Energy Saving: Scaling Network Energy Efficiency Faster than Traffic Growth

Yan Chen, Oliver Blume, Azeddine Gati, Antonio Capone, Chi-En Wu,
Ulrich Barth, Tom Marzetta, Haibin Zhang, and Shugong Xu

***Abstract*—As the mobile traffic is expected to continue its exponential growth in the near future, energy efficiency has gradually become a must criterion for wireless network design. Three fundamental questions need to be answered before the detailed design could be carried out, namely what energy efficiency network is, where the energy efficiency improvement may come, how the energy efficiency evaluation of a whole mobile network could be conducted. In this paper, we target to shed a light on the answers of the three questions by introducing the bottom-up energy efficiency evaluation methodology for large networks with the combination of different areas and time periods, as well as the basic traffic, deployment, and power models that will be applied therein. Moreover, we shall illustrate where the overall energy saving gain may come by analyzing the power-traffic relation and the three research projects in the GreenTouch Mobile WG.**

I. INTRODUCTION

Since 2006, data traffic on wireless networks has grown by approximately 400% and is expected to continue to increase rapidly in the coming years [1] [2]. The widespread use of complex, spectrum efficient techniques to support such high data volumes, the demand for higher data rates, and the ever-increasing number of wireless users, together translate to rapidly rising power consumption. Currently consuming 3% of the global energy and causing 2% of the CO2 emissions globally, the ICT industries are facing an increase in associated energy consumption of $16-20\%$ per year. Furthermore, the energy costs for mobile operators can be as high as half of their annual operating budgets. The foregoing considerations highlight the urgent need for energy efficient design of future wireless networks.

Here comes the first question: *what is energy efficient network?* We start to answer this question by defining the metric energy efficiency. Literally, it means the efficiency of using energy to deliver information, which is usually expressed in the unit of "bits per joule". Different from the widely known concept of spectrum efficiency, defined as the transmission capacity achieved in bits per second over the spectrum resource used, energy efficiency is not only related to the transmission capacity, but also the traffic demand. In particular, the real traffic need might be far less than the network transmission capacity, causing the energy consumption for over provision of capacity. Moreover, the traffic demand varies in time and space according to the user distribution and activities, which makes the idea of providing a fixed capacity close to the traffic demand impractical. As we will show later in the paper, increasing capacity alone will not be enough to boost the network energy efficiency, especially in the mid to low traffic scenarios. Therefore, the expected energy efficient network should not only have large enough capacity to efficiently accommodate fast increasing traffic, but also be capable of tuning its capacity according to the varying traffic need to reduce over provision in a dynamic manner, and together, minimizing the energy consumption per bit for both high and low traffic scenarios.

The second question that naturally follows up is, *how to design energy efficient networks?* To answer this question, we first need to understand where the gains may come from, which will be elaborated in section II by a closer look at the power-traffic relation. Secondly, we need to know how much freedom we have in the network design, according to which, we could roughly have two categories. The first one targets at the energy efficiency oriented optimization for existing networks (e.g., UMTS or LTE networks) with comparatively mature standards which allow only progressive modifications. A good example is the EARTH[1] project, which develops optimized deployment, radio management, and transmission strategies to reduce the energy consumption of the current LTE/LTE-A radio access networks by half [3]. The second one, on the other hand, takes a totally different approach and consider designing the networks from scratch. A typical example is the GreenTouch[2] (GT) Consortium, which is dedicated to fundamentally transforming communications and data networks towards energy-efficiency oriented design philosophy with a goal set to deliver the architecture, specifications and roadmap needed to increase end-to-end network energy efficiency by a factor of 1000 from current levels. In section IV, we shall go back to visit this question by introducing the research on green architectures and technologies that is ongoing in the GT Mobile WG.

Then the third question to ask is, *how to quantitatively measure the gain in network energy efficiency?* The wireless network contains rich diversity of the wireless access technolo-

This paper is supported by the joint research of the Architecture/Metric study group in the Mobile WG of GreenTouch.

Yan Chen (corresponding author, email: bigbird.chenyan@huawei.com) and Shugong Xu are with Huawei Technologies, Oliver Blume, Ulrich Barth and Tom Marzetta are with Alcatel-Lucent Bell Labs, Azeddine Gati is with Orange Labs, Antonio Capone is with Politecnico di Milano, Chi-En Wu is with Chunghwa Telecom, and Haibin Zhang is with Netherlands Organization for Applied Scientific Research.

