1

Green Network Fundamentals

Efficient energy usage in wireless networks has drawn significant attention from both academia and industry, mainly because of critical environmental, financial, and quality-of-experience (QoE) concerns. Research efforts have led to various solutions that allow efficient use of energy in wireless networks. Such approaches are referred to as *green wireless communication and networking*. Throughout this book, our main focus is on developing energy-efficient communication techniques in base stations (BSs) and mobile terminals (MTs), as they represent the major sources of energy consumption in wireless access networks, from the operator and user perspectives, respectively, while accounting for the heterogeneous nature of the wireless communication medium. Towards this end, the first two chapters of the first part of this book are dedicated to introducing the background concepts of green networking. The first chapter discusses the need for green (energy-efficient) communications, the modelling techniques used for energy efficiency and call traffic in wireless networks, and different conflicting performance metrics. Building on such a background, the second chapter reviews the state-of-the-art green communication solutions and analytical models proposed for network operators and mobile users at different traffic load conditions, and points out their major shortcomings.

1.1 Introduction: Need for Green Networks

In response to the increasing demand for wireless communication services during the past decade, there has been wide deployment of wireless access networks [1]. By definition, a wireless access network is a wireless system that uses BSs and access points (APs) to interface MTs with the core network or the Internet [2]. Hence, the main components of a wireless access network are BSs/APs and MTs [3]. BSs/APs are mainly in charge of radio resource control and user mobility management, and provide access to the Internet. MTs are equipped with processing and display capabilities, and provide voice services, video streaming, and data applications to mobile users. Currently, MTs are provided with multiple radio interfaces, and mobile users can connect to different networks, such as cellular networks, wireless local area networks (WLANs), and wireless metropolitan area networks (WMANs), and enjoy single-network and/or multi-homing services [4–6].

Green Heterogeneous Wireless Networks, First Edition. Muhammad Ismail, Muhammad Zeeshan Shakir,
Khalid A. Qaraqe and Erchin Serpedin.

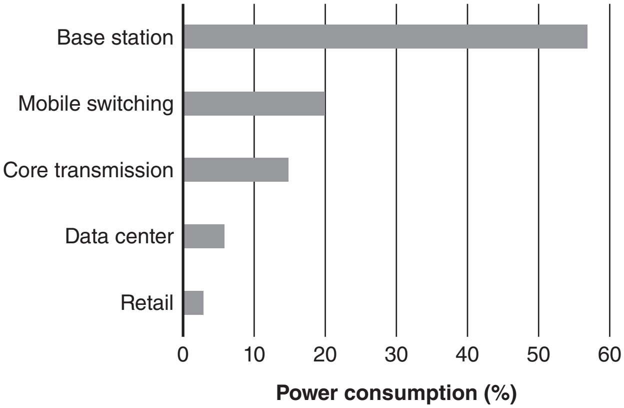


Figure 1.1 Breakdown of power consumption of a wireless cellular network [7]

From the network operator side, BS is the main source of energy consumption in the wireless access network [2]. The breakdown of a cellular network's typical power consumption is shown in Figure 1.1, which shows that almost 57% of the operator's total power consumption is in the BS [2, 8, 9]. Worldwide, there are about 3 million BSs, which consume in total 4.5 GW of power [10]. From the user side, it has been estimated that there exist roughly 3 billion MTs in the world with a total power consumption of 0.2–0.4 GW [11]. Such high energy consumption of wireless access networks has triggered environmental, financial, and QoE concerns for both network operators and mobile users.

From an environmental standpoint, the telecommunications industry is responsible for 2% of the total CO_2 emissions worldwide, and this percentage is expected to double by 2020 [12]. As shown in Figure 1.2, the mobile communications sector has contributed 43% of the telecommunication carbon footprint in 2002, and this contribution is expected to grow to 51% by 2020 [14]. Furthermore, the MT rechargeable batteries' expected lifetime is about 2–3 years and manifests in 25,000 t of disposed batteries annually, a factor that raises environmental concerns (and financial considerations for the mobile users as well) [15]. In addition, the high energy consumption of BSs and MTs is a source of high heat dissipation and electronic pollution [16]. From a financial standpoint, a significant portion of a service provider's annual operating expenses is attributed to energy costs [17, 18]. Technical reports have indicated that the cost of energy bills of service providers ranges from 18% (in mature markets in Europe) to 32% (in India) of the operational expenditure (OPEX) [19, 20]. The energy expenses reach up to 50% of the OPEX for cellular networks outside the power grid [21, 22]. Finally, from a user QoE standpoint, it has been reported that more than 60% of mobile users complain about their limited battery capacity [23]. In addition, the gap between the MT's offered battery capacity and the mobile users' demand for energy is growing exponentially with time [24]. Consequently, the MT's operational time between battery chargings has become a crucial factor in the mobile user's perceived quality-of-service (QoS) [25].

The aforementioned concerns have triggered increasing demand for energy-efficient solutions in wireless access networks. Research efforts carried out in this direction are referred to as *green network solutions*. The term 'green' confirms the environmental dimension of the proposed approaches. Therefore, a cost-effective solution that is not eco-friendly is

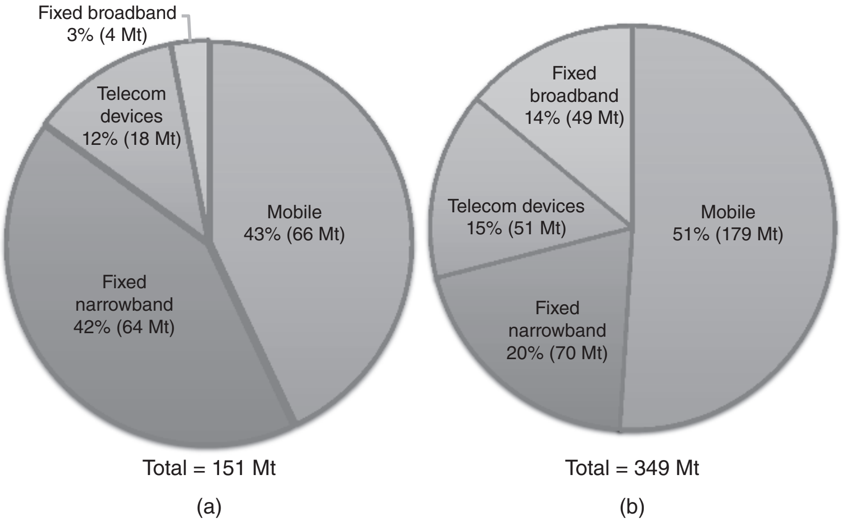


Figure 1.2 Carbon footprint contribution by the telecommunications industry: (a) 2002 and (b) 2020 [13]

not attractive. For instance, having a cost-effective electricity demand schedule for a network operator that relies on different electricity retailers, in a liberated electricity market, is not considered a green solution if it does not ensure that the proposed solution is also eco-friendly in terms of the associated carbon footprint [26]. The objectives of the green wireless communications and networking paradigm are, therefore, (i) reducing energy consumption of communication devices and (ii) taking into account the environmental impacts of the proposed solutions.

In order to develop/analyse a green networking solution, an appropriate definition of energy efficiency/consumption for network operators and mobile users should be formulated. This definition should account for the power consumption, throughput, traffic load models, and conflicting performance metrics for network operators and mobile users. The first chapter of this book is dedicated to building this necessary background.

1.2 Traffic Models

Some energy-efficiency and consumption models are defined on the basis of the temporal fluctuations in the traffic load. In addition, different green approaches can be adopted at different traffic load conditions. Furthermore, some green approaches rely on the temporal and spatial fluctuations in the traffic load to save energy. For instance, in order to determine the sleep duration of a BS or MT, traffic models are used to probabilistically predict the idle period duration, as will be presented in Chapter 2. Moreover, the performance evaluation of the green approaches should be carried out using an appropriate traffic model. Consequently, it is necessary to gain a better understanding of the different traffic load models proposed in the literature before introducing energy efficiency and consumption models as well as green solutions.

Table 1.1 Summary of different traffic models [27]

Model					Comments	References
Static					It does not capture the MT mobility and the traffic dynamics	[23, 28–34]
Dynamic	Spatial	Regional traffic load density			It defines a location -based traffic load density	[35]
		Stochastic geometry			BSs and MTs are located according to a homogeneous Poisson point process	[18]
		FSMC			It models the spatial distribution of MTs within a cell	[36]
	Temporal	Long-scale			The model captures traffic fluctuations over the days of the week	[17, 37–39]
		Short scale	Flow-level	Poisson-exponential	It models call arrivals as a Poisson process and call departures as an exponential distribution	[12, 40–42]
				FSMC	The number of calls within a cell is represented by a state in a Markov chain	[36]
			Packet-level	Infinite buffer	It models the number of backlogged packets in an MT buffer with infinite capacity	[43]
				Finite buffer	It models the number of backlogged packets in an MT buffer with finite capacity	[44]

Overall, the traffic modelling can be categorized into two classes, as shown in Table 1.1. The first class is referred to as the *static model* and assumes a fixed set of MTs, $\mathcal{M}$, that communicate with a fixed set of BSs, $\mathcal{S}$ [23, 28–34, 45]. The static model suffers from several limitations. First, it does not consider the mobility of MTs in terms of their arrivals and departures. Second, it does not capture the call-level or packet-level dynamics in terms of call duration, packet arrival, and so on. On the other side, the second class, which is referred to as the *dynamic model*, captures the spatial and temporal fluctuations of the traffic load, and is discussed next in detail.

