Energy Efficiency Solutions for the Mobile Network Evolution Towards 5G: an Operator Perspective

Damiano Rapone, Dario Sabella, Maurizio Fodrini
Wireless Access Innovation
Telecom Italia
Turin, Italy
{damiano.rapone, dario.sabella, maurizio.fodrini}@telecomitalia.it

***Abstract*—This paper focuses on techniques or actions that the network operator should consider for the mobile network's evolution towards 5G from a sustainability point of view. To this end, the energy consumption of the Radio Access Network (RAN) is assessed when considering different load conditions (obtained by means of actual daily traffic profiles extracted from the live network) and different yearly traffic forecasts. An overview of different "*what-if*" scenarios towards 2020 and beyond is provided and energy consumption of the network is assessed by considering evolutionary Power Models (PMs) of future Base Stations (BSs) related to different Radio Access Technologies (RATs). Together with evolutionary PMs, many Energy Efficiency (EE) features that may be implemented within the network are also considered, with particular focus on EE features evaluated at higher time scale (from minutes to years), i.e., related to progressive network renewal with traffic steering options and legacy RATs' phase-off policies. These have been evaluated by using a software tool fed by actual traffic profiles, aiming at assessing EE performance at network level, by also considering the European Telecommunications Standards Institute (ETSI) EE specification ES 203 228, which is based on homogeneous clusters' evaluations that may be extrapolated at country level, providing useful information on the possible evolutions of the mobile networks in the view of future 5G systems.**

Keywords*—2G (GSM), 3G (UMTS), 4G (LTE), 5G, RAT (Radio Access Technology), RAN (Radio Access Network), EE (Energy Efficiency), PM (Power Model), traffic steering, cell DTX (Discontinuous Transmission), phase-off, green mobile networks, evolution, C-RAN (Cloud-RAN).

I. Introduction

The evolution of mobile networks from the introduction of the first generation systems until today and the forecasts for the next decade clearly indicate a growth of both the network itself (in terms of installed equipment) and carried traffic (in terms of transmitted bits) [2]. In fact, main network operators nowadays can still count on their 2G (GSM) network for revenues coming from voice traffic, while they offer broadband and more advanced services based on their 3G (UMTS) network. Recently, they began to deploy 4G (LTE) technology, promising faster broadband connections to their customers. At the same time, innovation sectors of telecommunications' industry started to think about new scenarios and new features for what the next generation of mobile systems for the year 2020 and beyond will be, i.e., 5G technology.

In the context of this evolution it is noteworthy to consider the continuous current growth of traffic demand as previously stated and the fact that network evolution towards 5G should follow criteria of sustainability and affordability, especially from the Energy Efficiency (EE) point of view. In addition to that, it is important to consider the continuous presence of new incoming technologies together with the concurrent tendency of network technologies to coexist through the years, giving rise to the need of designing a 5G architecture able to efficiently manage this heterogeneous scenario with flexibility in terms of both multi-Radio Access Technology (multi-RAT) management and usage of network resources, but also by taking into account the EE in Base Stations (BSs) and the scarcity of resources of the mobile devices (both smartphones, with ever increasingly sophisticated applications, and Machine-to-Machine (M2M) devices, with low power and computational capacity characteristics).

A lot of previous studies already analysed energy consumption of the overall telecommunication network (including fixed and mobile infrastructures) [17], also evaluating related operational costs [18]; scientific literature also studied EE as an additional requirement for the design and optimization of network performance [19], thus operators [21] are now considering this important aspect with increasing attention. Nevertheless, up to now none of all these works targeted long term evolutionary evaluations at network level, which instead would be useful to provide a long term view of the energy consumption, giving useful information to operators (from a sustainability perspective) on the possible evolutions of the mobile networks towards 5G systems.

Finally, so far EE assessments performed at network level [20] were not based on a clear and standardized methodology, mainly due to the absence of international specifications that could act as a reference for the evaluations. Only recently ETSI EE group introduced the ES 203 228 specification [22], with the aim to provide a useful reference for operators for their network-level EE assessments.

In our paper we perform a preliminary evaluation based on this specification, by assessing an homogeneous cluster of mobile network sites with detailed modelling of BSs (in terms

978-3-901882-70-8 ©2015 IFIP

of both traffic and energy consumption points of view). Our study also considers few selected techniques or actions that the network operator should consider for the mobile network's evolution towards 5G from a sustainability point of view, especially taking into account the energy consumption of the Radio Access Network (RAN) when considering different load conditions, obtained by means of actual daily traffic profiles extracted from the live network, and different years, according to current traffic forecasts [16]. The aim is to provide an overview of different "*what-if*" scenarios towards 2020 and beyond, as useful insights for operators aiming at analysing evolutionary steps of their mobile networks in the view of the introduction of future 5G systems.

To estimate the energy consumption of the network, a model of the power consumption related to future BSs is needed thus, following the methodology reported in [1], results from the EARTH project [2] will be used, by considering different Power Models (PMs) for different types of BSs. Evolutionary PMs will also be considered for each Radio Access Technology (RAT) through the years, together with many EE features that may be implemented within the network and which can be classified in shorter time scale features (e.g., fast/long cell Discontinuous Transmission (DTX), on a millisecond basis) and higher time scale (from minutes to years) ones (e.g., progressive network renewal with traffic steering options and legacy RATs' phase-off policies). The EE features' evaluation here provided is focused on higher time scale techniques only and performance assessment is done by means of a software tool fed by actual traffic profiles.

The rest of the paper is organized as follows: in section II both baseline system assumptions and considered EE performance metrics are reported while, in section III, cell DTX and traffic steering are described as possible EE features to be considered by the network operator. For traffic steering options only, evaluation methodology and results are shown in sections IV and V, respectively. Finally, conclusions and future works are presented in section VI.

II. REFERENCE SYSTEM

A. Baseline system assumptions

Energy consumption for the mobile operator is a practical problem involving not only new technologies and RATs but also legacy networks, hence an evaluation of 5G systems in a *green* perspective should consider a comprehensive assessment through the years of the overall network's EE performance (from today to 2020 and beyond). On top of that, the operator will carefully evaluate the opportunity and convenience of new RATs' introduction in future systems (e.g., for the satisfaction of increasing traffic demand).

