1   How Much Energy is Needed to Run a Wireless Network ?

Gunther Auer[†], Vito Giannini[∗], István Gódor[‡], Magnus Olsson[§],
Muhammad Ali Imran[¶], Dario Sabella[♯], Manuel J. Gonzalez[◇]
Oliver Blume[♣], Albrecht Fehske[♠], Jose Alonso Rubio[§], Pål Frenger[§],
Claude Desset[∗]

[†]DOCOMO Euro-Labs, D-80687 Munich, Germany

[∗]IMEC, Leuven, Belgium

[‡]Ericsson Research, Budapest H-1117, Hungary

[§]Ericsson Research, SE-164 80 Stockholm, Sweden

[¶]CCSR University of Surrey, Guildford GU2 7XH, UK

[♯]Telecom Italia, 10148 Turin, Italy

[◇]TTI (Technologies of Telecommunication and Information), 39011 Santander, Spain

[♣]Alcatel-Lucent, Bell Labs Germany, D-70435 Stuttgart, Germany

[♠]Technische Universität Dresden, D-01062 Dresden, Germany

1.1   Introduction

The global mobile communication industry is growing rapidly. Today there are already more than 4 billion mobile phone subscribers worldwide [1], more than half the entire population of the planet. Obviously, this growth is accompanied by an increased energy consumption of mobile networks. Global warming and heightened concerns for the environment of the planet require a special focus on the energy efficiency of these systems [2]. The EARTH[1] project [3] is a concerted effort to achieve this goal and as part of its objectives, a holistic framework is developed to evaluate and compare the energy efficiency of several design approaches of wireless cellular communication networks.

For the quantification of energy savings in wireless networks, the power consumption of the entire system needs to be captured and an appropriate energy efficiency evaluation framework (E^3F) is to be defined. The EARTH E^3F presented in Section 1.2 provides the key levers to facilitate the assessment of the overall energy efficiency of cellular networks over a whole country. The E^3F primarily builds on well-established methodology for radio network performance evaluation developed in 3GPP; the most important addendums, introduced in Sections 1.3 and 1.4, are to add a sophisticated power model of the base stations

---

[1] EU funded research project EARTH (Energy Aware Radio and neTwork tecHnologies), FP7-ICT-2009-4-247733-EARTH, Jan. 2010 to June 2012. https://www.ict-earth.eu
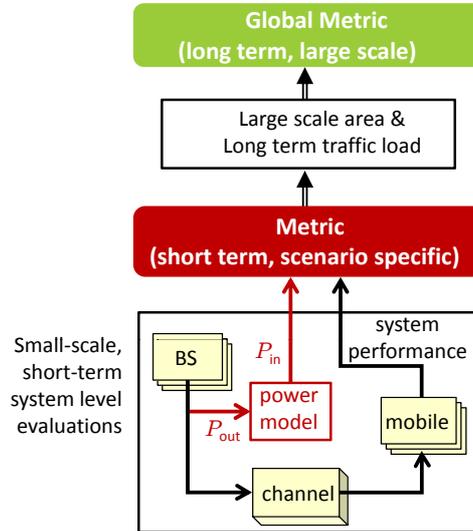
Figure 1.1 EARTH Energy efficiency evaluation framework (E^3F).

(BSs) as well as a large-scale long-term traffic model extension to existing 3GPP traffic scenarios. Then, using the metrics defined in Section 1.5, in Section 1.6 the E^3F is applied in order to provide an assessment of the BS energy efficiency of a 3GPP LTE network deployed within an average European country. The energy efficiency of LTE is compared to that of already deployed networks is discussed in Section 1.7, and targets for the energy efficiency of future wireless networks are given.

1.2   Energy Efficiency Evaluation Framework (E^3F)

The widely accepted state-of-the-art to evaluate the performance of a wireless network is to simulate the relevant aspects of the radio access network (RAN) at system level. The computed results are, e.g. the system throughput measured in bit/s, quality of service (QoS) metrics, and fairness in terms of cell-edge user throughput. In order to ensure that the results generated by different RAN system simulation tools are comparable, well defined reference systems and scenarios are specified. This is an outcome of extensive consensus work from standardization bodies, such as 3GPP [4], and international research projects, such as the EU project Wireless World Initiative New Radio (WINNER) [5], with partners from academia as well as from industry. The most recent example is the global effort in ITU to evaluate system proposals for compliance with IMT-Advanced requirements [6]. In that direction, the EARTH E^3F builds on the 3GPP evaluation framework for LTE [4].

Fig. 1.1 shows the necessary enhancements over existing performance evaluation frameworks, such that the energy efficiency of the entire network, comprising component, node and network level, over an extended time frame can be quantified. The EARTH E^3F illustrated in Fig. 1.1, identifies the essential building blocks that are necessary for an accurate holistic assessment of energy efficiency enhancements. Although the specific realization of a system level simulation tool largely depends on the specific problem at hand, as well as the chosen software implementation, it is envisaged that for the assessment of combinations of energy efficiency enhancements integrated into one holistic system concept, the E^3F should capture the following aspects:

- A sophisticated power model (specified in Section 1.3), that maps the RF output power radiated at the antenna elements to the total supply power of a BS site. The power model maps the gains on the component level (e.g. an improvement of the energy efficiency of the power amplifiers) to energy savings on the entire network.
- Long-term traffic models (established in Section 1.4), that describe load fluctuations over a day and complement the statistical short-term traffic models.
- Large-scale deployment models (developed in Section 1.4) of large geographical areas are considered to extend the existing small-scale deployment scenarios.

1.2.1 Small-Scale, Short-Term System Level Evaluations

Statistical traffic models (e.g. FTP file download or VoIP calls), specific small-scale deployment scenarios (e.g. urban macro-cell consisting of 57 hexagonal cells with uniformly distributed users), and power models that quantify the power consumption of components within a node, constitute small-scale, short-term system level evaluations (bottom block in Fig. 1.1). The small-scale, short-term system level evaluations are carried out by a system level simulation platform, augmented by a model capturing the BS power consumption.

1.2.2 Global E^3F

In order to extend small-scale, short-term evaluations to a global scale, covering countrywide geographical areas and ranging over a full day or week, long-term traffic models and large-scale deployment maps are to be integrated into the E^3F, as illustrated in Fig. 1.1. The global assessment of network energy efficiency comprises the following steps:

1. Small-scale, short-term evaluations are conducted for all scenarios (dense urban, urban, suburban and rural) and for a representative set of traffic loads, which captures the range between the minimum and the maximum load observed in a certain deployment.
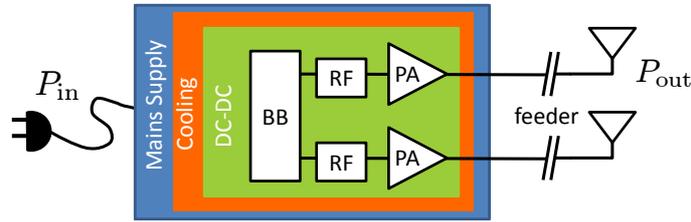
Figure 1.2 Block diagram of a base station transceiver.

