1

Fundamentals

1.1   4G and the Book Layout

The research community and industry in the field of telecommunications are considering continuously the possible evolution of wireless communications. This evolution is closely related to the future concept of the Internet. With the advances in multihop cellular networks (relaying) and the integrated elements of ad hoc and cellular networks the border between the Internet and wireless networks is disappearing rapidly. Instead of having *wireless access to Internet* we will see the extension of the Internet over wireless networks resulting in a *Wireless Internet*. For this reason an understanding of the future trends in the evolution of the Internet is necessary so as to be able to plan the necessary development of wireless networks in order to enable the closer integration of the two systems. This chapter will start with a generic 4G system concept that integrates the available advanced wireless technologies and will then focus on system adaptability and reconfigurability as a possible option to meet a variety of service requirements, available resources and channel conditions. The elements of such a concept can be found in [1–51]. This presentation will also try to offer a vision beyond the state of the art with an emphasis on how advanced technologies can be used for efficient 4G multiple access. The second part of the chapter will discuss the future evolution of the Internet, especially the concepts of the resource clouds and smart grids. Amongst a number of relevant issues the focus will be on:

- adaptive and reconfigurable coding and modulation including distributed source coding which is of interest for data aggregation in wireless sensor networks;
- adaptive and reconfigurable space time coding including a variety of turbo receivers;
- channel estimation and equalization and multiuser detection;
- orthogonal Frequency Division Multiple Access (OFDMA), Multicarrier CDMA (MC CDMA) and Ultra Wide Band (UWB) radio;
- antenna array signal processing;
- convex optimization based linear precoding for MIMO systems;
- channel sensing for cognitive radio;
- cooperative transmit diversity and relaying;
- biologically inspired paradigms in wireless networks;
- user location in 4G;
- reliability and redundancy design in communication networks;
- cross-layer optimization including adaptive and power efficient MAC layer design, adaptive and power efficient routing on IP and TCP layer including network coding and concept of green wireless network;
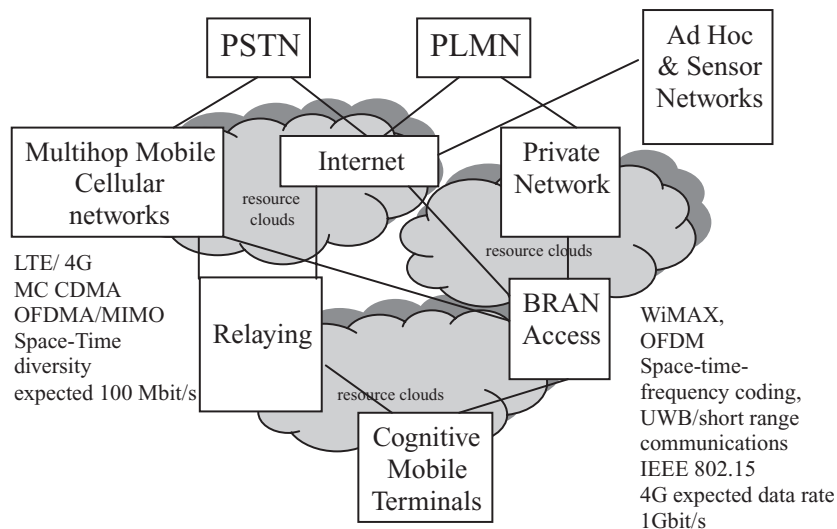- cognitive networks modeling based on game theory.

Figure 1.1   Cellular networks and WLAN convergence.

An important aspect of wireless system design is power consumption. This will be also incorporated in the optimization process of most of the problems considered throughout the book.

At this stage of the evolution of wireless communications there is a tendency towards even closer integration of mobile communications as specified by the International Mobile Telecommunications (IMT) standards and Wireless Local Area Networks (WLAN) or in the general Broadband Radio Access Networks (BRAN) specified by IEEE 802.xx. The core network will be based on Public Switched Telecommunications Network (PSTN) and Public Land Mobile Networks (PLMN) based on Internet Protocol (IP) [13, 16, 19, 24, 32, 41, 51]. This concept is summarized in Figure 1.1. Each of the segments of the system will be further enhanced in the future. The inter-technology roaming of the mobile terminal will be based on a reconfigurable cognitive radio concept presented in its generic form in Figure 1.2.

The material in this book is organized as follows:

Chapter 1 starts with the general structure of 4G signals, mainly Advanced Time Division Multiple Access – ATDMA, Code Division Multiple Access – CDMA, Orthogonal Frequency Division Multiplexing – OFDM, Multicarrier CDMA (MC CDMA) and Ultra Wide Band (UWB) signal. These
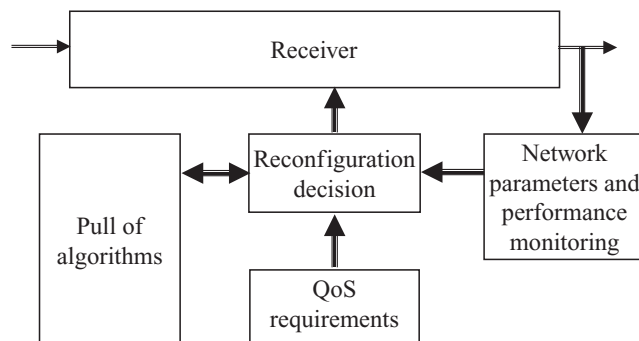

Figure 1.2   Reconfigurable cognitive radio concept intersystem roaming & QoS provisioning.

signals will be elaborated upon later in the book in more detail. In the second part of the chapter we discuss the future evolution of the Internet, especially the concepts of resource clouds and smart grids.

Chapter 2 introduces adaptive coding. The book is not intended to cover all the details of coding but rather to focus on those components that enable code adaptability and reconfigurability. Within this concept the chapter covers: adaptive and reconfigurable block and convolutional codes, punctured convolutional codes/code reconfigurability, maximum likelihood decoding/Viterbi algorithm, systematic recursive convolutional code, concatenated codes with interleaver, the iterative (turbo) decoding algorithm and a discussion on adaptive coding practice and prospects. The chapter also includes a presentation of distributed source coding which is of interest in data aggregation in wireless sensor networks.

Chapter 3 covers adaptive and reconfigurable modulation. This includes coded modulation, Trellis Coded Modulation (TCM) with examples of TCM schemes such as two, four and eight state trellis and QAM with 3 bits per symbol transmission. The chapter also discusses signal set partitioning, equivalent representation of TCM, TCM with multidimensional constellation, adaptive coded modulation for fading channels and adaptation to maintain a fixed distance in the constellation.

Chapter 4 introduces Space Time Coding. It starts with a discussion on diversity gain, the encoding and transmission sequence, the combining scheme and ML decision rule for a two-branch transmit diversity scheme with one and M receivers. Next, it introduces a general discussion on *space time coding* within a concept of space time trellis modulation. The discussion is then extended to introduce space-time block codes from *orthogonal design*, mainly linear processing orthogonal designs and generalized real orthogonal designs. The chapter also covers channel estimation imperfections. It continuous with quasy orthogonal space time block codes, space time convolutional codes and algebraic space time codes. It also includes differential space-time modulation with a number of examples.

Layered space – time coding and *concatenated* space time block coding are also discussed. Estimation of MIMO channel and space-time codes for frequency selective channels are discussed in detail. MIMO system optimization including gain optimization by singular value decomposition (svd) is also discussed. This chapter is extended to include a variety of turbo receivers.

Chapter 5 introduces multiuser detection starting with CDMA receivers and signal subspace-based channel estimation. It then extends this approach to iterative space time receivers. In Chapter 7 this approach is extended to OFDM receivers.

Chapter 6 deals with equalization, detection in a statistically known time-varying channel, adaptive MLSE equalization, adaptive joint channel identification and data demodulation, turbo-equalization Kalman filer based joint channel estimation and equalization using higher order signal statistics.

Chapter 7 covers orthogonal frequency division multiplexing (OFDM) and MC CDMA. The following topics are discussed: Timing and frequency offset in OFDM, fading channel estimation for OFDM systems, 64-DAPSK and 64-QAM modulated OFDM signals, space time coding with OFDM signals, layered space time coding for MIMO-OFDM, space time coded TDMA/OFDM reconfiguration efficiency, multicarrier CDMA system, multicarrier DS-CDMA broadcast systems, frame by frame adaptive rate coded multicarrier DS-CDMA system, intermodulation interference suppression in multicarrier DS-CDMA systems, successive interference cancellation in multicarrier DS-CDMA systems, MMSE detection of multicarrier CDMA, multiuser receiver for space-time coded multicarrier CDMA systems and peak to average power ratio (PAPR) problem mitigation.

Chapter 8 introduces *Ultra Wide Band Radio*. It covers topics such as: UWB multiple access in Gaussian channel, the UWB channel, UWB system with M-ary modulation, M-ary PPM UWB multiple access, coded UWB schemes, multiuser detection in UWB radio, UWB with space time processing and beamforming for UWB radio.

Chapter 9 covers linear precoding for MIMO Channels. This includes space–time precoders and equalizers for MIMO channels, linear precoding based on convex optimization theory and convex optimization-theory-based beamforming.

Chapter 10 discusses issues related to channel sensing for cognitive radio including optimal channel sensing in cognitive wireless networks, optimal sequential, parallel multiband channel and collaborative spectrum sensing and multichannel cognitive MAC.

Chapter 11: Introduces cooperative transmit diversity as a power efficient technology to increase the coverage in multihop wireless networks. It is expected that elements of this approach will be used in 4G cellular systems also, especially relaying which represents a simple case of this approach.

Chapter 12 covers a biologically inspired model for securing hybrid mobile ad hoc networks, biologically inspired routing in ad hoc networks, swarm intelligence based routing, analytical modeling of antnet as adaptive mobile agent based routing, biologically inspired algorithm for optimum multicasting, ant colony system (ACS) model, biologically inspired distributed topology control, optimization of mobile agent routing in sensor networks, epidemic routing, nanonetworks and genetic algorithm based dynamic topology reconfiguration in cellular multihop wireless networks.

Chapter 13 is modified significantly to include more detail on positioning. This is the result of a prediction that this technique will gain an increasing role in advanced wireless communications. This is also supported by activities within the Galileo program in Europe.

Chapter 14 covers survivable wireless networks design, survivability of wireless ad hoc networks, network dimensioning, genetic algorithm based network redundancy design, integer programming method, simulated annealing and survivable network design under general traffic.

Chapter 15 discusses conventional routing versus network coding, a max-flow min-cut theorem, algebraic formulation of network coding, random network coding, gossip based protocol and network coding, network coding with reduced complexity, multisource multicast network switching, conventional route packing problem, multicast network switching as matrix game, computation of maximum achievable information rate for single-source multicast network switching, optimization of wireless multicast ad-hoc networks, matrix game formulation of joint routing and scheduling, interference controlled scheduling, extended fictitious playing and dominancy theory for *AMG* games, optimization of multicast wireless ad-hoc network using soft graph coloring and non-linear cubic games, matrix game modeling for optimum scheduling, cubic matrix game modeling for joint optimum routing, network coding and scheduling and joint optimization of routing and medium contention in multihop multicast wireless network.

Chapter 16 covers network formation games including stability and efficiency of the game, traffic routing utility model, network formation game dynamics, knowledge based network formation games, network formation as a non-cooperative game, network formation as a cooperative game, dynamic network formation and topology control, greedy utility maximization, non-cooperative link utility maximization, cooperative link utility maximization via utility transfer, preferential attachment with knowledge of component sizes, preferential attachment with knowledge of neighbor degrees, link addition/deletion algorithm, coalition games in wireless ad hoc networks, stochastic model of coalition games for spectrum sharing in large scale wireless ad hoc networks and modelling coalition game dynamics.

The evolution of common air interface in wireless communications can be presented in general as in Table 1.1. The coding and modulation for 4G air interface are more or less defined. These problems are addressed in Chapters 1–9. The work on a new multiple access scheme still remains to be elaborated. In this segment the new generation of wireless networks will be significantly different from the solutions seen so far. A part of the solution is what we refer to as *intercell interference coordination (IIC) in MAC layer (IIC MAC)*, as a new multiple access scheme for 4G systems.

In multihop networks this problem will be addressed through different forms of joint optimization of scheduling, routing and relaying topology control. This is addressed in Chapters 10–16. Most of the material in Chapters 10–16 of the third edition of the book is new as compared with the previous edition.

Table 1.1   Evolution of common air interface in mobile communications

generation	2G	3G	4G
coding	convolutional coding	turbo coding	turbo coding & space time coding
modulation	GMSK, BPSK, QPSK	BPSK, QPSK, mQAM	NmQAM (OFDM)
multiple access	TDMA	CDMA	IIC MAC & joint optimization of scheduling/routing & topology control (new Chapters 10–16)

1.2   General Structure of 4G Signals

In this section we will summarize the signal formats used in the existing wireless systems and point out possible ways of evolution towards the 4G system. The focus will be on OFDMA, MC CDMA and UWB signals.

1.2.1   Advanced Time Division Multiple Access – ATDMA

In a TDMA system each user is using a dedicated time slot within a TDMA frame as in GSM (Global System of Mobile Communications) or in ADC (American Digital Cellular System). Additional data about the signal format and system capacity are given in [54]. The evolution of ADC system resulted in the TIA (Telecommunications Industry Association) Universal Wireless Communications (UWC) standard 136 [54]. The evolution of GSM resulted into a system known as Enhanced Data rates for GSM Evolution (EDGE) with parameters that are summarized in [54].