The starting point of evaluations here provided is that multi-RAT environments should always be taken into account with realistic traffic assumptions and in accordance with methodologies currently considered as a reference for operators. For this reason, the baseline system is characterized by the following aspects:

- presence of 2G, 3G and 4G (in which potentially higher order Multiple Input Multiple Output (MIMO) systems and Carrier Aggregation (CA) are possible) and modelling of 5G (in terms of capacity and energy performance) RATs;
- alignment with methodology currently under discussion and finalization in the framework of ETSI EE (EEPS subgroup) for the definition of the ES 203 228 specification, in which clustered assessments based on measurements from the live network are taken into account;
- homogeneous network layout within assessed clusters, as an easy way to model the network in contrast with exhaustive assessment methodologies, often too costly;
- performance evaluated by considering commonly accepted EE metrics, by exploiting the outcomes of previous works (e.g., EARTH project) and current standardization work in ETSI EE.

B. Energy Efficiency performance metrics

Mobile network's EE can be assessed by considering the following two performance metrics [3], [4]:

bit/Joule: by denoting with Ψ the *bit/Joule* efficiency of the network, this can be written as in (1):

$$\Psi = \frac{C_{net}}{P_{net}} \tag{1}$$

where C_{net} is defined as the aggregate network capacity in *bit/s* and P_{net} is the power consumption of the network in *W*.

W/km²: the area power consumption, here denoted by Ω, relates the total power consumption of the network P_{net} to the size of the covered area A as in (2):

$$\Omega = \frac{P_{net}}{A} \tag{2}$$

Note that the optimal EE is achieved when the *bit/Joule* metric is maximized or the power per unit area W/km^2 is minimized.

Reference [5] reports the variation of both the *bit/Joule* and W/km^2 metrics with respect to the number of BSs when the capacity requirement is not considered: while *bit/Joule* increases monotonically with network densification, the W/km^2 metric indicates that reduced transmit power cannot compensate the additional power consumption of the BSs in idle state. Therefore, the W/km^2 metric increases with the number of BSs after reaching the optimum point. This suggests that maximizing the EE is not always equivalent to minimizing the energy consumption and that is the reason why the capacity requirement must be considered in order to prevent contradictory conclusions with different metrics. Otherwise the usage of *bit/Joule* might be misleading, since adding more capacity into the network will always reflect an increase in EE.

Here both the *bit/Joule* and W/km^2 metrics can be applied to all simulations performed with the evaluation tool described in section IV.C, with particular reference to minimization of area power consumption since "*what-if*" scenarios analyzed in section V are compared in terms of power consumption of a single cluster of mobile network sites (which is the same for every simulation).

III. Energy Efficiency Features

Today's mobile networks are designed and deployed based on the peak traffic demand and kept active regardless of the low utilization during different times of the day. Even at a very busy traffic hour of the day it has been shown that there are few cells that experience high load [1].

In order to scale the power consumption to the traffic dynamicity, adaptive network operation features are considered within this study, which are practical to apply and demonstrate even in today's mobile network sites. In the following subsections traffic steering techniques are described as an EE enabler for operators, especially in multi-RAT network environments and in relationship to the future introduction of 5G systems, whilst DTX techniques are reported only (hence, not evaluated), in order to provide a more complete overview of the different possible features applicable to the mobile networks.

A. Cell DTX

DTX has been used for a long time in mobile terminals to achieve long battery life. The idea is to transmit when only needed, putting the transmitter in a low power state otherwise. At the network side this technique is referred to as *cell DTX* and it is based on the hardware component de-activation feature which facilitates low power states. Two cell DTX versions can be distinguished: *fast cell DTX* and *long cell DTX.*

Fast cell DTX acts on slot/subframe level and exists in some different versions as well. The first one is *cell micro DTX*, also known as *micro sleep*, and means that, when there is not any user data to transmit, the radio is put into DTX among transmissions of Cell-specific Reference Signal (CRS). The second one is known as *MBSFN-based DTX*, in which Multicast-Broadcast Single-Frequency Network (MBSFN) subframes are used to make room for longer sleep periods, since CRS are not transmitted in these subframes. Both cell micro DTX and MBSFN-based DTX are possible in LTE Release-8 networks. Finally, in *cell short DTX*, CRS are assumed to be removed and replaced by DeModulation Reference Signal (DMRS), hence leaving room for even longer sleep periods, even if this option is not considered in today's LTE standard, but may be in the future (e.g. [6]). All these fast cell DTX versions are most effective in low load scenarios, but there is not any drawback at high load either.

Long cell DTX, on the contrary, acts on a slower time scale and it refers to the cell being put into a low-activity mode [7]. As such, it can be seen as a cell sleep (on/off functionality) and it is based on a deeper sleep state (lower power consumption) than the low power state considered above for the fast cell DTX versions. In principle, long cell DTX can be activated when there is not traffic demand or, alternatively, when, in low-traffic periods, cells are at low load and then traffic steering techniques are easily applicable without many "*ping-pong*" effects during the day. One use case or strategy for long cell DTX is in relatively densely deployed networks (e.g., dense urban, urban, suburban scenarios), where there is good coverage and capacity nodes can be put in long cell DTX during low-traffic periods. Hence, long cell DTX can be seen as an enabler/tool for operators of great EE potential, especially in multi-RAT network environments.

B. Intra-sector traffic steering

An exemplary EE feature has been developed by Telecom Italia and evaluated by means of the MATLAB-based evaluation tool described in section IV.C.

Basically the traffic in charge of either 2G, 3G or both RATs' frequency layers within a sector of a trisectorial site is steered towards the 4G frequency layers of the same sector and, as a consequence, the steered layers are switched-off (i.e., put in sleep state). This is done in order to exploit the higher 4G network capacity and the more energy-efficient 4G equipment compared to the 2G/3G one, in the view of achieving as much energy saving as possible.

Candidate 2G/3G layer(s) for traffic steering is (are) identified according to load conditions, that is (condition C_1):

- the normalized traffic (with respect to the peak throughput of the considered RAT) in charge of the layer(s) must be less than or equal to a predefined threshold th_1 (e.g., 50%). If steering is performed the 2G/3G layer(s) is (are) switched-off, that is, the power consumption(s) of that (those) layer(s) is (are) the one(s) that can be achieved when a sleep mode is activated.

Furthermore, another condition must be satisfied in order to actually perform the traffic steering (condition C_2):

- the candidate 4G layer(s), which is (are) able to handle the original 2G/3G traffic, is (are) chosen in such a way that it (they) will not saturate its (their) capacity after traffic steering. This means that the 4G normalized traffic must remain less than or equal to a predefined threshold th_2 (e.g., 90%).