[1]EARTH is an EU IP project with the full name Energy Aware Radio TecHnolgies, see https://www.ict-earth.eu/.

[2]GreenTouch is a consortium of leading industry, academic and non-governmental research experts in ICT area for future end-to-end green design, see https://www.greentouch.org.

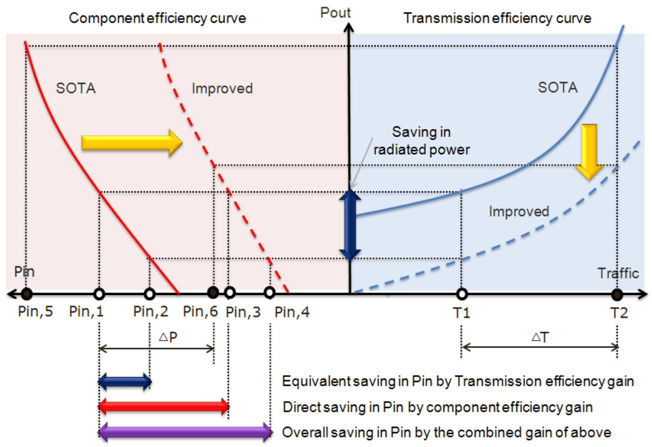


Fig. 1. The illustrative sketch of power-traffic relations.

gies, formed by various wireless base stations, controllers, and wireless access points, operating on different frequency bands and providing different service coverage. To characterize the energy efficiency performance of the wireless network, three fundamental models are needed, namely the traffic model, the deployment model, and the power model. However, it is comparatively easier to specify some number for spectrum efficiency requirements. This is mainly because the Shannon formula provides us with the upper bound of transmission capacity with given number of antennas, transmit power, and system bandwidth, and as the coding and modulation schemes develop, the system spectrum efficiency can be very close to that limit already. On the contrary, as we will show in section III, it is hard to give such values to the energy efficiency, even if for a single base station, due to its dependency on traffic models and power models [4].

II. Where Comes Energy Saving Gain

In this section, we take a closer look at the power-traffic relation to get some insight of where the energy saving gain comes from. The overall relation between the total input power and the traffic served could further be divided into two sub-relations, as shown in Fig. 1. In particular, on the right hand side, the transmission efficiency curve indicates the relation between traffic and output power, while on the left hand side, the component efficiency curve illustrates the relation between output power and input power.

The improvement in the two efficiency curves will results in different gains in the overall input power consumption. For instance, given the traffic demand T_1, the original required input power is $P_{in,1}$, with the improvement in the transmission efficiency curve (from blue solid line to blue dashed line), the required input power reduced to $P_{in,2}$, and the difference between the two $P_{in_1} - P_{in,2}$ is the equivalent saving by transmission efficiency gain. On the other hand, with the improvement in the component efficiency curve (from red solid line to red dashed line), the input power can be reduced from $P_{in,1}$ to $P_{in,3}$. Combining the improvement in both curves, the overall input power shrank from $P_{in,1}$ to $P_{in,4}$ by the combined component and transmission efficiency gains. Note the overall power saving from the combined gain, $P_{in,4} - P_{in,1}$, is not the simple summation of the power saving from the two gains, $P_{in,2} - P_{in,1}$ and $P_{in,3} - Pin,1$, separately.

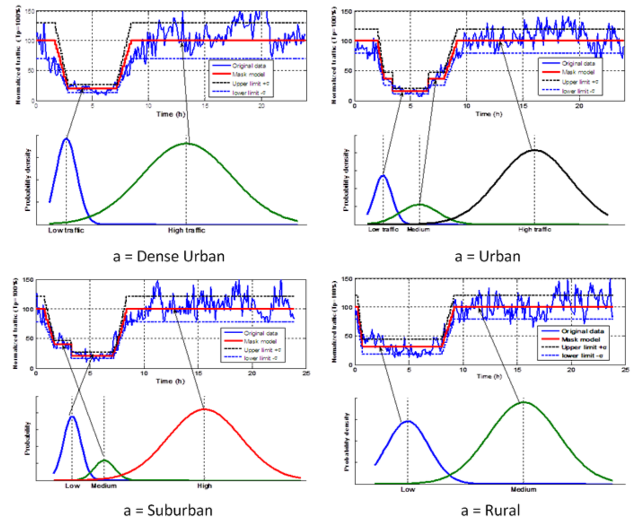


Fig. 2. The relative traffic profile from Orange Lab's measurement [4].

