Emerging Technologies for 4G

4.1 Introduction

OFDM-based access techniques are the most appropriate candidate for 4G, as it has been discussed in Chapter 3. In order to provide the required quality of services to users of 4G, several key technologies will be used with OFDM. Therefore, the next task is to present and update information on key technologies, namely MIMO technology, radio resource management, software defined radio (SDR) communication system, mobile IP, and relaying techniques [1-175]. This chapter starts with the multiantenna technologies in Section 4.2. Multiantenna is one of the key technologies for mitigating the negative effects of the wireless channel, providing better link quality and/or higher data rate without consuming extra bandwidth or transmitting power. The use of multiple antennas at either receiver, transmitter, or both ends provides several benefits: array gain, interference reduction, diversity gain, and/or multiplexing gain. Section 4.3 focuses on the main radio resource management functions. They constitute a fundamental aspect for the provision of a variability degree of the quality of service and the effective explanation of the limited radio resource, thus representing a hot research area. The concept of SDR is one of the bridges between software programming and real hardware implementation. Section 4.4 defines radio communication systems and summarizes the advantages of SDR. In addition, this section discusses the problems that must be overcome to realize the SDR communication system and explains the remarkable technologies developed to realize such a system. Section 4.5 deals with IP network issues: Mobile IP architecture, proposed for supporting mobility in Internet, is presented together with guidelines behind mobility. Section 4.6 introduces the concepts of relaying techniques for 4G, and other enabling techniques are listed in Section 4.7.
4.2 Multiantenna Technologies

Recent years have witnessed an explosive growth of interest in studying and using multiple antenna techniques for wireless communication systems. Such a trend is motivated by the fact that several of the specifications foreseen for future wireless systems appear to be difficult, if not impossible, to fulfill with conventional single antenna systems. It is widely accepted that multiple antennas have the potential to increase the achievable data throughput, to enhance link quality (BER, QoS), and to increase cell coverage and network capacity, among others. Such a promising array of enhancements has contributed to speeding up the development in the field, both at academic levels, where countless techniques are developed, and in the industry, where solutions based on these techniques are being rapidly adopted in real systems.

Multiple transmit and receive antennas can be combined with various multiple access techniques such as TDMA, CDMA, and OFDM to improve the capacity and reliability of communications. Multiple-input multiple-output (MIMO) communication systems are regarded as an effective solution for future high-performance wireless networks. The use of multiple antennas at transmitter and receiver, popularly known as MIMO, is a promising cost-effective technology that offers substantial leverages in making the anticipated 1-Gbps wireless links a reality.

Several efforts are currently underway to build non-line-of-sight (NLOS) broadband wireless systems. A MIMO wireless system (physical layer and MAC layer technology) using OFDM modulation for NLOS environments was successfully developed by NTT DoCoMo. In mobile access, there is an effort under the ITU working group to integrate MIMO techniques into the high-speed downlink packet access (HSDPA) channel, which is part of the Universal Mobile Telecommunications System (UMTS) standard. Lucent Technologies recently announced a chip for MIMO enhancement of UMTS/HSDPA but has released no further details. Preliminary efforts are also underway to define a MIMO overlay for the IEEE 802.11 standard for WLANs under the newly formed Wireless Next Generation (WNG) group. Moreover, many other companies also proposed advanced MIMO schemes for the IEEE 802.16 standard.

In this chapter, we provide an overview of MIMO technology and explain recent research directions and results.

4.2.1 Overview of MIMO Technology

Depending on the geometry of the employed antenna array, two basic multiantenna approaches can be considered: a beamforming approach for closely separated antenna elements (interelement separation is at most $\lambda/2$, where $\lambda$ is the carrier wavelength) or a diversity approach for widely separated
antenna elements (typical interelement spacing is at least a few $\lambda$). In this chapter, we will explore the latter approach where the fading processes associated with any two possible transmit-receive antenna pair can be assumed to be independent. The fact that a MIMO system consists of a number of uncorrelated concurrent channels has been exploited from two different perspectives. First, from a pure diversity standpoint, one can enhance the fading statistics of the received signal by virtue of the multiple available replicas being affected by independent fading channels. By sending the same signal through parallel and independent fading channels, the effects of multipath fading can be greatly reduced, decreasing the outage probability and hence improving the reliability of the communication link [1]. In the second approach, referred to as spatial multiplexing [2], different information streams are transmitted on parallel spatial channels associated with the transmit antennas. This could be seen as a very effective method to increase spectral efficiency. In order to be able to separate the individual streams, the receiver has to be equipped with at least as many receive antennas as the number of parallel channels generated by the transmitter in general. For a given multiple antenna configuration, one may be interested in finding out which approach would provide the best possible or desired performance.

4.2.1.1 Diversity

Space-time coding (STC) is a hybrid technique that uses both space and temporal diversity in a combined manner. There are two forms of STC namely space-time block code (STBC) and space-time trellis code (STTC). STBC efficiently exploits transmit diversity to combat multipath fading while keeping decoding complexity to a minimum. Tarokh et al. [3] showed that no STBC can achieve full-rate and full-diversity for more than two transmit antennas, and proposed a 3/4 rate, full-diversity code for four transmit antennas. A full-rate quasi-orthogonal (QO) STBC was proposed by Jafarkhani [4] for four transmit antennas based on Alamouti orthogonal STBC [1]. In this case, the transmission matrix is given by

$$
C = \begin{bmatrix}
A_{12} & A_{34} \\
-A_{34} & A_{12}^*
\end{bmatrix} =
\begin{bmatrix}
x_1 & x_2 & x_3 & x_4 \\
x_2^* & x_1^* & -x_4^* & x_3^* \\
x_3^* & -x_4^* & x_1^* & x_2^* \\
x_4 & -x_3^* & -x_2^* & x_1
\end{bmatrix}
$$

(4.1)

where $A_{12}, A_{34}$ are the Alamouti codes [1]. It is noted here that since the channel matrix of the QO-STBC is not full-rank, full-diversity gain cannot be attained. To achieve the full-diversity and full-rate (FDFR) property, a new FDFR STC approach was recently proposed.
4.2.1.2 Mixed (Hybrid Diversity and Spatial Multiplexing)

This mode combines diversity and spatial multiplexing by transmitting from four transmit antennas, each space-time block coded with the basic Alamouti scheme of order two [1]. The transmission matrix of the space-time block coding for the \( i \)th data stream, \( i = a, b \) is given by

\[
A_i = \begin{bmatrix}
  x_1(i) & x_2(i) \\
  -x_2(i)^* & x_1(i)^*
\end{bmatrix}
\]  

(4.2)

To decode the data, minimum mean square error (MMSE) and zero forcing (ZF) receivers can be employed. For the MMSE receiver, we assume that the transmitted matrix is \( \left[ a_{n1}(k), a_{n1+1}(k), b_{n2}(k), b_{n2+1}(k) \right]^T \), where \( a \) and \( b \) indicate different signal streams. First, the tap weight vector and decoding layer order are determined. If the first decoding layer is \( a \), the procedure can be represented by

\[
\begin{bmatrix}
  \hat{a}_{n2}(k) \\
  \hat{a}_{n2+1}(k)
\end{bmatrix} = \text{decision}\left\{ \begin{bmatrix}
  w^H_i(k) \\
  w^H_2(k)
\end{bmatrix} \cdot y_n(k) \right\}
\]  

(4.3)

The interference from the original signal can be subtracted using \( \hat{a}_{n2}(k) \) and \( \hat{a}_{n2+1}(k) \); accordingly, the other stream can be decoded as follows:

\[
\begin{bmatrix}
  y'_n(k) = y_n(k) - \left[ h_1(k) \right. \\
  \left. h_2(k) \right] \cdot \hat{a}_{n2+1}(k)
\end{bmatrix}
\]

(4.4)

Note that for comparative purposes we can also employ maximum likelihood (ML) decoding (explained in next section) to obtain the optimum performance, which was used as our baseline reference.

4.2.1.3 Spatial Multiplexing

In this section, we briefly review the spatial multiplexing scheme. The vertical Bell Lab layered space time (V-BLAST) architecture has been recently proposed for achieving high spectral efficiency over wireless channels characterized by rich scattering [2]. In this approach, one way of detection is to use conventional adaptive antenna array (AAA) techniques (i.e., linear combining and nulling). Conceptually, each stream (i.e., layer) in turn is considered to be the desired signal, while regarding the remaining signals as interference. Nulling is performed...
by linearly weighting the received signals so as to satisfy some performance related criterion, such as ZF or MMSE. This linear nulling approach is viable, but superior performance is obtained if nonlinear techniques are used. One particularly attractive nonlinear alternative is to exploit symbol cancellation as well as linear nulling to perform detection. By using symbol cancellation, the interference from the already-detected components is subtracted from the received signal vector, effectively reducing the overall interference. Here we will consider ordered successive interference cancellation with ZF and MMSE. Also, a ML decoding receiver will be used as a reference.

It is assumed that the \( H_{ij}(k) \) is the channel coefficient from \( j \)th transmit antenna to \( i \)th receive antenna and \( w \) is white Gaussian noise with covariance matrix \( C_w = \mathbb{E}[ww^H] = \sigma^2 I_{N_R} \) where \( N_R \) is the number of received antennas. Then, the received signal vector can be written as follows:

\[
y_n(k) = H(k)x_n(k) + w(k)
\]  

(4.5)

where the index \( k \) denotes the \( k^{th} \) subcarrier, \( y(k) = [y_1(k) \cdots y_{N_R}(k)]^T \), \( x(k) = [x_1(k) \cdots x_{N_R}(k)]^T \), and \( w(k) \) is the \((N_R \times 1)\) noise vector.

**Maximum Likelihood Decoding (Optimal Solution)**

The ML detection of \( x(k) \) can be found by maximizing the conditional probability density function and this is equivalent to minimizing the log-likelihood function:

\[
\hat{x}(k) = \min_{x(k)} \left\{ y(k) - Hx(k) \right\}^H \left\{ y(k) - Hx(k) \right\}
\]  

(4.6)

where, \( x(k) \in \) all possible constellation sets.

It is well known that ML decoding is characterized by a high implementation complexity, and thus, suboptimal but practically implementable solutions are considered next.

**Ordered Successive Interference Cancellation (OSIC)**

Instead of the ML decoding approach, linear detection techniques can be used (i.e., ZF and MMSE). To improve the linear detection techniques, we try to decode according to received signal strength, and extract the decoded signal from the received signal. This approach is referred to as D-BLAST or V-BLAST [2] according to the transmitted signal structure, where D stands for diagonal and V for vertical. For simplicity, we consider the OSIC.

The receiving operation of OSIC can be summarized as follows:
• Step 1: Compute the tap weight matrix $W$.
• Step 2: Find the layer with maximum SNR.
• Step 3: Detection

$$z_k(n) = W_k^H y(n)$$

$$\hat{x}_k(n) = \text{decision}[z_k(n)]$$

• Step 4: Interference cancellation

$$y(n) = y(n) - \hat{h}_k(n)$$

$$H = [h_1, \ldots, h_{k-1}, 0, h_{k+1}, \ldots, h_T]$$

• Step 5: Repeat Step 1 to 5 until all symbols are detected.

