

Network of Excellence

**NEWCOM#**

Network of Excellence in Wireless Communications#

**FP7 Contract Number: 318306**

---



**WP1.1 – PERFORMANCE LIMITS OF WIRELESS  
COMMUNICATIONS**

**D11.2**

**Consolidated Results on the performance limits of wireless  
communications**

<b>Contractual Delivery Date:</b>	October 31, 2014
<b>Actual Delivery Date:</b>	November 21, 2014
<b>Responsible Beneficiary:</b>	CNRS/SUPELEC
<b>Contributing Beneficiaries:</b>	CNRS, CNIT, CTTC, UPC, UOULU, INOV, IASA, Bilkent, ULUND, CNIT, TUD, TECHNION, UCAM, Aalborg
<b>Estimated Person Months:</b>	29
<b>Dissemination Level:</b>	Public
<b>Nature:</b>	Report
<b>Version:</b>	1.0

PROPRIETARY RIGHTS STATEMENT

This document contains information, which is proprietary to the NEWCOM# Consortium.



European Commission  
Information Society and Media



This page is left blank intentionally

## Document Information

<b>Document ID:</b>	<b>D11.1</b>
<b>Version Date:</b>	November 13, 2014
<b>Total Number of Pages:</b>	114
<b>Abstract:</b>	The report presents the Intermediate Results of N# JRAs on Performance Limits of Wireless Communications and highlights the fundamental issues that have been investigated by the WP1.1. The report illustrates the Joint Research Activities (JRAs) already identified during the first year of the project which are currently ongoing. For each activity there is a description, an illustration of the adherence and relevance with the identified fundamental open issues, a short presentation of the preliminary results, and a roadmap for the joint research work in the next year. Appendices for each JRA give technical details on the scientific activity in each JRA.
<b>Keywords:</b>	Sparse Bayesian Learning, Cooperation, Physical layer Security, Multi-User Communications, Large Dimensional Systems, Network coding, routing, Spatially coupled codes, Binary and non binary codes, stochastic geometry, large random matrices.

## Authors

**IMPORTANT:** The information in the following two tables will be directly used for the MPA (Monitoring Partner Activity) procedure. Upon finalisation of the deliverable, please, ensure it is accurate. Use multiple pages if needed. Besides, please, adhere to the following rules:

- **Beneficiary/Organisation:** For multi-party beneficiaries (CNIT) and beneficiaries with Third Parties (CNRS and CTTC), please, indicate beneficiary *and* organisation (e.g., CNIT/Pisa, CNRS/Supelec).
- **Role:** Please, specify: Overall Editor / Section Editor / Contributor.

Full Name	Beneficiary / Organisation	e-mail	Role
Romain Couillet	CNRS/Supelec	romain.couillet@supelec.fr	JRA 1.1.1.3 contributor
Mérouane Debbah	CNRS/Supelec	merouane.debbah@supelec.fr	JRA 1.1.1.3 contributor
Marco di Renzo	CNRS/Supelec	marco.di.renzo@gmail.com	JRA 1.1.1.3 contributor
Abdellatif Zaidi	CNRS	abdellatif.zaidi@u-pem.fr	JRA 1.1.1.3 contributor
Pablo Piantanida	CNRS/Supelec	pablo.piantanida@supelec.fr	JRA 1.1.1.3 contributor
Luca Sanguinetti	CNIT/UniPI	luca.sanguinetti@iet.unipi.it	JRA 1.1.1.3 contributor
Stephan Pfletschinger	CTTC	stephan.pfletschinger@cttc.cat	JRA 1.1.2.1 contributor

Monica Navarro	CTTC	monica.navarro@cttc.cat	JRA 1.1.2.1 contributor
Carmine Vitiello	CNIT/UniPI	carmine.vitiello@for.unipi.it	JRA 1.1.2.1 contributor
Giacomo Bacci	CNIT/UniPI	giacomo.bacci@iet.unipi.it	JRA 1.1.2.1 contributor
Marco Luise	CNIT/UniPI	marco.luise@iet.unipi.it	JRA 1.1.2.1 contributor
Jordi Pérez-Romero	UPC	jorperez@tsc.upc.edu	JRA 1.1.2.2 contributor
Juan Sánchez-González	UPC	juansanchez@tsc.upc.edu	JRA 1.1.2.2 contributor
Ramon Agustí	UPC	ramon@tsc.upc.edu	JRA 1.1.2.2 contributor
Beatriz Lorenzo	UOULU	lorenzo@ee.oulu.fi	JRA 1.1.2.2 contributor
Savo Glisic	UOULU	savo.glisic@ee.oulu.fi	JRA 1.1.2.2 contributor
Sina Khatibi	INOV	sina.khatibi@inov.pt	JRA 1.1.2.2 contributor
Luisa Caeiro	INOV	lcaeiro@est.ips.pt	JRA 1.1.2.2 contributor
Luis M. Correia	INOV	luis.correia@inov.pt	JRA 1.1.2.2 contributor
Aris L. Moustakas	IASA	arism@phys.uoa.gr	JRA 1.1.2.3 and JRA 1.1.1.3 contributor
Erdal Arikan	Bilkent	arikan@ee.bilkent.edu.tr	Task 1.1.3 and JRA 1.1.3.3 Editor
Michael Lentmaier	ULUND	michael.lentmaier@eit.lth.se	JRA 1.1.3.1 Editor
Guido Montorsi	CNIT/Polito	montorsi@polito.it	JRA 1.1.3.2 Editor
Iryna Andriyanova	CNRS	iryna.andriyanova@ensea.fr	Contributor
Pierre Duhamel	CNRS	pierre.duhamel@lss.supelec.fr	Contributor
Najeeb Ul Hassan	TUD	najeeb.ul.hassan@ifn.et.tu-dresden.de	Contributor
Shlomo Shamai	TECHNION	sshlomo@ee.technion.ac.il	Contributor
Jossy Sayir	UCAM	Jossy.sayir@eng.cam.ac.uk	Contributor
Bernard Henry Fleury	AAU	fleury@es.aau.dk	JRA 1.1.1.1 Contributor
Pascal Larzabal	CNRS	Pascal.larzabal@satie.ens-cachan.fr	JRA 1.1.1.1 Contributor
Rémy Boyer	CNRS	remy.boyer@lss.supelec.fr	JRA 1.1.1.1 Contributor
Nabil ElKorso	CNRS	m.elcorso@u-paris10.fr	JRA 1.1.1.1 Contributor
Myriam Bennamar	CNRS	Myriam.bennamar@supelec.fr	JRA 1.1.1.2 Contributor

## Reviewers

Full Name	Beneficiary / Organisation	e-mail	Date
Pierre Duhamel	CNRS	Pierre.duhamel@lss.supelec.fr	29/10/2014
Marco Luise	CNIT/UniPI	Marco.luise@cnit.it	13/11/2014

## Version history

Issue	Date of Issue	Comments
0.1	15/07/2014	Table of content
0.2	20/08/2014	Instructions sent out to take into account the Advisory Board comments
0.3	15/09/2014	First draft
0.4	15/10/2014	Second round of contributions
0.9	27/10/2014	Final version for internal review
1.0	13/11/ 2014	Final version

## Executive Summary

This WP addresses the theoretical aspects related to the performance limits of Wireless Communications. The objectives are in particular to:

- 1- Estimate the ultimate limits of communications and networking
- 2- To optimize the design of relay networks
- 3- To develop capacity achieving channel codes
- 4- To propose optimal distributed optimization and signal processing techniques
- 5- To develop novel network models and analyze their performance

As described in the project's Description of Work, the WP is divided into three Tasks, each one with specific scope and objectives: Task 1.1.1 "Theoretical Limits of Communications and Networks"; Task 1.1.2 "Relaying and Resource Allocation in Wireless Networks "; Task 1.1.3 "Capacity-reaching channel codes ".

The deliverable presents the intermediate results in the thematic areas addressed in WP1.1 and highlights the issues that are currently under investigation.

For each Task, specific JRAs are addressed. In particular a short description clarifies the targets and the adherence and relevance with the identified fundamental open issues; technical details are reported in the annexes.

## Table of Contents

<b>1. Introduction .....</b>	<b>9</b>
<b>1.1 Glossary .....</b>	<b>9</b>
<b>1.2 List of Joint Research Activities (JRAs) .....</b>	<b>10</b>
<b>1.3 Description of the Main WP Achievements in the Reporting Period.....</b>	<b>10</b>
<b>2. Detailed Activity and Achieved Results .....</b>	<b>13</b>
<b>2.1 JRA # 1.1.1.1 Performance Limit of Sparse Bayesian Learning with Application to Wireless Communications Systems .....</b>	<b>13</b>
2.1.1 Description of Activity .....	13
2.1.2 Relevance with the identified fundamental open issues.....	14
2.1.3 Main Results Achieved in the Reporting Period and planned activities .....	14
2.1.4 Publications .....	14
<b>2.2 JRA #1.1.1.2 An Information-Theoretic Perspective of Cooperation and Secrecy in Multi-User Communications .....</b>	<b>15</b>
2.2.1 Description of Activity .....	15
2.2.2 Relevance with the identified fundamental open issues.....	15
2.2.3 Main Results Achieved in the Reporting Period and planned activities .....	15
2.2.4 Publications .....	17
<b>2.3 JRA # 1.1.1.3: Communications Performance of Large Dimensional Systems</b>	<b>19</b>
2.3.1 Description of Activity .....	19
2.3.2 Relevance with the identified fundamental open issues.....	19
2.3.3 Main Results Achieved in the Reporting Period and planned activities .....	19
2.3.4 Publications .....	20
<b>2.4 JRA # 1.1.2.1: Network Coding Schemes for Relay Channels .....</b>	<b>21</b>
2.4.1 Description of Activity .....	21
2.4.2 Relevance with the identified fundamental open issues.....	21
2.4.3 Main Results Achieved in the Reporting Period and planned activities .....	21
2.4.4 Publications .....	22
<b>2.5 JRA 1.1.2.2: Optimization approaches for heterogeneous networks.....</b>	<b>23</b>
2.5.1 Description of Activity .....	23
2.5.2 Relevance with the identified fundamental open issues.....	24
2.5.3 Main Results Achieved in the Reporting Period and planned activities .....	24
2.5.4 Publications .....	25
<b>2.6 JRA 1.1.2.3 Traffic Dynamics: Routing and Topology Reconfiguration .....</b>	<b>26</b>
2.6.1 Description of Activity .....	26
2.6.2 Relevance with the identified fundamental open issues.....	26
2.6.3 Main Results Achieved in the Reporting Period and planned activities .....	27
2.6.4 Publications .....	27
<b>2.7 JRA # 1.1.3.1: Spatially Coupled Codes .....</b>	<b>28</b>
2.7.1 Description of Activity .....	28
2.7.2 Relevance with the identified fundamental open issues.....	28
2.7.3 Main Results Achieved in the Reporting Period and planned activities .....	28
2.7.4 Publications .....	29
<b>2.8 JRA # 1.1.3.2: Non Binary Codes .....</b>	<b>30</b>
2.8.1 Description of Activity .....	30
2.8.2 Relevance with the identified fundamental open issues.....	30
2.8.3 Main Results Achieved in the Reporting Period and Planned Activities.....	30
2.8.4 Publications/Presentations .....	32
<b>2.9 JRA 1.1.3.3 Coding for Multiterminal Communication Systems.....</b>	<b>34</b>
2.9.1 Description of Activity .....	34
2.9.2 Relevance with the identified fundamental open issues.....	34
2.9.3 Main Results Achieved in the Reporting Period and planned activities .....	34

---

2.9.4 Publications .....	38
<b>3. General Conclusions and Prospects.....</b>	<b>39</b>
<b>4. Annex I: Detailed Description of Main technical WP Achievements.....</b>	<b>40</b>

## 1. Introduction

### 1.1 Glossary

CRB	Cramér-Rao bounds
CSIT	channel state information
SNR	signal-to-noise
BC	Broadcast Channel
NID	Non Interference Decoding
ID	Interference Decoding
MD	Multiple Description
CD	Common Description
CIFC	Cognitive Interference Channel
MISO	Multiple Input Single Output
WBC	Wiretap Broadcast Channel
BS	binary symmetric
BEC	binary erasure code
UE	user equipment
BS	base station
SIC	successive interference cancellation
AP	Access Point
DNA	Dynamic Network Architecture
RAN	Radio Access Network
VNO	Virtual Network Operator
RAT	Radio Access Technology
VRRM	Virtual Radio Resource Management
RRU	Radio Resource Unit
SINR	Signal-to-interference-plus-noise ratio
LDPC	low-density parity-check
EMS	Extended Min Sum
ADBP	Analog-Digital Belief Propagation
FFT	Fast-Fourier-transform
BP	Belief Propagation
FEC	forward error correction
AF	Amplify and Forward
MARC	Multi Access Relay Channel
SDoF	secure degrees of freedom
MIMO	multi-input multi-output

## 1.2 List of Joint Research Activities (JRAs)

---

This work-package contains three tasks, with the corresponding JRAs as described below:

Task 1.1.1 : Theoretical Limits of Communications and Networks; Task Leader : Merouane Debbah (CNRS/SUPELEC)

JRA 1.1.1.1 : Performance limits of Sparse Bayesian Learning with application to wireless communication systems

JRA 1.1.1.2: An Information-Theoretic Perspective of Cooperation and Secrecy in Multi-User Communications

JRA 1.1.1.3: Communications Performance of Large Dimensional Systems

Task 1.1.2 : Relaying and Resource Allocation in Wireless Networks ; Task leaders : Beatriz Lorenzo and Savo Glisic (UOLU)

JRA 1.1.2.1: Network Coding schemes for relay channels

JRA 1.1.2.2: Optimization approaches for heterogeneous networks

JRA 1.1.2.3: Traffic dynamics - routing and topology reconfiguration

Task 1.1.3 : Capacity-reaching channel codes ; Task leader : Erdal Arikan

JRA 1.1.3.1: Spatially Coupled Codes

JRA 1.1.3.2: Non Binary Codes

JRA 1.1.3.3: Coding for Multiterminal Communication Systems

## 1.3 Description of the Main WP Achievements in the Reporting Period

---

Main achievements for task 1.1.1:

Task 1.1.1. targets the performance limits of networks using information theoretic, random matrix based and sparse bayesian techniques. The tools tackle some important problems in the design of wireless networks.

In particular, the tools were adapted to the specific requirements of networks considering asymptotic performance. The asymptotic limits provide useful bounds to understand the behaviour of the system in the finite regime. In some case, the gap between the asymptotic limit and the finite regime has been analyzed carefully to show the applicability of the results. The Task 1.1.1 has achieved in particular the following results:

1. In the large system limit, it was shown that the expected CRB, rather than the Bayesian CRB, seems to be the appropriate bound to benchmark sparse estimators.
2. The optimal power control strategies in the multicell was derived for large dimensional communication channels
3. The optimal probability of error in point to point communication in the MIMO setting, assuming large number of antennas but finite communication blocklength was derived
4. The SINR for large multi-cellular environments was derived using stochastic geometry tools
5. Secrecy capacity regions of Some Classes of Wiretap Broadcast Channels were derived
6. Dirty-paper Coding Techniques for Compound MISO Broadcast Channels were developed

Main Achievements for Task 1.1.2:

Task 1.1.2: Optimal Design of Wireless Networks investigates several techniques such as network coding (joint optimization of the network and physical layer), or adaptive relaying

strategies with power consumption and complexity constraints. The development of efficient resource allocation strategies in the heterogeneous network environment is also one of the objectives in this task. The dynamical nature of the network is also addressed as well as the speed of convergence to the desired operating points of the different algorithms. The main achievements in task 1.1.2 are summarized below:

1. A new distributed strategy based on Q-learning and softmax decision making has been proposed as a means to implement the connections between the terminals and the APs in a heterogeneous network with D2D capabilities. The proposed distributed strategy achieves a performance very close to the optimum one, and allows significant power consumption reduction when compared to the classical approach in which the terminals connect directly to the cellular base station.
2. A new Dynamic Network Architecture optimization is proposed that is designed to satisfy the QoS requirements from the users and to minimize the network cost. Economic model is proposed to award the users for acting as APs and to share their remaining prepaid traffic volume.
3. A comprehensive analytical model for VRRM is proposed which has two main parts: realistic estimation of the available virtual radio resources and the allocation of the estimated resources considering priority of different services. The proposed model supports the traffic offloading and the resource shortage situation.
4. Joint and separate decoding strategies have been defined for trellis codes, turbo codes and non-binary LDPC codes as the most general case and their performance have been evaluated for the AWGN and for fading channels. Joint decoding was shown to provide significant performance benefits for the fast fading channel and for higher-order modulations in the block fading channel while for the AWGN channel the gains are moderate.
5. The benefits of joint quaternary decoding applied to off-the-shelf LTE turbo channel codes in the TWRC have been shown
6. Non-binary channel coding has been shown to combine well with the principles of network coding.
7. Joint decoding has been shown to exhibit better error performance compared to separate decoding and SIC, approaching single user performance for symmetric channel conditions.
8. A new decoding scheme has been proposed that has been shown to achieve substantial gains compared to separate decoding and to SIC.

#### Main Achievements for Task 1.1.3:

As higher spectral efficiency and throughput targets are set for future communication systems, existing capacity-approaching channel coding schemes need to be developed further to work under more demanding scenarios. In general, Task 1.1.3. Capacity-Reaching Channel Codes aims to develop capacity-approaching channel codes for diverse set of future application scenarios.

In Task 1.1.3, the main achievements in the reporting period have been as follows:

1. It has been shown that spatially coupled LDPC codes are remarkably effective on block-fading channels
2. A method has been found to design spatially coupled LDPC codes with flexible rates, thus overcoming a practical hurdle associated with ordinary LDPC codes
3. A VLSI hardware implementation of a novel Belief Propagation (BP) algorithm, called *Analog Digital Belief Propagation* (ADBP) has been completed. The VLSI implementation

confirms that ADBP allows the construction of efficient decoders for digital transmission systems with *unbounded* spectral efficiency.

4. The capacity region of cooperative multiple-access channel has been determined. A two-user multiple access channel is used where channel state is known noncausally to one of the encoders and only strictly causally to the other encoder. Both encoders transmit a common message and the the encoder that knows the state noncausally transmits an extra individual message. The channel noise is gaussian. This is a rare example of a cooperative communication situation where the channel capacity is fully determined.
5. The relationship between polar coding and the varentropy parameter has been investigated. Varentropy stands for “variance of entropy” and is an important parameter for determining the finite-length performance of source and channel codes. It has been shown that varentropy decreases strictly monotonically under each step of the polar transform. Polar coding and varentropy are two recent research subjects that bring information theory to bear more heavily on construction of practical codes; this result establishes a link between these two subjects.
6. A rigorous mathematical analysis of two communication strategies, namely, soft decode-and-forward (soft-DF) for relay channels and soft partial interference-cancelation (soft-IC) for interference channels, has been given. One of the major findings is the proof that optimal point-to-point codes are unsuitable for soft-IC, as well as for all strategies that apply partial decoding to improve upon single-user detection (SUD) and multiuser detection (MUD), including Han-Kobayashi (HK).
7. Sparse sampling of coded signals at sub-Landau sampling rates has been investigated and it has been shown that with coded signals the Landau condition may be relaxed and the sampling rate required for signal reconstruction and for support detection can be lower than the effective bandwidth. This work establishes a link between sparse (compressive) sensing literature and communication systems.
8. Secure transmission over a two-user multi-input multi-output (MIMO) X-channel with noiseless local feedback and delayed channel state information (CSI) available at transmitters has been investigated and the optimal sum secure degrees of freedom (SDoF) region has been determined. It has been shown that, in presence of local feedback and delayed CSI, the sum SDoF region of the MIMO X-channel is same as the SDoF region of a two-user MIMO BC with  $2M$  antennas at the transmitter and  $N$  antennas at each receiver.

## 2. Detailed Activity and Achieved Results

### 2.1 JRA # 1.1.1.1 Performance Limit of Sparse Bayesian Learning with Application to Wireless Communications Systems

Leader : Rémy Boyer (CNRS/UPS)

Researchers involved : Bernard Henry Fleury (Aalborg Univ.), Pascal Larzabal (CNRS/ENS-Cachan), Nabil ElKorso (CNRS/Univ. Paris X), Romain Couillet (CNRS/Supelec), Yonina Eldar (Technion).

#### 2.1.1 Description of Activity

In the first part of this JRA we have derived fundamental lower bounds (specifically, Cramér-Rao bounds (CRBs)) on the MSE of sparse estimators for linear models in. A particular interest was devoted to sparse Bayesian learning in which a hierarchical probabilistic model is defined that produces sparsity-inducing priors.

We focused on two types of CRBs, the Bayesian CRB and an expectation of the deterministic CRB, the expectation being over the a-priori distributions of the random matrix the parameter vector of the linear model. It can be easily shown that the latter is above the former.

We have now extended these investigations to systems of large dimensions. This assumption allows us to exploit some recent results from the Random Matrix Theory. The derived lower bound takes a very simple expression parametrized by the noise variance, the sparsity-ratio and the aspect ratio of the dictionary. A first observation is that for large dimensions, the lower bound is only a function on the dimensions of the system and not of the active sources. This expression given for a known support size (number of active sources) has been generalized in the case where the support size follows a Binomial distribution. In this case, a second compact expression has been obtained.

The variance of the oracle-LMMSE estimator has been also derived in the context of the Random Matrix Theory. The first conclusions are that the oracle LMMSE reaches the Expected CRB but not the BCRB of VanTrees in the high SNR regime. At low SNR, the problem is different because the ECRB is no longer a lower bound, so we currently investigate the interest of the BCRB (which remains for all SNR a lower bound) regarding the oracle-LMMSE variance in the low SNR regime. The recent tracks are conducted with the collaboration of Romain Couillet.

Another topic is the performance (in the sense of the MSE) for asymptotic and non-asymptotic conditions of sparse estimators assuming hierarchical priors on the sources and the noise. This subject is investigated with the collaboration of Nabil ElKorso. Some close-form expressions of the BCRB have been derived but not totally interpreted.

In the context of the PhD thesis of Stephanie Bernhardt, we have developed two aspects:

1. Theoretical analysis of the off-grid problem for sparse estimators  
The off-grid problem appears in the high SNR regime where the continuous parameters does not belong to the discrete parameter set used in sparse estimators (LASSO, OMP, ...). In this context, we have derived some theoretical expressions of the Bayesian MSE for the azimuth of a source in the context of array processing. Currently, we study the off-grid problem in the context of large systems for the estimation of the amplitudes of the active sources.
2. Limit in sampling of Finite Rate of Innovation Signals  
This topic is derived with the collaboration of Yonina Eldar. We have derived some expressions for the estimation of the time-delay and amplitude for non-bandlimited

signals as the FRI signals. Currently, we look for an optimal sampling kernel in the sense of lower bounds on the MSE or in the sense of the Statistical Resolution Limit.

### **2.1.2 Relevance with the identified fundamental open issues**

The derived bounds can be used to benchmark the performance of sparse estimators. This JRA addresses the following fundamental open issues:

- Bayesian lower bound for a large random dictionary with known sparsity-parameter
- Bayesian lower bound for a large random dictionary with random sparsity-parameter
- Theoretical variance of the oracle LMMSE estimator for sparse estimation
- Bayesian lower bound for sparse sparse estimators assuming hierarchical priors on the sources and the noise
- Analysis of the off-grid problem: theoretical BMSE expression
- Performance analysis for estimation of FRI Signal

### **2.1.3 Main Results Achieved in the Reporting Period and planned activities**

We have shown that in the large system limit, the expected CRB, rather than the Bayesian CRB, appears to be the appropriate bound to benchmark sparse estimators. Further investigations will aim at better understanding this finding. Especially, the role and importance of the Bayesian CRB in the context of large systems is not clear yet. We also target sparse estimators that achieve these bounds.

### **2.1.4 Publications**

- [1] S. Bernhardt, R. Boyer, B. Zhang, S. Marcos, P. Larzabal, « Performance Analysis for Sparse Biased Estimator : Application to Line Spectra Analysis », Invited article during the IEEE SAM'14 conference.
- [2] S. Bernhardt, R. Boyer, S. Marcos, Y. Eldar, P. Larzabal, « Cramer-Rao Bound for Finite Streams of Multiple Filtered Pulses », Proceedings of the EUSIPCO'14 conference.

## **2.2 JRA #1.1.1.2 An Information-Theoretic Perspective of Cooperation and Secrecy in Multi-User Communications**

Leader : Pablo Piantanida (CNRS-Supelec)

Researchers involved : Meryem Benammar, Pablo Piantanida (CNRS-Supelec), Shlomo Shamai (Technion)

### **2.2.1 Description of Activity**

This JRA investigates several information-theoretic secrecy scenarios that prove the existence of coding guaranteeing a level of secrecy without pre-shared secret keys. The research aims at demonstrating through simple models of sources and channels that secrecy can be provided at no total rate cost at all. Another challenging goal is to study the impact of partial channel state information that is available to the transmitter (CSIT). Indeed, perfect secrecy cannot be established without CSIT unless the legitimate and the eavesdropper terminals observe asymmetrical channel statistics (e.g. asynchronous fading variation, better signal-to-noise (SNR), number of antennas), or at least partial CSIT. Moreover, secrecy capacity of MIMO Gaussian wiretap channels with (noisy) feedback, delayed CSIT, as well compound settings, remains open.

### **2.2.2 Relevance with the identified fundamental open issues**

This JRA addresses the following fundamental open issues:

- The optimal distortion-equivocation tradeoff of secure transmission of analog sources over noisy channels.
- The capacity of the Broadcast Wiretap Channel, including the time-varying channel models with state-feedback (delayed channel state information) at the transmitter.
- The capacity of the Compound Broadcast Channel for which Multiple Description Coding and Interference Decoding techniques are investigated.
- The capacity of the Multicast Cognitive Interference Channel.

Except for special cases, the optimal single-letter characterizations and the fundamental principles governing multiterminal communications under channel uncertainty, state-feedback, user's interference and secrecy constraints are not well understood, but several novel results which have been recently emerged may lead to promising techniques to deal with these problems.

### **2.2.3 Main Results Achieved in the Reporting Period and planned activities**

- Secure Transmission of Sources Over Noisy Channels With Side Information at the Receivers

Partners: S. Shamai (TECHNION) and P. Piantanida (CNRS/SUPELEC)

This work investigates the problem of source-channel coding for secure transmission with arbitrarily correlated side informations at both receivers. This scenario consists of an encoder (referred to as Alice) that wishes to compress a source and send it through a noisy channel to a legitimate receiver (referred to as Bob). In this context, Alice must simultaneously satisfy the

desired requirements on the distortion level at Bob and the equivocation rate at the eavesdropper (referred to as Eve). This setting can be seen as a generalization of the problems of secure source coding with (uncoded) side information at the decoders and the wiretap channel. A general outer bound on the rate-distortion-equivocation region, as well as an inner bound based on a pure digital scheme, is derived for arbitrary channels and side informations. In some special cases of interest, it is proved that this digital scheme is optimal and that separation holds. However, it is also shown through a simple counterexample with a binary source that a pure analog scheme can outperform the digital one while being optimal. According to these observations and assuming matched bandwidth, a novel hybrid digital/analog scheme that aims to gather the advantages of both digital and analog ones is then presented. In the quadratic Gaussian setup when side information is only present at the eavesdropper, this strategy is proved to be optimal. Furthermore, it outperforms both digital and analog schemes and cannot be achieved via time-sharing. Through an appropriate coding, the presence of any statistical difference among the side information, the channel noises, and the distortion at Bob can be fully exploited in terms of secrecy.

Planned activities:

This work is planned to be continued in the context of secure broadcasting where a common source has to be broadcasted (loosely) to two legitimate receivers in presence of side information at the decoders, while keeping them ignorant of irrelevant information.

- On the Compound Broadcast Channel: Multiple Description Coding and Interference Decoding

Partners: S. Shamai (TECHNION) and M. Benammar and P. Piantanida (CNRS/SUPELEC)

This work investigates the general two-user compound Broadcast Channel (BC) where a source wishes to transmit two private messages and a common message to two receivers of a BC, while being oblivious to the two channels realizations controlling the communication. The focus is on the characterization of the largest achievable rate region resorting to more evolved encoding and decoding schemes than standard coding techniques for the BC, i.e. Marton's coding. First, an achievable rate region is derived based on the principle of Interference Decoding (ID) where each receiver chooses to decode its intended message as well as (or not) non-uniquely decode the interfering message. The proposed inner bound is shown to be capacity achieving for a class of non-trivial compound BEC/BSC BCs while the worst-case of Marton's inner bound, which is based on Non Interference Decoding (NID), fails at achieving the capacity region. On the encoder's side, we study a more evolved encoding scheme based on Multiple Description (MD) coding where the source transmits a common description as well as multiple dedicated private descriptions to the many instances of the channels of the same user. It turns out that MD coding outperforms the single description scheme, i.e. Common Description (CD) coding, for a class of compound Multiple Input Single Output (MISO) BC.

Planned activities:

This work is planned to be continued in the context of the Multicast Cognitive Interference Channel (CIFC), where many primary users are interested in the same message and where there is only one secondary transmitter. A possible deployment scenario consists of several users (spectators) in a football stadium being able to have access to instantaneous replay of the most important scenes on their cell phone while a nearby base station helps with the communication. Our aim is to characterize the capacity region in the light of the results on the CIFC, and to derive the optimal communication strategies.

- Multi-User MISO wiretap channels with delayed CSIT

Partners: S. Shamai (TECHNION) and M. Benammar and P. Piantanida (CNRS/SUPELEC)

The multiple-input single-output (MISO) wiretap channel with  $K$  legitimate single-antenna receivers, one single-antenna eavesdropper in presence of delayed channel state information at the transmitter (CSIT) is considered. The transmitter is equipped with  $(K+1)$  antennas and has independent messages intended for each one of the  $K$  legitimate receivers. While for the case of a single receiver wiretap channel, i.e.,  $K = 1$ , the optimal secure degrees of freedom (SDoF) with delayed CSIT was recently characterized in [1] and shown to be  $2/3$ , the extension to the multi-receiver case ( $K > 1$ ) is far from straightforward. We present new results and insights for the simplest non-trivial extension of this problem, i.e., for the case of  $K = 2$  receiver MISO wiretap channel with delayed CSIT. The contribution of this paper is two fold: a) an asymptotic achievable scheme is presented for the case of two legitimate receivers which achieves a sum-SDoF of  $36/37$  and b) a novel converse proof is presented which shows that the sum-SDoF is upper bounded by  $16/15$ .

#### Planned activities:

We will pursue these theoretical directions as to enhance further insights on related problems for moderate SNR regimes. Evidently, theoretically addressing these problem is of fundamental impact, as if indeed in different wireless systems, security for part of the data can be provided at minimal or even no cost in the total throughput, then implementing physical-layer security in practical systems be most appealing.

- Secrecy Capacity Region of Some Classes of Wiretap Broadcast Channels  
Partners: M. Benammar and P. Piantanida (CNRS/SUPELEC)

This work investigates the secrecy capacity of the Wiretap Broadcast Channel (WBC) with an external eavesdropper where a source wishes to communicate two private messages over a Broadcast Channel (BC) while keeping them secret from the eavesdropper. We derive a non-trivial outer bound on the secrecy capacity region of this channel which, in absence of security constraints, reduces to the best known outer bound to the capacity of the standard BC. An inner bound is also derived which follows the behavior of both the best known inner bound for the BC and the Wiretap Channel. These bounds are shown to be tight for the deterministic BC with a general eavesdropper, the semi-deterministic BC with a more-noisy eavesdropper and the Wiretap BC where users exhibit a less-noisiness order between them. Finally, by rewriting our outer bound to encompass the characteristics of parallel channels, we also derive the secrecy capacity region of the product of two inversely less-noisy BCs with a more-noisy eavesdropper. We illustrate our results by studying the impact of security constraints on the capacity of the WBC with binary erasure (BEC) and binary symmetric (BSC) components.

#### **2.2.4 Publications**

- [1] Villard, J.; Piantanida, P.; Shamai, S., "Secure Transmission of Sources Over Noisy Channels With Side Information at the Receivers," *Information Theory, IEEE Trans. on*, vol.60, no.1, pp.713,739, Jan. 2014
- [2] M. Benammar, P. Piantanida, and S. Shamai (Shitz), "Multiple Description Coding for the Compound Broadcast Channel," in *Information Theory Workshop (ITW)*, 2014 IEEE, 2-5 November in Hobart, Tasmania, Australia.
- [3] M. Benammar, P. Piantanida, and S. Shamai (Shitz), "On the Compound Broadcast Channel: Multiple Description Coding and Interference Decoding," *Information Theory, IEEE Trans. on*, September 2014, (Submitted)

- 
- [4] M. Benammar, P. Piantanida, and S. Shamai (Shitz), "Dirty-Paper Coding Techniques for Compound MISO Broadcast Channels: A DoF Analysis," in 9th International Conference on Cognitive Radio Oriented Wireless Networks, Oulu, Finland, June 2014, (invited paper)
  - [5] R. Tandon, P. Piantanida, and S. Shamai (Shitz), "On Multi-User MISO Wiretap Channels with Delayed CSIT," in Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on, 2014
  - [6] M. Benammar and P. Piantanida, "Secrecy capacity region of some classes of wiretap broadcast channels," Information Theory, IEEE Trans. on, July 2014, (Submitted)
  - [7] M. Benammar and P. Piantanida, "On the secrecy capacity region of the wiretap broadcast channel," in Information Theory Workshop (ITW), 2014 IEEE, November
  - [8] M. Benammar and P. Piantanida, "On the secrecy capacity region of a class of parallel wiretap broadcast channel," in 52th Annual Allerton Conference on Communication, Control, and Computing, 2014

## 2.3 JRA # 1.1.1.3: Communications Performance of Large Dimensional Systems

Leader : Romain Couillet,

Researchers involved : Mérouane Debbah, Marco di Renzo, Pablo Piantanida (CNRS/Supélec), Aris Moustakas (IASA), Luca Sanguinetti (CNIT/UniPI)

### 2.3.1 Description of Activity

This JRA is at the border between information theory and random matrix theory/stochastic geometry to study the rate performance of large dimensional systems of the multi-cell type, along with the error rate performance of finite block length communications in MIMO fading channels (following the works in the SISO AWGN case).

### 2.3.2 Relevance with the identified fundamental open issues

This JRA addresses the following open issues:

- Determining optimal power control strategies in multicell settings for large dimensional communication channels
- Determining the optimal probability of error in point-to-point communications in the MIMO setting, assuming large number of antennas and large but finite communication blocklengths
- Modelling and analyzing in a tractable way the distribution SINR of large multi-cellular environments from a stochastic geometry approach.

### 2.3.3 Main Results Achieved in the Reporting Period and planned activities

Aspect 1 was dealt with in a work that considers the downlink of a single-cell multi-user MIMO system in which the base station (BS) makes use of  $N$  antennas to communicate with  $K$  single-antenna user equipments (Ues), the UEs moving around in the cell according to a random walk mobility model. The aim was to determine the energy consumption distribution when different linear precoding techniques are used at the BS to guarantee target rates within a finite time interval  $T$ . The analysis is conducted in the asymptotic regime where  $N$  and  $K$  grow large with fixed ratio. Recent results random matrix theory were used to provide concise formulae for the asymptotic transmit powers and beamforming vectors for all considered schemes, along with a deterministic approximation of the energy consumption, given in closed form expressions. In practice, these results were used to determine an approximation of the probability that a battery-powered BS runs out of energy and also to design the cell radius for minimizing the energy consumption per unit area.

Regarding Aspect 2, a finite block-length  $n$  regime of an  $N \times K$  MIMO Rayleigh block-fading channel was studied in which the second-order coding rate for the channel was evaluated. The study is at the onset difficult because the channel does not satisfy the ergodic requirements to apply the usual tools for the analysis of the second-order coding rate. Nonetheless, taking a large dimensional approach, it was made possible to study the asymptotic behavior of the error probability when the coding rate is a perturbation of order  $O(1/\sqrt{nK})$  of the asymptotic capacity while the block-length  $n$ , and the number of transmit and receive antennas  $K$  and  $N$ , respectively, grow infinitely large at the same rate. In this asymptotic regime, new lower and

upper bounds on the optimal average error probability were established. These were then applied to obtain a tight closed-form approximation of an upper bound for finite  $n$ .

As for Aspect 3, a new analytical framework was proposed which, at the same time, reduces the usual computational effort of multiple integrals and is flexible enough for application to arbitrary fading distributions (including correlated composite channel models). The framework leverages the application of recent results on the computation of the ergodic capacity in the presence of interference and noise. It is in particular applicable to multi-tier cellular networks with long-term averaged maximum biased-received-power tier association, with only a single or a two-fold numerical integral for arbitrary fading channels.

#### **2.3.4 Publications**

- [1] L. Sanguinetti, A. Moustakas, E. Bjornson, and M. Debbah 'Large System Analysis of the Energy Consumption Distribution in Multi-User MIMO Systems with Mobility', submitted to IEEE Trans. Wireless Commun., May 2014.
- [2] L. Sanguinetti, A. Moustakas, and M. Debbah 'Interference Management in 5G Reverse TDD HetNets: A Large System Analysis', submitted to IEEE J. Select. Areas Commun. (Special Issue on HetNets), July 2014.
- [3] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, M.-S. Alouini 'Power Efficient Low Complexity Precoding for Massive MIMO Systems', submitted to IEEE Global Conference on Signal and Information Processing (GlobalSIP), June 2014.
- [4] J. Hoydis, R. Couillet, P. Piantanida, "The Second-Order Coding Rate of the MIMO Rayleigh Block-Fading Channel", (submitted to) IEEE Transactions on Information Theory.
- [5] J. Hoydis, R. Couillet, P. Piantanida, "Bounds on the Second-Order Coding Rate of the MIMO Rayleigh Block-Fading Channel", IEEE International Symposium on Information Theory, Istanbul, Turkey, 2013.
- [6] M. Di Renzo, A. Guidotti, and G. E. Corazza, "Average Rate of Downlink Heterogeneous Cellular Networks over Generalized Fading Channels – A Stochastic Geometry Approach".

## **2.4 JRA # 1.1.2.1: Network Coding Schemes for Relay Channels**

Leader: Stephan Pfletschinger (CTTC)

Researchers involved: Monica Navarro (CTTC), Carmine Vitiello (CNIT/UniPI), Giacomo Bacci (CNIT/UniPI), Marco Luise (CNIT/UniPI)

### **2.4.1 Description of Activity**

We consider different decoding options for the joint decoding of the channel and the network code in the multi-way relay channel, starting with the important special case of two-way relaying. For this particular case, it has been found earlier that joint decoding of both the network-coded as well as the individual messages, joint decoding can perform significantly better than individual decoding and for low rates even shows better results than lattice coding. We are considering decoding options for different channel codes, which is particularly interesting for the practically relevant cases of asymmetric channel conditions. In this respect, we define joint and separate decoding strategies for trellis codes, turbo codes and non-binary LDPC codes as the most general case and evaluate their performance for the AWGN and for fading channels. We extend these decoding strategies to more than two users and consider the related scenario of uncoordinated multiple access schemes, among which the most prominent ones are known as coded slotted ALOHA.

### **2.4.2 Relevance with the identified fundamental open issues**

This JRA addresses the following open issues:

- Optimum decoding strategies for uplink of the two-way relay channel, in particular for realistic modulation and coding schemes
- Achievable rates with functional decoding, which recovers the network-coded packet at the relay without necessarily decoding the involved packets individually
- How to best combine channel and network coding with higher-order modulations

### **2.4.3 Main Results Achieved in the Reporting Period and planned activities**

We have applied non-binary coding schemes to the two-way relay channel and have exploited principles from physical-layer network coding and multi-user detection. It was found that joint decoding provides significant performance benefits for the fast fading channel and for higher-order modulations in the block fading channel while for the AWGN channel the gains are moderate. We have shown how non-binary channel coding combines well with the principles of network coding.

We have shown the benefits of joint quaternary decoding applied to off-the-shelf LTE turbo channel codes in the TWRC, especially in case of a symmetric channel. Joint decoding exhibits better error performance compared to separate decoding and SIC, approaching single user performance for symmetric channel conditions.

