Integrated Cross-Layer Energy Savings in a Smart and Flexible Cellular Network

Weisi Guo, Siyi Wang, Charles Turyagyenda, Tim O'Farrell
Department of Electronic and Electrical Engineering
University of Sheffield, United Kingdom
Email: {w.guo, siyi.wang, c.turyagyenda, t.ofarrell}@sheffield.ac.uk

Abstract— **A key challenge for mobile operators is how to reduce the operational energy and cost expenditure, whilst meeting the growing demand for throughput. In recent years, individual research techniques have shown that significant savings can be made. The majority of savings are achieved in the signal transmission stage and are obtained under certain modeling conditions and assumptions. How the gains can be combined together to yield higher total operational savings is largely unexplored, especially under a realistic multi-cell multi-user environment.**

This paper employs an integrated analysis of the cross-layer techniques that reduce energy consumption or improve the spectral- and energy- efficiency tradeoff. The research is part of the key integration process of the MVCE Green Radio (GR) programme, which combines architecture, transmission technique, resource management, and hardware research. The integrated operational energy savings have been shown to be above 90% and the associated cost savings are up to 34%. Furthermore, the paper discusses the impact of machine-learning and energy harvesting on the energy and cost consumption, to create a smart and flexible cellular network.

I. INTRODUCTION

It is expected that there will be over 5 billion mobile subscribers globally by 2015 and the mobile data usage is to grow by 50 fold from 2010 to 2015. This is primarily attributed to the increased usage of data intensive mobile applications and devices, as well as the increase population mobility. Furthermore, the migration of population from rural to urban areas means that an increasing majority of subscribers live in dense urban environments (up to 80% in developed nations).

In order to meet the growing traffic load, Mobile Network Operators (MNOs) have traditionally increased the number of base-stations (BSs) deployed (*densification*). This solution results in increased overall energy consumption and higher operational expenditure costs (e.g., fossil fuel and electricity bills) and eventually leads to adverse effects on the environment through increased carbon footprint (CO_2 emissions).

A. Environmental and Business Impact

The Information Communication Technologies (ICT) infrastructure currently consumes approximately 3.5-4% of the global energy consumption. Of which, wireless communication systems consume roughly 15% of this amount. The majority of the expenditure is consumed within the outdoor BSs of the cellular network. Thus the challenge to MNOs is how to provide sufficient network capacity that meets growing traffic demands while preventing the operational expenditure (OPEX) and energy requirements from scaling directly with traffic volumes. This challenge has attracted the attention of international MNOs [1], equipment vendors, and the research community [2] [3]. The primary benefits of saving energy consumption are two fold:

1) Improve commercial competitiveness and the average revenue per user (ARPU) by reducing costs. Up to 40% of operation and maintenance (O&M) costs are attributed to energy consumption.
2) Meet environmental targets by reducing the carbon footprint. MNOs such as Vodafone have pledged to reduce their carbon footprint by up to 50% by 2020 [1].

B. Review

When considering a wireless multi-cell network, the achievable system throughput performance typically consumes 3 valuable resources: *spectrum*, *energy*, and *cost*. Wireless research has primarily focused on maximizing the link-layer spectral efficiency (SE) and the system layer throughput performance through radio-resource-management (RRM). In the past decade, some of the research focus has been on improving energy efficiency (EE). Generally speaking, there is a tradeoff between improving the total network throughput and consuming more energy. For a given target throughput, different architectures types can meet the traffic load with different deployment solutions. This consequently incurs different energy consumption levels. The relationship between energy consumption and throughput was investigated for noise-limited channels in [4] and for interference-limited channels in [5].

Within the MVCE Green Radio (GR) programme [6] and the international research community, the research can be divided into 3 categories:

- Heterogeneous Architectures: relaying [7], multi-tier networks, multi-cell processing (CoMP), network reconfiguration.
- Transmission Techniques: MU-MIMO, energy efficient scheduling, traffic-offloading [8], and mechanical-relaying [9].
- Radiohead Hardware: power amplifier efficiency [10] and antenna efficiency.