**Zero-Forcing.** The cost function can be expressed as

$$J_{ZF} = \{y(k) - H\hat{x}(k)\}^H \{y(k) - H\hat{x}(k)\}$$

(4.7)

Since $J_{ZF}$ is a convex function over $\hat{x}(k)$, $\hat{x}(k)$ can be determined by using the minimum limit. Then, the tap weight vector is given by

$$W = \left\{H^H H\right\}^{-1} H^H$$

(4.8)

Minimum Mean Square Error. To take into account the noise variance, the cost function can be expressed as

$$J_{MMSE} = E\left[\{y(k) - H\hat{x}(k)\}^H \{y(k) - H\hat{x}(k)\}\right]$$

(4.9)

Using a similar method as in the ZF detection method, the weight vector results in

$$W = \left\{H^H + \sigma^2 I\right\}^{-1} H^H$$

(4.10)

Note that the noise variance $\sigma^2$ has to be estimated in order to use the MMSE approach.
4.2.2 Adaptive Multiple Antenna Techniques

Recently some authors have considered the diversity-spatial multiplexing problem. In [5], the fundamental trade-off between diversity and spatial multiplexing is explored by Zheng and Tse. A scheme based on switching between diversity and spatial multiplexing is proposed by Heath and Paulraj [6]. Authors have considered a fixed rate system in which the receiver adaptively selects one of the two transmission approaches based on the largest minimum Euclidean distance of the received constellation. The receiver informs its selection to the transmitter via a 1-bit feedback channel. To ensure a fixed bit rate, the diversity scheme uses modulation with a higher order than that used by its counterpart spatial modulation case. Skjevling et al. presented a hybrid method combining both diversity and spatial multiplexing [7]. The proposed approach optimally assigns antennas to a given (fixed) transmission scheme combining diversity and spatial multiplexing. Antenna selection is based either on full channel feedback or long-term statistics. Gorokhov et al. studied the relationship between multiplexing gain and diversity gain in the context of antenna subset selection [8], thereby extending the recent result by Zheng and Tse [5].

4.2.3 Open-Loop MIMO Solutions

Alamouti developed a remarkable orthogonal FDFR code for $N_T = 2$ transmit antennas [1], requiring a simple linear decoder at the receiver. Tarokh et al. [3] proved that a FDFR orthogonal code only exists for $N_T = 2$ and proposed some space-time block codes for $N_T > 2$ attaining full-diversity but not full-rate. In [4] a quasi-orthogonal full-rate code is proposed by Jafarkhani, though full-diversity gain cannot be attained. Based on space-time constellation rotation, Xin et al. [9] and Ma et al. [10] proposed a FDFR encoder for an arbitrary number of transmit antennas. For an even number of transmit antennas, Jung et al. [11] obtained coding gain with a FDFR space-time block code by serially concatenating the Alamouti scheme with the constellation rotation techniques used in [9, 10]. Although the Alamouti-based space-time constellation rotation encoder (A-ST-CR) of [11] can effectively achieve full-diversity and full-rate, the decoding complexity is an issue and its practical implementation becomes prohibitive, even for a small number of transmit antennas (e.g., $N_T = 4$). This is in virtue of the high computational complexity required by the ML decoding algorithm.

In addressing the complexity problem, this chapter further extends the results of [11] by considering a system based on the serial concatenation of a new rotating precoding scheme with the basic Alamouti codes of order two. A proper process of puncturing and shifting after the actual constellation-rotation operation can conveniently decompose the encoding process into rotation operations carried out in a lower order matrix space. The impact of this puncture and shift
rotation coding scheme is very significant at the receiver, where, due to the provided signal decoupling, the decoding complexity is significantly reduced. It is shown in this chapter that the proposed method attains the same performance as the scheme presented in [11] with a substantial complexity reduction.

References [9-11] use a precoder based on the Vandermonde matrix for attaining a FDFR system. After multiplying the received signal $x$ by the Vandermonde matrix, each component of vector $r$ combines all the symbols as can be observed in the next basic precoder equation.

$$
 r = \Theta x = \frac{1}{\sqrt{4}} \begin{bmatrix}
 1 & \alpha_0^1 & \alpha_0^2 & \alpha_0^3 & x_1 \\
 1 & \alpha_1^1 & \alpha_1^2 & \alpha_1^3 & x_2 \\
 1 & \alpha_2^1 & \alpha_2^2 & \alpha_2^3 & x_3 \\
 1 & \alpha_3^1 & \alpha_3^2 & \alpha_3^3 & x_4 \\
\end{bmatrix}
 = \begin{bmatrix}
 r_1 \\
 r_2 \\
 r_3 \\
 r_4 \\
\end{bmatrix}
$$

(4.11)

where $\alpha_i = \exp\left(j2\pi(1+1/4)/N\right), \ i = 0, 1, \ldots, N - 1$.

Xin [9] and Ma [10] use a diagonal channel matrix after multiplying the information symbols by the Vandermonde matrix. This linear precoding is referred to as the constellation rotation operation. Notice that the coding advantage of [9, 10] is not optimized, although the schemes successfully achieve FDFR. Jung [11] improves the coding advantages by concatenating the constellation rotating precoder with the basic Alamouti scheme, resulting in the following transmitted signals:

$$
 S = \begin{bmatrix}
 r_1 & r_2 & 0 & 0 \\
 -r_2^* & r_1^* & 0 & 0 \\
 0 & 0 & r_3 & r_4 \\
 0 & 0 & -r_4^* & r_3^* \\
\end{bmatrix}
$$

(4.12)

At the receiving end, the signal can be written as

$$
 y = \begin{bmatrix}
 y_1 \\
 y_2^* \\
 y_3 \\
 y_4^* \\
\end{bmatrix}
 = \frac{1}{\sqrt{2}} \begin{bmatrix}
 h_1 & h_2 & 0 & 0 & r_1 \\
 h_2^* & -h_1^* & 0 & 0 & r_2 \\
 0 & 0 & h_3 & h_4 & r_3 \\
 0 & 0 & h_3^* & -h_4^* & r_4 \\
\end{bmatrix}
 + \begin{bmatrix}
 n_1 \\
 n_2^* \\
 n_3 \\
 n_4^* \\
\end{bmatrix}
 = Hr + n
$$

(4.13)

Note that since $r_1, r_2, r_3, r_4$ already sums over $x_1 \sim x_4$ through the Vandermonde matrix, each symbol experiences the channel twice. We can now point out that (4.12) can be separated into two parts: $(r_1, r_3$ and $r_2, r_4)$ or $(r_1, r_4$...
and \( r_2, r_3 \). Consequently, the Vandermonde matrix for the precoder need not be of size 4, but smaller. Based on this observation, we can use a puncturing and shifting operation after the constellation rotation process resulting in a new precoder.

\[
\begin{align*}
\mathbf{r}_{1,3} &= \begin{bmatrix} r_1 \\ r_3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \alpha_0^1 \\ 1 & \alpha_2^1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} \\
\mathbf{r}_{2,4} &= \begin{bmatrix} r_2 \\ r_4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \alpha_1^1 \\ 1 & \alpha_3^1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_4 \end{bmatrix}
\end{align*}
\]

or

\[
\begin{align*}
\mathbf{r}_{1,4} &= \begin{bmatrix} r_1 \\ r_4 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \alpha_0^1 \\ 1 & \alpha_3^1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_4 \end{bmatrix} \\
\mathbf{r}_{2,3} &= \begin{bmatrix} r_2 \\ r_3 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \alpha_1^1 \\ 1 & \alpha_2^1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}
\end{align*}
\]

(4.14)

After puncturing and shifting, the encoder can be defined as

\[
\begin{align*}
\begin{bmatrix} 1 & \alpha_0^1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_1^1 \\ 0 & 0 & 1 & \alpha_3^1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & \alpha_0^1 & 0 & 0 \\ 0 & 0 & 1 & \alpha_1^1 \\ 0 & 0 & 1 & \alpha_2^1 \\ 1 & \alpha_3^1 & 0 & 0 \end{bmatrix}
\end{align*}
\]

(4.15)

Recently, Zafar et al. proposed a low decoding complexity (symbol by symbol decoding) improved space-time code with full-diversity for three and four transmit antennas configurations [13]. The following is the format obtained after modifying the transmission matrix:

\[
\begin{bmatrix}
x_1 + jy_3 & -x_2 + jy_4 & 0 & 0 \\
x_2 + jy_4 & x_1 - jy_3 & 0 & 0 \\
0 & 0 & x_3 + jy_1 & -x_4 + jy_2 \\
0 & 0 & x_4 + jy_2 & x_3 - jy_1
\end{bmatrix}
\]

(4.16)

where \( x_i = s_{il} \cos \theta - s_{iq} \sin \theta, y_j = s_{il} \cos \theta - s_{iq} \sin \theta, \) and \( \theta = \tan^{-1}\left(\frac{1}{3}\right) \). The complex symbols \( s_i \) take values from a QAM signal set. Note that we already separated the encoder into two parts, so a Vandermonde matrix of order two should be used. The decoding computational complexity is significantly reduced compared to that of Jung’s in [11].

### 4.2.4 Closed-Loop MIMO Solutions

In this section we explain closed-loop MIMO solutions, which consist of two parts: antenna grouping and codebook based schemes, which use feedback information from a mobile station.
4.2.4.1 Antenna Grouping

The rate 1 transmission code for four transmit-antenna base stations in the IEEE 802.16e is

\[
A = \begin{bmatrix}
    s_1 & -s_2^* & 0 & 0 \\
    s_2 & s_1^* & 0 & 0 \\
    0 & 0 & s_3 & -s_4^* \\
    0 & 0 & s_4 & s_3^*
\end{bmatrix}
\] (4.17)

Note that this scheme does not achieve full-diversity.

The effective channel model \( H_{\text{eff}} \) is orthogonal and it can be written as follows:

\[
H_{\text{eff}}^H H_{\text{eff}} = \begin{bmatrix}
\rho_1 & 0 & 0 & 0 \\
0 & \rho_1 & 0 & 0 \\
0 & 0 & \rho_2 & 0 \\
0 & 0 & 0 & \rho_2
\end{bmatrix}
\] (4.18)

where \( \rho_1 = |h_1|^2 + |h_2|^2 \) and \( \rho_2 = |h_3|^2 + |h_4|^2 \), being \( h_k, k = 1, 2, 3, 4 \) the channel coefficient associated with the \( k \)th transmit antenna.

If the BS can use channel state information, the performance of the existing matrix \( A \) approaches the performance of the FDFR STC in section 4.2.3 by using the following equation:

\[
\arg \min_{\text{antenna pair}} |\rho_1 - \rho_2|
\] (4.19)

Let \( d_{\text{min}}^2 \) be the corresponding minimum distance of the normalized unit energy constellation. The \( 2^R - \)QAM Euclidean distance equation \( d_{\text{min}}^2 = 12/(2^R - 1) \) will be used, corresponding to QAM modulation for diversity. Using this Euclidean distance equation, we can estimate the error probability as

\[
P_e \leq N_e Q \left( \sqrt{ \frac{E_i}{N_0} d_{\text{min}}^2 } \right)
\] (4.20)

where \( d_{\text{min}}^2 \) is the squared Euclidean distance of the received signal, and \( N_e \) is the number of nearest neighbors in the constellation, which can be found for each proposed mapping scheme based on the channel coefficient matrix \( H \).
\[ Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right) \]

where \( \text{erfc} \) is the complementary error function. For STC, the minimum distance of the diversity constellation at the receiver can be shown to be

\[
d_{\min}^2(H) \leq \frac{\min\left(\|H\|_F^2(a,b), \|H\|_F^2(c,d)\right)}{N_T} d_{\min}^2
\]

(4.21)

where \((a, b)\) and \((c, d)\) are antenna grouping index and \(\|H\|_F\) is the Frobenius norm of matrix \(H\). The details for derivation follow the derivation procedure of the maximum SNR criterion for code design.

Figure 4.1 shows the system block diagram, which makes use of a grouper to select the antenna pair based on feedback channel information from the MS.

The performance of the proposed scheme is shown in Figure 4.2. At \(\text{BER} = 10^{-3}\) point, the proposed scheme outperforms the conventional STC without antenna grouping by 3.5 dB.