1.2.2   Code Division Multiple Access – CDMA

CDMA technique is based on spreading the spectra of the relatively narrow information signal S^n by a code c, generated by much higher clock (chip) rate. Different users are separated by using different uncorrelated codes. As an example the narrowband signal in this case can be a PSK signal of the form

$$S_n = b(t, T_m) \cos \omega t \tag{1.1}$$

where $1/T_m$ is the bit rate and $b = \pm 1$ is the information. The baseband equivalent of (1.1) is

$$S_n^b = b(t, T_m) \tag{1.1a}$$

Spreading operation, presented symbolically by operator $\varepsilon(\ )$, is obtained if we multiply narrowband signal by a pseudo noise (PN) sequence (code) $c(t, T_c) = \pm 1$. The bits of the sequence are called chips and the chip rate $1/T_c \gg 1/T_m$. The wideband signal can be represented as

$$S_w = \varepsilon(S_n) = cS_n = c(t, T_c)b(t, T_m) \cos \omega t \tag{1.2}$$

The baseband equivalent of (1.2) is

$$S_w^b = c(t, T_c)b(t, T_m) \tag{1.2a}$$

Despreading, represented by operator $D(\ )$, is performed if we use $\varepsilon(\ )$ once again and bandpass filtering, with the bandwidth proportional to $2/T_m$, represented by operator BPF( ) resulting into

$$D(S_w) = \text{BPF}(\varepsilon(S_w)) = \text{BPF}(cc\ b\ \cos \omega t) = \text{BPF}(c2\ b\ \cos \omega t) = b\ \cos \omega t \tag{1.3}$$

The baseband equivalent of (1.3) is

$$D(S_w^b) = \text{LPF}(\varepsilon(S_w^b)) = \text{LPF}(c(t, T_c)c(t, T_c)b(t, T_m)) = \text{LPF}(b(t, T_m) = b(t, T_m) \tag{1.3a}$$

where LPF( ) stands for low pass filtering. This approximates the operation of correlating the input signal with the locally generated replica of the code $\text{Cor}(c, S_w)$. Nonsynchronized despreading would result in

$$D_\tau(\ ); \quad \text{Cor}(c_\tau, S_w) = \text{BPF}(\varepsilon_\tau(S_w)) = \text{BPF}(c_\tau\ c\ b\ \cos \omega t) = \rho(\tau)\ b\ \cos \omega t \tag{1.4}$$

In (1.4) BPF would average out the signal envelope $c_\tau c$ resulting in $E(c_\tau c) = \rho(\tau)$. The baseband equivalent of (1.4) is

$$D_\tau(); \quad \mathrm{Cor}\left(c_\tau, S_w^b\right) = \int_0^{T_m} c_\tau S_w^b dt = b(t, T_m) \int_0^{T_m} c_\tau c\, dt = b\rho(\tau) \tag{1.4a}$$

This operation would extract the useful signal b as long as $\tau \cong 0$ otherwise the signal will be suppressed because, $\rho(\tau) \cong 0$ for $\tau \geq T_c$. Separation of multipath components in a RAKE receiver is based on this effect. In other words if the received signal consists of two delayed replicas of the form

$$r = S_w^b(t) + S_w^b(t - \tau)$$

the despreading process defined by (1.4a) would result into

$$D_\tau(); \quad \mathrm{Cor}(c, r) = \int_0^{T_m} cr\, dt = b(t, T_m) \int_0^{T_m} c(c + c_\tau) dt = b\rho(0) + b\rho(\tau)$$

Now, if $\rho(\tau) \cong 0$ for $\tau \geq T_c$ all multipath component reaching the receiver with a delay larger than the chip interval will be suppressed.

If the signal transmitted by user y is despread in receiver x the result is

$$D_{xy}(); \quad \mathrm{BPF}(\varepsilon_{xy}(S_w)) = \mathrm{BPF}(c_x\, c_y\, b_y \cos \omega t) = \rho_{xy}(t) b_y \cos \omega t \tag{1.5}$$

So in order to suppress the signals belonging to other users (Multiple Access Interference – MAI), the crosscorrelation functions should be low. In other words if the received signal consists of the useful signal plus the interfering signal from the other user

$$r = S_{wx}^b(t) + S_{wy}^b(t) = b_x c_x + b_y c_y \tag{1.6}$$

despreading process at receiver of user x would produce

$$D_{xy}(); \quad \mathrm{Cor}(c_x, r) = \int_0^{T_m} c_x r\, dt = b_x \int_0^{T_m} c_x c_x dt + b_y \int_0^{T_m} c_x c_y dt = b_x \rho_x(0) + b_y \rho_{xy}(0) \tag{1.7}$$

When the system is synchronized properly $\rho_x(0) \cong 1$, and if $\rho_{xy}(0) \cong 0$ the second component representing MAI will be suppressed. This simple principle is elaborated in WCDMA standard resulting in a collection of transport and control channels. The system is based on 3.84 Mcips rate and up to 2 Mbits/s data rate. In a special downlink high data rate shared channel the data rate and signal format are adaptive. There shall be mandatory support for QPSK and 16 QAM and optional support for 64 QAM based on UE capability which will proportionally increase the data rate. For details see www.3gpp.com.

1.2.3 Orthogonal Frequency Division Multiplexing – OFDM

In wireless communications, the channel imposes the limit on data rates in the system. One way to increase the overall data rate is to split the data stream into a number of parallel channels and use
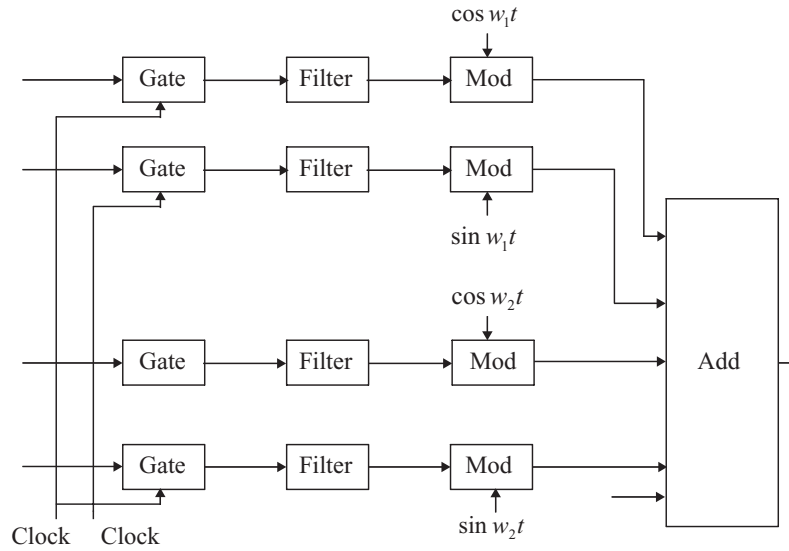
Figure 1.3   An early version of OFDM.

different subcarriers for each channel. The concept is presented in Figures 1.3 and 1.4 and represents the basic idea of OFDM system. The overall signal can be represented as

$$x(t) = \sum_{n=0}^{N-1}\left\{ D_n e^{j2\pi \frac{n}{N} f_S t}\right\}; \qquad -\frac{k_1}{f_s} < t < \frac{N+k_2}{f_s} \tag{1.8}$$

In other words complex data symbols $[D_0, D_1, \ldots, D_{N-1}]$ are mapped in OFDM symbols $[d_0, d_1, \ldots, d_{N-1}]$ such that

$$d_k = \sum_{n=0}^{N-1} D_n e^{j2\pi \frac{kn}{N}} \tag{1.9}$$
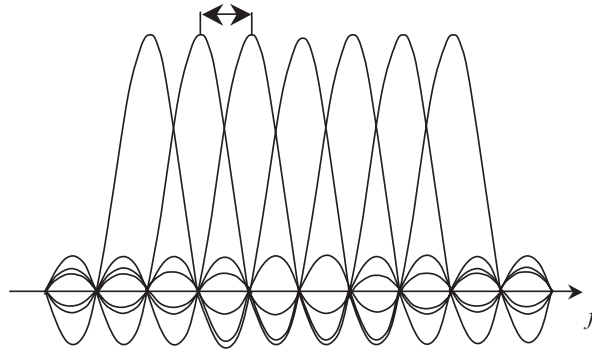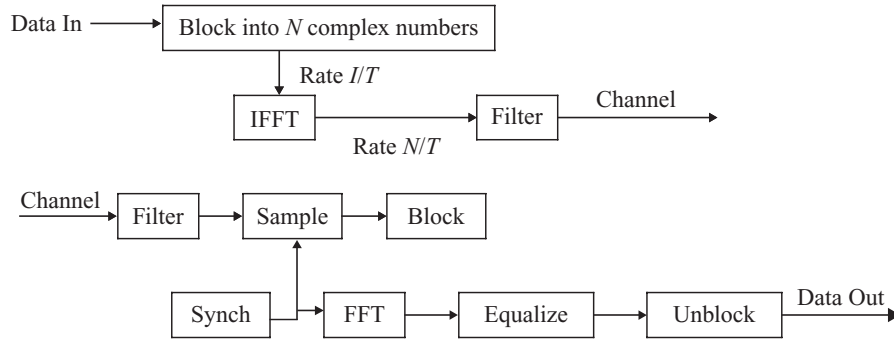

Figure 1.4   Spectrum overlap in OFDM.

Figure 1.5   Basic OFDM system.

The output of the FFT block at the receiver produces data per channel. This can be represented as

$$\tilde{D}_m = \frac{1}{N} \sum_{k=0}^{N-1} r_k e^{-j2\pi m \frac{k}{2N}}$$

$$r_k = \sum_{n=0}^{N-1} H_n D_n e^{j2\pi \frac{n}{2N}k} + n(k) \tag{1.10}$$

$$\tilde{D}_m = \begin{cases} H_n D_n + N(n), n = m \\ N(n), n \neq m \end{cases}$$

The system block diagram is given in Figure 1.5.

In order to eliminate residual intersymbol interference a guard interval after each symbol is used as shown in Figure 1.6.
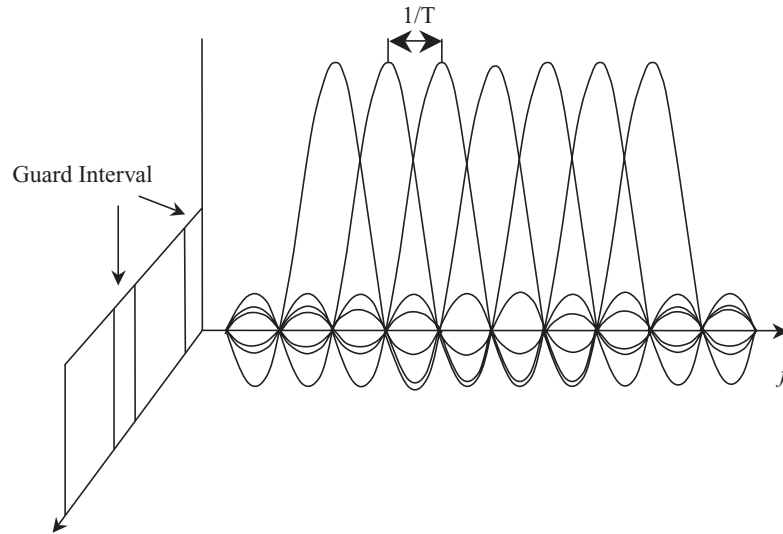

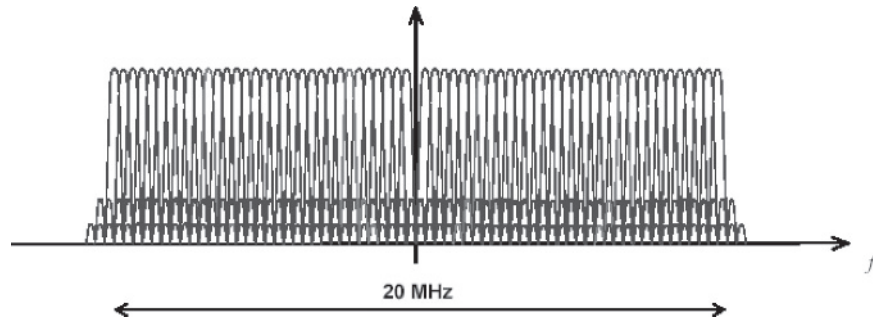**Figure 1.6**   OFDM time and frequency span.

Figure 1.7 802.11a/HIPERLAN OFDM.

An example of an OFDM signal specified by the IEEE 802.11a standard is shown in Figure 1.7. The signal parameters are: 64 points FFT, 48 data subcarriers, 4 pilots, 12 virtual subcarriers, DC component 0, Guard interval 800 ns. A discussion on OFDM and an extensive list of references on the topic are included in Chapter 7.

1.2.4 Multicarrier CDMA (MC CDMA)

Good performance and the flexibility to accommodate multimedia traffic are incorporated in MC CDMA which is obtained by combining CDMA and OFDM signal formats.

Figure 1.8 shows the DS-CDMA transmitter of the j-th user for binary phase shift keying/coherent detection (CBPSK) scheme and the power spectrum of the transmitted signal, respectively, where $G_{DS} = T_m/T_c$ denotes the processing gain and $C^j(t) = [C_1^j C_2^j \cdots C_{G_{DS}}^j]$ the spreading code of the j-th user.

Figure 1.9 shows the MC-CDMA transmitter of the j-th user for CBPSK scheme and the power spectrum of the transmitted signal, respectively, where G_{MC}, denotes the processing gain, N_C the number of subcarriers, and $C^j(t) = [C_1^j C_2^j \cdots C_{G_{MC}}^j]$ the spreading code of the j-th user. The MC-CDMA scheme is discussed assuming that the number of subcarriers and the processing gain are all the same.

However, we do not have to choose $N_C = G_{MC}$, and actually, if the original symbol rate is high enough to become subject to frequency selective fading, the signal needs to be first S/P-converted before
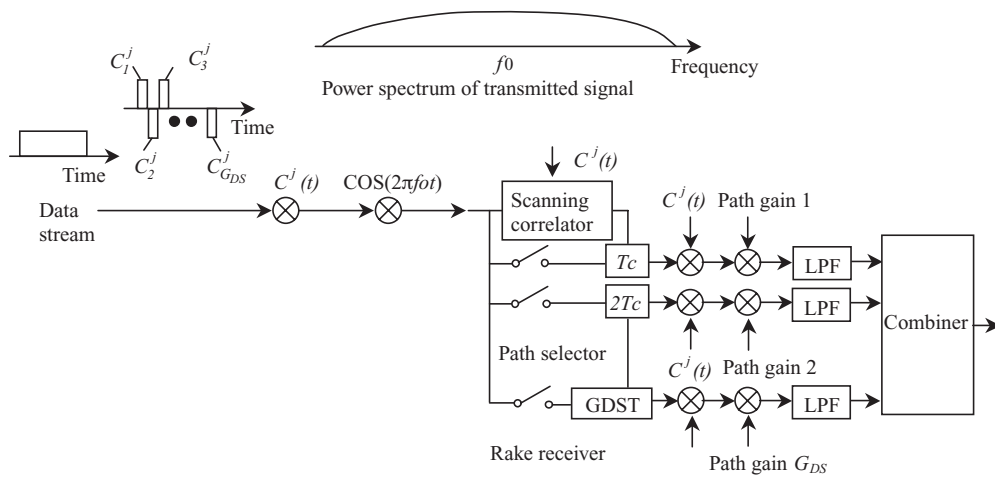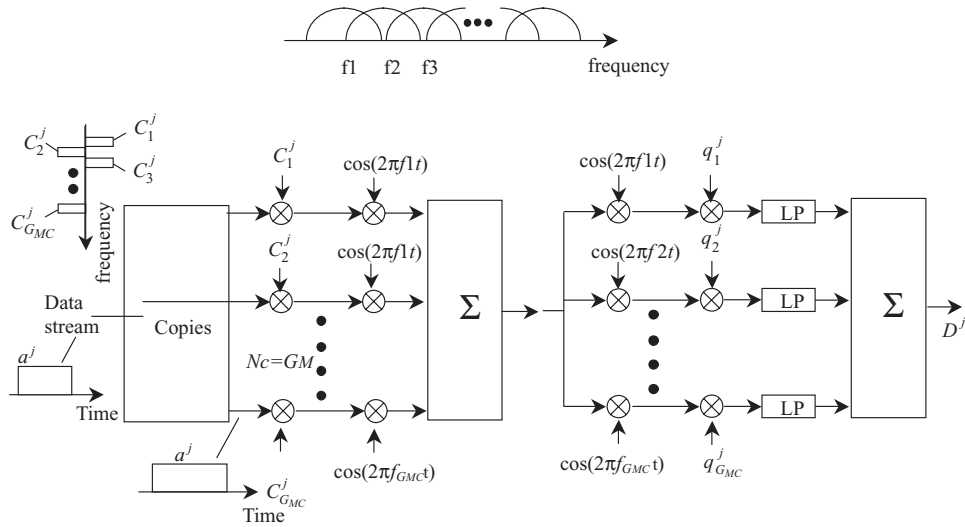

Figure 1.8 DS-CDMA scheme.