A key shortfall in all current research areas is the lack of an integrated RAN design. The isolated energy savings

978-1-4673-2997-2/12/$31.00 ©2012 IEEE

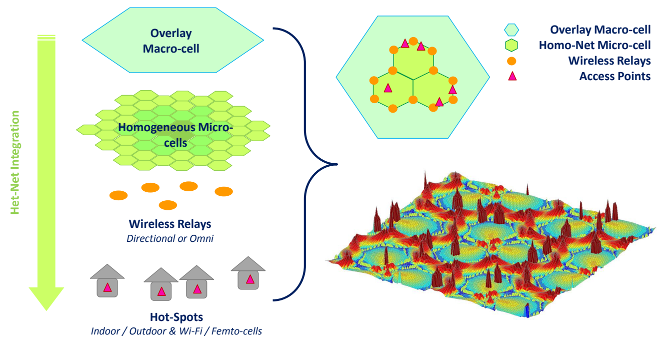


Fig. 1. Simulation Platform Modeling a Heterogeneous Network Architecture: Overlay Macro-BS, Micro-BSs, Relays and Indoor Access-Points.

for individual researches are *highly dependent* on modeling assumptions and the isolated nature of the investigation. A sensitivity analysis performed in [5] found that a large variety of factors can affect the energy savings of many techniques by an order of magnitude. The most sensitive assumptions include: interference, propagation channel and power consumption modelling.

C. Contribution

The stand-alone research in existing literature yields no clear integration methodology between the various techniques. This paper demonstrates how the authors have tackled the challenge of integration and proposes directions on how the integration process can be further enhanced. The analysis considers two forms of integrated analysis:

- **Dynamic Programming**: each technique is enhanced in isolation by the researcher and integrated under a common set of assumptions and simulation platform. There is a limited *human intelligence* involved in articulating the integration process and which techniques can work in synergy.
- **Machine Learning**: each technique is integrated blindly and the network nodes employ *artificial intelligence* to determine what combination of techniques are most suitable for the current traffic environment. Given that each network node's performance is dependent on the actions of itself and other nodes, there is the issue of unsupervised learning and coordination.

The methodologies and results presented can be used as guidance for integrated studies in other research areas.

II. RADIO-ACCESS-NETWORK (RAN) MODEL

A. Common Simulation Platform

The common simulation platform considers a Multi-Cell-Multi-User (MCMU) network, where the base-stations (BSs) are heterogeneously deployed in the following manner:

- 19 Homogeneous Micro-BSs with an optional Macro-BS overlay for high mobility users
- Heterogeneous Outdoor Relays that assist the Micro-BSs for improved cell-edge performance
- Heterogeneous Indoor Access-Points (APs) that assist the Micro-BSs for improved indoor performance
- BSs can dynamically reconfigure based on the local traffic load scenario

TABLE I
SYSTEM PARAMETERS FOR COMMON SIMULATOR PLATFORM

Parameter	**Value**
LTE Operating Frequency	2100MHz
LTE System Bandwidth	20MHz
HSPA Operating Frequency	2400MHz
HSPA System Bandwidth	5MHz
Path-loss Model	3GPP [11]
Traffic Intensity	30-120 Mbit/s/km^2
Shadow Fading variance	9dB
Transmission Scheme	AMC
Transmission Techniques	SISO, 2x2 MIMO
LTE subcarrier size	15 kHz
AWGN power per subcarrier	6×10^{-17} W
Power Consumption Model	[5] [12]

An illustration of the architecture is shown in Fig. 1 and additional transmission and RRM techniques are embedded within the transmission nodes, which will be explained in greater detail later on. The system layer simulation results are derived from the MVCE GR's proprietary **VCESIM** LTE Dynamic System Simulator [5], which is bench-marked against 3GPP tests and has been verified by industrial sponsors Fujitsu and Nokia Siemens Networks. Each BS's throughput considers 2-tiers of inter-BS interference (wrap-around model), which is sufficiently accurate.