By denoting non-normalized traffic data information of all RATs' frequency layers within sector s_1 of a trisectorial site by means of the 6-element row vector $\underline{\mathbf{t}}_{0,s1}^{(1)}$ (3)

$$\underline{\mathbf{t}}_{0,s1}^{(1)} = \begin{bmatrix} t_{0,f2G\_1}^{(1)} & t_{0,f2G\_2}^{(1)} & t_{0,f3G\_1}^{(1)} & t_{0,f3G\_2}^{(1)} & t_{0,f4G\_1}^{(1)} & t_{0,f4G\_2}^{(1)} \end{bmatrix} \quad (3)$$

where the first two elements are the traffic data of the 2G frequency layers, the third and fourth elements are the traffic information of the 3G frequency layers and the last two ones are related to the traffic on the 4G frequency layers. The superscript $^{(1)}$ refers to the considered time sample (i.e., 1^{st} time sample). The vector indicating how the traffic in s_1 is distributed after steering is $\underline{\mathbf{t}}_{1,s1}^{(1)}$ and is obtained by relating $\underline{\mathbf{t}}_{0,s1}^{(1)}$ to a 6 x 6 permutation matrix $\underline{\underline{\mathbf{M}}}_{s1}^{(1)}$ which represents the mathematical form of the EE feature for sector s_1 at the 1^{st} time sample, as in formula (4):

$$\underline{\mathbf{t}}_{1,s1}^{(1)} = \underline{\mathbf{t}}_{0,s1}^{(1)} \cdot \underline{\underline{\mathbf{M}}}_{s1}^{(1)} = \tag{4}$$
$$= \begin{bmatrix} t_{1,f2G\_1}^{(1)} & t_{1,f2G\_2}^{(1)} & t_{1,f3G\_1}^{(1)} & t_{1,f3G\_2}^{(1)} & t_{1,f4G\_1}^{(1)} & t_{1,f4G\_2}^{(1)} \end{bmatrix}$$

$\underline{\underline{\mathbf{M}}}_{s1}^{(1)}$ and $\underline{\mathbf{t}}_{1,s1}^{(1)}$ are computed iteratively by checking that the two conditions C_1 and C_2 are satisfied when steering the traffic from the 2G/3G (or both) frequency layers towards the 4G ones. In first step of the iterative process the permutation matrix $\underline{\underline{\mathbf{M}}}_{s1}^{(1)}$ is equal to the 6 x 6 identity matrix as in (5), i.e.:

$$\underline{\underline{\mathbf{M}}}_{s1}^{(1)} = \underline{\underline{\mathbf{M}}}_{1}^{(1)} = \underline{\underline{\mathbf{I}}}_{6} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where the subscript "1" in $\underline{\underline{\mathbf{M}}}_{1}^{(1)}$ refers to the first step of the iterative algorithm. By indicating the normalized pre-steering traffic data on the i^{th} frequency layer of the *X*G RAT (*X*=2,3,4) at the 1st time sample as $x_{0,fXG\_i}^{(1)}$ and, similarly, the normalized post-steering traffic data on the i^{th} frequency layer of the *X*G RAT (*X*=2,3,4) at the 1st time sample as $x_{1,fXG\_i}^{(1)}$, if the traffic data originally handled by the 2G frequency layers only is assumed to be hypothetically steered and managed by the 4G ones (provided that conditions C_1 and C_2 are satisfied) at the 2nd step of the algorithm, the normalized traffic data of the first 2G frequency layer (i.e., $x_{0,f2G\_i}^{(1)}$) must be less than or equal to th_1 and, at the same time, the normalized post-steering traffic data in charge of one between the two 4G frequency layers (i.e., $x_{1,f4G\_1}^{(1)}$ or $x_{1,f4G\_2}^{(1)}$, chosen according to post-steering saturation level) must be less than or equal to th_2. In formula (6):

$$\text{if } \left\{ x_{0,f2G\_1}^{(1)} \le th_1 \text{ AND } \left(x_{1,f4G\_1}^{(1)} \le th_2 \text{ OR } x_{1,f4G\_2}^{(1)} \le th_2 \right) \right\} \rightarrow$$
$$\rightarrow t_{1,f4G\_i}^{(1)} = t_{0,f4G\_i}^{(1)} + t_{0,f2G\_1}^{(1)} \tag{6}$$

If the above statement is fulfilled the traffic data of the first 2G frequency layer is steered towards one (and only one) of the two available 4G frequency layers. Assuming that the first 4G frequency layer can accept the traffic originating from the first 2G layer, the permutation matrix at the 2nd step of the algorithm (i.e., $\underline{\underline{\mathbf{M}}}_{2}^{(1)}$) is expressed as in (7):

$$\underline{\underline{\mathbf{M}}}_{2}^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7}$$

and, consequently, the post-steering non-normalized traffic data vector of sector s_1 at the 2nd step of the iterative process is as follows in (8):

$$\underline{\mathbf{t}}_{1,s1\_step2}^{(1)} = \underline{\mathbf{t}}_{0,s1}^{(1)} \cdot \underline{\underline{\mathbf{M}}}_{2}^{(1)} = \tag{8}$$
$$= \begin{bmatrix} 0 & t_{1,f2G\_2}^{(1)} & t_{1,f3G\_1}^{(1)} & t_{1,f3G\_2}^{(1)} & t_{1,f4G\_1}^{(1)} + t_{0,f2G\_1}^{(1)} & t_{1,f4G\_2}^{(1)} \end{bmatrix}$$

Equation (8) shows that the first 2G frequency layer in sector s_1, for the 1st time sample, can be switched-off since its traffic is now managed by, e.g., the first 4G layer.