For the uncoordinated multiple-access channel, we have proposed a new scheme based on PNC and MUD that aims at retrieving innovative packets from collisions in slotted ALOHA systems. Starting from two well-known techniques, namely separate decoding and successive interference cancellation (SIC), we have presented an additional decoding scheme which tries to decode the largest number of innovative packet combinations by simply modifying the detector at the receiver side, without making any modification to the channel decoder. We also evaluated

the joint decoding of all collided packets and the subsequent detection of packet combinations, which constitutes the optimum approach, but is also characterized by a high complexity and requires modifications at the decoder. Simulation results show that the new scheme achieves substantial gains compared to separate decoding and to SIC.

#### **2.4.4 Publications**

- [1] Stephan Pfletschinger, Carmine Vitiello, Monica Navarro , “Decoding Options for the Symmetric and Asymmetric Turbo-Coded Two-Way Relay Channel”, invited paper at *European Conference on Networks and Communications (EuCNC)*, Bologna, 23-26 June 2014.
- [2] Stephan Pfletschinger, Dirk Wübben, Giacomo Bacci, “Physical-Layer Network Coding with Non-Binary Channel Codes”, *invited paper at 31th URSI General Assembly and Scientific Symposium*, Beijing, China, 16-23 August 2014.
- [3] Stephan Pfletschinger, Monica Navarro, Giuseppe Cocco, “Interference Cancellation and Joint Decoding for Collision Resolution in Slotted ALOHA”, *International Symposium on Network Coding (NetCod)*, 27-28 June 2014.

## 2.5 JRA 1.1.2.2: Optimization approaches for heterogeneous networks

Leaders : Beatriz Lorenzo, Savo Glisic (UOULU)

Researchers involved: Jordi Pérez-Romero, Juan Sánchez-González, Ramon Agustí (UPC), Sina Khatibi, Luisa Caeiro, Luis M. Correia (INOV)

### 2.5.1 Description of Activity

The general framework for this JRA focuses on heterogeneous networks comprised of multiple technologies such as cellular, wireless local area networks, etc., and including also the possibility of resource virtualization. Under this main general framework the global objective of this JRA is the development of efficient resource allocation strategies to ensure the QoS requirements and the efficiency in the resource usage. In this respect, the activity has been organized around the following scenarios:

- Scenario 1: Heterogeneous Wi-Fi and Cellular with Multi-Hop capabilities. It assumes a network architecture in which a given terminal can receive a mobile service either by connecting directly to the cellular base station or by connecting through another terminal that acts as Access Point (AP) and relays the traffic to/from the cellular base station (e.g. through Device-To-Device technologies such as LTE D2D, Wi-Fi Direct, etc. 0[2][3]). In this scenario, given the randomness associated to the propagation, as well as the variability in the generation of data traffic, there will be situations in which it may be more efficient for a certain mobile terminal to connect to one or another terminal acting as AP or to connect directly to the infrastructure. In this respect, the activity investigates the optimization of the connectivity of the different terminals with the target to minimize the total transmission power in the scenario.
- Scenario 2: Dynamic Network architecture scenario: In this scenario, we consider an advanced wireless technology where certain class of wireless terminals can be turned temporarily into an access point any time while connected to the Internet. This creates a Dynamic Network Architecture (DNA) since the number and location of these access points vary in time. DNA provides significant soft capacity in the network enabling wide range of adaptations to traffic demand variations without any change in the fixed network infrastructure. An optimization framework is developed to optimize different aspects of this new architecture. First, we optimize the network by choosing the most convenient set of available APs to provide the QoS levels demanded by the users. To exploit the soft capacity provided by the DNA, an economic model is developed to award the users by adding a credit (negative price) to their bills while acting as APs for other users in their vicinity. This serves as an incentive for the users to efficiently use the network resources. Different options for the pricing mechanism are presented for wired and wireless Internet backhaul. As the change in the terminal's role (from user to AP) can make the system prone to eavesdropping, the user's requirements in terms of security are also considered in the selection of the AP.
- Scenario 3: Radio Resource Management for Virtual Radio Access Networks: In this scenario, it is assumed that multiple Virtual Network Operators (VNOs) with different requirements, objectives, and SLAs (Service Level Agreements) co-exist on the same physical Radio Access Network (RAN) infrastructure. The existing solutions, e.g., RAN sharing, divide the radio resources either statically or dynamically among VNOs. In contrast, the approach considered in this section is to aggregate and manage all the radio resources while offering the VNOs isolation in addition to simplicity-of-use (i.e., non-transparency), multi-RATs (Radio Access Technologies) and network element abstraction. By means of virtualisation of radio resources, the VNOs are served by connectivity per service class upon their request. Each VNO may have different

requirements and objectives. Meeting these objectives while optimising the resource usage and the other additional objectives (e.g., guaranteeing fairness among multiple instances) make Virtual Radio Resource Management (VRRM) an elaborated procedure. The importance and complexity of VRRM is the reason to count it as the milestone in realisation of end-to-end virtual radio network.

### **2.5.2 Relevance with the identified fundamental open issues**

This JRA addresses the following fundamental open issues:

- Development of novel architectural techniques and optimization approaches for efficiently providing ubiquitous broadband access.
- Power efficient optimization of the connectivity criteria in heterogeneous cellular network with D2D capabilities.
- Optimization of the resource management in heterogeneous virtual Radio Access Networks.
- Economic models for the cooperation between multiple operators

### **2.5.3 Main Results Achieved in the Reporting Period and planned activities**

The progress achieved during the period November 2013 - October 2014 can be summarized in the following points, which are further detailed in the annexes of section 4:

- Scenario 1: A new optimization framework has been elaborated to decide the best connectivity option for each terminal in a heterogeneous network with D2D capabilities so that the total power consumption in the scenario is minimized while ensuring the bit rate requirement associated to the mobile broadband services of each terminal. A new distributed strategy based on Q-learning and softmax decision making has been proposed as a means to implement the presented optimization framework. In this approach, each terminal autonomously decides the most appropriate AP (or the cellular base station) to receive the required service based on its previous experience of using the different APs. The main advantage of this type of distributed approaches is that they allow for a reduction in complexity and signaling overhead in comparison to centralized approaches that address the global optimization considering jointly all the APs and terminals. Therefore, they can scale better when increasing the network size. The proposed strategy has been evaluated under different conditions, revealing that it achieves a performance very close to the optimum one, and that it can allow significant power consumption reductions in the order of 20-50% depending on the specific configuration with respect to the classical approach in which the terminals connect directly to the cellular base station. The proposed approach is also benchmarked against a centralized genetic algorithm revealing that it can provide a very similar performance in spite of operating in a decentralized way. The details with respect to these results can be found in the Appendix of section 0.
- Scenario 2: A revision of the models, optimization framework and results has been done based on the reviews received on the submitted journal paper. In particular, more details are now provided on the practical implementation of the algorithms. Further comments are included to justify the election of the genetic algorithm to track the changes in the optimum topology. As we consider a high dense network, to be able to handle the optimization in such network and solve the problems efficiently, we suggest dividing the DNA macronetwork into clusters. As result of the optimization problems, the optimum topology which provides the data for *intra-cluster* reallocation (handover) is obtained. Further comments are included to explain how the admission control mechanism (2L-AAC) provides the access on a cluster basis. The *inter-cluster* handover may be handled by applying clustering/re-clustering algorithms after a change in the

traffic occurs. More details are provided in the Section **Errore. L'origine riferimento non è stata trovata.** of the appendix.

- Scenario 3: A comprehensive analytical model for VRRM is proposed which has two main parts: estimation of available virtual radio resources and resource allocation. In the first part, the set of available Radio Resource Units (RRUs), i.e., timeslot in GSM, code in UMTS, and radio block in LTE, is formed. Based on probability function of input Signal-to-interference-plus-noise ratio (SINR), the probability functions related to throughput of each RRU, each RAT and the network are estimated. Having a realistic estimation of the network capacity, the model optimises the weighted throughput by considering priority of different services. The optimisation problem of VRRM is linear programming with a set of constraints such as the minimum and maximum guaranteed data rate, the total network capacity, and fairness. The proposed model is also can support the traffic offloading and the resource shortage situation. The former is done by considering session average data rate to minimise collision rate in Wi-Fi network and the latter is support by introduction of violation. This suggested model is evaluated by a practical heterogeneous access network. The detail of this model in addition to numerical results is presented in Section **Errore. L'origine riferimento non è stata trovata.**

As for the planned activities during the next year, it is expected on the one hand to further evaluate and consolidate the developed approaches. Moreover, it is also expected to work on the integration of the three scenarios into a single framework that includes the concept of network virtualization.

#### **2.5.4 Publications**

The following papers have been submitted:

- [1] A.S. Shafigh, B. Lorenzo, S. Glisic, J. Pérez-Romero, L.A. DaSilva, A. B. MacKenzie, J. Röning, "A Framework for Dynamic Network Architecture and Topology Optimization", IEEE/ACM Transactions on Networking, Revision submitted on August 2014.
- [2] J. Pérez-Romero, J. Sánchez-González, R. Agustí, B. Lorenzo, S. Glisic, "Power Efficient Resource Allocation in a Heterogeneous Network with Cellular and D2D Capabilities", submitted to IEEE/ACM Transactions on Networking, July, 2014.

## 2.6 JRA 1.1.2.3 Traffic Dynamics: Routing and Topology Reconfiguration

Leader: P. Mertikopoulos (CNRS)

Researcher involved: A. L. Moustakas (IASA)

### 2.6.1 Description of Activity

This JRA focuses on the increase in the performance of a wireless network that results from optimizing packet traffic routing - possibly over several different paths, or by adapting to a changing network topology (such as the one induced by users going online and offline). Specifically, we focus on multi-path routing techniques that are known to yield significant increase in terms of packet latencies and achieved throughput. Despite their projected efficiency, such techniques are hampered by the massive size of current wireless networks, so it is not clear how an optimized multi-path routing scheme may be implemented in a decentralized way; furthermore, such considerations should also take into account the fact that wireless users are also interested in maximizing the rate of their transmissions (and not only their latencies). In view of the above, we examine distributed learning schemes for throughput-efficient multi-path routing. Our approach incorporates techniques from game theory, learning and online optimization and we focus on the development of algorithms that optimize the traffic distribution in the network in a robust and decentralized way. In these considerations, a major challenge occurs when delays and/or packet loss probabilities fluctuate unpredictably due to random exogenous factors (such as packet drops, fluctuations in link quality, etc.). In the presence of such perturbations, we examine the robustness of the proposed algorithms and determine conditions under which this distribution is near-optimal with high probability.

During the current reporting period, OULU expressed interest in JRA 1.1.2.3 and potential learning and game-theoretic approaches for efficient routing and scheduling in cognitive radio environments were identified after a series of Skype meetings; work on this direction is ongoing and is currently at an exploratory phase. Regarding work that was started during the previous reported period, a semester-long visit of the JRA leader was organized to IASA in order to finalize the development of fully decentralized and asynchronous algorithms for multi-path routing in networks with possibly changing topologies [MM14]. This work was based on recent learning techniques based on powerful stochastic approximation tools that were developed by CNRS in the framework of this JRA during the current and previous reporting period [CGM14]. The research output of these activities is summarized in the publications section below and consists of 1 accepted journal paper, 1 submitted and 1 more that is currently in preparation.

### 2.6.2 Relevance with the identified fundamental open issues

Throughput maximization and latency minimization is one of the core requirements for the efficient operation of wireless networks, especially when the network in question is deployed at a very large scale that precludes centralized control. In turn, these requirements both rely on the efficient distribution of packet traffic over the network, so any attempt at distributed network optimization must overcome several major challenges, such as unpredictable packet drops, "black swan" catastrophes in the network (link breakdown, stampedes of elephant packets, etc.). The development of robust algorithms for efficient network operation is one of the core open issues of N#, and the overarching objective of JRA 1.1.2.3 is to apply tools and techniques from the theory of learning in order to develop distributed optimization algorithms and techniques.

### 2.6.3 Main Results Achieved in the Reporting Period and planned activities

Much of the work in this JRA rests on a class of general game-theoretic learning algorithms whose development started during the first year of N# and was finalized in the current reporting period [CGM14]. Specifically, starting from a heuristic learning scheme for strategic person games with a large number of players, we derived a novel class of continuous-time learning dynamics consisting of an exponential learning driver adjusted by a penalty term that allows learners to efficiently explore the game's strategy space. These penalty-regulated dynamics are equivalent to players keeping an exponentially discounted aggregate of their on-going payoffs and then using a regularized best response to pick an action based on these performance scores. Owing to this inherent duality, the proposed dynamics converge to (arbitrarily precise) approximations of Nash equilibria in potential games, a property which we exploit further to design a discrete-time, learning algorithm which retains these convergence properties and only requires the agents to observe their realized payoffs. Importantly, the algorithm remains robust and retains its convergence properties in the presence of stochastic perturbations and observation errors of arbitrary magnitude, and it does not require any synchronization (or other form of coordination) between players.

The derived class of learning algorithms is of fundamental importance to this JRA because packet drops, lag, and delayed receipt of acknowledgments are issues that cause stochastic fluctuations and delays/asynchronicities to any attempt at optimizing the traffic distribution of the network. In view of all this, we address the problem of efficient traffic routing in networks by means of a distributed "exponential learning" scheme based on the Gibbs distribution of statistical physics. The resulting decentralized algorithm converges to a traffic distribution that minimizes aggregate latency in the network, and this convergence is exponentially fast: the network gets within  $\epsilon$  of an optimum traffic assignment in time which is  $O(\log(1/\epsilon))$ . Of course, a major challenge occurs when delays fluctuate unpredictably due to random exogenous (or endogenous) factors. Despite these stochastic perturbations, we find that optimum traffic distributions where the traffic of an origin-destination pair is not split over different routes remain stochastically stable irrespective of the fluctuations' variance, and users converge to this optimum traffic distribution almost surely. More generally, if users split traffic over several paths in the optimum traffic distribution, the proposed learning algorithm converges in probability to a steady-state distribution which is concentrated around this optimum point.

In the next phase of N#, we plan to finalize the work above and submit it for publication to the IEEE/ACM Transactions on Networking. One journal paper is already in preparation and we expect to submit it in early 2015. In addition, one final semester-long research visit of P. Mertikopoulos (CNRS) to IASA is planned to take place in the spring semester of 2015 in order to proceed further with issues having to do with topology reconfiguration. Finally, the research directions that have been identified jointly by CNRS and OULU will be further explored with the aim of establishing concrete use-cases for joint scheduling/routing optimization.

### 2.6.4 Publications

[1] P. Coucheney, B. Gaujal, and P. Mertikopoulos, "*Penalty-regulated dynamics and robust learning procedures in games*," Mathematics of Operations Research, to appear. ArXived at <http://arxiv.org/abs/1303.2270>.

[2] P. Mertikopoulos and A. L. Moustakas, "An efficient learning scheme for routing in the presence of stochastic fluctuations," with A. L. Moustakas, submitted.

## 2.7 JRA # 1.1.3.1: Spatially Coupled Codes

Leader: Michael Lentmaier (ULUND)

Researchers involved: Iryna Andriyanova (CNRS), Najeeb ul Hassan (TUD)

### 2.7.1 Description of Activity

This JRA investigates spatially coupled codes in wireless communication scenarios. Three activities were ongoing during this reporting period: i) Spatially coupled codes for block-fading channels, ii) Spatially coupled code design for flexible rates, and iii) Non-uniform windowed decoding schedules for spatially coupled codes.

### 2.7.2 Relevance with the identified fundamental open issues

The mobile-radio channel for multicarrier communications can be modelled on each carrier as a slow, flat fading together with additive noise. In many cases, the channel coherence time is (much) longer than one symbol duration. Thus several symbols are affected by the same fading coefficient. An example of such a channel model is the *block-fading* channel. In the block-fading channel, coded information is transmitted over a finite number of fading blocks to provide diversity. The diversity order of the code is an important parameter that gives the slope of the word error rate of the decoder. Convolutional codes, in general, are known to be suitable for transmission over block-fading channels and the diversity can be increased by increasing the constraint length of the code. Whereas, strong error correction codes like low-density parity-check (LDPC) codes fail to provide diversity gains without a special structure of the code. An example of such a structure is given by the root-LDPC codes introduced in [BG<sup>+</sup>10]. However, designing root-LDPC codes with diversity order greater than 2 requires codes with rate less than 1/2. The special structure of the codes makes it a complicated task to generate good root-LDPC codes with high diversity (and thus low rate). Rate flexibility and efficient window decoding, on the other hand, are mentioned in a recent survey among the open research problems related to the practical realization of SC-LDPC codes [CD<sup>+</sup>14].

### 2.7.3 Main Results Achieved in the Reporting Period and planned activities

#### Spatially Coupled Codes for Block-Fading Channels

Partners: M. Lentmaier (ULUND), I. Andriyanova (CNRS), N. ul Hassan (TUD)

We consider spatially coupled LDPC (SC-LDPC) codes to take advantage of the convolutional structure. The first results in [HL<sup>+</sup>12] show the density evolution outage probabilities for random SC-LDPC codes. The results suggest that increasing the coupling parameter (constraint length) of the SC-LDPC code improves the diversity order of the code. For a (3,6)-regular SC-LDPC code with rate  $R = 1/2$  and memory 4 a remarkable diversity of  $d = 10$  is achieved without the need for any specific code structure. The SC-LDPC codes show more robustness against the variation in the channel parameter compared to root-LDPC codes, i.e., designing a code for a specific value of channel coherence time is not required and the diversity order strongly depends on the memory of the code. We observed that as long as the constraint length of the code contains more than one block, the SC-LDPC code can provide a diversity order greater than 1. To the best of our knowledge, this was the first work that considers spatially coupled codes in such a setting. We consider this activity as very relevant for NEWCOM# and plan to continue in this direction. In particular, we are interested in a deeper understanding of the mechanisms that allow exploiting code diversity, which will hopefully also lead to more general design rules.

## Spatially Coupled Code Design for Flexible Rates

Partners: M. Lentmaier (ULUND), Walter Nitzold (TUD)

We investigate the suitability of coupled regular LDPC code ensembles with respect to rate-flexibility. Regular ensembles with good performance and low complexity exist for a variety of specific code rates. On the other hand it can be observed that outside this set of favorable rational rates the complexity and performance penalty become unreasonably high. We therefore propose ensembles with slight irregularity that allow us to smoothly cover the complete range of rational rates. Our simple construction allows a performance with negligible gap to the Shannon limit while maintaining complexity as low as for the best regular code ensembles. At the same time the construction guarantees that, asymptotically, the minimum distance grows linearly with the length of the coupled blocks. We plan to continue this activity and address the issue of incorporating rate-compatibility (incremental redundancy) into the design.

## Non-Uniform Windowed Decoding Schedules for Spatially Coupled Codes

Partners: M. Lentmaier (ULUND), N. ul Hassan (TUD)

In the first reporting period of NEWCOM# the results for non-uniform window decoding schedules were reported for the binary erasure channel (BEC). The results showed that the decoding complexity can be reduced up to 60% when non-uniform improvement based schedules are applied within a window compared to conventional uniform schedules. Now the analysis is extended for the AWGN channel and similar results are obtained. An article for possible publication in IEEE Transactions on Communications is currently in preparation.

### 2.7.4 Publications

- [1] N. ul Hassan, M. Lentmaier, I. Andriyanova, G. P. Fettweis, "Improving code diversity on block-fading channels by spatial coupling," in Proceedings of the IEEE International Symposium on Information Theory (ISIT), pp. 2311-2315, Jun. 2014
- [2] W. Nitzold, M. Lentmaier and G. P. Fettweis "Spatially-Coupled Nearly-Regular LDPC Code Ensembles for Rate-flexible Code Design" in Proceedings of the IEEE International Conference on Communications (ICC), Jun. 2014
- [3] N. ul Hassan, A.E. Pusane, M. Lentmaier, G.P. Fettweis, D.J. Costello, "Decoding schedules for Spatially Coupled LDPC Codes" in preparation, to be submitted to IEEE Transactions on Communications.

### Bibliography

- [BG<sup>+</sup>10] J. Boutros, A. Guillen i Fabregas, E. Biglieri, and G. Zemor, "Low-density parity-check codes for nonergodic block-fading channels," IEEE Trans. on Inf. Theory, vol. 56, no. 9, pp. 4286–4300, 2010.
- [CD<sup>+</sup>14] D. Costello, L. Dolecek, T. Fuja, J. Kliever, D. Mitchell, and R. Smarandache, "Spatially coupled sparse codes on graphs: theory and practice," IEEE Communications Magazine, vol.52, no.7, pp.168 – 176, July 2014.
- [HL<sup>+</sup>12] N. ul Hassan, M. Lentmaier, I. Andriyanova, and G.P. Fettweis "Improving Code Diversity on Block-Fading Channels by Spatial Coupling", accepted for publication in International Symposium on Information Theory (ISIT) 2014.
- [LF<sup>+</sup>09] M. Lentmaier, G. Fettweis, K. Zigangirov, and D. Costello, "Approaching capacity with asymptotically regular LDPC codes," in Proc. Information Theory and Applications Workshop (ITA), Feb. 2009, pp. 173 –177
- [NLF14] W. Nitzold, M. Lentmaier and G. P. Fettweis "Spatially-Coupled Nearly-Regular LDPC Code Ensembles for Rate-flexible Code Design" in Proceedings of the IEEE International Conference on Communications (ICC), Jun. 2014.

## 2.8 JRA # 1.1.3.2: Non Binary Codes

Leader: Guido Montorsi (CNIT/POLITO),

Researchers involved : Guido Masera (CNIT/POLITO), Muhammad Awais (CNIT/POLITO), David Declercq (CNRS), Florence Alberge (CNRS), Jossy Sayir (UCAM)

### 2.8.1 Description of Activity

As higher spectral efficiency and throughput targets are required for future communication systems, existing capacity-approaching channel coding schemes need to be developed further to work under more demanding scenarios. In general, Task 1.1.3 aims to develop capacity-approaching channel codes for diverse set of future application scenarios. The JRA 1.1.3.2 within this task is to consider efficient decoding algorithms (message-passing) for non-binary LDPC and related codes and the design of good class of non binary codes. Non binary codes are a natural choice for systems achieving large spectral efficiencies employing high cardinality modulation sets. Recently considerable steps forward have been made in this field with the introduction of decoding algorithms like the Extended Min Sum (EMS) or the Analog-Digital Belief Propagation (ADBP) which are competitive in term of complexity with those of binary codes. These algorithms use messages relative to non binary quantities allowing to easily couple the iterative decoders with non binary detectors that typically are present in receiver for highly bandwidth efficient systems. In this activity, starting from EMS and ADBP solutions, we develop several density evolution algorithms achieving different trade-offs between performance and complexity to further improve the efficiency of the non-binary decoding algorithms and to provide design and analysis tools for the correspondent encoders. We also investigate a new class of non-linear iteratively decodable non-binary codes that show promise with respect to performance attained for high spectral efficiency at low block lengths.

### 2.8.2 Relevance with the identified fundamental open issues

Non-binary LDPC codes [1][2][3] have shown improved performance over binary LDPC codes especially for short-length frames. Even if more complex than the binary case, the decoding of non-binary LDPC codes remains tractable. An iterative decoding based on the sum-product algorithm was proposed by Davey and MacKay in [1]. MacKay and Davey introduced a Fast-Fourier-transform (FFT) in the decoding process to reduce computational complexity [4]. This contribution was further improved in [5], and [6] and motivated more research effort for reaching good trade-off between complexity and efficiency [7][8].

Message passing for decoding non-binary LDPC shares common features with extrinsic information passing used in turbo-codes. Another amount of research focused on the design of non-binary LDPC codes by extending standard tools such as EXIT charts from binary to non-binary codes. In [12], an EXIT analysis is provided to visualize the exchange of information between the variable node decoder and the check node decoder leading to better codes compared to [13]. A similar work is proposed in [14] using an index-based approach. The extrinsic information can be computed with low complexity thanks to an alternative to the histogram approach.

### 2.8.3 Main Results Achieved in the Reporting Period and Planned Activities

#### Iterative Decoding Algorithm for Non Binary Encoders

Partner: G. Montorsi (CNIT/POLITO)

In the previous reporting period new iterative decoding algorithms for non-binary codes have been introduced and developed. This year, in the publication "Awais, M.; Masera, G.; Martina, M.; Montorsi, G., "VLSI Implementation of a Non-Binary Decoder Based on the Analog Digital Belief Propagation," Signal Processing, IEEE Transactions on , vol.62, no.15, pp.3965,3975, Aug.1, 2014", we finalized the work started in the previous reporting period. We presented the VLSI hardware implementation of a novel Belief Propagation (BP) algorithm introduced in [9] and named as Analog Digital Belief Propagation (ADBP). The ADBP algorithm works on factor graphs over linear models and uses messages in the form of Gaussian-like probability distributions by tracking their parameters. In particular, ADBP can deal with linear systems where variables can be discrete and/or wrapped. A variant of ADBP can then be applied for the iterative decoding of a particular class of non binary codes and yields decoders with complexity independent of alphabet size, thus allowing the construction of efficient decoders for digital transmission systems with **unbounded** spectral efficiency. In [RD- 1] we proposed some simplifications to the updating rules for ADBP algorithm that are suitable for hardware implementation. In addition, we analyzed the effect of finite precision on the decoding performance of the algorithm. A careful selection of quantization scheme for input, output and intermediate variables allowed us to construct a complete ADBP decoding architecture that performs close to the double precision implementation and shows a promising complexity for large values of the alphabet cardinality  $M$ . Finally, synthesis results of the main processing elements of ADBP are reported for 45 nm standard cell ASIC technology.

It should be noticed that ADBP cannot be applied to all types of linear non binary codes over  $M$  as multiplications by elements different from  $\pm 1$  are not allowed in the graph. This code ensemble has been analyzed in [10] (named "binary" LDPC ensemble over  $M$ ) and [11] (named modulo- $M$  LDPC ensemble). In both papers it is shown that the ensemble is capacity achieving as its distance spectrum approaches that of random codes while the underlying graph connectivity grows.

The exceptional complexity reduction achieved from using the ADBP, together with asymptotic results of [10] and [11] motivated us for further research effort in the design of good non binary LDPC encoders within the class. The introduction of ADBP in fact converts the decoding complexity drawback associated to non binary codes into an advantage and the problem of the applicability of non-binary coding techniques back to an encoding design problem. It is worth noting that the scaling of complexity with graph density is clear from hardware implementation results. Thus, a trade-off between graph density and performance and code construction issues have to be further addressed.

The activity of designing good non binary codes suitable for ADBP is currently ongoing. The main design techniques that are currently under investigation are those that are used for the design of classical binary codes:

- Design of degree distributions for left and right irregular bipartite graphs through density evolution and EXIT charts techniques
- Minimization of graph girth and Progressive Edge Growth Tanner graphs

These techniques have to be suitably modified and adapted to the considered class of non binary codes. The final planned activity for this JRA is that of providing a complete survey of binary and non binary coding techniques for realizing high spectrally efficient transmission systems ( $\gg 1$  bit/dimension). The survey will consider the performance/ flexibility/ complexity trade-offs of the following most promising techniques, namely:

- Serially Concatenated and Parallel Concatenated TCM

- Multistage encoding and SIC
- Bit Interleaved Coded Modulation and pragmatic decoder
- Non binary LDPC codes coupled with high order modulations and suitable decoder (FFT, log-FFT, EMS, Min-Max)
- mod- $M$  LDPC encoders and ADBP

Although this theoretical activity is not join across different partners, it is preliminary to a planned inter-track activity regarding implementation and testing of ADBP.

### **Non-linear iteratively decodable binary codes**

*Partner:* Jossy Sayir (UCAM)

We studied class of non-linear codes with local constraints where each constraint requires participating code symbols to take on distinct values over a finite alphabet. The constraint node operation in belief propagation decoding for this type of code was shown to be equivalent to the computation of a Cauchy permanent. A trellis-based decoder was devised whose complexity, while still of exponential in the alphabet size, is much reduced for alphabet sizes of interest. Transmission over erasure channels was investigated and the constraint node operation was modified into a partial trellis search method with improved complexity with respect to other known methods. Finally, density evolution was developed and results analysed. The study of non-linear codes with local constraints is of potential interest for wireless communications because their performance could outstrip that of linear codes for short block lengths and they are iteratively decodable using the well studied belief propagation algorithm. However, there are some open problems to solve before such codes become practical, including how they can be efficiently encoded, how their rate can be determined, and how their performance compares to linear codes once their design has been optimised and refined.

#### **2.8.4 Publications/Presentations**

Awais, M.; Masera, G.; Martina, M.; Montorsi, G., "VLSI Implementation of a Non-Binary Decoder Based on the Analog Digital Belief Propagation," Signal Processing, IEEE Transactions on , vol.62, no.15, pp.3965,3975, Aug.1, 2014, doi: 10.1109/TSP.2014.2330804.

C. Atkins, J. Sayir, " Density Evolution for SUDOKU Codes on the Erasure Channel", accepted for publication at the 8th International Symposium on Turbo Codes and Iterative Processing (ISTC 2014), Bremen, Germany, August 18-22, 2014.

J. Sayir, C. Atkins, "Erasure channel decoding and density evolution for a class of non-linear codes with local constraints", presentation (with no associated publication) at EuCNC 2014 in Bologna, Italy, June 2014.

### **References**

- [1] M. Davey and D. MacKay, "Low-density parity check codes over  $GF(q)$ ," IEEE Commun. Lett., vol. 2, no. 6, pp. 165–167, June 1998.
- [2] R. Gallager, "Low density parity check codes," IRE Trans. Info Theory, vol. IT-8, no. 1, pp. 21–28, Jan.1962.
- [3] D. Sridhara and T. Fuja, "LDPC over rings for PSK modulation," IEEE Trans. Info Theory, vol. 51, no. 9, pp. 3209–3220, Sep. 2005.
- [4] D. J. C. MacKay and M. C. Davey, "Evaluation of Gallager codes for short block length and high rate application," in Proc. IMA International Conference on Mathematic and its

- Applications: Codes, Systems and Graphical Models, Springer-Verlag, Ed., New York, 2000, pp. 113–130.
- [5] L. Barnault and D. Declercq, "Fast decoding algorithm for LDPC over  $GF(2q)$ ," in Information Theory Workshop, 2003. Proceedings. 2003 IEEE, 2003, pp. 70–73.
  - [6] D. Declercq and M. Fossorier, "Decoding algorithms for nonbinaryldpc codes over  $GF(q)$ ," IEEE Trans. Commun., vol. 55, no. 4, pp. 633–643, Apr. 2007.
  - [7] N. Mobini, A. H. Banihashemi, and H. Hemati, "A differential binary message-passing LDPC decoder," IEEE Trans. Commun., vol. 57, no. 9, pp. 2518–2523, Sep. 2009.
  - [8] Q. Huang, J. Y. Kang, L. Zhang, S. Lin, and K. A.-G. r, "Two reliability-based iterative majority-logic decoding algorithms for ldpc codes," IEEE Trans. Commun., vol. 57, no. 12, pp. 3597–3606, Dec. 2009.
  - [9] Montorsi, G., "Analog Digital Belief Propagation," *Communications Letters*, IEEE , vol.16, no.7, pp.1106,1109, July 2012, doi: 10.1109/LCOMM.2012.020712.112133
  - [10] Erez, U.; Miller, G., "The ML decoding performance of LDPC ensembles over  $Z_q$ ," *Information Theory*, IEEE Transactions on , vol.51, no.5, pp.1871,1879, May 2005 doi: 10.1109/TIT.2005.846431
  - [11] Bennatan, A; Burshtein, D., "On the application of LDPC codes to arbitrary discrete-memoryless channels," *Information Theory*, IEEE Transactions on , vol.50, no.3, pp.417,438, March 2004, doi: 10.1109/TIT.2004.824917
  - [12] G. Byers and F. Takawira, "EXIT charts for non-binary LDPC codes," in Proc. International Conf. Commun. (ICC), Seoul, Korea, May 2005, pp. 652–657.
  - [13] M. Davey, "Error-correction using Low-Density-Parity-Check Codes," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K, Dec. 1999.
  - [14] J. Kliewer, S. Ng, and L. Hanzo, "Efficient computation of EXIT functions for nonbinary iterative decoding," IEEE Trans. Commun., vol. 54, no. 12, pp. 2133–2136, Dec. 2006.

## 2.9 JRA 1.1.3.3 Coding for Multiterminal Communication Systems

Leader: Erdal Arıkan (BILKENT),

Researchers involved: Shlomo Shamai (TECHNION), Saygun Onay (BILKENT), Pierre Duhamel (CNRS), Pablo Piantanida (CNRS/SUPELEC), Abdellatif Zaidi (CNRS), F. Alberge (CNRS/UPSud), Luc Vandendorpe (UCL), Mohieddine El Soussi (UCL).

### 2.9.1 Description of Activity

This JRA comprises two main research subjects.

- Capacity and coding for multiterminal systems
- Secure transmission over networks

The common thread of these subjects is the quest for understanding characteristics of capacity-achieving code constructions in the multiterminal scenarios. One of the methods used for this purpose is the information-estimation paradigm which investigates the basic features of 'good' (capacity approaching) codes operating on a Gaussian channel. A second method is polar coding which has proven very effective in constructing capacity achieving codes for a number of multiterminal coding scenarios. In addition to these general methods, we employ specific coding techniques to construct novel coding schemes that can be used in practical cooperative coding scenarios such as relays and multiple access channels. Under the secure transmission research theme, we seek to develop physical layer security methods over networks.

The methods investigated in this JRA aim to improve the efficiency of wireless networks through more advanced signalling and coding techniques. The methods are developed with an overriding concern for low-complexity and suitability for practical implementation. At the moment, no joint activity with Track 2 has taken place; however, feasibility studies are underway for demonstration of relaying techniques developed under this JRA using Track 2 facilities.

### 2.9.2 Relevance with the identified fundamental open issues

The ultimate goal of this JRA is to design capacity-achieving signalling and coding schemes in the multi-terminal settings. Thus the work in this JRA lies at the main research frontier in wireless communications. The methods used in this JRA combine basic tools of multi-user information theory with models relevant to wireless networks envisioned for present or future applications. Understanding fundamental principles of good system architectures and design of low-complexity schemes in accordance with such architectures is the main aim. These aims are fully consistent with the goals of this Work Package on "Performance Limits in Wireless Networks".

### 2.9.3 Main Results Achieved in the Reporting Period and planned activities

In this part, we list the main results achieved in the reporting period, grouped by main themes of the task. Details of these achievements are given in the Appendix.

#### Capacity and coding for multiterminal systems

##### **Capacity Region of Cooperative Multiple-Access Channel With States**

Partners: S. Shamai (TECHNION), A. Zaidi (CNRS), P. Piantanida (CNRS/SUPELEC)

A two-user state-dependent multiaccess channel is considered in which the states of the channel are known noncausally to one of the encoders and only strictly causally to the other encoder. Both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. We find explicit characterizations of the capacity region of this communication model in both discrete memoryless and memoryless Gaussian cases. The following publication was made on this topic: "A. Zaidi, P. Piantanida and S. Shamai (Shitz), Capacity Region of Cooperative Multiple-Access Channel With States, IEEE Trans. on Information Theory, vol. 59, no. 10, pp. 6153-6174, October 2013." This work will be continued in the next reporting period by considering more general multiterminal cooperation scenarios.

### **Multiple Description Coding with Polar Codes**

Partners: E. Arıkan (BILKENT), S. Onay (BILKENT)

In the reporting period, we developed a polar coding method for multiple description coding. Multiple description coding set-up consists of a single source, two encoders and three decoders. Decoder 1 has access to the description generated by encoder 1 only and reconstructs the source with distortion  $D_1$ . Similarly, decoder 2 has access to the description generated by encoder 2 only and reconstructs the source with distortion  $D_2$ . There is a third decoder (decoder 0) in the setting which has access to both descriptions and constructs the source with distortion  $D_0$ . The capacity region for this problem is not known in general, however, there is an inner bound called the El Gamal-Cover inner bound. The polar coding method that we developed reaches any point on the dominant face of this bound. A conference paper and a journal publication is being prepared. For future work, we plan to investigate the possibility of formulating a general theorem translation method so that any coding theorem for multiterminal proved by using random coding methods can be translated into a proof by polar coding.

### **Varentropy under Polarization**

Partner: E. Arıkan (BILKENT)

This research which started in the present reporting period investigates the behavior of entropy of data elements as they are operated on by polar transforms. The data elements under consideration are pairs  $(X, Y)$  where  $X$  is a binary random variable and  $Y$  is some side information about  $X$ . The entropy random variable for such a data element is defined as  $h(X|Y) = -\log p(X|Y)$ . The variance of entropy (varentropy) is defined as  $\text{var}(h(X|Y))$ . Varentropy is an important determinant of performance in source and channel coding. A polar transform of order two is a mapping that takes two independent data elements and produces two new data elements (that are in general correlated). We show that the sum of the varentropies of the output data elements is less than or equal to the sum of the varentropies of the input data elements, with equality if and only if at least one of the input data elements has zero varentropy. For further work, we plan to the result here, which has been derived for single-user settings to multi-user scenarios, such as Slepian-Wolf coding. A paper has been submitted to IEEE Trans. Information Theory on the subject. A preliminary version of this work, entitled "A Note on Polarization Martingales", has been presented at the 2014 IEEE 2014 Int. Symp. Information Theory (ISIT 2014).

### **Soft-Decoding-Based Strategies for Relay and Interference Channels**

Partner: S. Shamai (TECHNION)

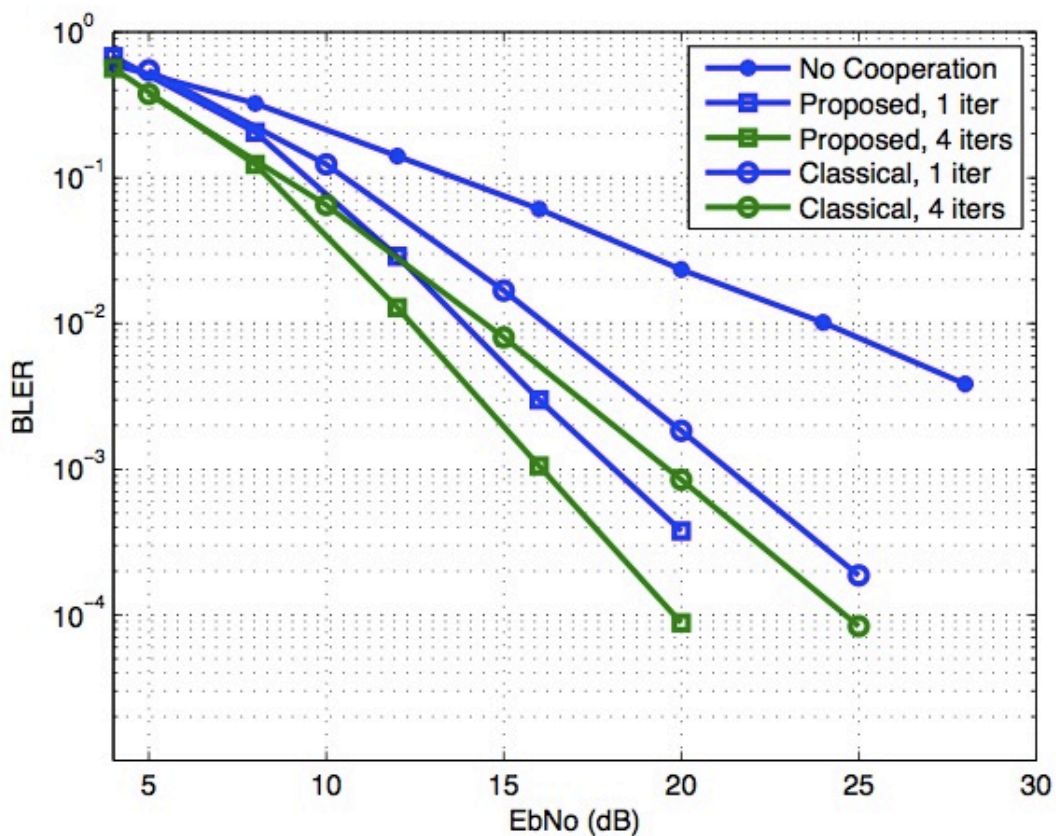
We provided a rigorous mathematical analysis of two communication strategies: soft decode-and-forward (soft-DF) for relay channels, and soft partial interference-cancellation (soft-IC)

for interference channels. Both strategies involve soft estimation which assists the decoding process. We prove that optimal point-to-point codes are unsuitable for soft-IC, as well as for all strategies that apply partial decoding to improve upon single-user detection (SUD) and multiuser detection (MUD), including Han-Kobayashi (HK). The following publication was made: "A. Bennatan, S. Shamai (Shitz), and A. R. Calderbank, "Soft-Decoding-Based Strategies for Relay and Interference Channels: Analysis and Achievable Rates using LDPC Codes", IEEE Trans. on Inform. Theory, vol. 60, no. 4, pp. 1977-2009, April 2014." Research on this work has been completed.