B. Common Environment

The simulation environment assumes a *dense urban scenario*, where an increasing proportion of subscribers and mobile data transfer occurs. The following key aspects of a dense urban scenario are worth highlighting: offered traffic rate varies between 30 to 120 Mbit/s/km^2 depending on time of the day, and the propagation model is an urban micro model [11]. In order to satisfy the high traffic rate demand, typically a dense number of BSs are deployed to exploit spectrum reuse and the low user mobility in city centers. As previously mentioned, such a deployment is the **reference deployment** for 3G HSPA and 3.9G LTE, which incurs a high operational energy and cost consumption. A list of simulation parameters employed can be found in Table I and a full range of power consumption modeling can be found in [5] [12].

III. INTEGRATION: DYNAMIC PROGRAMMING

Dynamic programming is a general methodology for solving complex network problems by breaking the problem down into simpler sub-problems, and then combine the solutions of the sub-problems into an overall solution. Some key aspects of dynamic programming are:

- *No Iteration*: solve each sub-problem once and combine their solutions under the common simulation platform.

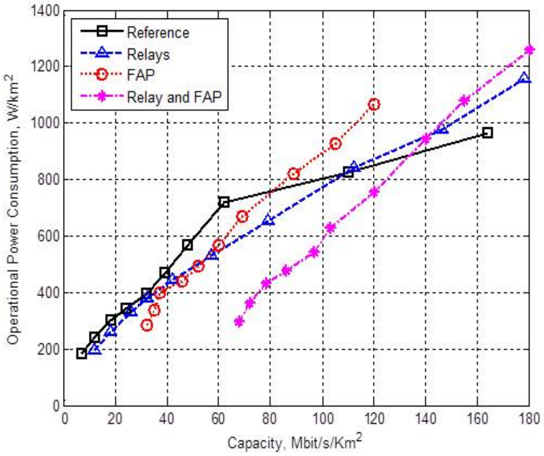


Fig. 2. Peak Capacity and Operational Power Tradeoff for different LTE RAN deployments: Homogeneous Reference, Heterogeneous Relay- and AP-assisted BSs.

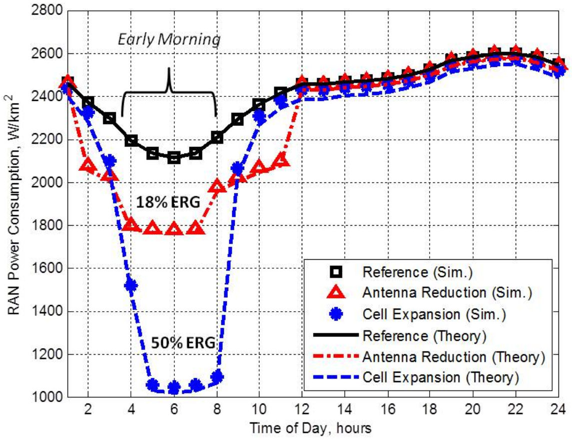


Fig. 3. Operational Power Consumption for Different Times of the Day, for: Reference RAN, Antenna Reduction Techniques [8], Cell Expansion Technique [3].

This is so that for a multi-cell multi-user network, the number of repeating sub-problems grows exponentially and the solution space is largely constant.

- *Offline*: store each sub-problem's solution as an offline look-up-table. In the live online network, the traffic environment is classified into categories and the look-up-table is used to devise a solution.

The paper will now present how the look-up-table is devised and how the live network can potentially use it to reduce energy consumption, whilst satisfying the traffic rate offered.

A. Static Deployment

In static deployment, the RAN is deployed to meet the peak traffic intensity at the lowest energy consumption level. The primary mechanism for energy reduction in a relay- or AP-assisted BS, is to transmit the same data at a lower power via the relay nodes or APs. However, as the BS technology shifts from a sparse distribution of high-powered macro-BSs in low traffic rural areas to a dense distribution of low-powered micro-BSs in high traffic urban areas, the energy saving is eroded.