2. The system level evaluations provide energy consumption and other performance metrics (e.g. throughput, QoS) for each small-scale deployment and a certain traffic load.
3. Given the daily/weekly traffic profile of each deployment, the power consumption over a day/week is generated by weighted summing of the short-term evaluations.
4. Finally, the mix of deployment scenarios that quantify the area covered by cities, suburbs, highways and villages, yield the global set of the large-scale system energy consumption.

1.3   Power Model

This section provides a power model for various types of LTE Base Stations. The power model constitutes the interface between component and system level, which allows quantifying how energy savings on specific components enhance the energy efficiency at the node and network level.

1.3.1   Base Station Power Consumption Breakdown

Fig. 1.2 shows a simplified block diagram of a complete BS that can be generalized to all BS types, including macro, micro, pico and femto BSs. A BS consists of multiple transceivers (TRXs), each of which is serving one transmit antenna element. A TRX comprises a Power Amplifier (PA), a Radio Frequency (RF) small-signal transceiver section, a baseband (BB) interface including a receiver (uplink) and transmitter (downlink) section, a DC-DC power supply, an active cooling system, and an AC-DC unit (mains supply) for connection to the electrical power grid. In the following the various TRX parts are analyzed.

Antenna Interface: The influence of the antenna type on power efficiency is modeled by a certain amount of losses, including the feeder, antenna band-pass filters, duplexers, and matching components. Since macro BS sites are often situated at different physical locations as the antennas a feeder loss of about $\sigma_{\text{feed}}=3\,\text{dB}$ needs to be added. The feeder loss of a macro BS may be mitigated by introducing a remote radio head (RRH), where the PA is mounted at the same

physical location as the transmit antenna. Likewise, feeder losses for smaller BS types are typically negligible.

Power Amplifier (PA): Typically, the most efficient PA operating point is close to the maximum output power (near saturation). Unfortunately, non-linear effects and OFDM modulation with non-constant envelope signals force the power amplifier to operate in a more linear region, i.e., 6 to 12 dB below saturation [7]. This prevents Adjacent Channel Interference (ACI) due to non-linear distortions, and therefore avoids performance degradation at the receiver. However, this high operating back-off gives rise to poor power efficiency η_{PA}, which translates to a high power consumption P_{PA}. Digital techniques such as clipping and digital pre-distortion [8, 9] in combination with Doherty PAs [7] improve the power efficiency and linearizes the PA, while keeping ACI under control, but require an extra feedback for pre-distortion and significant additional signal processing [9]. While these techniques are necessary in macro and micro BSs, they are not used in smaller BSs, as the PA power consumption accounts for a smaller percentage of the power breakdown, allowing for a higher operating back-off.

The *Small-Signal RF Transceiver (RF-TRX)* comprises a receiver and a transmitter for uplink (UL) and downlink (DL) communication. The linearity and blocking requirements of the RF-TRX may differ significantly depending on the BS type, and so its architecture. Typically, low-IF (Intermediate-Frequency) or super-heterodyne architectures are the preferred choice for macro/micro BSs, whereas a simpler zero-IF architecture are sufficient for pico/femto BSs [10]. Parameters with highest impact on the RF-TRX energy consumption, P_{RF}, are the required bandwidth, the allowable Signal-to-Noise And Distortion ratio (SiNAD), the resolution of the analogue-to-digital conversion, and the number of antenna elements for transmission and/or reception.

Baseband (BB) Interface: The baseband engine (performing digital signal processing) carries out digital up/down-conversion, including filtering, FFT/IFFT for OFDM, modulation/demodulation, digital-pre-distortion (only in DL and for large BSs), signal detection (synchronization, channel estimation, equalization, compensation of RF non-idealities), and channel coding/decoding. For large BSs the digital baseband also includes the power consumed by the serial link to the backbone network. Finally, platform control and MAC operation add a further power consumer (control processor).

The silicon technology significantly affects the power consumption P_{BB} of the BB interface. This technology scaling is incorporated into the power model by extrapolating on the International Technology Roadmap for Semiconductors (ITRS). The ITRS anticipates that silicon technology is replaced by a new generation every 2 years, each time doubling the active power efficiency but multiplying by 3 the leakage [11]. The increasing leakage puts a limit on the power reduction that can be achieved through technology scaling. Apart from the technology, the main parameters that affect the BB power consumption are related to the signal bandwidth, number of antennas and the applied signal process-

ing algorithms. While the consumed power scales linearly with the bandwidth; MIMO signal detection scales more than linearly with the number of antennas.

Power Supply and Cooling: Losses incurred by DC-DC power supply, mains supply and active cooling scale linearly with the power consumption of the other components, and may be approximated by the loss factors σ_{DC}, σ_{MS}, and σ_{cool}, respectively. Note that active cooling is only applicable to macro BSs, and is omitted in smaller BS types. Moreover, for RRHs active cooling is also obsolete, since the PA is cooled by natural air circulation, and the removal of feeder losses σ_{feed} allow for a lower PA power consumption, $P_{\text{PA}} = \frac{P_{\text{out}}}{\eta_{\text{PA}} \cdot (1 - \sigma_{\text{feed}})}$, where η_{PA} denotes the PA power efficiency.

Assuming that the BS power consumption grows proportionally with the number of transceiver chains N_{TRX}, the breakdown of the BS power consumption at maximum load, $P_{\text{out}} = P_{\text{max}}$, amounts to

$$P_{\text{in}} = N_{\text{TRX}} \cdot \frac{\frac{P_{\text{max}}}{\eta_{\text{PA}} \cdot (1 - \sigma_{\text{feed}})} + P_{\text{RF}} + P_{\text{BB}}}{(1 - \sigma_{\text{DC}})(1 - \sigma_{\text{MS}})(1 - \sigma_{\text{cool}})} \tag{1.1}$$

The *efficiency* is defined by $\eta = P_{\text{out}}/P_{\text{in}}$, whereas the *loss* factor is defined by $\sigma = 1 - \eta$. Note that the maximum RF output power per transmit antenna, P_{max}, is measured at the input of the antenna element, so that losses due to the antenna interface (other than feeder losses) are *not* included in the power breakdown.

Table 1.1 summarizes the state of the art power consumption of various LTE BS types as of the year 2010. By introducing RRHs in macro BS sites, so that feeder losses σ_{feed} and active cooling are avoided by mounting the PA close to the transmit antenna, the power savings exceed 40%.

1.3.2   BS Power Consumption at Variable Load

In a conventional BS, the power consumption depends on the traffic load; it is mainly the PA power consumption that scales down due to reduced traffic load. This mainly happens when, e.g., the number of occupied subcarriers is reduced in idle mode operation, and/or there are subframes not carrying data. Naturally this scaling over signal load largely depends on the BS type; for macro BSs the PA accounts for 55–60% of the overall power consumption at full load, whereas for low power nodes the PA power consumption amounts to less than 30% of the total.

Fig. 1.3 shows BS power consumption curves for a LTE system with 10 MHz bandwidth and 2×2 MIMO configuration. Three sectors are considered for macro BSs, whereas omni-directional antennas are used for the smaller BS types. While the power consumption P_{in} is load dependent for macro BSs, and to a lesser extent for micro BSs, there is a negligible load dependency for pico and femto BSs. The reason is that for low power BSs, the impact of the PA is diminishing. Other components hardly scale with the load in a state of the art implementation; although some more innovative designs could lead to an improved power scaling at low loads. As can be seen in Fig. 1.3, the relations between relative RF output

Table 1.1. Base station power consumption at maximum load of a LTE system with 2×2 MIMO for different BS types as of 2010.

