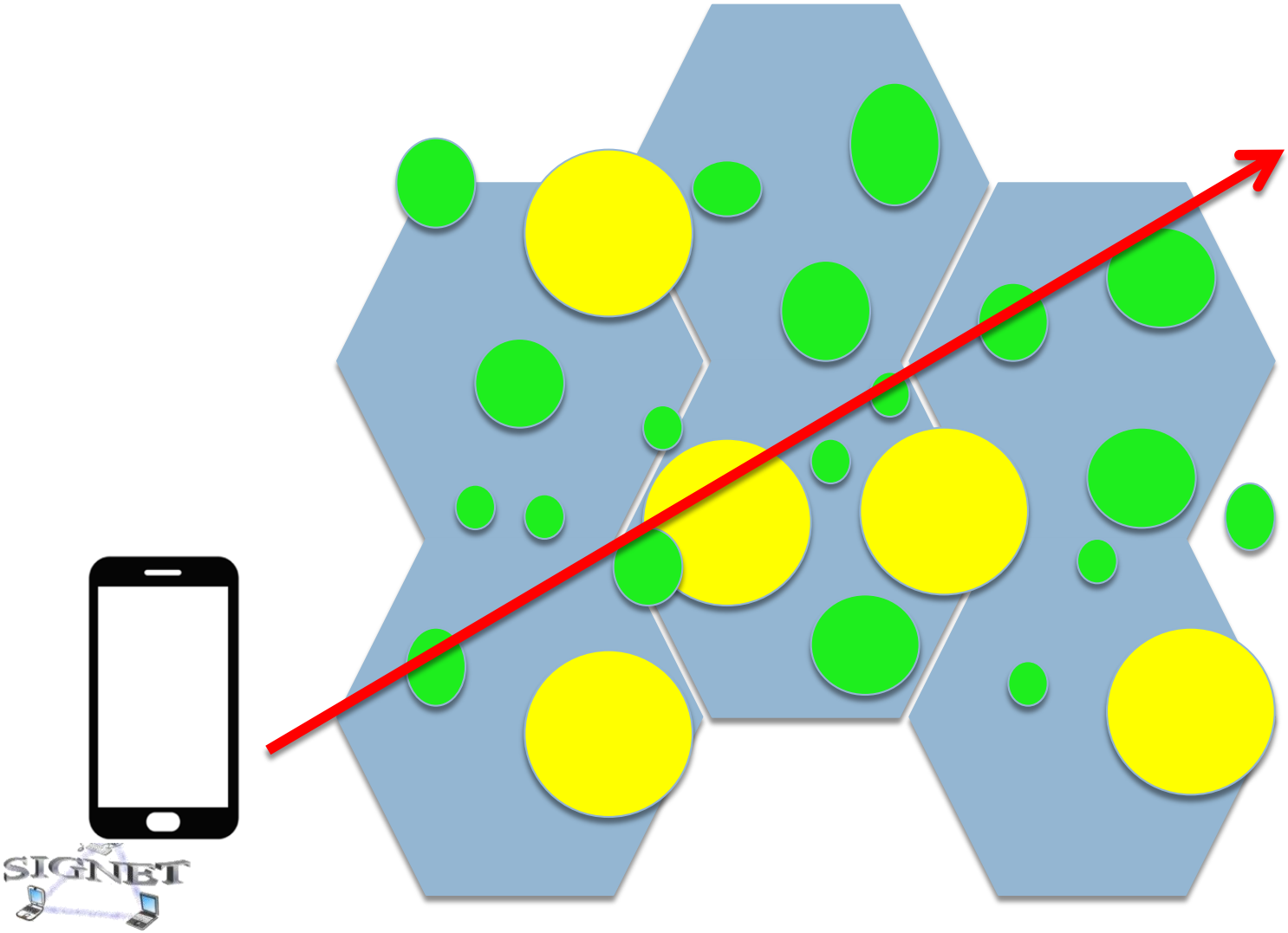
- In cellular telephone systems, the term **handover** refers to the process of keeping a mobile active user connected to the BS that offers the strongest radio signal