1.2.1 Traffic Spatial Fluctuation Modelling

Studies have indicated that traffic is quite diverse even among closely located BSs, as shown in Figure 1.3, [37, 38]. As a result, different models have been proposed in the literature to reflect the spatial fluctuations in call traffic load [18, 35, 36].

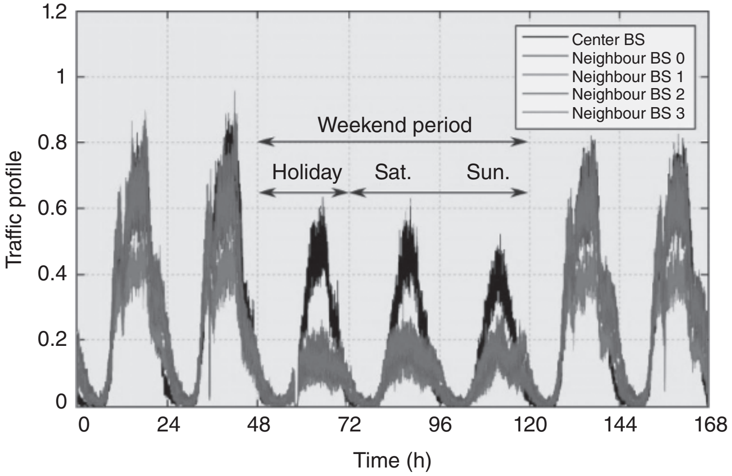


Figure 1.3 Spatial and temporal traffic fluctuations [38]

Location-based traffic load density is one approach to capture traffic spatial fluctuations [35]. In this context, a geographical region is covered by a set $\mathcal{S}$ of BSs and the region is partitioned into a set of locations. In a given location x, the file transfer request arrivals follow an inhomogeneous Poisson point process (PPP) with an arrival rate $\lambda(x)$ per unit area. The file sizes are independently distributed with mean $1/\mu(x)$ at the location. Consequently, the traffic load density is given by $\varrho(x) = \lambda(x)/\mu(x) < \infty$, which is used as a measure of the spatial traffic variability.

The aforementioned approach adopts a pre-defined set of BSs, $\mathcal{S}$, with specific locations. An alternative approach, which is more suitable for a design stage, defines the locations of BSs based on the stochastic geometry theory [18]. Hence, the network's n BS locations follow a homogeneous PPP, Θ_n, with intensity θ_n in the Euclidean plane. Similarly, MTs are located according to a different independent stationary point process with intensity θ_m. According to the stationary PPP Θ_n, the distance between an MT and its serving BS, D_m, follows the same distribution regardless of the MT's exact location. The probability density function (PDF) of D_m is expressed as [18]

$$f_{D_m}(d) = 2\pi\theta_n d \exp(-\theta_n \pi d^2), \quad d > 0. \tag{1.1}$$

The aforementioned models reflect the spatial variability of the traffic among different cells. To capture the spatial distribution variability of MTs within a given cell i, a finite-state Markov chain (FSMC) model is adopted [36]. This model classifies the MTs into G groups according to cell i's radius. Assuming there are M MTs in cell i, a G^M spatial location distribution is considered within the cell. Thus, the FSMC model presents $\mathcal{L} = \{L_1, \ldots, L_{G^M}\}$ states. The state transition probability $\Pr\{L_i(t+1) = v_i | L_i(t) = u_i\}$ is the probability of the spatial distribution of the MTs within the cell i at time slot $t+1$ to assume v_i, given that it was u_i at time slot t, where $u_i = \{u_{i,1}, \ldots, u_{i,M}\}$ and $v_i = \{v_{i,1}, \ldots, v_{i,M}\}$. Following this model, the dynamic fluctuations in the number of MTs in different regions within the cell can be captured.

1.2.2 Traffic Temporal Fluctuation Modelling

Two different time scales can capture the temporal fluctuations in the traffic load [12, 39]. The first time scale is a long-term one that reflects the traffic variations over the days of the week. Such a model can help in evaluating different energy-efficient approaches for network operators, as it captures both high and low call traffic load conditions. The second time scale is a short-term one that reflects the call (packet) arrivals and departures of the MTs. Such a model plays a vital role in evaluating energy-efficient resource allocation schemes for MTs and BSs. In the following subsections, we describe the two scales.

1.2.2.1 Long-Term Traffic Fluctuations

Real call traffic traces demonstrate a sinusoidal traffic profile in each cell, as shown in Figure 1.3, [17, 38]. During daytime (11 am–9 pm), traffic is much higher than that during nighttime (10 pm–9 am) [17, 37]. Furthermore, during weekends and holidays, the traffic profile, even during the peak hours, is much lower than that of a normal week day [17]. The traffic profile during a weekday is 10% less than its peak value 30% of the time, and this increases to 43% of the time during weekends [17]. This behaviour can be captured using an activity parameter $\psi(t)$, which specifies the percentage of active subscribers over time t, as shown in Figure 1.4 [39]. Denote p as the population density of users per km^2, N as the number of operators (each being able to carry $1/N$ of the total traffic volume), and M_k as the fraction of subscribers with an average data rate r_k for terminal type k (e.g. smart phone and tablet). Hence, the traffic demand, in bits per second per km^2, is given by

$$A(t) = \frac{p}{N}\psi(t)\sum_{k} M_k r_k. \tag{1.2}$$

Studies have indicated that the traffic load difference between two consecutive days for 70% of the BSs is less than 20% [37]. As a result, the long-term fluctuations in call traffic load can

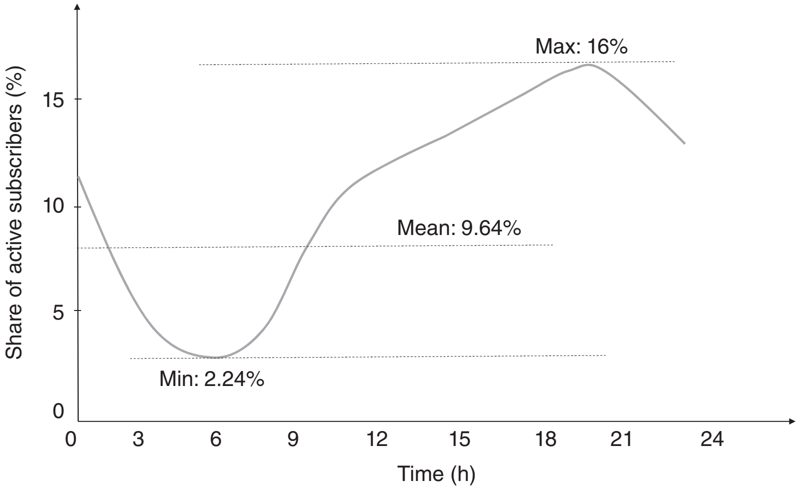


Figure 1.4 Average daily data traffic profile in a European country [39]

be estimated from the historical mobile traffic records; that is, the activity parameter $\psi(t)$ and the average data rate r_k can be inferred in practice from historical data.

1.2.2.2 Short-Term Traffic Fluctuations

Two categories can be distinguished for short-term traffic fluctuation models, namely call (flow)-level and packet-level models. Call (flow)-level models are useful in designing and evaluating green resource scheduling mechanisms at both BSs and MTs under high call traffic load. For myopic resource allocation solutions, the call arrivals are modelled using a Poisson process with rate λ, and the call durations are represented by an exponential distribution [12, 40–42]. Dynamic resource allocation solutions rely on FSMC to model traffic dynamics in terms of call arrivals and departures [36]. In this model, the number of calls in a given cell i is captured by an M-state Markov chain, with the state set $\mathcal{M} = \{0, 1, \ldots, M-1\}$. The state transition probability $\Pr\{M_i(t+1) = m_i | M_i(t) = \widetilde{m}_i\}$ is the probability of having m_i MTs within cell i at time slot $t+1$, given that there were $\widetilde{m}_i$ MTs at time slot t, where m_i, $\widetilde{m}_i \in \mathcal{M}$.