The rate 2 transmission code for four transmit antennas in the current IEEE 802.16 standard [1] is

![System block diagram](image-url)
Transmission matrix $B$ for rate 2 can be improved with antenna grouping information. The BS can group antennas 0 and 1 for the first diversity pair and antennas 2 and 3 for the second diversity pair. In matrix form, this can be expressed as follows:

$$B = \begin{bmatrix}
    s_1 & -s_2^* & s_5 & -s_7^* \\
    s_2 & s_1^* & s_6 & -s_8^* \\
    s_3 & -s_4^* & s_7 & s_5^* \\
    s_4 & s_3^* & s_8 & s_6^*
\end{bmatrix} \quad (4.22)$$

Transmission matrix $B$ for rate 2 can be improved with antenna grouping information. The BS can group antennas 0 and 1 for the first diversity pair and antennas 2 and 3 for the second diversity pair. In matrix form, this can be expressed as follows:

$$B_1 = \begin{bmatrix}
    s_1 & -s_2^* & s_5 & -s_7^* \\
    s_2 & s_1^* & s_7 & s_5^* \\
    s_3 & -s_4^* & s_6 & -s_8^* \\
    s_4 & s_3^* & s_8 & s_6^*
\end{bmatrix} \quad (4.23)$$

Figure 4.2 BER versus SNR with and without antenna grouping.
Considering different grouping index, the transmission matrix \( B \) can also be expressed as

\[
B_2 = \begin{bmatrix}
  s_1 & -s_2^* & s_5 & -s_7^* \\
  s_2 & s_1^* & s_7 & s_5^* \\
  s_4 & s_3^* & s_8 & s_6^* \\
  s_3 & -s_4^* & s_6 & -s_8^*
\end{bmatrix}
\]

\[
B_3 = \begin{bmatrix}
  s_1 & -s_2^* & s_5 & -s_7^* \\
  s_3 & -s_4^* & s_6 & s_8^* \\
  s_2 & s_1^* & s_7 & s_5^* \\
  s_4 & s_3^* & s_8 & s_6^*
\end{bmatrix}
\]

\[
B_4 = \begin{bmatrix}
  s_1 & -s_2^* & s_5 & -s_7^* \\
  s_4 & s_3^* & s_8 & s_6^* \\
  s_2 & s_1^* & s_7 & s_5^* \\
  s_3 & -s_4^* & s_6 & -s_8^*
\end{bmatrix}
\]

\[
B_5 = \begin{bmatrix}
  s_1 & -s_2^* & s_5 & -s_7^* \\
  s_3 & -s_4^* & s_6 & s_8^* \\
  s_2 & s_1^* & s_7 & s_5^* \\
  s_4 & s_3^* & s_8 & s_6^*
\end{bmatrix}
\]

\[
(4.24)
\]

\[
B_6 = \begin{bmatrix}
  s_1 & -s_2^* & s_5 & -s_7^* \\
  s_4 & s_3^* & s_8 & s_6^* \\
  s_3 & -s_4^* & s_6 & -s_8^* \\
  s_2 & s_1^* & s_6 & s_5^*
\end{bmatrix}
\]

At the mobile, the optimum transmission matrix is determined based on the following criteria. Let \( Y_{ri} \) be the received signal at the \( i \)th symbol time at the \( r \)th receive antenna, and \( h_{t,r} \) denote the channel parameter between the \( t \)th transmit and \( r \)th receive antenna. When the number of receive antennas is two, the received signal can be represented as

\[
y = X(\mathbf{H} \mathbf{W}) \mathbf{s} + \mathbf{v}
\]

(4.25)

where \( y = [y_{1,1} \ y_{1,2}^* \ y_{2,1} \ y_{2,2}^*]^T \), \( s = [s_1 \ s_2 \ s_3 \ s_4]^T \), \( \mathbf{v} \) is the noise vector, \( \mathbf{H} = [h_{1,1} \ h_{1,2} \ h_{2,1} \ h_{2,2} \ h_{3,1} \ h_{3,2} \ h_{4,1} \ h_{4,2}] \), \( X(\cdot) \) is a function of 2-by-4 input matrix, which is defined as

\[
X \left( \begin{bmatrix} a & b & c & d \\ e & f & g & h \end{bmatrix} \right) = \begin{bmatrix} a & b & c & d \\ b^* & -a^* & d^* & -c^* \\ e & f & g & h \\ f^* & -e^* & h^* & -g^* \end{bmatrix}
\]

(4.26)

and \( \mathbf{W} \) is a permutation of matrix \( \mathbf{B} \).
At the mobile station, the index of the transmission matrix $B_q$, $q = 1, 2, 3, \ldots , 6$ is determined based on the following criteria:

$$q = \arg \min_{l=1, \ldots , 6} \left[ \text{abs}(\det(H_{l,1}) + \det(H_{l,2})) \right]$$

(4.27)

where $H_{l,1}$ is the first two columns of $HW_l$, and $H_{l,2}$ is the last two columns of $HW_l$. Note that the antenna grouping matrix selection rule in (4.27) is equivalent to the following rule:

$$q = \arg \min_{l=1, \ldots , 6} \left[ \text{trace} \left( \left( (X(HW_l))^H X(HW_l) \right)^{-1} \right) \right]$$

(4.28)

Alternate criteria for the antenna grouping can be applied to determine antenna group index. For example, minimize BER, MMSE, and so forth.

We compare the proposed antenna grouping-based closed-loop STC with open-loop STC for four transmit-antenna rate 2 STC. In Figure 4.3, packet

![Figure 4.3](image)

**Figure 4.3** PER versus SNR with and without antenna grouping when the correlation coefficient is 0.7.
error rates (PERs) of the proposed antenna grouping method and the conventional open-loop STC (matrix B) method are compared in the pedestrian A channel with 3 km/h. One frame feedback delay is reflected in the simulation, and an MMSE linear detector is used at the receiver.

When the correlation coefficient is 0.7 (Figure 4.3), the proposed antenna grouping with 3-bit feedback outperforms the conventional STC without antenna grouping more than 1.8 dB at PER $= 10^{-2}$ (1.8 dB for 1/2 rate QPSK, 2.5 dB for 1/2 rate 16QAM, 2.4 dB for 1/2 rate 64QAM, and 3.2 dB for 2/3 rate 64QAM). As a higher MCS level is used, the performance gain is increased.

4.2.4.2 Codebook-Based Closed-Loop MIMO

The codebook words are employed in the feedback from mobile station (MS) to base station (BS). The MS learns the channel state information from downlink and selects a transmit beamforming matrix for the codebook. The index of the matrix in the codebook is then fed back to the BS. Each codebook corresponds to a combination of $N_t$, $N_d$, and $N_i$, which are the numbers of BS transmit antennas, available data streams, and bits for the feedback index, respectively. Once $N_t$, $N_d$, and $N_i$ are determined in the MS, the MS will feed back the codebook index of $N_i$ bits. After receiving an $N_i$ bit index, the BS will look up the corresponding codebook and select the matrix (or vector) according to the index. The selected matrix will be used as the beamforming matrix in MIMO precoding.

4.3 Radio Resource Management

As the evolution of wireless networks allows for an increasingly wide range of services, the Radio Resource Management (RRM) function, which divides the available resources amongst competing applications, is receiving increased attention. There are several reasons that render RRM very important [49-75]:

1. RRM functions allow the support of a range of different requirements from the various services that the wireless network is required to support.
2. RRM may ensure the planned coverage (i.e., the area where the service is supported) for each service.
3. RRM may optimize capacity utilization.

Basically, the RRM has the complex task of maximizing the number of users that can be served satisfying their different service requirements, in time varying radio conditions and dynamic traffic behavior. In this section we will focus on the main RRM functions.
4.3.1 QoS Requirements

In order to guarantee a satisfactory end user quality, the transmission of a data flow, which is originated by the application, has to satisfy certain requirements that define the QoS profile for the information data stream of interest. Usually, the QoS attributes for a particular application/service are: required throughput, maximum acceptable delay, maximum acceptable delay jitter, and maximum acceptable bit error rate.

From the end user point of view, the following issues must be taken into account [53]:

- End users only care about the degree of QoS, and not about how it is provided.
- Only the QoS perceived by end user matters.
- The number of “user defined/user controlled” parameters has to be at a minimum.
- A derivation/definition of QoS attribute from the application requirements has to be simple.
- End-to-end QoS has to be provided.

A very frequent classification of the service class is based on the application/service’s delay requirement. For example, in the UMTS standardization the following four service classes are defined: background class, interactive class, streaming class, and conversational class. The main difference between these classes is the delay sensitivity of the traffic. The conversational class is the most delay sensitive, while the background class is the most delay tolerant. Table 4.1 summarizes the UMTS QoS classes’ main characteristics.

In the standardization the definition of QoS attributes are the same for GPRS Release 99 and UMTS. The QoS attributes for General Packet Radio Service (GPRS) Release 97/98 can be mapped on the Release 99 UMTS attributes as specified in [53]. However, the set of QoS attributes in UMTS is much larger than the set specified in GPRS Release 97/98 (Figure 4.4).

The throughput requirement is dependent on the information source. A throughput of 12 Kbps would be enough in order to transfer speech with GSM quality. For an audio stream with stereo quality, the throughput requirement is higher than 32 or 64 Kbps, while for a video communication it is higher than 128 Kbps.

The BER depends on the service class. For background (e-mail, file transfer protocol (FTP)) and interactive (web browsing) services, the received data should be error free. Due to the less stringent delay requirements of this service class, higher reliability can be achieved by error correction techniques (e.g.,
packet retransmissions). The use of error correction mechanisms is rather limited by the delay requirement (e.g., for the interactive class). For the streaming and conversational classes, the acceptable BER depends on the type of information; for speech an acceptable BER is on the order of $10^{-2}$ or $10^{-3}$ and it may be even smaller for video transfer.

The application may specify its QoS requirements to the network by requesting a radio access bearer (RAB) with any of the specified traffic type, maximum transfer delay, delay variation, bit error rates, and data rates. In

<table>
<thead>
<tr>
<th>Traffic Class</th>
<th>Conversational</th>
<th>Streaming</th>
<th>Interactive</th>
<th>Background</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fundamental</td>
<td>Preserve time</td>
<td>Preserve time</td>
<td>Request response</td>
<td>Destination is not expecting the</td>
</tr>
<tr>
<td>Characteristics</td>
<td>relation (variation)</td>
<td>relation (variation)</td>
<td>pattern</td>
<td>data within a certain time</td>
</tr>
<tr>
<td></td>
<td>between information</td>
<td>between information</td>
<td>Preserve payload</td>
<td>Preserve payload content</td>
</tr>
<tr>
<td></td>
<td>entities of the stream</td>
<td>entities of the stream</td>
<td>content</td>
<td>content</td>
</tr>
<tr>
<td>Example of the</td>
<td>Voice, video</td>
<td>Streaming video</td>
<td>Web browsing,</td>
<td>Background download of e-mail</td>
</tr>
<tr>
<td>Application</td>
<td>games, voice</td>
<td></td>
<td>network games</td>
<td></td>
</tr>
<tr>
<td></td>
<td>telephony</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1
UMTS QoS Classes

Figure 4.4  Mapping of QoS parameters.
practice, it should be possible to define the main RAB characteristics from the service quality requirements:

- The transmission rate of RAB should be determined by the bandwidth requirement of the information source.
- The choice of dedicated or shared RAB should be based on the requirement for the maximum delay and delay jitter.
- The SNIR requirement, channel coding, and interleaving for the RAB should be based on the BER requirement.

4.3.2 General Formulation of the RRM Problem

Let us assume a cellular network with M mobiles in the service area and denote with $B = \{1, 2, \ldots, B\}$ the set of all BSs used to provide the necessary coverage. Denote with C the number of available orthogonal channels in the system (i.e., the system capacity). The numbered set of all available channels is $C = \{1, 2, \ldots, C\}$. The channels orthogonality could be established in different ways, such as in time and frequency domain in GSM or in the code domain in Wideband (W-CDMA). In GSM, as a representative of 2G systems, there is an intrinsic upper limit on the system capacity since the upper bound is the number of frequencies multiplied with eight time slots. On the other hand, in the WCDMA CDMA scheme, the set of orthogonal channels $C$ is practically infinite and the capacity is determined by the interference condition in the system. A WCDMA system is, therefore, interference limited.

The link (power) gain matrix $G$ characterizes the radio conditions in the system:

$$ G = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1M} \\ G_{21} & G_{22} & \cdots & G_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ G_{B1} & G_{B2} & \cdots & G_{BM} \end{bmatrix} $$

(4.29)

The matrix element $G_{ij}$ represents the link gain between the BS $i$ and MS $j$; $M$ represents the number of active mobiles. The gain matrix $G$ is dynamic, the dimension $M$ is changing, based on the offered load, and each element $G_{ij}$ changes with the mobile movement. The radio resource management, taking into account the link gain matrix $G$, assigns [51]:

1. One or more access points from the set $B$;
2. A channel from the set $C$;
3. The transmit power of the BS and of the mobile.