The possibility to steer the traffic data of the second 2G frequency layer towards one (and only one) of the available 4G frequency layers is analyzed in a similar manner. The overall algorithm must be performed for the remaining sectors, leading to a set of permutation matrices $\underline{\underline{\mathbf{M}}}_{s\_i}^{(1)}$ (with *s_i = 1,…, N*, where *N* represents the number of sectors within the cluster of network sites) for the 1st time sample only; the same approach must be followed for the remaining time samples within the observation period. A description of the intra-sector traffic steering algorithm in a pseudo-code form is reported in Fig. 1.

```
for each time sample        % cycle on time samples (1 ≤ time sample ≤ 96)
   for each sector          % cycle on each sector of the evaluated cluster of sites (1 ≤ sector ≤ 57)
         M^{time sample}_{sector} = I_6      % initialization of the permutation matrix
      for each 4G layer             % (two 4G frequency layers in each sector)
         for each xG layer               % (x = 2, 3)
               % identification of the candidate xG frequency layer for traffic
               steering towards one of the available 4G frequency layer within the
               considered sector (condition C1)
            if normalized traffic X(time sample, xG layer) < th1
                     % identification of the candidate 4G frequency layer able to
                     handle the xG frequency layer's traffic to be steered
                     (condition C2)
               if normalized traffic X(time sample, 4G layer) =
                     = {[T(time sample, 4G layer) + T(time sample, xG layer)] /
                     Capacity4G} < th2
                        % traffic steering of the xG frequency layer towards
                        the 4G one (with subsequent normalization) is
                        actually performed
                        normalized traffic X(time sample, 4G layer) =
                        = [T(time sample, 4G layer) + T(time sample, xG
                        layer)] / Capacity4G
                        % xG frequency layer's traffic zeroing
                        normalized traffic X(time sample, xG layer) = 0
                        % update of the permutation matrix
                        M^{timesample}_{sector}(xG layer, xG layer) = 0
                        M^{timesample}_{sector}(xG layer, 4G layer) = 1
               end
            end
         end
      end
   end
end
```

Fig. 1 – Intra-sector traffic steering algorithm.

IV. EVALUATION METHODOLOGY

EE evaluation conducted in this study is based on the assessment of performance of a homogeneous cluster of sites, by using actual traffic profiles extracted from the live network and by computing the cluster's energy consumption under different situations, according to proper PMs of the considered BSs.

A. *Network data traffic profiles*

In order to provide useful insights for operators enabling them to evaluate suitable actions in the actual network, energy performance of the cluster have been evaluated in realistic conditions. To be specific, a set of data traffic profiles related to a BS currently running in the Telecom Italia mobile network was extracted by a network monitoring system having a fixed time resolution. In particular, for each sector, a single data sample automatically provided by the monitoring system represents the average value of a specific Key Performance Indicator (KPI) over a 15 minutes time interval (e.g., for data bandwidth, a single sample represents the amount of data transferred in 15 minutes), hence a daily profile is represented by 96 consecutive values. Each sample is also averaged over the 5 working days of the week, in order to filter effects due to spurious data peaks and increase the reliability of the provided profile.

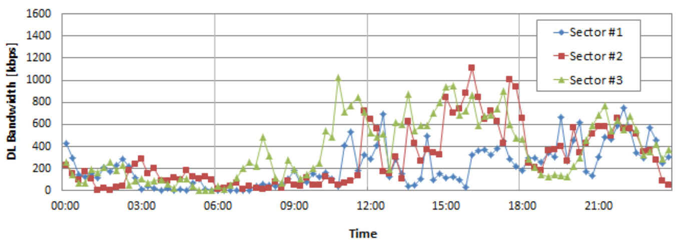


Fig. 2 – Example of daily traffic profiles extracted from the live network (Downlink bandwidth, 3 HSDPA sectors).

Fig. 2 shows an example of the data profiles used, related to an installed BS running in the Telecom Italia mobile network. The considered BS for data profiles' retrieving is a 3-sectorial macro site located in an urban environment, equipped with GSM 900, GSM 1800, UMTS, HSDPA and LTE 1800 frequency layers, capturing all operator's traffic demand in the covered area (2G, 3G, 3.5G and 4G). In the plot, downlink bandwidth for the 3 sectors is showed only, even if other traces can be extracted from the live network in order to analyze other KPIs. The trend of extracted data profiles showed a typical daily oscillation of the traffic, consistent with average profiles contained in [10] and [11], but with the important difference that while literature curves are averaged over an entire network, data profiles used in this paper are related to a single site, in order to better show particular burst effects of the traffic variation. The resulting curves are then suitable for a specific evaluation of the site's energy consumption, based on actual traffic load. Moreover, these profiles are also suitable for the evaluation of EE features (e.g. like those proposed in [8] and ON-OFF schemes tested in [9] with commercial BSs).

B. *Power consumption evaluation*

Power consumption of a BS is determined by means of a proper PM of the BS itself. Basically, a PM is a set of parameters describing the BS's power consumption P_{in} as a function of the Radio Frequency (RF) output power P_{out}, which is directly related to the normalized traffic load x.

In [1] PMs have been evaluated for macro BSs and they are available for years coming from 2010 to 2020; furthermore, provided PMs are RAT-based (i.e., there is a specific PM for a 2G/3G/4G macro BS) and defined by means of three parameters termed A, B and C, as depicted in Fig. 3.

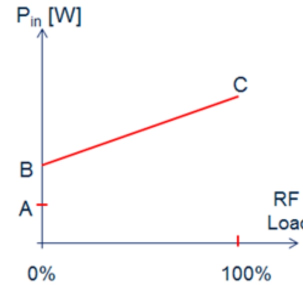


Fig. 3 – PM parameters definitions.

Now, since for our study a mathematical expression to compute the BS's power consumption P_{in} as a function of the normalized traffic information is needed, the original formula developed by EARTH project in [12] has been considered and properly modified as follows:

$$P_{in} = \begin{cases} N_{TRX} \cdot P_0 + N_{TRX} \cdot \Delta P \cdot P_{out}, & 0 < P_{out} \le P_{MAX} \\ N_{TRX} \cdot P_{sleep}, & P_{out} = 0 \end{cases} \quad (9)$$

where P_{MAX} denotes the maximum RF output power at maximum load, N_{TRX} is the number of transceiver chains, P_0 is the linear model parameter representing the power consumption at the zero RF output power, ΔP is the slope of the load-dependent power consumption and P_{sleep} is the sleep mode power consumption (i.e., the BS's power consumption achieved when there is nothing to transmit, so fast deactivation of the BS's components is selected). Under the hypothesis that P_{out} varies linearly with the throughput, the normalized throughput x with respect to the peak throughput of each considered RAT is introduced in the EARTH formula (9), leading to the following expression (10):

$$P_{in} = \begin{cases} N_{TRX} \cdot P_0 + N_{TRX} \cdot \Delta P \cdot P_{MAX} \cdot x, & 0 < x \le 1 \\ N_{TRX} \cdot P_{sleep}, & x = 0 \end{cases} \quad (10)$$

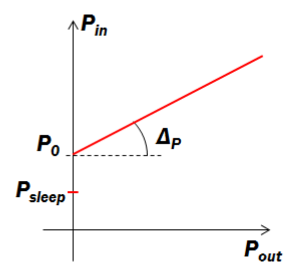


Fig. 4 – Original EARTH PM definition.