### Towards a backward compatible relaying scheme

Partners. P. Duhamel (CNRS), F. Alberge (CNRS/UPSud), E. Arikan (BILKENT), L. Vandendorpe (UCL), M. El Soussi (UCL)

In an activity that started towards the end of this reporting period, we started investigating a relaying scheme in which forward error correction (FEC) is used as an essential tool in addition to conventional methods such as Amplify and Forward (AF), Decode and Forward (DF), etc. The basic scenario that is studied is as follows: two users (A and B) communicate with a base station on orthogonal channels, and a relay node C helps A and B using only the transmission resources allocated to A and B. Hence the process does not require additional resource allocation to C. This corresponds to a semi-orthogonal Multi Access Relay Channel (MARC) studied in prior work. The relay operates in half duplex mode, and the users protect the transmitted data via convolutional codes of rate 1/3. The figure below depicts the block error rate of non cooperative schemes, of classical (orthogonal) MARC, and of the proposed scheme for the same spectral efficiency (in terms of information bits per channel use) and same total power consumption.



It is seen that the proposed semi-orthogonal MARC scheme clearly outperforms both reference situations. In this initial study, convolutional codes were chosen for reusing previous blocks which allowed some analytical study. In the next period, we will extend the situation to the use of more efficient channel codes, namely LDPCs and polar codes.

### **Sparse Sensing and Sparse Sampling of Coded Signals**

Partner: S. Shamai (TECHNION)

Advances of information-theoretic understanding of sparse sampling of continuous uncoded signals at sampling rates exceeding the Landau rate were reported in recent works. In this work we examined sparse sampling of coded signals at sub-Landau sampling rates and showed that with coded signals the Landau condition may be relaxed and the sampling rate required for signal reconstruction and for support detection can be lower than the effective bandwidth. The following publication was made: "M. Peleg and S. Shamai, "On sparse sensing and sparse sampling of coded signals at sub-Landau rates", Trans. Emerging Tel. Tech., vol. 25, pp. 259-272, 2014." For future work, we plan on taking an information theoretic perspective on a classical sparse-sampling noisy linear model. Our goal is to derive analytical expressions for mutual information terms which play central role in a variety of communications/processing problems. Such expressions were addressed previously either by bounds, by simulations or by the (non-rigorous) replica method from statistical physics. We plan to address the problem rigorously via partition functions and invoke classical random matrix results, namely the Stieltjes and Shannon transforms, as well as the information-estimation relation. Such expressions will facilitate an analytical analysis of coding problems related to sparse and compressed sampling models.

### **Secure transmission over networks**

#### **Secure Degrees of Freedom of MIMO Channels**

Partners: S. Shamai (TECHNION), A. Zaidi (CNRS), L. Vandendorpe (UCL)

We investigated the problem of secure transmission over a two-user multi-input multi-output (MIMO) X-channel with noiseless local feedback and delayed channel state information (CSI) available at transmitters. We characterized the optimal sum secure degrees of freedom (SDoF) region. We showed that, in presence of local feedback and delayed CSI, the sum SDoF region of the MIMO X-channel is same as the SDoF region of a two-user MIMO BC with  $2M$  antennas at the transmitter and  $N$  antennas at each receiver. The following publication was made on the topic: "A. Zaidi, Z. H. Awan, S. Shamai (Shitz) and L. Vandendorpe, Secure Degrees of Freedom of MIMO X-Channels with Output Feedback and Delayed CSI, IEEE Trans. on Information Forensics & Security, vol. 8, no. 11, pp. 1760-1774, November 2013." The theoretical basis established by this work will be extended to more general scenarios in the next reporting period.

#### **Secure Transmission of Sources over Noisy Channels**

Partners: S. Shamai (TECHNION) and P. Piantanida (CNRS/SUPELEC)

We investigated the problem of source-channel coding for secure transmission with arbitrarily correlated side information at both receivers. This scenario consists of an encoder (referred to as Alice) that wishes to compress a source and send it through a noisy channel to a legitimate receiver (referred to as Bob). A general outer bound on the rate-distortion-equivocation region, as well as an inner bound based on a pure digital scheme, is derived for arbitrary channels and

side informations. The following publication was made on the subject: "J. Villard, P. Piantanida, and S. Shamai (Shitz), Secure Transmission of Sources over Noisy Channels with Side Information at the Receivers, IEEE Trans. on Information Theory, vol. 60, no. 1, pp. 713-739, January 2014." This work is planned to be continued in the next reporting period with the inclusion of some practical implementation methods, such as polar codes.

#### **2.9.4 Publications**

- [1] A. Zaidi, P. Piantanida and S. Shamai (Shitz), Capacity Region of Cooperative Multiple-Access Channel With States, IEEE Trans. on Information Theory, vol. 59, no. 10, pp. 6153-6174, October 2013.
- [2] A. Zaidi, Z. H. Awan, S. Shamai (Shitz) and L. Vandendorpe, Secure Degrees of Freedom of MIMO X-Channels with Output Feedback and Delayed CSI, IEEE Trans. on Information Forensics & Security, vol. 8, no. 11, pp. 1760-1774, November 2013.
- [3] J. Villard, P. Piantanida, and S. Shamai (Shitz), Secure Transmission of Sources over Noisy Channels with Side Information at the Receivers, IEEE Trans. on Information Theory, vol. 60, no. 1, pp. 713-739, January 2014.
- [4] M. Peleg and S. Shamai, "On sparse sensing and sparse sampling of coded signals at sub-Landau rates", Trans. Emerging Tel. Tech., vol. 25, pp. 259-272, 2014.
- [5] A. Bennatan, S. Shamai (Shitz), and A. R. Calderbank, "Soft-Decoding-Based Strategies for Relay and Interference Channels: Analysis and Achievable Rates using LDPC Codes", IEEE Trans. on Inform. Theory, vol. 60, no. 4, pp. 1977-2009, April 2014.
- [6] E. Arıkan, "A Note on Polarization Martingales," *Proc. IEEE International Symposium on Information Theory*, pp. 1466-1468, 29 June - 4 July 2014, Honolulu, Hawaii, USA.

### 3. General Conclusions and Prospects

This deliverable provides a hierarchical presentation of the work undertaken within WP11 of NEWCOM# during Y2 of the project. The main results are first listed, then described in a compact way, demonstrating that new research areas and results were produced by joint research, a fact which resulted in a large number of joint publications. More details are provided in appendixes (section 4) allowing to obtain a deeper understanding of the outcomes.

From the reported work, it can be observed that the trend in these fundamental results is to address more and more complex situations, such as the study via large random matrix assumptions of the performance improvement brought by a large number of antennas, study of the average behaviour of a large part of the wireless network via stochastic geometry, etc... a fact which makes difficult the connection with EuWin, (track 2). As a result, the topics which are more easily connected to Track 2 are those involving more focused devices. As an example,, JRA 1.1.3.3 Coding for Multiterminal Communication Systems studies methods that are developed with an overriding concern for low-complexity and suitability for practical implementation, and feasibility studies are underway for demonstration of relaying techniques developed under this JRA using Track 2 facilities.

However, even if this workpackage was initially designed as being the most theoretical one, it should be noted that many results try to take into account practical constraints, which constitutes the first step before being able to study their practical implementation.

## 4. Annex I: Detailed Description of Main technical WP Achievements

### 4.1 ACHIEVEMENTS JRA 1.1.1.1

- Bernhardt S.; Boyer R.; Zhang B.; Marcos S.; Larzabal P.; “Performance Analysis for Sparse Biased Estimator : Application to Line Spectra Analysis.” Invited paper during the IEEE SAM’14 conference.

Dictionary based sparse estimators are based on the matching of continuous parameters of interest to a discretized sampling grid. Generally, the parameters of interest do not lie on this grid and there exists an estimator bias even at high Signal to Noise Ratio (SNR). This is the off-grid problem. In this work, we propose and study analytical expressions of the Bayesian Mean Square Error (BMSE) of dictionary based biased estimators at high SNR. We also show that this class of estimators is efficient and thus reaches the Bayesian Cramér-Rao Bound (BCRB) at high SNR. The proposed results are illustrated in the context of line spectra analysis and several popular sparse estimators are compared to our closed-form expressions of the BMSE.

- Bernhardt S.; Boyer R.; Zhang B.; Marcos S.; Eldar Y.C; Larzabal P.; “Cramer-Rao Bound for Finite Streams of Multiple Filtered Pulses” Published in the proceedings of the EUSIPCO’14

Sampling a finite stream of filtered pulses violates the bandlimited assumption of the Nyquist-Shannon sampling theory. However, recent low rate sampling schemes have shown that these sparse signals can be sampled with perfect reconstruction at their rate of innovation which is smaller than the Nyquist’s rate. To reach this goal in presence of digital noise, an estimation procedure is needed to estimate the time-delay and the amplitudes of each pulse. To assess the quality of any estimator, it is standard to use the Cramér-Rao Bound (CRB) which provides a lower bound on the Mean Squared Error (MSE) of any estimator. In this work, analytic expressions of the Cramer-Rao Bound is proposed for an arbitrary number of filtered pulses. Using the orthogonal properties of the kernels and its first-order derivative, an approximated compact expression of the CRB is provided. The choice of the kernel is discussed from the point of view of estimation accuracy.

### 4.2 ACHIEVEMENTS JRA 1.1.1.2

#### Secure Transmission of Sources Over Noisy Channels

We have investigated the general problem of source-channel coding for secure transmission of sources over noisy channels with side information at the receivers. This setting can be seen as a generalization of the problems of secure source coding with side information at the decoders, and the wiretap channel. A general outer bound on the corresponding achievable region has been derived, as well as two inner bounds based on (i) a pure digital scheme which combines secure source coding with coding for broadcast channels with confidential messages, and (ii) a novel hybrid digital/analog scheme (in the matched-bandwidth case). The proposed bounds do not match in general, but the digital scheme turns out to be optimal under some less noisy conditions. However, a simple counterexample shows that a joint source-channel scheme may achieve better performance in some other cases. At first look, this is not surprising since it is well-known that joint source-channel coding/decoding are well-suited for broadcast channels without secrecy constraints, when all decoders must perfectly reconstruct the source. But the secure setting is rather different because Alice only wants to help one receiver (Bob), while she

wants to blur the other one (Eve). Therefore, the intuition indicates that the optimal strategy would be the opposite i.e., separation between source and channel encoders, as in Propositions 1 and 2. On the other hand, the proposed hybrid digital/analog scheme can be useful in terms of secrecy. In a quadratic Gaussian setup when side information is only present at the eavesdropper, this strategy turns out to be optimal. However, in a more general case where both receivers have side information, a new scheme seems to be needed. For a detailed description, please refer to the published papers:

- Villard, J.; Piantanida, P.; Shamai, S., "Secure Transmission of Sources Over Noisy Channels With Side Information at the Receivers," *Information Theory, IEEE Trans.* on , vol.60, no.1, pp.713,739, Jan. 2014 (Available at: <http://arxiv.org/abs/1201.2315>)

### The Compound Broadcast Channel

In this work, we explored a decoding and a encoding technique for the two-user memoryless Compound Broadcast Channel (BC). We first studied the role of Interference Decoding (ID) where an achievable rate region is derived by using "single per-message description" codes constructed via superposition coding and random binning. At the decoders, the constraint of decoding only the intended message is alleviated to allow each of the users to decode or not the other user's (interference) message. Unlike for the standard two-user BC, this strategy proves to be useful in compound setups, where channel uncertainty prevents the encoder from coding optimally for each possible BC formed by all pairs of channels in the set. A simple outer bound is also derived based on the best outer bound hitherto known on the capacity region of the two-user BC. This outer bound captures one of the most stringent effects of simultaneity of users over the random codes constructed: antagonist coding strategies. Surprisingly enough, ID not only outperforms Non-Interference Decoding (NID) technique, i.e., Marton's worst-case rate region, but also allows to achieve the capacity of a class of non-trivial BC while NID stays strictly suboptimal. Thus, though the coding scheme is simple (in terms of the number of auxiliary variables involved and of the complexity of the encoding operation) the decoders' optimization allows to palliate the uncertainty at the source.

Later, we studied an encoding technique with a more evolved coding strategy, namely Multiple Description (MD) coding. The source transmits to the group of users, interested in the same message, common and private descriptions. For the specific case of the Compound MISO BC, resorting to MD is essential since a common description, i.e., applying DPC with a single description cannot accommodate the interference seen by each instance of the users channels in the set, unless combining it with a time-sharing argument. The key point in the MISO BC setting is that using a fraction of power to transmit the private descriptions is useful for all SNR ranges while turns out to be DoF optimal. Indeed, each private description creates an interference free link and thus each user can recover a part of its rate interference free.

Finally, we addressed the question of whether MD or ID may generalize each other. It appears that none of these schemes can perform well for ordered and non-ordered class of Compound BCs at once, mainly because the two strategies strongly rely on two different interference mitigation techniques: precoding against interference and decoding interference. The first results in a rate loss tantamount to a correlation cost while the latter results in an extra sum-rate constraint.

As a conclusion, it would be worth mentioning the benefits of combining these two schemes to yield a larger inner bound, and thus, full advantage would be taken from the joint optimization of the encoding technique (MD coding) and the decoding technique (ID). For a detailed description, please refer to the published papers:

- M. Benammar, P. Piantanida, and S. Shamai (Shitz), "Multiple Description Coding for the Compound Broadcast Channel," in Information Theory Workshop (ITW), 2014 IEEE, 2-5 November in Hobart, Tasmania, Australia.
- M. Benammar, P. Piantanida, and S. Shamai (Shitz), "On the Compound Broadcast Channel: Multiple Description Coding and Interference Decoding," Submitted to Information Theory, IEEE Trans. on, September 2014, (Available at: <http://arxiv.org/abs/1410.5187>)
- M. Benammar, P. Piantanida, and S. Shamai (Shitz), "Dirty-Paper Coding Techniques for Compound MISO Broadcast Channels: A DoF Analysis," in 9th International Conference on Cognitive Radio Oriented Wireless Networks, Oulu, Finland, June 2014.

### **The Wiretap Channel with State-Feedback**

The MISO wiretap channel with two legitimate receivers and one eavesdropper in presence of delayed channel state information at the transmitter (CSIT) was studied. We derived an inner bound based on an asymptotic achievable scheme and an outer bound showing that  $36/37$  is less than the optimal sum-SDoF which is less than  $16/15$ . The fundamental new issue that arises in this problem is that of securely multicasting a common message to two receivers with delayed CSIT. We believe that a better understanding of this sub-problem can lead to closing the gap between the bounds for the original problem of secure broadcasting with delayed CSIT and is currently under investigation. For a detailed description, please refer to the published paper:

- R. Tandon, P. Piantanida, and S. Shamai (Shitz), "On Multi-User MISO Wiretap Channels with Delayed CSIT," in Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on, 2014

### **Secrecy Capacity Region of Wiretap Broadcast Channels**

In this work, we investigated the secrecy capacity region of the general memoryless two-user Wiretap Broadcast Channel (WBC). We derived a novel outer bound which implies, to the best of our knowledge, all known capacity results in the corresponding setting while by removing secrecy constraints it performs as well as the best-known outer bound for the general Broadcast Channel (BC). An inner bound on the secrecy capacity region of the WBC was also derived by simply using existent encoding techniques based on random binning and stochastic encoders. These bounds allowed us to characterize the secrecy capacity region of several classes of channels, including the deterministic BC with a general eavesdropper, the semi-deterministic BC with a more-noisy eavesdropper and the less-noisy BC with a degraded eavesdropper, as well as some classes of ordered BCs previously studied. Furthermore, the secrecy capacities of the BC with BEC/BSC components and a BSC eavesdropper, as well as the product of two inversely ordered BC with a degraded eavesdropper were also characterized. For a detailed description, please refer to the published papers:

- M. Benammar and P. Piantanida, "Secrecy capacity region of some classes of wiretap broadcast channels," Submitted to Information Theory, IEEE Trans. on, July 2014, (Available at: <http://arxiv.org/abs/1407.5572>).
- M. Benammar and P. Piantanida, "On the secrecy capacity region of the wiretap broadcast channel," in Information Theory Workshop (ITW), 2014 IEEE, November

- M. Benammar and P. Piantanida, "On the secrecy capacity region of a class of parallel wiretap broadcast channel," in 52th Annual Allerton Conference on Communication, Control, and Computing, 2014

#### 4.3 ACHIEVEMENTS JRA 1.1.1.3

##### **Power efficiency in large networks**

Various considerations of large dimensional networks under power control constraints were studied in this subtask. The interest of considering here large networks, aside from obvious practical interest for future use in wireless technologies, lies in the channel hardening effect brought by large dimensional channels. Leveraging this aspect allows for a deterministic hardening of the power management put in place. In particular, it is shown in this achievement that, in a power control context, the powers to allocate to each user terminal converges in the large system limit to a deterministic quantity that can be simply mapped to the channel long term statistics. The solution found takes the form of a fixed-point equation whose solution uniqueness is well understood. In technical terms, the high dependence induced by the power control mechanism leads to unnatural random matrix considerations where the supposedly "deterministic" powers are in fact correlated to the channels in a quite intricate manner, not allowing for standard random matrix results to be used. A main success of this work consisted in relating the model under study to signal processing models of robust estimates of scatter that lead to similar random matrix expressions; since it is shown in these works that the aforementioned performance becomes asymptotically weak, the study simplifies as if the power controls were truly independent of the channels, leading to simplified treatment.

We hereafter detail the major contributions that led to two journal papers and one conference article.

[J1] L. Sanguinetti, A. Moustakas, E. Bjornson, and M. Debbah 'Large System Analysis of the Energy Consumption Distribution in Multi-User MIMO Systems with Mobility', submitted to IEEE Trans. Wireless Commun., May 2014.

This work considers the downlink of a single-cell multi-user MIMO system in which the base station (BS) makes use of  $N$  antennas to communicate with  $K$  single-antenna user equipments (UEs). The UEs move around in the cell according to a random walk mobility model. We aim at determining the energy consumption distribution when different linear precoding techniques are used at the BS to guarantee target rates within a finite time interval  $T$ . The analysis is conducted in the asymptotic regime where  $N$  and  $K$  grow large with fixed ratio. Both recent and standard results from large system analysis are used to provide concise formulae for the asymptotic transmit powers and beamforming vectors for all considered schemes. These results are eventually used to provide a deterministic approximation of the energy consumption and to study its fluctuations around this value in the form of a central limit theorem. Closed-form expressions for the asymptotic means and variances are given. Numerical results are used to validate the accuracy of the theoretical results and to make comparisons. We show how the results can be used to approximate the probability that a battery-powered BS runs out of energy and also to design the cell radius for minimizing the energy consumption per unit area. A typical performance of the Optimal Linear Precoder (OPF) against a mere Zero Forcing (ZF) is demonstrated below.

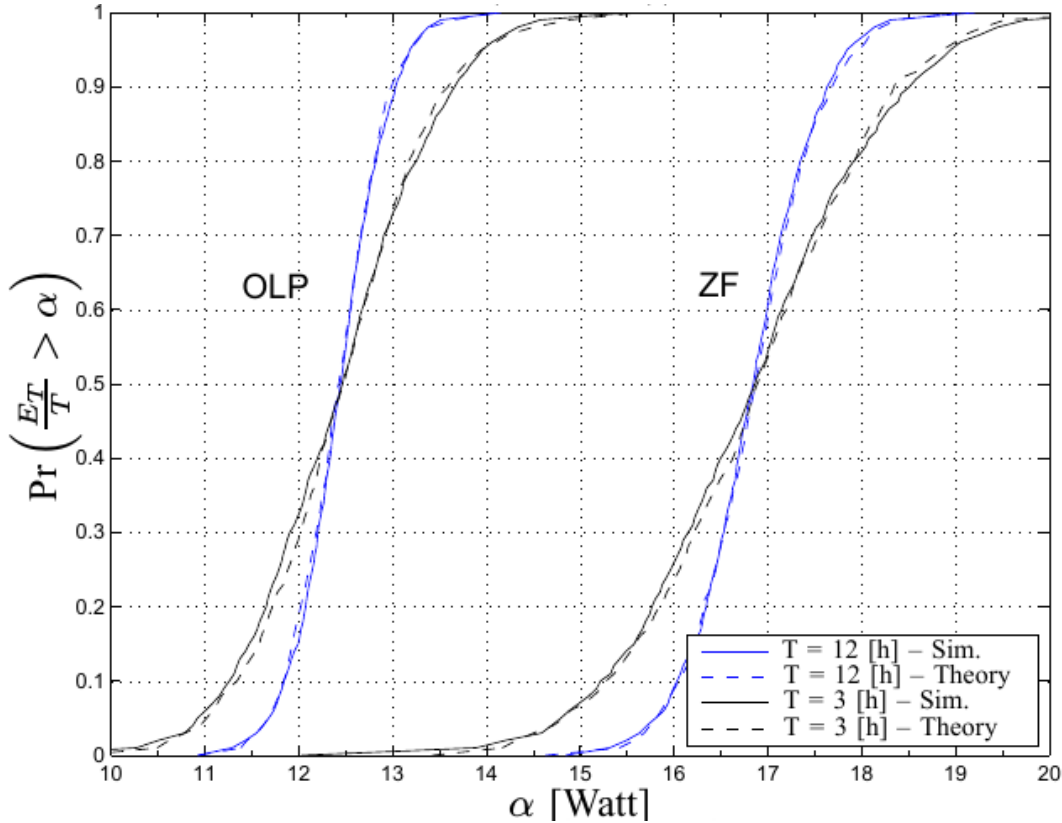


Figure: Outage probability for two different battery requested time intervals T

[J2] L. Sanguinetti, A. Moustakas, and M. Debbah 'Interference Management in 5G Reverse TDD HetNets: A Large System Analysis', submitted to IEEE J. Select. Areas Commun. (Special Issue on HetNets), July 2014.

This work analyzes a heterogeneous network (HetNet), which comprises a macro base station (BS) equipped with a large number of antennas and an overlaid dense tier of small cell access points (SCAs) using a wireless backhaul for data traffic. The static and low mobility user equipments (UEs) are associated with the SCAs while those with medium-to-high mobility are served by the macro BS. A reverse time division duplexing (TDD) protocol is used by the two tiers, which allows the BS to locally estimate both the intra-tier and inter-tier channels. This knowledge is then used at the BS either in the uplink (UL) or in the downlink (DL) to simultaneously serve the macro UEs (MUEs) and to provide the wireless backhaul to SCAs. A concatenated linear precoding technique employing either zero-forcing (ZF) or regularized ZF is used at the BS to simultaneously serve MUEs and SCAs in DL while nulling interference toward those SCAs in UL. We evaluate and characterize the performance of the system through the power consumption of UL and DL transmissions under the assumption that target rates must be satisfied and imperfect channel state information is available for MUEs. The analysis is conducted in the asymptotic regime where the number of BS antennas and the network size (MUEs and SCAs) grow large with fixed ratios. Results from large system analysis are used to provide concise formulae for the asymptotic UL and DL transmit powers and precoding vectors under the above assumptions. Numerical results are used to validate the analysis in different settings and to make comparisons with alternative network architectures.

[C1] H. Sifaou, A. Kammoun, L. Sanguinetti, M. Debbah, M.-S. Alouini 'Power Efficient Low Complexity Precoding for Massive MIMO Systems', submitted to IEEE Global Conference on Signal and Information Processing (GlobalSIP), June 2014.

This work is focused on precoding design in the downlink of a large-scale multiple-input multiple-output (MIMO) system in which the base station (BS) is equipped with  $M$  antennas to serve  $K$  single- antenna user equipments. In particular, we focus on designing the precoding scheme so as to minimize the transmit power while satisfying a set of target signal-to-interference-plus-noise ratio (SINR) constraints. The analysis is conducted in the asymptotic regime in which  $M$  and  $K$  tend to infinity with the same pace. A similar problem has been recently investigated under the assumption that a regularized zero-forcing (RZF) precoding scheme is used at the BS. As is known, in large-scale systems the computational complexity of RZF can be prohibitively high as it involves the computation of a matrix inverse. To address this issue, we adopt a precoding technique based on truncated polynomial expansion (TPE) and makes use of the asymptotic analysis to optimize the TPE weights so as to minimize the asymptotic transmit power while meeting the SNR constraints. Numerical results are used to compare the performance of the proposed solution to that of RZF. It turns out that TPE precoding requires the same amount of power of RZF but with a substantial computational reduction

### Finite Blocklength Error Probability

This achievement concerns finite block-length  $n$  regime of the  $N \times K$  MIMO Rayleigh block-fading channel and specifically the second-order coding rate for this channel. The study is made difficult because the channel does not satisfy the ergodic requirements to apply the usual tools for the analysis of the second-order coding rate. Nonetheless, it is possible to study the asymptotic behavior of the error probability when the coding rate is a perturbation of order  $O(1/\sqrt{nK})$  of the asymptotic capacity while the block-length  $n$ , and the number of transmit and receive antennas  $K$  and  $N$ , respectively, grow infinitely large at the same rate. In this asymptotic regime, new lower and upper bounds on the optimal average error probability are established. These are then applied to obtain a tight closed-form approximation an upper bound for finite  $n$ .

In a first article, an upper bound on the error probability is obtained assuming an average power constraint on the transmitted data. The advantage of this approach is that it lets one consider Gaussian codes, which let the system model more amenable theoretical analysis. The main tools being employed are standard results from fluctuations of functionals of the log-determinant form using random matrix theory, along with classical characteristic function approaches.

[C2] J. Hoydis, R. Couillet, P. Piantanida, M. Debbah, "A Random Matrix Approach to the Finite Blocklength Regime of MIMO Fading Channels", IEEE International Symposium on Information Theory, Boston, Massachusetts, USA, 2012.

In a second time, a lower bound on the error rate performance is considered, this time for amplitude code constraints. This study is made more complex as it involves assuming deterministic codes instead of random ones. The study makes then use of more advanced techniques of the random matrix framework. To this end, new tools were introduced to derive central limit theorems over multiple independent random matrices, along with a trace-normalized (and not spectral norm constrained) signal matrix. The difficulties are alleviated by the use of the Gaussian tools that are the Stein approach and the Poincare-Nash inequality. This work was proposed in:

[C3] J. Hoydis, R. Couillet, P. Piantanida, "Bounds on the Second-Order Coding Rate of the MIMO Rayleigh Block-Fading Channel", IEEE International Symposium on Information Theory, Istanbul, Turkey, 2013.

A joint article on the upper and lower bounds of the second order regime, both for amplitude input constraints was then developed. This article requires different tools for the (new) upper bound than proposed in the conference article [C2]. Since Gaussian inputs can no longer be considered here, the initial approach of [C2] is now invalid. Nonetheless, following the approach considered for the lower bound, the upper bound can be obtained using recent information theoretic considerations of the finite blocklength regime. The upper bound thus obtained is now a tighter approximation to the lower bound than was originally the bound in [C2]. The study was then compared to practical communications in the MIMO setting with usual LDPC codes. It was astonishingly shown that LDPC codes present error rate slopes as a function of the SNR that match closely the anticipated results for optimal (and thus non LDPC) codes. This result is of high interest for practical considerations as, aside from an SNR-gap related to the suboptimality of LDPC codes to reach the Shannon rate, probabilities of error can be fairly well approximated in practice by our theoretical results.

Visual details of our results are given below:

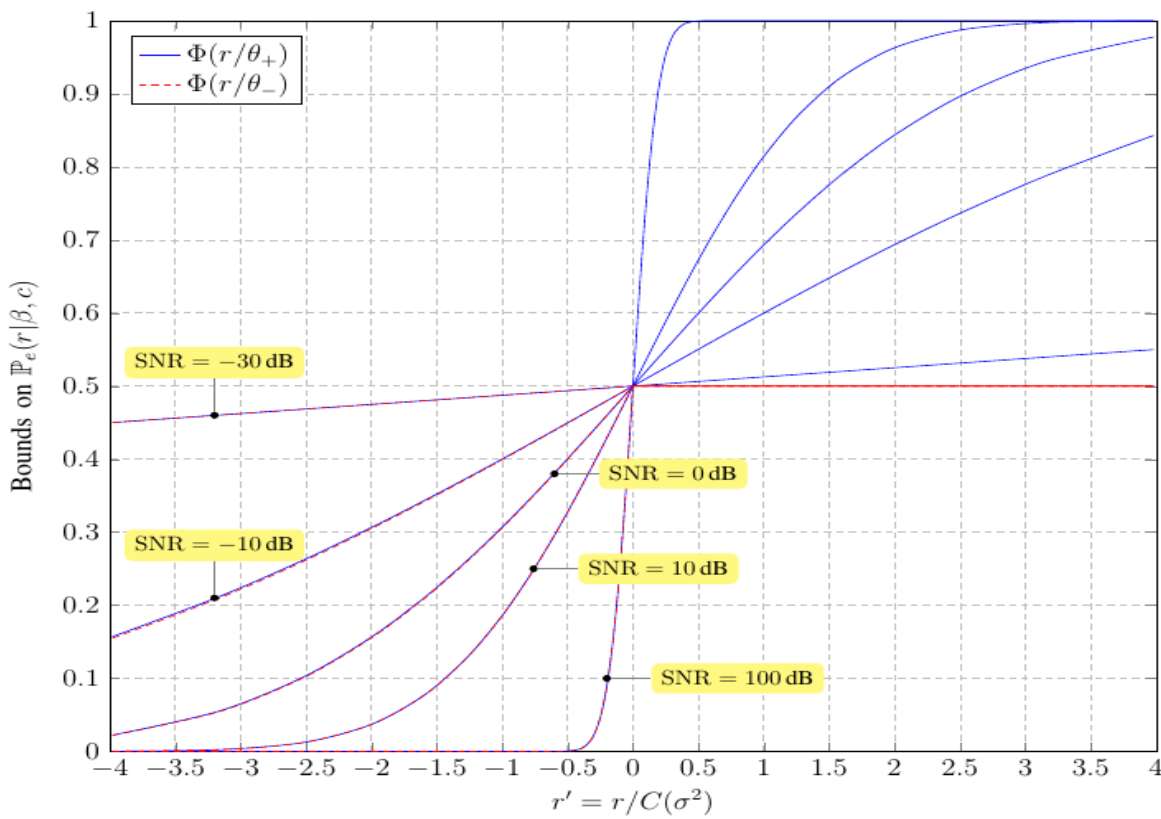


Figure: Upper (blue) and lower (red) bounds for various SNRs as a function of the second order coding rate  $r'=r/C$  with  $C$  the capacity at the given SNR.

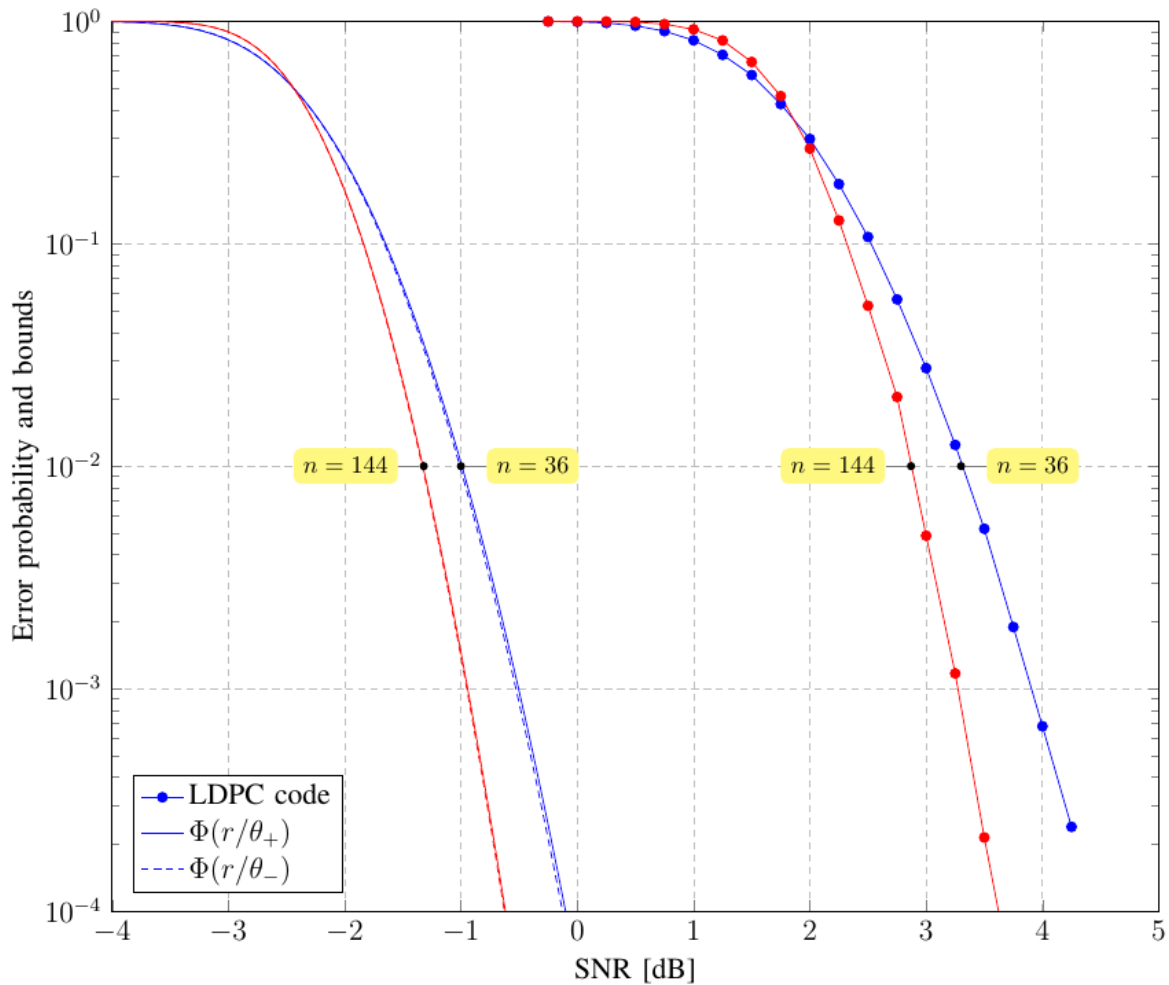


Figure: Performance comparison against LDPC code. Gaussian approximation with lower and upper bounds in plain and dashed lines, against LDPC codes in marked lines.

[J3] J. Hoydis, R. Couillet, P. Piantanida, "The Second-Order Coding Rate of the MIMO Rayleigh Block-Fading Channel", (submitted to) IEEE Transactions on Information Theory.

#### Average Rate of Downlink Heterogeneous Cellular Networks over Generalized Fading Channels – A Stochastic Geometry Approach

[J4] M. Di Renzo, A. Guidotti, and G. E. Corazza, "Average Rate of Downlink Heterogeneous Cellular Networks over Generalized Fading Channels – A Stochastic Geometry Approach", IEEE Transactions on Communications.

It is well-known that tractable and accurate mathematical frameworks for analysis, design, and optimization of cellular networks are difficult to be obtained. The main difficulty originates from the mathematical intractability of modeling other-cell interference, which highly depends on the irregular deployment of real cellular systems, e.g., the spatial positions of the Base Stations (BSs), and on the non-deterministic nature of the mobile channel. This analytical complexity is exacerbated in emerging cellular deployments, where new infrastructure elements, e.g., femto/pico BSs, relays, cognitive radios, distributed antennas, are being deployed making

cellular systems more heterogeneous. The mathematical modeling of cellular networks is usually conducted through abstraction models, which resort to simplified spatial models for the positions of the BSs. Common approaches include the Wyner model, the single-cell interfering model, and the hexagonal grid model. These abstraction models are, however, either inaccurate for many operating conditions or still need intensive numerical computations. Furthermore, they do not scale well for application to heterogeneous cellular deployments. As a result, until recently, analysis and design of these networks were done almost exclusively via simulation over selected scenarios representing specific arrangements of Bss. To circumvent this problem, a new abstraction model is currently emerging and gaining popularity for its analytical tractability and for its flexibility to accommodate various homogeneous and heterogeneous cellular deployments. In this new abstraction model a generic heterogeneous cellular network is modeled as the superposition of many tiers of BSs having different transmit power, density, path-loss exponent, fading parameters and distribution, and possibly unequal biasing for flexible tier association. The positions of the BSs in each tier are modeled as points of an independent Poisson Point Process (PPP). For this reason, this model is today known as PPP-based abstraction. Its popularity originates from the possibility of leveraging powerful tools from stochastic geometry for tractable analytical modeling and performance evaluation. A comprehensive study based on real BSs deployments has confirmed that the PPP-based abstraction model can be used for accurate performance analysis in major cities worldwide [34]. Recent results about the validation of the PPP-based abstraction model for real BSs deployments are available, where data collected from Ofcom, i.e., the independent regulator and competition authority in the UK, is used.

By exploiting the PPP-based abstraction model, we propose a new mathematical methodology to compute the downlink average rate over general fading channels. Recent papers have developed frameworks to compute the average rate for single-tier downlink, multi-tier downlink, downlink multi-cell coordination, and single-tier uplink cellular networks. All these works use the same two-step methodological approach to compute the average rate and exploits the Plancherel-Parseval theorem: i) first, the Coverage Probability ( $P_{cov}$ ) is computed; and ii) then, the average rate is obtained by integrating  $P_{cov}$  over the positive real axis. This methodology is denoted by  $P_{cov}$ -based approach. Although this technique avoids systemlevel simulations, it requires, for general fading channels, the computation of a four-fold integral. For this reason, many authors often limit the analysis to Rayleigh fading channels and/or to interference-limited networks, where simplified frameworks can be obtained.

To overcome this limitation, we proposed in this achievement a new analytical framework which, at the same time, reduces the number of integrals to be computed, and, similar to the  $P_{cov}$ -based approach, is flexible enough for application to arbitrary fading distributions (including correlated composite channel models). The framework leverages the application of recent results on the computation of the ergodic capacity in the presence of interference and noise. It avoids the computation of  $P_{cov}$ , and needs only the Moment Generating Function (MGF) of the aggregate interference at the probe mobile terminal. This framework is denoted by MGF-based approach. We show that it is applicable to multi-tier cellular networks with long-term averaged maximum biased-received-power tier association, and that either a single or a two-fold numerical integral need to be computed for arbitrary fading channels.