The assignments 1 through 3 should maximize the number of users with a sufficient QoS. As outlined in the previous section, providing a stringent definition of the QoS for a communication service is a complex problem. A simple measure of the QoS-namely, the signal-to-interference + noise ratio (SNIR)-is here considered. This measure is strongly connected with the performance measures as the bit or frame error probability. Therefore, assignments 1 through 3 aim at maximizing the number of users for which the following inequality holds, for both the uplink (mobile-to-access port) and the downlink (access port-to-mobile):

\[
\text{SNIR}_j = \sum_{\substack{m \neq j}}^{M} \frac{P_j G_{j} j}{\theta_{jm}} \geq \gamma_j \quad j = 1, \ldots, M
\] (4.30)

In (4.30), \(\text{SNIR}_j\) denotes the SNIR at the receiver; \(P_j\) is the transmitter power used by the end user \(j\); \(\theta_{jm}\) is the normalized cross-correlation between the signal of interest and the interfering signal from the mobile \(m\) (other than \(j\)); \(N\) denotes the thermal noise power at the access port; and while \(\gamma_j\) is the target SNIR of the service that is being used by the mobile \(j\).

### 4.3.3 RRM in Future Wireless Systems

In this section, future RRM developments in wireless LANs and general RRM issues in mobile ad hoc networks (MANETs) are presented.

The Broadband Radio Access Network (BRAN) working group in ETSI is standardizing a wireless LAN for broadband radio access up to 54 Mbps. This new standard is called HIPERLAN/2 and includes physical layer and radio link control and data link control standards (see [61-65]). The interfaces toward other networks (e.g., UMTS) are made via specific design of convergence sublayers. The physical layer of HIPERLAN/2 is aligned with the physical layer of the IEEE 802.11a wireless LAN system [65]. Note here that wireless broadband networks based on the IEEE 802.11a standard became commercially available in 2002. Among the most important RRM functions in these WLAN systems are the link adaptation function and the radio resources allocation in the MAC frame.

In HIPERLAN/2 and IEEE 802.11 systems there are different pairs of modulation and coding schemes possible. Each pair results in different transmission rate and PER performance, depending on the radio channel quality (i.e., signal-to-interference ratio). The link adaptation scheme can dynamically change
the pair modulation/coding scheme to optimize the throughput based on measured PER, signal level, packet size, and so forth. Extensive analysis of the physical layer performance of these two WLAN systems can be found in [63].

The MAC layer at the two WLAN standards is different. HIPERLAN/2 uses TDMA/TDD medium access with frame structure, as presented in Figure 4.5.

The allocation of radio resources is centrally scheduled by the access point (AP), which allows for implementation of scheduling algorithms, QoS differentiation, and resource reservation for services with stringent delay and delay variation requirements.

The duration of the broadcast channel (BCH) is fixed. Through these channels the relevant system information is conveyed to the terminals. The duration of the frame control channel (FCH), downlink (DL) and uplink (UL) phase, direct link (DiL) phase, and random channel (RCH) is dynamically adapted to the current load conditions of the AP. DiL phase is present if there are mobile terminals directly communicating on a peer-to-peer basis. Requests for the radio resources are signaled via the RCH, where contention for time slots is present. If scheduled data is transmitted, then FCH is present and it signals the frame structure. The transmission in the DL phase (from the AP to the terminals) and UL phase (from terminals to the AP) is contention free. The allocation of resources is signaled via the access feedback channel (ACH), as an answer to the resource request received via the RCH from the previous MAC frame.

The IEEE 802.11a standard, however, has a MAC scheme based on carrier sense multiple access with collision avoidance (CSMA/CA). A mobile terminal before transmission of data senses the radio channel. If the channel is free then the transmission can commence. Otherwise, an exponential back-off period is implemented before attempting the following packet transmission. This type of MAC (also known as distributed coordination function) makes the IEEE 802.11a standard more suitable for ad hoc wireless networks and non-real-time (background or interactive) type of applications. It should be mentioned that the standard also has contention-free MAC via the point coordination function (PCF), but this alternative is optional even though it could support real-time services. The advantage of the CSMA/CA is, however, the avoidance of a centralized scheduler that coordinates the radio transmissions.

![Figure 4.5](image-url) HIPERLAN/2 MAC frame structure.
Mobile ad hoc network, (MANETs) are also receiving attention in recent years. In Internet Engineering Task Force (IETF) there is a special working group (MANET working group) that investigates the routing protocols in ad hoc network. These networks are described by the IETF MANET group (source IETF MANET group):

A MANET is an autonomous system of mobile routers (and associated hosts) connected by wireless links, the union of which forms an arbitrary graph. The routers are free to move randomly and organize themselves arbitrarily; thus, the network’s wireless topology may change rapidly and unpredictably. Such a network may operate in a stand-alone fashion, or may be connected to the larger Internet.

Currently, there are many research activities in the RRM field for MANETs. Due to the specific characteristics of MANETs, such as dynamic topology, limited node performance, distributed algorithms, and so forth, the development of RRM functions is a difficult task. The most important factors in MANETs are the coverage (or connectivity) and the capacity [66, 67]. An extensive field of research is QoS-aware routing (see [68] for one typical example) and MAC with QoS support [69]. A common MAC design for MANETs is driven by the hidden/exposed terminal problem. However, the QoS constraints for particular applications require from the MAC layer additional functionalities to provide certain guarantees and to make distinctions among different types of connections. Furthermore, MAC has to interact with QoS-aware routing in an appropriate way to provide the required communication quality over the whole path from source to destination (i.e., over the multiple hops). The current MAC proposals [69] that support QoS provisioning and differentiation are based on resource reservation along the connection’s path, and differentiation between non-real-time and real-time connection establishment and maintenance.

The recent important development in the RRM field for multimedia wireless communication is that for systems beyond UMTS. In these systems the last wireless hop towards the end user could be carried over different radio access networks. Here, the interworking of UMTS, WLAN or HIPERLAN/2, and GSM/GPRS networks will play an important role. For example, the wireless operators could have in their coverage areas multiple possibilities for wireless communications via different radio access networks, as presented in Figure 4.6.

In this type of wireless networks, the setup of the Common RRM (CRRM) functions will play an important role in the efficient utilization of the available radio bandwidths. To achieve this goal the CRRM will perform traffic addressing towards less loaded wireless access systems. CRRM will choose the right radio access technology based on service requirements, current wireless system load, propagation conditions, interference, and capacity cost induced in the wireless network. This field is very challenging and interesting for future RRM research.
4.4 Software Defined (SDR) Radio Communication Systems

As the number of wireless communication systems that users can use increase, there has been increasing demands for the coexistence of several mobile telecommunication services; for example, GSM, or IS-54, or IS-136, or IS-95, Japanese Personal Digital Cellular (PDC) or Personal Handy Phone System (PHS) or IMT-2000 system. Currently, if one wants to be globally connected, more than one terminal may be needed, though it is becoming more common to find multistandard terminals.

A person working on SDR is expected to be familiar with any communication system whether it is for radio transmission or multiple access purposes. Such schemes can be developed and written in programming languages like MATLAB or C. These computer simulation languages have a good relationship with software languages that configure digital signal processing hardware (DSPH) such as digital signal processors (DSP), field programmable gate arrays (FPGA), and application-specific integrated circuits (ASIC). A typical software language for DSPH is very high speed integrated circuit hardware description language (VHDL) and Verilog-HDL. DSPH has been utilized to configure the mobile terminals and base stations of mobile communication systems. DSPH is
a general-purpose language, and the configuration can be programmed by downloading digital signal processing software (DSPS). This means that users can download DSPS describing the desired elemental components into the DSPH of only one terminal. This is the basic concept of an SDR communication system [76-90].

Software defined radio in theory will allow mobile terminal manufacturers to design and manufacture products that are independent of any particular specifications or standard. This means that users or the terminal itself can select the most appropriated air interface to be used based on channel conditions, traffic, cost, and so forth. There are benefits from the ecological point of view as well: SDR reduces the amount of hardware, which, in turn, reduces the amount of industrial waste.

Research into radio communication systems based on DSPH began in early 1990. These early studies, however, did not examine DSP-based radio communication systems from the viewpoint of reconfigurable radio communication systems. As far as we know, the first paper to coin the phrase “SDR” is that of Mitola [76]. Mitola described the fundamental and functional architecture of SDR. One prototype, the military SDR, SPEAKeasy, was introduced in [77]. SPEAKeasy can use several voice and data military services; the frequency band is 2 MHz to 2 GHz.

Most of the research towards realizing the ultimate communication system has focused on [78-82]:

1. The architecture of the DSPH;
2. The configuration of the analog signal processing hardware;
3. The method of downloading software to the hardware;
4. The method of application of software defined radio.

### 4.4.1 Definition of SDR Communication System

Figure 4.7 shows the basic configuration of the SDR of [76]. The radio consists of eight units: (1) antenna unit, (2) radio frequency signal processing unit (RFU), (3) intermediate frequency signal processing unit (IFU), (4) analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC) unit, (5) baseband signal processing unit (BBU), (6) transmission control unit (TCU), (7) input/output (I/O) processing unit (IOU), and (8) end-to-end timing processing unit (TPU). Each unit will be described in detail.

1. Antenna unit. An omnidirectional, low loss, and broadband antenna is required because it can be used in a variety of wireless communication systems. Moreover, signal processing technology based on array
antennas makes it possible to select the performance of the SDR according to the surroundings and to perform optimum selection of the algorithm by using SDR technology. Such an antenna is called a smart antenna or software antenna \cite{83, 84}. The software antenna is capable of space division multiple access (SDMA), in which the antenna steer the beam in the direction of selected users by computing appropriate weight coefficients for the antenna elements. Multiple access is achieved by changing the direction of the antenna or beams, or by interference cancellation, in which the antenna configures its direction to the desired user or allocates null points to the direction of undesired users or signals.

2. Radio frequency signal processing unit. In the transmitter’s RFU, the signals coming from the IFU or BBU are up-converted to the radio frequency band signals, amplified, and transmitted to the antenna unit. At the receiver, the signals received by antenna unit are amplified to a constant level that is suitable for signal processing and directly down-converted to a lower frequency band such as IF band or baseband. The signal processing is done by an analog circuit. The linearity or efficiency of the RF amplifier and the conversion method to the lower frequency band at the receiver are the main discussion points.

3. Intermediate frequency signal processing unit. In this unit, the signals from ADC/DAC unit are up-converted to the IF band signal, amplified, and transferred to the RFU of the transmitter. At the receiver, the signals from the RFU unit are amplified to an adequate level for signal processing in the IFU and directly down-converted to a suitable frequency for the ADC/DAC unit or baseband unit. When the signals of
several systems are received at the receiver, the required frequency band must be selected by using a filter.

4. Analog-to-digital conversion and digital-to-analog conversion unit. In this unit, the digital signal from the baseband unit is converted to an analog signal by using a DAC and the converted signal is transferred to the upper frequency band unit (IFU or RFU). At the receiver, the signals from the IFU or RFU are amplified to an adequate level for analog-to-digital conversion. The stabilized signal is then sampled by ADC and converted to a digital signal. In this unit, the sampling method is a key technology.

5. Baseband signal processing unit. In this unit, data is digitally modulated and transferred to the ADC/DAC unit of the transmitter. Transmitted data is recovered by using the sampled signal from the ADC/DAC unit and digital signal processing at the receiver. The basic configuration of the BBU is shown in Figure 4.8 [77]. In the BBU of the transmitter, frame, coding, mapping and modulation, and transmission filter blocks are the key blocks. On the other hand, in the BBU of the receiver, receiver filter, code and symbol timing, sampling rate conversion (resample), demapping and demodulation, and decoding blocks are the key blocks. Moreover, in the BBU of the receiver, the fading compensation (equalization) block and the interference cancellation block for eliminating undesired signals are present. In most cases, the BBU is configured by several DSPH such as DSP and/or FPGA and/or ASIC. All component blocks are described using DSPS written in VHDL or Verilog-HDL and compiled. The BBU’s configuration can be modified by changing the DSPS.