Fig. 4 shows physical meaning of the BS's PM parameters as provided by EARTH.

By comparing graphs in Fig. 3 and Fig. 4 equalities in expression (11) can be written:

$$\begin{cases} A = N_{TRX} \cdot P_{sleep} \\ B = N_{TRX} \cdot P_0 \\ C = N_{TRX} (P_0 + \Delta P \cdot P_{MAX}) = \\ \quad = N_{TRX} \cdot P_0 + N_{TRX} \cdot \Delta P \cdot P_{MAX} = B + N_{TRX} \cdot \Delta P \cdot P_{MAX} \end{cases} \quad (11)$$

Hence, the formula to be used in order to compute the BS's power consumption according to the selected PM is as in (12):

$$P_{in} = \begin{cases} B + (C - B) \cdot x, & 0 < x \leq 1 \\ A, & x = 0 \end{cases} \tag{12}$$

A PM for future 5G BSs has also been derived for performance evaluation when considering 2G/3G phase-off and contextual new 5G radio interface introduction. To this end, the PM of a 2010 pico BS reported in [12] has been considered as a starting point and, by assuming an yearly improvement of 8% (in accordance with [13]) for both A and B PM's parameters (while the C parameter is not subjected to any kind of improvement), the PM of a theoretical 5G BS can be obtained. This average annual improvement (taken as a reference throughout the study as baseline continuous improvement) can be attributed to the technology scaling of semiconductors, as well as to improved RATs. BSs PMs have been used within the evaluation tool described in the section below, together with the daily traffic profiles extracted from the live network.

C. Description of evaluation methodology and related tool

Fig. 5 shows a graphical representation of the MATLAB-based evaluation platform used for EE performance evaluation. The method consists in evaluating the performance of baseline and frontline systems, both assessed by feeding a cluster of sectors with daily traffic profiles and by modeling energy consumption with PMs previously described.

Then, EE benefits can be computed as relative gains obtained in terms of energy consumption; this because both baseline and frontline systems are fed by the same traffic amount, hence it is straightforward to notice that relative difference in terms of EE is equivalent to the relative difference in terms of energy consumption.

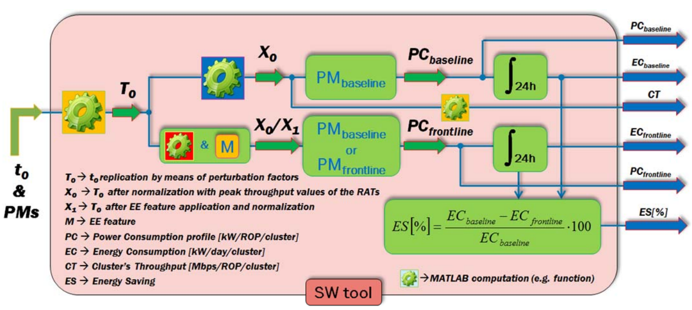


Fig. 5 – Graphical representation of the MATLAB-based platform.

In more detail, the following processing of traffic profiles is performed by the evaluation tool. Traffic data matrix for all RATs' frequency layers related to a single trisectorial site, that is $\underline{\mathbf{t}}_0$, and the BSs' PMs represent the inputs of the evaluation platform. The former is then used to generate the non-normalized traffic data matrix $\underline{\underline{\mathbf{T}}}_0$ of the whole cluster of network sites exploiting a proper set of perturbation factors chosen in such a way that the resulting traffic distribution among all sites in the cluster is, on average, as much homogeneous as possible. Since the PM application requires the traffic data information to be normalized with respect to the peak throughput values of the considered RATs, the normalized version of $\underline{\underline{\mathbf{T}}}_0$, indicated as $\underline{\underline{\mathbf{X}}}_0$, is computed by relating each entry of $\underline{\underline{\mathbf{T}}}_0$ to the peak throughput value of the RAT it belongs to. Hence, by definition, matrix $\underline{\underline{\mathbf{X}}}_0$ has the same dimensions of $\underline{\underline{\mathbf{T}}}_0$. Both $\underline{\underline{\mathbf{X}}}_0$ and the PMs of the currently deployed BSs allow for performance evaluation of the so-called baseline system, in which neither EE features nor evolved BSs' PMs are taken into account. When an EE feature (indicated in Fig. 5 as $\underline{\underline{\mathbf{M}}}$) is chosen to be evaluated, a new normalized traffic data matrix, namely $\underline{\underline{\mathbf{X}}}_1$, is determined starting from $\underline{\underline{\mathbf{T}}}_0$; this matrix can be used for frontline systems' performance assessment of various scenarios in which evolved PMs can be jointly considered or not.

Note that another frontline scenario is the one in which any kind of EE feature is taken into account (i.e., the normalized traffic data matrix used for performance assessment is still $\underline{\underline{\mathbf{X}}}_0$) but evolved PMs are considered only, allowing for the achievable energy saving to be quantified as a consequence of new network equipment deployment.

For baseline and each frontline systems a daily power consumption profile ($PC_{baseline}$ and $PC_{frontline}$, respectively) is determined and, by integrating it on a daily basis, the related energy consumption ($EC_{baseline}$ and $EC_{frontline}$, respectively) is assessed. Energy consumptions values are then compared by computing an Energy Saving metric (*ES*). Graphically shown performance metrics, all determined on a daily basis, are:

- total cluster's power consumption (in *kW*)
- total cluster's throughput (in *Mbps*)
- RAT-based power consumption of each frequency layer (in *kW*)
- daily total cluster's energy consumption (in *kWh*)

At the end of each simulation all the computed metrics and variables within the MATLAB workspace are saved in a proper ".mat" file.