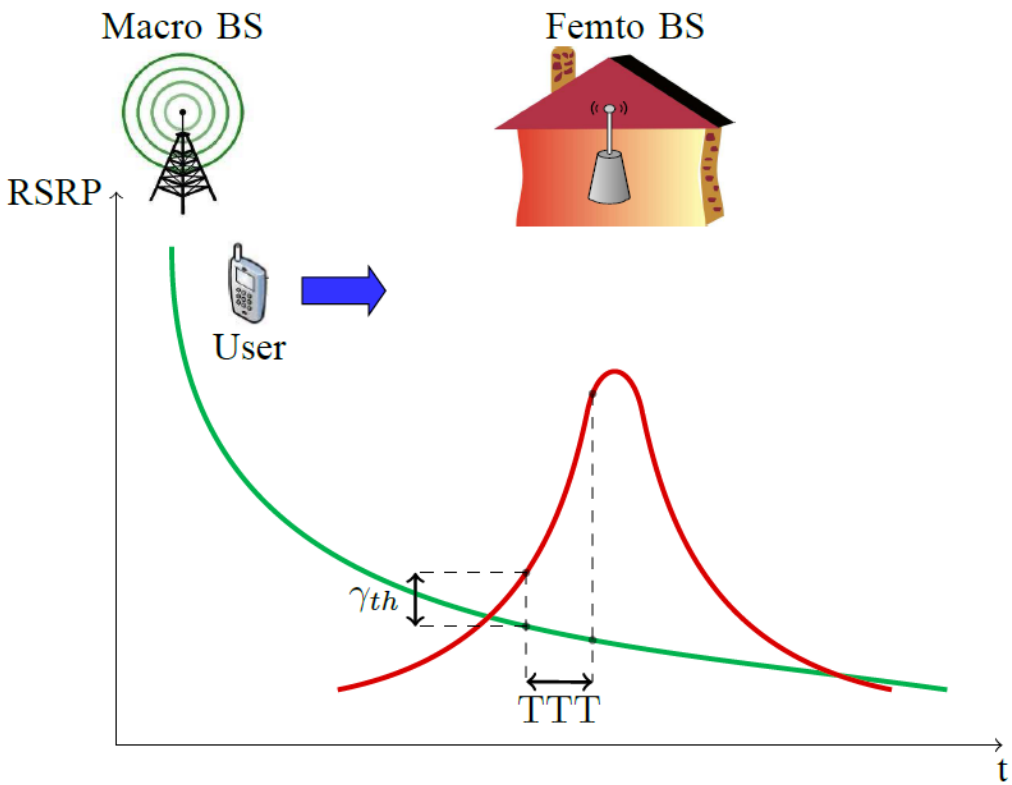
Traditional Cellular network



Heterogeneous Networks



Standard 3GPP Handover Procedure



Long **TTT** → Handover Failure

Short **TTT** → Ping-Pong

Objective

- Develop context-aware cognitive handover strategies to maximize the performance of each single mobile user and of the system as a whole in a HetNets scenario

Steps

① Develop a theoretical model that describes the evolution of the UE state along its trajectory by means of a non homogeneous Markov Chain