			Macro	RRH	Micro	Pico	Femto
BS	Max Tx power	[dBm]	43.0	43.0	38.0	21.0	17.0
	(average) P_{max}	[W]	20.0	20.0	6.3	0.13	0.05
	Feeder loss σ_{feed}	[dB]	3	0	0	0	0
PA	Back-off	[dB]	8.0	8.0	8.0	12.0	12.0
	Max PA out (peak)	[dBm]	54.0	51.0	46.0	33.0	29.0
	PA eff. η_{PA}	[%]	31.1	31.1	22.8	6.7	4.4
	Total PA, $\frac{P_{\mathrm{max}}}{\eta_{\mathrm{PA}} \cdot (1 - \sigma_{\mathrm{feed}})}$	[W]	**128.2**	**64.4**	**27.7**	**1.9**	**1.1**
RF	P_{TX}	[W]	6.8	6.8	3.4	0.4	0.2
	P_{RX}	[W]	6.1	6.1	3.1	0.4	0.3
	Total RF, P_{RF}	[W]	**12.9**	**12.9**	**6.5**	**1.0**	**0.6**
BB	Radio (inner Rx/Tx)	[W]	10.8	10.8	9.1	1.2	1.0
	Turbo code (outer Rx/Tx)	[W]	8.8	8.8	8.1	1.4	1.2
	Processors	[W]	10.0	10.0	10.0	0.4	0.3
	Total BB, P_{BB}	[W]	**29.6**	**29.6**	**27.3**	**3.0**	**2.5**
DC-DC, σ_{DC}		[%]	7.5	7.5	7.5	9.0	9.0
Cooling, σ_{cool}		[%]	10.0	0.0	0.0	0.0	0.0
Mains Supply, σ_{MS}		[%]	9.0	9.0	9.0	11.0	11.0
Total per TRX chain		[W]	**225.0**	**125.8**	**72.3**	**7.3**	**5.2**
# Sectors		#	3	3	1	1	1
# Antennas		#	2	2	2	2	2
# Carriers		#	1	1	1	1	1
Total N_{TRX} chains, P_{in}		[W]	**1350.0**	**754.8**	**144.6**	**14.7**	**10.4**

power P_{out} and BS power consumption P_{in} are nearly linear. Hence, a linear approximation of the power model is justified:

$$P_{\mathrm{in}} = \begin{cases} N_{\mathrm{TRX}} \cdot \left(P_0 + \Delta_{\mathrm{p}} \, P_{\mathrm{out}} \right) , & 0 < P_{\mathrm{out}} \leq P_{\mathrm{max}} \\ N_{\mathrm{TRX}} \cdot P_{\mathrm{sleep}} , & P_{\mathrm{out}} = 0 \end{cases} \qquad (1.2)$$

where P_{max} denotes the maximum RF output power at maximum load, P_0 is the power consumption calculated at the minimum possible output power, assumed to be 0.1% of P_{max}, and Δ_{p} is the slope of the load dependent power consumption.

Also indicated in Fig. 1.3 and Table 1.2, is a sleep mode power consumption, P_{sleep}. In future base stations, fast deactivation of components, i.e. to put them into sleep when there is nothing to transmit, is believed to be an important solution to save energy. The sleep mode power consumption is introduced here to capture such solutions.
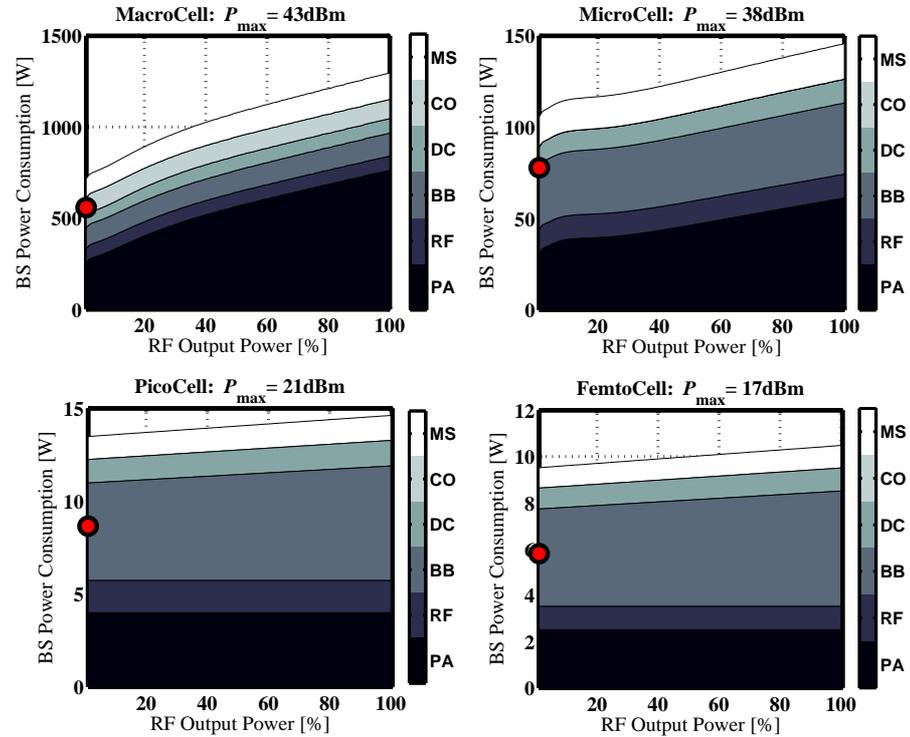
Figure 1.3 Power consumption for various BS types as a function of the RF output power. An LTE system with 10 MHz system bandwidth and 2×2 MIMO configuration is considered. Macro BSs employ 3 sectors per site. Legend: PA: power amplifier, RF: small signal RF transceiver, BB: baseband processor, DC: DC-DC converters, CO: active cooling (only applicable to macro BS), MS: mains power supply.

Table 1.2. Power model parameters for different BS types

BS type	N_{TRX}	$P_{\max}$ [W]	P_0 [W]	Δ_{p}	P_{sleep} [W]
Macro	6	20.0	118.7	5.32	93.0
RRH	6	20.0	??	??	??
Micro	2	6.3	53.0	3.1	39.0
Pico	2	0.13	6.8	4.0	4.3
Femto	2	0.05	4.8	7.5	2.9

1.4   Traffic Model

In order to provide a realistic analysis of the energy efficiency of wireless networks, it is essential to know the traffic demand to be served by the network. Thus, it is important to identify the spatial and temporal variation of the traffic demand both on large as well as small scale.

1.4.1 Deployment Areas of Europe

The geographical distribution as well as the population densities are fairly similar for most European countries; however, the Nordic countries (Finland, Norway and Sweden) and Russia substantially deviate from the European average. Let a_d, $0 \leq a_d \leq 1$, denote the share of the area covered by deployment scenario d normalized to the total area of a given country or region. The European average of the geographical distribution a_d of the considered deployment areas and the corresponding population densities p_d in citizen/km^2 shown in Table 1.3 therefore excludes the Nordic countries and Russia. Note that in central districts of a metropolis, the population density may exceed $p_d = 20,000$ citizen/km^2, but these are omitted due to their negligible covered area a_d.