In a low call traffic load condition, packet-level traffic models are useful in designing and evaluating green resource solutions (on–off switching) at the BSs and MTs, through modelling the BS/MT buffer dynamics in terms of packet arrival and transmission [43, 44]. For an infinite buffer size, the MT buffer dynamics can be expressed as

$$o_m(t+1) = \max\{o_m(t) + a_m(t+1) - z_m(t), a_m(t+1)\}, \tag{1.3}$$

where $o_m(t)$, $a_m(t)$, and $z_m(t)$ are the numbers of backlogged packets in the buffer, arriving packets, and transmitted packets, for MT m in time slot t, respectively. For a buffer with a finite size F, the MT buffer dynamics can be represented by

$$o_m(t+1) = \min\{o_m(t) + a_m(t+1) - z_m(t), F\}. \tag{1.4}$$

The models (1.3) and (1.4) are used to investigate the optimal on–off switching mechanisms for the radio interfaces of MTs to achieve energy-efficient (green) communications at a low call traffic load condition, a topic that will be addressed in Chapter 2.

1.3 Energy Efficiency and Consumption Models in Wireless Networks

Following the temporal and spatial fluctuations in traffic load, this section summarizes different definitions that have been proposed in the literature to assess energy consumption/efficiency of wireless networks. Towards this end, we first present different throughput and power consumption models for BSs and MTs.

1.3.1 Throughput Models

The utility obtained from the wireless network in exchange for its consumed power is expressed most of the time in terms of the achieved throughput. In this context, we first introduce the concepts of aggregate BS capacity C_s, area spectral efficiency T_s, and user-achieved data rate R_m, which will be used in the energy efficiency definitions to be presented later.

1.3.1.1 Network Side

The BS aggregate capacity C_s for BS s is measured using Shannon's formula as follows [26]

$$C_s = B_s \log_2 \det(I + PH), \tag{1.5}$$

where B_s denotes the total bandwidth of BS s, I represents the unit matrix, P is the transmission power vector of BS s to every MT m in service, and H stands for the channel gain matrix between BS s and each MT m, which accounts for the channel's fast fading, noise, and interference affecting the radio transmission. The BS capacity C_s in (1.5) is measured in bits per second (bps).

At a low call traffic load condition, the area spectral efficiency T_s provides a better representation of the BS's attained utility than the BS's aggregate capacity since it accounts for the coverage probability, which matters the most at such a condition [18]. Specifically, T_s measures the BS throughput while considering the coverage probability. Denote $\Pr\{\gamma_{x \to u} > \zeta\}$ as the success probability of the signal-to-noise ratio (SNR) γ received by an MT at location u from a given BS at some location x satisfying a certain QoS threshold ζ. Averaging the success probability $\Pr\{\gamma_{x \to u} > \zeta\}$ over the propagation range to location u yields the coverage probability $\mathbb{P}_s(\zeta)$. For BS s, the area spectral efficiency T_s measured over a unit area is expressed as

$$T_s = \mathbb{P}_s(\zeta) \log_2(1 + \zeta). \tag{1.6}$$

1.3.1.2 Mobile Terminal Side

While the definitions in (1.5) and (1.6) are mainly from the operator side, two definitions can be used to quantify the mobile user's attained utility (in terms of the achieved data rate R_m in the uplink by MT m) in exchange for the MT power consumption. Given the instantaneous channel state information (CSI), the achieved data rate R_m in bps can be expressed as [16, 28, 45, 46]

$$R_m = B_m \log_2 \left(1 + \frac{\gamma_m}{\Gamma}\right), \tag{1.7}$$

where B_m stands for the uplink allocated bandwidth to MT m, γ_m represents the SNR of MT m received at the destination, and Γ denotes the SNR gap between the channel capacity and a practical coding and modulation scheme. For the Shannon formula, $\Gamma = 1$. Reporting instantaneous CSI from each MT to the serving BS, in order to determine (1.7), leads to a large signalling overhead. In order to reduce the associated signalling overhead, a statistical CSI is used. Consequently, R_m in bps is expressed as

$$R_m = \mathbb{E}_H \left[B_m \log_2 \left(1 + \frac{\gamma_m}{\Gamma}\right)\right], \tag{1.8}$$

where $\mathbb{E}_H$ represents the expectation over the channel state H.

1.3.2 Power Consumption Models

In order to attain the aforementioned utilities in (1.5)–(1.8), power is consumed at both the network side and user side. In the literature, different models are proposed to capture such a power consumption, as summarized in Table 1.2. These models are next discussed.

Table 1.2 Summary of different power models proposed in the literature [27]

	Model			Comments	References
BS	Operation only	Large-cell	Ideal	The BS consumes no power when idle, that is, the BS consists only of energy proportional devices	[35]
			Realistic	The model captures the BS traffic load independent power consumption	[12, 18, 21, 26, 35, 39]
		Femto-cell	Load independent	The BS power consumption does not depend on the offered traffic load	[47]
			Load dependent	The BS power consumption relies on traffic load, packet size, and has an idle part	[48]
		Including temporal fluctuations		The model accounts for full load, half load, and idle traffic conditions	[7]
		Backhaul power consumption		The model defines power consumption for micro-wave and optical fiber backhaul links	[49]
	Operation and embodied			Besides the operation power, it accounts for the consumed energy in BS manufacturing and maintenance	[19]
MT	Transmission power only	Without power amplifier efficiency		The model does not account for the transmitter power amplifier efficiency	[29, 46, 50]
		With power amplifier efficiency		The model accounts for the transmitter power amplifier efficiency	[16, 30, 45, 51]
	Including circuit power	Constant		The circuit power consumption is given by a constant term independent of the bandwidth and data rate	[16, 23, 28, 31, 45, 52]
		Bandwidth scale		The circuit power consumption scales with the MT assigned bandwidth	[53]
		Data rate scale		The circuit power consumption scales with the MT achieved data rate	[51]
	Including reception power			Besides the transmitter and circuit power consumption, the model also accounts for the receiver power consumption	[16,53]

1.3.2.1 Network Side

The total power consumption P_n of a wireless access network n, from the network operator perspective, can be captured using the aggregate power consumption of the network BSs. Recently, in addition to the BS power consumption, more emphasis is put on the

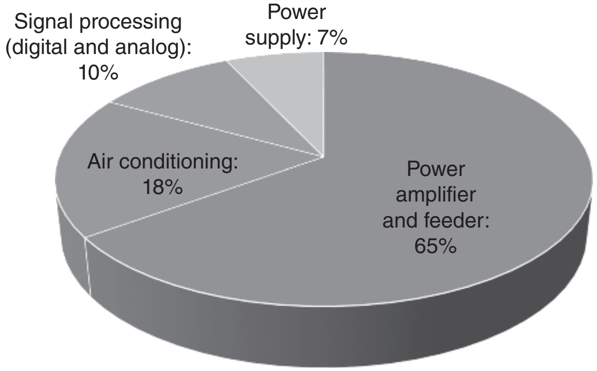


Figure 1.5 Percentage of power consumption at different components of a large-cell BS [27]

Table 1.3 Power consumption profile for a femto-cell BS [27]

Hardware component	Power consumption (W)	Percentage (%)
Microprocessor Associated memory Backhaul circuitry	1.7 0.5 0.5	26.4
FPGA Associated memory Other hardware functions	2 0.5 1.5	39.2
RF transmitter RF receiver RF power amplifier	1 0.5 2	34.3

backhaul power consumption, due to the information exchange among BSs for cooperative transmission/networking. Next, we will outline the different power consumption models proposed for BSs and backhauls.

For a large-cell BS (macro- and micro-BS), Figure 1.5 illustrates the power consumption percentage of different components of the BS. Furthermore, the power consumption profile of a femto-cell BS is shown in Table 1.3. According to Figure 1.5 and Table 1.3, the following facts turn out:

- The signal processing part is responsible for most of the power consumption in a femto-cell BS as opposed to a large-cell BS (namely, 65.6% and 10% for femto and large-cell BSs, respectively).
- The radio frequency (RF) transmission/reception power consumption in a femto-cell BS is almost half of that of a large-cell BS, with only 19.6% of the power consumed in the femto-cell BS power amplifier as opposed to 65% in a large-cell BS.