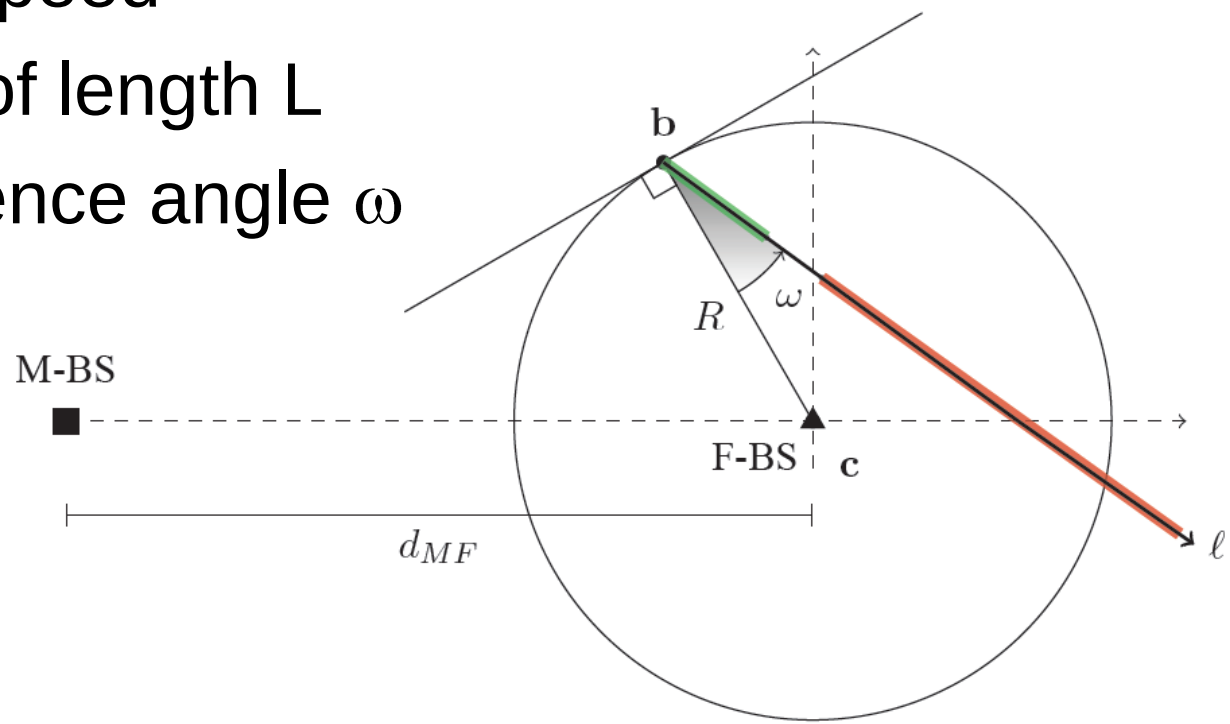
② Express the average UE performance as a function of the context parameters

③ Derive a context-aware HO policy (CAHP) that selects the best HO strategy for a UE approaching a femtocell

④ Develop cognitive mechanisms to infer context parameters [still in progress]

System Model

- 1 Master (M-BS) and 1 Femto Base Stations (F-BS), placed at distance d_{MF}
- Femtocell coverage area as a circle of radius R
- UE constant speed
- Straight path of length L
- Uniform incidence angle ω



Propagation & handover models

- RSRP from the h-BS: $\Gamma_h(\mathbf{a}, t) = \Gamma_h^{tx}\, g_h(\mathbf{a})\, \alpha_h(t) \qquad h \in \{M, F\}$
- where

Γ_h^{tx} : h-BS transmit power

$g_h(\mathbf{a})$: Pathloss gain

$\alpha_h(t)$: Fast-fading channel gain

- Handover model

T : Time-To-Trigger

T_H : Time for Handover signaling and BS switching

γ_{th} : SINR threshold to trigger handover

Average UE capacity

Sampling at time Tc > channel coherence time → independent fading samples → average Shannon capacity experienced by UE along the trajectory is

$$\bar{\mathcal{C}}_L = \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \frac{1}{N_L}\sum_{k=1}^{N_L} \sum_{S\in\{M,F,H\}} \bar{C}_S(\mathbf{a}_k(\omega))\mathrm{P}_S[\mathbf{a}_k(\omega)]\,\mathrm{d}\omega$$