a = Dense urban	$T_{a,m}$ (%)	$\sigma_{a,m}$ (%)	$l_{a,m}$ (hours)
m = Low	20	6	5.5
m = High	100	30	18.5

a = Urban	$T_{a,m}$ (%)	$\sigma_{a,m}$ (%)	$l_{a,m}$ (hours)
m = Low	20	15	3
m = Medium	36	10	3.5
m = High	10	20	17.5

a = Suburban	$T_{a,m}$ (%)	$\sigma_{a,m}$ (%)	$l_{a,m}$ (hours)
m = Low	21	21	4
m = Medium	40	10	2
m = High	100	20	18

a = Urban	$T_{a,m}$ (%)	$\sigma_{a,m}$ (%)	$l_{a,m}$ (hours)
m = Low	31	12	8
m = High	100	20	16

TABLE I
Table of mask model parameters for different areas.

Moreover, different sets of the efficiency curves give different capabilities for the system to scale the network efficiency compared with traffic growth. When traffic scales from T_1 to T_2 with the increase $\Delta T = T_2 - T_1$, if there is no improvement in the set of efficiency curves, the input power will increase from $P_{in,1}$ to $P_{in,5}$. With the combined efficiency gains, on the other hand, the system power consumption will go down to $P_{in,6}$ with the reduction $\Delta_P = -|P_{in,1} - P_{in,6}|$.

Let's assume the original traffic and input power of the system are T and P, and $\eta_T = \Delta_T/T$ and $\eta_P = \Delta_P/P$ are the ratio of increase in traffic and input power, respectively, then the system energy efficiency gain could be expressed as

$$\eta_{EE} = \frac{T + \Delta_T}{P + \Delta_P} \cdot \frac{P}{T} - 1 = \frac{1 + \eta_T}{1 + \eta_P} - 1. \quad (1)$$

It is easy to see that, in order to achieve energy saving gain, i.e., $\eta_P < 0$, it is required that $\eta_T < \eta_{EE}$, namely the network efficiency scales faster than the traffic growth.

III. Network Energy Efficiency Evaluation

GreenTouch is set to model the power consumption of the wireless network on a country wide or even global scale at a realistic traffic load (i.e. non-full buffer model), comprising different areas (like dense urban or rural areas with very different user densities), daily duty cycles and even long term traffic increase over years. It is not feasible to compute the performance or power consumption on this scale. Instead, we will adopt a methodology that abstracts the diversity of real networks into a limited set of representative scenarios. The spatial diversity is captured by the four widely used scenarios of Dense Urban, Urban, Suburban and Rural areas. Each of these are studied for different times of the day representing the busy hour, night time and a range of intermediate traffic levels. The temporal resolution here is on the order of an hour, averaging over the bursty momentary traffic load. It is then feasible to compute the energy efficiency for the snapshots defined by a spatial scenario and a relative daily load situation. By aggregating these computations the energy efficiency of a countrywide or global network can be found.

A. Traffic Model

Let a denote the area type, i.e.,

$$a = \{\text{dense urban}, \text{urban}, \text{suburban}, \text{rural}\},$$

and let t denote the time instance, so the set of (a, t) defines a spatial and temporal snapshot. Let $T_a(t)$ represent the daily traffic profile for area a. In Fig. 2, we show the relative traffic profile normalized by the average peak hour traffic $T_{p,a}$ based on the traffic measurement from Orange Labs [4].

Statistically, we propose to use the mask model to describe the traffic behavior, which divides the profile curve into typical zones based on the traffic behavior and describe each zone by a Gaussian model with mean value and variance. For the application of the traffic model in the simulations, the daily profile in hours can be taken as hourly quantized input so that we have 24 snap shots to simulate. Similarly, the mask model can be interpreted as the reduced status model for the continuous or hourly daily profile, with more limited number of snap shots, each standing for a set of similar behavior hourly snapshots in the original profile. Typically, the daily profile curve for a given area type can be divided into the low, medium, and high traffic zones, represented by m. Note that for some typical areas, there might be only two zones instead of three. Three parameters are associated with the definition of the zones, they are the average traffic value for the zone, the traffic variance within the zone, and the time duration of the zone, denoted by $T_{a,m}$, $\sigma_{a,m}$, and $l_{a,m}$, respectively. The granularity of the sample points for each zone's calculation depends on the measurement available, could be hours, minutes, or seconds. The smaller the granularity is, the larger the variance could be. Corresponding to Fig. 2, example values for the mask model are given in Tab. I.