Figure 1.9   MC-CDMA scheme.

spreading over the frequency domain. This is because it is crucial for Multicarrier transmission to have frequency non-selective fading over each subcarrier.

Figure 1.10 shows the modification to ensure frequency non-selective fading, where T_S denotes the original symbol duration, and the original data sequence is first converted into P parallel sequences, and then each sequence is mapped onto G_{MC} subcarriers ($N_C = P \times G_{MC}$).
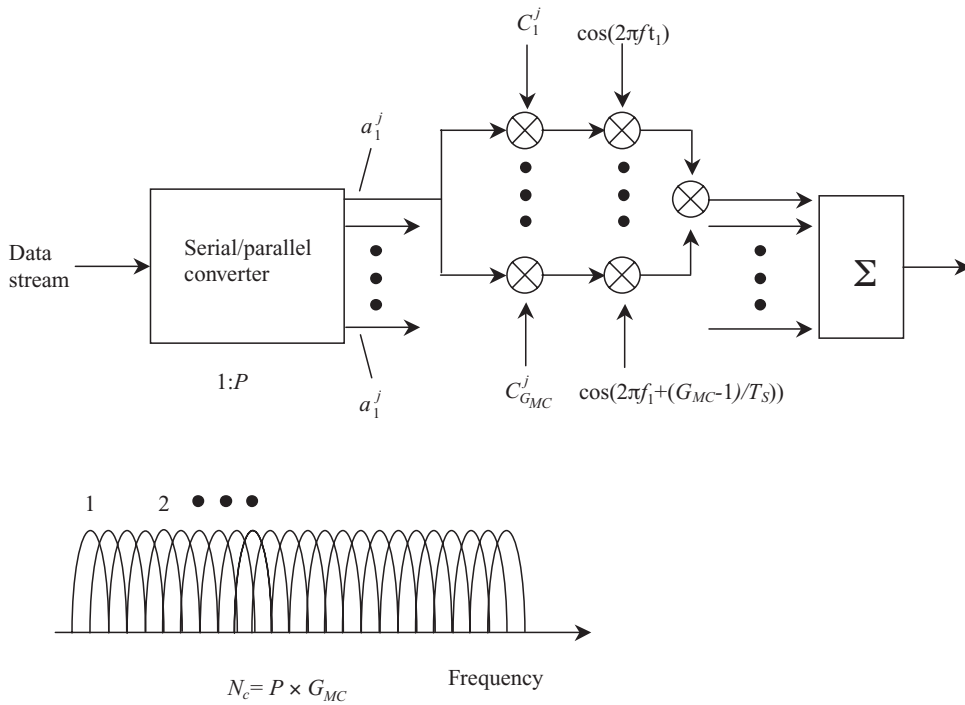

Figure 1.10   Modification of MC-CDMA scheme: spectrum of its transmitted signal.
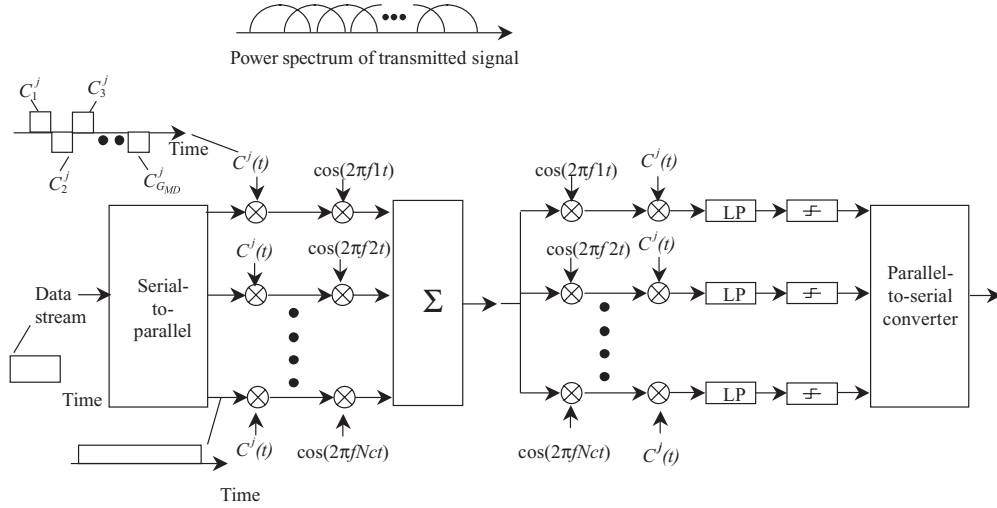
Figure 1.11 Multicarrier DS-CDMA scheme.

The Multicarrier DS-CDMA transmitter spreads the S/P-converted data streams using a given spreading code in the time domain so that the resulting spectrum of each subcarrier can satisfy the orthogonality condition with the minimum frequency separation. This scheme was proposed originally for a uplink communication channel, because the introduction of OFDM signaling into a DS-CDMA scheme is effective for the establishment of a quasi-synchronous channel.

Figure 1.11 shows the Multicarrier DS-CDMA transmitter of the *j*-th user and the power spectrum of the transmitted signal, respectively, where G_{MD} denotes the processing gain, N_C the number of subcarriers, and $C^j(t) = [C_1^j C_2^j \cdots C_{G_{MD}}^j]$ the spreading code of the *j*-th user.

The Multitone MT-CDMA transmitter spreads the S/P-converted data streams using a given spreading code in the time domain so that the spectrum of each subcarrier prior to the spreading operation can satisfy the orthogonality condition with the minimum frequency separation. Therefore, the resulting spectrum of each subcarrier no longer satisfies the orthogonality condition. The MT-CDMA scheme uses longer spreading codes in proportion to the number of subcarriers, as compared with a normal (single carrier) DS-CDMA scheme, therefore, the system can accommodate more users than the DS-CDMA scheme.

Figure 1.12 shows the MT-CDMA transmitter of the *j*-th user for CBPSK scheme and the power spectrum of the transmitted signal, respectively, where G_{MT} denotes the processing gain, N_C the number of subcarriers, and $C^j(t) = [C_1^j C_2^j \cdots C_{G_{MT}}^j]$ the spreading code of the *j*-th user.

All these schemes will be discussed in details in Chapter 7.

1.2.5 Ultra Wide Band (UWB) Signal

For the multipath resolution in indoor environments a chip interval of the order of few nanoseconds is needed. This results into a spread spectrum signal with the bandwidth in the order of few GHz. Such a signal can also be used with no carrier resulting in what is called impulse radio (IR) or Ultra Wide Band (UWB) radio. The typical form of the signal used in this case is shown in Figure 1.13. A collection of pulses received on different locations within the indoor environment is shown in Figure 1.14 and the corresponding delay profiles is presented in Figure 1.15. The ultra wideband radio will be discussed in detail in Chapter 8. In this section we will define initially only a possible signal format.

A typical time-hopping format used in this case can be represented as

$$s_{\mathrm{tr}}^{(k)}(t^{(k)}) = \sum_{j=-\infty}^{\infty} \omega_{\mathrm{tr}}(t^{(k)} - jT_f - c_j^{(k)}T_c - \delta d_{[j/N_s]}^{(k)}) \tag{1.11}$$
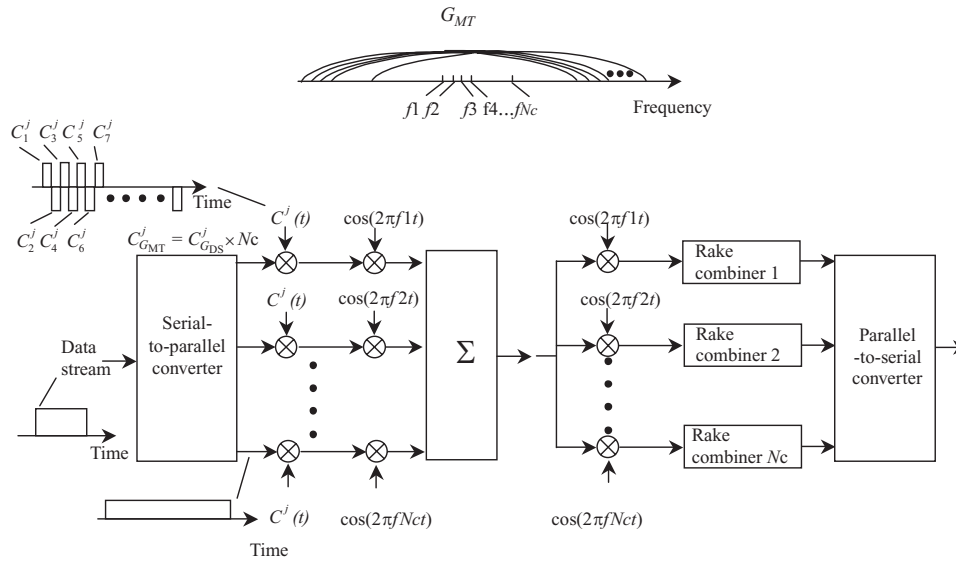
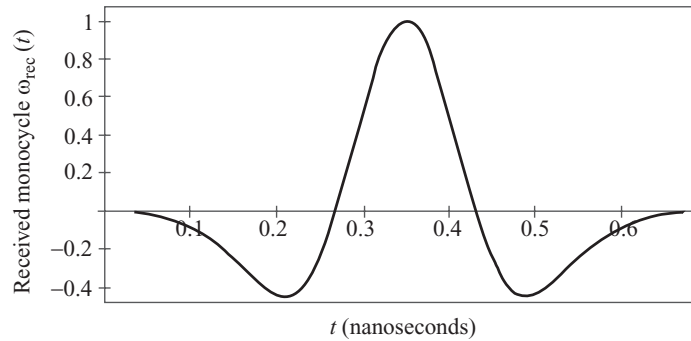**Figure 1.12**   MT-CDMA scheme.

Figure 1.13   A typical ideal received monocycle $\omega_{rec}(t)$ at the output of the antenna subsystem as a function of time in nanoseconds.
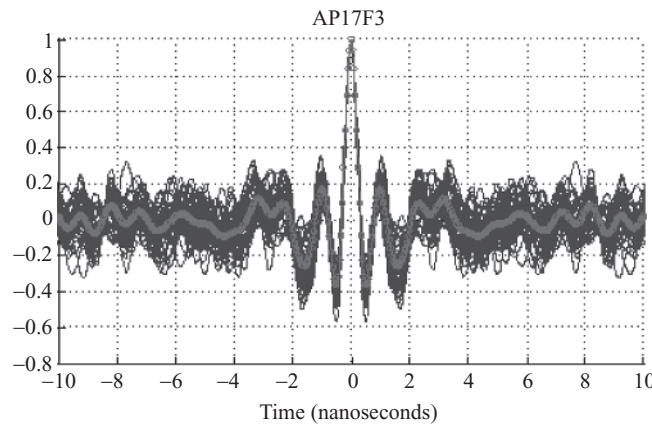

Figure 1.14   A collection of received pulses in different locations [53] © IEEE 2007.
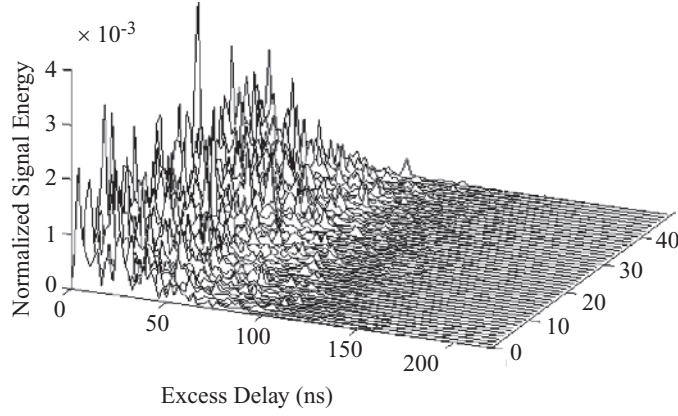
Figure 1.15   A collection of channel delay profiles [52] © IEEE 2002.

where $t^{(k)}$ is the kth transmitter's clock time and T_f is the *pulse repetition time*. The transmitted pulse waveform ω_{tr} is referred to as a *monocycle*. To eliminate collisions due to multiple access, each user (indexed by k) is assigned a distinctive time-shift pattern $\{c_j^{(k)}\}$ called a *time-hopping sequence*. This provides an additional time shift of $c_j^{(k)}T_c$ seconds to jth monocycle in the pulse train, where T_c is the duration of addressable time delay bins. For a fixed T_f the *symbol rate* R_s determines the number N_s of monocycles that are modulated by a given binary symbol as $R_s = (1/N_s T_f)s^{-1}$. The modulation index δ is chosen to optimize performance. For performance prediction purposes, most of the time the data sequence $\{d_j^{(k)}\}_{j=-\infty}^{\infty}$ is modeled as a wide-sense stationary random process composed of equally likely symbols. For data a pulse position data modulation is used.

When K users are active in the multiple-access system, the composite received signal at the output of the receiver's antenna is modeled as

$$r(t) = \sum_{k=1}^{K} A_k s_{rec}^{(k)}(t - \tau_k) + n(t) \tag{1.12}$$

The antenna/propagation system modifies the shape of the transmitted monocycle $\omega_{tr}(t)$ to $\omega_{rec}(t)$ on its output. An idealized received monocycle shape $\omega_{rec}(t)$ for a free-space channel model with no fading is shown in Figure 1.13.

The optimum receiver for a single bit of a binary modulated impulse radio signal in additive white Gaussian noise (AWGN) is a correlation receiver

"decided$_0^{(1)}$* = 0" if*

$$\underbrace{\sum_{j=0}^{N_s-1} \overbrace{\int_{\tau_1+jT_f}^{\tau_1+(j+1)T_f} r(u,t)\upsilon\left(t - \tau_1 - jT_f - c_j^{(1)}T_c\right)dt}^{\text{pulse correlator output} \triangleq \alpha_j(u)}}_{\text{test statistic} \triangleq \alpha(u)} \tag{1.13}$$

$$> 0$$

where $\upsilon(t) \triangleq \omega_{rec}(t) - \omega_{rec}(t - \delta)$.