The network planning policy of European operators concentrate to serve the most of the population and not on the amount of area covered [12]. That is, 2G area coverage is almost 100%, while 3G coverage is below 40%. This reflects that sparsely populated areas are served by the minimum service level as defined by national telecommunication authorities, i.e., voice (2G) and low speed data connection (GPRS). Following this trend, we assume that LTE is deployed in dense urban, urban, suburban and rural areas only. For instance, German regulation forces to serve "only" 90% of the population with broadband access [13], which practically allows to skip scarcely populated areas.

1.4.2 Long-Term Large-scale Traffic Models

The objective for the long-term large-scale traffic models is to determine the average served traffic on a certain time of day for a given deployment scenario. Abstracting the models from current European cell planning maps, the following methodology allows to deduce the daily traffic variations in terms of the actual traffic demand per unit area:

1. define the average traffic demand per active subscriber for different terminal types;

Table 1.3. Deployment Areas in Europe (Excluding Nordic and Russia)

Deployment d	Population density p_d [citizen/km^2]	Covered area a_d
Dense urban	3000	1%
Urban	1000	2%
Suburban	500	4%
Rural	100	36%
Sparsely populated & wilderness	25	57%

Table 1.4. Estimated traffic demand ranges of terminal types in Europe in 2015

Terminal type k	Average rate r_k in [Mbps]	Daily[†] [MB]	Monthly[†] [GB]
PC	0.5 $\rightarrow$ 2	512 $\rightarrow$ 2048	16 $\rightarrow$ 64
Tablet	0.25 $\rightarrow$ 1	256 $\rightarrow$ 1024	8 $\rightarrow$ 32
Smartphone	0.0625 $\rightarrow$ 0.25	64 $\rightarrow$ 256	2 $\rightarrow$ 8
Reference PC (2010)	0.03125 $\rightarrow$ 0.125	32 $\rightarrow$ 128	1 $\rightarrow$ 4

[†] For notational convenience the exact values are rounded to the closest power of 2.

2. by virtue of a daily traffic profile the traffic volumes per subscriber are obtained;
3. define relevant scenarios of different terminal/subscriber mixes;
4. determine the number of active users per unit area for the considered terminal/subscriber mixes;
5. given the population densities for the respective deployments, the scenario specific network traffic per unit area in [Mbps/km^2] can be derived;
6. the total network traffic of an average European country is obtained by weighted summing of the scenario specific network traffic

1.4.2.1 Traffic demand per active subscriber

The user generated data volume is tightly connected to operator policies and data subscriptions plans. While the amount of traffic varies from country to country, studies within the EARTH project revealed that the average rates per active subscriber are independent of the deployment scenario d. For the envisaged terminal types k we therefore propose to define a range of prospective traffic demands per subscriber r_k (in a representative Western European country) for the year 2015, which applies to all considered deployments:

- *PC users* demand for bandwidth equivalent of providing, e.g., SDTV or even HDTV for all active users, which translates to a data range of $r_k = [0.5, 2]$ Mbps/user [14]. This range is equivalent of average and high-end DSL demands of 2010.
- *Tablet users* demand for half of the bandwidth of PC users, $r_k = [0.25, 1]$ Mbps/user.
- *Smartphone users* demand for a quarter of the bandwidth of a tablet user [14], $r_k = [67.5, 250]$ kbps/user.
- As a benchmark *reference PC users* in 2010 demand for half the traffic volume of smartphone users in 2015, $r_k = [31.25, 125]$ kbps/user. This range is equivalent to average and high-end HSPA traffic in 2010 [16, 17, 18].

The data rate requirements per subscriber for the considered terminal types are listed in Table 1.4. We emphasize that the figures listed in Table 1.4 represent average traffic demands; typically strong temporal and geographical deviations with respect to these average values are experienced, e.g., one or two so-called heavy users may fully utilize a cell even for extended time periods.
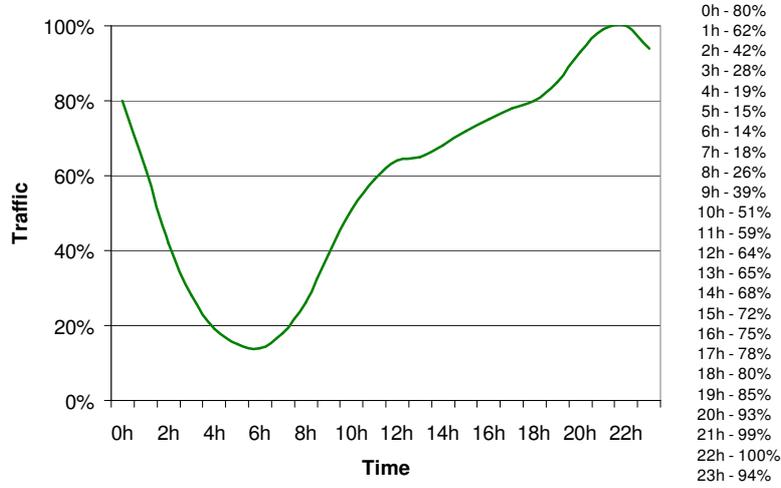
Figure 1.4 Normalized average daily data traffic profile $\tilde{\alpha}(t) = \alpha(t)/\widehat{\alpha}$, taken as a reference for an European country.

1.4.2.2 Daily traffic variations

Clearly, not all subscribers are always active; rather the number of active subscribers changes between busy and off-peak hours. In today's networks 10–30% of the data subscribers are active in the busy/peak hours; as an European average we assume that at peak hours a mobile broadband subscriber is active with probability $\widehat{\alpha}$=0.16. Thus, on average the generated traffic volume of terminal type k at peak hours of duration T yields $\widehat{v}_k(T) = \widehat{\alpha} r_k T$.

Let $\alpha(t)$, with $0 \leq \alpha(t) \leq \widehat{\alpha}$, define the daily variation of active users that captures the variations in average network traffic over a day. Based on internal surveys on operator traffic data within the EARTH project and the Sandvine report [15], the average network traffic follows the normalized daily variation $\tilde{\alpha}(t) = \alpha(t)/\widehat{\alpha}$ illustrated in Fig. 1.4. The EARTH project found that the daily traffic variation $\alpha(t)$ is independent of the deployment d; hence the daily traffic profile in Fig. 1.4 is valid for all deployments. Given the traffic variation over time $\alpha(t)$, the generated data volume of a mobile subscriber over duration T amounts to

$$v_k(t, T) = r_k \int_t^{t+T} \alpha(\tau) \, \mathrm{d}\tau \tag{1.3}$$

The obtained traffic volumes of the considered terminal types are summarized in Table 1.4.