In the literature, different models are adopted to represent the BS power consumption P_s. For a large-cell BS, the simplest model is an ideal load-dependent representation, which assumes that the BS consumes no power in its idle state, that is, the BS consists of energy-proportional devices [35]. Hence, the BS power consumption can be expressed as

$$P_s = \rho P_{ts}, \tag{1.9}$$

where ρ stands for the system traffic load density, and P_{ts} denotes the BS's transmitted power. The major limitation with such a model is that it is unrealistic, as the power consumption of some BS components in reality is not load-dependent, as shown in Figure 1.5 (e.g. power supply and air conditioning). To capture the power consumption of both load-dependent and load-independent components in the BS, a more sophisticated model assumes the following expression [39]

$$P_s = \frac{\dfrac{P_{ts}}{\xi(1-\sigma_{\text{feed}})} + P_{\text{RF}} + P_{\text{BB}}}{(1-\sigma_{\text{DC}})(1-\sigma_{\text{MS}})(1-\sigma_{\text{cool}})}, \tag{1.10}$$

where P_{RF} represents the RF power consumption, P_{BB} denotes the baseband unit power consumption, ξ is the power amplifier efficiency, and σ_{feed}, σ_{DC}, σ_{MS}, and σ_{cool} stand for the losses incurred by the antenna feeder, DC–DC power supply, main supply, and active cooling, respectively. The model (1.10) is further approximated using a linear (affine) function for simplicity [12, 18, 21, 26, 35]. The affine function consists of two components to represent P_s. The first term is denoted by P_f and represents a fixed (load-independent) power component that captures the power consumption at the power supply, cooling, and other circuits. The second term is a load-dependent component. The affine model is expressed as

$$P_s = \Delta_s P_{ts} + P_f, \tag{1.11}$$

where Δ_s is the slope of the load-dependent power consumption.

For a femto-cell BS, the power consumption model is described by Deruyck et al. [47]

$$P_s = P_{\text{mp}} + P_{\text{FPGA}} + P_{\text{tx}} + P_{\text{amp}}, \tag{1.12}$$

where P_{mp}, P_{FPGA}, P_{tx}, and P_{amp} denote the power consumption of the microprocessor, field-programmable gate array (FPGA), transmitter, and power amplifier, respectively. While the power consumption model in (1.12) captures most of the components in Table 1.3, it does not exhibit any dependence on the call traffic load. Experimental results in [48] have pointed out the dependence of the femto-cell BS power consumption on the offered load and the data packet size. Consequently, the power consumption model for a femto-cell BS is expressed by Riggio and Leith [48]

$$P_s = P_d(q, l) + P_f, \tag{1.13}$$

where $P_d(q, l)$ represents the BS power consumption, which depends on the traffic load q [Mbps] and packet size l [bytes], and P_f stands for the idle power consumption component.

In order to capture the temporal fluctuations in the call traffic load, as discussed in Section 1.2.1, a weighted sum of power consumptions at different traffic load conditions (full load, half load, and idle conditions) is considered [7]

$$P_{s,\text{total}} = 0.35P_{\max} + 0.4P_{50} + 0.25P_{\text{sleep}}, \tag{1.14}$$

where $P_{\max}$, P_{50}, and P_{sleep} denote the full rate, half rate, and sleep mode power consumption, respectively. The weights in (1.14) are determined statistically based on the historical traffic records.

Recently, cooperative networking among different BSs and APs in the heterogeneous wireless medium is regarded as an effective approach to enhance the network's overall capacity and reduce the associated energy consumption [1, 4–6, 54]. However, this approach relies on information exchange among different BSs and APs, such as CSI, call traffic load, and resource availability, which are carried mainly over the backhaul connecting these BSs and APs together. Hence, more emphasis is given to the backhaul design and its power consumption. Three types of backhaul solutions can be distinguished, namely copper, microwave, and optical fibre. The most common choice for backhaul is the copper lines [49]. Microwave backhauls are deployed in locations where it is difficult to deploy wired (copper) lines. Also, optical fibre backhauls are mainly used in locations with high traffic due to their high deployment cost. Current research is focusing mainly on the power consumption of microwave and optical fibre links, as they can support the current high data rates. In its simplest form, the microwave (wireless) backhaul power consumption is expressed as [49]

$$P_{\mathrm{BH}} = \frac{C_{\mathrm{req,s}} P_{\mathrm{mw}}}{C_{\mathrm{mw}}}, \tag{1.15}$$

where $C_{\mathrm{req,s}}$ and C_{mw} represent the BS's required backhaul capacity and the microwave backhaul total capacity (100 Mbps), respectively, and P_{mw} denotes the associated power consumption (50 W). However, the model in (1.15) does not account for many features of the backhaul. To gain a better understanding of the power consumption of backhauls, we first provide a brief description of the backhaul structure and associated topologies.

As shown in Figure 1.6, each BS is connected to one or more BSs via a backhaul link. All traffic from BSs is backhauled through a hub node (traffic aggregation point) [55]. Any BS in the network can serve as such a hub node. In general, more than one aggregation level (hub node) can be present. Each hub node is connected to a sink node, which, in turn, is connected to the core network. A BS is equipped with a switch if more than one backhaul link originates or terminates at this BS. Following this description, the microwave backhaul power consumption is expressed as [49]

$$P_{\mathrm{BH}} = P_{\mathrm{sink}} + \sum_{s=1}^{S} P_{\mathrm{BH},s}, \tag{1.16}$$

where P_{sink} is the power consumption at the sink node, $P_{\mathrm{BH},s}$ denotes the power consumption associated with the backhaul operations at BS s, and S stands for the total number of BSs. The following relationships hold

$$P_{\mathrm{BH},s} = P_s(C_{\mathrm{req},s}) + P_{\mathrm{switch,s}}(A_s, C_{\mathrm{req},s}), \tag{1.17}$$

$$P_{\mathrm{sink}} = P_{\mathrm{sink}}(C_{\mathrm{req,sink}}) + P_{\mathrm{switch,sink}}(A_{\mathrm{sink}}, C_{\mathrm{req,sink}}), \tag{1.18}$$

where $C_{\mathrm{req},s}$ and $C_{\mathrm{req,sink}}$ represent the required backhaul capacity for BS s and the sink node, respectively. The variable A denotes the number of microwave antennas, P_s and P_{sink} represent the power consumed for transmitting and receiving backhaul traffic for BS s and the sink node, respectively, and P_{switch} models the BS/sink switch power consumption. On the other hand, for an optical fibre backhaul, the power consumption is expressed as [49]

$$P_{\mathrm{BH}} = \left\lceil \frac{S}{\max N_{\mathrm{DL}}} \right\rceil P_{\mathrm{switch}} + S P_{\mathrm{DL}} + N_{\mathrm{UL}} P_{\mathrm{UL}} + \sum_{s=1}^{S} c_s, \tag{1.19}$$

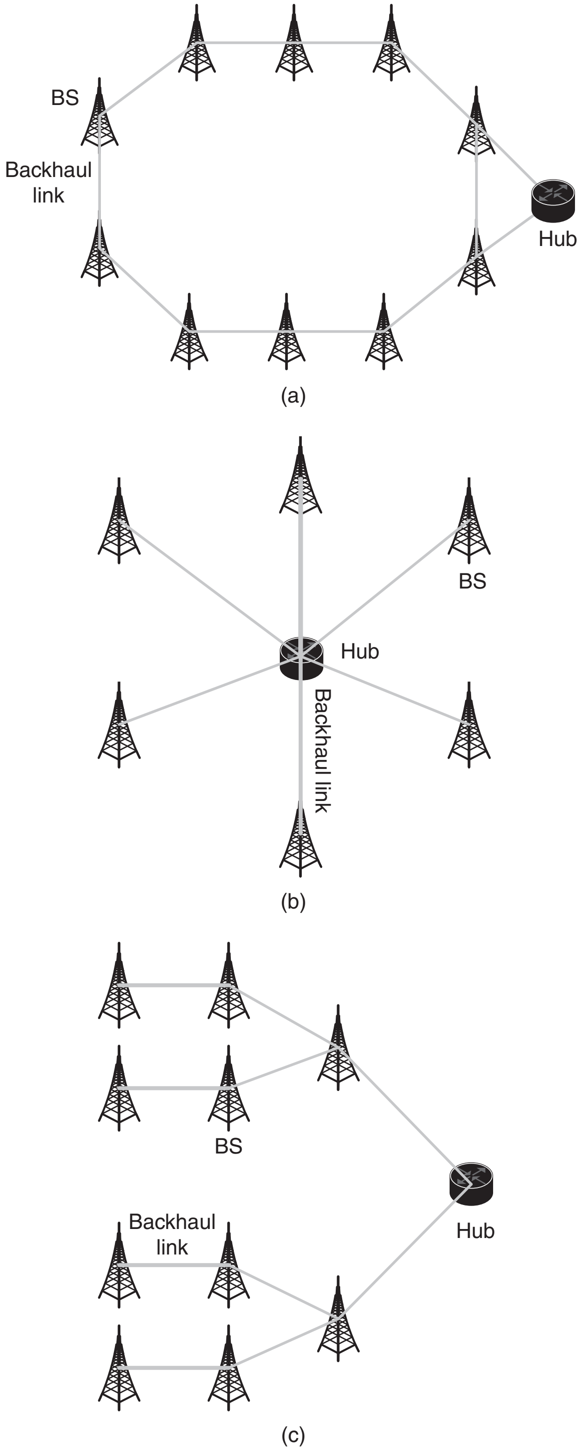


Figure 1.6 Different backhaul topologies [55]: (a) ring topology, (b) star topology, and (c) tree topology

where $\max N_{\mathrm{DL}}$ stands for the maximum number of downlink interfaces available at one aggregation switch, P_{DL} denotes the power consumption due to one interface of a switch, N_{UL} and P_{UL} represent the total number of uplink interfaces and power consumption of one uplink interface, and c_s denotes the power consumption of a pluggable optical interface, which is used to connect a BS to the switch at the hub node.