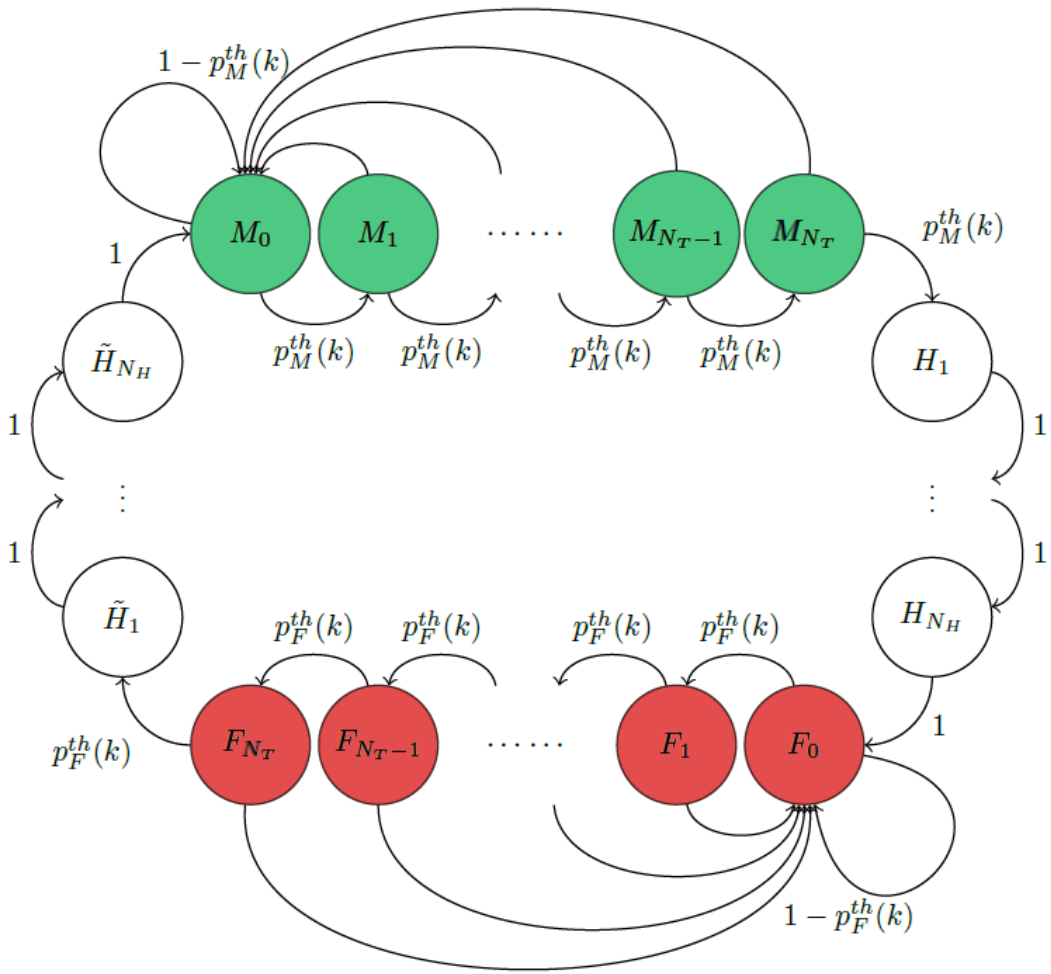
where

$$\bar{C}_S(\mathbf{a}_k(\omega)) = \mathbb{E}\left[\log_2\left(1+\gamma_S(\mathbf{a}_k, kT_c)\right)\right] = \log_2\left(\bar{\gamma}_S(\mathbf{a}_k)\right)\frac{\bar{\gamma}_S(\mathbf{a}_k)}{\bar{\gamma}_S(\mathbf{a}_k)-1} \qquad S\in\{M,F\}$$

$$\bar{C}_H(\mathbf{a}_k(\omega)) = 0$$

Average SINR: $\bar{\gamma}_M(\mathbf{a}_k) = \dfrac{\Gamma_M^{tx}\, g_M(\mathbf{a}_k)}{\Gamma_F^{tx}\, g_F(\mathbf{a}_k)}$ $\qquad \bar{\gamma}_F(\mathbf{a}_k) = \dfrac{1}{\bar{\gamma}_M(\mathbf{a}_k)}$

UE's state: *non homogeneous* Markov Chain



M_j: connected to the M-BS

F_j : connected to the F-BS

H_j : switching from M-BS to F-BS

$\tilde{H}_j$: switching from F-BS to M-BS

zorzi@dei.unipd.it

Solving the MC

Transition probabilities

$$p_M^{th}(k) = \mathrm{P}\left[\gamma_M(\mathbf{a}_k, kT_c) < \gamma_{th}\right] = \frac{\gamma_{th}}{\gamma_{th} + \bar{\gamma}_M(\mathbf{a}_k)}$$

$$p_F^{th}(k) = \mathrm{P}\left[\gamma_F(\mathbf{a}_k, kT_c) < \gamma_{th}\right] = \frac{\gamma_{th}}{\gamma_{th} + \bar{\gamma}_F(\mathbf{a}_k)}$$