The results in Fig. 2 shows different deployment solutions that can achieve different levels of peak capacity. For a high peak traffic intensity (120-160 Mbit/s/km^2), the relay- or AP-assisted RAN actually consumes more energy than the reference. However, as the peak traffic intensity is reduced, the energy saving benefits of a combined relay- and AP-assisted RAN is up to **60%** at 70 Mbit/s/km^2. The energy saving can be further enhanced by 15% by employing more efficient BS hardware in terms of increased power-amplifier efficiency [10] and antenna polarization efficiency.

The tradeoff presented in Fig. 2 can be seen as a form of look-up-table generated offline, whereby the system designer can select the lowest energy system for a given peak traffic intensity expected. What has not been considered, is how the system can reconfigure itself as the traffic intensity varies with time. Clearly the whole RAN can not re-deploy to a different solution in Fig. 2, as that would involve relocating the BSs and changing the hardware. Given that the network has been deployed for a certain peak traffic load, how can it scale down efficiently?

B. Flexibility: Reconfiguration

In order to increase the flexibility of the RAN in accordance with the traffic load offered, several reconfigurable RAN architectures and base-station designs have been proposed. The research in [3] proposes a RAN sleep mode operation that can meet the range of dynamic traffic loads at the lowest energy consumption value. This is known as *cell-zooming* or *cell-expansion*. The principle of scaling energy efficiently with traffic variations is to switch-off entire BSs, as opposed to just switching off the transmitting antennas [8]. The results in Fig. 3 shows the associated power consumption variation with offered traffic intensity, which changes with the time of the day. The data is for a typical European urban environment given in [13]. The results show that at peak traffic loads, the energy saving is negligible, but for low traffic loads the saving can be up to 18% for antenna reduction and 50% for cell expansion (sleep mode).

C. Transmission Techniques

A key research area in improving transmission spectral efficiency is MIMO. Over the past decade, Multi-User-MIMO (MU-MIMO) with Spatial-Multiplexing (SM) has attracted significant research attention for being able to obtain higher spectral efficiency compared to conventional Space-Frequency-Block-Coding (SFBC) MIMO, especially at high SINR regimes. Whilst greater spectral efficiency generally implies greater transmission energy efficiency, the inclusion of overhead power consumption in a BS means that more MIMO antennas can in fact lead to greater overall energy consumption [5].

The results in Fig. 4 show the CDF of the user throughput achieved, as well as the optimistic and pessimistic energy