Visual results are presented below:

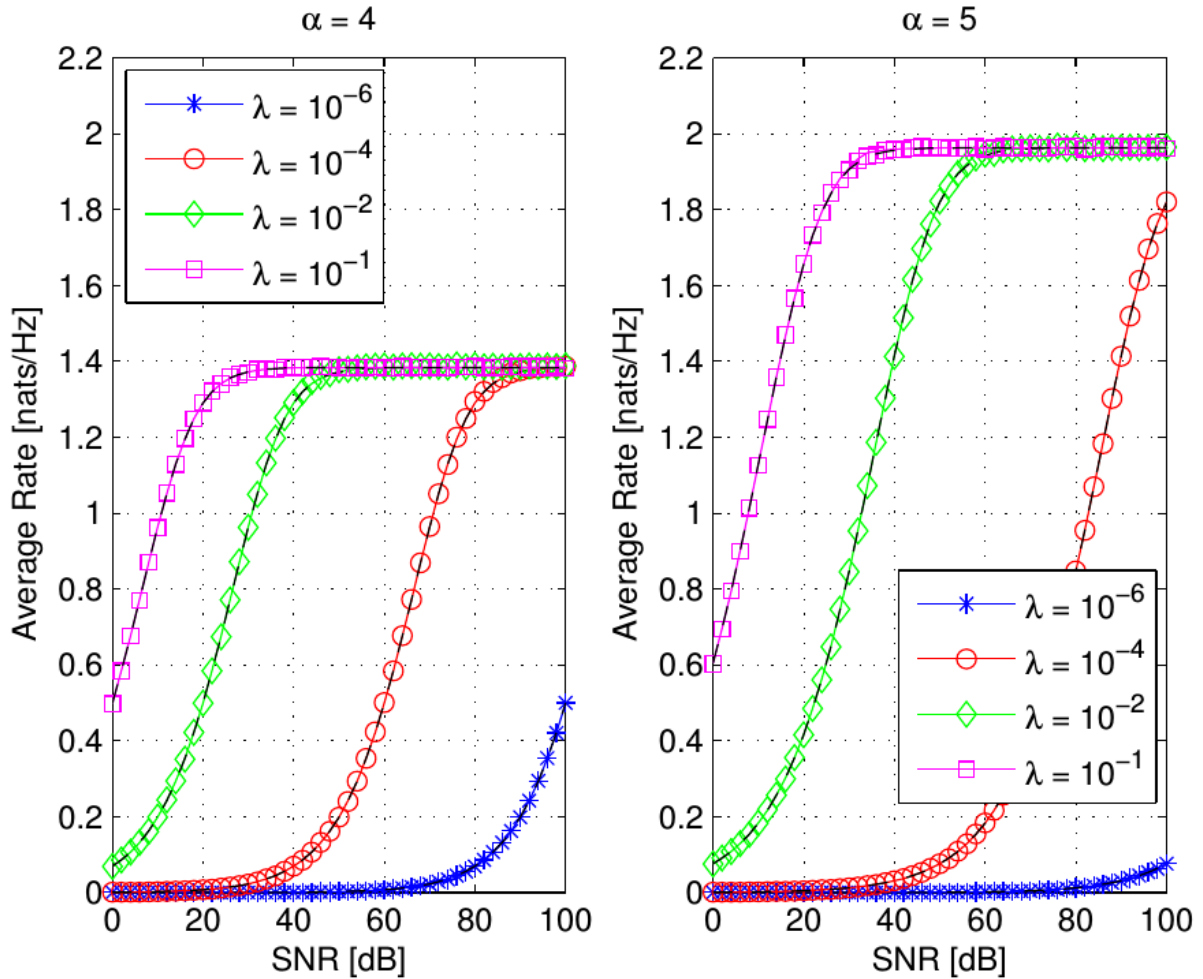


Figure: Average rate of a single-tier cellular network over composite Nakagami-m and Log-Normal fading. Markers show Monte Carlo simulations. Solid lines show the analytical framework, which is computed by using the main theorem of the article [J4].

#### 4.4 ACHIEVEMENTS JRA 1.1.2.1

##### Turbo-Coded Two-Way Relay Channel

For this work item, we extended our recent work on joint decoding of trellis codes for the two-way relay channel to quaternary decoding of turbo codes and evaluate two approaches for rate adaptation. More specifically, we consider the uplink phase, which has been identified as the bottleneck, and apply a quaternary joint turbo decoder for both packets. For asymmetric channels, we evaluate two methods of adapting the code rate while keeping the standard LTE turbo codes. The first approach applies puncturing which is known in LTE as rate matching and adapts the codeword lengths while the second, less well-known, method reduces the message length and is known as expurgation.

For a detailed description, please refer to the published paper

- Stephan Pfletschinger, Carmine Vitiello, Monica Navarro, "[Decoding Options for the Symmetric and Asymmetric Turbo-Coded Two-Way Relay Channel](#)", invited paper at

---

*European Conference on Networks and Communications (EuCNC), Bologna, 23-26 June 2014.*

### **Physical-Layer Network Coding with Non-Binary LDPC Codes**

We applied non-binary coding to the two-way relay channel and evaluated two decoding schemes for the multiple-access phase at the relay: joint decoding for both users and separate decoding for a linear combination of both packets. We evaluated several efficient modulation schemes which combine favourably with non-binary coding and found that joint decoding can offer significant performance benefits.

For a detailed description, please refer to the published paper

- Stephan Pfletschinger, Dirk Wübben, Giacomo Bacci, "[Physical-Layer Network Coding with Non-Binary Channel Codes](#)", *invited paper at 31th URSI General Assembly and Scientific Symposium*, Beijing, China, 16-23 August 2014.

### **Joint Decoding for Uncoordinated Multiple Access**

For the application in uncoordinated multiple-access protocols, we presented a novel decoding scheme for slotted ALOHA which is based on concepts from physical-layer network coding (PNC) and multi-user detection (MUD). In addition to recovering individual user packets from a packet collision as it is usually done with MUD, the receiver applies PNC to decode packet combinations that can be used to retrieve the original packets using information available from other slots. We evaluate the novel scheme and compare it with another scheme based on PNC that has been proposed recently and show that both attain important gains compared to basic successive interference cancellation. This suggests that combining PNC and MUD can lead to significant gains with respect to previously proposed methods on either one or the other.

For a detailed description, please refer to the published paper

- Stephan Pfletschinger, Monica Navarro, Giuseppe Cocco, "[Interference Cancellation and Joint Decoding for Collision Resolution in Slotted ALOHA](#)", *International Symposium on Network Coding (NetCod)*, 27-28 June 2014.

## **4.5 ACHIEVEMENTS JRA 1.1.2.2**

### **Power Efficient Resource Allocation in a Heterogeneous Network with Cellular and D2D Capabilities**

This work focuses on a cellular scenario where the mobiles have the capability to act as APs for other mobiles and relay their traffic to the cellular infrastructure. For that purpose, Device To Device (D2D) communications in which mobile terminals are able to directly communicate can be envisaged. Initiatives in this direction are carried out by the 3rd Generation Partnership Project (3GPP) in Long Term Evolution (LTE) Release 12 [2] and by the Wi-Fi Alliance that has recently developed the Wi-Fi Direct technology [3].

In such scenario, there are several challenges that need to be addressed in order to achieve an efficient provision of the mobile broadband services. In particular, given the randomness associated to the propagation in mobile environments, as well as the variability in the generation of data traffic, there will be situations in which it may be more efficient for a certain mobile terminal to connect to one or another terminal acting as AP or to connect directly to the infrastructure, leading to a dynamic network architecture in which the terminals can dynamically

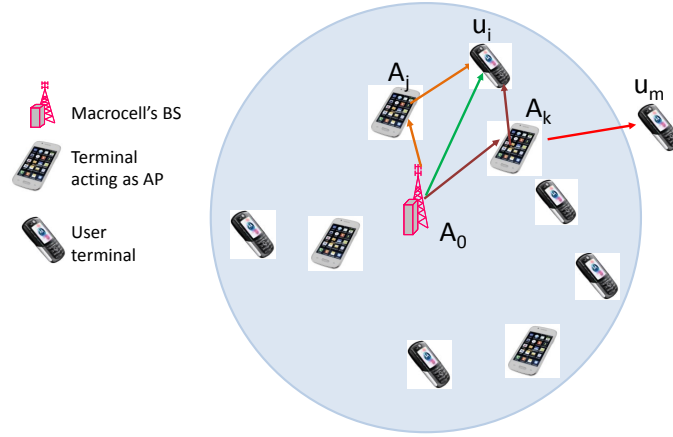
change the way they connect to the infrastructure. Consequently, it becomes crucial to have smart decision mechanisms to determine the best connection for each mobile terminal. Such decision needs to consider different elements such as the propagation conditions of the different links, the load existing in the macrocell and in each AP, the bit rate requirements of each terminal or the total power consumption. In this context, this work considers as a main objective the minimization of the total transmit power in the scenario, targeting efficient solutions from the perspective of energy consumption.

### System model and problem formulation

The scenario considered in this work is represented in Figure 4-1. It assumes a cellular network where each terminal (e.g. current smartphones) can be turned into an access point (AP) and can be used to provide wireless connectivity to other terminals. There are  $K$  terminals acting as APs denoted as the set  $\Lambda=\{A_1,...,A_K\}$ , one Base Station (BS) of a macrocell  $A_0$  with a cellular technology (e.g. LTE or LTE-A) and  $N$  user terminals denoted as  $U=\{u_1,...,u_N\}$ . User terminal  $u_i$  needs to have wireless connectivity to receive a data service with a certain bit rate requirement  $R_i$ . For that purpose, it can connect either directly to the BS of the macrocell  $A_0$  or to one of the APs in the set  $\Lambda$ . A smart selection of the AP or the macrocell would impact on the total resource consumption in the scenario (e.g. required transmitted power). This work assumes communication in the downlink direction, i.e. from the BS/APs to the terminals, although it could easily be extended to consider also the uplink direction.

Each AP has the capability to provide wireless Internet access to other terminals. In general, different possibilities exist for this access. This work assumes that a terminal acting as AP is also connected to the cellular infrastructure, so that the AP relays the data traffic of the macrocell to the terminal using a 2-hop approach. In the example of Figure 4-1, terminal  $u_i$  has three possibilities for getting service: direct connection to  $A_0$ , 2-hop connection  $A_0 \rightarrow A_k \rightarrow u_i$  and 2-hop connection  $A_0 \rightarrow A_j \rightarrow u_i$ . In contrast, the terminal  $u_m$ , out of the coverage area of the macrocell, only has one possibility  $A_0 \rightarrow A_k \rightarrow u_m$ . It is worth mentioning that this work could be extended to consider also multiple hops in which a terminal acting as AP would relay the traffic to other APs in order to finally reach the macrocell. However, this is left for future work. This work could also be easily extended to consider other possibilities for providing access through the APs, for example in the case that the APs are fixed and have wired connection to the Internet, so that in this case no relaying would be needed.

As for the terminals acting as APs, it is considered that, in addition to relaying the traffic from other terminals, they may have also their own service requirements. In particular, the service requirement of the  $k$ -th terminal acting as AP will be expressed in terms of a required bit rate  $R_{A,k}$ .



**Figure 4-1: Considered scenario: Heterogeneous network with D2D capabilities**

The communication between the terminals and the terminals acting as APs makes use of a device-to-device (D2D) technology<sup>1</sup>. It is assumed that all the APs operate in different frequencies so they do not interfere mutually, otherwise a careful frequency assignment would be needed, which is out of the scope of this work. The D2D link between the AP and the terminals is assumed to have a bandwidth  $B_A$  that is shared on the time domain between the users connected to the AP, e.g. by means of a scheduling algorithm or a Medium Access Control (MAC) protocol. It is also assumed that the terminals acting as APs can use simultaneously the D2D interface and the cellular interface to connect with the other terminals and with the infrastructure, respectively. Correspondingly, when the APs relay data from the infrastructure to other terminals, full duplex relay is assumed.

The macrocell is assumed to have  $M$  available Resource Blocks (RBs) of bandwidth  $B$  that can be assigned to the  $N$  terminals (if directly connected to the macrocell) or to the  $K$  access points. The maximum power available at the BS of the macrocell is  $P_{max}$ . It is assumed that the maximum power is split equally between the different RBs. Then, the power per resource block is given by  $p_{RB} = P_{max}/M$ .

$r_{T,n}$  is the capacity that the terminal  $u_n \in U$  can obtain when allocated to one RB of the macrocell, and it will depend on the propagation conditions and the transmit power. Assuming Shannon bound it can be computed as  $r_{T,n} = B \log_2 \left( 1 + p_{RB} / (L_{T,n} P_N) \right)$ , where  $L_{T,n}$  is the propagation loss between terminal  $u_n$  and the macrocell and  $P_N$  is the noise power measured over the bandwidth  $B$ , i.e.  $P_N = N_o \cdot B$  and  $N_o$  is the noise power spectral density. We assume no interference from other cells exists in the scenario. In any case, the above term could easily be extended to include inter-cell interference if multiple macrocells were considered.

Similarly,  $r_{A,k}$  is the capacity that the access point  $A_k \in \Lambda$  can obtain in one RB when connected to the macrocell. It is given by  $r_{A,k} = B \log_2 \left( 1 + p_{RB} / (L_{A,k} P_N) \right)$  where  $L_{A,k}$  is the propagation loss between the AP  $A_k$  and the macrocell.

It is worth mentioning that this work assumes that both  $L_{T,n}$  and  $L_{A,k}$  are average propagation losses, i.e. the fast fading effects are averaged. As a result, the propagation loss for a given user terminal or AP is the same in all RBs of the macrocell. Correspondingly, we are only concerned with the average number of RBs that will be required by each user terminal or AP to get their

<sup>1</sup> We assume the work at this stage to be technology-agnostic. In practice, there could be different possibilities such as IEEE 802.11, LTE D2D, etc.

average bit rate requirements, but not on the specific scheduling process that will decide which RB is allocated to each terminal/AP in the short term.

Regarding the D2D link between APs and terminals, the achievable capacity by the  $n$ -th terminal when connected to the access point  $A_k$ ,  $k=1, \dots, K$ , will be given by  $r_{k,n} = B_A \log_2 \left( 1 + p_A / (L_{k,n} P_{N,A}) \right)$  where  $L_{k,n}$  is the average propagation loss between the terminal  $u_n$  and the access point  $A_k$ ,  $P_{N,A} = N_o B_A$  is the noise power at the terminal and  $p_A$  is the transmit power of the access point, assumed to be constant and equal for all the access points as it usually happens with Wi-Fi links.

Moreover, it is assumed that a control mechanism exists at the AP so that each terminal only receives at most its bit rate requirement  $R_n$ . Correspondingly, if the achievable bit rate  $r_{k,n}$  is higher than the requirement  $R_n$ , the AP will only be transmitting data for the  $n$ -th user during a fraction of time  $\theta_{k,n}$  given by:

$$\theta_{k,n} = \min \left( 1, \frac{R_n}{r_{k,n}} \right) \quad (1)$$

In this way, the total average power transmitted by access point  $A_k$  to provide the service to the  $n$ -th terminal would be  $p_A \cdot \theta_{k,n}$ , so that the AP would spend only the minimum power needed to provide the terminal with its bit rate requirement. It is worth mentioning that in the case  $r_{k,n} < R_n$ , the  $n$ -th terminal would not be able to get its required bit rate through the access point  $A_k$ .

Based on the above, and considering that a given access point  $A_k$  will provide service to a number of users, the total fraction of time that the access point will be active is denoted as  $\Theta_k$ , given by:

$$\Theta_k = \sum_{n=1}^N b_{k,n} \theta_{k,n} \quad (2)$$

where  $b_{k,n}$  is a binary indicator that takes the value 1 if the  $n$ -th user terminal is connected to the access point  $A_k$  and 0 otherwise.

Clearly, it should be fulfilled that  $\Theta_k \leq 1$  in order that all the users connected to the access point  $A_k$  are able to get their bit rate requirements.

Thanks to the use of the relaying capabilities at the APs, it is expected that a more efficient resource usage can be achieved and that the total transmit power in the scenario will be reduced in comparison with the case when the terminals are directly connected to the macrocell. In particular, terminals with very high bit rate requirements located at the edge of the macrocell would require a large amount of RBs and, correspondingly, a large total power if connected directly to the macrocell. On the contrary, if connected through another AP with better propagation conditions to the macrocell, this may lead to less RBs/power allocated in the macrocell for the same bit rate requirement, at the expense of some additional power transmitted in the link with the AP. Clearly, a trade-off will exist between the usage of resources in the macrocell and the usage of resources in the D2D links, which leads to an optimization problem to identify the most suitable way to assign resources.

Specifically, the optimization problem considered in this work intends to determine the optimum assignment of terminals to APs or to the macrocell such that the total average transmitted power in the scenario is minimized. The total average transmitted power is given by:

$$P_{TOT} = M_{req} \cdot p_{RB} + \sum_{k=1}^K \Theta_k \cdot p_A \quad (3)$$

The first term is the total power transmitted by the BS of the macrocell, expressed in terms of the average number of required RBs,  $M_{req}$ , while the second term is the total power transmitted by all APs.

The average number of RBs required by the macrocell to serve its users can be expressed as follows:

$$M_{req} = \sum_{n=1}^N b_{0,n} \cdot \frac{R_n}{r_{T,n}} + \sum_{k=1}^K \frac{R_{A,k} + \sum_{n=1}^N b_{k,n} \cdot R_n}{r_{A,k}} \quad (4)$$

where  $b_{0,n}$  is a binary indicator that takes value 1 if the  $n$ -th terminal is directly connected to the macrocell and 0 otherwise. Then, the first term in (4) corresponds to the average number of RBs required by the terminals that are connected directly to the macrocell to achieve their bit rate requirements  $R_n$ , while the second term corresponds to the RBs required by the APs to achieve their own requirements plus those of the terminals that they are serving.

Therefore, the considered optimisation problem can be formulated as follows:

$$\min_{b_{k,n}} P_{TOT} = \min_{b_{k,n}} \left( M_{req}(b_{k,n}) \cdot p_{RB} + \sum_{k=1}^K \Theta_k(b_{k,n}) \cdot p_A \right) \quad (5)$$

subject to the following constraints:

$$\sum_{k=0}^K b_{k,n} \leq 1 \quad n=1, \dots, N \quad (6)$$

$$M_{req}(b_{k,n}) \leq M \quad (7)$$

$$r_{k,n} \geq b_{k,n} R_n \quad n=1, \dots, N, \quad k=1, \dots, K \quad (8)$$

$$\Theta_k = \sum_{n=1}^N b_{k,n} \theta_{k,n} \leq 1 \quad k=1, \dots, K \quad (9)$$

Constraint (6) reflects the fact that a terminal  $n$  can only be connected to one AP or to the macrocell, so at most one of the values of  $b_{k,n}$  should equal 1. Note that all the values  $b_{k,n}$  could be 0 if there is no possibility of connection for the  $n$ -th terminal. In turn, the total number of RBs required by the macrocell should be below the number of available RBs,  $M$ , as represented in constraint (7). Moreover, constraint (8) reflects that any user  $n$  can only be connected to an AP  $k$  in which the available bit rate in the D2D link,  $r_{k,n}$ , is higher or equal than the required bit rate  $R_n$ . Finally, the bandwidth sharing of all the users connected to access point  $k$  should allow that all of them are able to receive their required bit rate as stated in constraint (9).

It is worth mentioning that the fulfilment of constraints (7)-(9) ensures that all the users and APs are served with their required bit rates  $R_n$  and  $R_{A,k}$ , respectively. However, note that, depending on the number of available RBs  $M$  and the specific propagation conditions, it is possible that no solution exists that fulfils all the considered conditions. In such a case, either some users should not be admitted in the system or their achieved bit rate would be below the minimum requirements.

The considered optimization problem has focused on the association between terminals and access points assuming constant transmit powers per RB in the macrocell and the transmit power of the access point. It would also be possible to extend it to consider the optimization of the transmit power levels in each RB and in the access points, and to consider also the short term scheduling of users/APs to the RBs, but this is left for future work.

### Solution for Distributed AP Selection

The problem presented in previous section could be solved by different methods (e.g. branch and bound, genetic algorithms, etc.). However, these solutions would require considering simultaneously all the terminals and APs at a given time and perform the optimization in a centralized way. This would in general lead to high complexity as the number of terminals/APs increases. In order to overcome this limitation, in the following a distributed approach will be envisaged, in which the different terminals decide autonomously the access point (or macrocell) where they will be connected. The main advantage of using distributed approaches is that they allow for a reduction in complexity and signaling overhead. More specifically, the proposed distributed approach will be based on Q-learning [4], where each terminal  $n$  keeps a record of the experience when using the access point  $k=1, \dots, K$  or the macrocell  $k=0$ , stored in a value  $Q_n(k)$ .

Whenever an access point  $k$  or the macrocell has been used by a terminal  $n$ , the new value of  $Q_n(k)$  will be updated following a single state Q-learning approach with null discount rate given by:

$$Q_n(k) \leftarrow (1 - \alpha)Q_n(k) + \alpha W_n(k) \quad (10)$$

where  $\alpha \in (0,1)$  is the learning rate and  $W_n(k)$  is the reward resulting from the use of the access point  $k$ . The reward  $W_n$  should reflect the degree of fulfillment of the optimization target as well as the different constraints. In that respect, if we consider that the target to minimize is the total power transmitted by the macrocell and the access point to serve the terminal bit rate requirements  $R_n$ , the reward will be based on this total power and on the achieved bit rate  $\hat{r}_{k,n}$  during the connection. In this way, the idea is that those access points that lead to lower power consumption levels provide larger rewards, and correspondingly larger values of  $Q_n(k)$ . The computation of the reward is detailed in the following.

### Reward computation

The computation of the reward makes the following distinction depending on whether the terminal is connected through an access point or through the macrocell's BS.

#### 1) Reward for the case when the terminal is connected through an access point ( $k=1, \dots, K$ )

In this case the total power results from two contributions:

- $p_A \cdot \theta_{k,n}$  is the total power of the access point  $k$  devoted to serve the  $n$ -th terminal. This can be computed by the terminal from (1) and the value of  $r_{k,n}$  obtained from the Shannon bound by making use of e.g. the path loss between the terminal and the access point, or by measuring the signal to noise ratio in the link. Moreover, the value of  $p_A$  could be broadcast by the access point.
- $P_{k,n}$  is the power of the macrocell devoted to deliver the traffic of the  $n$ -th terminal through the link between the macrocell and the access point  $k$ . It is given by:

$$P_{k,n} = M_{k,n} \cdot P_{RB} \quad (11)$$

where  $M_{k,n}$  is the average number of RBs in the macrocell required by the terminal when connected through the access point  $A_k$ , given by:

$$M_{k,n} = \frac{R_n}{r_{A,k}} \quad (12)$$

Based on the above discussion, the reward function when the terminal is connected to an access point  $k=1, \dots, K$  is defined as:

$$W_n(k) = \begin{cases} 0 & \text{if } \hat{r}_{k,n} < R_n \\ 1 - \frac{p_A \cdot \theta_{k,n} + P_{k,n}}{p_A + P_{\max}} & \text{otherwise} \end{cases} \quad (13)$$

Note that this expression assigns a value of 0 whenever the achieved bit rate (i.e. measured bit rate) during the connection  $\hat{r}_{k,n}$  is below the requirement  $R_n$ . On the contrary, if the service requirement has been successfully fulfilled, the reward is a value between 0 and 1 that decreases when the required power consumption increases.

It is worth mentioning that the condition  $\hat{r}_{k,n} < R_n$  reflects that the terminal is not getting its required bit rate. This could occur because of three different reasons:

- The propagation conditions in the D2D link do not allow achieving  $R_n$ . When this occurs, constraint (8) wouldn't be fulfilled.
- There is an excessive load in the access point  $A_k$ . When this occurs, constraint (9) wouldn't be fulfilled.
- The lack of resource blocks at the macrocell to provide the service through the access point  $A_k$ . When this occurs, constraint (7) wouldn't be fulfilled.

Consequently, the formulation of the reward function takes into account in a natural way the constraints of the optimization problem.

## 2) Reward for the case when the terminal is connected directly through the macrocell ( $k=0$ )

In this case the total power consumption is in the macrocell. By making similar considerations as in the previous case, the transmitted power from the macrocell's BS devoted to the  $n$ -th terminal is now given as:

$$P_{0,n} = M_{0,n} \cdot P_{RB} \quad (14)$$

where  $M_{0,n}$  is the number of resource blocks that the macrocell would need to serve the requirements of the  $n$ -th terminal, and is given by:

$$M_{0,n} = \frac{R_n}{r_{T,n}} \quad (15)$$

Based on this, the reward function when the terminal is connected to the macrocell  $k=0$  is defined as:

$$W_n(k) = \begin{cases} 0 & \text{if } \hat{r}_{k,n} < R_n \\ 1 - \frac{P_{0,n}}{P_{\max}} & \text{otherwise} \end{cases} \quad (16)$$

### Computation of the $Q_n(k)$ value at initialization

At initialization, i.e. when the access point  $k$  has not been used before by terminal  $n$ , the value of  $Q_n(k)$  can be computed making use of similar expressions as the reward, by replacing the first condition (because there is no measured value of  $\hat{r}_{k,n}$ ) as explained in the following.

For the case  $k=1, \dots, K$  the initial value of  $Q_n(k)$  is defined as:

$$Q_{n,initial}(k) = \begin{cases} 0 & \text{if } (\theta_{k,n} > 1) \text{ OR } (M_{k,n} > M) \\ 1 - \frac{p_A \cdot \theta_{k,n} + P_{k,n}}{p_A + P_{\max}} & \text{otherwise} \end{cases} \quad (17)$$

The first condition in (17) reflects the case that the access point  $k$  is not appropriate to serve the terminal  $n$  because the propagation conditions in the link between the AP and the terminal are not able to provide the service requirements (i.e.  $\theta_{k,n} > 1$ ) or because the link between the AP and the macrocell would require more resource blocks than those available to provide the service (i.e.  $M_{k,n} > M$ ).

Similarly, for the case  $k=0$ , the initial value of  $Q_n(k)$  is given by:

$$Q_{n,initial}(k) = \begin{cases} 0 & \text{if } (M_{0,n} > M) \\ 1 - \frac{P_{0,n}}{P_{\max}} & \text{otherwise} \end{cases} \quad (18)$$

### Selection criterion

At the time when the  $n$ -th terminal aims to establish a new connection, it uses the available values of  $Q_n(k)$  to choose the access point to be connected by applying a softmax selection policy [4], in which the access point  $k=1, \dots, K$  or the macrocell  $k=0$  is randomly selected by terminal  $n$  with probability:

$$\Pr(k, n) = \frac{e^{\frac{Q_n(k)}{\tau}}}{\sum_k e^{\frac{Q_n(k)}{\tau}}} \quad (19)$$

where  $\tau$  is the so-called temperature parameter. High temperature causes the different options to be all nearly equiprobable. On the contrary, low temperature leads to a greater difference in selection probability for access points that differ in their  $Q_n(k)$  value estimates, and the higher the value of  $Q_n(k)$  the higher the probability of selecting access point  $k$ . Softmax decision making is a popular means of balancing the exploitation and exploration dilemma in reinforcement learning-based schemes. It exploits what the terminal already knows in order to obtain reward (i.e. selecting access points that have provided good results in the past), but it also explores in order to make better actions in the future (i.e. the selection must try first a variety of combinations and progressively favor those that appear to be the best ones) [4].

In order to facilitate the algorithm convergence as the best actions are being identified by the algorithm, a cooling function will be also considered here to reduce the value of the temperature  $\tau$  as time evolves. Specifically, the following logarithmic cooling function will be considered:

$$\tau = \frac{\tau_0}{\log_2(1+t)} \quad (20)$$

where  $\tau_0$  is the initial temperature and  $t$  is the time elapsed since the terminal makes the first AP selection.

### Admission control

Given that the access point selection is carried out by the terminal, it may occur that the load in the selected access point is already too high to support the new user. Consequently, it will be assumed that an admission control will be used at the selected access point to ensure that the number of connected users is sufficiently low to ensure that their required bit rates can be provided. This is in fact the condition captured by constraint (9) of the optimization problem. Then, if the resulting value of  $\Theta_k$  after including the new user is higher than 1 (or in general than a certain threshold) the new user is not admitted in this access point. When this happens, the reward for the selected access point will be set to 0 and another access point will be selected.

### Evaluation

The scenario considered for the evaluation of the proposed approach assumes a square area of 400 m x 400 m, subdivided into pixels of size 10m. The macrocell's BS is located at the upper left corner as shown in Figure 4-2. The following general propagation model is assumed for computing the propagation losses between the terminals/APs and the BS (i.e.  $L_{T,n}$ ,  $L_{A,n}$ ) and between the terminals and the APs (i.e.  $L_{k,n}$ ):

$$L(dB) = K_p + \beta_p \log f(GHz) + \alpha_p \log d(km) + S \quad (21)$$

Based on [5] and references therein, the considered parameters in (21) are  $K_p=122.1$  dB,  $\beta_p=21$ ,  $\alpha_p=37.6$ . Moreover,  $f=2.6$  GHz has been considered for the propagation loss between the macrocell and the terminals or APs, and  $f=2.4$  GHz for the propagation loss between the APs and the terminals.  $S$ (dB) is the shadowing, which follows a Gaussian distribution with mean 0 and standard deviation  $\sigma=6$  dB. Two dimensional spatially correlated shadowing is considered with exponential autocorrelation and decorrelation distance  $d_{corr}=10$  m. The shadowing of the links BS-AP and AP-terminal are assumed to be independent.

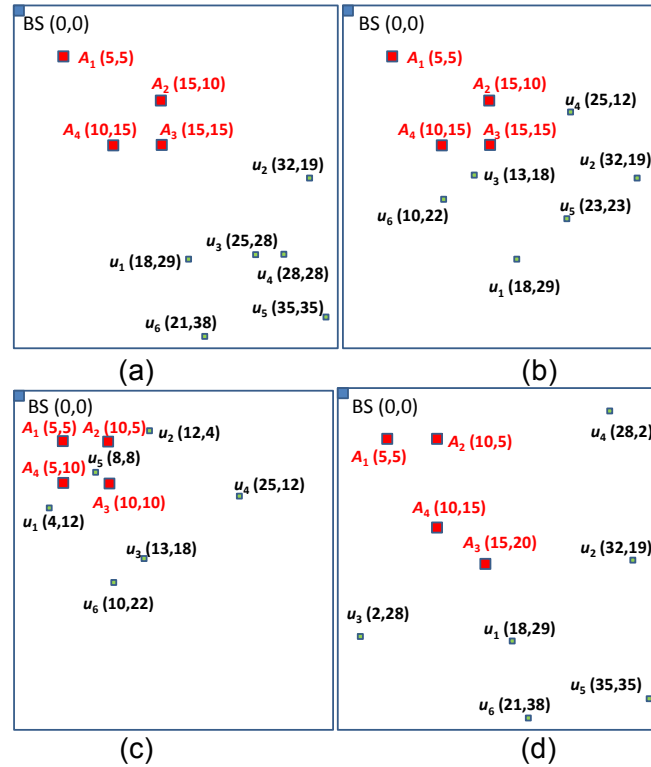
The macrocell has  $M=25$  RBs of bandwidth  $B=180$  kHz. The transmit power per RB is  $p_{RB}=29$  dBm. As for the access points, they have bandwidth  $B_A=20$  MHz, and transmit power  $p_A=20$  dBm. The noise power spectral density is  $N_0=-164$  dBm/Hz, and the required bit rates by the different terminals and APs are  $R_n=5$  Mb/s and  $R_{A,k}=5$  Mb/s. The terminals and the terminals acting as APs generate sessions whose duration is geometrically distributed with average 30 time steps. The time between the end of one session and the beginning of next one is also geometrically distributed with average 30 time steps. Each time that a new session starts, the terminal performs the AP selection process explained in Section **Errore. L'origine riferimento non è stata trovata.**

Each simulation experiment is run for a total of 10000 time steps. The Q-learning algorithm has been configured with learning rate  $\alpha=0.1$ , while different values of the temperature parameter  $\tau$  will be analyzed, including the logarithmic cooling case given by (20).

### Performance in terms of transmit power consumption

To gain a first insight into the behavior of the proposed strategy, initially a set of  $K=4$  access points and  $N=6$  terminals are considered, located in different positions of the scenario. In these configurations, it is easy to compute the optimum solution to the problem (e.g. by testing all the

possible combinations), and thus it is possible to analyze the capability of the proposed algorithm to converge to this optimum solution. Four different scenarios with different locations of the APs and terminals are considered, as depicted in Figure 4-2. In scenarios A and B, the APs are located at the same positions while terminals are far from the macrocell's BS in scenario A and at a closer distance in scenario B. In turn, in scenario C both terminals and APs are located close to the BS while in scenario D there are APs at close and medium distances from the BS and terminals are far from the BS. The shadowing term for the propagation between the macrocell and the different positions is the same in all the four scenarios. Moreover, in order to get the global performance for different positions of the users, scenario E is also considered that corresponds to 200 runs of experiments with different uniform random distributions of the 6 users and the same AP positions as in scenarios A and B.



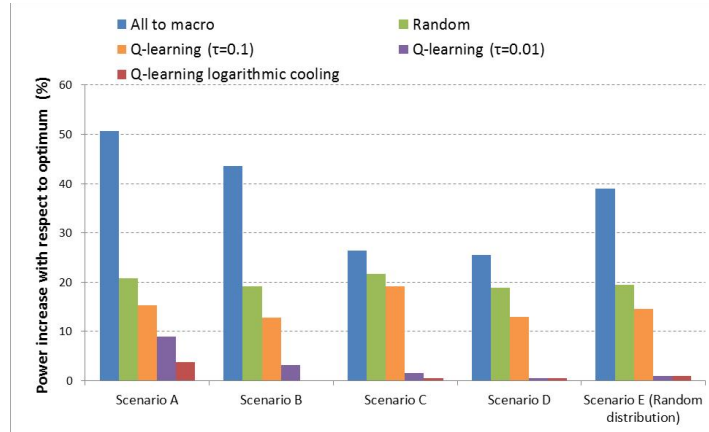
**Figure 4-2: Terminals and AP distributions in the considered scenarios. (a) Scenario A, (b) Scenario B, (c) Scenario C, (d) Scenario D. Coordinates are measured in pixels.**

As a relevant performance metric of the behavior of the proposed algorithm, Figure 4-3 presents the total transmit power increase with respect to the optimum solution for different settings of the temperature parameter in the different scenarios. In particular, two fixed values  $\tau=0.01$  and  $\tau=0.1$  are compared against the logarithmic cooling function given by (20) with  $\tau_0=0.1$ . As mentioned, for these scenarios with a reduced number of terminals and APs the optimum solution used as a reference for comparison can be easily obtained at each simulation time step by evaluating all the possible combinations. It is worth mentioning that the optimum solution depends on the terminals that are active at each time step, so it can change along the simulation. Presented results correspond to the average values along the whole simulation time.

As a reference for comparison, Figure 4-3 also includes the results with two simpler strategies. The first is the case in which all the terminals are connected to the macrocell (denoted as “All to Macro”). This would correspond to the classical approach in which no relaying through the APs

is used. The second is the random case in which each terminal selects randomly the AP or macrocell to receive service.

From Figure 4-3 it can be observed that the proposed approach with  $\tau=0.01$  and with the logarithmic cooling is able to achieve a performance very close to the optimum in all the considered scenarios. Moreover, very significant transmit power reductions can be achieved with the proposed strategy in comparison to the classical approach in which all the users are connected to the macrocell. Such approach requires in the order of 20%-50% higher transmit power depending on the scenario, demonstrating the efficiency of the proposed method to reduce the power consumption and thus contributing to overall energy savings.

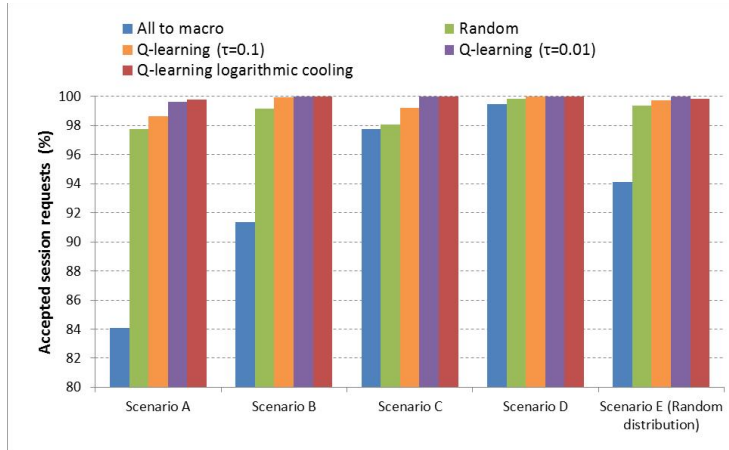


**Figure 4-3: Increase of the total transmitted power with respect to the optimum solution for the different strategies.**

The relevance of the temperature parameter  $\tau$  can also be observed in Figure 4-3. In fact, this parameter controls the trade-off between exploration and exploitation existing in the learning mechanism. Looking at the softmax criterion of (19), low temperature results into a greater difference in selection probability for access points that differ in their  $Q_n(k)$  value estimates, and the higher the value of  $Q_n(k)$  the higher the probability of selecting the  $k$ -th AP. As a result, with low values of  $\tau$  the system tends to converge quickly towards appropriate solutions with selection probabilities close to 1. In this way, the system can quickly exploit what has learnt by selecting the APs that provide the largest reward, at the expense that it will have less exploratory capability to identify other solutions in case the conditions change. On the contrary, with large values of the temperature, the differences in the selection probabilities for the access points become smaller, even if their  $Q_n(k)$  values are different. As a result, terminals will need longer times to identify the APs providing the largest reward so they will have less exploitation capability, but they will have a higher exploration capability to react to the changes. In the results presented in Figure 4-3 it can be observed that the choice  $\tau=0.01$  achieves a much better performance than the case  $\tau=0.1$  because in this later case there is an excessive exploration leading to the selection of non-optimal solutions in some cases. In fact, additional results not shown in the figure for the sake of simplicity have revealed that increasing  $\tau$  to larger values such as  $\tau=1$  leads to a performance very close to the random case, meaning that the selection probabilities of (19) become very similar for all the APs. In turn, when considering the logarithmic cooling, it can be observed in Figure 4-3 that the performance improves with respect to the fixed case  $\tau=0.01$ , mainly because the logarithmic cooling tends to reduce the temperature values as time elapses, so that the best solutions found are progressively selected with higher probability.

Results in Figure 4-3 indicate that the required transmit power along the simulation time with the proposed approach and logarithmic cooling is only 3.8% higher than the optimum one in scenario A while in the rest of scenarios this value reduces to less than 1.5%.

As another performance metric of interest, Figure 4-4 plots the rate of accepted requests by the admission control described in Section **Errore. L'origine riferimento non è stata trovata.** It can be observed that the proposed approach with  $\tau=0.01$  and with logarithmic cooling achieves the best performance, with almost 100% of acceptance in all the scenarios. On the contrary, for the rest of strategies, including the case in which all the terminals are connected to the macrocell, the acceptance ratio degrades significantly.



**Figure 4-4: Fraction of accepted session requests by the admission control.**

### Convergence analysis

This section analyses the convergence behavior of the proposed approach towards the optimum solutions. For this purpose, let consider first scenarios A and B in Figure 4-2. Since the optimum solution depends on the number of terminals that are active at each time, the analysis done in this section considers that all the terminals are continuously generating sessions of average duration 30 time steps without any inactivity period between sessions. In this way, all the terminals are active and the optimum solution to be found does not change during the simulation. In addition, no traffic is generated in this study by the terminals acting as APs.

To better analyze the characteristics of these scenarios,

Table **4-1** presents the total transmit powers that would be required for each terminal to get the required service through each of the APs and through the macrocell's BS. Moreover,

Table **4-2** presents the values of the fraction of times  $\theta_{k,n}$  for the different users when connected to the different APs. Focusing on Scenario A, from

Table **4-1** it can be observed that the minimum required transmit power is achieved in all the cases when the terminals are connected through the access point  $A_1$ . Then, if no other constraints were considered in the problem, the solution would be simply that all users connect to  $A_1$ . However, when considering the constraints of the problem, it can be noticed in

Table **4-2** that the solution in which all the users are connected to access point  $A_1$  does not fulfil constraint (9) with all the users active at the same time. By an exhaustive analysis of the combinations it can be observed that the optimum solution for scenario A is that users  $u_1, u_2, u_5$

and  $u_6$  connect to  $A_1$ , user  $u_3$  connects to  $A_2$  and user  $u_4$  connects to  $A_3$ . However, it can also be noticed that the differences between this global optimum solution and the variation of this combination when user  $u_3$  connects to either  $A_3$  or  $A_4$  and when user  $u_4$  connects to either  $A_2$  or  $A_4$  are very small (i.e. the power in any of these variations is only up to 0.08% higher than in the optimum case). This reflects that scenario A is characterized in practice by multiple optima. A similar analysis for scenario B yields that the optimum solution is that all users have to connect to  $A_1$ , so in this case there is a clear single optimum.