1.4.2.3 Scenarios of terminal & subscriber mixes

According to the expectations towards wireless Internet services, the fraction of broadband data subscribers of the whole population will increase from year to year and in the most mature European markets may reach 25% by 2015; however,

the European average might be somewhat lower. Moreover, tablets, smartphones and other mobile equipment, that are becoming increasingly popular already today, are expected to stimulate additional traffic demand in Europe; and even more so in North America, due to much lower mobile PC traffic compared to Europe. Let s_k denote the fraction of broadband data subscribers of the whole population for terminal type k, different scenarios can be constructed that reflect the expected share of mobile broadband subscribers in 2015:

- *Scenario #1*: s_{PC}=20% of the population are heavy PCs users, requesting an average data rate of r_{PC}=2 Mbps/user.
- *Scenario #2*: s_{PC}=20% of the population are PC users, s_{PC}=5% of the population are tablet users, and s_{PC}=50% of the population are smartphone users, all of which are classfied as heavy users requesting an average data rate of r_{PC}=2 Mbps, r_{tab}=0.5 Mbps and r_{fon}=0.25 Mbps. This scenario serves as an upper bound on the envisaged traffic demain in 2015.
- *Scenario #4*: s_{PC}=20% of the population are PC users, s_{PC}=5% of the population are tablet users, and s_{PC}=50% of the population are smartphone users, of which one half is classfied as heavy users with rates r_{PC}=2 Mbps, r_{tab}=0.5 Mbps and r_{fon}=0.25 Mbps, whereas the other half are classfied as average users with 50% lower rates, which is r_{PC}=1 Mbps, r_{tab}=0.25 Mbps and r_{fon}=0.125 Mbps. Based on [14], we consider this as the most relevant European scenario for 2015.
- *Scenario #5*: serves as reference scenario for the contemporary traffic demand, where s_{rPC}=10% of the population are reference PC users in 2010, requesting a rate of r_{rPC}=0.125 Mbps.

1.4.2.4   Active subscribers

Given the geographical population data, the daily traffic variations and the terminal & subscriber mixes established in Sections 1.4.1, 1.4.2.2 and 1.4.2.3, the average number of active subscribers of a European country at a given time of day can be quantified. Provided that the population is served by N_{op} operators, each of which having a $1/N_{\mathrm{op}}$ share of the total traffic volume, the average number of active subscribers of deployment d for terminal type k at time t is given by

$$U_{k,d}(t) = \alpha(t) p_d \frac{s_k}{N_{\mathrm{op}}} \quad \text{in [subscribers/km}^2] \tag{1.4}$$

The number of active users $U_{k,d}(t)$ is scaled with the population density p_d. That is, taking dense urban as reference, the number of active users in urban, suburban and rural deployments are 0.33, 0.16 and 0.033 times the dense urban deployment, respectively.

1.4.2.5   Aggregated Traffic Demand

As the data volume per subscriber does not depend on the deployment scenario, the generated network traffic is proportional to the population density. Given

Table 1.5. Estimated area throughput and number of active users at busy hours in a dense urban cell in Europe of 2015

	Scenarios & mixes	Peak area throughput [Mbps/km^2]	Active users per cell
#1	Heavy PCs	192	$2 - 3$
#2	Heavy PCs & tablets and smartphones	168	11
#4	mix of heavy & average PCs, tablets and smartphones	153	11
#5	Reference PC from 2010	6	$1 - 2$

the terminal specific data rates r_k, the number of active users in a given deployment $U_{k,d}(t)$, and the mix of terminal types, the generated traffic over a network is determined by

$$R_d(t) = \sum_k r_k U_{k,d}(t) = \sum_k r_k \, \alpha(t) p_d \frac{s_k}{N_{\mathrm{op}}} \quad \text{in } [\mathrm{bit/s/km}^2] \qquad (1.5)$$

Aggregation over time yields the total traffic per unit area served during a duration T, that is

$$R_d = \frac{1}{8} \sum_k v_k(t,T) p_d \frac{s_k}{N_{\mathrm{op}}} \quad \text{in } [\mathrm{Byte/km}^2] \qquad (1.6)$$

where the terminal specific data volume $v_k(t,T)$ is given by (1.3). Table 1.5 shows the area throughput R_d and number of active users per cell at peak traffic for dense urban environment with 3000 citizen/km^2 and 500 m inter-site distance between base stations [4]. Given the geographical distribution of the respective deployments a_d, the total traffic of a country-wide network is obtained by weighted summing

$$R_{\mathrm{tot}} = \frac{\sum_d a_d R_d}{8 \sum_d a_d} = \frac{\sum_d \sum_k v_k(t,T) a_d p_d \frac{s_k}{N_{\mathrm{op}}}}{8 \sum_d a_d} \quad \text{in } [\mathrm{Byte/km}^2] \qquad (1.7)$$

1.4.3 Statistical Short-Term Traffic Models

In order to model the fluctuation of the traffic in short-time scale, the packet distribution generated by the different type of applications is modeled statistically. Since the same short-term traffic models per active user should be applied in all deployment areas, the traffic demands in different deployments are derived from the differences in the corresponding user density figures. A detailed description of the traffic models can be found in [4].

1.5   Green Metrics

What metric we use to capture efficiency determines how we think and act. Therefore, it is important that we use metrics that guide us in the right direction. When we discuss energy efficiency may refer to how much energy it takes to achieve a certain amount of work, or we may refer to how much work we can get achieved by using a certain amount of energy. The difference is subtle, but important, as discussed in Section 1.5.1.

In many areas where efficiency is important, such as transportation, the definition of "work" is straightforward; e.g. a vehicle, a person, or a unit of weight is moved a certain distance. In cellular networks it is not as easy to define what exactly one unit of "work" constitutes. The network provides connectivity over a certain area and it transports bits to mobile users. Users pay not only for the served number of bits but also for the possibility to use the network everywhere and anytime. Hence the area coverage provided by the network is important, even when no user is transferring any bits. In order to capture both of these aspects two different metrics will be introduced in Section 1.5.2.

1.5.1   Efficiency Metrics vs Consumption Metrics

A metric of energy usage can be expressed either as a *consumption index* or as an *efficiency index*. In the automobile industry both the "miles per gallon" or MPG metric and the "liters per 100 km" metric are commonly used. The MPG metric is an efficiency index, i.e., a car that consumes less fuel will have a higher metric compared to another car with higher fuel consumption. The "liters per 100 km" on the other hand is consumption metric, i.e., a car that consumes less fuel will have a smaller metric. In essence these two indexes both contain the same information, and hence it is straightforward to convert any efficiency metric to the equivalent consumption metric and vice versa. However there are subtle differences that may lure an observer to misleading conclusions.