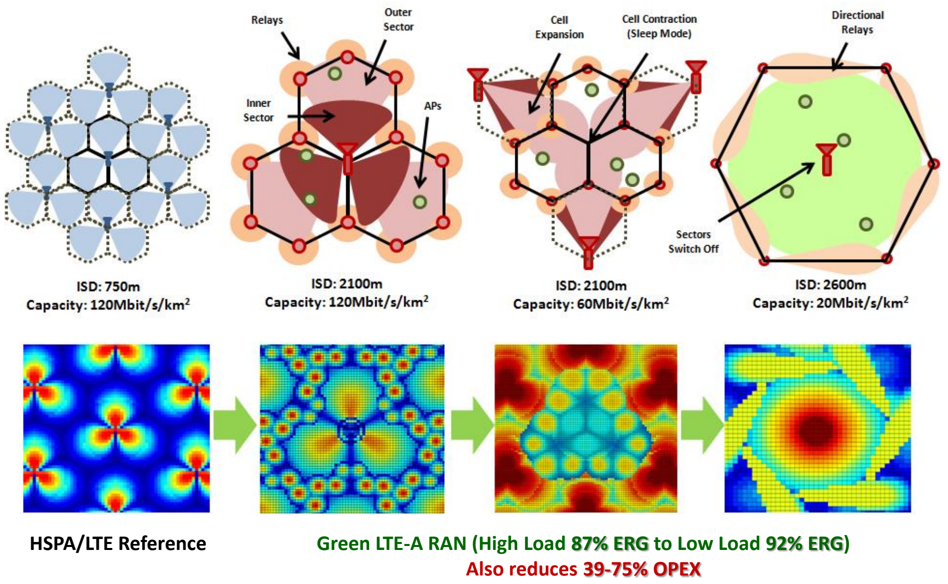


Fig. 5. Integrated Green RAN across a variety of traffic loads: Energy Saving Summary of different architectures, transmission and hardware research techniques. Top set of diagrams illustrate the dynamic network architecture and the bottom set of diagrams show the mean received SINR for each state.

savings of different MIMO techniques compared to SISO transmission. The optimistic results assume that a perfect cell expansion (sleep mode) technique can coordinate the BSs to keep the minimum number of BSs active to meet the traffic load. The pessimistic results assume that all BSs must remain active at all times. The results show that even under the optimistic assumptions, 2x2 MU-MIMO can save at most 4% energy compared to SISO. This is due to the fact that increasing the number of BS antennas, increases both the radiohead and overhead power consumption. On the other hand, 1x2 SIMO transmission can achieve 14-23% energy saving.

D. Integrated RAN

The final integrated RAN also included research from offloading data traffic to nearby Wi-Fi APs [8], mechanical relaying [9], CoMP, and other energy efficient scheduling techniques. Referring to Fig. 5, the figure shows the illustrated dynamic RAN and the corresponding mean capacity plots for:

- **Reference RAN at Peak Load**: homogeneous high density of micro- or pico-BSs with 3 sectors and 2x2 MIMO transmission [11].
- **Green RAN at Peak Load**: heterogeneous network with relay- and AP-assisted macro-BSs with 6 sectors and 1x2 SIMO transmission [5].
- **Green RAN at Medium Load**: dynamically reconfigure low load BSs to sleep and expand the coverage of neighbouring BSs to compensate for coverage loss [3]. Employ mechanical relaying [9] and coordinated multi-cell interference mitigation for users in the compensation zone.
- **Green RAN at Low Load**: dynamically switch the multi-sector BS to a single-sector omni-directional BS and switch the relays to directional relays [7].

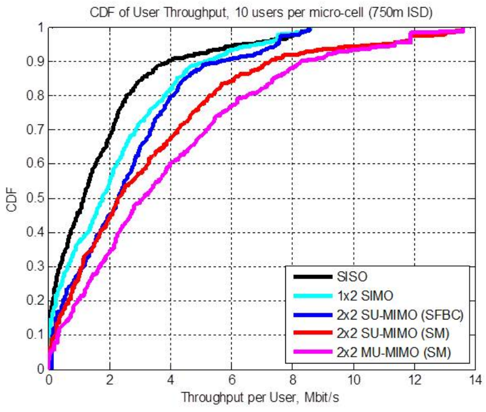
The combined energy saving achieved was approximately 87% at peak loads and 92% (12.5 fold) at low loads. This demonstrates that the Green RAN is **flexible** through its ability to transform its BS configuration depending on the regional traffic load intensity. The results in [14] have also shown that the proposed architecture can save the O&M costs by up to 34%. In order to achieve the 100 fold energy saving, a further 8 fold energy reduction is needed from an approach that is largely independent of the already considered techniques. In the next section the paper considers the possibility of improving on the dynamic programming approach of offline optimization by using machine learning and energy harvesting.