To illustrate how the algorithm is able to identify the optimum solution in the two scenarios,

Table 4-3 presents the selection probabilities where the system has converged at the end of the simulation. The Q-learning approach with logarithmic cooling is considered. Focusing on scenario A, it can be observed how the selection probabilities for all the users converge to the optimum solutions mentioned above. Specifically, users  $u_1$ ,  $u_2$ ,  $u_5$  and  $u_6$  select the access point  $A_1$  with very high probabilities above 98%. In turn, users  $u_3$  and  $u_4$  identify that connecting to  $A_2$ ,  $A_3$  or  $A_4$  give in practice almost the same result in terms of power consumption, so the three access points are selected with approximately the same probabilities of 33%. In turn, for scenario B it can also be observed that for all the users the probability of connecting to the optimum access point  $A_1$  becomes also above 98%.

**Table 4-1: Transmit power values (W) required by each terminal when connected to each of the different APs and to the macrocell in Scenarios A, B**

	Scenario A					Scenario B				
	macro	$A_1$	$A_2$	$A_3$	$A_4$	macro	$A_1$	$A_2$	$A_3$	$A_4$
$u_1$	3.990	1.605	2.399	2.399	2.402	3.990	1.605	2.399	2.399	2.402
$u_2$	3.192	1.6111	2.416	2.415	2.413	3.192	1.611	2.416	2.415	2.413
$u_3$	3.192	1.695	2.399	2.400	2.403	2.394	1.604	2.398	2.396	2.396
$u_4$	3.192	1.636	2.409	2.403	2.404	2.394	1.603	2.398	2.399	2.399
$u_5$	3.990	1.625	2.447	2.403	2.404	3.192	1.619	2.400	2.398	2.402
$u_6$	3.192	1.644	2.429	2.416	2.404	3.192	1.605	2.400	2.397	2.398

**Table 4-2: Fraction of times  $\theta_{k,n}$  for the different users when connected to the different APs in Scenarios A, B**

	Scenario A				Scenario B			
	$A_1$	$A_2$	$A_3$	$A_4$	$A_1$	$A_2$	$A_3$	$A_4$
$u_1$	0.084	0.051	0.045	0.082	0.084	0.051	0.045	0.082
$u_2$	0.148	0.222	0.206	0.186	0.148	0.222	0.206	0.186
$u_3$	0.987	0.050	0.057	0.085	0.082	0.036	0.016	0.017
$u_4$	0.402	0.144	0.088	0.093	0.068	0.036	0.045	0.047
$u_5$	0.285	0.532	0.085	0.097	0.231	0.057	0.035	0.080
$u_6$	0.475	0.352	0.213	0.220	0.092	0.054	0.030	0.042

**Table 4-3: Selection probabilities  $Pr(k,n)$  for the different users and APs in Scenarios A, B**

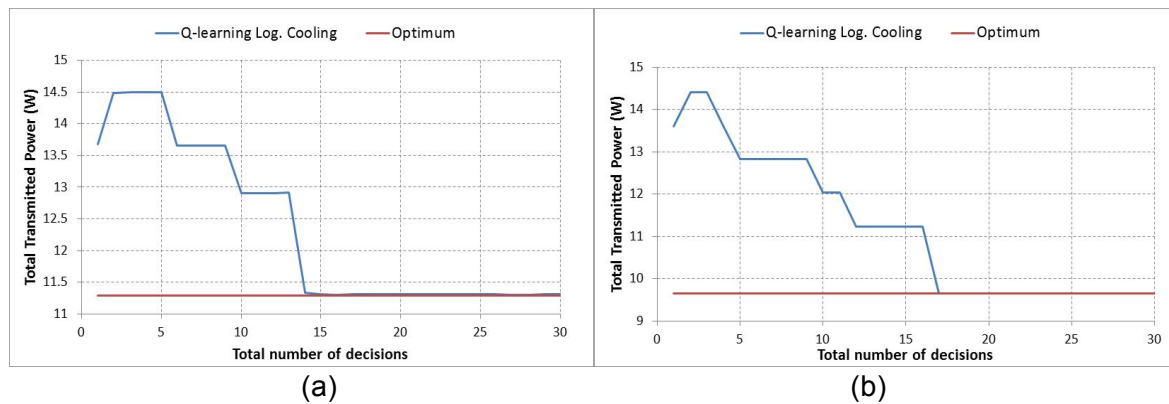
	Scenario A					Scenario B				
	macro	$A_1$	$A_2$	$A_3$	$A_4$	macro	$A_1$	$A_2$	$A_3$	$A_4$
$u_1$	0	0.985	0.005	0.005	0.005	0	0.984	0.005	0.005	0.005
$u_2$	0	0.989	0.003	0.003	0.004	0	0.989	0.003	0.003	0.004
$u_3$	0.001	0	0.336	0.334	0.328	0.005	0.980	0.005	0.005	0.005
$u_4$	0.001	0	0.324	0.338	0.336	0.003	0.987	0.003	0.003	0.003
$u_5$	0	0.989	0.003	0.004	0.004	0	0.989	0.004	0.004	0.003

$u_6$	0	0.992	0.004	0.004	0	0	0.988	0.004	0.004	0.004
-------	---	-------	-------	-------	---	---	-------	-------	-------	-------

Figure 4-5 depicts the evolution of the total transmitted power in scenarios A and B as a function of the total aggregated number of decisions (i.e. AP selections) made by all the terminals in the scenario. This total number of decisions is a representative metric of the convergence because the algorithm learns progressively the optimum solution as new decisions are being made by the different terminals. It can be observed that the total power is progressively decreasing until reaching the optimum value after a total of 14 decisions in the case of scenario A and 17 decisions in scenario B. Table 4-4 summarizes the total number of decisions made by the terminals before converging to the optimum solutions in the different scenarios of Figure 4-2. As it can be observed, in all the cases the convergence of the algorithm is quite fast. Note also that the presented numbers correspond to the total number of decisions for all the terminals. Then, considering that there are 6 terminals in the scenarios, this means that every terminal needs on average between 2 and 3 decisions to identify the proper access point.

**Table 4-4: Total number of decisions needed to converge to the optimum in the different scenarios**

Scenario	Total number of decisions
A	14
B	17
C	17
D	14



**Figure 4-5: Evolution of the total transmit power for the Q-learning approach with logarithmic cooling ( $\tau_0=0.01$ ) as a function of the total number of decisions made by the terminals in (a) Scenario A, (b) Scenario B.**

### Comparison against a centralized Genetic Algorithm

Results in previous sub-sections have compared the performance of the proposed approach directly against the optimum solution, since for the considered scenarios with a reduced number of access points and users, this optimum could easily be found e.g. by testing all combinations. When considering more complex scenarios, other optimization strategies need to be considered as a reference for comparison. In this respect, this section intends to benchmark the

performance achieved by the proposed distributed approach against a genetic algorithm. It is worth mentioning that such genetic algorithm will act by jointly considering all the different terminals and APs together in the optimization process. So its implementation would necessarily require a centralized approach, as opposite to the proposed Q-learning that is executed in a distributed way at each terminal. In this respect, the objective of this study is just to benchmark how far from such centralized technique the proposed distributed approach can be, but not to discuss on the implementation considerations associated to the comparison between the two techniques.

The genetic algorithm used for benchmark is triggered each time that a user begins or ends a session, in order to consider the possible reconfigurations that may be required as a result from these events. The genetic algorithm operates iteratively by evaluating in each iteration (also known as generation) a population of  $N_{pop}$  individuals or chromosomes, each of them corresponding to a candidate solution of the optimization problem [6]. The number of genes in each chromosome is  $N_{ACT}$  corresponding to the number of users in the scenario with an active session at the time the algorithm is triggered. Then, the  $i$ -th gene is associated to the user  $u_i$  and takes the value  $k=0$  if the user is connected to the BS of the macrocell or the value  $k \in \{1, \dots, K\}$  if the user is connected through the access point  $A_k$ . The chromosomes considered in each generation correspond to solutions that fulfill the constraints (6)-(9) of the problem (5).

Each chromosome is evaluated in terms of a cost or fitness function that captures the total required transmitted power associated to the solution represented by this chromosome. In particular, the cost function  $C(i)$  corresponding to the  $i$ -th chromosome is given by (3).

Based on the above considerations, the operation of the genetic algorithm is as follows:

- 1.- At initialization a set of  $N_{pop}$  chromosomes that fulfill the constraints (6)-(9) are randomly generated.
- 2.- The cost function  $C(i)$  is evaluated for each chromosome  $i$ .
- 3.- The following operators are applied over the different chromosomes to obtain the new set of  $N_{pop}$  chromosomes that will constitute the next generation:
  - 3.1.- Selection: The algorithm selects two chromosomes (parents) for the generation of two new chromosomes (children) for the subsequent generation. The parents are selected according to a roulette wheel process in which chromosomes with lower cost are selected with higher probability. Specifically, the probability of selecting the chromosome  $i$  is given by:

$$P_{Sel}(i) = \frac{\frac{1}{C(i)}}{\sum_{j=1}^{N_{pop}} \frac{1}{C(j)}} \quad (22)$$

3.2.- Recombination: The two selected chromosomes are recombined following the one-point-crossover methodology (see [6] for details) in order to obtain a new chromosome.

3.3.- Mutation: It consists in changing the value of a particular gene belonging to the new chromosomes resulting from the recombination step. The probability of mutating one gene is given by  $P_{mut}=1/N_{ACT}$ . When a gene is mutated, its new value is selected randomly among the values  $k=0, \dots, K$  excluding the current value of the gene.

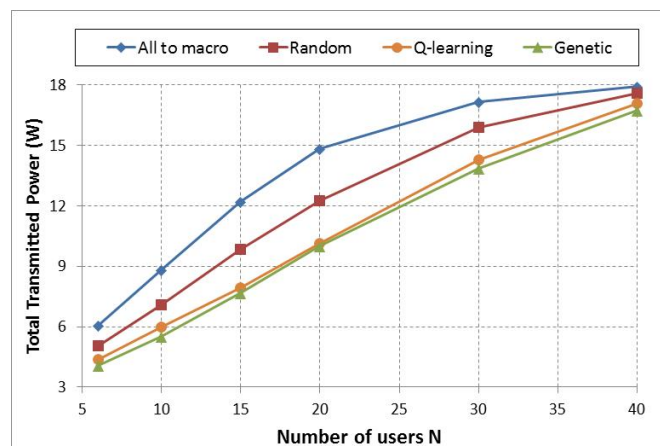
3.4.- It is checked that the resulting chromosome fulfills the constraints (6)-(9). If they are fulfilled, the chromosome is kept, and on the contrary, it is discarded. The overall selection,

recombination and mutation steps are repeated until obtaining a total of  $N_{pop}$  chromosomes for the next iteration.

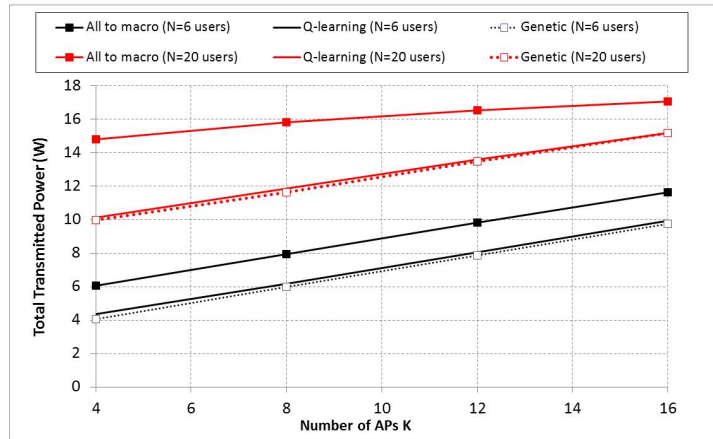
4.- The steps 2 and 3 are repeated iteratively until reaching a specific maximum number of iterations/generations. The solution of the algorithm corresponds to the chromosome with the minimum cost that has been found throughout all the generations.

The scenario considered for the evaluation in this case assumes the same square area of 400m x 400m with the BS located in the upper left corner as in Figure 4-2, but with random positions of the users and the APs. In particular, users are randomly distributed in the whole area, while the APs are also randomly distributed but only in the range [0-200m, 0-200m] which is closer to the base station. This is just to reflect the fact that the APs that are located very far away from the base station will not be useful for relaying traffic of terminals located closer to the base station. For each number of terminals and APs a total of 200 random realizations are simulated associated to different positions. Each execution of the genetic algorithm consists of 100 generations with a population of  $N_{pop}=30$  individuals. The rest of simulation parameters are the same as described at the beginning of this section with the only exception of the required bit rate that has been set to  $R_n=R_{A,k}=2$  Mb/s in order to have sufficient resources to serve a larger number of users in the scenario.

Figure 4-6 shows a comparison of the proposed Q-learning approach with logarithmic cooling and  $\tau_0=0.1$  with respect to the other considered strategies in terms of total transmit power in the scenario as a function of the number of user terminals  $N$  when there are  $K=4$  APs in the scenario. As shown, the results observed with the Q-learning approach clearly improve the results obtained by the “All to macro” and “random” strategies by significantly reducing the total transmitted power. Moreover the proposed Q-learning approach provides very similar performance as that of the centralized approach based on the genetic algorithm. Similarly, Figure 4-7 illustrates the impact of increasing the number of APs in the scenario in terms of the total transmitted power for the cases with  $N=6$  and  $N=20$  users. It can be noticed that the transmitted power increases with the number of APs because of the own transmission requirements associated to these terminals acting as APs. Note that in all cases, the proposed approach provides results very close to the genetic algorithm. Although this does not mathematically prove the guaranteed convergence to the optimum solution as in the previous sub-section, because the genetic algorithm could converge to either a global or a local optimum, it actually reveals that the proposed distributed approach is able to achieve a very close performance to a classical optimization approach such as the genetic algorithm, in spite of being much less complex.



**Figure 4-6: Total transmitted power in the scenario when increasing the number of terminals  $N$  for  $K=4$  access points.**



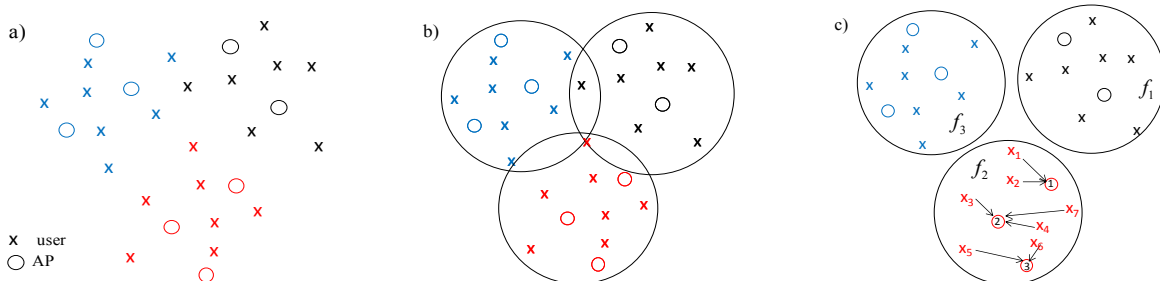
**Figure 4-7: Total transmitted power in the scenario when increasing the number of access points  $K$  for  $N=6$  and  $N=20$  users.**

### Dynamic Network Architecture and Topology Optimization

We consider an advanced wireless technology where a subset of terminals can be turned into an access point. With a slight modification of the existing technology a PC connected to the Internet can serve as an access point. Mobile phones have been recently designed to have such futures [7]. Such technology creates the possibility that a number of potential access points can be activated to serve a set of wireless terminals in their vicinity in an optimal way in accordance with some optimization criteria. This results into a Dynamic Network Architecture (DNA).

#### Network model

A high dense network, as shown in Figure 4-8a, is considered. To be able to handle such a dense network and solve efficiently the problems posed in the next sections, the DNA macro network is divided into clusters as shown in Figure 4-8b. The optimization will be solved per cluster as indicated in Figure 4-8c. The inter-cluster interference is eliminated by using frequency reuse factor as shown in the same figure. So, the problems addressed in the sequel will be solved first per DNA cluster and later on, further comments will be given on how to solve those problems at the level of the macro network. For clarity of presentation, we will refer by DNA to a DNA cluster within the the DNA macro network.



**Figure 4-8: Dynamic Network Architecture: a) Possible realization of DNA macro network; b) Clustering DNA macro network; c) Formal cluster separation in DNA network for**

### different frequencies ( $f_k$ , $k=1,2,3$ ) and, example of transmission between a cluster of DNA network.

We assume that there are  $K$  available potential access points and  $N$  users placed randomly in the DNA. Both, the location of the users and access points changes over time. We assume that the users in a DNA share a given channel to transmit uplink to the different APs. The resource allocation problem consists of allocating to  $N$  users,  $M$  out of  $K$  ( $M \leq K$ ) possible access points in an optimal way.

As the focus of this work is on high dense networks, the channel model considered includes the propagation losses, but not the effects of fading due to the proximity between the users [8]. So, if we denote by  $P_i$  the transmission power of user  $i$  then, the received power at AP  $j$ , is

$$P_j = h_{ij} P_i \approx d_{ij}^{-\alpha} \cdot P_i \quad (23)$$

where  $h_{ij}$  is the channel gain between user  $i$  and AP  $j$ ,  $d_{ij}$  is the transmission distance between them and  $\alpha$  is the path loss factor. The short transmission distance enabled by the availability of additional access points will significantly reduce the impact of fading. The issue of channel defading has been recently discussed in [8]. It was argued that in multihop wireless networks by reducing the distance between the receiver/transmitter pairs a point should be reached where the multipath component can be neglected and the general channel coefficient  $h_{ij}$  that includes the fading can be approximated by the attenuation due to distance only :  $d^{-\alpha}$ . The same principle applies to the case considered in this paper where additional access points are chosen on the distance that will enable channel defading. Elaboration of the fading issues as a standard problem has been minimized in order to make space to unconventional issues created by the new architecture presented in this paper. In the general case, instead of distance, the channel state information should be available in the optimization process.

We assume that user  $i$  can successfully connect to AP  $j$  only if the power received at  $j$  exceeds the receiver sensitivity  $\tau$ . As result, the distance between them must satisfy  $d_{ij} \leq (P_i / \tau)^{1/\alpha}$ . For a given distance  $d_{ij}$ , the minimum transmission power for user  $i$  to reach the access point is denoted as  $P_{iu} = \tau \cdot d_{ij}^\alpha$ . If the available power at user  $i$ ,  $P_{ia}$  is less than  $P_{iu}$  then the connection will not be established. We also define  $r_{iu} = (P_{iu} / \tau)^{1/\alpha}$  as the required communication range of user  $i$ .

In a high dense network, the aim is to keep the transmission power  $P_{iu}$  as low as possible, so that the interference is reduced and higher number of simultaneous transmissions can coexist in the network.

We denote by  $\mathbf{T} = [T_{ij}]$  the network topology matrix with entries  $T_{ij} = 1$  if user  $i$  transmits to AP  $j$  or zero otherwise.  $\mathbf{T}$  provides information of which user is transmitting to which AP for all pairs  $(i,j)$  of user  $i$ ,  $1 \leq i \leq N$  and AP  $j$ ,  $1 \leq j \leq K$  and it has dimensions  $N \times K$ .

We assume the following connectivity constraints:

- Each AP can serve one user at a time,

$$\mathbf{1} \cdot \mathbf{T} = \mathbf{a}, \quad a_j \leq 1 \quad (24)$$

where  $\mathbf{1}$  is a  $1 \times N$  vector with the value 1 in every entry.

- Each user will select one AP at a time,

$$\mathbf{T} \cdot \mathbf{1} = \mathbf{u}, \quad u_i \leq 1 \quad (25)$$

where  $\mathbf{1}$  is a  $K \times 1$  vector with the value 1 in every entry.

The topology  $\mathbf{T}$  is a *feasible* topology when the above constraints are satisfied and provides connectivity for all users  $N$  through the available access points  $K$  (i.e.,  $\text{Rank}(\mathbf{T})=N$ ). The set  $\mathbf{T}$  represents the set of all possible *feasible* topologies  $\mathbf{T}$ , for all possible pairs  $(i,j)$ .

Under the condition that we want to allocate  $K$  access points to all users in the same time slot ( $N=K$ ), the capacity of an uplink between the user  $i$  and access point  $j$  for certain topology  $\mathbf{T} \in \mathbf{T}$  is

$$c_{ij}(\mathbf{T}) = \rho_{ij} \log \left( 1 + \frac{h_{ij} \cdot P_i}{N_0 + \sum_{k \neq i} h_{kj} \cdot P_k} \right) \quad (26)$$

where it is assumed that all APs work on the same frequency channel and  $\rho_{ij}$  is a binary variable indicating the accessing status of user  $i$ . We have  $\rho_{ij}=1$  if the user  $i$  can communicate with the AP  $j$ ; otherwise,  $\rho_{ij}=0$ . That is by using (23)

$$\rho_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq r_{iu} \text{ and } P_i = P_{iu} \leq P_{ia} \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

where  $d_{ij}$  is the distance between the user  $i$  and its access point  $j$ . The interfering cochannel signal at access point  $j$ ,  $d_{kj}^{-\alpha} \cdot P_k$ , is generated by user  $k$ ,  $k \neq i$  while transmitting to its own access point. The background noise power is denoted by  $N_0$ . The overall network capacity is then defined as

$$C = \sum_{i=1}^N c_{ij}(\mathbf{T}) \quad (28)$$

If  $N > K$  or the spatial distribution of potential access points and users is such that none of the possible topologies  $\mathbf{T} \in \mathbf{T}$  can provide satisfactory performance then, some subsets of users might be scheduled in different time slots on the TDMA principle.

For this purpose, we redefine the topology as a block matrix  $\mathbf{T} = [\mathbf{T}^1 \mathbf{T}^2 \dots \mathbf{T}^\Delta] = [\delta \mathbf{T}]$  where each sub-matrix is a partial topology  $\delta \mathbf{T}$  per slot  $\delta$ ,  $\delta = 1, \dots, \Delta$  and  $\Delta$  is the length of the scheduling cycle.  $\delta \mathbf{T}$  provides the information of simultaneous transmissions in slot  $\delta$ . The dimension of  $\mathbf{T}$  is now  $\Delta \times N \times K$ .  $\mathbf{T}$  is a feasible topology if  $\delta \mathbf{T}$  satisfies the connectivity constraints given by (24) and (25) for every  $\delta$ ,  $\delta = 1, \dots, \Delta$  and provides connectivity to  $N$  users through  $K$  access points during the scheduling length  $\Delta$  (i.e.,  $\sum_{\delta} \text{Rank}(\delta \mathbf{T}) = N$ ). It is worth noticing that not all APs,  $K$ , need to be used for the topology to be feasible. Actually, we will be mostly interested to find a subset of the available APs that provide the users' QoS requirements.

The scheduling set  $S$ , is defined as the set  $S = \left\{ \Delta / \mathbf{T} \in \mathbf{T}, \mathbf{T} = [\delta \mathbf{T}], \delta \mathbf{T} \in \Pi, \delta = 1, 2, \dots, \Delta \right\}$  where  $\Pi$  denotes the collision-free set of transmissions. By definition, there is no collision in the transmission between user  $i$  and AP  $j$ , and interference user  $k$  and its AP  $j_k$ , if  $d_{ik} > d_{ij}$  and  $d_{ij_k} > d_{ij}$ .

The topology matrix, for the example presented in Figure 4-8c, is shown below where users 1 and 5 transmit on the first slot to access points 1 and 3, respectively. Users 2 and 4 share the second slot transmitting to the APs 1 and 2, respectively. Users 3 and 6 transmit in the third slot to APs 2 and 3, respectively. Finally, user 7 transmits on the fourth slot to AP 2.

$$\mathbf{T} = \begin{matrix} & i \backslash AP & 1 & 2 & 3 \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

The overall network capacity can be obtained as in (28) with  $\mathbf{T} = [\delta \mathbf{T}]$ ,  $\delta = 1, 2, \dots, \Delta$ .

The focus of this work is on uplink transmission, but the same model could be used for downlink as well by considering the AP transmission range, AP transmission power and SINR received by the users. It is worth noticing that the connectivity constraints for downlink transmission may result into different feasible topologies.

### Network Dynamics and Approximation for High Dense Scenarios

The network architecture and thus, the topology in DNA will change in time due to new/ended calls or new/ended service of APs. To reconfigure the optimum topology to the traffic changes in a high dense network, some simplifications are needed for practical implementation.

If we denote by  $\lambda_m$  the call arrival rate, the computational time  $T_c$  needed to obtain the new optimum topology after a traffic change should satisfy the following constraint

$$T_c < 1/\lambda_m \quad (29)$$

In this way, the new topology can track the network dynamics. To keep the computational complexity under that threshold, the size of the DNA macro network should be scaled down accordingly. For this purpose, we assume that the DNA macro network is divided into clusters where its size, in terms of  $N$  and  $K$ , is such that constraint (29) holds. The concept of DNA clustering is shown in Figure 4-8. More detailed comments on how to optimize different aspects of the DNA macro network will be given in the following section. Notes on the practical implementation will be provided in Section **Errore. L'origine riferimento non è stata trovata.**

### Acquisition of the Optimum Network Architecture

In this section, a number of optimization problems are presented for the DNA paradigm. In the first step, the aim is to develop a basic algorithm to find the optimum topology  $\mathbf{T} \in \mathbf{T}$  in accordance with some utility function. Later on, this algorithm will be used to solve the topology optimization in more sophisticated problems. The QoS requirements will be included as constraints of the optimization problems with the objective to find the minimum required number of APs. Then, an economic model is developed to compensate the users for acting as APs and thus, contribute to increase the network resources. Finally, the security requirements are considered.

### Topology Optimization for fixed number of APs

We start by considering that the DNA network consists of  $N$  users and  $K$  access points which will be referred to as DNA( $N, K$ ). There are a number of possibilities to allocate all these  $N$  users to  $K$  access points. Each option defines a feasible topology  $\mathbf{T} \in \mathbf{T}$ . The utility function is defined as the sum of the utilities per user  $i$

$$U = \sum_{i=1}^N U_i = \sum_{i=1}^N \frac{c_{ij}(\mathbf{T})}{\Delta \cdot P_i} \quad (30)$$

and includes the link capacity  $c_{ij}(\mathbf{T})$  between user  $i$  and AP  $j$  when transmitting by topology  $\mathbf{T}$  defined by (26), the scheduling length  $\Delta$  and the power consumption  $P_i$ . All these parameters have impact on the election of the optimum topology. In order to keep the overall transmission power as low as possible, we assume that each user transmits with the minimum transmission power needed to reach the access point  $P_i = P_{iu}$ . We also assume that the users' available power  $P_{ia} \geq P_{iu}$ . Then, the optimum topology is obtained by solving the following optimization problem

$$\begin{aligned} P_1 : \quad & \underset{\mathbf{T}}{\text{maximize}} \quad \sum_{i=1}^N \frac{c_{ij}(\mathbf{T})}{\Delta \cdot P_i} \\ & \text{subject to} \quad \mathbf{T} = [\delta \mathbf{T}], \delta = 1, 2, \dots, \Delta \\ & \quad \mathbf{T} \in \mathbf{T}, \Delta \in S \\ & \quad \sum_{\delta} \text{Rank}(\delta \mathbf{T}) = N \\ & \quad P_i = P_{iu}, P_i \leq P_{ia} \end{aligned} \quad (31)$$

As result, the optimum topology  $\mathbf{T}^*$  that satisfies the previous constraints is obtained where  $\Delta \in S$ . Although this problem is NP-hard [9], the utility considered allows certain simplifications.

The dependency between the utility and power follows the relation  $\log(P)/P$  so, lower power translates into higher utility. At the same time, lower power will imply lower  $\Delta$  as more simultaneous transmissions can coexist in the network. The load distribution is considered in (31) through the scheduling length  $\Delta$ . As we have assumed that users will transmit one at a time to a particular AP, higher number of users allocated to the same AP will result into higher  $\Delta$ , and thus lower utility. For these reasons, this optimization will provide as result for optimum topology  $\mathbf{T}^*$  the one that connects the users to its closest available AP. Based on this result, a Minimum Distance Clustering/Scheduling (MDCS) scheme can be used for topology optimization. In MDCS, the users transmit to its closest AP on a cluster basis. Although, there are many options to perform the scheduling, this scheme significantly reduces the topology search space. The scheduling between adjacent clusters in DNA can be performed with a temporal offset in terms of slots or spatial offset by allocating different frequencies to each cluster. This will result into different reuse factor  $F$  for the DNA macro network,  $\text{DNA}(N, K, F)$ , as shown in Figure 4-8c for  $F=1/3$ .

In order to provide the incentive for the terminals to serve as access points in a given time period, the network will have to compensate such a service by paying them a certain amount in a normalized currency that will be discussed later. As the cost of having large number of access points  $K$  may be too high, in the sequel we study the possibility of activating  $M \leq K$  available APs while still satisfying the users' QoS requirements.

## Topology and Architecture Optimization

The optimization problem defined in (31) can be modified to include the cost incurred by having certain amount of APs. The aim is now to obtain the optimum topology  $\mathbf{T}$  and the number of APs,  $M$ , that guarantees the connectivity for all users  $N$ , and provides the maximum utility and minimum cost.

Let  $\sigma$  be a  $1 \times K$  vector where each component  $\sigma_j$  is a binary variable indicating the selection of AP  $j$  with entries  $\sigma_j = 1$  if the AP  $j$  is elected; otherwise,  $\sigma_j = 0$ . The vector where each component is the opposite to each component of  $\sigma$  is denoted by  $\bar{\sigma}$ . Its transpose is given as  $(\bar{\sigma})^T$ .

The topology and architecture optimization problem is defined as follows

$$\begin{aligned} P_2: \underset{\mathbf{T}, \sigma}{\text{maximize}} \quad & \sum_{i=1}^N \sum_{j=1}^K \sigma_j \left( \frac{c_{ij}(\mathbf{T})}{\Delta \cdot P_i} - \nu \text{cost}_j \right) \\ \text{subject to} \quad & \mathbf{T} = [\delta \mathbf{T}], \delta = 1, 2, \dots, \Delta \\ & \mathbf{T} \in \bar{\mathbf{T}}, \Delta \in S \\ & \mathbf{T} \cdot (\bar{\sigma})^T = \mathbf{0} \\ & \sum_{\delta} \text{Rank}(\delta \mathbf{T}) = N \\ & P_i = P_{iu}, P_i \leq P_{ia} \end{aligned} \quad (32)$$

where  $\nu$  is a scaling parameter. The first and second constraints state that the topology  $\mathbf{T}$  should be a feasible topology, the third constraint indicates that the users should be allocated just to active APs, the fourth constraint indicates that  $\mathbf{T}$  should provide connectivity for all users  $N$  and finally, the fifth constraint establishes the power limits.

As result, the feasible set of topologies  $\mathbf{T}$  can be obtained by MDCS scheme for each  $\sigma$ . Then, the optimum  $\mathbf{T}^* \in \mathbf{T}$  and  $\sigma^*$  that maximize the utility for certain cost are jointly chosen. The number of selected APs,  $M$ , that maximizes the utility is obtained as  $M = \sum_{j=1}^K \sigma_j$ . If the  $\text{cost}_j$  is the same for every AP then, the previous optimization will provide the minimum  $M$ . Otherwise, it will lead to configurations that avoid using high cost APs.

### QoS requirements

We consider that the users' QoS requirement is given in terms of throughput by the following constraint

$$c_{ij}(\mathbf{T}) / \Delta \geq \gamma \quad (33)$$

where  $\gamma$  is a constant. The optimization problem defined by (32) can be reformulated to minimize the number of APs,  $M$ , needed to satisfy the QoS requirements. This can be formulated as

$$\begin{aligned} P_3: \underset{\mathbf{T}, M}{\text{minimize}} \quad & M = \sum_{j=1}^K \sigma_j \\ \text{subject to} \quad & \mathbf{T} = [\delta \mathbf{T}], \delta = 1, 2, \dots, \Delta \\ & \mathbf{T} \in \bar{\mathbf{T}}, \Delta \in S \\ & \sum_{\delta} \text{Rank}(\delta \mathbf{T}) = N \\ & \mathbf{T} \cdot (\bar{\sigma})^T = \mathbf{0} \\ & P_i = P_{iu}, P_i \leq P_{ia} \\ & c_{ij}(\mathbf{T}) / \Delta \geq \gamma \end{aligned} \quad (34)$$

where the constraints are the same as in (32) plus the QoS constraint given by (33).

As before, if we assume that the cost is the same for every AP then the previous optimization will implicitly minimize the cost. In the sequel, different options for the cost will be studied together with its impact on the AP selection.

### Economic Model for Resource Harvesting in DNA

As described in the previous sections, in DNA a terminal can operate either as a user consuming the network resources or as AP augmenting the network resources. In the former case, the terminal will pay to the network an amount proportional to the resource consumption while in the latter case, the network will pay to the terminal an amount proportional to its contribution to the overall augmentation of the network resources. To reduce the cost that APs bring to the network, the users will choose the AP that offers the minimum price. For this reason, we need to specify in more details the contract between the terminal and the operator(s). In the sequel, we present different options for these contracts.

#### T/W(r/q)/I(R/Q) contract

In this option the terminal,  $T$ , when acting as a user, has a contract with the wireless operator,  $W$ , and separate contract with the Internet operator,  $I$ . Both contracts may be based on the rate  $r$  provided by the wireless operator (rate  $R$  provided by Internet operator) or upload traffic volume  $q$  in the wireless connection (traffic volume  $Q$  on the Internet connection).

The pricing mechanism may be designed in such a way that the price is proportional to the real rate  $r(t)$  and time  $T_r$  the channel is used. This is designated as

$$price(r, T_r) = \alpha_r \int_0^{T_r} r(t) dt \quad (35)$$

where  $\alpha_r$  is a proportionality constant. If the operator provides a fixed guaranteed rate  $r_0$  then,

$$price(r_0, T_r) = \alpha_r \int_0^{T_r} r_0 dt = \alpha_r r_0 T_r \quad (36)$$

Instead of rate, the pricing might be based on the maximum upload traffic volume  $q$  during time  $T_q$  resulting in fixed price

$$price(q, T_q) = \alpha_q \int_0^{T_q} r(t) dt \quad (37)$$

where  $\alpha_q$  is a proportionality constant, or for the fixed rate

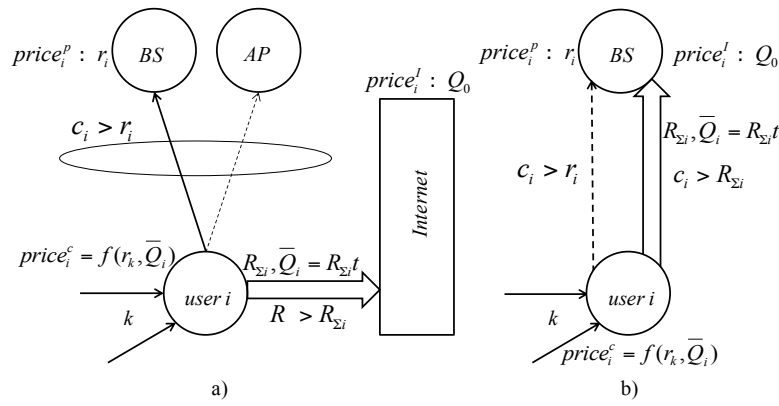
$$price(q_0, T_q) = \alpha_q \int_0^{T_q} r_0 dt = \alpha_q r_0 T_q \quad (38)$$

Similar expressions can be used for the Internet where instead of  $r$  and  $q$  notation,  $R$  and  $Q$  are used. Depending on the pricing mechanism used, a specific contract may have four different options:  $T/W(r)/I(R)$ ,  $T/W(r)/I(Q)$ ,  $T/W(q)/I(R)$  and  $T/W(q)/I(Q)$ . Besides, on the Internet segment there is the possibility that the AP has wired or wireless Internet through the cellular network.

To illustrate the previous discussion, an example for  $T/W(r)/I(Q)$  contract is shown in Fig. 4-9 including different options for DNA infrastructure for wired and wireless Internet. In particular, in Figure 4-9a, the user can transmit wireless through the base station (BS) or AP, and wired to the Internet. The price that terminal  $i$  will pay for the wireless connection,  $price_i^p$ , depends on its transmission rate  $r_i$ , while the price of the Internet service,  $price_i^l$ , depends on the prepaid traffic volume,  $Q_0$ . The price that the terminal will charge to any adjacent user  $k$  for the connection is referred to as  $price_i^c$  and depends on its rate  $r_k$  and the traffic volume  $\bar{Q}_i$  previously used at that

terminal  $i$ . We denote by  $R_{ei}(I_{ki}) = \sum_k I_{ki} r_k$  the overall external transmission rate of terminal  $i$  due to the adjacent users  $k$  that transmit to  $i$  when acting as AP. The indicator  $I_{ki}=1$  if user  $k$  transmits through  $i$  and zero, otherwise. The overall rate on the Internet connection  $R$  is constrained by the transmission rate of terminal  $i$ ,  $R > R_{\Sigma i} = R_i + R_{ei}$  where  $R_{\Sigma i}$  is the aggregated rate of terminal  $i$  and its adjacent users. The rate  $R_i$  will be referred to as the internal rate of the terminal  $i$  on the Internet and  $R_{ei}$  the external transmission rate of the adjacent users. The traffic volume transmitted by  $i$  up to time  $t$  is given by  $\bar{Q}_i = R_{\Sigma i} t$ . The same notation applies for Figure 4-9b, with the only difference that the Internet connection is wireless. Thus, the overall rate on the wireless Internet connection is constrained by the link capacity,  $c_i > R_{\Sigma i}$ .

In order to analyze the balance on the terminal's account and thus, the price that the terminal will charge to its adjacent users when acting as an AP, we need to average out the terminal operation in time. We will assume that a terminal has a contract of the type  $T/W(r)/I(Q)$ , as shown in Figure 4-9, signed with a single operator, which means that the same operator provides both the  $W$  and the  $I$  service.



**Figure 4-9: DNA network model with  $T/W(r)/I(Q)$  contract with: a) wired Internet and b) wireless Internet.**

Let assume that the user pays to the Internet operator a fixed price  $price_i^l = \chi_0$  for transmitting a maximum traffic volume  $Q_0$  on the Internet connection during period  $T_0$ . We denote by  $p^l$  the probability of having the Internet connection available at a given moment. The probability that the user transmits its own traffic on the Internet connection is given by  $p_i^l$ , and the probability of using instead the wireless connection is  $p_i^w$ .