The spectra of a signal using TH is shown in Figure 1.16. If instead of TH a DS signal is used the signal spectra is shown in Figure 1.17(a) for pseudorandom code and Figure 1.17(b) for a random code.
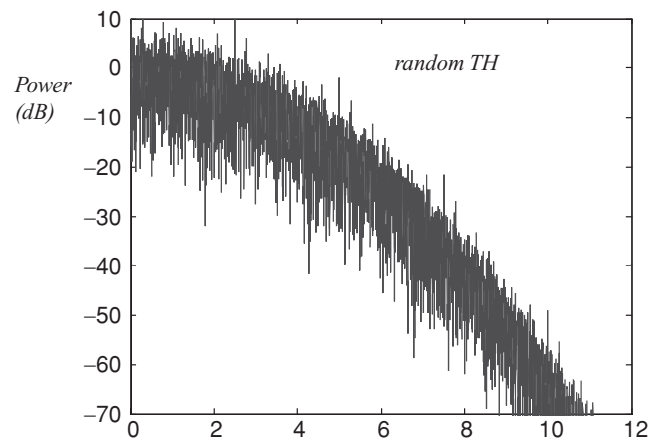
Figure 1.16 Spectra of a TH signal.
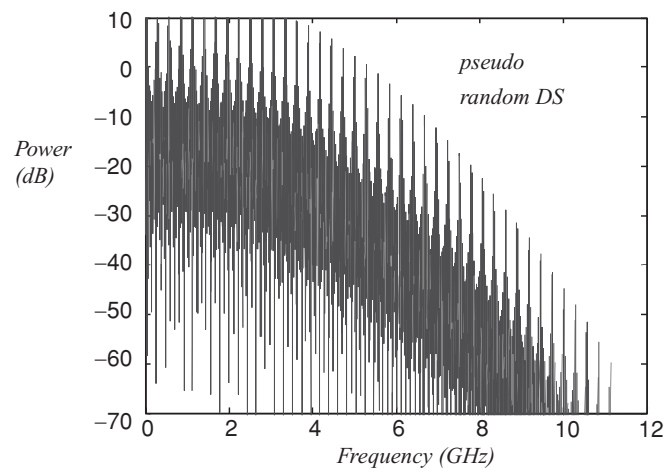
(a)
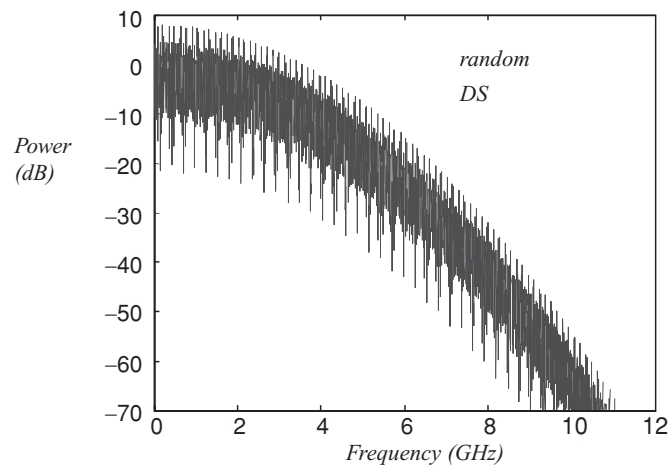
(b)

Figure 1.17 Spectra of pseudorandom DS and random DS signal.
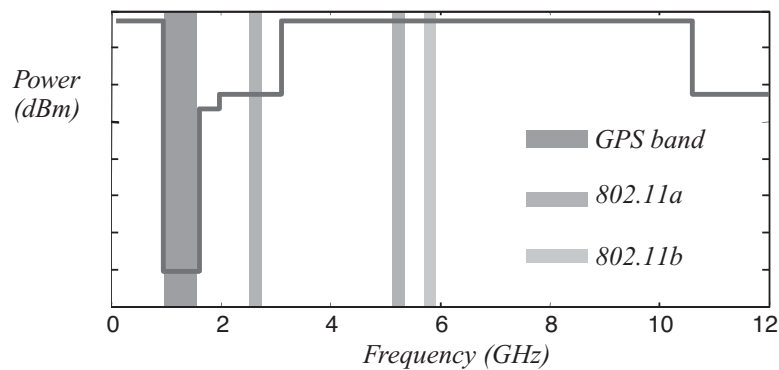
Figure 1.18   FCC frequency mask.
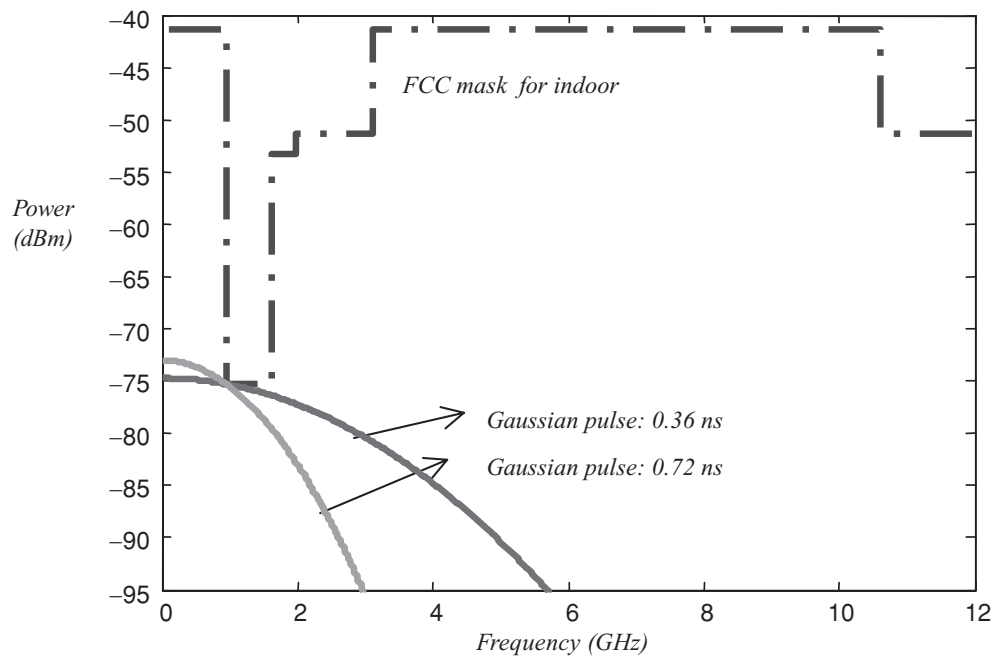
Figure 1.19   FCC mask and possible UWB signal spectra.

The FCC (Frequency Control Committee) mask for indoor communications is shown in Figure 1.18. Possible options for UWB signal spectra are given in Figures 1.19 and 1.20 for a single band and Figure 1.21 for a multiband signal format. For more detail see www.uwb.org and www.uwbmultiband.org.

The optimal detection in a multiuser environment, with knowledge of all time-hopping sequences, leads to complex parallel receiver designs [2]. However, if the number of users is large and no such multiuser detector is feasible, then it is reasonable to approximate the combined effect of the other users' dehopped interfering signals as a Gaussian random process. All of the detail regarding system performance will be discussed in Chapter 8.
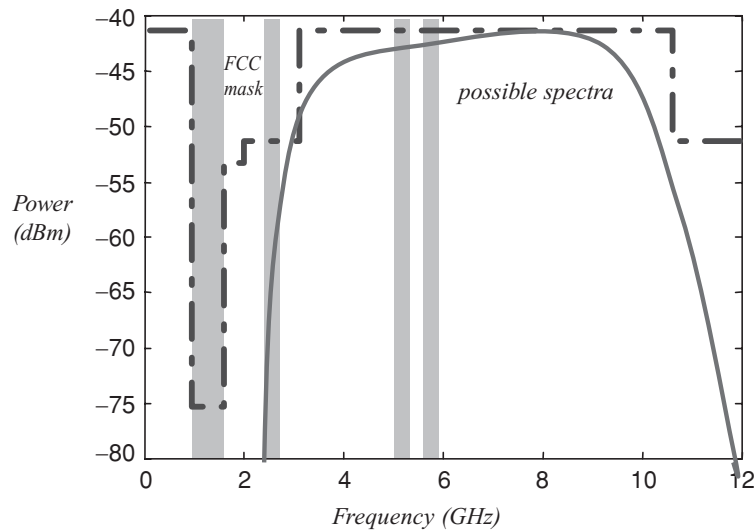
Figure 1.20   Single band UWB signal.
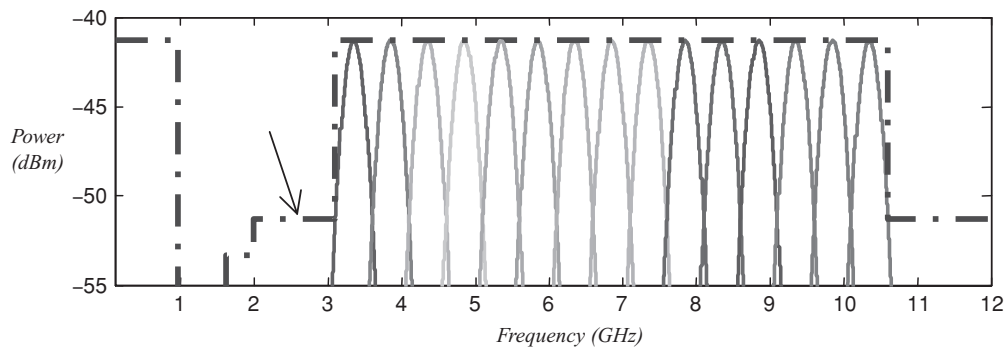
Figure 1.21   Multi band UWB signal.

1.3   Next Generation Internet

As already mentioned in the introduction to this chapter the evolution of wireless communications will be closely related to the evolution of the Internet. For this reason in this section we discuss this issue in more detail. The Internet architecture was developed almost 30 years ago and its basic framework has remained resistant to major changes. In order to predict future evolution within the Internet some authors [55] use the general theory of innovations [56–58] developed in economics.

Sustainable and Disruptive Innovation are the two important categories driving the market in two different ways along different economic dimensions. At some point in time in any industry, incumbent firms providing products and services in a market are competing by working to improve their offerings along a few narrowly defined dimensions. Typically, most products or services start out to be 'not good enough' in functionality, reliability and performance, and companies improve the products along one or more of these dimensions. The innovations involved in improving along these existing metrics are called *sustaining innovations*, even if such innovations from a technology standpoint may be quite radical.

Firms base their Sustaining innovation product improvement programs on the needs of their most demanding customers. The improvements at some point outstrip the needs of their low end customers.

These customers become 'overshot', while the needs of the high end customers remain 'undershot'. At this point, the low end customers begin to value *convenience, customizability*, and *price* more than functionality, reliability and performance.

When the number of low end, overshot customers becomes numerous enough, the industry becomes ripe for the entry of a collection of low-end innovators to take customers away from the incumbent firms. The entry firms are able to make good profits by providing a basic service or product that is unattractive to the high end customers but 'good enough' for low end customers due to lower price *and/or* much improved ease of use. If the entry firms are able to decouple their value chains completely from those of the incumbents, these entry firms can often force incumbent firms to abandon the low end customers and flee 'up market'. The innovations used by the entry firms in this kind of situation are called *disruptive*.

This entry and flight strategy only works for some time. Because the needs of the high end customers don't change fast enough, at some point, the entrant firms and the incumbent firms end up competing for the same pool of high end customers. At that point, the incumbent firms often go bankrupt or are merged with the entrants because the entrants' business models are honed to make money at a lower price point than the incumbents. On the other hand, disruptive entrants can fail if the incumbents are motivated to fight because the entrants' value chains overlap with theirs, or when an incumbent crams a technology successfully with disruptive potential into an existing business model.

There is also another, parallel disruptive innovation path, in which a firm offers a product or service that is not available within the existing market. The product or service often looks primitive or cheap because, initially, it does not provide the same level of performance as mainstream products. This strategy is referred to as *new market disruption* [56–58].

Conservation of Integration refers to the case when a market is in the initial stages of competing along the metrics of functionality, reliability and performance and companies need to control as many steps in the product architecture as possible in order to improve the product or service along the competition metrics. Firms that build their products or services around proprietary, integrated architectures have an advantage, because they can optimize along all metrics without having to compromise. Architectures that are modular – that is, which have well defined interfaces between components, allowing components to be independently developed – invariably fail to deliver along the competitive metrics since some aspects of functionality, reliability or performance must be compromised in order to achieve modularity.

However, once the market has flipped to competing along the metrics of convenience, customizability and price, firms which base their products on modular components have a competitive advantage. For an architecture to be modular, clients and implementers on both sides of the interface must agree on the *specifiability, verifiability* and *predictability* of the components. Modular interfaces allow flexibility in picking component suppliers, distributors and other participants in the ecosystem, while sacrificing some performance. Modular architectures therefore deliver convenience and customizability to customers much more quickly than integrated architectures, and at a lower price.

Internet architecture is a modular architecture with the Internet Protocol (IP) and the various IP transport protocols (such as TCP and UDP) as the modular interface between physical and data link layers and the upper layers. HTTP performs the same function for the application layer. The end-to-end principle, which is the fundamental architectural principle underlying the Internet architecture, is basically a modularity argument: the interface between applications and the network should be clearly defined and the functions specific to the communication system (i.e. transport and routing) are the only functions that should be within the network. IP and the base Internet transport protocol suite provide the specifiability, verifiability and predictability for applications and lower layer transport to agree on how packets get from one end to the other, and what the reliability characteristics of the transmission are. These characteristics are supported by the transparency of the Internet architecture: what goes in one end comes out the other.

The global communication networks replaced by the Internet (the circuit switched telephone network) were in contrast integrated, since applications embedded knowledge of the network operation within them and were deployed within the network. While the assumption among many in the technical community is that the end-to-end principle and the Internet superseded the circuit switched telephone network solely due to technical superiority, the Internet architecture would not have achieved widespread deployment over the earlier integrated architectures without a suitable economic driver. The economic driver was a change

in the desired customer performance metrics. The basic network performance metrics that customers care about (namely bandwidth and latency) became optimized to the point that the performance losses that came from modularization by moving to IP no longer mattered, and customers began to value applications other than simple voice. As a result, the economic benefits to the customers of a modular architecture (many suppliers, price reduction, ease of customization, etc.) outweighed the costs.

It is *because* the Internet architecture has not changed much over the last 30 years that massive innovation has been possible above and below the network layer, and, more recently, above the HTTP layer. By *Conservation of Integration* principle, the data transport equipment (such as routers) below the IP layer should all exhibit integrated architectures, since they are not yet good enough, along their particular metrics of competition, for modularization to occur. Similarly, applications should exhibit the same characteristics above the HTTP layer, but we focus on IP and the lower layers in the rest of this section.