Transition matrix

$$\mathbf{P}(k) = \begin{bmatrix} \mathbf{M}(k) & \mathbf{V}_M^H(k) & \varnothing & \varnothing \\ \varnothing & \mathbf{H}(k) & \mathbf{V}_H^F(k) & \varnothing \\ \varnothing & \varnothing & \mathbf{F}(k) & \mathbf{V}_F^{\tilde{H}}(k) \\ \mathbf{V}_{\tilde{H}}^M(k) & \varnothing & \varnothing & \tilde{\mathbf{H}}(k) \end{bmatrix}$$

State probability vector

$$\mathbf{p}(k) = \mathbf{p}(0) \prod_{i=0}^{k-1} \mathbf{P}(i) \qquad \mathbf{p}(0) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}$$

UE state probabilities

$$\mathrm{P}_S[\mathbf{a}_k(\omega)] = \sum_{i \in \{S_j\}} p_i(k) \qquad \textbf{with} \qquad S \in \{M, F, H \cup \tilde{H}\}$$

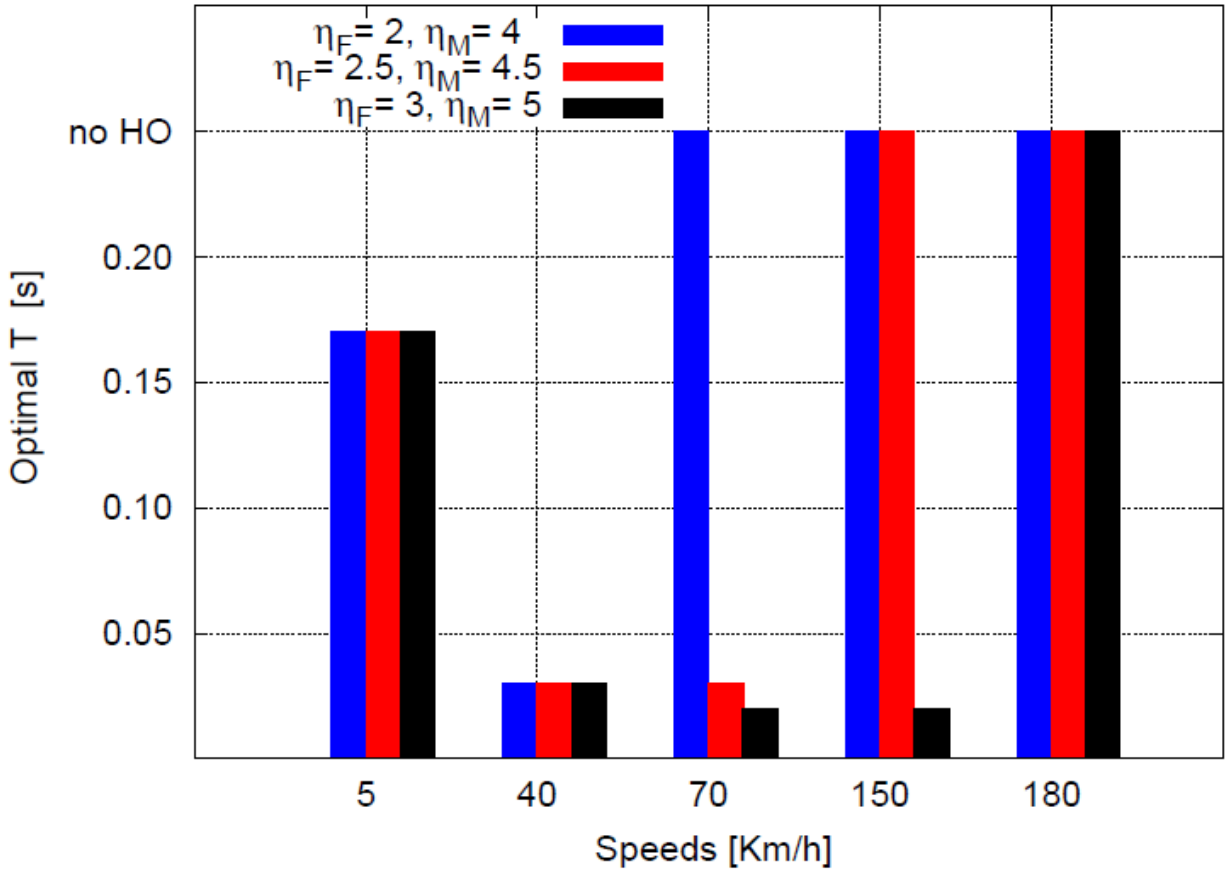
Simulation parameters

Parameter	Value
M-BS/F-BS Transmit Power	46 dBm/24 dBm
Distance between BSs	500 m
Macro/Femto Pathloss exponent	4.5/2.5
Fading	Rayleigh distribution
Handover execution time	200 ms
SINR threshold	0 dB