A limitation with the models (1.9)–(1.16) is that they focus mainly on the BS's operation power. In a more general model, the BS's total consumption is described in terms of the BS's operating energy and embodied energy, E_o and E_e, respectively. The BS's embodied energy represents 30–40% of the BS's total energy consumption [19] and accounts for the energy consumed by all the processes associated with the manufacturing and maintenance of the BS. Over the BS's lifetime, the embodied energy is calculated as 75 GJ [19]. It consists of two components. The first component refers to the initial embodied energy E_{ei}, while the second one stands for the maintenance embodied energy E_{em}. The initial embodied energy comprises the energy used to acquire and process raw materials, manufacture components, and assemble and install all BS components. The initial embodied energy is accounted for only once in the initial BS manufacturing process. The maintenance embodied energy includes the energy associated with maintaining, repairing, and replacing the materials and components of the BS throughout its lifetime. Thus, the BS's total energy consumption (in joules) throughout its lifetime is given by Humar et al. [19]

$$E_b = E_e + E_o = (E_{ei} + E_{em}) + E_o, \tag{1.20}$$

where $E_{em} = P_{em} T_{\mathrm{lifetime}}$, with P_{em} and T_{lifetime} representing the BS's maintenance power and lifetime, respectively. $E_o = P_o T_{\mathrm{lifetime}}$, where P_o is defined in terms of the BS's operating power described by (1.9)–(1.14). The model in (1.20) is useful in quantifying the BS's total power consumption during the network design stage, for example, while designing a multi-tier wireless network. Also, a similar expression can be derived for the backhaul energy consumption in (1.15)–(1.19), which when added to (1.20) can be used to calculate the overall network energy consumption.

1.3.2.2 Mobile Terminal Side

In the literature, different models have been proposed for the MT's power consumption P_m. In the simplest form, P_m captures only the MT's transmission power P_{tm} [29, 46, 50]. To account for the power amplifier efficiency, the MT's power consumption is expressed as [16, 30, 45, 51]

$$P_m = \frac{P_{tm}}{\xi_m}, \tag{1.21}$$

where ξ_m represents the power amplifier efficiency for MT m, $\xi_m \in (0, 1]$. According to this power consumption model, two conclusions can be drawn in terms of the employed modulation and coding schemes (MCS):

- The minimum energy consumption for a data call is attained by using the modulation of the lowest order while satisfying the QoS constraints (e.g. time delay) [28].
- Adopting M-ary frequency shift keying (MFSK) is more energy efficient than adopting M-ary quadrature amplitude modulation (MQAM), since for a given bit error probability, the SNR per bit requirement increases with M for MQAM while it decreases with M for MFSK [20].

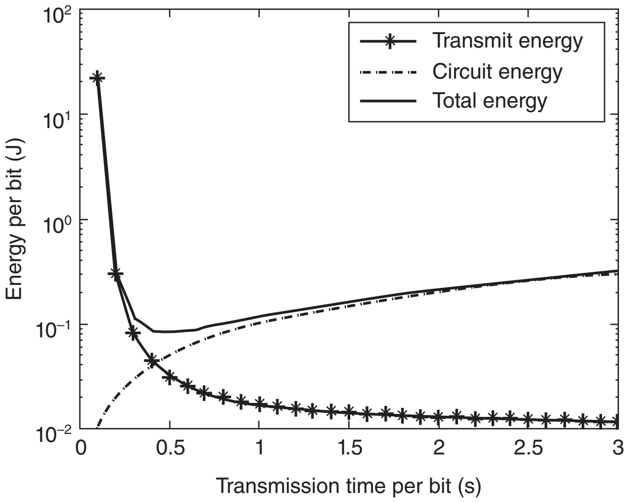


Figure 1.7 MT circuit and transmit energy consumption [56]

However, in practice, the MT circuit's power consumption plays a vital role in the MT's total power consumption, and therefore it should be captured in the power consumption model P_m. Figure 1.7 shows the transmit, circuit, and total energy consumption per bit performance for the MT versus the transmission time per bit [20]. While transmitting using the lowest modulation order, and hence over a long transmission duration, decreases the transmission energy consumption, this is not true for the circuit power consumption. Consequently, the total energy consumption per bit exhibits a minimum value, which corresponds to the optimal MCS. As a result, it is imperative to capture the MT's circuit power consumption in the MT power consumption model. In the literature, three different models have been proposed to represent the MT's circuit power Q_m. The first model assumes that the circuit power consumption is a constant value, independent of the achieved data rate R_m [16, 23, 28, 31, 45, 52]. Such a model changes the aforementioned conclusions regarding the optimal MCS as follows:

- With constant circuit power consumption, adopting the lowest modulation order is no longer the best transmission strategy since energy consumption is directly proportional to the transmission duration [28].
- The optimal MCS is based on the relation between the transmit and circuit power consumption. For long-range applications, the transmission power dominates the circuit power consumption, and hence MFSK is more energy efficient than MQAM, while the opposite is true for short-range applications where the circuit power dominates the total power consumption [20].

One limitation with the constant power consumption model is that it does not reflect the effect of transmission bandwidth and data rate on the MT's circuit power consumption. According to Table 1.4, it is clear that different radio interfaces consume different circuit powers. One reason for such a behaviour is the different operating bandwidths. To account for the effect of the allocated bandwidth, the circuit power consumption presents two terms [53].

Table 1.4 MT power consumption for different technologies [27]

Technology	Action	Power (mw)
WiFi IEEE 802.11 (infrastructure mode)	In connection	868
	In disconnection	135
	Idle	58
	Idle in power save mode	26
	Downloading at 4.5 Mbps	1,450
WiFi IEEE 802.11 (ad hoc mode)	Sending at 700 kbps	1,629
	Receiving	1,375
	Idle	979
2G	Downloading at 44 kbps	500
	Handover to 3G	1,389
3G	Downloading at 1 Mbps	1,400
	Handover to 2G	591

Table 1.5 MT power consumption for different data rates of audio streaming and downloading a 200-MB file using WiFi [27]

Bit Rate (kbps)	Nokia E-71 (mW)	Nexus S (mW)	Samsung Galaxy S3 (mW)
128	990	350	419
192	1,004	390	440
256	1,007	390	452
File download	1,092	998	1,012

The first term represents the digital circuit power consumption, which is modelled as a linear function of the transmission bandwidth (as the bandwidth increases, more computations and baseband processing are required), that is

$$P_{\mathrm{cm}} = P_m^{\mathrm{ref}} + \sigma \frac{B_m}{B_{\mathrm{ref}}}, \tag{1.22}$$

where P_m^{ref} [W] refers to the reference digital circuit power consumption for a reference bandwidth B_{ref}, and σ denotes a proportionality constant. The second term represents the power consumption of the RF chain and accounts for the power consumption in the digital-to-analog converter, RF filter, local oscillator, and mixer. A limitation of the model in (1.22) is that it does not reflect the effect of transmission data rate on the power consumption, which is evident from Table 1.5. To capture the transmission data rate's impact on the circuit power consumption, a linear function of the achieved data rate is assumed for the circuit power consumption following the fact that the clock frequency of the MT's digital chips scales with the achieved data rate [51]. Consequently, the circuit power consumption is expressed as

$$P_{\mathrm{cm}} = \beta_1 + \beta_2 R_m, \tag{1.23}$$

where β_1 and β_2 are two appropriately chosen constants, measured in watts and watt per bit per second, respectively. In addition to the transmission and circuit power modelling in P_m, a constant term is introduced to reflect the MT's receiver circuit power consumption [16, 53].

For orthogonal frequency division multiple access (OFDMA) networks, R_m and P_m are defined as the sum of the corresponding terms over multiple sub-carriers assigned to MT m [16, 28, 29, 45, 46, 50]. Similarly, for an MT m enjoying a multi-homing service, R_m and P_m are defined as the sum of corresponding terms over multiple radio interfaces [23, 31].

1.3.3 Energy Efficiency and Consumption Models

Following the aforementioned throughput and power consumption models, we next present several energy efficiency and consumption terms. A summary of these terms proposed in the literature is given in Table 1.6.