Then, the average traffic volume sent by terminal  $i$  on the Internet connection until time  $t$  is given by

$$\bar{Q}_i(t) = p^l (p_i^l R_i t + (1 - p_i^l) R_{ei} t) \quad (39)$$

where  $R_i$  is the transmission rate of terminal  $i$  and  $R_{ei}$  is its average external rate. To define this parameter, we first define the probability  $p_{ki}$  that  $i$  is in the transmission range of user  $k$ . This probability can be expressed as  $p_{ki} = A_k / A_c$  where  $A_k$  and  $A_c$  are the transmission area of user

$k$  and area of the DNA cluster, respectively. Now, the average overall external rate at terminal  $i$  can be presented as

$$R_{ei} = r \sum_{k=0}^N k(e_i p_{ki})^k \cdot (1 - e_i p_{ki})^{N-k} \quad (40)$$

where  $r = \bar{r}_k$  is the average user rate on the wireless connection and  $e_i$  is the probability that terminal  $i$  is elected as AP. The probability  $e_i$  is supposed to be inversely proportional to the price that the terminal  $i$  will charge to its adjacent users for the connection which is referred to as  $price_i^c$ . This price depends on the fixed price  $\chi_0$  paid for the Internet and remaining traffic volume available at  $i$ ,

$$price_i^c = \frac{\chi_0}{1 + Q_0 - \bar{Q}_i(t)}, \quad Q_0 \geq \bar{Q}_i(t) \quad (41)$$

where  $\bar{Q}_i(t)$  is the cumulative traffic transmitted until time  $t$ . As the available traffic volume at terminal  $i$  decreases, the higher will be the price it will charge to its adjacent users for the connection. Then, by modeling  $e_i = \alpha_e / price_i^c$  where  $\alpha_e$  is a proportionality constant, so that  $0 \leq e_i \leq 1$ , and using it in (40),  $\bar{Q}_i(t)$  can be evaluated. For simplicity, we assume that only one user can access terminal  $i$ , thus (39) can be rewritten for  $R_{ei}=r$  as

$$\begin{aligned} \bar{Q}_i(t) &= p^I (p_i^I R_i t + (1 - p_i^I) r t \alpha_e / price_i^c) \\ &= p^I (p_i^I R_i t + (1 - p_i^I) r t \alpha_e (1 + Q_0 - \bar{Q}_i(t)) / \chi_0) \end{aligned} \quad (42)$$

At  $t = T_0$ , we have

$$\begin{aligned} \bar{Q}_i(T_0) &= p^I (p_i^I R_i T_0 + (1 - p_i^I) r T_0 \alpha_e / price_i^c) \\ &= p^I (p_i^I R_i T_0 + (1 - p_i^I) r T_0 \alpha_e (1 + Q_0 - \bar{Q}_i(T_0)) / \chi_0) \end{aligned} \quad (43)$$

Alternatively to the previous definitions, a dynamic model for the consumed traffic volume can be defined as

$$\begin{aligned} Q_i(t+1) &= p^I (p_i^I R_i(t) \Delta t + (1 - p_i^I) e_i r \Delta t) \\ &= p^I (p_i^I R_i(t) \Delta t + (1 - p_i^I) r \Delta t \alpha_e (1 + Q_0 - \bar{Q}_i(t)) / \chi_0) \end{aligned} \quad (44)$$

where  $Q_i(t+1)$  is the volume of traffic transmitted in time  $(t, t+1)$ ,  $\Delta t$  is the duration of the transmission and  $\bar{Q}_i(t)$  is the cumulative traffic transmitted until time  $t$  obtained by (44) as

$$\bar{Q}_i(t) = \sum_{j=1}^t Q_i(j) \quad (45)$$

Finally, the overall traffic transmitted during  $T_0$ ,

$$\bar{Q}_i(T_0) = \sum_{t=1}^{T_0} Q_i(t) \quad (46)$$

As performance measure, the following parameters are defined:

- The Internet *contract utilization*  $\xi_i$  defined as

$$\xi_i = \bar{Q}_i(T_0) / Q_0 \quad (47)$$

where  $\bar{Q}_i(T_0)$  is the volume of traffic transmitted during the length of the contract,  $T_0$ , and  $Q_0$  is the initial amount of traffic volume available.

- The *contract price recovery*  $\varepsilon_i$  obtained as

$$\varepsilon_i = p^i \frac{1}{\chi_0} \left( \frac{\chi_0 p_i^i R_i T_0 + price_i^c (1 - p_i^i) e_i r T_0}{Q_0} \right) \quad (48)$$

where the first term is the percentage of price spent by terminal  $i$  on transmitting its own traffic and the second term is the percentage of price gained when acting as an AP.

The previous equations can be easily extended for other type of contracts (e.g.,  $T/W(r)/I(R)$ ,  $T/W(q)/I(R)$  and  $T/W(q)/I(Q)$ ). This framework provides many opportunities for further extension of the models, especially in multi-operator scenarios. The reward for serving as an access point may also depend on the capacity of the terminal so that the weaker terminals (like smart phones) could be awarded more than PCs or conventional access points with higher capacity.

## Security Investment

In this section, we are interested to evaluate how the user's security investment affects the performance and consequently, the network utilization.

We assume that the user invests in security measures, such as purchasing software and configuring it on its system. This provides a security level  $L_i$  and brings an increased direct cost  $S_i$ . Higher cost results into higher security level. The security investment reduces the probability that the user would be vulnerable to an attack and also the probability that other users will be attacked. When the terminal acts as an AP for other users in its vicinity it will benefit from its security investment and the probability of being elected as an AP,  $e_i$ , will be higher. The security level demanded by an adjacent user  $k$  when connecting to terminal  $i$  is

$$D_{ki} \geq L_k \quad (49)$$

so, that its investment will not be wasted. Otherwise, the user will not be interested on that connection. Then, the probability that terminal  $i$  will be elected as an AP can be approximated as

$$e_i = \frac{\alpha_e}{price_i^c} p(L_i \geq D_{ki}) = \frac{\alpha_e}{price_i^c} p(L_i \geq L_k) \quad (50)$$

where  $e_i$  is inversely proportional to the  $price_i^c$  charged for the connection by terminal  $i$  and given by (41) and, proportional to the probability that  $i$  will satisfy the security level demanded by user  $k$ .

If the difference between the security level of user  $k$  and terminal  $i$ , is very large  $L_i - L_k \gg 0$  then,  $i$  will charge an extra price,  $price_i^s = \alpha_s (S_i - S_k)$ , to serve user  $k$ , where  $\alpha_s$  is a proportionality constant. This cost will compensate that  $i$  will need to scan the system or taking extra security measures after serving user  $k$ .

As result, the price that terminal  $i$  will charge to  $k$  for a *secure* connection is

$$price_i^{cs} = price_i^c + price_i^s \quad (51)$$

The economic model discussed in Section 4.2.3.3 should be modified to include  $e_i$  and the price defined by (50) and (51), respectively.

## Dynamic Tracking of the Optimum Topology and Architecture by Genetic Algorithm

In a real network when the traffic changes in time and space, an efficient mechanism is needed to reconfigure the optimum topology to the traffic variations in the network. The optimization problems defined in this section can be solved in a dynamic environment, where the topology in the observation instant  $t$  is denoted by  $\mathbf{T}^t$ . As already mentioned, the time to reconfigure the topology is limited by the network dynamics. For these reasons, a genetic algorithm is developed to track the changes in the optimum topology due to the traffic variations.

### Genetic Algorithm (GA)

Genetic Algorithm is a computational mechanism inspired by natural evolution where stronger individuals are more likely to survive in a competitive environment. GA has been shown to be a useful alternative to traditional search and optimization methods, especially for problems where the space of all potential solutions is too high to be searched exhaustively in any reasonable amount of time [10][11].

The first step in GA is to encode the problem as a chromosome or a set of chromosomes that consist of several genes. Next, a pool of feasible solutions to the problem, called initial population, is created. A fitness value, calculated using a fitness function, is associated with each chromosome and indicates how good the chromosome is. Genetic operators' *selection*, *crossover* and *mutation* operate on the population to generate a new generation of population, i.e., a new set of feasible solutions from the old ones. Good feasible solutions are selected with higher probability to the next generation, in line with the idea of survival of the fittest. As the algorithm continues and newer generations evolve, the quality of solutions improves. The success of GA and its applications are outlined in [12].

In this paper, we formulate a genome as a feasible topology  $\mathbf{T} \in \mathcal{T}$  which consists of a block of chromosomes  ${}^\delta \mathbf{T}$  (partial topologies) that provide connection for  $N$  users to a  $K$  access points in a scheduling period  $\Delta$ . We model the traffic variations in our DNA network by considering that the topology variations are due to the changes in the APs and users availability.

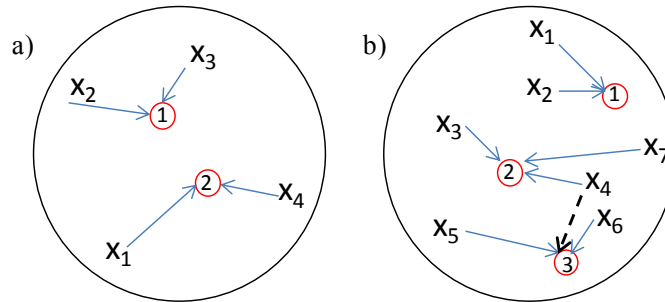


Figure 4-10: Illustration of the genetic operators: a) crossover and b) mutation.

### Encoding and initial population

We encode the topologies  $\mathbf{T}$  as a block of chromosomes where each chromosome defines a partial topology  $\mathbf{T} = [{}^\delta \mathbf{T}]$ ,  $\delta = 1, 2, \dots, \Delta$  and  $\Delta$  is the scheduling length. Each entry of the chromosome  ${}^\delta \mathbf{T} = [{}^\delta T_{ij}]$  represents a gene with genotype  $i$  (user) and phenotype  $j$  (access point). So, each gene defines a connection between user  $i$  and AP  $j$ .

We assume that the initial topology of the network is known and denoted by  $T_0$ . Later on, we will provide details on the robustness of our algorithm to the initial state of the network. This topology needs to be modified accordingly based on the following possible changes in the network:

- a) If a new user arrives to the network then, the new user will be initially assigned to its closest AP.
- b) If a user leaves the network then, its entry will be zero in the topology matrix.
- c) If a new AP appears in the network then, its closest user will be reassigned to this new AP.
- d) If a new AP leaves the network then, its users will be reassigned to the remaining closest AP.

There is also the possibility to detect more than one change at a time. This is the case when:

- e) A user turns into an AP. In this case, the topology should be modified as indicated in b) and c).
- f) An AP turns into a user. Then, the topology should be modified as indicated in a) and d).

The update of the initial topology  $T_0$  results into  $T_0^+$ . At any time instant  $t$ ,  $g$  topologies (genomes) are contained in the population pool  $P(t)$ . The initial population  $P(0)$  consists of topology  $T_0^+$  and  $g-1$  topologies randomly generated,  $P(0) = \{T_0^+, T_1, \dots, T_{g-1}\}$ . The next generation of population is obtained by applying *genetic operators* over the topologies of the current generation. The fitness function used to evaluate the topologies is given by the utility function defined by (30), (32) or (34).

## Genetic Operators

We consider the operations of *selection*, *crossover* and *mutation*.

The *selection* operation: consists of choosing 40% of the topologies that produce the highest fitness among the current population to survive to the next generation. Then, the rest of the new population is obtained by applying crossover and mutation operations to those selected topologies (30% topologies generated by each operator). This provides a compromise between the level of elitism and diversity to generate new topologies with the past of the generations.

The *crossover* operation: consists of shifting two genes between two different chromosomes from the same genome to generate an offspring topology. In particular, we randomly choose one of the selected topologies at a time and then we randomly choose two genes and apply the crossover to generate a new topology as

$$\delta_1 T(i_1, j_1) \text{ } \ddot{\wedge} \text{ } \delta_2 T(i_2, j_2)$$

A topology can be repeatedly selected, but if the offspring topology results into an existing one then it will be removed from the pool and another topology will be randomly chosen. The new topology resulting from this operation will be always a feasible topology.

The purpose of this operation is to reduce the interference resulting from concurrent transmissions. In this case, we can move one of the interference users (the whole gene) to transmit in a different slot. This operation is illustrated with the scenario shown in Figure 4-10a, where initially users 1 and 2 share the first slot and users 3 and 4 transmit in the next slot. After the crossover, the genes 1 and 4 will be shifted resulting into the following new topology

$$\mathbf{T} = \begin{matrix} & i \backslash \text{AP} & \begin{matrix} 1 & 2 \end{matrix} & & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{matrix} \delta=1 \\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=2 \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix} \end{matrix}; \quad \mathbf{T}_{new} = \begin{matrix} & i \backslash \text{AP} & \begin{matrix} 1 & 2 \end{matrix} & & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{matrix} \delta=1 \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=2 \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \end{matrix} \end{matrix}$$

$$\delta_1 T(1,2) \rightleftharpoons \delta_2 T(4,2);$$

where users 2 and 4 will share the first slot and users 1 and 3 will transmit in the second slot. As we can see in Figure 4-10a, the interference in this topology will be reduced.

The *mutation* operation: is performed to facilitate jumping of solutions to new unexplored regions of the search space. It consists on allocating the user to a more convenient AP. This can be achieved by mutating the phenotype of an individual gene. An example of the mutation operation is shown in Figure 4-10b, where the phenotype of user 4 is mutated from 2 to 3. The new topology results into

$$\mathbf{T} = \begin{matrix} & i \backslash \text{AP} & \begin{matrix} 1 & 2 & 3 \end{matrix} & & \begin{matrix} 1 & 2 & 3 \end{matrix} & & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{matrix} \delta=1 \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=2 \\ \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=3 \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix} \end{matrix}; \quad \mathbf{T}_{new} = \begin{matrix} & i \backslash \text{AP} & \begin{matrix} 1 & 2 & 3 \end{matrix} & & \begin{matrix} 1 & 2 & 3 \end{matrix} & & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{matrix} \delta=1 \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=2 \\ \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix} & \begin{matrix} \delta=3 \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{matrix} \end{matrix}$$

$$\delta_1 T(4,2) \rightarrow \delta_2 T(4,3);$$

### Convergence and complexity of GA

After a new generation of population is created by using the genetic operators described above, the fitness function is used again to evaluate the generated topologies and the process is repeated in the same fashion. As 40% of the best topologies are kept in the pool, the fitness of the best topology in each generation will always be better or equal than in the previous generation.

If we denote by  $f'_n(\mathbf{T}'_n)$  the best fitness function in generation  $n$  obtained for topology  $\mathbf{T}'_n$  then, GA converges to the solution when  $|f'_{n+1}(\mathbf{T}'_{n+1}) - f'_n(\mathbf{T}'_n)| \leq \varepsilon$ , where  $\varepsilon > 0$ . The topology  $\mathbf{T}'_{n+1}$  will be the optimum topology  $\mathbf{T}^*$  with certain probability that we denote probability of success or success ratio,  $p_{sus}$ . This probability is obtained as the ratio between the number of times the optimum topology has been found with respect to the number of runs of the GA. In this process, exhaustive search is used to confirm the optimum topology. The optimum fitness function  $f^*$  is obtained for topology  $\mathbf{T}^*$  when  $p_{sus} = 1$ . If  $p_{sus} < 1$  then,  $|f^* - f'| > 0$  and the topology  $\mathbf{T}'_{n+1}$  is a suboptimum solution. To achieve the optimum solution, the size of the population  $g$  and the number of generations  $N_g$  considered for the GA must be adjusted to the size of the DNA. This will be shown in the next section through simulations.

The complexity of the GA is given by the following parameters:

- Total number of generated topologies  $G$  needed to obtain the optimum solution,  $G = g \cdot N_g$  where  $g$  is the number of topologies per generation and  $N_g$  is the number of generations.
- Computational time which indicates the time it takes to obtain the optimum topology (CPU time).

## 2-Level Access Admission Control (2L-AAC) Scheme

The conventional admission control mechanisms aim at maintaining the required QoS by the users by limiting the number of new users that access the network at a given time. In this paper, we present a 2-level access admission control (2L-AAC) protocol that regulates the access of new users and APs to the network in order to keep the level of utility per user  $U_i$  above certain threshold  $U_0$ .

If we assume that the network efficiency is measured by the utility defined by (30) and there is a user with utility  $U_i < U_0$  then, the 2L-AAC can improve  $U_i$  by performing one of the following actions:

- Allow a new AP  $k$ , located in a closer distance to user  $i$  than the actual AP  $j$ ,  $d_{ik} < d_{ij}$ , to access the network. In this case, the power consumption  $P_i$  will be reduced as the new AP is closer.
- Reduce the number of users that are transmitting to AP  $j$  so that, the scheduling cycle  $\Delta$  will be reduced.
- Reduce the number of users that are sharing the slot with user  $i$  which will increase the capacity  $c_{ij}(\mathbf{T})$ .

All the previous options will also increase the overall utility. After there is a traffic changed, the new topology should be reconfigured to provide the optimum performance according to (30).

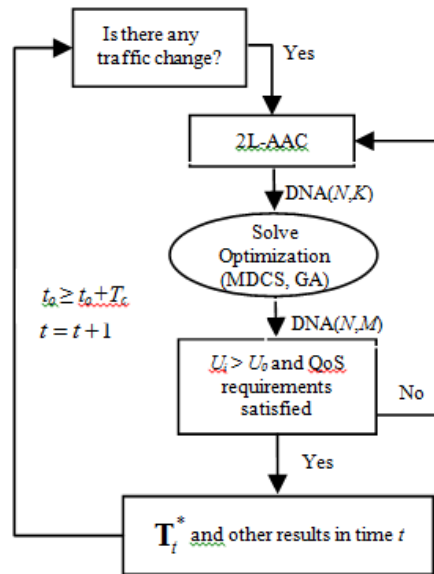


Figure 4-11: Interaction of the 2L-AAC scheme and the optimization problem.

If the utility also includes the cost as in (32) and the maximum utility at certain state of the network  $U_i < U_0$  then, the same actions as before can be applied. The only difference is that

introducing a new AP will increase the cost and this increment should compensate the gain on the utility. Otherwise, the new AP  $k$  may replace the actual AP  $j$ .

The illustration of the 2L-AAC scheme is shown in Figure 4-11 where it is assumed that the arrival rate of users/APs,  $\lambda_m$ , and the computational time,  $T_c$ , satisfy (29). The observation instant  $t_o$  is assumed to be multiple of  $T_c$ .

## Implementation

The optimization problems will be run at the BS or equivalent network controller who will keep track on the existing traffic in each DNA cluster. The BS will assign the users to the most appropriate cluster and the 2L-AAC mechanism will provide the access on a cluster basis.

As result of the optimization problems, the optimum topology which provides the data for *intra-cluster* reallocation (handover) is obtained. The terminal status (user or AP) is communicated on the conventional uplink signaling (control) channel. Then, the network controller will assign each user to the most convenient AP according to a given utility function. We have assumed that the user-AP allocation is fixed during the scheduling cycle. Besides, we assumed that the computational time  $T_c$  needed to obtain the new optimum topology after a traffic change is  $T_c < 1/\lambda_m$  where  $\lambda_m$  is the call arrival rate. In this way, the new topology can track the network dynamics. The observation instant  $t_o$  is assumed to be  $t_o \geq T_c$  since the system cannot react to the changes faster than  $T_c$ . The size of the network cluster considered should be scaled accordingly to keep the computational complexity under that threshold.

At the DNA macro network, the resulting utility is obtained as  $U = \Gamma \cdot \sum_{i=1}^{\Gamma} U_i$ , where  $\Gamma$  is the reuse factor.

The *inter-cluster* handover may be handled by applying clustering/re-clustering algorithms [13][14] after a change in the traffic occurs.

## Performance Evaluation

In this section, the performance of the network is evaluated through extensive computer simulations conducted using MATLAB. The scenario considered is shown in Figure 4 12, where we assume that  $N$  nodes and  $K$  access points are randomly placed in an area of  $1000 \times 1000$  m<sup>2</sup>. This scenario corresponds to a DNA( $N,K$ ). The results for the DNA macro network can be obtained as explained in the previous section. The simulation parameters are summarized in Table 4 5.

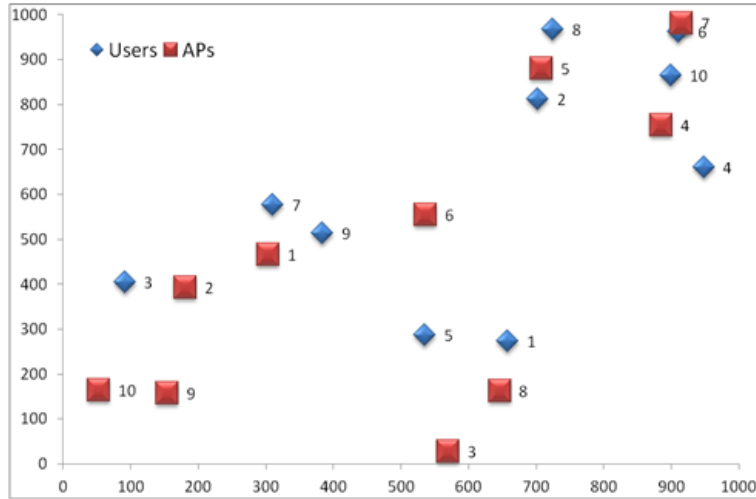


Figure 4-12: Simulation scenario with index of user  $i=1,\dots,N$  and AP  $j=1,\dots,K$ .

Table 4-5: Simulation parameters

Simulation	1000 x 1000
$\alpha$	3
$N_0$	1
$\tau$	10mW
$p_j$	0.5
$\chi_0/Q_0$	1
$R_i/R_{ei}$	1,...,4
$\lambda_m$	0.01calls/s

In Figure 4-13, the utility obtained for the optimum topology is presented as result of the optimization problem defined in (31) when there are  $K = 4$  and 5 APs, and  $N = 6, \dots, 10$  users. The location of those users and APs is shown in Figure 4-12. When  $K=4$  is used, APs with higher index from the figure are inactive. The same principle applies for  $K=5$ . The utility is higher for  $K = 5$  as the transmission power and the number of slots needed to complete the transmission is lower than for  $K = 4$ . The optimization has been solved by MDCS and GA. To obtain the optimum topology by GA, the size of the population and the number of generations was scaled with the size of the DNA. The values used for the previous parameters will be explained later in Figure 4-19.

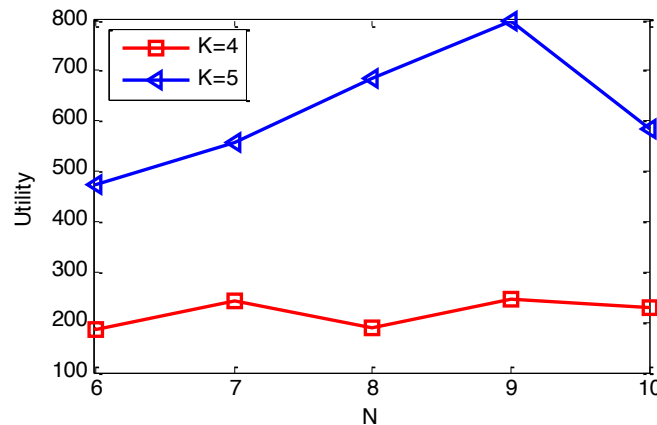


Figure 4-13: Utility defined by (29) for the optimum topology vs.  $N$ .

In Figure 4-14 and Figure 4-15, the number of generated topologies  $G$  and the running time to obtain the optimum topology are shown, respectively, for MDCS and GA algorithm. A processor Intel(R) Core(TM) i5-2400 CPU @3.10GHz with 8GB RAM memory has been used for the simulations.

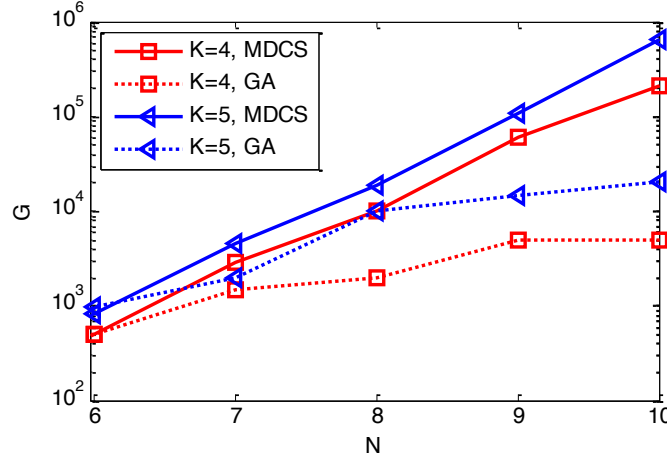


Figure 4-14: Number of topologies  $G$  generated to solve (31) vs.  $N$

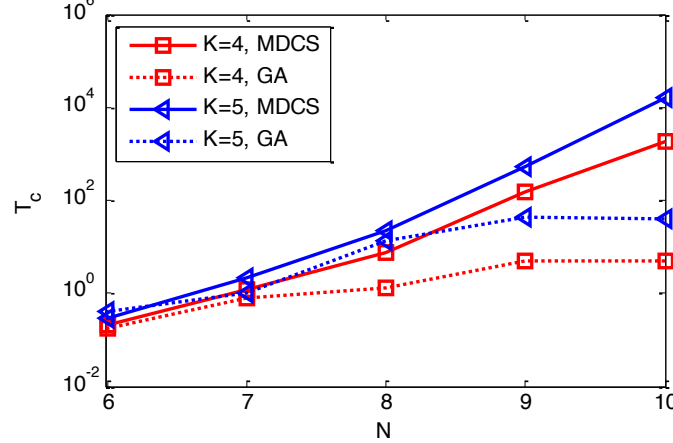
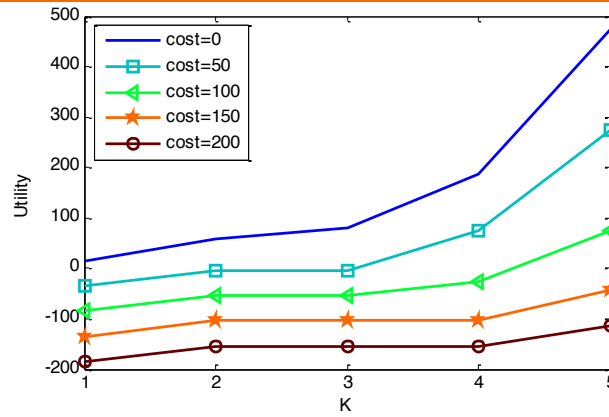


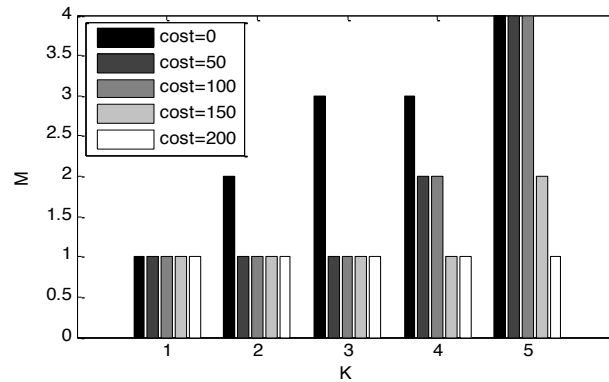
Figure 4-15: Running time  $T_c$  needed to obtain the optimum topology by (31) vs.  $N$ .

As we can see in those figures, for a given value of  $K$ , the improvement obtained by GA increases with  $N$ . It is worth noticing that the number of topologies generated by MDCS,  $G_{\text{MDCS}}$ , and thus the time,  $T_{c\text{MDCS}}$ , linearly increases with  $N$ . By GA, the increase of  $G_{\text{GA}}$  and  $T_{c\text{GA}}$  is more moderate. In particular, for DNA(4,8),  $G_{\text{GA}}$  is one order of magnitude smaller than  $G_{\text{MDCS}}$ . For each new user ( $N=9$ ,  $N=10$ ), one more order of magnitude is obtained as improvement by GA. The running time  $T_{c\text{GA}}$  for the scenarios considered is below 100s, so for a typical value of the arrival rate  $\lambda_m=0.01$  calls/s [15] and the size of the DNA considered, GA can track the changes in the optimum topology.

In Figure 4-16, the utility is shown as result of the optimization problem defined by (34) when the cost = 0, ..., 200. The DNA considered consists on  $N=6$  users and  $K=1, \dots, 5$  available APs. The scenario is shown in Figure 4-12 where users with index higher than 6 are inactive and the same applies for the APs. As expected, the utility is lower for higher values of the cost and is higher for higher  $K$ . In Figure 4-17, the optimum value of  $M$  is obtained for different values of the cost and  $K$ .



**Figure 4-16: Utility defined by (30) for the optimum topology vs.  $K$  and  $N=6$ .**



**Figure 4-17: Optimum  $M$  versus  $K$  for the optimization problem defined by (30).**

The results for the optimization problem (32) are presented in Figure 4-18 for  $N=4, \dots, 6$  and different values of  $K$ . The optimum  $M$  is obtained for each scenario for different values of the QoS constraint,  $\gamma$ . As before, the scenarios considered are shown in Figure 4-12.

In Figure 4-19, the success ratio of GA is shown for the scenarios DNA(7,5) and DNA(10,5) versus the number of generated topologies  $G$ . In this case, we consider that the initial state of network is unknown so this parameter is also an indication of the robustness of GA to the initial state of the network. The initial population consists on a number of feasible topologies randomly generated. The number of topologies generated  $G$  is obtained as the product of the size of the population  $g$  and the number of generations  $N_g$ . In particular, the values of  $g$  and  $N_g$  used in the results are shown in Table 4-6. The simulations have been generated 50 times, and the success ratio is defined as the number of times the optimum topology has been found with respect to the number of runs. We can see that for DNA(7,5),  $G=2000$  topologies are generated where  $g=20$  and  $N_g = 50$  are used to obtain the optimum topology with success ratio 1. In a bigger network, such as DNA(10,5),  $G=8000$  are needed to obtained the optimum one with  $g=40$  and  $N_g = 200$ .

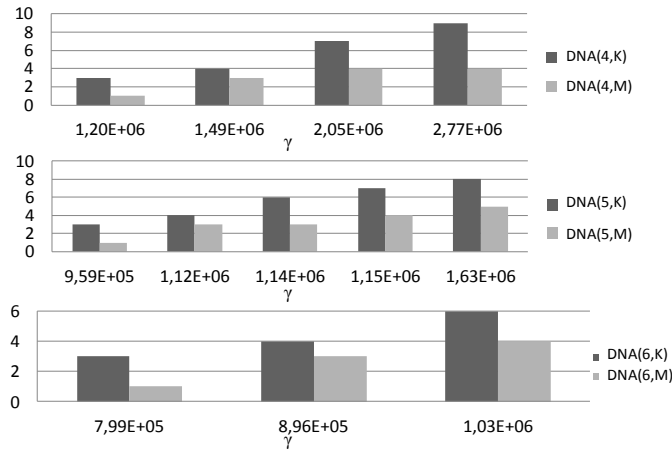


Figure 4-18:  $M$  and  $K$  vs. the QoS constraint  $\gamma$ .

Table 4-6: GA parameters

$g$	$N_g$	$G$
10	50	500
20	50	1000
20	100	2000
30	100	3000
30	200	6000
40	200	8000
50	200	10000
60	200	12000

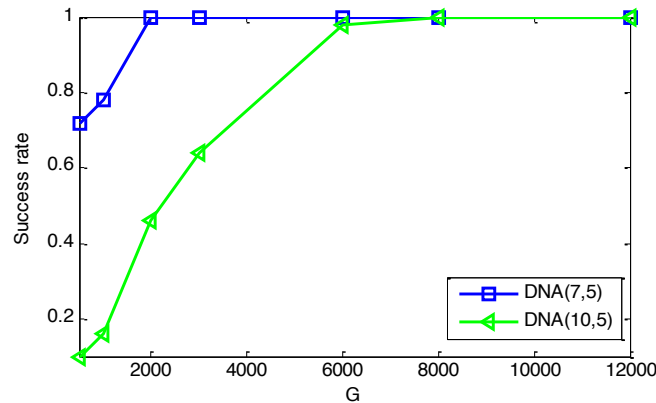


Figure 4-19: Success rate vs.  $G$  when the initial population is randomly chosen.

In Figure 4-20, the Internet contract utilization  $\xi_i$  defined by (47) is shown for different values of  $p_i^l$  and, different ratios of  $R_i$  and  $R_{ei}$ . Both, static and dynamic models for the traffic volume are considered. The scenario used for this simulation is shown in Figure 4-9 for the contract  $T/W(r)/I(Q)$ . The probability of Internet connection was set to  $p^l=0.5$  and the relation between the price  $\chi_0$  and the initial prepaid traffic volume  $Q_0$  has been normalized to 1, so that a monetary unit is charged for each unit of traffic volume. As result, we can see that the utilization  $\xi_i$  increases with  $p_i^l$  and with the rate of user  $i$ ,  $R_i$ . Besides,  $\xi_i$  is larger when the dynamic model is used. The reason for this relies on the fact that the  $price_i^e$  used in the static model is the one obtained at the end of the contract  $T_0$ , so it is higher than for the dynamic model which is calculated in every time  $t$ . As a consequence, the probability that user  $i$  will be elected as AP,  $e_i$ , is lower and thus, there are more chances that the user will transmit its own traffic. For larger  $R_i$ , the higher is the difference between  $\xi_i$  for both models.

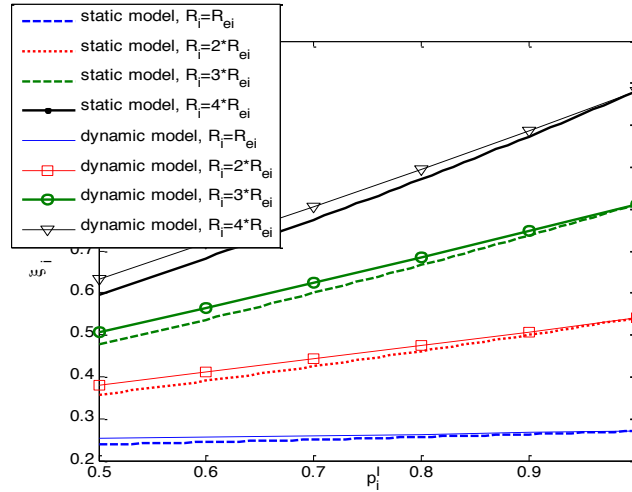


Figure 4-20:  $\xi_i$  vs.  $p_i^I$  for static and dynamic model of traffic volume.

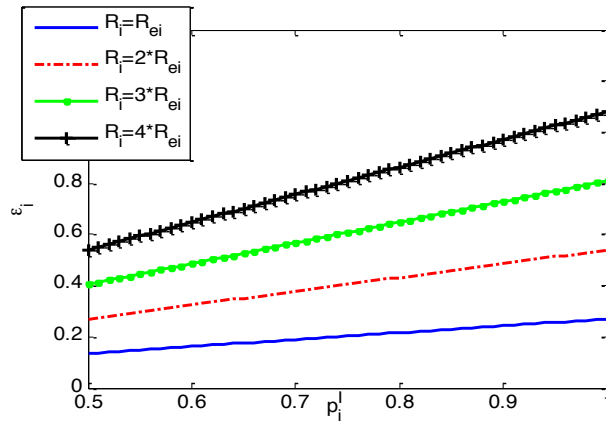


Figure 4-21: Contract price recovery  $\varepsilon_i$  vs.  $p_i^I$ .

The same scenario is used in Figure 4-21 to present the contract price recovery  $\varepsilon_i$  versus  $p_i^I$ . As we can see in the figure,  $\varepsilon_i$  is larger for higher values of  $p_i^I$  and increases with the utilization of user  $i$ .

A set of simulations have been generated to show the effects of the security requirements on parameters  $\xi_i$  and  $\varepsilon_i$ . Those results are not included for space constraints. The conclusions obtained can be easily justified from (50)(51). The higher is the probability that the user  $i$  will satisfy the security requirement of adjacent user  $k$ , the higher will be the probability of being elected as an AP. If the difference between the security level of user  $k$  and the one provided by the AP is very high, further security investment will be needed from the user  $k$  to be accepted by the AP.

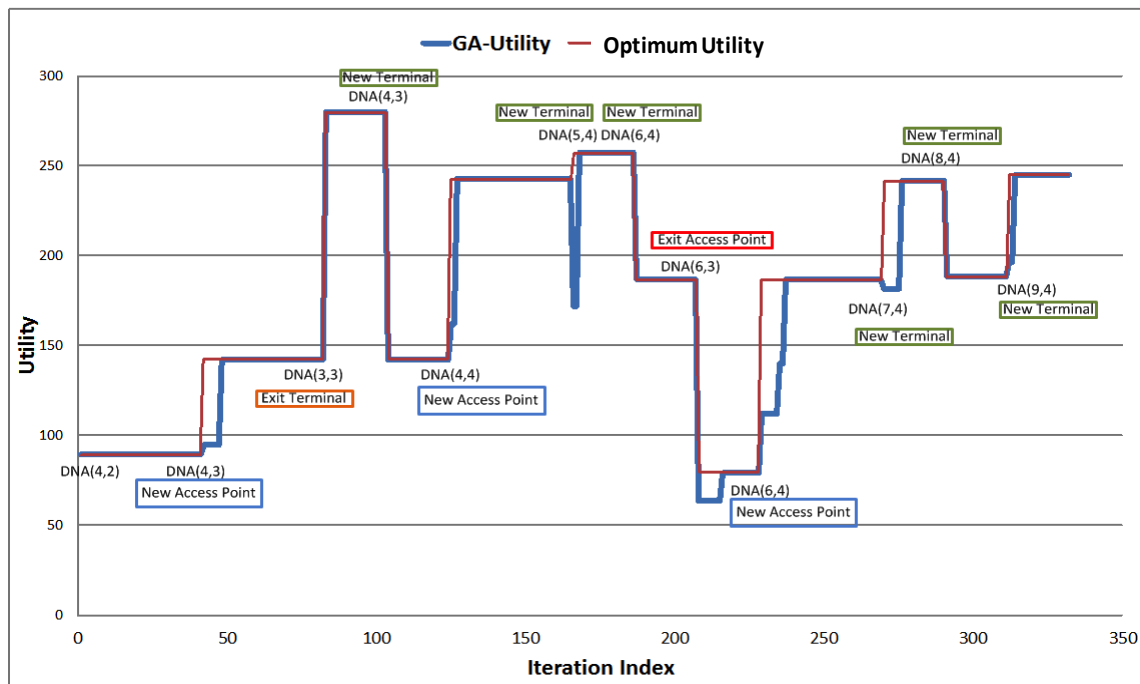
Finally, some results are provided to show the efficiency of GA in a dynamic environment. In Table 4-7, a number of scenarios are presented where the efficiency of GA in tracking the network dynamics is shown. The first column, DNAt represents the current network situation in time  $t$ . In the second column, DNAt+1 describes the network in the next time instant after the topology has changed. The description of the change is outline in the third column. The fourth column presents the value of the optimum utility (fitness) after the traffic changed. Column 5 and 6 show the number of topologies generated by MDCS and GA respectively to obtain the optimum one. The computational time in seconds is shown in the last two columns. We can see

that  $G_{GA} \ll G_{MDCS}$  for almost all scenarios. The only exception is when the size of DNA considered is small (e.g, DNA(4,3)). In that case, as the number of possible combinations is low, the number of topologies generated by MDCS is lower than GA. For the rest of the scenarios, the order of improvement by GA regarding the number of topologies generated and the computational time can reach up to 2 orders of magnitude compared to MDCS. The number of topologies generated by MDCS,  $G_{MDCS}$ , increases exponentially with the size of DNA. The values of  $g$  and  $N_g$  to run the GA in this scenario resulted into a maximum  $G=110$ . For this reason, the last three results of the table which correspond to DNA(7,4), DNA(8,4) and DNA(9,4) have reached that number of topologies. The error in the optimum utility for these scenarios has been in the order of  $10^{-3}$ . It is worth mentioning that for DNA(9,4) the computational time  $T_c=154.052s > 100s$ . So, the size of this DNA should be reduced in order to track the dynamics if the arrival rate is set to  $\lambda_m=0.01$ calls/s. For the same scenarios considered in Figure 4-14 and Figure 4-15,  $G$  and  $T_c$  are now much lower. This is because in these scenarios the previous state of the network, given by topology  $T_0$ , is known (topology reconfiguration) while in Figure 4-14 and Figure 4-15 the optimization was run starting from any random topology.

The representation of the dynamics is shown in Figure 4 22 for the scenarios described in Table 4 7. In this chart, we can see how GA tracks the changes in the network and reaches the optimum value of the utility computed by MDCS.