IV. TOWARDS 100X ENERGY SAVING

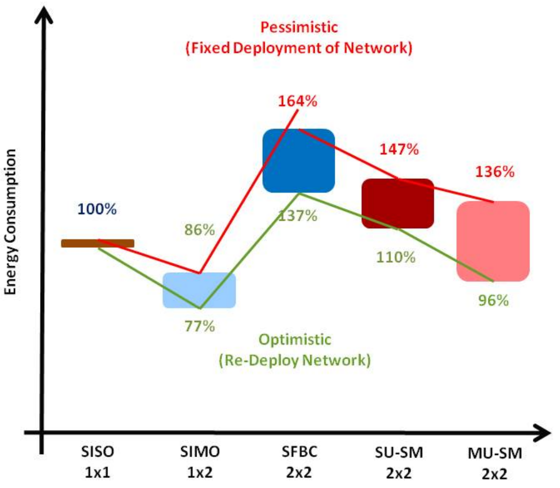
A. Integrated RAN: Smart & Flexible

The existing integrated architecture considered in Fig. 5 has a limited level of intelligence in terms of dynamic adaptivity to the local traffic variations. This is primarily due to the fact that any action a BS takes, not only affects its own performance, but the performance of other BSs. Therefore, it is challenging for a BS or indeed the network engineer to know what action can best serve a BS's own interests or the interests of the whole RAN. Without supervision and a *correct answer*, unsupervised learning or reinforced learning methods can help to achieve near-optimal performances [15]. This is part of the Self-Organizing Network (SON) theme of 3GPP standards to create **smart** automata networks.

In the absence of a full state transition map (Markov-Decision-Process (MDP)), there is often a tradeoff between the rate of *exploitation* and the rate of *exploration*. As shown in Fig. 6a, over-exploration of strategies can lead to greater understanding of the MDP, but not being able to sufficiently exploit the most beneficial strategy. Under exploration will lead to rapid convergence to a single strategy. A consequence



a) CDF of User Throughput for Different Transmission Schemes



b) Energy Consumption vs. SISO Reference for Different Transmission Schemes

Fig. 4. Throughput and energy consumption results compared to SISO, for different MIMO techniques: SFBC, SM and MU-MIMO.

can be that the RAN repeatedly employs a sub-optimal strategy and is not able to discover new and better strategies, as shown in Fig. 6a. A variety of exploration methodologies exist, and a popular one is the *Boltzmann Exploration*, where the chance of randomly exploring a strategy is weighted by the expectation of reward of all strategies.

The paper formally defines the following:

- *Agents*: the BSs, where the observed BS is $n \in N_{\text{BS}}$. Each BS can be assisted by a number of APs and relays.
- *Environment*: the heterogeneous RAN containing all the BSs, which are subject to a certain offered traffic rate that has a uniform temporal and spatial distribution.
- *Action*: each agent can take an action to save energy, $a_{n,t} \in \mathcal{A}$.
- *Strategy*: a strategy ($s_n \in \mathcal{S}$) is a set of actions: for a given state $L_{n,t}$, what set of actions $a_{n,t}$ to take in order maximize reward.
- *State*: is the defined by the current traffic load ($L_{n,t,s_{t-1}} \in \{0, 1\}$) experienced by the agent concerned, after taking an action in the previous time frame. The definition of a load is the ratio between the traffic rate offered to a BS and the maximum achievable system-level capacity of the BS: $L = \frac{R_{\text{traffic}}}{R_{\text{BS}}}$
- *Feedback*: by taking an action that leads to a state, the BS has an interference impact on the network containing other BSs.
- *Observed Reward*: defined by $\Re_s$, which is the energy saved without sacrificing the Quality-of-Service (QoS) throughput delivered to the bottom 5% of UEs.

A variety of exploration methodologies exist, and a popular one is the **Boltzmann Exploration**, where the chance of randomly exploring a strategy is weighted by the expectation of reward ($\Re_s$) of all strategies. Under the Boltzmann Exploration algorithm, the strategy selection is based on a weighted probability (p) biased in favor of likelihood to yield high rewards:

$$p_s = \frac{e^{\frac{\mathbb{E}[\Re_s]}{\Delta}}}{\sum_{s' \in \mathcal{S}} e^{\frac{\mathbb{E}[\Re_{s'}]}{\Delta}}}, \qquad (1)$$

where the parameter Δ adjusts the level of *exploration*: a small Δ favors exploitation and large Δ favors exploration. Generally, a large Δ can guarantee asymptotic optimality. However, there is a hidden variance tradeoff, which from Fig. 6a shows that higher rewards generally mean higher reward variance.