A generic definition that can be used regardless of the traffic load condition is referred to as the *energy consumption gain* (ECG), which is defined as the ratio of the energy consumed by a base system (BS, MT, or entire network) to the energy consumed by the system under test, assuming the same conditions [8, 57]. Formally, this is expressed as follows:

$$\mathrm{ECG} = \frac{E_{\mathrm{base}} - E_{\mathrm{test}}}{E_{\mathrm{base}}}, \tag{1.24}$$

where E_{base} and E_{test} represent the consumed energy measured in joules. The ECG definition in (1.24) is a relative definition that is measured as a percentage. It can be misleading if it is used to compare systems with different characteristics [49].

For access nodes (i.e. BSs and MTs), two definitions can be distinguished. The first definition is referred to as the *energy efficiency index* (EEI), which is defined as the ratio of the attained utility to the consumed energy. On the other hand, the second definition is referred to as the *energy consumption index* (ECI), which represents the reciprocal of the EEI, that is, the ratio of the consumed energy to the attained utility. Overall, both definitions capture the same information; however, they lead to different interpretations [49]. For instance, Figure 1.8 compares the behaviour of an EEI (a) with an ECI (b) [60]. As shown in Figure 1.8a, for the EEI in the low power region, a small improvement in energy saving will lead to a high gain in the EEI; that is, a minor energy saving improvement in a system that is already energy efficient will be interpreted as a high improvement in the achieved energy efficiency. In the medium power region, a high energy saving translates into a small EEI gain, that is, a large energy saving improvement in an energy-inefficient system is interpreted as a small improvement in the achieved energy efficiency. On the other hand, the ECI exhibits a linear relationship between the energy saving improvement and the attained gain in the ECI; that is, a large energy saving improvement for an energy-inefficient system leads to a high gain in the attained ECI. Consequently, for an energy-inefficient system, ECI is more intuitive than EEI. Next, we focus on the EEI and ECI definitions for BSs and MTs subject to different traffic load conditions.

Under low traffic load, it is not required that the BS operates at its full power due to low service demands. Hence, one way to represent the EEI for a given BS at a low call traffic load condition is by means of the ratio between the BS's output power (energy) and the total input power (energy) [2, 22]. That is, the EEI η_s for BS s is expressed as

$$\eta_s = \frac{P_t}{P_s}, \tag{1.25}$$

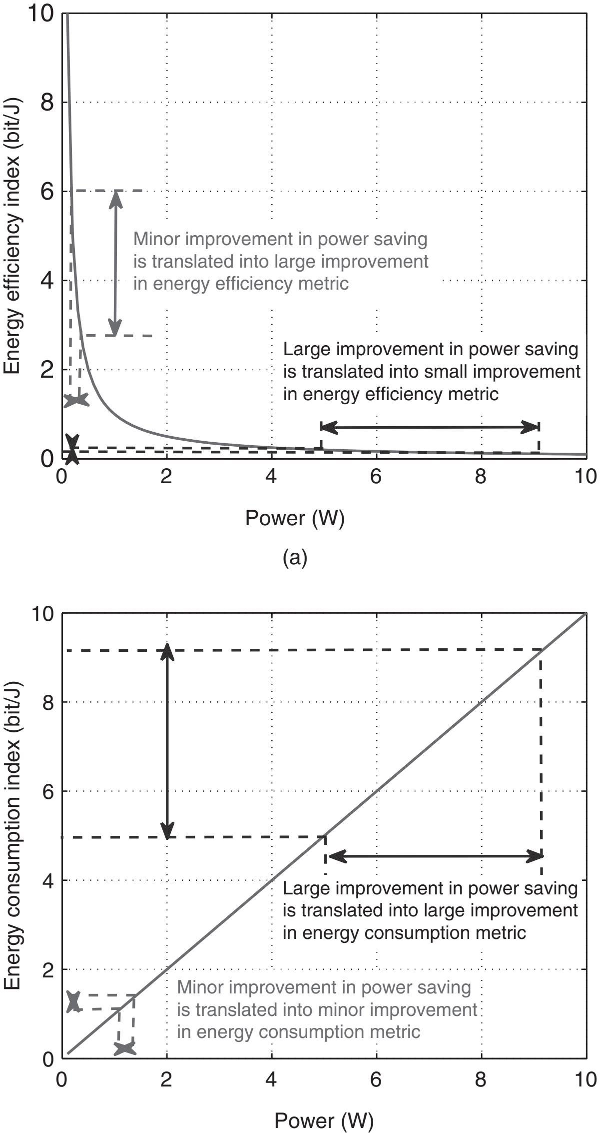
where P_t and P_s are the BS's output power (i.e. the power of the RF transmitted signal) and input (consumed) power, respectively. Therefore, η_s is unitless. In addition, at a low traffic

Table 1.6 Summary of different energy efficiency and consumption definitions proposed in the literature [27]

	Model		Comments	References
BS/MT	Energy consumption gain		A ratio of the energy consumed by a base system to the energy consumed by the system under test. It is a relative measure that can be used at any traffic load	[8, 57]
BS	Low traffic load	Output – input power	A ratio of BS output to input power. It is an EEI	[2, 22]
		Area spectral efficiency – input power	The definition measures the power consumed for a certain area coverage. It is used at a low traffic load. It is an EEI	[18, 22]
	High traffic load Network capacity – input power		A ratio of the aggregate BS capacity to the total power consumed by the BS. It is an EEI	[2, 10]
	Temporal fluctuations (ECRW, TEEER, ECRVL)		It uses a weighted sum of power consumption (and throughput, as in ECRVL) at different traffic load conditions. It is an ECI	[7, 49, 58]
	Absolute ECR		It accounts for the absolute temperature. It is an ECI	[7]
MT	Single-user system	Without error consideration	A ratio of throughput to power consumption. It is an EEI	[16, 23, 31, 45, 50]
		With error consideration	A ratio of goodput to power consumption. It is an EEI	[29, 50]
	Multi-user system	Without fairness consideration	It can be the sum rate of all MTs to total power consumption or sum of energy efficiency for individual MTs. It is an EEI	[28–30, 46, 52]
		With fairness consideration	It is the geometric mean of energy efficiencies of all MTs. It is an EEI	[28]
NW	Traffic load independent	APC	A ratio of total power consumption and network coverage area. It is an ECI	[59]
	Traffic load dependent	Rural definition	A ratio of coverage area to power consumption. It is an EEI	[7]
		Urban definition	A ratio of number of users to the total power consumption. It is an EEI	[7]

condition, it is not necessary for the BS to provide a full coverage. It is sufficient to achieve an acceptable coverage probability. As the definition in (1.25) does not capture the BS's achieved coverage, another definition for energy efficiency is proposed to measure the power consumed to cover a certain area [18, 22]. Consequently, the BS's EEI is defined as [18]

$$\eta_s = \frac{T_s}{P_s}. \tag{1.26}$$



(a)

(b)

Figure 1.8 Comparison of (a) energy efficiency and (b) energy consumption indices [60]

In (1.26), η_s has the unit of watt^{-1}, and T_s denotes the area spectral efficiency given in (1.6). Under high traffic load conditions, the BS's EEI is defined as the ratio of the aggregate BS capacity to the total power consumed by the BS [2, 10]. Hence, under high traffic load conditions, EEI of BS s is expressed in bits per second per watt as

$$\eta_s = \frac{C_s}{P_s}, \tag{1.27}$$

and C_s is given by (1.5). In (1.25)–(1.27), P_s is usually represented by one definition from (1.9)–(1.13). For MTs, EEI is defined as a measure of the maximum number of bits that can be delivered per joule of consumed energy [16, 23, 31, 45, 50]. EEI is expressed for MT m as

$$\eta_m = \frac{R_m \Delta T}{\Delta E_m} = \frac{R_m}{\Delta E_m / \Delta T} = \frac{R_m}{P_m}, \tag{1.28}$$

where ΔE_m denotes the energy consumed during the time interval ΔT by MT m. However, the expression in (1.28) does not consider the energy consumed for the correct reception of data. Another definition measures the net number of information bits that are successfully transmitted without error per joule [29, 50], and it is expressed as

$$\eta_m = \frac{R_m f(\gamma_m)}{P_m}, \tag{1.29}$$

where $f(\gamma_m)$ represents the packet transmission success rate for a given SNR γ_m for MT m. The expression in (1.29) assumes a ratio of the goodput to power consumption, as compared to the expression in (1.28), which assumes the ratio of throughput to power consumption. The packet transmission success rate $f(\gamma_m)$ follows an S-shaped (sigmoidal) function, exhibiting an increasing trend with respect to γ_m, approaching zero as γ_m approaches zero, and approaching unity as γ_m approaches infinity [29, 50]. The unit of η_m in (1.28) and (1.29) is bits per second per watt. The definitions in (1.28) and (1.29) are proposed for a single-user scenario [16, 23, 31, 45, 50]. In practice, a multi-user system is considered due to the competition over bandwidth [28] and the impact of interference [46] caused by simultaneous transmissions. In a multi-user system, EEI is defined as the ratio between the sum rate of all MTs to the total power consumption [30]