Routers are, in fact, highly integrated in the sense that they are complex integrated *hardware/software* products which include proprietary features and interfaces especially for management. The complexity and integration throw up high barriers to entrant firms. Routers also tightly integrate the control software with the data switching hardware, and both change in new versions. Thus operators are pressed to deploy control software upgrades (which reinforce the lock-in) when all they really need is additional switching capacity.

The evolution from an integrated to a modular architecture is not unidirectional, systems can evolve in the other direction. Something similar happened in the early 1990s when Microsoft integrated desktop office applications more tightly to the operating system in Windows in response to a change in the preferred customer performance metrics. Prior to that, MS-DOS provided a modular interface and desktop apps were not well integrated, either with each other or with the operating system.

This suggests that the most likely path to a major, incompatible shift in the Internet architecture is something that would cause the IP modular interface to lose its attractiveness to customers and cause an integrated architecture to come back into favor, since simply providing a slightly better modular interface is unlikely to win out against the enormous installed base of IP. A shift in the Internet architecture could occur if customers began to value some other metric than the current one of cheap, high bandwidth. The shift must cause the modular IP interface to deliver suboptimal performance along the new metric, requiring reintegration across the IP interface.

In the sequel, we discuss two technology areas that have the potential to foster radical innovation in the Internet in the short term.

Cloud Computing and network virtualization is a technology trend that is currently attracting a lot of attention. The *control/data* plane split is a technology trend that has yet to really develop, but holds potential. Although these innovations leave the basic Internet architecture untouched, they have the potential to catalyze radical change in the way Internet infrastructure and applications are deployed, following in the trend of many previous innovation waves in telecom and data networking.

1.4 Cloud Computing and Network Virtualization

In principle, enough unused computing capacity exists on the desktop so the computing needs handled by Cloud Computing could be handled by consolidating processing on desktops. The difficulty of software installation (that we can't locally execute the code we want) and of replicating large databases make providing cloud-like services difficult with the current service deployment model. By consolidating server capacity into large data centers, economies of scale allow companies hosting Web services to achieve much more cost effective hosting and data storage [59].

Servers in a Cloud Computing facility are virtualized, meaning that they can run multiple customer operating system images with different applications at the same time.

Consolidation of servers and virtualization simplifies management, both from the customers' viewpoint and from the cloud provider's viewpoint. While there are still many technical and business issues surrounding Cloud Computing, the ultimate vision is *utility computing*: providing resources for

Fundamentals   **19**

processing, bandwidth and storage in the same way that a utility provides electricity, water or telephone service.

Network virtualization is complementary technology. The idea is to divide the network into slices that are separate and run separate applications within the slices, with each slice allocated bandwidth and processing on network elements. The isolation between slices is used for privacy, security and guaranteed bandwidth. The goal is an on-demand network service that can provide a particular class of service (best effort, expedited forwarding, etc.) for reasonable cost.

Flexible connectivity into the cloud completes the end-to-end connection. The combination of Cloud Computing and network virtualization would allow a business to define a collection of applications and services in the compute cloud that could be accessed end-to-end within a network from virtual machines. The experience would be similar to a corporate *WAN/LAN* environment, except the compute and network resources would reside in the cloud and be rented by the business rather than owned.

For service providers, Cloud Computing is more likely to be a sustaining than a disruptive innovation. The large operators are likely to have the motivation and skills to master Cloud Computing and successfully offer their own cloud services. As an example, AT&T already has a Cloud Computing service [60] through their regional data centers. Cloud Computing is sustaining for incumbent service providers because any cloud provider must connect up to the network, so a disruptive entrant would have a hard time building an ecosystem independent from the incumbents. An incumbent service provider has an advantage, since it can also provide network virtualization as part of its cloud service.

The business models supported by incumbents and entrants are currently somewhat different. AT&T's service requires a contract including network SLAs to be negotiated between the business and cloud provider, whereas new entrant services such as Amazon's EC2 require a credit card for anonymous payment, and network bandwidth only is guaranteed to and from the data center. Initially, there may be room for both business models to grow but at some point the overlap in ecosystems may result in clashes.

For equipment vendors, the network equipment and server business seems to be realigning itself along a new performance metric, causing a re-integration across communication performance within data centers. Currently, standard L2 switches are used to build the switching fabric within data centers for communication between servers. Standard server blades are used to form the computational infrastructure. The cost advantage comes from using commercial off-the-self hardware optimized for deployment in individual chassis.

Optimization of communication between servers by integrating the communication more tightly with the server hardware seems a likely step.

The reintegration across the server backplane is enabled by a migration of the modular interface between the server and the network out to the virtualization layer. Newer releases of the virtualization systems, such as VMWare and Xen, include a virtualized switch that insulates running virtual machine images from the details of their location in the data center switching fabric. This allows the running images to be moved around the data center to different servers without changing the location of the running image in the IP address topology. The virtual machine images become independent of the switch hardware, opening a space for innovation to improve the performance of data center networking.

The basic idea behind the *control/data* plane split or split routing architecture is that IP subnet topologies, while quite useful in access networks, provide too much functionality in operator core networks at a high cost of configuration complexity. Most core network links are point to point, often switched, and therefore don't require the rich many-to-many provisioning capabilities offered by IP subnet topologies. In addition, basic IP routing protocols don't really take advantage of richly interconnected topologies anyway, since the shortest path first algorithm used by the interior gateway routing protocols concentrates traffic on one path.

MPLS [61], Carrier Ethernet [62] and GMPLS [63] are examples of split routing architecture technologies. Split routing architecture refers to splitting the router into two pieces in separate network nodes with a modular interface between: a route controller that handles the distribution of routing information and policy to control where packets go (the control plane) and a collection of routing switches where the actual forwarding decisions are made on individual traffic packets (the data plane). Another way to think of it is as a separation of policy-path setup, traffic engineering, etc. – from mechanism – the

forwarding action. As they currently stand, split routing architecture technologies don't appear particularly threatening to the basic infrastructure providers. Expensive core routers running IP-MPLS make up a good chunk of their high margin business. MPLS still requires routers, and Carrier Ethernet requires the control plane software to be integrated with the switches.

Information-Centric Networking (ICN), or Content Centric Networking (CCN), is a relatively new research trend [14]. The basic idea is for networking to address the content, or information objects, and not the containers, or network nodes and connections. Receivers therefore get to choose what senders they get content from, rather than allowing senders to address receivers without permission, as in the current Internet architecture. From the innovation theory point of view, ICN requires integrating across the IP layer and creating a new modularity interface in the network stack. This new modularly interface creates a new 'waist' for the protocol stack, a role similar to the role played by IP protocol for the last 30 years. However unlike HTTP, which recently established an application 'waist' in the IP stack, ICN fundamentally changes the architecture to remove end node addressability. Three main metrics seem to be pushing the world towards ICN: need for reduced latency, dropping storage space/transmission price ratio, and ease of use. ICN, once developed to the full, is likely to address all of these metrics.

1.5 Economics of Utility Computing

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide those services [59] (see Figure 1.22). The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is referred to as a *Cloud*. A Cloud, made available in a pay-as-you-go manner to the general public, is called Public Cloud; the service being sold is *Utility Computing*. The term Private Cloud is used to refer to internal data centers of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. From a hardware point of view, three aspects are new in Cloud Computing.

1. The illusion of infinite computing resources available on demand, thereby eliminating the need for Cloud Computing users to plan far ahead for provisioning.
2. The elimination of an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
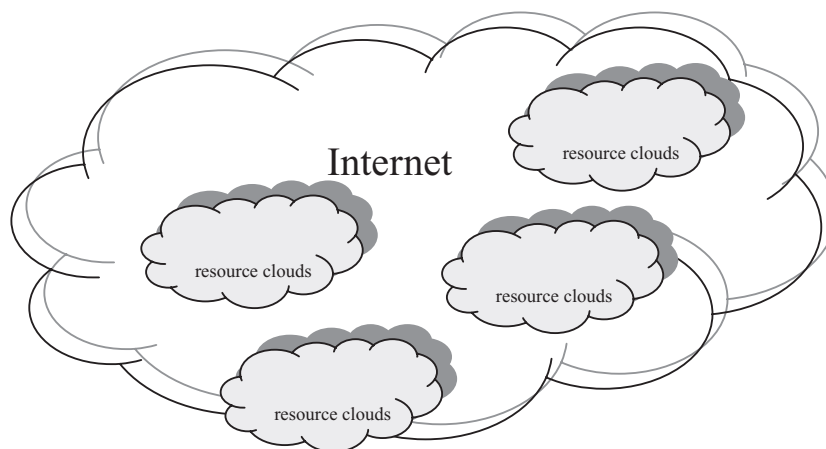
Figure 1.22   Utility computing concept.

3. The ability to pay for use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

Any application needs a model of computation, a model of storage, and a model of communication. The statistical multiplexing necessary to achieve elasticity and the illusion of infinite capacity requires each of these resources to be virtualized to hide the implementation of how they are multiplexed and shared.

Regarding Cloud Computing economic models the following observations can be made:

a) Economic models enabled by Cloud Computing make tradeoff decisions about whether hosting a service in the cloud makes sense over the long term, more fluid, and in particular the elasticity offered by clouds serves to transfer risk.
b) Although hardware resource costs continue to decline, they do so at variable rates. Cloud Computing can track these changes and potentially pass them through to the customer more effectively than building one's own datacenter, resulting in a closer match of expenditure to actual resource usage.
c) In making the decision about whether to move an existing service to the cloud, one must additionally examine the expected average and peak resource utilization, especially if the application may have highly variable spikes in resource demand; the practical limits on real-world utilization of purchased equipment; and various operational costs that vary depending on the type of cloud environment being considered.

The economic appeal of Cloud Computing is based on converting capital expenses to operating expenses and the phrase 'pay as you go' is often used to capture directly the economic benefit to the buyer. Hours purchased via Cloud Computing can be distributed non-uniformly in time (e.g., use 80 server-hours today and no server-hours tomorrow, and still pay only for what you use); in the networking community, this way of selling bandwidth is already known as usage-based pricing. In addition, the absence of up-front capital expense allows capital to be redirected to core business investment.

Therefore, even though pay-as-you-go pricing could be more expensive than buying and depreciating a comparable server over the same period, it is argued that the cost is outweighed by the extremely important Cloud Computing economic benefits of elasticity and transference of risk, especially the risks of overprovisioning (underutilization) and underprovisioning (saturation).

The elasticity refers to the Cloud Computing's ability to add or remove resources at a fine grain (one server at a time) within minutes rather than weeks allowing matching resources to workload much more closely. Real world estimates of server utilization in datacenters range from 5% to 20% [65, 66]. This is consistent with the observation that for many services the peak workload exceeds the average by factors of 2 to 10. Few users deliberately provision for less than the expected peak, and therefore they must provision for the peak and allow the resources to remain idle at non peak times. The more pronounced the variation, the more the waste.

While the monetary effects of overprovisioning are easily measured, those of underprovisioning are harder to measure yet potentially equally serious since not only do rejected users generate zero revenue, they may never come back due to poor service.

In the concept of Cloud Computing the risk of mis-estimating workload is shifted from the service operator to the cloud vendor.

There are two additional benefits to the Cloud Computing user that result from being able to change their resource usage on the scale of hours rather than years.

First, unexpectedly scaling down (disposing of temporarily underutilized equipment), due to a business slowdown, or due to improved software efficiency, normally carries a financial penalty.

Second, technology trends suggest that over the useful lifetime of some purchased equipment, hardware costs will fall and new hardware and software technologies will become available. Cloud providers, who already enjoy economy-of-scale buying power, can potentially pass on some of these savings to their customers.

The previous discussion tried to quantify the economic value of specific Cloud Computing benefits such as elasticity. In the sequel we will extend our discussion to the equally important but larger question of whether or not it is more economical to move the existing data center-hosted service to the cloud, or to keep it in a data center?

This simple analysis includes several important factors.

First, most applications do not make equal use of computation, storage and network bandwidth; some are CPU-bound, others network-bound and so on, and may saturate one resource while underutilizing others. Pay-as-you-go Cloud Computing can charge the application separately for each type of resource, reducing the waste of underutilization. While the exact savings depends on the application, suppose the CPU is only 50% utilized while the network is at capacity; then in a data center you are effectively paying for double the number of CPU cycles actually being used.

The costs of power, cooling, and the amortized cost of the building are missing from our simple analyses so far. It is estimated that the costs of CPU, storage and bandwidth roughly double when those costs are amortized over the building's lifetime.

Today, hardware operations costs are very low. Rebooting servers is easy and minimally trained staff can replace broken components at the rack or server level. On one hand, since Utility Computing uses virtual machines instead of physical machines, from the cloud user's point of view these tasks are shifted to the cloud provider. On the other hand, depending on the level of virtualization, much of the software management costs may remain, e.g. upgrades, applying patches, and so on.

1.6   Drawbacks of Cloud Computing

In this section, we discuss the obstacles to the growth of Cloud Computing as classified in [59].

Availability of service is what organizations are concerned about. So far, existing SaaS products have set a high standard in this regard.

Just as large Internet service providers use multiple network providers so that failure by a single company will not take them off the air, it is believed that the only plausible solution to very high availability is multiple Cloud Computing providers.

Distributed Denial of Service (DDoS) attacks is another availability obstacle. Criminals threaten to cut off the incomes of SaaS providers by making their service unavailable. As with elasticity, Cloud Computing shifts the attack target from the SaaS provider to the Utility Computing provider, who can more readily absorb it and is also likely to have already DDoS protection as a core competency.

Data lock in is the next concern. Software stacks have improved interoperability among platforms, but the *application programming interfaces* (APIs) for Cloud Computing itself are still essentially proprietary, or at least have not been the subject of active standardization. Thus, customers cannot easily extract their data and programs from one site to run on another. Concern about the difficult of extracting data from the Cloud is preventing some organizations from adopting Cloud Computing. Customer lock-in may be attractive to Cloud Computing providers, but Cloud Computing users are vulnerable to price increases, to reliability problems, or even to providers going out of business.

Data Confidentiality and Auditability is also a concern. Current Cloud offerings are essentially public (rather than private) networks, exposing the system to more attacks. It is believed that there are no fundamental obstacles to making a Cloud-Computing environment as secure as the vast majority of in-house IT environments, and that many of the obstacles can be overcome immediately with well understood technologies such as encrypted storage, Virtual Local Area Networks, and network middleboxes (e.g. firewalls, packet filters). For example, encrypting data before placing it in a Cloud may be even more secure than unencrypted data in a local data center.

Data Transfer Bottlenecks are also a concern. Applications continue to become more data-intensive. If we assume applications may be 'pulled apart' across the boundaries of Clouds as indicated in Figure 1.22, this may complicate data placement and transport. The concern becomes of paramount importance in the case of *Wireless Internet*. This will be discussed in more details in the next section. One opportunity to overcome the high cost of Internet transfers is to *ship disks*. Jim Gray found that the cheapest way to

send a lot of data is to physically send disks or even whole computers via overnight delivery services [67]. A second opportunity is to find other reasons to make it attractive to keep data in the Cloud, for once data is in the Cloud for any reason it may no longer be a bottleneck and may enable new services that could drive the purchase of Cloud Computing cycles.