6. Transmission control unit. In this unit, the input bit stream format for the BBU is configured at the transmitter by adjusting the transmission protocol of the MAC layer, and at the receiver, the detected data from the BBU is checked according to the data format of the transmission protocol of the MAC layer. If the number of bit errors in the detected data is large, retransmission is requested. In addition to this transmission control, this unit can manage cryptograph. In most cases, TCU can be configured by a range of DSPHs, and all the component blocks are also described using DSPS. By changing DSPS, the TCU can configure the transmission protocol as needed.

7. Input/output processing unit. In the mobile station, all information data comes from a handset, PDA terminal, or personal computer, and all received data comes back to these terminals or computers. The I/O and timing are managed so as to connect with the external terminals flexibly.
Figure 4.8 Basic BBU configuration.
8. End-to-end timing processing unit. This unit controls the transmission delay between transmitter and receiver. For example, the transmission delay for voice must be shorter than 150 ms (typically).

In most SDR systems, several software programs, which describe all telecommunication components in DSPS language, are used to configure the components on the DSPH. This software can be readily changed to suit the requirements of a particular system. Such a SDR system is called a full-download-type SDR system. Figure 4.9 shows the configuration of a full-download-type SDR system. The system has an RFU, IFU, and BBU as in the basic system. Moreover, it has a TX module, which is related to the transmitter, and an RX module, which is related to the receiver.

Implementing a specific telecommunication system with a full-download-type SDR system requires that all necessary DSPS be downloaded to the BBU before starting communication. DSPS blocks, including frame block, encoder block, mapping and modulation block, and filter block (Figure 4.8) are downloaded to the BBU of the TX module. DSPS blocks, including filter block, equalizer block, detector and decoder block (Figure 4.9), are also described in DSPS and downloaded to the BBU of the RX module. After software has been downloaded, the configuration check program is executed. The BBU then

![Figure 4.9 Configuration of a full-download-type SDR system.](image-url)
configures the required baseband modulation and demodulation circuit. Then, transmitted data can then be fed into the BBU of the TX module.

In the BBU of TX module, this data are formatted into frames, modulated and converted into two signals: in-phase channel (Ich) and quadrature-phase channel (Qch) signals by the DSP blocks mentioned above. These signals are fed into the IFU of the TX module.

In the IFU of the TX module, the digitally modulated Ich and Qch signals are converted from digital to analog by a D/A converter block. These signals are quadrature modulated on the IF band and sent to the RFU of the TX module. In the RFU of the TX module, the quadrature-modulated signal is up-converted to the RF band by power control part before being transmitted.

To receive the RF signal, the received signal is fed into the RFU of the RX module. Here, the received RF data is bandpass-filtered, which eliminates spurious signals, and down-converted to the IF band. The automatic gain control (AGC) block keeps the power of the down-converted signal at a constant level. This power-controlled signal is fed into the IFU of the RX module.

In the IFU of the RX module, the received signal from the RFU is split into Ich and Qch signals by using a quadrature demodulator block. The A/D converter block then oversamples and transfers them into the BBU of the RX module.

In the BBU of the RX module, all telecommunication component blocks have been implemented into the DSPH before starting communication, and the configuration of what has been checked by a test program. The oversampled Ich and Qch signals are filtered and equalized by a customized method, and they are detected and decoded by using the filter, equalizer, and decoder blocks in the BBU of RX module.

### 4.4.2 Advantages of SDR Communication Systems

In full-download-type SDR, the system configuration can be changed on demand. There are many advantages not only for operators and service providers but also for government and commercial customers.

In particular, for commercial operators and service providers the advantages are:

1. Global roaming services;
2. Upgradeable terminals;
3. New services added without having to change the terminal;
4. Bugs fixed without the need to recall the product;
5. Versatile software (i.e., wireless communication software that can be installed in other electrical products as well as in the mobile terminal).
For government agencies, the advantages are:

1. Global roaming services can be offered to customers;
2. SDR reduces the variety of hardware, since several standards can be implemented in a single mobile terminal.

For customers, the advantages are:

1. Unlimited global roaming;
2. One terminal for many services;
3. New services provided without needing to upgrade hardware;
4. New services added to the terminal without changing the terminal;
5. Bugs fixed without needing to recall the product.

### 4.4.3 Problems in SDR Communication Systems

As shown in Section 4.4.2, SDR has many advantages, but the technology must overcome the following problems.

1. The volume of software downloaded to the DSPH increases as the contents of the required telecommunication component blocks become more complicated. As a result, the download time is lengthened. In addition, the software files to be transmitted need to be protected by an adequate channel coding scheme to make the transmission more robust to fading and interference. In this case, the download time becomes even longer [87, 88].

2. The period for the configuration check of the DSPH also increases since the contents of the required telecommunication component blocks become more complicated. The problem also affects the stability of the operating characteristic of the DSPH when there is not a sufficient period of time for the configuration check [87, 88].

3. In the downloaded software, there are often several component blocks containing manufacturer-specific know-how (e.g., the optimization method or calculation algorithm for some special blocks). There is the possibility that this know-how may leak out when the software is downloaded. There is also the possibility that it may be tampered with [87, 88].

In addition to the above problems, if SDR controls RF and IFU as well as the baseband unit, the following issues must be considered.
1. By using SDR, a user must download only the software describing the elemental components to the hardware for realizing a particular communication system. However, if several communication systems operating on different frequency bands are integrated into one mobile communication hardware by SDR technology, several antenna units, RFUs, IFUs, and ADC/DAC units are still needed in the mobile communication hardware. Moreover, a user may want to use several systems simultaneously [89], in which case several ADC/DAC units, baseband units, transmission control units, I/O processing units, and end-to-end timing processing units must be prepared. The method of managing several units must be considered.

2. If several communication systems operating on different frequency bands are integrated into one mobile communication hardware by SDR technology, a broadband antenna must be used. Since the bandwidth of antenna is limited, however, several antennas are needed. The placement of these antennas in the terminal is an important design issue to be taken into consideration.

Moreover, from the viewpoint of the entire system, the following questions arise:

1. Which services are to be integrated in one SDR terminal? Say that a multimode terminal can easily be realized by developing a specific ASIC for each communication system and planning these ASICs on one circuit board. In this case, SDR technology is unnecessary. Therefore, SDR technology should be targeted to application fields in which many communication systems are required.

2. How and in which field is the SDR technology socially recognized? There are many applications for SDR technology. These applications must be identified.

3. How and in which field is standardization to be done? How should the software be protected against viruses or hackers? In particular, SDR’s capability of international roaming by changing software instead of changing terminal becomes a serious threat, if a virus is included that can reconfigure a wireless communication system. A simple example is the reconfigured terminal in which the transmitter no longer matches the receiver. Moreover, a virus altering the transmission power to a much higher level than the regulation threatens to shut down all wireless communication systems in the world. Reference [90] describes wireless terrorism. In wireless terrorism, all users of SDR effectively become potential terrorists. To prevent such an
incident from occurring, standardization must be done. In addition, the organizations and radio that certify SDR terminals must be developed. Software radio thus presents a new paradigm in radio law.

4.4.4 Future Applications of SDR Communication Systems

4.4.4.1 Future Telecommunication Applications
The following applications are envisioned for the future.

Mobile Communication Terminal. A SDR technology-based mobile communication terminal allows users to select a system by changing the software of the DSPH. Moreover, users can select the provider company that they like. However, to implement them will require miniaturization, reduction of the power consumption and cost of the hardware, as well as reduction in software download time. An efficient system selection procedure is also required.

Mobile Communication Base Station. The base station that has SDR technology would make it easy to implement new communication systems and fix bugs. In addition, a base station could use more than one algorithm to fight against fading. Moreover, by using an adaptive array antenna, the antenna beam shape could also be controlled with software. To realize this application, we must make a rule for programming software that configures components for the mobile communication system to change the program easily and to make the connection with other components. Moreover, a fast ADC or digital signal processor is needed to perform real-time transmission.

Broadband Radio Access System. Software radio technology can enhance the flexibility of a broadband radio access system, which operates between base stations and buildings. For example, software can be used for components such as distortion compensators and interference suppressors. The modulation scheme also can be changed, and the best components selected. Using an adaptive array antenna would allow the beam shape to be controlled with software. To do so, each module should be implemented by software, and also, the connections between modules should be programmable by software. A high speed ADC or digital signal processor must also be used to perform real-time transmission.

Intelligent Transport System. The communication systems utilized in the intelligent transport system (ITS) are shown in Table 4.2. They fall in three categories: communication-based systems, control-based systems, and broadcasting-based systems. Representative examples of communication-based systems are Personal Digital Cellular (PDC), PHS, and digital cellular using CDMA.
The global positioning system (GPS) system, vehicle information and communication system (VICS) system, and radar are representative examples of control-based systems. The Advanced Cruise-Assist Highway System (AHS)—which provides information on traffic accidents, road obstructions, and meteorological phenomenon—and the electronic toll collection (ETC) system will be mounted in cars. Representative examples of broadcasting-based systems are radio broadcasting systems, analog-television broadcasting systems, satellite broadcasting systems, and digital television systems. As mentioned above, the number of communication systems appearing in the ITS system will likely increase. This means that space-saving and integration of system components must be chief concerns if so many services are to be provided in the limited space of a car. SDR technology envisions that these services will be provided by only one small terminal. Figure 4.10 shows the potential application. The method of introducing SDR technology is the same as described in the first three applications above. In the ITS, the number of services that can be used in a car is more than 10. Therefore, we urge implementation of SDR technology as soon as possible.

4.4.4.2 Broadcasting Applications

Software radio technology can realize automatic switching between terrestrial wave/BS/CS/CATV. It is easy to introduce new services on existing hardware because only the software needs to be changed, as was discussed before. Moreover, receivers compatible with the standards of many countries or regions can be realized. This would have an impact on the promotion and distribution of international products. Such a multimode, multistandard receiver must have a broadband antenna and software describing the components of the particular communication system, and this means that the connectivity between components must be increased.

<table>
<thead>
<tr>
<th>System</th>
<th>PHS</th>
<th>GPS</th>
<th>ETC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>1.9-GHz band</td>
<td>1.5-GHz band</td>
<td>5.8 GHz</td>
</tr>
<tr>
<td>Modulation</td>
<td>p/4DQPSK</td>
<td>BPSK +</td>
<td>ASK</td>
</tr>
<tr>
<td>TDMA-TDD</td>
<td>TDMA-TDD</td>
<td>DS-Spread Spectrum</td>
<td>(Manchester code)</td>
</tr>
<tr>
<td>Data Rate</td>
<td>384 Kbps</td>
<td>50 bps</td>
<td>1,024 Kbps</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>300 kHz/1ch</td>
<td>1.023 MHz</td>
<td>Less than 8 MHz</td>
</tr>
</tbody>
</table>
Figure 4.10 Applications of SDR for future ITS system.
4.4.4.3 Private Networks

Nowadays, many private networks exist for diverse purposes, like education, office, community, ITS, emergency, and commercial use (Figure 4.11). Current terminals for these networks do not have any connectivity with each other. Software radio can be used to create a universal terminal that works on any private network. In the near future, these huge networks will not only complement but also compete with conventional public networks.

4.4.4.4 Certification Method of SDR

One problem facing practical implementation of SDR is wireless terrorism. The potential types of damage are shown in Figure 4.12. In some cases, downloaded data may be illegally altered or have embedded in it a computer virus. Altered or virus software may cause unexpected and even dangerous changes in the base station (e.g., the BS could be commanded to shut down or increase transmission power over permitted levels, or it could be permanently damaged). In order to avoid such wireless terrorism, a new certification method for the systems using SDR must be required in the certification organizations.

An example of certification is shown in Figure 4.13. The certification method has two stages. In the first stage, the certification organization checks the relationship between input data and output data by changing the software that configures several systems. In this case, for each system, the transmission power, transmission bandwidth, and electric power leakage in the adjacent

Figure 4.11  Private network applications of SDR.
frequency band are measured and evaluated by comparing the radio law of each country.

If the integrity of the software is verified and the new reconfiguration is acceptable, the certification organization gives a certification password to the software. The certification password is used in the handshake between the

Figure 4.12 Wireless terrorism.