$$\eta_{\text{total}} = \frac{\sum_m R_m}{\sum_m P_m}. \tag{1.30}$$

The definition in (1.30) treats all MTs as a single unit, and takes into account only the total achieved throughput and power consumption. In order to model the system as a set of distinct MTs, an alternative definition is used, which represents the total EEI as the sum of the energy efficiency for each individual MT [28, 29, 46, 52]

$$\eta_{\text{total}} = \sum_m \eta_m. \tag{1.31}$$

The unit of η_{total} in (1.30) and (1.31) is bits per second per watt. However, the definitions in (1.30) and (1.31) do not ensure energy efficiency fairness among different MTs. Therefore, some MTs might exhibit high energy efficiencies while others might present low energy efficiencies very close to zero. To promote fairness among MTs, the geometric mean of energy efficiencies of all MTs is used [28]

$$\eta_{\text{total}} = \sum_m \log(\eta_m). \tag{1.32}$$

Unlike (1.30) and (1.31), η_{total} in (1.32) has a unit of $\log(\text{bps/W})$. In (1.28)–(1.32), R_m is described using (1.7) or (1.8), and P_m is described using (1.21)–(1.23).

While the definitions (1.25)–(1.32) represent the EEI for BSs and MTs at different traffic load conditions, the following definitions are used to represent the ECI. For BSs and MTs, the energy consumption rating (ECR) is defined as the ratio of the power consumption to the achieved capacity [7, 49, 58]. Hence, for BSs, it is the reciprocal of the definition in (1.27), while for MTs, it is the reciprocal of the definitions in (1.28) or (1.29). For BSs, to account for the temporal fluctuation in traffic load, a weighted ECR definition is introduced (ECRW), which assumes the reciprocal of (1.27) while using the weighted sum of power consumption at different traffic load conditions (full load, half load, and idle conditions) to represent P_s as in (1.14) [7, 49, 58]. The telecommunications equipment energy efficiency ratio (TEEER) is calculated as a logarithmic function of ECRW. The ECRW assumes a weighted sum of only power consumption; yet a variable load ECR (ECRVL) definition follows the same expression as ECRW and involves also a weighted sum of the achieved throughput at different traffic loads, using similar weights as in (1.14) [7, 49, 58]. Finally, an absolute ECR definition that accounts for the absolute temperature of the medium, T, is given by Hasan et al. [7]

$$\tilde{\eta}_s = 10 \log \left(\frac{P_s/C_s}{kT \ln\ (2)} \right), \tag{1.33}$$

where k stands for the Boltzmann constant. According to [7], including the temperature in the analysis follows from the classical thermodynamics theory.

All the aforementioned definitions in (1.25)–(1.33) target access nodes such as BSs and MTs. A definition proposed in the literature for ECI at network level is referred to as the *area power consumption* (APC) [59], which is expressed as the ratio of the total power consumption to the network coverage area, and is measured in W/km^2. The APC is further extended to include the ratio of the total power consumption to achieve a given throughput in a given area, and it is expressed in W/Gbps/km^2. However, for such a metric to be valid in the comparison of different networks, it must be applied to networks with a similar number of sites in a given area [7]. Two EEI definitions can be distinguished at the network level based on the traffic load conditions. The first definition is for rural areas, that is, in a low traffic load condition, and it assumes the expression [7]

$$\eta_n = \frac{\text{Total coverage area}}{\text{Total power consumption at the site}}\ [\text{km}^2/\text{W}]. \tag{1.34}$$

The rationale behind such a definition is that, under low traffic load condition, the main objective is to reduce the total power consumption to cover a specific region. On the other hand, for urban areas with high traffic load, the objective is to reduce the power consumption to achieve a given capacity, and hence the energy efficiency is expressed as [7]

$$\eta_n = \frac{M_{\text{busy hour}}}{\text{Total power consumption at the site}}\ [\text{users/W}], \tag{1.35}$$

where $M_{\text{busy hour}}$ stands for the number of users in an average busy hour traffic.

1.4 Performance Trade-Offs

Improving energy efficiency of wireless networks is achieved at the cost of some performance degradation. Usually, a threshold level is specified for some target (acceptable) QoS. Green solutions aim to achieve the maximum energy saving while satisfying the QoS threshold. Overall, the performance trade-offs can be divided into two main categories, namely at the network and at the mobile user side, respectively. These trade-offs are discussed next.

1.4.1 Network-side Trade-Offs

The two performance metrics, namely the spectral efficiency and network coverage, conflict with the energy efficiency from a network operator perspective. Both metrics directly affect the network operator's investments, as they are related to the network's available resources in terms of bandwidth (for spectral efficiency) and number and types of deployed BSs (for network coverage).

1.4.1.1 Spectral Efficiency

By definition, spectral efficiency quantifies the system throughput per unit of bandwidth. Such a metric is a key performance indicator for the third-generation partnership project (3GPP) and reflects how efficiently the network bandwidth is utilized in the uplink and downlink. However, the energy efficiency (in the uplink, i.e., from the user perspective, and the downlink, i.e., from the network operator perspective) and spectral efficiency conflict with each other. This conflict is a direct consequence of the relation between bandwidth and power, as illustrated by Shannon's formula in (1.5) for the BSs and in (1.7) for the MTs, respectively. For instance, from Shannon's formula, the relationship between the transmission power and the allocated bandwidth for a given transmission rate is given by

$$P = BN_0 2^{\frac{R}{B}-1}. \tag{1.36}$$

In (1.36), for uplink communications, P stands for the MT's transmission power and B denotes the allocated bandwidth by the BS on the uplink, while for downlink communications, P represents the BS's transmission power and B denotes the allocated bandwidth by the BS on the downlink. The expression in (1.36) is shown in the top left sub-plot of Figure 1.9, which indicates a monotonic relation between the transmission power and the allocated bandwidth. From the top left sub-plot of Figure 1.9, it turns out that, for a given transmission rate R, in order to save transmission power and hence improve the resulting energy efficiency (for the BS on the downlink or for the MT on the uplink), a large transmission bandwidth should be used. In turn, this will reduce the achieved spectral efficiency. Using the energy efficiency definition in (1.27) for BSs or in (1.28) for MTs (and defining P_s or P_m using only the transmission power definitions in (1.9) or (1.21), respectively), the energy efficiency–spectral efficiency relation is expressed as [9]

$$\eta_{\mathrm{EE}} = \frac{\eta_{\mathrm{SE}}}{(2^{\eta_{\mathrm{SE}}} - 1)N_0}. \tag{1.37}$$

The relation in (1.37) is shown in the top right sub-plot of Figure 1.9, and it exhibits a similar performance to the one depicted in the top left sub-plot of Figure 1.9. Hence, η_{EE} converges to the minimum value of $1/(N_0 \ln\ 2)$ when $\eta_{\mathrm{SE}} = 0$, and $\eta_{\mathrm{EE}} = 0$ when $\eta_{\mathrm{SE}} = \infty$.

However, it should be noted that the expressions in (1.36) and (1.37) account only for the transmission power and do not consider the circuit power component as in (1.10)–(1.14) for the BS and in (1.22) or as in (1.23) for the MT. Accounting for such a circuit power consumption component yields a relationship that exhibits a minimum value, as shown in the bottom left sub-plot of Figure 1.9 for the power–bandwidth relationship and a maximum value in the bottom right sub-plot of Figure 1.9 for the energy efficiency–spectral efficiency relation. From (1.36), it turns out that an infinite bandwidth allocation leads to the minimum

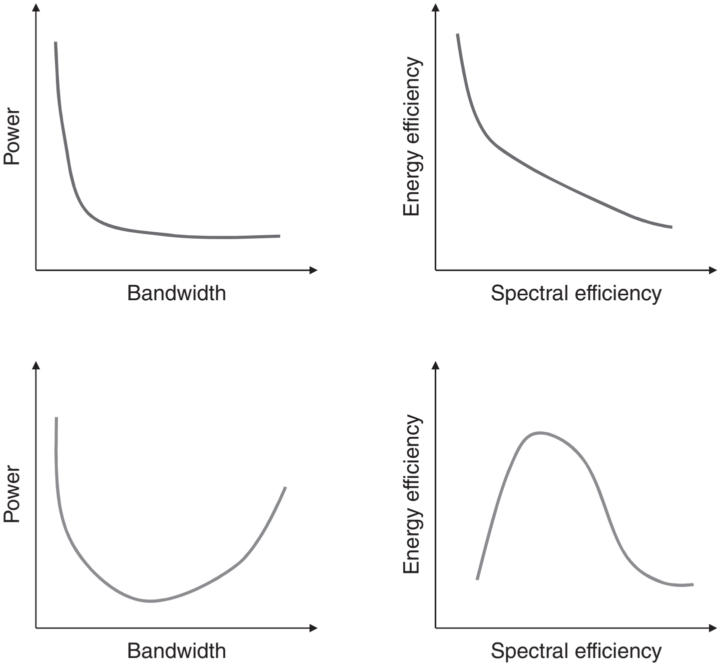


Figure 1.9 Performance trade-offs [9]

power consumption level $N_0 R \ln\ 2$. However, this is not true for the practical scenario, which involves the device (BS or MT) circuit power consumption. In this case, the green solutions aim to work on the minimum/maximum values in the bottom sub-plots of Figure 1.9 to reach a good power–bandwidth trade-off, and hence a good energy efficiency–spectral efficiency compromise.