The exact values for the traffic load are confidential for operators, but can be roughly estimated from the total traffic volume projection. Specifically, we can first convert the data obtained from [1] [2] in $GB/user/month$ to $Mbps/user$. Then by multiplying the user density in different areas, we get the area traffic in $Mbps/km^2$ ($Mbps/user * user/km^2$). For instance, given the projection of $90.66MB/day$ for general western European countries, we can calculate the average traffic for Germany in general and Nuremberg City Center as $2.1Mbps/km^2$ and $210Mbps/km^2$, given the user density $250user/km^2$ and $25,000user/km^2$, respectively.

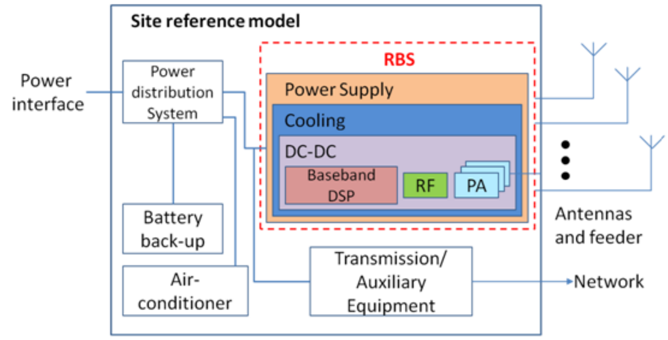


Fig. 3. The top down power consumption diagram for a site.

B. Deployment Model

The future network will have multiple layers, where each layer is constituted of a single base station (BS) type (e.g. a macro BS layer or pico BS layer), denoted by k. Depending on the geographically distribution of the traffic requirement, the network layout of each layer could be homogeneous and continuous (usually for marco BS to provide ubiquitous coverage), or inhomogeneous and/or discontinuous (usually for hotspot small cells to enhance capacity).

Associated with each layer, there is a set of basic parameters for the BSs therein. If we assume homogenous configurations for the BSs in the same layer, we have for each layer k

$$\{r_{k,a}, \lambda_{k,a}, \Pi_{k,a}, \Phi_{k,a}\}$$

where $r_{k,a}$ represents the cell radius, $\lambda_{k,a}$ is the deployment density, $\Pi_{k,a}$ is the set of BS locations, and $\Phi_{k,a}$ is the set of parameters related to the capability of the BS, including the configurations such as the number of sectors and antennas, frequency bands, transmit power, etc. Different deployment will result in different propagation conditions and also different network interference conditions, which in turn, result in different network capacity and energy consumption.

Therefore, the selection of possible deployments is equivalent to the joint work of the following actions: 1) selecting a combination of layers k for each area type a; 2) optimizing the densities $\lambda_{k,a}$ and locations $\Pi_{k,a}$ of the BSs in each layer; 3) optimizing the capability/configurations $r_{k,a}$ and $\Phi_{k,a}$ for the BSs in each layer. Note that these three actions may not be carried out sequentially but interact with each other. In general, there are two basic principles for any feasible solutions.

- *Coverage Principle.* The total area of interest $\mathcal{A}$ should be fully covered by the union of all layers, i.e. $\mathcal{A} = \sum_k \sum_j \mathcal{A}_{kj}$, where $\mathcal{A}_{kj}$ is the coverage provided by the j-th BS in the k-th deployment layer. The word coverage can further be defined by the minimum received signal strength at the cell edge. If there is a specified layer k^* that takes the responsibility for the overall coverage (e.g. the macro BS layer), then we have a stronger requirement that $\mathcal{A} = \sum_j \mathcal{A}_{k^*j}$. In this continuous deployment case, inter-site distance (ISD) is commonly used instead of the cell radius and density of the BS.

BS types	N_{TRX}	P^{out}	Δ_1	Δ_0	P^{BS}_{in}	P^{BS}_{sleep}
macro BS	$3*2$	20	4.7	780.0	1350.6	450.0
macro + RRH	$3*2$	20	2.1	504.0	754.8	336.0
micro BS	2	6.3	2.6	112.0	144.6	78.0
pico BS	2	0.13	4.3	13.6	14.7	8.6
femto BS	2	0.05	8.0	9.6	10.4	5.8

TABLE II
EXAMPLE DATA FOR THE LINEAR POWER MODEL WITH FULL LOAD (MAXIMUM OUTPUT POWER) [4]. THE POWER VALUES ARE IN WATT.