V. Evaluation Results

Several simulation campaigns have been performed in order to obtain different evolutionary scenarios of the network's power consumption through the years. In particular, Fig. 6 reports the daily average energy consumption (in *kWh*) related to the considered cluster of mobile network sites for the three baseline systems under evaluation. Daily data traffic (in *Pb*) is also shown, whilst Table 1 lists the baseline evaluation scenarios, in which different BSs replacement plans are considered and any kind of traffic steering option is taken into account. As can be observed, main improvements in terms of energy savings, with respect to the baseline system #1 ("Business as Usual", where the replacement of highly power consuming BSs with "*greener*" ones is not considered at all), can be achieved when considering network equipment replacement, as supposed in baseline system #2 ("Network renewal in 2017", where stepwise replacement of older BSs with new ones whose power consumption is modelled by means of the PM of the year 2016 is considered). In other words, a complete network equipment replacement in 2017 with BSs performing better in terms of power consumption and modelled with the PM of the year 2016 will result in high savings also in subsequent years. Furthermore, a possible decision on a second round of network equipment replacement from 2021 with BSs modelled with the PM of the year 2020 (indicated as baseline system #3, "Continuous network renewal") can be postponed at a time when sufficiently high energy saving can be reached, thus motivating a second round of investment. Besides, in 2020 or later on, the introduction of a new 5G RAT would represent another investment concurrent to the renewal of legacy equipment (thus, it is not likely to foresee at this time baseline #3 as the most probable scenario, due to the possible presence of two concurrent investments, not necessarily affordable for the operator). As a consequence, the baseline system #2 can be considered as the best solution, since it provides short term benefits with less costs when compared to the baseline system #3 and a 40% energy saving with respect to the baseline system #1, to be achieved in 2020 (note: in any case energy consumption of these baselines in 2020 is still higher than energy consumption in 2014).

In Fig. 7 evolutionary comparisons of different traffic steering solutions with respect to the evaluated baseline system #2 ("Network renewal in 2017") are shown. It can be observed that benefits due to the application of traffic steering from a single legacy RAT (2G or 3G) to 4G are similar, due to similar performance of the legacy RAT (according to the related PM). Furthermore, by jointly steering the traffic of both 2G and 3G RATs towards 4G, additional energy savings can be reached, with similar trends in terms of energy consumption decrease. Note that these traffic steering solutions do not require any investment for the operator and they may be implemented since 2014, allowing for short term benefits' achievement. When considering traffic steering options in combination with network renewal as thought for the baseline system #3 (i.e., continuous network renewal), the yearly energy consumption show similar trend as in Fig. 7. Hence, as previously stated, this kind of solution can be discarded since, at higher costs for the operator, it does not provide additional benefits.

In Fig. 8 cluster's energy consumption performance obtained by jointly considering network renewal in 2017 and traffic steering options implemented since 2014, with respect to the baseline system #1 performance, are reported.

It is worth noting that traffic steering options introduced and analyzed in Fig. 7 and Fig. 8 are evaluated by assuming, in the considered cluster, the presence of terminals equipped with all legacy radio chains (i.e. 2G, 3G and 4G). Of course, the presence of different kind of terminals and the related impact of such a constraint as a limiting factor for EE gains can be considered in further operator's evaluations (which is not taken into account in the present study). These evaluations could be done, e.g., with the aim to perform a fully compliant implementation of the methodology reported in ETSI ES 203 228, where different kind of clusters are assessed and the single evaluation is obtained by performing a weighted addition of different clusters' energy consumptions for country-wide EE assessments.

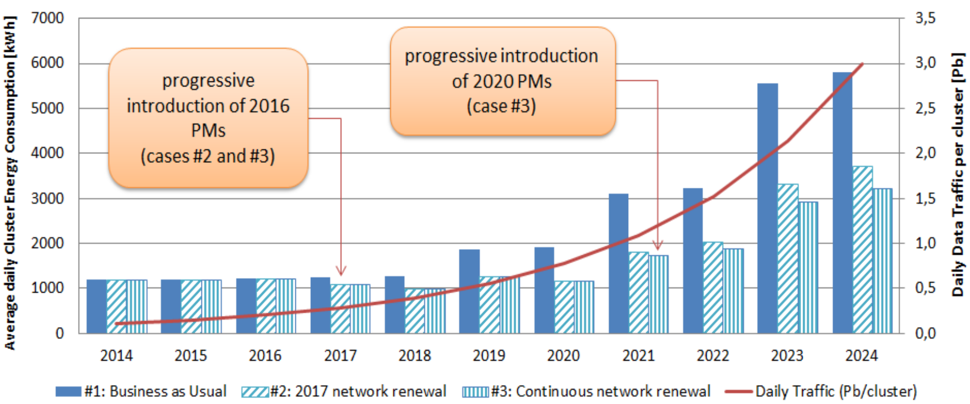
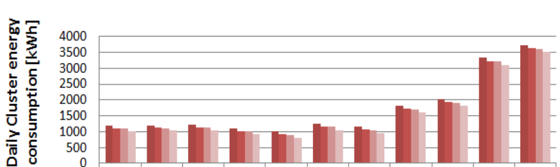


Fig. 6 – Evolutionary scenarios of network power consumptions; daily traffic evolution according to [16].

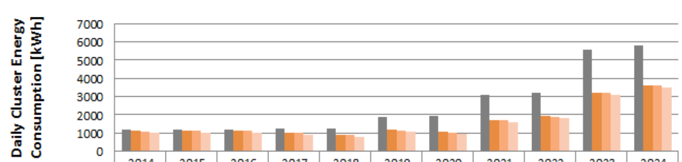
Table 1 – Baseline evaluation scenarios.

Baseline system #1 (Business as Usual)	• from 2014 to 2024: BSs' power model of the year 2012 is always considered
Baseline system #2 (Network renewal in 2017)	• from 2014 to 2016 as in "Baseline system #1" • from 2017 to 2020: new BSs modeled with the power model of the year 2016 are progressively deployed • from 2021 to 2024: BSs' modeled with the power model of the year 2016 are considered only
Baseline system #3 (Continuous network renewal)	• from 2014 to 2020 as in "Baseline system #2"; • from 2021 to 2024: newer BSs modeled with the power model of the year 2020 are progressively deployed



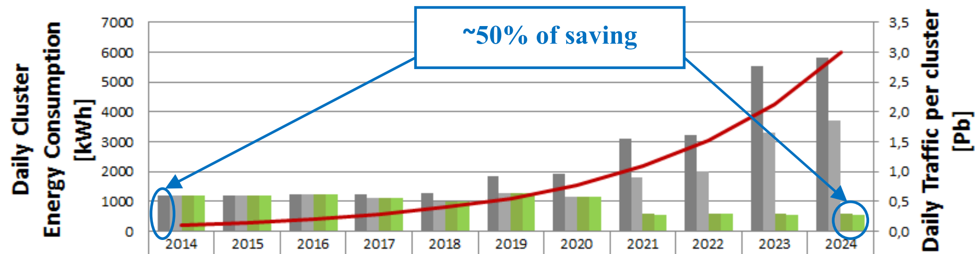
	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
baseline 2	1191,07	1203,10	1219,95	1102,49	1004,39	1260,71	1164,25	1818,17	2021,19	3323,22	3721,15
2G->4G traffic steering	1110,86	1122,90	1139,74	1017,89	915,38	1167,31	1066,45	1720,37	1923,39	3225,42	3623,35
3G->4G traffic steering	1095,48	1107,51	1124,35	1001,35	897,70	1148,47	1046,46	1700,38	1903,40	3205,43	3603,36
2G&3G->4G traffic steering	1015,27	1027,30	1044,15	916,75	808,69	1055,07	948,66	1602,58	1805,60	3107,63	3505,56

Fig. 7 – Comparison of traffic steering options Vs baseline system #2.