**Table 4-7: Topology reconfiguration scenarios.**

DNA <sub>t</sub>	DNA <sub>t+1</sub>	Traffic changed	Optimum fitness	$G_{MDCS}$	$G_{GA}$	Computation time(sec) MDCS	Computation time(sec) GA
DNA(4,2)	DNA(4,3)	new AP	142.0225	14	30	0.0520	0.0450
DNA(4,3)	DNA(3,3)	exit user	279.5851	33	5	0.0800	0.0530
DNA(3,3)	DNA(4,3)	new user	142.0225	14	45	0.0520	0.0620
DNA(4,3)	DNA(4,4)	new AP	242.5647	43	15	0.0780	0.0510
DNA(4,4)	DNA(5,4)	new user	257.1448	153	15	0.1	0.0680
DNA(5,4)	DNA(6,4)	new user	186.6262	503	35	0.21	0.0690
DNA(6,4)	DNA(6,3)	exit AP	79.1647	189	65	0.171	0.0630
DNA(6,3)	DNA(6,4)	new AP	186.6262	503	85	0.21	0.0790
DNA(6,4)	DNA(7,4)	new user	241.4456	2919	110	1.2	0.0940
DNA(7,4)	DNA(8,4)	new user	188.144	9996	110	7.48	0.1030
DNA(8,4)	DNA(9,4)	new user	245.1063	58964	110	154.052	0.1090



**Figure 4-22: Dynamic Topology and Architecture reconfiguration scenarios.**

### **Radio Resource Management for Virtual Radio Access Networks.**

The goal of Virtual Radio Resource Management (VRRM) is realising the co-existence of multiple instances of Virtual Radio Access Network (V-RAN) on the same physical infrastructure. Meeting the instances' objectives while optimising the resource usage and the other additional objectives (e.g., guaranteeing fairness among multiple instances) make Virtual Radio Resource Management (VRRM) an elaborated procedure. The importance and complexity of VRRM is the reason to count it as the milestone in realisation of end-to-end virtual radio network.

#### **System model**

The management of virtual radio resources, as it is presented in [16], is a hierarchical management consist of key actors as Virtual Network Operators (VNOs), VRRM, Common Radio Resource Management (CRRM), and local RRM.

At the highest level of this hierarchy, VNOs are placed. These operators, which do not own RAN infrastructure, have to ask for wireless connectivity in term of capacity to serve their subscribers. The SLA between them and infrastructure provider defines the details and quality of offered services.

VRRM, the highest resource manager, is in charge of translating VNOs requirements and Service Level Agreements (SLAs) through sets of polices for lower levels. In other words, the key role of VRRM is to map virtual radio resources into physical ones in addition to issuing policies to manage the virtual resources. VRRM does not handles the physical radio resources, nevertheless, reports and monitoring information (e.g. estimated remained capacity) received from Common Radio Resource Management (CRRM) enable it to improve the policies. It should be reminded that in virtual network, in addition to radio resources, there are processing, capacity, and fibre resource as well. These resource also have to be managed where only management of radio resources are the scope of this framework.

CRRM, in the next level, is the highest resource manager of the heterogeneous access network. It manages the resources of different Radio Access Technologies (RATs) under its control to meet the requests of VRRM. In other words, CRRM has to provide the requested capacity by demanding each RAT to deliver a portion of it. It also optimises its decision based on information from local RRM and reports to VRRM.

Local RRM, in the last level, are liable for optimising radio resource usage in a single access technology. They are in charge of assigning physical radio resource parameters e.g. power, frequency bandwidth, time slots, etc. to the end-users upon receiving request. The policies issued by CRRM for each local RRM is used as decision guidelines for that RRM. In addition to policies of CRRM, the resource allocation in each local RRM has to meet QoS requirement of each service.

In addition to QoS requirement of each service session, VNOs request for a level of service quality from VRRM. Capacity per service, average response time and average serving time are the terms based on which the SLAs can be defined.

The key term is capacity per service. In order to offer a service (e.g. VoIP) to subscribers, a VNO has to ask capacity (with certain QoS requirement) from VRRM. When the capacity is granted to VNO, it can admit UEs requesting that service and VRRM make sure that the virtual wireless links carrying traffic of the service can meet the QoS requirement as long as service is going on.

Average response time, the second metrics, is referred to average of time interval between VNO requesting for virtual link to serve a subscriber's session (e.g. FTP session) and request admission. Using this term, VNO can define average acceptable response time for each service.

At last but not at least, average serving time or its equivalent average data rate, is another term in VNOs' SLAs. Although the admission of subscribers is done by VNO, itself, it has request for virtual resources. Using this term, VNO asks the ability to serve average data rate to subscribers on a service (e.g. FTP). It should be reminded that Average response time and Average data rate are dependent to the number of active sessions. So practically, these terms defined to a maximum number of active sessions.

These SLAs can be summarised into three types of contract, which are:

- Guaranteed Bitrate (GB) in which VNO is guaranteed a minimum as well as a maximum level of service (e.g., data rate) to be served on-demand. The operator, who chooses guaranteed contract, receives the guaranteed capacity regardless of network situation. Subscribers also experience better quality of service.
- Best effort with minimum Guarantees (BG) is the second type of contract in which serving the VNO with a minimum level of service is guaranteed. The request for better services (e.g., higher data rates) is served in a best effort manner. The VNO does not invest as much as operator with guaranteed contract but it can guarantee to its subscriber the minimum quality of service, Subscribers of this type of VNOs experience acceptable network performance but not as good as subscribers in guaranteed VNO. However, it is expected that they pay relatively lower price for their services since the RAN costs less for operator.
- Best effort in which the VNO is served in a pure best effort manner. This type of contract has the lowest cost for operators. However, the operators and consequently, their subscriber may suffer from low service quality during busy hours.

## Estimation of Available Virtual Radio Resources

The goal in the estimation of available virtual radio resources is to obtain a relation in form of PDF from the set of resources to the total network capacity. In general, the applicable data rate by assigning a unit of radio resource (e.g. time slot, radio block or code block) to a UE varies between the minimum and maximum data rate based on various parameters such as Radio Access Technology, modulation, coding schemes, etc. In a certain configuration (e.g., certain modulation) the data rate of a single RRU is function of channel quality, SINR, varies between the minimum and maximum levels as follows:

$$R_{b_{RAT_i}[\text{Mbps}]}(\rho_{in}) \in [0, R_{b_{RAT_i}[\text{Mbps}]}^{max}] \quad (52)$$

where:

- $R_{b_{RAT_i}}$ : data rate of a single RRU of  $i$ -th RAT,
- $\rho_{in}$ : input SINR,
- $R_{b_{RAT_i}}^{max}$ : the maximum data rate of a single RRU of  $i$ -th RAT.

In [17] based on real logs, data rate as function of input SINR and vice versa for various access technologies have been presented. These functions, in next step, for the sake of simplicity, have been approximated to the equivalent polynomial of degree 5 described. Therefore, the SINR as function of data rate is as follows:

$$\rho_{in[\text{dB}]}(R_{b_{RAT_i}}) = \sum_{k=0}^5 a_k \left[ \frac{\text{dB}}{\text{Mbps}^k} \right] R_{b_{RAT_i}[\text{Mbps}]}^k \quad (53)$$

where:

- $a_k$ : factor of polynomial approximation of SINR as function of data in each RAT.

Based on [18], PDF of  $R_{b_{RAT_i}}$  can be achieved as follows:

$$p_{R_b}(R_{b_{RAT_i}[\text{Mbps}]}) = \frac{\frac{0.2}{\alpha_p} \ln(10) \left( \sum_{k=1}^5 k a_k (R_{b_{RAT_i}})^{k-1} \right) e^{-\frac{0.2}{\alpha_p} \ln(10) \sum_{k=0}^5 a_k (R_{b_{RAT_i}})^k}}{e^{-\frac{0.2}{\alpha_p} \ln(10) a_0} - e^{-\frac{0.2}{\alpha_p} \ln(10) \sum_{k=0}^5 a_k (R_{b_{RAT_i}}^{max})^k}} \quad (54)$$

where:

- $\alpha_p$ : the path loss exponent where  $\alpha_p \geq 2$ .

The next step is to extend the model to estimate the overall performance of a single RAT. The total data rate from a single RAT pool is:

$$R_{b_{tot}}^{RAT_i} = \sum_{n=1}^{N_{RRU}^{RAT_i}} R_{b_n}^{RAT_i} \quad (55)$$

where:

- $N_{RRU}^{RAT_i}$ : number of Radio Resources Unite (RRUs) in  $i$ -th RAT,
- $R_{b_{tot}}^{RAT_i}$ : data rate from a  $i$ -th RAT pool,
- $R_{b_n}^{RAT_i}$ : data rate from  $n$ -th RRU of  $i$ -th RAT.

Based on the assumption that channels are independent, random variables of achieved data rates,  $R_{bi}$ , are also independent. The PDF of the total data rate is also equal to the PDF of the sum of all the random variables. Let  $p_{R_{bi}}(R_{b_n}^{RAT_i})$  be PDF of  $i$ -th radio resource unit. Based on [19],  $p_{R_b}(R_{b_{tot}}^{RAT_i})$  is equal to the convolution of all the RRUs' PDFs.

$$p_{R_b}(R_{b_{tot}}^{RAT_i}) = p_{R_{b_1}}(R_{b_1}^{RAT_i}) * p_{R_{b_2}}(R_{b_2}^{RAT_i}) * \dots * p_{R_{b_n}}(R_{b_n}^{RAT_i}) \quad (56)$$

In the deployed heterogeneous access networks of today, the resource pools of RATs can be aggregated under the supervision of CRRM. The total data rate from all the RATs is the summation of the total data rate from each of them.

$$R_{b \text{ [Mbps]}}^{CRRM} = \sum_{i=1}^{N_{RAT}} R_{b_{tot} \text{ [Mbps]}}^{RAT_i} \quad (57)$$

Having a realistic estimation of the total network capacity, a portion of it is assigned to each service of each VNO based on their SLAs and priorities.

## Resource Allocation

After estimating the available resources, VRRM has an estimation about the total network capacity, which is going to be used in the allocation procedure. The key goal in the allocation procedure is to increase the total network throughput while considering the priority of different services of different VNOs and the other constraints. On the ground of this fact, the objective function for VRRM,  $f_{R_b}^v$ , is the total weighted network data rate in Mbps and it can be written as:

$$f_{R_b}^v(R_b) = W^{cell} f_{R_b}^{cell}(R_b^{cell}) + W^{AP} f_{R_b}^{AP}(R_b^{AP}) - W^f f_{R_b}^f(R_b^f) \quad (58)$$

where:

- $R_b$ : vector of serving data rates,

$$R_b = \{R_{b_{ji}}^{srv} | j = 1, 2, \dots, N_{srv} \text{ and } i = 1, 2, \dots, N_{VNO}\} \quad (59)$$

- $f_{R_b}^{cell}$ : objective function for cellular RATs,
- $W^{cell}$ : weight for allocating capacity from cellular RATs,
- $R_b^{cell}$ : vector of serving data rates from cellular network which is:

$$R_b^{cell} = \{R_{b_{ji}}^{cell} | j = 1, 2, \dots, N_{srv} \text{ and } i = 1, 2, \dots, N_{VNO}\} \quad (60)$$

- $W^{AP}$ : weight for allocating capacity from APs,
- $f_v^{AP}$ : objective function for APs,
- $R_b^{AP}$ : vector of serving data rates from APs, which is:

$$R_b^{AP} = \{R_{b_{ji}}^{AP} | j = 1, 2, \dots, N_{srv} \text{ and } i = 1, 2, \dots, N_{VNO}\} \quad (61)$$

- $f_{R_b}^f$ : fairness function,
- $W^f \in [0,1]$  is fairness weight in objective function indicates how much weight should be put on the fair allocation,
- $R_b^f$ : vector of intermediate fairness variable which is:

$$R_b^f = \{R_{b_{ji}}^f | j = 1, 2, \dots, N_{srv} \text{ and } i = 1, 2, \dots, N_{VNO}\} \quad (62)$$

The objective function for cellular RATs addressed in (48) is given by:

$$f_{\mathbf{R}_b}^{cell}(\mathbf{R}_b) = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^{srv} R_{bji}^{cell} [\text{Mbps}] \quad (63)$$

where:

- $N_{VNO}$ : number of VNOs served by this VRRM,
- $N_{srv}$ : number of services for each VNO,
- $W_{ji}^{srv}$ : weight of serving unit of data rate for service  $j$  of the VNO  $i$  by VRRM where  $W_{ji}^{srv} \in [0,1]$ ,

The weights in the equation above are used to prioritise the allocation of data rates to different services of different VNOs. The choice of these weights are done base on the SLAs between VNOs and VRRM, while the summation of all of them is equal to unit (i.e., they are normalised weights). The Services of VNOs, which are classified as background traffic, for instance, have the lowest weight in (63). It is desired that the service with the higher serving weights relatively receive higher data rate than the services with the lower serving weights.

The network architecture and the throughput dependencies are two reasons, based on which the cellular access network and Wi-Fi are considered separately in this work. From network architectural point of view, Access Points (APs) are connected to core network in different way than cellular RATs. The cellular network (i.e., GSM, UMTS, and LTE), considered as 3GPP trusted networks, are connected to S-GW but the Wi-Fi network is considered either as trusted or untrusted non-3GPP networks. It connect to P-GW instead of S-GW.

In addition, the throughput of Wi-Fi network is depended on collision rates and the channel quality while it can be claimed that the throughput in cellular networks is almost independent of the number of connected end-users. Granted that increasing the number of end-user leads to the increment of the interference level, the network throughput still depends only on SINR. In the contention-based networks such as Wi-Fi, however, the increment of connected end-users not only decreases SINR but also increases the conflict rates. Terminals have to retransmit their packets whenever a conflict occurs. Not taking the back-off time interval after each conflict, the retransmission of packets, itself, leads to the lower level of network throughput. Thus, the objective function has to minimise the number of connected users to APs, given by:

$$N_u^{AP} = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} N_{uji}^{AP} \quad (64)$$

where:

- $N_u^{AP}$ : total number users connect to APs,
- $N_{uji}^{AP}$ : number of users of  $i$ -VNO using service  $j$  connected to APs.

Considering average data rate of a service and allocated capacity, the number of connected user is:

$$N_u^{AP} = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{bji}^{AP}}{R_{bj} [\text{Mbps}]} \quad (65)$$

where:

- $R_{bji}^{AP}$ : allocated capacity for service  $j$  of the VNO  $i$  from Wi-Fi network,

- $\overline{R_{bj}}$  : average data rate for service j.

Based on (65), it is obvious that allocating bigger portion of network capacity to services with higher average data rates leads to having the smaller number of users in the network. Hence, there are two key objectives in modelling the traffic offloading capability, which are increment of network throughput and decrement of connected users. Although these two objectives are functions of allocated data rate, they are not the same type. In order to be able to optimise the both objectives at the same time, maximum summation of weighted data, an equivalent objective to minimum number of connected users, is introduced. The weighted data rate of a service in Wi-Fi network is defined by:

$$R_{bji[\text{Mbps}]}^{WAP} = \frac{\overline{R_{bj[\text{Mbps}]}}}{\overline{R_b^{max}[\text{Mbps}]}} R_{bji[\text{Mbps}]}^{AP} \quad (66)$$

where:

- $R_{bji}^{WAP}$ : weighted data rate for service j of the VNO i,
- $\overline{R_b^{max}}$  : maximum average data rate among all the network services (i.e., video streaming).

By normalising the average data rate using the maximum average data rate, the weights in (66) are bound to unit.

$$\frac{\overline{R_{bj[\text{Mbps}]}}}{\overline{R_b^{max}[\text{Mbps}]}} \leq 1 \quad (67)$$

where:

- $\overline{R_b^{max}}$  : maximum average data rate among all the network services (i.e., video streaming),
- $\overline{R_{bj}}$  : average data rate for service j.

Obviously, for two services with the same allocated capacity from Wi-Fi, the one with higher average data rate has the higher weighted data rate and lower number of connected users. Maximising the summation of weighted data rates as function allocated data rate for each service in the Wi-Fi network is equivalent of minimising the number of connected users.

$$\min \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} N_{uji}^{AP} \Leftrightarrow \max \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} R_{bji[\text{Mbps}]}^{WAP} \quad (68)$$

On the ground of this discussion, objective function for APs is:

$$f_{R_b}^{AP}(\mathbf{R}_b^{AP}) = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} W_{ji}^{Srv} R_{bji[\text{Mbps}]}^{AP} + W^{SRb} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \frac{\overline{R_{bj[\text{Mbps}]}}}{\overline{R_b^{min}[\text{Mbps}]}} R_{bji[\text{Mbps}]}^{AP} \quad (69)$$

where:

- $W^{SRb}$ : weight for session average data rate where  $W^{SRb} \in [0, 1]$ .

In (69), the  $W^{SRb}$  is used to control the weight of session average data rate in allocation. The average data rate of services in the allocation procedure with higher values of this weight has relatively higher effect than the same situation with lower weights. Obviously, assigning zero to

this weight, completely eliminates the average data rate effects and converts the objective function of APs addressed in (69) to cellular one presented in (63).

Fairness is the other objective in the allocation procedure. It is desired that services are served according to their serving weights where services with higher serving weights to be allocate comparatively bigger portion of resources. It is neither fair nor desirable to not allocate (or allocate the minimum possible) capacity to services with lower serving weight. The ideal case is when the normalised data rate (i.e., data rate dived by the serving weight) of the all services, and consequently the normalised average, has the same value. This can be expressed as:

$$\frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} = 0 \quad (70)$$

Nevertheless, the resource efficiency and the fair allocations are two contradict goals. The increment in the one of them leads to the decrement of the other one. Hence, instead of having the fairest allocation (i.e., the deviation of the all normalised data rate from the normalised average is zero), the minimisation of the total deviation from the normalised average is used, given by:

$$\min_{R_{bji}^{Srv}} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} R_{bji}^D [\text{Mbps}] \right\} \quad (71)$$

where:

- $f_{R_b}^{fr}$ : fairness objective function,
- $R_{bji}^D$ : max. deviation from the normalised average for service  $j$  of VNO  $i$ , given by:

$$R_{bji}^D [\text{Mbps}] = \left| \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} \right| \quad (72)$$

It is worth noting that the fairness for services with minimum guaranteed data rates, applies only to the amount exceeded than the minimum guaranteed level.

However, the geometrical representation of (71) is the minimum possible distance from the ideal case for each service (i.e., shown by  $R_b^f$ ). Equation (71) can be also reformulated using the epigraph technique [20] as:

$$\begin{aligned} & \min_{R_{bji}^{Srv}, R_{bji}^f} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} R_{bji}^f [\text{Mbps}] \right\} \\ & s. t. \left| \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} \right| \leq R_{bji}^f [\text{Mbps}] \end{aligned} \quad (73)$$

where:

- $R_{bji}^f$  is an intermediate variable used to simplify the problem.

In order to achieve Liniar Programming (LP), (73) has to change as follows:

$$\min_{R_{bji}^{Srv}, R_{bji}^f} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} R_{bji}^f [\text{Mbps}] \right\}$$

$$s. t. \begin{cases} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{Srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} \leq R_{bji}^f [\text{Mbps}] \\ -\frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} + \frac{1}{N_{VNO} N_{Srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \frac{R_{bji}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} \leq R_{bji}^f [\text{Mbps}] \end{cases} \quad (74)$$

As the network capacity increases, the weighted summation of (63) increases. Therefore, to combine it with (74), the two goals of optimisation, the fairness intermediate variable,  $R_{ji}^f$ , has to adopt with the network's capacity as in the objective function for fairness is given by:

$$f_{R_b}^{fr}(\mathbf{R}_b^f) = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \left( \frac{R_{bji}^{CRRM} [\text{Mbps}]}{R_{bji}^{min} [\text{Mbps}]} R_{bji}^f [\text{Mbps}] \right) \quad (75)$$

where:

- $\overline{R_b^{min}}$  is minimum average data rate among all the network services (i.e., VoIP).

The division of network capacity by the minimum average data rate of services gives the maximum possible number of users in the network with given network capacity and service set. By multiplying the fairness variable by the maximum number of users, the balance of these two objectives (i.e., the network throughput and the fairness) can be kept.

In addition, there are more constraints for VRRM to allocate data rates to various services, which should not be violated. The very fundamental constraint is the total network capacity estimated in last section. The summation of the entire assigned data rates to the all services should always be smaller than the total estimated capacity of the network given by:

$$\sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} R_{bji}^{Srv} [\text{Mbps}] \leq R_b^{CRRM} [\text{Mbps}] \quad (76)$$

The offered data rate to the guaranteed and the best effort with min. guaranteed services imposes the next constraints. The allocated data rate related to these services have to be higher than minimum guaranteed level (for guaranteed and best effort with min. guaranteed) and lower than maximum guaranteed (for guaranteed services only):

$$R_{bji}^{Min} [\text{Mbps}] \leq R_{bji}^{Srv} [\text{Mbps}] \leq R_{bji}^{Max} [\text{Mbps}] \quad (77)$$

where:

- $R_{bji}^{Min}$ : minimum guaranteed data rate service  $j$  of the VNO  $i$ ,
- $R_{bji}^{Max}$ : maximum guaranteed data rate service  $j$  of the VNO  $i$ .

In (77), the allocated data rate for a specific service is defined as:

$$R_{bji}^{Srv} [\text{Mbps}] = R_{bji}^{cell} [\text{Mbps}] + R_{bji}^{AP} [\text{Mbps}] \quad (78)$$

Due to changes in the physical infrastructure, users' channel, etc., in practice, there are situation where the resources are not enough to meet all guaranteed capacity and the allocation

optimisation, as addressed in previous sections, is no longer feasible. A simple approach in these cases is to relax the constraints by introduction of violation (also known as slack) variables. The new objective function contains the objective function of the original problem plus the penalty for the violations. In case of VRRM, the relaxed optimisation problem can be considered by adding a violation parameter to (77) given by:

$$\begin{aligned} R_{bji}^{Min} [Mbps] &\leq R_{bji}^{Srv} + \Delta R_{bji}^v [Mbps] \\ \Delta R_{bji}^v &\geq 0 \end{aligned} \quad (79)$$

where

- $\Delta R_{bji}^v$ : violation variable for minimum guaranteed data rate service  $j$  of the VNO  $i$ .

By introducing the violation parameter, the former infeasible optimisation problem turns into a feasible one. The optimal solution maximises the objective function and minimises the average constraints violations. The average constraints violation is defined as follows:

$$\overline{\Delta R_b^v} [Mbps] = \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^v \Delta R_{bji}^v [Mbps] \quad (80)$$

where

- $\overline{\Delta R_b^v}$ : average constraint violation,
- $W_{ji}^v$ : weight of violating minimum guaranteed data rate service  $j$  of the VNO  $i$  where  $W_{ji}^v \in [0,1]$ .

The objective function presented in (58) also has to be altered. The new objective function, the relaxed one, has to contain the minimisation of violations in addition to the maximisation of former objectives. Although the average constraint violation has direct relation with the allocated data rate to services where the increment in one leads to decrement of the other one, it does not have the same relation with fairness. It can be claimed that the maximisation of fairness and minimisation of constraints violations are independent. Therefore, the final objective function considering both issues has to consider the same approach for minimisation of the violations the same as fairness. In better words, the fairness variable is weighted as it is presented in (76) to compensate with summation of weighted data rate of various services. The derivation from fair allocation, which is desired to be as minimum as possible, gain relatively higher weighted in the objective function and may confiscate the constraints violation strategies. Therefore, minimising the average constraint violation in addition to maximising the fairness and the network's weighted throughput have to be considered together as follows:

$$f_{R_b}^v(\mathbf{R}_b) = W^{AP} f_{R_b}^{AP}(\mathbf{R}_b^{AP}) + W^{cell} f_{R_b}^{cell}(\mathbf{R}_b^{cell}) - f_{R_b}^{vi}(\overline{\Delta R_b^v}) - W^f f_{R_b}^f(\mathbf{R}_b^f) \quad (81)$$

where  $f_{R_b}^{vi}$  is the constraint violation function:

$$f_{R_b}^{vi}(\overline{\Delta R_b^v}) = \frac{R_b^{CRRM} [Mbps]}{R_b^{min} [Mbps]} \overline{\Delta R_b^v} \quad (82)$$

The definition of fairness, however, in congestion situation is not the same. The fairness objective in the normal case is to have the same normalised data rate for all the services. As a reminder, when the network faces with congestion, there are not enough resources to serve all the services with the minimum acceptable data rates. Therefore, not only all best effort services are not allocated any capacity, some violation is also introduced to the guaranteed data rates.

Obviously, the violation data rates for the best effort services are always zero. Consequently, (74) is altered as follows:

$$\begin{cases} \frac{W_{ji}^v \Delta R_{bji}^v [\text{Mbps}]}{W_{ji}^{Usg}} - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{W_{ji}^v \Delta R_{bji}^v [\text{Mbps}]}{W_{ji}^{Usg}} \leq R_{bji}^f [\text{Mbps}] \\ -\frac{W_{ji}^v \Delta R_{bji}^v [\text{Mbps}]}{W_{ji}^{Usg}} + \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{W_{ji}^v \Delta R_{bji}^v [\text{Mbps}]}{W_{ji}^{Usg}} \leq R_{bji}^f [\text{Mbps}] \end{cases} \quad (83)$$

### VRRM Interaction with CRRM and Local RRM

Although each local RRM works independent of the other, one, CRRM can coordinate them through the set of policies. The main objective of these policies is to consider the problem of allocating RRUs from access technique point of view with various criteria such as load balancing among resource pools. In brief, it can be said that CRRM is in charge of RAT selection and vertical handover. The key part of objective function for the CRRM can be written as follows:

$$f_{\mathbf{R}_b}^c(\mathbf{R}_b) = \sum_{i=1}^{N_{VNet}} \sum_{j=1}^{N_{srv}} \sum_{k=1}^{N_{RAT}} (W_c^{srv} R_{bji}^k [\text{Mbps}]) \quad (84)$$

where:

- $W_c^{srv}$ : Weight of serving unit of data rate by CRRM,
- $R_{bji}^k$ : Data rate served for service j of the VNO i from RAT k.

CRRM applies techniques such as load balancing among different RAT resource pools to maximise the objective function under some constraints. The main constraint is the limited RRUs. The summation of all data rates assigned to all services of all VNOs from a RAT resource pool should always be less than or equal to expected total data rate for that RAT.

$$\sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} R_{bji}^k [\text{Mbps}] \leq \tilde{R}_{bT}^k [\text{Mbps}], k = 1, \dots, N_{RAT} \quad (85)$$

where  $\tilde{R}_{bT}^k$  is expected total data rate for the RAT k.

Furthermore, the CRRM has to abide VRRMs policy set which imposes some more constraints. Starting with data rate, the total data rate offered to a service of a VNO from various RATs should be equal or higher than requested rate asked by VRRM,  $R_{bji}^{srv}$ .

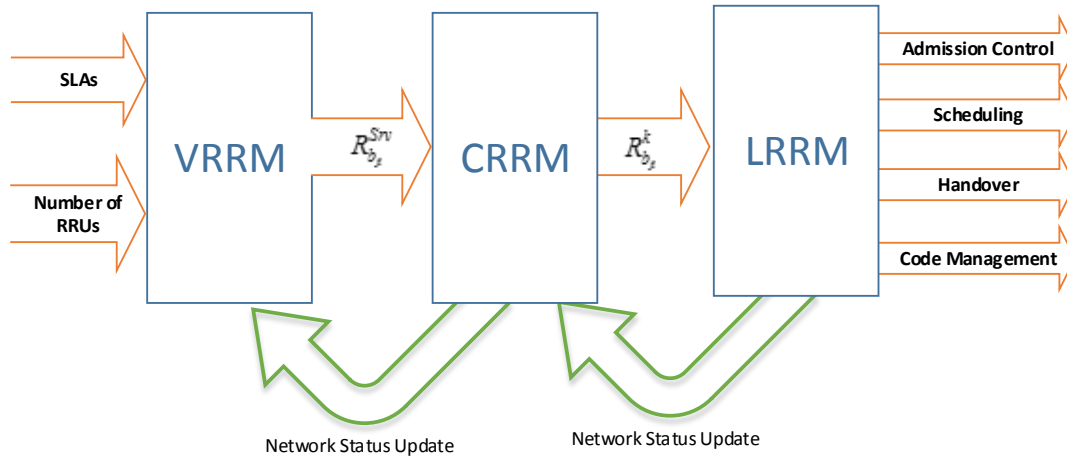
$$R_{bji}^{srv} [\text{Mbps}] \leq \sum_{k=1}^{N_{RAT}} R_{bji}^k [\text{Mbps}] \quad (86)$$

CRRM, like local RRM, provides higher level, VRRM, with periodic reports. The report of CRRM contains general information on QoS metrics and estimated remaining data rate.

VRRM, receiving these reports, tries to allocate the available capacity to various services of various VNOs. The allocation should be done so that the guaranteed level of service to be met.

Concluding this section, Figure 4-23 presents the relation among the three key management entity. VRRM on the first step uses the SLAs and the available physical radio resource to allocated data rates to the services of the VNOs. These data rate are passed to CRRM as the

management policies to help it in its decisions. CRRM has to provide these data rate through different RATs and balance the load among them. It sends its decisions, which is data rate from each RAT for each service, to each local RRM. The local RRM receiving the information, coordinates its functionality (e.g., admission control and scheduling) to satisfy the requested data rate.

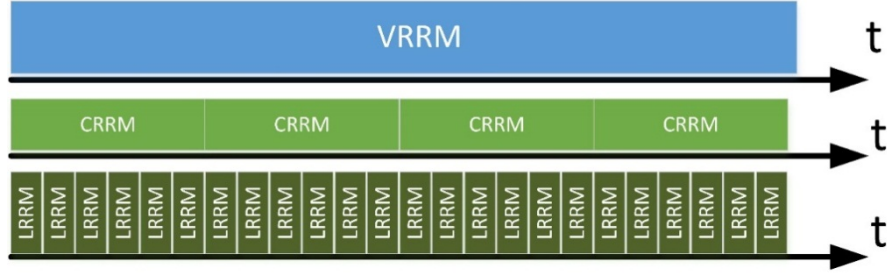


**Figure 4-23: VRRM, CRRM, and LRRM relation.**

### Partial VRRM

The former section presented the modelling of virtual radio resource management addressing the objectives in form of optimisation problem. Since the network status and constraints varies during the time, this problem is an uncertain non-convex optimisation problem. However, there are techniques to solve these kinds of problems where “Partial VRRM” among various approaches is one of the simplest one. In this approach, the optimisation problem is tackled by breaking the main problem into multiple sub-problems. In the case of virtual radio resource management, time axis is divided into decision windows and the objective function is maximised in each of these intervals, independently. However, it is worth noting that decisions in each interval affect directly on network state and the outcome of a policy at a certain point is dependent of the decisions and states in former intervals. The optimal solution has to take this dependency of decisions also into consideration. On the ground of this discussion, the output of partial VRRM may be only a local minimum and not the global one. Nevertheless, the partial VRRM is a simple solution, which can be used as the starting step and reference point. Outcome of the more sophisticated approaches, then, can be compared to these results.

Figure 4-24 illustrates decision window of VRRM, CRRM, and LRRMs. The VRRM decision window contains multiple CRRM ones during which CRRM applies the decided policy set. On the next decision window of VRRM, after multiple network stages, the VRRM updates network situation and makes the new decision for the next time interval.



**Figure 4-24: Decision window of VRRM and CRRM.**

Consider the objective function presented in previous sections, it can be considered as summation of objective functions over time intervals in partial VRRM as follows:

$$f_{\mathbf{R}_b}^v(\mathbf{R}_b) = \sum_{t_i=0}^{n_t} f_{\mathbf{R}_b}^v(\mathbf{R}_b[t_i]) \quad (87)$$

where:

- $n_t$ : number decision window,
- $\mathbf{R}_b[t_i]$ : allocated data rate vector at time interval  $t_i$
- $R_{bji}^{Srv}[t_i]$ : Served (allocated) data rate for service  $j$  of the VNO  $i$  at the time frame  $t_i$ ,

The goal is to maximise the objective function for all running time. Considering objective function in each time interval independent of the other time intervals, it can be written:

$$\max_{\mathbf{R}_b} f_{\mathbf{R}_b}^v(\mathbf{R}_b) \Leftrightarrow \sum_{t_i=0}^n \max_{\mathbf{R}_b[t_i]} f_{\mathbf{R}_b}^v(\mathbf{R}_b[t_i]) \quad (88)$$

As the VRRM starts to work, it has to gather information about serving VNOs plus their services and SLAs. This stage is referred as “Network Initiation” in the Figure 4-25. Estimating total systems (V-RAN) capacity is the next step. VRRM uses the information about number of available RRUs in different RATs and estimate the PDF of (57) using (56). The allocation of the available capacity to VNOs services is done by solving Linear Programming (LP) problem. In each decision window, the network status has to be update. The network status at time interval  $t_i$  contains the following items:

- Remained network capacity,  $R_{b[\text{Mbps}]}^{CRRM}[t_i]$ , which is the data rate achievable from unassigned RRUs.
- Freed service data rate,  $R_{bji[\text{Mbps}]}^{fr}[t_i]$ , which is referred to RRUs freed by termination of service  $j$  of VNO  $i$  in time interval  $t_i$ . It should be reminded that this value is considered from the estimation of data rate based on number of freed RRUs. The actual data rate offered to the subscribers may not be the same value.
- On use service data rate,  $R_{bji[\text{Mbps}]}^{Us}[t_i]$ , which is referred to RRUs exhausted by service  $j$  of VNO  $i$  in time interval  $t_i$ . Like the freed service data rate, this value also is achieved based on the estimation of data rate of used RRUs.
- Actual freed service data rate,  $R_{bji[\text{Mbps}]}^{Afr}[t_i]$ , which is referred to actual data rate of service  $j$  of VNO  $i$  terminated in this time interval.

- Actual on used service data rate,  $R_{bji}^{AUS}[t_i]$ , which is the actual data rate exhausted by ongoing service  $j$  of VNO  $i$ .

According to these definitions the allocated data rate to for each service in each decision is:

$$R_{bji}^{Srv}[t_{i+1}] = R_{bji}^{Srv}[t_i] + R_{bji}^{fr}[t_i] - R_{bji}^{US}[t_i] \quad (89)$$

Using the updated information of network status, the new network capacity in time interval  $i+1$  is defined as follows:

$$R_b^{CRRM}[t_{i+1}] = R_b^{CRRM}[t_i] + \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} (R_{bji}^{fr}[t_i] - R_{bji}^{US}[t_i]) \quad (90)$$

Relatively, the minimum and maximum guaranteed data rate on update as follows:

$$R_{bji}^{Min}[t_{i+1}] = R_{bji}^{Min}[t_i] + R_{bji}^{Afr}[t_i] - R_{bji}^{AUS}[t_i] \geq 0 \quad (91)$$

$$R_{bji}^{Max}[t_{i+1}] = R_{bji}^{Max}[t_i] + R_{bji}^{Afr}[t_i] - R_{bji}^{AUS}[t_i] \geq 0 \quad (92)$$

The maximum and minimum guaranteed data rate, obviously, cannot be less than zero. A service with zero (or negative) minimum guaranteed data is going to be served in best effort manner and the one with zero maximum guaranteed data is no longer going to be served. Using this technique, VRRM algorithm repeatedly optimises the objective function presented in (58) as follows:

$$\begin{aligned} & \max_{\mathbf{R}_b^{cell}[t_i], \mathbf{R}_b^{AP}[t_i], \mathbf{R}_b^f[t_i]} \left\{ W^{AP} f_{\mathbf{R}_b}^{AP}(\mathbf{R}_b^{AP}[t_i]) + W^{cell} f_{\mathbf{R}_b}^{cell}(\mathbf{R}_b^{cell}[t_i]) - W^f f_{\mathbf{R}_b}^f(\mathbf{R}_b^f[t_i]) \right\} \\ \text{s. t. : } & \begin{cases} R_{bji}^{Srv}[t_i] = R_{bji}^{cell}[t_i] + R_{bji}^{AP}[t_i] \\ 1^T \mathbf{R}_b[t_i] \leq R_b^{CRRM}[t_i] \\ R_{bji}^{Min}[t_i] \leq R_{bji}^{Srv}[t_i] \\ R_{bji}^{Srv}[t_i] \leq R_{bji}^{Max}[t_i] \\ \frac{R_{bji}^{Srv}[t_i]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{Srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \frac{R_{bji}^{Srv}[t_i]}{W_{ji}^{Srv}} \leq R_{bji}^f[t_i] \\ \frac{-R_{bji}^{Srv}[t_i]}{W_{ji}^{Srv}} + \frac{1}{N_{VNO} N_{Srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{Srv}} \frac{R_{bji}^{Srv}[t_i]}{W_{ji}^{Srv}} \leq R_{bji}^f[t_i] \end{cases} \quad (93) \end{aligned}$$

For the resource shortage situations, the objective function is (81) and the optimisation is:

$$\max_{\mathbf{R}_b^{cell}[t_i], \mathbf{R}_b^{AP}[t_i], \mathbf{R}_b^f[t_i], \Delta \mathbf{R}_b^v[t_i]} \left\{ W^{AP} f_{\mathbf{R}_b}^{AP}(\mathbf{R}_b^{AP}[t_i]) + W^{cell} f_{\mathbf{R}_b}^{cell}(\mathbf{R}_b^{cell}[t_i]) - f_{\mathbf{R}_b^v}^{vi}(\Delta \mathbf{R}_b^v[t_i]) - W^f f_{\mathbf{R}_b}^f(\mathbf{R}_b^f[t_i]) \right\} \quad (94)$$

$$s. t: \left\{ \begin{array}{l} R_{b_{ji}[\text{Mbps}]}^{Srv}[t_i] = R_{b_{ji}[\text{Mbps}]}^{cell}[t_i] + R_{b_{ji}[\text{Mbps}]}^{AP}[t_i] \\ 1^T \mathbf{R}_b[t_i] \leq R_b^{CRRM}[\text{Mbps}][t_i] \\ R_{b_{ji}[\text{Mbps}]}^{Min}[t_i] \leq R_{b_{ji}[\text{Mbps}]}^{Srv}[t_i] \\ R_{b_{ji}[\text{Mbps}]}^{Srv}[t_i] \leq R_{b_{ji}[\text{Mbps}]}^{Max}[t_i] \\ 0 \leq \Delta R_{b_{ji}}^v \\ W_{ji}^v \Delta R_{b_{ji}[\text{Mbps}]}^v[t_i] - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^v \Delta R_{b_{ji}[\text{Mbps}]}^v[t_i] \leq R_{b_{ji}[\text{Mbps}]}^f[t_i] \\ -W_{ji}^v \Delta R_{b_{ji}[\text{Mbps}]}^v[t_i] + \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^v \Delta R_{b_{ji}[\text{Mbps}]}^v[t_i] \leq R_{b_{ji}[\text{Mbps}]}^f[t_i] \end{array} \right.$$

In the provisioning phase, system is in initial status and VRRM outputs leads to set of policies for system runtime. In this situation, VRRM divides the total network capacity among the services using (58) when all RRUs are available and there is no on used data rate. As the requests arrive and network starts to work, system states (e.g. remaining RRUs and served data rate) will change according to (89)-(92). The algorithm updates network capacity estimation and guaranteed data rates. Then it solve again optimisation problem of (58) again. The outcome is sent for CRRM as the policy update. Figure 4-25 presents partial VRRM flowchart in provisioning and runtime of virtual RAN.