The results in Fig. 6b show how the Boltzmann Exploration can be used in learning sleep mode strategies to tradeoff energy saving and outage occurrences. In cell expansion, a BS can go into sleep mode only to find that an unacceptable high level of outage has been caused. Given that the actions of neighbouring BSs and the local traffic intensity both affect its decision process, the MDP is an infinite state model with a changing probability model. The results in Fig. 6b show that over exploitation of the MDP (learning rate Δ=2) can cause a low outage and low energy saving. However, sufficient exploration of the MDP (learning rate Δ=15) can cause an acceptable increase in outage and a much greater energy saving. The energy savings can be further enhanced by adding intelligence to the proposed RAN, which can yield up to a total of 95% energy saving.

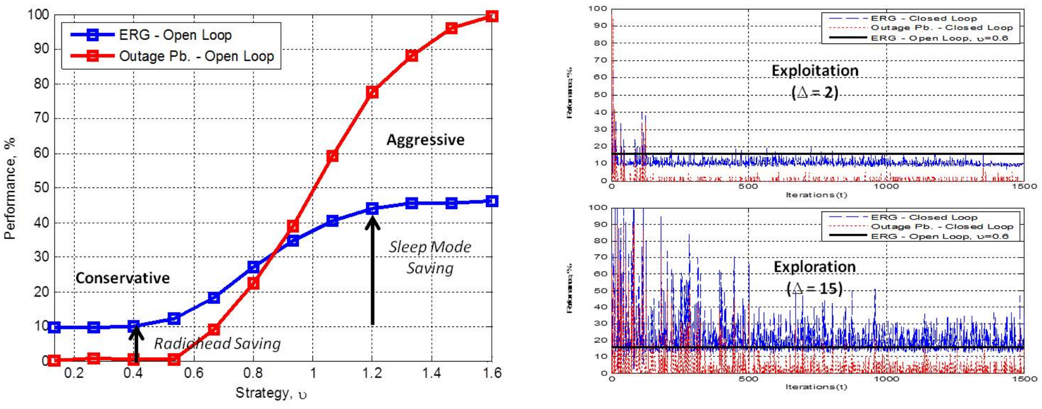
The key *advantages* of machine learning over off-line programming shown previously are:

- No local- or network-wide knowledge of RAN is required
- No coordination between transmission nodes is required
- Self-organizing ability without supervision

The *disadvantage* is that tuning the learning rate (Δ) is not trivial and can not be explicitly analyzed. This is an area of further study to examine how machine-learning can further enhance the energy savings for other integrated techniques.

B. Energy Harvesting and Cloud Networks

Another area of promising research is energy harvesting and being able to reduce the grid energy consumption by using solar energy generation at the BSs. Typically a 12m^2 solar panel can generate sufficient electricity for a 100-200W pico-BS in the regions of: Europe, North America, China and



(a) Open-Loop Distributed Coordination with Fixed Strategy, υ

(b) Closed-Loop with Random Greedy Algorithm with Exploration factor Δ = 2 and 15.

Fig. 6. Self-Organizing-Network (SON) with Reinforced Learning: a) Tradeoff between Exploitation and Exploration, b) Energy Saving and Outage Performance (%) for Exploitative and Explorative Behaviors in Boltzmann Exploration.

Japan. Alternatively, cloud-networks can also reduce energy consumption of the whole radio-access-network by collecting the overhead processing and cooling at a single central and remote location, especially one that has access to energy harvesting potential or cheap grid electricity pricing (i.e., the Sahara for Europe or Canada for North America).