	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
baseline 1	1191,07	1203,10	1219,95	1243,53	1276,54	1860,84	1925,54	3092,29	3219,12	5548,99	5797,56
2G->4G traffic steering	1110,86	1122,90	1139,74	1017,89	915,38	1167,31	1066,45	1720,37	1923,39	3225,42	3623,35
3G->4G traffic steering	1095,48	1107,51	1124,35	1001,35	897,70	1148,47	1046,46	1700,38	1903,40	3205,43	3603,36
2G&3G->4G traffic steering	1015,27	1027,30	1044,15	916,75	808,69	1055,07	948,66	1602,58	1805,60	3107,63	3505,56

Fig. 8 – Comparison of traffic steering options Vs baseline system #1.



	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
baseline 1	1191,07	1203,10	1219,95	1243,53	1276,54	1860,84	1925,54	3092,29	3219,12	5548,99	5797,56
baseline 2	1191,07	1203,10	1219,95	1102,49	1004,39	1260,71	1164,25	1818,17	2021,19	3323,22	3721,15
PhaseOff 2G & 5G introduction @2021 (PM2016 from 2017 on)	1191,07	1203,10	1219,95	1102,49	1004,39	1260,71	1164,25	608,78	609,13	609,62	610,31
PhaseOff 3G & 5G introduction @2021 (PM2016 from 2017 on)	1191,07	1203,10	1219,95	1102,49	1004,39	1260,71	1164,25	566,76	576,12	567,61	568,3
Daily Traffic (Pb/cluster)	0,1035	0,1448	0,2028	0,2839	0,3974	0,5564	0,7789	1,0905	1,5267	2,1374	2,9924

Fig. 9 – Phase-off of old RAT and progressive 5G introduction; daily traffic evolution according to [16].

As expected, the combination of both traffic steering and stepwise network renewal permit to achieve short term benefits in terms of energy consumption by minimizing, at the same time, the investment for the network operator. The holistic solution, of course, is still able to satisfy all traffic demand in the network, allowing for huge energy savings with respect to baseline #1 (38% for 2G/3G traffic steering and 40% for 2G&3G traffic steering, respectively, in 2024), even if power consumption in 2024 is still higher, in absolute terms, than the one in 2014. This can represent a motivation for 5G introduction, in order to further save energy by considering 4G offloading towards the new RAT, satisfying the additional traffic.

Finally, in Fig. 9, the cluster's energy consumption performance obtained by jointly considering network renewal in 2017 and 5G introduction from 2021 with contemporary phase-off of a legacy RAT, with respect to both baseline systems #1 and #2, are provided. When 5G is deployed, the power consumption of remaining RATs (2G/3G and 4G) is still computed by considering BSs' PM of the year 2016.

It's worth noting that evaluation results in Fig. 9 are expressed in terms of cluster's energy consumption; when comparing the evaluated scenarios also in terms of area power consumption as EE metric (described in section II), it can be found that, starting from an area power consumption of 12.1 kW/km^2 in 2014 (obtained by considering a dense urban scenario with inter-site distance of 500 *m* and all deployed RATs, i.e., 2G, 3G and 4G), a reduction of nearly 50% of the adopted EE metric can be achieved in 2024, with full deployment of 5G and phase-off of a legacy RAT (either 2G or 3G). In fact, the area power consumption reduces to 6.2 (5.8) kW/km^2 with 5G fully *on-field* and 2G (3G) RAT phased-off. In addition, it should be highlighted that evaluation results reported within the paper have been obtained by keeping both cluster's area and served traffic as constants, hence evaluations have simply been compared in terms of energy consumption only, in order to put the year-by-year evolutionary trend on evidence.

Note that all previous evaluations have been performed by computing the RAN power consumption only, even if, in practical cases, operator costs are also influenced by infrastructure within mobile sites. Moreover, it is likely that the introduction of 5G networks will be influenced by the introduction of Cloud-RAN (C-RAN) as a new architectural paradigm able to save energy thanks to the adoption of a *greener* centralized infrastructure. Thus, the contribution to the energy consumption by considering not only the RAN part, but also infrastructure equipment present in mobile sites (air conditioning, power supply, and so on) has also been preliminary assessed.

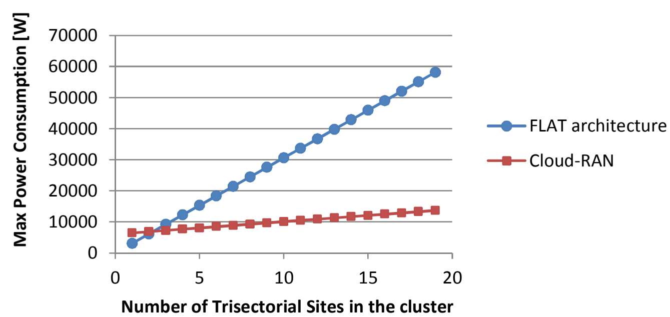


Fig. 10 – First exemplary comparison of cluster power consumption between 4G flat and C-RAN architectures (LTE MIMO 2x2 systems).

Results of this preliminary evaluation, based on further PMs elaborated in [14], are shown in Fig. 10, where an exemplary comparison in terms of power consumption between a traditional deployment of 4G sites (flat architecture without any centralization) and an equivalent cluster in C-RAN architecture (with a power consumption of the infrastructure equipment of about 2 *kW*, as in [15]) is presented. Note that, by increasing the number of cells in the cluster, the benefits of C-RAN become evident and high savings are essentially due to the infrastructure part, which is concentrated in the data center.