Algorithms

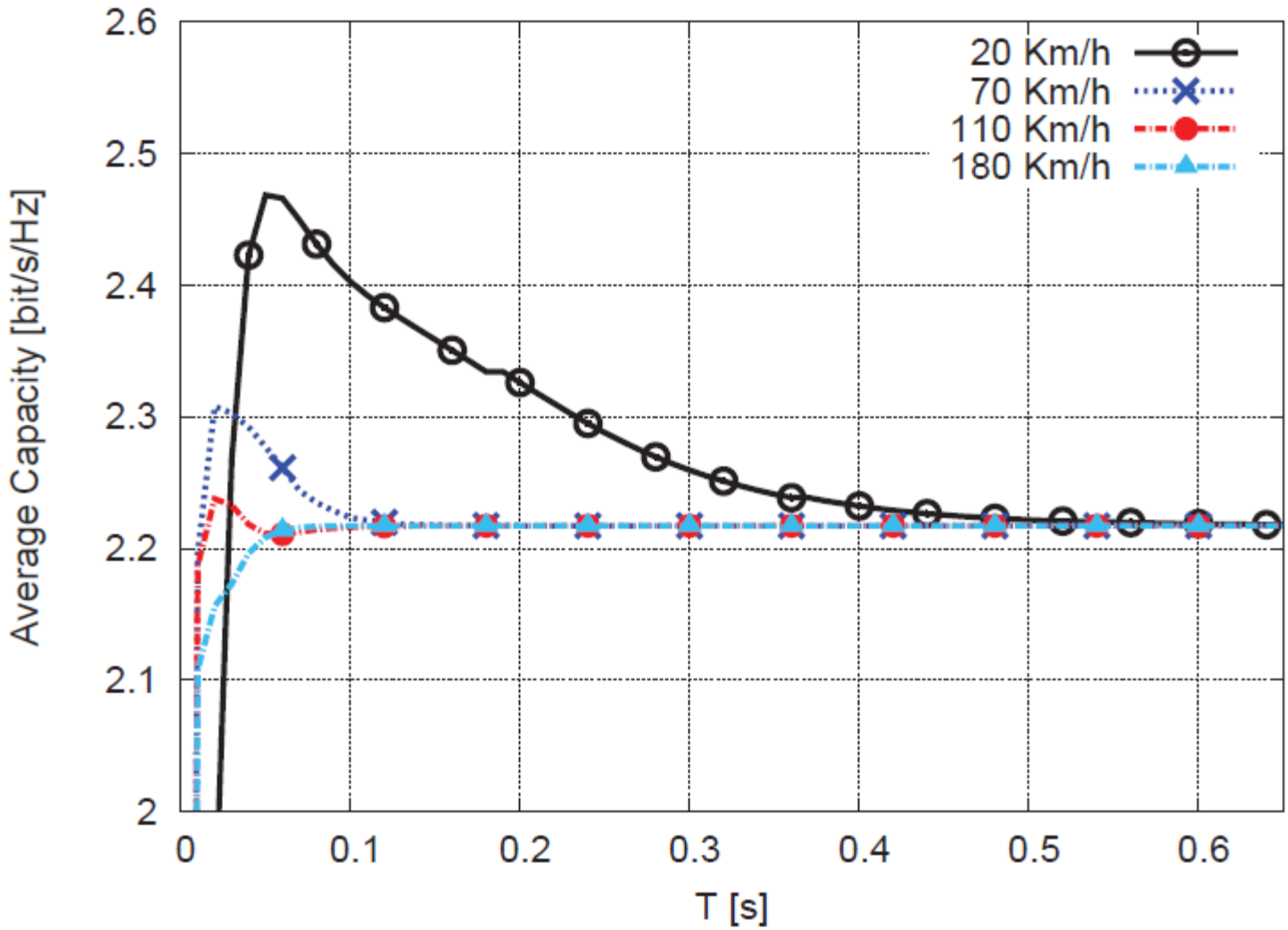
- **CAHP**: Context-Aware Handover Policy:
  - For a certain scenario, and according to the UE speed, it either selects the value of the Time-To-Trigger that gives the maximum average capacity or avoids the HO procedure
- **FIX**: Fixed Time-To-Trigger Policy:
  - A static value T of TTT is used
    - 100 ms, 256 ms, 512 ms