The main benefit with an efficiency index, where the useful work is in the nominator and the consumed energy is in the denominator (work/energy), is that a larger metric means better performance, which many people find intuitive. The drawback of an efficiency index is that it is difficult to relate an efficiency increase with the achieved energy savings. As an example, 20% better MPG translates to only 16.7% less fuel (20% is 1.2 times larger distance correspond to $1/1.2 = 0.83$ of the original fuel consumption). The interpretation of an efficiency metric becomes yet more difficult if two different upgrade options are to be compared. This is visualised in Figure 1.5, drawing the efficiency metric (left) and the consumption metric (right) over the consumed energy E (for a fixed provided work). Fig. 1.5(a) illustrates that for a given reduction in energy consumption $\Delta E = E_1 - E_2$, the lower the absolute energy consumption E_1, the larger the increase of the efficiency metric $f(E) \propto 1/E$. This implies that improving
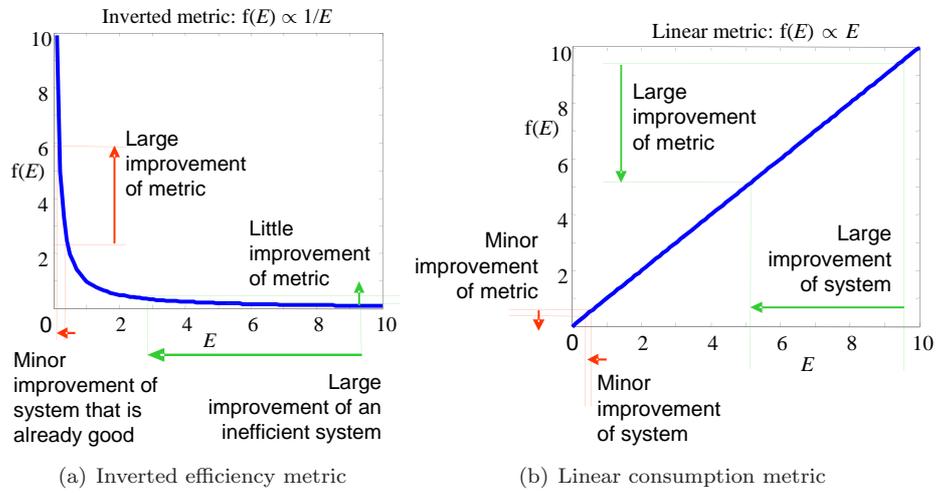
Figure 1.5 A linear consumption metric is easier to interpret than an inverted efficiency metric.

a more inefficient system, with larger absolute energy consumption, will result in a relatively small improvement of the efficiency metric. This may lead to an underestimation of the improvements that actually provide the highest energy savings ΔE. In contrast, a consumption metric where the energy use is in the numerator, $f(E) \propto E$, will always show the same improvement for a given ΔE. Consumption metrics are therefore preferred, since they guide people to take rationale decisions.

1.5.2 Energy Consumption Metrics in Cellular Networks

To capture the energy consumption perspective in the analysis, we employ the two energy consumption indices:

- Power per area unit, measured in $[\mathrm{W/m^2}]$;
- Energy per bit, measured in $[\mathrm{J/bit}]$.

The reason for arguing for two different metrics instead of only one is that they both are relevant and they provide complementary information about the how efficient the energy use in a network is.

The power per area unit metric is defined as the network average power usage divided by the coverage area of the network, P/A, and is measured by the unit $[\mathrm{W/m^2}]$. The metric is a measure for the total energy consumption and is closely related to the CO_2 emissions and the associated carbon footprint. Power per area unit is particularly relevant at low traffic loads, as in this case the network is coverage limited rather than capacity limited. Moreover, since the coverage area A, for which the system is to be evaluated, is typically a predefined constant the metric avoids quotient of variables. This prevents misleading conclusions since

when forming the quotient of variables it is impossible to understand whether an increase of the metric is due to the increase of the numerator, and/or the decrease of the denominator.

In order to be meaningful the power per area metric needs to be complemented by a quality metric. When using the $[W/m^2]$ metric it is important not to directly compare the power per area values of two systems, if they deliver vastly different performance. For instance, a dense network may consume more power per area unit than a sparse network, but the dense network may provide significantly higher user performance. However, given a minimum performance requirement, e.g. the 10 th percentile user throughput shall be at least 1 Mbps, the $[W/m^2]$ metric provides a meaningful comparison of any two systems that both fulfill these requirement(s). This metric is also proposed in [19]; however, in ETSI this quantity is expressed as an efficiency metric, A/P, instead of the consumption metric adopted here.

The metric [W/user] is also commonly used and may be derived from the $[W/m^2]$ metric by applying the conversion factor $[m^2/user]$. If the focus is on evaluating the energy consumption by using a radio network system simulator, which is the case here, then the $[W/m^2]$ metric is more convenient, since the system load is typically adjusted by varying the number of active users. If the purpose instead is to express the performance of a real network based on real measurements, then the W/user metric might be the better choice (since the area covered by a real network is difficult to estimate accurately).

The second metric, which measures the energy per bit in [J/bit], is defined as the network energy consumption E during the observation period T, divided by the total number of bits RT that were correctly delivered in the network. Since the number of successfully transmitted bits is the average rate R, multiplied by the observation period T, this metric could equivalently be described as the average network power $P=E/T$ in relation to the average data rate R, expressed in [W/bps]. The energy consumption metric J/bit focuses on the amount of energy spent per delivered bit and is hence an indicator of network bit delivery efficiency, which may be important especially in scenarios where the traffic load is high. The [J/bit] metric is commonly used in the literature, especially for theoretical studies and single link evaluations. However, in the literature the metric is often used as an efficiency metric in [bit/J], instead as the [J/bit] metric proposed here.

Since both the numerator (the network energy consumption) and the denominator (the number of delivered bits) are typically variable, the metric is affected both by changes in the energy consumption and by changes in the number of delivered bits. Some caution is therefore required when using the J/bit metric. The first thing to note is that, this metric approaches infinity as the traffic load goes to zero. It is also interesting to note that the metric improves in case the amount of delivered bits RT grow faster than the required energy use E. In this respect the metric may be criticized as being self-optimizing. If we e.g. consider a newly deployed system built for area coverage that has only few users then the

energy used per bit will initially be extremely high. As time passes more users will join the network and the bits they generate will drive the metric down, since the extra users do not require any new base stations. Hence an operator may claim that the energy efficiency in terms of J/bit is improving at an impressing rate, while in fact the total energy consumption has increased and nothing has been done to reduce the total energy consumption of the network. Basing decisions only on this metric is certain to result in the conclusion that high data volumes are the best way to improve efficiency. Small cells will be considered more efficient than large cells, 4G more efficient than 3G and 2G, etc., simply because of the higher bit-rates provided.

1.6　Case Study: Energy Efficiency of LTE

For short-term, small-scale evaluations a macro-cellular network with regular hexagonal cell layout is implemented. 19 sites, each with 3 sectors, 10 MHz bandwidth operating at 2.1 GHz carrier frequency is assumed. Moreover, 2×2 MIMO transmission with adaptive rank adaption is assumed. The inter-site distance (ISD) for the dense urban and urban environments is set to 500 m, whereas the ISD for suburban and rural areas is set to 1732 m. The users are uniformly distributed, with population densities corresponding to the respective deployment scenarios. The simulation parameters are taken from 3GPP specifications [4]. Two cases are simulated: a) Base stations without sleep mode, and b) Base stations with micro sleep during idle transmit intervals, i.e, neither data nor control signals are transmitted.