- *Capacity Principle.* The total traffic requirement in the whole area of interested $\mathcal{T}$ should be served by the union of all layers, i.e. $\mathcal{T} = \sum_k \sum_j \mathcal{T}_{kj}$, where $\mathcal{T}_{kj}$ is the traffic served by the j-th BS in the k-th layer. Moreover, the traffic served by each BS should not exceed its capacity, i.e. $\mathcal{T}_{kj} < \Theta \cdot \mathcal{C}_{kj}$, where $\mathcal{C}_{kj}$ represents the capacity of the BS and $0 < \Theta < 1$ stands for the variation protection. This capacity of a BS can be derived from system level simulations under given configurations and network load.

Once the deployment model is given, combined with the area traffic profile, we have the traffic distribution among layers, represented by $T_{a,m,k}$ and $\sigma_{a,m,k}$ for the k-th layer. Moreover, under the assumption of homogeneous traffic distribution in each layer and with the Gaussian mask model, the mean and variance of the traffic in each type k BS are $T^{BS}_{a,m,k} = T_{a,m,k}/N_{k,a}$ and $\sigma^{BS}_{a,m,k} = \sqrt{\sigma^2_{a,m,k}/N_{k,a}}$, respectively, where $N_{k,a} = \lambda_{k,a}\tau_a\mathcal{A}$ is the number of type k BS in area a, with $\mathcal{A}$ and τ_a denote the area of the overall playground of interest and the surface ratio of area a, respectively.

C. Power Model

The total mobile network (MN) power consumption (input power) includes power consumption in radio access network (RAN), backhaul (BH), and the mobile core (MC), i.e.

$$P^{MN} = P^{RAN} + P^{BH} + P^{MC}. \tag{2}$$

Since the power consumption in BH and MC parts is comparatively low, we hence set our focus on the RAN part. The RAN is consisted of carefully deployed sites, each containing units of BS, air-conditioner (AC), power distribution system (PDS), backup battery (BU) and auxiliary equipment (AE), as shown in Fig. 3 for a complete reference. Hence,

$$P^{site} = P^{BS} + P^{AC} + P^{PDS} + P^{BU} + P^{AE}. \tag{3}$$

The other parts of the RAN power consumption are comparatively static, except the BS part. In the following, we shall introduce two types of BS power modeling methodologies.

- *Top-down model.* This model is based on the measurement of each part of the existing BS equipments and is widely adopted [**?**]. In particular, a BS is consisted of baseband (BB) processor, transcievers (TRX) each associated with radio frequency (RF) circuits and power amplifier (PA), as well as power supply unit, cooling, and

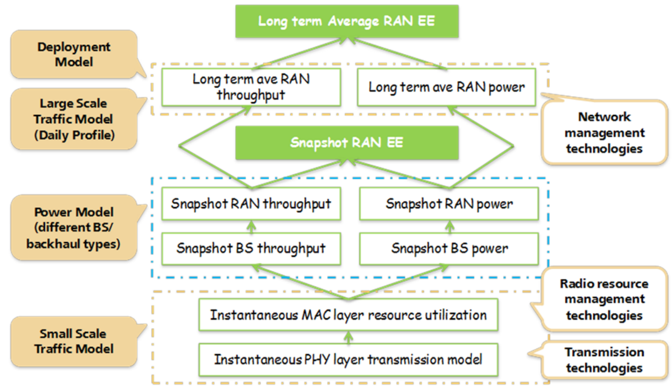


Fig. 4. The bottom up evaluation methodology.

DC-DC convertors, i.e,

$$P^{BS} = \beta \left[N_{TRX}\left(P^{RF} + P^{PA}\right) + P^{BB}\right], \tag{4}$$
$$\beta = \frac{1}{\eta_{DC}\cdot\eta_{PS}\cdot(1-\epsilon_{cool})},$$
$$P^{PA} = P^{PA,s} + \gamma P^{out}, \gamma = \frac{1}{\eta_{PA}\cdot(1-\epsilon_{feeder})}.$$