V. CONCLUSIONS

This paper has tackled the challenge of how to reduce energy consumption in cellular networks, whilst meeting the growing demand for capacity. The authors have used an integrated approach to different cross-layer techniques to propose a new low energy green RAN, that is both flexible and smart.

The **flexibility** aspect of the RAN comes from the dynamic base-station designs that allows them to change their operation mode in accordance with the traffic load. The **smart** aspect of the RAN comes from the machine learning algorithms that allows self-organization and optimal operational strategy usage. The achieved energy savings for the proposed RAN is up to 87% at high loads and 95% at low loads compared to the reference RAN. This leads to a 34% operational cost saving for the network operator. The remaining challenge is how to reach the 99% saving mark by employing additional artificial intelligence techniques and utilizing cloud-network and energy harvesting research.

Acknowledgement

The work reported in this paper has formed part of the Green Radio Core 5 Research Programme of the Virtual Centre of Excellence in Mobile and Personal Communications, Mobile VCE. Fully detailed technical reports on this research are available to Industrial Members of the Mobile VCE. www.mobilevce.com

REFERENCES

[1] "Sustainability Report 2010-2011," Vodafone Group Plc," Technical Report, 2011.

[2] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P. Grant, H. Haas, J. Thompson, I. Ku, C.-X. Wang, T. A. Le, M. Nakhai, J. Zhang, and L. Hanzo, "Green radio: radio techniques to enable energy-efficient wireless networks," in *Communications Magazine, IEEE*, vol. 49, June 2011, pp. 46–54.

[3] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," in *Communications Magazine, IEEE*, Nov. 2010, pp. 74–79.

[4] Y. Chen, S. Zhang, S. Xu, and G. Li, "Fundamental trade-offs on green wireless networks," in *Communications Magazine, IEEE*, vol. 49, no. 6, June 2011.

[5] W. Guo and T. O'Farrell, "Green cellular network: Deployment solutions, sensitivity and tradeoffs," in *Wireless Advanced (WiAd), IEEE Proc.*, London, UK, June 2011.

[6] "Mobile Virtual Centre of Excellence (VCE) - Green Radio Programme," in *http://mobilevce.com/*.

[7] W. Guo and T. O'Farrell, "Relay deployment in cellular networks: Planning and optimization," in *Selected Areas in Communications (JSAC), IEEE Journal on*, Sept. 2012.

[8] O. Holland, V. Friderikos, and A. H. Aghvami, "Energy efficient cross band spectrum management for mobile operators," in *Proc. IEEE Globecom*, Dec. 2010, pp. 2020–2040.

[9] P. Kolios, V. Friderikos, and K. Papadaki, "Load balancing via store-carry and forward relaying in cellular networks," in *Proc. IEEE Globecom*, Feb. 2010.

[10] K. Mimis, K. Morris, and J. McGeehan, "A 2GHz GaN class J Power Amplifier for Basestation Applications," in *IEEE Symp. for Radio and Wireless (RWS)*, Phoenix, USA, Jan. 2011.

[11] 3GPP, "TR36.814 V9.0.0: Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," 3GPP," Technical Report, Mar. 2010.

[12] G. Auer, V. Giannini, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, C. Desset, and O. Blume, "Cellular energy efficiency evaluation framework," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE*, May 2011, pp. 1–6.

[13] EARTH, "WP2.D2.3: Energy Efficiency Analysis of the Reference Systems," Energy Aware Radio and Network Technologies (EARTH)," Technical Report, Dec. 2010.

[14] W. Guo and T. O'Farrell, "Capacity-Energy-Cost Tradeoff for Small-Cell Networks," in *Proc. IEEE International Workshop on Green Wireless Communications and Networks (GreeNet)*, Yokohama, Japan, May 2012.

[15] A. Galindo-Serrano and L. Giupponi, "Distributed Q-Learning for Aggregated Interference Control in Cognitive Radio Networks," in *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, May 2010, pp. 1823–1834.