A third, more radical opportunity is to try to reduce the cost of WAN bandwidth more quickly.

Performance Unpredictability is also a concern. The experience is that multiple Virtual Machines can share CPUs and main memory surprisingly well in Cloud Computing, but that I/O sharing is more problematic. One opportunity is to improve architectures and operating systems to virtualize interrupts and I/O channels efficiently.

Technologies such as PCIexpress (Peripheral Component Interconnect) are difficult to virtualize, although they are critical to the Cloud. Fortunately IBM mainframes and operating systems largely have overcome these problems, so we have successful examples from which to learn. Another possibility is that flash memory will decrease I/O interference. Flash is semiconductor memory that preserves information when powered off like mechanical hard disks, but since it has no moving parts, it is much faster to access (microseconds vs. milliseconds) and uses less energy. Flash memory can sustain many more I/Os per second per gigabyte of storage than disks, so multiple virtual machines with conflicting random I/O workloads could coexist better on the same physical computer without the interference we see with mechanical disks. Another unpredictability obstacle concerns the scheduling of virtual machines for some classes of batch processing programs, specifically for high performance computing.

Scalable Storage is one of the system parameters too. Early in this section, we identified three properties whose combination gives Cloud Computing its appeal: short-term usage (which implies scaling down as well as up when resources are no longer needed), no up-front cost, and infinite capacity on-demand. While it's straightforward what this means when applied to computation, it's less obvious how to apply it to persistent storage. It includes questions of the performance guarantees offered, and the complexity of data structures that are directly supported by the storage system. The opportunity, which is still an open research problem, is to create a storage system that would not only meet these needs but combine them with the Cloud advantages of scaling arbitrarily up and down on-demand, as well as meeting programmer expectations in regard to resource management for scalability, data durability and high availability.

Bugs in Large-Scale Distributed Systems is the next concern. One of the difficult challenges in Cloud Computing is removing errors in these very large scale distributed systems. A common occurrence is that these bugs cannot be reproduced in smaller configurations, so the debugging must occur at scale in the production data centers.

One opportunity may be the reliance on virtual machines (VM) in Cloud Computing.

Scaling Quickly is also important. Pay-as-you-go certainly applies to storage and to network bandwidth, both of which count bytes used. Computation is slightly different, depending on the virtualization level. Google AppEngine scales in response to load increases and decreases automatically, and users are charged by the cycles used. AWS (*Amazon Web Services*) charges by the hour for the number of instances you occupy, even if your machine is idle. The opportunity is then to scale quickly up and down automatically in response to load in order to save money, but without violating service level agreements.

Cloud Computing providers already perform careful and low overhead accounting of resource consumption. By imposing per-hour and per-byte costs, utility computing encourages programmers to pay attention to efficiency (i.e., releasing and acquiring resources only when necessary), and allows more direct measurement of operational and development inefficiencies.

Reputation Fate Sharing is a part of the overall system performance too. Reputations do not virtualize well. One customer's bad behavior can affect the reputation of the Cloud as a whole. An opportunity would be to create reputation-guarding services similar to the 'trusted email' services currently offered (for a fee) to services hosted on smaller ISP's, which experience a microcosm of this problem. Another legal issue is the question of transfer of legal liability. Cloud Computing providers would want legal liability to remain with the customer and not be transferred to them.

Software Licensing should be also considered. Current software licenses commonly restrict the computers on which the software can run. Users pay for the software and then pay an annual maintenance

fee. Hence, many Cloud Computing providers originally relied on open source software in part because the licensing model for commercial software is not a good match to Utility Computing.

The primary opportunity is either for open source to remain popular or simply for commercial software companies to change their licensing structure to better fit Cloud Computing. For example, Microsoft and Amazon now offer pay-as-you-go software licensing for Windows Server and Windows SQL Server on EC2.

1.7   Wireless Grids and Clouds

In this section we continue our discussion on Cloud Computing with emphasis on wireless networks. *Wireless grid* is a wireless network based virtual system that consists of wireless-connected different types of electronic devices and computers [68–77]. It has broad application prospects in *e*-learning, mobile *e*-business, modern healthcare, smart home, wireless sensor networks and disaster management. Wireless grids are based on wireless networks but the communication infrastructure is not the only difference between wireless grids and traditional wired grids. Wireless grids have the following distinguishing features:

Ad hoc mode: Wireless networks, especially wireless ad hoc networks, facilitate the on-demand integration of heterogeneous devices. The ad hoc connected devices have no centric control. It makes the wireless grid more dynamic and brings more convenience to resource sharing and coordination. At the same time, the ad hoc mode also brings more challenges regarding security, resource discovery and trusted resource management to wireless grids.

Resource constrained devices like high performance servers, mass storage devices, are very common in computing grids, data grids, access grids and other traditional wired grids. But in wireless grids, the low-powered small devices, like sensors, smart phones, digital cameras, PDAs and laptops, are more popular. Most of them are battery-powered. Their storage and computing capabilities are also limited.

Heterogeneous components: Grid facilitates sharing of heterogeneous resources [78]. The evolution from PCs to smaller and mobile devices with computing capabilities brings more heterogeneous devices and applications into wireless grids. Wireless grids have to manage and support the inter-organizational applications that consist of more heterogeneous components from autonomic devices.

Dynamic resources and requirements: In wireless grids, the users, resources, requirements and computing environments are dynamic. The wireless networks are inherently volatile, due to their unstable wireless communication medium. In addition, the changes of natural environment also influence the wireless networks. In wireless grids, mobile users are common. And the users use mobile devices increasingly. Due to the ad hoc mode, the users and resources can join and quit the grid conveniently. The flexible resource sharing and coordination mode inspire the users to spontaneously raise their requirements for new applications.

Wireless grid architectures could be classified into the following categories according to the predominant devices and device mobility [70]: (a) Wireless sensor grid architecture; (b) Mobile wireless grid architecture; (c) Fixed wireless grid architecture.

The wireless sensor grid integrates wireless sensor networks and traditional grid computing technologies [71]. The applications are designed around the data from the sensors. And the sensors play an important role in this kind of wireless grid architecture. On the basis of wireless sensor network infrastructures, it adopts the technologies of data grids, computing grids and access grids to storing, processing, presenting and sharing the data collected from the sensors.

The mobile wireless grid has much more mobile devices, such as mobile phones and PDAs. In addition, the users are also mobile. The network infrastructure is a mobile ad hoc network. The devices participate in the mobile wireless grid as data collectors or client nodes. In most cases, these devices are mobile wireless access clients. They can freely enter and leave the wireless grid together with their mobile users.

The fixed wireless grid has no difference with traditional wired grid, except the communication medium is wireless.
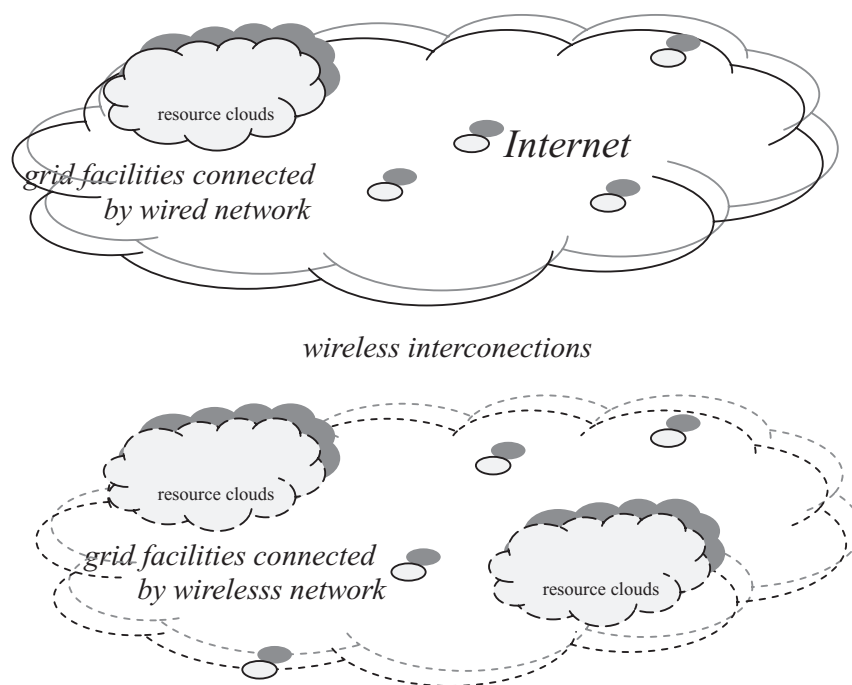
Figure 1.23   Hybrid wireless grid architecture.

In practice, besides the pure wireless network infrastructure, a hybrid network infrastructure is more commonly used in wireless grids, where the wireless networks are connected with wired networks.

Figure 1.23 depicts an architecture that integrates the wireless grid with wired network backbones.

In this architecture, the high resource-consuming devices, such as high performance servers and mass storage devices, are connected by high speed wired networks; the small or mobile devices are integrated by wireless ad hoc networks. This hybrid architecture makes up the lack of resources in wireless grids.

Taking one with another, the architecture of a wireless grid can be divided into four layers: physical layer, communication layer, middleware layer and application layer.

The physical layer covers all the devices involved in the wireless grid. The communication layer chains the devices with wireless networks or connects the wireless-linked devices with wired networks. The middleware layer takes charge of resource description, resource discovery, coordination and trust mechanism, and supports development of wireless grid applications. On the basis of middleware layer, the application layer includes a variety of wireless grid applications, such as health care systems, environment surveillance systems, seismic monitoring systems and so on.

Many wireless communication technologies are used in wireless grids, to provide wireless interfaces to connect high-bandwidth powered PCs and low-bandwidth powered smart devices. These technologies include IEEE 802.llx, IEEE 802.16x, IEEE 802.15x, IEEE 802.20x, Zigbee, Bluetooth, wireless ad hoc networks, etc.

IEEE 802.11 is a set of protocol standards proposed by the IEEE *LAN/MAN* Standards Committee. The protocol family includes 802.11a~z. They support wireless local area network communications in the 2.4, 3.7 and 5 GHz frequency bands.

The 802.11 protocols are very similar to the 802.3 protocols. In fact, they are wireless Ethernet protocols. The 802.11 protocols cover the physical layer (PHY) and the media access control layer (MAC) of the data link layer. Different from the 802.3 protocols, the 802.11 protocols use *CSMA/CA*

(Carrier Sense Multiple Access with Collision Avoidance) technologies instead of *CSMA/CD* (Carrier Sense Multiple Access with Collision Detection) at the MAC layer, to meet the requirements of wireless communications.

Wi-Fi (Wireless Fidelity) is a trademark of the Wi-Fi Alliance. The Wi-Fi technologies cover the protocols of 802.11a, 802.11b, 802.11g and 802.11n, and can be used in mobile phones, home networks and other devices that are connected by wireless networks. With Wi-Fi, one can construct flexible wireless networks in infrastructure mode or ad hoc mode. These 802.11 based networks equip wireless grids with fundamental communication facilities.

The protocol family of IEEE 802.16 was proposed by the IEEE *LAN/MAN* Standards Committee also. It includes a group 802.16a–k, which are the standards of broadband wireless access. The IEEE 802.16 protocols support wireless communications in the frequency bands of 10–66 GHz.

The IEEE 802.16 defines the standards of protocols at the physical layer (PHY) and the media access control layer (MAC). These protocols support both mobile access and fixed access. At PHY layer, the 802.16 adopts the technologies of OFDM (Orthogonal Frequency Division Multiplex) and OFDMA (Orthogonal Frequency Division Multiple Access). OFDMA can adapt flexibly to requirements changes of bandwidth. With fixed sub-carrier frequency interval and symbol time, it also reduces the impact of changes at PHY layer on the MAC layer.

Different from the IEEE 802.11 protocols that support WLAN (Wireless Local Area Networks), the IEEE 802.16 protocols focus on WWAN (Wireless Wide Area Network). They use TDMA (Time Division Multiple Access) at MAC layer, instead of CSMA. In addition, the protocols have three sub-layers at MAC layer: service specific convergence sub-layer, common part sub-layer and privacy sub-layer.

In addition, the IEEE 802.16 protocols have more choices with regard to bandwidth and frequency and have higher transmission power. It makes them more suitable for the wireless grids that need wide area wireless networks.

WiMAX (Worldwide Interoperability for Microwave Access) is a technology based on the IEEE 802.16. It supports 802.11d and 802.11e. Similar to Wi-Fi, *WiMAX* can connect handsets, PC peripherals and embedded devices. As a backhaul technology for 3*G* (Third Generation) and 4*G* (Fourth Generation), *WiMAX* is a promising technology for constructing wireless grid.

The ad hoc is an organization mode of the wireless network nodes. A wireless ad hoc network consists of a group of autonomous nodes that communicate with each other in a decentralized manner. Each node of the network connects others with wireless links, and acts as both a host and a router. The nodes can join or quit the network freely. Therefore, the topology of wireless ad hoc network is dynamic. The decentralized nature makes the ad hoc network more suitable for wireless grid.

According to applications, wireless ad hoc networks are classified into three categories: (a) Mobile ad hoc networks where mobile devices connect with each other in a self-configuring way; (b) Wireless mesh networks where nodes are organized in mesh topology and in most cases, the nodes of wireless mesh networks are static; and (c) Wireless sensor networks where nodes are distributed autonomous devices that can monitor environmental conditions. Wireless sensor networks are typical tools for data collection in wireless grid.

Wireless ad hoc networks provide wireless grids with ubiquitous untethered communication, and allow wireless devices to discover and interact with each other in a *P2P* manner. With a special purpose gateway, an ad hoc network can bridge to wired LANs or to the Internet. It facilitates the integration of wireless grids and traditional grids.

Discovering and maintaining available routes is a key issue of wireless ad hoc networks. Although there are several routing protocols (such as AODV, DSR, ZIP, AOMDV et al.) [79], it still needs more efforts to provide adaptable routes for resource discovery in wireless grids. Traditional wired grids suffered the problems caused by heterogeneous resources and requirement changes. To solve the problems, many efforts have been made on traditional grid middleware. Regarding wireless grids, the problems are more serious, due to the dynamic characteristics. Meanwhile, the wireless grid middleware faces other challenges, for example, limited power, high-latency connectivity and unstable wireless connection.

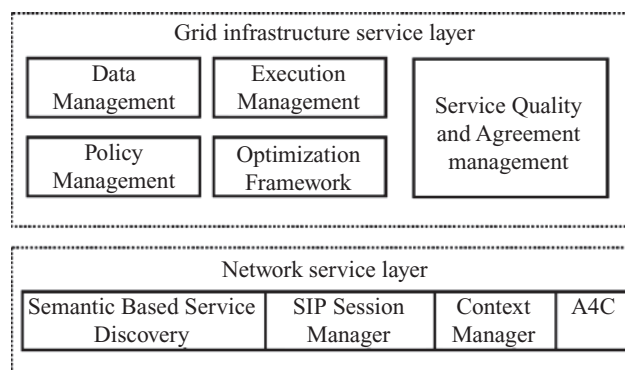In the following we review the typical work on wireless grid middleware.