The η_i and ϵ_i stand for the efficiency and loss factor of the i-th component, $P^{PA,s}$ is the static power of PA, and P^{out} is the output power per antenna (uniform power allocation case). Note that P^{BB} and P^{out} both depend on the system configurations such as the bandwidth, the transmission scheme, the duty-cycle in time and frequency domain, which in turn, depend on the traffic load and scheduling algorithms. A linear approximation[3] of P^{BS} can be expressed as

$$P^{BS} = \Delta_0 + \Delta_1 P^{out} \tag{5}$$
$$\Delta_0 = \beta P^{BB} + \beta N_{TRX}(P^{RF} + P^{PA,s}),$$
$$\Delta_1 = \beta N_{TRX}\gamma.$$

Examples are given in Tab. II for different BS types. For energy saving, it is motivated to switch off unnecessary components of the system when a BS is not transmitting. This is generally referred to as sleep modes. When the BS is in sleep mode, the output power of the system is zero. Correspondingly, the PA can be switched off. Moreover, even for micro sleep, some part of the RF chain or BB components can be further switched off, depending on their ability to be wake up from sleep (the depth of the sleep). Define $P^{BS,sleep}$ as the general power in sleep mode, we shall have

$$P^{BS,sleep} \leq \beta P^{BB} + \beta N_{TRX}P^{RF} < \Delta_0. \tag{6}$$

- *Bottom-up model.* This model starts from scratch by estimating the basic power requirement in the radiation and in the RF/BB processing for given rate requirements and specific ways of transmission. For instance, the BB power is estimated by the ratio of the amount of calculations in operations such as FFT, precoding, and

[3] In practice, η_{PA} is function of P^{out}, usually behaving concavely as P^{out} increases. Moreover, it is also controlled by the bias voltage of the PA hardware, which can be tuned to match the maximum output power level.

FEC (GOPs) and the capability of the chips (GOPs/Watt). Also, the components in the RF chain such as DPD may be removed and the one-to-one mapping between the RF chain and the antenna may be changed, introducing new models to replace the traditional ones. This is needed especially for those brand new systems that haven't been carried out completely in practice, such as the large-scale antenna system that we shall introduce later on. The gaps between the two models remain big and open challenges that need further investigation.

To match the power model with the traffic and deployment model, we can express the power consumption of a type k BS in the area a during m zone as

$$P^{BS}_{k,m,a} = \begin{cases} \Delta_{0,k} + \Delta_{1,k} P^{out}_{k,m,a}\left(\frac{T_{a,m,k}}{N_{k,a}}, \sqrt{\frac{\sigma^2_{a,m,k}}{N_{k,a}}}\right) & \text{active} \\ P^{BS,sleep}_{k} & \text{sleep} \end{cases} \quad (7)$$

Note the calculation of $P^{out}_{k,m,a}$ depends not only on the system configurations, but also the transmission schemes and the resource management strategies (including the application of sleep mode), as well as the interference conditions from the neighboring BS transmitting in the same band. The detailed expression or values can be obtained either analytically with more assumptions or through sophisticated simulations.

D. Evaluation Methodology

Fig. 4 depicts a bottom-up approach to carry out the network energy efficiency calculation. The approach starts with a system level simulation of a snapshot at the physical and MAC layer. Given the radio transmission technique and the instantaneous resource utilization status, we can calculate the instantaneous power consumption of each transceiver system in the snapshot based on the system power model. The instantaneous power consumption of each base station is obtained by summing all transceiver and the overhead power for baseband processing, power supply and cooling. The total network power of a snapshot is the sum over all base stations of different types. Note that in snapshots with low traffic load some base stations may be switched off by network management and just contribute very small sleep mode power consumption. Finally, based on the long-term traffic model and the large scale deployment model, the long-term average power of the whole access network is yielded by weighting of the snapshots power values in time and space. Towards this end, we can convert the network power together with the total served traffic into the energy efficiency metric.

Given the logic and based on the models described earlier, the total network energy consumption[4] can be expressed as

$$\begin{aligned} E^{net} &= \sum_a \sum_m \sum_k P_{a,m,k}\left(T_{a,m,k}, \sigma_{a,m,k}\right) l_{a,m} \quad (8) \\ &= \sum_a \sum_m \sum_k N_{k,a} P^{BS}_{a,m,k} \left(\frac{T_{a,m,k}}{N_{k,a}}, \sqrt{\frac{\sigma^2_{a,m,k}}{N_{k,a}}}\right) l_{a,m}, \end{aligned}$$

[4]For illustration purpose, only the BS power consumption (the most complicated part) is considered. However, it could be easy to extend to the overall network power consumption.