Figure 4.13 Certification method of SDR equipments: (a) first phase; and (b) second phase.
Figure 4.13 continued.
DSPH and memory before downloading the software to the DSPH. During the handshake, if the password is certificated, the software is downloaded from memory to the DSPH, otherwise the SDR equipment issues a warning. The public key cryptography technique is a candidate for the certification password. Moreover, the download protocol between the memory and the DSPH must be standardized and the protocol must be certificated.

The second stage of certification involves the modules that configure SDR. To certify each component, the configuration of SDR must be changed. An example of configuration is shown in Figure 4.13(b). Here, switching or interface modules are inserted between functional modules to configure the SDR equipment. These switching or interface modules can be controlled from outside of the DSPH. By switching modules, each functional module is certificated. The switching or interface module must be standardized for such a certification procedure to work.

4.4.5 Summary

One of the future applications of the software programming method is in the SDR communication systems. The concept of SDR bridges the software programming and hardware implementation.

This section has described the SDR communication system, its advantages and its problems, as well as some possible future applications. This technology will likely become key to the success of forth generation mobile communication systems, because nowadays there are many systems in the world and these systems must be integrated.

4.5 IP Network Issues

The evolution towards a common, flexible, and seamless IP-based core network that will connect heterogeneous networks gives rise to several issues at the network layer [91-160]. The need for ensured QoS is the key of this evolution. Fundamentally, the day that packet-switched networks can credibly approach the QoS of circuit-switched networks is the day customers stop paying for two networks. The QoS problem involves integrating delay-sensitive applications such as voice, audio, and video onto a single network with delay-insensitive applications, such as e-mail, fax, and static file transfer. That network must be able to discriminate, differentiate, and deliver communications, content, and commerce services. Moreover, it needs to support communication services creation, modification, bundling, and billing in a way that is unobtrusive yet powerful for end users, especially business users. The scenario becomes more challenging as QoS and security issues are faced in a mobile environment. The
Internet has not been designed with mobility in mind and lacks mechanisms to support mobile users. Some architectures have been proposed for supporting mobility in the Internet. The most important is Mobile IP, which will be discussed in Section 4.5.2. Before that, some guidelines behind mobility management are discussed in Section 4.5.1. Proposals to solve some of the open issues in Mobile IP are described in Section 4.5.3.

4.5.1 Mobility Management

When a mobile node is roaming through one or more service areas, mobility management mechanisms may be required to locate it for call delivery and maintenance of its connections. Generally, in cellular systems, mobility management is performed through two main mechanisms:

1. **Location management** is used for discovering the current attachment point of the mobile user for call delivery. It consists of two phases. In the first phase, called the location registration (or location update), the mobile terminal periodically notifies the network for its new access point, allowing the network to authenticate the user and revise the user’s location profile. The second phase is call delivery. The network is queried for user location profile and the current position of the mobile host is found.

2. **Handoff management** enables the user to keep its connection alive as it moves and changes its access point to the network. This is performed in three steps: initiation, connection generation, and data flow control.

When the user moves within a service area or cell and changes communications channels allocated by the same BS, the mobility management procedure is called intracell handoff. Intercell handoff occurs when the user moves into an adjacent service area or cell for which all mobile connections are transferred to a new BS. While performing handoff, a mobile terminal can be simultaneously connected to two or more BSs and use some kind of signaling diversity to combine the multiple signals. This condition is called soft handoff. In hard handoff, the mobile device switches from one base station to another with active data being forwarded on only one path at a time.

4.5.1.1 Mobility Classes

In general, there are four categories that support IP mobility:

1. Pico-mobility is the movement of a mobile node (MN) within the same BS. The operating space is the space around person that typically
extends up to 10m in all directions and envelopes the person, being either stationary or in motion.

2. Micro-mobility is the movement of an MN within or across different BSs within a subnet; this occurs very rapidly. Management of micro-mobility is accomplished using link-layer support (layer 2 protocol), which is already implemented in existing cellular networks.

3. Macro-mobility is the movement of an MN across different subnets within a single domain or region; this occurs relatively less frequently. This is currently handled by Internet mobility protocols such as Mobile IP.

4. Global mobility is the movement of an MN among different administrative domains or geographical regions. This is also handled by layer 3 techniques such as Mobile IP.

Mobility management has a responsibility of providing uninterruptible connectivity during micro- and macro-mobility, which usually occur over relatively short time scales. Global mobility, on the other hand, usually involves longer time scales. Therefore, the goal is just to ensure that mobile users can reestablish communication after they change the domain, but not necessarily to provide uninterruptible connectivity.

4.5.1.2 Architectures for Mobility Supporting

There are several frameworks that support mobile users, and the IETF standardizes two of them: Mobile IP and Session Initiation Protocol (SIP).

1. Mobile IP [91] supports application-layer transparent IP mobility. The basic Mobile IP protocol does not require protocol upgrades in stationary correspondent nodes (CNs) and routers. Its drawback is that it does not consider the integration of additional functions such as authentication and billing, which are critical for successful adoption in commercial networks.

2. SIP [92] is an application-layer control (signaling) protocol that can establish, modify, and terminate multimedia sessions or calls. The main disadvantage of SIP mobility is that it cannot support transmission control protocol (TCP) connections and is also not an appropriate solution for micro- or macro-mobility.

SIP mobility will not be a subject of further discussion in this chapter; instead the focus will be on Mobile IP.
4.5.2 Mobile IP

Mobile IP refers to a set of protocols, developed and still under development by the IETF to allow the Internet Protocol to support the mobility of a node [91]. The idea for Mobile IP first emerged in 1995 and since then it has undergone some changes.

First of all, it is necessary to understand what makes the IP mobility complicated. An IP address consists of two parts: a prefix that identifies the subnet in which the node is located, and a part that identifies the node within the subnet. Routers use look-up tables to forward incoming packets according to their destination addresses. A router does not store the addresses of all computers in the Internet, which is not feasible. Only prefixes are stored in the routing tables and some optimizations are applied. Therefore, as a receiver moves outside the original subnet, it can no longer be reached. A new IP address should be assigned to the mobile node, but this operation requires time. Specifically, the Domain Name System (DNS) needs time to update its internal tables for mapping a logical name to an IP address. Therefore, this approach cannot work if the node moves quite often. Moreover, a browser and a Web server being, respectively, client and server in communicating, generally use the TCP/IP protocol suite to establish a reliable end-to-end communication. Using TCP means that a virtual connection must be established before data transmission and reception. When a logical circuit is created, both connection sides must be assigned port numbers in order to let application layers keep track of the communications. Every end-to-end TCP connection is identified by four values: client IP address, client TCP port, server IP address, and server TCP port. These values must be constant during the conversation. Therefore, a TCP connection will not survive any address change.

Allowing the mobile nodes the use of two IP addresses provides the solution to the above problem. In Mobile IP, home address remains unchanged regardless of where the node is attached to the Internet and it is used to identify TCP/IP connections. Another IP address, which is dynamic, is assigned to the mobile node when it moves to another network different than its home network. This address is called the care-of address (COA) and it is used to identify the mobile node point of attachment in the network topology. Using its home address, the mobile node is seen by other Internet hosts as a part of its own home network. In other words, other Internet hosts do not have to know the actual mobile node’s location. The home agent (HA) is located in the home network. Whenever the mobile node is not attached to its home network but to another network, called the foreign network (FN), the home agent gets all packets destined for the mobile node and arranges their delivering to the foreign agent (FA). The FA then sends the packets to the mobile node. An HA can be implemented on a router that is responsible for the home network. This position is
quite appropriate because without any optimization to mobile IP, all packets for the MN have to go through the router anyway. An alternative solution consists in making the router behave as a manager for MNs belonging to a virtual home network. In this case, all MNs are always in a foreign network. The HA could also be implemented on an arbitrary node in the subnet. This solution is necessary when the router software cannot be changed, but it has the disadvantage that there is a double-crossing of the router by the packet if the MN is in a foreign network.

In general, the Mobile IP concept can be considered as a combination of three major functions:

1. **Discovery mechanism**, which allows mobile computers to determine their IP address as they move from network to network;
2. **Registering** of the new IP address with its home agent;
3. **Tunneling** (delivery) of packets to the new IP address of the mobile node.

### 4.5.3 Evolution of Mobile IP

Mobile IPv4 is insufficient to meet the requirements of the evolving communication scenario for several reasons.

**Macro-Mobility.** Open issues in Mobile IPv4 related to macro-mobility management include the following:

- **Asymmetric (Triangle) routing.**
- **Inefficient direct routing:** The routing procedure in Mobile IPv4, with respect to the number of hops or end-to-end delay, is inefficient.
- **Inefficient home agent notification:** When a mobile node hands off from one network to another, it has to notify the home agent about that. This operation is inefficient in Mobile IPv4.
- **Inefficient binding deregistration:** In Mobile IPv4, if a mobile node moves to a new FA, then the previous FA could not release the resources occupied by the mobile node. The previous FA must wait until a binding registration lifetime expires.

**Micro-Mobility.** Mobile IPv4 is not concerned with micro-mobility issues. Since it is expected that in the near future, Mobile IP will be a core wireless IP architecture interconnecting different wireless networks, the interoperability...
between macro-mobility and micro-mobility issues is very important. There are several open issues in this context:

- **Local management of micro-mobility events:** In micro-mobility dimension, handoffs occur very frequently. Therefore the handoff procedures should be managed as locally as possible.

- **Seamless intradomain handoff:** After intradomain handoff, the IP data stored into the previous entities (e.g., base stations) should be transferred to the new BS.

- **Mobility router crossings in an intranet:** During intradomain handoff, the router crossings should be avoided as much as possible.

### 4.5.3.1 Mobile IPv6

The IETF began work in 1994 on IPv6, a proposed standard designed to address various problems with Mobile IPv4. The IETF skipped the IPv5 standard because network operators had already adopted many of the protocol’s potential provisions before it could be adopted.

Mobile IPv6 is supposed to replace Mobile IPv4. One of the main advantages of IPv6 over IPv4 is the higher number of IP addresses. IPv4’s 32-bit addressing scheme can support a theoretical maximum of 4.29 billion IP addresses. However, due to the operational inefficiency, useful IP addresses are about 200 million. IPv6 offers a 128-bit addressing scheme that permits about $2^{1033}$ useful IP addresses.

Other major differences between Mobile IPv4 and the Mobile IPv6 [146] are:

- Mobile IPv6 supports the mechanism of route optimization. This feature is already an integral part of the Mobile IPv6 protocol. In Mobile IPv4 the route optimization feature is just a set of extensions that may not be supported by all IP nodes.

- Mobile IPv6 specifies a new feature that allows mobile nodes and Mobile IP to coexist efficiently with routers that perform ingress filtering [147]. The packets sent by a mobile node can pass normally through ingress filtering firewalls. This is possible because the COA is used as the source address in each packet’s IP header. Also, the mobile node home address is carried in the packet in a **home address destination option** [146]. In this way, the use of the COA in the packet is transparent above the IP layer.

- The use of the COA as the source address in each packet’s IP header simplifies the routing of multicast packets sent by a mobile node. Thus,
the tunneling of the multicast packets to its home agent will no longer be necessary in Mobile IPv6. The home address can still be used and it is compatible with multicast routing that is based in part on the packet source address.

- Neighbor discovery [148] and address autoconfiguration [149] enable the functionality of the foreign agents instead of using special routers. The foreign agents are not required any more in Mobile IPv6. Thus, the issue of mobility router crossings in an intranet is resolved.

- The Mobile IPv6 uses IPsec for all security requirements (e.g., sender authentication, data integrity protection, and replay protection for binding updates—which serve the role of both registration and route optimization in Mobile IPv4). Mobile IPv4 is based on its own security mechanisms for each function, based on statically configured mobility security associations.

- Mobile IPv6 provides a mechanism for supporting bidirectional (i.e., packets that the router sends reach the mobile node, and packets that the mobile node sends reach the router) confirmation of a mobile node ability to communicate with its default router in its current location. This bidirectional confirmation can be used to detect the black hole situation, where the link to the router does not work equally well in both directions. Unlike Mobile IPv6, Mobile IPv4 does not support bidirectional confirmation. Only the forward direction (packets from the router are reaching the mobile node) is confirmed, and therefore the black hole situation may not be detected.