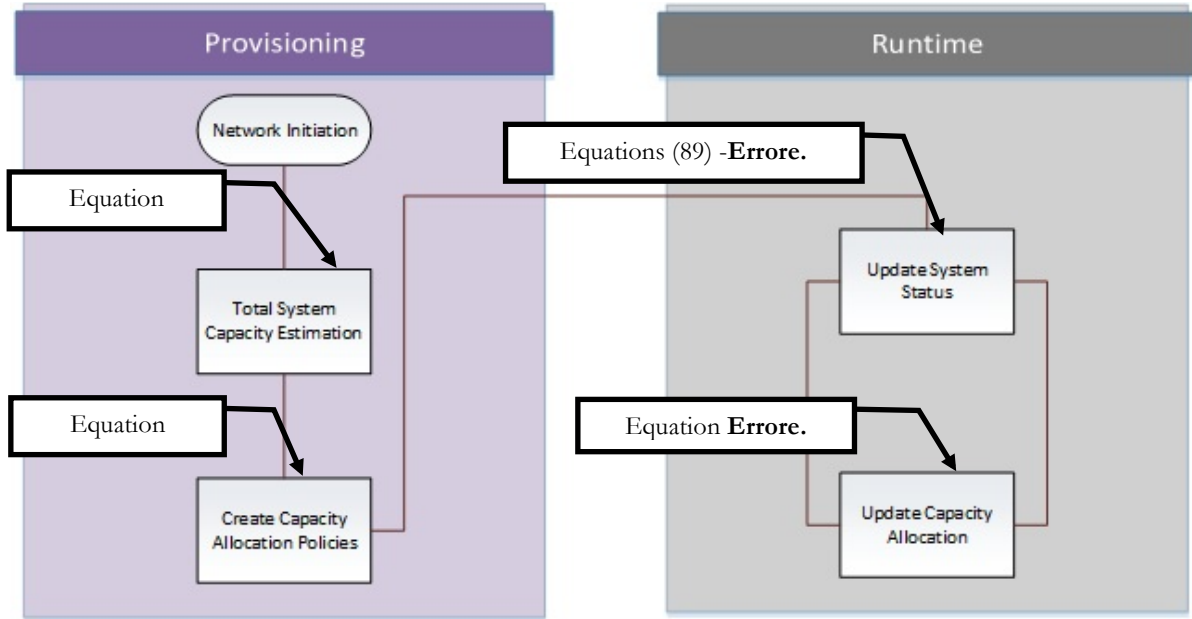


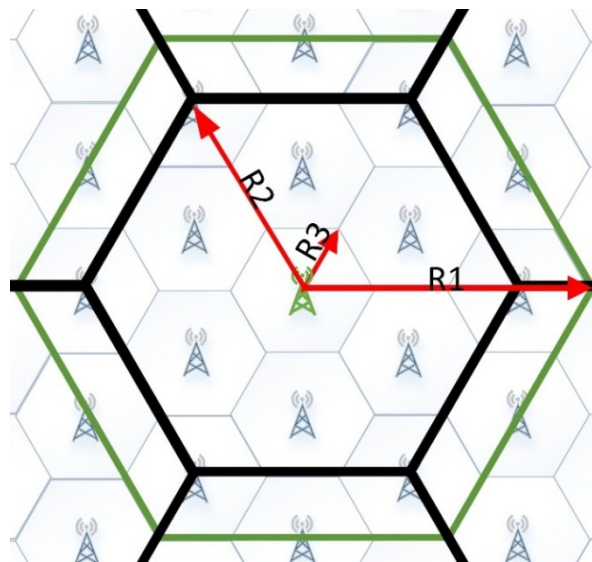
Figure 4-25: Partial VRRM flowchart.

### Reference Scenarios - Urban Hotspot

The performance of VRRM is simulated and evaluated in an urban hotspot scenario. This section describes this reference scenario, based on which other variation of scenarios (e.g., different number of user) are considered. The key parts of this scenario are RATs, subscribers, services, and VNOs.

Regarding to the RATs, both cellular RAN and Wi-Fi APs are considered in this scenario. The RAN is based on a set of RRHs capable of supporting multiple RATs, which are OFDMA (e.g. LTE), CDMA (e.g. UMTS), and FDMA/TDMA (e.g. GSM). Although flexibility of these RRHs offers various cell layouts, the configuration illustrated in Figure 4-26, is used.

- The OFDMA RAT assumption is based on the 100 MHz LTE-Advanced. The cells of this RAT are the smallest ones with the radius of 400 meters. Each cell has 500 RRUs, which can be assigned for traffic bearers.
- The configurations for CDMA cells are chosen according to UMTS (HSPA+) working on 2.1 GHz. Each cell of this RAT with the radius of 1.2 km has 3 carriers and each carrier has 16 codes. Only 45 codes out of all 48 codes in each cell can be assigned to users' traffic.
- The biggest cell size with the radius of 1.6 km is configured for FDMA/TDMA. Based on GSM900, each cell has 10 frequency channels and each channel has 8 timeslots. It is assumed that 75 timeslots out total 80 available timeslots in each cell can be used for to users' traffic.



**Figure 4-26: Network Cell Layout (R1=1.6 km, R2=1.2 km, R3=0.4km)**

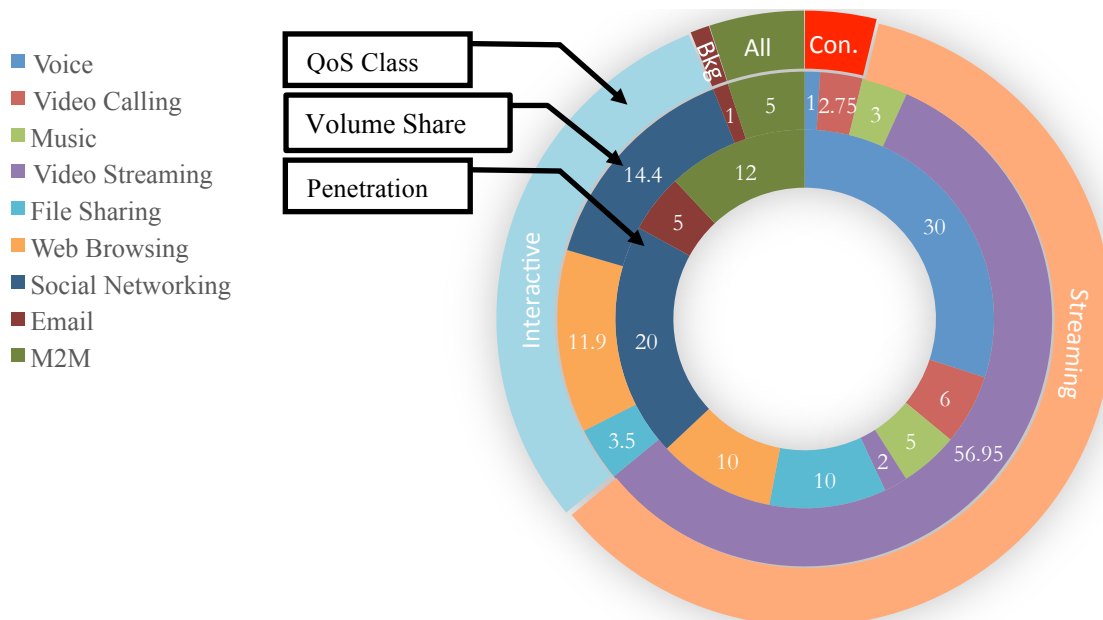
In addition to cellular access network, in this scenario full coverage of Wi-Fi (OFDM) is provided by means of IEEE802.11ac standard APs, configured to work 80 MHz channel bandwidth. It is assumed that each access point covers a cell with radius 80 meters and they are facilitated with beamforming and MU-MIMO technology to support up to eight spatial streaming. Due to European Union rules and regulations, there are only five available channels for 80 MHz Aps [21]. In contrast to former RATs, the Wi-Fi APs use the same set of links for upload and download streams. To achieve coherency among various RATs, the total throughput of APs is equally divided between downlink and uplink. Therefore, in Table 4-8, where coverage information is summarised, the number of RRUs in each Wi-Fi cell is indicated as half of total number of available channels. This table also presents the maximum data rate for each RAT in downlink. In this scenario, it is assumed that only 5% of W-Fi AP can be used for mobile subscribers and the rest belongs to private fixed users. Regarding the number APs, a derivation of this reference scenario is also considered, in which the APs are only deployed on the OFDM base station and there is not full coverage. However, in this case, the entire Wi-Fi capacity can be used for traffic offloading.

Table 4-8: Different RAT cell radius (based on [21]).

RAT	Number Cells	Cell Radius [km]	System	Downlink				
				Number of RRU/Cell	$N_{RRU}^{RAT_i}$	$R_{b_{RAT_i}}^{max}$ [Mbps]	$R_{b_{tot}}^{RAT_i}$ [Mbps]	$R_{b_{tot}}^{RAT_i}$ [Gbps]
OFDMA	16	0.4	LTE	500	8000	0.7	5600	5.47
CDMA	1.7	1.2	UMTS	45	80	43	3440	3.36
TDMA	1	1.6	GSM	75	75	0.05833	4.37	4.37 [Mbps]
OFDM	420	0.08	Wi-Fi	2.5	1024	1300	1300 [Tbps]	50.78

The terminals of subscribers can be smartphone, tablets, laptops, and M2M devices. According to [22] and [23], the average subscribers' traffic is related to the type of terminals they use as it is presented in **Errore. L'origine riferimento non è stata trovata.** Using this table, the VNOs' contracted capacity can be estimated based on number of subscribers. In this reference scenario, it is assumed that all the UEs are smart phones. . Based on [24], each smart phone terminal requires an average speed of 6.375 Mbps.

The service set in this scenario is assumed based on Figure 4-26. The volume share of services, which is presented in outer circle in the figure, is the percentage of the specific service traffic volume from total operator's traffic. The service penetration, however, is the percentage of active subscribers using that service. These charts clearly present that the video streaming holds the highest volume where VoIP is the most requested service. These charts may be altered where different type of terminals mixture (and consequently service profile) is chosen. Table 4-9 present more details on these services and their QoS requirements.



**Figure 4-27: Various service volume share and penetration.**

Three VNOs are assumed to operate in this area, VNO GB, BG, and BE. All of these VNOs offer the same set of services to their users. These services and their volume share of an operator traffic is listed in Table 4-7. The order of service weights base on general service classes are conversational (0.4), streaming (0.3), interactive best effort (0.2), and best effort (0.05). To not compromise the objective function for having higher fairness, the fairness weight,  $W^f$ , is heuristically chosen to be equal to the lowest serving weight (0.05).

Each VNO have 300 subscribers where the average data rate for each of them is 6.375 Mbps [23]. Hence, the contracted data rate for whole operator is 1912.5 Mbps and each service receives a portion based on the volume percentage in Table 4-9. The SLAs of these VNOs are summarised as follows.

- VNO GB, the allocated data rates for services are guaranteed to be between 50% - 100% of the service data rate.
- VNO BG has best effort with minimum 25% of service data rate guaranteed SLA.
- Services of VNO BE are served all in best effort manner

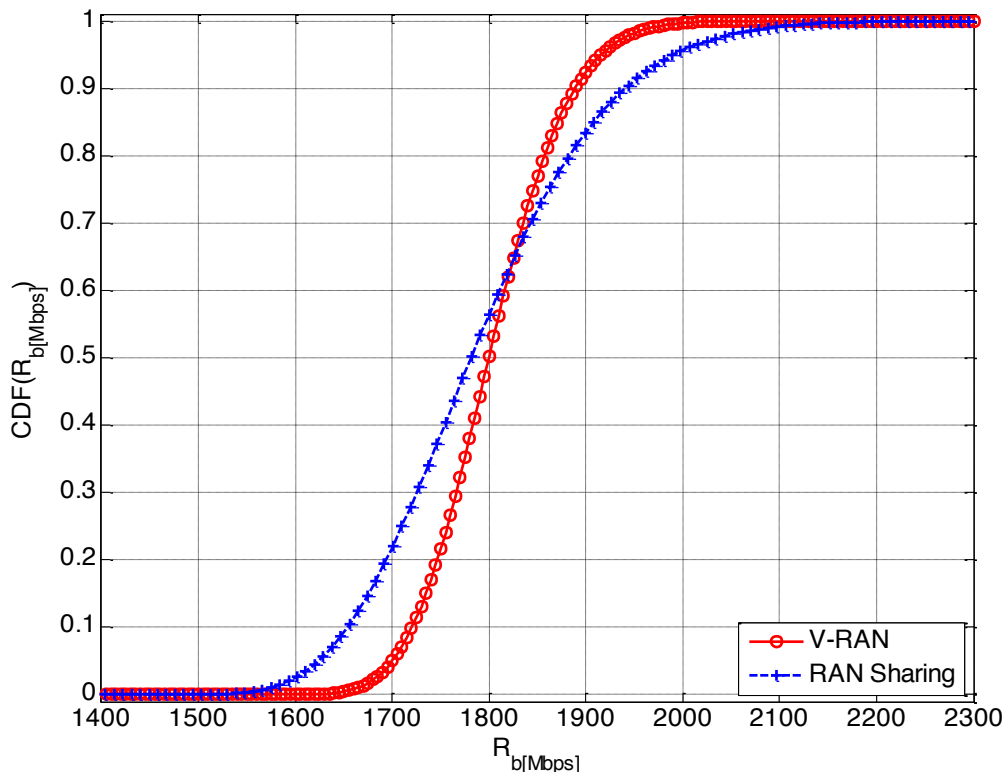
**Table 4-9: Services characteristics for multi-service VNOs based on [25].**

Service		Service Class	Max Data Rate [kbps]			Duration [s]		Size[kB]
			Min.	Average	Max.			
Voice		Conversational	5.3	12.2	64	60		–
Music		Streaming	16	64	160	90		–
File Sharing		Interactive	384	1024	–	–		2042
Web Browsing		Interactive	30.5	500	–	–		180
Social Networking		Interactive	24	384	–	–		45
Email		Background Traffic	10	100	–	–		300
M2M	Smart Meters	Background Traffic	–	200	–	–	–	2.5
	e-Health	Interactive	–	200	–	–	–	5611.52
	ITS	Conversational Real-Time	–	200	–	–	–	0.06
	Surveillance	Streaming Real Time	64	200	384	–	–	5.5
Mobile Video	Video Calling	Conversational Real Time	64	384	2048	60		–
	Video streaming	Streaming Real Time	500	5120	13000	3600		–

## Results

To estimate total data rate of the network described in last section, PDF of different RATs as it is presented in (57), is calculated. By means of implementation of numeric convolution, the PDF and CDF of data rate for total network are achieved. Besides, to have a comparison between virtualisation of radio resources with RAN sharing concept, CDF of total network for RAN sharing and V-RAN approach is illustrated in Figure 4-.

In the V-RAN, all resources are aggregated where in the RAN sharing approach, each operators has on third of resources and total network data rate is three times of a single operator. It can be seen that it more probable to serve all three operator when the resources are aggregated than when the resources are divided. For instance, for 50% of time (when CDF is equal to 0.5) the total network capacity in V-RAN is 1799.62 Mbps where RAN sharing offers 1782 Mbps. The highest difference can be seen where CDF is equal to 0.1. In this case, the relative data rate for V-RAN is 1725 Mbps where RAN sharing offers only 1656 Mbps.



**Figure 4-28: CDF of network capacity for V-RAN and RAN sharing**

In the next step, allocation of resources is done for the case where CDF is equal to 0.5. The VNO GB received the biggest portion (59%) of resources since it is the VNO with the guaranteed SLA. VNO BE, the VNO to be best effort manner, is allocated only 7% of capacity. The rest of resources (34%) are assigned to VNO BG so that it is served better than the minimum guaranteed.

The optimisation problem stated in (93) is solved by MATLAB linear programming problem solver (i.e., linprog function) [25]. The results for various services of different VNOs, based on [16], are presented in **Errore. L'origine riferimento non è stata trovata.** As it can be seen, VRRM manages to meet the guaranteed level of services. Based on serving weight, as it was

expected, the conversational services (e.g. VoIP and Video call) have the highest data rate while background services (e.g. Email and Smart Meter) have the lowest data rate. The M2M services present this issue better since they all have the same volume percentage but different serving weight. Services of VNO GB and BG have minimum guaranteed level so the assigned data rates cannot go below a certain level. This is the reason for having high data rate assigned to video streaming in these two VNOs. In general, it can easily be seen that the same service in the VNO GB and BG received more capacity in comparing to VNO BE.

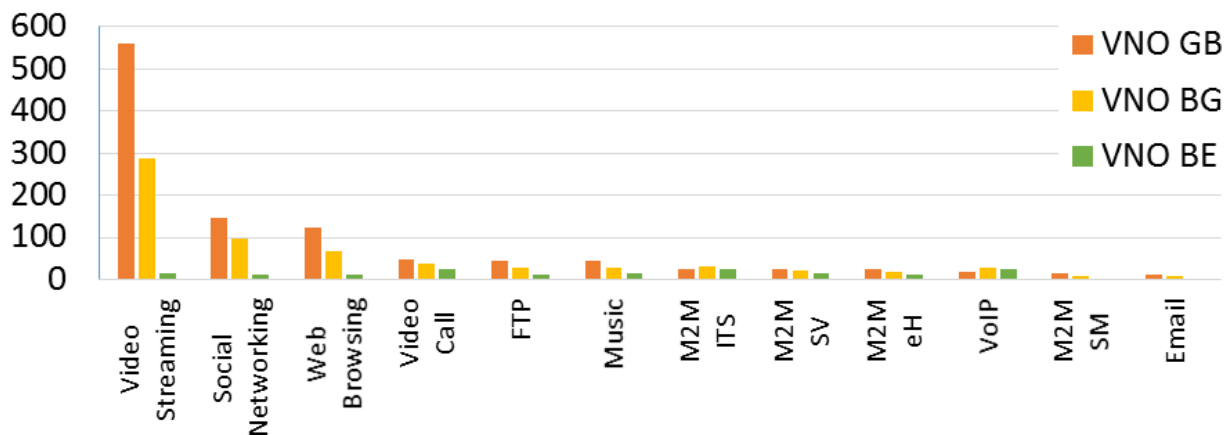


Figure 4-29: Data rate allocated to each service

## References

- [1] M. N. Tehrani, M. Uysal, H. Yanikomeroglu, "Device-to-Device Communications in 5G Cellular Networks: Challenges, Solutions and Future Directions", IEEE Communications Magazine, Vol. 52, No. 5, May, 2014, pp. 86-92.
- [2] 3GPP TR 36.843 v1.2.0 "Study on LTE Device to Device Proximity Services; Radio Aspects (Release 12)", February, 2014.
- [3] D. Camps-Mur, A. Garcia-Saavedra, P. Serrano, "Device-to-Device Communications with WiFi Direct: Overview and Experimentation", IEEE Wireless Communications, June, 2013, pp. 96-104.
- [4] R. S. Sutton, A. G. Barto, Reinforcement Learning: an Introduction, MIT Press, 1998.
- [5] J. Nasreddine, O. Sallent, J. Pérez-Romero, R. Agustí, "Positioning-based Framework for Secondary Spectrum Usage", Physical Communication Journal (PhyCom), Elsevier, June 2008, Vol.1, pp: 121-133.
- [6] L. Davis (ed.), Handbook of genetic algorithms, International Thomson Computer Press, 1996.
- [7] [http://help.telecom.co.nz/app/answers/detail/a\\_id/30335/~/setup-device-as-wireless-modem-for-pc---nokia-lumia-920](http://help.telecom.co.nz/app/answers/detail/a_id/30335/~/setup-device-as-wireless-modem-for-pc---nokia-lumia-920)
- [8] B. Lorenzo, S. Glisic, "Context Aware Nano Scale Modeling of Multicast Multihop Cellular Network," IEEE/ACM Trans. on Networking, vol. 21, no. 2, pp. 359 – 372, April 2013.
- [9] E. Arikan, "Some complexity results about packet radio networks," IEEE Trans. Intell. Transp. Syst., vol. IT-30, pp. 910–918, July 1984.
- [10] L. Davis, Handbook of Genetic Algorithms. Van Nostrand Reinhold, 1991.

- [11] B. Lorenzo, S. Glisic, "Optimal Routing and Traffic Scheduling for Multihop Cellular Networks Using Genetic Algorithm," IEEE Trans. on Mobile Computing, vol. 12, no.11, pp. 2274–2288, Nov. 2013.
- [12] J. J. Grefenstette, "Genetic algorithms for changing environments," in 2nd Int. Conf. Parallel Problem Solving Nature, 1992, pp. 137–144.
- [13] Zaheer Khan, Savo Glisic, Luiz A. DaSilva, Janne Lehtomäki, "Modeling the Dynamics of Coalition Formation Games for Cooperative Spectrum Sharing in an Interference Channel", IEEE Trans. Computational Intelligence and AI in Games, vol. 3, no. 1, March 2010, pp. 17-31.
- [14] E. Karami, S. Glisic, "Stochastic Model of Coalition Games for Spectrum Sharing in Large Scale Interference Channels", in Proc. International Conference on Communications, ICC2011. Kyoto, Japan, June 2011.
- [15] Y.H. Tam, R. Benkoczi, H.S. Hassanein, S. G. Akl, "Channel assignment for Multihop Cellular Networks: Minimum Delay," IEEE Trans. on Mobile Comp., vol. 9, no. 7, pp. 1022 – 1034, 2010.
- [16] Khatibi, S. and Correia, L. M., "Modelling of Virtual Radio Resource Management for Cellular Heterogeneous Access Networks", in Proc. of PIMRC'14 - IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, Washington, DC, USA, Sep. 2014.
- [17] Jacinto, N. M. d. S., "Performance Gains Evaluation from UMTS/HSPA+ to LTE at the Radio Network Level," Master of Science, Department of Electrical and Computer Engineering, Instituto Superior Técnico - Universidade Técnica de Lisboa, 2009.
- [18] Dhillon, H. S., Ganti, R. K., Baccelli, F., and Andrews, J. G., "Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks", IEEE Journal on Selected Areas in Communications, Vol. 30, No. 3, Apr. 2012.
- [19] Papoulis, A. and Pillai, S. U., Probability, random variables, and stochastic processes. McGraw-Hill, New York, USA, 2002.
- [20] ERICSSON, ERICSSON Mobility Report, ON THE PULSE OF THE NETWORKED SOCIETY, ERICSSON, Stockholm, Sweden, Nov. 2013 Available: <http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-november-2013.pdf>
- [21] Cisco Systems, The Zettabyte Era - Trends and Analysis, from Visual Network Index (VNI) White Paper, Cisco Systems, California, USA, May 2013 Available: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI\\_Hyperconnectivity\\_WP.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.pdf)
- [22] Cisco Systems, Global Mobile Data Traffic Forecast Update, 2012 - 2017, from Visual Network Index (VNI) White Paper, Cisco Systems, California, USA, Feb. 2013 Available: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-520862.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf)
- [23] Cisco Systems, Cisco VNI Adoption Forecast, 2012 - 2017, from Visual Network Index (VNI) White Paper, Cisco Systems, California, USA, May 2013
- [24] Roberts, J. W., "Internet traffic, QoS, and pricing", Proceedings of the IEEE, Vol. 92, No. 9, Sept. 2004.
- [25] Clark, C., Fraser, K., Hand, S., Hansen, J. G., Jul, E., Limpach, C., Pratt, I., and Warfield, A., "Live migration of virtual machines", in Proc. of - Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation Boston, Massachusetts, USA, May 2005.

#### 4.6 ACHIEVEMENTS JRA 1.1.2.3

Communication networks often span many hundreds of thousands of users with competing objectives, all selfishly looking to optimize their individual objectives (latency, drop probability, throughput, etc.). As a result, we are witnessing a paradigm shift towards distributed optimization and control, essentially letting users reach an optimal state by themselves. For this reason, a key element in the design of decentralized networks is the implementation of tools and ideas from game-theoretic learning and online optimization.

To that effect, we examine the problem of efficient multi-path routing in wireless networks by means of a distributed “exponential learning” scheme based on the Gibbs distribution of statistical physics. The attracting set of this learning algorithm consists of those traffic distributions that minimize aggregate latency in the network, and we find that this set is a convex polytope with dimension determined by the network’s degeneracy index (a notion which measures the overlap of utilized paths). Despite this abundance of stationary points, exponential learning always converges to a well-defined state and not merely to the set of optimum distributions. Furthermore, this convergence is exponentially fast: the network gets within  $\epsilon$  of an optimum traffic assignment in time which is  $O(\log(1/\epsilon))$ .

On the other hand, a major challenge occurs when delays fluctuate unpredictably due to random exogenous factors. Despite these stochastic perturbations, we find that “non-mixing” optimum distributions where the traffic of an origin-destination pair is not split over different routes remain stochastically stable irrespective of the fluctuations’ variance, and patient users converge to it almost surely. Alternatively, if the network only admits a fully-mixed optimum, the proposed learning scheme converges to a steady-state distribution which is concentrated around this optimum point – with the degree of concentration depending on the strength of the delay fluctuations.

This issue is treated at length in the working paper: P. Mertikopoulos and A. L. Moustakas, “An efficient learning scheme for routing in the presence of stochastic fluctuations”. Available online: <http://mescal.imag.fr/membres/panayotis.mertikopoulos/files/RobustRouting.pdf>

## 4.7 ACHIEVEMENTS JRA 1.1.3.1

### Spatially Coupled Codes for Block-Fading Channels

Partners: M. Lentmaier, I. Andriyanova, N. Hassan

We use density evolution to analyze the exact performance of a random LDPC block code and SC-LDPC code for a block-fading channel. Density evolution tracks the probability density function of the messages exchanged between the check and variable nodes in the bipartite Tanner graph. The worst channel parameter for which the bit error probability converges to zero is called the threshold of an ensemble. The threshold of an ensemble for the block-fading channel depends on the channel realization and hence does not exist. In order to characterize such a channel, the outage probability serves as a lower bound on the word error probability for any coding scheme.

An outage occurs when instantaneous input-output mutual information is less than the transmission rate  $R$ . In terms of density evolution, a density evolution outage (DEO) is defined as an event when the bit error probability does not converge to zero for a fixed value of SNR after a finite or an infinite number of decoding iterations are performed [BG<sup>+</sup>10]. The probability

of density evolution outage, for a fixed value of SNR, can then be calculated using a Monte Carlo method considering significant number of fading coefficients.

The component matrices for a  $(3,6)$ -regular SC-LDPC code with an increasing memory from 1 to 4 are considered in Table 10. The diversity order achieved for a random  $(3,6)$ -regular LDPC code is also given for reference. The edge spreading defines the interconnections between the blocks at different times for SC-LDPC code. For more detail on edge spreading, reader is referred to [LF<sup>+</sup>09]. The diversity order for the ensembles in Table 10 is numerically computed. It can be observed that the LDPC block code does not achieve the maximum diversity of  $d = 2$  and its diversity order is  $d = 1.3$ . However, the diversity of the SC-LDPC code increases with the coupling parameter  $m_{cc}$ . We observe that even a coupling to one neighboring block ( $m_{cc} = 1$ ) gives a diversity order of 3, which is more than twice as compared to the LDPC block code. Furthermore, increasing the memory of the code from 3 to 4 does not give any significant improvement in the diversity order. This is due to the fact that the maximum number of codewords connected to a check node is limited by the memory and the node degree. Hence, only simultaneously increasing node degree and memory would result in a further increase of the diversity order of the ensemble. In order to achieve a diversity order of  $d = 10$  using root-LDPC code, a special structure defined to achieve full diversity, a rate  $R = 1/d = 1/10$  is required. Whereas, diversity  $d = 10$  is achieved using rate  $R = 1/2$  SC-LDPC code with memory 4 and SC-LDPC codes do not require channel specific design and the increase in diversity order depends on the coupling parameter  $m_{cc}$ .

We further present the word error rate for finite length codes generated randomly while avoiding the cycles of length 4 in [HL<sup>+</sup>12]. The results show that the simulated word error rate for different ensembles defined in Table 10 matches well with the density evolution results. Note that, density evolution gives the results for the asymptotic case, i.e., an infinite block length is assumed.

**Table 10: The edge spreadings used for  $(, )$ -regular SC-LDPC code and the corresponding diversity values.**

Ensemble	$m_{cc}$	Edge spreading	$d$
LDPC-BC	0	$B = B_0 = [3, 3]$	1.3
EnsA1	1	$B_0 = [2, 2], B_1 = [1, 1]$	3
EnsA2	2	$B_{0,1,2} = [1, 1]$	6
EnsA3	3	$B_{0,3} = [1, 1], B_1 = [1, 0], B_2 = [0 \ 1]$	10
EnsA4	4	$B_0 = [1, 1], B_{1,3} = [1, 0], B_{2,4} = [0 \ 1]$	10

### **Spatially Coupled Code Design for Flexible Rates**

Partners: M. Lentmaier, N. Hassan

The capacity achieving property of regular SC-LDPC codes raises the question whether irregularity is still needed at all. In principle, it is possible for any arbitrary rational rate to construct regular codes that guarantee a vanishing gap to capacity with BP decoding. On the other hand, for some specific code rates, the required node degrees and hence the decoding complexity increase drastically. But even if we neglect the complexity, there exists another problem of practical significance that so far has not received much attention in the literature: for large node degrees  $J$  and  $K$  the threshold saturation effect will only occur for larger values of the coupling width  $w$ , as illustrated for the BEC in Figure 1 [NLF14]. The parameter  $w$  denotes the number of neighboring blocks that are coupled together and hence determines the minimal latency required by a sliding window decoder. We can see that for a given coupling width  $w$  the

gap to capacity becomes small only for certain code rates  $R$ , and it turns out that these rates correspond to the ensembles for which the variable node degree  $J$  is small.

also gives an upper bound  $C_{Max}$  on the complexity ensembles which is additionally shown in Fig. 2 by

$$C_{Max} = \frac{1}{A} \frac{1-R}{R}.$$

#### B. Thresholds of Coupled Regular Codes

Using spatial coupling of regular codes, the Sh can be achieved with increasing degrees. In contrast this stands the need for low complexity that is connected to low node degrees. We investigate the performance of code ensembles with  $(J_{Min}, K_{Min})$  from the start when spatial coupling is applied. The smoothing parameter

Figure 27: Density evolution thresholds  $\varepsilon^{BP}$  for  $(J,K)$ -regular SC-LDPC ensembles with coupling widths  $w=3$  and  $w=10$ , in comparison with the Shannon limit  $\varepsilon^{Sh}$ . Blue squares show LDPC code ensembles with  $J=3$ .

Motivated from this observation, we introduce some nearly-regular SC-LDPC code ensembles that are built upon the mixture of two favorable regular codes of same variable node degree. The key is to allow for a slight irregularity in the code graph to add a degree of freedom that can be used for supporting arbitrary rational rates as accurately as needed while keeping the check and variable degrees as low as possible. These codes exhibit performance close to the Shannon limit for all rates in the considered rate interval, while having a decoder complexity as low as for the best regular codes. The exclusion of variable nodes of degree two in the construction ensures that the minimum distance of the proposed ensembles increases linearly with the block length, i.e., the codes are asymptotically good.

### 4.8 ACHIEVEMENTS JRA 1.1.3.2

Partners involved : Guido Masera (CNIT/POLITO), Muhammad Awais (CNIT/POLITO), David Declercq (CNRS), Florence Alberge (CNRS), Jossy Sayir (UCAM)

#### Iterative Decoding Algorithm for Non Binary Encoders

Partner: G. Montorsi (CNIT/POLITO)

A VLSI implementation of analog digital belief propagation (ADBP) has been completed and reported in the publication "Awais, M.; Masera, G.; Martina, M.; Montorsi, G., "VLSI Implementation of a Non-Binary Decoder Based on the Analog Digital Belief Propagation," Signal Processing, IEEE Transactions on , vol.62, no.15, pp.3965,3975, Aug.1, 2014". An exceptional complexity reduction has been achieved by using the ADBP.

#### Non-linear iteratively decodable binary codes

Partner: Jossy Sayir (UCAM)

The class of non-linear codes with local constraints has been studied where each constraint requires participating code symbols to take on distinct values over a finite alphabet. The constraint node operation in belief propagation decoding for this type of code was shown to be equivalent to the computation of a Cauchy permanent. A trellis-based decoder was devised

whose complexity, while still of exponential in the alphabet size, is much reduced for alphabet sizes of interest.

#### 4.9 JRA 1.1.3.3

##### **Capacity Region of Cooperative Multiple-Access Channel With States**

Partners: Shlomo Shamai, A. Zaidi, P. Piantanida.

We consider a two-user state-dependent multiaccess channel in which the states of the channel are known noncausally to one of the encoders and only strictly causally to the other encoder. Both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. We find explicit characterizations of the capacity region of this communication model in both discrete memoryless and memoryless Gaussian cases. In particular, the capacity region analysis demonstrates the utility of the knowledge of the states only strictly causally at the encoder that sends only the common message in general. More specifically, in the discrete memoryless setting, we show that such a knowledge is beneficial and increases the capacity region in general. In the Gaussian setting, we show that such a knowledge does not help, and the capacity is same as if the states were completely unknown at the encoder that sends only the common message. Furthermore, we also study the special case in which the two encoders transmit only the common message and show that the knowledge of the states only strictly causally at the encoder that sends only the common message is not beneficial in this case, in both discrete memoryless and memoryless Gaussian settings. The analysis also reveals optimal ways of exploiting the knowledge of the state only strictly causally at the encoder that sends only the common message when such a knowledge is beneficial. The encoders collaborate to convey to the decoder a lossy version of the state, in addition to transmitting the information messages through a generalized Gel'fand-Pinsker binning. Particularly important in this problem are the questions of 1) optimal ways of performing the state compression and 2) whether or not the compression indices should be decoded uniquely. By developing two optimal coding schemes that perform this state compression differently, we show that when used as parts of appropriately tuned encoding and decoding processes, both compression a?-la noisy network coding by Lim or the quantize-map-and-forward by Avestimeher, i.e., with no binning, and compression using Wyner-Ziv binning are optimal. The scheme that uses Wyner-Ziv binning shares elements with Cover and El Gamal original compress-and-forward, but differs from it mainly in that backward decoding is employed instead of forward decoding and the compression indices are not decoded uniquely. Finally, by exploring the properties of our outer bound, we show that, although not required in general, the compression indices can in fact be decoded uniquely essentially without altering the capacity region, but at the expense of larger alphabets sizes for the auxiliary random variables.

##### **Publication**

A. Zaidi, P. Piantanida and S. Shamai (Shitz), Capacity Region of Cooperative Multiple-Access Channel With States, IEEE Trans. on Information Theory, vol. 59, no. 10, pp. 6153-6174, October 2013.

##### **Secure Degrees of Freedom of MIMO Channels**

Partners: S. Shamai, A. Zaidi, L. Vandendorpe

This work investigates the problem of secure transmission over a two-user multi-input multi-output (MIMO) X-channel in which channel state information is provided with one-unit delay to both transmitters (CSIT), and each receiver feeds back its channel output to a different transmitter. We refer to this model as MIMO-X channel with \textit{asymmetric} output feedback and delayed CSIT. The transmitters are equipped with  $M$  antennas each, and the receivers are equipped with  $N$  antennas each. For this model, accounting for both messages at each receiver, we characterize the optimal sum secure degrees of freedom (SDoF) region. We show that, in presence of asymmetric output feedback and delayed CSIT, the sum SDoF region of the MIMO X-channel is \textit{same} as the SDoF region of a two-user MIMO BC with  $2M$  antennas at the transmitter,  $N$  antennas at each receiver and delayed CSIT. This result shows that, upon availability of asymmetric output feedback and delayed CSIT, there is no performance loss in terms of sum SDoF due to the distributed nature of the transmitters. Next, we show that this result also holds if only output feedback is conveyed to the transmitters, but in a symmetric manner, i.e., each receiver feeds back its output to both transmitters and no CSIT. We also study the case in which only asymmetric output feedback is provided to the transmitters, i.e., without CSIT, and derive a lower bound on the sum SDoF for this model. Furthermore, we specialize our results to the case in which there are no security constraints. In particular, similar to the setting with security constraints, we show that the optimal sum DoF region of the  $(M, M, N, N)$ -MIMO X-channel with asymmetric output feedback and delayed CSIT is same as the DoF region of a two-user MIMO BC with  $2M$  antennas at the transmitter,  $N$  antennas at each receiver, and delayed CSIT. We illustrate our results with some numerical examples.

## Publications

MIMO X-Channels with Output Feedback and Delayed CSI, IEEE Trans. on Information Forensics & Security, vol. 8, no. 11, pp. 1760-1774, November 2013.

A. Zaidi, Z. Awan, S. Shamai (Shitz) and L. Vandendorpe, "Secure Degrees of Freedom of X-Channel with Output Feedback and Delayed CSIT", IEEE Int. Workshop on Information Theory, ITW, accepted for publication, Seville, Spain, Sep. 2013.

Zaidi, Z. H. Awan, S. Shamai (Shitz) and L. Vandendorpe, Secure Degrees of Freedom of

## Secure Transmission of Sources over Noisy Channels

Partners: S. Shamai and P. Piantanida,

This paper investigates the problem of source-channel coding for secure transmission with arbitrarily correlated side informations at both receivers. This scenario consists of an encoder (referred to as Alice) that wishes to compress a source and send it through a noisy channel to a legitimate receiver (referred to as Bob). In this context, Alice must simultaneously satisfy the desired requirements on the distortion level at Bob and the equivocation rate at the eavesdropper (referred to as Eve). This setting can be seen as a generalization of the problems of secure source coding with (uncoded) side information at the decoders and the wiretap channel. A general outer bound on the rate-distortion-equivocation region, as well as an inner bound based on a pure digital scheme, is derived for arbitrary channels and side informations. In some special cases of interest, it is proved that this digital scheme is optimal and that separation holds. However, it is also shown through a simple counterexample with a binary source that a pure analog scheme can outperform the digital one while being optimal. According to these observations and assuming matched bandwidth, a novel hybrid digital/analog scheme that aims

to gather the advantages of both digital and analog ones is then presented. In the quadratic Gaussian setup when side information is only present at the eavesdropper, this strategy is proved to be optimal. Furthermore, it outperforms both digital and analog schemes and cannot be achieved via time-sharing. Through an appropriate coding, the presence of any statistical difference among the side informations, the channel noises, and the distortion at Bob can be fully exploited in terms of secrecy.

#### **Publication**

J. Villard, P. Piantanida, and S. Shamai (Shitz), Secure Transmission of Sources over Noisy Channels with Side Information at the Receivers, IEEE Trans. on Information Theory, vol. 60, no. 1, pp. 713-739, January 2014.

#### **Sparse Sensing and Sparse Sampling of Coded Signals**

Partner: S. Shamai

Advances of information-theoretic understanding of sparse sampling of continuous uncoded signals at sampling rates exceeding the Landau rate were reported in recent works. This work examines sparse sampling of coded signals at sub-Landau sampling rates. It is shown that with coded signals the Landau condition may be relaxed and the sampling rate required for signal reconstruction and for support detection can be lower than the effective bandwidth. Equivalently, the number of measurements in the corresponding sparse sensing problem can be smaller than the support size. Tight bounds on information rates and on signal and support detection performance are derived for the Gaussian sparsely sampled channel and for the frequency-sparse channel using the context of state dependent channels. Support detection results are verified by a simulation. When the system is high-dimensional the required SNR is shown to be finite but high and rising with decreasing sampling rate, in some practical applications it can be lowered by reducing the a-priori uncertainty about the support e.g. by concentrating the frequency support into a finite number of subbands.

#### **Publication**

M. Peleg and S. Shamai, "On sparse sensing and sparse sampling of coded signals at sub-Landau rates", Trans. Emerging Tel. Tech., vol. 25, pp. 259-272, 2014.

#### **Soft-Decoding-Based Strategies for Relay and Interference Channels**

Partner: S. Shamai

We provide a rigorous mathematical analysis of two communication strategies: soft decode-and-forward (soft-DF) for relay channels, and soft partial interference-cancelation (soft-IC) for interference channels. Both strategies involve soft estimation which assists the decoding process. We consider LDPC codes, not because of their practical benefits, but because of their analytic tractability, which enables an asymptotic analysis similar to random coding methods of information theory. Unlike some works on the closely-related demodulate-and-forward, we assume non-memoryless, code-structure-aware estimation. With soft-DF, we develop simultaneous density-evolution to bound the decoding error probability at the destination. This result applies to erasure relay channels. In one variant of soft-DF, the relay applies Wyner-Ziv coding to enhance its communication with the destination, borrowing from compress-and-forward. To analyze soft-IC, we adapt existing techniques for iterative multiuser detection, and focus on binary-input additive white Gaussian noise (BIAWGN) interference channels. We prove that optimal point-to-point codes are unsuitable for soft-IC, as well as for all strategies that apply

---

partial decoding to improve upon single-user detection (SUD) and multiuser detection (MUD), including Han-Kobayashi (HK).

**Publication**

A. Bennatan, S. Shamai (Shitz), and A. R. Calderbank, "Soft-Decoding-Based Strategies for Relay and Interference Channels: Analysis and Achievable Rates using LDPC Codes", IEEE Trans. on Inform. Theory, vol. 60, no. 4, pp. 1977-2009, April 2014.

Comments and suggestions for the improvement of this document are most welcome and should be sent to:

[project\\_office@newcom-project.eu](mailto:project_office@newcom-project.eu)



<http://www.newcom-project.eu>