1.4.1.2 Network Coverage

Another important metric that conflicts with energy efficiency is the network coverage. High network coverage performance can be achieved in two ways, namely cell stretching and small-cell deployment. The first approach relies on a small number of BSs to cover a large area by stretching the cell coverage as much as possible [9]. While such an approach can reduce the capital expenditure (CAPEX), it leads to high BS transmission power to support MTs at the cell edge. Specifically, it has been shown that for a path loss exponent of 4, the path loss between the BS and cell edge MTs will be reduced by 12 dB if the cell radius is doubled [9]. Consequently, this leads to a 12 dB increase in the BS transmit power to satisfy the target QoS. When only transmission power is considered, energy efficiency scales (degrades) continuously and proportionally with the cell radius. However, accounting for the BS circuit power consumption as in (1.10)–(1.14) leads to a more complex relationship between the cell radius and the achieved energy efficiency. Therefore, green solutions aim to determine the optimal energy efficiency–network coverage (BS cell radius) compromise. The second approach that is adopted to achieve a high network coverage performance relies on small-cell deployment. The main advantage of such an approach is that a low BS transmission power is expected in this case due to the short distance between MTs and the BS. However, it should be noted that using a large number of small-cells might not eventually lead to an improvement in energy efficiency. This is due not only to the BS circuit power consumption but also to the BS embodied energy (1.20). Consequently, green solutions aim to balance the energy efficiency with the network coverage (in terms of specifying the optimal number of small-cells).

Furthermore, the energy efficiency–network coverage trade-off plays a vital role during low traffic load conditions. Specifically, the energy efficiency of wireless networks can be improved by switching off some BSs at a low call traffic load, as will be explained in the next chapter. However, this can result in an increased call blocking probability. Therefore, green solutions aim to achieve energy saving while maintaining the call blocking probability below a certain threshold [12, 26]. In some cases, BSs do not need to be completely switched off, but rather they are allowed to shrink their coverage area by reducing their transmission power, a technique referred to as *cell zooming* [27]. However, this technique may lead to failures in service coverage. Consequently, green solutions aim to enhance the network energy efficiency while maintaining a target performance level in terms of coverage probability $\mathbb{P}_n(\zeta)$.

1.4.2 Mobile User Trade-Offs

The main performance metric conflicting with the energy efficiency from the mobile users' perspective is related to the quality of the ongoing application. Different equivalent measures exist to quantify the quality of the ongoing application, including SNR, data rate, delay (latency), and video quality.

For instance, consider SNR as a QoS indicator [36, 37]. Using the energy efficiency definition in (1.28) for MTs (and defining P_m using only the transmission power definition as in (1.21)), an inverse relationship exists between energy efficiency and SNR, as shown in Figure 1.10a. Specifically, achieving a high SNR requires a high transmission power, which in turn leads to low energy efficiency. However, when the circuit power is accounted for in the MT's total power consumption, as expressed in (1.23), a different relation can be observed, as shown in Figure 1.10b. In this case, two regions can be distinguished in the energy efficiency–SNR relationship. In the first region, the circuit power consumption dominates the MT total power consumption. Consequently, increasing the transmission power (and hence the SNR), will lead to an increased data rate (and hence lower transmission delay). In turn, this will reduce the circuit energy consumption and, as a result, improve the overall energy efficiency. On the other hand, in the second region, the transmission power dominates the MT total power consumption. Consequently, increasing the transmission power (and hence SNR) will lead to high energy consumption and hence reduced energy efficiency (similar to the performance in Figure 1.10a). In practice, the receiver requires a target SNR in order to be able to decode the transmitted signal. Green solutions aim to find the optimal point that balances the achieved energy efficiency–target SNR compromise. A similar argument holds for the BS energy efficiency in the downlink while considering a target SNR threshold.

The same arguments stand for the data rate when it is considered as a QoS metric. When circuit power is dominating the total power consumption, it is more energy efficient to transmit with high power, and hence achieve high data rates. However, as transmission dominates the total power consumption, the high data rate requirement results in reduced energy efficiency. For some applications, a minimum required data rate should be achieved

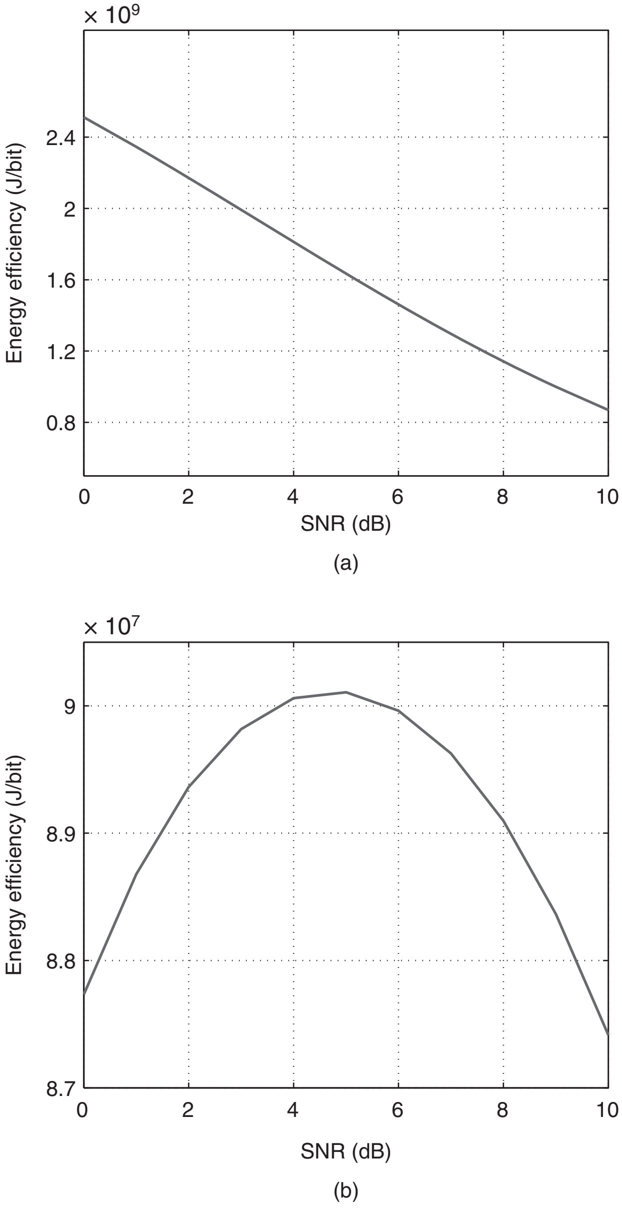


Figure 1.10 Energy efficiency versus SNR (a) with and (b) without MT circuit power consumption

[31, 36, 37] or a constant required data rate should be satisfied [23, 33]. Therefore, green solutions aim to balance the achieved energy efficiency–target data rate compromise.

An equivalent representation to ensure a minimum required data rate is not to violate a maximum delay bound for data transmission. Delay bound has been used as a QoS indicator in [44] for data calls. Similarly, for video streaming applications, in order to maintain a high video quality, video packets should be transmitted before a given delay deadline as in [61, 62]. Stringent delay (latency) requirement calls for high transmission power, which affects the energy efficiency based on which part (circuit power or transmission power) dominates the total power consumption. Green solutions aim to balance the achieved compromise between energy efficiency and target delay bound (video quality).

1.5 Summary

In order to develop and analyse a green network solution, an appropriate definition of energy efficiency and consumption for network operators and mobile users should be adopted. Such a definition is based on the traffic load condition, power consumption, and throughput for network operators and mobile users. In addition, the green network solution should satisfy some target (and possibly conflicting) performance metrics. Therefore, this chapter was dedicated to energy efficiency and consumption definitions, as well as power consumption, throughput, and traffic load models for network operators and mobile users, along with conflicting performance metrics.

After having introduced the necessary background concepts in this chapter, the next chapter will focus on state-of-the-art green network solutions and projects along with the analytical models employed by network operators and mobile users at different traffic load conditions.