Figure 1.24   Akogrimo architecture.

Akogrimo, supported by the FP6-IST program, enables access to knowledge through wireless grids. Some systems of *e*-learning, disaster management or *e*-health used the middleware Akogrimo.

Akogrimo supports wireless grids with mobile grid clients. It provides fixed, nomadic and mobile users with mobile grid services. On the basis of Akogrimo, grid services can be composed dynamically in an ad hoc way. Figure 1.24 presents the architecture of Akogrimo.

The middleware consists of two logical layers: network service layer and grid infrastructure service layer. The network service layer is constructed on a mobile IPv6 based infrastructure. It provides network services like Authentication, Authorization, Accounting, Auditing and Charging (*A4C*). With a *QoS* broker, the network service layer enables context driven selection of different bandwidth bundles. On the basis of the network service layer, the grid infrastructure service layer provides service oriented facilities for wireless grid applications [80]. It includes the business partner management, service quality management, business process execution management and mobile virtual organization management.

Akogrimo adapts many traditional grid technologies (such as virtual organization management, grid workflow) to wireless mobile environments, and enables the awareness of the mobility of virtual organization members.

MORE is the middleware which adapts web services for embedded systems. The middleware facilitates construction of mobilized enterprise information systems.

It combines technologies of embedded systems and web services to facilitate the integration of heterogeneous devices. It hides the complexity of heterogeneous embedded systems with unified and simplified interfaces. It also supports scalable group communication, where the small devices such as sensors and smart phones can connect with each other through wireless networks. With *MORE*, application developers can deploy ubiquitous services on embedded devices, and share the services within a group.

The key blocks of *MORE* include core management service and application enabling services. The core management service adopts group management concept, and uses XML-based policies to create and configure groups to control the behavior of group participants. The enabling services implement the service functionalities and provide fundamental supports for the applications [81].

It uses connectors for the communication between the enabling services. The connectors include internal connectors, proprietary connectors, SOAP connectors and *μSOA* connectors. The proprietary connectors and *μSOA* connectors support the embedded device involved services. The *μSOA* connectors enhance the efficiency of mobile embedded devices. By the *μSOA* approaches, *MORE* reduces the resource consumption in wireless grids.

MoGrid is the middleware, which supports grid services in wireless ad-hoc networks. The middleware has much application potential in mobile collaborations. It orchestrates the distributed grid tasks among mobile devices in a *P2P* manner, and uses a resource discovery protocol referred to as *P2PDP* to coordinate tasks among the most resourceful and available mobile devices. *MoGrid* offers
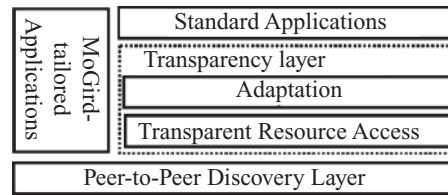
Figure 1.25   MoGrid architecture.

application-level mobility transparency, and supports the development of context-sensitive applications for mobile collaborations [72].

It leverages the resource coordination in wireless ad hoc networks by allocating the tasks for a group of mobile collaborators. The coordination consists of two phases. First, the *P2PDP* protocol discovers the resources that the coordination needs. Second, a collaborator submits the tasks to the participants according to the available resources they have [72].

As shown in Figure 1.25, the *MoGrid* architecture comprises a *P2P* discovery layer and a transparency layer. The applications are categorized into standard applications and *MoGrid*-tailored applications. The *P2P* discovery layer includes entities of collaborators, coordinators and initiators, and supports the resource registration and announcement, context definition, and resource discovery. After the devices register resources in *MoGrid*, they become collaborators. The registered resources are accessible to other devices in the *MoGrid*-enabled wireless grids. The initiators request the collaborators for task processing. The coordinators broadcast the request of initiators for resource discovery and coordinate the collaborators and initiators to complete the tasks. The transparency layer is responsible for the resource coordination among collaborators. It consists of two sub-layers: adaptation sub-layer and transparency resource access sub-layer (TRAS). The latter provides application-independent transparency for resource utilization. The adaptation sub-layer handles the mobility and connection-related events for each specific application.

MoGrid handles the issues of resource discovery and collaboration in wireless ad hoc networks. But its resource discovery protocol *P2PDP* should be improved further, in order to solve the problems caused by device mobility and transparency contexts [7].

MiPeG middleware enhances classic grid environments with wireless mobile mechanism. This middleware can be used in pervasive grid applications.

It enables integrating traditional grids and wireless devices. With a group of fundamental services, it enables deploying grid service clients on wireless mobile devices, and provides context-awareness for the applications. In addition, it can allocate user's tasks to different mobile devices also. In a *MiPeG*-powered grid, the mobile devices can be either active resources or grid service clients. When the *MiPeG*-powered mobile devices enter, the grid can automatically recognize the devices and coordinate them in applications.

MiPeG is a service-oriented middleware. As shown in Figure 1.26, it consists of a set of basic services which are compliant with the OGSA specifications.

In *MiPeG*, the asynchronous communication broker implements the WS-Brokered Notification specification, and provides an interaction mechanism based on the event publish-subscribe paradigm. It
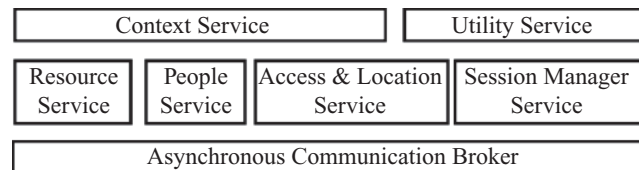

Figure 1.26   *MiPeG architecture*.

dispatches asynchronous messages in the grid. The resource service manages the grid resource registration, and enables the integration of wireless devices (such as PDAs and sensors) in the grid. The people service provides the mobile users with basic authentication mechanisms. The access and location service supports accessing to 802.11-enabled mobile devices. In addition, it can locate the position of the mobile devices. The session manager service handles the sessions for mobile users. The context service offers the context information including the state of resources, mobile user locations and users' profiles. The utility service implements the utility computing model and accepts the tasks submitted by users for execution. With the utility service, the users can complete a task without considering the issues of *reserving/releasing* resources et al.

MiPeG focuses mainly on the integration of mobile devices and traditional grids. It does not solve the problems of dynamic device changes well. In addition, *MiPeG* uses service-oriented technology, but its support for deploying the services on the resource-constrained mobile devices is still weak.

Applications of wireless grid technologies have broad prospects in many fields. This includes wireless sensor grids, environment data collection, *e*-health, e-learning, wireless-linked device integration and so on.

Wireless sensor grids combine wireless grids with sensor networks. It facilitates the transfer of information about the physical world around us to a plethora of web-based information utilities and computational services [76]. Wireless sensor grids can be used widely in disaster early warning and environment surveillance et al.

GridStix [82], Equator [83] and Floodnet [84] are three systems that use technologies of sensors, wireless networks and grids to monitor flood and provide early warning. In these systems, a variety of sensors equipped with wireless networking technology form a lightweight grid which is capable of collecting and transmitting data about flood disaster [82]. The lightweight sensor grid connects with traditional grid infrastructure which analyses the data gathered by the sensors for flood warning.

There are also wireless grid based supply-chain management system, the Smart Warehouse system. The Smart Warehouse is a prototype of a supply chain management system equipped with a wireless sensor grid [76]. With the smart sensor, it can track the items in the warehouse and identify problems in the supply chains.

e-Healthcare wireless grids can integrate a variety of medical devices (such as vital signs sensors) and provide a personalized healthcare service delivery paradigm which always requires a one-to-one connection between the patient and the medical expert [85]. Besides the personalized services, the wireless grid based *e*-healthcare systems can integrate the healthcare services into the patient's environment, forming a network of mobile and stationary monitoring and diagnosis facilities around the patient. By these means, medical professionals can establish an inter-organizational virtual hospital, and provide efficient detection of health emergencies and pervasive healthcare.

The Heart Monitoring and Emergency Service (HMES) system is an Akogrimo-based *e*-healthcare platform. It integrates the patient monitoring, emergency detection and subsequent rescue management, and enables an early recognition of heart attacks or apoplectic strokes and a proper treatment of patients as fast as possible [74].

Besides HMES, there are other wireless grid powered *e*-healthcare systems, for example, MobiHealth [86], Rural *e*-Healthcare Framework [87] and Hourglass [77].

e-Learning wireless grid technologies facilitates *e*-learning. Akogrimo enables almost universal availability of mobile devices to education and training. In an Ahogrimo-based *e*-learning system, the learners can use handheld devices to get learning contents from the wireless grid. And they can engage them in purposeful activity, problem solving, collaborations, interactions and conversations [88].

e-Campus system is another example of wireless grid powered *e*-learning. It adopts wireless grid technologies and supports mobile learning [89]. Based on grid service technology, the mobile learning combines traditional grids with mobile devices, and facilitates the sharing of learning resources distributed on different *e*-Learning platforms [75].

Wireless-linked device integration focuses on integrating wireless-linked small devices. *IMOGA* supports integrating mobile devices into a grid system, and sharing/producing information in a cooperative manner [76]. In IMOGA, the mobile devices, such as PDAs and smart phones, employ different kinds of

sensors to collect data, such as Global Positioning System (GPS), temperature monitoring components, health monitoring systems and pollution monitoring components. The gathered data is shared between huge numbers of mobile devices and high capacity grid infrastructures where the developers can construct applications of *e*-healthcare, *e*-crisis management and *e*-government [76].

References

[1] Adachi, F. (2002) Evolution towards broadband wireless systems, *The 5th International Symposium on Wireless Personal Multimedia Communications*, **1**, 27–30 Oct. 2002, 19–26.

[2] Jun-Zhao Sun, Sauvola, J. and Howie, D. (2001) Features in future: 4G visions from a technical perspective, *IEEE Global Telecommunications Conference. GLOBECOM '01*. 25–29. Nov. 2001, **6**, 3533–3537.

[3] Axiotis, D. I., Lazarakis, F. I., Vlahodimitropoulos, C. and Chatzikonstantinou, A. (2002) 4G system level simulation parameters for evaluating the interoperability of MTMR in UMTS and HIPERLAN/2, *4th International Workshop on Mobile and Wireless Communications Network, 2002*, 9–11 Sept., 559–563.

[4] Mihovska, A., Wijting, C., Prasad, R., Ponnekanti, S., Awad, Y. and Nakamura, M. (2002) A novel flexible technology for intelligent base station architecture support for 4G systems, *The 5th International Symposium on Wireless Personal Multimedia Communications, 2002*, **2**, 27–30 Oct., 601–605.

[5] Kitazawa, D., Chen, L., Kayama, H. and Umeda, N. (2002) Downlink packet-scheduling considering transmission power and QoS in CDMA packet cellular systems, *4th International Workshop on Mobile and Wireless Communications Network*, 9–11 Sept., 183–187.

[6] Dell'Uomo, L. and Scarrone, E. (2002) An all-IP solution for QoS mobility management and AAA in the 4G mobile networks, *The 5th International Symposium on Wireless Personal Multimedia Communications*, **2**, 27–30 Oct, 591–595.

[7] Wallenius, E. R. (2002) End-to-end in-band protocol based service quality and transport QoS control framework for wireless 3/4G services, *The 5th International Symposium on Wireless Personal Multimedia Communications*, 2, 27–30 Oct, 531–533.

[8] Benzaid, M., Minet, P. and Al Agha, K. (2002) Integrating fast mobility in the OLSR routing protocol, *4th International Workshop on Mobile and Wireless Communications Network*, 9–11 Sept., 217–221.

[9] Kambourakis, G., Rouskas, A. and Gritzalis, S. (2002) Using SSL/TLS in authentication and key agreement procedures of future mobile networks, *4th International Workshop on Mobile and Wireless Communications Network*, 9–11 Sept., 152–156.

[10] van der Schaar, M. and Meehan, J. (2002) Robust transmission of MPEG-4 scalable video over 4G wireless networks, *2002 International Conference on Image Processing. 2002. Proceedings*, **3**, 24–28 June, 757–760.

[11] Qing-Hui Zeng, Jian-Ping Wu, Yi-Lin Zeng, Ji-Long Wang and Rong-Hua Qin (2002) Research on controlling congestion in wireless mobile Internet via satellite based on multi-information and fuzzy identification technologies, *Proceedings. 2002 International Conference on Machine Learning and Cybernetics*, **4**, 4–5 Nov., 1697–1701.

[12] 5th International Symposium on Wireless Personal Multimedia Communications. Proceedings (Cat. No. 02EX568), *The 5th International Symposium on Wireless Personal Multimedia Communications, 2002*, **1**, 27–30 Oct.

[13] Sukuvaara, T., Mahonen, P. and Saarinen, T. (1999) Wireless Internet and multimedia services support through two-layer LMDS system, *1999 IEEE International Workshop on Mobile Multimedia Communications, (MoMuC '99)* 15–17 Nov., 202–207.

[14] Martin, C. C., Winters, J. H. and Sollenberger, N. R. (2000) Multiple-input multiple-output (MIMO) radio channel measurements, *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop 2000*, 16–17 March, 45–46.

[15] Pereira, J. M. (2000) Fourth generation: now, it is personal! *The 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000*, **2**, 18–21 Sept., 1009–1016.

[16] Otsu, T., Umeda, N. and Yamao, Y. (20010 System architecture for mobile communications systems beyond IMT-2000, *IEEE Global Telecommunications Conference, GLOBECOM '01*, **1**, 25–29 Nov., 538–542.

[17] Yi Han Zhang, Makrakis, D., Primak, S. and Yun Bo Huang (20020 Dynamic support of service differentiation in wireless networks *IEEE CCECE 2002*, *Canadian Conference on Electrical and Computer Engineering*, **3**, 12–15 May, 1325–1330.

[18] Jun-Zhao Sun and Sauvola, J. (2002) Mobility and mobility management: a conceptual framework, *10th IEEE International Conference on Networks, ICON 2002*, 27–30 Aug., 205–210.

[19] Vassiliou, V., Owen, H. L., Barlow, D. A., Grimminger, J., Huth, H.-P. and Sokol, J. (2002) A radio access network for next generation wireless networks based on multi-protocol label switching and hierarchical Mobile IP, *Proceedings of IEEE 56th Vehicular Technology Conference, 2002 VTC 2002-Fall*, **2**, 24–28 Sept., 782–786.

[20] Nicopolitidis, P., Papadimitriou, G. I., Obaidat, M. S. and Pomportsis, A. S. (2002) 3G wireless systems and beyond: a review, *9th International Conference on Electronics, Circuits and Systems, 2002*, **3**, 15–18 Sept., 1047–1050.