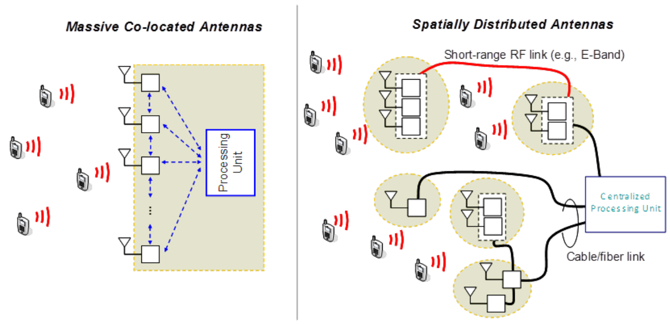


Fig. 5. The centralized and distributed deployment scenarios in LSAS project.

while the total network traffic load is

$$T^{net} = \sum_a \sum_m \sum_k T_{a,m,k} l_{a,m} = \sum_a \sum_m T_{a,m} l_{a,m}. \quad (9)$$

So the reverse of the average network energy efficiency is

$$\begin{aligned} \frac{1}{\eta^{net}} &= \frac{E^{net}}{T^{net}} = \frac{\sum_a \sum_m \sum_k N_{k,a} P^{BS}_{a,m,k} l_{a,m}}{\sum_a \sum_m T_{a,m} l_{a,m}} \quad (10) \\ &= \sum_a \sum_m \sum_k \tau_{k,m,a} \cdot \frac{1}{\eta^{BS}_{k,m,a}}, \end{aligned}$$

$$\tau_{k,m,a} = \frac{T_{a,m,k} l_{a,m}}{\sum_a \sum_m T_{a,m} l_{a,m}}, \qquad \eta^{BS}_{k,m,a} = \frac{T^{BS}_{k,m,a}}{P^{BS}_{k,m,a}}.$$

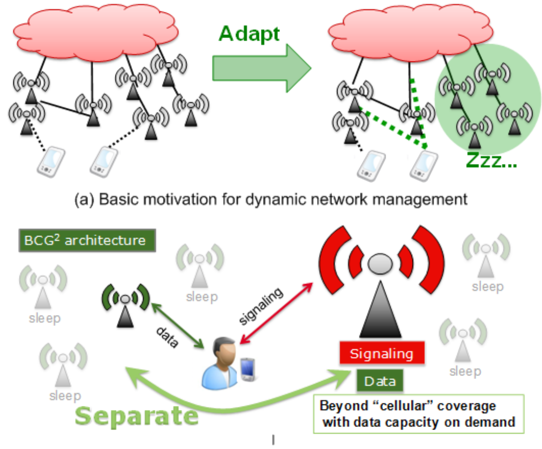
From the equation above, we see that the inverse of the network energy efficiency is the linear combination of the inverse of the BS energy efficiency of various types, weighted by the traffic serviced by the type of the BS. Therefore, the impact that the enhancement of the energy efficiency for one type of BS would have on the overall network energy efficiency scales by the traffic this layer of BSs serves.

IV. GreenTouch Wireless Architecture

The overall GreenTouch Mobile Architecture consists of the architectures and optimization framework from its three umbrella projects, namely the large-scale antenna system (LSAS) project, the beyond cellular green generation (BCG2) project, and the green transmission technologies (GTT) project. The first two projects suggest two sub-architectures for the future wireless networks, dense deployed with antennas and small data base stations, respectively. The GTT project, on the other hand, serves as the optimization engine for the architectures, with advanced transmission schemes designed and radio resource management optimized. In the following, we shall give brief introduction of each of the project, as well as how they jointly contribute energy saving gains.

A. Large-Scale Antenna System

LSAS [5] is a radical concept that is distinguished by having a large number of antennas compared with terminals under service. Compared with 4G technology, LSAS promises orders-of-magnitude increases in both radiated energy-efficiency and spectral-efficiency. Defining a LSAS architecture remains an elusive goal. The traditional macro-cellular scenario is generally ill-suited for LSAS quite simply because, with some



(a) Basic motivation for dynamic network management

(b) Illustration of the architecture with BCG2 concept

Fig. 6. The illustration of the BCG2 architecture to support dynamic network management for energy saving.

exceptions such as roof-top antennas in a city, most macro-cellular towers cannot physically support the one- or two-order-of-magnitude increase in the number of antennas. Another issue for LSAS is that, to meet the ambitious GreenTouch goal of 1000-fold improvement in overall energy-efficiency it is necessary to match large reductions in radiated power with commensurate reductions in internal power that is expended on signal processing and analog and digital electronics.