- In Mobile IPv6, the correspondent node sends packets to a mobile node while it is away from its home network using an IPv6 routing header rather than IP encapsulation, whereas Mobile IPv4 must use encapsulation for all packets. In this way ReSerVation Protocol (RSVP) operation in Mobile IP is enabled and also the problem of ingress filtering is partially solved. In Mobile IPv6, however, the home agents are allowed to use encapsulation and tunnel the packets to the mobile node.

- In Mobile IPv6, the home agent intercepts the packets, which arrive at the home network and are destined for a mobile node that is away from home, using IPv6 neighbor discovery [148] rather than Address Resolution Protocol (ARP) [100] like in Mobile IPv4.

- IPv6 encapsulation (and the routing header) removes the need to manage tunnel soft state, which was required in Mobile IPv4 due to limitations in ICMP error procedure. In Mobile IPv4, an ICMP error message that is created because of a failure to deliver an IP packet to the COA will be returned to the home network, but it will not contain the
IP address of the original source of the tunneled IP packet. This is solved in the home agent by storing the tunneling information (i.e., which IP packets have been tunneled to which COA), called *tunneling soft state*.

- Mobile IPv6 defines a new procedure, called _anycast_. Using this feature, the dynamic home agent discovery mechanism returns one single reply to the mobile node, rather than the corresponding Mobile IPv4 mechanism that used IPv4 directed broadcast and returned a separate reply from each home agent on the home network. The Mobile IPv6 mechanism is more efficient and more reliable. In this way, only one packet requires to be replied to the mobile node.

- In Mobile IPv6, an advertisement interval option on router advertisements (equivalent to agent advertisements in Mobile IPv4) is defined. This allows a mobile node to decide for itself how many router advertisements (agent advertisements) it is ready to miss before declaring its current router unreachable.

- The IPv6 destination options permit all Mobile IPv6 control traffic to be piggybacked on any existing IPv6 packets. Mobile IPv4 and its route optimization extensions require separate UDP packets for each control message.

### 4.5.3.2 Macro/Micro-Mobility Extensions to Mobile IP

Several protocols and frameworks have been proposed to extend Mobile IP to better support micro- and macro-mobility in next generation wireless cellular environments.

**Fast and Scalable Handoffs for Wireless Internetworks**

This is an extension to Mobile IP [150] that uses hierarchical FAs to handle macro-mobility. This architecture assumes BSs to be network routers; for that reason, it is not compatible with current cellular architectures, where BSs are simply layer 2 forwarding devices. Moreover, deploying a hierarchy of FAs imposes complex operational and security issues (especially in a commercial multiprovider environment) and requires multiple layers of packet processing over the data transport path. The presence of multiple layers of mobility-supporting agents also significantly increases the possibility of communication failure, since it does not exploit the inherent robustness of Internet routing protocols.

**Mobile-IP Local Registration with Hierarchical Foreign Agents**

This protocol is an IETF proposal [151], and it solves the issues of triangle routing and inefficient direct routing. The basic mechanism of this architecture
deploys hierarchical FAs for seamless mobility within a domain. During the COA discovery procedure multiple FAs are advertised using the agent advertisement message. The COA registration will be provided for the FA that is the lowest common FA ancestor at the two points of attachment of interest. The requirement for hierarchical agents in Internet mobility architecture remains an open issue. Even though it does not appear to be a critical consideration in the immediate future, it is possible that hierarchical mobility management will become more attractive as the IP security infrastructure matures and deployment of mobile multimedia terminals gets much larger.

**TeleMIP**

TeleMIP stands for Telecommunication Enhanced Mobile IP [152]. It achieves smaller handoff latency by localizing the scope of most location update messages within an administrative domain or a geographical region. However, this architecture faces the problem of inefficient home agent notification. TeleMIP introduces a new logical entity, called the mobility agent (MA), which provides a mobile node with a stable point of attachment in a foreign network. While the MA is functionally similar to conventional FAs, it is located at a higher level in the network hierarchy than the subnet-specific FAs. Location updates for intradomain mobility are localized only up to the MA. Global location updates are necessary only when the mobile changes the administrative domains. The TeleMIP allows efficient use of public address space, by permitting the use of private addresses for handling macro-mobility. Reduction of the frequency of global update messages overcomes several drawbacks of existing protocols—such as large latencies in location updates, higher probability of loss of binding update messages, and loss of in-flight packets—and thus provides better mobility support for real-time services and applications. The dynamic creation of mobility agents (in TeleMIP) permits the use of load balancing schemes for the efficient management of network resources. Its drawback is the potential nonoptimal routing within the domain.

**Wireless IP Network Architecture by TR45.6**

Another framework for IP-based mobility management was developed by the Telecommunications Industry Association (TIA) Standards Subcommittee TR45.6 [153] to target 3G cellular wireless systems. This architecture is consistent with the requirements set by the ITU for IMT-2000. Therein, solutions are provided to the issue of inefficient direct routing. The framework uses Mobile IP with fast rerouting for global mobility. For macro-mobility, the scheme proposes the use of dynamic HAs (DHAs), which reside in the serving network and are dynamically assigned by the visited authentication, authorization, and accounting server. The DHA allows the roaming user to gain service with a local access service provider while avoiding unnecessarily long routing. The
architecture defines a new node called a **packet data-serving node** (PDSN) (which contains the FA), and uses visitor location register(VLR)/home location register (HLR) (ANSI-41 or GSM-MAP) authentication and authorization information for the access network. The mobile node is identified by a network access identifier (NAI) [154] in the visiting or foreign network. Within the registration process, an MN sends the registration message to the FA, which in turn interacts with an authentication, authorization, and accounting server residing in that network or uses the broker network for authentication with the home network.

**Micro-Mobility Extensions to Mobile IP**

Due to the fact that the basic Mobile IP protocol [91] is only concerned with the macro-mobility management, some other solutions are required to enhance the Mobile IP functionality to support micro-mobility. The following overview shows the results of several current research activities in this area.

**Wireless Network Extension Using Mobile IP.** This micro-mobility management framework [155] is combined with Mobile IP. It provides solutions to the issue of local management of micro-mobility events. The development of this scheme is realized in the Motorola iDEN architecture. Micro-mobility events should be managed more efficiently than macro-mobility events, because they can happen with relatively high frequency. Therefore, the procedures and participants are being kept as local as possible. The micro-mobility procedures are managed by a data gateway, thus achieving the previous condition. The macro-mobility between iDEN subnetworks and other subnetworks is accomplished by implementing Mobile IPv4 in the FA and HA (Figure 4.14).

**HAWAII.** The Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [156] proposes a method for using a separate binding protocol to handle micro and macro-mobility. For global mobility, it uses Mobile IP. Using this architecture, a solution to the issue of local management of micro-mobility events is provided. It uses a two-layer hierarchy for mobility management. When the MN moves into a foreign domain, it is assigned a collocated COA from that domain, and the MN retains its COA unchanged while moving within the foreign domain. Thus, the movement of the MN within a domain is transparent to the HA. This protocol uses path setup messages to establish and update host-based routing entries for MNs in some specific routers within the domain; other routers not in the path are kept in the dark about the MN’s new COA. When a CN sends packets to a roaming user, it uses the MN home IP address. The HA intercepts the packets and sends the encapsulated packet to the MN’s current border router. The border or root router decapsulates and again encapsulates the packet to forward it to either the intermediate router or BS, which decapsulates the packet and finally delivers it to the MN (Figure 4.15).
**Cellular IP.** Cellular IP [157, 158] proposes an alternative method for providing mobility and handoff support in a cellular network, which consists of interconnected cellular IP nodes. Solutions to issues like local management of micro-mobility events and seamless intradomain handover are provided. This

---

**Figure 4.14** Mobile IP wireless network extension.

**Figure 4.15** HAWAII architecture.
protocol uses Mobile IP for global mobility. It is very similar to the host-based routing paradigm of HAWAII. Specifically, Cellular IP provides local mobility (e.g., between BSs in a cellular network) (Figure 4.16).

The architecture uses the home IP address as a unique node identifier, since MN addresses have no location significance inside a cellular IP network. When an MN enters a Cellular IP network, it communicates the local gateway (GW) address to its HA as the COA. Nodes outside the Cellular IP network do not require any enhancements to communicate with nodes inside the network. When a CN sends packets to a roaming user, it uses the MN’s home IP address. As in conventional Mobile IP, the HA intercepts the packets and sends the encapsulated packet to the MN’s current GW. The GW decapsulates the packet and forwards it to the MN’s home address using a node-specific route. Thus, the nodes sending or receiving packets to and from the MN remain unaware of the node location inside the Cellular IP network.

HAWAII and Cellular IP are based on very similar concepts. There is a major difference between them, however. In the HAWAII protocol, most of the intelligence is in the network part, while in Cellular IP most of the intelligence is in the mobile node. Therefore, Cellular IP is not optimal for management of the security and quality of service. However, the network equipment is simpler and therefore cheaper.

4.5.3.3 RSVP Support for Mobile IPv6

The solution for the RSVP operation over IP tunnels is provided for Mobile IPv6 as well [159]. The specification in the draft, proposes three solutions:

Figure 4.16  Cellular IP.
1. The first solution requires modifications in RSVP at both mobile and correspondent nodes, so that they have to be aware of Mobile IPv6 addressing.

2a. In order to enhance the performance and make handoffs smooth and seamless, optional *triggers/objects* are added to RSVP messages. The RSVP PATH messages are triggered on bindings updates and home address objects that are contained in RSVP RESV messages. Thus, intermediate routers are enabled to recognize connections and to use resources even when the COA changes.

2b. A *flow extension* mechanism is provided. This mechanism is able to extend the existing RSVP flows (i.e., flow_ids) that are applied on typical IP routers, to the new Mobile IP router. It is combined with a simultaneous binding option that has to be applied for the roaming mobile node. The mobile node receives packets on both previous and current COA.

To make Mobile IPv6 and RSVP interoperable, the minimal solution (1) is a requirement [159]. This requires the modification and the interfacing of the RSVP daemon and Mobile IP binding cache at both CN and MN.

The latter two solutions (2a or 2b) can provide uninterrupted operation since they support fast reestablishment or preservation of resource reservations when mobile nodes move. Table 4.3 presents a qualitative comparison of the latest two approaches.

It should be noted that triggers/objects is a quick solution with low complexity, which is able to provide sufficiently good service. The flow extension approach is a little more complex but has the advantage of faster deployment. In multiprovider environments, where the whole end-to-end path could not be controlled, a solution that modifies only CNs, access network routers, and MNs has a big advantage.

### 4.6 Relays for Next Generation 4G Systems

#### 4.6.1 Introduction

Characteristics of future mobile/wireless radio networks will be quite different to those of current networks. Compared to a *uniform* spatial distribution of, say, voice traffic in a GSM cell, a much more nonuniform traffic will be witnessed in future networks in terms of spatial distribution within the cell of voice and data, in particular with even higher demands for a much broader bandwidth stretching from low to high bit rates for higher mobility. At the same time, the more we move to higher frequencies, the more the range of the cell shrinks compared
to current systems (e.g., GSM cells). This means that in order to cover the same geographical area, more BSs need to be deployed, which leads to much higher deployment costs for operators. Apart from those new needs, future network deployments will need to solve more adequately the problems that exist in current conventional network deployments and that are related to the issue of coverage; either extend/stretch the coverage of cells or provide better coverage in shadowed areas within a cell in terms of, say, better QoS. The above needs/requirements will need to be addressed either by new technologies (for instance, new air interface with conventional cellular architectures) or by novel concepts that will introduce new network elements and architectures.