The power per area unit P/A, expressed in [kW/km^2], is depicted in Fig. 1.6. As can be seen, the power consumption increases with the served traffic in the network. In an urban scenario (see Fig. 1.6(a)), with an ISD of 500 m corresponding to a coverage area of 0.2165 km^2 per site, the power per area unit is around 4.15 kW/km^2 at low loads, whereas it approaches 5.1 kW/km^2 at high loads. For the network with micro sleep capable base stations, the corresponding figures are 3.5 kW/km^2 at low loads and above 5 kW/km^2 at high loads. For comparison, an empty network when only control channels are transmitted, but no user data, the power consumption equals 885 W per site, which corresponds to a power per area unit of 4.1 kW/km^2. In the (hypothetical) extreme case, when nothing at all is transmitted (i.e., no data and no control channels) so that the RF output power is 0 W, we obtain P/A=3.3 kW/km^2. The power consumption per area unit for suburban and rural areas, shown in Fig. 1.6(b), is substantially lower, which is due to the increased ISD of 1732 m, which corresponds to a coverage area of 2.6 km^2. However, the system throughput per area unit decreases accordingly, due to the increased site coverage area.

Fig. 1.7 contains the energy consumption per delivered bit, $E/(RT)$ in [kJ/Mbit], over the served data rate R in [Mbps]. Even though the total energy consumption increases with the traffic load, the energy consumption per bit

LTE Downlink

(a) Dense urban/urban scenarios
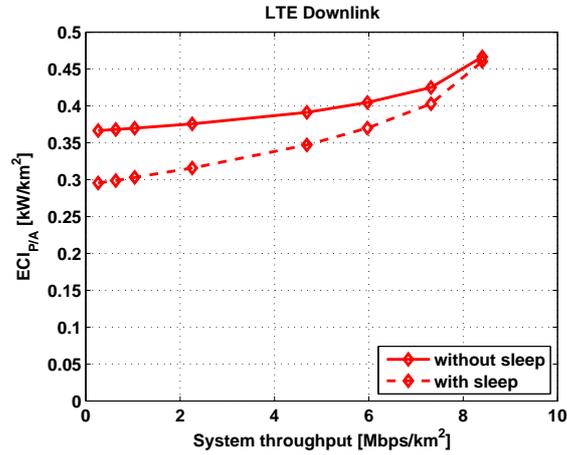
LTE Downlink

(b) Suburban/rural scenarios

Figure 1.6 Power per area unit, P/A, of the downlink of a LTE radio access network as a function of the system load.

decreases with traffic. That is, the number of delivered bits increases faster than the network energy consumption. The dominating reason for this is the fact that the power model (1.2) is associated with a fixed cost at $P_{\text{out}}{=}0\,\text{W}$ RF output power and when the traffic increases, this fixed cost is shared over a larger number of bits, which results into the energy per bit decrease.

In order to assess the expected performance of a country-wide area over a day, short-term, small-scale evaluations are combined with the long-term traffic models and the geographical distribution presented in Section 1.4. It is assumed that no coverage is provided in the sparsely populated and wilderness areas and hence, these areas are not included in the analysis. Moreover, the global traffic
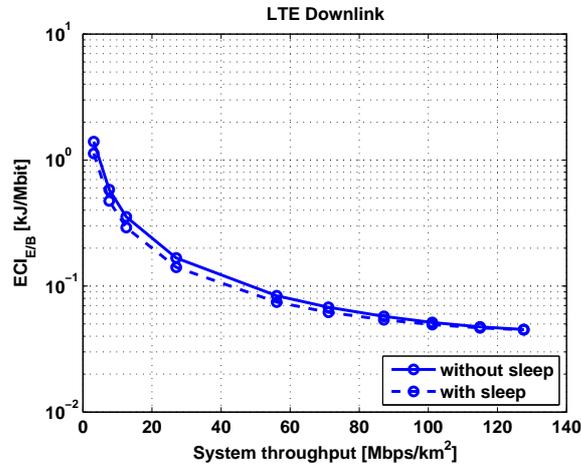
(a) Dense urban/urban scenarios
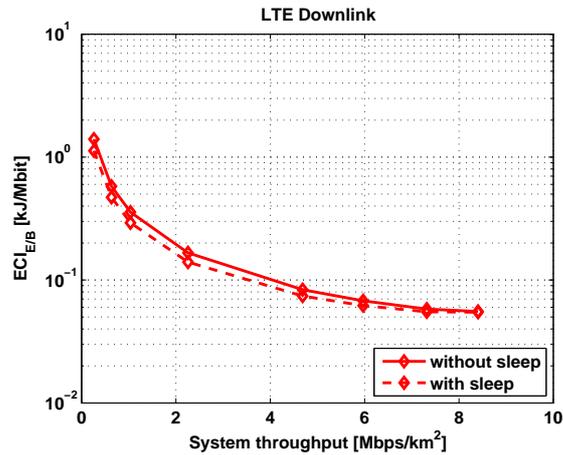
(b) Suburban/rural scenarios

Figure 1.7 Energy per bit of the downlink of a LTE radio access network as a function of the system load.

model in Section 1.4 is for the entire user population, whereas the short-term, small-scale evaluations are performed for a single carrier, which is served by a single operator. It is here assumed that the market share of the studied operator is 30 % and that this operator also carries 30 % of the total data traffic.

The outcome of the aggregation indicates that with the models and assumptions used in this evaluation, the average power per area unit is about $0.6\,\text{kW/km}^2$ without sleep modes, and slightly above $0.5\,\text{kW/km}^2$ for the network with micro sleep capable base stations, corresponding to an energy saving of 15–20%. These values are almost independent of whether high, medium, or low traffic density is assumed, i.e., the power consumption is mostly insensitive to

the traffic load. One reason for this is that given the traffic and deployment mod-
els used here, the network typically operates in the very low load regime, where
transmissions of control signals dominate over data. These evaluations highlight
the fact that on average cellular networks are primarily providing coverage, and
therefore mainly operate at low traffic loads.

1.7   Relation of E^{3}F to the Carbon Footprint of Mobile Communication Systems

Statistical analysis and projections presented in [2] suggest that global radio
access networks might increase from 3.3 million BS sites in 2007 to more than
11.2 million BS sites in 2020 and the total power consumption is expected to
grow from 49 TWh in 2007 to 98 TWh in 2020. During the same period the
average traffic per site will increase from 62 kbit/s to somewhere between 11
to 18 Mbit/s. This implies that by 2020 the radio access network will have to
improve its energy efficiency from about 28 J/kbit to 0.1–0.06 J/kbit.

The simulation results for a 10 MHz LTE system as presented in Section 1.6
suggest for the dense urban case that data rates of up to 26 Mbit/s/site are
attainable with acceptable user rates on average as well as on the cell edge. In the
ideal case where all the traffic is carried by a single operator the efficiency of the
LTE system is about 0.03 J/kbit. In the more realistic case where one operator
has only 30% market share, i.e., there are more than three operators present, all
sites operate in lower load conditions, increasing the power consumtion in Joule
per bit by a factor of three. The overall efficiency obtained then is at about
0.1 J/kbit. This compares well with the targeted efficiency values of 0.1 J/kbit
obtained as a global average from the statistical models and, we can conclude
that LTE already provides the potential for the required reduction in Joule per
bit efficiency.

The simulation based analysis in Section 1.6 assumes a pure LTE network,
which is hardly present in real deployments. Instead, the currently installed base
of equipment will account for the major part of energy consumption and user
demands for data services will to a large extent be fulfilled by heterogeneous
deployments of different standards and cell sizes. In order to achieve the targeted
efficiency values further optimization work is required, e.g. by smart deployment
and dynamic resource and network management. In particular, there is a vast
potential in improving the energy efficiency when the network is *not* transmitting
at maximum load, which is almost always encountered in practice, as a wireless
network is primarily providing coverage. In this case, however, the energy effi-
ciency is particularly poor, as vigorously highlighted by the system simulation
results in Section 1.6.