Optimal TTT values

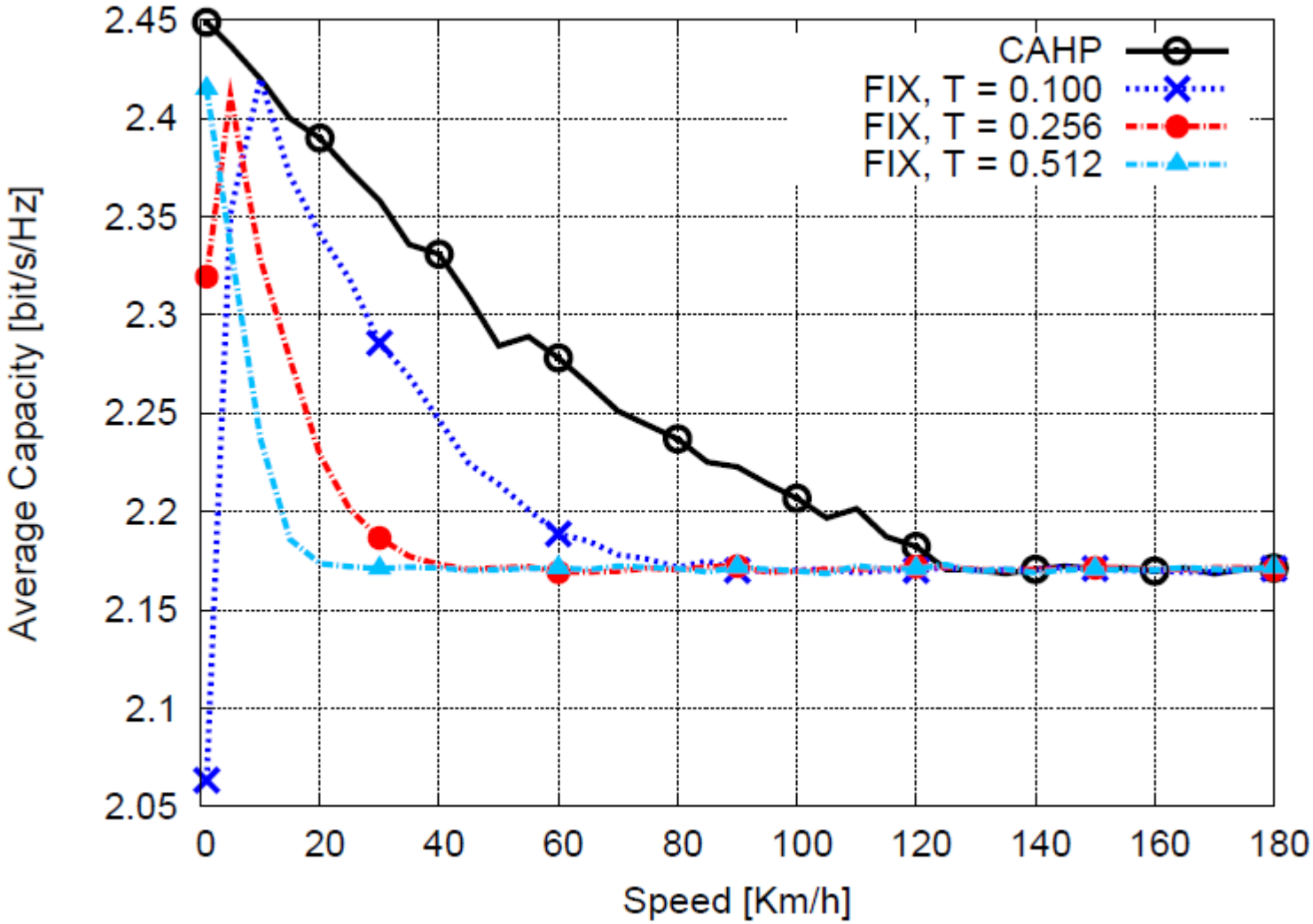


Optimal T values obtained for different speeds and scenarios according to CAHP approach