One of those concepts that can address capacity and coverage issues is that of relaying. Relaying has attracted a lot of interest in recent years mainly through research projects of companies and universities, particularly the Wireless World Initiative New Radio (WINNER) project part of the EU IST FP6 [161] (see also Section 5.2.5.2). Relaying techniques can be classified according to several

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Triggers/Objects</th>
<th>Flow Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changes to CN</td>
<td>Yes (required for minimal solution)</td>
<td>Yes (required for minimal solution)</td>
</tr>
<tr>
<td>Changes to Intermediate</td>
<td>Yes (RSVP MIP object extension and reuse of flow’s resources)</td>
<td>No</td>
</tr>
<tr>
<td>Routers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changes to MIP-Router</td>
<td>No (forwarding of late packets is also an option here)</td>
<td>Yes (binding update interception, flow forwarding)</td>
</tr>
<tr>
<td>Changes to MN</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Changes to HA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Supports Multicast Delivery</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bandwidth efficient</td>
<td>Yes</td>
<td>Yes (it is assumed efficient overprovisioning in the access network)</td>
</tr>
<tr>
<td>End-to-End Delay</td>
<td>Always shortest path (but reestablishment of resources requires a round-trip)</td>
<td>Slightly increased delay</td>
</tr>
<tr>
<td>Lossless HO</td>
<td>Yes (with forwarding of late packets)</td>
<td>Yes</td>
</tr>
<tr>
<td>HO Delay</td>
<td>Roundtrip</td>
<td>Faster</td>
</tr>
<tr>
<td>Implementation Complexity</td>
<td>Moderate</td>
<td>Higher</td>
</tr>
</tbody>
</table>

Table 4.3
RSVP Support for Mobile IPv6: Qualitative Comparison of Approaches 2a and 2b
standpoints. Analog relaying refers to the simple case where the signal to be repeated is simply amplified and forwarded, possibly with a different frequency (frequency translation). The counterpart is digital relaying, where the signal is fully regenerated before being retransmitted (e.g., decode-and-forward). From a different perspective, fixed or infrastructure relays refer to repeating stations specifically put on fixed locations for that purpose. Relay stations can also be infrastructureless—that is, movable or mobile. The former refers to stations that can be moved to a specific location to help with some temporary requirements of coverage, or mounted on moving vehicles, while the latter refers to the use of other terminals as relaying stations. The relaying operation can be done in two hops, involving a single repeater, or in a truly multihop fashion, with several relaying stations being used. Fixed relaying techniques have been extensively studied (see [162, 163] and the references therein). Mobile relaying is currently being studied in the WINNER project and most of this section will deal with this subject. Mobile relays are effectively fixed-relay logical elements incorporating the mobility factor (see Figure 4.17). Although more complex in comparison to the fixed relaying approach, mobile relaying can offer incremental gains

Figure 4.17 Mobile relay concept (MRi and UTi are the MR/UT positions for i=1,2,3,4 time instances).
in future networks, exactly due to that mobility. The main advantages are the plurality of types of mobile relays and the multiple locations where they can be found, which allow the network the opportunity to act on a more ad hoc basis to deal and address more nondeterministic needs.

4.6.2 Mobile Relay Types/Deployment Concepts

The first classification of mobile relays (MRs) (which can be done with reference to the ownership of the MR) is as follows:

- Dedicated mobile relays, with elements being built only for relaying purposes (Type I/II in Figure 4.18);
- User terminals (UT), which can act additionally as mobile relays (Type III in Figure 4.18).

Dedicated MRs are expected to be fitted on top of moving carriers (e.g., cars, ships, trains, or any other vehicle). As such, and with reference to the mobility of those MRs in relation to the mobility of the target area/UT population to cover, dedicated MRs can be split into two further categories:

- Dedicated MRs (Type I) to cover the UT population of the carrier on which they are fitted, either on the top or the inside (MR-UT mobility

![Figure 4.18 Types of mobile relays.](image-url)
correlated; normally the same) (e.g., MR on a train to provide coverage inside the train);

- Dedicated MRs (Type II) to cover UT population outside the carrier they are fitted on (MR-UT mobility uncorrelated) (e.g., MR fitted on bus to provide coverage in a park).

Thus, in Figure 4.18 we present the main classification of MRs, as described in [164].

### 4.6.3 Rationale for Mobile Relays

Mobile relays in effect can address the same problems that fixed relays address: coverage and capacity. However, they are envisaged to cover cases that cannot or may not be adequately covered by fixed relays. In that sense fixed and mobile relay can be thought as complementing each other. Thus, the rationale for MR deployment can be summarized in the following points:

- To provide a cheaper solution (in terms of CAPEX/OPEX/backhaul costs) compared either to conventional topologies or fixed relay-based topologies.

- To provide coverage in unexpected events where a high concentration of UTs occurs (e.g., road accident/traffic jam). Those relays could be fitted on police cars or ambulances.

- To provide coverage for events that occur infrequently, but are known of in advance (e.g., a sporting event or concert). It might not be economically wise to deploy BSs next to a football stadium. In these cases MRs could be reused on demand.

- To provide coverage where fixed relays cannot go (e.g., provide coverage to terminals on top of ships). Thus, we rely on fixed relays, which, on such a carrier, become mobile relays.

- To reduce BS power transmission levels (by evenly distributing this power among mobile relays). Thus, we gain in increased capacity in, for example, a CDMA-based interference-limited system.

- To reduce the impact on the environment/humans. A plethora of GSM BSs and UMTS Node Bs has triggered a lot of controversy and complaints from communities and environmental organizations. Thus, through mobile relays we make deployment easier and cheaper (less cost is required to incorporate them on carriers rather than deploy BSs on top of buildings), and we provide less risk for humans in terms of power...
of RF emission, better aesthetical intervention, and less impact to the environment.

4.6.4 Applicability to Environments

Each of the aforementioned types of mobile relays is expected to have different characteristics and capabilities, especially in terms of mobility, complexity, and coverage. We expect them to be applicable or best suited to a number of different scenarios and environments [164].

- **Type I**: The main characteristics for Type I are high mobility, small coverage, and high complexity. Thus, they are envisaged to be typically applicable to scenarios of wide or rural areas.
- **Type II**: The main characteristics are medium/low mobility, medium/large coverage, and medium complexity. Thus, the applicability is towards hotspots and mostly wide area deployments.
- **Type III**: Typical characteristics are low/zero mobility, small coverage, and low/medium complexity. Thus, they are appropriate for indoors, hotspots, and possibly wide area deployments.

4.6.5 Parallels with Other Technologies

Technologies and concepts similar to the mobile relay concepts have been proposed in several fora, for existing and short/medium-term to-come systems. Extrapolating from these we can find commonalities in order to highlight the applicability of MR-based concepts on realistic scenarios.

- **Positioning**: A mobile relay in several locations can mimic a BS, which means that certain limitations of, say, using only round-trip time (RTT)/Cell Id positioning techniques in isolated sites are overcome and more techniques can be applicable [e.g., RTTs and observed time difference of arrival (OTDOA)]. At the same time, due to the favorable characteristics of the MRs (e.g., shorter range and better channel conditions), positioning accuracy is expected to be better compared to the same positioning techniques being applied on larger cells [165].
- **In 3g Partnership Proejct (3GPP)** the concept of repeaters has been proposed [166]. It will be interesting to see what type of simple mobile repeaters could be deployed and under what requirements.
- **Vehicular networks** are proposed as part of future networks for the automotive industry. Although more of an ad hoc/mesh-type networks compared to the MR-based hierarchical networks, interesting areas with regard
to interworking of those networks could be investigated. Effectively, MR-based networks could be seen as some kind of higher level layer (in the architecture concept) compared to the vehicular networks and effectively promote interworking between a mesh and a cellular network (e.g., mesh-network with gateway devices communicating with the MRs).

- Moving networks. For train companies/network operators the need for coverage inside trains is important. Thus, Type I MRs are very applicable for these types of deployments, and projects addressing this issue (the Moving Networks concept) are already underway [167].

- The concept of opportunity driven multiple access (ODMA), providing coverage through UTs to out-of-coverage UTs, was proposed in 3GPP [168]. It seems that now that networks can support more robust functionalities, faster signaling, ODMA could be extended to the Type III concept.

- Multimedia Broadcast/Multicast Services. MR-based networks could be simplified if MRs aimed to cover certain types of services. For instance, in the case of MBMS [169], the MRs do not need to maintain dedicated links with the terminals. As such, the whole deployment is much easier to implement. Furthermore, by allocating MBMS for the edge of cell on MRs, we free capacity in the BS.

In order to be able to propose specific solutions, a number of issues need to be resolved like the multiple access (MA) technique (e.g., OFDMA based), the duplex scheme (e.g., TDD), the level of complexity of a mobile relay (layer 1 or layer 2 mobile relays), the number of hops (preferably 2), routing and forwarding issues, security, signaling issues, types of protocols (e.g., nonstatic protocols, adaptive or feedback based), power control, and connectivity issues. Some of these have already been addressed, but further studies are required [167, 170]. However, in future networks it is envisaged that more complex functionalities and processes could be supported (e.g., faster signaling, higher processing power, shorter delays), which means that mobile relay-based concepts would be more easily supported compared to the point of view we have by looking at their applicability on current networks.

### 4.6.6 Cooperative Mobile Relaying

Cooperative relaying (CR) is another concept within the relaying field. The main idea is to use multiple relays to enhance the received signal in the UT. CR schemes could be seen under the mobile relaying concept, and thus, they give rise to the cooperative mobile relaying (CMR) concept [167]. Again, due to the different types of mobile relays, different levels of applicability exist for CMR...
schemes for a number of scenarios. Specifically, and with reference to the criteria of coverage, velocity, and types of applications, we prioritize the three types of MR with reference to the CMR concept.

- For short-type/fast applications/services, supporting two-hop strategy, Type III MRs are promising in providing CMR. The main advantage is the plurality of terminals in multiple locations out of which we can select the optimum to provide cooperative schemes. The main restrictions seem to be the frequent positioning, the low computational power, the possible reduction of the user experience (e.g., battery life), and the small coverage, which is, however, compensated by the low/zero mobility.

- For Type II, the problem is the relatively high mobility (average 40 km/h) but this can be compensated by the large coverage of the MR (up to 500m), which means that connectivity will be in general similar to that of Type III. So, these types of MRs could be applicable for more bandwidth-hungry applications/services due to the higher complexity incorporated into them.

- CMR may not be that applicable for Type I, due to the very good links we assume between MR-BS and MR-UTs, which is the initial reason for applying CMR (i.e., poor links). So, even though CMR could be supported it is not envisaged to be of high incremental gain to the performance. However, under a possible MIMO approach (multiply the bit rates by x times) it could be applicable.

4.6.7 Conclusions

Mobile relays have been proposed as part of the relay-based topologies for future network deployments for 4G systems in order to provide better coverage and increased capacity in cellular/wireless environments. A number of potential gains can be extracted by the use of mobile relays, as a complementary solution (e.g., on an ad hoc basis) to that of fixed relays. As we have seen, mobile relays pose a number of challenges, mainly those related to the issue of mobility and extending to routing, forwarding, positioning, connectivity, security, and complexity. At the same time, though, as we have analyzed, mobile relays could potentially better cover certain needs compared to conventional or fixed relay-based topologies by addressing and incorporating other concepts (e.g., reconfigurability, to be able to cope in several locations for multiple RATs).
4.7 Other Enabling Technologies

Other technologies have been identified as fundamental for allowing the development of future 4G. We now list some other important techniques and technologies not considered in this chapter: adaptive modulation and coding, techniques for seamless vertical and horizontal handovers, cross-layer design and optimization, multidimensional scheduling (time, frequency, space), techniques for reducing the PAPR problem typical of multicarrier systems, sensor networks, battery technology, and access techniques specifically tailored for short-range communications [171]. The latter techniques have received considerable attention lately. These are basically ultra-wideband (UWB) and optical wireless communication techniques. Interested readers are referred to [172, 173] and the references therein for a comprehensive introduction to UWB. The role of optical wireless systems within future 4G communications is discussed in detail in [174]. Optical wireless communications is in fact a very attractive complementary technology that has been extensively considered recently due to its unmatched advantages [175].

References


[58] 3GPP TS 25.211, “Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD),” v.5.1.0, June 2002.


[61] ETSI TS 101 475, “Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Physical (PHY) layer,” v1.1.1, April 2000.


[64] ETSI TS 101 761-2, “Broadband Radio Access Network (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 2: Radio Link Control (RLC) Sublayer,” v1.1.1, April 2000.


[169] 3GPP TS 22.146, “Multimedia Broadcast/Multicast Services.”