[21] 2002 IEEE Wireless Communications and Networking Conference Record, WCNC 2002 (Cat. No.02TH8609), *IEEE Wireless Communications and Networking Conference, WCNC2002*, **1**, 17–21 March.

[22] Borras-Chia, J. (2002) Video services over 4G wireless networks: not necessarily streaming, *2002 IEEE Wireless Communications and Networking Conference, WCNC2002*, **1**, 17–21 March, 18–22.

[23] Evans, B. G. and Baughan, K. (2000) Visions of 4G, *Electronics & Communication Engineering Journal*, **12**(6), 293–303.

[24] Kim, J. and Jamalipour, A. (2001) Traffic management and QoS provisioning in future wireless IP networks, *IEEE Personal Communications* [see also *IEEE Wireless Communications*], **8**(5), Oct., pp. 46–55.

[25] Aghvami, A. H., Le, T. H. and Olaziregi, N. (2001) Mode switching and QoS issues in software radio *IEEE Personal Communications*, [see also *IEEE Wireless Communications*], **8**(5), 38–44.

[26] Kanter, T. (2001) An open service architecture for adaptive personal mobile communication, *IEEE Personal Communications* [see also *IEEE Wireless Communications*], **8**(6), 8–17.

[27] Sampath, H., Talwar, S., Tellado, J., Erceg, V. and Paulraj, A. (2002) A fourth-generation MIMO-OFDM broadband wireless system: design, performance, and field trial results, *IEEE Communications Magazine*, **40**(9), 143–149.

[28] Kellerer, W. and Vogel, H.-J. (2002) A communication gateway for infrastructure-independent 4G wireless access, *IEEE Communications Magazine*, **40**(3), 126–131.

[29] Huang, V. and Weihua Zhuang (2002) QoS-oriented access control for 4G mobile multimedia CDMA communications, *IEEE Communications Magazine*, **40**(3), 118–125.

[30] Smulders, P. (2002) Exploiting the 60 GHz band for local wireless multimedia access: prospects and future directions, *IEEE Communications Magazine*, **40**(1), 140–147.

[31] Raivio, Y. (2001) 4G-hype or reality, *Second International Conference on 3G Mobile Communication Technologies, 2001.* (Conf. Publ. No. 477), 26–28 March, pp. 346–350.

[32] Becchetti, L., Delli Priscoli, F., Inzerilli, T., Mahonen, P. and Munoz, L. (2001) Enhancing IP service provision over heterogeneous wireless networks: a path toward 4G, *IEEE Communications Magazine*, **39**(8), 74–81.

[33] Abe, T., Fujii, H. and Tomisato, S. (2002) A hybrid MIMO system using spatial correlation, *The 5th International Symposium on Wireless Personal Multimedia Communications*, **3**, 27–30 Oct., 1346–1350.

[34] Lincke-Salecker, S. and Hood, C. S. (2003) A supernet: engineering traffic across network boundaries, *36th Annual Simulation Symposium*, 30 March–2 April, 117–124.

[35] Yamao, Y., Suda, H., Umeda, N. and Nakajima, N. (2000) Radio access network design concept for the fourth generation mobile communication system, *Proceedings of IEEE 51st Vehicular Technology Conference*, *VTC 2000-Spring Tokyo*, **3**, 15–18 May, 2285–2289.

[36] Ozturk, E. and Atkin, G. E. (2001) Multi-scale DS-CDMA for 4G wireless systems, *IEEE Global Telecommunications Conference, GLOBECOM '01*, **6**, 25–29, Nov., 3353–3357.

[37] Dell'Uomo, L. and Scarrone, E. (2001) The mobility management and authentication/authorization mechanisms in mobile networks beyond 3G, *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, **1**, 30 Sept.–3 Oct., C-44–C-48.

[38] Kumar, K. J., Manoj, B. S., Murthy, C. S. R. (2002) On the use of multiple hops in next generation cellular architectures, *10th IEEE International Conference on Networks, 2002. ICON 2002.* 27–30 Aug, 283–288.

[39] Wang, S. S., Green, M. and Malkawi, M. (2002) Mobile positioning and location services, *IEEE Radio and Wireless Conference, RAWCON 2002*, 11–14 Aug., 9–12.

[40] Motegi, M., Kayama, H. and Umeda, N. (2002) Adaptive battery conservation management using packet QoS classifications for multimedia mobile packet communications, *Proceedings of IEEE 56th Vehicular Technology Conference, VTC 2002-Fall*, **2**, 24–28 Sept., 834–838.

[41] Ying Li, Shibua Zhu, Pinyi Ren and Gang Hu (2002) Path toward next generation wireless internet -cellular mobile 4G, WLAN/WPAN and IPv6 backbone, *Proceedings of 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering TENCOM '02*, **2**, Oct. 28–31, 1146–1149.

[42] Qiu, R. C., Wenwu Zhu and Ya-Qin Zhang (2002) Third-generation and beyond (3.5G) wireless networks and its applications, *IEEE International Symposium on Circuits and Systems, ISCAS 2002*, **1**, 26–29 May, I-41–I-44.

[43] Bornholdt, C., Sartorius, B., Slovak, J., Mohrle, M., Eggemann, R., Rohde, D. and Grosskopf, G. (2002) 60 GHz millimeter-wave broadband wireless access demonstrator for the next-generation mobile internet, *Optical Fiber Communication Conference and Exhibit, OFC 2002*, 17–22 March, 148–149.

[44] Jianhua He, Zongkai Yang, Daiqin Yang, Zuoyin Tang and Chun Tung Chou (2002) Investigation of JPEG2000 image transmission over next generation wireless networks, *5th IEEE International Conference on High Speed Networks and Multimedia Communications*, 3–5 July, 71–7.

[45] Baccarelli, E. and Biagi, M. (2003) Error resistant space-time coding for emerging 4G-WLANs, *IEEE Wireless Communications and Networking, WCNC 2003*, **1**, 16–20 March, 72–77.

[46] Mohr, W. (2002) WWRF – the Wireless World Research Forum *Electronics & Communication Engineering Journal*, **14**(6), Dec., 283–291.

[47] Otsu, T., Okajima, I., Umeda, N. and Yamao, Y. (2001) Network architecture for mobile communications systems beyond IMT-2000, *IEEE Personal Communications*, [see also *IEEE Wireless Communications*], **8**(5), Oct., 31–37.

[48] Bria, A., Gessler, F., Queseth, O., Stridh, R., Unbehaun, M., Jiang Wu; Zander, J., and Flament, M. (2001) 4th-generation wireless infrastructures: scenarios and research challenges, *IEEE Personal Communications* [see also *IEEE Wireless Communications*], **8**(6), Dec., 25–31.

[49] Fitzek, F., Kopsel, A., Wolisz, A., Krishnam, M. and Reisslein, M. (2002) Providing application-level QoS in 3G/4G wireless systems: a comprehensive framework based on multirate CDMA, *IEEE Wireless Communications* [see also *IEEE Personal Communications*], **9**(2), April, 42–47.

[50] Classon, B., Blankenship, K., and Desai, V. (2002) Channel coding for 4G systems with adaptive modulation and coding, *IEEE Wireless Communications* [see also *IEEE Personal Communications*], **9**(2), April, 8–13.

[51] Yile Guo and Chaskar, H. (2002) Class-based quality of service over air interfaces in 4G mobile networks, *IEEE Communications Magazine*, **40**(3), 132–137.

[52] Cassioli, D., Win, M. Z. and Molisch, A. F. (2002) The ultra-wide bandwidth indoor channel: from statistical model to simulations, *IEEE Journal on Selected Areas in Communications*, **20**(6), 1247–1257.

[53] Win, M. Z. and Scholtz, R. A. (2002) Characterization of ultra-wide bandwidth wireless indoor channels: a communication-theoretic view, *IEEE Journal on Selected Areas in Communications*, **20**(9), 1613–1627.

[54] Glisic, S. (2004) *Advanced Wireless Communications, 4G Technology*, John Wiley and Sons, London.

[55] Kempf, J., Nikander, P. and Green, H. (2010) Innovation and the Next Generation Internet, *INFOCOM IEEE Conference on Computer Communications Workshops*, 1–6.

[56] Christensen, C. (2003) *The Innovator's Dilemma*, Harper Business, New York.

[57] Christensen, C. and Raynor, B. (2003) *The Innovator's Solution*, Harvard Business School Publishing Corp., Boston.

[58] Christensen, C., Anthony, S. and Roth, E. (2004) *Seeing What's Next*, Harvard Business School Publishing Corp., Boston.

[59] Armburst, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwmski, A., Lee, G., Patterson, D., Rabkm, A., I. Stoica and Zaharia, M. (2009) Above the clouds: a Berkeley view of cloud computing, Technical Report No. UCB/EECS-2009-28, EE-CS Dept., University of California Berkeley, February 10.

[60] Synaptic Hostmg Enterprise, www.business.att.com/enterprise/Family/application-hosting-enterprise/synaptic-hosting-enterprise/.

[61] Rosen, E., Viswanathan, A. and Callon, R. (2001) Multiprotocol label switching architecture, RFC 3031, Internet Engineering Task Force, January.

[62] Green, H., Monette, S., Olsson, J., Saltsidis, P. and Takács, A. (2007) Carrier Ethemet: The native approach, *Ericsson Review*, **3**.

[63] Farrel, A. and Bryskm, I. (2006) *GMPLS: Architecture and Applications*, Morgan Kaufman, San Francisco.

[64] Jacobson, V. (2006) in Stanford clean slate seminar, September, www.cleanstate.stanford.edu/seminars/jacobson.pdf.

[65] Rangan, K. (2008) The Cloud Wars: $100+ billion at stake. Tech. Rep., Merrill Lynch, May.

[66] Siegele, L (2008) Let It Rise: A Special Report on Corporate IT, *The Economist* (October).

[67] Gray, J., and Patterson, D. (2003) A conversation with Jim Gray, *ACM Queue* **1**(4), 8–17.

[68] Foster, I., Kesselman, C. and Tueck, S. (2001) The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of Supercomputer Applications*, **15**(3), 200–222.

[69] McKnight, L. W., Howison, J. and Bradner, S. (2004) Wireless Grids: Distributed Resource Sharing by Mobile, Nomadic, and Fixed Devices, *IEEE Internet Computing*, **8**(4), 24–31.

[70] Ahuja, S. P., and Myers, J. A. (2006) A Survey on Wireless Grid Computing, *Journal of Supercomputing*, **37**(1), 3–21.

[71] Gaynor, M., Moulton, S. L., Welsh, M., LaCombe, E., Rowan, A. and Wynne, J. (2004) Integrating Wireless Sensor Networks with the Grid, *IEEE. Internet Computing*, **8**(4), 32–39.

[72] Lima, L. M., Gomes, A. T. A., Ziviani, A., Endler, M., Soares, L. F. G. and Schulze, B. (2005) Peer-to-Peer Resource Discovery in Mobile Grid, *Proceedings of the 3rd International Workshop on Middleware for Grid Computing*, ACM, New York, 1–8.

[73] Coronato, A. and De Pietro, G. (2008) MiPeG: A middleware infrastructure for Pervasive grids, *Future Generation Computer Systems*, **24**(1), 17–29.

[74] Liao, C., Chi, Y., and Ou Yang, F. (2003) A Grid Service Oriented Platform for Mobile Learning of English, *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003*, 2265–2268.

[75] Ozturk, E. and Altilar, D. T. (2007) IMOGA: An Architecture for Integrating Mobile Devices into Grid Applications, *Proceedings of Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking& Services*, 1–8.

[76] Gaynor, M., Moulton, S. L., Welsh, M., LaCombe, E. and Rowan (2004) Integrating Wireless Sensor Networks with the Grid, *IEEE Internet Computing*, **8**(4), 32–39.

[77] McKnight, L. W. (2007) The future of the internet is not the internet: open communications policy and the future wireless grid(s), http//www.cecd.org/dataoecd/18/42/3 8057172 pdf.

[78] Li, G., Han, Y. B., Wang, J., Zhao, Z. F. and Wagner, R. M. (2006) Facilitating Dynamic Service Compositions by Adaptable Service Connectors, *International Journal of Web Services Research*, **3**(1), 68–84.

[79] Glisic, S. and Lorenzo, B. (2009) *Advanced Wireless Networks, 4G technology*, 2e John Wiley and Sons, London.

[80] Jähnert, J. M., Wesner, S. and Villagrá, V. A. (2008) The Akogrimo Mobile Grid Reference Architecture-Overview, Whitepaper, www.akogrimo.org/download/White_Papers_and_Publications/Akogrimo_WhitePaper_Architecture_v1-1.pdf.

[81] Wolff, A., Michaelis, S., Schmutzler, J. and Wietfeld, C. (2007) Network-centric Middleware for Service Oriented Architectures across Heterogeneous Embedded Systems, *Proceedings of the Eleventh International IEEE EDOC Conference Workshop*, IEEE Press, Washington, 105–108.

[82] Hughes, D., Greenwood, P., Coulson, G. and Blair, G. (2006) GridStix: Supporting Flood Prediction using Embedded Hardware and Next Generation Grid Middleware, *Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, IEEE Press, Washington.

[83] Humble, J., Greenhalgh, C., Hampshire, A., Muller, H. L. and Egglestone, S. R. (2005) A Generic Architecture for Sensor Data Integration with the Grid, *Proceedings of Scientific Applications of Grid Computing 2004*, Springer-Verlag, Berlin.

[84] Roure, D. D. Improving Flood Warning times using Pervasive and Grid Computing, http://envisense.org/floodnet/ingenia/ingenia.htm.

[85] Daramola, J. O., Osamor, V. C. and Oluwagbemi, O. O. (2008) A Grid-Based Framework for Pervasive Healthcare Using Wireless Sensor Networks: a Case for Developing Nations, *Asia Journal of Information Technology*, **7**(6), 260–267.

[86] Mei, H., Widya, I. A., van Halteren, A. T. and Erfianto, B. (2006) A Flexible Vital Sign Representation Framework for Mobile Healthcare, *Proceedings of the 1st International Conference on Pervasive Computing Technologies for Healthcare*, IEEE Press, Washington, 1–9.

[87] Oludayo, O. Olugbara, Sunday, O., Ojo, Mathew, O., Adigun, Justice, O. (2007) Emuoyibofarhe and Sibusiso S. Xulu, An Architectural Framework for Rural e-Healthcare Information Infrastructure with Web Service-Enabled Middleware Support, *Proceedings of HELINA 2007*.

[88] Laria, G. (2008) Mobile and nomadic user in e-learning: the Akogrimo case, Whitepaper, http://www.akogrimo.org/download/White_Papers_and_Publications/Akogrimo_WhitePaper_eLearning.pdf.

[89] Chang, R. S. and Chang, J. H. (2006) E-campus system based on the pervasive grid, *Proceedings of TANET 2006*.