VI. CONCLUSIONS AND FUTURE WORKS

Within this paper, techniques or actions that the network operator should consider for the mobile network's evolution towards 5G from a sustainability point of view, especially taking into account the yearly RAN's energy consumption in different load conditions, have been evaluated. The evolutionary energy consumption of the RAN has been assessed by considering RAT-based PMs of future BSs, following the methodology used in [1] and results from the EARTH project [2], together with EE features evaluated at higher time scale (from minutes to years) e.g., related to progressive network renewal with traffic steering options and legacy RATs' phase-off policies. Evaluations have been conducted by using a software tool fed by actual traffic profiles from the live network, assessing EE performance at cluster level, giving useful information on the possible evolutions of the mobile networks in the view of future 5G systems. Possible future evolution of the present study may include the modeling of DTX techniques at faster time scale (i.e. up to milliseconds, not evaluated in the present paper), in order to provide a more complete overview of the different possible features applicable to the mobile networks.

Moreover, an evolution of this study may include the full implementation of the methodology introduced in ETSI ES 203 228 specification, with a complete assessment of different clusters of cells in order to provide country-wide EE evaluations. In addition, the preliminary results obtained so far are also encouraging us to further work on yearly evolutions of mobile network's EE towards 5G and a possible future step could be a proper modelling of the resource pooling in C-RAN systems, that may offer further benefits in terms of energy savings. Finally, in the view of 5G systems, also M2M infrastructure effects could be considered in order to have a more complete view. It should be stressed that traffic steering among frequencies and RATs, aiming at reducing the energy consumption, can be already exploited in several real scenarios. For example, in business areas, the deployments have to be performed in a way to guarantee high data traffic, but outside office hours such resources are pretty much unused, e.g., allowing for a switch-off of 4G layers. Another interesting use case is related to M2M devices used for data reporting, like in smart metering scenarios. This type of terminal needs to report specific measured parameter without constraints in terms of latency and timing. Hence, the reporting can be enabled according to the switch-on time of the frequency layer/cell depicted for such a scope. Moreover, many of these devices works on legacy network like 2G system, hence the related reporting can be synchronized with the relevant switch-on/off of such a layer.

ACKNOWLEDGMENTS

This work has been carried out within the framework of the 5GrEEN project (http://wireless.kth.se/5green/), funded by EIT ICT Labs. We gratefully recognize the great collaboration of 5GrEEn partners, who contributed with valuable insights, surveys and vision. Special thanks to Magnus Olsson, Pal Frenger and Cicek Cavdar.

REFERENCES

[1] P. Frenger, Y. Jading, J. Turk, "A Case Study on Estimating Future Radio Network Energy Consumption and CO_2 Emissions," in Proc. PIMRC 2013.

[2] EARTH project deliverable D6.4, June 2012, "Final integrated concept", https://bscw.ict-earth.eu/pub/bscw.cgi/d49431/EARTH_WP6_D6.4.pdf

[3] EARTH project deliverable D2.4, June 2012, "Most suitable efficiency metrics and utility functions", https://bscw.ict-earth.eu/pub/bscw.cgi/d70454/EARTH_WP2_D2.4.pdf

[4] S. Tombaz, "On the Design of Energy Efficient Wireless Access Networks", PhD Thesis in Information and Communication Technology, Stockholm, Sweden 2014, p. 24, http://www.diva-portal.org/smash/get/diva2:714913/FULLTEXT02

[5] S. Tombaz, K.W. Sung, J. Zander, "On Metrics and Models for Energy Efficiency in Wireless Access Networks", Comm. Letters, August 2014.

[6] Ericsson, "Further details on design principle for a CRS-free additional carrier type", R1-121017, 3GPP RAN1 #68bis, 26-30 Mar 2012

[7] EARTH project deliverable D3.3, June 2012, "Final Report on Green Radio Technologies", https://bscw.ict-earth.eu/pub/bscw.cgi/d70472/EARTH_WP3_D3.3.pdf

[8] D. Sabella et al.: "Evaluation of ON-OFF schemes and Linear Prediction methods for increasing energy efficiency in Mobile Broadband Networks", Networking 2012, Prague, Czech Republic, May 25, 2012

[9] W. Tomaselli, V. Palestini, D. Sabella, V. Squizzato: "Energy efficiency performances of selective switch OFF algorithm in LTE mobile networks", London, PIMRC 2013

[10] I. Godor et al.: "Green Communications: Theoretical Fundamentals, Algorithms and Applications", book chapter on "Green Wireless Access Networks"; Publisher: CRC Press; 1 edition (September 20, 2012)

[11] G.Auer, et al. "Cellular Energy Efficiency Evaluation Framework", Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd

[12] EARTH project deliverable D2.3, January 2012, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown", https://bscw.ict-earth.eu/pub/bscw.cgi/d71252/EARTH_WP2_D2.3_v2.pdf

[13] EARTH project deliverable D2.1, April 2011, "Economic and ecological impact of ICT", https://bscw.ict-earth.eu/pub/bscw.cgi/d38532/EARTH_WP2_D2.1_v2.pdf

[14] D. Sabella et al., "Energy Efficiency benefits of RAN-as-a-Service concept for a cloud-based 5G mobile network infrastructure", IEEE Access, December 2014

[15] China Mobile Research Institute: "C-RAN: The Road Towards Green RAN", white paper, http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN.pdf

[16] Ericsson, "Ericsson Mobility Report – On the pulse of the networked society", Nov. 2014, http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-november-2014.pdf

[17] Laetitia Souchon Foll, PhD Thesis: "TIC et Énergétique : Techniques d'estimation de consommation sur la hauteur, la structure et l'évolution de l'impact des TIC en France", Laboratoire TECH de Orange Labs, October 2008, http://www.biblioite.ethz.ch/downloads/Souchon_these_version-publique.pdf

[18] S. Tombaz, A. Vastberg and J. Zander, "Energy and Cost Efficient Ultra High Capacity Wireless Access," IEEE Wireless Communication Magazine, Vol. 18, No. 5, 2011

[19] C. Han, T. Harrold and S. Armour, "Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks", IEEE Communications Magazine, Volume 49, Issue 6, p.p.: 46 - 54, May 2011

[20] Ericsson: "Ericsson Energy and Carbon Report - On the impact of the networked society", June 2013 http://www.ericsson.com/res/docs/2013/ericsson-energy-and-carbon-report.pdf

[21] GSMA Mobile Energy Efficiency, http://www.gsma.com/publicpolicy/mobile-energy-efficiency

[22] ETSI EE European Specification, "ES 203 228 - Assessment of mobile network energy efficiency", January 2015 http://www.etsi.org/deliver/etsi_es/203200_203299/203228/01.00.00_50/es_203228v010000m.pdf