B. Beyond Cellular Green Generation

BCG2 project investigates the potential energy savings that can be achieved from a new system architecture based on the separation between signaling and data networks at the radio interface of mobile access networks. This separation allows mobile terminals to request service through a base station (signaling base station) and get the radio resources for data transmission from a different one (data base station). The new architecture goes beyond traditional cellular paradigm since only signaling network provides full coverage and always-on connectivity, while data coverage and communication capacity can be supplied on-demand activating portions of data network only when necessary. Fig. 6 illustrates the basic idea of BCG2 and explains where the first order of gain comes, namely from more flexible switching off data base stations that are not in serving states so as to save the static power of the whole network. The efficiency of the network can further be improved by advanced transmission technologies and resource management technologies studied in the GTT project.

C. Green Transmission Technologies

The improvement of energy efficiency and the reduction of energy consumption are usually associated with the cost in other operational metrics. This means the energy efficiency of the system will have tradeoffs. The key issue is when and how to tradeoff what while maintaining satisfactory network service [6]. Shannon's groundbreaking work on reliable communication over noisy channels [**?**] showed that higher bandwidth leads to lower transmitted power for the same data rate. It also suggests the benefit of transmitting a packet over a longer period of time to save transmitter energy. Inspired by

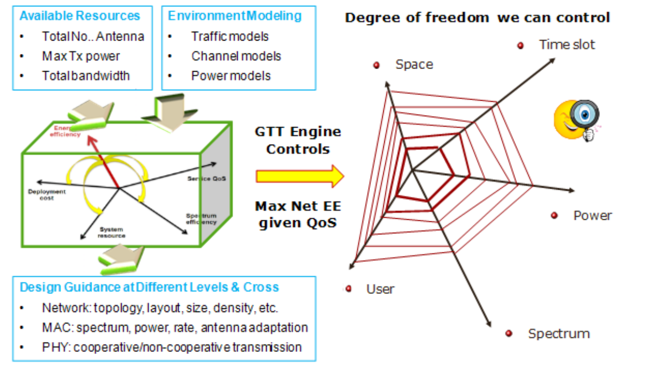


Fig. 7. The illustration of the working philosophy of the GTT engine.

the Shannon's theory, the GTT project will exploit the various dimensions of freedom in any given architecture to improve the network energy efficiency based on the fundamental tradeoffs, such as the tradeoffs between spectrum efficiency and energy efficiency, bandwidth and power, as well as service delay and power. The working philosophy of the GTT project is illustrated in Fig. 7.

D. Joint Improvement from GreenTouch Efforts

Referring to Fig. 1, the LSAS architecture helps to improve the transmission efficiency at both system level and network level by orderly reducing the transmit power and the also the neighboring interference level. Its impact on the component efficiency remains unclear due to the potential disruptive design in the RF and PA architecture. The BCG2 architecture foresees the improvement in the component efficiency at the network level by enabling fast switching on/off capability of the data BSs. Joint transmission and component efficiency on top of LSAS and BCG2 may be further enlarged by the optimized resource management strategies within and between layers, as well as the various kinds of cooperative transmission techniques developed by the GTT framework.

V. Conclusion

Measured by bit/joule, the energy efficiency of a network depends heavily on the power models and traffic models. To have positive energy saving gain, it is required to scale the network energy efficiency faster than the traffic growth, which shall be achieved by jointly improving the transmission efficiency and component efficiency of the network, and the efficiency improvement in a single BS type will be scaled by the traffic offloaded by the layer of that BS type.

References

[1] *Cisco Visual Networking Index: Forecast and Methodology, 2010-2015.*
[2] *Projections of IP Traffic to 2020 Based on Regression Analyses of Historical Trends and Nearer-Term Forecasts*, GreenTouch SAT Committee.
[3] EARTH project deliverable, *D2.3 Energy Efficiency Analysis of the Reference Systems, Areas of Improvements and Target Breakdown*, 2012.
[4] GreenTouch Mobile WG, *GreenTouch Mobile Architecture Doc 1 - Models and Methodology*, 2012.
[5] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, Nov 2010.
[6] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental tradeoffs on green wireless communications," *IEEE Commun. Mag.*, June 2011.