1.8   Fundamental Challenges And Future Potential

Traditionally, radio access research both in the academy and the industry focuses on the challenge to achieve as high data rates as possible for a given maximum transmission power. Closer analyzing the challenges and the possibilities for the research area of network energy efficiency it is found that there is a second challenge, which is still not widely addressed or even accepted in the community. This second challenge addresses the power consumption of the system when it is not transmitting any data. The first challenge is thoroughly addressed in research and specifications to allow for high peak data rates and capacity, for instance in LTE and HSPA. This very important work, which has been a main driver for the success of mobile telecommunication, as well as rapidly and steadily decreasing energy consumption per bit for the 3GPP technologies, will need to be continued although it is clear that it will be more and more difficult to maintain the rate of improvement with sinking energy consumption per bit. The second challenge, exploiting system operation whilst not transmitting, either by improving the energy efficiency during idle operation, or by finding efficient solutions to eliminate empty resource block transmissions, has so far been a neglected research area. Consequently, this is where the big unexplored potential lies. It is important to understand that this second challenge does not only address empty cells and no load scenarios. In fact the potential of the non-transmitting scenario depends strongly on the considered time scale. Considering a traditional O&M time scales of 15 minutes there may not be many periods, if any, without any transmissions at all. However, LTE scheduling decisions are made per ms, i.e. per every LTE subframe; when addressing this time scale instead, the possibility for idle subframes becomes considerable, even in fairly loaded cells. EARTH-reference scenario simulations of a network covering the dense urban, the urban and the sub-urban areas of a country or region, and assuming the medium traffic profile yield that more than 98% of the subframes do not contain any transmitted data: this number can be seen as a form of theoretical limit for how large the potential is for the second challenge. As in this 98% of time about 97% of energy is consumed by the system, mechanisms addressing individual subframes without data transmissions could deliver up to 97% of power savings. This is in addition to the power reduction achievable by features addressing subframes that are utilized for data transmissions. Some examples of future potentials in energy saving are given by improvements at component level, e.g. introducing new power management concepts able to adapt to varying traffic load. In fact introducing scalability into hardware components, and supporting them by dynamic power management, enables the adaptation of energy consumption to actual performance requirements. Further power savings are facilitated by the deactivation of components in time periods of no operation. These hardware characteristics could be exploited at packet scheduling level, that has in charge to efficiently allocate radio resources, in order to properly

manage empty subframes and drive components deactivation. Finally, it should be made clear that different solutions can be envisaged by considering different time perspectives: starting from what energy saving can be achieved on ms, or even mu s scale, up to the possibility of savings when taking down cells or switching off whole BSs, something which will require actions on a time scale in the order of seconds or even minutes. As a matter of fact, all solutions on different time scales constitute part of the road toward the future, ultimately energy-efficient radio-access networks.

1.9   Conclusions

In order to identify the key levers for energy savings the power consumption of mobile communication systems needs to be quantified. This includes sophisticated power models that map the radiated RF power to the supply power of a BS site, as well as traffic and deployment models that extend short-term small-scale evaluations to the country wide power consumption of a network over a whole day or week. Numerical results reveal that for current network design and operation, the power consumption is mostly independent of the traffic load. This highlights the vast potential for energy savings by improving the energy efficiency of BSs at low load.

1.10   Acknowledgments

References

[1]   International Telecommunication Union, "Worldwide mobile cellular subscribers to reach 4 billion mark late 2008," Press release, 2008.

[2]   A. Fehske, J. Malmodin, G. Biczok, G. Fettweis, "The Global Carbon Footprint of Mobile Communications - The Ecological and Economic Perspective", *IEEE Communications Magazine*, to appear, 2011.

[3]   L.M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Godor, G. Auer, L. Van der Perre, "Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks," *IEEE Communications Magazine* special issue on green radio, pp. 66-72, Nov. 2010.

[4]   3GPP TR 36.814 v9.0.0, "Further advancements for E-UTRA. Physical layer aspects (Release 9)," 3GPP, Technical Specification Group Radio Access Network, Mar. 2010.

[5]   WINNER II, "Deliverable D6.13.7: Test Scenarios and Calibration Cases Issues 2", IST-4-027756 WINNER, Dec. 2006.

[6]   International Telecommunication Union, Report ITU-R M.2134, "Requirements related to technical performance for IMT-Advanced radio interface(s)," 2008. `http://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2134-2008-PDF-E.pdf`

[7]   S.C. Cripps, "RF Power Amplifiers for Wireless Communications", Artech House Microwave Libreary, 2nd Edition, 2006.

[8]   J. Xu, "Practical Digital Pre-Distortion Techniques for PA Linearization in 3GPP LTE", Agilent Technologies, 2010.

[9]   W. Kim, S.P. Stapleton, J.H. Kim and C. Edelman. "Digital Pre-Distortion Linearizes Wireless Power Amplifiers", *IEEE Microwave Magazine*, Sep. 2005.

[10]   B. Debaillie, A. Giry, M.J. Gonzalez, L. Dussopt, M. Li, D. Ferling, V. Giannini, "Opportunities for Energy Savings in Pico/Femto-cell Base-Stations," in *Proc. Future Networks and Mobile Summit*, Warsaw, Poland, 2011.

[11]   S. Borkar, "Design Challenges of Technology Scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23-29, July 1999.

[12]   National Media and Infocommunications Authority Hungary, "Report on mobile internet coverage," 2010. `http://www.nmhh.hu/dokumentum.php?cid=25074`

[13]   "Entscheidung der Präsidentekammer der Bundesnetzagentur fur Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen" (in German), October 2009. `http://www.bundesnetzagentur.de/cae/servlet/contentblob/138464/publicationFile/2807/PraesKammerEntschg_Id17404pdf.pdf`

[14]   UMTS Forum, "Mobile Traffic Forecasts 2010-2020", Report no. 44, May 2011 `http://www.umts-forum.org/component/option,com_docman/task,cat_view/gid,485/Itemid,213/`

[15] Sandvine, "Mobile Internet Phenomena Report", 2010. `http://www.sandvine.com/` `downloads/documents/2010` Global Internet Phenomena Report.pdf

[16] Akamai, "State of the Internet" report, 2010. `http://wwwfp.akamai.com/dl/` `whitepapers/akamai_state_of_the_internet_q1_2010.pdf`

[17] National Media and Infocommunications Authority of Hungary, "Report on mobile data usage", 2010. `http://www.nhh.hu/dokumentum.php?cid=24769`

[18] The Swedish Post and Telecom Agency, "Telecommunication Markets in the Nordic Countries", 2010. `http://statistik.pts.se/pts1h2010e/download/PTS_ER_` `2010_26_Svensk_Telemarknad_2010_1h_en.pdf`

[19] ETSI TS 102 706 v1.1.1, "Environmental Engineering (EE); Energy Efficiency of Wireless Access Network Equipment," Aug. 2009.