Analytical average capacity



Average capacity with different HO policies

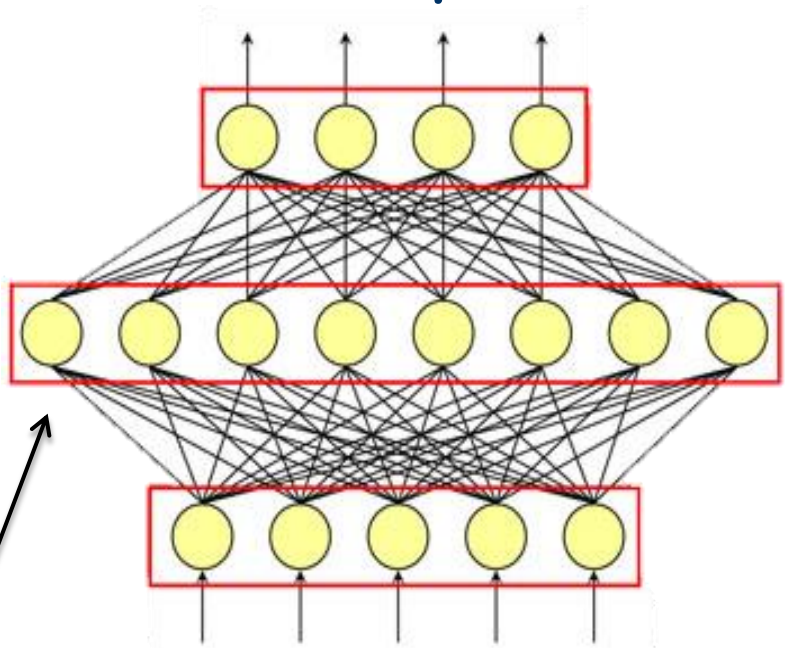


Conclusions

- In heterogeneous cellular scenarios, a context-aware Handover policy is beneficial for UEs with respect to conventional Handover management techniques
- Our optimal context-dependent Handover criterion is based on a solid mathematical analysis and validated by means of simulations
- Future work:
  - SINR threshold parameter optimization
  - More complex scenarios (multiple BSs and UEs, load)
  - **Machine-learning context estimator**

Future work: Neural Networks

- Parameter estimation:
  1. UE speed *v*
  2. Pathloss exponents
  3. Transmit power
  4. Distance M-BS - F-BS
- Prediction:
  - Handover? Yes/No



Publications

- Francesco Guidolin, Irene Pappalardo, Andrea Zanella, Michele Zorzi , "Context-Aware Handover in HetNets" in the Proceedings of the European Conference on Networks and Communications 2014, June 23/26, 2014, Bologna, Italy. **[Best paper Award]**

- Francesco Guidolin, Irene Pappalardo, Andrea Zanella, Michele Zorzi, "A Markov-based Framework for Handover Optimization in HetNet" in the Proceedings of IEEE IFIP Annual Mediterranean Ad Hoc Networking Workshop, Med-Hoc-Net 2014, June 2-4, 2014, Piran, Slovenia

Conclusions

- We have discussed the concept of cognition-based networking:
  - Holistic approach (network-wide, end-to-end, cross-layer)
  - Using the most advanced understanding drawn from cognitive science
  - Machine learning a key ingredient
- Most of the current approaches are too limited and do not address the essence of cognition
- Initial examples show the potential gains
- Although we have been hearing about cognitive radio and networks for years, now is the time to do it

Thanks to our sponsors

- Several agencies have funded our work in this area, including
  - The European Commission (ARAGON project)
  - The US Army Research Office (ARO)
  - The CaRiPaRo Foundation in Padova

A new venue for cognet research

- COMSOC has just started a new journal
- IEEE Transactions on Cognitive Communications and Networks
  - Started as a JSAC series by Y.C. Liang
  - Now greatly expanded in scope
- I am the founding EiC of TCCN
- It will soon start accepting submissions
- Stay tuned…

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Cognition-based networks: applying cognitive science to wireless networking

Michele Zorzi
zorzi@dei.unipd.it

University of Padova (ITALY)
and UC San Diego (USA)